Construction of height and weight growth charts for Iran, with an investigation of appropriate statistical methods

by

Mostafa Hosseini

Thesis submitted for the degree of Doctor of Philosophy
Faculty of Science
(Field of study: Statistics)

Medical Statistics Unit
Department of Epidemiology and Population Sciences
London School of Hygiene & Tropical Medicine
University of London
February 1997

Abstract

Data on weights and heights of children 2-18 years old in Iran were obtained in a National Health Survey of families in 1990-2, with a sampling ratio of 1 in 1000. In total 10660 households were surveyed in random cluster samples of households in all 24 provinces of Iran. The data are hierarchically structured with children in families, families in clusters and clusters in provinces. **The main aim of the thesis is to use the survey data on height and weight to model growth pattern of children and adolescents in Iran and the construction of growth charts for height and weight.**

After removing the outliers in the data by multivariate analysis, regional variation in growth patterns were studied by constructing multilevel models. The results of these analyses showed that the data from Urban Tehran can be used as a reasonable baseline for the country, and further investigations of distributions of different centiles confirmed this.

Three recently developed techniques of chart construction were compared with a chart constructed from a multilevel model to see which method produces centiles which fit the data best. the data structure has little effect on estimates of population centiles. The HRY method using spline procedure is shown to produce the best fit, and this technique has been used to construct the growth charts of weight and height for Iranian boys and girls. Checks confirm that all these curves fit the data well. However, growth of rural children differs significantly from that of children in urban Iran; a practical solution enabling one set of charts to be used for both groups of children is proposed. In view of the difference between the Tehran charts and the NCHS reference centiles, it is concluded that charts presented here should be adopted as the new reference curves for children in Iran.

To my parents,

my brothers, my sister, and to Nasrin and Ali.

To Robert and his family; to Kazem and his family, and to the memory of Firozeh.

And to my supportive friends especially Mohammad Bagher, Bahman, Shams, and Hassan.

Acknowledgements

In the name of Almighty Allah, the most beneficent and the merciful, who bestowed me the strength and courage to complete this thesis.

First and foremost I would like to extend my sincere thanks to my supervisor Dr R G Carpenter. His guidance, and encouragement played a vital role in the completion of this research. I am deeply indebted to him for all his kindness and help.

I would like to express my thanks to Professor K. Mohammad for providing me with this invaluable data set, and also to the staff, families and children from all over the country who participated in the National Health Survey 1990-2. This valuable study is one of the most important scientific projects that has been undertaken in Iran.

I am very grateful to the members of staff of the Medical Statistics Unit and also to those in the EPS department who taught me, for their help in many ways and their friendship during the course of this study. I thank especially Tom Marshall who was my course tutor in 1992-3. His advice was invaluable and his friendship has been of great support all the way through this work. I also thank Professor Pocock and Professor Thompson for help and support.

I would also like to extend my very special thanks to my parents, to my brothers and to my sister, to whom I am forever indebted for their support during the entire phase of this study. I am also grateful to my kind and supportive friends in England who treated me like a member of their family, especially Craig, Heather, Sarala, Ed, Jane, Vicky, Maria, Najwa, Emilia and Giota. Also, I am grateful to all my fellow PhD students especially Abbas and all the kind people in the Department and School whose friendship made this hard work easier for me. I also thank my kind office-mates Sue and Ana for all their support and encouragement. I would also like to thank Dr Michael Jones for reading the draft of this thesis, for his scientific advice, and for his friendship and support.

Finally, I wish to extend my acknowledgements to the Ministry of Health and Medical Sciences of Iran for granting my scholarship for this research.

CONTENTS

List of Tables

Chapter 1

Chapter 3

Chapter 5

Chapter 8

Appendix C

Appendix D

List of Figures

Chapter 7

Chapter 8

Appendix C

Appendix D

Abbreviations

AED	Academy for Educational Development
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
CDC	Centre for Disease Control
CHW	Community Health Worker
cm	Centimetre
COME	Community Oriented Medical Education
ESCDP	Economic, Social and Cultural Development Plans
FAO	The United Nation Food and Agriculture Organization
fp	Fractional Polynomial(s)
GMP	Growth Monitoring and Promotion
GMT	Greenwich Mean Time
GSO	Gita Shenasi Organization, Iran
HRY	Healy, Rasbash, Yang method of growth chart construction
ICMR	Indian Council of Medical Research
IMR	Infant Mortality Rate
IRI	Islamic Republic of Iran
IGLS	Iterative Generalized Least Squares
kg	Kilogram
LMS	Lambda, Mu, Sigma (Cole's method of growth chart construction)
MD	Mahalanobis Distance
MLE	Maximum Likelihood Estimation
MOH&ME	Ministry Of Health & Medical Education
MVE	Minimum Volume Ellipsoid
NCHS	National Health for Health Statistics
NED	Normal Equivalent Deviate
PHC	Primary Health Care
PGY	Pan, Goldstein, Yang method of chart construction

RIGLS	Restricted Iterative Generalized Least Squares
SCI	Statistical Centre of Iran
SD	Standard Deviation
SE	Standard Error
SDS	Standard Deviation Score
SEDP	Socio Economic Development Plan
U5MR	Under Five years old Mortality Rate
UNICEF	United Nations Children's Found
UNU	United Nations University
UPGMA	Unweighted paired-group method using Arithmetic Average
WHO	World Health Organization

CHAPTER ONE

INTRODUCTION AND BACKGROUND

1.1 Introduction

The health of a population is most accurately reflected in growth of its children. Naturally, child growth, in all its aspect, is the priority concern to all (Petros-Barvazian, 1990). A child who is growing well is likely to have healthy immunological defenses against infection. Hence, better growth means decreased risk of severe infections, case fatality rates, and child mortality. Promotion of healthy growth in childhood will result in youths with a greater potential for being productive members of society. Good physical growth results in increased human capital.

As the infant mortality rates decrease during a country's development, governments and agencies which deal with child welfare are becoming more concerned with quality of life. Quality of life can be broadly defined as physical and mental well-being. Health-related indicators of physical well-being include adequate nutritional status, as measured by anthropometric, biochemical, and clinical indicators. So, the importance of monitoring growth increases.

Growth monitoring is a useful measure which can significantly contribute to the promotion of child health and nutrition. Growth monitoring is potentially extremely useful in Primary Health Care (PHC) programs, particularly for education, motivation, and promotion of other health services. Since the mid-1980s there has been remarkable improvement in design and implementation of growth monitoring, which has led to a greater emphasise on community-based programs. Active involvement of mothers, families, and communities at all stages of design, implementation, and evaluation of the program, as well as intensive training and regular supervision of the health workers, appears to be key to success.

It is widely accepted that for practical purposes anthropometry is the most useful tool for assessing the nutritional status of children (WHO, 1986). In this context, the individual child growth chart is broadly approved as an important and sensible instrument that can contribute significantly to achieve better child health and growth. The growth chart offers a very simple and inexpensive means of monitoring child health and nutritional status and can be used by community workers with very little instruction and supervision. The chart represents a convenient means of presenting basic health data and permits the assessments of current status as well as the observation of trends in growth. Hence, I shall discuss growth monitoring and promotion and the corresponding technology (growth charts) with particular reference to Iran in more detail in the following chapter.

Extensive studies of the growth of children have been undertaken in developed countries (notably the USA and UK), and have resulted in the production of growth charts and their use as reference standards has been encouraged widely. On the other hand, there has been a long-running discussion amongst auxologists, nutritionists and public health workers about the desirability or otherwise of a single universal reference or standard for growth. There are certainly large differences between populations, and there is no guarantee that all the populations have the same growth potential (Eveleth and Tanner, 1990). Consider for example, differences in height and weight and the age of puberty; it is now clear that a portion of these differences is genetic in origin, and a portion (in developing countries a large portion) environmental. It simply will not do to use an American or a British standard to judge the growth of Japanese or Hong Kong infants or children (Davies and Yamamuro, 1985; Leung and Davies, 1989; Baldwin and Sutherland, 1988; and Goldstein and Tanner, 1980). Therefore, the use of American or European norms in clinical work in developing countries may be seriously misleading.

In some developing countries, to determine local growth pattern and identify undernourished communities, nation-wide studies on growth and development have been undertaken. For instance, in India (ICMR, 1972), Bahrain (Mater et al., 1990), Pakistan (Akram and Agboatwala, 1991), China (Lin et al., 1992), and Kuwait (Al-Isa and Bener, 1995). In Iran, at present there is no reliable and up to date information regarding the growth pattern of children and adolescents at the national level. Previous sparse studies were taken from small selected clinics, groups, or small districts. These studies do not provide a basis for constructing population norms (Ayatollahi, 1991).

The existing deficiencies in different studies, especially non reliability in sampling and out of datedness, and the need for reliable *'norms'* reflecting the reality of the growth patterns of healthy children and adolescents were the main motivations for doing this research study. The National Health Survey 1990-2 provides a wealth of data on development of the population of interest. Therefore, the objectives of this study are the construction of the norms for assessing the growth pattern of Iranian children and adolescents. We have also investigated different approaches to growth chart construction and have compared them.

1.2 Background

1.2.1 General information about Iran

1.2.1.1 Geography and Climate

Iran[1] is in southwest Asia and is one of the Middle Eastern countries located at 25 degree 3 minutes to 39 degree 47 minutes north latitude and 44 degree 5 minutes to 63 degree 18 minutes east longitude. Iran (Figure 1.1) is bounded on the north by the Republics of the former USSR and the Caspian sea; on the south by the Persian Gulf and the Sea of Oman; on the east by Afghanistan and Pakistan; and on the west by Turkey and Iraq.

The total area of the country is 1,648,000 square kilometres, and is the 16th largest country in the world. About one-half of Iran's land consist of mountains and one quarter desert, leaving only 25 percent as arable land. Two-thirds of the land is situated on a high plateau with an average altitude of 1,150 meters above sea level (MOH&ME, 1995).

The Alborz mountains cross the northern part of the country, separating the Caspian sea from the plateau and includes Mt. Damavand, the highest peak in the country

1: Official Name : Islamic Republic of Iran
International Name : Iran
Former International Name: Persia, Perse, Pars (GSO, 1993)

(5671 m). The Zagros mountain range stretches from the northwest of the Caspian sea to the Persian Gulf. The great mountain ranges of Alborz and Zagros are believed to be central in shaping Iranian's varied climate. This diversity of environment has resulted in dramatic variations in the socio-economic status of the people, along with differences in health, ecological and epidemiological conditions; from the prosperous, fertile green fields and forests of the humid north, to the arid uncompromising deserts of the central belt.

Iran has a variable climate ranging from sub-tropical to sub-polar. In winters, a high pressure belt slashes the north-east and north-west and the central parts of the plateau while low pressures develop over the warm waters of the Caspian Sea and the Persian Gulf where the weather is mild and of a mediterranean type. In summers, one of the lowest pressure centres in the world prevails in the southern parts of the country. Rainfall varies widely, from less than an average of 50 mm per year in south-east to about 1,900 mm in the Caspian Region. The annual average of rainfall for whole country is 355 mm (UNICEF, 1992).

1.2.1.2 Population

According to the results obtained from the 1991 census, Iran's population was 55,837,163 with a sex ratio of 106.

Fifty seven percent of the population live in urban areas, 42% live in rural communities and 1% are nomads (SCI[2], 1991). The population of Iran is extremely young with over 44% below 15 years of age and only 3.5% over 65. The population growth rate was among the highest in the world, estimated 2.3% per year in 1991 and decreased to 1.75% in 1995 with a rate of 1.5% for urban and 2.0% for rural areas respectively (MOH&ME[3], 1995).

Table 1.1 illustrates the population growth rate for urban and rural areas for 1986 compared with the 1991 census. The urban population growth rate is approximately three times that of rural areas which is partly because of rural urban migration. Based on the analysis of the 1991 census, the average household size has increased from 5.11 persons in 1986 to 5.31 in 1991. Figure 1.2 shows the pyramid of age groups in Iran based on the results of the 1991 census.

1.2.1.3 General Characteristics

The administrative structure of Iran is characterized by its centralized system of control. At the time of the National Health Survey there were 24 provinces, further divided into 227 districts, 602 subdistricts, 512 cities

2&3 : The SCI (Statistical Centre of Iran) and MOH&ME (Ministry Of Health and Medical Education) are the most valid references of information on Iran and over 70,000 villages. In November of 1992, and early 1996 two of the provinces were divided and now there are 26 provinces. Our analyses are based on the original 24 provinces and boundary changes should not affect our findings.

The major sources of financing for Iran's rapid modernization and construction is the country's vast oil resources. Other leading exports include fruits and nuts, carpets, minerals, wool and textiles, some manufactured goods (for example trucks and buses), and caviar.

Iran has a multi-lingual and diverse cultured society. The official language and script is Persian. But other languages such as Azari, Kurdish, Luri, Arabic, Baluchi, Armenian and Assyrian are also spoken (UNICEF, 1992). 99.6% of the population are Moslem, and 91% of them belong to Shia branch. There are minorities such as Armenians, Jews, Assyrians and Zoroastrians who freely practice their religion in Iran.

The importance of formal as well as non-formal education is clearly indicated in the prominent document of post-war Economic, Social and Cultural Development Plans (ESCDP). The government of Iran has declared the provision of primary education, for all school age children (6-10), as one of its major qualitative objectives (IRI, 1990). The education plan further declares compulsory general education as a means to eradicating illiteracy, and also significant attention has been paid to needs of girls and women in the formal and non-formal system of education.

1.2.2.1 Health Policies

In order to provide an overview of the public health situation and corresponding activities for its improvement, it is important to explain briefly about the general policy related to health which has been planed by the government of Iran. The health policies as stated in the second Socio-Economic Development Plan (SEDP) are set by the Ministry of Health and Medical Education (MOH&ME) and overall can be summarized as below:

- To keep the priority of public health and preventive care over curative practice, emphasizing the expansion of the former priority to prevent physical and mental disabilities;
- To expand the Primary Health Care (PHC) of the country and strengthen the referral system while using the support of health insurance policies;
- To promote an equitable distribution of health resources with special emphasis on the vulnerable groups, rural areas, and peri-urban areas;
- To seek the support of all government sectors in

conjunction with mass media in order to promote community participation and to raise the awareness level of the community;

- To continue the policies leading to reduction of the population growth rate;
- To upgrade the quality as well as quantity of the PHC service delivery by involving medical universities and promoting Community Oriented Medical Education (COME);
- To promote the use of applied research in health delivery systems;
- To emphasize the policy of decentralization, giving decision making capabilities to the districts and other implementing units;
- To improve and strengthen the information system related to health statistics and computerization of the system (MOH&ME, 1995).

1.2.2.2 Some aspects of health in Iran

In the last decade the infant mortality rate (IMR) and child mortality rate (U5MR) have fallen. In the years of 1980 to 1991 and 1995 the IMR fell from 89 (per 1,000 live birth) to 37 and 31, and the U5MR from 114 to 43 and then to 35; there is no noticeable difference between the rates for boys and girls. Life expectancies at birth in 1991 for males and females were 61 and 62 years respectively (Shadpour, 1993). Also, diarrhoeal disease-related mortality rates have fallen off considerably since the mid-eighties; from 1985 to 1991 total diarrhoeal disease deaths fell from 25% of under-five mortality to 17%. Breast-feeding practices have been encouraged, and the MOH&ME survey (1995) revealed that 79% of urban and 85% of rural children are breast-fed for a period of at least one year (in whole country 83%).

1.2.3 The National Health Survey 1990-2 in Iran

Research is the fundamental basis for national development. Recently in Iran research activities are considered to have top priority, with the expansion of its programs by provision of adequate resources and equipment in order to accelerate our national development.

The necessity for research in the health system of our society and consideration of specific problems like population growth, endemic, tropical disease and other major conditions including, cardiovascular disease, thalassaemia, cancer, diabetes, etc., were the principal motivations for the Under Secretary of Health of the Ministry of Health and Medical Education to conduct a nationwide health survey in Iran. Without such data health programs are likely to be misdirected and wasteful.

The main purpose of the survey of national health can be

summarised as follows:

1) To promote awareness and knowledge about national health and medical problems;
2) To obtain quantitative and qualitative assessment of the health and medical problems in different areas of the country, in relation to population, geography, social and other factors;
3) To establish policy and strategies in accordance with the needs of the country;
4) To make available baseline information in order to evaluate the implementation of programmes.

In order to achieve these goals, a six part questionnaire was constructed to obtain data under the following headings:

a) Comprehensive information related to the family and health;
b) Food consumption over two days;
c) Reproductive information on 15-49 year old married women;
d) Individual information and medical examination including weight (kg) and height (cm) measurements of children and adolescents aged 2-18 year olds;
e) Dental Health;
f) Laboratory results (for example, from the analysis of blood samples).

Preliminary studies and preparation of the questionnaire (Appendix A) took four months. Finally in August 1990, the project was started in all provinces with the help of medical universities, and the Under secretary for Primary Health and Blood Transfusion Centres. The pilot study was carried out in Semnan province, and later in other provinces. For certain information, the unit of study was the family and for others individuals aged 2 to 69 years. Table 1.3 presents the time schedule of the Survey; it took over 22 months to collect the data in all provinces of Iran.

Data collection was carried out in the form of cluster sampling, each cluster containing seven families. The selection of seven families for each cluster was dependent on the ability of examining group (composed of 5 persons, 2 physicians, 1 dental hygienist, 1 laboratory technician, and 1 interviewer) to be able to contact them during one working day.

The sample size was 1/1000 of the total population. The sampling framework was based on the regularly updated family index that was available in each health department of every province (except for a few large cities like Tehran, Mashhad, and Shiraz). Cluster selection was carried out by systematically selecting 1 in 7000 families from these family indexes. The cluster comprised of the selected families together with the six nearest neighbouring families.

In cities, where an index of families was not available, the number of families was approximately known from the 1986 census data. A list of 1 in 7000 was obtained by selecting randomly the required number of families from a list of all mothers having their first or second child in the city's hospitals over a 48-hour period. These addresses provided a geographical starting point for each cluster of seven families. It should be noted that the size of samples obtained in this way was approximately proportional to the city's population.

Table 1.2 presents the distribution of Iranian children in each age band in our data. The result of survey in Iran provided data on about 10,660 families involved in the study. These families included 22,349 children on whom anthropometric measurements were made which form the basis of this study[4].

It is worthwhile mentioning that the structure of data in this survey is hierarchial, with 4 levels of nested membership. The highest level is province followed by clusters within provinces, families in clusters, and

4: The number of children with both weight and height measurements after exclusion of discordant measurements (to be discussed in chapter 3) children in families. The statistical analysis takes account of this hierarchial structure (to be discussed in chapters 4 and 5).

1.3 Aims of the present research

Age-related centiles are commonly used in the routine monitoring of individuals, where interest lies in the detection of extreme values, possibly indicating abnormality. Such charts are used widely in paediatrics, for measurements related to growth and development. If the population centile corresponding to the subject's value is atypical this may indicate an underlying pathological condition. The chart can also provide a background for monitoring the measurement as it changes with time (Cole, 1988a).

Such is the background to the importance of an appropriate growth study and the necessity for local standards. The availability of the representative data from National Heath Survey 1990-2 provides the opportunity to address these needs. Hence the main objectives of this research study can be outlined as follows:

1) Demonstration of growth patterns of children and adolescents in different provinces of Iran
2) A study of the way in which differing growth patterns in Iran may be grouped
3) A search for a reasonable base line to represent the

growth patterns in Iran

4) Comparisons of different approaches to the construction growth charts

5) Consideration of assessing the fit of the centile curves

6) Construction of weight-for-age[5] and height-for-age[6] norms for children and adolescents in Iran.

In summary the overall aim of this research is to construct charts that will provide a norm for assessing the growth of children and adolescents, which can be used in the health system (PHC) in Iran.

1.4 Thesis outline

This chapter has provided an overview of the importance and advantages of growth monitoring and promotion, and gave some background information about Iran and the National Health Survey 1990-2. Also the necessity of studying the growth pattern of children and adolescents and need for local norms, which is the primary objective of this study, has been addressed. Although, no overview can adequately summarize the contents of each chapter, some highlights are given below and may be useful for the reader. Chapter two discuses the importance and benefits of growth monitoring

5&6: Instead of weight-for-age and height-for-age charts throughout this thesis for simplicity, the words weight and height chart have been used as a key component of promoting healthy growth. Chapter two reviews the usage and interpretations of anthropometric indicators as a widely accepted tools for assessing the nutritional status of individuals and populations, and continues with a discussion of why centiles curves may be the most appropriate system of reporting these indices. Moreover, the necessity of using local standards instead of international references is discussed, and a general overview of Primary Health Care (PHC) and growth monitoring in Iran is provided.

Since the problem of misrecording and error in measuring in any observational study is unavoidable, especially studies of nation-wide magnitude, chapter three reviews different methodological approaches for detection of discordant or outlying observations in multivariate data. A technique for transformation of the data to Multivariate Normality is presented and the exclusion of measurements with too large a distance measure (outlying observations) is discussed. Checking the original records indicates that the applied method is highly specific in labelling the measurements as outilers.

Chapter four gives a review of different approaches to growth chart construction with a discussion of the methods which have been used throughout this dissertation. Since taking account of the structure of the data is very important in any appropriate analysis, construction of multilevel models within the frame work of the structure of underlying data has been briefly explained.

Results of a preliminary study of growth patterns of children and adolescents is presented in chapter five. The features of the preliminarily analysis of growth are illustrated across provinces and for the whole country. Then the comparison of a 4-level analysis of growth patterns with the results of 3-level modelling within provinces are shown and discussed. Chapter six reviews various scenarios for producing growth charts for children in Iran, and analytically considers the possibility of using regional growth charts in different parts of the country. Then since the results of grouping the patterns of growth were not found practical, chapter six presents an analysis of the plausibility of urban Tehran being a base line for the urban areas of Iran. The conclusion is reconfirmed by further exploration of the distribution of different centiles.

Four recently developed methods of growth chart construction are compared in chapter seven. Healy's (1988) method using spline procedures is shown to produce the best fit. In consideration of different techniques for assessing goodness of fit, a new test statistic for comparing the fit of different models to the raw centiles is introduced and the corresponding distribution is also derived.

The growth charts of weight and height for Iranian boys and girls are constructed using Healy's method and presented in chapter eight. For each chart, goodness of fit was verified by several tests. Also, a suggestion is made for achieving an appropriate fit to the centiles by optimum selection of polynomials on age and Z in Healy's method. Comparisons of our norms with a previous study in Iran and also with the NCHS data is presented. The results show why our norms are realistic for the growth of Iranian children. However, growth of rural and urban children differs; a practical solution enabling one set of charts to be used for both groups of children is suggested.

A short review of the relevant literature is included in most chapters, because these chapters form a distinct pieces of work in their own right.

Table 1.1 The population and growth rate of Iran; 1986-91

	1986	1991	Population Growth Rate (%) 1986-91
Urban	26,844,561	31,836,598	3.5
Rural & Nomad	22,600,449	24,000,565	1.2
Total	49,445,010	55,837,163	2.5

Source: MOH&ME, 1995

Table 1.2 Sample distribution and ratio of age groups

Age group (years)	No. of sample	Sample fraction (10^{-3})
2-4	4394	0.90
5-9	7989	0.90
10-14	6501	0.90
15-18	3465	0.73
Total	22349	0.85

Table 1.3 The data collection time schedule of National Health Survey 1990-2 in Iran according to provinces

Code	Name of province[7]	Date of Survey Month[8]	Year
1	Semnan	August	90
2	Chaharmahal-Bakhtiari	October	90
3	East Azarbaijan	November	90
4	Isfahan	November	90
5	Kordestan	December	90
6	West Azarbaijan	December	90
7	Hamadan	December	90
8	Bakhtaran	December	90
9	Kerman	February	90
10	Tehran	March	90
11	Fars	July	91
12	Kohkiloyeh-Boyerahmad	July	91
13	Boushehr	July	91
14	Mazandaran	September	91
15	Khorassan	September	91
16	Gilan	October	91
17	Zanjan	October	91
18	Lorestan	October	91
19	Yazd	November	91
20	Ilam	January	91
21	Sistan-Balouchestan	January	91
22	Markazi	February	91
23	Khouzestan	February	91
24	Hormozgan	June	92

Source: Zali et al., 1993

7: There are minor differences between the spelling of the names of the provinces in English. In all Tables and all through this thesis, the spelling of names of provinces is taken from UNICEF (1992; p.30)

8: The middle of each month of Persian month of data collection was translated to the corresponding English month

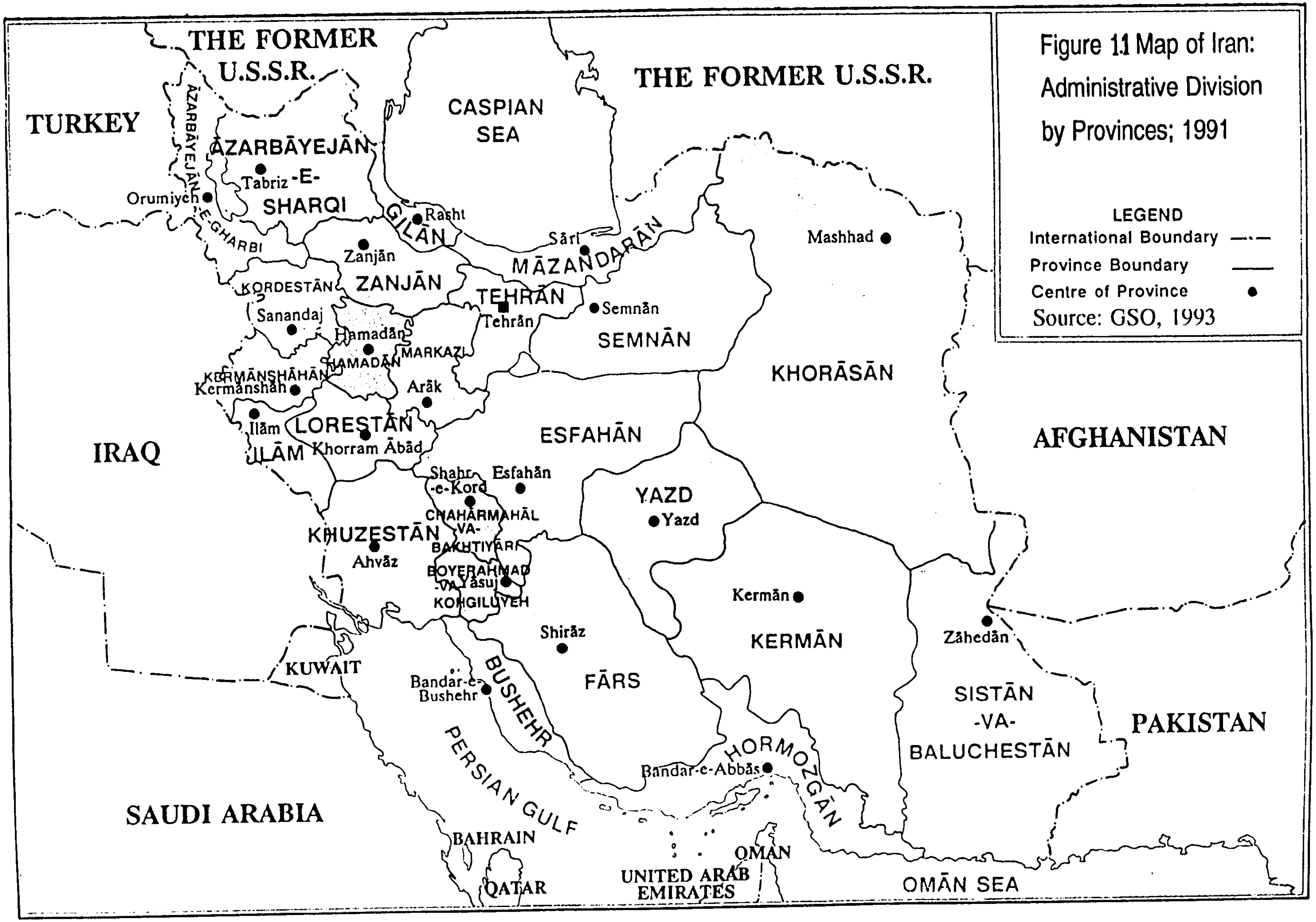
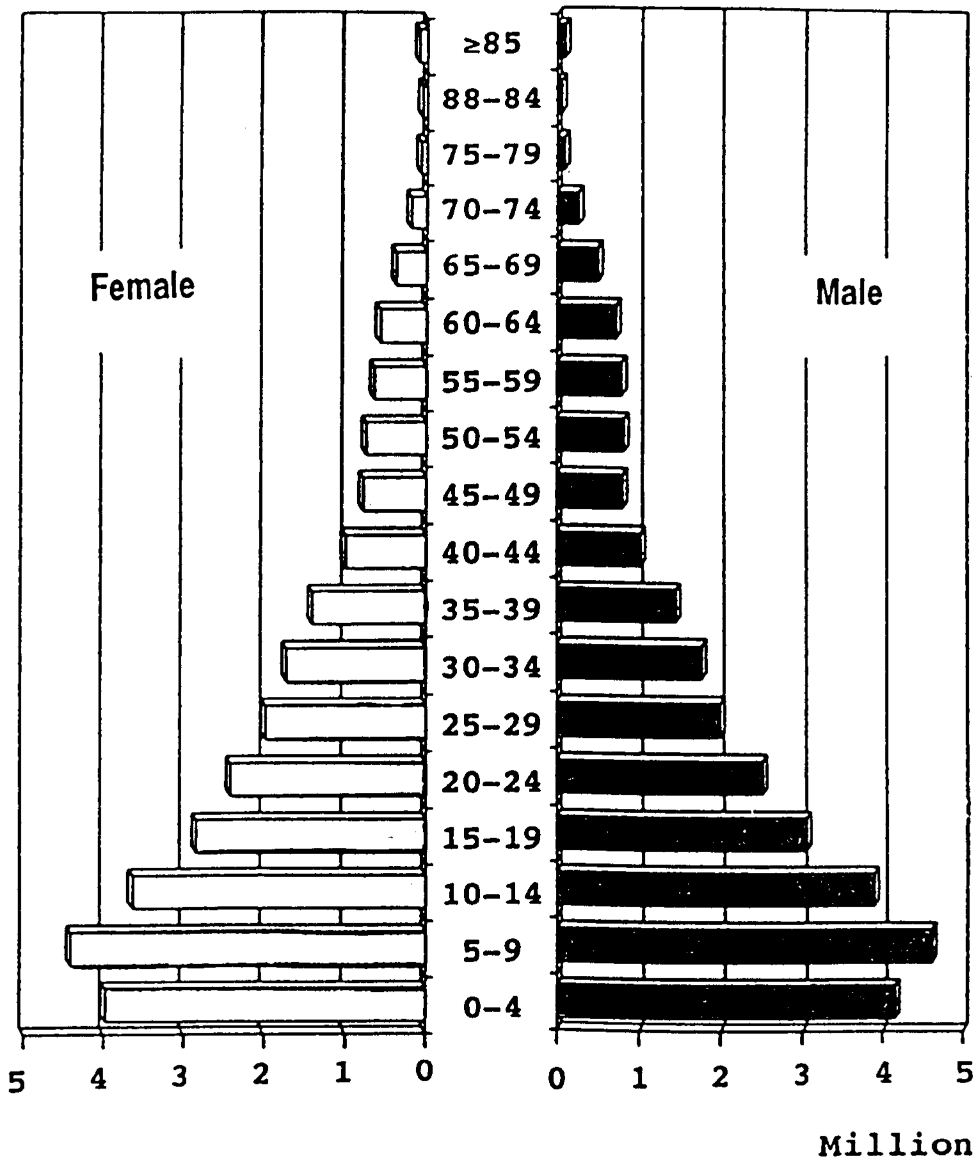


Figure 1.1 Map of Iran: Administrative Division by Provinces; 1991

Figure 1.2 Population age pyramid of Iran, 1991



Source: SCI, 1991

CHAPTER TWO

THE IMPORTANCE OF GROWTH CHARTS AND MONITORING

2.1 Growth monitoring and promotion

In recent years, there have been many publications on growth monitoring and promotion (GMP) with contrary opinion and 'evidence' concerning the usefulness, the effectiveness, and the necessity of including GMP in primary Health care (PHC) programs (Ruel, 1995). Hall (1989) concludes growth monitoring plays a very important role in preventive and curative medicine and its impact on child health and their mental, social and psychological health in puberty and adulthood.

Growth monitoring which can promote child health, human development, and quality of life, in practical terms is a process of measuring children's physical growth periodically, and when necessary initiating actions that can promote normal growth. The term 'promotion' is added to 'growth monitoring', and designated GMP, to emphasize the action component of the activity. This description makes it clear that GMP, which consists of measuring children, is not an intervention in itself but rather a process or strategy to generate action.

The health of the adult is in part determined by his/her health as a child; the growth and development of one generation can affect the next generation. Physical growth is a key indicator of child health and is recognized as a sensitive index of the health and nutrition of the population. Monitoring growth of young children helps to detect growth failure at an early stage; and if addressed immediately adequate growth will be maintained and survival and health will be assured. For example, a child whose height is not increasing as it should or who suffers weight loss may have an underlying medical problem, an emotional problem or even a genetic defect (Tanner, 1986).

The purposes of GMP can be divided into three broad categories: (1) an educational and promotional tool; (2) an integrating strategy , and (3) a source of information. GMP is a tool that aims to improve processes, motivate individuals or groups to take action, and/or provide information that will guide decision making.

First, as an educational tool, which includes breast-feeding support, timing and selection of complementary food and appropriate weaning foods, GMP makes malnutrition visible to both health workers and mothers, and facilitates the formulation of individualized advice to meet the specific need of each child (Griffiths, 1988). GMP makes nutrition education action-oriented, specific, and relevant; and thus more effective. By strengthening the communication component of nutrition education, GMP improves the process of nutrition education and its impact on behaviour change (Hendrata, 1987). A study by Ruel et al. (1990) supports this hypothesis in that Lesotho mothers receiving high-quality nutrition education in combination with growth charts improved their nutrition knowledge to a far greater extent that those receiving nutritional education alone. Indeed, this effect was even greater in the high risk segment of the sample (lower-educated, primiparous mothers with malnourished children).

Also, as a motivational tool, GMP can be used to promote growth in children and make healthy growth a goal to be achieved by mothers, health workers, communities, and governments: GMP can 'create a mass of awareness of the importance of the growth problem and a mass demand for its solution' (Grant, 1987). For GMP to create mass awareness about growth in children, active community participation by members and their leaders is highly recommended (AED, 1989). Both of these advantages, however, also require a well-organised program to support the education and community development activities, apart from the support and supervision of the measuring and plotting related tasks.

Second, as an integrating strategy, 'growth monitoring offers a way of uniting the low cost actions outlined by child survival initiatives into a synergistic whole which could improve child survival' (Grant, 1987). As an integrating strategy GMP has the potential to improve child survival and growth by reducing morbidity as a result of more efficient delivery of health services and through more appropriate food intake as a consequence of the more effective delivery of nutrition education and of its greater impact on maternal feeding practice. Improvements can be expected from increased coverage and quality of services delivered and potentially from a synergism among the various interventions. The process indicators of GMP's effectiveness as an integrating strategy include change in immunization coverage, delivery of oral rehydration solution packages, use of family planning, number of sick children referred and treated by medical staff, attendance and regularity of attendance at clinics or health posts (where applicable), and others. However, two conditions are necessary for GMP to be an effective promoter of other health and nutrition services: (1) the rate and regularity of attendance at the health delivery point must be improved by the inclusion of GMP and (2) other health and nutrition services must be available and properly delivered to the intended beneficiaries (Ashworth and Feachem, 1986).

Third, the data generated by GMP activities can be used for various purposes. At the individual level, they can be used to assess the adequacy of a child's growth and guide the targeting of appropriate actions. This use of GMP for screening and targeting is universal. At the program or national level, the data can be aggregated and used for monitoring, evaluation, and surveillance purposes. The important difference between the use of GMP as screening or monitoring strategy and its use as an integrating or motivational tool is in the quality of data required. For GMP to be a useful screening tool, children must be measured with sufficient accuracy to minimize misclassification and maximize the cost-effectiveness of targeting. Similarly, if the data are to be aggregated and used for evaluation and surveillance, higher accuracy in the measurements will ensure higher precision of the estimates of prevalence and of the changes in prevalence over time. By contrast, if GMP is used to create interest and awareness about growth or as an integrating strategy, the data do not need to be of such high quality for GMP to achieve its motivational objectives.

In 1993, an evaluation of the UNICEF-supported growth monitoring activities was conducted by UNICEF in seven case study countries: China, Ecuador, Indonesia, Malawi, Thailand, Zaire, and Zambia. The mandate of the evaluation was to document whether the information generated by growth monitoring assessment and analysis did, in practice, lead to action. In the context of UNICEF's nutrition strategy, growth monitoring was seen as a strategy to assist families and communities in getting the necessary information to help them with their decision-making process. UNICEF's evaluation confirms the general success achieved in most GMP programs in teaching mothers and health workers the appropriate use of the technology involved (e.g. scales and growth charts). The evaluation concludes, however, that in all seven study countries 'few action – either at household or at community levels – aimed at improving the nutritional status of children, were reported by caretakers of children and monitors' (Pearson, 1993). Other limitations of GMP related to use of the data for decision making at regional and central levels are discussed in the report, namely the problem of coverage and lack of representativeness of the data. But overall, UNICEF'S evaluation corroborates most of the discussion presented here.

It is worthwhile mentioning that critics argue that the role of GMP should be seriously reconsidered and that more research is needed on its feasibility and cost-effectiveness (Gerein, 1988; Gopalan and Chatterjee, 1985; Nabarro and Chinnok, 1988), while supporters believe that current evidence is sufficient to support continued practice and implementation of GMP in a growing number of programs worldwide (Rohde, 1988; Rohde and Northrup, 1988; Griffiths, 1988; Hendrata, 1987). Part of this controversy seems to result from lack of consensus on what GMP really is, what it can potentially achieve, and how it should be evaluated.

2.2.1 Use and interpretation of anthropometric measurements

Anthropometric measures are generally the best global indicators of physical well-being in children because inadequate food intake, poor nutritional quality of the diet, and various infections affect growth. Poor growth is also a predictor of other undesirable outcomes such as increased morbidity and mortality in early childhood. However, one way of measuring the health and nutritional status of children as a whole in communities in the developing world – despite inadequacies in medical records – might be by the number of cases of kwashiorkor, marasmus and xerophthalmia. But this would not be very useful because such cases are the tip of the iceberg of the conditions they represent. Alternatively various anthropometric measurements can be used to assess individual or group growth or nutritional status. Among the measurements most studied are: weight, height, arm circumferences, skin-fold thickness, chest circumference, and head circumference. Gorstein and Akre' (1988) concludes they might be less accurate than clinical and biochemical techniques when it comes to assessing individual nutritional status. But in many field situations, where resources are limited, it is possible to use anthropometry as a screening device to identify individuals at risk of undernutrition.

Weight is a measure of total body mass, and hence is sensitive to change in body fluids, fat, muscle mass, the skeleton, and other organs. Arm circumferences assess the degree of muscle and fat (around the bone) in the mid upper-arm area. Skin-fold thickness is an indication of body fat reserves and gives information about body composition which is additional to that given by weight. Height is a measure of the linear growth of body – the degree of skeletal development and its deficits indicate long term, cumulative inadequacies of growth or nutrition. Two related terms – length and stature – are also used. Length refers to the measurement in a recumbent position, and is often used for children under 2-3 years of age who can not stand well. Standing height measurement is often referred to as stature. As the age range of our study was 2-18 years old, our measurements of height are of stature and the term height is used throughout the thesis.

Anthropometric *'indicies'* are combinations of measurements. They are useful for the interpretation of measurements: it is evident that a value for body weight alone has no meaning unless it is related to an individual's age or height (WHO, 1986). An *'indicator'* is often constructed from indices; thus , the proportion of children below a certain level of weight is widely used as an indicator of community status. Indices have been suggested, sometimes to distinguish *'types'* of malnutrition

(Seone, and Latham 1971). Also, Gorstein et al. (1989) stated each index expresses a distinct biological process and their use has permitted a distinction between different types of undernutrition which have different etiologies. This distinction is quite important for public health purposes and for the epidemiological assessment of nutritional status. Waterlow et al. (1977) believed that weight and height measurements together are needed to understand the dynamics of malnutrition and distinguish between current malnutrition and long-term or chronic malnutrition. Visweswara Rao et al. (1977) argued persuasively that these measures are adequate for assessing nutritional status and that not much is gained by additional forms of measurements. Gopalan (1985) stated that weight, height, and arm circumference have come to be considered the most *'sensitive'* parameters in under five year old children's health status, and most practical for monitoring of individual children, or of a population.

It is important to note that all indices derived from age-specific reference data depend for their precision on exact knowledge of age; when this information is not available, use of age-based indices such as height may result in misclassification (Gorstein, 1989). So, other indicators such as weight-for-height, and arm circumferences-for-height have been proposed for growth monitoring (APHA, 1981) which is useful when age is not known. They may be a sufficient tool for screening in emergencies, that is for counting the undernourished (WHO, 1986). Weight-for-height reflects body weight relative to height. However, it is important to note that weight-for-height does not serve as a substitute for height or weight, since as it is stated each index reflects a different combination of biological processes. An FAO/UNICEF/WHO (1976) expert committee on nutritional surveillance recommended the use of height and weight-for-height as primary indicators of nutritional status in children.

The choice of indices and indicators is subject to constraints, there are practical limits to feasibility, accuracy, and precision of all measurements, including that of age. The size of the sample and the number of measurements that can be made are constrained by the resources available. For children the use of two indices weight-for-height and height, my be recommended for most purposes but not necessarily for all. In certain instances the combined index weight (weight represents the sum of the information given by the other two indices (Keller, 1983)), is more practical for giving an overview of the distribution of nutritional problems in a country, or the direction of change.

The anthropometric indices can be expressed in terms of Z-scores, percentiles, or percent of median. These reporting systems are as follows: (1) Z-scores (or standard deviation scores) - the deviation of the value for an individual from the median value of the *reference population*, divided by the standard deviation for the reference population. A fixed Z-scores interval implies a fixed height or weight difference for children of a given age. An advantage of this system is that, for population-based applications, it allows the mean and standard deviation to be calculated for a group of Z-scores. (2) Percentiles - the rank position of an individual in a given reference distribution, stated in terms of what percentage of the group the individual equals or exceeds. Percentiles are commonly used in clinical settings because their interpretation is straightforward. (3) Percent of median - the ratio of a measured value in the individual, for instance weight, to the median value of the reference data for the same age or height, expressed as a percentage. The main disadvantage of this system is the lack of an exact correspondence with a fixed point of the distribution across age or height status. For example, depending on the child's age, 80% of the median weight might be above or below -2 Z-scores; in terms of health, this would result in different classifications of risk. In addition, typical cut-off for percent of median are different for the different anthropometric indices (Dibley et al., 1987; Gorstein et al., 1994).

Anthropometric measurements, recorded together with age and sex, provides a valuable profile of body composition and physical development that is an expression of nutritional status. These indices, in turn, are compared with a reference standard in order to assess the relative status of individuals or groups. However, the question of which reference (standard) these indices should be compared with is a matter of controversy which is discussed in the later section.

2.2.2 Adolescent anthropometry

It was stated the National Health Survey in Iran provided the wealth of data on adolescents. Anthropometry is especially important during adolescence because it allows the monitoring and evaluation of the hormone-mediated changes in growth and maturation during this period. Also, since growth may be sensitive to nutritional deficit and surfeit, adolescent anthropometry provides indicators of nutritional status and health risk, and may be diagnostic of obesity. Many important changes in psychological and social development take place during adolescence. Potential for pregnancy and parenthood, educational choices, occupational commitment, and interpersonal relationship are only a few of the concerns and responsibilities that challenge adolescents. While adolescence is clearly an important period in human development it has often failed to receive the attention given to the earlier period in childhood with regard to health-related uses interpretations of anthropometry.

In addition, because of the variable timing of the pubertal growth spurt the NCHS[1] reference charts for weight and height include this information for children over 10 years of age but do not recommend its use. The interpretation of indices in adolescents is complicated by the fact that body composition is more variable at this age than in healthy younger children and difference in fatness and muscle mass between the two sexes increases with age. Therefore, having these data available it is fascinating to investigate the pattern of change of weight and height in Iranian adolescents.

2.3 Growth charts

For effective growth monitoring, a child's indices (for example, weight) must be recorded against his age and serial weight-for-age must be readily comparable. Plotting weight data on a growth chart permits this. It also enables comparison of the child's status with the desired norms of reference standards and thereby, estimation of the degree of growth deficit. The idea that monitoring the growth of the individual child on a long-term basis would be useful in the provision of child health care gave rise to the

1: National Centre for Health Statistics "growth chart" pioneered by Morley (1973) as a result of a long term study of growth in Imesi, Nigeria. He produced his growth chart on the simple principle of regular monthly weighing of children under five. Each child's weight is plotted every month against his or her age, giving a 'weight-for-age' graph, which is regarded as most 'dynamic' and sensitive to change.

The growth chart incorporates in it a reference growth trajectory, allowing comparison of a child's growth curve, with that of the reference population. This is the 'road-to-health' concept envisioned by Morley. His chart was standardised for use with both sexes by employing the 3rd centile of data on girls as the lower reference line and the 50th centile of boys as the upper reference line. Since then there has been a proliferation of charts, using various reference curves, cut-offs and so on. The WHO produced a chart based on the Morley chart (for example, see Figure 2.1), using the NCHS standard (WHO, 1978) with the aim of putting order in this puzzle.

The major role of the growth chart is to focus attention on promotion of growth. Morley suggests that in the case of attempting to provide health care for children, observing the chart of each child, and monitoring adequate growth is important. He stresses that the reference lines are intended to illustrate the direction in which a child's growth curve should travel, i.e., providing a way of comparing the child's rate of growth with good growth. A growth disturbance is noted by comparing the gradient or slope of the child's growth over two or more measurements with that of the reference lines drawn on the chart. The child's own earlier growth trajectory are also important.

Morley (1973) affirms that the slope of the curve is more important than its position in relation to the reference lines. The corollary of this is that looking at actual position of weight at any age by itself may be misleading. In other words, children's growth charts should simply interpret the direction of the curve derived from the individual child's own measurements; up (gaining) relative to the standard trajectory is good. Down (loosing) in relation to the standard trajectory is a warning; actual loss of weight is alarming.

The improved charts are now integral parts of several health care systems. They also function as a readily available record of age, immunisation status, past illness, and at risk status, as a ready made referral document for primary or secondary care elsewhere, and a concurrent denominator-based tool for research and evaluation (Cunningham, 1978). Growth charts were originally developed for use in under-five clinics and so only included the first five years (Morley, 1973). The WHO modification added the sixth year (Shadpour, 1994; Figure 2.1) so emphasis has been placed on weighing all children under six. The effectiveness of nutritional interventions can be apprised by monitoring the growth and return to normal growth pattern in individuals or groups. Relative increases in weight-for-age in serial measurements recorded on the chart are a measure of the *'success'* of programs in improving nutritional status. It is obvious that growth charting based on weight-for-age is only one method of growth monitoring. Other indices such as height-for-age, weight-for-height, arm circumferences have been proposed as well. As was stated, the last two are particularly useful when age is not known and they may be more appropriate for rapid screening of malnourished children. For example, Nabarro and McNab (1980) devised a weight-height chart for field use which is known as the 'Thinness Chart' since it establishes when children are low weight-for-height, i.e., 'thin' or 'wasted'.

Newell (1975) states that 'the growth chart has been found to be appropriate in a number of community programs'. *In conclusion, the value of growth charts can be summarised as follows:*

1) It provides a continuous record of health. When a child is sick, there is valuable background information to help treatment.
2) It is the basis for comprehensive health monitoring and

care. It avoids the dangers of giving a single response to a specific need, without consideration of the total picture.

3) It encourages positive action for all children. Promotion of adequate growth is a more useful target than prevention of malnutrition.

4) The chart is of the greatest value in terms of action as it makes the serious disease of malnutrition easy to diagnose. The role of the weight chart is to identify the potential danger of malnutrition before management becomes too difficult at the primary health care level. The weight curve may identify trouble six months or even a year before the child has obvious signs of malnutrition (Morley, and Woodland, 1979). The opposite is also true. A child who is gaining weight regularly and whose curve is parallel with the curves on the chart is unlikely to develop malnutrition.

2.4 Why is a national growth chart required?

A child's growth data are usually compared with that of a *'reference'* population to evaluate his/her health status. The question of which population provides the most relevant *'reference'* for groups of undernourished children in developing countries has engendered much debate (Gopalan, and Chatterjee, 1985). Moreover, misunderstanding particularly arises over the meaning of *'normal'* growth.

Since the weight and height of a child at any point in time will depend on his genetic potential (Morley, 1977) as well as environmental factors.

The term *'normal'* carries two distinct meanings. The first aims to relate the child to some kind of perfect or ideal standard; the *'norm'* in this sense is regarded as target (standard), quite possibly an inaccessible one which no child actually reaches, and opposite to *'normal'* is *'subnormal'*. In the second sense, *'normal'* is roughly equivalent to *'commonly occurring'* or *'ordinary'*. The *'normal'* is now simply the situation which most commonly occurs, and the opposite of *'normal'* is *'abnormal'*. It is important to state that *'normality'* in human growth studies (as throughout most of medicine) has in practice the second of these two senses. Hence, the normality of a child (or more precisely of some measured aspect of that child) is assessed by the frequency of occurrence of his measurement in an appropriate *'standardizing group'* or reference (Healy, 1986). Large numbers of studies refer to growth in *'normal'* children or growth in *'non-normal'* children and use their own definition for their study; *'normal'* could be thought of as *'expected'*. This moved people to attempt to define what might be expected. Both Hauspie (1980) and Marubini (1980) discussed normal and abnormal growth and related this to appropriate standards for the population under study. Bell (1989) uses the term *'failure to thrive'* as describing an infant who fails to achieve the growth or weight gain that is normally expected.

The most frequently used reference standards are derived from studies of growth in healthy children from the US and UK The 'Harvard Standard' of weight-for-age were obtained by a study of well-nourished Caucasian children in Boston in the 1930s (Stuart, and Stevenson, 1959) and have been used extensively through the world. Weight-for-height have also been calculated from these data. Reference data were also collected on British children (Tanner et al., 1966) and these have been used in the development of the *'Road-to-Health'* card (Gopalan, 1985). More recently, the (US) National Centre for Health Statistics (NCHS) has collected data on weight, height, arm circumference, skin-fold thickness, and head circumference on a large, economically and ethnically heterogenous sample of the US child population (NCHS, 1977). Despite the heterogenous sample used, compared with the homogeneity of Boston and British groups, the previous standards have been lowered only minimally.

The world health organization (WHO) has adopted the references curves of the NCHS for worldwide use (1978). These curves are based on several sources of data from the United States (Hamill et al., 1979). For children under two years of age the data are from Fels Research institute in

Yellow Springs, Ohio, and come from studies of a white, middle-class population. For older children the data come from nationally representative surveys of children in the United States and include all ethnic groups and social classes. These curves and the CDC (Center for Disease Control) programs for calculation of corresponding Z-scores have a number of technical problems because two distinct data sets were used to construct the reference curves (Dibley et al., 1987; Yip and Trowbridge, 1989). The implication of these problems for analyses and interpretation of anthropometry data in developing countries are not fully known. Concern also has been expressed that the NCHS curves are inappropriate for healthy, breast-fed infants. Some authors believe that these infants tend to be lighter, nearly as long, but similar, in head circumferences when compared to the NCHS references (Butte et al., 1984; Neville and Oliva-Rasbach, 1989; Dewey et al., 1992).

The justification for using the NCHS data is the evidence collected by some studies, like those of Habicht et al. (1974) and Graitcer et al. (1981), that in populations the effect of ethnic differences on the growth of young children is small compared with the effects of the environment, as they adhere to the view of equal growth potential during early childhood for children of all races. However, the assumption that all populations are genetically the same so far as growth is concerned is unfounded (Eveleth and Tanner, 1990). Davies (1988) in different studies on the importance of genetic influence in early childhood stated that 'it is difficult to avoid concluding that the early growth pattern of Asiatic children is different from that of the NCHS reference'. And catch-up of faltered growth, which would have been expected with the provision of a well-balanced diet does not take place. Although, it is accepted by WHO (1986) that there may be some ethnic differences between groups, just as there are genetic differences between individuals, but for practical purpose they are not considered large enough to invalidate the general use of NCHS data both as a reference and as a standard (WHO, 1985).

On the other hand, some investigators have argued that the use of a reference standard derived from a developed country's population, or even from well-nourished groups in developing countries, sets impossibly high 'standards' (Eusebio, and Nube, 1981). Seckler (1982) suggested that not only the Harvard or NCHS standard but even the 'best indigenous standard' derived from growth measurements of their affluent sections would also be 'abnormally large' for some populations, for example India. It has been proposed that developing countries should evolve their own reference standards in order to be more 'realistic' about potential growth achievements of their child populations

(Goldstein and Tanner, 1980; Seth et al., 1979).

Goldstein and Tanner (1980) provide an in-depth discussion of this topic. They conclude when screening children it is vital that the standards with which they are compared are actually derived from that population. For example it would not be sensible to use height standards derived from data on a tall population to assess the normality of children who belonged to a short population as otherwise one would conclude that many of the children were short and possibly suffering from some illness. ***However, there is a balance to be struck between constructing a chart for any given population and not generating thousands of standards for every sub group of children.***

Waterlow et al. (1977) stated that because the reference chart cannot be used as a universal target, the question of what is a realistic goal in any particular situation does become important. If it is felt that the growth of children in an industrialized country is not a realistic target in another country in which the population has a different genetic and environmental background, two courses are possible: the first to construct a local standard, the second is to make an arbitrary adjustment in the cut-off points derived from the reference chart. For example, if it is felt that in a particular population even well-nourished children are shorter in stature than children in the North

American reference population, then it might be reasonable to set the target for height as 95% of the reference height rather than 100%. Waterlow continues that decisions of this kind have to be taken locally, and it is not possible to make international recommendations about them. Relating to this Morley (1973) stresses that the trajectory of growth not the position by itself is important, but this is a fairly crude measure because if the measurements is some distance away from the 'normal' trend line it is difficult to assess trend accurately.

Relating to Iran, previous sparse studies from small selected clinics or groups do not provide a population norm, but have suggested that children in Iran may be a good deal smaller than in the United States (Ayatollahi, 1993a). As it will be shown later in our study all Iranian growth centiles are lower than NCHS centiles, eg. the 25th Iranian centile is about the 3rd centile of NCHS. So, for instance, how does one monitor a child who is healthy and whose growth follows our 10th centile? In some areas of Iran there are many healthy children who would be classified 'malnourished' using the NCHS standard. If this is so, the use of American or European norms in clinical work may be seriously misleading. Where could the facilities and money for treating these 'malnourished' children come from? More importantly, we need to recognize those children whose weight or height are in lower centiles of our 'normal' range – for example, below 5th percentile – in order to provide facilities or specific care or any help which is necessary to promote their normal development. One may also ask whether a few children from high socio-economic background whose weight curves follow 50th or 75th centiles of the NCHS standard are representative of our 'normal' and healthy children or whether they are in fact obese?

Another example of this drawback is a follow-up study of one hundred and seventy four 'healthy' Chinese children (Leung and Davies, 1989) over a period of two years. The median weight of Chinese girls and boys showed a distinct fall away from NCHS median from 6 months onwards. The median of length of Chinese infants runs just beneath the NCHS median. At two years Chinese infants were in general lighter and shorter than American infants from whom the NCHS data were derived. Leung concludes Southern Chinese who are genetically shorter and lighter, would readily be misdiagnosed as suffering nutritional problems if the American-derived reference was used. Thus although anthropometric measurement can be a good indicator of nutritional status, an appropriate standard reference is essential.

Eveleth (1978) also discusses the effect of environmental and genetic factors on population differences in growth.

Cameron (1986) categorised a sample of black south African children using US centiles and found that too many boys and girls fell below the 16th centile and not enough above the 84th centile. The normal healthy population were different from the NCHS reference and use of NCHS standard would have resulted in a large number of children being classified as being too short when they were normal for their own population. He recommended the construction of charts based on appropriate children. Meredith (1971) showed that head circumferences varied considerably between different races, at birth, and during childhood and also adulthood. Likewise, Eusebio and Nube (1981) stressed the importance of having growth standards appropriate to the population under study, and showed increasing differences between standard charts e.g. NCHS and a number of country or population specific charts.

The necessity of assessing the growth of a child using an appropriate standard for that population has led to construction of growth standards for a large number of populations. Terada and Hoshi (1965) summarise a longitudinal study measuring chest and head circumference for Japanese babies for the first 3 years of life. Low (1970) collected heights for Chinese children, aged 6 to 20 years, over a three year period, and constructed height velocity charts. Chen (1985) presents growth charts for Malaysian children, aged birth to six years. Bhargava et al. (1980) reported a longitudinal study from birth to 6 years on Asian children residing in India, while Brooke and Wood (1980) presented the result of a longitudinal study of British Asians, collected from birth to one year. Molteno et al. (1991) presented the results of a study of South African children from birth until 5 years, and Guo et al. (1990) presented growth charts based on monthly growth data from 1 to 18 months on Canadian infants, and also Ayatollahi and Carpenter (1991) present the results of a growth study in one of the biggest cities in the south of Iran.

In summary, while the 'road to health' charts can make a valuable contribution to health in under-developed countries, this is a crude instrument, and there comes a time when clinicians and health workers require a standard against which to assess individuals. Waterlow's suggestion of using a standard international chart with local correction is appropriate when differences are small and fine tuning is only occasionally necessary. When differences are large, such as those between NCHS and Shiraz charts reported by Ayatollahi and Carpenter (1991), international standard charts are virtually useless in the local situation because large adjustments have to be made to interpret any observation accurately. Furthermore, to determine the necessary adjustments, a local chart has to be constructed. Having constructed such a chart, why not use it? Also, it is stated (WHO, 1995) that large, representative sample sizes and sophisticated statistical expertise are required in constructing the curves. Not all developing countries have these resources, but in this study both are available.

The situation is somewhat analogous to time. When watches are scarce, sundials will serve for general use and exact time can be recorded in GMT. When clocks become generally available GMT is inconvenient because it must be corrected to apply locally. So a time zone is set up, for general use in the region. Within such a zone, sun rise and sun set may vary a little and occasionally adjustments have to be made. In the same way because growth in Iran differs substantially from NCHS standards, a chart appropriate for the region is required, to which adjustments can be made if necessary for special situations.

In the same way that population specific data are required, to allow appropriate monitoring of growth, the use of disease specific growth data are required in order to help identify children or groups with particular disability or illness (Goldstein and Tanner, 1980; Tanner, 1981). Such studies are however beyond the scope of this thesis.

2.5 An overview of Primary Health Care (PHC) and Growth Monitoring in Iran

Iran's Primary Health Care (PHC) movement arose out of the attempt to satisfy the public health need, increase coverage, lessen disparity, reduce health cost and improve the health delivery system. The constitution of Iran recognizes the right of all citizens to health and the equitable distribution of health services. The main features of the Master Plan of Health (1983-2002) reflects the commitment of the government to achieve Health For All by year 2000 and clear acceptance by the government of the PHC approach.

Iran has been a pioneer in terms of the adoption of PHC approaches, and has had a fully functioning multi-level health care network in place since mid 1980s. The PHC network in Iran is composed of a multi-level system of fixed health facilities at the base of which is the Rural Health House. At present, The PHC network consist of 26 provincial health centres, district centres, urban and rural health centres, midwifery facilities, and health houses, through which health services are provided to about 66% of the total population. There are some villages which are not covered by this network. Most of these villages receive preventive health services through mobile teams. The community Health Workers (CHW) in villages are native to the area and communities they serve, having had two years of special training after finishing primary school. There are two CHWs one male and one female in each health house (MOH&ME, 1995). Figure 2.2 presents the PHC network expansion in Iran in the period of 1985-93.

As an example, the sharp decline in infant and child mortality rates during the last decade are the immediate outcome of PHC program in Iran although the contribution of other developments such as water supply, roads and education should not be underestimated. Also the PHC network has improved the health status of women (especially in relation to the period of maternity and delivery) and children (e.g. vaccination needs, breast feeding, growth monitoring and control of diarrhoeal disease) at the rural as well as urban level (UNICEF, 1992).

Growth monitoring which is integrated in PHC network activities is one of the responsibilities of health workers. According to UNICEF (1992) 65% of boys under-3 and 63% of girls under-3 in urban areas and 55% of boys under-3 and 56% of girls under-3 in rural areas have growth monitoring charts. It seems the impact of growth monitoring on the improvement of children's nutrition has not been studied as it should have been. There is still poor understanding of the concept of GMP among the health workers as well as in the community. Observation revealed that the emphasis of the program is more on technical aspects of GMP than the nutritional education of the mother and the community. It will be recalled that the growth monitoring and promotion program can only be useful if it provides information on growth status, cause of growth faltering and appropriate intervention to counteract malnutrition. Clear understanding of the GMP concept by all health personnel and the community, particularly mothers, is the key to its success.

As a result of the importance of growth monitoring in reduction of severe and moderate malnutrition of children, the following activities have been implemented by MOH&ME (1995) in the workplan for achieving the mid-decade goals to increase access of children to effective growth monitoring by 30% by 1997:

- Providing training and refresher courses for community health workers aimed at promoting a better understanding of the concept of growth monitoring, and at raising awareness on the cause of growth retardation and appropriate measures to combat growth faltering;
- Support the production of mass media spots and other health education activities aimed at raising maternal awareness of the importance of growth monitoring for the early identification and proper treatment of malnutrition;

- Evaluation and reviewing growth monitoring promotion (GMP) activities, and in assessing the impact of growth monitoring on malnutrition;
- Increase public awareness of the importance of proper weaning practices, train mothers on GMP and other relevant subjects.

2.6 Summary

Growth monitoring and promotion (GMP) is a strategy to detect early growth retardation, promote optimum growth, create awareness about growth among mothers, enhance delivery of primary health care (PHC), and identify those at risk of malnutrition. The 'P' relates to the term 'promotion' in GMP which emphasizes the promotive, preventive and curative actions which accompany the monitoring process. Therefore, growth monitoring is not just periodic nutritional assessment, it involves improving interactions between mothers and community, education about nutrition, and improving child health. Overall the most common ways in which GMP is expected to lead to action can be divided in three categories: (1) an educational and promotional tool; (2) an integrating strategy; and (3) a source of information. So whatever is it's specific purposes, GMP is consistently intended to lead to action, and is a tool to motivate individuals or groups to take action, and/or provide information that will guide decision making.

Anthropometric measurements are generally the best global indicators of physical well-being in children because inadequate food intake, poor nutrition quality of the diet, and various infections affect growth. Poor growth is also a predictor of other undesirable outcomes such as increased morbidity and mortality in early childhood. Finally, the end result of poor growth in early childhood, small adult body size, has functional consequences such as diminished work capacity and increased obstetric risk for woman. It is possible to use a variety of anthropometric measures to assess child growth. Among the most studied are: weight, height, arm circumference, and head circumference. Each index express a distinct biological process and their use has permitted a distinction between different types of undernutrition which have different etiologies. For children the use of two indices weight-for-height and height, may be recommended for most purposes however in certain instances the combined index weight is more practical because of the measure's sensitivity to change.

The growth chart is widely accepted as an important and practical tool that can contribute significantly to child health. It can provide a continuous record of health which encourages positive action and makes the serious disease of malnutrition easy to diagnose. With basic instruction it offers a simple and inexpensive monitor that can be used by community heath workers or mothers.

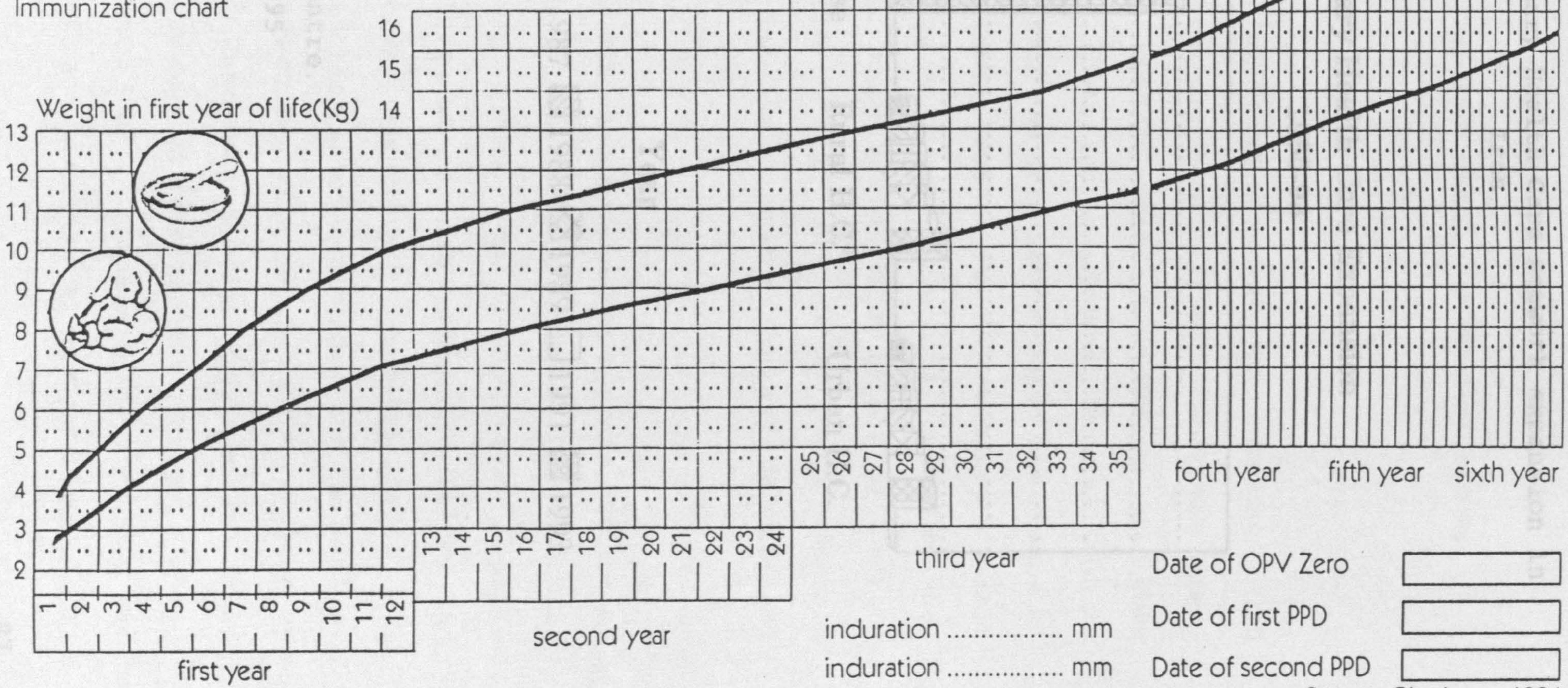
Since there is no guarantee that all populations have the same growth potential (there are certainly large differences between populations in height and weight and the age of puberty, for example) percentile growth standards should be derived from the population to which the children belong. Now it is clear that a portion of these differences is genetic in origin, and in developing countries a larger portion is environmental. Therefore, it may be inappropriate and even harmful to use standards derived from an economically privileged group. Also, it may be inappropriate in one country to use standards derived from another. So the concept of using imported standards derived from populations of developed industrialized countries is unsound.

Iran has been a pioneer in terms of the adoption of PHC approaches. Growth monitoring is integrated in PHC network activities and is one of the responsibilities of health workers. It seems the impact of growth monitoring on the improvement of children's nutrition has not been studied as it should have been. There is still work to be done for a better understanding of the concept of GMP among the health workers as well as the community. As a result of the importance of growth monitoring in reducing severe and moderate malnutrition of children, the MOH&ME has drawn up a workplan for achieving the mid-decade goals. This plan proposed activities to increase the access of children to effective growth monitoring by 30% by 1997.

Figure 2.1 The Growth Chart from birth to six years old

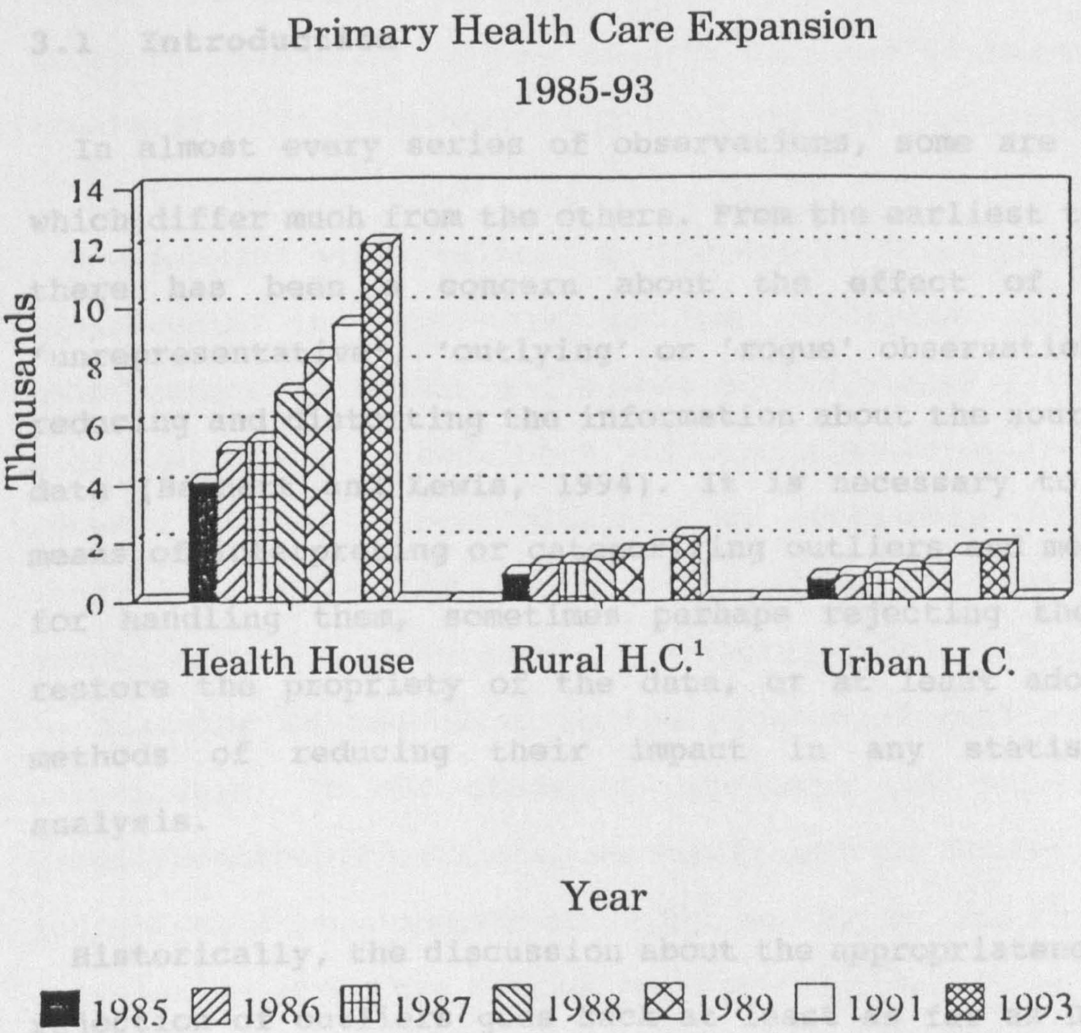
DOSE / VACCINE	1ST	2ND	3RD	BOOSTERS FIRST	BOOSTERS SECOND
BCG					
DPT					
OPV					
DT					
Measles					

Immunization chart



Source: Shadpour, 1994

Figure 2.2 The Primary Health Care network expansion in Iran



1) H.C.: Health Centre

Source: MOH&ME, 1995

CHAPTER THREE

Data cleaning and handling the outliers

3.1 Introduction

In almost every series of observations, some are found which differ much from the others. From the earliest times, there has been a concern about the effect of these 'unrepresentative', 'outlying' or 'rogue' observations in reducing and distorting the information about the source of data (Barnett and Lewis, 1994). It is necessary to seek means of interpreting or categorizing outliers and methods for handling them, sometimes perhaps rejecting them to restore the propriety of the data, or at least adopting methods of reducing their impact in any statistical analysis.

Historically, the discussion about the appropriateness or rejection of outliers goes back at least as far as Daniel Bernolli (1777) and Bessel and Baeuer (1838), while Boscovich (1755) is known to have rejected outliers and this course of action, according to Bernolli (1777), was already a common practice among astronomers of their time. The first formal rejection rules apparently were given by Peirce (1852) and Chauvenet (1836), followed by Stone (1868), who uses a 'moduls of carelessness', Wright (1884), Irwin (1925), Student (1927), Thompson (1935), Pearson and Chandra Sekar (1936), and many others. However it should be mentioned that outlying observations are not necessarily 'bad' or 'erroneous', and the experimenter may be tempted in some situations not to reject an outlier but to welcome it as an indication of some unexpectedly useful industrial treatment or surprisingly successful agricultural variety.

The present study relates to the growth of children and adolescents in Iran, and we are concerned with the measurements of weight and height of individual 2-18 year olds. As stated in section 1.2.3, these measurements were rounded to the nearest kilogram and centimetre. In this study like any another large study, there are sources of error (error of measurement, misrecording, etc) that lead to mistakes in the data. So the problem of outliers was unavoidable. In our study, to optimise the problem of identification of outliers, we simultaneously looked at an individual's measurements on weight and height and used the correlation structure of the data. We considered the joint distribution of weight and height of children at different ages to detect which observations were discordant.

The basic notions, of an outlier as an observation which engenders surprise owing to its extremeness, and of its discordancy in the sense of that 'extremeness' being statistically unreasonable in terms of some basic model, are not constrained by the dimensionality of the data. However, Gnanadesikan and Kettenring (1972) remark, a multivariate outlier no longer has a simple manifestation as an observation which 'sicks out at the end' of the sample. The sample has no 'end'! But, notably in bivariate data, we may still perceive an observation as suspiciously aberrant from the main bulk of the data, particularly so if the data are represented in the form of a scatter diagram.

For example see observations A and B in Figure 3.1. Observation A happens to be an extreme in height (e.g. of a clearly unacceptable value such as a height of 182 cm at 6 year olds), but it is not an extreme in weight. Similarly for B a six year old child with a weight of 91 kg. But a multivariate outlier need not be an extreme in any of its components. Someone who is short and fat need not be the shortest, or the fattest, person around. But that person still can be an 'outlier' (might be misrecording of age).

In general, for univariate Normally distributed observations the standardized distance of observation from the mean is the traditional method of detecting outliers. The method is satisfactory for large samples which includes only a very small proportion of outliers. For Multivariate Normal data, the comparable statistic is the Mahalanobis distance of observation from the sample mean. The formal justification of the use of this statistic was given by Barnet and Lewis (1994), chapter seven, and since it is not well known their presentation is largely reproduced in section 3.2. The results of applying these method to our data after appropriate Normalising transformations is presented in section 3.5. The discussion presents a comparison of using the traditional Mahalanobis distance with recent modifications which aim to avoid the influence of the unknown outliers on the estimates of the mean and variance. The chapter concludes with a brief account of the results of comparing the analytically detected outliers against the original records.

3.2 Outliers in Multivariate Normal samples

The study of outilers and the notion of a test of discordancy for multivariate data is as important as it is for a univariate sample, but the conceptual and manipulative difficulties have limited the number of formal and specific proposals. So as may be anticipated, it will come as no surprise that most of the work on outliers in multivariate data deals with the case of an underlying Normal distribution.

Suppose $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is a sample of n observations of a p-component Normal variable $\mathbf{X}$. We initially assume these to have arisen at random from a p-dimensional Normal distribution, $\mathbf{N}(\mu, V)$, where μ is the p-vector of means, and V the p×p variance matrix. A possible alternative model which would account for a single outlier is the slippage alternative, obtained as a multivariate adaptation of the univariate *models* **I** (slippage of the mean) and **II** (slippage of the variance) discussed by Ferguson (1961). Specifically, the alternative hypotheses are:

model I $E(\mathbf{X}_i)=\mu+\mathbf{a}$ (some i) $(\mathbf{a}\neq 0)$

$$E(\mathbf{X}_j)=\mu \quad (j\neq i) \tag{3.1}$$

with variance-covariance matrix $V(\mathbf{X}_j)=V$ $(j=1,2,\ldots,n)$.

model II $V(\mathbf{X}_i)=bV$ (some i) $(b>1)$

$$V(\mathbf{X}_j)=V \quad (j\neq i) \tag{3.2}$$

with mean vector $E(\mathbf{X}_j)=\mu$ $(j=1,2,\ldots,n)$.

Initially, we shall examine a test of discordancy based on the *two-stage maximum likelihood ratio principle*, which declares X_i to be an outlier if the difference between the likelihoods of the sample under the null hypothesis and the alternative hypothesis is surprisingly large. We consider *models* **I** and **II** separately, with and without the assumption that parameter values are known.

3.2.1 Model I, V known

The likelihood of the sample $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n$ of the basic model is proportional to

$$P_\mu(\mathbf{x}|V)=|V|^{-\frac{n}{2}}\exp\{-\frac{1}{2}\sum_{j=1}^{n}(x_j-\mu)'V^{-1}(x_j-\mu)\}$$

The maximized log-likelihood (apart from the constant factor) is

$$L(\boldsymbol{x}|V) = -\frac{1}{2}\sum_{j=1}^{n} (x_j-\bar{x})' V^{-1}(x_j-\bar{x})$$

The corresponding maximized log-likelihood under the alternative (*model* I) hypothesis of a single outlier is:

$$L_A(\boldsymbol{x}|V) = -\frac{1}{2}\sum_{j\neq i}^{n} (x_j-\bar{x}_{(i)})' V^{-1}(x_j-\bar{x}_{(i)})$$

Where $\bar{\mathbf{x}}_{(i)}$ is the sample mean of (n-1) observations excluding $\mathbf{x}_i$ and i is chosen to maximize

$$L_A(\mathbf{x}|V) - L(\mathbf{x}|V).$$

Hence we are led to declare as the outlier $\mathbf{x}_{(n)}$ that observation $\mathbf{x}_i$ for which $R_i(\bar{\mathbf{x}}, V)=(\mathbf{x}_i-\bar{\mathbf{x}})'V^{-1}(\mathbf{x}_i-\bar{\mathbf{x}})$ is a maximum, so that implicitly the observations have been ordered in terms of reduced form of sub-ordering based on the distance measure $R(\mathbf{x}; \bar{\mathbf{x}}, V)$. Furthermore we will declare $\mathbf{x}_{(n)}$ a *discordant* outlier if

$$R_{(n)}(\bar{\mathbf{x}},V)=(\mathbf{x}_{(n)}-\bar{\mathbf{x}})'V^{-1}(\mathbf{x}_{(n)}-\bar{\mathbf{x}})=\max_{j=1,2,\ldots,n} R_j(\bar{\mathbf{x}}, V) \qquad (3.3)$$

is significantly large .

The null distribution of $R_{(n)}(\bar{\mathbf{x}}, V)$ is not readily determined because the exact form is not very tractable. However, Siotani (1959) reviews the problems associated with calculating percentage points of $R_{(n)}(\mathbf{x}_0,V)$ when $\mathbf{x}_0$ is either 0, μ or $\bar{\mathbf{x}}$. And for the latter case he presents approximate upper 5, 2.5 and 1 percent points of $R_{(n)}(\bar{\mathbf{x}}, V)$ for p=2(1)4 (i.e. p=2, 3, 4) and n=3(1)10(2)20(5)30 (i.e.

n=3, 4, 5,..., 10, 12, 14,...,20, 25, 30). But if μ were known, the corresponding $R_j(\mu,V)$ would be independent χ_p^2 variates. We would then have to relate their maximum $R_{(n)}(\mu,V)$ to the distribution of maximum observation in a random sample of size **n** from a χ_p^2 distribution.

Specifically, in the case of a bivariate sample (p=2), $R_{(n)}(\mu, V)/2$ has the distribution of the maximum of **n** independent exponential variates (mean 1) and its percentage points are easily determined. For a level α-test we would conclude that $\mathbf{x}_{(n)}$ (the observation $\mathbf{x}_i$ yielding $R_{(n)}(\mu, V)$) is a discordant outlier if $R_{(n)}(\mu, V)>\xi_\alpha$ where

$$\alpha=P\{R_{(n)}(\mu, V)>\xi_\alpha\}=1-\{F(\xi_\alpha/2)\}^n$$

with

$$F(x)=1-e^{-x}$$

Thus

$$\xi_\alpha=-2\ln[1-(1-\alpha)^{1/n}] \qquad (3.4)$$

provides an explicit value for use in the test.

The assumption of known V is in general impractical. We therefore continue to examine the two-stage maximum likelihood ratio test for *model* I when both μ and V are unknown.

3.2.2 Model I, V unknown

With V unknown (as well as μ) The maximized log-likelihood under the basic model is (apart from the constant term; Barnett and Lewis, 1994)

$$L(\mathbf{x}) = -n/2 \log|\mathbf{A}|$$

where $|\mathbf{A}|$ is the determinant of the matrix of sums of squares and cross-products of the observations about the component sample means: that is

$$\mathbf{A} = \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})'.$$

Under the *model I* alternative the maximized log-likelihood is

$$L_A(\mathbf{x}) = -n/2 \log|\mathbf{A}^{(i)}|$$

where $\mathbf{A}^{(i)}$ is the restricted matrix obtained on omission of $\mathbf{x}_i$, and i is chosen to maximize

$$L_A(\mathbf{x}) - L(\mathbf{x}).$$

Hence at first view when V is unknown it appears that quite a different principle is advanced for declaration of an outlier $\mathbf{x}_i$ and for the assessment of its discordancy. Here we are implicitly ordering the multivariate observations in terms of an *aggregated* form of reduced sub-ordering based on the values of $|\mathbf{A}^{(j)}|$. The $|\mathbf{A}^{(j)}|$ are ordered, and the observation corresponding to the smallest value of $|\mathbf{A}^{(j)}|$ is declared an outlier. Equally, if we denote

$$\Re_j = \frac{|A^{(j)}|}{|A|}$$

The sample points are 'ordered' in accord with the ordered

$\mathfrak{R}_j$ and the outlier is that observation corresponding with the smallest $\mathfrak{R}_j$, $\mathfrak{R}_{(1)}$. If $\mathfrak{R}_{(1)}$ is significantly low in value the outlier is adjudged discordant. Thus the outlier is that observation whose removal from the sample effects the greatest reduction in the 'internal scatter' of the data set. But the distinction of principle for declaring an outlier in the case of unknown V, compared with the case where V is known, is less profound than might appear at first sight. Clearly $\mathfrak{R}_j$ can be written

$$\mathfrak{R}_j = \frac{|A-(\frac{n}{n-1})(x_j-\bar{x})(x_j-\bar{x})'|}{|A|} = 1-(\frac{n}{n-1})R_j(\bar{x},A) \qquad (3.6)$$

and minimization of $\mathfrak{R}_j$ becomes equivalent to maximization of

$$R_j(\bar{x},\ A) = \frac{R_j(\bar{x},\ S)}{(n-1)} \qquad (3.7)$$

Thus the outlier is again that observation whose 'distance' from the body of the data set measured in terms of a familiar quantity (Mahalanobis distance) is a maximum. It should be remarked that the various formal tests of discordancy for the Multivariate Normal outlier discussed in this part have the desirable property of being invariant with respect to the location and scale of measurement of the observations.

<u>For *model* II</u>, Ferguson (1961) has derived a multi-decision procedure with certain optimal properties for a single possible discordant value in a Normal sample. Assume again $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is a random sample initially imagined to arise from $\mathbf{N}(\mu,\ V)$, with μ and V unknown. Under the

alternative, *model* **II**, hypothesis of a single outlier, we have

$$
\begin{aligned}
&E(\mathbf{x}_j)=\mu && (j=1,2,\ldots,n),\\
&V(\mathbf{x}_i)=bV && (\text{some } i;\ b>1), \qquad (3.8)\\
&V(\mathbf{x}_j)=V && (j\neq i).
\end{aligned}
$$

Denoting by $\mathbf{D}_i$ the decision to regard $\mathbf{x}_i$ as the discordant value $(i=1,2,\ldots,n)$ with $\mathbf{D}_0$ the decision to declare no discordant value, Ferguson considers those decision rules which satisfy four conditions:

i) each is invariant under the addition to $\mathbf{X}_i$ of a constant vector;

ii) each is invariant under the multiplication of $\mathbf{X}_i$ by a common non-singular matrix;

iii) the probability of $P_i(\mathbf{D}_i)$ of declaring $\mathbf{x}_i$ the discordant value when this is true is independent of **i**:

iv) the probability of correctly declaring no discordant value is $1-\alpha$, for a pre-assigned α in (0, 1); that is, the procedure has size α.

He seeks that decision rule which maximizes the value $p_i(\mathbf{D}_i)$. It turns out to have a familiar form: the optimum rule is to reject $\mathbf{x}_i$ if $\mathbf{x}_i$ yields the maximum value $R_{(n)}(\bar{\mathbf{x}}, \mathbf{S})$ and

$$R_{(n)}(\bar{\mathbf{x}}, \mathbf{S})>K$$

Where K is chosen to satisfy the test size condition (iv).

Hence, as for model I, we again declare the outlier to be the observation with maximum generalized distance $R_j(\bar{\mathbf{x}}, \mathbf{S})$, and assess it as discordant if that maximum, $\mathbf{R}_{(n)}(\bar{\mathbf{x}}, \mathbf{S})$, is sufficiently large. Additionally, however, Ferguson demonstrates that this procedure is *uniformly best over all values of b>1.*

Thus when μ and V are unknown it is immaterial whether we adopt the model I or model II formulation of the alternative hypothesis describing the occurrence of a single outlier. In either case the test has the same form, and can be implemented by using some critical value (Barnett and Lewis, 1994).

It should be pointed out that the procedure of omitting several observations with $R_{(n)}(\bar{\mathbf{x}}, \mathbf{S})>k$ at one time, gives similar results as if we were excluding outliers consecutively from the remaining $n'=n-1,n-2,...$ observations with significant $R_{(n')}(\bar{\mathbf{x}}, \mathbf{S})>k'$.

Also, for testing outlying observations in a sample from Multivariate Normal distribution with unknown mean and variance-covariance matrix, Wilks (1963) proposes an intuitively based representation of the sample in terms of sum squares of the volumes of all simplexes that can be found from p of the sample points augmented by the sample mean $\bar{\mathbf{x}}$. Wilks (1962) shows that this is just $(p!)^{-2}|\mathbf{A}|$, where A is the matrix defined above. He calls $\mathbf{A}$ the internal scatter of the sample and suggests that a sensible criterion for the declaration of an outlier is to choose that sample member whose omission leads to the least value for the so-called one-outlier scatter ratio

$$\mathfrak{R}_j = \frac{|A^{(j)}|}{|A|}$$

But this is precisely the likelihood ratio criterion and corresponding test statistic described above. Wilks shows that the $\mathfrak{R}_j$ are identically distributed *Beta* variates $B(\frac{(n-p-1)}{2}, \frac{p}{2})$ with a joint distribution asymmetric over R^n subject to

$$\Sigma\mathfrak{R}_j = n(1-\frac{p}{n-1})$$

$$0 \le \mathfrak{R}_j \le 1 \qquad (j=1,2,\ldots,n).$$

The joint distribution is intractable, but Wilks uses Bonferroni inequalities to obtain an upper bound for the distribution function of $\mathfrak{R}_1$ (which he denotes r_1) and hence lower bounds for the lower percentage points of $\mathfrak{R}_1$ thus enabling conservative tests of significance for a single outlier to be conducted. For the case p=1 the approximation values seems reasonable, in comparison with exact results due to Grubbs (1950). Though it must be stressed that for p>1 it has not been thoroughly assessed since there is at present no yardstick (in terms of exact probabilities) for comparison. However experience elsewhere with Bonferroni inequalities (e.g. for outliers in linear model data) suggests that the approximations should be good (Barnett and Lewis, 1994).

