Anthrop. Közl. *31*; 17–20 (1987/88)

GROWTH STANDARDS FOR NON-NORMALLY DISTRIBUTED VARIABLES – HEIGHT AND AGE AT MENARCHE

J. L. Boldsen

Department of Social Medicine, Institute of Community Health, Odense University, Denmark

Abstract: It appears impossible to describe the distribution of many growth-related variables by Gaussian curves. However, it is often possible to fit to them a mixture of two or three normal distributions with equal variances but different means. The parameters in the fitted distributions show great consistency over geographical and time differences. In order to decide whether to intervene in the growth of a child showing a variant growth pattern it is advantageous to know if the child occupies an extreme position at the tail of the composite distribution, or actually belongs to some other distribution.

Key words: Growth standards, Non-normally distributed variables, Height, Age at menarche.

Introduction

The most important usage of growth standards is to provide guidelines for decisions whether to intervene in the development of a given child be means of individual medical and/or nutritional treatment. It is impracticable to treat a large fraction of the population in this way, hence the application of growth standards most often implies extreme-value statistics.

There is growing evidence that most growth related physical variables follow non-normal distributions and that the shape and certain other aspects of these distributions are consistent throughout Northern Europe (Boldsen, 1983; Boldsen and Kronborg, 1984; Boldsen and Mascie-Taylor, 1985; Mascie-Taylor and Boldsen, in press).

It takes very large samples to establish standards which are reliable at the extreme ends of the spectrum from the empirical distribution of the variable. This is the main reason for the desire work out standards based on a thoroughly tested classes of parametrical distributions. The establishment of the type of the distribution and a probable parameter-space for the given variable in the specified sex and at the relevant stage of development or age must be carried out in a very large sample; but having established this frame the parameters can be estimated from far smaller samples.

It is the aim of this presentation to give and discuss examples of non-normally distributed growth-related variables in relation to problems in the establishment of standards.

Material and Methods

The discussion in this paper is based on reanalysis of three previously analyzed sets of data.

The height-distribution among 22 year old Danish males who were submitted to the examination of men liable for compulsory military service in the periode 1852 to 1856 in Viborg County in central Jutland (n=2343) is analyzed as an example of height-standards for adults. These conscription data were first reported in "Meddelelser fra det Statistiske Bureau (1859)" and they have resently been analyzed by Boldsen and Kronborg (1984) ("H22" subsequently). They found that a mixture of two normal distributions with equal variances but different means could describe these data as well as data from all other Danish counties in the same periode, i.e. height-variation in all Danish Counties in the middle of the 19th century could be described by the same class of distributions and the regional differences between the parameter estimates were limited.

Standards for height in connection with the growth spurt is illustrated by data on the 16-year old British males ("H16" subsequently) from the National Child Development Study (n=5755). H16 could neither be described by one normal distribution nor by the type of normal distribution mixtures which fitted H22. However, a hyperbolic distribution could describe these data satisfactory.

In order to evaluate the physical development of children/youth it is impossible to rely solely on continuous variables it is necessary ascertain the stage of maturity by recording whether the child has reached a given developmental stage. In order to illustrate standards for such variables the distribution of age at menarche has been analyzed. Among a large number of other questions the girls of the NCDS were at the 1974 follow up asked to state the age in years at which they had their first menstruation, 4518 did so or reported not to have had it as yet. These data ("M16" subsequently) could be described by at mixture of three normal distributions (cf. Mascie-Taylor and Boldsen, in press).

In estimating the parameters in the normal distribution mixtures a method developed by Sundberg (1972) and described for anthropometric analysis by Boldsen and Kronborg (1984) was used. The parameters in the hyperbolic distribution was estimated using a procedure and program developed and described by Jensen (1983).

Results

The classical way of constructing growth standards use the normal distribution as basis for predicting the fractiles and thus the expected frequency of more extreme observations. Hence the ratio between observed and expected frequencies of observations more extreme than a given value gives information on the appropriateness of the normal distribution approximation to the data. Table 1 gives this ratio for different empirical fractile values for each of the three variables.

