Assessing Familiality of Cognitive Ability*

by

Leon Jay Gleser

Technical Report #85-9

Department of Statistics
Purdue University

May, 1985

ASSESSING FAMILIALITY OF COGNITIVE ABILITY*

by

Leon Jay Gleser
Purdue University

ABSTRACT

In a recent paper in this journal, Benbow, Zonderman and Stanley (1983) conclude that intellectually precocious children resemble their parents to a lesser extent than do children of lesser ability. In reply, Vining (1985) asserts that Benbow, Zonderman and Stanley's results are artifacts of selection and their statistical metholology, and that a more appropriate statistical methodology yields quite the opposite conclusion. The present paper has two purposes: (i) to show that Vining's criticism is misdirected, stemming from a misunderstanding of how Benbow, Zonderman and Stanley selected their subjects, and (ii) to point out some problems in the model, indices of familiality and design used by Benbow, Zonderman and Stanley which need to be addressed before future comparative studies of familiality are attempted.

1. INTRODUCTION

Benbow, Zonderman and Stanley (1983; hereafter abbreviated BZS) have studied the association of cognitive abilities between intellectually extremely able children and their parents, and have compared such associations to those found for children and their parents in the total population. Their conclusion is that over a variety of tests of cognitive ability, the regression slopes of child's scores Y on parent's scores X are typically lower for extremely able children than for children in the population as a whole ("unselected children").

The extremely able children in the BZS study had scores on the mathematics or verbal portions of the Scholastic Aptitude Test (SAT-M, SAT-V) in the top 1% for their age among all children who participated in the Study of Mathematically Precocious Youth (SMPY). Since children in the SMPY themselves score in the top 1/30 of their respective age groups, the children studied by BZS can certainly be described as being extremely able.

Selected children participating in the study, along with one or both of their parents, took a battery of nine tests measuring a range of general and specialized cognitive abilities (BZS, 1983; see also Benbow, Stanley, Kirk and Zonderman, 1983). These tests were designed to be sufficiently difficult that "ceiling effects" would be unlikely. The classical least squares regression slopes b of child's score Y on parents score X (X = father's score, mother's score, or mid-parent score) were calculated for all nine tests, and for each definition of X. The median slopes over these tests were .17, .09, .11 for X = father's score, X = mother's score, X = midparent score, respectively. These slopes were compared to previously published child-parent regression slopes for unselected populations, which ranged from .42 to .60 (BZS, p. 158). Subject to errors resulting from the voluntary participation of the SMPY subjects and the small sample sizes, BZS tentatively conclude that there is less resemblance between the cognitive abilities of extremely able children and their parents than there is for the child-parent population as a whole.

Vining (1985) has questioned this claim, and the methodology upon which this claim is based. His objections are based upon a misunderstanding both of the way BZS selected their data and of the nature of the population defined by the term "extremely able." In the next section, Vining's criticisms are discussed and shown to be irrelevant to the problem studied by BZS. However, there are some problems with the model, index of familiality, and design used by BZS. These are discussed in Section 3.

2. VINING'S CRITICISM

Vining's (1985) criticism of BZS's methodology is based on the assumption that the pairs Y = child's score, X = parent's score come from a bivariate normal population with mean vector (μ_y, μ_x) and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_y^2 & \rho\sigma_y\sigma_x \\ \rho\sigma_y\sigma_x & \sigma_x^2 \end{pmatrix} , \tag{1}$$

and that BZS are attempting to estimate the slope

$$\beta = \rho\frac{\sigma_y}{\sigma_x} \tag{2}$$

of the linear regression of Y on X in this population. He also assumes that the child-parent pairs of scores (Y,X) are sampled from this population under the restriction that

$$Y \geq a, \tag{3}$$

where a is the 99.97th percentile of the distribution of Y. If these assumptions are correct, the usual least squares estimator b of the slope of the regression of Y on X calculated from this restricted sample is well-known to be biased downwards (Skinner, 1984). Vining then proposes several ad-hoc estimators of β which are unbiased. [a/] These estimators (predictably) yield values for the slope which are considerably larger (in magnitude) than the values for b given in BZS, and indeed reverse the direction of BZS's conclusions about familiality. Thus, Vining concludes that extremely able children more closely resemble their parents than do children in the population at large.