Wilks (1963) tabulates lower bounds to the lower 10, 5, 2.5, and 1 percent of $\mathfrak{R}_1$ for p=1(1)5 and n=5(1)30(5) 100(100)500. For closer comparability the values can be transformed via 3.6 and 3.7 into approximate upper percentage points for $R_{(n)}(\bar{\mathbf{x}}, \mathbf{S})$. Barnett (1994) on page 517 presents critical values for 5 and 1 per cent of $R_{(n)}(\bar{\mathbf{x}}, \mathbf{S})$ for p=2(1)5 and n=5(1)10(2)20(5)50, 100, 200, 500. Extended tables with 20, 10, 5, 2.5, 1, and 0.5 percent critical values for p=2(1)10, 12, 15, 20 and a similar range of sample size appear in Jennings and Young (1988). These were obtained by simulation.

Also Wilks adopts a similar approach to the testing en bloc of 2, 3, or 4 outliers in a multivariate sample, by considering for the s-outlier case (s=2, 3, 4) the s-outlier scatter ratios

$$\mathfrak{R}_{j_1, j_2, \ldots, j_s} = \frac{|A^{(j_1, j_2, \ldots, j_s)}|}{|A|}$$

where $|\mathbf{A}^{(j_1, j_2, \ldots, j_s)}|$ is the internal scatter when x_{j_1}, $x_{j_2}, \ldots, x_{j_s}$ are omitted from the sample. Again it is the subset of observations that minimizes $\mathfrak{R}_{j_1, j_2, \ldots, j_s}$, which is declared the outlying subset and their discordancy must be assessed in terms of how small is

$$r_s = \min \mathfrak{R}_{j_1, j_2, \ldots, j_s}.$$

3.3 Informal methods for multivariate outliers

A host of informal proposals have been made for detecting outliers in multivariate data by quantitative or graphical methods. These cannot be regarded as tests of discordancy; they may be based on derived reduction measures (but with no supporting distribution theory) or, more commonly, they are represented as aids to intuition in picking out multivariate observations which are suspiciously aberrant from the bulk of the sample.

As already noted in multivariate data the concept of extremeness is a nebulous one, unlike the case in univariate samples where outliers can be declared by simple inspection. Various forms of initial processing of the data, involving transformation, study of individual marginal components of the observations, judicious reduction of the multivariate observations to scalar quantities in the form of reduction measures or linear combinations of components, changes in the coordinate bases of the observations, and appropriate methods of graphical representation, can all help to identify or highlight a suspicious observation. If several such procedures are applied simultaneously (or individually) to a set of data they can help to overcome the difficulty caused by the absence of a natural overall ordering of the sample members. An observation which clearly stands out on one, or preferably more, processed re-representations of the sample becomes a firm candidate for identification as an outlier.

An early example of an informal graphical procedure is described by Healy (1968) who proposes plotting the ordered $R_j(\bar{\mathbf{x}}, \mathbf{S})$ against the expected values of order statistics of a sample of size n from χ_p^2. He is principally concerned with bivariate data (p=2). Outliers are detected as observations yielding values $R_{(n)}(\bar{\mathbf{x}}, \mathbf{S})$, $R_{(n-1)}(\bar{\mathbf{x}}, \mathbf{S})$ ... lying above the expected straight line (if $\mathbf{X}$ were Normal and μ, V known). Healy considers $R_j(\bar{\mathbf{x}}, \mathbf{S})$ on intuitive grounds, rather than from any formal (alternative model, or test-construction) principle. This idea has been supported by the formal analysis presented in section 3.2.

3.4 Transformation to Normality

If Normality is not a viable assumption, as in the case of joint distribution of weight and height, what should be done? If we look at the distributions of weight across ages it is skewed, and the skewness is changing with age. An appropriate suggestion is to make non-Normal data more 'Normal looking' by considering transformations of the data so that normal theory analyses can be applied. With multivariate observations (say p dimensional), at the first step a transformation must be selected for each of the variables. Then, if the obtained joint distribution is not Normal with a generalization of the first step (as described below) one can try for Multivariate Normality of observations.

In many instances the choice of a transformation to improve the approximation to Normality is not obvious. For such cases it is convenient to let the data suggest a transformation. A useful family of transformation for this purpose is the family of power transformations. In the univariate case a convenient analytical method is available for choosing a power transformation. Box and Cox (1964) describe a family of power transformations for eliminating skewness

$$x^{(\lambda)} = \begin{cases} \dfrac{x^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \ln(x) & \lambda = 0 \end{cases} \tag{3.9}$$

which is continuous in λ for $x>0$ (this is not restrictive, because a constant can be added to each observation if some of the values are negative). Given the observations $x_1, x_2, \ldots, x_n$, the Box-Cox solution for the choice of an appropriate power λ is the one which maximizes the expression

$$\ell(\lambda) = -\frac{n}{2}\ln\left[\frac{1}{n}\sum_{j=1}^{n}(x_j^{(\lambda)}-\bar{x}^{(\lambda)})^2\right] + (\lambda-1)\sum_{j=1}^{n}\ln x_j \tag{3.10}$$

it should be noticed that $x_j^{(\lambda)}$ is defined in **3.9** and

$$\bar{x}^{(\lambda)} = \frac{1}{n}\sum_{j=1}^{n} x_j^{(\lambda)} = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{x_j^{\lambda}-1}{\lambda}\right)$$

is the arithmetic average of transformed observations. The first term in 3.10 is the logarithm of a normal likelihood function (apart from a constant), after maximising it with respect to the population mean and variance parameters. The calculation of $\ell(\lambda)$ for many values of λ is an easy task for a computer. The procedure described should be used for each of the k measured characteristics (k=1,2,...,p) to make it Normal.

The Normal marginals, however are not sufficient to ensure that the joint distribution is Normal, but in practical applications it may be good enough. If not, we could start with the values $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_p$ obtained above and iterate toward the set of values $\lambda' = [\lambda_1, \lambda_2, \ldots, \lambda_p]$, which collectively maximizes

$$\ell(\lambda_1, \lambda_2, \ldots, \lambda_p) = -\frac{n}{2}\ln|\mathbf{s}(\lambda)| + (\lambda_1 - 1)\sum_{j=1}^{n}\ln x_{1j} + (\lambda_2 - 1)\sum_{j=1}^{n}\ln x_{2j}$$

$$+ \ldots + (\lambda_p - 1)\sum_{j=1}^{n}\ln x_{pj} \qquad (3.11)$$

where $\mathbf{S}(\lambda)$ is the sample covariance matrix computed from

$$x_j^{(\lambda)} = \begin{pmatrix} \dfrac{x_{1j}^{\lambda_1} - 1}{\lambda_1} \\ \dfrac{x_{2j}^{\lambda_2} - 1}{\lambda_2} \\ \vdots \\ \dfrac{x_{pj}^{\lambda_p} - 1}{\lambda_p} \end{pmatrix}^T, \qquad j = 1, 2, \ldots, n.$$

The selection method based on equation 3.11 is equivalent to maximising a multivariate likelihood over μ, Σ, and λ (Johnson, and Wichern, 1988).

3.5 Results

Results are presented in six sections. First, we describe the results of trimming the data of gross outliers. Second, the power transformations required to transform height and weight to marginal Normality are presented. Then, in sections (c) and (d) it is shown that there is little advantage in transforming to Bivariate Normality. Fifth, it is shown that the criteria given in (3.4) for testing outliers is appropriate to our data. Finally, the number of outliers detected and excluded is presented in section (f); also a review on the pattern of missing observations is provided.

a) Trimming the data

At the first stage, as it is perfectly plausible for contamination to occur in one of the marginal variables alone: for example by misrecording or an error of data processing. By intuitive consideration of scatter diagrams of height and weight across ages, some of the obvious gross marginal outliers were removed. Basically, we looked at the scatter diagrams of height and weight of individuals at each year of age from two to eighteen year olds (as age of individuals in our study was recorded in complete years). For instance, the measurements with extreme marginal values like six year old individual A in Figure 3.1 with a height of 182 cm, or individual B whose weight is recorded 91 kg are outliers. These misrecorded observations are due to one or a combination of the errors addressed above. Kempthorne and Mendel (1990) offer some interesting comments on the intuitive stimulus to the declaration of a bivariate outlier. In this example, individual A with height of 182 cm could have a height of 82 cm which during the data entry was recorded by mistake as 182, or just a mark of a pen while filling in the questionnaire (Appendix A) in the left box of three pre-specified boxes for recording height was read as 1. From now on 'data' relates to 'trimmed data'. Table 3.4 presents the percentage of trimmed data across years of ages.

b) Normalising the data

If we look at the joint distribution of height and weight of data (trimmed data) across ages, there are still bivariate outliers which can be seen more easily by their Mahalanobis distances from the centre of the data. Figure 3.2 illustrates the scatter plot of weight and height of two year old children in our data and their distance measures. Now, the crucial question is what cut off value to choose for excluding the outliers? It should be mentioned that we considered the extremeness, and their corresponding test of discordancy in Multivariate Normal data. Hence, at the first step we tried to Normalise the joint distribution of weight and height by transformations.

Since for transformation of multivariate data, initially the marginals, must be transformed, we started by Normal transformation of the distributions of weight and height at each age separately. There is a macro written in GLIM (1992, V.4) which does this. Basically it calculates -2ℓ for a range of different values of λ. The increments in λ can be adjusted as necessary to obtain the required precision. Figure 3.3 presents the graphical and tabular form of the output for this procedure in determination of λ for weight of 2 year old children in our data.

As another example of the result of using these transformations, Figure 3.4 shows the normal plots of weight and height of twelve year old girls, before and after transformations. The Normal plots of the transformed weight and height (right panel in the Figure) are very close to straight lines. After transformation the correlations of transformed weight and height with their corresponding Normal scores increased, and also the Wilks' test of Normality (STATA, 1993) of transformed weight and height were not significant (p-values for test of Normality for transformed weight and height of 12 year old girls were 0.39, and 0.66; but for untransformed data were <0.001 and 0.01 respectively).

Fisher's ML scoring method can be used for derivation of variance of λ. Then one can also compute the confidence interval (CI) easily. Royston (1992) wrote a macro in STATA which calculates the appropriate λ by iteration and also computes the corresponding confidence interval ('support interval'). The availability of different options in his macro makes the work even easier. Although, we initially used Glim 4 for determination of appropriate values of λ later we confirmed the consistency of our findings in GLIM 4 with the outputs of Royston's macro when the corresponding facility on our computer network system became available. Table 3.1 presents the values of λ for marginal transformation of weight and height for different ages, and their 95% CI which were used to Normalise the marginal distributions of height and weight. Also, in order to take account of the difference in the pattern of growth of boys and girls at puberty the data for boys and girls has been checked for outliers separately for eleven year olds and older. Table 3.1 shows the power λ for transformation to Normality from 11 years old for each sex.

c) Transformation to Bivariate Normality

Transformation to Bivariate Normality was attempted through the application of the bivariate form of the Box-Cox power transformation. For this a MINITAB (1991) macro was written which computes $\ell(\lambda_1,\lambda_2)$ for different values of λs in a reasonable neighbourhood, and finds those which maximize the expression 3.11. As an example, the pair of λs for marginal Normal transformations of weight and height at age two year olds was (0.4, 2.00). The pair of λs obtained for Bivariate Normality was (0.4, 2.15) which are very close to the transformation to marginal Normality and well within the regions of support for those values. Also the corresponding deviances (-2ℓ) for finding marginal Normals and the expression 3.11 for finding λ_1, λ_2 for transformation to Bivariate Normality are almost identical. Using the resulting Bivariate Normal distribution of weight and height one can exclude those measurements whose distances measures are greater than the critical value of ξ_α given by (3.4).

d) No advantage in attempting transformation to Bivariate Normality

In order to investigate whether marginal Normality is sufficient for excluding the outliers in our study or not, we compared the outcomes of using both forms of transformations in some of age groups of children. First, for two year old children in our data, we looked at the outliers after transformation to marginal Normality as well as after the transformation to Bivariate Normality. Observation were declared outliers, when $R_{(n)}>\xi_\alpha$. For this age group n=1176 then ξ_α=20.1. Figure 3.5a and Table 3.2 present the results of this comparison. Figure 3.5a shows

Z-scores of the measurements in both forms of transformations. In this Figure the joint Z-scores of measurements after marginal Normal transformations are shown by '+'. The Z-scores of the Bivariate Normal transformed measurements are shown by '×'. The same individual's measurements are joined by a line. In Figure 3.5b the square area shown in Figure 3.5a is enlarged, these points are far from the centre of the data. It can be seen in the Figure 3.5b that after both forms of transformation the points appear identical and that is why the joining lines are not recognizable in either figures. This shows that even for points far from the centre the results of marginal transformations are very similar to bivariate transformation.

In the Figure 3.5a the measurements M_1, M_2,..., M_6 are those who were identified as outliers when using both forms of transformation, Table 3.2. The first observation, B, was the only pair of measurements which was recognised as an outlier when using the transformation to Bivariate Normality and which is borderline when the data are transformed to marginal Normality, and conservatively would not have been declared an outlier. Thus as one can see six out of seven (86%) of the outliers were identified using the results of marginal Normal transformations.

Also as another example the corresponding pairs of λ for marginal and Bivariate Normal transformations of weight and height of three year old children in our data are $\lambda_1(0.15, 2.00)$ and $\lambda_2(0.20, 2.05)$ respectively. Interestingly the same measurements were identified as outliers for both forms of transformation. Similar results were observed when we formed this comparison for some of the other ages. On all occasions the small differences in determination of outliers were related to those measurements whose distance R_j were on the borderline in comparison with their critical values, and they had not been excluded in one or other form of transformation. So the precision of using marginal Normals seems adequate as might be expected since optimal values of λ for Bivariate Normality are all well within the interval of support shown in Table 3.1.

Therefore, on the basis of this finding, we excluded the outlying observations in the joint distributions of weight and height of individuals across ages using transformation to marginal Normality of weight and height. The values of λ chosen for this purpose are shown in Table 3.1. As a way of confirmation of appropriateness of results, for example, Figure 3.6 presents the gamma plot of squared generalized distances R_j, after excluding the bivariate outliers in transformed data for twelve year old girls. The close to straight line of gamma plot suggests that in the remaining observations there is no evidence of any remaining outlying observations.

e) Critical values in test of discordancy

It was mentioned that the critical values of the test for outliers in our data were computed from formula 3.4. The reasons are: first, Barnett (1994) only computed the critical values for p=2(1)5 and n=5(1)10(2)20(5)50, 100, 200, 500. Second, because almost all samples in our data in different ages include about 1000 children or more. So the $\bar{x}$ and S estimated from these transformed samples in different ages are reasonable estimates of corresponding μ, and V, and can be used to calculate the critical values of $R_{(n)}(\mu, V)$ in Bivariate Normal data, as given by 3.4.

As a check on this assumption Table 3.3 compares the results of applying the formula 3.4 in calculating critical values for p=2 at test levels 5%, 1%. As it can be seen, with increasing sample size, these values become closer to those computed by Barnett (1994), even for sample sizes less than 500. So for sample sizes with which we are concerned the assumption that μ and V are known is acceptable. It is noted that unlike the univaritae case, in which for smaller n when the S is unknown the t value is bigger than z for a level of α, in comparison of the critical values it is observed that the values computed from formula 3.4 (μ and V known) are larger than those Barnett (1994) calculated for the cases when μ and V are unknown.

f) Descriptive presentation of discordants

Table 3.4 shows the results of two stages of exclusion of outlying measurements in our data. It can be seen from Table 3.4 that for the first two age groups (2 and 3 years old) the percentage of outliers is relatively high (1.7% and 2% respectively) and drops to a minimum between the ages of 8 to 10 years old where the percentages of observations excluded by trimming and multivariate analysis are fairly similar. Then from 11 the percentage of exclusion increases and is maximum at age 14 years (2.3%), and is roughly the same up to 17 years. Finally, at 18 years the exclusion rate drops to 1.4%. However, the number of children in the sample at the older ages were less than average. It should be noted that the proceeding analysis could only be applied when both measurements on height and weight were present.

It is worthwhile mentioning that the pattern of missing values in the observations was studied in total and within age groups (Table 3.5). It was found that there was an association between the number of missing values in each category (both measurements available, both measurements missing, only weight available, only height available) and age (χ^2_{48}=1142.3; p=0<0.001). It can be seen from the Table 3.5 that the proportion of missing values in each category is similar at each age up to about 12 years. However beyond age 12, the number with both missing rises substantially. Also, in a log-linear model analysis the effect of sex on this association was found significant ($p=0.02$) showing that the proportion of missing at older age was higher in boys.

Furthermore, since the data on some of the laboratory results of individuals in the study were available, it was noted that among children who had any kind of height and/or weight measures missing, 60% still had blood taken for testing showing that generally the co-operation with medical examinations was higher than with anthropometry. The value of monitoring anthropometric measurements was discussed in chapter two. Therefore the observed situation on the pattern of missing data, for example, having more missings in some of ages such as younger children (2 and 3 years) or older boys (13 years and older) in this survey, and the importance of the anthropometric measurements, are points that should be mentioned to the staff of survey teams in any future studies in Iran.

3.6 Discussion

3.6.1 Comparison with other methods

The method used for identification of outliers, which is based on Mahalanobis Distance (MD), is among the powerful methods of the detection of the outliers in multivariate data. Wilks method leads to similar criterion in a test of discordancy, and both methods declare observations with too large $R_{(n)}(\bar{\mathbf{x}}, \mathbf{S})$ to be discordant.

Other methods in this area have used similar ideas but use alternative statistics in estimation of location and dispersion in computation of distance measures. For instance, using median instead of mean or using different definitions in estimation of variation in the data, may give results which are much less influenced by the outlying observations.

In detection of outliers in multivariate data, the most important problems are the problems of masking and swamping. For example, in some cases a small cluster of outliers will attract $\bar{\mathbf{x}}$ and inflate $\mathbf{S}$ in its direction, yielding small values for MD. This problem is known as masking because the presence of one outlier masks the appearance of another outlier. Second, a small cluster of outliers will attract $\bar{\mathbf{x}}$ and inflate $\mathbf{S}$ in its direction and away from some other observations which belong to the pattern suggested by the majority of observations, thus yielding large MD values for these observations. This problem known as swamping. These difficulties arise because $\bar{\mathbf{x}}$ and $\mathbf{S}$ are not robust to large outliers. Some qualitative effects of outliers on estimation of V, and corresponding attitudes to robust estimation, are considered by Devlin et al. (1981), and by Campbell (1980) who claims that outliers tend to deflate correlations and possibly inflate variance. One way to avoid such difficulties is to use more robust estimators of location and covariance. Several such estimators have been suggested for example, Donoho (1982), Hampel et al. (1986), and Rousseeuw and van Zomeren (1990).

Campbell (1980) discusses robust procedures in multivariate analysis and robust covariance estimation. He suggests a robust weighted **MD** for identification of outliers. A similar method to his approach may use different weights. Rousseeuw (1985) uses the minimum volume ellipsoid (MVE) that covers at least half of the observations to construct robust estimators. The centre and covariance matrix of the observation included in the MVE are robust location and covariance matrix estimators. Lopuhaä and Rousseeuw (1991) have shown that MVE estimators are less affected by contamination than Campbell's (1980) M-estimators.

Hadi (1992) proposed a method for identification of multiple outliers in multivariate data and later modified it (1994). He claims that his method is effective in dealing with masking and swamping problems. His proposed procedure in detection of outliers for n observations on p variates can be summarized in the following steps. First, the n observations are ordered using a suitably chosen robust measure of outlyingness, then divide the data set into two initial subsets: a 'basic' subset which contains p+1 'good' observations and a 'non-basic' subset which contains the remaining n-p-1 observations. Second, he computes the relative distance from each point in the data set to the centre of the basic subset, relative to the covariance matrix of the basic subset. Third, he rearranges the n observations in ascending order of revised distance then divides the data set into two subsets: a basic subset which contains the first p+2 observations and a non-basic subset which contains the remaining n-p-2 observations. He repeats this process until an appropriately chosen stopping criterion is met. The final non-basic subset of observations is declared an outlying subset. Modification may be made to Hadi's method by using the median instead of mean or using iteration until convergence within the third step.

Penny (1995) in an interesting review of some of the methods of detection of outliers in multivariate data considers the strength of different methods in a Monte Carlo simulation study. Penny compares the precision of different methods in identification of planted outliers in different forms of slippage (slippage from the mean, slippage from the variance). She looks at the combination of different percentages of outliers in different dimensional data, using Multivariate Normal and bootstrapped data. Penny concludes that some methods do better than others depending on:

1. Whether the data is Multivariate Normal
2. The dimension of the data set
3. The type of outlier
4. The proportion of outliers
5. the degree of outlyingness.

Penny finds that some of the methods are more promising, such as Hadi's method and **MD** for higher dimensions in data from a symmetric Normal distribution.

In 1993, Gould and Hadi wrote a macro which is now a routine in STATA which can be used for identification of multiple outliers in multivariate data. Hadi's (1994) simulation study shows, for large sample sizes, the results of the routine macro and his modification are almost identical. For comparison of the methodology we repeated our analysis using Hadi's procedure in STATA. Since both methods use measures of distance of observations from the population mean, the same test criteria, ξ_α, given by 3.4 is suitable for both methods. It is observed that the results are almost consistent across all of the age groups, and the problems of masking and swamping which Hadi is very concerned about have not been troublesome in our work. This is because the proportion of outliers is small and extreme values have been trimmed from the data. So our distribution parameters were estimated on the basis of large samples with a nearly Normal data.

In order to examine the reason for some small differences in some age groups, we review our routine example about the two year old children. Table 3.6 presents the results of both methods, as one can see both methods result in approximately the same observations as outliers. It seems that Hadi's procedure identifies two more outliers than the one we used. But from the columns of distance measures, we see that for these two sets of observations $R_1=R_3=20.1$ (when rounded). So these two measurements are borderline in comparison with the critical value $\xi_{0.05}=20.1$, and in order to save information we had previously decided not to exclude them. Otherwise both methods excludes all the same data points. Overall in our data we found that if there is a difference between results, it is due to the inclusion or exclusion of borderline cases.

When excluding the outliers, we were comparing not only distance measures with the critical values but looked for a gap in comparison with other observations with $R_{(n)}$ values near the critical value. Also we took into consideration whether the observations seemed surprising in relation to existing reference values for height and weight (NCHS, 1977). It should also be noted that there is no guarantee that even the best choice of λ_i will produce a transformed set of values that adequately conform to the Normal distribution. If the data had not been rounded the transformations might have worked even better because the rounding of the measurements to the nearest integer has created irregularities in the distributions.

However Hadi's method is a powerful method and practical with the availability of the corresponding macro which is simple to use (given that you are a STATA or S-plus user). But the most important drawback of this method is that, the technique does assume the underlying variables are Multivariate Normal or at least elliptically symmetric (STATA Reference Manual 1993). So if the data is not reasonably Normally distributed, the reviewed techniques for transformation or some kind of short cut like log transformation are technically crucial. As an interesting example about this point we considered the first example of Hadi (1992) about finding outliers in data on brain weight and body weight of 28 species which he took from Rousseeuw and Leroy (1987; p.57). Hadi worked with log(brain weight) and log(body weight) which are reasonable transformations and identified 4 observations as outlying. But if one uses the original measurements which do not satisfy the assumptions of his method, his macro declares 13 out of 28 observations (nearly 50%) as outliers. Similarly we observed that in our study whenever we analyzed untransformed data with his method a large number of observations were identified as outliers.

So, unless the data are Normal or transformed to Normal, one may incorrectly exclude a lot of data. In most situations this is not only loss of information but may also waste a lot of money (e.g. on laboratory experiments). Another drawback is that the macro automatically excludes all points whose distances are greater than a certain critical value, but most of the time the points with borderline distances really need careful consideration as to whether they should or should not be excluded.

3.6.2 How did outlying observations arise?

As may be expected in a large national survey of this nature, a variety of practical problems can cause errors to arise in the data despite the care taken in the planning and training of the survey teams. However, checking procedures are available to reduce errors but limitations of time and financial constraints restrict the amount of error correction that could be done (especially in a study of this magnitude).

Amongst the various sources of error in the data, the ones most commonly responsible for producing 'surprising' measurements of weight and height in our data are listed below:

i- error in the measurement of weight and height of individuals

ii- error in age recording (even though the survey protocol specified that age was to be checked against the subjects identity card)

iii- error in completing the questionnaire

iv- error in data entry.

There was no previous experience in our country of using a computer and data base of this magnitude. In addition, the level of experience and familiarity of the staff with data management was such that a sophisticated computer program for data checking could not be employed. But despite these difficulties the rate of error was very low and the data appears to be of high quality.

3.6.3 Checking the original records

The data set was made available to me by my department (Biostatistics and Epidemiology) in the Tehran University of Medical Sciences who were one of the most important groups responsible for the design and conduct of the survey. I received the data in England August 1993, and most of the data validation and management was completed by December 1993. Due to difficulties in contacting the appropriate people and the lengthy official procedure involved, getting access to the original records for checking the questionnaires during the data validation stage was impossible from London. So, I located and deleted the outliers in my data as described. However, I was anxious to consult the original questionnaires to see how the method worked.

In June 1994, when I went back to my country for a short visit, I managed to check some of the original questionnaires personally. A full account of this experience and what I saw as practical problems in a large study, is beyond this short note. It could be organised as instructions for staff of a survey team, and could be taught during preparation classes for similar surveys. The problem should also be discussed with the people who are responsible for data base management. At each stage data can be screened for errors and relative precision will increase. While in Tehran I discussed the problems that were identified with the people responsible for the data, and I pointed out the possible ways of avoiding errors in the future. A few of these will be discussed after describing the review of the original records. It should be noted at the outset that data cleaning had been carried out without first hand observation of the original records, so my study of the records only relates to errors that had been analytically detected. There are doubtless many minor errors which did not result in observations being identified as outliers.

Here some of my observations are briefly described. As a first example, from consideration of data on height and weight of two year old children, I excluded observations (2 kg, 591 cm) and (95 kg, 74 cm) as outliers. When I looked at these records, the first measurement was (25 kg, 91 cm); data entry was in free format style; this error resulted from a space being typed in the wrong place. The second child's weight was 9.5 kg and should have been rounded to 10 kg, but it was misrecorded as 95 kg. It should be noted that during data cleaning I avoided any personal judgement about different possibilities of errors and treating the data myself since none of the questionnaires were available.

Further examples at this age: the measurements (25 kg, 61 cm) were identified as an outlier (Table 3.2); when I considered his record, I realised that the real measurements were 25 kg, and 91 cm. The age of a child with measurements (19 kg, 110 cm) who was identified outlier, in fact was 3 years, and was misrecorded as 2 years. When examining data of three year old children, the observation (30 kg, 91 cm) was identified as an outlier; when his record was checked the real measurements were 10 kg, and 93 cm.

It is very difficult to explain some of the errors clearly because the questionnaires were completed in Farsi, and the characters are very different. So explanations of the similarity in Farsi which were the cause of error, when translated to English, look strange and unlikely to be a cause of confusion. So here I have tried to explain some which are not dependent on language or are simple to understand despite differences in languages. Also, some of the questionnaires were filled in using English numbers. This could cause confusion as well, for instance, the height of a 3 year old child with measurements (12 kg, 90 cm) was recorded (12 kg, 40 cm), because a badly written 9 was read as a 4. Similarly, 7 can be easily read as 1 and even 2.

Some of the errors were caused by the similarity of different written numbers in Farsi. For example, the measurements (14 kg, 154 cm) and (18 kg, 154 cm) of five year old children were identified as outliers. The reason was that a badly written 0 in Farsi was misread as 5. In fact the measurements were (14 kg, 104 cm) and (18 kg, 104 cm) respectively. Another example is the similarity of writing of 2 and 3 in Farsi which can be misrecorded easily. For instance, the measurements of a five year old child with weight and height of 26 kg, and 112 cm was misread and misrecorded as 36 kg, and 112 cm. On the whole, after examining many records in different age groups, it was found that all of the measurements which our multivariate analysis identified outliers, were really errors caused by one or a combination of errors listed above. Thus the criteria employed is highly specific in that no records were incorrectly labelled outliers. The number of erroneous records that have not been detected, and hence the sensitivity of the screening procedure could only be determined by a repeat survey of a sample of families included in the survey.

There are a number of issues which need to be considered in reducing the errors described here and other types of error. First, data can be entered using data base software which is either a dedicated data entry system or has a programming language attached to it. Entered values can then be checked against other related information. For example, not only the maximum and minimum values of weight and height could be fixed using a range command also their values can be checked against another related factor(s) in one of previous question(s) (e.g. age, or sex and age). Therefore, one can see that by employing a sophisticated program the chance of observed errors might have been reduced.

Second, with procedures like double entry the systematic error and or random error of the data entry can be examined. Also, it is worth mentioning that this is possible with some current software such as EPI INFO. With software of this kind most of the explained aims in reducing the errors might be achieved. Third, it is very important to be consistent all through the work, for instance, using both Farsi and English characters in different places may increase the errors. So, whilst stressing the use of using Farsi script and ensuring accuracy in filling in the questionnaires the chance of misreading and misentering will decrease.

In addition, it is very important to have a clear form layout for the questions. For instance, with a pre-specified place for numbers which was done in the design of the questionnaire of our survey (Appendix A). Also, if a database system such as FOXPRO is used it is possible to design the computer to be screen similar to the forms. A screen having the same appearance as the forms should reduce confusion. Also, regarding the accuracy of the measurements, the equipment should be examined regularly during the survey. For example, the weighing scale or the tools for height measuring (portable stadiometer) in the case of our work regarding the growth of children. Furthermore, it is very important to check the accuracy of the response to the questions with possible simple aids; a good example regarding our work can be the case of age which was supposed to be checked against the identity cards. An error in age recall may result in having some unusual measurements for a specific age group. Also, in surveys it is very important to manage the time plan of daily work appropriately because if there is too much to do the tiredness of the survey staff will affect both the precision of measuring and the completion of the questionnaires.

In summary, the occurrence of errors in a study of this magnitude is unavoidable. The rate of errors detected in weight and height measurements of children in our study is low, and did not exceed 2.3%. All outliers detected by the analysis were found to have resulted from errors in data processing. However, since most data precessing errors are likely to generate gross errors or at least outliers the small proportion of data that had to be discarded suggests that data processing errors were comparatively rare.

Table 3.1 Estimated λ (Box-Cox power transformation to Normality) and corresponding support intervals of weight and height by age and sex for ages 11 and older

Age(years)	Weight (both sexes)	Height (both sexes)
2	0.4 (0.2 , 0.6)	2.0 (1.6, 2.4)
3	0.15 (0.0 , 0.3)	2.0 (1.6, 2.2)
4	0.5 (0.3 , 0.6)	2.45(2.1, 2.8)
5	0.5 (0.3 , 0.6)	1.9 (1.6, 2.2)
6	0.2 (0.0 , 0.3)	1.4 (1.1, 1.7)
7	0.1 (-0.1, 0.3)	2.0 (1.6, 2.4)
8	0.0* (-0.2, 0.2)	1.0 (0.4, 1.4)
9	0.1 (-0.1, 0.3)	2.0 (1.3, 2.6)
10	-0.3 (-0.5,-0.1)	1.3 (0.7, 1.9)

Age(years)	Boys	Girls	Boys	Girls
11	0.0 (-0.2, 0.4)	-0.2 (-0.6,-0.1)	1.5 (0.4, 2.2)	1.0 (0.0, 1.7)
12	-0.85 (-1.2,-0.5)	-0.25(-0.5, 0.0)	-0.4 (-1.2, 0.5)	2.5 (1.5, 3.4)
13	-0.2 (-0.5, 0.1)	0.0 (-0.3, 0.3)	0.4 (-0.5, 1.4)	2.8 (1.7, 3.8)
14	-0.15 (-0.5, 0.2)	0.45(0.1, 0.8)	-0.3 (-1.3, 1.0)	2.7 (1.4, 4.0)
15	0.15 (-0.2, 0.5)	0.3 (-0.1, 0.6)	2.1 (0.9, 3.4)	2.0 (0.9, 3.6)
16	0.5 (0.1, 1.0)	0.3 (-0.1, 0.7)	3.5 (2.2, 5.0)	0.7 (-0.6, 2.0)
17	0.5 (0.0, 0.9)	-0.3 (-0.7, 0.1)	4.0 (2.9, 6.0)	1.0 (-0.8, 2.4)
18	-0.3 (-0.9, 0.3)	-0.2 (-0.6, 0.2)	2.0 (0.4, 3.9)	0.7 (-1.2, 2.6)

* if λ=0 then ln(x) was the transformation

Table 3.2 Identified outliers after two forms of transformation of weight and height of 1176 two year old children

Subject Label	Measurements	$R_j(\bar{x}, S)$ (T_1)	$R_j(\bar{x}, S)$ (T_2)
B	24 kg , 87 cm	20.1	20.2
M_1	15 kg , 54 cm	21.6	20.8
M_2	19 kg , 110 cm	20.6	21.3
M_3	11 kg , 110 cm	20.6	21.4
M_4	10 kg , 111 cm	24.2	25.3
M_5	12 kg , 115 cm	29.4	31.0
M_6	25 kg , 61 cm	43.7	43.2

Critical value $\xi_{0.05}=20.1$

T_1: Marginal normal transformations $\lambda(0.4, 2.00)$

T_2: Bivariate normal transformation $\lambda(0.4, 2.15)$

Table 3.3 Comparison of critical values tabulated by Barnett[†] with values derived from 3.4 for 5% and 1% tests of discordancy of a single outlier in a Bivariate Normal sample for the distance measure $R_{(n)}(\bar{x}, \mathbf{S})$[‡] when parameters are unknown and sample size is large

n	p=2			
	Barnett[2]		Formula 3.4	
	5%	1%	5%	1%
100	14.22	16.95	15.15	18.41
200	15.99	18.94	16.54	19.80
500	18.12	21.22	18.37	21.63

n = number of observation ; p = dimension

† - Barnett (1994, p.517)

‡ - $R_{(n)}(\bar{x}, \mathbf{S}) = \max_{j=1,2,\ldots,n} R_{(j)}(\bar{x}, \mathbf{S})$

Table 3.4 Number of data, gross outliers (trimmed observations), and identified bivariate outliers of measurements of height and weight; National Health Survey 1990-2, Iran

Age (years)	No. with both measurements	Exclusion				No. Used in analysis
		Trimmed data No.	%	Bivariate outliers No.	%Trimmed*	
2	1190	14	1.2	6	0.5	1170
3	1580	14	1.5	8	0.5	1558
4	1680	9	0.5	5	0.3	1666
5	1708	9	0.5	12	0.7	1687
6	1561	7	0.4	16	1.0	1538
7	1703	6	0.4	12	0.7	1685
8	1618	7	0.4	8	0.5	1603
9	1486	8	0.5	2	0.1	1476
10	1517	6	0.4	8	0.5	1503
11	1447	13	0.9	5	0.3	1429
12	1330	19	1.4	5	0.4	1306
13	1228	21	1.7	3	0.2	1306
14	1084	23	2.1	2	0.2	1204
15	1042	19	1.8	0	0.0	1023
16	951	17	1.8	3	0.3	931
17	841	9	1.1	1	1.1	831
18	690	9	1.3	1	0.1	680
Total	22656	220	1.0	97	0.4	22349

*: Percentage of outliers of trimmed data

Table 3.5 Number of children surveyed, children with both weight (Wt) and height (Ht) measurements, missing (did not attend measuring), and children with only Wt or Ht measurements; National Health Survey 1990-2, Iran

Age	Total No.	No. with both measurements						Did not attend in Wt & Ht measuring						Attended in Wt measuring						Attended in Ht measuring					
(years)	surveyed	Boys	%	Girls	%	Total	%	Boys	%	Girls	%	Total	%	Boys	%	Girls	%	Total	%	Boys	%	Girls	%	Total	%
2	1301	620	92.1	570	90.7	1190	91.4	31	4.6	34	5.4	65	5.0	4	0.6	6	1.0	10	0.8	18	2.7	18	2.9	36	2.8
3	1678	820	93.9	760	94.4	1580	94.2	33	3.8	28	3.5	61	3.6	5	0.6	8	1.0	13	0.8	15	1.7	9	1.1	24	1.4
4	1784	858	93.7	822	94.7	1680	94.2	32	3.5	31	3.6	63	3.5	4	0.4	3	0.3	7	0.4	22	2.4	12	1.4	34	1.9
5	1814	872	93.4	836	95.0	1708	94.2	35	3.7	20	2.3	55	3.0	7	0.8	8	0.9	15	0.8	20	2.1	16	1.8	36	2.0
6	1674	840	93.1	721	93.4	1561	93.2	37	4.1	32	4.1	69	4.1	6	0.7	5	0.7	11	0.7	19	2.1	14	1.8	33	2.0
7	1849	888	91.5	815	92.7	1703	92.2	58	6.0	37	4.2	95	5.1	4	0.4	8	0.9	12	0.6	20	2.1	19	2.2	39	2.1
8	1761	820	91.8	798	91.9	1618	91.9	54	6.0	42	4.8	96	5.4	6	0.7	11	1.3	17	1.0	13	1.5	17	2.0	30	1.7
9	1615	719	91.9	767	92.1	1486	92.0	36	4.6	46	5.5	82	5.1	9	1.2	7	0.8	16	1.0	18	2.3	13	1.6	31	1.9
10	1652	755	91.0	762	92.7	1517	91.8	56	6.7	40	4.9	96	5.8	5	0.6	9	1.1	14	0.9	14	1.7	11	1.3	25	1.5
11	1596	741	90.6	706	90.8	1447	90.7	53	6.4	43	5.5	96	6.0	12	1.5	5	0.6	17	1.1	12	1.5	24	3.1	36	2.2
12	1500	685	86.9	645	90.6	1330	88.7	73	9.3	37	5.2	110	7.3	13	1.6	11	1.5	24	1.6	17	2.2	19	2.7	36	2.2
13	1390	603	86.4	625	90.3	1228	88.3	75	10.7	51	7.4	126	9.1	6	0.9	4	0.6	10	0.7	14	2.0	12	1.7	26	1.9
14	1229	544	87.0	540	89.4	1084	88.2	68	10.9	50	8.3	118	9.6	3	0.5	5	0.8	8	0.7	10	1.6	9	1.5	19	1.5
15	1231	472	80.1	570	88.9	1042	84.7	105	17.9	58	9.0	163	13.2	2	0.3	8	1.2	10	0.8	10	1.7	6	0.9	16	1.3
16	1147	424	77.8	527	87.5	951	82.9	104	19.1	59	9.8	163	14.2	7	1.3	7	1.2	14	1.2	10	1.8	9	1.5	19	1.7
17	1068	364	73.1	477	83.7	841	78.7	109	21.9	83	14.6	192	18.0	7	1.4	3	0.5	10	1.0	18	3.6	7	1.2	25	2.3
18	1076	285	57.9	405	69.3	690	64.1	193	39.2	163	27.9	356	33.1	9	1.9	8	1.4	17	1.6	5	1.0	8	1.4	13	1.2
Total	25365	11310	88.2	11364	90.5	22656	89.3	1152	9.0	854	6.8	2006	7.9	109	0.8	116	0.9	225	0.9	255	2.0	223	1.8	478	1.9

Table 3.6 Comparison of distance statistic $R_{(n)}$ and Hadi's D^2 for 8 outliers identified by D^2 in 1176 measurements of weight and height of two years old children; outliers arranged in order of D^2

No	$R_j(\bar{x}, S)(M_1)$	Observations	$D^2_i(C_B, C_{np} S_B)(M_2)$
1	20.1*	24 kg , 87 cm	21.4
2	20.6	19 kg , 110 cm	21.9
3	20.1*	17 kg , 61 cm	22.8
4	20.6	11 kg , 110 cm	23.2
5	21.6	15 kg , 54 cm	24.6
6	20.6	10 kg , 111 cm	27.4
7	29.4	12 kg , 115 cm	33.0
8	43.7	25 kg , 61 cm	48.7

* Critical value $\xi_{0.05}=20.1$

M_1 marginal normal transformations $\lambda(0.4, 2.0)$

M_2 Hadi's Methods

Figure 3.1 Scatter plot of height against weight of 6 year old children

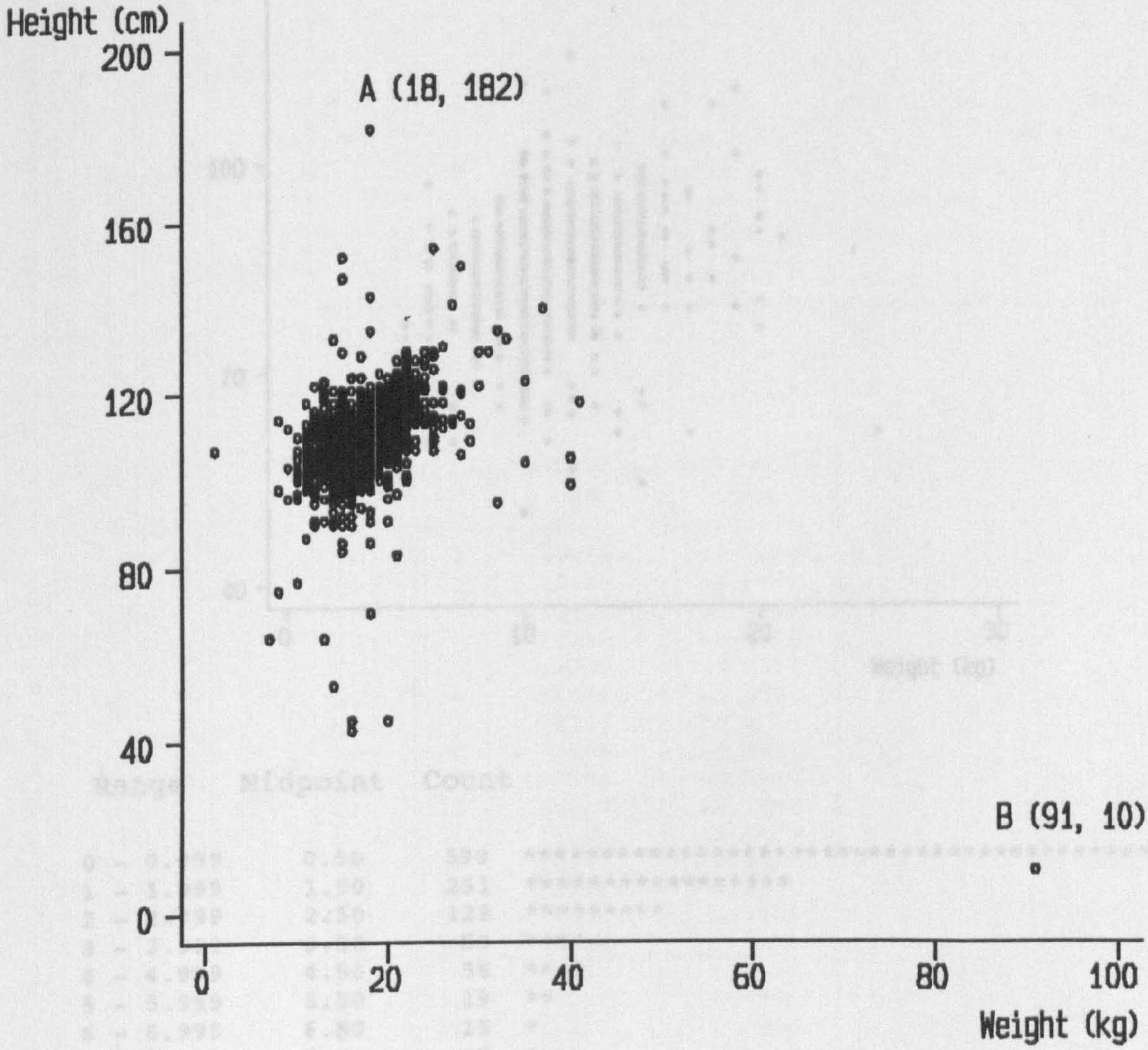
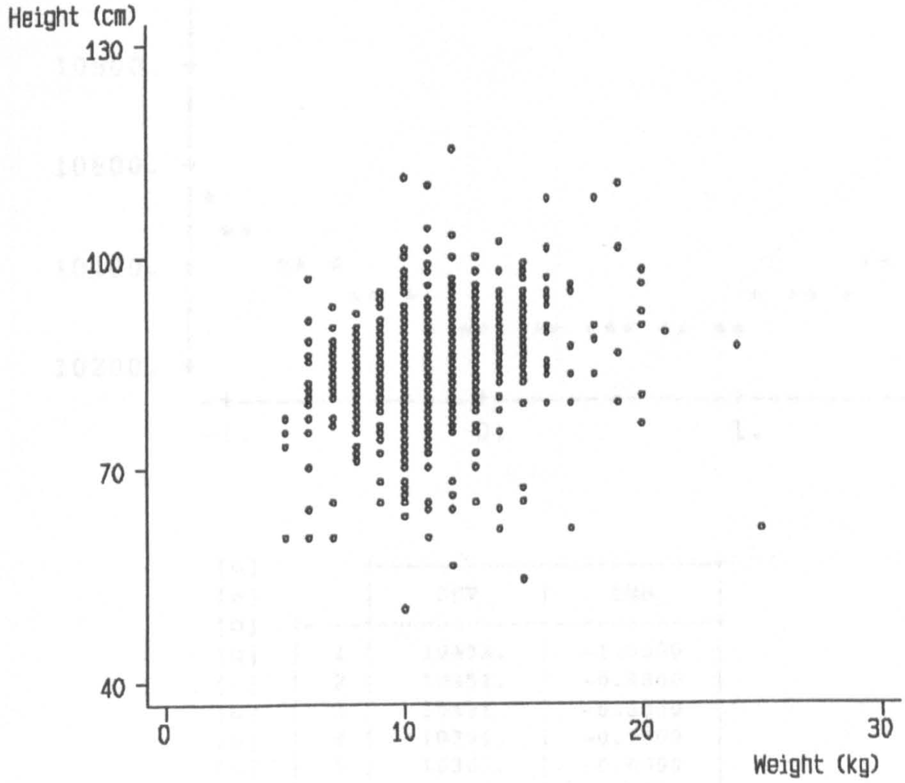


Figure 3.2

a) Scatter plot of height against weight of 2 year old children "trimmed data"

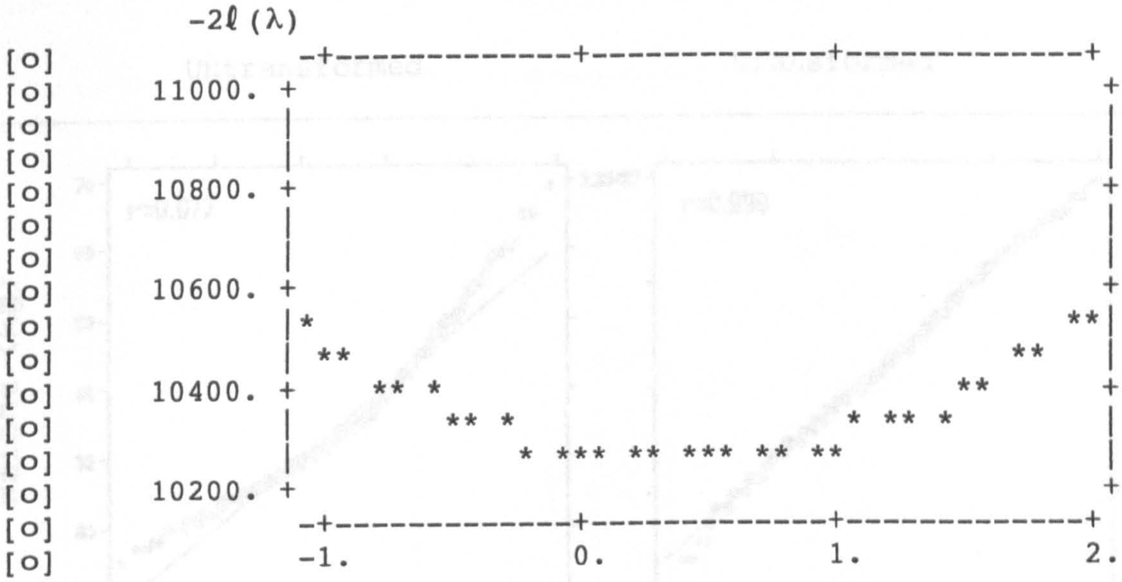


Range	Midpoint	Count	
0 - 0.999	0.50	599	**
1 - 1.999	1.50	251	*****************
2 - 2.999	2.50	122	*********
3 - 3.999	3.50	53	****
4 - 4.999	4.50	56	****
5 - 5.999	5.50	19	**
6 - 6.999	6.50	15	*
7 - 7.999	7.50	13	*
8 - 8.999	8.50	7	*
9 - 9.999	9.50	5	*
10-11.999	11.00	4	*
12-13.999	13.00	11	*
14-15.999	15.00	8	*
16-17.999	17.00	3	*
18-19.999	19.00	4	*
20-24.999	23.00	2	*
25-29.999	28.00	2	*
30-39.999	35.00	1	*
40-65.000	57.00	1	*

Each * represents up to 15 obs.

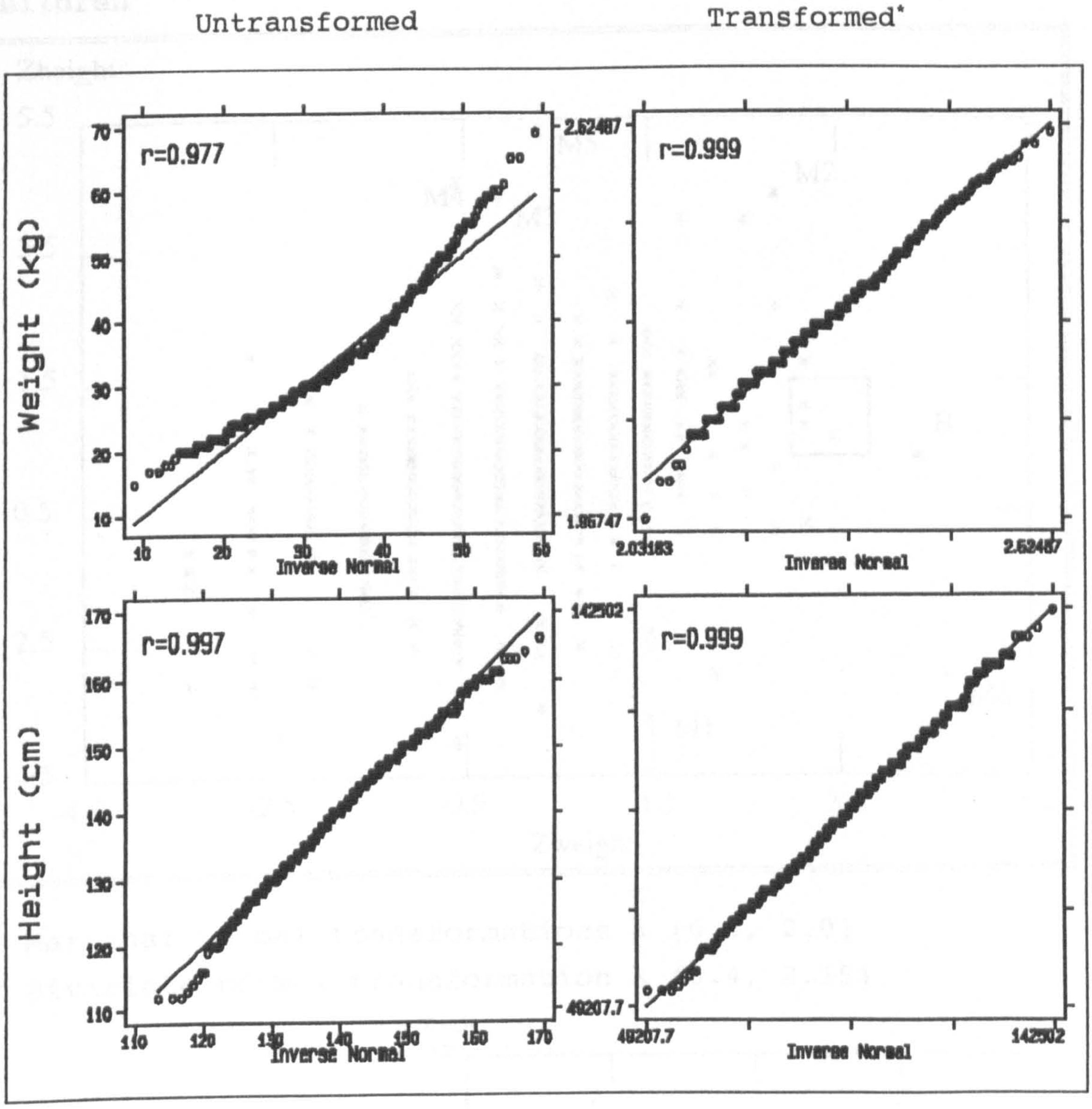
b) Mahalanobis distances from the centre of the measurements of weight and height of 2 year old children "trimmed data"

Figure 3.3 Box-Cox power transformation of weight measurements of 2 year old children, GLIM 4, 1992



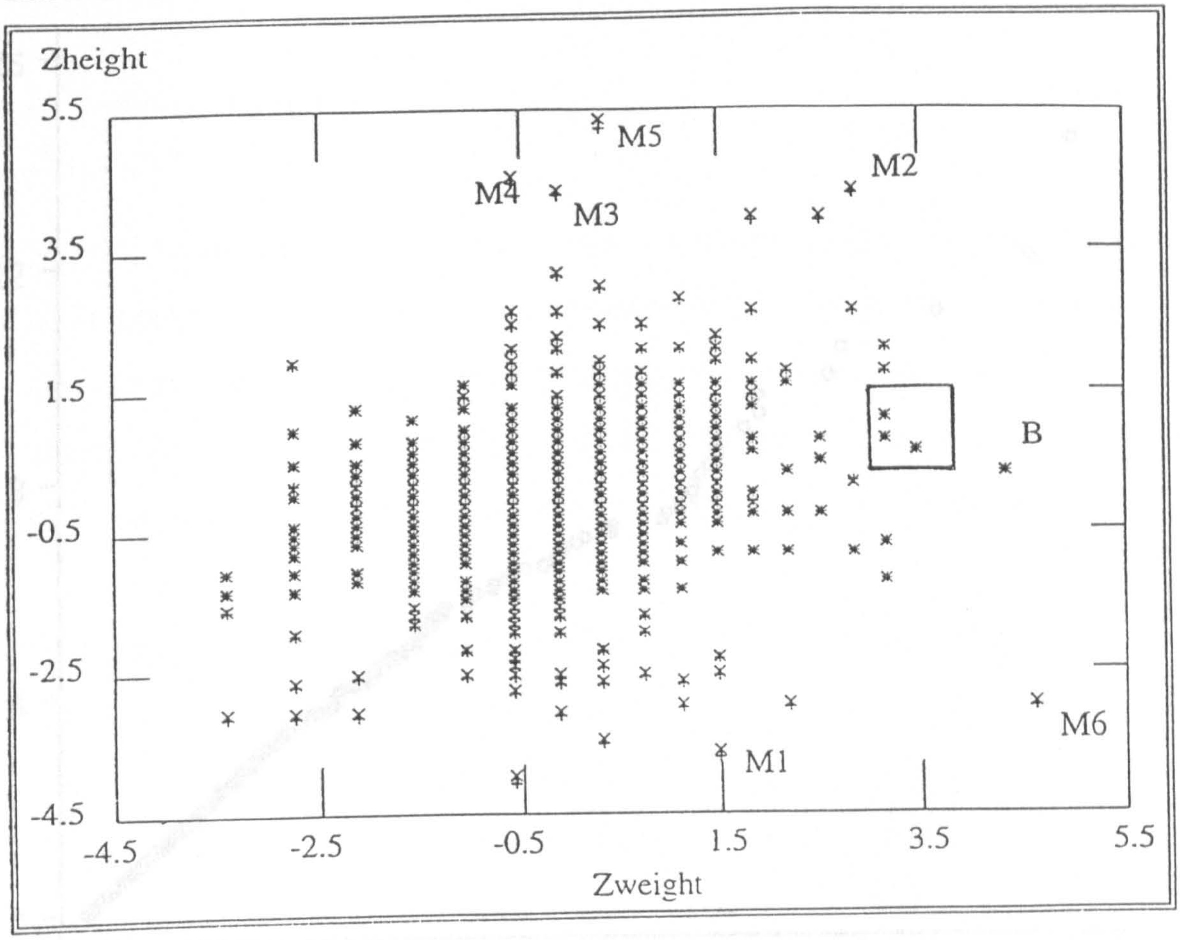
	DEV_	LMB_
1	10483.	-1.0000
2	10451.	-0.9000
3	10421.	-0.8000
4	10394.	-0.7000
5	10369.	-0.6000
6	10346.	-0.5000
7	10326.	-0.4000
8	10309.	-0.3000
9	10293.	-0.2000
10	10280.	-0.1000
11	10269.	0.0000
12	10261.	0.1000
13	10255.	0.2000
14	10251.	0.3000
15	**10250.**	**0.4000**
16	10251.	0.5000
17	10254.	0.6000
18	10260.	0.7000
19	10267.	0.8000
20	10277.	0.9000
21	10290.	1.0000
22	10304.	1.1000
23	10321.	1.2000
24	10341.	1.3000
25	10362.	1.4000
26	10386.	1.5000
27	10412.	1.6000
28	10441.	1.7000
29	10472.	1.8000
30	10505.	1.9000
31	10541.	2.0000

Figure 3.4 Normal plots of weight and height of 12 year old girls before and after transformation



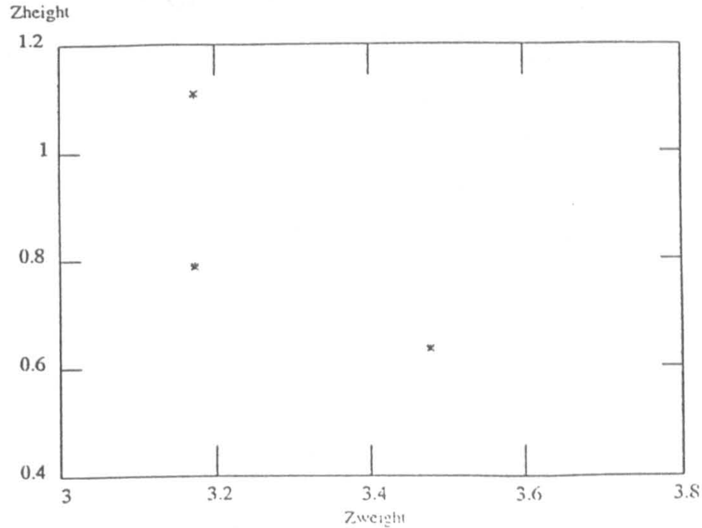
* *Wilks Normality test* { *transformed weight* $p=0.39$ / *transformed height* $p=0.66$

Figure 3.5.a Scatter plot of Z-scores of measurements of weight and height after both forms of transformation (marginal Normals and Bivariate Normal); 2 year old children



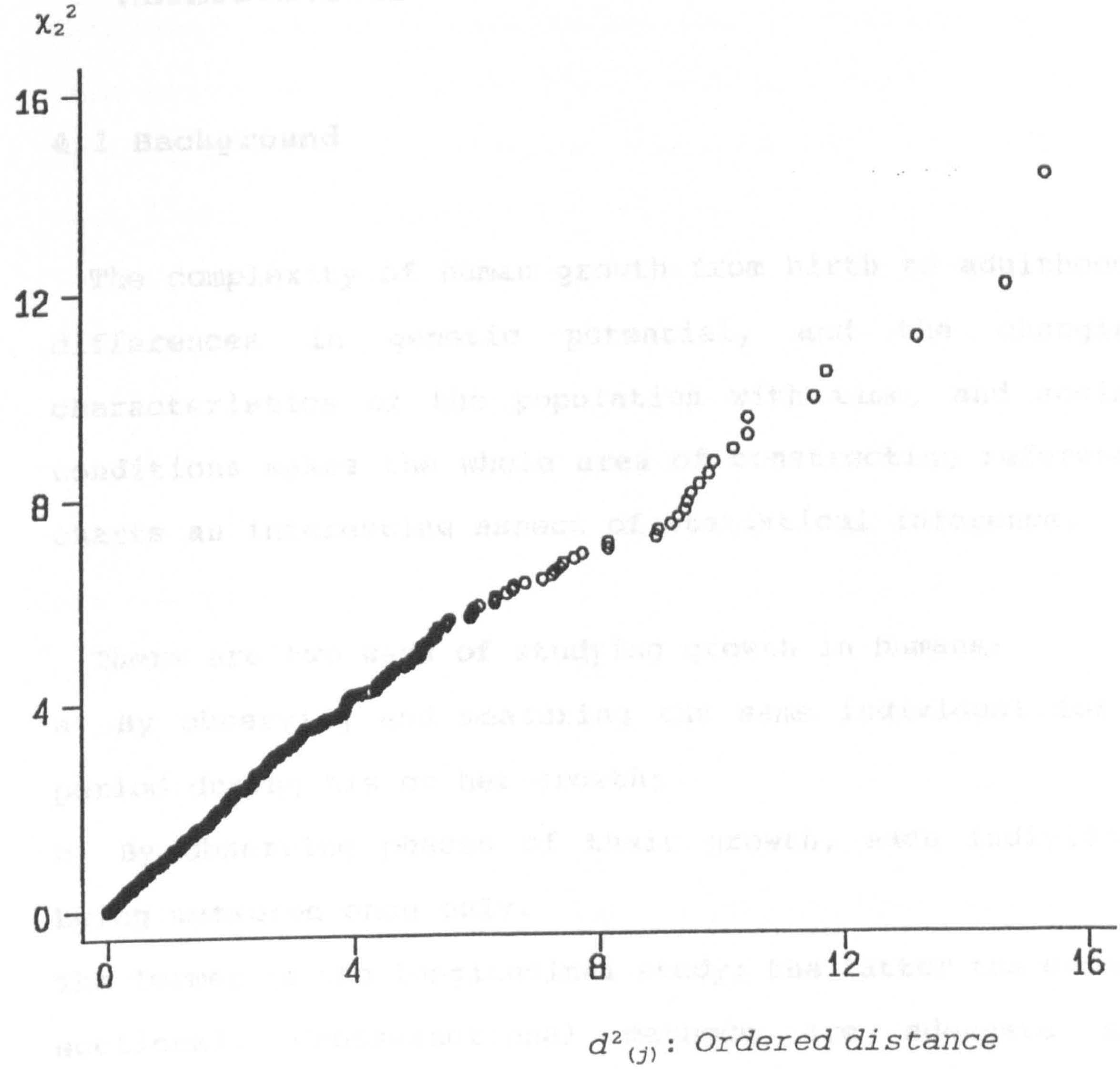
+ Marginal normal transformations λ (0.4, 2.0)

× Bivariate normal transformation λ (0.4, 2.15)



b) Enlarged section form Figure 3.5.a

Figure 3.6 Gamma plot (chisquared plot) of ordered distances of measurements (weight & height) of 12 year old girls after excluding outliers



CHAPTER FOUR

METHODOLOGIES FOR GROWTH CHART CONSTRUCTION

4.1 Background

The complexity of human growth from birth to adulthood, differences in genetic potential, and the changing characteristics of the population with time, and social conditions makes the whole area of constructing reference charts an interesting aspect of statistical inference.

There are two ways of studying growth in humans:

a) By observing and measuring the same individual for a period during his or her growth;

b) By observing phases of their growth, each individual being measured once only.

The former is the longitudinal study; the latter the cross-sectional. Cross-sectional methods are adequate for studying distributions of various measurements in different individuals at different ages and for constructing standards of growth attained, e.g. height and weight standards. Population 'standards' and 'growth charts' for clinical and field use are based very largely on cross-sectional data, although they may be modified in the light of information obtained by longitudinal studies (Marshall, 1977).

A longitudinal study is less satisfactory than a cross-sectional one for estimating the population mean at successive ages. This is due to the fact that the samples at each age in a longitudinal study are not independent of each other. But the longitudinal data are essential for the study of growth of individuals and for accurate assessments of the means and the standard deviations of growth increments, e.g. velocity and acceleration of growth. Sometimes, the mixed longitudinal study is carried out in which some subjects are measured repeatedly over a long period, some over short periods and some perhaps only once.

Several methods have so far been proposed for constructing smooth centiles curves or velocity standards for either cross-sectional or longitudinal studies. Healy (1989) outlined recent developments in the statistical handling of growth data. The type of distribution of measurements taken at each of a sequence of ages is of great relevance. While many linear measurements can be reasonably assumed to be Normally distributed (such as height), others of equal importance in anthropometry (such as weight, circumferences, skinfolds thickness) have distributions which clearly do not satisfy this assumption. This is why construction of age-related centiles has always been something of an art.

4.2 Traditional methods for constructing growth charts

4.2.1 Estimating Gaussian percentiles

Any data cross-sectional or longitudinal, may be considered as cross-sectional data and centiles derived using this information. However, information collected longitudinally may be used to construct a) longitudinal growth charts, b) velocity charts or c) individual growth curves. Suppose, we consider a single measurement made on a group of n individuals of the same age. Healy (1986) describes the method for constructing centiles assuming the data are distributed according to the Gaussian distribution. The p_{th} population centile $\mu+z_p\sigma$ may be estimated by the p_{th} sample centile $\bar{x}+z_p s$ where

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}, \qquad s=\sqrt{\frac{\sum_{i=1}^{n} (x_i-\bar{x})^2}{(n-1)}}$$

and z_p is the appropriate Normal equivalent deviate (eg, $z_p=0$ for the 50th centile and $z_p=-1.88$ for the 3rd centile). The larger the sample at each age group the more precise will be the population estimates.

The distribution of $\bar{x}+z_p s$ is found by observing that $\bar{x}$ and s^2 are independent and that:

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n}), \qquad \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

Noting that $V(f(y)) \approx (f'(\mu))^2 \times V(y)$, and setting $y=s^2$ and $f(y)=\sqrt{y}$, and $f'(y)=1/(2\sqrt{y})$, then by Taylor expansion

$$E(s) \approx \sigma(1-\frac{1}{4(n-1)})$$

$$V(s) \approx (\frac{1}{2\sigma})^2 \times \frac{\sigma^4}{(n-1)^2} \times 2(n-1) \approx \frac{\sigma^2}{2n}$$

Marubini (1980) shows that the distribution of s for large n is approximately Gaussian, so the distribution of $\bar{x}+z_p s$ is approximately Normal with a mean equal to the true centile, o(1/n), with a standard deviation given by:

$$\sigma\sqrt{\frac{1+\frac{1}{2}z_p^2}{n}}, \quad o(\frac{1}{n}) \tag{1}$$

As one generates centiles at either tail of the distribution, corresponding standard deviations become large unless the sample size is large. For example, the standard deviation for the 3rd centile is obtained by substituting $z_p=-1.88$ into equation (1) resulting in:

$$\sigma\sqrt{\frac{1+\frac{1}{2}(-1.88)^2}{n}}=\sqrt{\frac{2.77\sigma^2}{n}}$$

If the underlying population is not Gaussian there are two simple alternative methods to estimate the centiles. The first is to seek a transformation to the sample data which will render it Gaussian, allowing one to use the above method. If a suitable transformation cannot be found then a non-parametric approach is required.

4.2.2 Nonparametric method

The methodology described relies on the Gaussian shape of the standardizing population, but we must consider what to do when this assumption cannot be made. This is the case for weight, and many circumferential measurements and most biochemical measurements. In this approach the measurements are ranked from smallest to largest and then the required centiles can be read off. That is, in a sample of 99 the 3rd centile is the 3rd reading from the bottom and so on. For smaller samples some form of interpolation is required. Suppose, for example, given a sample of n ordered observations $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, the kth centile, Y_k, may be estimated as follows: let $p=k/100$ and $(n+1)p=J+G$, where J is the integer part and G the fractional part,

$$Y_k = (1-G)x_{(J)} + G\, x_{(J+1)}$$

Kendall and Stuart (1958) show that the sampling distribution of Y_k is approximately Gaussian and centred at the required population value, and the standard deviation for the $100p_{th}$ centile given by:

$$\sqrt{\frac{p(1-p)}{nf^2}} \qquad (2)$$

where f is the ordinate of the (noncumulative) distribution curve. If for a particular measurement the underlying distribution is approximately Gaussian, then for the 3rd centile, substituting $f=0.0681/\sigma$ into equation (2), gives the SD equal to

$$\sqrt{\frac{0.03(0.97)}{n(0.0681)^2}\sigma^2}=\sqrt{\frac{6.27\sigma^2}{n}}$$

Ignoring the Gaussian property of the data, for the 3rd centile, has an effect on precision which is similar to throwing away more than half the data. So there is a great loss of information with the ranking method and it is best avoided by searching for an appropriate transformation.

The discussion so far has assumed first, one requires centiles for a specific age and second, the available measurements with which to calculate the centiles are in fact all made at exact age. In practice centiles are required over ranges of ages, and children in an age group will not be exactly the same age when measured. Suppose one wishes to comment on the normality of a single child, we may usually assume that we know the child's age precisely; the implication is that we need centiles for a standardizing group of exactly the same age- 10.00 years old, say - whereas the available data will usually relate to a group whose ages range evenly from 9.5 to 10.5 years. Children are growing over this period which needs to be accounted for in the standard deviation, otherwise the SD will be overestimated. The lower centiles will contain children who have not actually reached their tenth birthday and the higher centiles contain children nearer to 10.5 years. Healy (1962) discusses a method to account for this where, assuming the ages vary uniformly over an interval h, centred on scheduled age, the calculated variance of the observation should be reduced by $(\frac{h^2}{12})(\frac{dy}{dx})^2$ where $\frac{dy}{dx}$ is the gradient of the 50th centile at the scheduled age.

Various suggestions for modelling the variance as a function of age have been made. Goldstein (1972) proposed to regress variance on age after dividing the data in several small age groups similar to the way that the mean of the data has been modelled. Royston (1991) gave attention to the possibility that the SD changes with age by dividing the data into three equal age groups and comparing the SDs in the youngest and the oldest, which is a test for linear trend in SD. While this method works well when the SD changes linearly with age, it can not detect non-linear (including non-monotonic) variation in the SD. A similar approach (Isaacs et al., 1983) allows the SD to be modelled as a (possibly non-linear) function of age but with rather more age groups. Moreover, Altman (1993) proposed a method based on absolute residuals to model SD with age, which works without grouping age (to be discussed later).

4.2.3 Curve smoothing

When constructing a growth chart, after appropriate age grouping a series of centile values may be constructed for a set of ages covering the required age range. Very often these series of crude centiles were joined by hand. Until recently most growth experts have smoothed their curves by hand to minimize distorting the data by unknown mathematical factors. Perhaps the chief disadvantage of smoothing by hand is that, like all great art, it is not quantifiable and not exactly reproducible.

The problem of drawing a smooth curve through a set of points that are affected by errors can be tackled in a number of ways. It may be possible to devise a mathematical formula relating the values of a particular centile to age; this can then be fitted to the data by ordinary statistical methods. This technique has not been very widely used in practice owing to the lack of generally applicable equations. In particular, no curve that can be fitted by ordinary least squares (such as polynomial curves) is likely to be satisfactory at ages approaching adulthood, where the curve tends to an asymptote. Methods using splines have been proposed, but these are inclined to introduce ripples into the fitted curve or to overdo the smoothing process. Moving average methods can be also useful (with equally spaced points, a simple moving average of pairs of consecutive values may work very well), but the end of the age range has to have a special treatment (Healy, 1986).

The NCHS (1977) curves were constructed by curve smoothing. When constructing these curves, two basic methods of systematic computerized curve smoothing were considered:

1-*spline polynomial smoothing of the observed percentiles*

2-*smoothing by means of the Pearson curve system using polynomials in age to estimate the first four moments.*

Splining is a mechanism used to fit curves to a body of data in such a way that no discontinuities (i.e., disjunctions or sharp angles at junctions) exist in the final curves. Mathematically, the first and possibly higher derivatives at the points where curves join are continuous. The second system, while very sensitive to the enormous amount of information contained in the median or central tendency, would possibly be oversensitive to outlying values (although many outlying values are valid, this region usually presents the greatest frequency of spurious data). The method also would have required much more developmental work before it could be used.

The cubic spline technique was employed to smooth the observed centiles individually to the eventual satisfaction of the NCHS task force. It was realized that the cubic spline technique, because of the strength of having two modes (fixed or variable) of placing the knots (NCHS, 1977), is even better than the quadratic one to express a finer interrelation between the percentile lines.

The traditional methods of construction that Healy (1986) described are simple to use but suffer from various limitations. For example, the actual ages of children are not used in the calculation of centiles, which is also the case in the estimation of age specific SD. Besides, the correction factor applied to the variance is based on assumptions that are not always true, such as: uniform ages over each age group and linear increase of variance and mean with age. Another drawback of these methods is that there is no technique for converting the raw data to Z-scores as is required in studying age independent growth patterns within and between children. This drawback applies to spline curves as well (e.g. NCHS curves). Section 2.4 describes the technical problems of using two distinct data set in deriving Z-scores for NCHS curves.

4.3.1 Distribution free estimation method for constructing growth chart (HRY method)

Healy et al. (1988) recognized the drawbacks and lack of flexibility in some of the available methods of producing growth centile charts. The data are often not Normally distributed and the measurements are rarely recorded at the scheduled age. Nor was it easy to convert a child's measurements into Z-scores, which enable a child's underlying growth pattern to be studied independent of the normal effects of age.

The HRY method (an acronym of the authors' initials) overcomes most of the shortcomings with the traditional methods, and makes few assumptions about the nature of the distribution of the measurements at fixed ages. The HRY method incorporates the Cleveland's (1979) procedure of robust locally weighted regression into the first of the following two steps:

1. *Obtaining the 'raw' centiles*

The observed measurements are first sorted according to the age (time). The first k measurements are then selected, where k is a fraction of the total number of observations n, typically 5-10%. A linear regression of these k measurements on age is formed and the residuals calculated. By sorting and ranking, selected centiles (usually the 3rd, 10th, 25th, 50th, 75th, 90th, and 97th) are derived for these k residuals. By adding them to the fitted value of the regression, the corresponding centiles at the median age $\tilde{t}_1$, of the k values of t are calculated. These provide crude estimates of the centiles of the data at age $\tilde{t}_1$, called 'raw' centiles. This procedure has used the first k points of the data; it is repeated successively using data points 2 to k+1, 3 to k+2,... until the whole span of age has been covered. Note the age range of the raw centiles will not include k/2 of time points at each end of the age-range.

2. *Smoothing*

The points arising from step 1 will be very irregular and need to be smoothed to provide usable centile curves. Not only should each centile follow a smooth curve, but the intervals between centiles at a given age should also change smoothly. The method proposed copes with both requirements through the fitting of a multivariate polynomial.

The fitted polynomial has two parts to it; the 50th raw centile is modelled over time by a polynomial of degree p in t, where t denotes age and Y_{50} the smoothed value of the 50th centile, shown in the equation

$$Y_{50} = a_0+a_1t+a_2t^2+...+a_pt^p. \qquad (3)$$

Next, at any given age t, it is supposed that the distribution of the measurements can be described by a polynomial in the Normal deviate **z** so that the ith centile Y_i is given by equation

$$Y_i = Y_{50}+b_1z_i+b_2z_i^2+...+b_qz_i^q \qquad (4)$$

where z_i is the Normal equivalent deviate corresponding to ith centile of the Normal distribution. For the 50th centile z=0, and for the 3th centile z=-1.88. If at this age the data are Normally distributed with standard deviation $\sigma(a)$ then $b_1=\sigma(a)$ and $b_i=0$ for i>1. The term z^2 and higher even powers of z can describe skewness and other forms of asymmetry while z^3 can account for kurtosis in the distribution of y.

As age changes, the variance, skewness and kurtosis of y may also change. To model such changes the coefficients b_i are made polynomials in t so that after collecting together terms in t^0, t^1, t^2 etc., we obtain from equations (3) and (4)

$$Y_i=a_0+b_{01}z_i+..+(a_1+b_{11}z_i+..)t+(a_2+b_{21}z_i+..)t^2+(a_p+b_{p1}z_i+..+b_{pq}z_i^q)t^p$$

Since the ages and the z's corresponding to the raw centiles are known, the model parameters can be calculated by standard least squares (e.g. using multiple regression procedures). The main advantage of this method is the great flexibility in allowing for

a) the centile curves to vary smoothly in a non-linear manner with age and

b) non-Normal cross-sectional distributions which may change shape as age changes, including changing the standard deviation (SD) and skewness.

The described procedure for fitting the model supposes that measurements are made at a continuum of ages which are not grouped. In our study of growth of Iranian children, data are grouped in complete years. The appropriate raw centiles for each age group were computed using SPSS software. These centiles were used in the second stage of the HRY analysis to derive the smoothed centiles.

4.3.2 Splining polynomials (PGY Method)

In practice, when the age span is wide, for example our data on weight and height of Iranian children aged 2-18 years old, or when the velocity of growth changes markedly over the age range, this method tends not to be satisfactory. The approach of simply increasing the order of the polynomial does not work because high order polynomials in age often do not provide a good fit at the extremes of the age range.

Pan, Goldstein and Yang (1990) proposed a method which allows a wide age range to be fitted using relatively low-order polynomials. The new method (**PGY**) extends the procedure of the HRY method by splining low-order polynomials each spanning a limited age range. As with the HRY method two steps are involved. The first is to obtain the 'raw' centile estimate as previously described for the whole age range. These estimate are irregular and need to be smoothed, to provide the required centile curves. It is assumed that the curves can be fitted by a series of splined polynomials of degree p. If t denotes age, and $Y_{i,t}$ the smoothed value of the ith centile, we may write

$$Y_{i,t} = a_{0,i}+a_{1,i}+\ldots+a_{p,i}t^{p}+a_{p+1,i}(t-c_1)^p_+ +\ldots+a_{p+m-1,i}(t-c_{m-1})^p_+ \quad (5)$$

$$\text{where} \quad (t-c_l)_+^p = (t-c_l)^p \quad \text{when } t>c_l$$
$$= 0 \quad \text{when } t\le c_l \quad l=1,2,..,m-1$$

and c_l is the lth cut or join point between the polynomials spanning successive age ranges with $c_1<c_2\ldots<c_{m-1}$. Of course, when m=1, the new method is the same as the HRY method. With one join point and a polynomial of degree 3 we obtain for the ith centile

$$Y_{i,t} = a_{0,i}+a_{1,i}t+a_{2,i}t^2+a_{3,i}t^3 \quad \text{when } t\le c_1 \quad (6)$$
$$= a_{0,i}+a_{1,i}t+a_{2,i}t^2+a_{3,i}t^3+a_{4,i}(t-c_1)_+^3 \quad \text{when } t>c_1$$

It is clear that the two components of the curve join smoothly at the join point c_l, and this is generally true for curves of this form, because not only are the two polynomials continuous at c_l but both 1st and 2nd derivatives are also continuous at c_i. The value m and the position of the joins can be chosen after inspection of the data and using existing knowledge.

The coefficient $a_{j,i}$ is now expanded as a polynomial function of the NEDs z_i of degree q:

$$a_{j,i} = b_{j,0} + b_{j,1}z_i + \ldots + b_{j,q_j}z_i^{q_j} \qquad j=0,1,\ldots,p+m-1 \qquad (7)$$

combining (1) and (3) leads to composite linear model for the centile curves whose coefficients can be estimated from the raw centiles using ordinary least-squares (OLS) analysis, as with the HRY method.

In general, the values of q_j will usually be higher for low-order coefficients of t and may be zero for the high-order ones. A test of fit can be obtained by comparing the percentages of data that fall between the centile curves for sub-ranges, with their expected values, as described below.

So, the PGY method makes it possible to construct centiles for a wide span of ages using comparatively low order polynomials in age. Each section is splined together so that each centile line is continuous and has p-1 continuous derivatives at the joins.

4.3.3 Assessing the fit of the model

How does one assess the fit of a selected polynomial fitted to the raw centile? Because of the way the raw centiles are constructed all the points are highly correlated. Consequently, it is not possible to use the residual mean square to assess the fit of the model. However there are alternative methods to aid in the assessment of fit.

First, one may assess the fit of a particular polynomial to the raw centile by inspection. Plotting the raw and smoothed curves together will identify any obvious shortcomings of the model. When assessing the smoothed model, one needs to remember that the aim of the polynomial model is to smooth out some of the irregularities that occur in raw centiles. So it is not necessary, and in fact undesirable, to seek to match exactly the raw centiles in the smoothing process. Hence, if a particular raw centile suddenly drops off for a short time and none of the other centiles follow suit, it is probably a spurious result and should not be reflected in the model.

Second, the authors of HRY method suggest that the data points falling between the centiles curves are counted and compared with the expected number. The expected figures are obtained simply by taking the complete data set, say 100 measurements, and calculating that there should be 3 data points above 97th centile, 15 between 75th and 90th centiles, etc. A single overall summary of the observed and expected data points over all ages may be obtained, or this summary may be divided into some sub age groups. Difference between the observed and expected counts indicate lack of fit between the data and smooth centiles for any model. This may mean the model is poor over a particular age range or it may mean that too much smoothing of the raw centile has taken place. Splitting the data up by ages groups provides more detailed information across the complete model. In practice the results of this procedure are similar to those obtained by inspection of the raw and smooth centiles (this will be discussed more in chapters 7 and 8).

Third, having obtained what appears to be an acceptable model to the original measurements. It is possible to transform the measurement value into centile values, and then convert the centile values into Z-scores, also referred to as Normal Equivalent Deviate (NEDs) or Standard Deviation Scores (SDS). If the model fits the data, the distribution of these Normal deviates should be Normal with mean 0, the SD=1 and should be constant with age. Visually, the Normal plots of these Z-scores should be a straight line or in examination of the SDS against age no pattern should be observed (Altman, 1993).

4.4 Methods based on Normalising transformations

4.4.1.1 Cole's method

A method for constructing population centiles based on the Box-Cox transformation to Normality was proposed by Cole (1988a). This was a wider development of the use of these transformations than suggested by Van't Hof et al. (1985). The concept is to use a Box-Cox transformation to Normalise the data, estimate the mean and SD of the transformed data as a function of age and then derive the required centiles.

Cole's method concentrates on the power transformation of observations y, $y^{(\lambda)}$, proposed by Box and Cox (1964) for correction of skewness (described in 3.4). Box and Cox (1964) showed that the maximum likelihood estimate of λ is that which minimizes the variance of the scaled variable

$$f^{(\lambda)}=\frac{y^{(\lambda)}}{y_m^{\lambda-1}}$$

where y_m is the **geometric** mean of y's. It is clear that $f^{(\lambda)}$ is in the same dimension as y whatever the value of λ, so that as λ varies $var(f^{(\lambda)})$ remains in the same units as var(y).

An alternative function for deriving the scaling transformation is given by:

$$g^{(\lambda)}=\left(\frac{y}{y_m}\right)^{(\lambda)}$$

which arises from applying the Box-Cox transformation to $\frac{y}{y_m}$, where y_m is the median of y. In this case $g^{(\lambda)}$ is dimensionless and its standard deviation is analogous to the coefficient of variation of y. The advantage of $g^{(\lambda)}$ over $f^{(\lambda)}$ is that for many variables where this technique is of value the standard deviation increases fairly steadily with the mean, while the coefficient of variation, c.v, does not. I.e., the c.v. is relatively independent of the mean.

In Cole's method as originally proposed the data need to be divided into distinct time groups, with means t_i ($i=1,2,\dots,p$). MLEs of λ_i are obtained for each group, which minimize σ (standard deviation of $g^{(\lambda)}$), by fitting a quadratic in λ to $\log[\mathrm{var}(g^{(\lambda)})]$. Within each group, various values of λ_i are used to seek this minimum, $\hat{\lambda}_i$. Corresponding estimates for the median of the y's, $\hat{\mu}_i$, SD of the transformed observations ($\hat{\sigma}$) are also obtained (Cole 1988a). Then $\hat{\lambda}_i, \hat{\mu}_i$, and $\hat{\sigma}_i$ plotted against t_i, and are fitted by smoothed curves L(t) (L for the lambda), M(t) (M for mu), and S(t) (S for sigma) respectively. This is the reason why the method is called **LMS**.

The smoothing can be done using polynomial regression or whatever method is convenient, e.g cubic splines (Silverman, 1985; as Cole used), or simply fitting by eye. Then together a smooth curve for the 100αth centile is given by

$$C_{100\alpha}(t) = M(t)[1 + L(t)S(t)z_\alpha]^{1/L(t)} \qquad (8)$$

where z_α is the Normal equivalent deviate for tail area α. The equivalent form if L(t) is zero is given by

$$C_{100\alpha}(t) = M(t)\exp[S(t)\ z_\alpha].$$

Rearranging equation 8 makes it possible to express measurements for individual subject in standard deviation score:

$$z=\frac{[y/M(t)]^{L(t)}-1}{L(t)S(t)} \qquad L(t)\neq 0. \qquad (9)$$

Also the term y/M(t), Y is expressed as a fraction of the median.

Cole (1989b) used this method on national weight-for-age standard data from UK (Tanner et al., 1966), US (Hamill et al., 1979) and the Netherlands (Roede et al., 1985) to re-construct growth standards. The agreement is reasonably close, particularly for the UK and Dutch standards, which were originally fitted by eye, whereas the US standard was fitted using cubic spline functions as described before. This method also was used to produce reference centiles for weight, height, body mass index, and head circumference for new national standards for the UK (Cole et al., 1995).

However the LMS method is a completely different approach to HRY method, it is a powerful method that has much in common with the latter in that it produces centile which are

a) *smooth*

b) *close to the data*

c) *constrained to accord with neighbouring centiles, and*

d) *data can be converted to SD or Z-scores.*

4.4.1.2 Modification of the LMS method using penalized likelihood

The LMS method makes the big distributional assumption that the Box-Cox transformation Normalises the data, although for modest departure from Normality (i.e some skewness) it behaves very similarly to the HRY method. If non-Normality were very marked (which Cole would not expect with anthropometry) this might be a serious problem with the LMS method. Apart from the Normality assumption, the main weakness of the LMS method was its need of distinct age groups. The choice of age cut-offs is arbitrary, and in theory could have influenced the final result. In addition, the value for the power L read off the smoothed curve was not always the value used to calculate the M and S values at the same age, so that the process did not iterate to convergence. Green (1988) highlighted the problem, and pointed out that penalized likelihood could be used to provide smooth estimates of the L, M, S curves directly.

For the case of n independent observations $\{y_i\}$ at corresponding covariate values $\{t_i\}$, the log-likelihood function (L^*) derived from equation 9 is (apart from the constant) given by

$$\boldsymbol{L}^*=\boldsymbol{L}^*(L,M,S)=\sum_{i=1}^{n}\left(L(t_i)\log\frac{y_i}{M(t_i)}-\log S(t_i)-\frac{1}{2}z_i^2\right) \qquad (10)$$

where $\{z_i\}$ are the SD scores corresponding to $\{y_i\}$. The curves L(t), M(t) and S(t) are estimated by maximizing the penalized likelihood

$$\boldsymbol{L}^* -\frac{1}{2}\alpha_\lambda\int\{L''(t)\}^2dt-\frac{1}{2}\alpha_\mu\int\{M''(t)\}^2dt-\frac{1}{2}\alpha_\sigma\int\{S''(t)\}^2dt \qquad (11)$$

where the α_λ, α_μ, and α_σ are the smoothing parameters. The three integrals provide roughness penalties according to the squared second derivatives of the L, M, S curves, so that maximizing (11) strikes a balance between fidelity to the data and smoothness of the L, M, S curves. These forms of the penalty lead to natural cubic splines with knots at each distinct value of t (Silverman, 1985). The parameters α_λ, α_μ, and α_σ need to be chosen in order to fit the model.

It is worthwhile mentioning that Cole (1988a) power transformed y ($y^{(\lambda)}$), and in calculation of var($g^{(\lambda)}$) worked with y^λ. But in the modification of his paper (Cole and Green, 1992), he worked with $x=(\frac{y}{\mu})^{(\lambda)}$ where this transformation maps the median μ of y to x=0. For $\lambda=1$ the SD of x is exactly the coefficient of variation (CV) of y, and this remains approximately true for all moderate λ. The optimal value of λ is that which minimizes the SD of **x**.

