Journal of Applied Statistics

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cjas20

Identification of outlying height and weight data in the Iranian National Health Survey 1990-92

M. Hosseini , R. G. Carpenter & K. Mohammad

Published online: 02 Aug 2010.

To cite this article: M. Hosseini , R. G. Carpenter & K. Mohammad (1998) Identification of outlying height and weight data in the Iranian National Health Survey 1990-92, Journal of Applied Statistics, 25:5, 601-612, DOI: 10.1080/02664769822855

To link to this article: http://dx.doi.org/10.1080/02664769822855

Journal of Applied Statistics, Vol. 25, No. 5, 1998, 601–612

Identification of outlying height and weight data in the Iranian National Health Survey 1990–92

M. HOSSEINI[1,2], R. G. CARPENTER[2] & K. MOHAMMAD[1], [1]*Department of Epidemiology and Biostatistics, Tehran University of Medical Sciences, Iran, and* [2]*Medical Statistics Unit, London School of Hygiene and Tropical Medicine, UK*

SUMMARY *Data on the weights and heights of children 2–18 yeas old in Iran were obtained in a National Health Survey of 10 660 families in 1990–92. Data were 'cleaned' in 1 year age groups. After excluding gross outliers by inspection of bivariate scatter plots, Box–Cox power transformations were used to normalize the distributions of height and weight. If a multivariate Box–Cox power transformation to normality exists, then it is equivalent to normalizing the data variable by variable. After excluding gross outliers, exclusions based on the Mahalanobis distance were almost identical to those identified by Hadi's iterative procedure, because the percentages of outliers were small. In all, 1% of the observations were gross outliers and a further 0.4% were identified by multivariate analysis. Review of records showed that the outliers identified by multivariate analysis resulted from data-processing errors. After transformation and 'cleaning', the data quality was excellent and suitable for the construction of growth charts.*

1 Introduction

This paper describes the 'cleaning' of a random cluster sample of measurements of the weight and height of about 25 000 children aged 2–18 years, recorded by the National Health Survey of Iran, in preparation for the construction of the first growth charts specifically for Iranian children. For such data, inspection of the frequency distributions and bivariate plots of data for each age group revealed gross outliers. For univariate normally distributed observations, the standardized distance of observations from the mean is the traditional method of detecting

Correspondence: R. G. Carpenter, Medical Statistics Unit, London School of Hygiene and Tropical Medicine, University of London, Keppel Street, London WC1E 7HT, UK. Tel: 0171 927 2259; Fax: 0171 637 2853.

0266-4763/98/050601-12 $7.00

outliers. The method is satisfactory for large samples that include only a very small proportion of outliers. For multivariate normal data, the comparable statistic is the Mahalanobis distance (MD) of observations from the sample mean. The formal justification of the use of this statistic was given by Barnett and Lewis (1994) and is briefly summarized here.

In the detection of outliers, the most important problems are those of masking and swamping, which occur when the estimates of the MD are significantly affected by the outliers, so that some outliers appear to be genuine and some genuine values appear to be outliers. These difficulties arise because estimates of the mean and variance are not robust to large outliers.

One way to avoid such difficulties is to use more robust estimators of location and covariance. Several estimators of this sort have been suggested (see, for example, Donoho, 1982; Hampel *et al.*, 1986). Campbell (1980) discusses robust procedures in multivariate analysis and robust covariance estimation. He suggests a robust weighted MD for the 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 observations included in the MVE are robust location and covariance matrix estimators.

Hadi (1994) proposes an iterative modification of the method of Campbell (1980) for the identification of multiple outliers in multivariate normal data, which he claims is effective in dealing with masking and swamping problems. His procedure for n observations on p variates can be summarized as follows.

(1) First, the n observations are ordered using the MD and the data set is then divided into two initial subsets: a 'basic' subset, which contains the first $p+1$ 'good' observations; and a 'non-basic' subset, which contains the remaining $n-p-1$ observations.
(2) Second, using the mean and variance of the basic data, recompute the distance of each observation from the centre of the basic data.
(3) Third, rearrange the n observations in ascending order of the revised distance, and then divide the data set into two subsets: a revised basic subset, which contains the first $p+2$ observations, and a non-basic subset, which contains the remaining $n-p-2$ observations.