However, Vining is wrong both about how the (Y,X) pairs are selected in BZS's study, and also about which population slope is being estimated. The children in BZS's study are not selected through Y, which is the child's score on one of the 9 cognitive tests used by BZS. Rather, selection is based on another variable Z, where Z is a "precocity quotient" based on a score on either SAT-M or SAT-V. Although Z is admittedly correlated with Y, so that reduction of the variance of the Y-scores may result from selection based on Z (Skinner, 1984), there is not the direct truncation of range for Y assumed by Vining's model. Indeed, it is quite possible for (Y,X) given $Z \geq a$ to have a (conditional) bivariate normal distribution. [b/]

Since BZS used cognitive tests with an extended upper range of difficulty, the fact of selection using Z may also not materially reduce the variance σ_y^2 of Y. BZS note that the standard deviations on one of the cognitive tests (CTMM) exhibited no reduction. [c/]

Let us suppose, however, that Vining's assumption is correct that BZS selected children by requiring that $Y \geq a$. Even so, Vining fails to realize that such selection is used to define the population of extremely able children. BZS's sample from this truncated population is actually unrestricted. Vining attempts to estimate the slope β of the untruncated population, but the population slope of concern in BZS's study is the slope β^* of Y on X in the truncated population. It is easily shown that the classical least squares estimator b of slope used by BZS is an unbiased estimator of β^* (while Vining's estimators, being estimators actually of β, are biased upwards). Thus, even if Vining were right as to how children were selected, BZS's estimator is the appropriate estimator, because Vining is estimating the slope of the wrong population.

It is worth noting for possible future use in familiality studies that there is a direct mathematical relationship between the slope β^* of the restricted population and the slope β of the unrestricted population in Vining's model. Assume that the critical point a in Inequality (3) is the $100(1-\alpha)$th percentile of the (unrestricted)distribution of Y. Thus,

$$a = \mu_y + \sigma_y Z_{1-\alpha} ,$$

where $Z_{1-\alpha}$ is the $100(1-\alpha)$th percentile of the standard normal distribution. Then,

$$\beta^* = \frac{v_\alpha \beta}{1 + \beta^2 \frac{\sigma_x^2}{\sigma_y^2} (v_\alpha - 1)} , \qquad (4)$$

where

$$v_\alpha = \int_{Z_{1-\alpha}}^{\infty} \frac{t^2 e^{-\frac{1}{2}t^2} dt}{(2\pi)^{\frac{1}{2}} \alpha} - \left[\int_{Z_{1-\alpha}}^{\infty} \frac{t e^{-\frac{1}{2}t^2} dt}{(2\pi)^{\frac{1}{2}} \alpha} \right]^2$$

$$= 1 - \frac{e^{-\frac{1}{2}(Z_{1-\alpha})^2}}{(2\pi)^{\frac{1}{2}} \alpha} \left[\frac{e^{-\frac{1}{2}(Z_{1-\alpha})^2}}{\alpha (2\pi)^{\frac{1}{2}}} - Z_{1-\alpha} \right] .$$

Assuming further, as Vining does, that $\sigma_y^2 = \sigma_x^2$, and that α =.0003 (as in BZS's study), we find that

$$\beta^* = \frac{.06\beta}{1-\beta^2(.94)} \quad .$$

Thus, β values of from .42 to .60 will yield β^* values between .03 and .055. Consequently, if BZS had used the values of Y to classify children as extremely able ($Y \geq a$), and then estimated the slope β^* of Y on X in the truncated (extremely able) population, the low slopes which they found would be explainable as a mathematical consequence of the truncation of the population, rather than because of a poor choice of estimator.