4.4.1.3 Advantages and limitations of the modified LMS method

The penalized likelihood approach avoids identifying individual age groups, and instead treats the entire data set as a single entity. The key assumption of the method is that after a suitable power transformation, the data are Normally distributed. Some anthropometric measurements, such as weight and height, tend to follow this pattern but the main problem with the assumption may be the presence of kurtosis for which the transformation does not adjust.

Another drawback of the modified method is the arbitrary choice of the three smoothing parameters α. There is no mathematical justification for reducing this arbitrariness by assessing goodness of fit. However the LMS method fitted by penalized likelihood provides a convenient 'black box' for the fitting of the smooth reference centiles curves. Also cubic splines require a table of values on the age grid, whereas the advantages of a functional form are that asymptotic behaviour can be modelled explicitly, and the summary curves are parsimoniously represented (Royston and Wright, 1996).

The main interest in the distribution is the region beyond the extreme observed centiles. The two requirements to predict this region at all accurately are a well-behaved distribution (such as is found with anthropometry) and an adequate summary of the distribution. The HRY method is theoretically more flexible and can do this by solving the polynomial of order q linking centiles and SD scores. So the presence of kurtosis also can be modelled easily.

For practical applications, a FORTRAN program has been developed to implement the modified LMS method (Cole and Green, 1992)[1]. On the other hand this advantage is met by

1: This program has generously been made available to other research workers, Royston (1996) adapted it so that it can be run from STATA (1995), and he has given a copy to the Medical Statistics Unit of London School of Hygiene and Tropical Medicine more flexibility in the HRY method by using GROSTAT II software. GROSTAT II which is a superset of GROSTAT I, is a program for estimating age-related centiles for growth and development data. It is useful for both distance and velocity data and also can handle piecewise polynomials based on the PGY method. Both cross-sectional and longitudinal data can be handled by this software. The software is equipped with a high resolution graphic facility which can be viewed and printed.

A good discussion on the comparison of the application of the two methods with an amalgamation of the two can be found in Ayatollahi et al. (1993b). They used both methods on three simulated sets of data and found that the performance of both methods was satisfactory when data was derived from distributions close to Normal. But practical problems emerge with data that are not close to Normal. More specifically in the LMS method the L curve is more difficult to smooth and it was the matter of high discrepancy with two of the data sets in their work. They therefore proposed to amalgamate both methods. This worked best. Their suggestion was to Normalise data in age groups using the L(t) transformation from the LMS method, or any other approach to Normalise the data, and then using the HRY method to derive centiles. This coincides with Healy's opinion, that HRY method will work best if the underlying distribution is approximately Normal. It is worth mentioning that Ayatollahi et al. (1993b) did not use the penalized log-likelihood method in their work.

If the data is ill behaved (as sometimes occurs), the HRY method is the approach to be preferred. Its flexibility and robustness makes it ideal for handling all kinds of data. As explained, it is computationally no more demanding because it has been successfully implemented into a statistical package which is fairly easy to use. The HRY method is probably more robust than LMS and may show better results than the LMS method, especially toward the tails. However in the latter case the magnitude of differences between the two methods is never dramatic (Ayatollahi, 1991).

4.4.2 Exponential models with log transformation

Recently, Royston and Wright (1996) proposed the EN method. The corresponding STATA procedures are available. Basically, this method is based on the idea of two stages of transformation of the data and modelling the mean and standard deviation (SD) using Fractional Polynomials (Royston and Altman, 1994; to be discussed in section 4.6), and the absolute residuals (Altman, 1993; section 4.4.3). Also, in order to improve the fit, an iterative procedure between estimations (Aitkin, 1987) of mean and SD is implemented. In this method, the non-Normal skewness or kurtosis is modelled at the second stage if necessary after log transformation of the variable Y at the first step. To accommodate skewness, they suggested a 3-parameter Exponential Normal (EN) model (Manly, 1976) as:

$$U=\frac{\exp(\gamma z)-1}{\gamma}; \qquad \text{where } \begin{cases} z=\dfrac{g(y)-\mu}{\sigma} \\ g(y)=\log(y). \end{cases}$$

The parameters μ, σ, γ are referred to as M, S, G curves. Also, a 4-parameter modulus exponential (MEN) distribution

$$V=sign(u)\frac{1}{\delta}\{(1+|U|)^{\delta}-1\}; \qquad \text{where } u=\frac{\left(\frac{y}{\mu}\right)^{\lambda}-1}{\lambda\sigma}$$

is proposed for modelling the kurtosis (δ: D curve). Then the percentiles in the Normalised scale are computed and the original centiles are obtained by algebraically inverting the formula.

4.4.3 Construction of the reference centiles using absolute residuals

Altman (1993) proposed a method of modelling variation in the SD which has several advantages over other methods when the data are Normally distributed or can be transformed to be so. Suppose the mean of the Y is modelled adequately by some function of age (for example, **fp**), the information about the SD of the measurements is contained in the residuals around the fitted curve. The parametric method of deriving a reference range is based on the assumption that the variable has a Normal distribution at all ages, so the residuals should have a Normal distribution at each age and absolute values of the residuals should have a half Normal distribution. The mean of a half standard Normal distribution is $\sqrt{(\frac{2}{\pi})}$ (Johnson and Kotz, 1970). Thus the mean of the absolute residuals multiplied by $\sqrt{(\frac{\pi}{2})}$ is an estimate of the SD of the residuals. It follows that, if the SD is not reasonably constant over ages, the predicted values from regression of the absolute residuals on age multiplied by $\sqrt{(\frac{\pi}{2})}$ will be age-specific estimates of the SD signed residuals, and hence of Y.

The proposed method does not involve the criterion of arbitrary age grouping and makes no additional assumptions (just conditional Normality of Y), and is computationally simple. Methods based on grouping across age involve separate analyses of summary statistics to get a regression equation which then has to be reintroduced into the main analysis. This two-step analysis is tedious and there is the risk of transcription error.

4.5 Multilevel modelling

4.5.1 Introduction

In the social and biological sciences, data are often hierarchical in the following sense: we have variables describing individuals, but the individuals also are grouped into larger units, each unit consisting of a number of individuals (Bryk and Raudenbush, 1992). The leading example, is perhaps in education. Pupils learn in classes; classes are taught within school; and schools may be administered within local authorities or school boards. The units in such a system lie at different levels of the hierarchy, or are "hierarchically" related. In other words the membership is nested, to which the term "multilevel" refers.

In a household survey like our data, the level 1 units are individual children, the level 2 units are households and level 3 units, area defined in different ways (for example, clusters in our data). And these areas (clusters) are from different administrative authorities (for example, provinces). Such a hierarchy is often described in terms of level's of units within higher level units, and the term *clustered population* sample is also used. In recent years many social researchers have become aware that much, if not most, of their data have an inherent hierarchical structure and that the structure affects the summary statistics of interest.

The term ***multilevel*** models discussed here, are referred to as ***mixed-effect* models** and ***random-effects* models** in biometric applications (cf.Elston & Grizzle, 1962; Laird & Ware, 1982). They are also called ***random-coefficient regression* models** in econometric literature (cf.Rosenberg,

1973) and in the statistical literature are often referred to as ***covariance components* models** (cf.Dempster, Rubin, & Tsutakawa, 1981; Longford, 1987). Because the models convey an important structural feature of data they are called ***hierarchical linear* models** by Bryk (1992).

A number of useful reviews of articles within the field of education have appeared which clearly present the foundation for multilevel analysis and several aspects of the theory of it, for example Goldstein (1995) and Bryk (1992) and Longford (1993). Also among a broad range of articles in this area Aitkin and Longford (1986) provide a comparison of the multilevel approach with single level techniques for analysing hierarchically structured data.

4.5.2 Characteristics of multilevel models

Some of the key features of multilevel analysis which have been found to be useful in extending understanding of social and biological processes beyond that which can be obtained through single linear modelling will be discussed here to demonstrate the potential of the methodology.

1) Coefficients in a linear model of a process occurring at one level of a hierarchial system can be viewed as variables of interest that are functions of characteristics of units at another level. Further, the variance and covariance of these coefficients are often

of direct interest.

2) Coefficients of within-unit relations among variables are generally estimated better than they would be if a single level analysis was conducted for each group.

3) Using the appropriate model specification resolves the problem of misestimated precision inherent in single level analyses of hierarchically structured data.

4) Longitudinal data have a measurement structure which is nested within individuals. Multilevel analysis permits individuals to have their own growth curves.

By focusing attention on the levels of hierarchy in the population, multilevel modelling enables the researcher to understand where and how effects are occurring. It provides better estimates in answer to the questions for which single-level analyses were once used and in addition allows more complex questions to be addressed.

4.5.3 The consequences of ignoring a multilevel structure

Having implied that a statistical model should recognise explicitly a hierarchical structure where one is present, we need to point out the consequences of failing to do this. There are two obvious procedures that have been somewhat discredited. The first is to disaggregate all higher order variables to the individual level. Teacher, class, and school characteristics are all assigned to the individual, and the analysis is done on the individual level data. The problem with this approach is that if we know that students are in the same class, then we also know that they have the same value on each of the class variables. Thus we cannot use the assumption of independence of observations that is basic for the classical statistical techniques. Also, carrying out an analysis which does not recognise the existence of clustering at all, creates serious problems. For example, clustering will generally cause standard errors of regression coefficients to be underestimated. Hence significance tests will too often reject the null hypothesis.

The other alternative is to aggregate the individual-level variables to higher characteristics over classes and do a class analysis, perhaps weighted with class size. The main problem here is that we throw away all the within group information, which may be as much as 80% or 90% of the total variation before we start the analysis. As a consequence, relations between aggregated variables are often much stronger, and they can be very different from the relation between the nonaggregate variables. Thus we waste information, and we distort interpretation if we try to interpret the aggregated analysis on the individual level. An empirical demonstration of this was given by Woodhouse and Goldstein (1988). Thus both aggregating and disaggregating may be unsatisfactory.

4.5.4 Model formulation

The key idea of theory of multilevel modelling will be discussed briefly to explain some general background, and this discussion will begin by presenting a two-level model, but the theory and interpretation generalize straightforwardly to three and higher levels. The presentation draws on Prosser et al. (1991) and Woodhouse (1993).

4.5.4.1 A basic two-level model

Suppose that data have been collected in J families, each of which contain n_j ($j=1,2,\ldots,J$) children and that the researcher is concerned with relationship between individual's height (Y_{ij}) and his or her age (X_{ij}). For child i in family j a linear relationship can be written as follows:

$$Y_{ij} = \beta_{0j} X_0 + \beta_{1j} X_{ij} + e_{ij} \qquad (X_0=1) \qquad (12)$$

The intercepts β_{0j} in this within-family relationship is expected height for child of average age (suppose X is centred), and the slope β_{1j} is the average change in height for each year of age. This is the basic relationship between X and Y, i.e., at level 1.

Since both β_{0j} and β_{1j} in general can vary across families, these coefficients can be viewed as random variables at the between-family level, level 2. One aim of the analysis might be to attempt to account for this variation in terms of one or more features, Z, of the family being studied; for example, income.

Between-unit models for the β_{0j} and β_{1j} in terms of Z can be written as follow:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j} \tag{13}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + u_{1j} \tag{14}$$

where γ's are fixed parameters and u_{0j} and u_{1j} are random variables.

In this model, the coefficients γ_{01} (γ_{11}) represent the rate of average change of the group mean (average height) or of the rate of growth (slope) with Z. There are many other possible models of the between-family variation as shown below. One is not required to model both β_{0j} and β_{1j} as functions of Z. Four other examples are:

A) $\beta_{0j} = \gamma_{00} + u_{0j}$
$\beta_{1j} = \gamma_{10}$

B) $\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j}$
$\beta_{1j} = \gamma_{10}$

C) $\beta_{0j} = \gamma_{00} + u_{0j}$
$\beta_{1j} = \gamma_{10} + u_{1j}$

D) $\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j}$
$\beta_{1j} = \gamma_{10} + u_{1j}$

where γ's are fixed and u_{0j}, u_{1j} are random variables with parameters

$$\begin{cases} E(u_{0j}) = 0 \quad , \quad E(u_{1j}) = 0 \\ var(u_{0j}) = \sigma^2_0, \ var(u_{1j}) = \sigma^2_1 \quad , cov(u_{0j}, u_{1j}) = \sigma_{01} \ . \end{cases}$$

When (A) or (B) is used, the model in equation 12 is often referred to as ***variance components*** model. Since the variance of the response, about the fixed component, the fixed predictor, is

$$Var(y_{ij}|\beta_{0j},\beta_{1j})=var(u_{0j}+e_{ij})=\sigma^2_0+\sigma^2_e$$

that is the sum of level 1 and level 2 variances. In both cases, slope β_{1j} is considered to have the same value in every family. The model consisting of (C) and the model in equation 12 is termed a ***random coefficient regression*** model. In (D) only simple random variation is present in slope, but variation in family means (when $X_1=0$) is partly accounted for by Z.

4.5.4.2 Distributional assumptions

The following assumption are usually made:

a) the e_{ij}s in group j are independently distributed with an expected value of 0 and variance of σ^2_e for all j.

b) the level 2 random terms u_{0j} and u_{1j} have a joint distribution with mean 0 and covariance matrix $\Omega_{(2)}$ (The subscript (2) refers to level 2).

$$\Omega_{(2)}=\begin{pmatrix}\sigma_0^2 & \sigma_{01}\\ \sigma_{01} & \sigma_1^2\end{pmatrix}$$

c) the level 1 random term (e_{ij}) is distributed independently from each level 2 random terms. Typically, (Multivariate) Normality is assumed for the random terms at each level. Here level 1 has a simple form: $\Omega_{(1)} = \sigma^2_e$.

d) X_{1ij} and Z_j are fixed, known variables.

Combining (12), (13) and (14) in section 4.5.4.a provides the single equation version of the model:

$$Y_{ij}=\gamma_{00}X_0+\gamma_{01}Z_j+\gamma_{10}X_{1ij}+\gamma_{11}Z_jX_{1ij}+(u_{0j}X_0+U_{1j}X_{1ij}+e_{ij}X_0) \qquad (15)$$

the fixed part of the model in equation 15 contains explanatory variables with γ's as coefficients. The random part of the model, in brackets, involves two explanatory variables – the constant X_0 and child's age X_1. Note in the random part X_0 is involved at both levels, whereas X_1 is multiplied by only a level 2 random variable because the effect of age has been considered only random between families. The most general two-level model permits the coefficients of any explanatory variable to be random at any level.

4.5.4.3 - Formulation of a two-level model

A general two-level model with a simple level 1 covariance structure is now described. Suppose $\mathbf{Y}_j$ is the vector of the n_j response variable values for group j. $\mathbf{X}_j$ is an $n_j \times r$ matrix of group members' values on a set of r explanatory variables (typically X_0, the unit vector, is in this set), β_j is a vector of the r coefficients for the group, and $\mathbf{e}_j$ $(=[e_{1j},\ldots,e_{n_jj}]^T)$ is a n_j-vector of level 1 random terms. The within-unit model for the jth level 2 unit can be written as

$$\mathbf{Y}_j = \mathbf{X}_j \, \beta_j + \mathbf{e}_j.$$

The between-unit model for the coefficients can be written as

$$\beta_j = \mathbf{Z}_j \, \Gamma + \mathbf{u}_j$$

Where $\mathbf{Z}_j$ is an r×q between-unit design matrix, Γ is a vector of the fixed coefficients (q of them in all), and $\mathbf{u}_j$ is an r-vector of random terms. Combining the within- and between-unit models for the group gives:

$$\mathbf{Y}_j = \mathbf{X}_j \, \mathbf{Z}_j \, \Gamma + \mathbf{X}_j \, \mathbf{u}_j + \mathbf{e}_j$$

The $\mathbf{X}_j \, \mathbf{Z}_j \, \Gamma$ term is called the fixed part and the latter two terms form the random part.

The variance of $\mathbf{Y}_j$ conditional on the fixed part is given by

$$\sum\nolimits_j = Var(X_j u_j + e_j) = X_j \Omega_{(2)} X_j^T + \sigma_e^2 I_{n_j} \qquad (16)$$

where $\Omega_{(2)}$=Var(u_j), and $\mathbf{I}_{n_j}$ is an $n_j \times n_j$ unit matrix. It is noted in passing that the assumption of independence of the level 1 and level 2 random terms is used in the derivation of this expression.

Stacking models of the J groups produces the following expression

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_J \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_J \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_J \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_q \end{pmatrix} + \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_J \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_J \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_J \end{pmatrix}$$

i.e. $$Y = X \, Z \, \Gamma + X \, U + e \qquad (17)$$

The variance of the random part $\mathbf{X} \, \mathbf{u} + \mathbf{e}$ can be expressed in a form corresponding to equation (16):

$$\sum = X\begin{pmatrix}\Omega_{(2)} & 0 & \cdots & 0 \\ 0 & \Omega_{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Omega_{(2)}\end{pmatrix}X^T + \begin{pmatrix}\sigma_e^2 I_{n_1} & 0 & \cdots & 0 \\ 0 & \sigma_e^2 I_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_e^2 I_{n_J}\end{pmatrix}$$

$\sum$ is an N×N diagonal matrix where $N=\sum_j n_j$. Each entry on the diagonal of $\sum$ is the total variance (level 2+ level 1) for an individual.

This basic two-level model imposes restrictions in several ways: all explanatory variables that have appeared in the random part of the model are also in the fixed part; and only the intercept can be considered to be random at level 1. A more complex variance structure will be presented in 4.5.5.

4.5.4.4 Three-level models

In the interest of clarity we introduce the three-level model in the context of a specific problem: a single cross-sectional data set with three level structure consisting of children (level 1) nested within families (level 2) nested within clusters (level 3). Within family j in cluster **k**, a relationship between height (Y) and age (A) might be modelled as follows for the ith child.

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk} A_{ijk} + e_{ijk}$$

The intercepts and slopes in this model are random, varying across families and clusters.

Simple models for the coefficients may be constructed using family and cluster characteristics to account for intercept and slope variation. One such model involving family income variable S and cluster feature R (for example, Urban or Rural) is presented here. First, the intercept for family jk might be expressed as a linear function of the average intercept for the kth cluster, π_{00k}, and S, and a family level random term, u_{0jk}:

$$\beta_{0jk} = \pi_{00k} + \pi_{01k} S_{jk} + u_{0jk}$$

The average intercept for cluster k could be modelled in terms of an overall average intercept (γ_{000}), R, and a cluster-level random term v_{00k}:

$$\pi_{00k} = \gamma_{000} + \gamma_{001} R_k + v_{00k}$$

Likewise the impact of A on Y in family jk might be expressed as a function of S_{jk} and a random component u_{1jk}:

$$\beta_{1jk} = \pi_{10k} + \pi_{11k} S_{jk} + u_{1jk}$$

The cluster level average slope (π_{10k}) could be viewed as the sum of all overall slope (γ_{100}) and a level 3 random variable (v_{10k}); i.e.,

$$\pi_{10k} = \gamma_{100} + v_{10k}$$

The effects of S on the between-family intercept and slope might be constant across all clusters.

$$\pi_{01k} = \gamma_{010}$$
$$\pi_{11k} = \gamma_{110}$$

The components of the model can be combined to produce the following single equation:

$$Y_{ijk} = \gamma_{000} + \gamma_{100} A_{ijk} + \gamma_{010} S_{jk} + \gamma_{001} R_k + \gamma_{110} A_{ijk} S_{jk} +$$
$$[u_{0jk} + v_{00k} + u_{ijk} A_{ijk} + v_{10k} A_{ijk} + e_{ijk}]$$

The terms on the first line belong to *fixed part* and those in brackets belong to *random part*.

Three-level models are often constructed by combining two-level models. In our study the three level models for each province are later combined to create a four level model.

4.5.5 Complex variance structure

So far we have assumed that a single variance describes the random variation at level 1. At level 2 we have introduced more complex variance structure (4.5.4.3). Now we look at how we can model the variation at level 1 explicitly as a function of explanatory variables. The same principles apply to higher levels. Such models are particularly relevant to the modelling of growth. Since the generalization to higher level model is similar, the idea will be presented in a two-level model.

4.5.5.1 A two-level model with complex level 1 variance

The symbol **W** will be used to denote the matrix of all explanatory variables that are used in the fixed part of the model. In equation 17, $\mathbf{W}=\mathbf{XZ}$. The matrix of explanatory variables whose coefficients are random at level 2 will be written as $\mathbf{X}_{(2)}$, and $\mathbf{X}_{(1)}$ will be a matrix of explanatory variables whose coefficient are random at level 1. The sets of variables in $\mathbf{X}_{(2)}$ and $\mathbf{X}_{(1)}$ will often intersect. For example, both will usually contain X_0. Further, $\mathbf{X}_{(2)}$ and $\mathbf{X}_{(1)}$ will often be submatrices of $\mathbf{W}$ (Prosser et al., 1991).

In general,

$$Y_{ij} = \mathbf{W}^T{}_{ij}\,\Gamma + \mathbf{X}^T{}_{(2)ij}\,\mathbf{u}_j + \mathbf{X}^T{}_{(1)ij}\,\mathbf{e}_{ij}$$

where $\mathbf{e}_{ij}$ is the vector of level 1 random terms for person i in group j, and $\mathbf{W}^T{}_{ij}$, $\mathbf{X}^T{}_{(2)ij}$, and $\mathbf{X}^T{}_{(1)ij}$ are rows of the three new matrices defined above. For group j,

$$Y_j = W_j\Gamma + X_{(2)}u_j + \begin{pmatrix} X^T{}_{(1)1j} & 0 & \cdots & 0 \\ 0 & X^T{}_{(2)2j} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X^T{}_{(1)n_jj} \end{pmatrix} \begin{pmatrix} e_{1j} \\ e_{2j} \\ \vdots \\ e_{n_jj} \end{pmatrix}$$

i.e. $$Y_j = W_j\Gamma + X_{(2)j}u_j + X_{(1)j}e_j$$

Here, $\mathbf{u}_j$ is a vector whose length is r_2, the number of explanatory variables whose coefficients are random at level 2, $\mathbf{e}_j$ is a vector of length n_jr_1, where r_1 is the number of explanatory variables with coefficients random at level 1. The fixed part matrix $\mathbf{W}_j$ is $n_j \times q$, $\mathbf{X}_{(2)j}$ is $n_j \times r_2$, and $\mathbf{X}_{(1)j} = \oplus_{i=1}^{n_j} \mathbf{X}^T{}_{(1)ij}$, where $\oplus$ denotes the 'direct' sum of the individual $\mathbf{X}^T{}_{(1)ij}$ row vectors. $\mathbf{X}_{(1)j}$ has dimensions $n_j \times n_jr_1$. The random parameters to be estimated are the elements of $\Omega_{(2)}$ (=Var $\mathbf{u}_j$) and $\Omega_{(1)}$ (=Var $\mathbf{e}_{ij}$). Thus the variance of $\mathbf{Y}_j$ is

$$\Sigma_j = \mathbf{X}_{(2)j}\, \Omega_{(2)j}\, \mathbf{X}^T_{(2)j} + \mathbf{X}_{(1)j}(\oplus_{i=1}^{n_j}\, \Omega_{(1)})\mathbf{X}^T_{(1)j}$$

and Σ, the variance of $\mathbf{Y}$, is a block diagonal matrix composed of the Σ_js. For the cases where the level 1 residuals are not independent (repeated measures) the idea of modelling residuals is described in Appendix B.

4.5.6 Modelling residual variation and growth chart construction

In our review of methods of chart construction we have shown that after proper modelling of the mean, modelling standard deviation is very crucial to obtain the appropriate growth chart. In multilevel modelling not only all suggested techniques of curve fitting such as fractional polynomials can be implemented but also standard deviation can be modelled appropriately. This can be done by suitable modelling of variation at different levels such as: variation between family within the clusters, variation between clusters within the provinces, and also, modelling variation of growth within the family. There are many advantages with this approach. For example, the underlying variation due to each level can be studied better. Second, the residuals at any level of hierarchy can be modelled with linear or nonlinear models which allows great flexibility in modelling variation in terms of age. Third, having the possibility of modelling of variation related to different sources in the data should make it possible to have a more efficient estimate of changing variation of growth over time.

4.5.7 Models for longitudinal growth data

Multilevel modelling is a method for analysing longitudinal observations (such as longitudinal growth study) which copes both with missing data and observations recorded at different times (ages), and allows other occasion-related factors to be evaluated. In relation to growth data a brief explanation of modelling longitudinal data is provided in Appendix B.

4.5.8 Estimation of parameters in the multilevel models

The Iterative Generalized Least Squares (IGLS) algorithm has been proposed for estimating the parameters of the models (Goldstein, 1987). Goldstein has shown that when distribution of the random terms follows a Multivariate Normal distribution, IGLS estimates are maximum likelihood. Under Normality assumption for the random terms, the log-likelihood function for the Γ and Σ given Y is:

$$l(\Gamma,\Sigma|Y) = (Y-XZ\Gamma)^{T}\,\Sigma^{-1}\,(Y-XZ\Gamma)$$

which is minimized using the IGLS algorithm. Fuller discussion can be found in Goldstein (1995).

Maximum likelihood estimation involving many parameters may give seriously biased results. This problem can be overcome by the use of restricted maximum likelihood methods. In the same way the IGLS algorithm is readily modified to produce the similar restricted estimates (RIGLS) (Goldstein, 1989) which are less open to bias. Because our data are very large these problems do not arise in the present context. Other estimation procedures regarding hierarchial data are discussed in Appendix B.

4.6.1 Fractional polynomials

In all these models the age related functions are usually modelled by polynomials, typically quadratic, or cubic. It has been recognized that conventional low order polynomials, which offer only a few curve shapes, do not always fit the data well. High order polynomials follow the data more closely but often fit badly at the extremes of the observed age range. Various attempts have been made to devise more acceptable models, for example, splining as described in 4.3.2. Using a large number of join points would lead to a localized smoothing procedure. Nonparametric scatter plot smoothers are an attempt to 'let the data show us the appropriate functional form' (Hastie and Tibshirani, 1990) rather than imposing a limited range of forms on the data. Typically, the smoother is constructed at each data point in turn by weighted regression within a neighbourhood of the corresponding covariate value.

The cubic spline may be seen as the link between conventional polynomials and the modern methods of nonparametric smoothing. Nonparametric and spline smoothers are powerful and flexible tools which indeed impose few limitations on the functional form. Although the methods are successful for describing data, a major drawback is that, since they use local models, they do not yield simple equations for prediction. Royston and Altman (1994) therefore extended the traditional polynomial curves with a set of curves which they call fractional polynomials, **fp**, whose power terms are restricted to a small predefined set of integer and non-integers. The powers include conventional polynomials as a subset of the family.

4.6.2 Definition

Royston and Altman propose to model the relationship between a response y and positive[2] observations x by

1: If non-positive values of X can occur, a preliminary transformation of X to ensure positivity is needed. One solution is to choose a non-zero origin τ<X and to rewrite definition X in (18) as X-τ. A simple choice of τ is minus the rounding interval of sample values of X.

$$y=\sum_{j=0}^{m}\beta_j H_j(X) \tag{18}$$

where $H_0(X)=1$, and for $j=1,\ldots,m$

$$H_j(X)=\begin{cases} X^{(p_j)} & if\ p_j\neq p_{j-1}, \\ H_{j-1}(X)\ln X & if\ p_j=p_{j-1}. \end{cases} \tag{19}$$

The β_j are real-valued coefficients, and the round bracket notation signifies the Box-Tidwell transformation,

$$X^{(p_j)}=\begin{cases} X^{p_j} & if\ p_j\neq 0 \\ \ln X & if\ p_j=0, \end{cases}$$

with $p_j=j$ restricted to the set $\Omega=\{-2, -1, -0.5, 0, 0.5, 1, 2, \ldots, \max(3,m)\}$. Expression (18) and (19) are full definition of a fractional polynomial of degree m. So, a specific fp of degree m ($\phi_m(\mathbf{p})$) is defined by its set of coefficients p. For example, $\phi_5(-0.5, 0, 1, 2, 2)$ will include a constant term ($H_0=1$) and terms , $H_1=X^{-1/2}$, $H_2=\ln X$, $H_3=X$, $H_4=X^2$ and $H_5=X^2(\ln X)$. Extending polynomial in this way introduces a wide range of potential curves. In particular the family includes curves with asymptotic properties, which are lacking from the traditional polynomial family.

4.6.3 Choice of the model

Royston and Altman suggest fractional polynomials are fitted in much the same way as ordinary polynomials are fitted, i.e., by starting with the simplest models, and building up more complex terms. Models involving different sets of p's, which are of the same order are compared by comparing the reduction in deviance, or 'gain' with the χ^2 distribution on m degrees of freedom. When comparing the best fitting model of order m with the best of order m+1, Royston and Altman suggest that the change in deviance should be compared to a χ^2_1 distribution. Fractional polynomials can be fitted step by step with any statistical spread sheet, but macros are available in STATA which automate the procedure. In this thesis fractional polynomials are used whenever appropriate.

CHAPTER FIVE

GROWTH PATTERNS OF CHILDREN AND ADOLESCENTS IN IRAN

5.1 - Introduction

Chapter two describes the major role that growth monitoring can play in preventive as well as curative medicine especially in developing countries. Growth studies have a relatively long history in western countries and were well known to scientists in the nineteenth century (Eveleth and Tanner, 1990). The earliest study on human growth in Iran dates back to 1966. Most of the early studies are based on a few large cities such as Tehran and Shiraz, and in total the number of them is not more than fifty (Ayatollahi, 1993a). Apart from Ayatollahi's survey of six to twelve year old children in Shiraz in 1988-9, almost all studies related to small selected groups of clinic attenders. None of these studies provide any reliable information regarding growth patterns of children in different provinces of the country. This research is a national study of growth in Iran.

In this chapter, analysis of growth of children and adolescents 2-18 years old using measurements on weight and height are shown across provinces of Iran. Multilevel models are constructed in order to take account of the data structure, and also a comparison of pooled estimation across 3-level models in provinces with a 4-level analysis is presented.

5.2 - Materials

It was stated that in the National Health Survey of 1990-2 in Iran, 10660 families were selected using a random cluster selection of families (chapter one). Also, in chapter one some descriptive statistics of the data as general information were presented. Now in relation to the aim of this part of the study of growth patterns of children, some of the other related summary statistics on weight and height of children according to age, sex and place of residence (province and area of residence) are presented. It should be noted that the ensuing analyses are based on the data set relating to subjects with measurements on both height and weight and from which outliers have been excluded using the methodology and procedures described in chapter three.

5.2.1 Description of weight and height differences

5.2.1.1 Sex difference

Table 5.1 and 5.2 present the mean and standard deviation (SD) of weight and height measurements of boys and girls in different ages (years) according to their area of residence. As Table 5.1 shows at the age of two years old boys are on average 0.3 kg heavier than girls and the difference is about 0.5 kg up to the age 10, Figure 5.1a. Then from 11 up to 15 the mean weight of girls is higher than boys. This is due to the early onset of puberty in girls. Then, from 16 and onwards the trend reverses and boys are heavier than girls, and remain heavier. The difference at 18 is about 6 kg. The pattern of height of boys and girls is shown in Figure 5.1b. The difference in height is about 0.5 cm at 2, Table 5.2, and generally boys are not more than 1.5 cm taller up to 10 years old. From 11 up to 13 this trend reverses and girls are on average about 1 cm taller than boys (Figure 5.1b and Table 5.2). However, from 14 and onwards boys grow up taller and steadily remain taller than girls and as Table 5.2 shows the difference is 13.5 cm at 18 years old.

5.2.1.2 Urban and rural difference

Generally, as can be seen in Figure 5.1c urban children are heavier than rural children. The difference in weight is about 0.5 kg at 2 years old and remains the same up to 7, Table 5.1. Then from 8 this difference starts to increase and at 11 is about 2.2 kg, and consistently increases up to 14 when urban children are on average 4.3 kg heavier than rural children. But from 15 this difference diminishes and at age 18 years the difference goes down to 2.3 kg because rural children catch up. The pattern is similar for the height of children. As can be seen in Figure 5.1d and Table 5.2 urban children are about 3 cm taller than rural children at 2 years old and this remains the same up to 8 years. From 9 this difference starts to increase up to 15 years old. Then, as the rural children catch up in growth the gap starts to decrease from 15 and onwards. On the whole, the urban children are growing taller and earlier than rural children.

Table 5.3 and 5.4 present the estimated yearly rate of growth of boys and girls according to area of residence. These rates have been derived from the present cross-sectional survey. The observed pattern of differences between sexes and urban and rural children can easily be quantified by looking at these tables. For example, at 13 the mean height of boys is about 1 cm less than girls, but from 14 and onward the yearly height velocity of boys is greater so they grow faster and become taller than girls. Similarly, for ages 11 to 15, girls are heavier than boys because their yearly weight velocity at ages 11 to 13 is greater than for boys.

The difference in growth of urban and rural children varies between boys and girls. As Tables 5.1 illustrates, and can more easily be seen in Figure 5.2a, for boys the difference in mean weight of urban and rural children is about 0.5 kg at 2 and does not exceed 0.5kg up to the age of

8 but at 9 this difference rises to 1.5 kg and continue to increase steadily up to 14 years when the average difference is 4.4 kg. From 15 since rural children catch up in growth this difference remains about 3.5 and is 4 kg at eighteen years old. The differences between urban and rural girls in the age range of 11 to 15 are bigger than those just described in boys. The greater gap in Figure 5.2b shows this clearly, and is probably due to earlier onset of puberty in urban girls. In other words there is a sex and urban-rural interaction effect. Then, at 16 and 17 the difference between urban and rural girls is 2.9 kg and 2.4 kg respectively, and reduces to 1.2 kg at 18 since the rates of growth of rural girls are greater at 16 and 17 than for their urban counterparts.

5.2.1.3 Provinces difference

The general observed differences in growth patterns of boys and girls and urban and rural children is explored over all the provinces of Iran. Table 5.5 presents the percentage of urban children, average age, and averages of weight and height for both sexes in each province of Iran. We saw that there are differences in growth patterns of urban and rural children but it can be seen from the table (5.5) that the percentage of urban area varies across the provinces of the country; the urban percentage varies from 21% in Kohkiloyeh-Boyerahmad (province number 12 in Table

5.5) to 94% in Tehran (province 10) which includes about 25.6% of the urban children of Iran. In other words some of the provinces are more urban than others and some of the provinces are more rural than others. Also, it can be seen from the table that the average size of children varies from province to province although the average age seems fairly similar. In addition, the average of size of children in some provinces, with a smaller percentage of urban children, is larger than in some of the provinces with higher percentages of urbanization. For example, on average the children in province Markazi with 45% urban percentage (Province 22; Table 5.5) are heavier and taller than Tehrani children where the urban percent is 94%. It seems unlikely that these difference are due to the small difference in the averages ages of children. Similarly, on average Khouzestani boys (province 23) are heavier than their Tehrani counterparts although Khouzestan is only 65 percent urbanized, and also Semnani (province 1) and Mazandarani girls (province 14) are taller than Tehrani girls although the urban percent in these provinces are less. Therefore, there are not just urban and rural effects in boys and girls weight or height differences, there may be differences in growth patterns from province to province.

It should be mentioned that so far the general exploration on growth patterns of children in Iran was based on the comparisons of the mean of weight and height. However, in monitoring growth of children the lower centiles of the underlying measurements are of more interest. This will be considered later. Also, the general preliminary investigations presented above did not account for the existing structure in the data due to cluster sampling across the provinces of the country. Therefore, the next section gives an account on the structure of the data, so that this may be taken into account when estimating the parameters of the models of the data.

5.3 - Structure of the data

The structure of the data is one of the most important aspect that should be accounted for in the analysis (4.5). Section 1.2.3. describes that in each province the updated lists of the families were used as the sampling framework, and that on the basis of these lists a random cluster selection of families was carried out each containing seven households. In each household the health indicators of the individuals were studied including weight and height of children and adolescents aged two to eighteen years old. Therefore our data has a hierarchial structure with four levels of variation. The highest level is provinces as the fourth level, followed by clusters within the provinces, families in clusters and children in families which are levels three, two and one respectively.

The following diagram presents the scheme of variation corresponding to the levels of nested membership in the structure of our data.

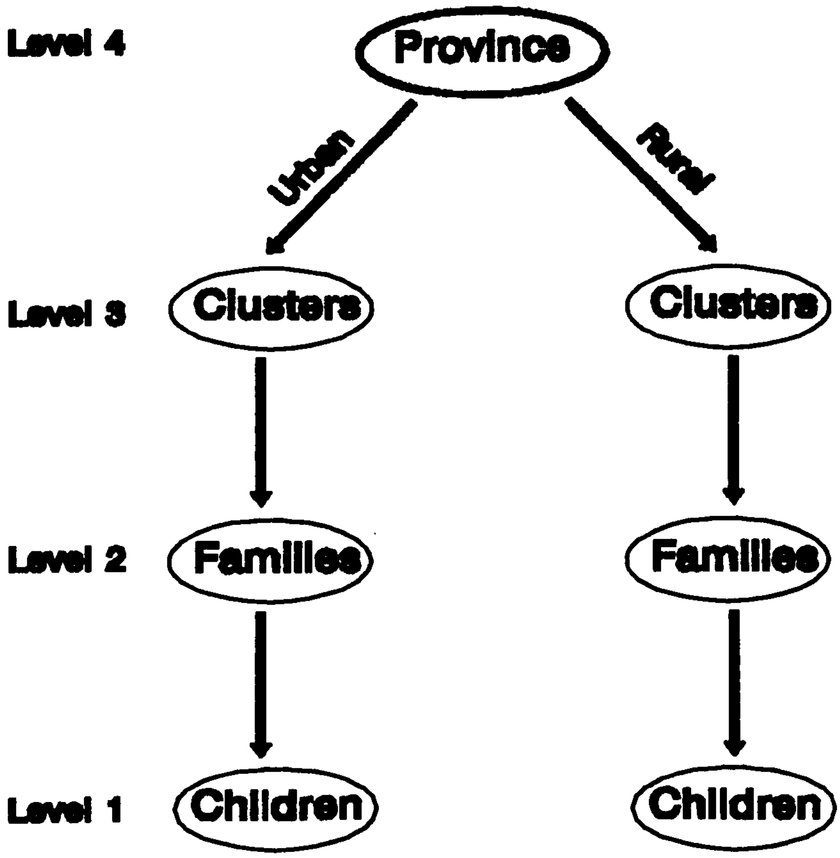


Figure 5.3 Hierarchial structure of the data

One possible way to study the difference between provinces, is to look at them as the fixed differences, say each province has its own effect on the growth pattern of children. But this brings some technical inconvenience, for instance, for a preliminarily investigation of province effect in any model we have to contribute 23 dummy variables corresponding to the effect of provinces. Also, any consideration of a province effect on growth with age, sex and so on needs a huge number of terms to be entered in the model for introducing the corresponding interactions. This is not only unwise but even with powerful computers, because of the magnitude of the data, the model construction may be impossible. And the results could be very difficult to present and interpret. Hence, as the number of the provinces was enough large, it was decided to look at the province differences as a random part, and try to model underlying variation. Then by looking at the structure of the data as levels of variation, it is possible to investigate different complex models by considering the effect of the corresponding variables in the random part of the model.

5.4 - Results of preliminary analysis

5.4.1 Selection of data for preliminary analysis

Inspection of the data and examination of the centiles 3rd, 10th,...,97th of weight and log(weight) showed that increments of log(weight) centiles with age were approximately linear up to the age fifteen, Figure 5.4a. After the age of fifteen centile curves begin to flatten off. This pattern was observed to be similar for both sexes. Also over the age range two to fifteen years the percentiles variation is fairly uniform. Some slight curvatures, especially in the height centiles Figure 5.4b, suggests the need for at least a quadratic component of age in modelling of the age-related centiles. Similar patterns were found when data from several of the provinces were examined separately.

Based on these findings it was decided that for a general preliminary study of growth patterns across provinces of Iran, the analysis would be carried out first in the wide age band of two to fifteen years old. Then, because at the time of this analysis (Spring 1994) only ML3E was available, and second as we were interested in exploring and comparing how the growth pattern varies within and across the provinces of the country, three-level analyses on weight and height of children were carried out for each province separately. Finally, a summary for the fourth level (provinces) was established by looking at the results of these analyses together.

5.4.2 - Choice of the model and centring

For the study of weight and height in different provinces of Iran and comparison of them, a basic multilevel model with three levels, clusters, families and children as levels 3, 2 and 1 was developed for each province. The following equation shows the fixed part of model that was tried in the three-level analyses of weight and height in all provinces:

$$\frac{LWT}{HT}\Big\} = \beta_0 Cons + \beta_1 Cage + \beta_2 SEX + \beta_3 UR + \beta_4 Cage^2 + \beta_5 S*CA + \beta_6 S*UR + \beta_7 UR*CA$$

(For simplicity the indices are avoided)

where LWT is the '100×log(weight)'. The logarithmic transformation of weight is a reasonable transformation to Normality for present purposes and is multiplied by 100 to avoid decimals. So, the results in Table 5.7 are in this scale. The other terms in the model are as bellow:

Cons : *Constant*
Cage : *Centred age (Age-8)*
Cage2 : *Cage squared*
UR : *Urban or Rural (Urban=0 , Rural=1)*
Sex : *Boys=0, Girls=1;*

the interaction terms are

S*CA : *Sex*Cage*
S*UR : *Sex*(Urban or Rural)*
UR*CA : *(Urban or Rural)*Cage*

The above model was selected for studying the general pattern of growth in provinces for the following reasons: first, because the pattern of change of the centiles of log(weight) and also height up to age of fifteen was observed to be approximately linear. The possibility of curvilinear change however, was also considered by looking at the quadratic term in the models; second based on a review of previous growth studies in Iran (Ayatollahi 1993a); third, to be able to examine the difference in growth pattern in relation to gender, place of residence (urban or rural) and to investigate how these variations interact with each other and how they change with age.

In the random part of these three-level models we explored whether the regression on age varied with the cluster or with the family. As Table 5.6 shows neither level 3 nor level 2 variation was found to be generally significant (p=0.20 and p=0.09 respectively). The variation in differences between boys and girls' growth at cluster level was similarly examined but was not significant (p=0.34). A random sex effect was not examined at family level because there were too few children in each family. Also random variation in the sex difference in growth of children with age was found not to be significant across the clusters (p=0.28). Thus the results of these investigations of the random variation of parameters of interest showed that they were not significant. So a simple structure ('Cons'/ 'Cons') was adapted for each level of the random part of the three-level models used for the provincial analysis.

For technical purposes the age was centred at eight (Cage=Age-8). It should be noted that theoretically there are advantages with centring the independent variables in regression procedure. For example, in simple case of linear regression, the estimated intercept and slope are negatively correlated if the mean of the independent variable is positive. But centring around the mean of independent variables produces independent estimates. Also, the estimate of intercept and its variance when the origin of the independent variable is outside the range of the data is not useful and can be misleading (in our data the children are 2 years old or older). Also, centring helps to insure numerical stability in estimating procedures.

Bryk and Rudenbush (1992) argue that centring explanatory variables will often be a sensible way of specifying multilevel models. They discussed the possibilities for different locations for the centre of **X** and conclude that in general, a reasonable choice of location depends upon the purpose of the research. It is often useful to centre the variable **X** around the grand mean. For example, in the case of one-way ANCOVA with random effects

$$Y_{ij}=\beta_{0j}+\beta_{1j}(X_{ij}-\bar{X}_{..})+r_{ij}; \qquad \begin{cases}\beta_{0j}=\gamma_{00}+U_{0j}\\ \beta_{1j}=\gamma_{10}\end{cases}$$

the estimate of intercepts is mean outcome for each group j adjusted for differences among these units in X_{ij}. This is the standard location for X_{ij} in the classical ANCOVA model. Similarly, the var(β_{0j}) is the between group variance of group means adjusted to $\bar{X}_{..}$.

Plewis (1985) argues that researchers should be very wary about their choice of centre. For instance in our study one suggestion might be to centre the data differently for each province, but this would cause inconvenience in comparisons and interpretations. Then, since the mean of age in the age range of 2 to 15 years old was 8.11 (SD=3.82) and the median was 8. The age 8 years was chosen for centring the age in provincial analysis. In the author's opinion centring which was used throughout the ensuing analysis was crucial.

5.4.3 Results of province by province analyses

The coefficients of the models fitted to the data on log(weight)[A] and height from the 24 provinces are shown in Table 5.7 and 5.8. From the results of these analyses the estimated average of weight and height of a child selected from any part of Iran easily can be determined from the corresponding tables. For example, one can easily see that the mean weight of eight year old boys in the urban area of province Fars (province number 11) is:

Exp('log(weight)'/100=exp(306/100)=21.33 kg,

and the mean height of these boys is 120.9 cm. And similarly the mean weight of fifteen year old girls in urban area of Tehran (province number 10) is 51.22 kg with the average height of 160.3 cm. Also at the age of eight years old, the average weight of East Azarbaijani (province 3) children is the highest and Ilam (province 20) has the lightest children. Similar interpretations of the results can be derived for different age and sex and place of residence (UR) in any of the provinces. It is vital to keep the location of the centre (8 years) in mind when computing

A: Natural logarithms (base 'e') are used throughout this thesis averages of weight or height for any child across the country.

Also consideration of the coefficients of the models in Tables 5.7 and 5.8 can give us many ideas about how the average log(weight) and height of Iranian children varies across the provinces with age, how it differs between provinces with sexes and place of residence (UR), and how these variations change as the children grow up. For example, generally eight years old boys are heavier and taller than girls across almost all the provinces of Iran (the coefficients of 'Sex' are negative). Similarly at the age of eight years, children living in rural areas are lighter and shorter than their urban counterparts (the coefficients of 'UR' are negative). Also, this difference in height and log(weight) is grater for girls than boys (since the coefficients of 'S*UR' are generally negative).

Then, since ML3E was only able to analyze the data up to three-levels, the estimates of four-level analysis can be found by looking at the results of the 3-level modelling of weight and height in the 24 provinces. So a weighted analysis of the coefficients of the models was carried out. The weighted average of the coefficients was computed taking $w_i=\frac{1}{(s.e)^2}$ as weights. Then

$\bar{b}_j=\frac{\sum_{i=1}^{n} w_i b_{ij}}{\sum_{i=1}^{n} w_i}$; where i represents the provinces: i=1 to 24.

The estimated coefficients of the fixed part of the model for four-level analyses of log(weight) and height is shown in Table 5.9. Since these coefficients are averaged over all the provinces they are the national statistics for Iran. For example, it can be seen from the table that log(weight) and height difference between boys and girls changes significantly with age ('S*CA'; $p<0.001$). Also, log(weight) and height difference between urban and rural children changes significantly as children grow up ('UR*CA'; $p<0.001$). In addition, the log(weight) and height differences between boys and girls and urban and rural children described by looking at the coefficients of the provincial models in Tables 5.7 and 5.8 are confirmed analytically by the above analysis, because the terms 'Sex', 'UR' and 'S*UR' are found to be highly significant, Table 5.9; $p<0.001$.

Furthermore, the variation of the coefficients in the models across the provinces of Iran were studied by looking at their heterogeneity between provinces. The heterogeneity of the coefficients across the provinces were examined by using the statistic:

$$Q = \sum_{i=1}^{n} w_i (b_{ij}-\bar{b}_j)^2$$

which follows $\chi^2_{(n-1)}$ under assumption of homogeneity of b_{ij} (coefficients of provincial regression models) (Armitage and Berry 1988). Table 5.10 presents the results of the tests of heterogeneity of the coefficients of modelling log(weight) and height across the provinces. For example, the average log(weight) and height of eight years old children ('Cons') varies significantly across the provinces ($p<0.001$). Also at the age of eight, rate of growth in height and log(weight) varies across the provinces of Iran ('Cage'; $p<0.001$). In addition, the difference in rate of growth of children in urban and rural areas varies from province to province as they grow up ('UR*CA'; $p<0.001$). Similarly, as Table 5.10 shows the log(weight) difference of boys and girls varies across the provinces as they grow up ('S*CA'; $p<0.001$) but not their height. According to Table 5.10 differences of log(weight) for eight year old urban and rural children do not vary significantly across the provinces, but their height difference varies significantly varies from province to province. A more detailed comparison with a four-level analysis which was carried out after the release of MLn is discussed later.

Investigation of the correlation structure of the coefficients resulting from the province analysis of log(weight) and height provide an interesting view of the growth pattern of children and adolescents in the country. Information on child mortality rates (q_{1-4}) in Iran which is an indicator of health status of children was available from previous research (Hosseini, 1988), and was used for further analysis.

The Pearson's coefficients of correlation of the major heterogeneous coefficients ('Cons' and 'Cage') resulting from modelling growth in the different provinces of Iran are shown in Table 5.11. As one can see at age of eight, the slope of log(weight) is significantly negatively correlated with average weight ($r=-0.70$, $p<0.001$) showing that generally across the provinces of Iran heavier boys are growing up slower than their lighter counterparts. The corresponding correlations for urban girls, rural boys, and rural girls were -0.47, -0.69, and -0.69 respectively ($p<0.05$ for all these three correlation coefficients). This shows that this finding on weight is similar for all children all over provinces in Iran. Also, the average log(weight) is positively correlated to average height and height development ($r=0.34$, $r=0.23$) showing that heavier children tend to be taller and gain height faster. In addition, the slope of log(weight) (Cage-Wt) of eight year old boys is positively correlated to the development of height (Cage-Ht) and average height ($r=0.25$, $r=0.18$). Also the mortality rate is lower across the provinces where the height and weight development is higher.

These results presented above can be visually inspected by looking at Figure 5.5. For example, for the case of boys' weight, generally points are along a negative trend showing that provinces where the boys are on average heavier at eight ('Cons') tend to have smaller rate of growth ('Cage-Wt'). In this figure (5.5a) the extreme point in the right corner of the graph represent the province Ilam (20) where children are light but put on weight fast; heavier children in Kohkiloyeh-Boyerahmad (12) or Semnani children (1) tend to put on weight slower. In contrast, 'Cons-Ht' and 'Cage-Ht' are weakly positively correlated, Table 5.11 and Figure 5.5b, where fairly tall children in Semnan (1) tend to grow faster than their shorter counterparts in Kohkiloyeh-Boyerahmad who are growing slower. Relationships between these variables are considered in more detail in the next chapter where possible provincial grouping is explored.

5.5 - Discussion

The estimates of the growth parameters presented for whole country obtained by weighted averaging the results of the provincial analysis provided a general picture of growth variation across the country which is reasonable for a preliminary investigation of the data. This analysis was a crude approach and more appropriate estimates might be obtained by looking at the whole data in four-level analysis[B].

B: In some of the preliminary analysis by combining family and children level, three-level analysis of whole data was tried and generally similar results obtained using ML3E

5.5.1 - Four-level analysis of the data

After MLn (1995) was released, it was possible to investigate in more detail the four-level variation by looking at the data all together and to compare the results with the previous work. Hence, a four-level analysis on weight data using MLn with provinces, clusters, families, and children as levels 4, 3, 2, and 1 respectively was carried out. The estimates of the fixed part of the model were similar to those found in the weighted average analysis. A comparison of findings in Tables 5.12 and 5.9 shows that all estimates are in a one s.e range, or at most less than 2*s.e apart. One may notice that intercept estimates from two approach (306.90 (1.21), and 308.44 (0.34); Tables 5.12 and 5.9) are not significantly different, but these estimates are not as close as the others. This is due to the point that the 4-level model exploits the fact that the coefficients we are estimating (for example, 'Cons') have a distribution across the provinces plus a random component associated with lower level random variation. The provincial means are stripped of this lower level variation and thereby shrunk toward the grand mean. Thus these coefficients are drawn towards the average values along the trajectory given by their distribution across provinces. But in provincial analysis the cluster and family estimates are shrunken towards the provincial mean. Therefore, the separate province estimates are more scattered (to be discussed in 5.5.4) than those found in 4-level analysis. So, the weighted average may be a bit different from the pooled analysis.

Then, in order to compare the result of modelling the random part with those found in tests of heterogeneity, we looked at the variation of growth parameters (the constant term, slope, and the quadratic term in age and so on) at province level. The result of this modelling can be seen in Table 5.12 and compared with previous results presented in Table 5.10. This table (5.12) shows some terms which were significant (heterogeneous) in the pooled analysis, Table 5.10, are no longer significant here. For instance, the difference in log(weight) of urban and rural children as they grow up ('UR*CA') was not found to vary significantly across the provinces ($p \approx 0.9$). Or log(weight) difference between boys and girls in urban and rural areas ('S*UR') are no longer significantly different from province to province ($p=0.16$). So by looking at all the subjects together some of the apparent differences between provinces appear to be artifacts due to estimating the coefficients independently. Differences become non significant when we looked at the covariance structure of the data more appropriately. In relation to this part, it should be noted that so far it is not possible to look at more than 6 variables in the random part in levels higher than three. This is one of the bugs in MLn that needs to be sorted out.

Generally the results in the four-level analysis is adjusted for the effect of other levels and is shrunken. But this is not the case when we are looking at the variation in the level of interest by the result of the test of homogeneity or by averaging over provinces. For example, one way of looking at the variation of the constant term (intercept) of the models across provinces, is to look at the variance of the difference of province estimate and the estimated average. This gives the value

$$\frac{\sum_{j=0}^{24}(b_j-\mu)^2}{23}=35.42; \qquad \mu=\bar{b}_j, \qquad \begin{cases}\bar{b}_j \text{ weighted average of } Cons_j \\ j\text{: province}\end{cases}$$

which is bigger than 30.26 which resulted in four-level analysis, Table 5.12. One can compare the situation with the simple case of one way ANOVA with random effect where the expectation of the mean squared error due to the subject is the summation of residual variance and the subjects variance. So the estimated variance with this approach in fact includes the extra variation. It is an approximation, and in our case it includes variation due to the other levels. But is a fairly close estimate.

Besides, since the standard deviation of the logarithmic transformation is the coefficient of variation (C.V) of the untransformed data, the variation in weight of children due to provinces, clusters, and families can be determined from estimated variance of the 'Cons' terms in different levels of the model. For instance, the variation in weight due to variation between the province, clusters and families are 6.6%, 8.6%, and 14.8% (Table 5.12).

Our previous study of the correlation of the models showed that the average weight ('Cons') and the rate of growth ('Cage') are negatively correlated. This can also be seen in Table 5.12 where the correlation of *'Cons'/ 'Cage'* (correlation of intercept and slope at eight) at level 4 is -0.85, showing that the small children in some of the provinces are putting on weight faster than their heavier counterparts in another provinces. The difference in the estimates (Table 5.12, r=-0.85; Table 5.11, r=-0.70) is due to the shrinkage of the provincial constant and slope estimates towards the overall mean which reduces the corresponding between province variances.

5.5.2 - Investigation of the assumptions

The reasonableness of the assumptions of the four level model were examined by looking at the corresponding residuals at different levels. Residual diagnostics for the model are shown in Figure 5.6. Normal plots (a)-(d) of the standardized residuals correspond to different levels and assess the assumptions of Normality. These are fairly straight lines. Also, the independence of error terms in different levels can be investigated by looking at the scatter plots of the residuals. Scatter plots (a)-(f) of the levels against each other in Figure 5.7 show no evidence against the assumption of independence of random terms at different levels. Amongst 1497 clusters (level 3) which were observed in this age band only the residual of one of the clusters was large. Checking the individuals' measurements of this cluster did not show them to be outliers. The corresponding residual is not shown in Figure 5.6c.

5.5.3 - Fixed provincial difference

As it is stated because of the flexibility of the underlying models in MLn and also the fact that the number of provinces were large enough, the provincial difference in our study has been considered as a random part of the model. It should also be mentioned that incorporating separate terms for modelling variation between provinces is inefficient because it involves estimating many times more coefficients than the multilevel procedures, and is also unnecessary since we can when necessary obtain estimates from the multilevel analysis. Besides, since the model uses the data from all children and provinces, the estimates will be more precise especially for provinces with smaller sample size.

To see the difference between the estimates for each province from a random effect model and a fixed effects model a model was fitted to the log(weight) data, but instead of including a random component level 4, constant were fitted for each province. In this 3-level model clusters, families, and children were as level three, two and one respectively. Table 5.13 shows the result of this modelling in the estimation of 'Cons' (average **LWT** at eight) in comparison with those estimates obtained for each province from the 4-level analysis. Column I is the results obtained by adding base line and the fixed province difference, column II is the result of adding up the fixed part of the model with the corresponding level 4 residual. It can be seen that the two approaches produce very similar results which confirms that our strategy of setting province variation as random at level 4 is reasonable. Also, note that the constants estimated from the random effect model are nearer the overall mean of 306.90 than the estimates in which these terms are fixed effects, this is the effect of shrinkage which in data of this size is comparatively small.

5.5.4 Provincial variation in growth pattern of children

Since the estimates from the 4-level analysis are more efficient. In this part the investigation of the difference between provinces are carried out on the basis of estimates obtained from the residuals of the random part derived from 4th (province) level. Hence, the provincial estimates of intercepts and slopes of growth and quadratic term in models of log(weight) and height were estimated from the corresponding models.

Figure 5.8 presents the weight and height development of urban boys in different provinces of Iran. As one can see in Figure 5.8a lighter children in provinces Ilam (green line) and Zanjan (brown) are growing up faster, and from 11 and onwards they start to catch up and at fifteen their sizes are in the middle range of variation of weight. Also in this figure, pink line shows the development of weight of urban Tehrani boys which are generally in the middle of the range of variation between provinces. It is worthwhile mentioning that if these line were drawn from provincial analyses they would have been more scattered. The multilevel model pulls the coefficients for the provinces towards the average coefficient values estimated by the fixed part of the model.

The variation of height attainments of urban boys are shown in Figure 5.8b. Smaller children in some provinces such as Kohkiloyeh-Boyerahmad (brown line) are not growing as fast as their taller counterparts in other provinces. It can be seen from the figure that on average Tehrani boys are taller than average (pink line). But how much taller and how these differences matters in practice and also how large these differences are if we look at lower centiles is discussed in chapter six. In addition, how these differences can be addressed is described in chapter eight.

Table 5.1 Mean and standard deviation (SD) of weight (kg), and the number of children by sex, age and urban-rural status; National Health Survey 1990-2, Iran

Age years*	Boys			Girls			Total		
	U	R	Total	U	R	Total	U	R	Total
2	11.73 (2.18) 343	11.28 (2.30) 270	11.53 (2.24) 613	11.47 (1.96) 286	10.94 (2.48) 271	11.21 (2.24) 557	11.61 (2.09) 629	11.11 (2.39) 541	11.38 (2.25) 1170
3	12.96 (2.18) 431	12.72 2.61 373	12.85 (2.59) 804	12.34 (2.23) 383	12.26 (2.52) 371	12.30 (2.38) 754	12.67 (2.43) 814	12.49 (2.58) 744	12.58 (2.50) 1558
4	14.57 (2.72) 457	14.55 (2.82) 394	14.56 (2.77) 851	14.43 (2.70) 457	13.66 (2.72) 358	14.10 (2.73) 815	14.50 (2.71) 914	14.13 (2.81) 752	14.33 (2.76) 1666
5	16.34 (2.77) 475	16.00 (3.00) 388	16.18 (2.88) 863	15.71 (2.77) 469	15.19 (2.59) 355	15.49 (2.70) 824	16.03 (2.79) 944	15.61 (2.84) 743	15.84 (2.82) 1687
6	17.93 (3.18) 462	17.68 (3.33) 363	17.82 (3.25) 825	17.44 (3.15) 399	16.75 (3.22) 314	17.13 (3.19) 713	17.70 (3.17) 861	17.25 (3.31) 677	17.50 (3.24) 1538
7	19.61 (3.53) 518	19.60 (3.59) 365	19.61 (3.55) 883	19.49 (3.43) 462	18.64 (3.45) 340	19.13 (3.46) 802	19.55 (3.48) 980	19.13 (3.55) 705	19.38 (3.51) 1685
8	22.00 (3.78) 459	21.37 (3.61) 352	21.72 (3.72) 811	21.78 (4.03) 448	20.26 (3.34) 344	21.12 (3.82) 792	21.89 (3.91) 907	20.82 (3.52) 696	21.42 (3.78) 1603
9	24.70 (4.08) 409	23.16 (3.91) 305	24.05 (4.07) 714	23.78 (4.48) 448	22.68 (4.40) 314	23.33 (4.47) 762	24.22 (4.31) 857	22.92 (4.17) 619	23.67 (4.30) 1476
10	27.52 (5.58) 433	25.98 (4.50) 316	26.87 (5.20) 749	26.73 (5.32) 431	25.50 (5.00) 323	26.21 (5.22) 754	27.13 (5.46) 864	25.74 (4.76) 639	26.54 (5.22) 1503
11	30.00 (5.45) 399	28.33 (4.65) 332	29.23 (5.17) 731	30.82 (7.13) 392	28.10 (6.00) 306	29.62 (6.79) 698	30.40 (6.35) 791	28.22 (5.33) 638	29.42 (6.02) 1429
12	33.43 (6.67) 361	31.00 (5.78) 315	32.30 (6.38) 676	35.61 (8.73) 341	32.03 (7.70) 289	33.97 (8.46) 630	34.49 (7.81) 702	31.50 (6.78) 604	33.11 (7.50) 1306
13	37.50 (8.03) 317	34.30 (6.69) 276	36.01 (7.60) 593	40.65 (8.79) 322	36.46 (8.19) 289	38.67 (8.76) 611	39.10 (8.56) 639	35.40 (7.57) 565	37.36 (8.31) 1204
14	43.39 (9.40) 296	38.97 (7.88) 237	41.42 (9.02) 533	45.03 (8.36) 289	40.74 (7.99) 237	43.10 (8.46) 526	44.20 (8.93) 585	39.86 (7.97) 474	42.26 (8.78) 1059
15	48.87 (9.77) 248	45.46 (10.5) 214	47.29 (10.20) 462	49.31 (9.00) 299	45.45 (8.25) 262	47.51 (8.86) 561	49.11 (9.35) 547	45.46 (9.31) 476	47.41 (9.50) 1023
16	53.78 (9.99) 215	50.31 (8.82) 197	52.12 (9.59) 412	50.27 (8.67) 261	47.38 (7.52) 258	48.83 (8.24) 519	51.86 (9.44) 476	48.65 (8.23) 455	50.29 (9.01) 931
17	56.49 (9.05) 203	53.06 (8.81) 155	55.01 (9.10) 358	52.67 (8.94) 255	50.31 (8.56) 218	51.59 (8.83) 473	54.37 (9.18) 458	51.46 (8.76) 373	53.06 (9.10) 831
18	60.68 (10.1) 149	56.77 (7.94) 132	58.84 (9.33) 281	52.55 (8.72) 212	51.38 (8.63) 187	52.00 (8.69) 399	55.91 (10.1) 361	53.61 (8.75) 319	54.83 (9.57) 680

* In all tables and figures age refers to age in complete years unless otherwise specified

Table 5.2 Mean and standard deviation (SD) of **height** (cm), and the number of children by sex, age and urban-rural status; National Health Survey 1990-2, Iran

Age years	Boys			Girls			Total		
	U	R	Total	U	R	Total	U	R	Total
2	85.8 (6.7) 343	83.3 (6.6) 270	84.7 (6.8) 613	85.7 (6.0) 286	82.3 (6.9) 271	84.1 (6.7) 557	85.8 (6.4) 629	82.8 (6.8) 541	84.4 (6.7) 1170
3	92.0 (7.8) 431	89.6 (8.0) 373	90.8 (8.0) 804	90.8 (7.3) 383	87.8 (7.9) 371	89.3 (7.7) 754	91.4 (7.6) 814	88.7 (8.0) 744	90.1 (7.9) 1558
4	98.9 (7.7) 457	97.0 (7.7) 394	98.0 (7.8) 851	97.9 (7.9) 457	95.6 (7.5) 358	96.9 (7.8) 815	98.4 (7.8) 914	96.3 (7.6) 752	97.5 (7.8) 1666
5	105.8 (7.9) 475	103.8 (7.4) 388	104.9 (7.7) 863	104.6 (7.2) 469	101.6 (7.5) 355	103.3 (7.5) 824	105.2 (7.6) 944	102.7 (7.5) 743	104.1 (7.6) 1687
6	111.4 (6.4) 462	109.7 (7.2) 363	110.7 (6.8) 825	110.9 (6.8) 399	107.6 (6.8) 314	109.4 (7.0) 713	111.1 (6.6) 861	108.8 (7.1) 677	110.1 (6.9) 1538
7	116.3 (6.8) 518	114.2 (7.0) 365	115.4 (7.0) 883	116.8 (6.7) 462	113.6 (6.7) 340	115.4 (6.9) 802	116.5 (6.8) 980	113.9 (6.8) 705	115.4 (6.9) 1685
8	122.1 (6.9) 459	119.5 (7.0) 352	121.0 (7.0) 811	122.1 (7.0) 448	118.6 (6.7) 344	120.5 (7.1) 792	122.1 (6.9) 907	119.1 (6.9) 696	120.8 (7.1) 1603
9	127.3 (6.8) 409	123.9 (7.5) 305	125.9 (7.3) 714	126.5 (7.0) 448	122.1 (7.6) 314	124.7 (7.5) 762	126.9 (6.8) 857	123.0 (7.6) 619	125.3 (7.4) 1476
10	132.5 (7.2) 433	129.0 (7.4) 316	131.1 (7.5) 749	132.1 (7.2) 431	128.3 (7.3) 323	130.5 (7.5) 754	132.3 (7.2) 864	128.7 (7.3) 639	130.8 (7.5) 1503
11	136.7 (8.3) 399	132.8 (7.4) 332	134.9 (8.2) 731	138.3 (8.5) 392	133.2 (8.4) 306	136.0 (8.8) 698	137.5 (8.4) 791	133.0 (7.9) 638	135.5 (8.5) 1429
12	142.4 (8.6) 361	138.2 (8.9) 315	140.4 (9.0) 676	143.9 (8.4) 341	138.5 (9.6) 289	141.4 (9.3) 630	143.1 (8.5) 702	138.4 (9.2) 604	140.9 (9.2) 1306
13	148.2 (9.3) 317	142.5 (9.2) 276	145.5 (9.6) 593	149.0 (8.6) 322	143.2 (9.0) 289	146.3 (9.2) 611	148.6 (9.0) 639	142.9 (9.1) 565	145.9 (9.4) 1204
14	154.6 (10.1) 296	148.8 (9.5) 237	152.1 (10.2) 533	153.4 (6.8) 289	147.9 (7.8) 237	150.9 (7.7) 526	154.0 (8.6) 585	148.4 (8.7) 474	151.5 (9.1) 1059
15	160.7 (9.8) 248	155.5 (10.6) 214	158.3 (10.5) 462	155.3 (6.6) 299	150.6 (7.3) 262	153.1 (7.3) 561	49.11 (9.4) 547	152.8 (9.3) 476	155.4 (9.2) 1023
16	166.0 (8.7) 215	161.5 (9.4) 197	163.9 (9.3) 412	156.0 (6.6) 261	152.5 (6.4) 258	154.3 (6.7) 519	160.5 (9.1) 476	156.4 (9.0) 455	158.5 (9.3) 931
17	167.4 (8.9) 203	163.9 (8.8) 155	165.9 (9.0) 358	156.4 (5.7) 255	153.9 (6.8) 218	155.2 (6.4) 473	161.3 (9.1) 458	158.0 (9.2) 373	159.8 (9.3) 831
18	170.3 (8.0) 149	167.1 (8.6) 132	168.8 (8.4) 281	156.6 (6.0) 212	153.7 (6.9) 187	155.3 (6.5) 399	162.3 (9.7) 361	159.2 (10.0) 319	160.9 (10.0) 680

Table 5.3 Estimated yearly velocity of ***weight*** (kg/year) of boys and girls by age and urban-rural status; National Health Survey 1990-2, Iran

Average age+	Boys			Girls		
	U	R	Tot	U	R	Tot
3.0	1.23	1.44	1.32	0.87	1.32	1.09
4.0	1.61	1.83	1.71	2.09	1.40	1.80
5.0	1.77	1.45	1.62	1.28	1.53	1.40
6.0	1.59	1.68	1.64	1.73	1.56	1.65
7.0	1.68	1.92	1.79	2.05	1.89	1.99
8.0	2.39	1.77	2.11	2.29	1.62	1.99
9.0	2.70	1.79	2.33	2.00	2.42	2.21
10.0	2.82	2.82	2.82	2.95	2.82	2.88
11.0	2.48	2.35	2.36	4.09	2.60	3.42
12.0	3.43	2.67	3.07	4.79	3.93	4.35
13.0	4.07	3.30	3.71	5.04	4.43	4.69
14.0	5.89	4.67	5.41	4.38	4.28	4.43
15.0	5.48	6.49	5.87	4.28	4.71	4.41
16.0	4.91	4.85	4.83	0.96	1.93	1.32
17.0	2.71	2.75	2.89	2.40	2.93	2.75
18.0	4.19	3.71	3.83	*	1.07	0.42

* : No Estimate reported, because of the artificial negative estimate due to lack of adequate sample.

+ : The mid point between the average age of children in succesive age groups

Table 5.4 Estimated yearly velocity of ***height*** (cm/year) of boys and girls by age and urban-rural status; National Health Survey 1990-2, Iran

Average	Boys			Girls		
age	U	R	Tot	U	R	Tot
3.0	6.2	6.3	6.1	5.1	5.5	5.3
4.0	6.9	7.4	7.2	7.1	7.8	7.6
5.0	6.9	6.8	6.9	6.7	6.0	6.4
6.0	5.6	6.1	5.8	6.3	6.0	6.1
7.0	4.9	4.3	4.7	5.9	6.0	6.0
8.0	5.8	5.3	5.6	5.3	5.0	5.1
9.0	5.2	4.4	4.9	4.5	3.5	4.2
10.0	5.2	5.1	5.2	5.6	6.2	5.8
11.0	4.2	3.8	3.9	6.2	4.9	5.5
12.0	5.7	5.4	5.5	5.6	5.3	5.4
13.0	5.8	4.3	5.1	5.1	4.7	4.9
14.0	6.4	6.3	6.6	4.4	4.7	4.6
15.0	6.1	6.7	6.2	1.9	2.7	2.2
16.0	5.3	6.0	5.6	0.7	1.9	1.2
17.0	1.4	2.4	2.0	0.4	1.4	0.9
18.0	2.9	3.2	2.9	0.2	*	0.1

* : No Estimate reported, because of the artificial negative estimate due to lack of adequate sample.

Table 5.5 Sample size, percent urban, average age, average weight and height by sex across 24 provinces of Iran, National Health Survey 1990-2

Province Code	No	% Urban	Average age (yr.) (SD)	Boys Av*. weight (SD)	Boys Av. height (SD)	Girls Av. weight (SD)	Girls Av. height (SD)
1	159	69	9.37 (4.25)	27.08 (13.89)	126.8 (23.3)	30.08 (12.71)	131.5 (21.3)
2	320	39	8.57 (4.48)	25.15 (12.76)	121.4 (24.9)	24.44 (12.88)	118.7 (24.6)
3	1805	42	9.44 (4.62)	28.46 (13.56)	124.1 (25.2)	30.12 (14.97)	124.9 (24.5)
4	1305	63	8.99 (4.48)	24.14 (13.87)	123.1 (24.8)	25.86 (14.34)	125.4 (22.9)
5	462	49	8.55 (4.58)	24.12 (14.37)	121.6 (24.6)	24.70 (14.51)	121.8 (24.2)
6	813	42	9.27 (4.49)	28.74 (14.42)	124.7 (24.4)	29.94 (14.55)	125.0 (23.4)
7	848	43	8.81 (4.52)	25.10 (14.19)	123.1 (25.4)	24.04 (12.60)	122.0 (23.4)
8	637	50	8.82 (4.59)	25.78 (14.18)	125.5 (25.7)	25.02 (14.51)	122.4 (24.1)
9	635	38	8.46 (4.42)	25.19 (12.64)	120.8 (25.3)	23.90 (12.86)	117.6 (24.9)
10	3525	94	8.96 (4.50)	27.91 (15.05)	127.1 (25.9)	28.13 (15.23)	126.1 (24.3)
11	1561	46	9.32 (4.43)	25.76 (13.22)	123.4 (24.6)	27.05 (13.70)	124.7 (23.2)
12	225	21	8.77 (4.60)	25.88 (12.97)	121.1 (25.0)	23.24 (12.96)	114.3 (25.4)
13	292	52	8.99 (4.47)	25.62 (13.00)	122.5 (24.0)	28.24 (15.10)	124.2 (23.7)
14	1363	42	9.48 (4.60)	27.90 (14.33)	125.8 (25.7)	29.39 (15.53)	126.3 (23.5)
15	2747	51	8.94 (4.62)	25.37 (13.13)	122.0 (24.7)	27.00 (13.79)	123.7 (23.7)
16	708	40	9.49 (4.38)	27.75 (15.82)	126.8 (25.2)	28.29 (15.16)	127.0 (22.9)
17	776	40	9.39 (4.48)	25.08 (14.76)	123.6 (25.2)	26.23 (14.33)	124.9 (22.8)
18	608	49	9.00 (4.64)	25.67 (13.29)	120.4 (24.8)	29.17 (15.34)	124.0 (25.2)
19	327	59	9.24 (4.45)	27.52 (12.69)	126.2 (24.2)	28.32 (15.05)	124.9 (23.9)
20	185	48	9.29 (4.50)	25.63 (15.45)	126.5 (25.9)	26.13 (15.28)	124.9 (22.5)
21	704	37	8.77 (4.47)	23.98 (11.53)	120.1 (24.1)	25.61 (12.69)	121.5 (24.1)
22	485	45	9.52 (4.69)	29.29 (15.57)	128.4 (25.4)	30.39 (15.92)	127.8 (24.8)
23	1449	65	8.87 (4.59)	28.89 (15.71)	127.0 (26.2)	27.98 (14.93)	123.9 (24.7)
24	410	58	9.07 (4.25)	24.52 (11.50)	122.1 (22.7)	25.92 (11.89)	125.6 (21.8)
Total	22349	55	9.07 (4.53)	26.60 (14.15)	124.3 (25.2)	27.44 (14.54)	124.4 (24.0)

Province code:
1- Semnan
2- Chaharmahal-Bakhtiari
3- East Azarbaijan
4- Isfahan
5- Kordestan
6- West Azarbaijan
7- Hamadan
8- Bakhtaran
9- Kerman
10- Tehran
11- Fars
12- Kohkiloyeh-Boyerahmad
13- Boushehr
14- Mazandaran
15- Khorasan
16- Gilan
17- Zanjan
18- Lorestan
19- Yazd
20- Ilam
21- Sistan-Balouchestan
22- Markazi
23- Khouzestan
24- Hormozgan

* Av.: Average

Table 5.6 The level of significant of component of random variation in three-level analyses of log(weight) averaged over the 24 provinces; National Health Survey 1990-2, Iran

Parameter	Level 1	Level 2	Level 3
Cons	$p<0.001$	$p<0.001$	$p<0.001$
Cage	-	$p=0.09$	$p=0.20$
SEX	-	-	$p=0.34$
S*CA	-	-	$p=0.28$

- : Not fitted

Table 5.7 Coefficients of Multilevel Modelling Regression of *log(weight)'* across the 24 provinces of Iran

Province code	Cons (SE)	Cage (SE)	Sex (SE)	UR (SE)	Cage2 (SE)	S*CA (SE)	S*UR (SE)	UR*CA (SE)
1	311.7 (3.78)	10.62 (0.51)	-1.02 (2.91)	-2.84 (6.39)	0.15 (0.09)	2.13 (0.65)	3.878 (4.91)	0.51 (0.68)
2	309.9 (3.45)	11.08 (0.38)	0.55 (2.51)	-8.15 (4.27)	-0.03 (0.06)	0.51 (0.43)	-3.41 (3.25)	-1.74 (0.44)
3	314.6 (1.21)	9.83 (0.20)	-1.73 (1.28)	-1.30 (1.56)	0.02 (0.03)	1.08 (0.21)	-3.78 (1.67)	-0.60 (0.22)
4	295.9 (1.61)	11.68 0.23	-0.18 (1.37)	-3.85 (2.58)	0.09 (0.04)	0.78 (0.28)	0.06 (2.20)	-0.52 (0.29)
5	303.8 (4.71)	10.75 (0.46)	-5.77 (2.85)	-3.66 (6.55)	-0.07 (0.07)	0.30 (0.53)	-1.79 (3.88)	-0.01 (0.05)
6	313.9 (2.03)	10.03 (0.29)	1.70 (1.96)	0.17 (2.60)	0.064 (0.04)	0.84 (0.32)	-5.52 (2.56)	-0.80 (0.34)
7	301.0 (1.91)	11.25 (0.29)	-1.40 (1.81)	-5.51 (2.44)	-0.09 (0.04)	-0.15 (0.32)	-1.01 (2.43)	-0.20 (0.32)
8	305.2 (2.08)	11.21 (0.30)	-1.09 (2.0)	-3.87 (2.90)	-0.05 (0.05)	0.40 (0.35)	-6.02 (2.83)	-0.50 (0.36)
9	307.4 (2.19)	10.80 (0.31)	-1.14 (2.08)	3.37 (2.67)	-0.01 (0.05)	0.81 (0.34)	-1.70 (2.61)	-1.10 (0.35)
10	311.2 (0.75)	10.69 (0.12)	-1.80 (0.63)	-9.29 (2.81)	0.09 (0.02)	0.71 (0.16)	5.64 (2.63)	-0.59 (0.35)
11	306.0 (1.23)	10.55 (0.21)	-1.94 (1.27)	-5.94 (1.64)	0.02 (0.03)	0.59 (0.22)	0.51 (1.71)	-0.60 (0.23)
12	306.3 (5.76)	9.36 (0.70)	1.10 (4.85)	-6.27 (6.40)	0.07 (0.07)	0.44 (0.53)	-6.21 (5.40)	1.09 (0.68)
13	309.5 (2.36)	10.45 (0.41)	-5.84 (2.51)	-4.92 (3.12)	0.03 (0.06)	1.81 (0.46)	6.12 (3.61)	-0.73 (0.47)
14	306.6 (1.45)	10.13 (0.22)	2.30 (1.46)	0.06 (1.78)	0.11 (0.03)	1.46 (0.25)	-4.09 (1.92)	-0.34 (0.26)
15	307.0 (0.88)	10.05 (0.13)	0.59 (0.80)	-3.56 (1.22)	0.04 (0.02)	0.88 (1.15)	-3.04 (1.15)	-0.46 (0.15)
16	304.1 (2.97)	11.85 (0.40)	0.37 (2.35)	-2.57 (3.78)	0.17 (0.06)	0.29 (0.41)	-2.08 (3.05)	-0.62 (0.43)
17	298.5 (2.79)	11.82 (0.40)	-0.17 (2.23)	-7.45 (3.64)	0.0004 (0.05)	0.82 (0.38)	-2.63 (2.90)	-0.32 (0.41)
18	310.9 (1.86)	10.44 (0.33)	2.92 (2.03)	-3.38 (2.56)	0.009 (0.05)	0.79 (0.37)	-4.38 (2.84)	-0.73 (0.37)
19	308.0 (1.99)	10.76 (0.35)	0.05 (2.13)	2.20 (2.97)	0.10 (0.06)	0.41 (0.43)	-5.03 (3.26)	-1.04 (0.43)
20	290.1 (5.48)	11.99 (0.62)	9.20 (4.03)	5.17 (7.34)	0.04 (0.10)	1.10 (0.68)	-14.83 (5.16)	-1.39 (0.70)
21	303.6 (2.03)	10.56 (0.36)	2.76 (2.13)	-1.18 (2.46)	-0.02 (0.05)	0.56 (0.37)	-2.87 (2.72)	-1.03 (0.37)
22	313.8 (2.00)	10.43 (0.37)	0.27 (2.35)	-6.43 (2.62)	0.11 (0.06)	0.82 (0.42)	-3.24 (3.16)	-0.21 (0.40)
23	313.1 (1.26)	10.60 (0.18)	0.07 (1.11)	-3.35 (2.01)	0.12 (0.03)	0.88 (0.23)	-1.95 (1.85)	-1.25 (0.25)
24	309.2 (2.34)	10.16 (0.37)	-2.88 (2.30)	-8.87 (3.30)	-0.16 (0.06)	0.43 (0.46)	2.55 (3.42)	0.16 (0.47)

Province code:
1- Semnan
2- Chaharmahal-Bakhtiari
3- East Azarbaijan
4- Isfahan
5- Kordestan
6- West Azarbaijan
7- Hamadan
8- Bakhtaran
9- Kerman
10- Tehran
11- Fars
12- Kohkiloyeh-Boyerahmad
13- Boushehr
14- Mazandaran
15- Khorassan
16- Gilan
17- Zanjan
18- Lorestan
19- Yazd
20- Ilam
21- Sistan-Balouchestan
22- Markazi
23- Khouzestan
24- Hormozgan

†) In the above analyse, log(weight) is multiplied by a hundred, and age is centred at eight

Table 5.8 Coefficients of Multilevel Modelling Regression of ***height*†** across the 24 provinces of Iran

Province code	Cons (SE)	Cage (SE)	Sex (SE)	UR (SE)	Cage2 (SE)	S*CA (SE)	S*UR (SE)	UR*CA (SE)
1	125.1 (1.42)	5.58 (0.24)	-1.16 (1.37)	-5.14 (2.40)	-0.005 (0.04)	-0.56 (0.31)	0.33 (2.35)	0.26 (0.32)
2	120.7 (1.70)	5.92 (0.19)	-0.70 (1.24)	-3.10 (2.10)	-0.03 (0.03)	-0.08 (0.21)	-0.02 (1.60)	-0.776 (0.22)
3	121.7 (0.65)	5.44 (0.10)	-1.02 (0.61)	-2.78 (0.85)	-0.08 (0.01)	0.08 (0.10)	-1.18 (1.67)	-0.20 (0.10)
4	121.2 (0.46)	5.58 (0.08)	-0.1 (0.5)	-1.19 (0.68)	-0.07 (0.01)	-0.02 (0.10)	0.62 (0.80)	-0.08 (0.10)
5	123.3 (0.97)	5.35 (0.18)	-1.87 (1.09)	-3.28 (1.28)	-0.08 (0.03)	0.09 (0.20)	0.11 (1.49)	-0.26 (0.20)
6	121.8 (0.87)	5.35 (0.14)	-0.03 (0.91)	-1.79 (1.10)	-0.08 (0.02)	0.05 (0.15)	-2.11 (1.19)	-0.189 (0.16)
7	122.2 (0.67)	5.71 (0.12)	-0.43 (0.75)	-3.42 (0.82)	-0.08 (0.02)	-0.08 (0.13)	-0.45 (1.0)	-0.45 (0.13)
8	123.2 (0.91)	5.55 (0.12)	-0.64 (0.81)	-0.97 (1.29)	-0.10 (0.02)	-0.15 (0.14)	-2.33 (1.13)	-0.22 (0.14)
9	121.7 (1.24)	5.93 (0.16)	-1.20 (1.08)	-3.70 (1.53)	-0.11 (0.02)	0.13 (0.18)	1.65 (1.36)	-0.33 (0.18)
10	124.6 (0.32)	5.73 (0.06)	-1.03 (0.25)	-4.48 (1.11)	-0.07 (0.01)	0.01 (0.07)	0.30 (1.05)	-0.13 (0.14)
11	120.9 (0.51)	5.53 (0.08)	-0.53 (0.51)	-2.88 (0.68)	-0.08 (0.01)	-0.04 (0.09)	-0.24 (0.69)	-0.02 (0.09)
12	115.1 (2.39)	5.41 (0.31)	3.16 (2.19)	0.77 (2.63)	-0.07 (0.03)	0.14 (0.24)	-4.42 (2.43)	0.11 (0.31)
13	121.5 (0.91)	5.55 (0.16)	-2.15 (0.98)	-2.85 (1.19)	-0.06 (0.03)	0.19 (0.18)	1.68 (1.42)	-0.22 (0.19)
14	119.7 (0.66)	5.68 (0.10)	1.42 (0.64)	0.91 (0.82)	-0.05 (0.10)	0.06 (0.11)	-2.99 (0.85)	-0.37 (0.11)
15	121.7 (0.38)	5.36 (0.06)	-0.09 (0.35)	-3.21 (0.53)	-0.06 (0.01)	0.08 0.06)	-1.18 (0.50)	-0.14 (0.06)
16	123.8 (1.05)	5.70 (0.18)	-0.37 (1.05)	-3.28 (1.30)	-0.10 (0.03)	-0.02 (0.18)	-0.62 (1.37)	-0.09 (0.19)
17	121.0 (0.92)	5.49 (0.15)	-0.07 (0.85)	-3.90 (1.18)	-0.04 (0.02)	0.02 (0.15)	-0.57 (1.10)	-0.23 (0.16)
18	120.9 (0.78)	5.63 (0.16)	-0.07 (0.97)	-2.09 (1.10)	-0.10 (0.02)	0.16 (0.18)	-1.38 (1.36)	-0.35 (0.18)
19	121.5 (0.84)	5.94 (0.14)	0.86 (0.88)	-1.35 (1.26)	-0.09 (0.02)	-0.24 (0.17)	-2.36 (1.33)	-0.24 (0.18)
20	119.8 (1.45)	5.34 (0.25)	1.76 (1.57)	1.44 (1.79)	-0.07 (0.04)	0.08 (0.28)	-4.12 (2.07)	0.08 (0.28)
21	120.6 (0.99)	5.64 (0.16)	0.47 (0.97)	-1.75 (1.20)	-0.12 (0.02)	0.08 (0.16)	-1.05 (1.24)	-0.37 (0.17)
22	125.3 (0.87)	5.67 (0.17)	-0.51 (1.06)	-4.27 (1.23)	-0.10 (0.03)	0.03 (0.18)	-0.78 (1.42)	-0.19 (0.19)
23	124.3 (0.56)	5.76 (0.08)	-0.87 (0.50)	-3.20 (0.89)	-0.06 (0.01)	0.003 (0.10)	-0.59 (0.83)	-0.38 (0.11)
24	121.7 (0.95)	5.57 (0.17)	0.19 (1.04)	-1.08 (1.31)	-1.18 (0.03)	0.22 (0.21)	0.51 (1.54)	-0.64 (0.21)

Province code:
1- Semnan
2- Chaharmahal-Bakhtiari
3- East Azarbaijan
4- Isfahan
5- Kordestan
6- West Azarbaijan
7- Hamadan
8- Bakhtaran
9- Kerman
10- Tehran
11- Fars
12- Kohkiloyeh-Boyerahmad
13- Boushehr
14- Mazandaran
15- Khorassan
16- Gilan
17- Zanjan
18- Lorestan
19- Yazd
20- Ilam
21- Sistan-Balouchestan
22- Markazi
23- Khouzestan
24- Hormozgan

†) In the above analyse, age is centred at eight

Table 5.9 Estimated coefficients (and standard errors) of the fixed part of the four-level model using results of the three-level analyses on log(weight)[†] and height in the 24 provinces of Iran

Log(weight)		Height	
Parameter	Estimate(s.e)	Parameter	Estimate(s.e)
Cons	308.44 (0.34)	Cons	122.42 (0.14)
Cage	10.57 (0.51)	Cage	5.56 (0.01)
SEX	-0.54 (0.31)	SEX	-0.46 (0.13)
U/R	-3.47 (0.50)	U/R	-2.48 (0.21)
Cage2	0.05 (0.001)	Cage2	-0.07 (0.004)
S*CA	0.75 (0.06)	S*CA	0.02 (0.03)
S*UR	-2.20 (0.49)	S*UR	-0.92 (0.21)
UR*CA	-0.60 (0.07)	UR*CA	-0.09 (0.02)

†) In the above table, log(weight) is multiplied by a hundred, and in both analyses age is centred at eight

Table 5.10 Heterogeneity of coefficients of growth across the 24 provinces of Iran