Repeat steps (2) and (3) until an appropriate chosen stopping criterion is met. The final non-basic subset of observations is declared an outlying subset. Modifications may be made to Hadi's method, by using the median instead of the mean, or using iteration until convergence within the third step. STATA (1997) includes a routine that implements Hadi's procedure (Gould & Hadi, 1993).

Penny (1995) uses Monte Carlo methods to compare the precision of different methods in the identification of planted outliers with various forms of slippage (slippage from the mean or slippage from the variance). After looking at varying percentages of outliers in several dimensions, Penny concludes that the MD method and Hadi's (1994) method were the most promising in symmetric situations—especially Hadi's method for low-dimensional data.

This paper reports the results of applying standard methods to the Iranian data set after normalizing transformations, and compares these results with the results derived from the STATA routine of Gould and Hadi (1993), which later became available. Also, a brief account of the results of comparing the analytically detected outliers against the original records is included.

2 Material

The National Health Survey of Iran was carried out between August 1990 and June 1992, and comprised a random cluster sampling of one-in-1000 families in each of the 24 provinces of Iran. Data included measurements of the weight, to the nearest kilogramme, and height, to the nearest centimetre, for children aged 2–18 years. Age was derived from date of birth recorded on identification cards and was recorded in years. It is shown elsewhere (Hosseini, 1997) that there are significant differences between the sexes especially in the older age groups; that rural children are generally smaller than urban children; and that there are significant differences between the provinces in the patterns of growth. In addition, there are significant components of variance associated with the clusters and families within the clusters. However, compared with the overall standard deviation of the observations, these sources of variations are small and were not known when the data were cleaned.

3 Transformations to normality

Weight for age generally has a skew distribution, while heights are usually approximately normally distributed. After a normalizing transformation, kurtosis is seldom a problem with growth data (Cole, 1997). Both weight and height data for each age group were normalized using the Box and Cox (1964) power transformation defined by

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

which is continuous in λ for $x > 0$. Given the observations $x_1, x_2, \ldots, x_n$, the appropriate power λ is the one that maximizes the expression

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

where $\bar{x}^{(\lambda)}$ is the mean of the transformed observations. A routine is available in STATA based on the Fisher ML scoring method, for the derivation of λ (Royston, 1992) and the interval of support that corresponds to the 95% confidence interval for λ (Clayton & Hills, 1993).

In general, however, normal marginals are not sufficient to ensure that the joint distribution of the transformed observations is bivariate normal, although it may be good enough in practical applications. Johnson and Wichern (1988) have suggested that we could start with the values $\hat{\lambda}_1$, $\hat{\lambda}_2$ obtained from normalizing the marginal distributions of height and weight, and then iterate toward the set of values $\lambda^{\mathrm{T}} = (\lambda_1, \lambda_2)$, which collectively maximizes

$$l(\lambda_1, \lambda_2) = -\frac{n}{2}\ln\left|\mathbf{s}(\lambda)\right| + (\lambda_1 - 1)\sum_{j=1}^{n}\ln x_{1j} + (\lambda_2 - 1)\sum_{j=1}^{n}\ln x_{2j} \qquad (2)$$

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

$$\mathbf{x}_j^{(\lambda)} = \begin{pmatrix} \dfrac{x_{1j}^{\lambda_1} - 1}{\lambda_1} \\ \dfrac{x_{2j}^{\lambda_2} - 1}{\lambda_2} \end{pmatrix}, \qquad j = 1, 2, \ldots, n$$

The method is equivalent to maximizing a multivariate likelihood over μ, $\mathbf{\Sigma}$ and λ. A MINITAB (1991) macro was written to seek the maximum of $l(\lambda_1, \lambda_2)$ in the neighbourhood of the values of λ_1 and λ_2 obtained from the transformations to marginal normality.

4. Testing for outliers

The null hypothesis is that $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is an independent random sample of n observations on p variables, where $\mathbf{X} \sim \mathrm{MN}(\mu, V)$. The alternative models considered have $\mathbf{x}_i$ as an outlier from a population (I) $\mathrm{MN}(\mu + \tau, V)$ or (II) $\mathrm{MN}(\mu, kV)$ for some i.

