A CURVE THAT FITS CANADIAN PROVINCIAL BIRTHWEIGHTS

James Stafford
Lakehead University, Thunder Bay, Ontario, Canada

Résumé—Après avoir examiné les différentes distributions théoriques, on en est arrivé à la conclusion que la distribution géométrique type IV de Pearson reproduit d'une manière satisfaisante la distribution observée des poids de garçons à la naissance, pour le Canada et ses provinces.

Abstract—This paper explores distributions that will best fit Canadian male birthweights and finds that Pearson's Type IV Geometric Distribution fits the data of each province.

Introduction

Birthweight has become a focal point of interest in medical research because of its high negative correlations with infant mortality. The risk of neonatal death for the cohort of infants born in the United States in 1960 was about 34 times greater for those infants weighing less than 2,500 grams than for those weighing more (U.S. Public Health Service, 1972a:9). The risk of neonatal death as depicted in Figure 1 decreases rapidly as birthweight increases to 2,500 grams, continues to decrease slowly to about 4,000 grams where it proceeds to increase slowly.

Variation in weight is not in itself a cause of death. Numerous researchers have investigated the association of mortality and weight, together and separately with a large number of physiological and sociological factors. Dellaportas, investigating birthweights in Canada, discovered a significant difference between the mean weight of Quebec-born infants and those of the remainder of Canada (Dellaportas, 1971:159). Since most of the Quebec population is French in origin, while only a minority of the remainder of the Canadian population is French, the inference is that ethnicity is a factor of birthweight. He pursues this idea further by investigating the correlation between birthweight and proportion of British extraction for each province, and he finds the two variables highly correlated. The implication he draws is that infants of French ancestry weigh less at birth, a conclusion which may be correct but which does not follow from the statistical analysis when we consider that one-fifth of the population is neither British nor French. Dellaportas' methodology also falls prey to the ecological fallacy in that he applies correlations which have been obtained using areas as the unit of analysis to individuals (Robinson, 1950).

The influence of genetic factors on birthweight has been examined by a number of researchers, but evidence of the interrelationships is inconclusive because of the interference of socio-economic factors. Studies in Hawaii have documented differences in mean birthweight between several racial groups (Taff and Wilbar, 1953; Connor et al., 1957; Bennet and Louis, 1959) but the presence of other factors has mitigated the validity of the findings.

One of the major influences on birthweight is the gestation period: the longer the period of gestation, the heavier the infant (U.S. Public Health Service, 1970:27). The variable "period of gestation" has created problems for researchers because of the variation in the way the question is worded — "length of pregnancy in completed weeks" and "first day of last menstrual period" are two examples of questions that might contribute to variation in response. The errors in human recall, deliberate or unconscious, also add to the variation of this reponse.

birthweight. In a study by the U.S. Public Health Service, it was noted that when income was statistically controlled the average birthweight of infants born to white wives was higher than that of infants born to non-white wives (U.S. Public Health Service, 1968). This implies that ethnicity has a basic influence on birthweight. However, among white wives, birthweight varied directly with social class, indicating that socio-economic class also influences birthweight, although the assumption that all white ethnicities are equally distributed throughout the social classes is incorrect and weakens this argument. If certain European ethnic groups are concentrated in one part of the socio-economic continuum, the class-specific differences in birthweight among white women may have been caused by the ethnic factor. The old problem of separating nature and nurture has not been resolved.

Moving from socio-economic to environmental factors, we discover that children born to rural residents tend to be heavier than those of urban residents. The U.S. Public Health Service found that among infants born of rural residents, the proportion of immature births was lower, and the average weight was higher than among infants born of urban residents for both whites and non-whites (U.S. Public Health Service, 1964:31). This represents another anomaly as rural residents tend to be placed lower than urban dwellers on a socio-economic scale. If social class and birthweight are interrelated, we would expect that infants of rural residents would tend to weigh less at birth than those of urban residents.

Empirical investigation has revealed that beside the physiological, genetic, socio-economic and residential influences on birthweight, geography seems to be a factor. In the United States, the highest proportion of immature births occurs in the mountain states while the area of the lowest proportionate immaturity is in the West North Central States (U.S. Public Health Service, 1964:31). Such variations in birthweight might be caused by differences in altitude or climate but it might also be caused by differences in living conditions or by variation in proportionate ethnic distribution from region to region in the United States.