Parameter	$Q_{Log(weight)}$	Q_{Height}
Cons	190.29*	147.22*
Cage	105.10*	58.82*
SEX	34.64	31.88
U/R	27.87	40.99*
Cage2	50.88*	40.23*
S*CA	50.84*	13.52
S*UR	35.73*	26.55
UR*CA	37.88*	64.76*

* $p<0.05$

$\chi^2_{23}(0.05)=35.17$

Table 5.11 Pearson's coefficient of correlation between parameter estimates in modelling log(weight) and height, and child mortality (q_{1-4}) in the 24 provinces of Iran

Parameter	Cons-Wt	Cage-Wt	Cons-Ht	Cage-Ht	q_{1-4}
Cons-Wt[A]	1				
Cage-Wt[B]	-0.70**	1			
Cons-Ht[C]	0.34	0.25	1		
Cage-Ht[D]	0.23	0.18	0.23	1	
q_{1-4}	0.13	-0.21	0.12	-0.54*	1

*: $P<0.05$
**: $P<0.01$
A: Constant (intercept) in analysis of log(weight)
B: Slope in analysis of log(weight)
C: Constant (intercept) in analysis of height
D: Slope in analysis of height

Table 5.12 Random coefficient model of the National Health Survey data in Iran (4-level analysis of log(weight)[†])

Parameter	Estimate (s.e)	Corr	C.V
Fixed:			
Cons	306.90 (1.21)		
Cage	10.69 (0.13)		
SEX	-0.61 (0.32)		
U/R	-3.73 (0.56)		
Cage2	0.04 (0.016)		
S*CA	0.72 (0.08)		
S*UR	-1.94 (0.53)		
UR*CA	-0.59 (0.07)		
Random:			
level 4			
σ^2_{w0} (between provinces)	30.26 (9.47)		5.5%
σ_{w01}	-2.54 (0.86)	-0.85	
σ^2_{w2}	0.29 (0.10)		
σ^2_{w4}	0.003(0.002)		
σ^2_{w5}	0.05 (0.04)		
σ^2_{w6}	0.99 (1.02)		
σ^2_{w7}	0.01 (0.03)		
Level 3			
σ^2_{v0} (between clusters)	43.00 (2.97)		6.6%
Level 2			
σ^2_{u0} (between families)	74.49 (3.15)		8.6%
Level 1			
σ^2_{e0} (between children)	219.20 (2.72)		14.8%

C.V: Coefficient of variation; Corr: Correlation
σ_{w01}=Cov(Cons /Cage)
σ^2_{w2}=Var('Cage' /'Cage')
σ^2_{w4}=Var('Cage2' /'Cage2')
σ^2_{w5}=Var('S*CA' /'S*CA')
σ^2_{w6}=Var('S*UR' /'S*UR')
σ^2_{w7}=Var('UR*CA' /'UR*CA')

†) In the above analysis, log(weight) is multiplied by a hundred, and age is centred at eight

Table 5.13 Comparison of the estimation of 'Cons' (intercept) in 4-level analysis and the 3-level analysis of the data on weight with provinces as fixed differences

Province Code	I Intercept estimates 3-level analysis with fixed provincial difference	II Intercept estimates 4-level analysis
1	313.30	312.20
2	306.09	306.14
3	314.10	313.94
4	297.17	297.39
5	299.06	299.62
6	316.08	315.67
7	300.30	300.60
8	302.22	302.47
9	306.57	306.58
10	310.99	310.95
11	304.32	304.38
12	303.98	304.43
13	307.62	307.50
14	310.33	310.23
15	307.21	307.20
16	306.94	306.94
17	296.21	296.70
18	311.36	311.06
19	310.51	310.11
20	295.65	297.66
21	305.33	305.41
22	313.47	312.99
23	314.62	314.40
24	303.57	303.84

Figure 5.1 Mean weight (kg) and height (cm) according to sex and place of residence (urban/rural) in Iran

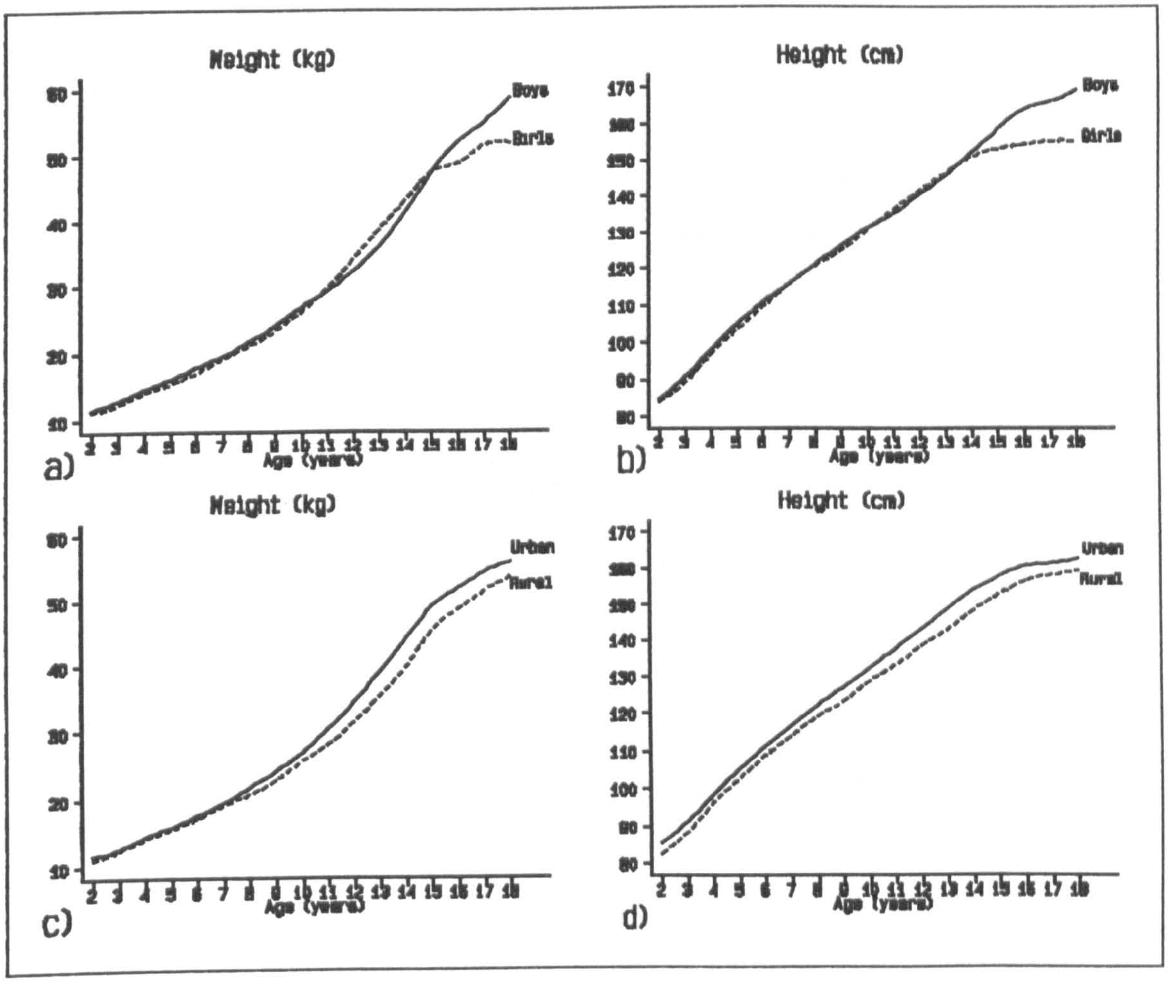


Figure 5.2 Mean height (cm) of urban and rural children for different sexes

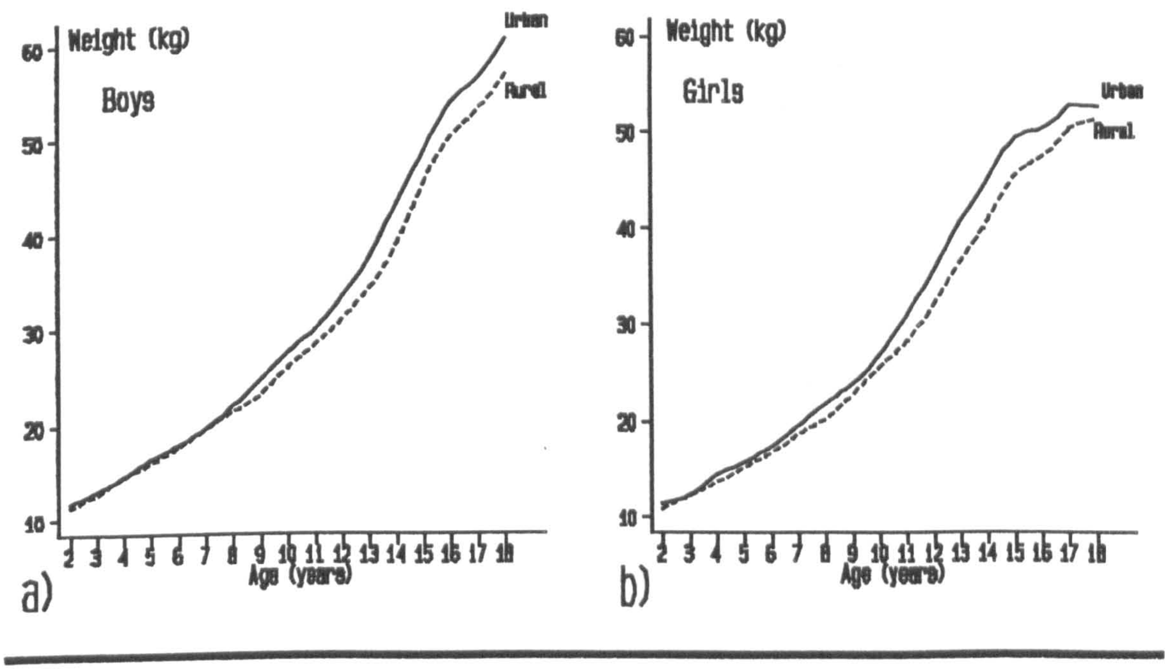
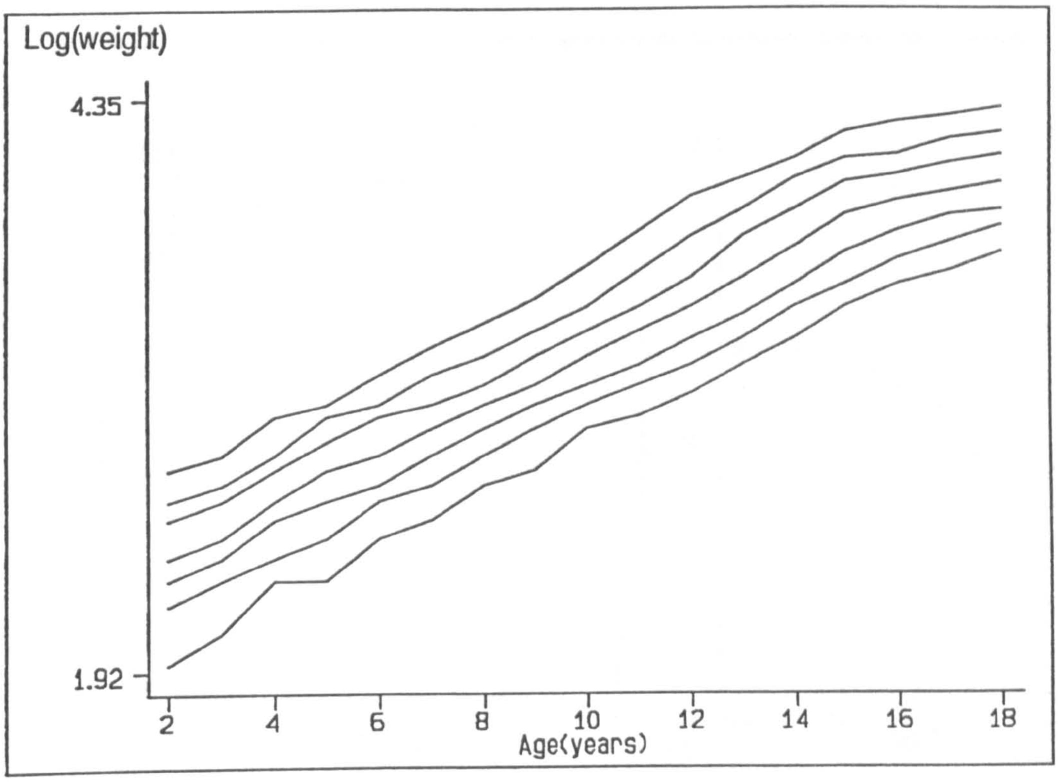
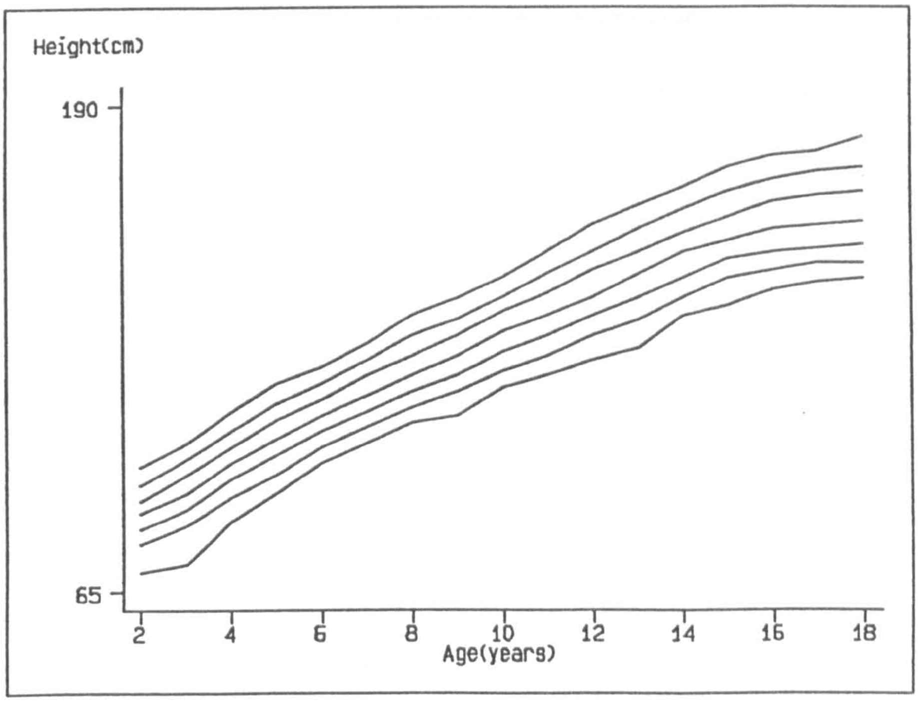


Figure 5.4 Percentiles of log(weight) and height of Iranian children and adolescents aged 2-18 (years); centiles 3rd, 10th, 25th, 50th, 75th, 90th, 97th

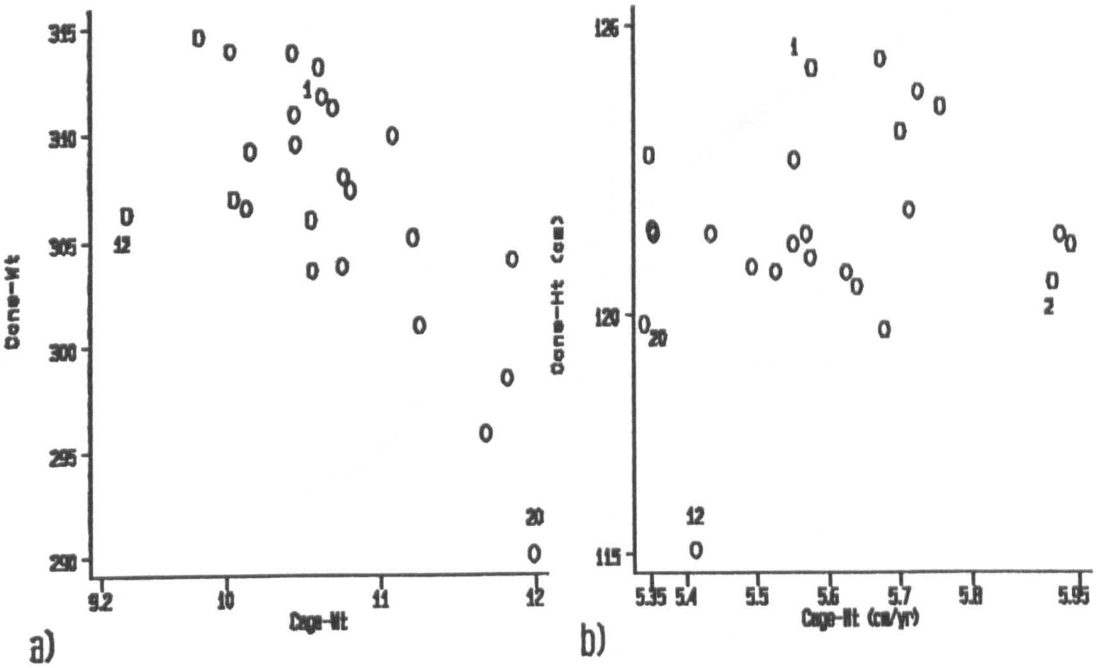


a) Weight



b) Height

Figure 5.5 Scatter plots of intercepts ('Cons') and slopes ('Cage') of 3-level models of log(weight) and height in the 24 provinces of Iran



a) Weight b) Height

1: Semnan
2: Chaharmahal-Bakhtiari
12: Kohkiloyeh-Boyerahmad
20: Ilam

Figure 5.6 Normal plot of residuals of four-level analysis of log(weight), National Health survey 1990-2, Iran

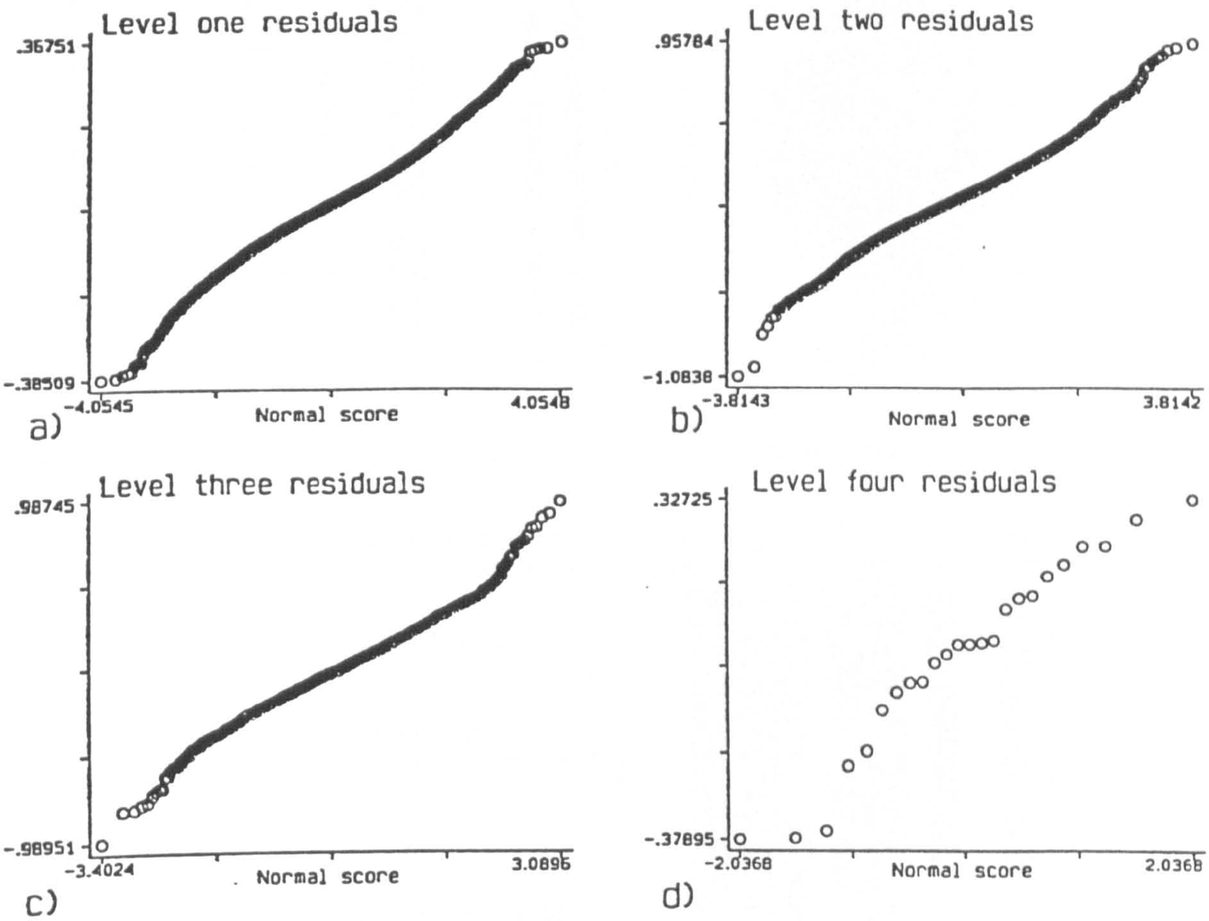


Figure 5.7 Scatter plots of residuals, four-level analysis of log(weight), National Health Survey 1990-2, Iran

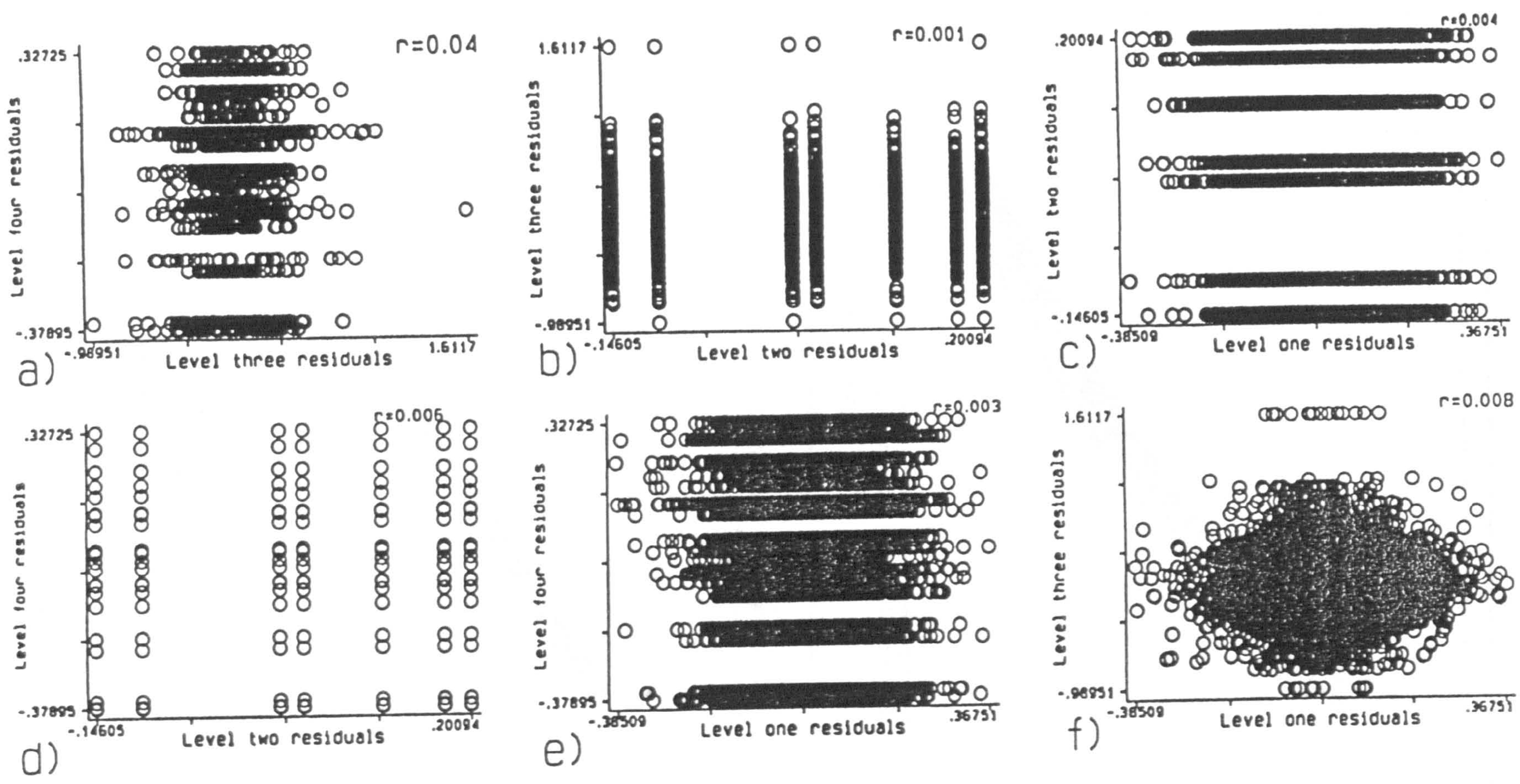
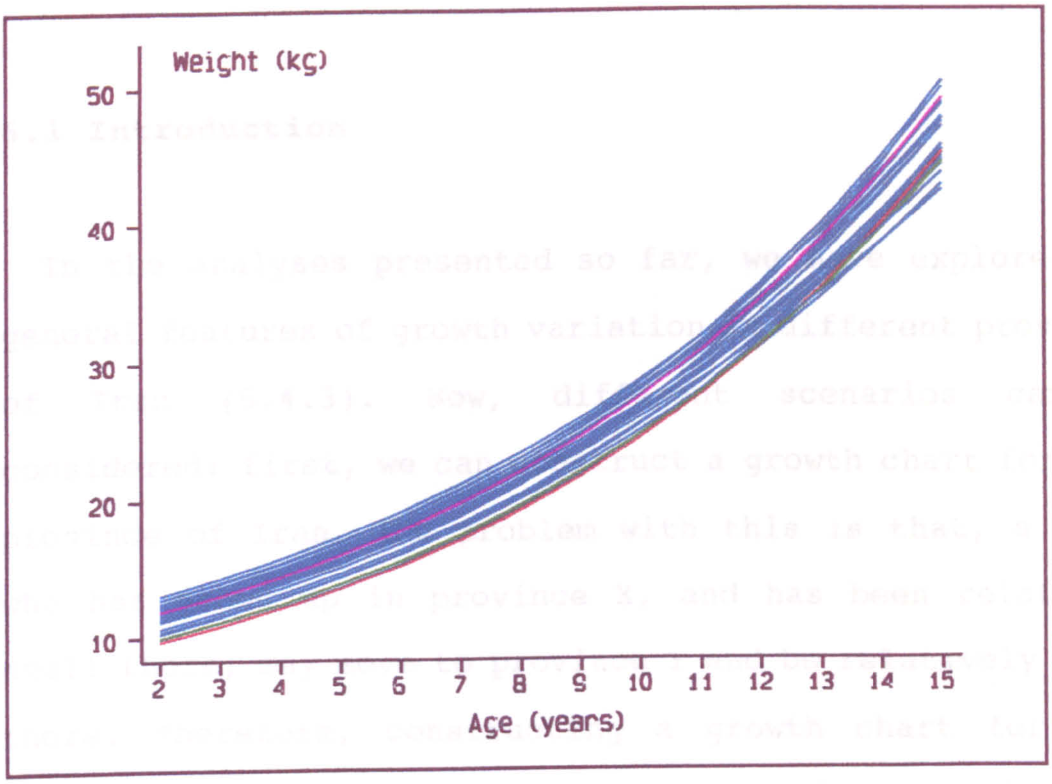
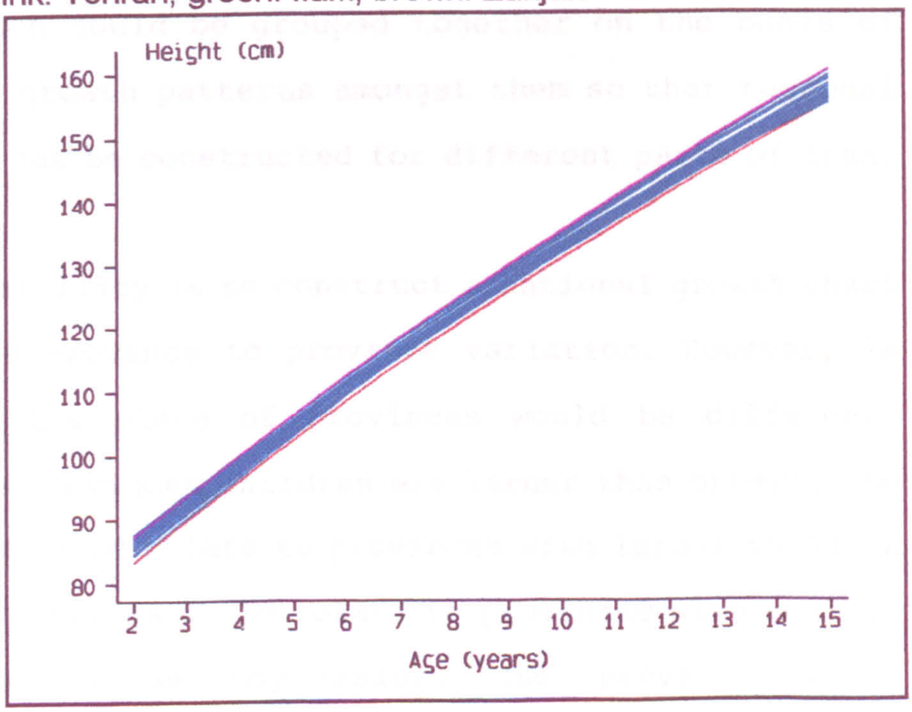


Figure 5.8 Provincial estimtes of growth patterns of urban boys across Iran derived from 4-level analysis of data on weight and height



a) Weight (kg); pink: Tehran, green: Ilam, brown: Zanjan



b) Height(cm); pink: Tehran, brown: Kohkiloyeh-Boyerahmad

CHAPTER SIX

CHOICE OF A BASELINE AREA

6.1 Introduction

In the analyses presented so far, we have explored the general features of growth variation in different provinces of Iran (5.4.3). Now, different scenarios can be considered: first, we can construct a growth chart for each province of Iran. The problem with this is that, a child who has grown up in province X, and has been relatively small there, may move to province Y and be relatively large there. Therefore, constructing a growth chart for each province of Iran is not reasonable. Second, we can look for provinces which could be grouped together on the basis of similarity of growth patterns amongst them so that regional growth charts can be constructed for different parts of Iran.

Another possibility is to construct a national growth chart which includes province to province variation. However, in such a chart the place of provinces would be different. Since, in some provinces children are larger than others, the upper centiles would relate to provinces with larger children and the lower centiles would relate to provinces with smaller children. Lastly, we may select one province as a representative for the country. This chapter examines the results of the preliminary analysis presented in chapter 5 to answer these questions.

6.2 Grouping the pattern of growth in provinces of Iran

6.2.1 Method

6.2.1.1 Cluster analysis

In statistics, the search for relatively homogeneous groups of objects is called *cluster analysis*. In biology, cluster analysis is used to classify animals and plants. In medicine, cluster analysis is used to identify diseases and their stages. For example, by examining patients who are diagnosed as depressed, one finds that there are several distinct subgroups of patients with different types of depression. Generally, the problem that cluster analysis is designed to solve is the following one: given a sample of n objects, each of which has a score on *p* variables, how to devise a scheme for grouping the objects in to classes so that *'similar'* ones are in the same class (Manly, 1994).

6.2.1.2 Types of the cluster analysis

Many algorithms have been proposed for cluster analysis. The two popular approaches are discussed below: firstly, there are hierarchic techniques. These methods start with the calculation of the distances of each individual to all other individuals. Groups are then formed by a process of agglomeration or division. With agglomeration all objects start by being alone in groups of one. Close groups are then gradually merged until finally all individuals are in a single group. With division all objects start in a single group. This is then split in to two groups, the two groups are then split, and so on until all objects are in groups of their own.

The second approach to cluster analysis involves partitioning, with objects being allowed to move in and out of groups at different stages of the analysis. To begin with, some arbitrary group centres are chosen and individuals are allocated to the nearest one. New centres are then calculated where these are at the centres of the individuals in groups. An individual is then moved to a new group if it is closer to that group's centre than it is to the centre of its present group. Groups 'close' together are merged; spread-out groups are split, etc. The process continues iteratively until stability is achieved with a predetermined number of groups. Usually a range of values is tried for the final number of groups (SPSS, 1990).

6.2.1.3 Hierarchic methods

Agglomerative hierarchic methods start with a matrix of 'distances' between individuals. All individuals begin alone in groups of size one and groups that are 'close' together are merged. (Measures of 'distance' will be discussed later.) There are various ways to define 'close'.

One of the simplest methods is *single linkage*, sometimes called 'nearest neighbour'. The first two cases combined are those that have the smallest distance (or largest similarity). The distance between the new cluster and individual case is then computed as the minimum distance between an individual case and a case in the cluster. At every step, the distance between two clusters is the distance between their two closest points. Another commonly used method is called *complete linkage*, or the 'furthest neighbour' technique. In this method, the distance between two clusters is calculated as distance between their two furthest points. With *group average linkage*, often called UPGMA (unweighted pair-group method using arithmetic average), two groups merge if the average distance between them is small enough.

Divisive hierarchic method have been used less often than agglomerative ones. The objects are all put into one group initially, and then this is split into two groups by separating off the object that is furthest on average from other objects. Individuals from the main group are then moved to the new group if they are closer to it than they are to the main group. Further subdivisions occur as the distance that is allowed between individuals in the same group is reduced. Eventually all objects are in groups of size one.

It has been mentioned that there are many algorithms for cluster analysis. However, there is no generally accepted *'best'* method. Unfortunately, different algorithms do not necessarily produce the same results on a given data set and there is usually rather a large subjective component in the assessment of the results from any particular method.

6.2.1.4 Measures of distances and similarities

There are many methods for estimating the distance or similarity between two cases. But even before these measures are computed, we must decide whether the variables need to rescaled. When the variables have different scales, and they are not standardized, any distance measure will reflect primarily the contributions of variables measured in the large units. One means of circumventing this problem is to express all variables in standardized form.

Based on the transformed data, it is possible to calculate many different types of distance and similarity measures. Different distance and similarity measures weight data characteristics differently. The choice among the measures should be based on which differences or similarities in the data are important for a particular application. The most commonly used distance measure is the squared Euclidean distance:

$$SECULID(X,Y)=\sum_{i=1}^{n}(X_i-Y_i)^2$$

or sometimes its square root (Euclidean distance). Some of the methods require particular types of distance measures. With the methods discussed different measures can be used, however, in the following work squared Euclidean distance using SPSS software was used to carry out the analysis.

6.2.2.1 Results of grouping

As was stated, one of the objectives of the study is to see if there are any distinct groups of provinces with similar growth patterns. If so, regional growth charts might be needed. However, when trying to find any grouping in growth pattern amongst the provinces of Iran, instead of working on the raw data on weight and height, we worked on the estimated coefficients found in modelling of log(weight) and height of individuals in provinces (section 5.4.3), because these coefficients summarise the growth characteristics in the different provinces of Iran. So our variables on each subject (province) are the estimated regression coefficients where each province contributed sixteen variables to the cluster analysis, eight from the analysis of weight ('Cons','Cage', ...,'UR*CA'; 5.4.3), and the other eight were from the analysis of height.

6.2.2.2 Principal component analysis

When the first two principal components (PC) account for a high percentage of variation in the data a plot of subjects (provinces) against these two components is a useful way for looking for any grouping (Manly, 1994). Also, Gower (1991) has commented that if there are real clusters in the data they can almost always be observed in such a scatter plot. Generally, in principal component analysis, linear combinations of the observed variables are formed. The first principal component is the combination that accounts for the largest amount of variance in the data. The second principal component accounts for the next largest amount of variance and is uncorrelated with the first. Successive components explain progressively smaller portions of the total sample variance, and all are uncorrelated with each other.

In principal component analysis of the coefficients obtained from the analyses on weight and height of children, it was observed that the first two components explain 50.4% of the total variation. Hence, a plot of these components by provinces would be a good summarisation of the data. Figure 6.1 displays the plot of the first two principal components by provinces. In this figure, the points labelled 1 (Semnan), 20 (Ilam), 12 (kohkiloyeh-Boyerahmad) , and 24 (Hormozgan) are far apart from the rest of the points and appeared as distinct provinces. This is consistent with the findings in chapter 5 that lighter children in province Ilam (20), for example, were putting on weight faster than children elsewhere and that the children in Semnan (1) were substantially taller than the relatively smaller children in Kohkiloyeh-Boyerahmad. However, Figure 6.1 shows there do not appear to be any big groups of provinces. Also, province Tehran in this figure (labelled as 10) is not far from the centre. One should bear in mind that the provincial models in this analysis were concerned with modelling average values and also that the first two PC only accounted for 50.4% of the variation. So investigation of variations of different growth centiles may clarify the similarities better. The next section explores the possibility of grouping the growth models analytically.

6.2.2.3 Results of cluster analysis

Table 6.1 presents the results of the cluster analysis using UPGMA method. In grouping procedures, it was desirable to avoid having too many groups. Therefore having two to four groups were considered. Often there is not one single meaningful cluster solution, but many, depending on what is of interest. Table 6.1 shows the cluster membership for the provinces at different stages of the solution. One can easily see which clusters, the provinces belong to in the two- to four-cluster solutions. For example, in the two-cluster solution we have Semnan (1) as one group and the rest of the country as the second cluster. Or in trying three-cluster, province Ilam (20) appeared as third group, Semnan (1) as one group, and the rest of Iran as the other cluster. Similarly, in four cluster-solution, province 20 (Ilam) appeared as the fourth cluster, Semnan (1), Kohkiloyeh-Boyerahmad (12), and the rest of the country as the other groups. It is interesting that these results are consistent with those were seen in Figure 6.1. For instance, Semnan (1) which is far apart from the rest of the points appears as a separate cluster in the two-cluster solution. Also, the provinces 12 and 20 (Kohkiloyeh-Boyerahmad and Ilam) which appeared in three and four-cluster solutions, can be visually seen to be distinct from the rest of the provinces in Figure 6.1.

In addition, when as complementary information child mortality rate (q_{1-4}) was included in the variables, the results of the analyses were the same indicating that the weak correlation between the mortality rate and the model parameters did not add to the classification. It should be noted that all the three methods that discussed in section 6.2.1.2 were tried, and it was observed that the result of the *group average linkage* (UPGMA) method and the *single linkage* method were almost identical. The above presentations are the results of the average linkage method.

Since in modelling height and weight across provinces the interaction between **UR** term (urban-rural) and the other variable were significant (Table 5.9), there might be a difference in clustering between the urban and rural areas of the country. To analyse the urban and rural data independently separate models of weight and height were constructed for urban and rural data for each province similar to the models in section 5.4.2 except that there were no terms in **UR**. After constructing these models, cluster analyses were carried out on the coefficients found in these new analyses. The result of the cluster analysis for urban areas was the same as that found for whole Iran. However the findings were slightly different when looking at the rural areas. Table 6.2 presents the result of the analysis for the rural areas of Iran. As can be seen from the Table 6.2 in two-cluster solution Yazd (19) joined Semnan as a group versus the rest of rural part of the country, and in three-cluster solution province Bakhtaran (8) and Kordestan (5) appeared as the third cluster instead of Ilam (20) in Table 6.1. In the four-cluster solution the number of provinces in the third and fourth group increased but again these provinces are far away from each other. For example, in the four-cluster solution the third group consists of provinces Isfahan (4), Hamadan (7), Kohkiloyeh-Boyerahmad (12), Gilan (16), Zanjan (3), and Ilam (20) which are geographically far from each other (Figure 1.1; map of Iran).

The results of grouping for the rural areas were slightly different when different methods of clustering (6.2.1.3) were attempted but the type of results, in the sense that geographically far apart provinces appeared in same groups or observing one province as a group, did not change. The presentation in Table 6.2 is also UPGMA approach. In summary, as in the cluster analysis of all the data, in neither urban nor rural cluster analyses, no big groups of geographically adjacent provinces were found, these results agree with PC analysis presented in Figure 6.1.

It is worthwhile mentioning that if the principal component scores of the coefficients were used in the cluster analysis, the squared Euclidian distance would be the Mahalanobis distance which is independent of both scale and correlation. So, after principal component analysis in MINITAB (1991) of the sixteen coefficients from provincial modelling of log(weight) and height, cluster analysis was carried out on the PC scores. It was observed that the results were similar to those derived above for the whole of Iran.

In addition, in order to see if centring affects the results of clustering, the provincial analyses of weight and height data were repeated without centring age. Then, cluster analysis were carried out on the coefficient found in this new uncentred age analysis and on their principal component scores. In both cases, it was observed that the general finding of having just a few provinces as separate groups, shown in Figure 6.1, is the same. However, this time the order in which provinces Semnan (1), Kohkiloyeh-Boyerahmad (12), Ilam (20), and Hormozgan (24) appeared in two to four-cluster solution were different from steps three to five. Tales 6.3 illustrates the result of this analysis. As one can see in this table, from third step Hormozgan (24) appeared as a separate group, and Kohkiloyeh-Boyerahmad (12) and Ilam (20) as groups fourth and fifth but again no big groups of provinces were found. Observation of growth patterns of children in Hormozgan showed that at younger ages the mean of weights and heights of children is in the middle of the range of variation but they are not growing as fast as children in some other provinces and at older ages such as fourteen or fifteen they are relatively smaller and shorter. This is apparent in PC analysis of scores shown in the Figure 6.1 where Hormozgan (24) appeared as a separate point. The new PC scores are different from the ones when the data were centred. Therefore, we speculate that the centring does not affect the result of clustering. The new scores might be a rotation of the previous ones in multi-dimensional space.

6.2.2.4 Conclusion of grouping the growth patterns

The analytical findings of the cluster analyses would not be practically applicable for Iran because in almost all of the analysis it was observed that one or a few provinces are in the extremes in comparison with the rest of the country and appeared as distinct clusters. Also the pattern of clustering changed when urban areas of the country were considered in comparison with rural areas. For construction of growth charts this kind of grouping, which is different for urban to rural areas, is impractical. A view of a map of Iran (Figure 1.1) shows that in all cluster-solutions the provinces in a cluster are located in different corners of the country whereas what might be expected was to have groups of provinces which were neighbours, or possibly geographically distinct such as north-south, east-west; or similar in ethnic group, or urbanisation (industrial development). However in none of the analyses were such results found. Therefore, on the base of these findings the possibility of developing separate regional growth charts appears neither practical nor useful. So more investigation are needed to attempt to find an appropriate solution.

6.3 Is there a baseline province which reasonably represents the country?

Since the investigation of whether any of the provinces could be grouped together to produce regional growth charts provided no practical solution, other ways of studying the data were explored. By considering the coefficients of growth patterns of children across provinces of the country, one can see that the intercepts and slopes of the models for the urban Tehran are close to the average of the values (Table 5.7, Tehran: Cons=311.2, Cage=10.69; Table 5.9, average values: Cons=308.44, cage=10.57)[1], indicating that the weight and height of the urban Tehrani children could be used as a reasonable average of urban Iran as a whole. Therefore the possibility of choosing urban Tehran as a baseline for the country was examined.

In order to study this possibility, in a 4-level analysis of the data, the difference between the urban Tehran (**UT**) and the rest of the urban part of the country was examined by introducing a variable (**UT**) in the model where

$$\boldsymbol{UT} = \left\{ \begin{array}{ll} 1 & \textit{if urban Tehran} \\ 0 & \textit{if elsewhere} \end{array} \right\}.$$

Table 6.4 (column A) presents the result of this analysis. As can be seen from the table, the **UT** difference was not significant (p=0.12), indicating that urban Tehran is not significantly different from the rest of the urban Iran. Also the age interaction term ('UT*CA') was included in the model to see wether the change with age in rate of growth for **UT** was significant, and it was not (p=0.15), implying that the non-significant difference across the urban areas is not changing as children grow up. It should be noted that in this analysis urban Tehran was included

1: In fixed part of a 4-level analysis of weight with simple variance structure: Cons=306.90, Cage=10.69. And shrunken estimate for urban Tehran Cons=310.29, Cage=10.64

both in the random and the fixed part of the model. This means that urban Tehran was viewed as one of the units in the urban part of Iran, by consideration of the existing variation between the provinces, and this difference was not significant ($p=0.12$). So as one of the parts of this set of heterogeneous urban areas, urban Tehran is not significantly different from the average of the rest.

Another possible way to investigate this situation is to remove urban Tehran from the variation of the province level (level four). In this view we may wish to look at urban Tehran as a different population and to investigate, as a separated unit, where it lies in comparison with the average of the rest of the country. In order to do this the explanatory variable defining the level four, province level variation, was taken as $\mathbf{UT}^{*}=1-UT$ rather that a column of ones (Goldstein, 1995). Part B of Table 6.4 presents the results of this analysis. Although treating **UT** in this way produced a significant result ($p<0.01$), when looking at the level four variation ($31.66=5.6^2$) which shows the variation between the other provinces, **UT** is still within the one SD (5.6) of the range of variation about the mean. Therefore, in the range of urban areas of the provinces, urban Tehran does not lie far from the average. Regarding the previous analysis (Table 6.4, column A), this means that the difference between the mean for urban Tehran and the reminder of urban Iran of 4.68 ($unit=100\times\ln(kg)$) is divided between a component of random variation at level 4 and a fixed effect of 2.59 which is not significant, $p=0.12$. On the whole it should be noted that Tehran as the capital of Iran is one unit of the existing variation between the provinces, so the first approach is more appropriate. It is concluded that the weight of urban Tehrani children was not found to be significantly different from the rest of urban Iran. Note the 'CA*UT' fixed effect is the same in both models because 'Cage' is not included in the random part of the model.

A similar analysis was carried out for the height of children, Table 6.5. For model A the height of urban Tehrani children was found to be non-significant at level of $\alpha=0.01$ ($p=0.016$). As with the weight analysis, the finding for height when urban Tehran was removed from the random part of the model was significant ($p<0.01$; Table 6.5, model B). However the estimate was still within the range of 2×SD of the variation between the rest of the urban areas. Now the 'CA*UT' term is also significant ($p<0.01$), which implies that urban Tehrani children are 50% nearer the norm for urban children at age 2 and that the difference increases as they grow older.

6.4 Discussion of the choice of urban Tehran as baseline

6.4.1 Comparison of centiles

So far in consideration of choosing a baseline for the country we have investigated the significance of urban Tehran in the models. However, these results relate to the median (50th centile) of weight and height but in growth monitoring of equal if not more importance is the situation where a child's measurement is at the extremes of the distribution, especially at the lower centiles. Therefore, further preliminary analyses were carried out by looking at all centiles of interest (3rd, 10th, 25th, 50th, 75th, 90th, 97th) over the age range.

In this section the centiles of weight and height of urban Tehrani children are compared to the 95% confidence intervals for corresponding centiles for all urban Irani children. The analyses were carried out for measurements on boys and girls separately. For each of the four data sets (height and weight of boys and girls) computations were as follows. First the seven centiles were computed from the raw data for each age group 2 to 15 for all urban Irani children using SPSS. See Table 6.6 for the centiles for girls weights. Corresponding centiles were also computed for the urban Tehrani children, also shown in table 6.6, together with the smoothed centiles derived in chapter 8.

It was stated that the variance of the pth centile (C_p) is $\frac{p\,q}{n\,f_p^2}$ where f_p is the ordinate of the distribution at the pth percentile (section 4.2.2). Log(weight) and height were observed to be approximately Normal, so in these analyses log(weight) was used in computation of confidence intervals and the results were transformed back. For calculation of the density function at C_p $\left(f(C_p)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{C_p-\mu}{\sigma})^2}\right)$, the mean $\bar{x}$ was taken as the mean of log(weight) or mean height of each age-sex group to estimate corresponding μ. Then, since age specific estimates of the SD of hierarchically structured data were required, the four level models described in the previous section were extended in modelling the random components so as to take account of age and sex, as in Table 6.7. An MLn macro was written to perform the calculation of the confidence intervals of the centiles for urban Iran children at different age-sex groups.

Figure 6.2 shows the resulting estimates of the SD of log(weight) by age for boys and girls. It is found that the variation is minimal at about the ages of 7 for girls and 8 for boys, and is slightly less for girls at age 2 (Figure 6.2). This agreed with the idea that girls had less weight variation at birth (Carpenter, 1994). Also, in this age band (2-15 years) girls have larger variation in older ages due to earlier puberty. In contrast, the pattern of variation in height attainment of boys was in the opposite direction, i.e. boys height variation was less at 2 years and larger at puberty. Also the variations were minimal at ages 6 and 7 for boys and girls respectively. These differences in the pattern of variation in weight and height of boys and girls imply the need for separate charts for the two sexes. Generally, the change in SD with age was small and is consistent with slight narrowing of the centiles around the age of 8 years, as can be seen in Figure 5.4.

Tables 6.7 and Figure 6.3 present the results of the analysis for the girls' weight (the example of girls' weight will be used throughout this thesis when comparing various techniques in construction of age-related centiles). As one can see from the table and figure, in almost all ages, the centiles 10th, 25th, 50th, 75th of urban Tehran and urban Iran are very compatible. Generally, the 3rd urban Tehran centile at the bottom and the upper urban Tehran centile at the upper end of the corresponding confidence interval. This is also true for the 3rd and 97th centiles. Although at some ages such as puberty, the 3rd centile for urban Tehran lies towards the lower part of the confidence interval, and the 97th centile lies towards the upper part of the confidence interval.

There are ages where the lower or upper centiles of urban Tehran do not lie within the confidence intervals of the urban Iran centiles. If we look at the lower centiles, which are of most concern, in most cases the difference is within the range of error of measurements as weight was rounded to the nearest kilogram. For instance, at the age of four years, the value of the 3rd raw centile of urban Iran is 10 kg with 95% confidence interval (9.61, 10.41) where urban Tehran's 3rd centile is 9.19 kg (Table 6.7). In some ages this discrepancy was resolved when the raw urban Tehran centile was smoothed (last column in Table 6.7). Then the smoothed 3rd urban Tehran centile at age 4 is 9.68 which is within the urban Iran confidence interval. Another example occurs at age of eight years; the 3rd centile of urban Iran is 14.99 kg (14.31, 15.71) and urban Tehran's 3rd centile is 13.90 kg but the smoothed centile value is 15.29 kg. Overall, the fit is very good despite the roughness introduced by rounding errors.

In some ages, the upper centile of girls' weight in urban Tehran is higher that for urban Iran. This could be due to some of the wealthier families in urban Tehran having children who grow up to be heavier than their counterparts in other cities of Iran. For example in some ages such as 14 or 15 years, the 90th and 97th centiles are higher. Although Tehran might have heavier children this would not greatly affect the issues of appropriateness of urban Tehran being chosen as the baseline for the country. It should be borne in mind however, due to the small size of the sample in some age groups in urban Tehran, that the extreme centiles are derived on the basis of comparatively few observations. The comparison of the centiles for boys' weight are presented in Table C.1 and Figure C.1 (Appendix C) with similar conclusions.

The comparisons of the heights of boys and girls for urban Tehran with those for urban Iran were carried out in a similar way to that for girls' weight; results are given in Appendix C, Tables C2-3 and Figures C2-3. Besides the issue of the rounding of measurements, the first point to be made regarding the significance of some of the results of these analyses is the question, how important this significant difference is in practice. For instance, our analysis on height showed that the fixed effect for **UT** is significant ($p<0.01$, Table 6.5a). However the actual values of the 50th centiles of boys' height in urban Iran and **UT** are 122 cm and 124 cm respectively. So, at the level of $\alpha=0.05$ a significant difference is obtained by a difference of only 2 cm, which in percentage terms, is less than 1.5% and is practically negligible. Also, although it is observed that the interaction of age and **UT** ('UT*CA') is significant ($p<0.01$; Table 6.5), consideration of height of children across the age range 2-15 (Appendix C) showed that the differences are less than 2%. The 50th centile of height of urban Tehran's children in these comparisons from age of eight years and onwards was slightly higher. In the case of the extreme centiles, if there is a difference, it is very small and is less for the lower centiles than the upper centiles. In addition, smoothing the urban Tehran centiles brought them closer to the urban Iran centiles.

6.5 Summary and conclusion

1. There is no useful provincial clustering, suggesting that there is a broad regional variation in measurements of height and weight.

2. Urban Tehran's intercept at eight years, and growth rate are within the range of provincial variation.

3. Centiles of height and weight for urban Tehran children are generally within the corresponding confidence intervals for urban Iran children. Also, the point that in some ages such as puberty the 90th and 97th centiles of urban Tehrani children are high and the 3rd centile is low but within the centile confidence intervals of urban Iran children, shows that urban Tehran's centiles provide a good coverage for children in urban Iran.

In conclusion the growth of urban Tehran children provide a good normal baseline for urban areas of Iran.

Table 6.1 Cluster analysis of the coeffients of the growth models (centred age)♣ in provinces of Iran; National Health Survey 1990-92

Code	Names of Province	Number of Clusters		
		4	3	2
1	Semnan	1*	1	1
2	Chaharmahal-Bakhtiari	2	2	2
3	East Azarbaijan	2	2	2
4	Isfahan	2	2	2
5	Kordestan	2	2	2
6	West Azarbaijan	2	2	2
7	Hamadan	2	2	2
8	Bakhtaran	2	2	2
9	Kerman	2	2	2
10	Tehran	2	2	2
11	Fars	2	2	2
12	Kohkiloyeh-Boyerahmad	3	2	2
13	Boushehr	2	2	2
14	Mazandaran	2	2	2
15	Khorasan	2	2	2
16	Gilan	2	2	2
17	Zanjan	2	2	2
18	Lorestan	2	2	2
19	Yazd	2	2	2
20	Ilam	4	3	2
21	Sistan-Balouchestan	2	2	2
22	Markazi	2	2	2
23	Khouzestan	2	2	2
24	Hormozgan	2	2	2

* : The number is the label for the cluster that the province belongs to

♣ : Age was centred at 8 years in modelling weight and height

Table 6.2 Cluster analysis of the coeffients of the growth models (centred age) for rural areas of the provinces of Iran; National Health Survey 1990-92

Code	Names of Province	Number of Clusters		
		4	3	2
1	Semnan	1*	1	1
2	Chaharmahal-Bakhtiari	2	2	2
3	East Azarbaijan	2	2	2
4	Isfahan	3	2	2
5	Kordestan	4	3	2
6	West Azarbaijan	2	2	2
7	Hamadan	3	2	2
8	Bakhtaran	4	3	2
9	Kerman	2	2	2
10	Tehran	2	2	2
11	Fars	2	2	2
12	Kohkiloyeh-Boyerahmad	3	2	2
13	Boushehr	2	2	2
14	Mazandaran	2	2	2
15	Khorasan	2	2	2
16	Gilan	3	2	2
17	Zanjan	3	2	2
18	Lorestan	2	2	2
19	Yazd	1	1	1
20	Ilam	3	2	2
21	Sistan-Balouchestan	2	2	2
22	Markazi	2	2	2
23	Khouzestan	2	2	2
24	Hormozgan	2	2	2

* : The number is the label for the cluster that the province belongs to

Table 6.3 Cluster analysis of the coeffients of the growth models (uncentred age) in provinces of Iran; National Health Survey 1990-92

Code	Names of Province	Number of Clusters			
		5	4	3	2
1	Semnan	1	1	1	1
2	Chaharmahal-Bakhtiari	2	2	2	2
3	East Azarbaijan	2	2	2	2
4	Isfahan	2	2	2	2
5	Kordestan	2	2	2	2
6	West Azarbaijan	2	2	2	2
7	Hamadan	2	2	2	2
8	Bakhtaran	2	2	2	2
9	Kerman	2	2	2	2
10	Tehran	2	2	2	2
11	Fars	2	2	2	2
12	Kohkiloyeh-Boyerahmad	3	3	2	2
13	Boushehr	2	2	2	2
14	Mazandaran	2	2	2	2
15	Khorasan	2	2	2	2
16	Gilan	2	2	2	2
17	Zanjan	2	2	2	2
18	Lorestan	2	2	2	2
19	Yazd	2	2	2	2
20	Ilam	4	2	2	2
21	Sistan-Balouchestan	2	2	2	2
22	Markazi	2	2	2	2
23	Khouzestan	2	2	2	2
24	Hormozgan	5	4	3	2

Table 6.4 Random coefficient models of the National Health Survey data in Iran (4-level analyses of log(weight)[†])

Parameter	Estimate (s.e) A	Estimates (s.e) B[1]
Fixed:		
Cons	306.40 (1.24)	306.40 (1.24)
Cage	10.60 (0.06)	10.60 (0.06)
SEX	-0.60 (0.32)	-0.60 (0.32)
U/R	-3.45 (0.57)	-3.45 (0.57)
Cage2	0.05 (0.008)	0.05 (0.008)
S*CA	0.74 (0.06)	0.74 (0.06)
S*UR	-1.96 (0.47)	-1.96 (0.47)
UR*CA	-0.60 (0.07)	-0.60 (0.07)
UT	2.59 (2.17)	4.68 (1.34)
UT*CA	0.10 (0.096)	0.10 (0.096)
Random:		
level 4		
σ^2_{w0} (between provinces)	31.46 (9.76)	31.66 (9.87)
Level 3		
σ^2_{v0} (between clusters)	42.71 (2.98)	42.71 (2.98)
Level 2		
σ^2_{u0} (between families)	75.1 (3.19)	75.09 (3.19)
Level 1		
σ^2_{e0} (between children)	223.60 (2.76)	223.60 (2.76)

1: In analysis B urban Tehran is removed from the random part of level 4 (section 6.3)

†) In these analyses log(weight) is multiplied by a hundred, and age is centred at eight

Table 6.5 Random coefficient models of the National Health Survey data in Iran (4-level analyses of height[†])

Parameter	Estimate (s.e) A	Estimates (s.e) B[1]
Fixed:		
Cons	122.00 (0.29)	122.00 (0.29)
Cage	5.57 (0.02)	5.57 (0.02)
SEX	-0.50 (0.14)	-0.50 (0.14)
U/R	-2.56 (0.23)	-2.56 (0.23)
Cage2	0.08 (0.004)	0.08 (0.004)
S*CA	0.02 (0.03)	0.02 (0.03)
S*UR	-0.75 (0.20)	-0.75 (0.20)
UR*CA	-0.23 (0.03)	-0.23 (0.03)
UT	1.80 (0.74)	2.36 (0.35)
UT*CA	0.16 (0.041)	0.16 (0.041)
Random:		
level 4		
σ^2_{w0} (between provinces)	1.19 (0.44)	1.23 (0.45)
Level 3		
σ^2_{v0} (between clusters)	6.01 (0.45)	6.01 (0.45)
Level 2		
σ^2_{u0} (between families)	10.82 (0.53)	10.82 (0.53)
Level 1		
σ^2_{e0} (between children)	41.54 (0.51)	41.54 (0.51)

1: In analysis B urban Tehran was removed from the random part of level 4 (section 6.3)

†) In these analyses age is centred at eight

Table 6.6 Comparison of the centiles of weight of girls in urban Iran and urban Tehran

Percent	Age	Centile urban Iran	[95% Confidence Interval] urban Iran		Centile urban Tehran	Smoothed urban Tehran
3	2	7.996	7.65	8.35	7.92	7.87
3	3	7.996	7.37	8.68	8.35	8.72
3	4	10.00	9.61	10.41	9.19	9.68
3	5	11.00	10.59	11.42	11.00	10.82
3	6	11.00	9.59	12.61	11.71	12.20
3	7	13.00	12.16	13.89	14.99	13.83
3	8	14.99	14.31	15.71	13.90	15.29
3	9	16.99	16.46	17.53	17.40	16.57
3	10	18.99	18.38	19.61	20.00	17.85
3	11	20.00	18.64	21.46	18.97	19.38
3	12	21.99	19.66	24.60	21.99	21.46
3	13	25.68	23.60	27.95	24.87	24.38
3	14	30.69	29.38	32.06	28.13	28.18
3	15	34.98	33.92	36.08	33.68	32.49
10	2	8.997	8.70	9.30	10.00	9.27
10	3	10.00	9.77	10.24	10.00	10.20
10	4	11.00	10.65	11.35	11.00	11.26
10	5	13.00	12.74	13.25	12.00	12.45
10	6	13.99	13.65	14.35	14.89	14.06
10	7	14.99	14.54	15.46	16.00	15.73
10	8	16.99	16.52	17.48	16.99	17.32
10	9	18.99	18.53	19.45	20.00	18.88
10	10	21.01	20.45	21.57	21.39	20.60
10	11	21.99	20.89	23.15	21.01	22.72
10	12	25.99	24.87	27.16	25.99	25.51
10	13	29.99	28.75	31.28	31.31	29.15
10	14	33.98	32.67	35.34	32.00	33.57
10	15	39.01	37.83	40.23	39.01	38.12
25	2	10.00	9.70	10.31	10.00	10.47
25	3	11.00	10.73	11.26	11.00	11.50
25	4	13.00	12.72	13.27	13.00	12.68
25	5	13.99	13.70	14.29	13.99	14.01
25	6	14.99	14.62	15.38	16.00	15.75
25	7	17.74	17.39	18.10	17.99	17.49
25	8	18.99	18.55	19.43	18.99	19.27
25	9	20.24	19.74	20.76	21.99	21.20
25	10	22.98	22.42	23.56	23.99	23.45
25	11	25.99	25.28	26.72	25.00	26.25
25	12	29.99	29.10	30.90	29.99	29.80
25	13	33.98	32.88	35.12	34.98	34.17
25	14	39.48	38.27	40.74	38.01	39.08
25	15	42.99	41.62	44.40	44.47	43.66
50	2	11.00	10.68	11.33	11.00	11.61
50	3	12.00	11.71	12.29	13.00	12.79
50	4	13.99	13.69	14.30	13.99	14.14
50	5	14.99	14.67	15.32	16.00	15.67
50	6	16.99	16.61	17.38	17.99	17.55
50	7	18.99	18.59	19.39	20.00	19.43
50	8	21.99	21.52	22.48	21.99	21.52
50	9	22.98	22.49	23.49	23.99	23.97
50	10	25.99	25.41	26.59	25.99	26.94
50	11	29.99	29.27	30.73	31.00	30.60
50	12	33.98	33.07	34.92	34.98	35.06
50	13	40.00	38.86	41.17	40.00	40.21
50	14	45.01	43.61	46.46	45.01	45.50
50	15	49.00	47.44	50.62	49.99	49.89
75	2	13.00	12.58	13.42	13.00	12.71
75	3	13.99	13.61	14.39	13.99	14.12
75	4	16.00	15.63	16.38	16.00	15.71
75	5	16.99	16.64	17.35	17.99	17.53
75	6	18.99	18.56	19.43	20.00	19.55
75	7	21.01	20.58	21.44	21.99	21.67
75	8	23.99	23.46	24.54	23.99	24.20
75	9	25.99	25.42	26.57	27.00	27.33
75	10	29.99	29.25	30.75	29.99	31.20
75	11	34.98	34.00	36.00	37.00	35.90
75	12	40.00	38.81	41.22	42.99	41.36
75	13	45.01	43.69	46.37	45.01	47.25
75	14	49.99	48.42	51.62	52.24	52.79
75	15	54.98	53.21	56.80	57.97	56.83
90	2	13.99	13.57	14.43	13.99	13.88
90	3	14.99	14.60	15.40	14.99	15.59
90	4	16.99	16.64	17.35	17.99	17.52
90	5	18.99	18.55	19.43	20.00	19.72
90	6	21.01	20.49	21.54	21.01	21.89
90	7	23.99	23.37	24.64	24.80	24.34
90	8	26.10	25.51	26.70	27.99	27.44
90	9	29.99	29.06	30.95	32.58	31.39
90	10	33.01	32.11	33.94	33.01	36.31
90	11	40.00	38.35	41.73	42.99	42.16
90	12	48.81	45.61	52.23	51.41	48.68
90	13	51.67	49.76	53.66	48.37	55.25
90	14	55.98	54.11	57.90	59.97	60.89
90	15	61.00	59.07	63.00	61.99	64.60
97	2	15.37	14.89	15.88	14.99	15.49
97	3	16.99	16.37	17.63	16.99	17.65
97	4	20.00	19.25	20.78	23.59	20.08
97	5	21.01	20.43	21.59	23.99	22.74
97	6	23.99	23.04	24.98	23.12	25.19
97	7	27.00	25.94	28.10	28.99	28.14
97	8	30.53	29.17	31.96	32.10	32.03
97	9	34.98	32.48	37.67	37.00	37.10
97	10	39.01	36.54	41.66	40.56	43.41
97	11	48.23	41.48	56.06	49.10	50.73
97	12	54.98	48.6	62.19	59.79	58.50
97	13	62.30	56.56	68.62	66.48	65.79
97	14	61.99	59.83	64.22	68.71	71.47
97	15	67.02	65.02	69.07	70.73	74.72

Table 6.7 Random coefficient model of the National Health Survey data in Iran (4-level analysis of weight and height)

Parameter	Estimate (s.e) Weight*	Estimates (s.e) Height
Fixed:		
Cons	306.80 (1.20)	122.00 (0.31)
Cage	10.65 (0.13)	5.57 (0.04)
Sex	-0.71 (0.31)	-0.44 (0.13)
Cage2	0.06 (0.011)	-0.07 (0.005)
S*CA	0.74 (0.09)	0.05 (0.04)
Random:		
level 4		
σ^2_{w0} (between provinces)	28.83 (9.60)	1.50 (0.61)
σ_{w01}	-2.11 (0.83)	0.035(0.048)
σ^2_{w1}	0.24 (0.10)	0.011(0.007)
Level 3		
σ^2_{v0} (between clusters)	40.98 (3.88)	6.33 (0.61)
σ_{v01}	-1.84 (0.47)	0.06 (0.07)
σ^2_{v1}	0.73 (0.11)	0.07 (0.02)
Level 2		
σ^2_{u0} (between families)	83.1 (4.3)	11.73 (0.69)
Level 1		
σ^2_{e0} (between children)	167.10 (5.62)	32.44 (1.05)
σ^2_{e01}	3.23 (0.70)	0.89 (0.14)
σ^2_{e2}	1.74 (0.29)	0.46 (0.05)
σ^2_{e03}	12.14 (3.49)	-1.69 (0.61)
σ^2_{e04}	3.76 (1.03)	-0.58 (0.19)

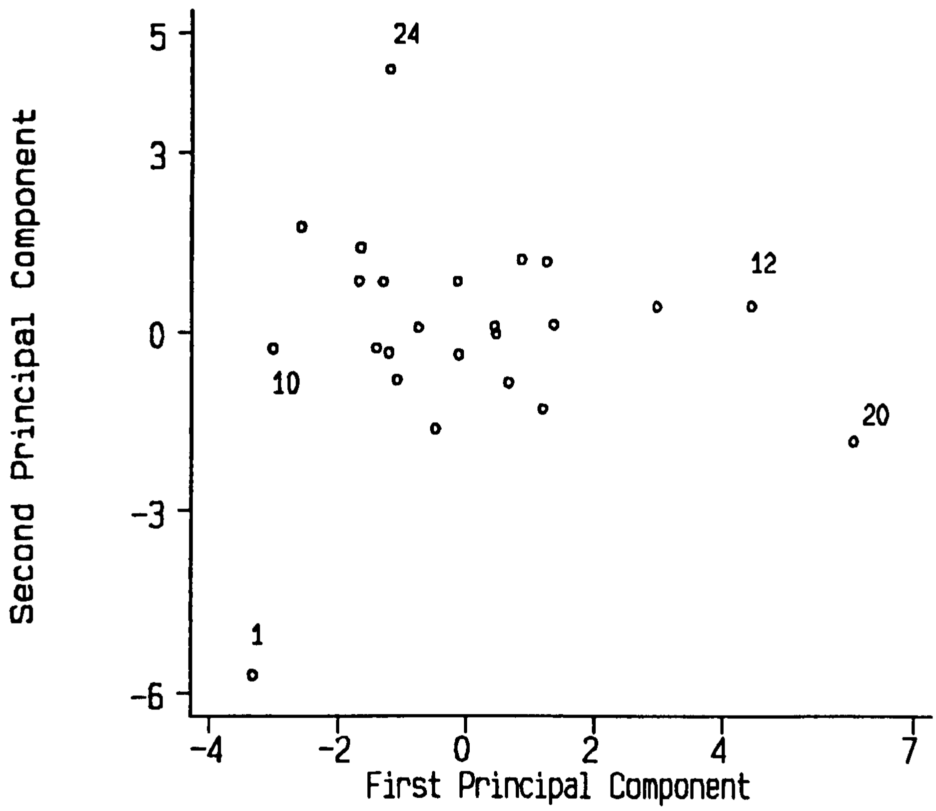
* In analysis on weight, Log-transformation of weight multiplied a hundred is the response, and in both analyses age is centered at eight

σ_{w01}=Cov('Cage' /'Cons')
σ^2_{w1}=Var('Cage')

σ_{v01}=Cov('Cage' /'Cons')
σ^2_{v1}=Var('Cage')

σ^2_{e01}=Cov('Cage' /'Cons')
σ^2_{e2} =Var('Cage')
σ^2_{e03}=Cov('Sex' /'Cons')
σ^2_{e04}=Cov('Sex' /'Cage')

Figure 6.1 Scatter plot of the first two principal components of coefficients of growth models in provinces of Iran



Each point represents a province:

1: Semnan
10: Tehran
12: Kohkiloyeh-Boyerahmad
20: Ilam
24: Hormozgan

Figure 6.2 Change of standard deviation of log(weight) of boys and girls with age in urban areas of Iran

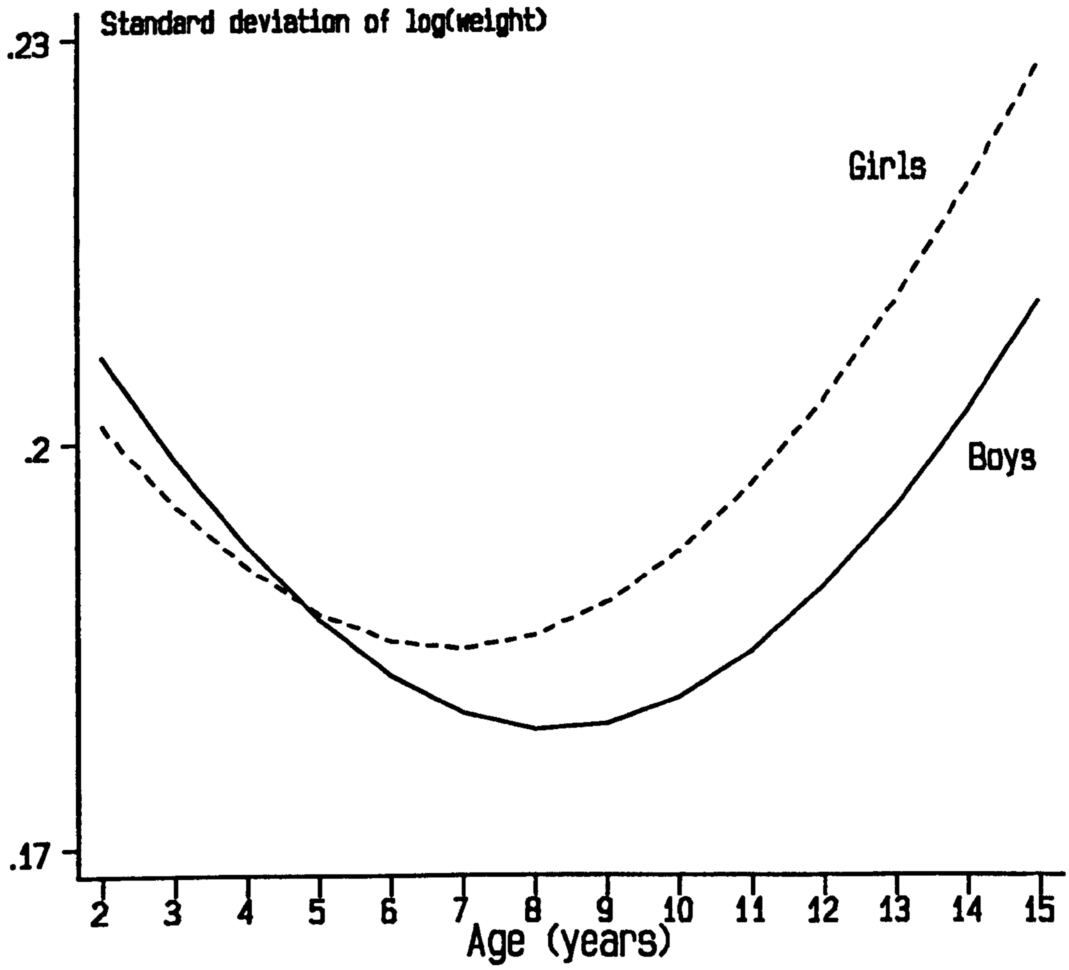
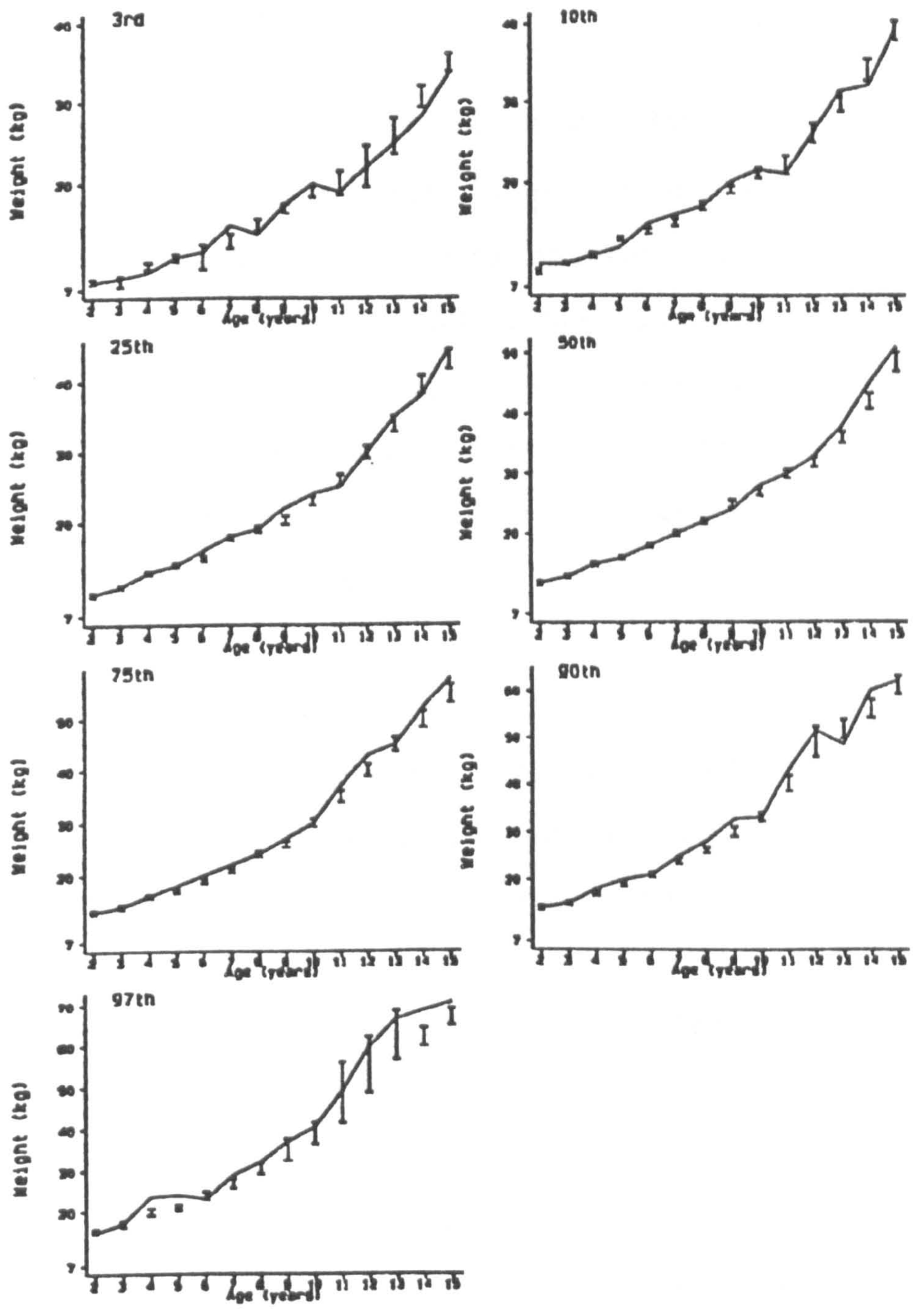


Figure 6.3 95% confidence intervals of **girls' weight** centiles in urban Iran and the corresponding raw centiles of weight of girls in urban Tehran. Centiles: 3rd, 10th, 25th, 50th, 75th, 75th, 90th, 97th



CHAPTER SEVEN

COMPARISON OF METHODS OF GROWTH CHART CONSTRUCTION

7.1 Introduction

The preliminary analysis of growth data for ages 2 to 15 presented in chapter five showed there are significant differences in growth pattern between the sexes, between urban and rural children and between provinces. In chapter six no natural clustering of growth data for different provinces was found and it was concluded that the urban Tehrani children can serve as a population norm, possibly with local adjustments.

Before presenting growth charts for urban Tehran children, the different techniques of growth chart construction that were discussed in chapter four are all used to construct the growth charts for girls' weights in urban Tehran. In order to see which of these approaches produce the best fit to the raw centiles these different techniques have been compared using a new test statistic developed for this purpose. It should be noted that the use of **fp** in modelling growth curves in this chapter was carried out in early 1995 and was independent of the EN method (Royston and Wright, 1996). Results presented here were later checked and were consistent with the results using their STATA (1995) procedures.

7.2 Transformation of the data to Normality

Apart from GROSTAT, the methods of chart construction described in chapter four start by transforming the measurements in each age group to Normality. So our first step was to use STATA to find the Box-Cox (1964) power transformation for the girls weights for each group. Then a smooth relationship between powers of λ and age (Cole, 1988a) was found using **fp** procedure in STATA.

A quadratic in age was found to be a good fit to the powers (λ). However $\phi_2(0, 0.5)$ has the highest gain, i.e. deviance reduction[A], but it was not significantly better than quadratic (p=0.7). The finding was reconsidered, weighting the **fp** regression by the number of girls in each age group, Table 7.1, but the quadratic was still suggested to be a reasonable fit to the λs. The smoothed values of λ were computed by using the equation:

$$\lambda = 1.474419 - 0.2502936 \text{ Age} + 0.0095074 \text{ Age}^2.$$

The λs at different age groups and the corresponding smoothed values are shown in Table 7.1.

Then for transformation of weight across ages, instead of

A) When working with fractional polynomial models, it is convenient to use the deviance D(1,1) associated with the straight line model $\phi(1)$ (i.e. m=1, p=1; section 4.6.1) as a baseline for reporting the deviance of other models. The Gain (G) for a model on a given data set is defined as G=G(m, **p**)= D(1,1) - D(m, **p**) transforming weight itself, weight/**M** for each year of age was transformed (**M** is the smoothed retransformed mean of $y^{(\lambda)}$). The advantage of using this transformation is that the expected value of the Box-Cox transform is then close to zero and variations in the mean of $y^{(\lambda)}$ due to changing powers of λ with age are eliminated.

The values of **M** used are the ***retransformations of the mean of*** $y^{(\lambda)}$ (m)[B] in each age group after smoothing. It is worthwhile mentioning that finding a reasonable fit to the ms is very crucial because if unsmoothed ms are used in modelling, retransformed models are unsmoothed centile curves. Also an inappropriate fit for **M** affects both the position of the 50th and all the other centiles. So different models were considered using **fp** to model m. Figures 7.1a and 7.1b show different **fp** fits (ϕ_3(3, 3) and ϕ_3(3, 3, 3) respectively) to m and shows that some of the models have problems with either the top end or at lower ages. Cubic spline at age of 12 (Figure 7.1c) was tried as well. This provides a good fit except at older ages such as 17 and 18 years, where the fit is poor. So a combination of the fits ϕ_3(3, 3) and ϕ_3(3, 3, 3), Figure 7.1d, was recognized appropriate. The values of **M** are from fit ϕ_3(3, 3) (Figure 7.1a) up to 11 years and from ϕ_3(3, 3, 3) (Figure 7.1b) from 13-18 years, the two fits give almost identical

B) m: Retransformed mean of $y^{(\lambda)}$ in each age groups
M: Smoothed m

values at 12 years. This was also practically supported by the results of modelling using other model for **M** where the fits were undesirable. Table 7.1 presents ms and the fitted values **M**, in each age group. In this chapter the transformed data are (weight/**M**)$^{(\lambda)}$ where the λs and **M**s are the smoothed values shown in Table 7.1.

7.3.1 Construction of age-related reference centiles using fp

In this section three different approaches are used to construct the reference centiles to the transformed data using **fp**. Firstly, the data is assumed to be unstructured and Altman's method (1993) is used to construct the charts. Secondly, the observed untransformed centiles were modelled individually using **fp** regression. And finally the hierarchy in the structure of the data has been taken into account using MLn (1995) to construct a growth chart by modelling the levels of variation in the transformed data.

7.3.2 Absolute residuals technique in construction of a growth chart

As described in chapter four, the absolute residuals can be used to model the standard deviation (SD) of the measurements with the covariate (age). So at the first step the mean of transformed weight/**M** was modelled by age using fractional polynomial regression. In modelling the mean of transformed measurements as the Table 7.2a shows two models $\phi_2(-2, -2)$ and cubic polynomial have the highest gains G=12.8, and G=10.8 respectively. Then as a further consideration for a better model, the weights $1/SD^2$ were used (Altman, 1993). Standard deviation (SD) was modelled using the absolute residuals as a cubic polynomial in age. Table 7.2b shows the result of this analysis for the weighted **fp**, where it can be seen that $\phi_2(-2, -2)$ appears to provide a better fit to the mean of the transformed data than cubic polynomial in age ($\chi^2_1=3.14$, p=0.09).

Cole (1994) suggested that one should consider higher degree fractional polynomials for more flexibility in modelling the ages including puberty and especially when the age band is wide. So different ϕ_3 were also examined, and $\phi_3(0, 0.5, 0.5)$ with the highest gain was chosen to model the data. Figures 7.2a and 7.2b present the fits $\phi_2(-2, -2)$, and $\phi_3(0, 0.5, 0.5)$ to the mean of the transformed weight/**M**. As can be seen in the figure, the mean of the transformed data (shown by '+', coloured red) are close to zero across different age groups. However, generally the fit $\phi_3(0, 0.5, 0.5)$ follows the mean of the data closer at most ages, but in order to see how these different models behave, it was decided to present results of both models in the original scale to see which model provides a better fit especially at puberty and in the tails of the age range.

After fitting each model the corresponding residual and absolute residuals were computed. Then the next step was to find out which model is suitable for modelling the relationship of absolute residuals to age. Because the absolute residuals have a skewed distribution, tests based on the standard error of the regression coefficients are not strictly correct. So the choice of the best model should not be based solely on P-values but should also take account of the visual acceptability of the fitted reference range and the goodness of fit. However, fractional polynomial regression, like ordinary regression is essentially an analysis of weighted means and the procedure will be robust (Box, 1962). This implies that the standard **fp** regression analysis can be used to guide us to a reasonable model for the relationship of absolute residuals to age. This gives the smoothed estimate of the SD of the transformed data by age.

For both of the above models (ϕ_2 and ϕ_3) of the mean of transformed data, the models with m=2 for the change of the absolute residuals with age were considered using **fp** procedure. In both cases, although the quadratic polynomial in age were reasonable, the two models with the highest gain were the $\phi_2(3, 3)$ and the cubic polynomial in age. These models were significantly better than quadratic models with $p<0.04$, and $p<0.001$ respectively for both models of the mean. Then, since the $\phi_2(3, 3)$ and cubic of absolute residual provide almost identical estimations of SD, the cubic polynomial of absolute residuals in age with the quadratic polynomial were chosen to present the results. Table 7.3 presents the mean and the fitted values $\phi_3(0, 0.5, 0.5)$ to the mean of the transformed data, and the cubic polynomial fit in age to the corresponding absolute residuals to smooth SD by age.

The results of the two fits to the mean of the transformed data and the corresponding fits of the standard deviation for each model are presented after back transforming the models to the original scale. Figure 7.3a shows the results of fitting $\phi_2(-2, -2)$ to the mean of transformed data and two different models of absolute residuals after back transformation to original scale. The pink colour shows the fit of cubic polynomial in age of absolute residuals and the green centiles corresponds to a quadratic model. As it can be seen clearly in this figure the cubic fit of SD is spread wider at puberty, and produces the curvature that follows the upper centiles more closely than the quadratic fit at puberty, but the problem at the top end is associated with the cubic fit to SD. Generally, the fits are almost identical at ages 3 to 9 years; Figure 7.3b shows that cubic centiles spread out more at age 2 than quadratic SD fitted centiles, and moreover, the 3rd centile shows a bend downward. The quadratic seems more reasonable at the top end and is also good at age 2 years. This might be the reason that Altman (1993) worked with a linear form when he was modelling SD in his example, although he mentioned that the higher degree polynomials were significant. He says it is unlikely that a curve more complex than a quadratic will be required to get a satisfactory fit to SD. This conclusion is borne out by our example.

Figure 7.4 shows fitted centiles when the transformed mean are modelled by $\phi_3(0, 0.5, 0.5)$; both models of SD were recomputed from the new residuals. Inspection of Figure 7.4 suggests that in some places the cubic fit of SD is better but there is still a problem at the top end of the age band. In this area probably the quadratic model is more suitable. Because the cubic model is generally a good fit, one might choose this model to fit the absolute residuals and treat the upper tail by hand.

It is worthwhile mentioning that in all of the comparisons the visual closeness of the fit to the raw centiles as well as the Chi-squared based criterion which is discussed later (7.4) were the tools for the general assessment between the models. Also, although the appropriateness of the choice of $\phi_2(-2, -2)$ and $\phi_3(0, 0.5, 0.5)$ were confirmed by refitting the weighted **fp** regression the results presented above were observed to be almost identical for both weighted and unweighted **fp** regression of the mean.

7.4 Modelling the untransformed centiles individually using fp

It is easy to fit curves to each centile separately (i.e for our study the 3rd, 10th, ..., 97th centiles at ages 2 to 18 years), as no extra assumptions are required, whereas treating the centiles as a group involves making assumptions about how the centiles are spaced relative to each other, which in turn depends on the frequency distribution of the measurements at each age. To construct a series of smooth centiles curves one can model all the 3rd centile points with an **fp** and then repeat the procedure for each of the other sets of centiles in turn. Raw centiles of girls' weight measurements in each year of age were calculated in SPSS and each set of centiles was modelled over ages with the **fp** procedure using STATA.

The result of these modelling analyses showed that among second-degree **fp**, $\phi_2(3, 3)$ has the highest gain in modelling each centile, and after that the cubic polynomial is a fit with the highest gain. However, for the 3rd and 10th centiles, ϕ_2 and the cubic polynomial do not have a gain significantly better than quadratic. For other centiles they are significantly better. Figure 7.5 shows the raw centiles and the $\phi_2(3, 3)$ centile models fitted to all centiles. This **fp** was chosen because of a higher gain in comparison with other models and because it avoids the 'end effect' of a cubic polynomial. Comparison of Figure 7.5 with Figures 7.3 and 7.4 shows a generally wider gap between 3rd and 10th centile and also 90th and 97th centile over most of the range of the data. This is especially noticeable at age 2, when the fitted 10th centile follows the observed centile which is closely grouped with the 25th and 50th centiles. The comparison suggests the possibility that the data after transformation may be leptokurtic and that this feature has not been modelled by Altman's method (1993). Overall Figure 7.5 shows that **fp** regression provides the possibility of a good smoothing of the observed raw centiles. The problem is that since the centiles are not independent the goodness of fit criteria used in the modelling are not independent and so the reference centiles obtained are in some sense an improper group of curves.

There are advantages and disadvantages to both combined and independent modelling of centiles. Combined modelling has one major advantage. The process of smoothing across centiles is effectively equivalent to estimating the cumulative distribution function of measurements at each age, and knowing this function allows each observation to be converted to its corresponding Normal Equivalent Deviate (and vice versa). This duality between centiles and Z-scores is of considerable practical value in the assessment of nutritional status (Waterlow et al., 1977).

