Educational and Psychological Measurement
1975, 35, 657–662.

POSSIBLE SAMPLING BIAS IN GENETIC STUDIES OF GENIUS

DANIEL P. KEATING[1]
University of Minnesota

The data from Terman's *Genetic Studies of Genius* (1925–1959) relating to sample size, mean IQ, and variance of IQ scores were analyzed in terms of their conformation to the theoretically projected statistics derived from a consideration of the normal curve. Deviations from the theoretical projections lead to the probable conclusion that the sample size was too small, with the IQ scores clustered more closely about a significantly higher mean than projected. Although the major findings of the "Genius" study are not cast into doubt by this analysis, caution is urged with respect to comparisons to a normal sample when the differences are not large.

The five-volume *Genetic Studies of Genius* (1925–1959), edited by the late Lewis Terman, has been widely and justifiably acclaimed as a landmark in longitudinal research. Its refutation of myths widely held at the time (e.g., that highly intelligent children are weak and sickly, that early ability is rarely maintained through adolescence and into maturity) was a starting point for work with gifted children, as well as for much research into the intellectual development of individuals. The study also illustrated the many difficult and often intractable problems of large-scale longitudinal research, one of which is examined here more closely.

In Terman's (1925) selection of his gifted group, he realized that his sample was not entirely correct. He stated:

[1] The author would like to thank Julian C. Stanley for his assistance in the preparation of this note and also Melita H. Oden for her cooperation in supplying data from the Terman Study of the Gifted. This note was completed in conjunction with a project sponsored by the Spencer Foundation.

> One may conclude that the method of selection employed, although far from ideal, probably led to the discovery of at least 80 percent and possibly 90 percent of all the cases who could have qualified in the school population canvassed (p. 33).

What he may not have realized was that his estimate of error might itself have been erroneous. There are a number of indications that it was.

This is most clearly seen from an examination of the normal curve. With a mean of 100 and a $\sigma = 16$ (Terman and Merrill, 1937) the percentage of cases falling beyond $+2.5\sigma$ (i.e., 140 IQ) $\cong$ .62%. Multiplying the population canvassed, 168,000 (Terman, 1925, p. 29), by this figure yields a projected sample of about 1042 cases over 140 IQ. Terman's (1925) actual yield was 649 cases (.38% of the population), or 61.22% of the projected sample. Further, the projected mean of the portion of the unit normal distribution beyond $+2.5\sigma$ (140 IQ) is, after Kelley (1947, p. 297),

$$\mu_{>2.5\sigma} = \frac{y_{2.5\sigma}}{1 - P_{<2.5\sigma}} = \frac{0.0175283}{0.0062097}$$
$$= 2.8227\sigma,$$

where $y_{2.5\sigma}$ is the height of the ordinate 2.5σ above the mean of the unit normal distribution, and $P_{<2.5\sigma}$ is the area of the distribution below 2.5σ.

The mean of this tail portion of the unit normal distribution is 2.82; thus the mean of scores beyond 140 IQ is

$$100 + (2.8227)(16) \cong 145,$$

where 100 is the overall IQ mean and 16 is the standard deviation of IQ scores. The actual mean of the gifted group was 151 (Terman, 1925, p. 45), a difference of 6; this is about $.4\sigma$ above the mean of the normal distribution beyond $+2.5\sigma$.

Jensen (1969) has pointed out the variations in the normal curve for IQ at the extremes. This casts some doubt on the reliability of the difference between the projected mean and the actual mean. However, it reinforces the difference between the projected and actual *size* of the sample. The proposed alteration of the normal curve would, if anything, increase the percentage of area under the curve beyond $+2.5\sigma$, thus increasing the size of the projected sample.

The theoretical standard deviation for the normal distribution of scores beyond $+2.5\sigma$ may also be calculated. The variance is, again after Kelley (1947, p. 298),

$$\sigma_{(>2.5\sigma)}{}^2 = 1 + \frac{(2.5)y_{2.5\sigma}}{1 - P_{<2\ 5\sigma}} - \mu_{>2.5\sigma}$$

$$= 1 + \frac{2.5(0.0175283)}{0.0062097} - (2.8227)^2$$

$$= 0.089187,$$

where $\sigma_{(>2.5\sigma)}{}^2$ is the variance of the unit normal distribution beyond $+2.5\sigma$ (values derived from Pearson and Hartley, 1956). For $\sigma = 16$, $\sigma^2 = 256$, and $\sigma_{(>2.5\sigma)}{}^2 = 0.089187$, the standard deviation for the portion beyond 140 IQ is

$$\sqrt{(.089187)(256)} = \sqrt{22.83187} \cong 4.8.$$

By comparison, the standard deviation of the obtained sample was 10.2 (Terman, 1925, p. 45). This suggests the picture of the theoretically projected sample having scores clustered much more closely around a significantly lower mean ($p < .001$).