It is obvious that the large ratios of observed to expected of more extreme observations occur near the limits of the distributions, most appearent at lower end of the distribution of H16.

Discussion

This paper analyzes growth standards for ordinary anthropometric variables. It is found that the traditional way of constructing standards, i.e. by applying the normal distribution theory gives systematically wrong standards and thus predicts wrong frequencies of selected persons if persons are chosen as being extreme to a given value (Table 1).

However, it has been possible to find statistical models which describe the data in the three examples satisfactory. These models might be interesting to academic anthropometry but they do also provide possibilities to predict the frequency of persons in certain intervals from a theoretical model.

The firm establishment of such non-normal models for anthropometric variation requires a very large number of observations. Once the model and probable area of variation for the parameters has been established it is possible to estimate the parameters from far smaller samples. In Table 2 the estimates for the parameters which fitted M16 is compared with estimates obtained from a far smaller Danish sample. It is evident that though there are significant differences between the two samples they are perfectly acceptable taking the geographical distance between United Kingdom and Denmark into consideration.

Standards calculated from the realistic model for the variation of age at menarche and using the DK estimates will give a far more efficient background for intervention into the individual development of Danish girls than would standards calculated using the normal distribution model.

Table 1. The ratio between observed and expected under the hypothesis that the variation can be explained within the normal distribution model. The X^2 tests have 1 degree of freedom they can be considered significant at the five percent level if larger than 3.84

Variable	Fractiles		Ratio	X^2
	observed	expected		
H22	0.005	0.00132	3.79	59.04
	0.01	0.00391	2.56	203.13
	0.05	0.0463	1.08	1.70
	0.95	0.9371	0.79	15.23
	0.99	0.98190	0.55	20.86
	0.995	0.99275	0.69	4.02
H16	0.005	0.00029	17.24	179.23
	0.01	0.00181	5.52	86.83
	0.05	0.0516	0.97	0.12
	0.95	0.9436	0.89	1.70
	0.99	0.98554	0.69	3.22
	0.995	0.99500	1.00	0.00
M16	0.01	0.00822	1.22	1.74
	0.025	0.0298	0.84	3.49
	0.1	0.0792	1.26	24.68
	0.9	0.9027	1.03	0.34
	0.95	0.9610	1.28	14.02
	0.975	0.9782	1.42	2.17

Table 2. Estimates for the parameters in the mixtures of three normal distributions which fitted M16 and a small contemporary sample of menarchal ages reported by young Danish women (DK)

Parameters	Estimates	
	M16	DK
n	4518	282
1st mean	12.090	12.199
2nd mean	13.699	13.892
3rd mean	15.789	16.515
1st frequency	0.323	0.234
2nd frequency	0.625	0.716
3rd frequency	0.052	0.050
variance	0.560	0.750

Realistic statistical modelling of sets of data is intellectually satisfactory but it is also a means to develop practical tool for the monitoring of growth.

ACKNOWLEDGEMENTS: Dr. Mascie-Taylor have allowed me to use H16 and M16 and Dr. B. Jeun gave me acces to the data for DK they are both greatfully acknowlegded for these contributions.

*

Received 26 June 1986.

References

Boldsen, J (1983): *Analyzing morphological polymorphisms.* – Symposium i anvendt statistik, 1983.

Boldsen, J and Kronborg, D (1984): The distribution of stature of Danish conscripts 1852–56. – Ann. Hum. Biol. *11;* 555–565

Boldsen, J and Mascie-Taylor, CGN (1985): Analysis of Height Variation in a Contemporary British Sample. – Hum. Biol., *57;* 473–480.

Boldsen, J and Mascie-Taylor, CGN (1986): Recalled age at menarchy in Britain. – Ann. Hum. Biol. in press.

Meddelelser fra det statistiske bureau 1859: femte samling. København: 159–190.
Sundberg, R (1972): *Maximum likelihood theory and applications for distributions generated when observing a function of an exponential family variable.* – Thesis, Stockholm University.

Mailing address: Dr. Jesper L. Boldsen
Department of Social Medicine
Institute of Community Health Odense University
J. B. Winslows Vej 17 DK–5000 Odense C Denmark