A final factor, illegitimacy, has been found to influence birthweight. Paktar and Nelson, studing illegitimacy in New York, found a higher rate of prematurity among illegitimate children (1965). This difference is probably a reflection of the limited medical care and lower socio-economic status of girls who have conceived out of wedlock.

The multitude of interrelated variables need to be unravelled before progress can be made in this area of research. One contribution would be a causal framework showing the interplay of relevant factors and the specification of indicators of these factors. Such a task would best fall on the shoulders of a medical person. A tentative framework might classify all factors of variation in birthweight into three categories: genetic, physiological and social. Genetic factors would determine the infant's sex, propensity to develop *in utero* and length of gestation period. Results of some medical research suggest that physiology may influence the sex of the child, but these are inconclusive and at best, probabilistic. Genetics and physiology interact to determine single or plural births. The environment of the foetus, influenced by the physiology and nutrition of the mother would have a direct affect on foetal development, as would the prenatal care practiced by the mother. This last factor would be a product of the mother's knowledge, values and beliefs regarding pregnancy-care, and would be strongly influenced by her class background.

The primary purpose of this paper is not to develop a meaningful causal framework because such a task requires investigation by a person with professional medical qualifications. Rather, the purpose of this paper is to contribute to such a paradigm by proposing a function of the frequency distribution of birthweights. Such a function would allow future investigations of variation in the frequency distribution rather than simple comparisons of arithmetic means and medians. It might also eliminate the practice of many researchers of comparing proportions of birthweights under 2,500 grams, a completely arbitrary cutoff point

function represents areas under sections of the curve. This is so because the number of elements in each class represent the area under the curve and within the limits of the independent variable which are determined by the class boundaries. The usual assumption that class frequencies are proportional to the length of the ordinate at the mid-point of the class is valid only when the elements are distributed equally within the category. Such an assumption concerning continuous data is valid only when the distribution is rectangular, when all class frequencies are approximately equal, or when the categories are narrow.

The fitted curve derived from grouped data is flattened with a consequent increase in its variation. This can be countered by Shephard's Correction which involves adjusting the values of the second and fourth moments of the grouped data. These two moments reflect the variance and flatness of the distribution, and were adjusted according to the formulas:

$$u_2 = V_2 - \frac{1}{12} h^2$$

$$u_4 = V_4 - \frac{1}{2} V_2 h^2 + \frac{7}{240} h_4$$

where: u_2 and u_4 are the adjusted moments;

V_2 and V_4 are the unadjusted moments;

and h represents the class size.

The procedure of applying corrections allows us to utilize grouped data, saving us time, without introducing a significant amount of error.

Using Pearson's Kappa we found a type IV geometric distribution applicable to 1969 Canadian male birthweights. We derived parameters of the following function by the method of moments:

$$Y = 85.5 \left[1 + \left(\frac{x}{1377}\right)^2\right]^{-4.7} \cdot e^{-3.1 \cdot tan^{-1}\left(\frac{x}{1377}\right)}$$

where x is the birthweight in grams $(\bar{x} = 0)$,

and Y is the frequency function.

The calculation of this function, which is the most difficult of all Pearson's frequency curves, is elaborated in the appendix.

Curves were fitted to 1970 Canadian male birthweights by provinces. In each case a type IV curve fitted the data. The parameters of the curves are listed in Table 2 allowing an increased number of ways of comparing the provincial birthweight distributions other than by their means.

A Z-test of provincial means revealed Z scores ranging from absolute values of 7 to 27, suggesting that factors affecting birthweights vary from province to province. The Z-test substantiates Dellaportas' contention that Quebec differs significantly from the rest of Canada (1971) but it also indicates that he would have obtained similar results using any one of the ten provinces.

The skewness of the distribution of birthweights is reasonable given that there appears to be a threshold weight below which births are not viable. Such weights may be the result of adverse genetic and/or in utero conditions. These often result in immature births, natural abortions, and stillbirths. This set of parturitions is not included in the distribution of birthweights because these are live birthweights. This minimum threshold varies according to a number of genetic factors but it does represent a force which is not present at the other extreme of the distribution. Thus we would expect the distribution of birthweights to have a negative skew.

37

and larger measures of skewness respectively. If we attribute the negative skewness to the disproportionate number of natural abortions of under-weight foetuses, then variations in abortions may cause variations in skewness, in which case, women in Manitoba may experience an above-average abortion rate while those in the Maritimes, a below-average rate. This is not likely.