It can be shown (Barnett & Lewis, 1994) that, regardless of whether μ and/or V are known or unknown, it is immaterial whether we adopt the model I or model II formulation of the alternative hypothesis that describes the occurrence of a single outlier. In either case, the test reduces to testing whether or not the largest MD D^2 for the data set is significantly large. For $p = 2$, Barnett and Lewis (1994) have tabulated critical values for D^2 for n between 5 and 500.

If μ and V were known, then the corresponding $D^2(\mu, V)$ would be independent χ_p^2 variates. We then have to relate the observed maximum to the distribution of the maximum observation in a random sample of size n from a χ_p^2 distribution. Specifically, in the case of a bivariate sample ($p = 2$), $D^2(\mu, V)/2$ has the distribution of the maximum of n independent exponential variates (mean 1). For a level α-test, we find that

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

giving

$$\xi_\alpha = -2 \ln [1 - (1 - \alpha)^{1/n}] \tag{3}$$

Since Hadi's method, as implemented in STATA, also uses the MD, ξ_α is the appropriate large-sample criterion for this method also.

5 Results

Table 1 shows the number of children in the survey, the number of children for whom both height and weight were recorded, and the percentages of outliers in each age group. The reduction in numbers in the older age groups is partly because fewer children were born 20 years ago, and partly because a large proportion of boys and girls over the age of 16 years old were not weighed and measured.

5.1 Trimming the data

By intuitive consideration of scatter diagrams of the heights and weights for each age group, outliers were removed. For instance, the measurements with extreme

TABLE 1. Number of subjects, gross outliers (trimmed observations) and identified bivariate outliers of measurements of height and weight, from 1990–92 National Health Survey of Iran

Age (years)	Total no. in survey	No. with both measurements	Exclusion: Trimmed data No.	Trimmed data %	Bivariate outliers No.	Bivariate outliers %	No. used in analysis
2	1301	1190	14	1.2	6	0.5	1170
3	1678	1580	14	1.5	8	0.5	1558
4	1784	1680	9	0.5	5	0.3	1666
5	1814	1708	9	0.5	12	0.7	1687
6	1674	1561	7	0.4	16	1.0	1538
7	1849	1703	6	0.4	12	0.7	1685
8	1761	1618	7	0.4	8	0.5	1603
9	1615	1486	8	0.5	2	0.1	1476
10	1652	1517	6	0.4	8	0.5	1503
11	1596	1447	13	0.9	5	0.3	1429
12	1500	1330	19	1.4	5	0.4	1306
13	1390	1228	21	1.7	3	0.2	1204
14	1229	1084	23	2.1	2	0.2	1059
15	1231	1042	19	1.8	0	0.0	1023
16	1147	951	17	1.8	3	0.3	931
17	1068	841	9	1.1	1	0.1	831
18	1076	690	9	1.3	1	0.1	680
Total	25 365	22 656	210	1.0	97	0.4	22 349

marginal values such as 6-year-old individual A in Fig. 1 with a height of 182 cm, or individual B, whose weight is recorded as 91 kg are obvious outliers. The height recorded as 182 cm could have been a height of 82 cm that, during the data entry, was recorded by mistake as 182, possibly because a pen mark, in the left-hand box of three prespecified boxes for recording height, was read as '1'. Table 1 presents the percentage of trimmed data by age. From now on, 'data' refers to 'trimmed data'.

5.2 Normalizing the data

Our methodology depends on the assumption of multivariate normality. Table 2 presents the values of λ for the Box–Cox transformations of the marginal distributions of weight and height to normality. Support intervals, corresponding to the 95% confidence interval for λ, are also shown. To take account of the difference in the pattern of growth of boys and girls at puberty, from the age of 11 years upwards, the analysis was carried out separately for the two sexes (Table 2).

