

ESTIMATION OF THE PROPORTION OF GENETIC DEVIATES
UNDER THE ASSUMPTIONS OF NORMALITY

BU-143-M

K. H. Thompson and W. T. Federer

February, 1962

ABSTRACT

A simple statistical procedure for estimating the proportion of plus (or negative) genetics deviates in univariate normal populations is described and illustrated with a numerical example. The procedure for the bivariate normal situation is discussed, but no numerical illustration of the results was obtained. The points of intersection for two bivariate normals projected on the plane defined by the two variables is an ellipse. The difference in areas beyond the points of intersection for the pair of bivariate normals yields the proportion of genetic deviates of the specified type.

*No. BU-143-M in the Biometrics Unit, Cornell University, Mimeo Series

ESTIMATION OF THE PROPORTION OF GENETIC DEVIATES
UNDER THE ASSUMPTIONS OF NORMALITY*

BU-143-M

K. H. Thompson and W. T. Federer

February, 1962

In order to make genetic advance in the mean value of a character under selection it is essential that genetic variation be present. Given the genetic and the environmental variances and that the yields are normally distributed it is possible to estimate the proportion of genetic deviates in a population which exceed some prescribed value, say Z_0 or Z_1 . Also, given two genetically variable populations with the same environmental variance it is possible to compare the relative proportion of genetic deviates exceeding some prescribed value, say Z_1 , in one population as compared to the other. In the past, geneticists have fitted normal distributions to each of the two populations and then have compared the relative proportions in given class intervals. This is a somewhat laborious procedure, and the purpose of this note is to describe a simplification which is simple computationally. The procedure described relies on the normality assumption concerning yields and on existing tables of the unit normal deviate. The results are illustrated with an example.

Selection on two characters simultaneously would involve use of the bivariate normal distribution and the corresponding tables. Some discussion of this problem is presented herein, but a full solution was not obtained. No numerical application of these results is given, as further work is required.

Univariate Normal Distribution

If the assumption of normality of the distribution of observations for a given character, say X_1 , is tenable, then a relatively simple procedure exists for estimating the proportion of genetic deviates in a genetically variable population which exceeds some prescribed value. Suppose that the estimated environmental variance for X_1 is s_e^2 and suppose that the estimated phenotypic variance is $s_p^2 = s_e^2 + s_g^2$ = environmental variance plus genetic variance. Now if two

* No. BU-143-M in the Biometrics Unit, Cornell University, Mimeo Series, February 1962.

normal curves are centered at μ and if one has variance s_e^2 and the other has variance $s_p^2 > s_e^2$, they could be represented as in Figure 1. These two curves

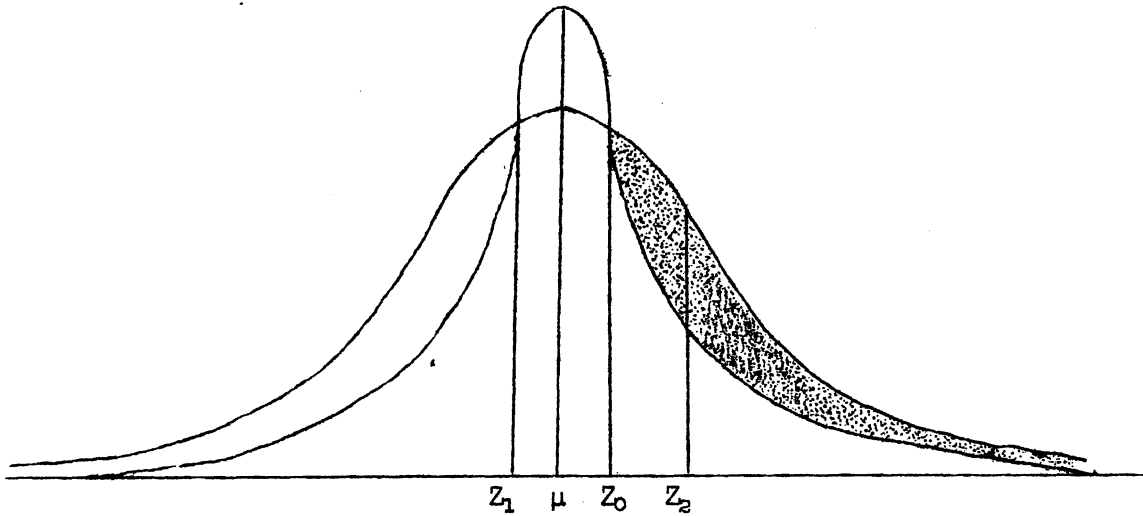


Figure 1. Two normal curves centered at μ and with variances s_e^2 and $s_p^2 > s_e^2$.