The negative skewness, characteristic of all the provinces, is probably the result of a genetic zero-boundary effect whereby the opportunities for variation above the mean are greater than for variation below the mean. The question why these two provinces vary from the others in the directions they do is not answerable without further investigation into factors contributing to variations in birthweight distributions.

The parameters of the birthweight distributions in Manitoba and British Columbia appear to diverge from the Canadian average more than any of the others. Manitoba's distribution has the highest mean, standard deviation, skewness, and kurtosis, while the parameters of British Columbia's Type IV distribution are highest, resulting in the lowest measure of kurtosis and the second lowest standard deviation. Our existing knowledge of variations in birthweights at the aggregate level precludes a presentation of a viable explanation for these variations.

Conclusion

The variation in birthweight is of concern because of its association with neonatal mortality. Birthweight is associated with a broad range of genetic, physiological, social, economic, and environmental factors, awareness of which may assist in a reduction of neonatal deaths.

The purpose of this paper is to investigate the methodology of developing a precise function to represent birthweight distributions. Pearson's Type IV Geometric distribution fits most closely the distribution of male birthweights in each Canadian province. Comparison of the provincial birthweight functions reveals considerable similarities in standard deviation and kurtosis. The mean birthweights of Quebec and Alberta are considerably less than those of the other provinces, while the skewness of the distributions of the Maritimes and Manitoba were larger and smaller, respectively, than those of the other provinces.

The birthweight functions of Manitoba and British Columbia deviate most among the provinces when all parameters are compared. Manitoba's distribution has the highest mean, standard deviation, skewness, and kurtosis, while that of British Columbia ranks lowest in kurtosis and second lowest in standard deviation. No explanation in the literature can account for these variations.

Appendix

The purpose of the appendix is to provide the reader with the formulas used in deriving birthweight functions. Data on the yearly distribution of birthweights published by the Vital Statistics Section of Statistics Canada are used to calculate second, third and fourth moments:

$$\mu'_2 = \frac{\Sigma (x - \bar{x})^2}{N}$$

$$\mu'_3 = \frac{\Sigma (x - \bar{x})^3}{N}$$

$$\mu'_4 = \frac{\Sigma (x - \bar{x})^4}{N}$$

References

Bennett, C. G., and L. S. K. Louis. 1959. Demographic Factors Influencing Birth Weight, Hawaii Medical Journal 18:239-244.

Connor, A., C. G. Bennett, and L. S. K. Louis. 1957. Birth Weight Patterns by Race in Hawaii, Hawaii Medical Journal 16:626-632

Dellaportas, G. J. 1971. Birth Weight, Ethnic Origin, and Perinatal Mortality, Social Biology 18:158-163.

Dominion Bureau of Statistics. 1972. Vital Statistics 1969. Cat. No. 84-202. Ottawa: Information Canada.

Elderton, William Palin, and Norman Lloyd Johnson. 1969. Systems of Frequency Curves. Cambridge, Massachusetts: Cambridge University Press.

Kendall, M. G. 1943. The Advanced Theory of Statistics. Vol. I. London: Charles Griffin & Co., Ltd.

Paktar, J. M. D., and F. Nelson. 1965. The Unmarried Mother and Her Child: The Problems and the Challenges, Illegitimacy. New York: National Council on Illegitimacy.

Pearson, Karl. 1924. (ed.) Tables for Statisticians and Biometricians. Cambridge: Cambridge University Press.

Robinson, W. S. 1950. Ecological Correlations and the Behavior of Individuals, American Sociological Review 15:351-357.

Statistics Canada. 1972. Vital Statistics 1970. Cat No. 84-202. Ottawa: Information Canada.

Taff, M. A., Jr., and C. L. Wilbar, Jr. 1953. Immaturity of Single Live Births According to Weight, With Particular Reference to Race, American Journal of Disabled Children 85:279-284.

United States Public Health Service. 1964. Natality Statistics Analysis. United States — 1962. Series 21, Number 1. Washington, D. C.

_____ .1968. Variations in Birthweight. Series 22, Number 8. Washington, D. C.

_____ .1970. Natality Statistics Analysis. United States, 1965-1967. Series 21, Number 19. Washington, D.C.

_____ .1972a. Comparison of Neonatal Mortality From Two Cohort Studies. Series 20, Number 13. Washington, D. C.

_____ .1972b. Infant Mortality Rates: Socio-Economic Factors. Series 22, Number 14. Washington, D.C.