Having said this, it should be repeated that BZS did not use Y to classify children as extremely able. Thus, Vining's criticisms are irrelevant.

3. CRITIQUE OF BZS

The purpose of this section is to point out some problems connected with the approach used by BZS, and to give some suggestions for future studies.

The Model

The criterion used by BZS for identifying "exceptionally able" children is quite arbitrary. One could just as easily define "extremely able" in terms of any other upper percentile of the distribution of the "precocity quotient" Z.[d/]

Further, for any classification rule which assigns a child to an "extremely able" subpopulation on the basis of a not perfectly reliable test, there is the possibility of misclassification error. That is, repetition of the study on the same cohort of children might result in a somewhat different set of children being chosen to represent the "extremely able" subpopulation.

Consequently, it is not clear that the "extremely able" subpopulation of BZS is well-defined for scientific purposes.[e/]

Measures of Familiality

BZS use the slope of Y = child's score on X = parent's score as an index, or measure of familiality, for a given child-parent population on a given test. In so doing, they appear to be following an already established tradition in this area of research.

In the terminology used by BZS, "familiality" refers to the amount of "resemblance" between parent and child. A natural measure of resemblance is the magnitude $|Y-X|$ of the difference in test scores between child and parent. For its greater mathematical convenience, $(Y - X)^2$ is usually used in place of $|Y - X|$. Within a given child-parent population, the mean of $(Y - X)^2$,

$$\Delta = E(Y - X)^2 = \text{Variance } (Y - X) + \mu^2_{(Y - X)} , \tag{5}$$

serves as an intuitively meaningful index of child-parent resemblance.

If in the ith child-parent population under study, Y and X have means μ_{yi}, μ_{xi} and variances σ^2_{yi}, $\sigma_{xi}{}^2$, respectively, and if the slope of the linear regression of Y on X is β_i, then our index Δ_i of child-parent resemblance in this population satisfies:

$$\Delta_i = \sigma_{yi}{}^2 + \sigma^2_{xi} - 2\beta_i \, \sigma^2_{xi} + (\mu_{yi} - \mu_{xi})^2. \tag{6}$$

In typical studies, the same test is given to all children and all parents in all populations. If the test is also similarly standardized in all populations, so that

$$\sigma^2_{yi} = \sigma^2_y, \; \sigma^2_{xi} = \sigma^2_x, \; \mu_{yi} = \mu_y, \; \mu_{xi} = \mu_x, \; \text{all } i , \tag{7}$$

then

$$\Delta_i = \sigma_y^2 + \sigma_x^2 - 2\beta_i\sigma_x^2 + (\mu_y - \mu_x)^2 \quad (8)$$

varies only as the slopes β_i in the different populations vary - the greater the slope, the greater the resemblance. Consequently, when comparing familiality across these populations for a given common test, the slope β_i can be used in place of the more obviously meaningful measure Δ_i as an index of child-parent resemblance.

Unfortunately, in the BZS study, the same test is not given to the two populations being compared. Further, BZS do not compare slopes test by test, but instead compare the median slope across a class of diverse tests in one population to the range of slopes over a different class of tests for the other population. It is therefore far from clear that BZS's comparisons yield any insight into differences of child-parent resemblance (familiality) between the two populations.

An alternative measure of child-parent resemblance in population i would be to compare the child's standardized score $\sigma_{yi}^{-1}(Y - \mu_{yi})$ to the parent's standardized score $\sigma_{xi}^{-1}(X - \mu_{xi})$. Thus, one could use the index

$$\xi_i = E\left(\frac{Y - \mu_{yi}}{\sigma_{yi}} - \frac{X - \mu_{xi}}{\sigma_{xi}}\right)^2 = 2(1 - \rho_i), \quad (9)$$

where ρ_i is the correlation between Y and X in population i, as a measure of familiality for population i. Note that the index ξ_i compares the child's relative position in the population of children to the parent's relative position in the population of parents, rather than comparing their actual scores. Arguments favoring each of the indices Δ_i, ξ_i as appropriate indices of familiality can be given. One great advantage of ξ_i is, of course, that it is scale-free. Note from (9) that ξ_i is a function only of the correlation ρ_i between Y and X in population i. Consequently, BZS might have avoided the criticism given above concerning their use of slopes as indices of familiality, by instead using correlations.