7.5 Multilevel modelling of the reference centiles using fp

In order to take account of the hierarchial structure of the data, the models ϕ_2 and ϕ_3 which were found suitable before were considered in MLn (1995), to see how the structure of the data and the flexibility in modelling the levels of variation will alter the findings. It should be noted that in modelling the transformed weights of girls in urban Tehran our three levels of variation are clusters, families and individuals as levels three, two and one respectively.

Firstly, for a general comparison between the models of the transformed weights, the deviances ($D=-2\times\log$–likelihood) of the models, without modelling the variation in the levels as a function of age, were compared. Also as a further point in these comparisons the variance of the constant terms in level one (individual level) which is similar to the residual error in ordinary multiple regression procedure were looked at as well. Table 7.4 shows the result of these comparisons. As it can be seen from the table the model $\phi_3(0, 0.5, 0.5)$ has the lowest deviance (D=7051.4). Also the two models $\phi_2(-2, -2)$ and the cubic polynomial have approximately equal deviances and level one variances. It should be noticed that for convenience, in all of these comparisons the transformed data was multiplied by ten.

In the MLn analyses of the girls' transformed weight that follows $\phi_3(0, 0.5, 0.5)$, model (a) in Table 7.5, and cubic polynomials, model (b) in Table 7.5, were chosen as models for the fixed part, i.e., the models for the transformed mean. Because the fractional polynomial uses log(age), age can not be centred for model (a), but the effect of centring age on model (b) were examined. In these modelling procedures variation at cluster level was modelled as a quadratic function of age, and residual variation varies as a cubic-linear function of age. The family level variation was found not to vary significantly with age. It can be seen from Table 7.5 that the deviance of model (a) is lower than that for model (b).

In considering the cubic model, splining at the age of 13 was tried. This was carried out by creating an extra variable $(age-13)^3$ for ages older than 13 and 0 elsewhere (section 4.3.2). This new variable enabled us to spline the mean curve at 13 but it was observed that the corresponding fit was not better and even the obtained centile curves were worse at the top end. For the cubic model age can be centred and for comparison model (b) was fitted using **CAGE** (age-10). With this model random variation at level 1, 2, and 3 were similar those of model (b) (the model with uncentred age), but the model with **CAGE** produces wider centiles at both ends of the age range. In addition, it was found that the model (b) has overall a closer fit to the observed centiles (smaller D^*; see section 7.6). In both analyses the variation at level 2 did not change significantly with age which means that, within families, the variation of girls' weight is not changing significantly in terms of transformed data as they grow up.

It is interesting to mention that **fp** regression of the residuals corresponding to different levels of the hierarchy can be considered with the covariates (for example, age) to see how these residuals can be modelled more appropriately at different levels. In order to do this, after fitting the model of interest to the fixed part a simple constant variance is set for the random part at each levels. Then residuals can be extracted from MLn for each level together with the other covariates, for example, age. These residuals at each level can be read in STATA as separate files to see which **fp** of covariate(s) is the best fit. In the above analyses this has been examined, and it was found that the cubic and quadratic function of age have the minimum deviances for modelling the variation at the cluster and individual level of the girls transformed weight. That is the models used for the random parts in Table 7.5 are optimal with the family of **fps** of order 2.

It is observed that while modelling data in MLn, the tails of the centiles seems to behave more smoothly than when we tried to model the SD using the absolute residuals. This is may be due to the correct weighting of the data when looking at variation of levels in MLn. Figure 7.6a shows the result of model (a) in MLn with the raw centiles superimposed. In consideration of Altman's suggestion (1994) for model checking, one can superimpose $\pm S_{pread}$ or, more helpfully, $\pm kS_{pread}$ on a scatter plot of the standardized residuals against age. Figure 7.6b shows the standardized residuals for the model shown in Figure 7.6a; there is no evidence of existence of any peculiarities. The proportion of observations that fall above 97th and below 3rd centiles are 3.6% (61 values) and 3% (51 values) respectively, also the corresponding percentage above 90th and below 10th centiles are 9% and 10%, and were similar in the following three age groups: 2 to 6, 7 to 12, and 13 to 18 years old.

Also, Figure 7.6c shows the Normal plot of SDS, and that the distribution of SDS is close to Normal. It is not wise to worry about moderately small p_value when we look at the distribution of SDS, because in large samples slight deviations from Normality can lead to significant test results. However, for more consideration of this the age band were divided to three groups: 2 to 6, 7 to 12, and 13 to 18 year olds. Then it was observed that the distribution of SDS in the last two age groups are fairly reasonable (except age 13; p=0.01) but there is some abnormality in the first age band. For that group the Shapiro-Wilk W test (STATA Reference Manual, 1993) was carried out on SDS in single years of age and there was evidence of departure from Normality in 4 and 6 year olds data ($p<0.001$, and $p=0.001$) which may be due to the existence of some undetected outlying observations in these age groups. Moreover, in order to check the assumption of Normality of the random terms at level two and level three, the diagnostic residuals were examined and found to be close to Normal.

7.6 Comparison of the different models

Of the methods of modelling centiles considered in this chapter, only the MLn model takes account of the hierarchical data structure. However, it is discussed in section 7.9 that provided intra-class correlations are less than 0.8 observed centiles provide unbiased estimates of true centiles and that for the purpose of modelling the centiles the data structure can be ignored. So we now compare four methods of modelling the girls weight data. These are Healy's method using spline procedure (Pan and Goldstein, 1990; this model is called GROSTAT in this part and fuller details is discussed in chapter 8); Cole's method (1992) using the STATA version of his program (section 4.4.1.3); fractional polynomials and Altman's technique for modelling the SD; and the MLn method. These methods are compared to find out which gives the closest fit to the observed centiles.

Royston and Thompson (1995) proposed a technique for comparison of non-nested models, but this is mainly concerned with the comparison of models of the mean of the data. Here the interest is to see which model produces the overall best fit to the set of observed centiles. This is done by visual inspection, examination of residuals, and by means of a new test statistic developed for this purpose.

Figure 7.7 presents the results of different modelling approaches with the raw centiles superimposed. The black curves show the fit corresponding to Cole's method (LMS) obtained using edfs (4.4.2) 4, 6, and 4 for L, M, S Curves respectively on log-transformed weights (the choice of edf controls the amount of smoothing). This choice provided the optimum fit, assessed by comparing penalized log-likelihoods. Comparing the fits: first, by visual inspection it appears that the GROSTAT model is a closer fit to the centiles especially at puberty where the bulge of the curves follows the upper raw centiles (cyan) as well as the curvature of the lower centiles. In particular the overall fit of the 50th centile is much closer over most ages where the other methods, for instance, over 12 to 16, produce poorer fits. This informal approach is very useful as an initial check but is not analytical, and more formal approaches may be preferable.

Next, it is reasonable to look at Z-scores of different fits to see which is the most appropriate one. Figure 7.8 illustrates the Q-Q plots of the four different models. As can be seen from the figures (7.8a-d), the Z-scores of the GROSTAT model is a fairly straight line and there is not any remarkable sign of departure from expected Normal scores. However, in the Z-score plots of the residuals of the other models there are sign of departure from expected Normal scores. This is due to the advantage of simultaneously modelling the skewness and *kurtosis* of distributions across the ages, which are involved in the GROSTAT model, but not in the other methods.

The correlation between Z-scores and the corresponding Normal scores provides a test of Normality (W test) which may be oversensitive to a few outlying observations. This test gives $p=0.14$ for the GROSTAT model and $p<0.001$ for the each of the other models. When computing Z-scores of observations in GROSTAT models, Z-score greater than +3 or less than -3 are rounded to the ±3. Figure 7.8a shows that only 5 of 1702 observations were truncated in this way to -3 and 1 Z-score to 3 and that this procedure has not affected the modelling (to be discussed in chapter 8).

Third, one might wish to compare different methods by comparing the numbers of observations falling between the estimated centiles curves with their model-based expectation (Grid test; Healy, 1988). If a model is better the variation between observed and expected is smaller. Generally, an extension of this assessment is to categorise the age range into *m* columns, say of roughly equal size or years in our study. After fitting k centiles, in m age groups we obtain a *m* by *(k+1)* table of observed and expected cells with *m* constraints due to fixing the totals in each age group. So the nominal degrees of freedom are m(k+1)-m=km. However, the estimation of parameters in the model from which the fitted centiles were derived will further constrain the χ^2 statistic and reduce the effective degrees of freedom, resulting in a conservative test if no adjustment is made. Royston (1995) reported that, if there are more age groups than parameters, (*p*), in the mean curve (M), that is *m*>*p*, simulations suggest an adjusted df of $[k(m-\frac{1}{2}p)]$ brings the type I error rate closer to nominal values.

We compared the GROSTAT model with the MLn model using the Healy's grid test with Royston's correction to the degrees of freedom. Table 7.6 shows the observed numbers for the seven centiles in each of the age groups from 2 to 18 and the corresponding cell expectations. The Pearson χ^2 values are 151.1 for the GROSTAT model and 247.8 for the MLn model. The conservative test on 7×17 df gives the

Normal approximation (Fisher and Yates[c], 1948) Z=1.99 for GROSTAT and 6.87 for the MLn model, p=0.023 and p<0.001 respectively. So without Royston's correction neither set of centiles fit well, but the GROSTAT model appears to fit substantially better than the MLn model. Reducing the degrees of freedom using Royston's correction increases Z for GROSTAT to 3.05 and for the MLn to 8.05. The grid test has not been reported for Cole's and Royston-Altman models since curves are so similar to MLn curves.

Although it is concluded that the GROSTAT model is a better fit, there are some drawbacks with the Grid test. First, the proposed adjustment for degrees of freedom only considers adjustment for modelling the mean of the data not the SD, and skewness which have been carried out in all the above methods and also kurtosis for GROSTAT. Second, in the case of no observations in the space between the fitted centiles from different models, grid test is not measuring which of the models is a closer fit to the raw centiles. For example, if the raw centile is 5kg and the fitted centiles for two models are 5.1 and 5.8 the grid test will only distinguish between them if there are observations in the interval (5.1 and 5.8). Third, in cases of rounding of measurements, as in our study, the borderline observation may be misclassified. Also, because of the grouping, small

---

C) $\sqrt{2\chi^2}-\sqrt{2n-1} \approx N(0,1)$; $n=df$

changes in the fitted curves may produce large changes in the test statistic. A new test statistic has been developed to overcome these problems.

7.7 The D* test statistic

A new 'Chi-squared type' criterion is introduced to compare different models. Basically, this criterion measures how close the fitted values, f_i, are to the observed centiles, C_i. In order to take account of the variation of the observed centiles each difference is weighted by $w_i=1/Var(C_i)$. When age is grouped, the criterion is defined as

$$D^*=\sum_j \sum_i w_{ij}(c_{ij}-f_{ij})^2$$

where i refers to the centile, i, and j to the age group. As stated before (4.2.2) the distribution of C_{100p} (<u>100 pth centile</u>) is approximately Gaussian with variance

$$VAR(C_{100p})=\frac{p(1-p)}{n\,f^2}$$

if the observations are Normally distributed the corresponding ordinate of the frequency density f is given by

$$f(p)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}z^2} \quad ; \quad z=\varphi^{-1}(p)$$

and σ is the standard deviation of the data in the corresponding age group j. When evaluating D*, in order to use the appropriate Normal estimate of frequency density at the ith centile and to provide the same scale for comparison, the raw centiles and the GROSTAT model were transformed to Normal using the transformations described in section 7.2.

The D^* criterion has many advantages over the others, for instance the problem of degrees of freedom is overcome because it depends on the joint distribution of observed centiles (see below), and also the point about the real place of rounded measurements in the contingency tables is no longer troublesome because this criterion is a comparison of closeness of the centiles instead of counts.

Table 7.7 presents the values of the D^* statistic for the seven models based on the four methods of modelling girls' weights that have been described. Table 7.7 shows that the GROSTAT model has substantially the lowest D^* which is an indication that the GROSTAT model is overall the most appropriate one, with the model based on Cole's method the next best choice.

7.7.2 Distribution of D^*

It can be shown that the joint distribution of percentiles C_i and $C_{i'}$ is Bivariate Normal in the limit (Kendall, 1952). Extending this idea for a set of centiles, one can conclude that the joint distribution of the centiles vector is Multivariate Normally distributed. If X_1, $X_2,..,X_n$ is a random vector following a Multivariate Normal distribution with expected value vector (ξ_1, $\xi_2,...,\xi_n$) and the variance-covariance matrix V, let the quadratic form associated with the symmetric matrix A be defined as

$$Q(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}X_iX_j.$$

Thus the terms $\sum_i W_{ij}(c_{ij}-f_{ij})^2$ relating to each of the age groups j, has a quadratic form of Normally distributed variables, which on the null hypothesis is distributed as a central quadratic form with $\xi_1=\xi_2=\cdots=\xi_n=0$. Suppose, we name the terms $W_{ij}(c_{ij}-f_{ij})^2$, z, then for age group j the corresponding quadratic component of D^* is of the form $Q(Z)=Z'AZ=\sum_i Z_i^2$ with symmetric identity matrix A=I. The probability distribution of Z is as

$$P_Z(z) = (2\pi)^{-\frac{1}{2}n}|V|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}z'V^{-1}z\right)$$

where $V=\rho$, the correlation matrix of the Zs. $|V|$ is the determinant of V, and Z′, z′ are the transpose of Z, z, respectively.

Distribution of Q is that of $\sum_{i=1}^{n}\lambda_i W_i^2$ where the W_i are independent unit Normal (N(0,1)) variables, and the numbers $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_n$, are the eigenvalues of $VA=VI=\rho$. It follows that attention can be focused on the distribution of $Q(W_i)=\sum_{i=1}^{n}\lambda_i W_i^2$ where the W_i's are mutually independent unit Normal variables. Mathai (1992) discusses different approximations for Q. The Patnaik's (1949) Chi-squared type approximation is one of the best where the distribution of Q is replaced by $c\chi^2_v$ which c and v are chosen so that Q and $c\chi^2_v$ have the same first two moments, that is:

$$E(Q)=E(c\chi^2_v),\ Var(Q)=Var(c\chi^2_v).$$

I.e. $\sum_{i=1}^{n}\lambda_i=cv,\ \sum_{i=1}^{n}\lambda_i^2=c^2v$. from these relations one gets

$$c=\frac{\sum_{i=1}^{n}\lambda_i^2}{\sum_{i=1}^{n}\lambda_i},\qquad v=\frac{(\sum_{i=1}^{n}\lambda_i)^2}{\sum_{i=1}^{n}\lambda_i^2}.$$

Hence, $$\Pr\ (Q\le x)\approx\Pr(\chi^2_v\le x/c). \qquad (7.4)$$

It can happen that v is fractional, then the notation χ^2_v will mean a gamma variable with parameters $\alpha=v/2$ and $\beta=2$. In sum, the distribution of D* over the age groups is the sum of these χ^2s.

We note that for the Z's corresponding to the centiles i and i': $Cor(C_i,C_{i'})=\frac{p_i q_{i'}}{\sqrt{p_i q_i p_{i'} q_{i'}}}$ (Kendall, 1952), i.e. is independent of f_i or $f_{i'}$ and is the same for all age groups. So that c and v for each age group depend only on V, i.e. on the choice of the centiles used in the modelling. For the 7 centiles that we use $c\approx2.7$ and $v\approx2.6$. Note that in the calculation of eigenvalues of V the λ_7 is zero due to the correlation of the centiles. From the distribution of the D* we have for the GROSTAT model for the ages up to 17 years (raw centile at eighteen are not reliable) df=16×2.6≈42, and χ^2 is D*/c=175.08/2.7=64.8 which gives p<0.02. Thus again the goodness of fit test for the GROSTAT model is significant at around the 2% level but the D* statistics for other models shows that the fit is much worse, as shown in the last two columns of Table 7.7. However, for the aim of comparison of closeness of the fits, it is of more interest to compare the magnitude of D*s which are weighted sum of squares of the differences between observed and fitted values. Due to the sampling scheme, observations are correlated in families and clusters but this will have very little effect on the independence of χ^2 across age groups in our study because any such correlations are very small.

7.7.3 Evaluation of distribution of D* and the effect of using centiles in estimation of standard deviation

In Healy's method (1988) the SD is calculated from the centiles. This suggests that D*, which measures the distance between observed and fitted centiles, may be biased in favour of Healy's method. To investigate this, a simulation study of random samples of the standard Normal distribution was set up, and sets of seven fitted centiles have been estimated by both approaches: First, by standard estimation: i.e. from the sample mean and SD, calculated in the usual way. Second by regressing the observed centiles on the corresponding z_i and deriving the estimated centiles as fitted values on the regression line. Note that the intercept and slope of the line estimate the 50th centile and the sample SD. Then, the D^* were calculated for both sets of fitted centiles. The procedure was replicated 1000 times for sample sizes n=99, 199, 499, 799, 999. These sample sizes were chosen to give simple unbiased raw centile estimates.

Table 7.8 shows the mean (SD) of D^* for different sample size. The mean of the distributions across all different sample size are very close although in some cases significantly different. This shows that there is no tendency for D^* from Healy's approach to be smaller than the standard method and confirms the robustness of the D^* statistic. The distributions of the D^* for both cases are shown in Figure 7.9. It can be seen from the figure that the distributions of D^* are like that of χ^2. Also, in most cases the means and the variances of the distributions are near 3 and 6 which suggests that the degrees of freedoms of these χ^2 distributions are nearly 3. The standard set of seven centiles were estimated in the calculation, derived from estimates of mean and standard deviation. The sample size in each simulation was fixed. Also another degree of freedom was lost because centiles are symmetric and correlated. So 3 seems a reasonable value for the df and this is consistent with the previous approximations (7.6.2;

where $v=(\sum_{i=1}^{n}\lambda_i)^2/(\sum_{i=1}^{n}\lambda_i^2)=49/18.9\approx 3$.

7.8 Further exploration of models

7.8.1 Further investigation in modelling using fp

Cole's (1994) suggestion of applying a higher degree **fp** when the age range is wide or includes puberty, as well as the expansion of the Ω when the powers of the suggested **fp** model is in the extremes of Ω like $\phi_2(-2, -2)$, have both been considered to see whether the proposed models is still reasonably appropriate.

First, for m=2 (ϕ_2), a wider range of powers from -9 to 9 for modelling the mean of transformed data were tried. Even with weighted **fp** procedure, the finding has not altered, and still the $\phi_2(-2, -2)$ has the highest gain amongst ϕ_2 models. This in fact supports Royston and Altman's idea (1994) about the appropriateness of suggested range for Ω.

Then the third-degree **fps** as well as fourth-degrees with the extended Ω={-9, -8,..., 9} in both unweighted and weighted **fp** regression were considered. Amongst the **fp** of degree three in expansion of Ω the $\phi_3(-5, -5, -5)$ and among the **fp** of degree four $\phi_4(2, 3, 3, 3)$ were the best in weighted **fp** regression using $1/SD^2$ as weights. Although the idea of increasing the degree of **fp** and the domain of Ω is theoretically worth pursuing to try to get a lower deviance (despite the cost of extra computing), practically the problem of over fitting at the end of the distribution or even sometimes over fitting within the range of the data like Figure 7.10a for $\phi_3(-5, -5, -5)$ at age 4 years is unavoidable, and generally is not going to provide an appreciably better fit. Figure 7.10b shows the model $\phi_4(2, 3, 3, 3)$ in both approaches: first, using MLn (green) and modelling residuals at different levels as Table 7.5 ; and then Altman's method with modelling SD as quadratic of age (blue). It can be seen from the figure that up to 16 years the centiles are almost identical and similar to $\phi_3{}^{D}$ (pink) but at the end of age band the high order fractional polynomial (ϕ_4) start to bend down.

Also, since the sample size at 18 year olds was less than the other groups (Table 7.1), **fp** regression for looking for the appropriate fit was carried out, excluding the last age group. And it was observed that still $\phi_2(-2, -2)$ has the highest gain amongst ϕ_2 which is consistent with the result of weighted **fp** regression. In addition, in MLn modelling of $\phi_3(0, 0.5, 0.5)$ at level one the variance is modelled as a linear-cubic function of age (Table 7.5) and there is no problem in the tails. However, **fp** regression of the absolute residuals did not show that linear-cubic has the highest gain, and when it was tried the upper tail of the

D) ϕ_3 (0, 0.5, 0.5); model **a**, Table 7.5

curves did not fit appreciably better, but it was better than the cubic fit of SD.

It is worth while mentioning that the work of this chapter using fp was done in early 1995 before the macros of EN method (Royston and Wright, 1996) became available. After receiving these macros from the authors, further examination of previous findings for girls' weights data were carried out in two ways. First, instead of log transformation of the raw data at the first step (section 4.6), the girls' transformed data was remodelled using XRIGLS macro and the appropriateness of models $\phi_2(-2, -2)$, and $\phi_3(0, 0.5, 0.5)$ in models with m=2, and m=3 were confirmed. Also, investigations using their suggested 3-parameter Exponential (EN) model and the **fp** $\phi_3(0, 0.5, 0.5)$ showed that the test of the hypothesis $\gamma=0$ is not significant (p=0.46), which means that the applied power transformations were reasonable for removing the skewness from the data. Furthermore, when fitting a 4-parameter modulus exponential (MEN) model the constant kurtosis was found significant (p<0.01), but linear and higher order trends in kurtosis were not significant. The resulting 3rd and 97th centiles were much too wide at ages 2, 3 and 4 years due to incorporating constant kurtosis in all age groups. In addition, the fitted curves start to bend down from 17 years.

The fourth-degree fractional polynomial $\phi_4(2, 2, 3, 3)$ was also suggested as a good fit using their macro's. This has almost identical deviance to our $\phi_4(2, 3, 3, 3)$. However, using their proposed 'Augmentation test[E]', the fit of the mean was not significant (p=0.11) for fp $\phi_4(2, 2, 3, 3)$, but like the blue coloured curves in Figure 7.10b the curves were overfitting the upper tail of the distribution. These analyses show that using their macro's and the iterative maximum likelihood procedure produces the same results as had previously been obtained.

Second, we examined whether the results would have been altered if Royston and Wright's (1996) two steps of transformations and modelling the time varying skewness and kurtosis using a 3-parameter exponential (EN) or a 4-parameter modulus exponential (MEN) had been used on the raw data. The model $\phi_3(-0.5, 3, 3)$ with S curve as a linear function of log(age) and skewness and kurtosis as linear functions of age was found to be an appropriate fit when using 4-parameter modulus model on log(weight). The D^* statistic for this model was 250.1 and the result is very similar to those of the LMS method when applied to the original data. Still GROSTAT model has the lowest D^*.

E) Augmentation test suggested by Royston and Wright (1996) for testing reasonableness of the M, S, G curves

7.8.2 More exploration using Cole's method

In addition, Cole's method was used for further investigation using girls' height data. First, regarding the example of girls' weight the obtained smooth curves from log-transformed data were better than curves derived from the original measurements showing that the power transformation does not always produce the best distribution to start modelling the data. Second, while working on girls' heights data, after finding that using edf's 4, 6, 4 optimizes the likelihood, it was observed that curves were overfitting the upper tail of the age band from age of 15 and onwards. Hence, in order to obtain an appropriate fit to the data, in each case careful consideration of the ways of modelling irregularities in cross sectional distributions of the measurements as well as the mean curves is required. These problems are discussed in more detail in chapter 8.

7.8.3 Appropriateness of 'weight/M' for transformation

It is stated in section 7.2 that the transformation of the data to Normality is the crucial assumption underlying the methods in which centiles are estimated from the mean and SD. Also, it is mentioned that it is better to avoid using power transformed weight itself, and the suggestion was to use a function of the variable which stabilizes the oscillation of the transformed mean like 'weight/**M**'. Then the mean of the transformed data is nearly zero in each age group (Table 7.3).

In order to examine whether this is really important, we modelled the data using weight$^{(\lambda)}$. Figure 7.11a shows the transformed data and the fit $\phi_2(0, 2)$ to the mean (in some examples the transformed mean of $y^{(\lambda)}$ may fluctuate widely). Then the absolute residuals (Figure 7.11b) were modeled using $\phi_2(0,1)$ and the final chart is shown in Figure 7.11c. These curves are not an appropriate fit, although the curves fitted to the mean and SD appear satisfactory. It is worthwhile mentioning that surprisingly another fit of SD like quadratic or cubic in this example produces centile curves that cross. Furthermore, a third-degree **fp** for the mean ($\phi_3(-1, -0.5, 3)$) and to SD ($\phi_3(-1, -0.5, -0.5)$) were tried but the result was no better (Figure 7.11d). This confirms the advantage of transforming y/**M** rather than the observations themselves.

In the analyses described in sections 7.3 onwards, after transformation of 'weight/**M**', we modelled the remaining variation of the mean of transformed data with age using **fp**. Then in order to see if using any other possible good **M** curve such as Cole's M curve or GROSTAT fit could have produced better results, the analyses were repeated using these different M curves in our transformation procedure.

The results in both cases were no better. The point is that even using an extended family of **fp** curves, modelling such a wide age range sometimes needs other methods like splining curves which are considered in chapter 8.

7.9 Discussion

The analyses in this chapter compare fitted centiles with observed centiles. But the data are hierarchically structured, this raises the question as to whether the structure affects the observed centiles.

Let α denote the α centile and z_α the αth centile of N(0,1) distribution. Suppose data are in I clusters with the J observations per cluster, and that an observation

$$y_{ij}= au_i+bv_{ij}$$

where U and V are independent random variables distributed as N(0, 1). And let $a^2+b^2=1$, so that y_{ij}~N(0, 1) for all i, and j. For a given k and j=k, the αth centile of the I observations y_{ij} is given by $y_{\alpha k}$ provided $I\alpha$ and $I(1-\alpha)>1$. Also, since y_{ik} are independent N(0, 1)

$$E\ (y_{\alpha k})\ =\ z_\alpha$$

So the αth centile of the all the data is an average of J estimates of the centile. So for fixed cluster sizes the sampling structure does not affect the estimates of the population centiles. It can, however, be shown that the SE's of the estimates are increased by the sampling structure. These conclusions also hold with variable cluster sizes as shown by simulations with intra-class correlation=$a^2/(a^2+b^2)$, ranging from 0.0 to 0.9 and cluster sizes varying randomly from 1 to 7, as shown in Figure 7.12. The simulations also show that using an estimate of SD which ignores the data structure underestimates the true SD by less than 1% when the intra-class correlation is less than 0.5 (Carpenter, 1996).

These findings validate the use of our D^* statistic to assess the goodness of fit of observed and fitted centiles. They also validate our efforts to model observed centiles which is the basis of the GROSTAT methodology.

To get some general idea of intra-cluster correlation plus intra-family correlation in our data we refer to Tables 6.4 and 6.5 which give average variances for a 4-level model for ages 2 to 15. These tables roughly suggest that the average of:

$$\frac{\text{Between cluster variance + between family variance}}{\text{Total (level 1 + level 2 + level 3)}}$$

is 0.35 for weight and less than 0.30 for height. These findings also were confirmed by a 3-level analysis of weight and height using urban Tehran's data where corresponding average intra-class correlations were 0.41

and 0.38 respectively.

Thus the relative weak data structure explains why centiles constructed using SD's which ignore the data structure, e.g. Royston and Wright method (1996), and Cole's (1992) method give results which are very similar to those derived from a multilevel model.

In conclusion it appears from these analysis that the GROSTAT method is most appropriate to structured data and has produced the model which best fits the observed centiles of girls' weight.

Table 7.1 The powers (λ) for Normal tranformation of Y=weight/M and their fitted values; retransformed mean of $Y^{(\lambda)}$, m, and M[1]; girls' weight, urban Tehran

Age in years	λ	Smoothed λ	m (kg)	M (kg)	No. in sample (n_i)
2	1.2650	1.0119	11.660	11.807	82
3	0.0814	0.8091	12.575	12.467	111
4	0.2483	0.6254	14.534	13.549	139
5	0.2900	0.4607	15.934	15.078	128
6	1.2199	0.3150	17.542	17.055	128
7	-0.2512	0.1883	19.800	19.458	121
8	0.3030	0.0806	21.832	22.250	129
9	-0.4434	-0.0081	24.504	25.379	123
10	-0.9112	-0.0777	26.634	28.780	113
11	0.1108	-0.1284	30.498	32.377	98
12	-0.2186	-0.1600	35.850	36.122	105
13	-0.1417	-0.1726	39.891	40.211	82
14	0.5061	-0.1662	44.253	44.132	70
15	0.3457	-0.1408	50.034	47.677	89
16	0.1913	-0.0963	50.697	50.618	56
17	-0.5183	-0.0329	52.196	52.698	74
18	-0.2524	0.0496	53.254	53.647	54

1: M= Smoothed value of m
$N=\Sigma n_i=1702$

Table 7.2 Modelling mean of transformed data by age using fractional polynomials (fp); girls' weight, urban Tehran

a) <<Unweighted fp>>

****

```
. fp transwt age
```

MODELS, POWERS (p), DEVIANCES (D) and GAINS (G) for:
Y = transwt, X = age.

	(*) Base model	Linear	Quadratic	Cubic	BoxTid	df(2)	df(4)
p	--	1	1, 2	1, 2, 3	1, 1	.5	-2, -2
D	-604.647	-612.192	-617.571	-623.031	-614.960	-612.975	-625.014
G		0.000	5.379	10.839	2.768	0.783	12.822

b) <<Weighted fp>>

****

```
. fp transwt age [weight=wt*]
```

MODELS, POWERS (p), DEVIANCES (D) and GAINS (G) for:
Y = transwt, X = age.

	(*) Base model	Linear	Quadratic	Cubic	BoxTid	df(2)	df(4)
p	--	1	1, 2	1, 2, 3	1, 1	.5	-2, -2
D	-615.222	-624.652	-627.794	-635.571	-625.772	-624.965	-638.716
G		0.000	3.142	10.920	1.121	0.313	14.064

*: wt=1/SD

SD was modelled using the absolute residuals as a cubic polynomial of age

Table 7.3 Mean of the transformed data[1] and the corresponding fit $\phi_3(0, 0.5, 0.5)$, standard deviation of the residuals and fitted standard deviation using cubic polynomials of absolute residuals in age; girls' weight, urban Tehran

Age in years	Mean of transformed data	Fitted mean $\phi_3(0,0.5,0.5)$	SD of the residuals	Fitted SD 'absolute residuals'
2	-0.0124254	-0.0247656	0.1558274	0.1746442
3	0.0086325	0.0443053	0.1790520	0.1701226
4	0.0717286	0.0538285	0.2148302	0.1695805
5	0.0559506	0.0438598	0.1983041	0.1722757
6	0.0282733	0.0280212	0.1789983	0.1774658
7	0.0174927	0.0117836	0.1787451	0.1844085
8	0.0270341	-0.0024895	0.1932938	0.1923616
9	-0.0351149	-0.0137745	0.1871490	0.2005828
10	-0.0777165	-0.0216746	0.1819195	0.2083296
11	-0.0599814	-0.0260990	0.2590309	0.2148599
12	-0.0075960	-0.0271100	0.2574696	0.2194314
13	-0.0079447	-0.0248449	0.2078519	0.2213017
14	0.0027793	-0.0194714	0.2328725	0.2197285
15	0.0480403	-0.0111710	0.1876085	0.2139696
16	0.0015204	-0.0001233	0.1717465	0.2032827
17	-0.0095601	0.0134968	0.1724551	0.1869254
18	-0.0073172	0.0295252	0.1974277	0.1641554

1: $(\text{Weight}/M)^{(\lambda)}$

Table 7.4 Comparison of deviance and residual variance of different models in MLn; girls' transformed weight[A], urban Tehran

Model of transformed mean	Deviance (D)	Level one variance σ^2_{e0}
$\phi_3(0, 0.5, 0.5)$	7051.5	2.45
$\phi_2(-2, -2)$	7069.5	2.49
Cubic polynomial	7071.9	2.51
Quadratic polynomial	7079.1	2.52

A) In these analyses transformed weight is multiplied by ten

Table 7.5 Coefficients of two models for which the fixed part is modelled by ϕ_3 (0, 0.5, 0.5) or a cubic polynomial and the random variation at level 1 is a cubic function of age and level 3 is a quadratic function of age; girls transformed weight[A], urban Tehran

a) Fixed part Model ϕ_3 (0, 0.5, 0.5)			b) Fixed part Cubic polynomial		
Parameter	Estimate	(s.e)	Parameter	Estimate	(s.e)
Fixed			**Fixed:**		
Constant	62.79	(9.952)	Age	0.2151	(0.0442)
Ln(Age)	28.46	(4.5)	Age^2	-0.0362	(0.0073)
$Age^{0.5}$	-66.24	(10.53)	Age^3	0.0014	(0.00029)
$Age^{0.5}(\ln(Age))^2$	11.12	(1.78)			
Random:			**Random:**		
Level 3			***Level 3***		
σ^2_{v0}	1.219	(0.3211)	σ^2_{v0}	1.192	(0.325)
σ_{v01}	-0.07553	(0.02695)	σ_{v01}	-0.0732	(0.00027)
σ^2_{v1}	0.00662	(0.00266)	σ^2_{v1}	0.0064	(0.0027)
Level 2			***Level 2***		
σ^2_{u0}	1.179	(0.158)	σ^2_{u0}	1.135	(0.1595)
Level 1			***Level 1***		
σ^2_{e0}	0.3651	(0.3368)	σ^2_{e0}	1.215	(0.2394)
σ_{e01}	0.1629	(0.0308)	σ^2_{01}	0.03678	(0.0076)
σ_{e03}	-0.00038	(0.00009)	σ_{e03}	-0.00099	(0.000218)
-2(loglikelihood)	7019.5		-2(loglikelihood)	7041.3	

σ_{e03}=Cov(Age^2, Age) which referes to Age^3
σ_{e01}=Cov(Age, cons)

σ_{v01}=Cov(Age, cons)
σ^2_{v1} =Var(Age)

A) In theses analyses transformed weight is multiplied by ten

Table 7.6 Comparison of number of observations falling between the estimated centiles curves for two models: GT(GROSTAT; 3 3 2 3 1 2) and MLn (ϕ, (0, 0.5, 0.5)) with the expected number (E); girls' weight, urban Tehran

Age (yr.) No.	2 82	3 111	4 139	5 128	6 128	7 121	8 129	9 123	10 113	11 98	12 105	13 82	14 70	15 89	16 56	17 74	18 54	Total 1702
Centiles	GT MLN E	GT MLN E	GT MLN E	GT MLN E	GT MLN E	GT MLN E	GT MLN E	GT MLN E	GT MLN E	GT MLN E	GT MLN E	GT MLN E	GT MLN E	GT MLN E	GT MLN E	GT MLN E	GT MLN E	GT MLN E
>97	1 1 2.5	2 2 3.3	7 9 4.2	4 6 3.8	1 1 3.8	5 5 3.6	3 3 3.9	2 2 3.7	1 1 3.3	2 6 2.9	4 8 3.1	3 5 2.5	0 3 2.1	2 2 2.7	0 0 1.7	2 2 2.2	3 3 1.6	42 59 51.1
90–97	13 4 5.7	6 2 7.7	9 7 9.7	13 11 9.0	10 5 9.0	7 4 8.5	10 9 9.0	11 10 8.6	3 3 7.9	8 6 6.9	12 11 7.4	3 1 5.7	5 6 4.9	5 6 6.2	4 4 3.9	3 2 5.2	1 0 3.8	123 91 119.1
75–90	13 9 12.3	12 16 16.7	19 9 20.9	21 21 19.2	23 16 19.2	19 13 18.2	16 8 19.4	17 13 18.5	15 10 17.0	17 13 14.7	16 15 13.8	2 9 12.3	12 13 10.5	17 27 13.4	7 8 8.4	6 5 11.1	6 7 8.1	238 212 255.5
50–75	13 26 20.5	36 17 27.7	33 43 34.8	30 30 32.0	34 46 32.0	33 42 30.3	43 36 32.3	38 17 30.8	34 26 28.3	24 19 24.5	15 13 26.3	31 30 20.5	16 17 17.5	24 21 22.3	15 21 14.0	17 19 18.5	17 11 13.5	453 434 425.5
25–50	21 21 20.5	24 43 27.7	46 46 34.8	26 35 32.0	32 32 32.0	34 34 30.3	24 40 32.3	28 54 30.8	36 32 28.3	19 23 24.5	37 30 26.3	28 22 20.5	19 15 17.5	20 17 22.3	8 12 14.0	25 28 18.5	10 16 13.5	437 500 425.5
10–25	14 14 12.3	14 14 16.7	11 11 20.9	21 12 19.2	16 20 19.2	12 12 18.2	20 20 19.4	21 18 18.5	17 30 17.0	14 12 14.7	14 14 15.8	11 9 12.3	9 7 10.5	13 12 13.4	17 7 8.4	13 12 11.1	9 9 8.1	246 233 255.3
3–10	5 5 5.7	14 14 7.7	10 10 9.7	11 11 9.0	7 5 9.2	9 9 8.5	6 6 9.0	3 6 8.6	7 9 7.9	11 8 6.9	5 8 7.4	2 4 5.7	7 6 4.9	7 3 6.2	3 3 3.9	7 5 5.2	6 6 3.8	120 118 119.1
≤3	2 2 2.5	3 3 3.3	4 4 4.2	2 2 3.8	5 3 3.8	2 2 3.6	7 7 3.9	3 3 3.7	0 2 3.3	3 11 2.9	2 6 3.1	2 2 2.5	2 3 2.1	1 1 2.7	2 1 1.7	1 1 2.2	2 2 1.6	43 55 51.1

Table 7.7 Comparison of different methods

Method	Model	D^*	χ^2_{42}	P-value
Absolute Residuals Technique	Mean: $\Phi_2(-2, -2)$ SD: Cubic	341.7	126.6	p<0.0001
	Mean: $\Phi_3(0, 0.5, 0.5)$ SD: Cubic	304.2	112.7	p<0.0001
Smoothing the Observed Centiles	$\Phi_3(3, 3)$	415.9	154.0	p<0.0001
MLn	Cubic polynomial[1]	358.4	132.7	p<0.0001
	$\Phi_3(0, 0.5, 0.5)$	300.8	111.4	p<0.0001
Cole	edfs: **4, 6, 4**	217.2	80.4	p=0.0003
GROSTAT	5 3 3 3 3 3 Spline at age 13	187.4	69.4	p=0.005

1: MLn models: Table 7.5

Table 7.8 Comparison of D^* derived from 1000 replicates for two approaches of centile estimation of the set of seven centiles, 3, 10, 25, 50, 75, 90, and 97, for the Normal distribution for sample sizes of 99 up to 999

Sample size	Mean (SD) of D^*		P-value of paired t-test
(n)	Standard Method	Healy's Methd	of mean differences
99	3.21 (2.14)	3.24 (2.55)	0.18
199	3.01 (1.92)	3.05 (2.57)	0.08
499	3.11 (1.99)	3.19 (2.33)	<0.01
799	2.98 (1.92)	3.02 (2.49)	<0.01
999	3.05 (2.02)	3.12 (2.40)	<0.01

Figure 7.1 Different fits of median (M), girls' weight urban Tehran

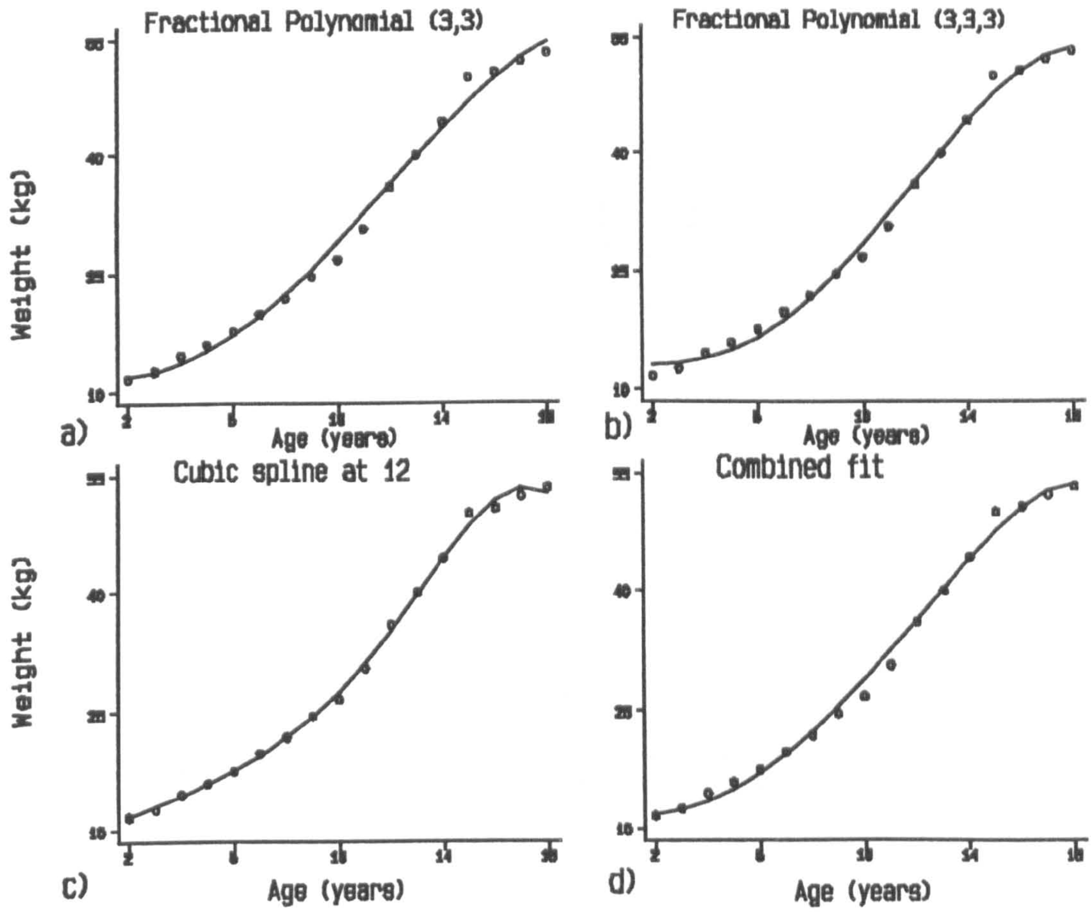
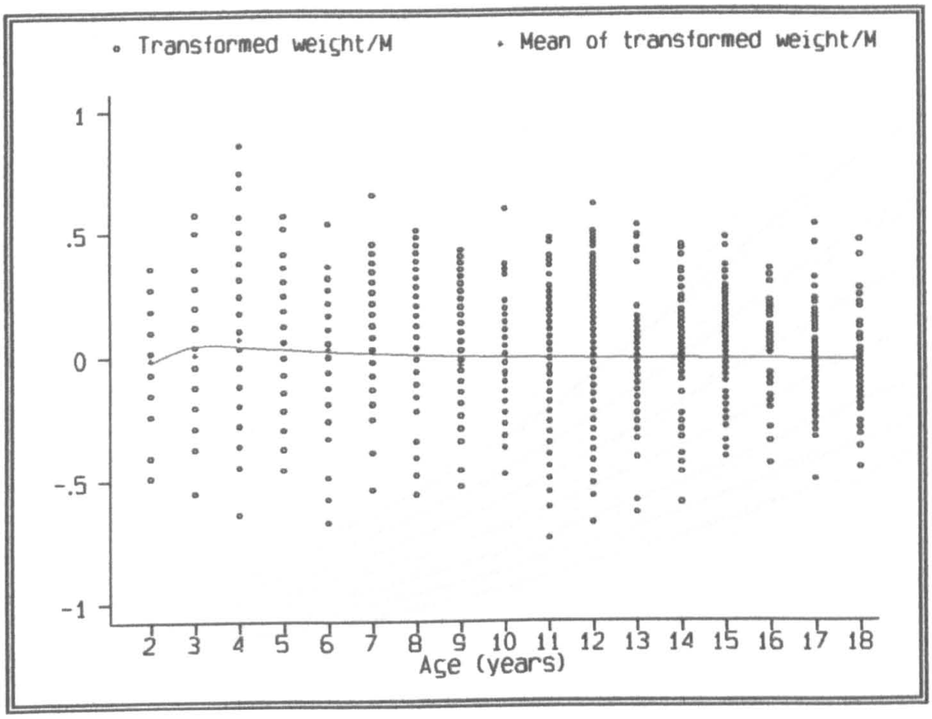
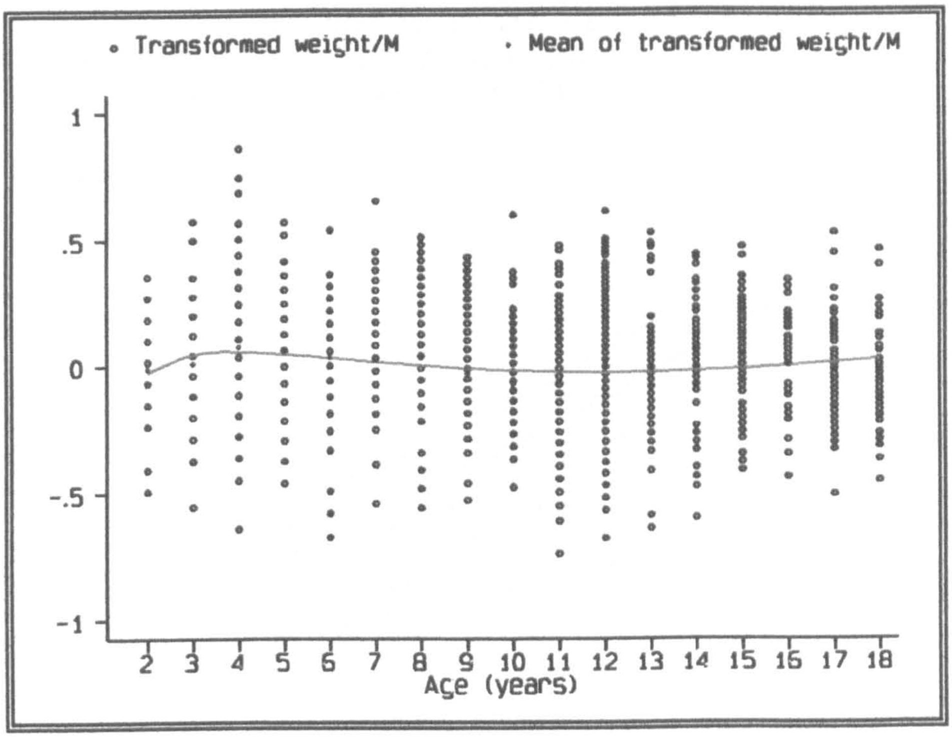


Figure 7.2 Two different fits to the mean of transformed weight/M, girls' weight, urban Tehran

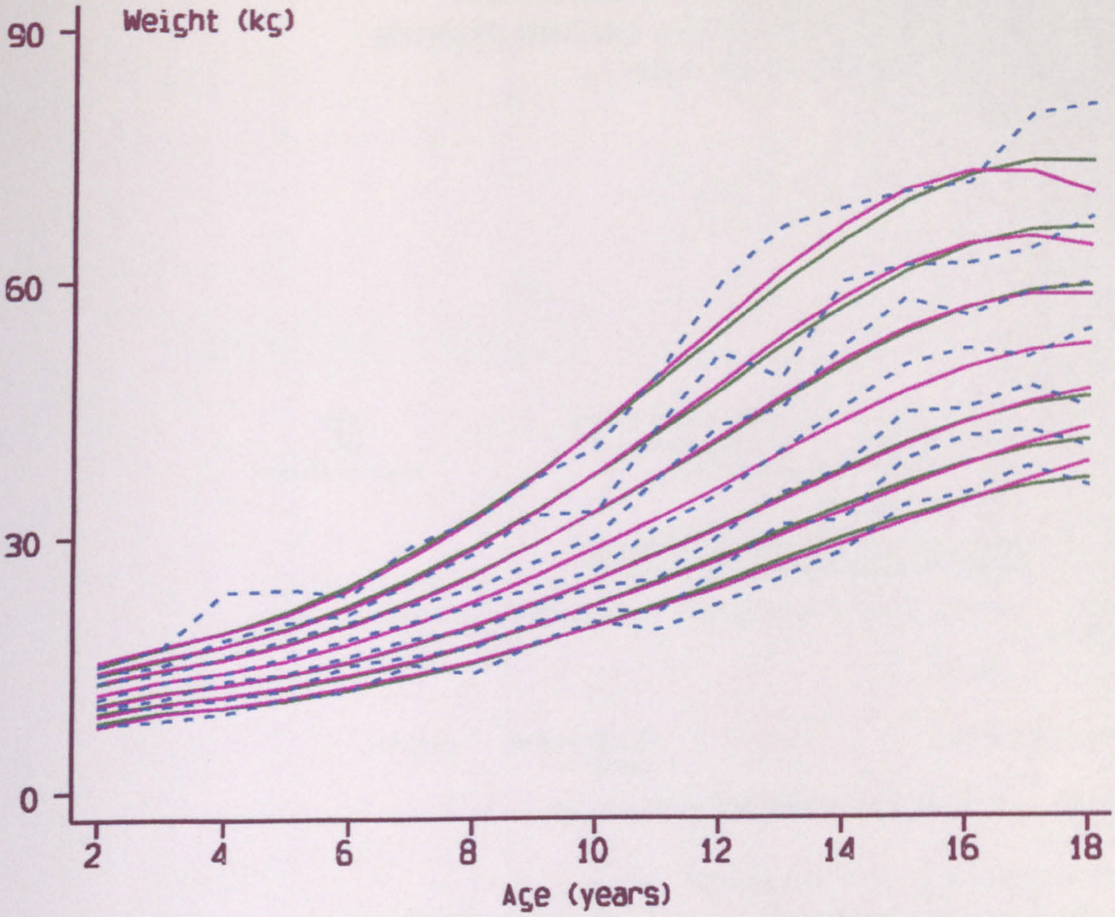


a) The fit $þ_2(-2, -2)$ to the mean of transformed weight/M

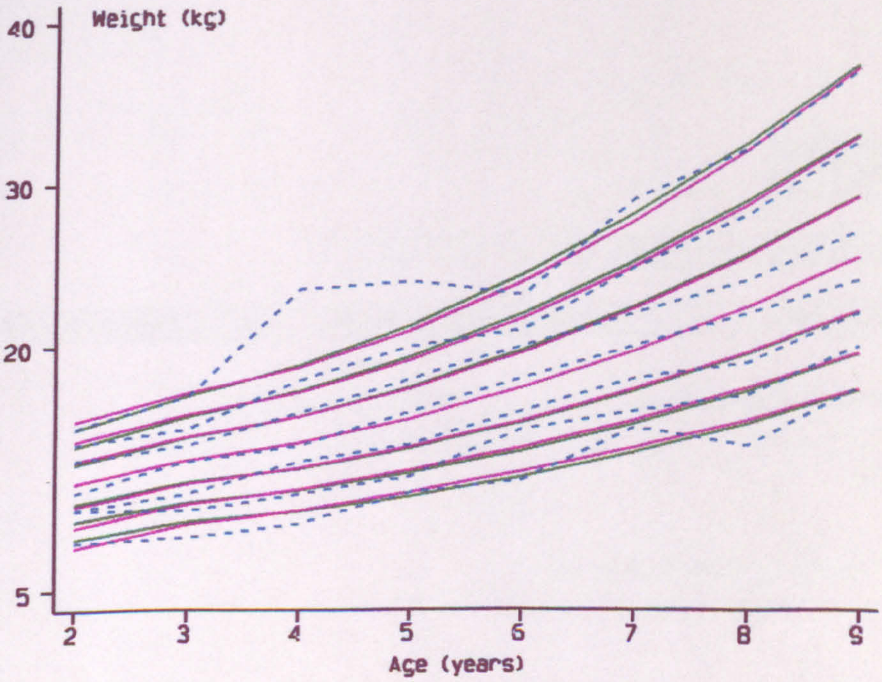


b) The fit $þ_3$ (0, 0.5, 0.5) to the mean of transformed weight/M

Figure 7.3

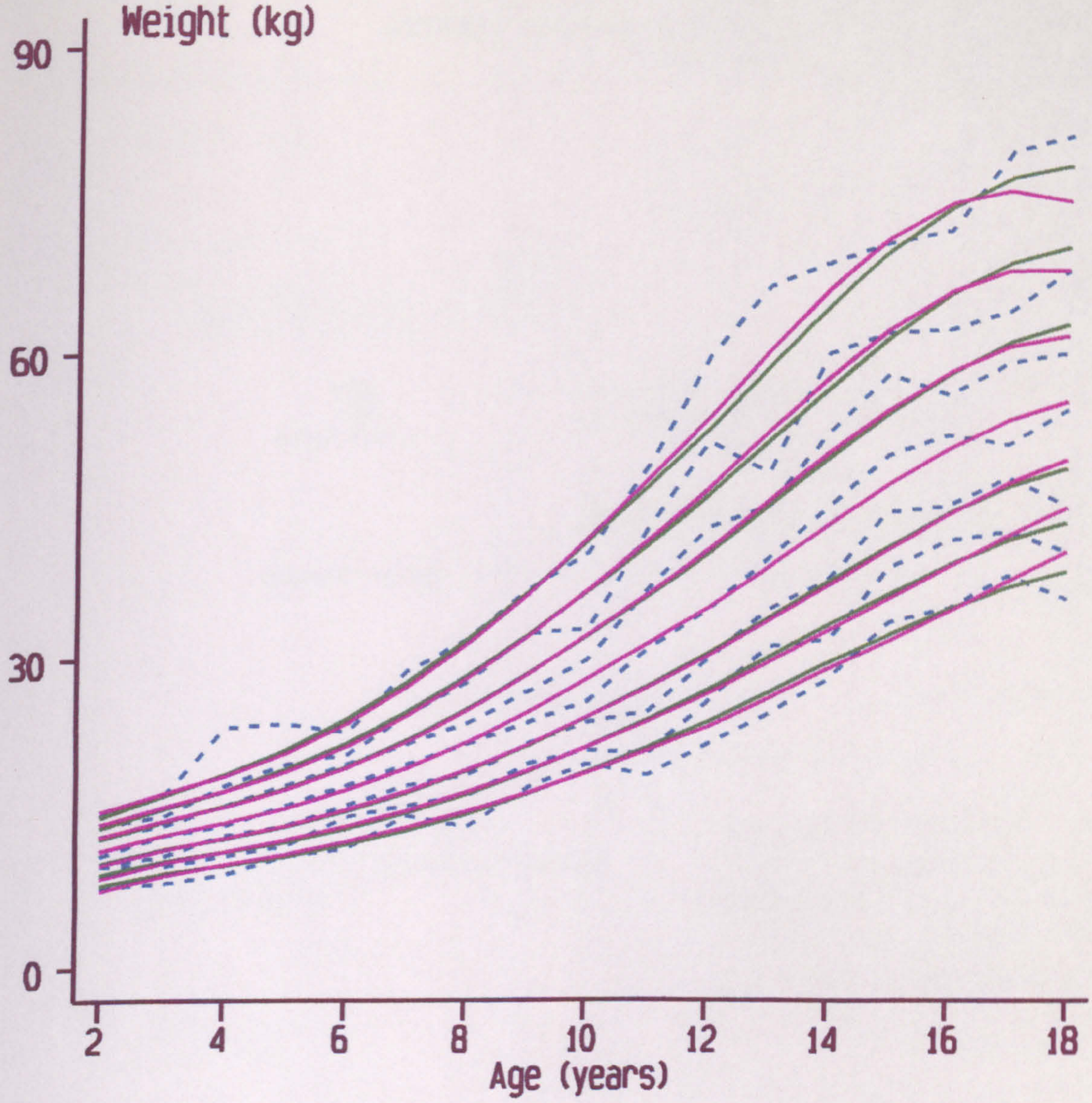


a) The fit Þ$_2$(-2, -2) of the mean of transformed data combined with two models of SD: cubic (pink), quadratic (green); Observed centiles (dotted cyan)



b) As a) with section corresponding to 2-9 year olds enlarged.

Figure 7.4



Reference centiles 3rd, 10th, ..., 97th derived by modelling mean of transformed data using fp $\text{Þ}_3(0, 0.5, 0.5)$ and modelling the absolute residuals by cubic (pink) and quadratic (green) polynomials of age

Figure 7.5 Fractional polynomial ϕ_3 (3,3) fits of the raw centiles, girls' weight, urban Tehran

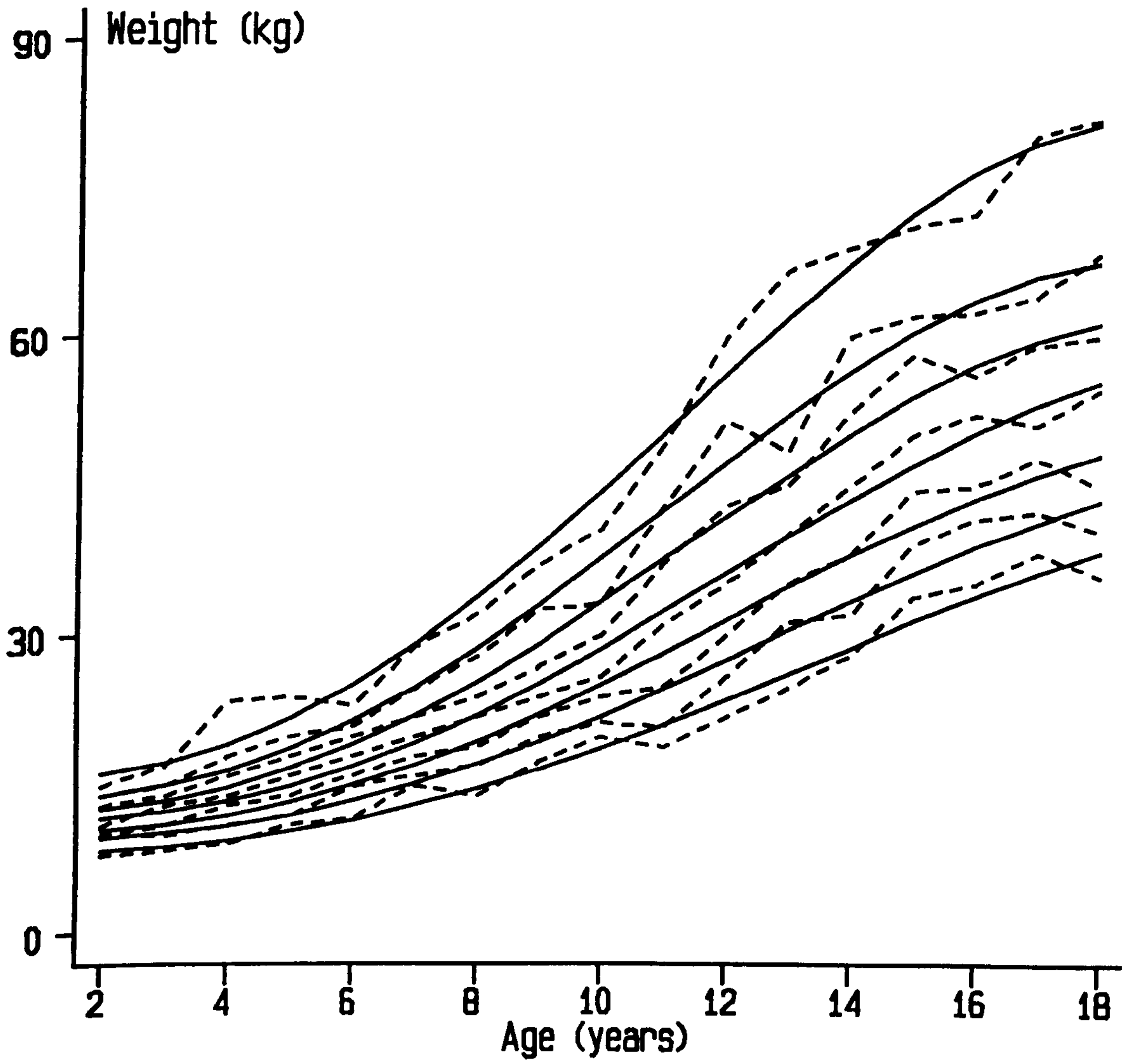
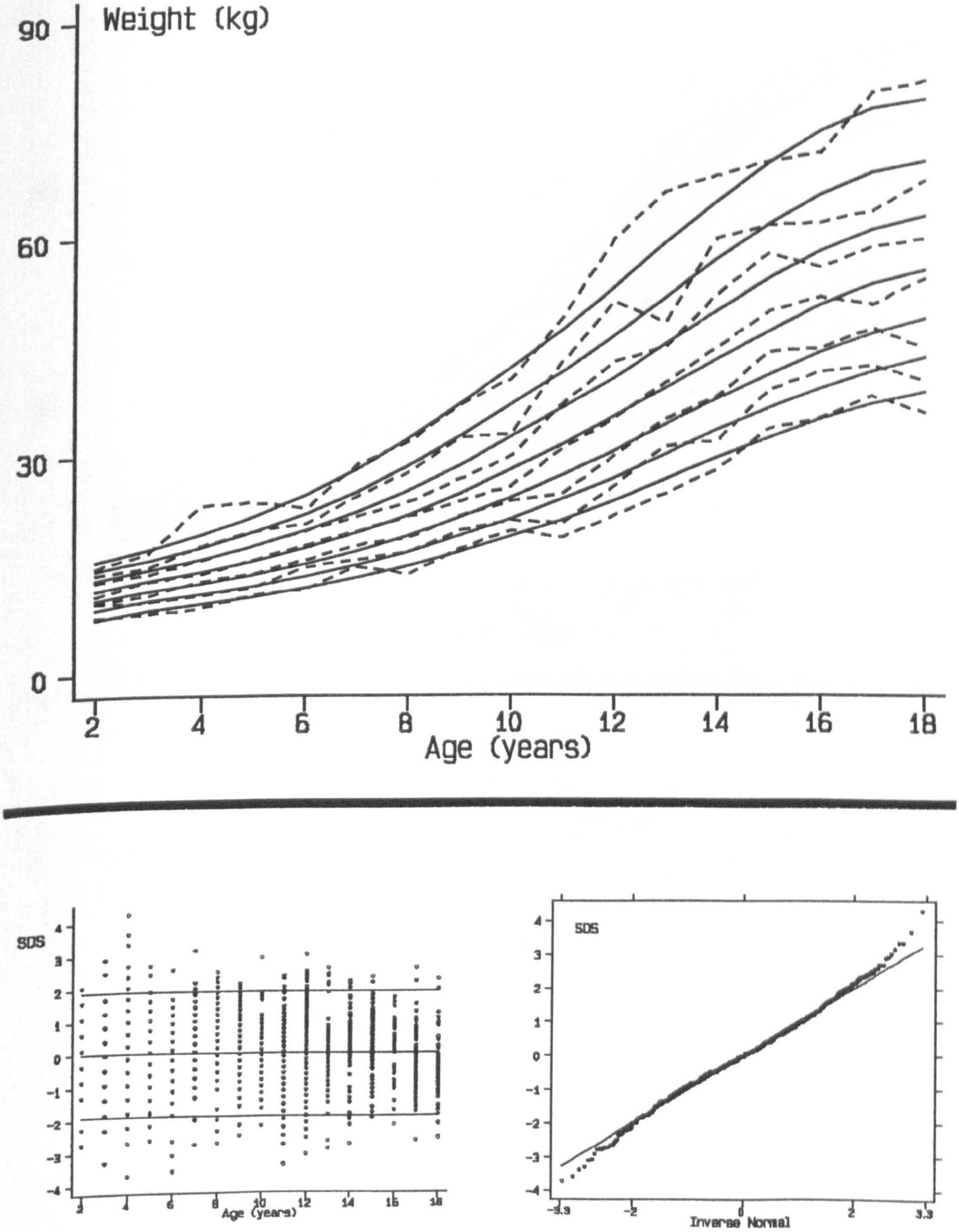


Figure 7.6

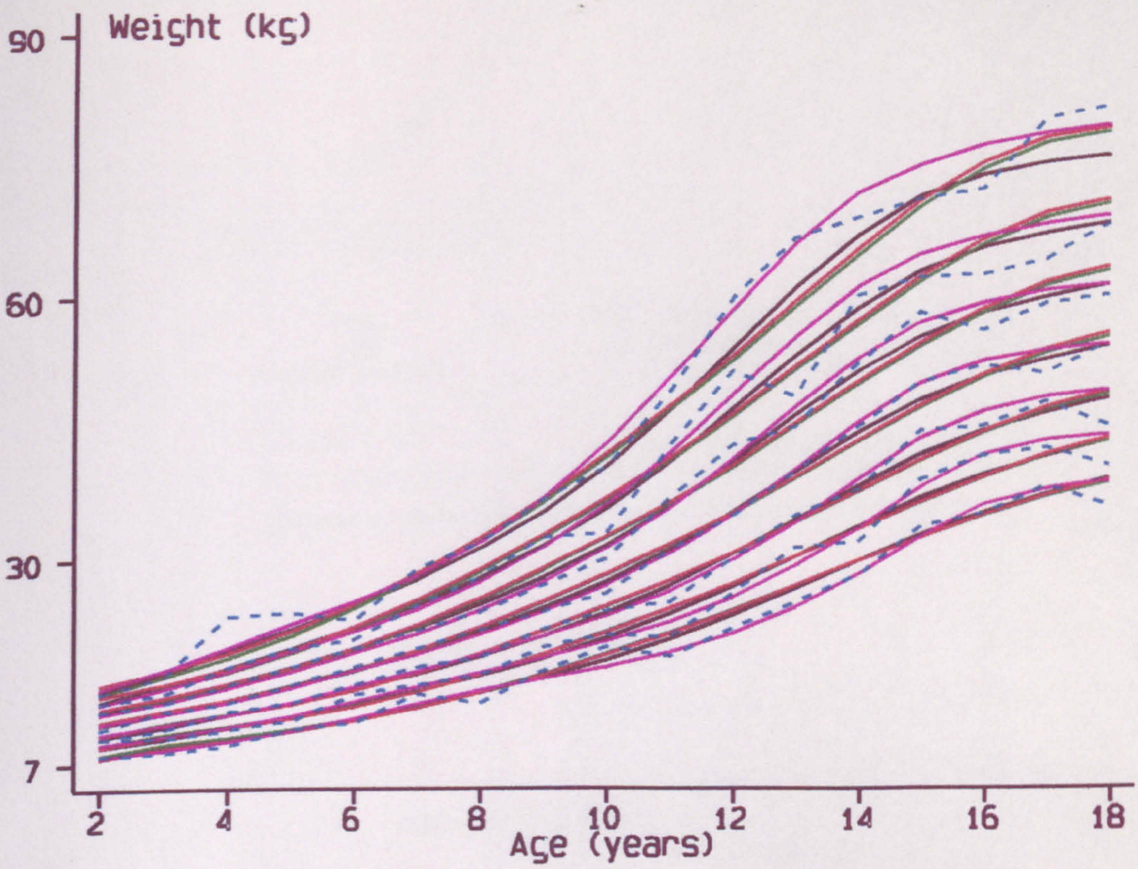
a. Urban Tehran girls' weight with superimposed 3rd, 10th,...,97th centiles derived using **fp** $\phi_3(0, 0.5, 0.5)$ to model transformed data in MLn



b. Residuals (SDS) from model **a**, with expected 3rd and 97th centiles

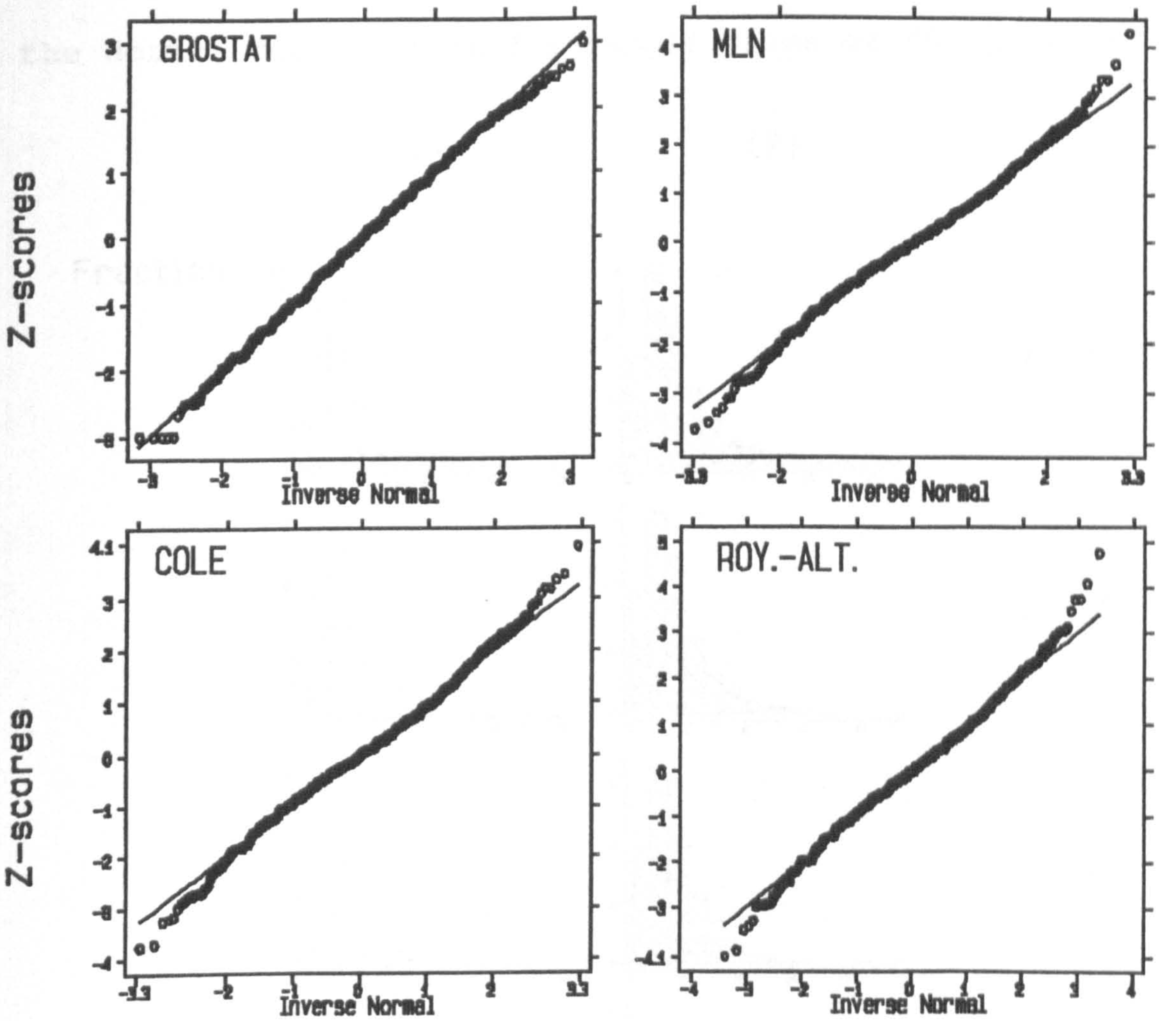
c. Normal plot of SDS from **a**

Figure 7.7 Comparison of different models, girls' weight, urban Tehran



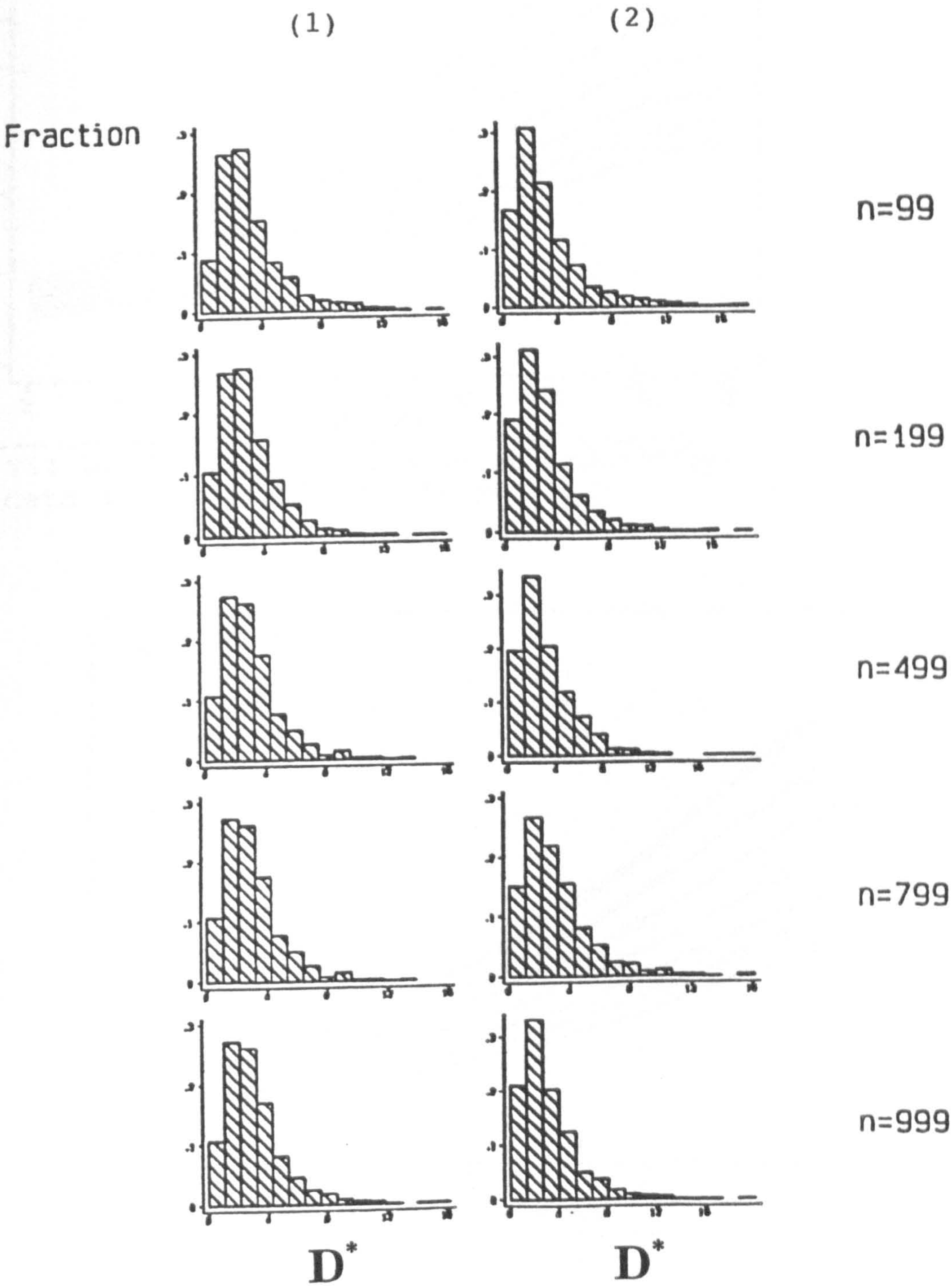
þ_3(0, 0.5, 0.5) in MLn: Brown,
þ_3(0, 0.5, 0.5) and SD quadratic: Green,
GROSTAT: Pink.
Cole's method; edfs 4,6,4: Black
Observed centiles: Dotted Cyan.

Figure 7.8 Comparison of Z-scores of different models



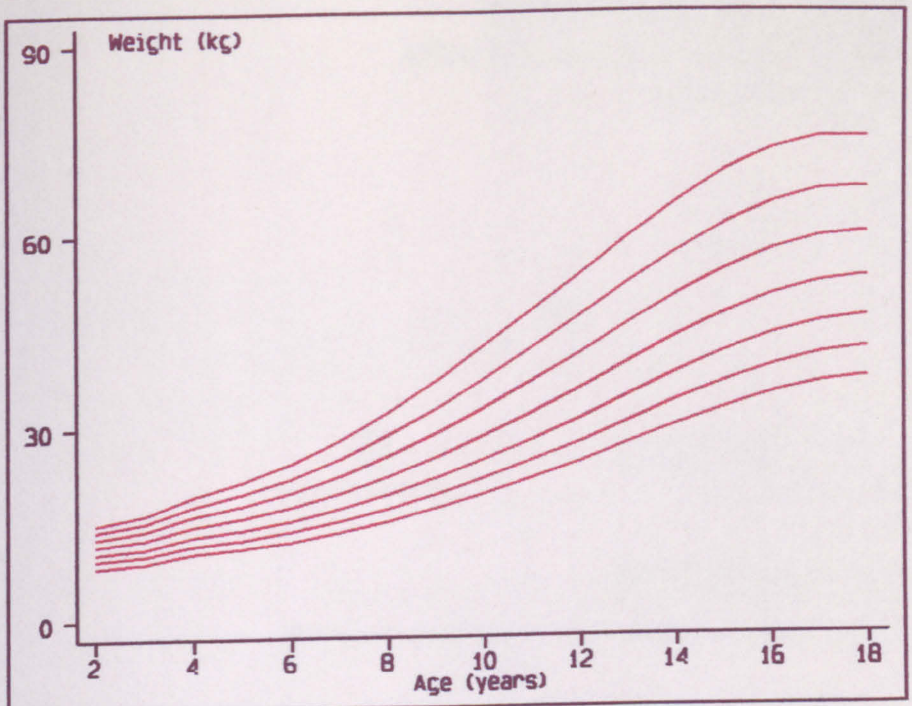
ROY.-ALT. : mean ϕ_3 (0, 0.5, 0.5) and SD cubic (7.4.2.1); Royston and Altman (1994)

Figure 7.9 Distribution of D^* for two approaches to centile estimation of the set of centiles, 3rd, 10th,..., 97th, for the Normal distribution for sample sizes of 99 up to 999

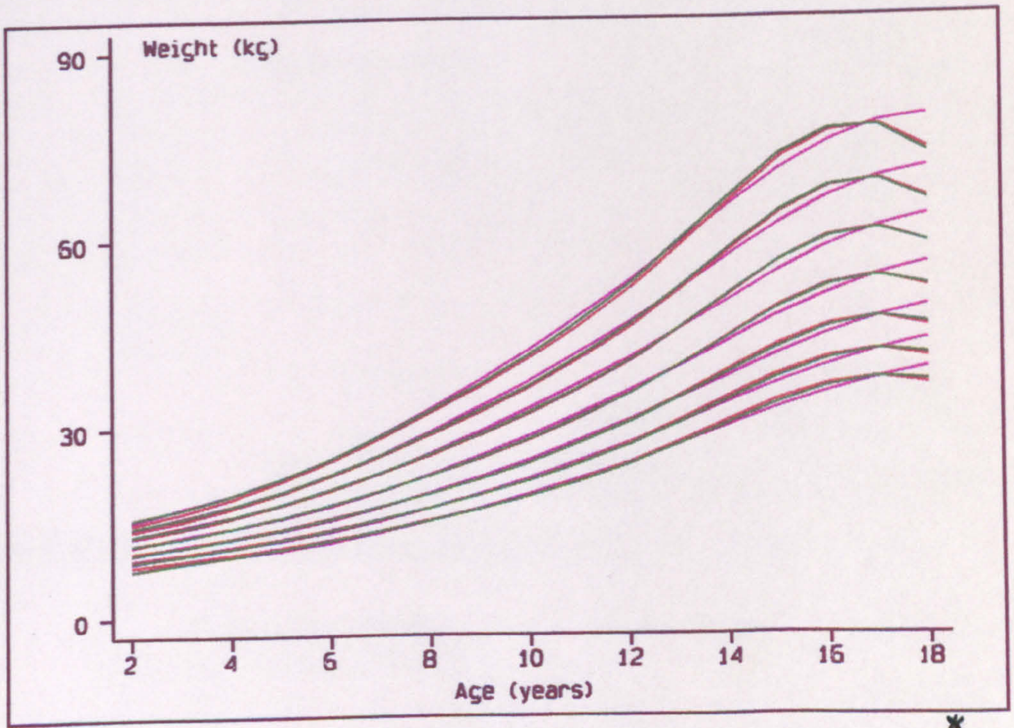


1: Standard method ($\bar{x}+z_{\alpha}s$)   2: Healy's method

Figure 7.10



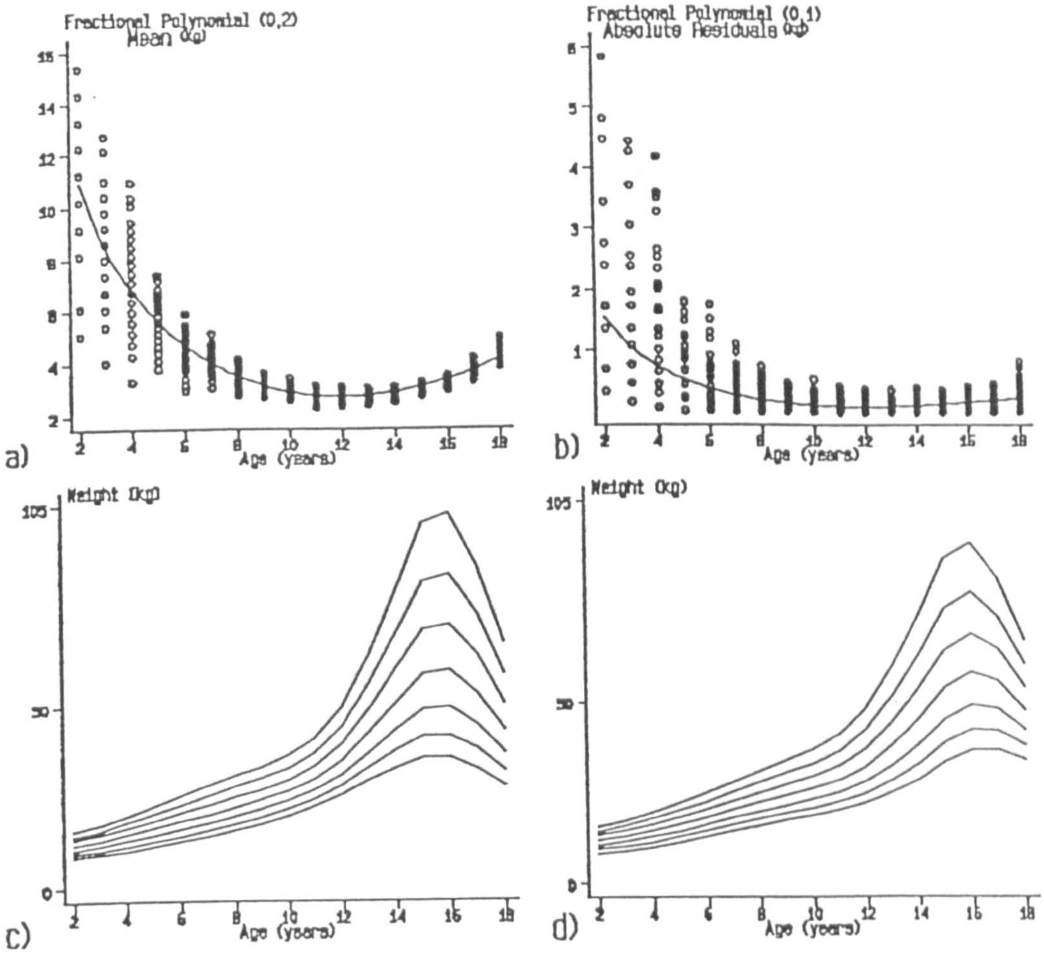
a) Fit $\text{þ}_3(-5, -5, -5)$ of mean of transformed data and quadratic fit to SD



b) Comparison of þ_4 (MLn: green), þ_4 (ROY.-ALT.* :brown), and þ_3 in MLn (pink)

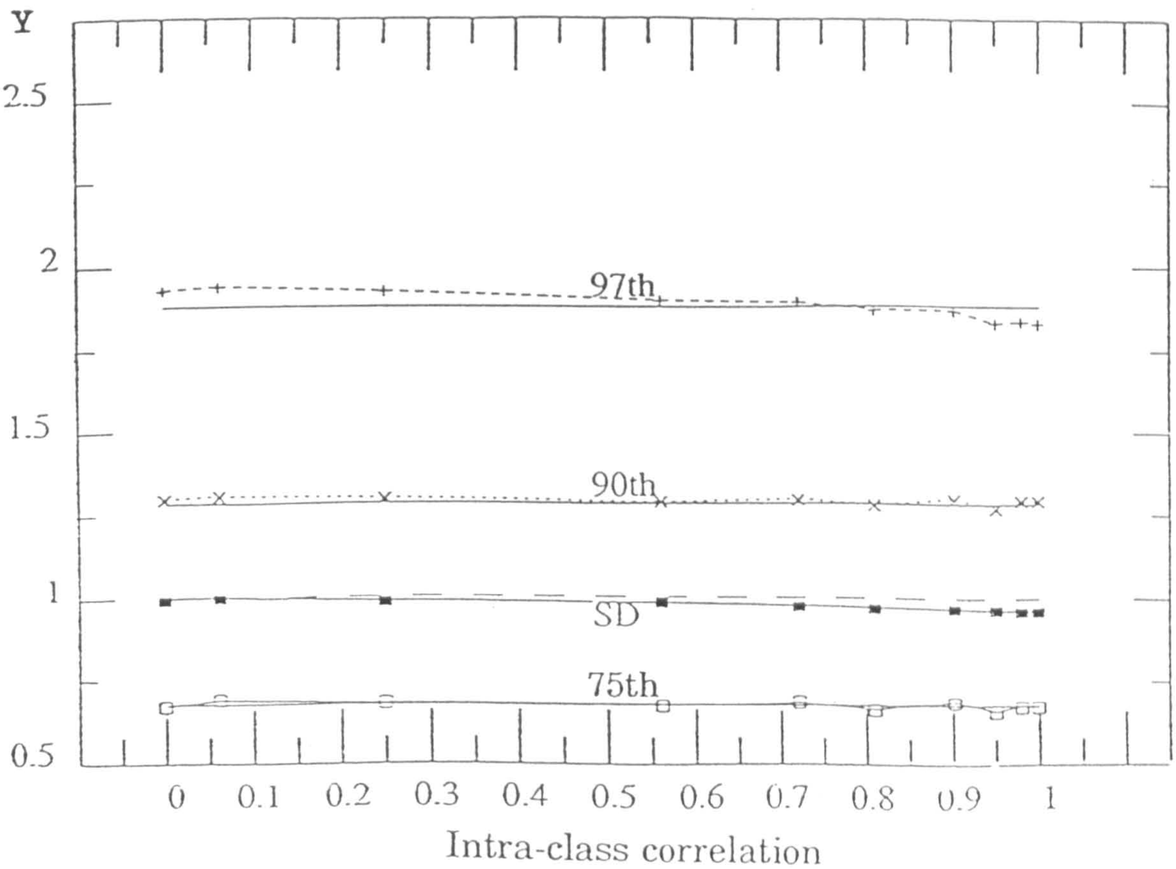
b) $\text{þ}_3(0, 0.5, 0.5)$, and $\text{þ}_4(2, 3, 3, 3)$
* Absolute Residuals modelled as quadratic of age

Figure 7.11 Fp ϕ_2, and ϕ_3 models of mean and SD of weight$^{(\lambda)}$; girls' weight, urban Tehran



c) $\phi_2(0,2)$ of mean and $\phi_2(0,1)$ to SD
d) $\phi_3(-1, -0.5, 3)$ of mean and $\phi_3(-1, -0.5, -0.5)$ to SD

Figure 7.12 Simulated estimates of 75th, 90th, 97th centiles and the SD of y ignoring structure, by intra-class correlation; y=au+bv, where U and V are N(0, 1) and $a^2+b^2=1$



CHAPTER EIGHT

GROWTH CHARTS CONSTRUCTED FOR URBAN TEHRAN USING GROSTAT II

8.1 Introduction

The interpretation of any measurement of an individual depends on the existence of a body of relevant information to which that measurement may be referred for comparison (Harris and Boyd, 1995). Centiles reference charts are commonly used in the routine monitoring of individuals, where interest lies in the detection of extreme values, possibly indicating abnormality.