intersect at the points Z_1 and Z_0 . Now at either point the ordinates of the curves are equal and the two points Z_1 and Z_0 are identical except for sign. Therefore, setting the ordinates equal we obtain:

$$\frac{1}{\sqrt{2\pi s_e^2}} e^{-(Z_0 - \mu)^2 / 2s_e^2} = \frac{1}{\sqrt{2\pi s_p^2}} e^{-(Z_0 - \mu)^2 / 2s_p^2}$$

Taking logarithms of both sides results in:

$$\log s_e^2 + (Z_0 - \mu)^2 / s_e^2 = \log s_p^2 + (Z_0 - \mu)^2 / s_p^2$$

Solving,

$$Z_0 - \mu = \sqrt{\frac{s_e^2 s_p^2}{s_e^2} \log(1 + s_e^2 / s_p^2)} = \frac{s_e s_p}{s_e} \sqrt{\log(s_p^2 / s_e^2)}$$

where $s_g^2 = s_p^2 - s_e^2$.

The difference in the tail areas between these two normal distributions may be expressed as

$$p \left\{ \frac{X_1 - \mu}{s_p} > \frac{Z_0 - \mu}{s_p} \right\} - p \left\{ \frac{X_1 - \mu}{s_e} > \frac{Z_0 - \mu}{s_e} \right\}$$

and obtained from normal deviate tables.

As a numerical illustration of the above suppose that $s_e^2 = 0.509$, $s_p^2 = 0.911$ and $s_g^2 = 0.402$. Then $Z_0 - \mu$ is computed as:

$$Z_0 - \mu = \sqrt{\frac{.509(.911)}{.402} \log(.911/.509)} = 0.82$$

Therefore, for $\mu = 2.60$, $Z_0 = 2.60 + 0.82 = 3.42$. The area to the right of $\frac{Z_0 - \mu}{s_e} = \frac{0.82}{\sqrt{0.509}} = 1.15$ is equal to 12.5% (from normal deviate tables) and the area to the right of $\frac{Z_0 - \mu}{s_p} = \frac{0.82}{\sqrt{0.911}} = 0.86$ is equal to 19.5%. Therefore, the latter area minus the former is an estimate of the proportion of positive genetic deviates in the population, i.e., $19.5\% - 12.5\% = 7.0\%$.

Taking any other point, say $Z_2 = 4.00$, the proportion of plus genetic deviates is estimated by

$$p \left\{ \frac{X_1 - \mu}{s_p} > \frac{Z_2 - \mu}{s_p} \right\} - p \left\{ \frac{X_1 - \mu}{s_e} > \frac{Z_2 - \mu}{s_e} \right\}$$

For the above example, the area to the right of $\frac{Z_2 - \mu}{s_e} = \frac{1.4}{\sqrt{.509}} = 1.96$ is 2.5% and the area to the right of $\frac{Z_2 - \mu}{s_p} = \frac{1.4}{\sqrt{0.911}} = 1.47$ is 7.1%. Therefore the estimated proportion of plus genetic deviates to the right of $Z_2 = 4.00$ is $7.1\% - 2.5\% = 4.6\%$.

Bivariate Normal Distribution

Suppose that we have a bivariate normal distribution for two characters X_1 and X_2 with parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$, and ρ_{12} and suppose that this curve represents the distribution when only environmental effects are present. Now suppose that both environmental and genetic effects are present and the parameters of the bivariate normal distribution for the two characters X_1 and X_2 are $\mu_1, \mu_2, \sigma_3^2 > \sigma_1^2, \sigma_4^2 > \sigma_2^2$, and ρ_{34} . The intersection of the two bivariate normals is obtained by equating the two bivariate normals, thus:

$$\begin{aligned} & \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} e^{-\frac{1}{2}\left\{\frac{(Z_1-\mu_1)^2}{\sigma_1^2} + \frac{(Z_2-\mu_2)^2}{\sigma_2^2} - 2\frac{\rho_{12}(Z_1-\mu_1)(Z_2-\mu_2)}{\sigma_1\sigma_2}\right\}} \\ &= \frac{1}{2\pi\sigma_3\sigma_4\sqrt{1-\rho_{34}^2}} e^{-\frac{1}{2}\left\{\frac{(Z_1-\mu_1)^2}{\sigma_3^2} + \frac{(Z_2-\mu_2)^2}{\sigma_4^2} - 2\frac{\rho_{34}(Z_1-\mu_1)(Z_2-\mu_2)}{\sigma_3\sigma_4}\right\}} \end{aligned}$$