The powers λ_1 and λ_2 to transform the data to bivariate normality obtained by minimizing equation (2) were close to the values of λ_1 and λ_2 given in Table 2. For example, for children aged 2 years, for bivariate normality, $(\lambda_1, \lambda_2) = (0.40, 2.15)$, which compares with (0.40, 2.00) given in Table 2. Similarly, those for 3 and 4 years old, the optimum power transformations for bivariate normality are estimated to be (0.20, 2.05) and (0.50, 2.70) compared with values of (0.15, 2.00) and (0.50, 2.45), respectively, for marginal normality. There was little increase in $-2l$ in moving from the optimum values for marginal normality to the optimum values for bivariate normality. This was true for these ages groups and for some others that were also tried; in every case, the optimum values of λ for bivariate normality

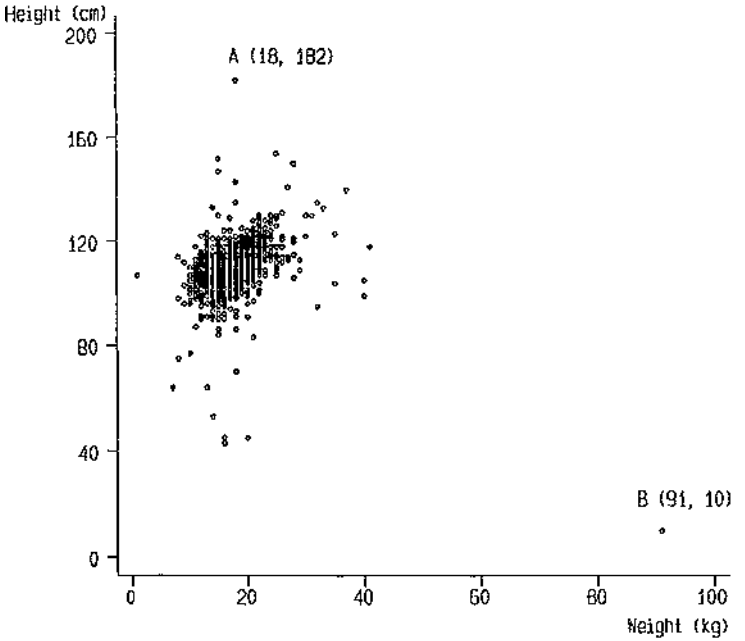

FIG. 1. Scatter plot of height against weight of 6-year-old children.

were well within the corresponding regions of support for λ_1 and λ_2 presented in Table 2.

5.3 Comparison of transformations for the identification of outliers

To investigate whether or not marginal normality is sufficient for excluding the outliers in our study, we compared the outcomes of using both forms of transformation in some age groups of children. For example, for the 4-year-old children, observations were declared outliers when $D^2 > \xi_\alpha$, with $\alpha = 0.05$. For this age group, $n = 1671$, so that $\xi_\alpha = 20.78$. Figure 2 and Table 3 present the results of this comparison. Figure 2 shows Z-scores of the measurements in both forms of transformation.

In Fig. 2, the measurements $M_1, M_2, \ldots, M_5$ are identified as outliers using both forms of transformation (Table 3). Observation N was the only pair of measurements that was not identified as an outlier when using the transformation to marginal normality, and is borderline when the data are transformed to bivariate normality. Thus, as one can see, five out of six of the outliers were identified using the results of marginal normal transformations.

For 2-year-old children, one borderline outlier after marginal transformation appeared not to be an outlier after bivariate transformation; for the 3 year olds, the same eight outliers were identified whether $\lambda = (0.20, 2.05)$ or $(0.15, 2.00)$ was used. Similar results were observed when this comparison was made for the other age groups. The small differences that were found related to datum points that were borderline in comparison with the critical values, and it was concluded that the transformation to marginal normality was adequate.

On the basis of this finding, we identified the outlying observations in the joint

TABLE 2. Estimated λ for the Box–Cox power transformation to normality and corresponding support intervals, of weight and height by age and sex

Area (years)	Weight			Height		
	Both sexes	Boys	Girls	Both sexes	Boys	Girls
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[a] (−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)		
11		0.0 (−0.2, 0.4)	−0.2 (−0.6, δ0.1)		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, 0.9)	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)

Note: [a]If $\lambda = 0$, then $\ln(x)$ is the transformation.