Reference curves are useful for monitoring growth and provide criteria for selection of children for further clinical investigation when attained growth deviates from that expected in a healthy population. In population studies growth reference curves are used to define the extent and severity of abnormal anthropometric status, to monitor trends in growth and nutritional status and to evaluate the impact of nutritional interventions. Also, growth reference curves are useful to describe nutritional outcomes in epidemiological studies. As was discussed in section 2.4, the standards should be derived from the population to which the children belong, and it is generally inappropriate to use imported standards

(Goldstein and Tanner, 1981). Therefore, in relation to the discussion in Chapter 6, the data from urban Tehran has been selected for investigation and the construction of a set of national growth charts for Iran in this chapter.

8.2.1 Spline

In Chapter 4 it was discussed that as an extension to the HRY method, when smoothing curves over a wide age range, spline techniques (Pan et al., 1990) can be used to obtain a more appropriate fit. It was assumed that all curves can be fitted by a polynomial of degree p. If t denotes age and Y_i the smoothed value of ith centile, we have

$$Y_{i,t} = a_{0,i}+a_{1,i}t+\ldots+a_{p,i}t^{p}+a_{p+1,i}(t-c_1)_+^p+\ldots+a_{p+m-1,i}(t-c_{m-1})_+^p$$

where a_{ij} are function of z_i, the ith Normal centile, and where

$$(t-c_l)_+^p = (t-c_l)^p \quad \text{when } t>c_l$$
$$= 0 \quad \text{when } t\leq c_l \quad l=1,2,\ldots,m-1$$

c_l is the lth join point in the age span and $c_1<c_2\ldots<c_{m-1}$. Of course, when m=1, the method is the same as that presented by Healy et al. (1988).

It is clear that all the centiles and their first p-1 derivatives join smoothly at the points c_l. The value of m and c_l can be chosen after inspection of the data and using existing knowledge, although, there exist a few strategies for the optimal selection of number and position of the join points (Saber and Wild, 1989). In the following presentation, for girls the age 13 years old and for boys 15 years, were chosen as the join points for splining curves.

Throughout this chapter, the overall polynomial, in above formula fitted to any set of raw centiles is presented in the form of:

$$p\ q_0\ q_1\ q_2 \ldots q_p$$

where p is the order of the overall polynomial in t (age)

q_0 is the order of polynomial in z in the constant term

q_1 is the order of polynomial in z in the coefficient of t

q_2 is the order of polynomial in z in the coefficient of t^2

....

q_p is the order of polynomial in z in the coefficient of t^p.

Hence, a *3 3 2 1 3* polynomial indicates that the overall polynomial level, *p*, is a cubic in t (age). The constant term in t is a cubic polynomial in z (q_0=3); the coefficient of t is a quadratic in z (q_1=2); the coefficient of t^2 is linear in z (q_2=1); and the coefficient of t^3 is a cubic polynomial in z (q_3=3). It is should be pointed out that if in the above case the smoothed curves are splined at say c_1, the coefficient of the '+' term is similar to the q_p (i.e. a cubic in z). So the overall polynomial in the above example would be

$$Y_i=a_0+b_{01}z+b_{02}z^2+b_{03}z^3+(a_1+b_{11}z+b_{12}z^2)t+(a_2+b_{21}z)t^2$$
$$+(a_3+b_{31}z+b_{32}z^2+b_{33}z^3)t^3+(a_4+b_{41}z+b_{42}z^2+b_{43}z^3)(t-c_1)_+^3$$

8.2.2 Data

Consideration of variation of different centiles of weight and height of children in urban areas of Iran, and comparison with the urban Tehran's centiles showed that urban Tehran's data can be used as reasonably representative of the growth pattern in urban Iran, section 6.3. Tehran is the capital of Iran, with an area of 28,225 sq.kms. and is one of the most populous provinces in Iran. Tehran is located on the southern slope of the Alborz Mountains and extends from Firouzkuh in the east, to Hashtgerd in the west and Qom township in the south. According to the 1992 census, Tehran province had a population of 10,337,826, had the highest increase in population and was the most urbanised province in Iran (GSO, 1993).

It might be expected, especially in developing countries, that the capital city is a mixture and representative of people from all part of the country. This is due to many factors such as better facilities and possibility for better jobs and business. The movement to the capital was accelerated during the war (1980-8) when many people migrated to central Iran from the west and southern borders. The analytical investigation of the data also confirmed that Tehran could reasonably be regarded as representative of Iran. In early 1996 Qom, as a new province, was separated from the large Tehran province but this would not alter the findings. The interesting practical feature of the representative nature of urban Tehran is that it can be very important for future research in Iran. The province is where the field work of training of medical students is conducted because the three major universities of medical sciences of the country are located in Tehran. Also, the Primary Health Care system is well established and growth monitoring is practised regularly by Health Centres and Health Houses there. So the collection of representative data for any future studies will be comparatively easy.

In the following sections the results of modelling weight and height measurements of children in urban Tehran are presented. The measurements of weight and height of 3,301 children aged 2-18 years old, 1,702 (51.6%) girls and 1,599 (48.4%) boys were used to construct the age-related centiles. Table 8.1 provides the general information about the distribution of these measurements in urban Tehran for both sexes.

8.3 Results of modelling using GROSTAT

8.3.1 Modelling weight of 2-18 years old children and adolescents

8.3.1.1 Girls' weight

In the following sections, **GROSTAT II** (Rasbash et al., 1993) was used in the computation of smoothed age-related centiles.

In the present data age was recorded in complete years, therefore, the first step was to compute the appropriate raw centiles at each year of age using the SPSS (1992) software package. Then, these raw centiles were fed into GROSTAT for computation of smooth centiles. These centiles span the full age range 2-18 and there was no loss of data at each end of the age range such as occurs if the CBOX procedure had been used to obtain the raw centiles (8.4.3). Although, in the HRY method there is no assumption about the distribution of the measurements across ages (4.3), we fitted curves after using a logarithmic transformation of the weight because the method works best when the data is approximately Normally distributed (Healy, 1991; Ayatollahi et al., 1993b).

It is known in the HRY method that the fit of the model to the data critically depends on the degree of polynomials in age (p) and z (q). Although there is no formal theory for determining these values, and they can only be found by experimentation and validation, in this work we have suggested a way of choosing the order of polynomial in age and z (8.5.1). It should be mentioned that the only models contained in this thesis are the ones considered to be optimal. However it took a long time generating and examining many models to find the best. There would be little benefit from including details of these searches in the thesis.

The actual centiles that are displayed on all graphs in this chapter are 3rd, 10th, 25th, 50th, 75th, 90th, 97th. Apart from the need for these centiles in our study the number of them is reasonable for obtaining the overall accuracy of the models. The smoothing polynomial found to fit the raw centiles best was a quintic of the form *5 3 3 2 3 1 2* which was splined at the age of 13 years. In fact, smoothed age-related centiles for girls' weights (Table 8.2) were obtained by fitting the following model:

$$Y_i = Exp[a_0 + a_1 Age + a_2 Age^2 + a_3 Age^3 + a_4 Age^4 + a_5 Age^5 + a_6 (Age-13)_+^5]$$

where,

$a_0 = 2.2690 - 2.9990\times10^{-1} z_i + 1.5668\times10^{-2} z_i^2 + 1.6589\times10^{-2} z_i^3$

$a_1 = 3.8906\times10^{-2} + 3.6575\times10^{-1} z_i - 1.9894\times10^{-2} z_i^2 - 1.0686\times10^{-3} z_i^3$

$a_2 = 3.2108\times10^{-2} - 1.0218\times10^{-1} z_i + 4.6292\times10^{-3} z_i^2$

$a_3 = -5.7549\times10^{-3} + 1.2710\times10^{-2} z_i - 3.0729\times10^{-4} z_i^2 + 2.366310^{-6} z_i^3$

$a_4 = 4.4274\times10^{-4} - 7.0870\times10^{-4} z_i$

$a_5 = -1.1965\times10^{-5} + 1.4457\times10^{-5} z_i + 3.6263\times10^{-7} z_i^2$

$a_6 = 1.2399\times10^{-4} - 1.2488\times10^{-5} z_i - 1.3352\times10^{-5} z_i^2$.

In addition, because the smoothed centiles over age range 2 to 5 overfitted the raw centiles it was decided to replace this tail for 5 and below with the tail of the model *2 3 3 2* which was similarly splined at 13. The corresponding model was

$$Y_i = Exp[\ a_0 + a_1 Age + a_2 Age^2 + a_3 (Age-13)_+^{\ 2}]$$

$a_0 = 2.2679 + 0.11681 z_i - 0.27544\times10^{-1} z_i^2 + 0.12543\times10^{-1} z_i^3$
$a_1 = 0.89095\times10^{-1} + 0.10002\times10^{-1} z_i + 0.75248\times10^{-2} z_i^2 - 0.25696\times10^{-3} z_i^3$
$a_2 = 0.15382\times10^{-2} - 0.30929\times10^{-3} z_i + 0.41444\times10^{-3} z_i^2$
$a_3 = -0.16330\times10^{-1} - 0.13177\times10^{-2} z_i + 0.14188\times10^{-2} z_i^2$.

To examine the fit of the model to the data the number of original observations falling between the smoothed centiles and the expected numbers were compared. Table 8.3 presents the distribution of the observed and the expected points falling between the smoothed centiles. Overall, the observed and the expected percentages were very close. With 2.5%, 7.0%, 14.5%, 25.7%, 26.6%, 14.0%, 7.2%, and 2.5% of the points observed to fall in the intervals formed by the 3rd, 10th, 25th, 50th, 75th, 90th and 97th centiles respectively, indicating a good fit to the data. Also, the observed and expected percentages of points across all age groups were fairly close, with the largest difference occurring between the 50th and 75th centile in age group 6 to 9 years where the observed percentage point was greater than expected, 29.5% compared with 25%. And between the 25th and 50th in age group 10 to 13, 30.2% compared with

25%. It should be noted that some of these differences are due to weight being recorded to the nearest kilogram, so that some of the observations which are in the borders might really belong to the other cells of the table.

A Pearson χ^2 statistic with 7 degrees of freedom may be used to assess whether the total frequencies in Table 8.3 of observations found within the 8 groups differ significantly from expected. For this comparison, the Pearson χ^2 statistic (7 df) was found to be 6.3 (p=0.51). Also, as an extension of the fit assessment in Table 8.3, the data are divided into four age groups: [2,5], [6,9], [10,13], [14,18] which were thought to be representative of changes in the pattern of growth of girls. Then the Pearson χ^2 test statistic (28 df) was computed to compare the observed and expected frequencies in the resulting contingency table and found to be 31.62 (p=0.29) showing that the fit is good. It should be pointed out that in the present case, because the number of groups (m=4) is less than the parameters of the M curve (pm=7), there is no need to use Royston's (1995) proposed adjustment (7.6). Moreover, the p-value is enough large to assure us of a good fit.

Figure 8.1a shows the raw and the smoothed centiles generated by the above model. The fit of the smoothed curves to the raw centile appeared to be excellent over most of the age range. The smoothed centiles passed through the middle of the raw centiles across the age range except for the extreme upper centiles at ages older than fifteen. This is as might be expected, because in these ages fewer observations were involved in the calculation of the extreme centiles, so to some extent the raw 90th and 10th but mainly the 97th and 3rd centiles are more variable than the 25th, 50th and 75th. Also due to the fact that in urban Tehran there are some overweight children in some of the well-off families the age-specific distributions in these ages may be more skewed. Therefore in smoothing curves for these ages not only the extrapolation of the underlying fit after seventeen but also general knowledge of the growth of children has been used to extend the curves to age 18 years by hand.

Overall, one would say from the Figure 8.1a that the raw and smoothed centiles are very close over most of the age range. However, looking at the upper extreme observed and smoothed centiles may give rise to this question: has the data been reasonably explained by the fit at the right end of age band? There are two points to note here. First, comparison of the percentages of observations falling between the smoothed centiles and the expected in the age band 14 to 18 showed that the distribution of the original observations is reasonably close to the expected. Second, one does not necessarily wish the smoothed centiles to match the raw ones, as one of the aims of the smoothing process is to iron out the irregularity in the raw centiles in order to produce a reasonable chart.

Moreover, Figure 8.1b illustrates the smoothed centiles plotted with the original data points. Some of the observed difference such as the one for the 97th at four years and 3rd at eight years can be explained by the large spread of raw points at these ages. There is evidence of increase of the spread in the data over time from ten years, i.e. the period of puberty. The upper centiles rise first and lower centiles come up later. The distance between the centiles steadily increase as the girls reach their adulthood. Then from sixteen curves start to flatten off. There is evidence of skewness as the lower centiles are closer to the 50th than the upper. The distribution of points below and above the 3rd and 97th centiles appear to be very even over the complete age range, with no ages where there is a large cluster of outlying points.

The next device to aid the assessment of the fit involves converting all measurements to Z-scores. If the model is correct the empirical quantile of the Z-scores should be approximately equal to -1.88,-1.28, ...,1.88 respectively, i.e. to the 3rd, 10th, ..., 90th and 97th centile of N(0,1). After converting all measurements to Z-scores using CVAL procedure in **GROSTAT II**, the quantiles were computed at each age. Figure 8.2a shows the quantile of the Z-scores of the model with the quantile of the Normal distribution. It is clear that there is no trend apparent in the Figure 8.2a. The Z-scores of the smoothed centiles oscillate around their expected values closely over the age range which is an indication of a good fit.

Finally, a Normal quantile-quantile (Q-Q) plot of the residuals (SDS), which is also an appropriate technique to assess the model, is presented (Figure 8.2b). The Normal plot of the Z-scores (SDS) has a reasonably close distribution to the Normal ($p=0.14$; W test). In conversion of the measurements to Z-scores in GROSTAT since the model assumes that the centile curves can be fitted as a polynomial function of Normal deviates, the equivalent standard deviation score for a value of y at age t is obtained by solving a polynomial equation. Because this equation is usually non-linear in z, there will in general be more than one solution for z, given y and t. To avoid this difficulty, GROSTAT only looks for a solution for z which is contained in the interval [-3, 3]; and if the solution is outside this range, z is reported as ±3.

It can be seen in Figure 8.2b that the Z-score of one point is computed as 3, and 5 points -3. For the remainder of the data the Normal fit is excellent. For these 6 points further analysis (Goldfrad, 1994) suggested that these are extreme values and probably represent outliers, which are revealed by accurate fit of the Normal curves to the data but were not revealed by the preliminary multivariate analysis of outliers described in chapter three. For instance a 27 kg child of 4 years old or the other children 8 kg or 9 kg and 6 years old. These few outlying points can be seen in Figure 8.1b in comparison with the rest of the data. The inclusion of these points in the analysis of 1702 records has not affected the resulting curves and they have been ignored.

Royston (1995) discusses a more formal test (Q-test) for investigation of significant age-related variation in location, scale and shape of the distributions of Z-scores. He suggests the age range is divided into k (e.g. k=10) groups with means $\bar{z}_1,\ldots,\bar{z}_k$, variances $s_1^2,\ldots,s_k^2$, group sizes $n_1,\ldots,n_k$, and the degrees of freedom $v_1,\ldots,v_k$ ($v_i=n_i-1$). Then the location test is, H_0: all mean of Z-scores equal to zero. A test statistic for testing H_0 is

$$Q_M = \sum_{i=1}^{k} \bar{z}_i^2 n_i$$

Q_M has distribution approximately χ^2 on k-1 df.

The scale test is, H_0: all variances are equal to one. A test statistic is based on the Wilson-Hilferty (1931) transformation:

$$\left(\frac{s_i^2}{\sigma_i^2}\right)^{\frac{1}{3}} \sim N\left(1-\frac{2}{9v_i}, \frac{2}{9v_i}\right)$$

and so Q_S, $Q_S=\sum_{i=1}^{k}\{s_i^{\frac{2}{3}}-(1-\frac{2}{9v_i})\}/\frac{2}{9v_i}$, also has an approximately

χ^2 distribution on k-1 df.

The shape test is based on Shapiro-Wilks W test. H_0: Z-scores in all groups are Normal. A test statistic based on Fisher's method for combining independent P-values is:

$$Q_W=\sum_{i=1}^{k} -2\ln P_i$$

where P_i is p-value for the W statistic in the ith group. Q_W has approximately a χ^2 distribution on 2k df. Therefore, it was decided to consider each year of age in our data as a group. In Table 8.4a the mean, standard deviation and p-value of the W test in all seventeen group of age are presented. Consideration of means, standard deviation, and p-values in each year of age in Table 8.4a shows that the distribution of Z-scores may be accepted as Normal with a mean and standard deviation close to 0 and 1 except possibly for the ages older than 16 which, as explained, is due to unreliability of the data in these ages. The results of the overall Q-tests for the smoothed models are presented in Table 8.4b. None of the location, scale and shape tests of Z-scores were found significant with the p-values p_M=0.56, p_S=0.16, and p_W=0.75 respectively.

It is worthwhile mentioning that, in order to compute the exact value of Z-scores in those ages where the curves are treated by hand; first a model accurately describing the smoothed curves for those ages was found, and then the Z-scores were computed from the corresponding model (Appendix D).

8.3.1.2 Boys' weight

A quartic model of the form 4 3 3 3 0 2 splined at fifteen was fitted to the raw centiles of boy's weight, and is presented in Figure 8.3a. The model parameters are presented in Appendix D. The smoothed centiles by age are shown in Table 8.5. The overall fit of the model to the data was found to be reasonable but there were minor problems. One of the expected problems was with the upper tail of the curves especially at eighteen. However, the fit to the observed raw centiles was very good but the data in this age is probably selected and unreliable. And this tail could not be accepted as a part of a reasonable chart. Next, over the ages 14 to 16 however the trend of the 25th, 50th, 75th was following the raw centiles closely but the smoothed curves were lower. Therefore, the outcome of another close fit (4 3 3 3 2 3) which had overfitting problems at lower ages, and the result of the present fit led us to treat these centiles manually. As the Figure 8.3a shows, the smoothed curves appear to match the raw centiles very closely overall although there are ages where raw centiles and the fitted curves diverge.

A plot of the smoothed centiles and the original measurements is displayed in Figure 8.3b. Since the distribution of observations at ages such as 10, and 15 is skewed, the smoothed 3rd centile at these ages appear too low. Otherwise, the raw data points appear to be quite evenly distributed about the smoothed centiles.

Table 8.6 summarises the percentages (numbers) of observations falling between the centiles over the complete age range and the subgroups. For boys, age was divided into four group as: [2, 6], [7, 10], [11, 15], [16, 18]. About 50% of points fall on or below the 50th centiles. Overall, the cumulative percentage of observations and the expected percentages are close; with 3.0%, 10.0%, 24.0%, 50.6%, 76.9%, 91.4%, and 97.5% below the smoothed 3rd, 10th, 25th, 50th, 75th, 90th, and 97th centiles respectively. In subgroups, the largest differences occur in the ages 7 to 10 between 10th and 25th, and also the 50th and 75th centiles. It seems that 25th centile is a bit low and in contrast the 75th is a bit high. The differences between observed and expected in the last subgroup (16 to 18 years old) is mostly due to the problem of few observations at 17 and mostly at 18 years where the observations are condensed around the mean. So between 25th and 50th, 50th and 75th as well as 75th and 90th we have more weight measurements than the expected; on the contrary less observations than expected fell between the 90th and 97th and above the 97th centile.

The overall Pearson χ^2 statistic was found to be non significant (χ^2_7=7.2, p=0.34). Moreover, after considering the proposed age subgroups, the difference between observed and expected numbers of observations between the centiles was not significant, χ^2_{28}=31.8, p=0.28.

In Figure 8.4a, the raw centiles of the Z-scores are plotted together with the expected ones (expected Z-scores for the 3rd,..., 97th of N(0, 1)). The absence of any systematic trend and closeness of the two sets of Z-scores is an indication of a good overall fit. In addition, more appropriate assessment of the fit can be carried out by investigation of variation of Z-scores. Q-tests were carried out to test the significance of variation in Z-scores in terms of location, scale, and shape. In this analysis, like girl's weight seventeen age groups were considered as the groups of age. Table 8.7b presents the results of Q-tests. The tests of location, scale, and the shape were not significant with p_M=0.60, p_S=0.52, and p_W=0.68 respectively. The Normal plots of Z-scores in subgroups of age and the corresponding p-value of W tests are displayed in Figure 8.4b.

8.3.2 Modelling height of 2-18 years old children and adolescents

8.3.2.1 Girls' height

The smoothing polynomial found to fit the raw centiles of girls' height overall was the cubic of the form 3 3 1 2 3 which was splined at the age of 13 years. The smoothed percentiles are presented in Table 8.8, and the model parameters are tabulated in Appendix D. Figure 8.5a illustrates the raw and smoothed centiles by the above model. On the whole, the fit of the smoothed centiles to the raw ones appears to be very good except for the last two years where the raw centiles themselves are not very reliable. It should be pointed out that over ages 13 to 15, the quartic model of the form 4 3 3 2 3 2 was a closer fit to the raw centiles. However, overall there were some drawbacks with this model. First, in younger ages (2 to 8) the upper centiles of quartic fit were lower than the raw centiles. Second, the lower centiles of the fit were much higher than the raw centiles at the upper end of the age band. Thus, it was decided to treat the result of cubic model over the ages 13 to 15 by averaging the two models. Also, the cubic curves were extended over 17 and 18 years manually. Again, the general behaviour of the splined cubic fit and knowledge of girls' growth was used to obtain a reasonable chart.

The plot of the original measurements and the smoothed centiles is presented in Figure 8.5b. There is evidence of increasing variance over ages 10 to 15 as the points fan outwards and the distance between the centiles steadily increase as the children grow. This could be a simple explanation of why technically allowing for the skewness (z^2), and kurtosis (z^3) in the coefficients of quartic model could produce a better fit over this period. The overall distribution of the points about the 97th and 3rd centiles appears to be fairly even over the complete age range, except for the ages 5 to 8 where the 97th centile appears to be low.

The distribution of the observations falling between the smoothed centiles is presented in Table 8.9. It is evident that overall the observed and expected percentages are close. Fifty two percent of points fall below the 50th centile. In subgroups of age, the largest differences between observed and expected occurred in ages 6 to 9 between centiles 50th and 75th, and the 25th and 50th, with 28.9% and 22.8% respectively in comparison with 25%. This indicates that either the smoothed 75th centile is slightly high, or the 50th centile slightly low, or both. The next biggest differences between observed and expected occurred between centiles 90th and 97th, and over 97th in age subgroup 14-18. This might be explained by the unreliability of the observations at ages 17 and 18. As Figure 8.5a shows, in these ages the fitted curves are higher than raw ones; therefore, there are relatively few observations above the upper centiles. The overall Pearson χ^2 statistic (7 df) without considering the age effect was found to be 7.58 ($p=0.43$). In addition, allowing for the grouping, the Pearson χ^2 statistic (28 df) was 47.05 (p=0.014). After excluding the ages 17 and 18 from the last age group, the Pearson χ^2 statistic (28 df) reduced to 35.25 (p=0.16).

Figure 8.6a presents the Z-scores of the smoothed centiles to the girls' height with the expected ones. Over all ages, the smoothed 10th, 25th, 50th, 75th, and 90th Z-scores are very close to the expected ones except for the last two years of age. There are some departures from the expected line for the 3rd and 97th centiles. For example, at the ages of 6 and 7, the 3rd centile is too low as can be seen in Figure 8.5a where the distribution of observations is skewed. Or, at 9 and 10 the 97th centiles are too high where the spread of observations suddenly has a drop in comparison with 7 and 8 (Figure 8.5b). A Normal quantile-quantile (Q-Q) plot is presented in Figure 8.6b, and the distribution of Z-scores is close to Normal (p=0.22; W test).

The variation in Z-scores across the years of age (considering each year of age as a group) was examined by Q-tests. The mean and SD of Z-scores and the p-values of W tests at each age for the girls' height model and the corresponding tests are displayed in Table 8.10a & b. As part a of Table 8.10 shows apart from the last two years of age, means and SD of Z-scores are close the 0 and 1. And the distributions of Z-scores across ages are approximately Normal. As could be expected, for achieving an overall good fit one would inevitably sacrifice a little exactness in fit to some of the individuals.

In Table 8.10b the results of Q-tests after excluding the ages 17 and 18 from the analyses are displayed. The p-values of test of location, scale, and shape are: p_M=0.22, p_S=0.06 and p_W=0.47. This indicates an overall good fit to the data.

8.3.2.2 Boys' height

In modelling boys' height it was observed that two models, one quartic and the other quintic, fitted the data reasonably well. However, as might be expected, both fits had problems in the upper tail where the observed centiles are not reliable. The result of the quartic model (4 3 3 2 0 2) were smooth curves which were a bit higher at the upper end of the age band than the trend of the raw centiles whereas the result of the quintic fit (5 3 3 2 0 2 0) was a bit lower than the general trend of the raw centiles. Since overall the quartic fit appeared better, it was decided to use this model and treat the upper tail of the age band by averaging the two models over the ages sixteen to eighteen. It should be noted that both smoothed models were splined at the age fifteen. The smoothed centiles of boys' height are presented in Table 8.11. The parameters of boys and girls' height models are displayed in Appendix D.

The smoothed centiles and the raw ones of boys' height are presented in Figure 8.7a. Overall, the smoothed curves are close to the raw centiles. Almost 50% of points lie on each side of the 50th centile, 51% below and 49% above. The fit of the 3rd centiles is reasonably close to the raw ones whereas the 97th centiles is a little too high. The 50th centile is above the raw one at ages 11 to 13 and afterwards, at 14 and 15, lower. Over these ages (11 to 15) due to the puberty growth spurt in boys the spread at some ages is increased (Figure 8.7b).

Table 8.12 summarises the distribution of observations between the centiles. The percentage of points falling above the 97th centile is 2.0%; the expected Figure is 3.0%. This indicates that the smoothed 97th centile is a fraction above the raw one. Overall, the percentages of the points falling between the centiles are fairly close to expected values. The percentages of points in the intervals defined by the 3rd, 10th,..., 90th, and 97th are 2.6%, 7.4%, 14.7%, 26.2%, 23.9%, 16.1%, and 6.8% respectively. In subgroups of age ([2, 6], [7, 10], [11, 15], [16, 18]) apart from ages 16 and over and noting that the 97th centile being a little high, there is no big difference between the observed and expected numbers. The overall test of significance of the model was not significant ($\chi^2_7 = 5.99$; $p=0.54$). Moreover, considering the age groups, the differences of observed and expected across the subgroups was not significant either ($\chi^2_{28} = 29.83$; $p=0.37$).

Figure 8.8a shows a plot of Z-scores based on the fitted model and the expected Z-scores from the Normal distribution. The 97th centiles over most ages are slightly below the expected value showing that the fitted ones are a bit too high. In general, the closeness of the two sets of centiles and the absence of any systematic trend are indications of a good overall fit. This can also be seen by looking at the Normal plots of the Z-scores in subgroups of age, illustrated in Figure 8.8b.

In addition, the variation in distribution of Z-scores across ages was investigated using Q-tests. Overall, there was no strong evidence against the hypotheses that the means and standard deviations of Z-scores in all ages is zero and one ($p_M=0.16$, and $p_S=0.09$ respectively). Also, the overall test of shape of the distribution of Z-scores being Normal is not significant ($p_W=0.17$). The mean, standard deviation, and the p-values of W tests across ages and the result of above Q-tests are presented in Tables 8.13a & b.

8.3.3 Conclusions

A comparison of growth charts of boys and girls are shown in Figure 8.9. As it is clear from the Figure 8.9a up to the age of 10 years old, the 50th centiles of girls' weight (pink) lies below the boys' one (cyan) but the spread of the weights are fairly similar. From 11 to 14 years old girls are generally heavier than boys and the upper centiles of girls' weights are higher than boys; whereas, due to wide spread of girls' weights the lower centiles are below the boys' and later on catch up with them. In contrast, from 15 and onwards all centiles of the boys' weights are higher than girls however the spread of both sets of centiles are still the same.

Figure 8.9b presents the centiles of height of boys (green) and girls (pink) together. According to Figure 8.9b up to age 7 years old, all of the boys' heights centiles lie above those of the girls'; however, from 8 to 13 years old this situation is reversed and the girls height centiles are higher. It can be seen from the figure that after 13 years old the boys are taller than girls and the spread of the centiles is a little wider than that of girls. In conclusion, it appears from the Figure 8.9 that in the period of puberty (10-14) on average girls are slightly larger than boys, as it is well known.

8.4 Discussion

8.4.1 A suggestion for finding the powers p and q in HRY method

The HRY method is a powerful technique but in practice, this valuable method has not been used as widely as it might have been, possibly due to the problems in choosing the powers p (t) and q (z). Since there is no formal theory for obtaining these values (p and q), the solution is mostly left as a matter of trial and error. Trying higher order polynomials has the disadvantage of overfitting the end tails, also it is not obvious how these coefficients should be expressed as a function of skewness (z^2) and kurtosis (z^3). So in general, the essential question is which combination of p and q produces the best fit. Experience in looking at the general change of the skewness and kurtosis across ages may help to achieve a reasonable fit.

In our study we considered a regression procedure to obtain the powers p and q, and examined it throughout the work. Generally, the procedure can be summarised as follows; first, divide the data to a reasonable number of groups (say 10 to 15 groups) – in our data each year of age was used as a group – and compute the Normal scores of the measurements in each group. Since in an appropriate fit Normal scores should be fairly close to the Z-scores (SDS) we use these term interchangeably. Then, as the terms z^2 and z^3 describe the skewness and kurtosis of the data, two columns should be constructed containing these values corresponding to each measurement. Afterwards, the median (or mean) of each age group can be chosen as the average age of individuals in that group and the columns of age^2, age^3, ... should be constructed. Next, each column of age and its powers should be multiplied by the terms z, z^2 and z^3 which later on can be used to study how the polynomials in z change in relation to age.

Finally, in a multiple regression procedure, regress the measurements of interest (weight, height) on the set of variables; age and its higher orders (age^2,...), z, z^2, z^3 and their interaction terms which have already been produced. Then, the powers of age and z with significant coefficients indicate the best choices of p and q which can be used to fit the models in **GROSTAT II**. As an example Table 8.14 and Figure 8.10 illustrate the results of this regression procedure for boys' height measures and the corresponding fit. As can be seen from the Table 8.14, the coefficient of $cage^4$ (Cage; since data from ages 2-18 used in this analysis age is centred at 10, see section 5.4.2) is significant; also the interaction terms for all degree of age are significant too. So the model *4 3 3 3 3 3* is a good model from which to begin.

There are a few points that should be noticed here:

first, the model shown in Figure 8.10 is not good at the higher ages partly because, the data in older ages especially 17 and 18 are not reliable (the sample size is low, and most raw centiles values at the age of 18 are lower than 17; Figure 8.10). In order to avoid problem associated with small sample sizes at older ages, all models were obtained using the above procedure on the data in age range 2 to 15 years old. Then the curves were splined at 13 and 15 for girls and boys respectively. Second, in finding the values of q, sometimes the higher order of z is significant whereas the lower is not. For example for a power of age, z^3 is significant but z^2 is not. If this is the case for the higher order polynomial on age, our practical experience showed that generally using a lower value of q for higher degrees of age (p) is a better choice. However, sometimes it is worth to trying all the possible combinations as well as looking at different polynomials order (p) with significant coefficients in the regression procedure. Third, if another form of asymmetry (higher order than z^3) is of interest, it can be investigated. However, in this example the forth degree of age was studied but a higher order polynomial could also be considered.

8.4.2 Piecewise polynomials (PCWS)

It is possible to fit smoothly joining piecewise polynomials using 'grafted polynomials' to obtain smoothly joining centiles curves. Details can be found in Pan et al. (1992), and Saber and Wild (1989). In general, a piecewise polynomial regression, or a segment polynomial regression is a special continuous case of piecewise regression with continuity constraints of various orders in which the individual phase models are polynomials. Briefly, in smoothing curves from age band a to b, suppose $a=\xi_0<\xi_1<...<\xi_m=b$, then the piecewise polynomials can be expressed as $f_j(x)=\sum_{i=0}^{p_j}\beta_{ij}x^i$, $\xi_{j-1}<x<\xi_j$, $j=1,2,...,m$; the ξ's are join points and we assume q_j continuous derivatives at ξ_j but discontinuous when $s\geq q_j$, for $j=1,...,m-1$, i.e.

$$\left[\frac{d^s}{dx^s}\sum_{i=0}^{p_j}\beta_{ij}x^i\right]_{x=\xi_j}=\left[\frac{d^s}{dx^s}\sum_{i=0}^{p_{j+1}}\beta_{i,j+1}x^i\right]_{x=\xi_j},\quad s=0,1,2,...,q_j-1.$$

Thus these constraints make the curves smooth at the join points.

The piecewise polynomials procedure is more general than the GROSTAT spline procedure because successive polynomials do not to be of the same order and there are no restrictions on the β_{ij} other than those imposed by the requirements of smoothness at ξ_{j-1} and ξ_j. The procedure PCWS has been implemented in **GROSTAT II**, and can be used to fit smoothly joined 'grafted polynomials'. Thus, after finding suitable models for the subgroups of ages for girls and boys separately using the procedure discussed in 8.4.1, PCWS procedure was used to smooth the curves. It was observed that mostly the curves were not smooth at join points especially those of the extreme centiles. However, some of the limitations such as: the maximum number of subgroups, maximum number of coefficients in the model, and failure in getting the smooth centile curves for some of the models has been mentioned by Rasbash et al. (1993). Some general suggestion like what orders of p and q's should be fitted across the centile distribution in neighbouring age ranges to join curves smoothly has been made but this was inconsistent with our desire for general models that in principle can be obtained in modelling using this procedure. In contrast to the difficulty experienced in implementing PCWS models in GROSTAT, the spline procedure works well. Therefore, we carried on using the spline procedure in **GROSTAT II**.

Furthermore, after finding appropriate suggestions for p and q for age subgroups and choosing a reasonable common model as a whole for the age range, splines with more than one join points were tried. In this modelling approach for girls ages 5, 9, 13, and for boys 6, 10, 15 were chosen as the join points; and splining the curves using lower degrees (quadratic: *2 3 3 3* and cubic: *3 3 3 3*) were tried but the results of these models were not better than those already described.

8.4.3 Estimation of the raw centiles

The essential difference between the HRY non-parametric approach and parametric approaches is that parametric models assume that after transformation data are Normally distributed in each age group, so that z's corresponding to each observation y can be estimated from mean and SD. The whole analysis therefore turns on the effectiveness of the Normalising transformation. Apart from the complication of finding the scale on which the observations are Normally distributed, one of the problems of this approach is that tests of Normality that are available are not particularly powerful (they often give 'non-significant' results when applied to samples of moderate size (Healy, 1994)), so relying on a P-value greater than 0.05 does not prove the truth of the null hypothesis of Normality.

In contrast, the HRY non-parametric approach is simpler and it rests on no particular mathematical model for the distribution of the variables in the population from which the reference sample has been drawn. Thus, since age was recorded in years, across each year of age the centiles were determined using ranking methods (4.2.2) and SPSS software. If the age data are not grouped the first stage (CBOX) of the HRY method derives the raw centiles by constructing a succession of age groups, correcting for small differences in age within the group by linear regression, and then computing 'raw' centiles by ranking.

8.4.4 Variance correction

Since our data are grouped in years the appropriateness of variance estimations were considered by looking at Healy's correction (Healy, 1962). From 4.2.2 it can be seen that the correction term is:

$$1-\frac{b^2}{12*var(Y)};\qquad b=\frac{dy}{dx}.$$

For both weight and height measurements the average values of b and variances were computed from the corresponding data. It was observed that for weight measurements the values of b for boys and girls were 0.098, and 0.096; with average variances 0.035, and 0.039 respectively (on Log scale). Therefore, the variance correction coefficients for boys and girls were about 98% and for the standard deviations about 99%, which is negligible.

Also, for height measurements of children the average values of b and variances were 5.17 cm/year, and 4.46 cm/year and 63.6 cm^2, and 53.5 cm^2 respectively. Consequently, the standard deviation correction factor is 98%, which is negligible.

8.4.5 Comparison of charts with data from the rest of urban Iran

In chapter 6 (section 6.3) we have showed that raw centiles for urban Tehran fell within the 95% confidence interval for urban Iran children. Figure 8.11 compares the raw centiles of height and weight of boys and girls for rest of urban areas of Iran with the fitted centile for urban Tehrani children. It can be seen from the graphs that for weights of girls and boys over all the ages the lower centiles especially the 3rd and 10th, which are more of interest in monitoring the nutritional status of children, are compatible and the urban Tehran's centiles covers the centiles of the rest of urban Iran reasonably well. The one exception is the 3rd centile for heights of boys under the age of 6. Here the raw centile appear to be very low although the 10th centile is reasonably close to the fitted curve.

For girls up to the age of 8 years old, all the centiles are compatible and also up to the age of 13 years old the median and the 75th centiles are following the urban Tehran's curves. However afterwards the upper centiles for urban non-Tehrani girls are lower than those presented in the charts. Furthermore, when the fit of the GROSTAT and the MLn model (Table 7.5a) for girls weight were compared using the data from the rest of urban Iran with the grid test, using 17 age groups, the Pearson χ^2 statistics with 119 df were 491.7 for GROSTAT and 616.5 for the MLn curves. This shows that the GROSTAT curves provide better fit. It has been mentioned before that sometimes test statistics are significant but the comparison of numerical values is a general indictor of closeness of the fit.

Similarly, for boys up to the age of 6 years old almost all the centiles are compatible. Also the lower centiles (3rd, 10th) are fairly compatible over all the ages, but from 11 and older the 50th and higher centiles of urban Tehrani children are above their counterparts from rest of the urban areas.

In comparisons of heights of girls and boys, it was observed (Figure 8.11) that the compatibility of lower centiles over all the ages was reasonably good. But, like weight, the upper centiles for urban Tehran were higher than the centiles for the rest of the urban areas of Iran. In conclusion, these findings show that GROSTAT models are acceptable models and confirms that the use of urban Tehran is an appropriate baseline for urban areas of the country.

8.4.6 Rural areas of Iran - What is a reasonable chart to be used?

We have seen that urban Tehran is a reasonable basis for the urban areas of Iran. On the other hand, in the analysis in chapter five it was shown that the urban children were heavier and taller than their rural counterparts. Now the question is how these significant differences can practically be addressed. The possible suggestion of separate charts for the urban and rural areas of the country is not practically appropriate for many reasons. For example, as a result of the recent development programmes by the government, the rural areas are developing and many villages are growing into small towns. Hence, if separate charts were used for rural and urban areas, the charts would be changed frequently from rural to urban. Such changes would cause a lot of confusion for the health workers, but would also be logistically difficult and expensive.

Since in growth monitoring the children whose position are in the extreme centiles of charts are of most interest, one suggestion is to study the charts of the urban Tehran which already have been constructed (8.4) in comparison with the charts representing the average of rural areas of Iran to see what these differences look like. Thus, for rural areas of Iran the growth charts of weight and height were constructed using the spline technique in the HRY method (Appendix D). The models for smoothing the raw centiles of weight of rural girls and boys were *4 3 3 3 1 1* and *4 3 3 2 3 1* respectively. In this smoothing procedure curves were splined at 13 and 15 years old for girls and boys.

Figure 8.12a-b present the comparisons of growth charts of girls and boys for urban Tehran (cyan) with the average growth charts of rural areas (pink). It can be seen from the Figure 8.12 that for both sexes the overall shape of the change of weight with ages are similar for urban and rural children. However the centiles of the rural areas in general are shifted lower down in comparison with their urban counterparts but these differences for the lower centiles do not seem large. The spread of rural centiles is wider in younger ages but is narrower in puberty and upper end of the age range.

As the extreme centiles are more of interest especially those for children whose nutritional status is recognized as deprived (i.e. lower centiles), these centiles were looked at first. According to the Figure 8.12 the 3rd and 10th centiles of weight of urban Tehrani children do not lie much above the 3rd and 10th of the rural ones for both sexes. Thus, it was decided to see to which centiles of the rural children, the lower centiles of girls and boys' weight of urban Tehrani children correspond. In order to study this the smoothed centiles of urban Tehran were fed into the rural models and it was observed that for weight, the 3rd and 10th centiles of urban Tehrani girls correspond on average to the 5.6th centile (SD=1.4%) and 18th centile (SD=3%) of rural children. Also, for boys these centiles correspond on average to the 5.5th (SD=1.4%) and 17th (SD=2.6%) respectively.

On the whole, the 25th, 50th, 75th, 90th, and 97th centiles of weight of urban Tehrani girls correspond to the 39.5th, 62th, 87.7th, 96th, and 99th of their rural counterparts. And for weight of urban Tehrani boys, these centiles correspond to the 38th, 65.7th, 85.5th 95th and 99.6th centiles of their rural counterparts. In conclusion, since there is not a great deal of difference between the lower centiles of weight of urban Tehrani girls and boys and their rural counterparts, the weight charts of urban Tehran, which before was recognized reasonable for urban areas of Iran, can be used all over the country. In the rural areas, a child whose weight is just below the 3rd centile should, when followed up, at least grow in a parallel line with the lowest chart centile. If required, the average centiles for rural children that correspond to the standard centiles on the weight charts could be added to the charts as has been done for height, see below.

The comparisons of height of urban Tehrani girls and boys and their rural counterparts are shown in Figure 8.13a-b. It can be seen in Figure 8.13a that the shape of the curves for heights of rural girls are very similar to their urban Tehrani counterparts, but the whole distribution is shifted down and roughly the 10th centile of height of rural girls lies on 3rd centile of urban Tehrani girls. In general, for boys the situation is the same, however, the systematic trend of increasing height of boys continues up to 18 whereas for girls it starts to flatten off around 15 years old. Then as with the weights of these children, it was of interest to see how the lower centiles of height of rural girls and boys compared with the centiles of their urban Tehrani counterparts. Hence, the smoothed centiles of height of rural children was looked at in the urban Tehran's models. And it was observed that the 3rd and 10th centiles of rural girls height approximately correspond to 1st centile (SD=0.1%) and 3rd centile (SD=0.5%). The 25th, 50th, 75th, 90th, and 97th centiles of height of rural girls correspond to the 10th, 23th, 45th, 70th, and 89th of their urban Tehrani counterparts.

In addition, a similar analysis was carried out for boys' height and it was observed that the 3rd and 10th centiles of height of rural boys approximately correspond to the 1st (SD=0.4%) and 3.7th (SD=1%) of their urban Tehrani counterparts. In general, the 25th,..,97th of rural boys' height (pink) correspond to the 11th, 26th, 49th, 73th, and 91th centiles of boys' height in urban Tehran (green). Thus, the results of these analyses show that the previous charts of height of girls and boys in urban Tehran with an extra 1st centile can be used in all areas of Iran where in fact the 1st centile corresponds to the 3rd centile of rural children.

Figure 8.14 presents the growth chart of girls' height for Iran where the previous chart is extended to include the 1st centile. This chart is applicable to all areas of the country, since it not only includes the pattern of height development of urban girls but also with this chart the corresponding centiles of rural children can be identified. After finding that on average what percentages point on urban Tehran's height charts the 3rd and 10th of rural children correspond to, further centiles for rural children corresponding the standard for urban Tehran were derived by computing the average position of percentiles of urban Tehran's height models in the models for rural children.

For better estimation of position of the centiles since the curves were splined this procedure has been repeated in different age ranges for boys and girls separately. The results of these analyses are presented in Appendix D. The approximate averages are shown in the chart of boys' height where any corresponding centiles of rural heights can be read off. For example, as was explained for the case of weight in rural areas, if a child's height is short, his/her height could be compared and followed up in comparison with the 1st centile of the charts which corresponds to 3rd centile of rural children and so on. The boys' height chart is presented as Figure 8.15. The added 1st centiles of height of girls and boys are also presented in Tables 8.8 and 8.11. In conclusion, the extended charts of height enable one set of charts to be used for both groups of urban and rural children. It is worthwhile mentioning that the models for smoothing the raw centiles of height of rural girls and boys were *3 3 3 3 3* and *4 3 3 2 1 1* respectively (Appendix D), which were splined at 13 and 15 years old.

In Chapter 5 it is shown that the children in urban Semnan are generally among the largest and children in Kohkiloyeh-Boyerahmad are among the smallest. Appendix D, Figure D.1 shows the weights and heights of boys in these two provinces plotted on the standard weight and height charts, also the corresponding grid tests are presented in Table D.5. As can be seen in the figure the points for the urban Semnani boys are clearly shifted towards the higher centiles and those for rural children in Kohkiloyeh-Boyerahmad are shifted downward. However, the extend of the shifts are clearly not sufficient to invalidate the charts for individual use since in no age group do a large number of points lie beyond the extreme centiles on the charts.

8.4.7.1 Comparison of centiles for urban Iran with other studies

In this section the results of the present study are compared with the previous work in Iran and with the international references. Ayatollahi (1993a) in a critical appraisal of human growth studies in Iran, compared his findings with previous studies in Iran. According to him children in Iran are now substantially bigger than they were some twenty years ago. It is, however, impossible to quantify these changes since the previous studies were based on children attending clinics, who are unlikely to form a representative sample.

School children 6-12 years in a city in south Iran (Shiraz) were studied by Ayatollahi (1991). In general, comparison of the results of our study with his work shows that the 50th centiles of weights of Shirazi boys and girls are close to ours. The lower centiles of weights in our study are lower than his, and the upper centiles of weights of boys and girls of urban Tehrani children are higher than in Shiraz. That is the median of the two distributions are close but the tails of the distribution of weights in urban Tehran are longer on both sides. A comparison of girls' height showed similar results; from 8 to 12 years the median of height of urban Tehrani boys and the upper centiles are higher than his findings.

Ayatollahi also found non-significant differences between the growth of children whose families immigrated during the war (1980-88) from the southern and western borders of Iran to Shiraz. From this observation he suggested that his results are applicable for urban areas of Iran. However, our data from urban Tehran covers the wider variation which exist in the growth pattern of children in Iran (5.4.3).

Eveleth and Tanner (1990) discussed the variation in growth worldwide. Here we compare the proposed growth charts with the NCHS standards. The NCHS reference centiles (Hamill et al., 1979) have been recommended by WHO to be used worldwide, and in most developing countries as well as Iran these charts are widely used in health fields (PHC). But due to differences in genetics and environmental factors, in different parts of the world children grow up differently. So the application of these standards can be misleading. As an example, the proposed growth charts of weights of girls and boys are compared with the corresponding NCHS references in Figure 8.16a-b. According to the Figure 8.16a median weights of urban Irani girls (cyan) up to the age of 13 years old is below the 25th centiles of NCHS standards (pink), and the 25th centiles of girls' weight is about the 3rd centile of the NCHS reference. Likewise, for boys (Figure 8.16b) up to the age of 15 years old the 25th centile of weight corresponds approximately to the 3rd centile of the NCHS data and similarly the median of boys' weight is lower than the 25th centile of the NCHS standard. For girls after the age of 13 years there is a catch up in weight, and the centiles of urban Iran start to cross the NCHS centiles but on the whole the set of centiles are substantially below the corresponding NCHS data.

An analytical comparison of these centiles was carried out to find out to which percentage point on our charts the NCHS centiles are correspond. After feeding the NCHS data in to the models of weight for girls and boys separately it was observed that for girls up to age of 13 and for boys up to 15 years old the 3rd, 10th, 25th, and 50th centiles of the NCHS references correspond on average to the 28th, 51th, 69th and 84th centiles of the Iranian charts.

A comparison of height chart of urban Irani girls and boys and the NCHS standards are presented in Figure 8.17a-b. Similar differences to weight were found when comparing heights of children (i.e. the 3rd NCHS centiles of height of girls and boys lie on 25th of our charts and the Iranian medians are lower than the 25th centiles of the NCHS data). Also, the results of the analysis of the NCHS centiles in the models of height of girls and boys showed that up to the age 13 years old for girls and 15 for boys the 3rd, 10th, 25th, 50th centiles of the NCHS references correspond on average to the 28th, 44th, 63th, 84th centiles of heights of Iranian children.

In summary, the comparison of charts show that the NCHS reference centiles are considerably higher than our centiles. Even some general chart like home-based weight chart (WHO, 1978) which consists of two lines; 3rd centiles of girls' weight and 50th centiles of boys' weight may be not an appropriate solution for solving the worldwide variation in growth since, as has been shown, the 3rd centiles of girls of the NCHS reference correspond approximately to 25th centiles of our children. So, for the smaller children, for whom monitoring is of greatest importance, the NCHS standard is not an appropriate standard to use.

8.4.7.2 A Z-scores based comparison

Traditionally, in the United States and some other countries, percentiles are used as cutoff points. In other parts of the world, either Z-scores or percent of median are used. However, the calculation of the percent of median does not take into account the distribution of the reference population around the median. Therefore, interpretation of the percent of median is not consistent across age and height levels nor across the different anthropometric indices (WHO, 1986). WHO favours the use of Z-scores which have the statistical property of being Normally distributed, thus allowing a meaningful average and standard deviation for a population to be calculated.

The Z-scores cut off point recommended by WHO, CDC, and others to classify anthropometric level is 2SD units below the reference median for the indices (weight-for-age (**WA**), height-for-age (**HA**)). The proportion that falls below a Z-score of -2 is generally compared with the reference population in which 2.3% fall below this cutoff. Thus, the

EPINUT (a programme for nutritional anthropometry), which is implemented in EPI INFO 6 (1996), was used to compute the Z-scores of weight and height of urban Tehrani children. Since EPINUT computes the Z-scores of these indices only for up to 114 months old, the results in this part relates to the age 2 to 9 years old. NCHS believes that the age of more than 10 years old should not be used for comparison of nutritional status (WHO, 1983; p. 62). The EPINUT results are summarised in Table 8.15a; the average percentages of Z-scores of weight less than -2SD for girls and boys were 23.3% and 26.4% respectively. Similarly, the average percentages of Z-scores of heights below -2 were 25.3% and 24.6%. These findings are similar to ours reported in the previous section of the relation between the Iranian and NCHS centiles.

Table 8.15b shows the WHO classification of prevalence of low anthropometric values (-2SD). On average the prevalence of low anthropometric values (<-2SD) for weight-for-age of our children is high, and their heights is a medium prevalence of stunting (Table 8.15b).

At first sight these findings imply that children in Iran are generally seriously malnourished, however, a comparison of our centiles for weight and height with NCHS centiles in Figures 8.16 and 8.17 show a general shift downwards but little changes in the spread of the centiles in each age group. If serious malnourishment were present in a proportion of the population the upper centiles could be normal and the lower centiles would be relatively much lower than expected. The general shift of the centiles suggests either that the usual diet of Iranian children does not promote growth as much as the usual diet of American children or that there are genetic differences between the populations. Either way the WHO reference is too high for identification of nutritional status of Iranian children.

WHO also recommended (EPI INFO Manual, 1994) that the prevalence of low anthropometric indices should be evaluated as in Table 8.15b by one-year intervals for children less than six years of age. However, the analysis in Table 8.15a gives an analytical overview of the criteria derived by EPINUT for ages 2 to 9. Table 8.15a shows that the average percentages used in the above presentation are generally valid for an overall overview, and that there is no remarkable difference in percentage of low Z-scores in different age groups.

It should be mentioned that in the above comparisons standards for urban Iran were compared with the NCHS reference standard. If data from rural Iran had been compared with this reference the differences would have been bigger, and larger percentages of children than presented above would have been classified as malnourished.

Table 8.1 Mean and SE of weight and height of girls and boys by age, urban Tehran, National Health Survey 1990-2

AGE*	Weight (kg)						Height (cm)					
	Girls			Boys			Girls			Boys		
	No.	Mean	SE	No.	Mean	SE	No.	Mean	SE	No.	Mean	SE
2	82	11.66	(0.20)	87	12.20	(0.23)	82	86.2	(0.68)	87	87.0	(0.68)
3	111	12.61	(0.21)	109	13.47	(0.24)	111	91.5	(0.70)	109	93.6	(0.70)
4	139	14.65	(0.26)	117	15.29	(0.26)	139	98.2	(0.64)	117	100.3	(0.63)
5	128	16.09	(0.28)	121	16.65	(0.28)	128	105.8	(0.66)	121	108.0	(0.73)
6	128	17.73	(0.27)	136	17.90	(0.28)	128	111.8	(0.55)	136	111.9	(0.53)
7	121	20.06	(0.34)	129	20.64	(0.34)	121	118.1	(0.67)	129	117.7	(0.61)
8	129	22.20	(0.38)	121	22.79	(0.37)	129	124.6	(0.65)	121	124.3	(0.58)
9	123	24.94	(0.44)	98	24.86	(0.45)	123	128.3	(0.65)	98	128.8	(0.68)
10	113	27.12	(0.50)	118	29.03	(0.60)	113	132.9	(0.67)	118	134.7	(0.75)
11	98	31.63	(0.82)	101	30.66	(0.59)	98	139.7	(0.93)	101	138.1	(0.91)
12	105	37.26	(0.96)	88	34.85	(0.88)	105	146.0	(0.85)	88	144.1	(1.05)
13	82	40.92	(0.98)	79	38.81	(1.01)	82	150.9	(0.97)	79	149.8	(1.14)
14	70	45.61	(1.23)	86	45.06	(1.04)	70	154.7	(0.95)	86	157.5	(1.15)
15	89	51.03	(1.01)	61	52.07	(1.45)	89	157.0	(0.68)	61	164.0	(1.09)
16	56	51.50	(1.18)	56	56.33	(1.38)	56	158.1	(0.88)	58	168.7	(1.17)
17	74	53.01	(1.12)	43	57.63	(1.35)	74	156.7	(0.71)	43	169.0	(1.18)
18	54	54.24	(1.49)	47	58.40	(1.20)	54	157.5	(0.81)	47	169.8	(1.16)

Table 8.2 Smoothed percentiles of girls' weight (kg) by age, urban Tehran, Iran

Age (years)	Smoothed percentiles 3rd	10th	25th	50th	75th	90th	97th
2	7.87	9.27	10.47	11.61	12.71	13.88	15.49
3	8.72	10.20	11.50	12.79	14.12	15.59	17.65
4	9.68	11.26	12.68	14.14	15.71	17.52	20.08
5	10.82	12.45	14.01	15.67	17.53	19.72	22.74
6	12.20	14.06	15.75	17.55	19.55	21.89	25.19
7	13.83	15.73	17.49	19.43	21.67	24.34	28.14
8	15.29	17.32	19.27	21.52	24.20	27.44	32.03
9	16.57	18.88	21.20	23.97	27.33	31.39	37.10
10	17.85	20.60	23.45	26.94	31.20	36.31	43.41
11	19.38	22.72	26.25	30.60	35.90	42.16	50.73
12	21.46	25.51	29.80	35.06	41.36	48.68	58.50
13	24.38	29.15	34.17	40.21	47.25	55.25	65.79
14	28.18	33.57	39.08	45.50	52.79	60.89	71.47
15	32.49	38.12	43.66	49.89	56.83	64.60	74.72
16	36.15	41.80	46.70	52.50	59.10	66.60	77.10
17	37.90	43.40	48.30	53.70	60.50	68.00	78.50
18	38.70	44.00	49.10	54.37	61.20	69.00	79.50

Table 8.3 Observed and expected number (percentage) of girls' weight measurements falling between centiles in subgroups of age, urban Tehran

Centile	2-5 years O	2-5 years E	6-9 years O	6-9 years E	10-13 years O	10-13 years E	14-18 years O	14-18 years E	Totals O	Totals E
	No %	No %	No %	No %	No %	No %	No %	No %	No %	No %
>97	14 3.0	13.7 3.0	11 2.2	15.0 3.0	10 1.5	11.9 3.0	7 2.0	10.3 3.0	42 2.5	51.1 3.0
90-97	41 9.0	32.3 7.0	38 7.6	35.1 7.0	26 6.5	27.9 7.0	18 5.2	24.0 7.0	123 7.2	119.1 7.0
75-90	65 14.1	69.0 15.0	75 15.0	75.2 15.0	50 12.6	59.7 15.0	48 14.0	51.5 15.0	238 14.0	255.3 15.0
50-75	112 24.4	115.0 25.0	148 29.5	125.2 25.0	104 26.1	99.5 25.0	89 26.0	85.7 25.0	453 26.6	425.5 25.0
25-50	117 25.4	115.0 25.0	118 23.5	125.2 25.0	120 30.2	99.5 25.0	80 23.4	85.7 25.0	435 25.6	425.5 25.0
10-25	60 13.0	69.0 15.0	69 13.8	75.2 15.0	56 14.1	59.7 15.0	62 18.1	51.5 15.0	247 14.5	255.3 15.0
3-10	40 8.7	32.2 5.0	25 5.0	35.1 7.0	25 6.3	27.9 7.0	31 9.0	24.0 7.0	121 7.1	119.1 7.0
≤3	11 2.4	13.7 3.0	17 3.4	15.0 3.0	7 1.7	11.9 3.0	8 2.3	10.3 3.0	43 2.5	51.1 3.0
Total	460		501		398		343		1702	

O : Observed
E : Expected

Table 8.4.a Mean, standard deviation (SD), and p-value of Z-scores of fit to the girls' weight, urban Tehran

Age	Mean	SD	P-values	No.
2	0.0293	0.9825	0.32	82
3	-0.1159	1.0049	0.99	111
4	0.1053	1.0365	0.65	139
5	0.0504	1.0206	0.67	128
6	-0.0064	0.9614	0.07	128
7	0.0675	0.9658	0.92	121
8	0.0388	1.0150	0.32	129
9	0.0695	0.9090	0.72	123
10	-0.0765	0.8087	0.79	113
11	-0.0206	1.0398	0.38	98
12	0.0914	0.9945	0.58	105
13	-0.0154	0.8128	0.05	82
14	-0.0767	0.9829	0.36	70
15	0.0547	0.9129	0.46	89
16	-0.1978	0.9038	0.60	56
17	-0.2376	0.9292	0.52	74
18	-0.1444	1.0808	0.83	54

b. Q-tests of Z-scores; girls' weight, urban Tehran

	Q_M	Q_S	Q_W
No. of groups	17	17	17
Statistic	14.46	21.60	28.20
d.f	16	16	34
P-value	0.56	0.16	0.75

Table 8.5 Smoothed percentiles of boys' weight (kg) by age, urban Tehran, Iran

Age (years)	Smoothed percentiles						
	3rd	10th	25th	50th	75th	90th	97th
2	7.80	9.39	10.89	12.38	13.76	15.00	16.39
3	8.99	10.64	12.16	13.65	15.08	16.49	18.21
4	10.21	11.93	13.48	15.01	16.54	18.16	20.30
5	11.46	13.27	14.87	16.48	18.17	20.08	22.73
6	12.77	14.68	16.37	18.12	20.04	22.31	25.59
7	14.13	16.18	18.01	19.96	22.19	24.92	28.96
8	15.59	17.82	19.84	22.06	24.69	27.99	32.92
9	17.17	19.62	21.90	24.47	27.61	31.58	37.55
10	18.92	21.65	24.24	27.26	31.00	35.76	42.90
11	20.90	23.85	26.93	30.48	34.92	40.55	48.94
12	23.17	26.30	30.04	34.19	39.40	45.96	55.60
13	25.70	29.30	33.63	38.60	44.46	52.00	62.63
14	28.90	32.90	38.00	44.50	51.00	59.83	69.72
15	33.05	38.10	44.00	50.60	57.30	66.40	76.10
16	37.17	42.63	48.50	55.50	62.50	70.60	80.30
17	41.30	46.60	52.00	58.65	65.40	73.20	82.70
18	42.95	48.40	53.40	59.73	66.36	74.20	83.59

Table 8.6 Observed and expected number (percentage) of boys' weight measurements falling between centiles in subgroups of age, urban Tehran

Centile	2-5 years O	2-5 years E	6-9 years O	6-9 years E	10-13 years O	10-13 years E	14-18 years O	14-18 years E	Totals O	Totals E
	No %	No %	No %	No %	No %	No %	No %	No %	No %	No %
>97	15 2.6	17.1 3.0	15 3.2	14.0 3.0	9 2.2	12.5 3.0	1 1.0	4.4 3.0	40 2.5	48.0 3.0
90-97	25 4.4	39.9 7.0	37 8.0	32.6 7.0	29 7.0	29.0 7.0	7 4.7	10.4 7.0	98 6.0	112.0 7.0
75-90	88 15.4	85.5 15.0	59 12.7	69.9 15.0	60 14.5	62.3 15.0	24 16.2	22.2 15.0	231 14.4	240.0 15.0
50-75	138 24.2	142.5 25.0	141 30.2	116.5 25.0	101 24.3	103.7 25.0	40 27.0	37.0 25.0	420 26.2	399.5 25.0
25-50	154 27.0	142.5 25.0	124 26.6	116.5 25.0	109 26.3	103.7 25.0	39 26.4	37.0 25.0	426 26.6	399.5 25.0
10-25	86 15.2	85.5 15.0	53 11.4	69.9 15.0	63 15.1	62.3 15.0	22 14.7	22.2 15.0	224 14.0	240.0 15.0
3-10	40 7.0	39.9 5.0	29 6.2	32.6 7.0	31 7.5	29.0 7.0	13 8.8	10.4 7.0	113 7.0	112.0 7.0
≤3	24 4.2	17.1 3.0	8 1.7	14.0 3.0	13 3.1	12.5 3.0	2 1.4	4.4 3.0	47 3.0	48.0 3.0
Total	570		466		415		148		1599	

O : Observed
E : Expected

Table 8.7.a Mean, standard deviation (SD), and p-value of Z-scores of fit to the boys' weight, urban Tehran

Age	Mean	SD	P-values	No.
2	-0.0514	0.9528	0.31411	87
3	-0.0688	1.0171	0.99953	109
4	0.0902	1.0347	0.64300	117
5	0.0009	1.0849	0.11409	121
6	-0.1400	1.0338	0.06910	136
7	0.0772	0.9812	0.34431	129
8	0.0623	0.9193	0.52180	121
9	-0.0339	0.9011	0.96982	98
10	0.1463	0.8940	0.33698	118
11	-0.1067	0.9267	0.75877	101
12	-0.0963	0.9989	0.27597	88
13	-0.1208	1.0077	0.97283	79
14	-0.0168	0.9386	0.21311	86
15	0.0553	0.9565	0.19188	61
16	0.0200	0.9571	0.91590	58
17	-0.1391	0.8232	0.94191	43
18	-0.1800	0.8623	0.51777	47

b. Q-tests of Z-scores; boys' weight, urban Tehran

	Q_M	Q_S	Q_W
No. of groups	17	17	17
Statistic	13.95	15.02	29.74
d.f	16	16	34
P-value	0.60	0.52	0.68

Table 8.8 Smoothed percentiles of girls' height (cm) by age, urban Tehran, Iran

Age (years)	Smoothed percentiles							
	1st	3rd	10th	25th	50th	75th	90th	97th
2	65.5	70.4	76.2	81.1	85.9	90.4	94.3	98.5
3	73.4	78.1	83.5	88.1	92.5	96.7	100.4	104.4
4	80.6	85.1	90.3	94.8	99.1	103.1	106.8	110.9
5	87.2	91.7	96.8	101.2	105.5	109.6	113.5	117.6
6	93.3	97.7	102.9	107.4	111.8	116.1	120.2	124.6
7	98.9	103.4	108.7	113.3	118.0	122.5	126.9	131.6
8	104.1	108.7	114.2	119.0	123.9	128.8	133.4	138.5
9	108.6	113.9	119.5	124.5	129.7	134.8	139.7	145.1
10	113.2	118.7	124.5	129.8	135.1	140.5	145.6	151.3
11	117.4	123.3	129.5	135.0	140.6	145.9	151.1	156.8
12	121.5	127.4	134.4	140.4	146.2	151.5	155.9	161.7
13	125.9	131.6	139.1	145.3	151.0	156.3	160.5	166.3
14	130.3	136.4	143.6	149.3	154.7	159.7	163.9	169.2
15	135.0	141.3	147.3	152.3	157.2	161.6	166.0	170.8
16	138.7	144.1	149.7	154.4	158.9	163.0	167.1	171.7
17	140.0	145.0	150.5	155.1	159.5	163.6	167.6	172.5
18	140.5	145.4	150.8	155.5	160.0	164.0	168.1	172.8

Table 8.9 Observed and expected number (percentage) of girls' height measurements falling between centiles in subgroups of age, urban Tehran

Centile	2-5 years O	E	6-9 years O	E	10-13 years O	E	14-18 years O	E	Totals O	E
	No %	No %	No %	No %	No %	No %	No %	No %	No %	No %
>97	14 3.0	13.7 3.0	17 3.4	15.0 3.0	7 1.8	11.9 3.0	6 1.7	10.3 3.0	44 2.6	51.1 3.0
90-97	34 7.4	32.3 7.0	29 5.8	35.1 7.0	29 7.3	27.9 7.0	15 4.4	24.0 7.0	107 6.3	119.1 7.0
75-90	74 16.1	69.0 15.0	70 14.0	75.2 15.0	53 13.3	59.7 15.0	47 13.7	51.5 15.0	244 14.3	255.3 15.0
50-75	109 23.7	115.0 25.0	145 28.9	125.2 25.0	97 24.4	99.5 25.0	75 21.9	85.7 25.0	426 25.0	425.5 25.0
25-50	121 26.3	115.0 25.0	114 22.8	125.2 25.0	102 26.6	99.5 25.0	90 26.2	85.7 25.0	427 25.0	425.5 25.0
10-25	55 12.0	69.0 15.0	88 17.6	75.2 15.0	62 15.6	59.7 15.0	75 21.9	51.5 15.0	280 16.5	255.3 15.0
3-10	33 7.2	32.2 5.0	31 6.2	35.1 7.0	38 9.5	27.9 7.0	29 8.5	24.0 7.0	131 7.7	119.1 7.0
≤3	20 4.3	13.7 3.0	7 1.4	15.0 3.0	10 2.5	11.9 3.0	6 1.7	10.3 3.0	43 2.5	51.1 3.0
Total	460		501		398		343		1702	

O : Observed
E : Expected

Table 8.10.a Mean, standard deviation (SD), and p-value of Z-scores of fit to the girls' height, urban Tehran

Age	Mean	SD	P-values	No.
2	0.0779	0.8487	0.26107	82
3	-0.0911	1.0326	0.50970	111
4	-0.0702	1.0969	0.38237	139
5	0.1063	1.0319	0.26339	128
6	0.0073	0.9112	0.14017	128
7	0.0447	0.9433	0.23207	121
8	0.0968	1.0346	0.86328	129
9	-0.1581	0.8920	0.30753	123
10	-0.2606	0.8452	0.61408	113
11	-0.0752	1.0622	0.15344	98
12	0.0367	0.9998	0.11219	105
13	0.0854	1.0252	0.51844	82
14	0.0749	1.0028	0.79624	70
15	0.0136	0.8764	0.60128	89
16	-0.0700	0.9006	0.99996	56
17	-0.4048	0.8771	0.14634	74
18	-0.2876	0.8911	0.98834	54

b. Q-tests of Z-scores; girls' height, urban Tehran

	Q_M	Q_S	Q_W
No. of groups	15	15	15
Statistic	17.76	22.80	30.00
d.f	14	14	30
P-value	0.22	0.06	0.47

Table 8.11 Smoothed percentiles of boys' height (cm) by age, urban Tehran, Iran

Age (years)	Smoothed percentiles							
	1st	3rd	10th	25th	50th	75th	90th	97th
2	64.4	70.1	76.6	82.1	87.1	91.5	95.3	99.3
3	74.7	79.9	85.7	90.4	94.9	98.8	102.3	106.0
4	83.2	87.9	93.1	97.5	101.5	105.2	108.6	112.3
5	90.2	94.6	99.5	103.5	107.4	111.0	114.5	118.6
6	96.1	100.3	104.9	108.9	112.8	116.6	120.4	124.8
7	101.2	105.3	109.9	113.9	117.9	122.1	126.3	131.3
8	105.8	109.8	114.5	118.7	123.0	127.6	132.4	138.0
9	110.0	114.2	119.1	123.5	128.2	133.3	138.6	144.9
10	114.1	118.5	123.7	128.4	133.6	139.2	145.1	152.0
11	118.2	122.9	128.4	133.6	139.2	145.3	151.6	159.0
12	123.0	127.4	133.4	138.9	145.0	151.4	158.0	165.8
13	128.0	132.7	138.7	144.6	150.9	157.5	164.2	171.9
14	133.5	138.6	144.5	150.8	157.0	163.7	170.2	177.4
15	139.0	144.4	150.7	157.3	163.5	169.5	175.4	181.6
16	143.5	149.0	155.4	162.1	168.0	173.7	179.0	184.4
17	146.0	151.7	158.0	164.3	170.2	175.6	180.5	185.5
18	147.5	152.8	158.8	165.0	171.0	176.1	180.9	186.1

Table 8.12 Observed and expected number (percentage) of boys' height measurements falling between centiles in subgroups of age, urban Tehran

Centile	2-5 years O	2-5 years E	6-9 years O	6-9 years E	10-13 years O	10-13 years E	14-18 years O	14-18 years E	Totals O	Totals E
	No %	No %	No %	No %	No %	No %	No %	No %	No %	No %
>97	16 2.8	17.1 3.0	6 1.3	14.0 3.0	8 1.9	12.5 3.0	2 1.4	4.4 3.0	32 2.0	48.0 3.0
90-97	42 7.4	39.9 7.0	34 7.3	32.6 7.0	23 5.5	29.0 7.0	9 6.1	10.4 7.0	108 6.8	112.0 7.0
75-90	82 14.4	85.5 15.0	83 17.8	69.9 15.0	75 18.1	62.3 15.0	23 15.5	22.2 15.0	263 16.4	240.0 15.0
50-75	132 23.2	142.5 25.0	123 26.4	116.5 25.0	92 22.2	103.7 25.0	35 23.6	37.0 25.0	382 23.9	399.5 25.0
25-50	152 26.7	142.5 25.0	124 26.6	116.5 25.0	101 24.3	103.7 25.0	42 28.4	37.0 25.0	419 26.2	399.5 25.0
10-25	84 14.7	85.5 15.0	62 13.3	69.9 15.0	65 15.7	62.3 15.0	24 16.2	22.2 15.0	235 14.7	240.0 15.0
3-10	43 7.5	39.9 5.0	28 6.0	32.6 7.0	37 8.9	29.0 7.0	10 6.8	10.4 7.0	118 7.4	112.0 7.0
≤3	19 3.3	17.1 3.0	6 1.3	14.0 3.0	14 3.4	12.5 3.0	3 2.0	4.4 3.0	42 2.6	48.0 3.0
Total	570		466		415		148		1599	

O : Observed
E : Expected

Table 8.13.a Mean, standard deviation (SD), and p-value of Z-scores of fit to the boys' height, urban Tehran

Age	Mean	SD	P-values	No.
2	0.0496	0.8384	0.36729	87
3	-0.1020	1.0174	0.99722	109
4	-0.1113	1.0415	0.82877	117
5	0.1754	1.1636	0.84110	121
6	-0.1407	0.9776	0.78071	136
7	-0.0214	0.9645	0.10717	129
8	0.1590	0.8866	0.94801	121
9	0.0500	0.8805	0.10456	98
10	0.0828	0.9342	0.82999	118
11	-0.1684	1.0015	0.75066	101
12	-0.1410	0.9743	0.07730	88
13	-0.1446	0.9865	0.35141	79
14	0.0330	1.0523	0.13554	86
15	0.0660	0.8743	0.00676	61
16	0.1186	0.9680	0.40244	58
17	-0.1117	0.8543	0.37265	43
18	-0.0626	0.8764	0.16285	47

b. Q-tests of Z-scores; boys' height, urban Tehran

	Q_M	Q_S	Q_W
No. of groups	17	17	17
Statistic	21.54	24.03	41.90
d.f	16	16	34
P-value	0.16	0.09	0.17

Table 8.14 Obtaining powers p and q in HRY method by regression procedure of height on powers of age and z; boys' height 2-18 years old, urban Tehran

Terms	coefficients	s.e	t
z	7.4688	0.13394	55.761
z^2	0.35349	0.057764	6.120
z^3	0.17266	0.039266	4.397
CAGE*	5.5913	0.028100	198.975
CAGE*z	0.79314	0.040022	19.818
CAGE*z^2	0.064678	0.017227	3.755
CAGE*z^3	-0.091810	0.012068	-7.608
$CAGE^2$	0.16441	0.0091524	17.963
$CAGE^2$*z	0.069682	0.012954	5.379
$CAGE^2$*z^2	-0.042881	0.0055841	-7.679
$CAGE^2$*z^3	-0.012808	0.0038627	-3.316
$CAGE^3$	-0.0042370	0.00063933	-6.627
$CAGE^3$*z	-0.012642	0.00091462	-13.822
$CAGE^3$*z^2	-0.0013632	0.00039478	-3.453
$CAGE^3$*z^3	0.0015141	0.00027901	5.427
$CAGE^4$	-0.0036286	0.00015013	-24.169
$CAGE^4$*z	-0.0016335	0.00021345	-7.653
$CAGE^4$*z^2	0.00041059	0.000092114	4.457
$CAGE^4$*z^3	0.00029161	0.000064292	4.536
Constant term =	133.41		

* Cage : age is centred at 10 years old

Table 8.15.a Percentage of Z-scores <-2SD of WA*, HA*, and WH* of urban Tehrani children in comparison with **NCHS**

Age	**WA**		**HA**		**WH**	
	Boys	Girls	Boys	Girls	Boys	Girls
2	23.0	25.6	24.1	23.2	9.2	15.9
3	33.0	27.9	27.5	47.7	15.6	21.6
4	20.5	30.9	28.2	36.0	16.2	14.4
5	41.5	26.6	24.8	25.0	24.8	20.3
6	33.1	21.9	32.4	25.8	24.3	15.6
7	20.9	19.0	18.6	14.0	12.4	15.7
8	18.2	17.1	17.4	14.7	14.9	19.4
9	18.4	17.9	23.5	16.3	16.3	9.8
Average	23.3	26.4	25.3	24.6	16.5	24.6

* WA : Weight-for-age

* HA : Height-for-age

* WH : Weight for Height

Table 8.15.**b** Prevalence of low anthropometric values (-2SD) compared with other surveys for children five years of age or younger♣

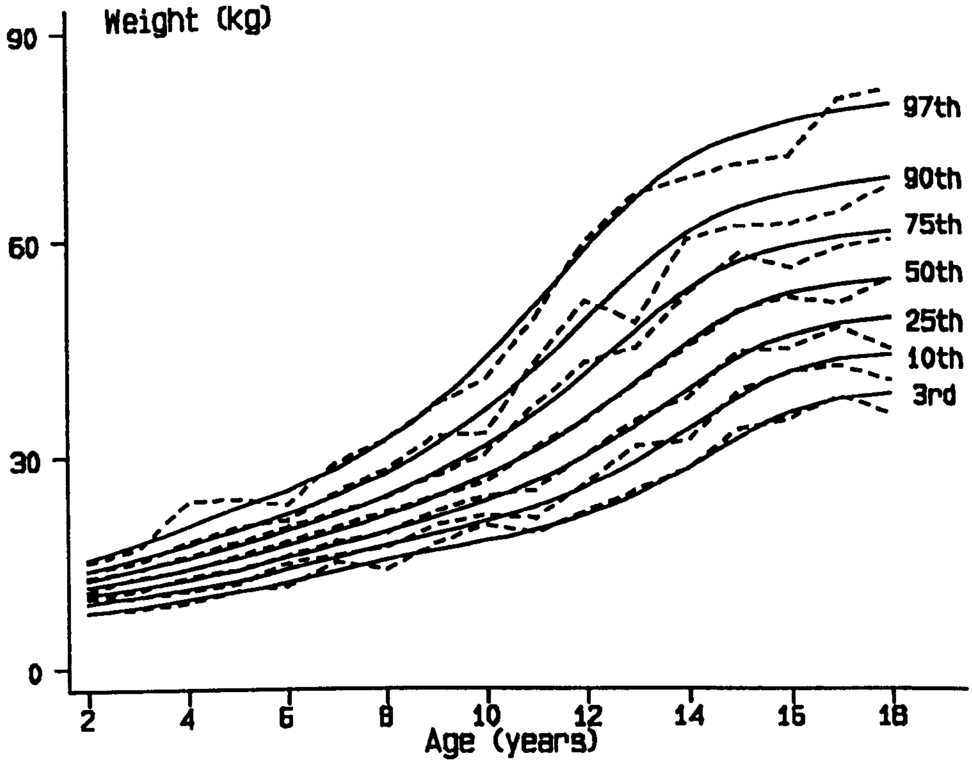
Relative Prevalence of Low Anthropometric Values

Index	**Low**	**Medium**	**High**	**Very High**
Low WH	<5.0%	5.0-9.9%	10.0-14.9%	≥15.0%
Low HA	<20.0%	20.0-29.9%	30.0-39.9%	≥40.0%
Low WA	<10.0%	10.0-19.9%	20.0-29.9%	≥30.0%

♣: EPI INFO 6 Manual (1994); p. 271

Figure 8.1

a) Weight of girls aged 2-18 years, urban Tehran: raw (--) and smoothed (-) centiles



b) Weight of girls aged 2-18 years, urban Tehran: smoothed centiles and original observations

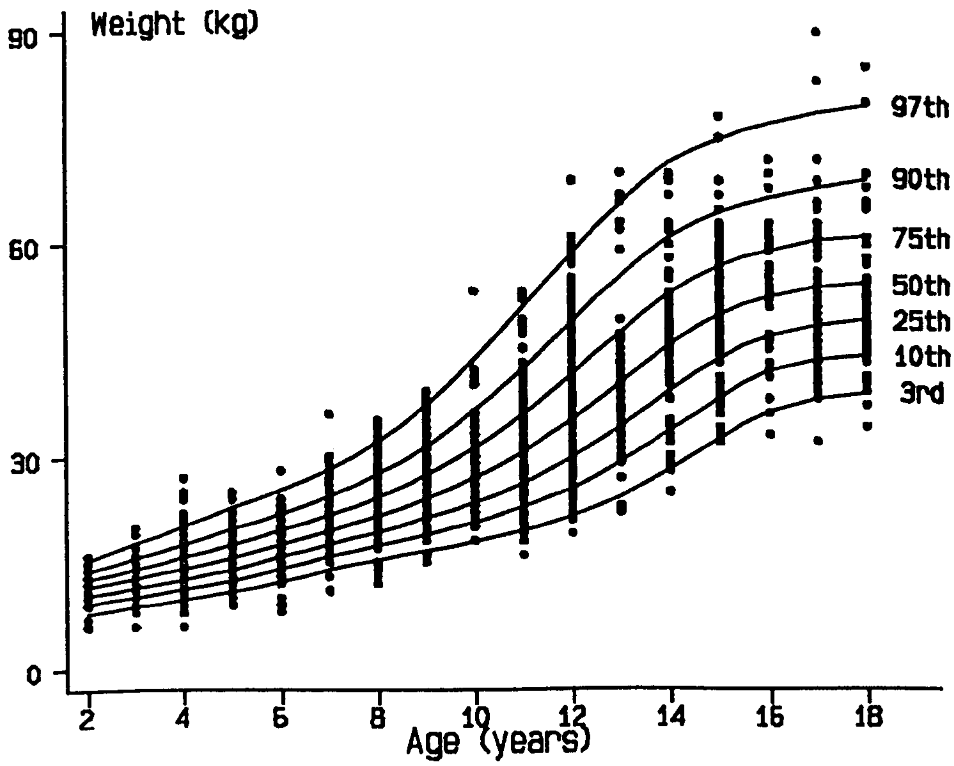
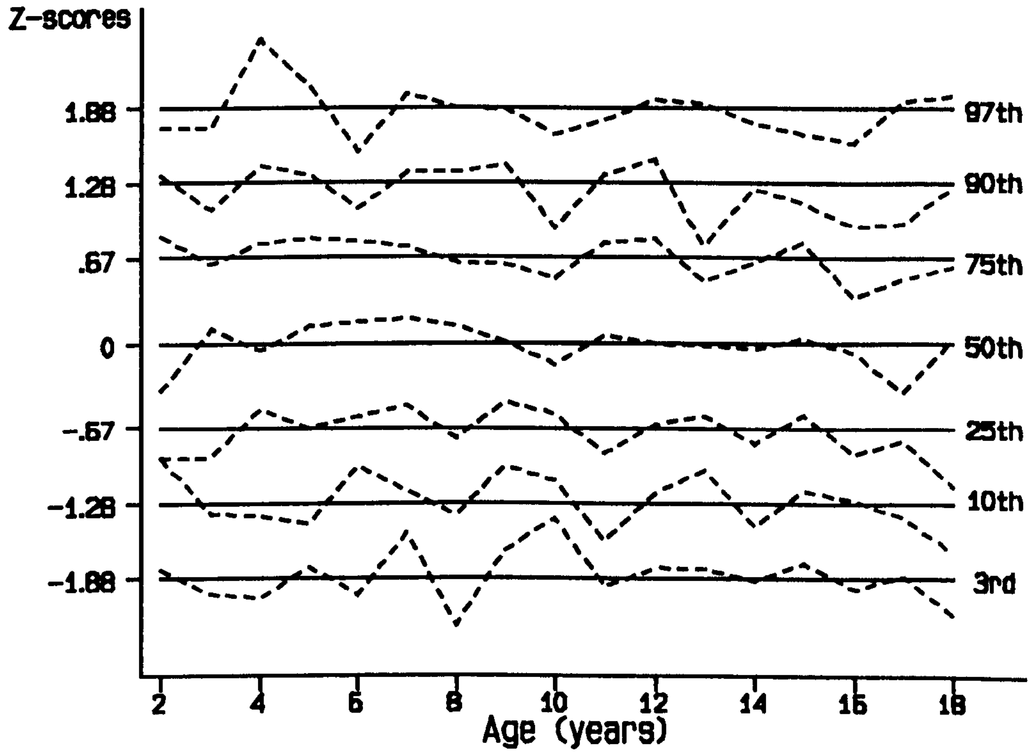
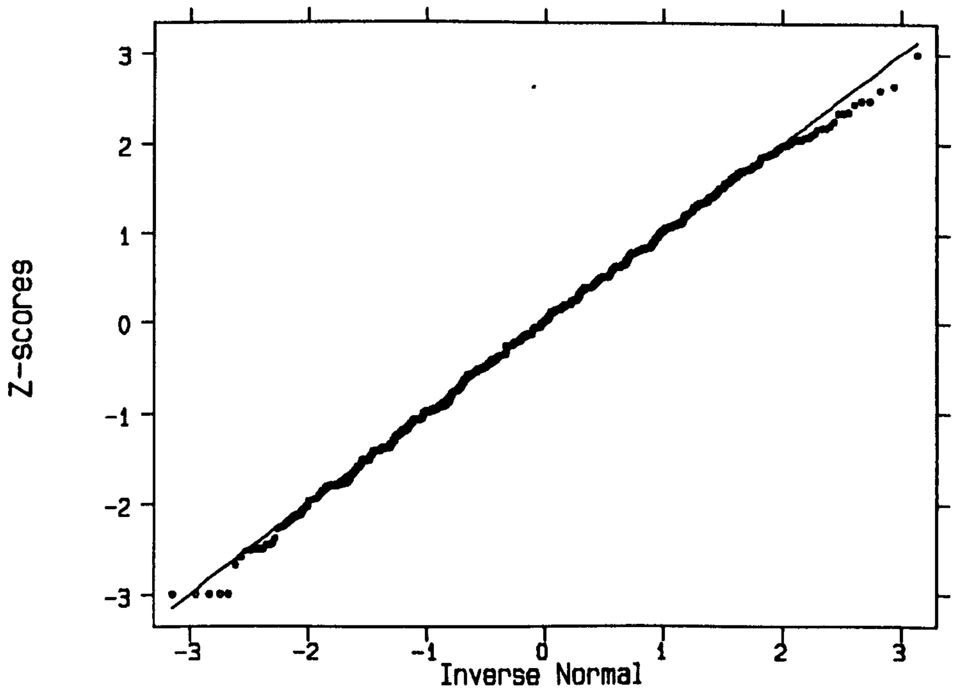


Figure 8.2

a) Weight of girls aged 2-18 years, urban Tehran: observed (--) and expected (-) Z-score centiles



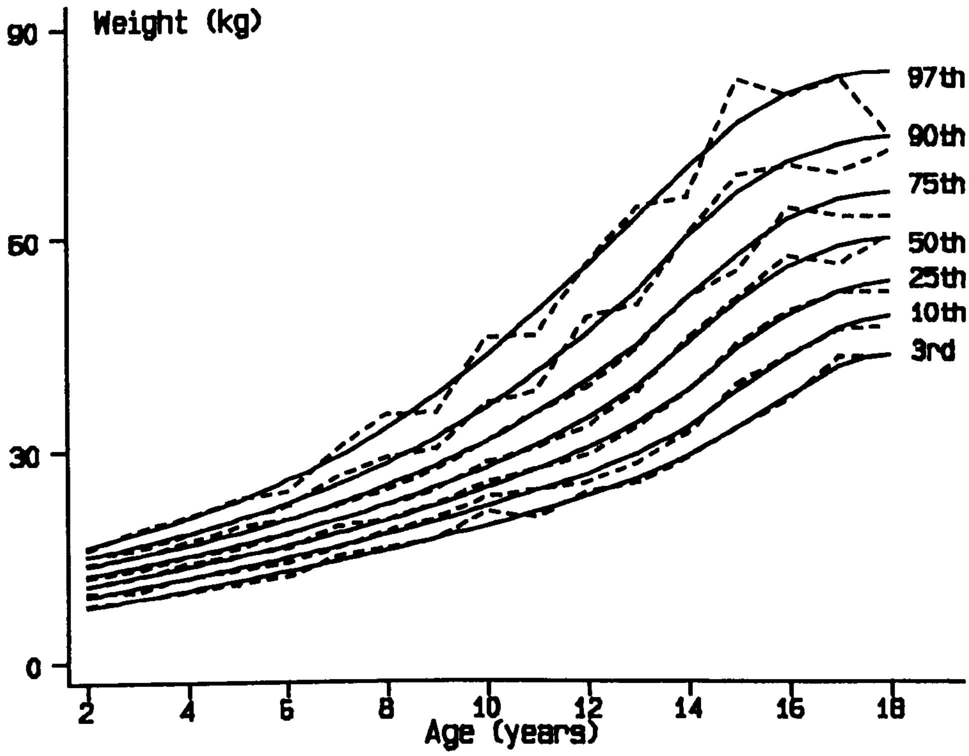
b) Weight of girls aged 2-18 years, urban Tehran: Normal plot of Z-scores



P=0.14 (W test)

Figure 8.3

a) Weight of boys aged 2-18 years, urban Tehran: raw (--) and smoothed (-) centiles



b) Weight of boys aged 2-18 years, urban Tehran: smoothed centiles and original observations