Consequences of Joint Normality Assumptions

Let us suppose for the moment that the test scores Y, X, Z are perfectly reliable. A common assumption in psychometric research is that Y, X, Z have a trivariate normal distribution. In this case, the conditional joint distribution of (Y, X) given Z = z is bivariate normal with means

$$\mu_{yz} = \alpha_y + \gamma_y z \ , \quad \mu_{xz} = \alpha_x + \gamma_x z \ ,$$

and variances and correlations independent of the value of z:

$$\sigma^2_{yz} = \sigma^2_{y\cdot} \ , \ \sigma^2_{xz} = \sigma^2_{x\cdot} \ , \ \rho_z = \rho_{\cdot} \quad , \text{ all } z.$$

Note that each value of z defines a subpopulation of child-parent pairs, and BZS's definition of "exceptionally able" children simply groups together all of the subpopulations for which z exceeds the 99.97th percentile z* of the distribution of Z.[f/]

For the subpopulation of child-parent pairs defined by the value z of Z, the index Δ_z of familiality defined by (5) is given by

$$\Delta_z = \sigma^2_{y\cdot} + \sigma^2_{x\cdot} - 2\rho_{\cdot}\, \sigma_{y\cdot}\, \sigma_{x\cdot} + [\alpha_y - \alpha_x + (\gamma_y - \gamma_x)z]^2 \ ,$$

while the index of familiality (9) is given by

$$\xi_z = 2(1 - \rho_{\bullet}).$$

Consequently, there are <u>no</u> differences in the indices of familiality ξ_z across the subpopulations indexed by z, and differences in the indices of familiality Δ_z across the subpopulations will exist if and only if the slope γ_y of child's score Y on child's precocity quotient Z differs from the slope γ_x of parent's score X on Z. It is hard to see what the difference $\gamma_y \neq \gamma_x$ has to do with familiality; it is more easily explained as a predictable difference between the association between two scores Y, Z obtained from the same person (the child), and the association between scores X, Z obtained from different persons.

The subpopulations of child-parent scores (Y, X) indexed by individual values z of Z are free of criticisms concerning arbitrariness. Certainly, any discussion of differences in familiality across child ability levels should start with this infinite collection of subpopulations. In their study, BZS have simply formed two larger populations out of these subpopulations by grouping z-values. However, we have seen above that if the usual assumptions of joint normality of the distribution of (Y, X, Z) is true, there can be no differences in relative-score familiality ξ_z across these subpopulations, and the only differences in absolute-score familiality Δ_z occur because of the fact that Y and Z are measured on the same person. Anyone, therefore, who believes that (Y, X, Z) are jointly normally distributed over the entire child-parent population must be prepared to deny that differences in familiality across ability levels exist. Any differences found when subpopulations are grouped (as in BZS's study) must be attributed to artifacts created by grouping.[g/]

Contrariwise, researchers who believe that differences in familiality exist across ability levels must avoid the assumption that (Y, X, Z) are jointly normally distributed. This will therefore require new models for the joint distribution of (Y, X, Z), which will need to be tested for goodness-of-fit against existing data. BZS avoid any explicit distributional assumptions, but the tests of correlation and slope in their Table 1 (p. 157) at least implicitly rest upon assumptions of normality.

In finding new models, it is important to remember that Y, X, Z are measured with error. Thus,

$$Y = y + e_1, \quad X = x + e_2, \quad Z = z + e_3 ,$$

and we are interested in modelling the conditional joint distribution of the true scores y, x given the true score z in such a way that the partial correlation ρ_z between y and x given z varies with z. Constructing such models, and validating them on data, will be an interesting and important area of research, which is likely to yield models of use in other scientific contexts.