The schematic comparison of the distribution of obtained IQ's above 140 in Terman's (1925) sample with the tail portion of the unit normal distribution in Figure 1 demonstrates the nature of the discrepancies calculated above. The shaded area represents the discrepancy leading to the inflated mean of the actual sample. This indicates, as can be seen from Figure 1, that too few "low" subjects and too many "high" subjects were included in the sample.

The calculations of the sample were performed on the grouped data found in Terman (1925). There is the possibility that the calculations might be affected by the grouping procedure. Unfortunately the original ungrouped data is not directly recoverable (Oden, personal communication), but the appropriate calculations were performed on the ungrouped data which were easily recoverable. No important differences were found between the grouped and ungrouped statistics.

It might be argued that the above statistical arguments fail because of the nature of the 1916 revision of the Binet-Simon (Terman, 1916). The mean IQ and σ were not calculated at that time, and there was no specification of the IQ distribution as a normal curve. One might infer from the technical monograph which accompanied the 1916 Stanford revision (Terman et al., 1917, p. 43) that $\sigma = 13.5$. Using the same reasoning as above, we obtain the following statistics for the theoretically projected sample: $N = 267$; mean $= 143$; standard deviation $= 3.6$. Thus the mean and standard deviation are even more discrepant from the obtained figures than with an assumption of $\sigma = 16$, but the obtained sample size is greater than the projected size rather than smaller.

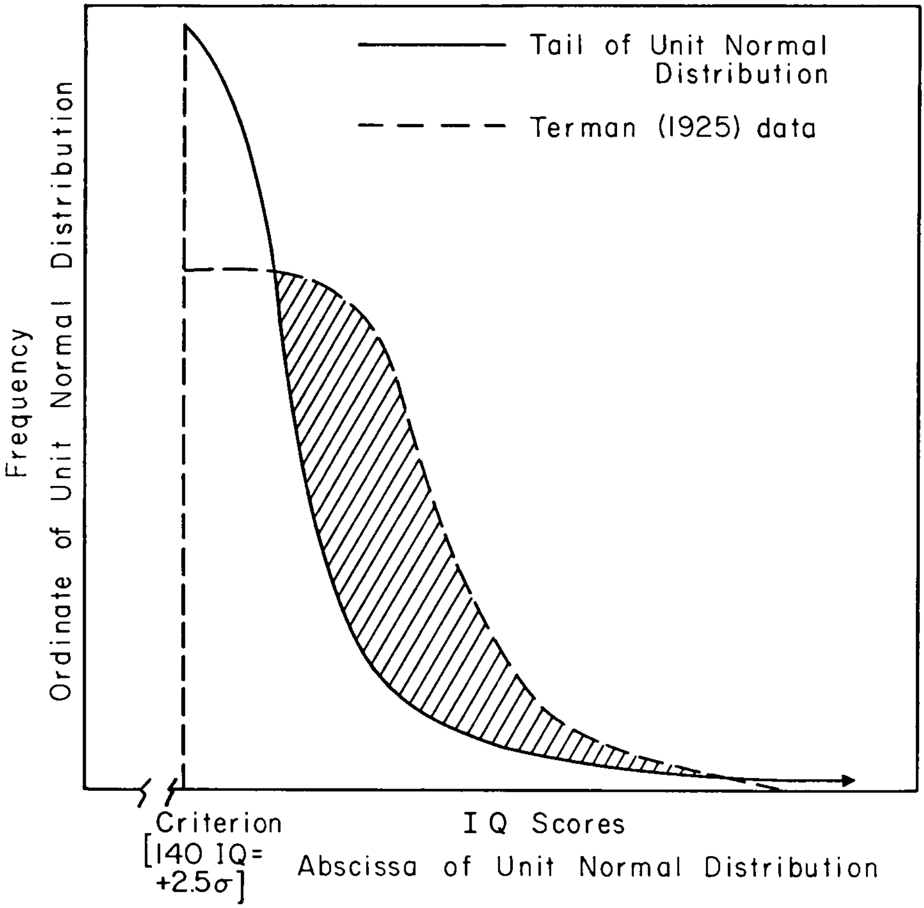

Figure 1. Schematic comparison of actual sample and theoretically projected sample.