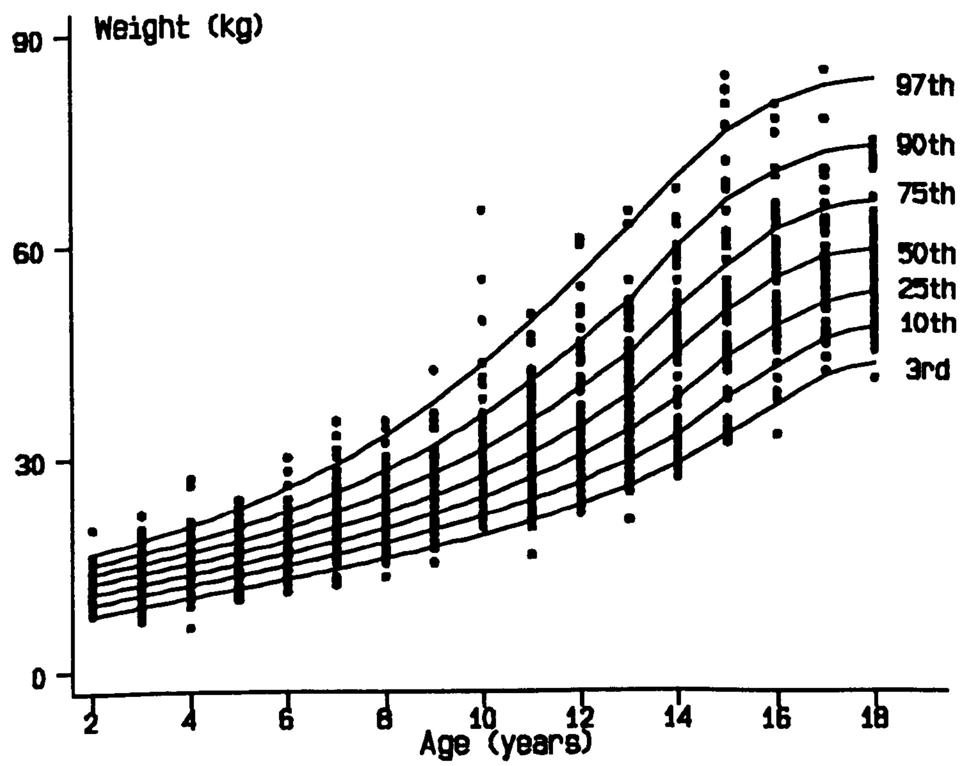
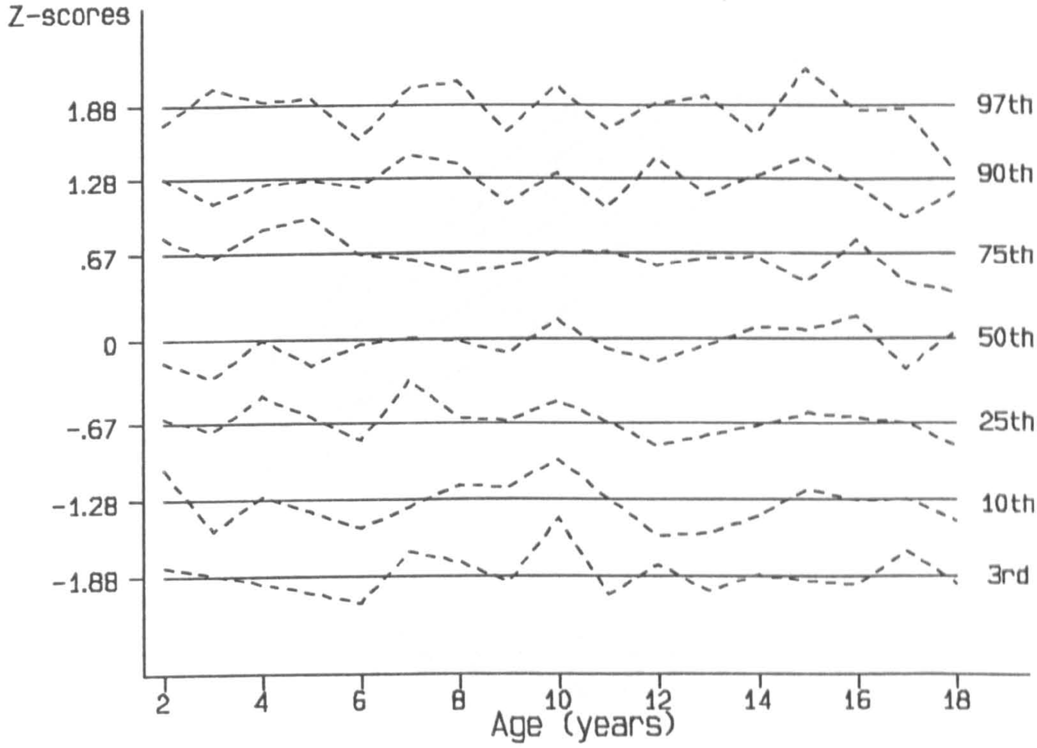


Figure 8.4

a) Weight of boys aged 2-18 years, urban Tehran: observed (--) and expected (-) Z-score centiles



b) Boys' weight, urban Tehran: Normal plots of Z-scores in age subgroups

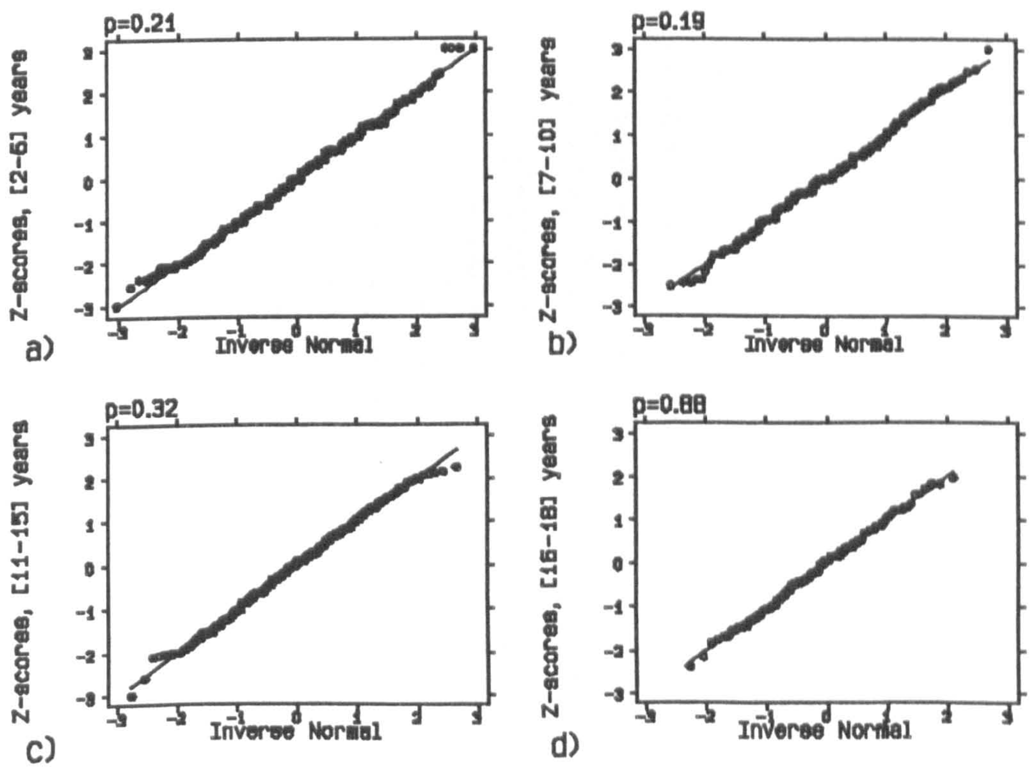
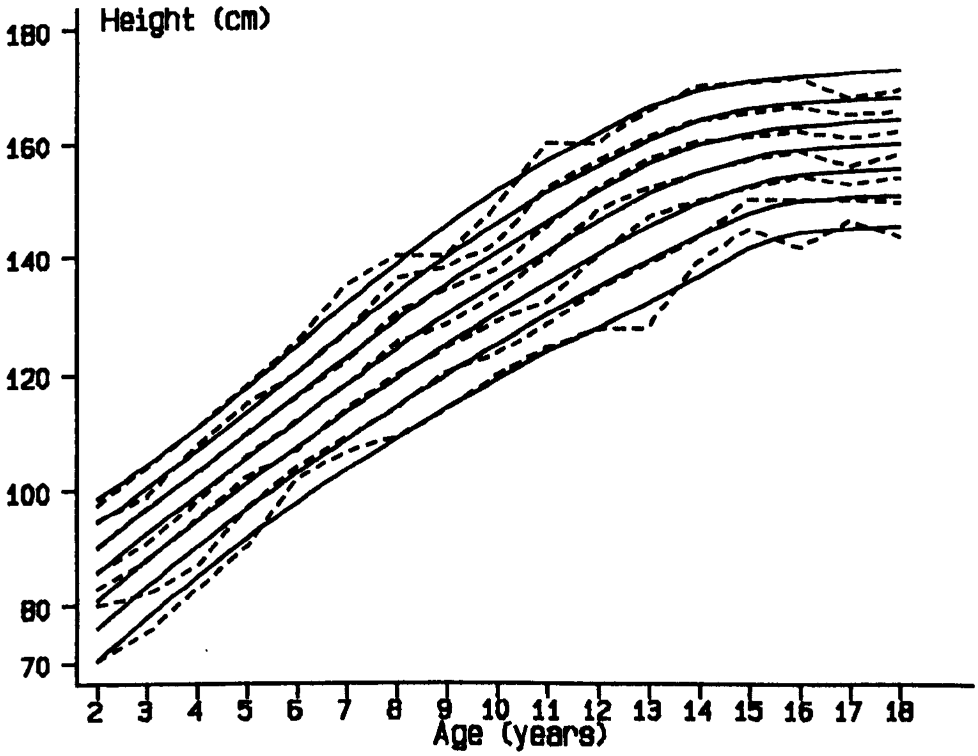
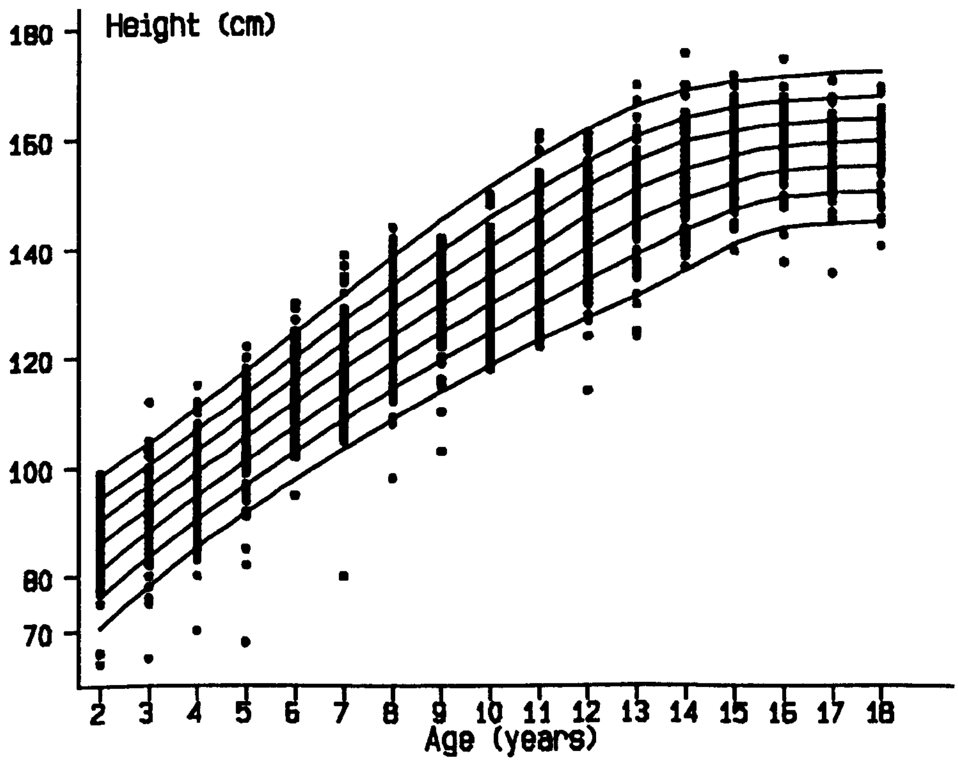


Figure 8.5

a) Height of girls aged 2-18 years, urban Tehran: raw (--) and smoothed (-) centiles*



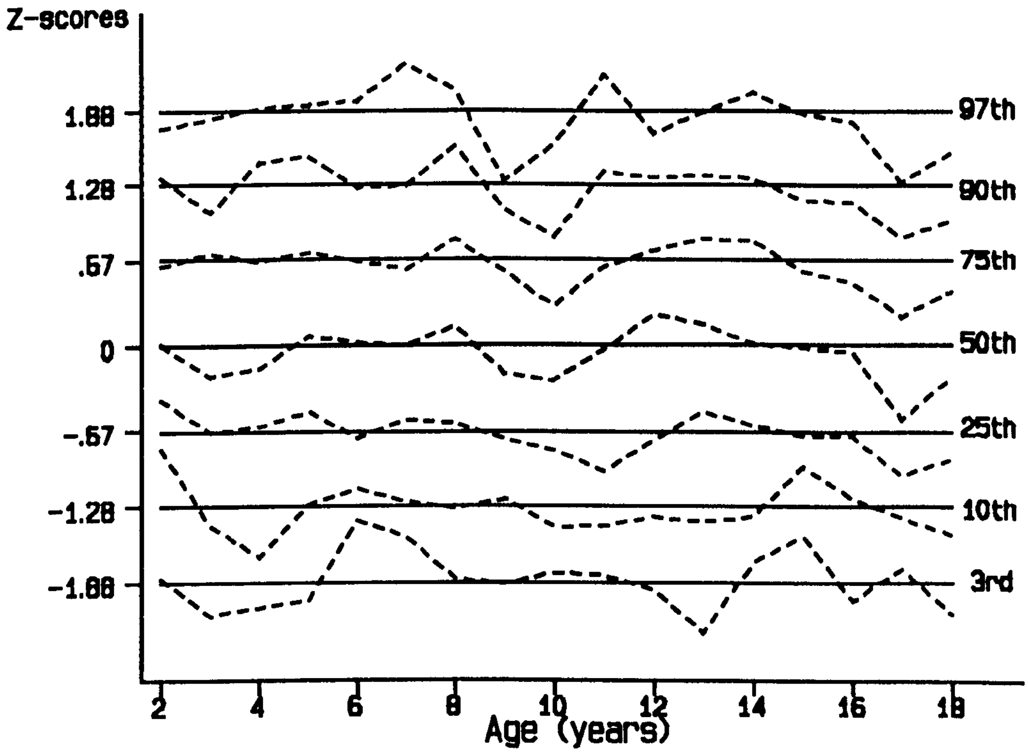
b) Height of girls aged 2-18 years, urban Tehran: smoothed centiles* and original observations



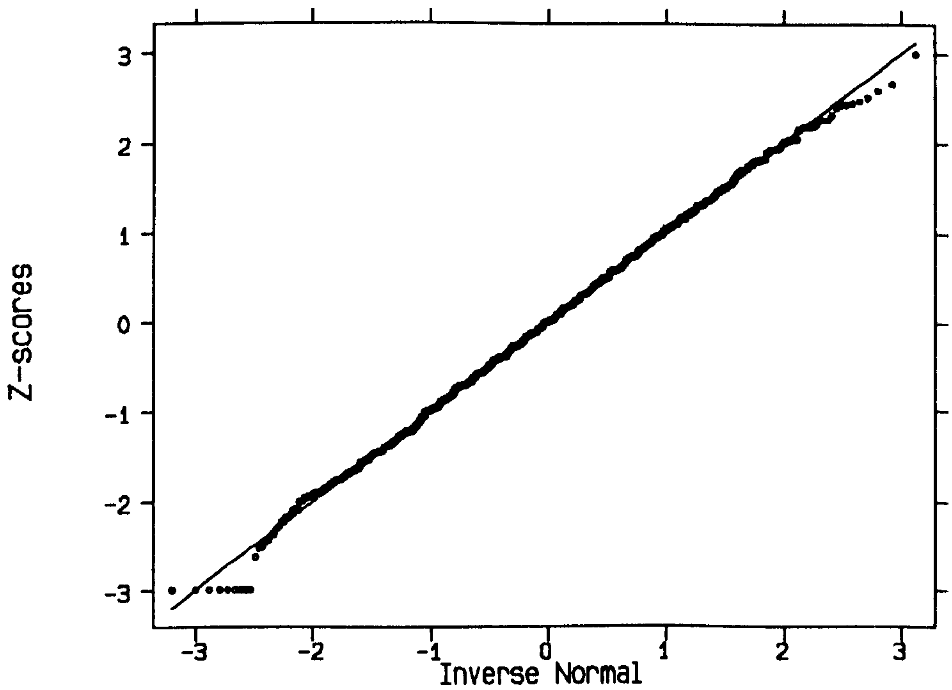
* : Seven centiles are shown; 3rd, 10th, 25th, 50th, 75th, 90th, and 97th

Figure 8.6

a) Height of girls aged 2-18 years, urban Tehran: observed (--) and expected (-) Z-score centiles



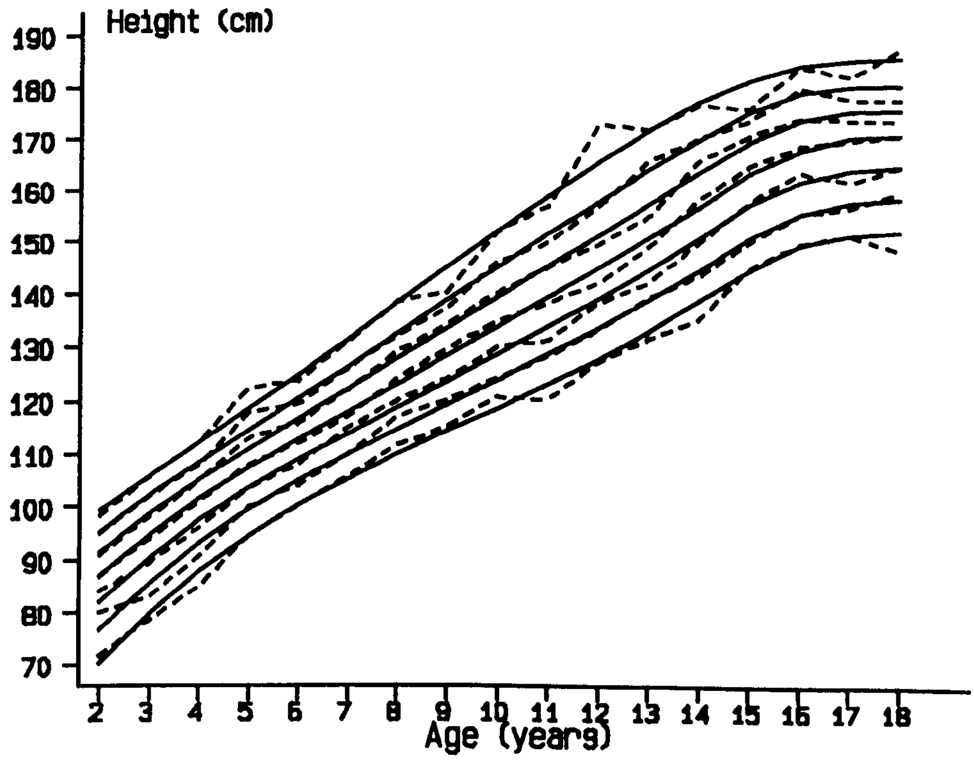
b) Height of girls aged 2-18 years, urban Tehran: Normal plot of Z-scores



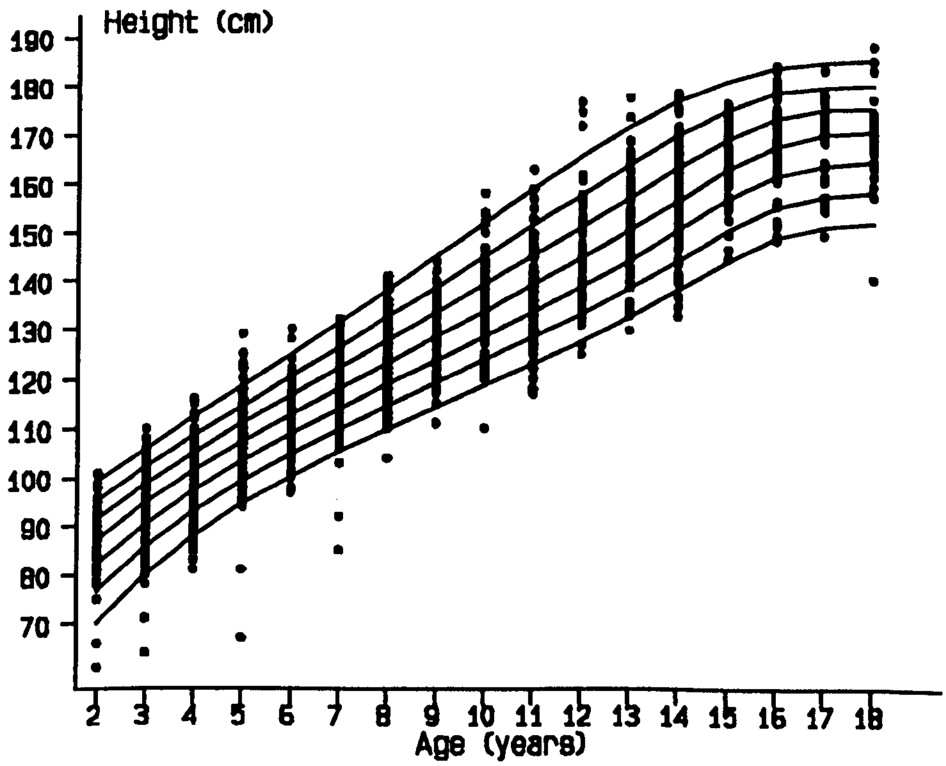
P=0.22 (W test)

Figure 8.7

a) Height of boys aged 2-18 years, urban Tehran: raw (--) and smoothed (-) centiles*



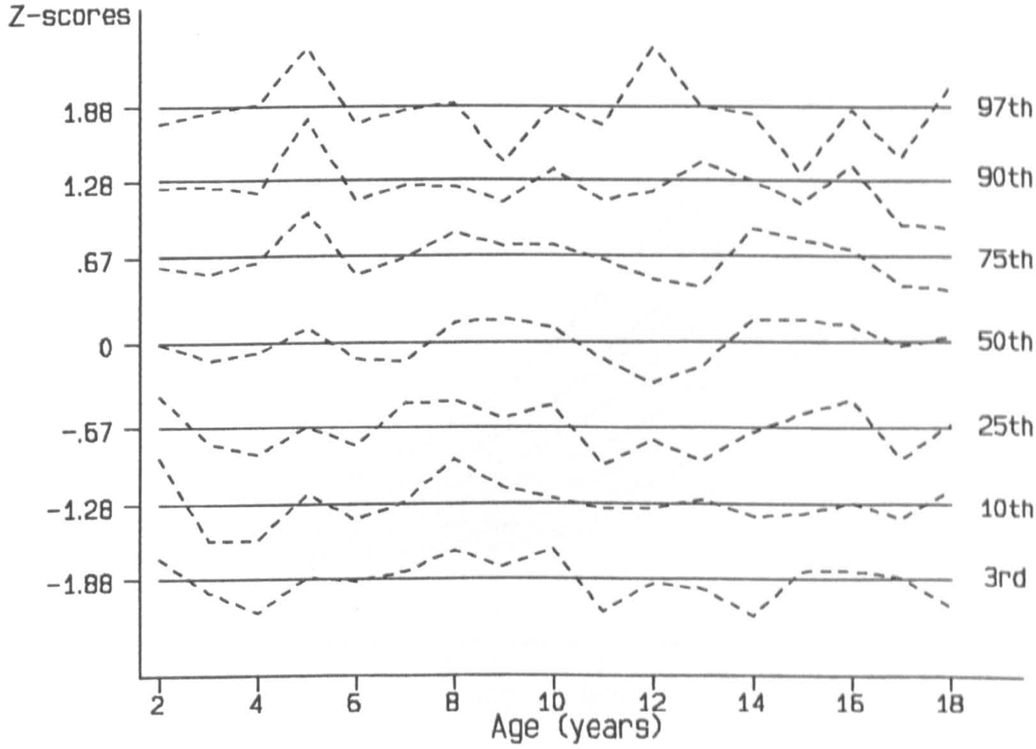
b) Height of boys aged 2-18 years, urban Tehran: smoothed centiles* and original observations



*: Seven centiles are shown; 3rd, 10th, 25th, 50th, 75th, 90th, and 97th

Figure 8.8

a) Height of boys aged 2-18 years, urban Tehran: observed (--) and expected (-) Z-score centiles



b) Boys' height, urban Tehran: Normal plots of Z-scores in age subgroups

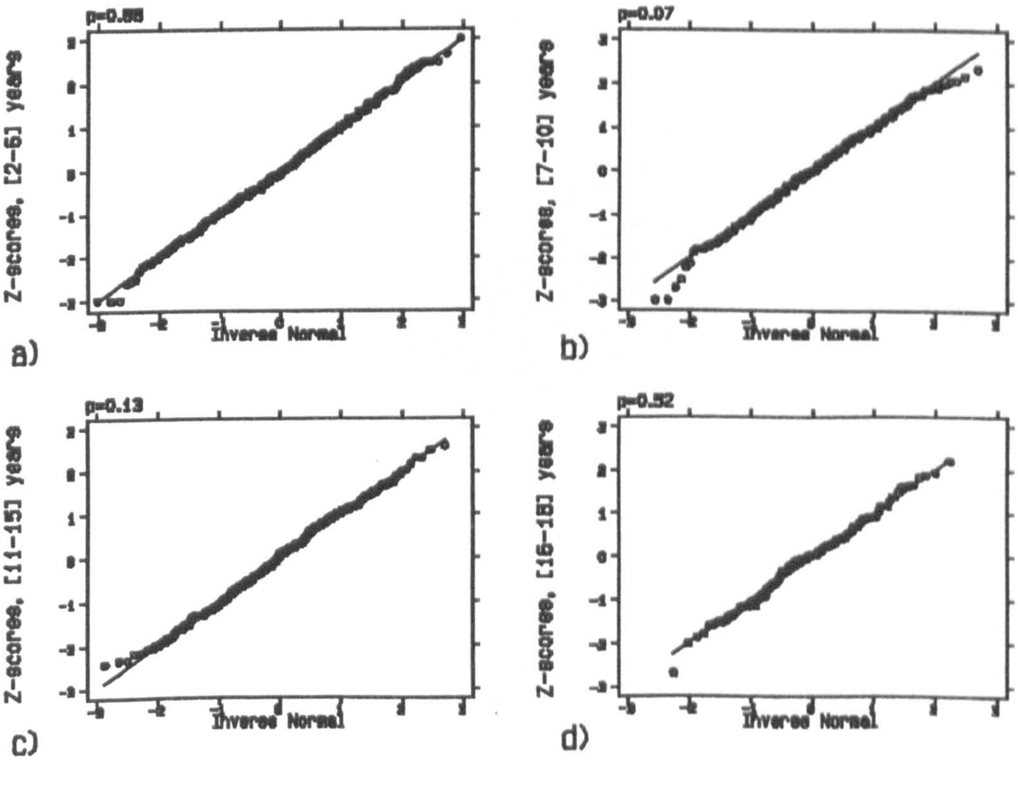
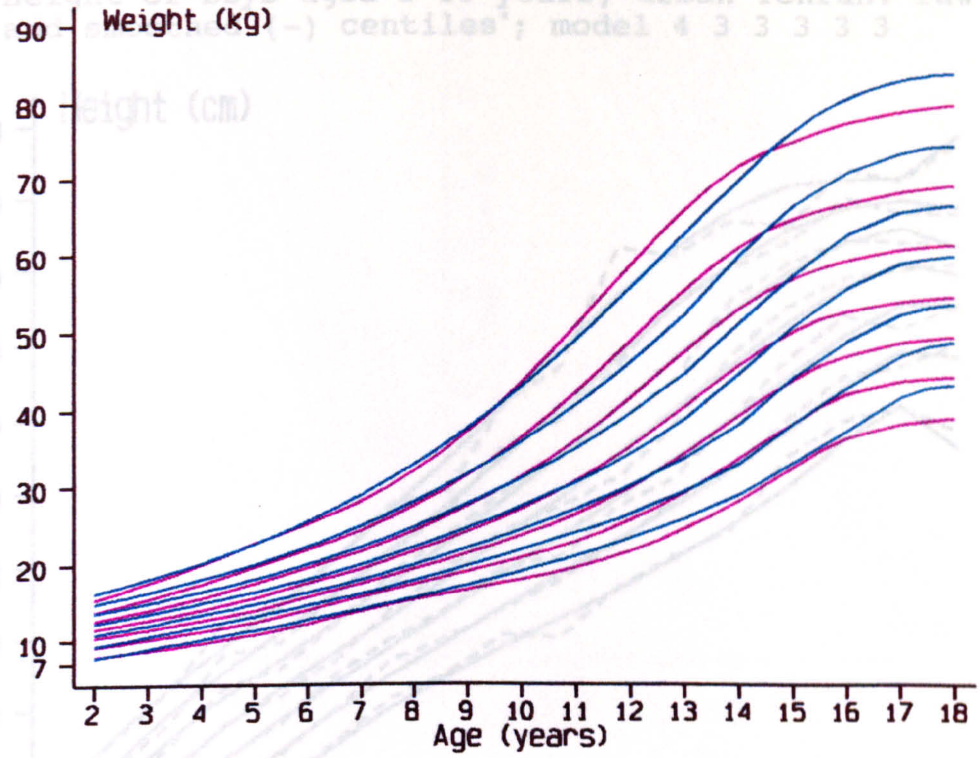
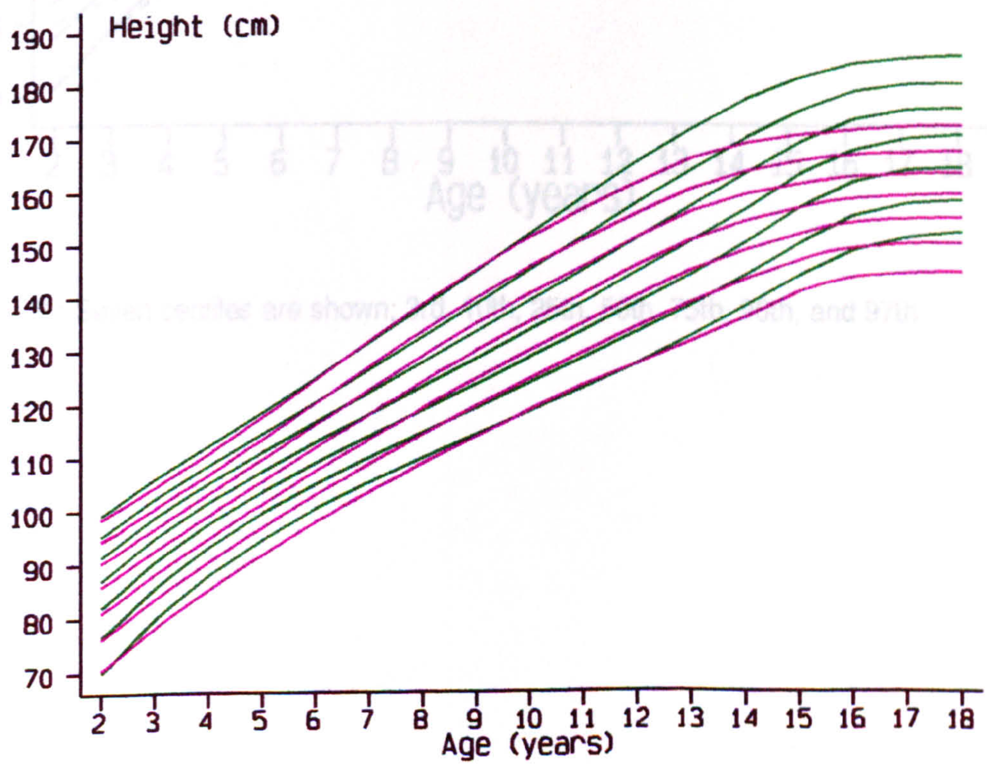


Figure 8.9

a) Comparison of weight charts of boys (cyan) and girls (pink), 2-18 years old, urban Tehran



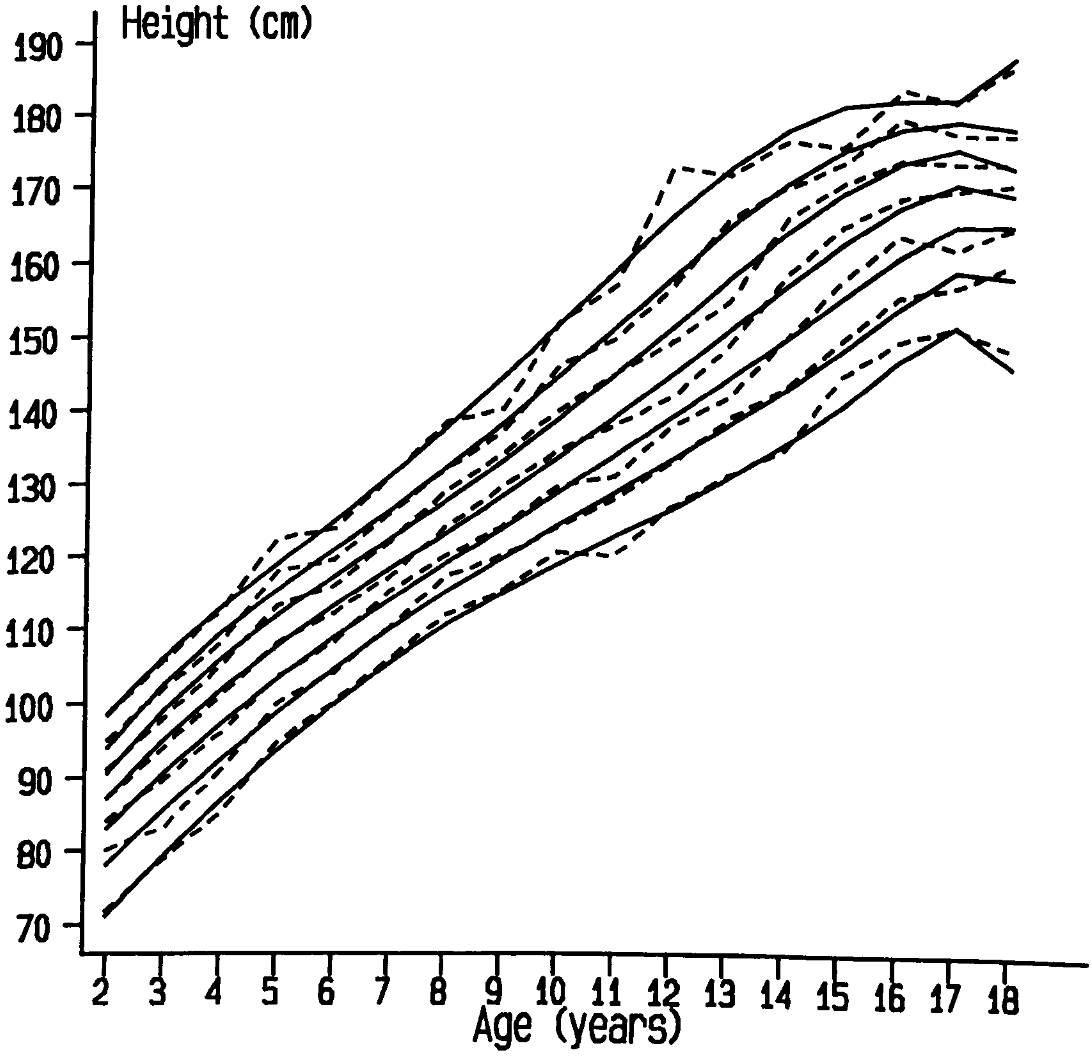
b) Comparison of height charts of boys (green) and girls (pink), 2-18 years old, urban Tehran



Seven centiles are shown; 3rd, 10th, 25th, 50th, 75th, 90th, and 97th

Figure 8.10

Height of boys aged 2-18 years, urban Tehran: raw (--) and smoothed (-) centiles*; model 4 3 3 3 3 3



* : Seven centiles are shown; 3rd, 10th, 25th, 50th, 75th, 90th, and 97th

Figure 8.11 Comparisons of growth charts of urban Tehran and raw centiles for the rest of urban areas of Iran (seven centiles are shown: 3rd, 10th, 25th, 50th, 75th, 90th, 97th)

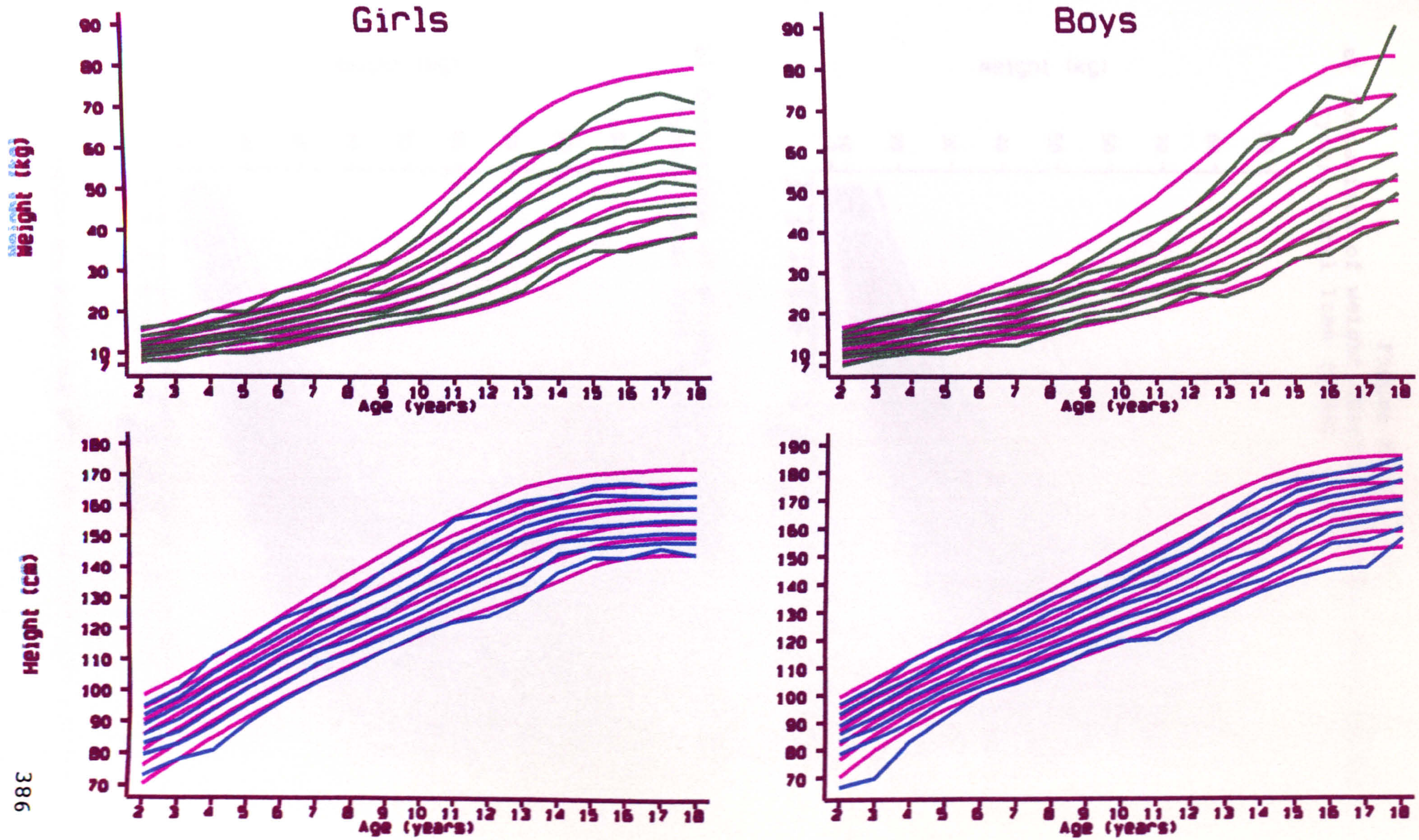
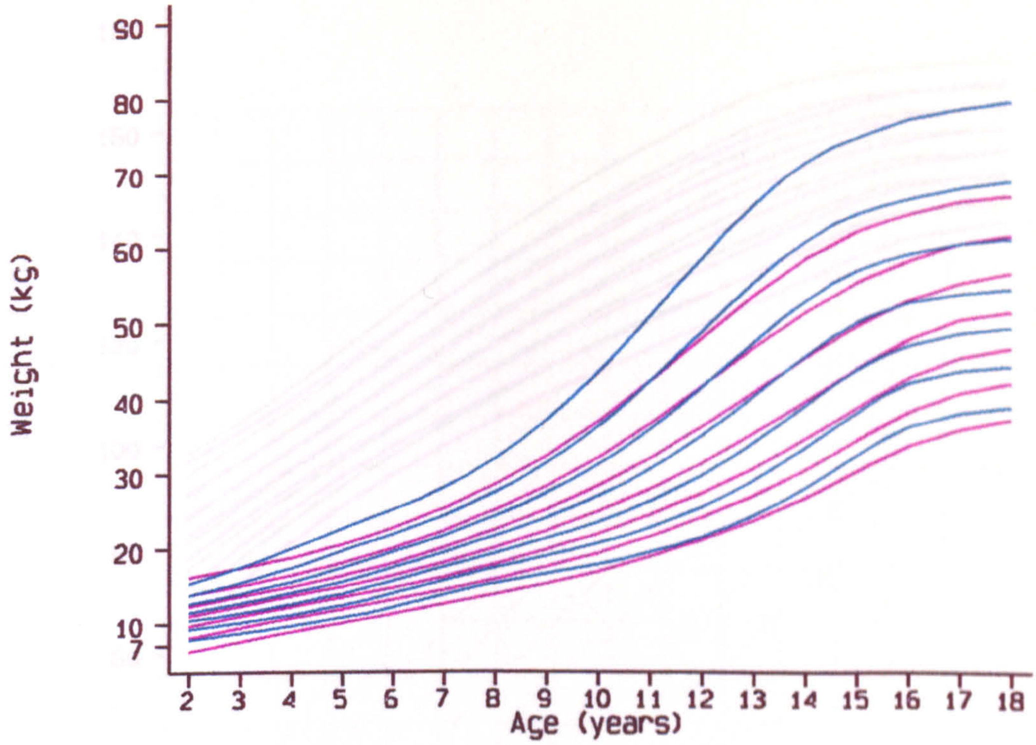
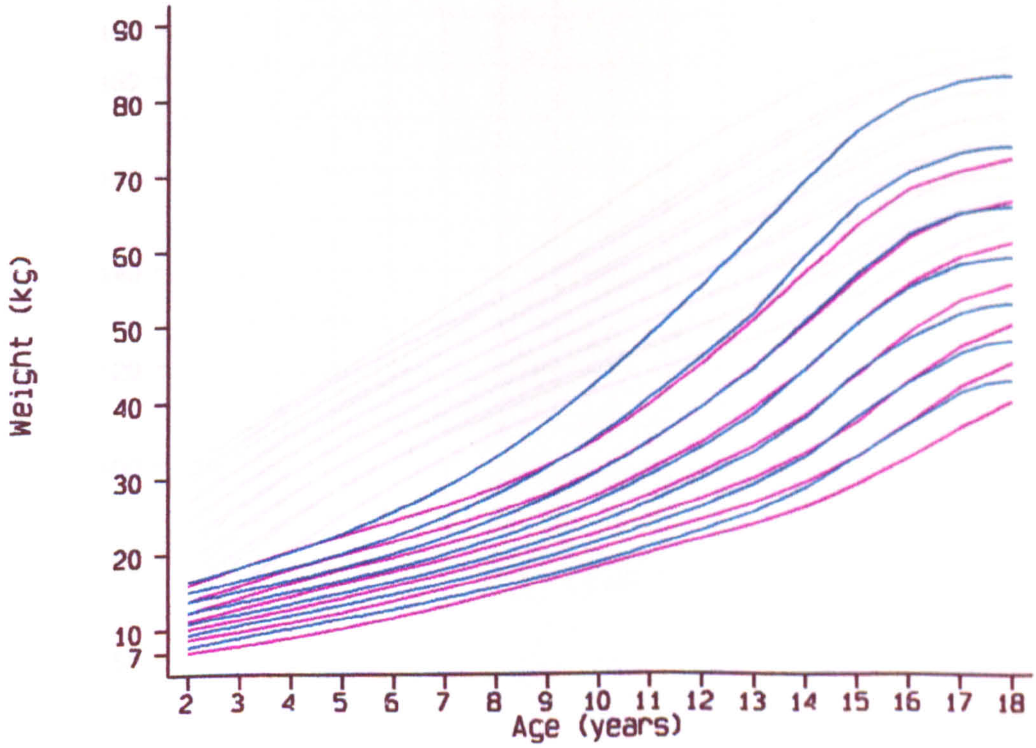


Figure 8.12

a) Comparison of weight charts of girls: urban Tehran (cyan), rural Iran (pink)



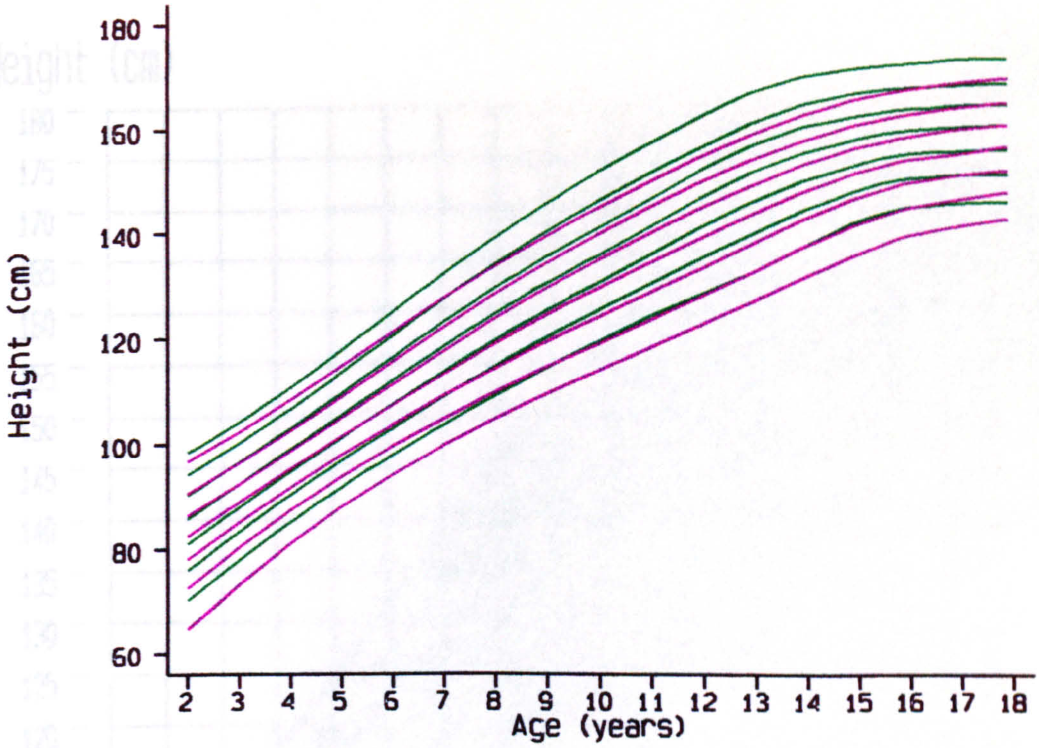
b) Comparison of weight charts of boys: urban Tehran (cyan), rural Iran (pink)



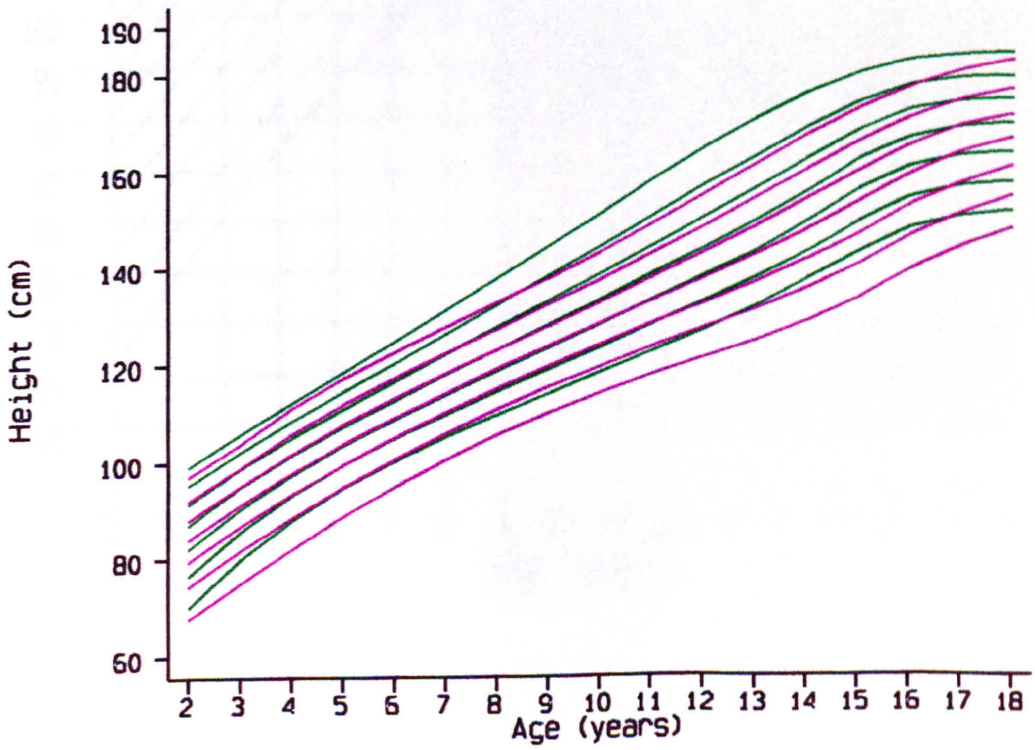
Seven centiles are shown; 3rd, 10th, 25th, 50th, 75th, 90th, and 97th

Figure 8.13

a) Comparison of height charts of girls: urban Tehran (green), rural Iran (pink)



b) Comparison of height charts of boys: urban Tehran (green), rural Iran (pink)



Seven centiles are shown; 3rd, 10th, 25th, 50th, 75th, 90th, and 97th

Figure 8.14

Iran: growth chart of girls' height aged 2-18 years old

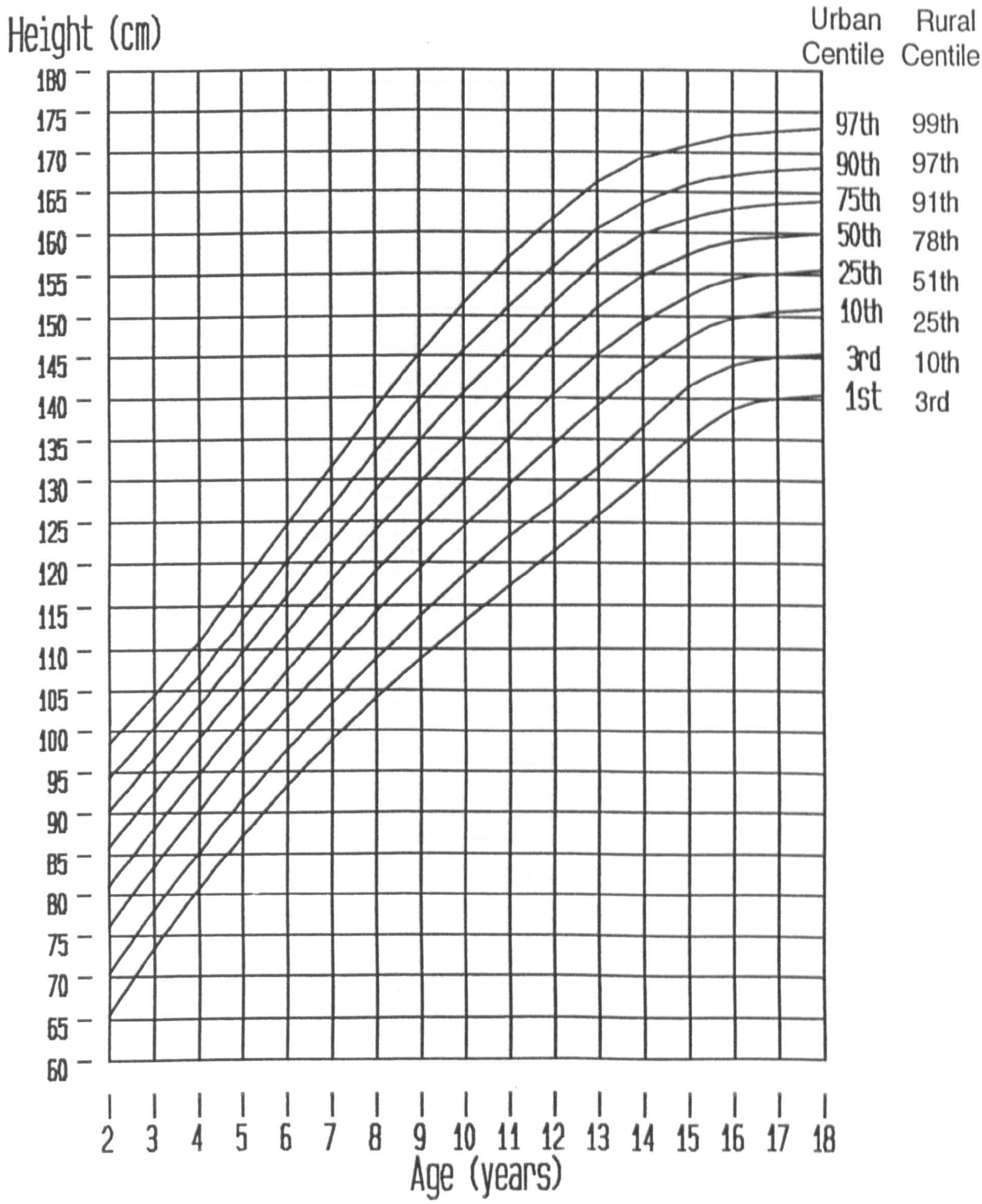


Figure 8.15

Iran: growth chart of boys' height aged 2-18 years old

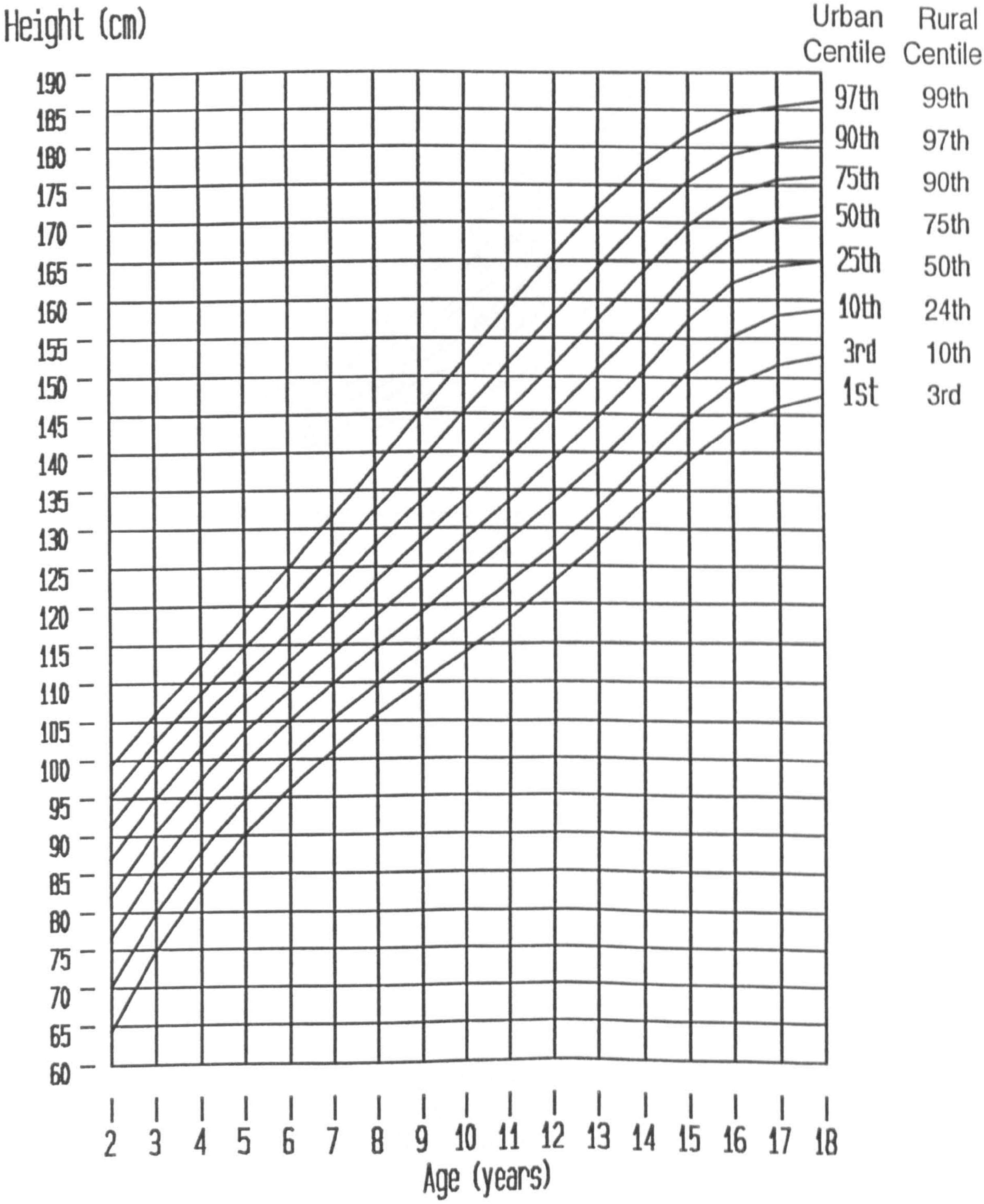
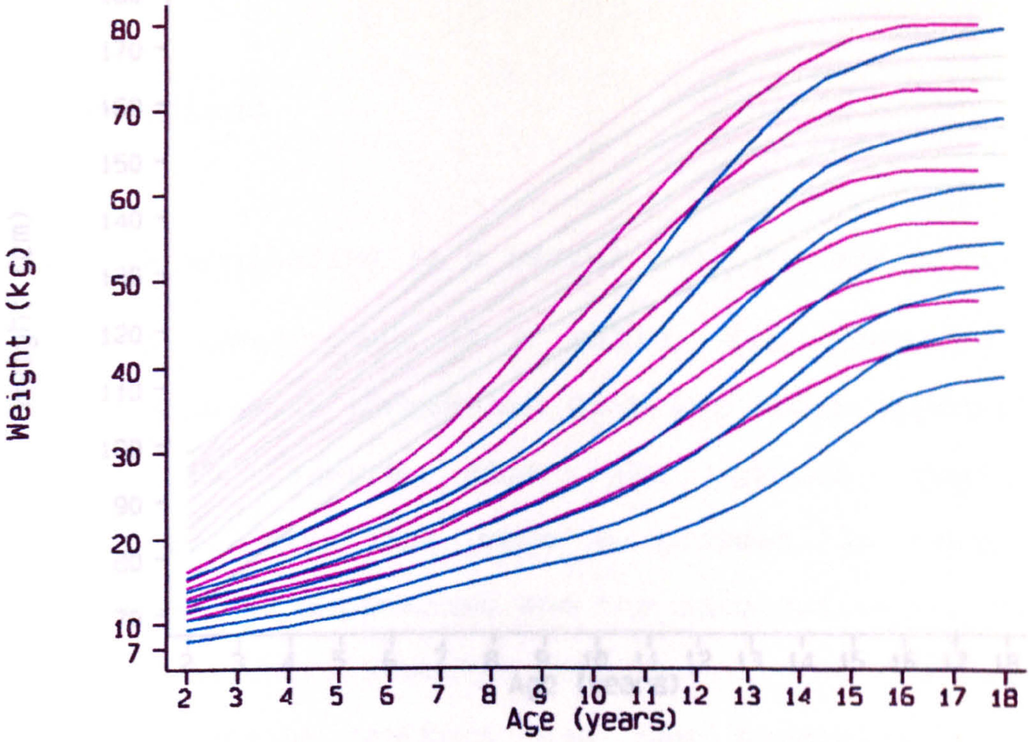
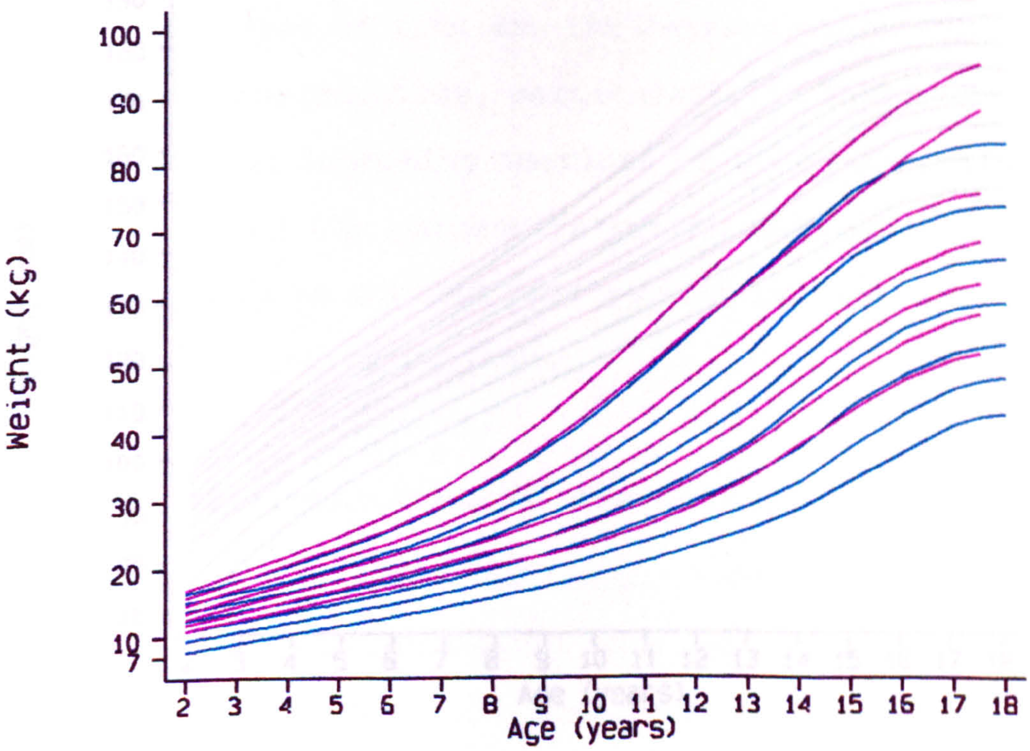


Figure 8.16

a) Comparison of weight charts of girls: Iran (cyan), NCHS (pink)



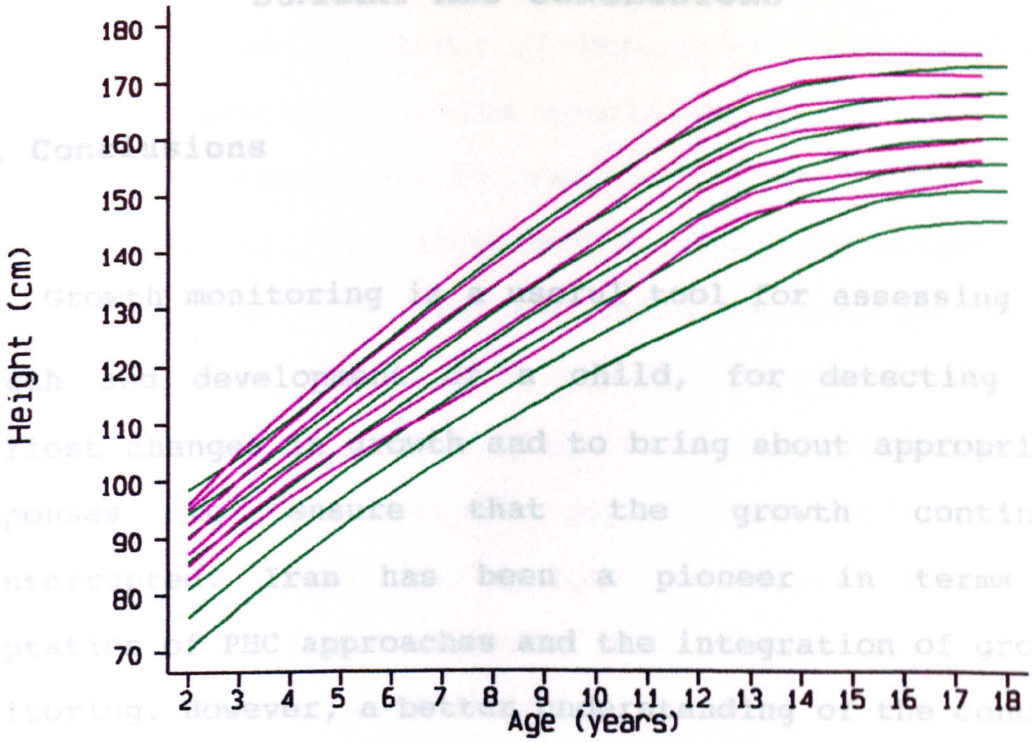
b) Comparison of weight charts of boys: Iran (cyan), NCHS (pink)



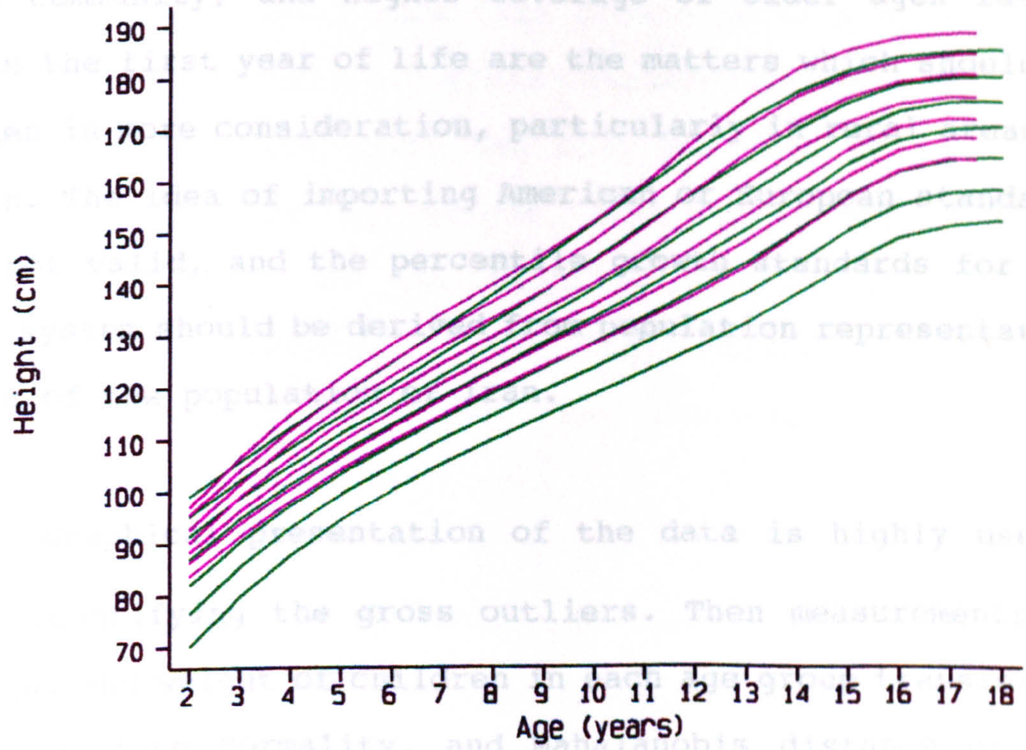
Seven centiles are shown; 3rd, 10th, 25th, 50th, 75th, 90th, and 97th

Figure 8.17

a) Comparison of height charts of girls: Iran (green) NCHS (pink)



b) Comparison of height charts of boys: Iran (green), NCHS (pink)



Seven centiles are shown; 3rd, 10th, 25th, 50th, 75th, 90th, and 97th

CHAPTER NINE

SUMMARY AND CONCLUSIONS

9.1 Conclusions

- Growth monitoring is a useful tool for assessing the growth and development of a child, for detecting the earliest changes in growth and to bring about appropriate responses to ensure that the growth continues uninterrupted. Iran has been a pioneer in terms of adaptation of PHC approaches and the integration of growth monitoring. However, a better understanding of the concept of GMP by health workers and the community, proper guidelines and educational material for health workers and the community, and higher coverage of older ages rather than the first year of life are the matters which should be taken in more consideration, particularly in rural areas of Iran. The idea of importing American or European standards is not valid, and the percentile growth standards for our PHC system should be derived from population representative data of the population of Iran.

- Graphical presentation of the data is highly useful for identifying the gross outliers. Then measurements of height and weight of children in each age group transformed to Bivariate Normality, and Mahalanobis distance of the observations from the means for age groups were computed.

Excluding observations with too large distance measures was found to be an efficient way of dealing with the problem of discordant observations. The specifity of the underlying method was observed to be very high, and none of the original records were incorrectly labelled outliers. In addition, it was shown that employing computer procedures of methods like Hadi's for identification of outlying observations which rely on Multivariate Normality or at least an assumption that the population is elliptically symmetric may lead to serious mistakes. Uncritical application of these methods can result in loss of information and waste of money.

o The results of analysis of the data from National Health Survey 1990-2 show that growth pattern of boys and girls are different, and girls' weight measurements are generally more variable than boys. The mean and rate of growth as measured by height and weight, also differ between urban and rural areas of the country. Shorter and lighter children are growing faster than taller and heavier children of the same age. And the variation between provinces diminishes with age.

o Various analysis on the results of modelling growth patterns of children across provinces of Iran did not suggest any regional grouping of provinces such as north-south, west-east or any another form of like groups which are similar in weather or urbanisation. Also, some analytical suggestions of having a one or two provinces different from the rest of the country are not practical. Also, the findings of separate cluster analyses of growth patterns of children in urban and rural areas were different. In the analysis of growth pattern urban Tehran was near the average of the models parameters. Further analysis on weight and height of children showed that urban Tehran is not significantly different from the rest of urban areas of Iran. Also the results was similar when urban Tehran was looked as a random/fixed element in variation of growth patterns between provinces of Iran. Further investigations of distribution of different centiles confirmed that the centiles of urban Tehran are within confidence intervals of urban Iran's centiles. Therefore, urban Tehran was recognized to be a reasonable baseline for the country. This also is practically very interesting for future studies of growth in Iran, because this province is where the field work training of medical students from the three most important universities of medical sciences in the country takes place. Also, the Primary Health Care system is well established and monitoring is practised regularly by Health Centres and Health Houses. So collection of data for any future cross sectional or longitudinal study will be comparatively easy.

- Using girls' weight data from urban Tehran different new approaches to growth chart construction were compared with chart constructed with GROSTAT. Using different ways of checking the models it was found that the GROSTAT model fits the data best. A new test statistic (D^*) for comparison of different fits to the centiles is introduced which measures the closeness of the model centiles to the raw centiles. The distribution of D^* is derived. D^* statistic overcomes the drawbacks of the grid test, which relies on counts, and quantifies differences between models. Also, since it is shown that in this data set centiles are not affected by the structure use of this statistics is valid.

- Growth charts of weight and height of boys and girls 2-18 years old using urban Tehran's data were constructed using Healy's method by splining of high order polynomials on age and Z. Goodness of fit analyses suggested that the models faithfully represent the data. Regarding Healy's methods, a suggestion is made as to how determine starting values of p and q when building GROSTAT models.

- Apart from the analytical finding regarding difference in growth patterns of urban and rural children, since in growth monitoring the children whose position are in the extreme centiles are of more interest, a practical solution enabling one set of charts to be used for both groups of children was proposed. Because there were negligible differences in lower centiles of weight of boys and girls between children from urban and rural areas, the derived centiles of weight of girls and boys may be used for both groups. Second, since generally urban children were taller than their rural counterparts, the 1st centiles of urban children were added to the height charts of girls and boys. These are shown to be approximately equal to the 3rd centiles of heights of rural children. Also, the correspondence of the other centiles were computed through reading the urban data into related rural models. It is suggested that the resulting charts are practically reasonable for monitoring of the children's height in urban and rural areas.

- Comparisons of our norms with NCHS data showed that our centiles are substantially below their corresponding centiles. This shows that local standards should be used for clinical work in Iran since they are realistic about the expected growth of Iranian children. Also, a comparison with a previous study in Iran showed that urban Tehran's data covers the wider range of variation which exist in growth patterns across provinces of the country. The immediate need is for a survey of growth in Iran covering the first two years to us help to complete our norms. The norms we propose will have to be updated periodically in order to take account of the changes that appear over time.

9.2 Possibilities for further works

- In addition to basic demographic data, information on accommodation (which could be used as a proxy for wealth) and occupation of parents is also available. So. identification of factors affecting the growth pattern of children in Iran is the next study that could be undertaken immediately. Lack of nation wide information in this field is a great motivation for the author to undertake this plan. Moreover, some sort of definition of social class in Iran needs to be established.

- Since the value of weight-for-height standards, as a screening tool is particularly marked in two areas of public health: protein energy malnutrition and obesity (Cole, 1979), there is a need to construct weight-for-height standards for Iranian children and adolescents taking the effect of age in to account (Cole, 1979; Ayatollahi, 1995). A nation wide study of obesity among children in Iran would then be straightforward.

- This large data set provides many opportunities for further research beyond the basic tabulations that have been published. For example, in chapter one it was stated that one section of the questionnaire of the Survey was related to the laboratory examination of individuals in the families. Serum cholesterol data is available on a sample from about 50,000 people aged 2-69 years old. Also, available are haematocrit and haemoglobin levels as well as blood pressure for those aged over 12 years old together with the information on cigarette smoking. Possibilities for further research are: a study of distributions of age and sex specific distributions (or reference ranges) of blood pressure from 12 and upwards. Investigation of associations between blood pressure and smoking. Study of relationship between obesity and blood pressure.

In addition, demonstration of age and sex distributions (reference ranges) for serum cholesterol from age 2 and upwards might be of value; also, an investigation of the association of height and weight with the distribution of cholesterol levels in children aged 2 to 18 years; an investigation of the association between cholesterol levels in married partnerships, taking into account ages of the individuals concerned. More interesting from an epidemiological point of view is the relationship between cholesterol levels in parents and children, taking into account the ages and sexes of those involved. More generally, the degree of clustering of cholesterol levels according to family, urbanization and geographical areas could be looked at via multilevel modelling. Relationships between occupation and accommodation and cholesterol levels could be examined, and each of the previous analyses could be looked at in relation to social or wealth indicators. These are just some of the possibilities for further research opened up by the unique data set of the National Health Survey of Iran.

REFERENCES

AED (Academy for Educational Development), (1988) 'Growth monitoring and promotion: issues and action.', A report of an advisory meeting sponsored by nutrition communication project of the Academy for Educational Development in Coordination with the Office of Nutrition of AID. *Washington, D.C.: AED.*

AITKIN M., and LONGFORD N.(1986) 'Statistical modelling in school effectiveness studies.', *J. Royal Statist. Soc.* Ser A, 144 (4), 419-461.

AITKIN M.(1987) 'Modelling variance heterogeneity in normal regression using GLIM.', *Appl. Statist.* 35, 332-339.

AKRAM D.S.,and AGBOATWALA A.(1991) 'Growth parameters of Pakistani children.', *Indian J. Pediatr.*, 58: 825-832.

AL-ISA A.N.,and BENER A.(1995) 'The growth pattern of Kuwaiti pre-school children compared to NCHS/CDC reference population.', *J. Roy Soc Health* June, 159-163.

ALTMAN D.G.(1993) 'Construction of age-related reference centils with absolute residuals.', *Stat in Med*, 12: 917-924.

ALTMAN D.G., and CHITTY L.S.(1994) 'Charts of fetal size: 1. Methodology.', *British J. Obstetrics and Gynaecology*, 101, 29-34.

APHA (1981) 'Growth monitoring.', Washington D. C., *American Public Health Association.*

ARMITAGE P., and BERRY G.(1988). *Statistical methods in medical research.* 2nd ed, Blackwell Scientific Publication.

ASHWORTH A., and FEACHEM R. G.(1986) 'interventions for the control of diarrhoeal disease among young children: Growth monitoring programmes.', *Unpublished document.*

AYATOLLAHI S.M.T., and CARPENTER R.G.(1991) 'Growth of school children of southern Iran in relation to NCHS standard.', *Annals of Human Biology*, 18 (6), 515-522.

AYATOLLAHI S.M.T.(1993a) 'A critical appraisal of human growth studies in Iran.' *M. J. I.R.Iran*, Vol. 6, No. 4, 311-319.

AYATOLLAHI S.M.T., COLE M., and MATTEWS J.N.S.(1993b) 'Comparison and amalgamation of two new methods for estimating age-related centiles.', *Submitted for Publication*.

AYATOLLAHI S.M.T.(1995) 'Age standardization of weight-for-height in children using a unified Z-score method.', *Annals of Human Biology*, Vol 22, 2, 151-162.

BALDWIN L.M., and SUTHERLAND S.(1988) 'Growth pattern of first-generation Southeast Asian infants.', *American Journal of Disease of Children*, 142, 526-531.

BARNETT V., and LEWIS T.(1994). *Outliers in statistical data*. 3rd ed, Wiley, New York.

BELL M.(1989) 'What to do when a child fails to thrive.', *MIMS*, 24, 1 63-68.

BESSEL F.W., and BAEUER J.J.(1838) 'Gradmessung in ostpreussen und ihre verbindung mit preussischen und russischen dreiecksketten.', Berlin, Reprinted in *Abhendlungen von*, F. W. Bassel, ed. Engleman R., Leipzing, 1876.

BERNOULLI D.(1777) 'Dijudicatio maxime probabilis plurium observationum discrepantium atque verisimillima inducatio inde formanda.', Acta academiae scientiorum petropolitanae, 1, 3-33. English translation by ALLEN C.G. (1961), *Biometrika*, 48. 3-13.

BHARGAVA S.K., KUMARI S., CHOUDHURY P. and LALL U.B.(1980) 'A longitudinal study of physical growth from birth to six year in children with birth weight of 2501 g. or more.', *Indian Paediatrics*, 33, 495-502.

BOSCOVICH R.J.(1757) 'De litteraria expeditione per pontificiam ditionem, et synopsis amplioris operis, ac habentur plura ejus ex exemplaria etiam sensorum impressa.', *Bononiensi Scientiarum et Artum Instuto Atque Academia Commentarii*, 4, 353-396.

BOX G. E. P., and TIDWELL P. W.(1962) 'Transformation of the independent variables.', *Technometrics*, 4, 531-550.

BOX G. E. P., and WATSON G. S.(1962) 'Robustness to non-Normality of regression tests.', *Biometrika*, 49, 93-106.

BOX G.E.P., AND COX D.R.(1964) 'An analysis of transformation.', *Journal of Royal Statistical Society*, series B, 26, 211-252.

BREIMAN L., and FRIEDMAN J.H.(1985) 'Estimating optimal transformations for multiple regression and correlation (with discussion).', *J. Am. Statist. Ass.*, 580-619.

BROOKE O.G., and WOOD C.(1980) 'Growth in British Asians: longitudinal data in the first year.' *Journal Of Human Nutrition*, 34, 355-359.

BRYK A.S, and RAUDENBUSH S.W.(1992). *Hierarchial linear models*. Newbury park: Sage Publications.

BUTEE N.F., GARZA C., SMITH E.O., and NICHOLS B.L.(1984) 'Human milk intake and growth performance of exclusively breast-fed infants.', Journal of pediatrics, 104 (2), 187-195.

CAMERON N.(1986) 'Standards for human growth - their construction and use.', *South African Medical Journal*, 70, 422-425.

CAMPBELL N.A.(1980) 'Robust procedures in multivariate analysis I: Robust covariance estimation.', *Applied Statistics*, 29, 231-237.

CARPENTER R.G.(1994). Personal communication.

CARPENTER R.G.(1996). Personal communication.

CHAUVENET W.(1863) 'Method of least square.', Appendix to *Manual of Spherical and practical astronomy*, Vol 2, Lippincott, Pheladelphia, pp. 469-566; tables 593-599. Reprinted (1960) 5th edn. Dover, New York.

CHEN S.T.(1985) 'Growth chart based on longitudinal study of Malaysian children from birth to six years.' *The Medical Journal Of Malaya*, 40 (2), 80-89.

CLEVELAND W.S.(1979) 'Robust locally weighted regression and smoothing scatter plots.', *Journal of the American Statistical Association*, 79: 829-836.

COLE T.J.(1979) 'A method for assessing age-standardized weight-for-height in children seen cross sectionally.', *Annals of Human Biology*, 13, 5: 433-451

COLE T.J.(1988a) 'Fitting smoothed centile curves to reference data.', *J. R. Statist. Soc. A*, 151, 3: 385-418.

COLE T.J.(1988b) 'Summarizing growth standards by LMS method.', *Proc. Nutr. Soc.*, 47:61A.

COLE T.J., and GREEN P.L.(1992) 'Smoothing reference centile curves: The LMS method and penalized likelihood.', *Stat. in Med.*, Vol 11, 1305-1319.

COLE T.J.(1994) 'Discussion of ROYSTON, and ALTMAN'S paper on fractional polynomials.', *Applied Statistics*, 43 (3), 459-460.

COLE T.J., FREEMAN, J.V., PREECE M.A.(1995) 'Cross sectional stature and weight reference curves for the UK, 1990.', *Archives of Disease in Childhood*, 73, 17-24.

CUNNINGHAM N.(1978) 'The under-five clinic - What difference does it make?, *J. Trop. Paediat.*, (Dec).

DAVIES D.P.(1988) 'The importance of genetic influence on growth in early childhood with particular reference to children of Asiatic origin.' In *Linear growth retardation in less developed countries*, edited by WATERLOW J.C., Vevey: New York, Nestle Foundation Publication Press.

DAVIES D.P., and YAMAMURO Y.(1985) 'Growth of Hong Kong infants during the first two years of life.', *Early Human Development*, 11, 247-254.

DEMPSTER A.P., RUBIN D.B., and TSUTAKAWA R.K.(1981) 'Estimation in covariance components models.', *J. Royal Statist. Soc.*, Ser B, 39, 1-8.

DEVLIN S.J., GNANADESIKAN R., and KETTENRING J.R.(1981) 'Robust estimation of dispersion matrices and principal components.', *J. Amer. Statist. Assn*, 76, 354-362.

DEWEY K.G., HEINIG M.J., NOMMSEN L. A. PEERSON L.A., PEERSON J.M., and ZONNERDAL B.(1992) 'Growth of breast-fed and formula-fed infants from 0-18 months: The Darling Study.', *Pediatrics* 89: 1035-41.

DIBLEY M.J., STAEHLING N.W., NIEBURG P., and TROWBRIDGE, F.L.(1987) 'Interpretation of Z-score anthropometric indicators derived from the international growth reference.', *American Journal of Clinical Nutrition*, 46 (5): 749-762.

DONOHO D.L.(1982). *Breakdown properties of multivariate location estimators*. PhD dissertation, Harvard University, Boston.

ELSTON R.C., and GRIZZLE J.E.(1962) 'Estimation of time response curves and their confidence bands.', *Biometrics*, 18, 148-159.

EPI INFO 6.03 (1996). 'A word processing, Database, and statistical program for public Health', WHO and CDC.

EPI INFO 6 Manual (1994). 'A word processing, Database, and statistical system for Epidemiology and Microcomputers, pp 267-289.

EUSEBIO J.S., and NUBE M.(1981) 'Attainable growth'. *Lancet*, II, 1223.

EVELETH P.B., 1978. 'Population difference in growth: Environmental and genetic factors.' In *Human Growth* Vol 3: edited by Falkner F., and Tanner J.M., New York: Plenum Press. pp 372- 393.

EVELETH P.B., and TANNER J.M.(1990). *Worldwide variation in human growth*. Cambridge University Press.

FAO/UNICEF/WHO, Joint Expert Committee on Nutrition, 1976. 'Methodology of nutrition surveillance.', Geneva: *World Health Organization*, technical reports, No. 593.

FERGUSON T.S.(1961) 'On the rejection of outliers.', *Proceedings of the fourth Berekely symposium on mathematical statistics and probability*, Vol 1, pp 253-287.

FISHER R. A., and YATES F.(1948). *Statistical tables for* *biological, agricultural, and medical research*. Oliver and Boyed, London.

GEREIN N. M.(1988) 'Is growth monitoring worthwhile? Review article.', *Health Policy and Planning*, 3 (3): 181-194.

GLIM (1992). *The Generalised Linear Iterative Modelling System. Release 4 UK: Royal Statistical Society.*

GNANADESIKAN R., and KETTERNING J.R.(1972) 'Robust estimates, residuals and outlier detection with multiresponse data.', *Biometrics*, 28, 81-124.

GOLDFRAD C.(1994). *Early doppler measurements of the utrine and umbilical blood flow for the prediction of : Pre-eclampsia, being small for gestational age, premature delivery and abruption*. London School of Hygiene and Tropical Medicine, Med. Stat. Msc Thesis.

GOLDSTEIN H.(1972) 'The construction of standards for measurements subject to growth.', *Human Biology*, 44 (3), 255-261.

GOLDSTEIN H., and TANNER J.M.(1980) 'Ecological considerations in the creation and use of child's growth standards.', *Lancet*, I, 582-585.

GOLDSTEIN H.(1986) 'Efficient statistical modelling of longitudinal data.', *Annals of Human Biology*, 13, 129-141.

GOLDSTEIN H.(1987). *Multilevel mixed models in educational and social research*. London: Griffin.

GOLDSTEIN H.(1989) 'Restricted unbiased iterative generalized least square estimation.', *Biometrika*, 76, 622-623.

GOLDSTEIN H., HEALY M.J., and RASBASH J.(1994) 'Time series models for repeated measures data.', *Stat. in Med.*, 30 (13): 1643-55.

GOLDSTEIN H.(1995). *Multilevel statistical models*. London, Edward Arnold: New York, Halsted Press.

GOPALAN C., and CHATTERJEE M.(1985). *Use of growth chart for promoting child nutrition: A review of global* *experiences*. Nutrition Foundation of India, Special Publication Series 2.

GORSTEIN J., and AKRE' J.(1988) 'The use of anthropometry to assess nutritional status.', *World Health Statistics Quarterly*, Vol 41, No 2, 48-60.

GORSTEIN J.(1989) 'Assessment of nutritional status: Effects of different methods to determine age on the classification of undernutrition.', *Bull of the World Health Organization*, 67: 143-150.

GORSTEIN J. et al.(1994) 'Issues in the assessment of nutritional status using anthropometry.', *Bull of the World Health Organization*, 72: 273-384.

GOWER J.(1991). Personal communication.

GSO (1993). *Iran today.*, 1st ed., Tehran: Gita Shenasi Organization, Iran.

GOULD W.W., and HADI A.S.(1993) 'Smv6: Identifying multivariate outliers.', *Stata Technical Bulletin*, 11: 28-32. Reprinted in Stata Bulletin Reprints, Vol 2, pp 163-168.

GRANT J.P.(1987). *The state of world's children 1987*. New York: UNICEF.

GRIFFITHS M.(1988) 'Growth monitoring: making it a tool for education.', *Indian Journal of Pediatrics*, 55 (1, suppl.): s59-s66.

GREEN P.L.(1988) 'Discussion of Cole's paper (1988a) on fitting smoothed centile curves to reference data.', *J. R. Statist. Soc. A*, 151, 3: 407-418.

GRUBBS F.E., 1950. 'Samples criteria for testing outlying observations.', *Ann. Math. Statist.*, 21, 27-58

GUO S., ROCHE A.F., and YEUNG D.L.(1990) 'Monthly growth status from a longitudinal study of Canadian infants.', *Canadian Journal Of Public Health*, 81, 215-221.

GRAITCER P.L., and GENTRY M.(1981) 'Measuring children: one reference for all.', *Lancet*, 2: 279-299.

HABICHT J.L., MARTORELL R., YARBROHGH C., MALINA R.M., AND KLEIN R.E.(1974) 'Height and weight standards for the pre-school children.', *Lancet*, I, 611.

HADI A.S.(1992) 'Identifying multiple outliers in multivariate data.', *J.R. Statist. Soc.*, B, 54, 761-771.

HADI A.S.(1994) 'A modification of a method for the detection of outliers in multivariate sample.', *J.R. Statist. Soc.*, B, 56, 393-396.

HALL D.M.B.(1989). *Health for all children- A programme for child health surveillance*. Oxford: British Paediatric Association.

HAMILL P.V.V., DRIZD T.A., JOHNSON C.L., REED R.B., and ROCHE A.F., and MOORE W. M.(1979) 'Physical growth: National Centre for Health Statistics.', *American Journal of Clinical Nutrition*, 32 (3): 607-629.

HAMPEL F.R., RONCHETTI E.M., ROUSSEEUW P.J., STAHEL W.A. (1986). *Robust statistics: The approach based on influence functions*. New York, Wiley.

HARRIS E.K., and BOYD J.C.(1995). *Statistical bases of reference values in laboratory medicine*, Marcel Dekker, Inc, New York.

HASTIE T.J., and TIBSHIRANI R.J.(1990). *Generalized additive models*. London: Chapman and Hall, p.1.