Controlling For Other Factors

We have previously noted that different collections of tests were used for the two populations (actually subpopulations) compared by BZS. This approach clearly has the potential of confounding differences in familiality with differences in tests! A quick look at Table 1 in BZS shows substantial differences in slopes across the 9 cognitive tests within the same subpopulation.

BZS can respond that to give the same test to both the average and the exceptionally able subpopulations would run the risk of "ceiling" or "floor" effects. That is, either the test would be too hard for the average subpopulation or too easy for the exceptionally able subpopulation. One way around this problem might be to use adaptive tests, such as those pioneered by F. M. Lord, where a test administrator sequentially administers items from a large pool of similar items which vary widely in difficulty. The initial choice of items is designed to determine the approximate level of ability of the subject; once this is done, all remaining items are chosen at an appropriate level of difficulty for that ability, enabling the test administrator to "home in" on a precise estimate of the subject's ability. Some adaptive tests have already been developed (some can be given utilizing interactive computer programs), and standardized.

Even if the same test cannot be used across subpopulations, investigators should certainly attempt to control the nature of the subject matter of the tests used. Observe from Table 1 in BZS that slopes for verbal tests are substantially higher than slopes for mathematics tests (with the interesting exception of the "Cubes" test). Also the CTMM-Language, General Information,and Cubes slopes are either in or nearly in the range of slopes (.42 - .60) reported for unselected populations. Although BZS report on these differences in slope by subject matter for the "exceptionally able" group, they unfortunately do not similarly separate slopes by subject matter for the unselected populations, forcing readers to seek out the original references to find such information.

Although BZS obviously realize that sex is a factor in these studies, they do not give slopes by sex, due to the small number of girls in the study. Yet for a study of familiality, it would certainly be of interest to compare the associations between boy-mother, boy-father, girl-mother, girl-father both within and across subpopulations, and also across a variety of subject matter tests.

Midparent Score

BZS define the midparent score as the simple average of X_1 = father's score and X_2 = mother's score. Since parents do not necessarily have equal influence on their child, it might be more appropriate to define the midparent score as the weighted sum

$$X = a_1X_1 + a_2X_2 \ , \ a_1 + a_2 = 1 \ ,$$

having maximum correlation with Y. If b_1 and b_2 are the slopes of the multiple regression of Y on X_1, X_2, then

$$a_i = \frac{b_i}{b_1+b_2} \ , \ i = 1,2,$$

and the slope b of Y on the midparent score X is

$$b = b_1 + b_2 \ .$$

Further, the correlation between Y and X is the multiple correlation $R_{y \cdot x_1, x_2}$ between Y and X_1, X_2.

It should be noted that in cases where b_1 and b_2 are close to equal, the correlation between Y and the usual midparent score $\frac{1}{2}(X_1 + X_2)$ will be higher than the correlation of Y with either parent score (X_1 or X_2), since the multiple correlation $R_{y \cdot x_1, x_2}$ is always larger than the correlations $\rho_{x_1 y}$, $\rho_{x_2 y}$. The reader should also recall that b_i will not be the same as the slope of the simple regression of Y on X_i , i = 1,2.

Summary

The above discussion has pointed out two key elements needed for any future studies of familiality:

(1) Clear and non-arbitrary definitions of populations or subpopulations to be compared, preferably defined by models giving the conditional joint distribution of child's and parent's true score x, y given the child's true ability score z,

(2) Use of common tests across the populations, or at a minimum, control of subject matter content for such tests.

More sophisticated designs and/or comparisons which isolate and compare effects due to test subject matter content, child's sex, and parental influences are also desirable. Although arbitrary groupings of subpopulations indexed by z (such as used by BZS) may be necessary in early studies, so that models can be tested for fit, one would hope that later studies would utilize models which clearly show familiality as a function of z = child's ability true score.