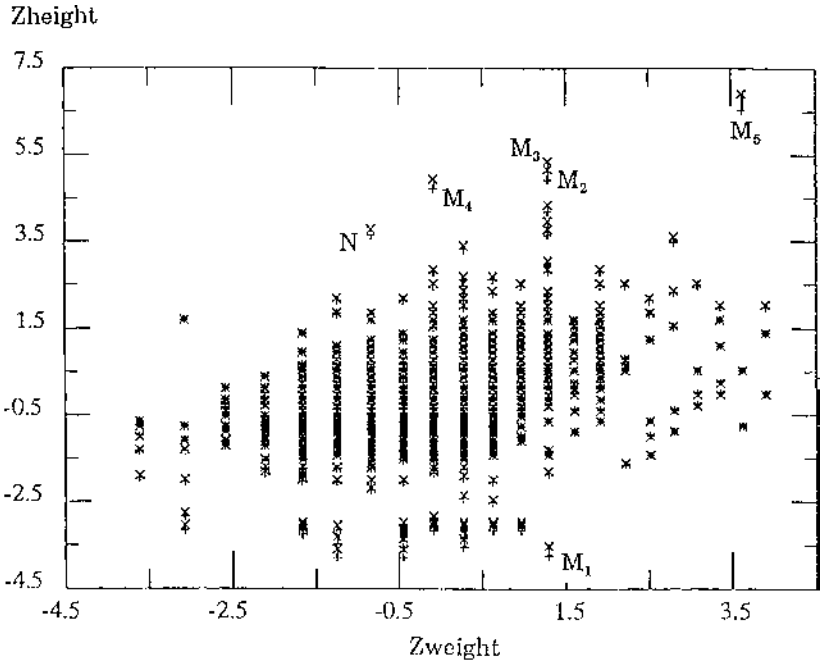

FIG. 2. Scatter plot of Z-scores of measurements of weight and height after both forms of transformation: +, joint Z-scores of measurements after marginal normal transformation $\lambda^T = (0.5, 2.45)$; ×, the Z-scores after bivariate normal transformation $\lambda^T = (0.5, 2.7)$.

TABLE 3. D^2 for outliers after two forms of transformation of weight and height of 1671 four-year-old children, and for Hadi's method applied to data transformed to marginal normality

Subject label	Measurements	Marginal normal D^2	Multivariate normal D^2	Hadi's D^2
N	12 kg, 122 cm	19.80	20.88	21.47
M_1	18 kg, 59 cm	22.62	21.06	24.16
M_2	18 kg, 129 cm	25.34	27.44	27.75
M_3	18 kg, 130 cm	27.44	29.79	30.05
M_4	14 kg, 128 cm	27.55	29.67	30.09
M_5	26 kg, 137 cm	44.72	48.75	48.31

Notes: critical value $\xi_{0.05} = 20.78$; for marginal normal transformations, $\lambda^T = (0.50, 2.45)$; for bivariate normal transformation, $\lambda^T = (0.50, 2.70)$.

distributions of weight and height by age, after transformations to marginal normality of weight and height using the values of λ shown in Table 2. The number of records identified as outliers is shown in Table 1.

The effectiveness of the normalizing transformation on the data for 12-year-old girls, i.e. $\lambda^T = (-0.25, 2.5)$, is shown in Fig. 3; Fig. 4 shows the associated gamma plot of the generalized distance for these data after excluding outliers. Both figures suggest that the data have been normalized and cleaned satisfactorily.

5.4 *Hadi's method*

Hadi's method, implemented in STATA, was applied to data for 4-year-old children after transformation to marginal normality. This analysis identified all the datum

