Available online at www.sciencedirect.com
ScienceDirect
Economics and Human Biology 6 (2008) 43–56
ECONOMICS AND HUMAN BIOLOGY
http://www.elsevier.com/locate/ehb

On estimators for truncated height samples

Jan Jacobs [a,*], Tomek Katzur [a], Vincent Tassenaar [b]

[a] *Faculty of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands*
[b] *Department of Economic and Social History, Faculty of Arts, University of Groningen, P.O. Box 716, 9700 AS Groningen, The Netherlands*

Received 31 July 2006; received in revised form 2 April 2007; accepted 3 April 2007

Abstract

Statistical inference from truncated height data is often based on distributional assumptions. In this paper we analyze a data set of over 23,000 conscript height observations, covering nearly all conscripts in Drenthe, a province of the Netherlands, over the period 1826–1860. The data do not satisfy the normality assumption. We demonstrate that the ML estimators of the mean proposed for normally distributed data do not yield satisfactory results. We propose a new estimator that exploits the relationship between the conditional mean of the observations above the minimum height requirement and the conditional mean and proportion of conscripts below the minimum height requirement.
© 2007 Elsevier B.V. All rights reserved.

JEL classification : 131; N33

Keywords: Anthropometric history; Physical stature; Truncated distribution; Maximum likelihood; Normality assumption; 19th century; Drenthe; The Netherlands

1. Introduction

The anthropometric approach to such issues as changes in the standard of living has become very popular during the last decades (Steckel, 1995). A clear link exists between changes in the average height attained by a population and changes in its living standards as reflected by nutrition, sanitary conditions, housing, schooling, social security and so on. Therefore, the average height of the population can be treated as an indicator of living conditions and economic development, especially in the absence of reliable figures on GDP per capita, a very likely situation in development economics or historical research.

* Corresponding author. Tel.: +31 50 363 3681.
E-mail address: j.p.a.m.Jacobs@rug.nl (J. Jacobs).

1570-677X/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.ehb.2007.04.001

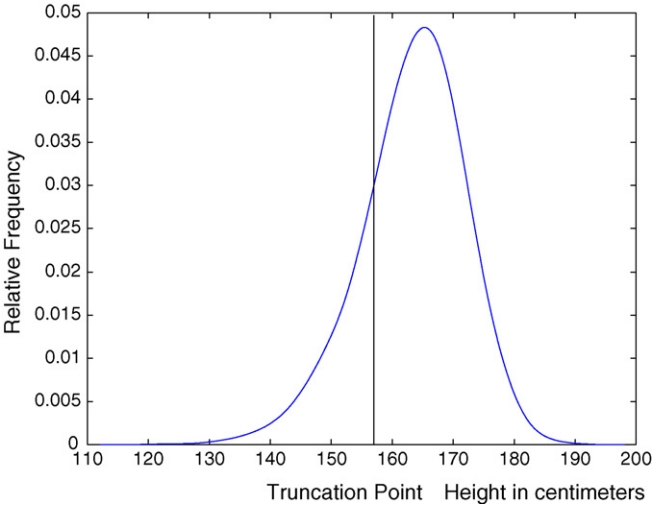

Fig. 1. Height distribution and truncation.

The empirical data to assess the development of average height over time often originate from military conscription registers. Conscription was introduced on a large scale throughout Europe in the beginning of the 19th century. Because of the recurrent character of these samples (in the form of yearly drafts) and their relative homogeneity (measurement at approximately the same age) these data are well suited for statistical comparisons. There is, however, one major drawback. Most armies only admitted conscripts whose height exceeded a certain threshold. The heights of undersized conscripts were rarely recorded.

This leaves the researcher without knowledge of the shape of the left tail of the height distribution. The problem is illustrated in Fig. 1, where the height distribution of a typical cohort of conscripts is plotted, smoothed with an algorithm proposed by Scott (1992; Chapter 6). If the sample is truncated, no information on any observation below the Minimum Height Requirement (MHR), 157 cm in the example, is available. The truncation of the distribution has important implications for the estimation of the parameter of interest, the mean, since the left tail of the distribution may contain considerable probability mass, the size of which is unknown.[1]

Estimating the moments of height distributions is a classic problem in the history of (applied) statistics, associated with great 19th century statisticians such as Quetelet, Galton, and Pearson (Stigler, 1986). Nowadays height samples are typically used to illustrate the normal distribution. As Meier (1982) put it: ''Although adult male heights in a relatively homogeneous healthy population really are very nearly normally distributed, hardly anything else one is likely to study shares this property.'' Recently, the statistical problem of estimating the mean of truncated height samples attracted much interest in the anthropometric literature (see, e.g., Komlos, 2004; A'Hearn, 2004; Penttinen et al., 2005). Several estimation procedures have been proposed, with a focus on maximum likelihood (ML) estimation based on a normal distribution.

The aim of this paper is to evaluate estimators for truncated height samples. To that purpose we employ height data for the Dutch province of Drenthe for the period 1826–1860 (Tassenaar,

[1] Truncation is be distinguished from censoring, where the number of observations below the MHR is known.

Download English Version:

https://daneshyari.com/en/article/5057311

Download Persian Version:

https://daneshyari.com/article/5057311

Daneshyari.com