There is also considerable deviation of the gifted group sample (Terman, 1925) from the sample projected by using Terman's (1916, p. 66) percentages. At least 0.55% of the standardization group score above 135 on the 1916 scale. Even allowing .05% for the 136–139 range, the projected sample size is 840. The actual sample (639) is thus 76.7% of the projected sample. The projected error, therefore, ranges from a low of nearly 24% to a high of almost 40%.

There are a number of plausible speculations regarding the source of this sampling error. First, the population from which the sample was drawn might have been markedly non-normal. Given the high number of students canvassed and the demonstrated normality of the

Stanford-Binet scores (Terman and Merrill, 1937), however, this seems unlikely.

The second speculation seems more likely. Taking together the facts of a too small sample, a (possibly) too high mean, and a (possibly) too large standard deviation, the intuitive inference is that too few of the "borderline" cases (140–150) were included. Terman (1925), after considering the defects of his selection techniques, conjectured about the nature of the cases he failed to locate:

> They would almost certainly have been found a little less accelerated in school. Some would be excessively shy, others lazy, and still others lacking in adaptability [p. 33].

A third possible source of sampling error was the concentration on urban and suburban canvassing, with the rural population being nearly ignored (Terman, 1925, p. 29). The concentration of talent (as measured by IQ) tends to be greater in metropolitan as opposed to rural areas (e.g., Terman and Merrill, 1937; von Fieandt, 1958). The overload of "high" cases may be partly attributed to this factor.

If the error is actually closer to the 25–40% we have suggested than to the 10–20% Terman (1925) estimated, and if his characterization of those not included is correct, then one may easily see the ramifications for the significance of a number of this conclusions. Many of the statistically significant differences between his gifted group and the "general population" which were reported in 1925 and subsequently throughout the longitudinal study (Burks, Jensen, and Terman, 1930; Terman and Oden, 1947; Terman and Oden, 1959; and Oden, 1968) may in fact lack significance for the *specific* sample which Terman originally prescribed, especially since the "missed" cases were likely to be less differentiated from the general population than the cases in the obtained sample. It is clear from both the number and the degree of differences obtained that the major conclusions of the "Genius Study" were warranted. Caution is urged, however, in the interpretation of data from the "Genius Study" where only a small difference was reported.

REFERENCES

Burks, B. S., Jensen, D. W., and Terman, L. M. *Genetic studies of genius.* Vol. III: *The promise of youth.* Stanford, Calif.: Stanford University Press, 1930.

Jensen, A. R. How much can we boost IQ and scholastic achievement? *Harvard Educational Review,* 1969, 39, 1–123.

Kelley, T. S. *Fundamentals of statistics.* Cambridge: Harvard University Press, 1947.

Oden, M. H. The fulfillment of promise: 40-year follow-up of the Terman gifted group. *Genetic Psychology Monographs,* 1968, 77, 3–93.

Pearson, E. S. and Hartley, H. O. (Eds.). *Biometrika tables for statisticians,* Vol. 1. Cambridge, England: Cambridge University Press, 1956.

Terman, L. M. *The measurement of intelligence.* Boston: Houghton-Mifflin, 1916.

Terman, L. M., Lyman, G., Ordahl, G., Ordahl, L. E., Galbreath, N., and Talbert, W. *The Stanford revision and extension of the Binet-Simon scale for measuring intelligence.* Baltimore: Warwick & York, 1917.

Terman, L. M. *Genetic studies of genius.* Vol. I: *Mental and physical traits of a thousand gifted children.* Stanford, Calif.: Stanford University Press, 1925.

Terman, L. M. and Merrill, M. A. *Measuring intelligence.* Boston: Houghton-Mifflin, 1937.

Terman, L. M. and Oden, M. H. *Genetic studies of genius.* Vol. IV: *The gifted child grows up.* Stanford, Calif.: Stanford University Press, 1947.

Terman, L. M. and Oden, M. H. *Genetic studies of genius.* Vol. V: *The gifted group at mid-life.* Stanford, Calif.: Stanford University Press, 1959.

von Fieandt, K. Psychological effects of urban and rural domiciles. *Acta Psychologica,* 1958, 14, 81–91.