Some discussion of indices of familiality has also been given. The index $\Delta = E(Y - X)^2$ may be most meaningful as a comparison of Y = child's score and X = parent's score, but is scale dependent. The scale free index

$$\xi = E\left(\frac{Y-\mu_y}{\sigma_y} \quad \frac{X-\mu_x}{\sigma_x}\right)^2$$

compares only relative scores, but is scale free. Since ξ is a function only of the correlation ρ between Y and X, investigators may prefer to report ρ as an index of familiality, rather than the slope of Y on X.

Finally, a definition of midparent score which accounts for the possibly differing influences of father and mother on children has been suggested.

REFERENCES

Benbow, C. P., Stanley, J. C., Kirk, M. K. & Zonderman, A. B.(1983). Structure of intelligence in intectually precocious children and in their parents. Intelligence, 7, 129-152.

Benbow, C. P., Zonderman, A. B., & Stanley, J. C. (1983). Assortative marriage and the familiality of cognitive abilities in families of extremely gifted students. Intelligence, 7, 153-161.

Skinner, C. J. (1984). On the geometric approach to multivariate selection. Psychometrika, 49, 383-390.

Vining, D. R. Jr. (1985). Familiality estimates from restricted samples. Intelligence, 9, 15-21.

FOOTNOTES

*My thanks to Professors Julian C. Stanley and Camilla Persson Benbow for bringing their papers and that of Vining to my attention, and for helpful discussion. Correspondence and requests for reprints should be addressed to Leon Jay Gleser, Department of Statistics, Purdue University, West Lafayette, IN 47907.

a/ Vining actually discusses estimates of ρ. However, since he assumes that Y and X-scales are standardized to have identical variances in the population, it follows that $\rho = \beta$, and hence Vining's estimates also estimate β.

b/ It would have been helpful if BZS had tested their sample of (Y,X) pairs for goodness-of-fit to the bivariate normal distribution.

c/ If the correlation of Z with Y, and the resulting reduction in variation of Y, were the sole factor in accounting for differences in the slope of Y on X over tests, then one would expect these slopes to vary inversely with the correlation between Y and Z over the 9 tests studied by BZS. That this is not the case can be seen by looking at Table 2 in Benbow, Stanley, Kirk and Zonderman (1983) to find the Z-Y correlations, and then looking at Table 1 in BZS. In fact, CTMM has the largest Z-Y correlation and one of the largest slopes for Y on X, while Semantic comprehension has the third largest Z-Y correlation and the smallest slope for Y on X.

d/ There is also the problem that the choice of which SAT test (SAT-V, SAT-M, SAT-M + SAT-V) to use to compute the "precocity quotient" Z varied with the child (see Benbow, Stanley, Kirk and Zonderman, 1983, pp. 130-131). Henceforth, in the hope of improved practice in studies of this sort, we will assume in our discussion that Z is a score obtained from a fixed test given to all children. (In their *subsequent research*, Benbow and Stanley are using this type of criterion, e.g., "700-800 on SAT-M before age 13.")

e/ BZS uses the terminology "gifted" as a synonym for "extremely able." Webster's New Collegiate Dictionary defines "gifted" as "possessing a special talent or aptitude." One can hardly assert that all children not classified as "extremely able" by BZS are not gifted (or even not "extremely gifted"). BZS are not unique in the use of the terminology "gifted." It would be desirable for researchers in the field of intelligence to use more precise, and less semantically loaded, adjectives.

f/ Note also that $\rho_{\cdot}$ is the partial correlation between Y and X given Z, that $\sigma^2_{y\cdot}$, $\sigma^2_{x\cdot}$ are respectively the conditional variances of Y, X given Z, and that γ_y, γ_x are the regression slopes of Y, X on Z.

g/ If Y and Z are the same test score (as assumed by Vining), this assertion still applies. Thus, Vining's model of bivariate normality for (Y, X) contradicts his later finding of differences in familiality.