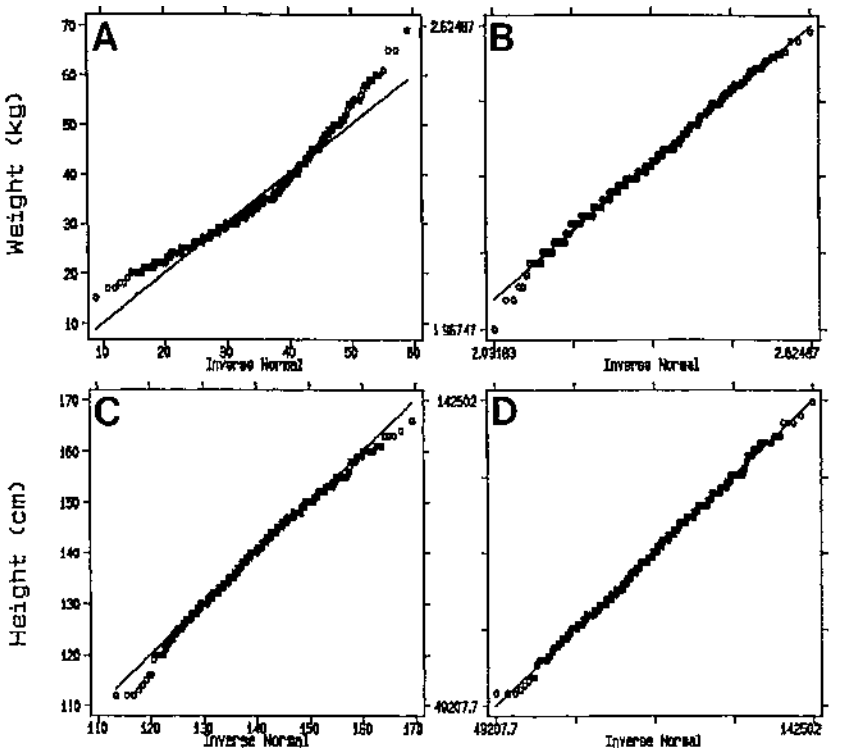

FIG. 3. Normal plots of weight and height of 12-year-old girls (a), (c) before and (b), (d) after transformation: (a) $r = 0.977$; (b) $r = 0.999$; (c) $r = 0.997$; (d) $r = 0.999$. For transformed plots, Wilks normality tests give transformed weight $p = 0.39$ and transformed height $p = 0.66$.

points M_1–M_5 as outliers. Point N in Fig. 2 was also identified as an outlier, whereas it was previously borderline (Table 3).

Hadi's analysis was similarly applied to data for other age groups. Overall, we found that, if there was a difference between the results, then it related to borderline cases.

5.5 Origin of some outliers

The data only became available for analysis some time after they had been processed, so that it was not possible to check and correct outliers before analysis. Later, however, during a visit to Tehran, one of the authors (MH) was able to check the outliers against the original records. Data had been recorded in a precoded form, with boxes for numerical observations. Some of the numerical data were in Arabic numerals but, mostly, numbers were written in Farsi script, in which '0' is similar to '5', and '2' is similar to '3'. Table 4 sets out some examples of the errors detected. All the observations detected as outliers were the result of errors during data entry. No retrospective checks on the accuracy of measurements and recording were possible.

6. Discussion

The first step in data cleaning is to eliminate gross outliers, which can be identified by tabulation or plotting the data. After that, all the more sophisticated methods

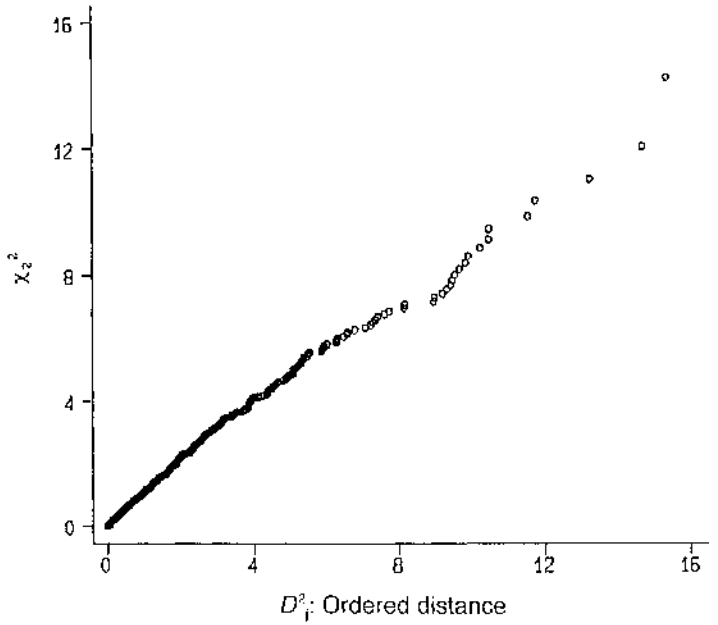

FIG. 4. Gamma plot (χ^2 plot) of ordered distances of measurements (weight and height) of 12-year-old girls after excluding outliers.