HAUSPIE R.(1980) 'Abnormal growth.' In *Human physical growth and maturation, Methodologies and factors*, edited by JOHNSON F.E., ROCHE A.F. and SUSANNE C., New York: Plenum Press, pp 193-201.

HEALY M.J.R.(1962) 'The effect of age-grouping on the distribution of measurements effected by growth.', *American of Physical Anthropology*, 20, 49-50.

HEALY M.J.R.(1968) 'Multivariate normal Plotting.', *Applied Statistics*, 17, 157-161.

HEALY M.J.R.(1986) 'Statistics of growth standards.' In *Human Growth*, Vol. 3: edited by Falkner F., and Tanner J.M., New York: Plenum Press, pp 47-57.

HEALY M.J.R., RASBASH J., and YANG M.(1988) 'Distribution-free estimation of age-related centiles.', *Annals of Hum. Biol.* 15, 1:17-22.

HEALY M.J.R.(1989) 'Growth curves and growth standards- The state of art.', *Auxology 88. Perspectives in the sciences of growth and development*, (ed.) TANNER, J.M., London: SMITH-GORDON.

HEALY M.J.R.(1991). Personal communication.

HEALY M.J.R.(1994) 'Statistics from the inside: Non-normal data.', *Arch. Dis. Child*, 70 (2), 158-163.

HENDRATA L.(1987) 'Growth monitoring and promotion.' guideline for programming. Unpublished.

HOSSEINI M.(1988). *Determination of child mortality rate (q_{0-5}) in Iran; 1973-88*. MSc thesis, Tehran University of Medical sciences, Iran (Farsi).

ICMR (1972). *Growth and physical development of Indian infants and children*. New Dehli, Indian Council of Medical Research.

IRI (1990). *Education in the Islamic Republic of Iran: now and in future*. Tehran, Ministry of Education.

IRWIN J.O.(1925) 'On a criterion for the rejection of outlying observations.', *Biometrika*, 17, 238-250.

ISAACS D., ALTMAN D.G., TIDMARSH, C.E., VALMAN, H.B., AND WEBSTER, A.D.B.(1983) 'Serum immunoglobulin concentration in preschool children measured by laser nephelometry: reference range for IgG, IgA, IgM.', *Journal of Clinical Pathology*, 36, 1193-1196.

JENNINGS L.W., and YOUNG D.M.(1988) 'Extended critical values of the multivariate extreme deviate test for detecting a single spurious observation.', *Commun. Statist. Sim. Comp.*, 17, 1359-73.

JOHNSON N.L., and KOTZ S.(1970). *Continuous univariate distribution*. Vol 1, Wiley, New York, p 81.

JOHNSON R.A., and WICHERN D.W.(1988). *Applied multivariate* *statistical analysis*, 2nd ed. Prentice-Hall, Englewood Cliff, NJ.

KELLER W.(1983) 'Choice of indicators of nutritional status'. In *Evaluation of nutrition education in third world communities*. Nestle Foundation Publication Series. Bern Hans Huber Publishers.

KEMPTHORNE P.J., and MENDEL M.B.(1990) 'Comments on Rousseeuw & Van Zomeren'.

KENDALL M.G.(1952). *The advanced theory of statistics*. Vol 1, 5th ed, London: Charles Griffin, ch 9.

KENDALL M.G., and STUART A.(1958). *The advanced theory of statistics*. Vol 1, London: Charles Griffin.

LAIRD N.M., and WARE H.(1982) 'Random-effects models for longitudinal data.', *Biometrics*, 38, 963-974.

LEUNG S.S.F, and DAVIES D.P.(1989) 'Anthropometric assessment of nutritional status: a need for caution.' In *Auxology 88 perspective in the science of growth and development*, edited by Tanner J. M, London: Gordon-Smith.

LIN W.S., CHEN A.C. SU J.Z., ZHU F.C. XING W.H., and YE G.S.(1992) 'The menarcheal age of Chinese girls.' Ann Human Biol. Sep-Oct; 19(5): 503-512.

LINDLEY D.V., and SMITH A.F.M.(1972) 'Bayes estimation for the linear model.', *Journal of Statistical Association*, 87, 1227-37.

LONGFORD N.T.(1987) 'A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects.', *Biometrika*, 74 (4), 817-827.

LONGFORD N.T.(1993). *Random coefficient models*. Oxford University Press.

LOPUHAÄ H.P., and ROUSSEEUW P.J.(1991) 'Breakdown points of affine equivariant estimators of multivariate location and covariance matrices.', *Ann. Statist.*, 19, 229-248.

LOW W.D.(1970) 'The cross-sectional, longitudinal and mixed longitudinal methods in the study of human growth.', *Z.*

Morph. Anthrop., 62 (3), 249-258.

MATHAI A.M, PROVOST, B.S.(1992). *Quadratic forms in random variables, theory and applications*. Marcel Dekker Inc, New York.

MANLY B.F.J.(1994). *Multivariate statistical analysis*, A primer, 2nd ed. London Chapman & Hall.

MARSHALL W.A.(1977). *Human growth and disorders*. London Academic Press.

MARUBINI E.(1980) 'General statistical consideration in analysis.' In *Human Physical growth and maturation, Methodologies and Factors*, edited by Johnston F. E., Roche A. F., and Susanne C., New York: Plenum Press, pp 101-117.

MARUBINI E., and MILANI S.(1986) 'Approaches to the analysis of longitudinal data.' In *Human Growth*, Vol 3: edited by Falkner F., and Tanner J.M., New York: Plenum Press, pp 79-92.

MATER A.M., ALEKRI, S.A., and MAHDI A.E.(1990) 'Growth pattern of adolescents in Bahrain.', *J.R.S.H*, 5, 179-180.

MEREDITH H.V.(1971) 'Human head circumferences from birth to early adulthood: racial, regional, and sex comparisons.', *Growth*, 35, 233-251.

MOH&ME (1991). *Education in the Islamic Republic of Iran: Now and in the future*. Ministry Of Health & Medical Education, Tehran, Iran.

MOH&ME (1995). *Workplan For Achieving the 1995 Global Mid-Decade Goals For Children*. Ministry Of Health & Medical Education, Tehran, Iran.

MINITAB (1991). *Minitab Statistical Software: Release 8.2, State College*.

MLn (1995). *Software for multi-level analysis, V 1.0a*. Multi-level models project, Institute of Education, University of London.

MOLTENO C.D., HOLLINGHEAD J., MOODIE A.D., BRADSHAW D., WILLOUGHBY W., BOWIE M.D., and SMALLMAN L.A.(1991) 'Growth of preschool coloured children in Cape Town. *South African Medical Journal*, 79, 670-676.

MORLEY D.(1973). *Pediatric priorities in the developing world*. London: Butterworth.

MORLEY D.(1977) 'Growth charts: curative or preventive?', *Arch. Dis. Child*, 52: 395-398.

MORLEY D., and WOODLAND M.(1979). *See how they grow - Monitoring child growth for appropriate health care in developing countries*. London: The Macmilan Press Ltd.

NABARRO D., and MCNAB S.(1980) 'A simple new technique for identifying thin children.', *J. Trop. Med. Hyg.*, 83, 21-23.

NABARRO D., and CHINNOCK P.(1988) 'Growth monitoring-Inappropriate of an appropriate technology.', *Social Science and medicine*, 26 (9): 941-948.

NCHS (1977). 'NCHS Growth curves for children birth 2-18 years, United States.', *US National Centre for Health Statistics*, Series 11, No. 165.

NCHS (1978). 'A growth chart for international use in maternal and child health care'; guidelines for Primary Health Care personnel, *WHO*.

NEVILLE M.C., and OLIVA-RASBACH, J.(1989) 'Is maternal milk production limiting for infant growth during the first year of life in breast-fed infants?' In *Human Lactation 3: the effects of human milk on the recipient infant*, ed Goldman A.S., Atkinson A.S., and Hanson L.A., New York: Plenum Press.

NEWELL K.W.(1975). *Health by the people*. Geneva, World Health Organization (WHO).

PAN H.Q., GOLDSTEIN H., and YANG Q.(1990) 'Non-parametric estimation of age-related centiles over wide age ranges.', *Annals of Human Biology*, 17, 6:475-481.

PAN H.Q., GOLDSTEIN H., and GUO D.(1992) 'A two-level cross-scetional model using grafted polynomials.', *Annals of Human Biology*, 19, 337-346.

PATNAIK'S P.B.(1949) 'The non-central and F-distributions and their applications.', *Biometrika*, 36, 202-232.

PEARSON E.S., and CHANDRA SEKAR, C.(1936) 'The efficiency of statistical tools and a criterion for the rejection of outlying observations.', *Biometrika*, 28, 308-320.

PEARSON R.(1993) 'Thematic evaluation of UNICEF support to growth monitoring.', New York: *UNICEF, Evaluation and Research Office*.

PETROS-BARVAZIAN A.(1990). Foreword of EVELETH P.B., and TANNER J.M. *Worldwide variation in human growth*. Cambridge University Press.

PLEWIS I.(1985). *Analysing change*. Chichester: Wiley.

PEIRCE B.(1852) 'Criterion for the rejection of doubtful observations.', *Astr. J.*, 2, 161-163.

PENNY K.I.(1995). *Multivariate outlier detection in laboratory safety data*. PhD dissertation, University of Aberden.

PROSSER R., RASBASH J., and GOLDSTEIN H.(1991). *ML3 Software for three-level analysis. User guide for V.2*, Institute of Education, University of London.

RASBASH J., PAN H., GOLDSTEIN H.(1993). *GROSTAT II: A program for estimating age related centiles using piecewise polynomials*. WHO collaborating centre on growth and development, and Institute of Education and Child Health, University of London.

RAUDENBUSH, S.W.(1994) 'Equivalence of Fisher scoring to iterative generalized least squares in the normal case with application to hierarchical models.', Unpublished.

ROEDE M.J., and VAN WIERINGEN J.C.(1985) 'Growth diagrams 1985: Netherlands third nation-wide survey.', *Tijdsch. Soc. Gezhdszq*, 63, Suppl, 1-34.

ROHDE J.E.(1988) 'Beyound survival, promoting healthy growth.', *Indian Journal of Pediatrics*, 55 (1 suppl.): s3-s8.

ROHDE J.E., and NORTHRUP R.S.(1988) 'Feeding, feedback and sustenance of Primary Health Care.', *Indian Journal of Pediatrics*, 55 (1 suppl.): s110-s123.

ROSENBERG B.(1973) 'Linear regression with randomly dispersed parameters.', *Biometrika*, 60, 61-75.

ROUSSEEUW P.J.(1985) 'Multivariate estimation with high breakdown point.' In *Math. Statist. Applic.*, V B, ed, GROSSMANN W., PFLUG G., VINCZE I., and WERTZ W., Riedel Publishing, pp. 283-297.

ROUSSEEUW P.J., and LEROY A.M.(1987). *Robust regression and outlier detection*. Wiley, New York, p 57.

ROUSSEEUW P.J., and VAN ZOMEREN B.C.(1990) 'Unmasking multivariate outliers and leverage points.', *J. Amer. Statist. Assoc.*, 85, 633-639.

ROYSTON P.(1991) 'Constructing time-specific reference ranges.', *Stat. in Med.*, 10, 675-690.

ROYSTON P.(1992) 'Approximating the Shapiro-Wilk W-test for non-normality.', *Statistics and Computing 2*.

ROYSTON P., and ALTMAN D.G.(1994) 'Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling (with discussion).', *Applied Statistics*, 43 (3), 429-467.

ROYSTON P., and THOMPSON S.(1995) 'Comparing non-nested regression models.', *Biometrics* 51, 114-127.

ROYSTON P.(1995) 'Assessing goodness of fit of models for reference centile estimation.', Unpublished.

ROYSTON P., and WRIGHT E.(1996) 'Statistical methods for the construction of age-related reference intervals: A review.', Unpublished.

RUEL M.T., PELLETIER D.L., HABICHT J.P., MASON J.B., CHOBOKOANE C.S.(1990) 'Comparison of Mothers' understanding of two child growth charts in Lesotho.', Bulletin of *World Heath Organization*, 68 (4): 483-491.

RUEL M.T.(1995) 'Growth monitoring as an educational tool, an integrating strategy, and a source of information: a review of experience.' In *Child Growth and Nutrition in Developing Countries*, edited by Pinstrup-Andersen P., Pelletier D., and Alderman H., Ithaca and London: Cornell University Press, pp 78-96.

SABER G.A.F., and WILD C.J.(1989). *Nonlinear regression*. Wiley, New York, pp 433-459.

SCI (1991). *Iran statistical yearbook 1373.*, [March 1994-March 1995]; Tehran: Statistical Centre of Iran.

SECKLER D.(1982) ''Small but healthy': A basic hypothesis in the theory, measuremens, and policy of malnutrition.' In *Newer concepts in nutrition and their implications for policy*, ed. SUKHATME P. V. Pune, India: Maharashtra association for the cultivation of science research institute.

SEONE N., and LATHAM M.C.(1971) 'Nutritional anthropometry in the identification of malnutrition in childhood.', *J. Top. Pediat.*, 17: 98.

SETH V., SUNDARAM K., and GUPTA M.(1979) 'Growth references standards for developing countries.', *J. Top. Pad. and Env. Child health*, 25-37.

SHADPOUR K.(1993). *Shabake-ye moraghebatha-ye avaleyeh-ye behdashti darmani dar Iran (The PHC experience in Iran)*. MOH&EM, and UNICEF, Tehran.

SHADPOUR K.(1994). *The PHC experience in Iran*. UNICEF, Tehran.

SILVERMAN B.W.(1985) 'Some aspects of the spline smoothing approach to non-parametric regression curve fitting.', *J. R. Statist. Soc.* B, 47, 1-52.

SIOTANI M.(1959) 'The extreme value of the generalized distances of the individual points in the multivariate normal sample.', *Ann. Inst. Statist. Math.* Tokyo, 10, 183-208.

SPSS/PC^{+} (1990). *Statistics guide manual*. Marija J. Norusis/SPSS Inc, Chicago.

STATACORP (1995). *Stata Statistical Software: Release 4.0, College Station, TX: Stata Corporation.*

STATA REFERENCE MANUAL 3.1 (1993). 'Shapiro-Wilks and Shapiro-Francia test for Normality'. Vol 3, pp 189-191.

STONE E.J.(1868) 'On the rejection of discordant observations.', *Monthly Notices Roy. Astr. Soc.*, 34, 9-15.

STUART H.C., and STEVENSON S.S.(1959) 'Physical growth and development.' In Nelson W. ed. *Textbook of Pediatrics*, Philadelphia, Sanders, 7th ed., p. 12-61.

STUDENT (1927) 'Errors of routine analysis.', *Biometrika*, 19, 151-164.

TANNER J.M., WHITEHOUSE R.H., and TAKAISHI M.(1966) 'Standardisation from birth to maturity for height, weight, height velocity, and weight velocity: British Children, 1965.', *Arch. Dis. Child*, 41, 454.

TANNER J.M.(1981). *A concise History of the study of human growth*. Cambridge University Press.

TANNER J.M. (1986) 'Use and abuse of growth standards.' In *Human Growth* Vol 3: edited by Falkner, F. and Tanner, J.M., New York: Plenum Press. pp 95-108.

TERADA H., and HOSI H.(1965) 'Longitudinal study on the physical growth in Japanese. (3) growth in chest and head circumferences during the first three years of life.', *Acta Arab Nippon*, 40, 368-380.

THOMPSON W.R.(1935) 'On a criterion for rejection of observations and the distribution of the ratio of the deviation to the sample standard deviation.', *Ann. Math. statist.*, 6, 214-219.

UNICEF (1992). *Situation analysis of women and children in the I.R.Iran.*

VAN'T HOF M.A., WIT J.M., and ROEDE M.J.(1985) 'A method to construct age references for skewed skinfold data using Box-Cox transformations to normality.', *Hum. Biol.*, 57, 131-139.

VISWESWARA RAO K., REDDY P.J., NARAYANAN T.P., and SUBHADARA DEVI V.(1979) 'Discriminant function analysis - A case study for an evaluation of various forms of protein-energy malnutrition.', *Ind. J. Med. Res.*, 49: 99-108.

WATERLOW J.C., BUZINA R., KELLER W., LANE J.M., NICHAMAN M.Z., and TANNER J.M.(1977) 'The presentation and use of height and weight data for comparing the nutritional status of groups of children under the age of 10 years.', *Bull. WHO*, 55, 489-98.

WHO (1983). *Measuring change in nutritional status:* Guidelines for assessing the nutritional impact of supplementary feeding programmes for vulnerable groups.

WHO Technical Report Series No. 593 (1976). Methodology of nutritional surveillance.

WHO (1978). A growth chart for international use in maternal and child care guidance for primary health care personnel. Geneva: *World Health Organisation*.

WHO Technical Report Series No. 724 (1985) Energy and protein requirements: report of a joint FAO/WHO/UNU Expert Consultation.

WHO (1986) 'Use and interpretation of anthropometric indicators of nutritional status.', *Bull WHO*, 64: 924-41.

WHO (1995) 'Physical status: the use and interpretation of anthropometry.', *World Health Organization technical report*, series 854.

WILKS S.S.(1962). *Mathematical Statistical*. Wiley, New York.

WILKS S.S.(1963) 'Multivariate statistical outliers.', *Sankhyā*, A, 25, 407-426.

WILSON, E.B., and HILFERTY M.M.(1931) 'The distribution of chi-square.', *Proceeding of the National Academy of Sciences* 17, 684-688.

WOODHOSUE G., and GOLDSTEIN H.(1988) 'Educational performance indicators and LEA league tables.', *Oxford review of education.*, Vol 14, 13, 310-320.

WOODHOUSE G.(1993). *A guide to ML3 for new user* 2nd ed. Institute of Education, University of London.

WRIGHT T.W.(1884). *A treatise on the adjustment of observations by the method of least squares*. Van Nostrand, New York.

YIP R., and TROWBRIDGE F.L.(1989) 'Limitation of the current international growth reference.', Paper presented at the 14th international congress of nutrition, Seoul, Korea, 20-25 Aug, abstract, P4-135: 800.

ZALI R., MOHAMMAD K. and MASJEDI M.R.(1993). *Bar-resi-ye salamat va bimari dar Iran (A survey of health and disease in Iran)*. Vol. 25, Under Secretory of Research, Ministry of Health and Medical Education of I.R.Iran.

ZEGER S.L., LIANG K.Y., AND ALBERT P.S.(1988) 'Models for longitudinal data: a generalized estimating equation approach.', *Biometrics*, 44, 1049-60.

ZEGER S.L., KARIM, M.R.(1991) 'Generalized linear models with random effects; a Gibbs Sampling approach.', *Journal of the American Statistical Society*, 86, 79-102.

APPENDICES

APPENDIX A

Questionnaire of the National Health Survey 1990-2,

Iran

MINISTRY OF HEALTH AND MEDICAL EDUCATION
UNDERSECRETARY OF RESEARCH

NATIONAL HEALTH SURVEY of
HEALTH AND DISEASE

Province .... County .... Cluster ID .... Urban/Rural[1]

Name of village or city .... Name of head of house ....

Address .... Date: / /199

Note: If resident is absent, leave a message about your future visit with a neighbour. Do not revisit the house if two visit are unsuccessful*.

Family ID	No. of visits	Interviewer's name	Date	Time	Result

Table of household members interviewed

No	Name	Sex	Date of Birth (with day & month if aged under 5 yrs.)	Relationship to head of the house	Completed questionnaire (please mark with X)								
					1	2	3	4	5	6			
										Blood		Faeces	
										Sample	Result	sample	Result
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													

*: To make up the required numbers choose another family from the end of the cluster.

1: Delete if inapplicable

MINISTRY OF HEALTH AND MEDICAL EDUCATION
UNDERSECRETARY OF RESEARCH

National Health Survey of
HEALTH AND DISEASE

Family and Health Supplies

Name of interviewer

Name of respondent:

1	2	3	4	5	6	7	8
							1
Area		Cluster		Family	Person		Questionnaire

1) No. of household members	☐☐	10) How do you wash vegetables when you eat them raw? 1) Inapplicable ☐ 2) With untreated water ☐ 3) With salt water ☐ 4) With disinfected water ☐	☐
1- Yes 2- No 2) Bathroom ☐ ☐ 3) Refrigerator ☐ ☐ 4) T.V ☐ ☐ 5) Radio ☐ ☐	☐ ☐ ☐ ☐		
6) Total floor area in house (M^2)	☐☐☐	11) If you have a 2-3 year old child at home, when was he/she weaned?	☐☐☐
7) Access to Health Care Centres 1) Easy ☐ 2) Difficult ☐	☐		
8) Health Care Centre visited 1) PHW or Health House ☐ 2) Public Health Centres ☐ 3) Welfare Health Centres ☐ 4) Other ☐	☐	12) Type of bathroom (toilet) 1) Indoor and sanitary ☐ 2) Indoor and semi-sanitary ☐ 3) Indoor but not sanitary ☐ 4) Outdoor ☐ 5) None ☐	☐
9) Water supply 1) Tab water ☐ 2) Clean well or quanat ☐ 3) Unclean well or quanat ☐ 4) River ☐ 5) Collected rain water ☐ 6) Other (please specify) ☐	☐		

MINISTRY OF HEALTH AND MEDICAL EDUCATION
UNDERSECRETARY OF RESEARCH

National Health Survey of
HEALTH AND DISEASE

Food consumption over two days

Name of interviewer:

Name of respondent:

1	2	3	4	5	6	7	8
							2
Area		Cluster		Family	Person		Questionnaire

No.	Food	1st day					2nd day					Total
		Breakfast	Mid morning	Lunch	Mid afternoon	Supper	Breakfast	Mid morning	Lunch	Mid Afternoon	Supper	
1	Bread											
2	Rice											
3	Macaroni											
4	Milk											
5	Cheese											
6	Yogurt											
7	Cream											
8	Butter											
9	Oil											
10	Sugar (p&c)											
11	Honey											
12	Jam											
13	Eggs											
14	Meat											
15	Chicken											
16	Fish											
17	Peas											
18	Beans											
19	Lentils											
20	Other Cereals											
21	Leafy Veg.											
22	Potato											
23	Other root Veg.											
24	Garden Veg.											
25	Citrus Fruit											
26	Other tree growing fruit											
27	Garden fruit											
28	Dates											
29	Dried fruit											
30	Nuts											

MINISTRY OF HEALTH AND MEDICAL EDUCATION
UNDERSECRETARY OF RESEARCH

National Health Survey of
HEALTH AND DISEASE

Reproductive information on 15-49 year old married women
Name of interviewer:

Name of respondent:

1	2	3	4	5	6	7	8
							3
Area		Cluster		Family	Person		Questionnaire

1) Age when interviewed ☐☐

2) Age at menarche ☐☐

3) Age at marriage ☐☐

4) Age at first pregnancy ☐☐

5) Number of live births ☐☐

6) Number of sons still alive ☐☐

7) Number of daughter still alive ☐☐

8) Do you agree with family planning?
1) Yes ☐ 2) No ☐ ☐

9) if disagree, why? ☐
1) Religious beliefs ☐
2) Economic issues ☐
3) Other reasons ☐

10) If agree, do you use contraceptives?
1) Yes ☐ 2) No ☐ ☐

11) If yes, which method(s) do you use? ☐
1) The pill ☐ 4) Rhythm ☐
2) I.U.D ☐ 5) Tubal ligation ☐
3) Condom ☐ 6) Other ☐

12) From where do you obtain your contraceptives? ☐
1) Health house ☐
2) Other public Centres ☐
3) Private sectors ☐

13) If you agree with family planning but do not use contraceptives, why? ☐
1) Pregnant ☐
2) Fear of side effects/complications ☐
3) Husband's disapproval ☐
4) Other reasons ☐

14) If you answered yes to Q8, who provided you with the relevant information regarding family planning? ☐
1) PHW ☐ 2) Husband ☐
3) Relatives ☐ 4) Pub. doctor ☐
5) Priv. doctor ☐ 6) Mass Media ☐
7) Nurse, midwife, family planning advisor ☐
8) Other ☐

15) If you use contraceptives do you have any problem obtaining them? ☐
Yes ☐ 1) No ☐
2) Expensive ☐
3) Centre is too far away ☐
4) Good quality not easily obtainable ☐
5) Long time delays ☐
6) Unco-oprative provider ☐
7) Other ☐

16) If you have been pregnant more than once, what was the time interval between your last two pregnancies? ☐☐☐

17) how long (months) has it been since you were last pregnant? ☐☐☐
* Zero for pregnant
* Duration since marriage if nulliparous

18) Education (no. of grades) ☐☐

MINISTRY OF HEALTH AND MEDICAL EDUCATION
UNDERSECRETARY OF RESEARCH

National Health Survey of
HEALTH AND DISEASE

Individual information and medical examination
Name of interviewer:

Name of respondent:

1	2	3	4	5	6	7	8
							4
Area		Cluster		Family	Person		Questionnaire

Part I: General information: Life style and medical history

1) Age	☐☐	9) Marital status (if aged 15 years or more) 1) Married ☐ 2) Single ☐ 3) Widowed ☐ 4) Divorced ☐	☐
2) Sex 1) Male ☐ 2) Female ☐	☐		
3) Education (no. of grades) 1) Able to read ☐ 2) Able to read and write ☐	☐☐	Tobacco consumption (if aged ≥15 years) 10) Do you smoke cigarettes? 1) Yes ☐ 2) No ☐ If yes, no. per day? 11) If yes, from what age? 12)Do you smoke a pipe or anything else? Yes ☐ 2) No ☐	☐☐ ☐☐ ☐
4) Occupation 1) Student ☐ 2) Unemployed ☐ 3) Housewife ☐ 4) Civil servant ☐ 5) Private sector ☐ 6) Unskilled labourer ☐ 7) Skilled labourer ☐ 8) Self-employed ☐ 9) Agricultural or livestock farmer ☐ 10) Retired ☐ 11) Other (please specify) ☐	☐☐	Answer if aged 2-18 years 13) Height (cm) 14) Weight (kg) Blood pressure (mmHg) 15) Systolic 16) Diastolic	☐☐☐ ☐☐ ☐☐☐ ☐☐☐
Personal hygiene 5) How frequently do you brush your teeth? 1) Never ☐ 2) Once a day ☐ 3) More than once daily ☐	☐	17) Eyesight 1) Good ☐ 2) Wear glasses ☐ 3) Do not wear glasses but have sight problems ☐ 4) Blind in one eye ☐ 5) Totally blind ☐	☐
Do you wash your hands 1- Yes 2-No 6) before meals ☐ ☐ 7) after going to the toilets ☐ ☐	☐ ☐		
8) How many bath do you have each week? (zero if not every week)	☐		

1	2	3	4	5	6	7	8
							4
Area		Cluster		Family	Person		Questionnaire

18) Hearing
1) Good ☐
2) Uses hearing aid ☐
3) Poor but does not use hearing aid ☐
4) Completely deaf in one ear ☐
5) Completely deaf ☐
☐

Teeth (please use the corresponding form)
19) D.M.F (milk)
20) D.M.F (permanent)

21) Do you have any neck pain?
1) Yes ☐ 2) No ☐
☐

22) Do you have any back pain or lower back pain?
1) Yes ☐ 2) No ☐
☐

Joint pain	1-No	2-Mild	3-Severe	
23) Hip or pelvic	☐	☐	☐	☐
24) Other	☐	☐	☐	☐

25) Do you have any physical disability or mental retardation?
1) No ☐ 2) Eye ☐
3) Ear ☐ 4) Arm ☐
5) Leg ☐
6) Other ☐
7) Several sites affected ☐
8) Retardation ☐
☐

26) If yes, how did you become disabled?
1) Congenital ☐
2) Polio ☐
3) Accident ☐
4) War ☐
5) Stroke ☐
6) Chemical exposure ☐
7) Other (please specify) ☐
☐

27) Do you have chronic cough and sputum?
1) Yes ☐ 2) No ☐
☐

28) Chronic constipation (lasting more than 3 days)
1) Yes ☐ 2) No ☐
☐

Medical history	1- Yes	2-No	
29) Renal & Urinary stones	☐	☐	☐
30) Asthma	☐	☐	☐
31) Peptic ulcer	☐	☐	☐
32) Hypersensitivity, Urticaria & Pruritis	☐	☐	☐
33) Seasonal nasal discharge	☐	☐	☐
34) Epilepsy/convulsions	☐	☐	☐
35) Hepatitis	☐	☐	☐
36) Cataract	☐	☐	☐
37) Cancer	☐	☐	☐
38) Malaria	☐	☐	☐
39) Diabetes Mellitus	☐	☐	☐
40) Hypertension	☐	☐	☐
41) Gas poisoning (specify)	☐	☐	☐
42) Other eye disorders	☐	☐	☐
43) Skin disorders	☐	☐	☐
44) Other Respiratory dis.	☐	☐	☐
45) Other digestive dis.	☐	☐	☐
46) Psychiatric dis.	☐	☐	☐

47) Age at menopause (women aged 50-59 years) ☐☐

48) Do you take any medication regularly?
1) Yes ☐ 2) No ☐
Reason :
Type of medication:
☐

49) Have you had a B.C.G vaccination (2-6 year olds)?
1) Yes ☐ 2) No ☐
☐
50) B.C.G scar
1) Yes ☐ 2) No ☐
☐
51) P.P.D Test result (mm) ☐☐

1	2	3	4	5	6	7	8
							4
Area		Cluster		Family	Person		Questionnaire

Part II: Medical Examination

52) Pulse
1) Regular ☐ 2) Irregular ☐ ☐

53) Appearance of skull
1) Normal ☐ 2) Abnormal ☐ ☐

Eyes

	1-Yes	2-No	
54) Pteryginn/Finguecula	☐	☐	☐
55) Exophthalmia	☐	☐	☐
56) Conjunctival redness	☐	☐	☐
57) Angular blepharitis	☐	☐	☐
59) dryness	☐	☐	☐
60) Abnormalities of the comea or iris	☐	☐	☐
61) Jaundice	☐	☐	☐
62) Bitot spots	☐	☐	☐
63) Keratomalacia	☐	☐	☐
64) Strabismus & Asymmetry	☐	☐	☐

65) Eyelids
1) Normal ☐ 2) Abnormal ☐ ☐

Ears
66) Appearance
1) Normal ☐ 2) Abnormal ☐ ☐

67) Ear canal
1) Patent ☐ 2) Closed ☐ ☐

68) Purulent discharge
1) Yes ☐ 2) No ☐ ☐

Lips

	1-Yes	2-No	
1) Cheilosis	☐	☐	☐
2) Cyanosis	☐	☐	☐
3) Cleft lip	☐	☐	☐
4) Nasolabial seborrhoea	☐	☐	☐

73) Tongue ☐
1) Atrophic ☐
2) Hypertrophic ☐
3) Geographical ☐
4) Grooved ☐
5) Beefy ☐
6) Normal ☐

74) Tonsils
1) Normal ☐ 3) Suppurative ☐ ☐
2) Large ☐ 4) Surgery ☐ ☐

Gums

	1-Yes	2-No	
75) Hemorrhage	☐	☐	☐
76) Swelling	☐	☐	☐
77) Periodental dis.	☐	☐	☐
78) Pale	☐	☐	☐
79) Scurbotic	☐	☐	☐

Nose

	1-Yes	2-No	
80) Stuffy	☐	☐	☐
81) Septal deviation	☐	☐	☐
82) Turbinate thickening	☐	☐	☐
83) Polyps	☐	☐	☐

84) Parotids
1) Normal ☐ 2) Abnormal ☐ ☐

1	2	3	4	5	6	7	8
							4
Area		Cluster		Family	Person		Questionnaire

85) Neck veins
1) Normal ☐
2) Congested ☐
☐

86) Neck lymphadenopathy
1) Yes, Anterior ☐
2) Yes, Posterior ☐
3) No ☐
4) Both 1 & 2 ☐
☐

87) Axillary lymphadenopathy
1) Yes ☐ 2) No ☐
☐

88) Groin lymphadenopathy
1) Yes ☐ 2) No ☐
☐

89) Thyroid
1) Group 0 ☐ 4) Group 2 ☐
2) Group A1 ☐ 5) Group 3 ☐
3) Group B1 ☐ 6) Group 4 ☐
☐

90) Thorax
1) Normal ☐
1) Pigeon chest ☐
2) Other deformity ☐
☐

Lung auscultation

	1-Normal	2-wheezing	3-Rales	
91) Right	☐	☐	☐	☐
92) Left	☐	☐	☐	☐

93) Heart
1) Normal ☐
1) Murmur ☐
2) Other abnormalities ☐
☐

94) Breasts
1) Normal ☐
1) Mass ☐
☐

Abdomen	1- Normal	2- Abnormal	
95) Liver	☐	☐	☐
96) Spleen	☐	☐	☐
97) Kidneys	☐	☐	☐

98) Abdominal mass
1/9- Yes ☐
2- No ☐
☐

1	2	3
4	5	6
7	8	9

99) abdominal tenderness
1/9- Yes ☐ 2- No ☐
☐

100) Surgical scar
1/9- Yes ☐ 2- No ☐
☐

101) Hernia
1) Inguinal ☐
2) Femoral ☐
3) Umbilical ☐
4) None ☐
☐

102) Joints
1) Normal ☐
2) Abnormal ☐
(stiffness, limitation of movement, swelling, deformity)
☐

If abnormal, locate: ☐☐

	Left	Right
Shoulder		
Elbow		
Wrist		
Finger		
Hip		
Knee		
Ankle		
Toes		

MINISTRY OF HEALTH AND MEDICAL EDUCATION
UNDERSECRETARY OF RESEARCH

National Health Survey of
HEALTH AND DISEASE

1	2	3	4	5	6	7	8
							4
Area		Cluster		Family	Person		Questionnaire

103) Appearance of back and lower back

1- Normal ☐ 2- Scoliosis ☐
3- Kyphosis ☐ 4- Lordosis ☐
☐

Legs
104) appearance
1- Normal ☐ 2- Genu Varum ☐
3) Genu Valgum ☐
4) Other abnormalities ☐
☐

105) Edema
1) Yes ☐ 2- No ☐
☐

106) Varices
1) Yes ☐ 2- No ☐

Skin

	1-Yes	2-No	
107) Impetigo	☐	☐	☐
108) Warts	☐	☐	☐
109) Scabies	☐	☐	☐
110) Psoriasis	☐	☐	☐
111) Atopic eczema	☐	☐	☐
112) Urticaria	☐	☐	☐
113) Alopecia	☐	☐	☐
114) Skin tumour	☐	☐	☐
115) Recent burn	☐	☐	☐
116) Vitiligo	☐	☐	☐
117) Albinism	☐	☐	☐
118) Xerotic skin	☐	☐	☐
119) Dryness & scaling	☐	☐	☐
120) Follicular hyperkeratosis	☐	☐	☐

Skin

	1-Yes	2-No	
121- Molluscum contagiosum	☐	☐	☐
122- Contact dermatitis	☐	☐	☐
123- Seborrhoeic dermatitis	☐	☐	☐

124) Acne
1) Purulent ☐
2) Non Purulent ☐
3) None ☐
4) Mixed ☐
☐

125) Nails
1) Spooned-shaped ☐
2) Grooved nails ☐
3) Normal ☐
☐

Fungal lesions

	1-Yes	2-No	
126) Scalp Fungi	☐	☐	☐
127) Tinea cruris or pedis	☐	☐	☐
128) Impetigo	☐	☐	☐

129) cyanosis
1) Yes ☐
2) No ☐
☐

1	2	3	4	5	6	7	8
							4
Area		Cluster		Family	Person		Questionnaire

Part III: Diagnosis of illness

130) Diseased 1) Yes ☐ 2) No ☐

	I	II	III	IV	V
If yes: name of disease(s)					
ICD Code	131 ☐☐☐	135 ☐☐☐	139 ☐☐☐	143 ☐☐☐	147 ☐☐☐
Diagnosed on the basis of	132 ☐	136 ☐	140 ☐	144 ☐	148 ☐
1) Medical history	☐	☐	☐	☐	☐
2) Examination	☐	☐	☐	☐	☐
3) Both	☐	☐	☐	☐	☐
Severity of disease	132 ☐	137 ☐	141 ☐	145 ☐	149 ☐
1) Mild	☐	☐	☐	☐	☐
2) Moderate	☐	☐	☐	☐	☐
3) Severe	☐	☐	☐	☐	☐
Certainty of diagnosis	132 ☐	138 ☐	142 ☐	146 ☐	150 ☐
1) Certain	☐	☐	☐	☐	☐
2) Highly probable	☐	☐	☐	☐	☐
3) Probable	☐	☐	☐	☐	☐

Code 029 is for chemical injury.
Code 039 is for a disability acquired during the war.
Code 308 is here defined as PTSD.
Code 280 is defined as 'Iron deficiency anemia' in the ICD-9 book but this diagnosis is based on clinical judgement.
Code 401 is defined as 'hypertension'.

Anemia (on the basis of Hb)			
	Mild	Moderate	Severe
Women	11-11.9	10-10.9	<10
Men	13-13.9	12-12.9	<10
6 month to 6 yrs.	9.5-10.4	8.5-9.4	<10
7 yrs. to 12 yrs.	10-10.9	9-9.9	<10

Hypertension	Diastolic
Mild	90-104
Moderate	105-114
Severe	≥115

Systolic over 160 & diastolic below 90 is isolated systolic hypertension[1]

1: Hypertension in the analysis of the Nation Health Survey data was defined according to WHO and Harrison (Zali et al., 1993; pp 253)

MINISTRY OF HEALTH AND MEDICAL EDUCATION
UNDERSECRETARY OF RESEARCH

National Health Survey of
HEALTH AND DISEASE

Name of examiner

Name of patient

1	2	3	4	5	6	7	8
							5
Area		Cluster		Family	Person		Questionnaire

Oral health assessment

Education (no. of grade) ☐ Age (yrs.) ☐ Sex 1) Male ☐ 2) Female ☐

1) Malocclusion

(24) ☐

0) None
1) Slight
2) Moderate or severe

Periodontal status (CPTIN)
0) Healthy
1) Bleeding
2) calculus
3) Pocket 4-5 mm (black band of probe partially visible)
4) Pocket 6 mm or + (blck band of prob not visible)
X) Excluded sextant

(25)				(27)
(28)				(30)

Dentition status and treatment needed

Permanent teeth	Primary teeth	Treatment
0) Sound	A	0) None
1) Decayed	B	1) Caries arresting or sealant care
2) Filled & decayed	C	2) One surface filling
3) Filled, no decayed	D	3) Two or more surface filling
4) Missing due to caries	E	4) Crown or bridge abutment
5) Missing for another reason	-	5) Bridge element
6) sealant, varnish	F	6) Pulp care
7) Bridge abutment or special crown	G	7) Extraction
8) Unerupted tooth	-	8) Need for other care
9) Excluded tooth	-	9) Please specify

				55	54	53	52	51	61	62	63	64	65				
	18	17	16	15	14	13	12	11	21	22	23	24	25	26	27	28	
Status (31)																	(46)
(47) Treatment																	(62)

				85	84	83	82	81	71	72	73	74	75				
	48	47	46	45	44	43	42	41	31	32	33	34	35	36	37	38	
Status (63)																	(78)
(79) Treatment																	(94)

Fluorosis

(95) ☐

0) Normal
1) Questionable
2) Very mild
3) Mild
4) Moderate
5) Severe

Denture wearing	Upper	Lower	Both
Complete	☐	☐	☐
Partial	☐	☐	☐

Result of the assessment:

Tooth	No. decayed	No. filled	No. extracted	Total
Permanent				
Primary				

MINISTRY OF HEALTH AND MEDICAL EDUCATION
UNDERSECRETARY OF RESEARCH

National Health Survey of
HEALTH AND DISEASE

Name of examiner:

Name of patient

1	2	3	4	5	6	7	8
							6
Area		Cluster		Family	Person		Questionnaire

Laboratory results

Faeces (Parasites)	1-Yes	2-No		16) Hep B surface Ag.		
1) Enterobius	☐	☐	☐	1) Positive	☐	☐
2) Ascaris	☐	☐	☐	2) Negative	☐	
3) Giardia	☐	☐	☐			
4) Entamoeba histolytica	☐	☐	☐	17) Serum ferritin		☐☐
5) Ancylostoma	☐	☐	☐			
6) Taenia	☐	☐	☐			
7) Trichocephalus	☐	☐	☐			
8) Strongyloides	☐	☐	☐			
9) Others (please specify)	☐	☐	☐	18) Cholesterol		☐☐☐
Blood						
10) Hemoglobin			☐☐☐			
11) Hematocrit			☐☐			
12) MCV			☐☐☐	19) Sputum results from suspect cases (with cough and productive/ bloody sputum lasting 4+ weeks)		
Hb Electrophoresis						
13) A2			☐☐			
14) F			☐☐			
15) Other Hemoglobin			☐	1) Positive	☐	☐
1) No 2) Yes (type)				1) Negative	☐	
				1) Not suspicious	☐	

APPENDIX B

B.1 Models for longitudinal growth study

Goldstein (1986) presents the concept of longitudinal growth data as a two-level model: level 2, the highest level, is the individual subject j ($j=1,\ldots,n$) while at level 1, within each level 2 unit, are the occasions i ($i=1,..,T_j$) when the subject is measured. So his or her growth can be modelled as a polynomial function of time t, which is suggested as a promising approach to summarize growth data (Goldstein, 1986)

$$Y_{ij} = \beta_{0j} + \beta_{1j}\, t_{ij} + \beta_{2j}\, t^2_{ij} + \ldots + e_{ij}$$

The degree of the polynomial being fitted for all individuals is K, say.

In this formulation, the shape of each person's curve is unique since each has its own sets of parameters $\{\beta_{kj}\}$. A simple between-person model is:

$$\beta_{kj} = \gamma_{k0} + u_{kj} \qquad k=0,\ldots,K$$

Also let $Var(\beta_k)=\sigma^2_k$, $Cov(u_{kj}\,,\,u_{k'j})=\sigma_{k'k}$, and it is assumed that level 1 random terms for an individual are distributed independently with mean and variance 0 and σ^2_e, and $Cov(e_{ij}\,,\,e_{i'j})=0$ for $i\neq i'$. It is also assumed that level 1 random terms for different individuals are distributed independently, and the level 1 and the level 2 random terms are independent.

Two features of real growth that this model incorporates are:

1) *between-person variation is a function of time*

2) *a person's response on different occasions are correlated, because of the subject specific regression parameters.*

This basic model can be elaborated in a number of ways. Two will be described here. It may be possible to account for growth curve coefficient variability in terms of differences between persons on some characteristic (Z, say) that remains fixed across occasions -sex, for example. Thus the βs might be modelled as:

$$\beta_{kj} = \gamma_{k0} + \gamma_{kj} Z_j + u_{kj} \qquad k=0,\ldots,K$$

Covariates that changes over time (say, W) can also be included in the model:

$$Y_{ij} = \beta_{0j} + \alpha_j W_{ij} + \beta_{1j} t_{ij} + \beta_{2j} t^2_{ij} + \ldots + e_{ij}$$

A second elaboration involves modelling level 1 dispersion as a function of time, which is discussed in section 4.5.5.1.

B.2 Autocorrelation structure at level 1

In some situations the assumption of the independency of level 1 residuals may be false. For growth data if the measurements on an individual are obtained very close together in time, they will tend to have similar departures from that individuals underlying growth curve. That is, there will be 'autocorrelation' between level 1 residuals. Goldstein et al. (1994) discuss multilevel time series in both the discrete case, where the measurements are made at the same set of equal intervals for all level 2 units, and the continues time case where the time intervals can vary. A general model for the level 1 residuals can be written as

$$Cov(e_t, e_{t-s}) = \sigma^2_e \, f(z,s) .$$

The covariance between two measurements depends on a variance function of the time (or age) difference between the measurements but is not a function of t, the age at which the measurements are made. A possible form of this function is negative exponential reflecting the common assumption that with increasing time difference the covariance tends to fixed value, $\alpha \, \sigma^2_e$. Typically this is assumed to be zero and

$$f(z,s) = \alpha + \exp(-g(\beta, z, s))$$

where β is a vector of parameters for explanatory variables z. Goldstein (1995) explains some choices for **g**. These models are not discussed in length here because our data are entirely cross-sectional.

B.3 Other parameter estimation procedures in analysing hierarchial data

Longford (1987) developed a procedures based upon a 'Fisher scoring' algorithm and Raudenbush (1994) showed that it is formally equivalent to IGLS. A program VARCL (Longford, 1987) uses this algorithm and also incorporates certain extensions, for example to handle discrete response data. A rather different approach is to view equation 12, and more general extensions, as a Bayesian linear model (Lindley and Smith, 1972) where the β_j are assumed to be exchangeable and to have a prior distribution with variance σ^2_{u0}. The full Bayes estimation then requires a prior distribution for the random parameters also, in this case the level 1 and level 2 variances. An alternative to the full Bayes estimation, known as 'empirical Bayes', ignores the prior distributions of the random parameters, treating them as known for purpose of inference. When Normality is assumed, these estimates are the same as IGLS or RIGLS. Bryk and Rudenbush (1992) describe the use of the EM algorithm to provide such estimates, and the program HLM use this algorithm.

Another approach which parallels all of these is that of the Generalized Estimating Equation (GEE) introduced by

Zeger et al. (1988). The principal difference is that GEE obtains the estimates of V (variance of the random part) using simple regression or 'moment' procedures based upon functions of the actual calculated raw residuals. It is concerned principally with modelling the fixed coefficients rather than exploring the structure of the random component of the model. While the resulting coefficient estimates are consistent they are not fully efficient. In some circumstances, however, GEE coefficient estimates may be preferable, since they will usually be quicker to obtain and they make weaker assumption about the structure of V.

More recently, the full Bayesian treatment has become computationally feasible with the development of 'Markov Chain Monte Carlo' (MCMC) methods, especially Gibbs sampling (Zeger and Karim, 1991). A Bayesian package using Gibbs sampling, BUGS, which might fit most of the discussed models is also available. This has the advantage, in small samples, that it takes account of the uncertainty associated with estimates of the random parameters and can provide exact measures of uncertainty. The maximum likelihood methods tend to overestimate precision because they ignore this uncertainty. In small samples this will be important especially when obtaining 'posterior' estimates for residuals. However Goldstein (1995) presents an alternative 'bootstrap' procedure for taking account of this uncertainty.

APPENDIX C

Comparisons of the centiles of weight and height of children in urban areas of Iran with urban Tehran according to sex and age

Table C.1 Comparison of the centiles of weight of boys in urban Iran and urban Tehran

Percent	Age	Centile urban Iran	[95% Confidence Interval] urban Iran		Centile urban Tehran	Smoothed urban Tehran
3	2	7.996	7.65	8.36	7.996	7.80
3	3	8.997	8.67	9.33	8.997	8.99
3	4	10.00	9.59	10.44	10.00	10.21
3	5	11.00	10.38	11.65	11.00	11.46
3	6	12.00	11.23	12.81	12.00	12.77
3	7	13.00	12.07	13.99	14.79	14.13
3	8	14.99	14.15	15.89	16.00	15.59
3	9	17.99	17.44	18.56	16.92	17.17
3	10	20.00	19.43	20.59	21.01	18.92
3	11	21.01	20.14	21.91	20.06	20.90
3	12	24.85	24.24	25.47	23.66	23.17
3	13	25.00	23.47	26.63	25.00	25.70
3	14	27.99	25.81	30.35	28.61	28.90
3	15	33.48	31.91	35.12	32.85	33.05
10	2	8.997	8.70	9.31	10.00	9.39
10	3	10.00	9.72	10.30	10.00	10.64
10	4	11.00	10.63	11.37	12.00	11.93
10	5	13.00	12.68	13.32	13.00	13.27
10	6	13.99	13.61	14.39	13.99	14.68
10	7	14.99	14.54	15.47	16.00	16.18
10	8	16.99	16.47	17.53	18.19	17.82
10	9	20.00	19.53	20.49	20.00	19.62
10	10	21.99	21.46	22.54	22.98	21.65
10	11	23.99	23.39	24.61	23.99	23.85
10	12	25.99	25.19	26.82	25.00	26.30
10	13	27.99	26.83	29.19	27.99	29.30
10	14	32.68	31.41	34.00	32.00	32.90
10	15	37.90	36.53	39.31	39.01	38.10
25	2	10.00	9.71	10.30	11.00	10.89
25	3	11.00	10.71	11.29	12.00	12.16
25	4	13.00	12.72	13.28	13.99	13.48
25	5	14.99	14.70	15.29	14.99	14.87
25	6	16.00	15.67	16.34	16.00	16.37
25	7	17.99	17.66	18.32	18.99	18.01
25	8	20.00	19.61	20.40	20.00	19.84
25	9	21.99	21.51	22.49	21.99	21.90
25	10	23.99	23.45	24.55	25.00	24.24
25	11	25.99	25.35	26.65	27.00	26.93
25	12	28.99	28.23	29.76	28.99	30.04
25	13	32.00	31.07	32.97	33.01	33.63
25	14	36.01	34.79	37.28	37.75	38.00
25	15	42.01	40.58	43.49	44.47	44.00
50	2	12.00	11.67	12.34	12.00	12.38
50	3	13.00	12.69	13.31	13.00	13.65
50	4	14.99	14.66	15.34	14.99	15.01
50	5	16.00	15.67	16.34	16.00	16.48
50	6	17.99	17.61	18.37	17.99	18.12
50	7	20.00	19.61	20.40	20.00	19.96
50	8	21.99	21.55	22.45	21.99	22.06
50	9	25.00	24.45	25.55	23.99	24.47
50	10	27.00	26.43	27.58	27.99	27.26
50	11	29.99	29.31	30.68	29.99	30.48
50	12	32.00	31.22	32.80	33.01	34.19
50	13	36.01	35.05	37.00	38.01	38.60
50	14	42.01	40.81	43.24	45.01	44.50
50	15	48.52	46.95	50.14	51.00	50.60
75	2	13.00	12.63	13.37	13.99	13.76
75	3	13.99	13.68	14.32	14.99	15.08
75	4	16.00	15.64	16.37	16.99	16.54
75	5	17.99	17.60	18.38	18.99	18.17
75	6	20.00	19.55	20.46	20.00	20.04
75	7	21.01	20.61	21.40	21.99	22.19
75	8	23.99	23.50	24.50	23.99	24.69
75	9	27.00	26.41	27.61	27.00	27.61
75	10	29.99	29.34	30.65	31.00	31.00
75	11	33.01	32.24	33.80	34.98	34.92
75	12	36.01	35.16	36.88	38.47	39.40
75	13	42.01	40.78	43.28	43.99	44.46
75	14	49.99	48.28	51.77	51.00	51.00
75	15	54.98	53.08	56.95	54.98	57.30
90	2	14.99	14.45	15.56	14.99	15.00
90	3	16.00	15.57	16.45	16.00	16.49
90	4	17.99	17.51	18.48	17.99	18.16
90	5	20.00	19.51	20.50	20.00	20.08
90	6	21.99	21.43	22.57	21.99	22.31
90	7	23.99	23.42	24.58	25.99	24.92
90	8	25.99	25.46	26.53	28.78	27.99
90	9	29.99	29.23	30.76	29.99	31.58
90	10	33.01	32.24	33.80	36.19	35.76
90	11	36.01	35.14	36.90	38.01	40.55
90	12	42.99	41.28	44.76	48.18	45.96
90	13	49.00	46.67	51.45	49.99	52.00
90	14	56.99	54.18	59.95	59.97	59.83
90	15	61.99	59.46	64.62	68.78	66.40
97	2	16.00	15.48	16.54	16.00	16.39
97	3	17.99	17.34	18.66	18.69	18.21
97	4	20.00	19.32	20.70	20.45	20.30
97	5	21.99	21.36	22.64	22.98	22.73
97	6	23.99	23.30	24.71	23.87	25.59
97	7	27.00	26.00	28.04	30.08	28.96
97	8	29.99	28.92	31.10	34.98	32.92
97	9	33.98	32.60	35.42	35.02	37.55
97	10	40.00	37.08	43.15	45.46	42.90
97	11	42.01	40.02	44.10	45.74	48.94
97	12	49.99	45.29	55.19	55.92	55.60
97	13	54.98	50.80	59.50	64.19	62.63
97	14	63.11	58.95	67.57	65.56	69.72
97	15	68.51	65.33	71.84	82.26	76.10

Table C.2 Comparison of the centiles of height of girls in urban Iran and urban Tehran

Percent	Age	Centile urban Iran	[95% Confidence Interval] urban Iran		Centile urban Tehran	Smoothed urban Tehran
3	2	73.6	72.4	74.9	70.4	70.4
3	3	77.5	76.0	79.0	75.4	78.1
3	4	82.7	80.3	85.2	83.2	85.1
3	5	90.0	87.8	92.3	90.2	91.7
3	6	98.0	96.4	99.6	102.0	97.7
3	7	103.8	102.2	105.3	106.6	103.4
3	8	109.0	107.4	110.6	109.0	108.7
3	9	114.0	112.6	115.3	113.6	113.9
3	10	119.0	117.5	120.5	119.4	118.7
3	11	123.0	120.4	125.5	123.9	123.3
3	12	125.5	119.7	131.3	127.1	127.4
3	13	130.0	124.2	135.8	127.4	131.6
3	14	139.7	138.1	141.2	139.1	136.3
3	15	144.0	143.0	144.9	144.7	141.3
10	2	80.0	79.1	80.9	80.0	76.2
10	3	83.0	82.0	84.0	82.2	83.5
10	4	89.0	87.9	90.1	87.0	90.3
10	5	96.0	95.0	97.0	96.9	96.8
10	6	103.0	102.0	103.9	104.0	102.9
10	7	109.0	108.1	109.8	109.0	108.7
10	8	113.0	111.8	114.1	114.0	114.2
10	9	118.0	116.9	119.0	120.0	119.5
10	10	123.0	121.8	124.1	123.0	124.5
10	11	128.0	126.5	129.4	127.9	129.5
10	12	133.0	131.3	134.6	134.0	134.3
10	13	137.0	135.0	138.9	138.6	139.1
10	14	144.0	142.6	145.3	143.1	143.6
10	15	147.0	145.8	148.1	150.0	147.3
25	2	83.0	81.9	84.1	83.0	81.1
25	3	87.0	86.1	87.94	88.0	88.1
25	4	94.0	93.2	94.84	95.0	94.8
25	5	101.0	100.2	101.8	102.2	101.2
25	6	107.0	106.1	107.8	107.0	107.4
25	7	113.0	112.2	113.8	114.0	113.3
25	8	118.0	117.1	118.8	119.5	119.0
25	9	122.0	121.1	122.8	124.0	124.5
25	10	128.0	127.1	128.8	128.5	129.8
25	11	132.2	131.0	133.3	132.0	135.0
25	12	139.0	137.9	140.0	140.0	140.4
25	13	144.0	142.8	145.1	146.7	145.3
25	14	149.0	147.7	150.2	149.7	149.3
25	15	150.0	148.6	151.3	152.0	152.3
50	2	86.0	84.9	87.2	86.0	85.9
50	3	91.0	90.04	92.0	91.0	92.5
50	4	98.0	97.16	98.8	98.0	99.1
50	5	105.0	104.1	105.8	106.0	105.5
50	6	111.0	110.1	111.8	112.0	111.8
50	7	117.0	116.2	117.8	118.0	118.0
50	8	122.0	121.1	122.8	125.0	123.9
50	9	126.5	125.6	127.3	128.0	129.7
50	10	132.0	131.1	132.8	133.0	135.1
50	11	138.0	137.0	138.9	140.0	140.6
50	12	145.0	143.9	146.0	148.0	146.2
50	13	150.0	148.8	151.1	152.0	151.0
50	14	154.0	152.7	155.2	154.5	154.7
50	15	155.0	153.7	156.2	157.0	157.2
75	2	89.0	87.9	90.1	90.0	90.4
75	3	95.0	94.0	96.0	97.0	96.7
75	4	102.5	101.6	103.3	103.0	103.1
75	5	109.0	108.1	109.8	110.0	109.6
75	6	114.0	113.1	114.8	116.0	116.1
75	7	120.0	119.2	120.7	122.0	122.5
75	8	126.0	125.1	126.8	130.0	128.8
75	9	131.0	130.1	131.8	134.0	134.8
75	10	136.0	135.1	136.8	137.5	140.5
75	11	144.0	142.9	145.0	145.2	145.9
75	12	150.0	148.7	151.2	152.0	151.5
75	13	154.0	152.8	155.1	157.0	156.3
75	14	158.0	156.7	159.2	160.2	159.7
75	15	160.0	158.7	161.2	161.0	161.6
90	2	93.0	92.0	94.1	94.7	94.3
90	3	99.0	98.0	100.0	99.0	100.4
90	4	106.2	105.2	107.1	108.0	106.8
90	5	113.0	112.0	113.9	115.0	113.5
90	6	119.0	117.9	120.0	120.0	120.2
90	7	124.0	123.1	124.8	126.8	126.9
90	8	131.0	129.8	132.1	136.0	133.4
90	9	135.0	133.9	136.0	138.0	139.7
90	10	141.0	139.9	142.0	142.0	145.6
90	11	149.0	147.4	150.5	152.0	151.1
90	12	154.0	152.5	155.4	156.8	155.9
90	13	159.0	157.5	160.4	161.0	160.5
90	14	162.0	160.7	163.2	164.0	163.9
90	15	165.0	163.6	166.3	165.0	166.0
97	2	96.0	95.1	96.9	97.5	98.5
97	3	103.0	101.8	104.1	104.0	104.4
97	4	111.0	109.5	112.4	111.0	110.9
97	5	117.0	115.7	118.2	118.0	117.6
97	6	124.0	122.3	125.7	125.2	124.6
97	7	129.0	127.7	130.2	135.0	131.6
97	8	137.0	134.2	139.7	140.1	138.5
97	9	140.0	138.2	141.7	140.0	145.1
97	10	147.0	144.5	149.4	148.5	151.3
97	11	156.0	150.7	161.2	160.0	156.8
97	12	159.7	156.9	162.4	160.0	161.7
97	13	164.0	161.9	166.0	165.5	166.3
97	14	165.0	163.9	166.0	170.0	169.2
97	15	168.0	166.8	169.1	170.6	170.8

Table C.3 Comparison of the centiles of height of boys in urban Iran and urban Tehran

Percent	Age	Centile urban Iran	[95% Confidence Interval] urban Iran		Centile urban Tehran	Smoothed urban Tehran	Percent	Age	Centile urban Iran	[95% Confidence Interval] urban Iran		Centile urban Tehran	Smoothed urban Tehran
3	2	68.0	62.9	73.0	71.8	70.1	50	9	128.0	127.1	128.8	129.5	128.2
3	3	72.0	61.0	82.9	78.6	79.9	50	10	133.0	132.1	133.8	134.5	133.6
3	4	85.0	83.2	86.8	85.0	87.9	50	11	137.0	136.0	137.9	138.0	139.2
3	5	93.3	92.0	94.6	94.7	94.6	50	12	141.0	139.9	142.0	142.0	145.0
3	6	100.0	99.0	101.0	100.1	100.3	50	13	148.0	146.8	149.2	149.0	150.9
3	7	104.0	102.8	105.1	105.7	105.3	50	14	155.0	153.7	156.3	158.5	157.0
3	8	110.0	108.8	111.1	111.6	109.8	50	15	160.0	158.5	161.5	165.0	163.5
3	9	115.0	113.7	116.2	115.0	114.2	75	2	89.0	88.0	90.0	91.0	91.5
3	10	119.0	117.5	120.4	120.5	118.5	75	3	97.0	96.0	98.0	98.0	98.8
3	11	120.0	116.9	123.0	120.1	122.9	75	4	103.0	102.1	103.8	105.0	105.2
3	12	127.0	124.9	129.0	127.0	127.4	75	5	110.0	109.1	110.8	113.0	111.0
3	13	131.5	128.9	134.0	131.2	132.7	75	6	116.0	115.1	116.8	115.7	116.6
3	14	136.8	133.9	139.8	135.0	138.6	75	7	120.0	119.2	120.7	122.0	122.1
3	15	144.0	141.7	146.3	145.0	144.3	75	8	126.0	125.1	126.8	129.0	127.6
10	2	79.0	78.1	79.9	80.0	76.6	75	9	132.0	131.0	132.9	134.0	133.3
10	3	84.0	83.1	84.9	83.0	85.7	75	10	137.0	136.0	137.9	140.0	139.2
10	4	90.0	88.9	91.1	90.8	93.1	75	11	142.0	140.9	143.0	145.0	145.3
10	5	98.0	97.1	98.9	100.0	99.5	75	12	148.0	146.8	149.1	149.5	151.3
10	6	103.0	102.0	103.9	104.0	104.9	75	13	153.0	151.7	154.2	155.0	157.5
10	7	108.0	107.0	108.9	110.0	109.9	75	14	162.0	160.4	163.5	166.0	163.7
10	8	113.0	111.9	114.0	117.0	114.5	75	15	169.0	167.1	170.8	171.0	169.5
10	9	119.0	117.9	120.0	120.0	119.1	90	2	93.0	92.1	94.0	95.0	95.3
10	10	123.0	121.8	124.1	124.0	123.7	90	3	100.0	99.1	100.9	102.0	102.3
10	11	127.0	125.7	128.2	128.0	128.3	90	4	107.0	106.0	107.9	108.0	108.6
10	12	132.0	130.5	133.4	133.0	133.3	90	5	115.0	113.9	116.1	117.8	114.5
10	13	137.0	135.3	138.6	139.0	138.7	90	6	119.0	118.1	119.8	119.3	120.4
10	14	141.7	139.5	143.8	143.3	144.5	90	7	125.0	124.0	125.9	126.0	126.3
10	15	147.0	144.5	149.4	150.0	150.7	90	8	130.0	129.0	130.9	132.0	132.4
25	2	83.0	82.1	84.0	84.0	82.1	90	9	136.0	134.9	137.0	137.0	138.6
25	3	88.0	87.1	88.9	89.5	90.4	90	10	141.0	140.0	142.0	146.0	145.1
25	4	95.0	94.2	95.8	96.0	97.5	90	11	147.0	145.6	148.3	149.8	151.6
25	5	102.0	101.2	102.8	103.5	103.5	90	12	153.0	151.5	154.4	157.0	158.0
25	6	107.0	106.1	107.8	108.0	108.9	90	13	160.2	158.3	162.0	166.0	164.2
25	7	112.0	111.2	112.8	115.0	113.9	90	14	168.0	165.7	170.2	170.3	170.2
25	8	118.0	117.1	118.8	120.0	118.7	90	15	173.0	170.9	175.0	173.8	175.4
25	9	123.0	122.0	123.9	124.0	123.5	97	2	98.0	96.8	99.2	98.4	99.3
25	10	128.0	127.0	128.9	129.7	128.4	97	3	104.0	102.8	105.1	105.7	106.0
25	11	131.0	129.9	132.0	131.0	133.6	97	4	112.0	110.5	113.4	112.4	112.3
25	12	136.0	134.7	137.2	138.0	138.8	97	5	120.0	118.0	121.9	122.3	118.6
25	13	142.0	140.6	143.3	142.0	144.6	97	6	122.0	121.1	122.8	123.7	124.8
25	14	147.0	145.3	148.6	150.0	150.8	97	7	129.3	127.9	130.8	131.0	131.3
25	15	153.0	151.2	154.7	158.0	157.3	97	8	136.0	134.1	137.8	138.3	138.0
50	2	86.0	85.0	87.0	87.0	87.1	97	9	140.0	138.6	141.3	140.0	144.9
50	3	92.0	91.1	92.9	94.0	94.9	97	10	146.0	144.5	147.4	152.0	152.0
50	4	100.0	99.2	100.8	101.0	101.5	97	11	153.0	150.2	155.7	157.0	159.0
50	5	106.0	105.2	106.8	108.0	107.4	97	12	159.1	156.2	161.9	172.8	165.8
50	6	112.0	111.1	112.8	112.0	112.8	97	13	167.0	162.6	171.3	172.0	171.9
50	7	116.0	115.2	116.7	117.0	117.9	97	14	175.0	169.6	180.3	176.7	177.4
50	8	122.0	121.1	122.8	124.0	123.0	97	15	176.5	174.5	178.4	176.1	181.6

Figure D.1 95% confidence intervals of **boys' weight** centiles in urban Iran and the corresponding raw centiles of weight of girls in urban Tehran. Centiles: 3rd, 10th, 25th, 50th, 75th, 75th, 90th, 97th

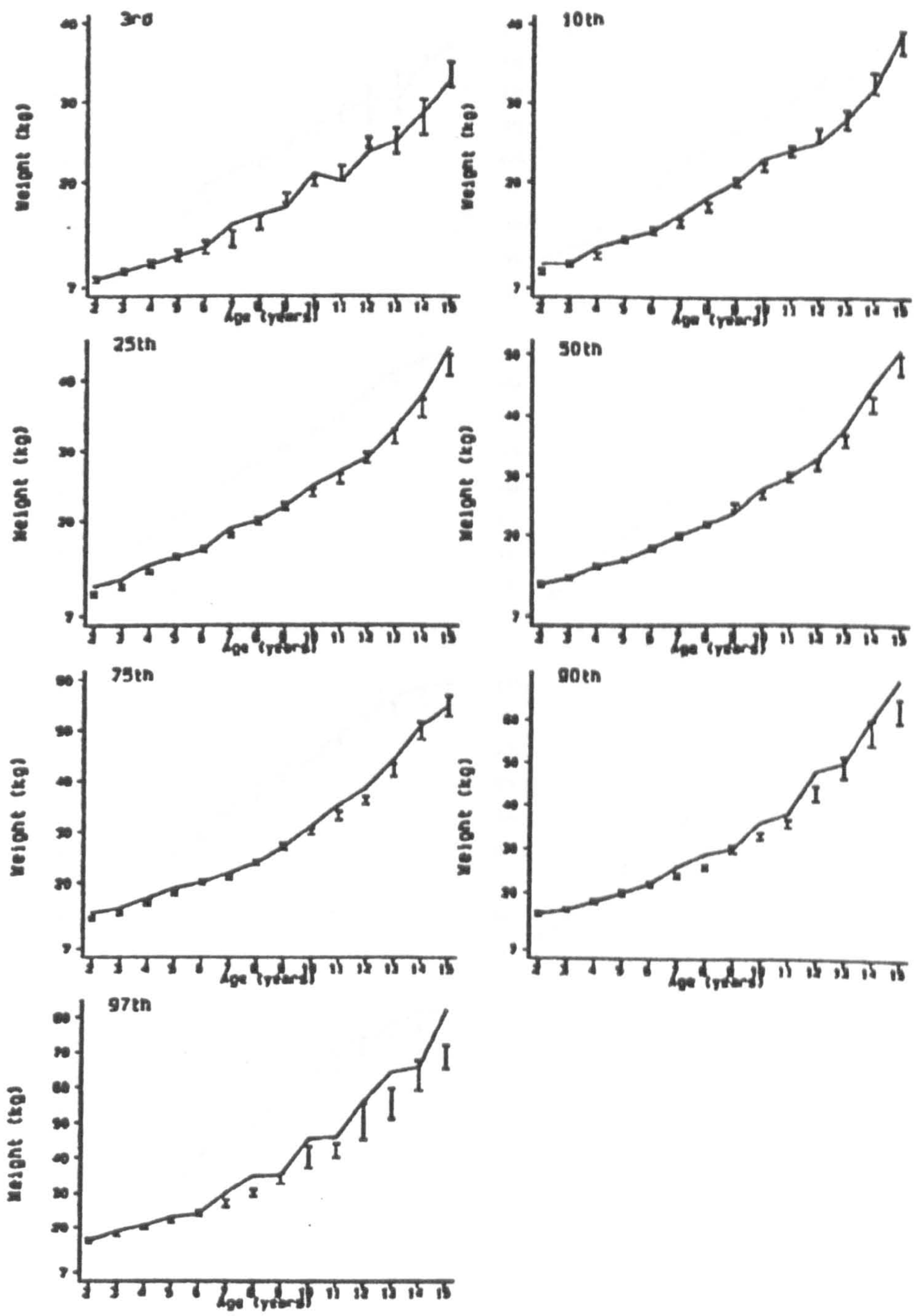


Figure D.2 95% confidence intervals of **girls' height** centiles in urban Iran and the corresponding raw centiles of weight of girls in urban Tehran. Centiles: 3rd, 10th, 25th, 50th, 75th, 75th, 90th, 97th

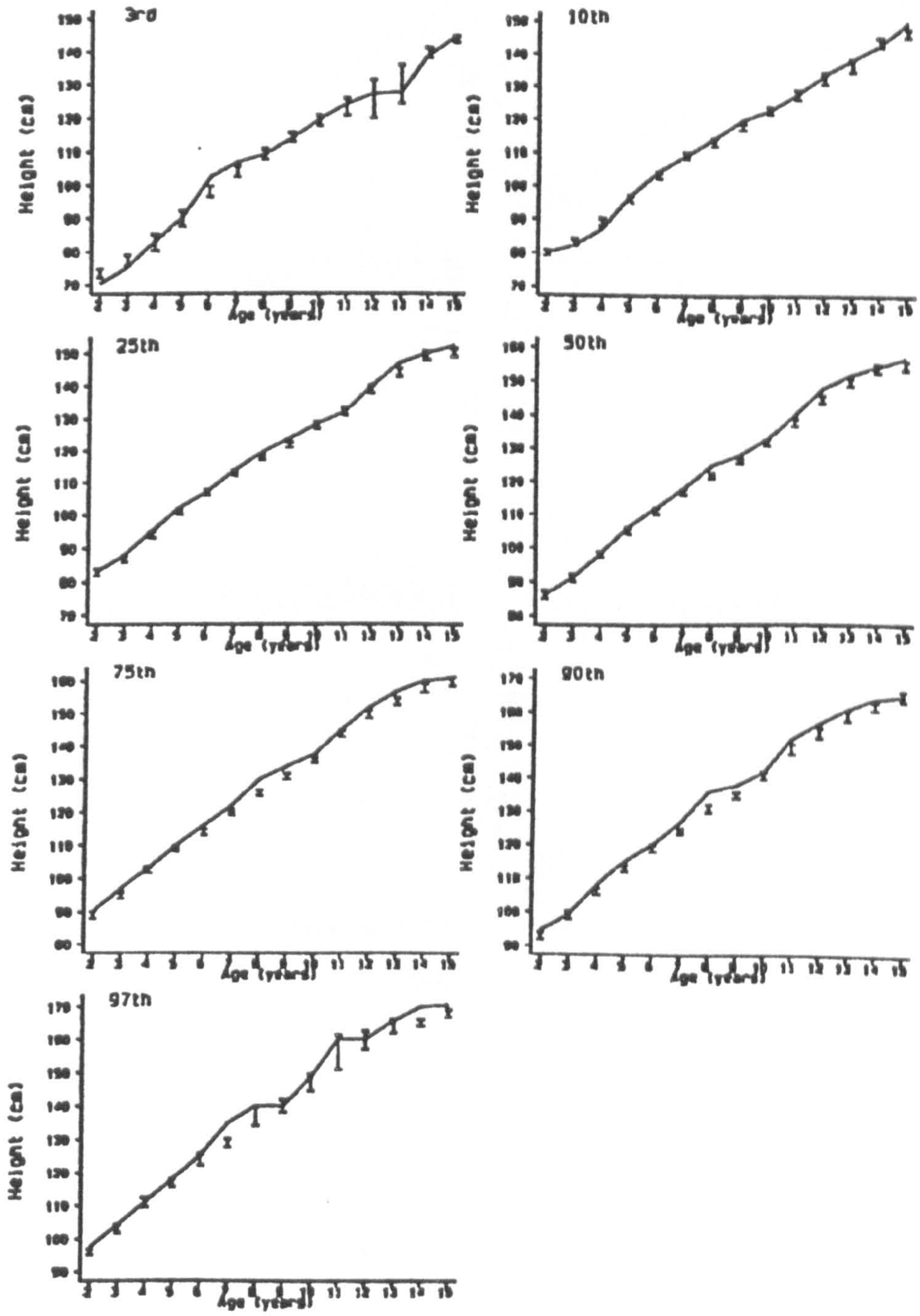
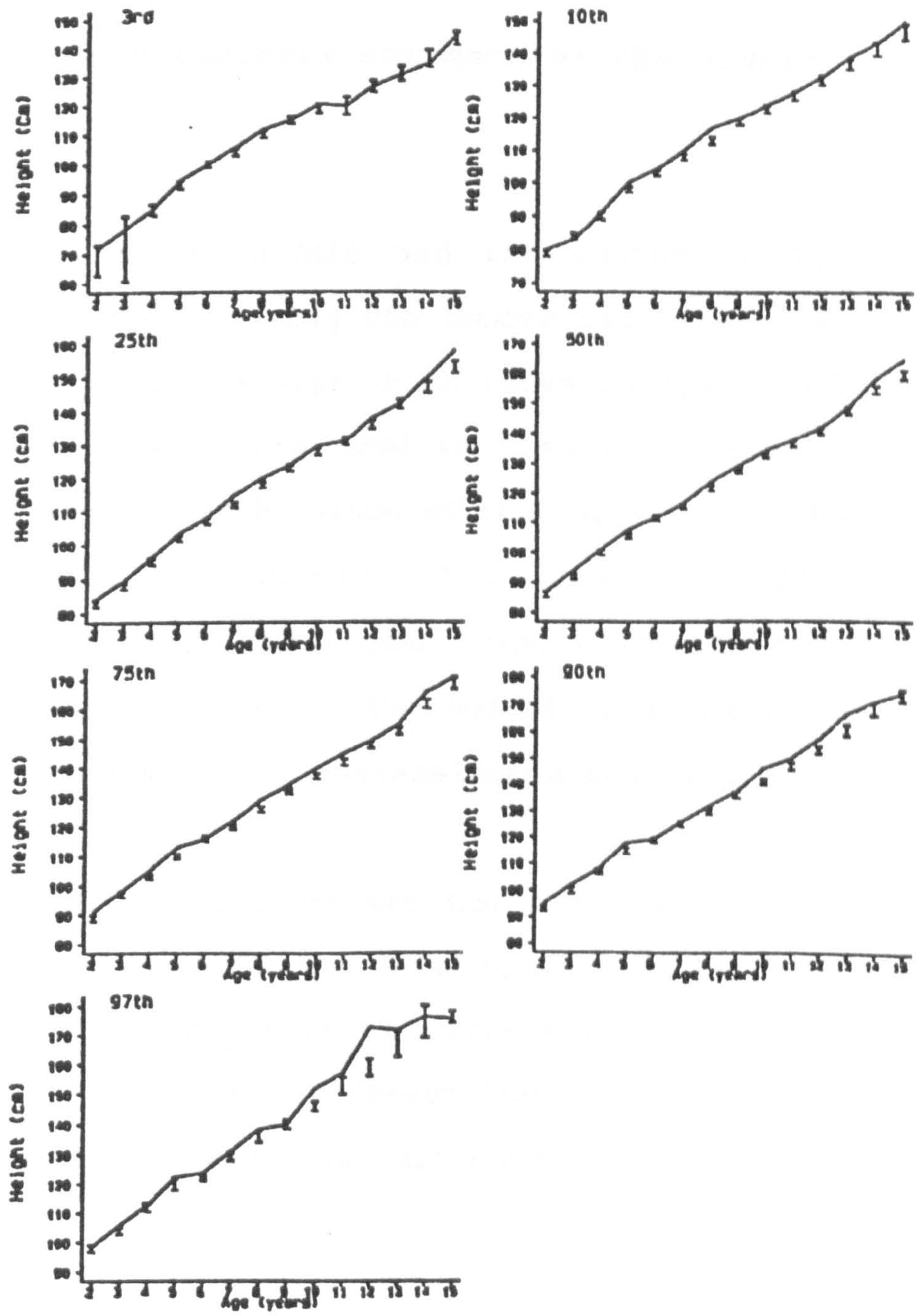


Figure D.3 95% confidence intervals of **boys' height** centiles in urban Iran and the corresponding raw centiles of weight of girls in urban Tehran. Centiles: 3rd, 10th, 25th, 50th, 75th, 75th, 90th, 97th



APPENDIX D

Parameters estimate of the models

Following the models and the corresponding parameter estimates for deriving the smooth centiles of weight-for-age and height-for-age of children in Iran are presented. The polynomials were used to obtain the charts to which they refer. In each column of the table, the order in which the set of coefficients of Z come is: first, for the intercepts, next for linear term, quadratic and so on, as explained in chapter 8. For weight measurements the models are related to the log-transformed observations.

The models presented are the basic polynomials used to construct the GROSTAT curves. Special procedures were used to modify these curves at certain points and to find the corresponding Z-scores (described in chapter 8). Table D.4 presents the models which are used for deriving Z-scores in the related ages.

D.1 Model parameter estimates for urban Tehran

	Girls	Boys
	5 3 3 2 3 1 2 *spline at 13 yr.*	*4 3 3 3 0 2* *Spline at 13 yr.*
Weight	2.2690 -2.9990E-01 1.5668E-02 1.6589E-02 3.3806E-02 3.6575E-01 -1.9894E-02 -1.0686E-03 3.2108E-02 -1.0218E-01 4.6292E-03 -5.7549E-03 1.2710E-02 -3.0729E-04 2.3663E-06 4.4274E-04 -7.0870E-04 -1.1965E-05 1.4457E-05 3.6263E-07 1.2399E-04 -1.2488E-05 -1.3352E-05	2.2999 2.1486E-01 -4.1827E-02 4.6922E-03 1.1799E-01 -2.8430E-02 8.7238E-03 1.7602E-03 -6.0874E-03 2.9707E-03 -3.0142E-04 -1.0804E-04 6.0084E-04 -1.6661E-05 -5.0669E-06 -2.7800E-07 -1.8252E-03 1.6652E-04 5.9636E-05
	3 3 1 2 3 *Spline at 13 yr.*	*4 3 3 2 0 2* *Spline at 15 yr.*
Height	72.590 8.8187 -4.4585E-01 2.0434E-01 6.6670 -1.4720 9.7238E-03 2.3897E-01 1.1852E-02 -5.0936E-03 -1.0149E-02 -7.8385E-04 3.1827E-05 -7.3242E-02 4.1419E-02 8.0736E-03 -1.3115E-03	66.606 9.6848 -8.9756E-01 2.1871E-01 12.381 -1.7696 8.3269E-02 7.6076E-03 -1.2313 1.9388E-01 1.1661E-02 8.8023E-02 -2.1706E-03 -3.5069E-04 -6.3880E-05 -7.9530E-02 5.3574E-02 2.2332E-02

D.2 Model parameter estimates for rural areas of Iran

	Girls	Boys
	4 3 3 3 1 1 *spline at 13 yr.*	*4 3 3 2 3 1* *Spline at 13 yr.*
	2.1177	2.0554
	0.24527	0.21369E-01
	-0.40719E-01	-0.10824E-01
	0.32744E-01	0.23847E-01
	0.17684	0.21150
	-0.51050E-01	0.92602E-01
	0.76289E-02	-0.46940E-02
	-0.45179E-02	-0.19492E-02
	-0.19371E-01	-0.19367E-01
Weight	0.71334E-02	-0.20223E-01
	-0.33867E-03	0.10298E-02
	0.17200E-03	0.12679E-02
	0.17907E-02	0.16962E-02
	-0.23104E-03	-0.44203E-04
	-0.52864E-04	0.12212E-05
	-0.25699E-05	-0.27350E-04
	0.32839E-04	-0.46933E-04
	0.21127E-03	-0.74436E-03
		0.69353E-03
	3 3 3 3 3 *Spline at 13 yr.*	*4 3 3 2 1 1* *Spline at 15 yr.*
	68.729	64.438
	8.7228	1.8762
	-1.0082	-1.1161
	0.69026	0.54415
	7.1462	11.108
	-2.1782	2.8629
	0.30482	0.23715
	0.10578	-0.14566E-01
	-0.12831	-0.87835
Height	0.36236	-0.66853
	-0.25419E-01	-0.11421E-01
	-0.33436E-01	0.50820E-01
	0.13459E-02	0.60175E-01
	-0.15106E-01	-0.99331E-03
	0.51021E-03	-0.17403E-02
	0.16398E-02	-0.48180E-01
	-0.85031E-01	0.41743E-01
	0.61335E-01	
	0.93810E-02	
	-0.10437E-01	

D.3 Model parameter estimates for urban Tehran from which the Z-scores derived after spline procedures

	Girls	Boys
	2 3 3 2 *14 to 18 yr.*	*2 3 3 2* *16 to 18 yr.*
Weight	-0.54376 1.3924 0.42391E-01 0.56135E-02 0.51940 -0.13835 -0.76728E-02 0.18903E-03 -0.14849E-01 0.38800E-02 0.31377E-03	-4.4925 2.1945 0.49889 0.57613E-02 0.96388 -0.22304 -0.58750E-01 -0.63685E-04 -0.27074E-01 0.61277E-02 0.17307E-02
	2 3 1 1 *14 to 18 yr.*	*2 3 3 2* *16 to 18 yr.*
Height	8.5273 35.151 -0.30046 0.30953 17.344 -3.3313 -0.49727 0.95074E-01	-154.11 28.383 18.708 0.19158E-01 36.697 -2.0848 -2.1779 0.12870E-01 -1.0344 0.53777E-01 0.61610E-01

The above models for the corresponding age range can be applied to weight and height measurements to obtain the fitted values or deriving Z-scores

D.4 Comparison of centiles of height of urban Tehran with rural areas of Iran by sex, National Health Survey 1990-2

Sex	urban Tehran centiles	Rural areas of Iran 2-13 (yr.) percent* (SD)	14-18 (yr.) percent (SD)	2-18 (yr.) percent (SD)
Girls	1	3.2 (0.5)	2.8 (0.2)	2.8 (0.3)
	3	10.4 (1.7)	7.9 (1.7)	8.3 (1.9)
	10	29.8 (3.8)	23.5 (5.0)	24.6 (5.4)
	25	55.5 (4.3)	50.2 (5.4)	51.0 (5.6)
	50	78.4 (3.3)	78.0 (2.2)	78.0 (2.4)
	75	91.1 (2.0)	91.7 (1.2)	91.6 (1.3)
	90	96.9 (0.9)	96.8 (0.9)	96.8 (0.9)
	97	99.1 (0.3)	98.8 (0.7)	98.8 (0.6)
Boys	1	3.5 (1.0)	3.8 (1.3)	3.7 (1.2)
	3	8.8 (2.4)	9.2 (2.7)	9.1 (2.6)
	10	22.7 (6.0)	24.4 (4.6)	24.1 (4.7)
	25	48.3 (10)	49.7 (5.2)	49.5 (5.9)
	50	73.7 (8.0)	76.4 (2.9)	75.9 (4.0)
	75	88.9 (4.8)	91.1 (2.1)	90.7 (2.7)
	90	95.6 (2.3)	96.7 (1.3)	96.5 (1.5)
	97	98.4 (0.8)	98.9 (0.7)	98.8 (0.7)

* : The corresponding centiles in rural areas' models

Table D.5 Observed and expected number of boys' weight and height measurements falling between urban Tehran standards centiles in subgroups of age, for provinces Semnan and Kohkiloyeh-Boyerahmad

Centile	Semnan urban boys				Kohkiloyeh-Boyerahmad rural boys			
	Weight		Height		Weight		Height‡	
	O	E	O	E	O	E	O	E
>97	0	1.9	1	1.9	0	3	1	3
90-97	1	4.3	6	4.3	0	7	4	7
75-90	16	9.3	12	9.3	7	15	7	15
50-75	19	15.5	20	15.5	17	25.5	24	26
25-50	16	15.5	15	15.5	27	25.5	22	25
10-25	9	9.3	5	9.3	26	15	24	15
3-10	1	4.3	3	4.3	16	7	11	7
≤3	0	1.9	0	1.9	8	3	8	3
χ^2_5†, P-value	14.40	p=0.01	10.04	p=0.07	44.6	$p<0.001$	20.8	$p<0.001$

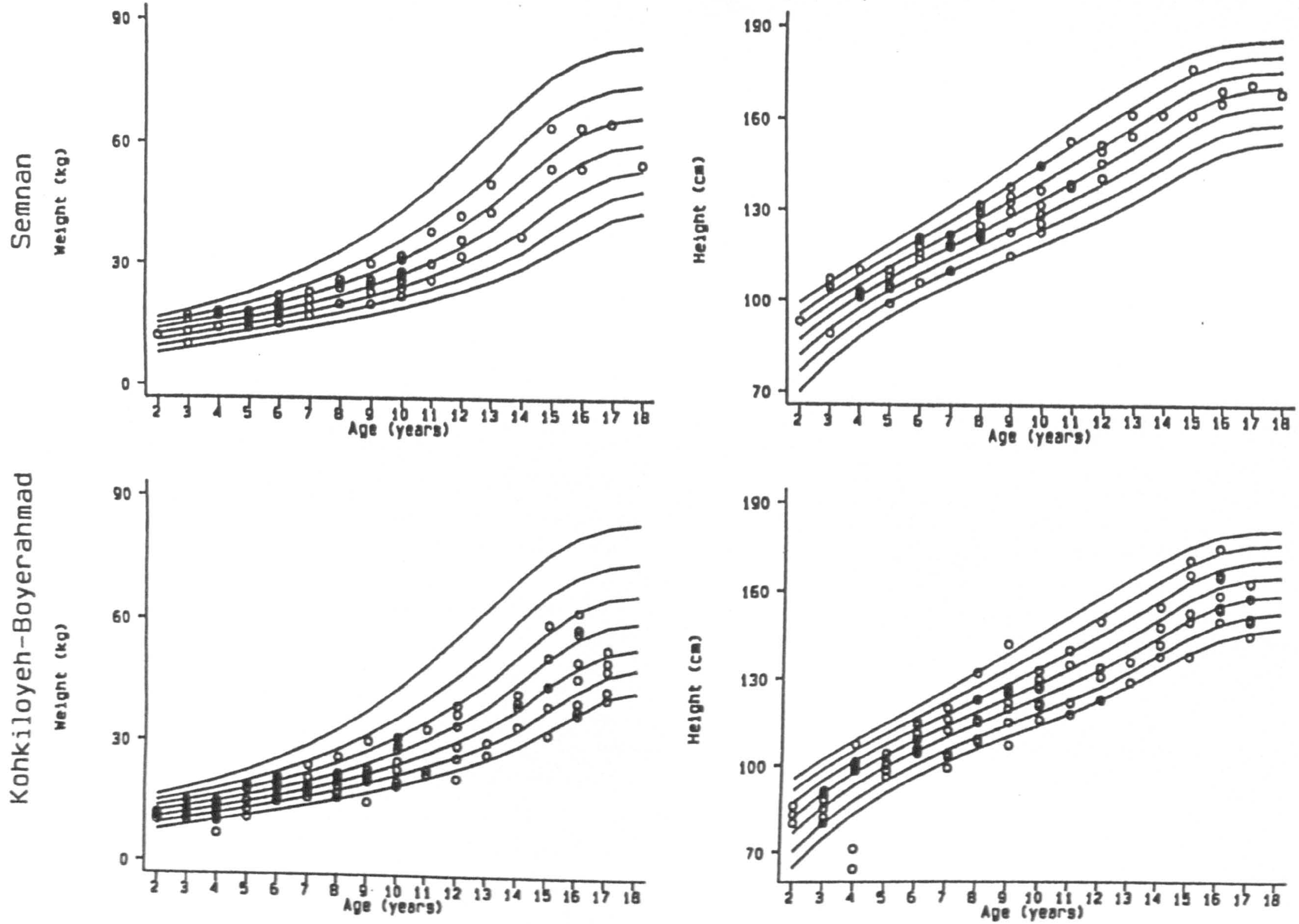
O : Observed

E : Expected

†) Grouping first two and last two centiles in each column to correct for small expectation

‡) Height of children from rural Kohkiloyeh-Boyerahmad compared with the corresponding average centiles shown on boys' height standard (Figure 8.15)

Figure D.1 Standard Growth charts for Iran with measurements of weight and height of children for urban Semnan and rural Kohkiloyeh-Boyerahmad superimposed (seven centiles* are shown 3rd, 10th,...,97th)



*) Height of rural Kohkiloyeh-Boyerahmad children compared with the corresponding average centiles shown on boys' height standard (Figure 8.15)