Taking logarithms of both sides and rearranging terms we obtain:

$$\begin{aligned} & (Z_1-\mu_1)^2 \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_3^2} \right) + (Z_2-\mu_2)^2 \left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_4^2} \right) \\ & - 2(Z_1-\mu_1)(Z_2-\mu_2) \left(\frac{\rho_{12}}{\sigma_1\sigma_2} - \frac{\rho_{34}}{\sigma_3\sigma_4} \right) + \log \sigma_1^2\sigma_2^2(1-\rho_{12}^2)/\sigma_3^2\sigma_4^2(1-\rho_{34}^2) = 0 \end{aligned}$$

or rewriting,

$$C_1 W_1^2 + C_2 W_2^2 - C_3 W_1 W_2 + C_4 = 0$$

where the constants C_i are functions of the variances and correlations and the $W_1 = Z_1 - \mu_1$. The projection of the intersection on the $X_1 X_2$ plane yields an ellipse. The area between the two bivariate normals in any given quadrant yields the proportion of genetic deviates of the desired type. Thus, if high values of both X_1 and X_2 were considered desirable then only positive roots for Z_1 and Z_2

in the preceding quadratic equation would be used; if low X_2 values and high X_1 values were desired then negative solutions for Z_2 and positive solutions for Z_1 would be the admissible values in the preceding quadratic equation, and so on.

As pointed out by Cramer, *Mathematical Methods of Statistics*, page 290, the mass in the first and third quadrants for the above pair of bivariate normals would be $\frac{1}{4} + \frac{1}{2\pi} \arcsin \rho_{12}$ and $\frac{1}{4} + \frac{1}{2\pi} \arcsin \rho_{34}$, respectively, and the mass in the second and fourth quadrants would be $\frac{1}{4} - \frac{1}{2\pi} \arcsin \rho_{12}$ and $\frac{1}{4} - \frac{1}{2\pi} \arcsin \rho_{34}$, respectively. Hence a rough approximation of the proportion of genetic deviates in a given quadrant would be obtained as the difference of means for the corresponding bivariate normals. A more precise evaluation would be obtained from

$$\begin{aligned}
 & + p \left\{ \frac{X_1 - \mu_1}{\sigma_1} \leq \frac{Z_1 - \mu_1}{\sigma_1}, \frac{X_2 - \mu_2}{\sigma_2} \leq \frac{Z_2 - \mu_2}{\sigma_2} \right\} \\
 & - p \left\{ \frac{X_1 - \mu_1}{\sigma_3} \leq \frac{Z_1 - \mu_1}{\sigma_3}, \frac{X_2 - \mu_2}{\sigma_4} \leq \frac{Z_2 - \mu_2}{\sigma_4} \right\} \\
 & + \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \int_{-\infty}^{W_1} \int_{-\infty}^{W_2=f(W_1)} e^{-\frac{1}{2}\left\{\left(\frac{u}{\sigma_1}\right)^2 + \left(\frac{v}{\sigma_2}\right)^2 - 2\frac{\rho_{12}uv}{\sigma_1\sigma_2}\right\}} \\
 & - \frac{1}{2\pi\sigma_3\sigma_4\sqrt{1-\rho_{34}^2}} \int_{-\infty}^{W_1} \int_{-\infty}^{W_2=f(W_1)} e^{-\frac{1}{2}\left\{\left(\frac{u}{\sigma_3}\right)^2 + \left(\frac{v}{\sigma_4}\right)^2 - 2\frac{\rho_{34}uv}{\sigma_3\sigma_4}\right\}}
 \end{aligned}$$

where $f(W_1) = C_3 W_1 \pm \sqrt{C_3^2 W_1^2 - 4C_2(C_1 W_1^2 + C_4)}/2C_2 = W_2$ and where only positive or negative roots of W_1 and W_2 are allowed depending on which quadrant one is interested in.

A numerical example has not yet been worked, but it should be possible to do this by using published results, e.g., *National Bureau of Standards Applied Mathematics Series 50* (U.S. Government Printing Office, 1958).