TABLE 4. Examples of data-processing errors that cause outliers

Record	Entry	Source of error
25 kg, 91 cm	2, 591	Misplaced space
9.5 kg, 74 cm	95, 75	9.5 not rounded to 10, omission of decimal place
25 kg, 91 cm	25, 61	Copying error
10 kg, 93 cm	30, 91	Digit interchange
12 kg, 40 cm	12, 90	Badly formed '4' read as '9' (Arabic numerals)
14 kg, 104 cm	14, 154	'0' and '5' are similar in Farsi script
18 kg, 104 cm	18, 154	'0' and '5' are similar in Farsi script
26 kg, 112 cm	36, 112	'2' and '3' are similar in Farsi script
19 kg, 110 cm	19, 110	At age 2 years; age error, correct age 3 years

assume that the data or the transformed data are normally distributed. Furthermore, when data are multivariate, it is important to use multivariate statistics such as the MD D^2 to detect outliers that are univariately shadowed by the distribution; in Fig. 2, observation N is an example of an outlier that would not be detected by univariate analysis.

Johnson and Wichern (1988) suggest using the multivariate version of the Box–Cox power transformation, given by maximizing equation (2), to transform data to multivariate normality when normalizing the data variable by variable fails to achieve the desired result. However, suppose that a multivariate Box–Cox power transformation to multivariate normality has been implemented. If z_i is the resulting power transform of x_i, then z_i will be normally distributed. This is because, if z is multivariate normal, then all the marginal distributions are normal. However, if $x_i^{(\lambda)}$ is normal, then it implies that $x_i^{(\lambda')}$ is not, unless $\lambda' = \lambda$. Hence, if we assume

that the data can be normalized by maximizing equation (2), then it is sufficient to normalize each variable individually using equation (1).

This explains why the values of λ_1 and λ_2 obtained from maximizing equation (2) are close to those obtained from transforming the data to marginal normality. The reason why the two procedures do not give identical results is that the likelihood surface is virtually flat in the neighbourhood of the maximum, which is illustrated by the wide intervals of support for λ_1 and λ_2 shown in Table 2. Therefore, it is not surprising that there is very little difference in the distribution of the data or in the number of outliers that are detected by the two different methods of normalizing the data. In the same vein, Johnson and Wichern (1988) comment that maximizing equation (2) is 'unlikely to yield remarkably better results' compared with normalizing the variables individually.

Table 3 shows that D^2 for the outliers, computed from all the data, gives smaller values than does the D^2 term derived from Hadi's iterative method. This is expected, because the outliers are not included when computing the mean and variance for Hadi's D^2. However, it appears that, in the present application, there are too few outliers for the difference to be of any practical importance. In general, Hadi's method is to be preferred, especially now that it is readily available in STATA. However, it should be noted that, as with the classical method, Hadi's method also requires that the data are normally distributed. Failure to transform the data can lead to serious errors. This can be demonstrated by applying the method to untransformed data, which may lead to half the data being declared to be outliers (Hosseini, 1997).

Our criteria for exclusion, i.e. ξ_α, are based on the assumption that the population mean and variance are known. This is conservative, because, for $p = 2$, ξ_α is larger than the critical values tabulated by Barnett and Lewis (1994) for $D^2_{\max}$. However, the number of children in each age group is sufficiently large for this to be of little consequence. For example, for $n = 500$, the 5% critical value for bivariate data using the estimated mean and variance is 18.12 (Barnett & Lewis, 1994) compared with 18.37 when the mean and variance are assumed to be known. Therefore, it is unnecessary to compute exact values for $D^2_{\max}$ as suggested by Fung (1996) to clean our data. Our criteria for exclusion are also conservative in that we have not adjusted ξ_α for the reduction of the sample size that results from the identification of the outliers. Again, when there are only a small proportion of outliers, the adjustment is small.

The investigation of the origin of the outliers reveals that they are typical data-entry errors, bearing in mind the similarity of some of the Farsi numerals. The number of errors appears to be small, given that sophisticated software were not used for automatic checking of data entry, such as routine comparison of double entry, as is available in EPI INFO (CDC/WHO, 1996). Plots of the data, illustrated in Figs 3 and 4, suggest that the transformation and cleaning operations were highly effective. It was also encouraging to note that the outliers that were found were the result of errors in data entry, and were not the result of suspect records. This supports the view that the data were of high quality and provide a sound basis on which to construct the required growth charts.

Acknowledgements

M. Hosseini was funded by Ministry of Health and Medical Education of Iran. M. Hosseini would like to thank Professor P. G. Smith for his support, and Dr M. Jones for his help at London School of Hygiene and Tropical Medicine.