their interaction with locations and years. Agron. J. 44, 462–466 (1952). — SPRAGUE, G. F.: Early testing of inbred lines of corn. J. Am. Soc. Agron. 38, 108–117 (1946). — SPRAGUE, G. F., and W. T. FEDERER: A comparison of variance components in corn yield trials II. Error, year × variety, location × variety and variety components. Agron. J. 43, 535–541 (1951). — 20. SPRAGUE, G. F., and L. A. TATUM: General and specific combining ability in single crosses of corn. J. Am. Soc. Agron. 34, 923–932 (1942). — 21. STRAIN, J. H.: Genotype-environmental interactions in caged layers. M. S. thesis, Iowa State Univ. Library (1960.)

# The Proportion of Genetic Deviates in the Tails of a Normal Population*[1]

D. S. ROBSON[2], LEROY POWERS[3] and N. S. URQUHART[4]

**Summary.** If genetic and environmental effects upon a quantitative phenotype $X = G + E$ are normally and independently distributed then the probability distribution of genetic value $G$ among individuals of fixed phenotypic value $X$ is likewise a normal distribution. The mean of this *a posteriori* distribution of genetic values is $\bar{g} + h^2 (X - \bar{g})$ and the variance is $\sigma_g^2 (1 - h^2)$, where $\bar{g}$ is the *a priori* mean of $X$, $h^2$ is the heritability ratio, and $\sigma_g^2$ is genetic variance. For any fixed values of $h^2$ and $\sigma_g^2$ the *a posteriori* probability that the genetic value $G$ associated with a given phenotype $X$ exceeds the population mean by any specified amount can therefore be read directly from the tables of the standard normal distribution. The expected proportion of these superior genetic deviates among individuals whose phenotypic value exceeds some specified constant may also be calculated (by numerical analysis) and is presented here in graphical form.

If phenotypic selection is practiced by choosing the best out of $N$ phenotypes then $N$ should be large enough to assure high probability of obtaining a superior genetic deviate. The operating characteristics of this type of selection are displayed in tabular form, again based upon numerical integration.

## Introduction

The phenotypic array exhibited by a segregating genetic population reflects both the genetic and the environmental variability within the population. As a consequence, an element of uncertainty attaches to selection for genetically superior individuals on the basis of their phenotypic traits. The latter may, by chance, be merely the result of an unusually favorable environment acting upon a genotype which under less favorable conditions would display only a mediocre or even undesirable phenotype. Chances for the occurrence of such phenotypic deception depend, of course, upon the magnitude of variability environmentally induced as compared to that caused by genetic differences.

Any mathematical formulation of this problem to enable the geneticist to evaluate numerically his chances for successful selection would require a detailed description of the phenotypic frequency distribution in the population. The total segregating population is a mixture of subpopulations, each of which represents the distribution of phenotypes produced by a single genotype under the existing range of environmental conditions. Each subpopulation or genotype contributes to the total population in proportion to its genotypic frequency. A mathematical description of the population consists, therefore, of specifying the relative frequency of each genotype and the exact form of its associated distribution of phenotypes.

Usually, when selection is practiced for economic purposes, a large number of both genetic and environmental factors operate at variable levels to determine the phenotypes in the population. Empirical evidence supports the belief that, in this case, the total-frequency distribution and also the component distributions for a quantitative phenotypic trait are approximately Gaussian in form. A standard population model has therefore come into use for such problems as the prediction of advancement under selection. It is EISENHART's Model II (1947) which represents, in the simplest case, a normal mixture of normal subpopulations with constant variance. Each genotype is assumed to generate a normal distribution of phenotypes under the existing range of environmental conditions. The distribution of phenotypic means (called genotypic values) is likewise normal.

## Graphs of the Proportion of Genetic Deviates

The phenotypic value $X$, for some quantitative trait of an individual selected at random from a genetic population, may be regarded, conceptually, as the sum of two components:

$G$ = average phenotype for the genotype of the chosen individual

$E$ = deviation of the particular phenotype of the chosen individual from the average phenotype ($G$) for the genotype of that individual = $X - G$

or

$$X = G + E .$$

The first component $G$ is conventionally called the genotypic value. $E$ is the environmental effect. If the population structure is a normal mixture of normal subpopulations having a common environmental variance, then the chance variables $G$ and $E$ follow independent normal distributions. $G$ has a mean value of $\bar{g}$ and variance $\sigma_g^2$. $E$ has a mean value of zero and variance $\sigma_e^2$. Thus $X$ itself follows a normal distribution, with mean $\bar{g}$ and variance $\sigma_g^2 + \sigma_e^2$.

* Dedicated to Dr. GEORGE F. SPRAGUE on the occasion of his 65th birthday.

[1] Cooperative investigations of the Colorado Agricultural Experiment Station, the Crops Research Division, Agricultural Research Service, U.S. Department of Agriculture, and the Beet Sugar Development Foundation. Approved by the Colorado Agricultural Experiment Station for publication as Scientific Series Article No. 880. Paper No. BU-78, Biometrics Unit, and Paper No. 529, Plant Breeding Department, Cornell University.

[2] Professor of Biological Statistics, Cornell University.

[3] Geneticist, Crops Research Division, Agricultural Research Service, U.S. Department of Agriculture; now deceased.

[4] Formerly Research Assistant, Colorado State University; now Assistant Professor of Biological Statistics, Cornell University.

A probability distribution of particular interest to the geneticist is the a posteriori distribution of $G$ among individuals of a fixed phenotype $x'$; that is, given that he has selected an individual of phenotype $x'$, the geneticist is then concerned with the probability that this chosen individual is of superior genotype, say greater than some standard value $g$. For any selected phenotypic value $x'$, the conditional distribution of genotypic values $G$ is in this case normal with mean $\bar{g} + h^2(x' - \bar{g})$ and variance $\sigma_g^2(1 - h^2)$ as shown in appendix 1. Here, $h^2$ is the *heritability ratio*

$$h^2 = \frac{\sigma_g^2}{\sigma_e^2 + \sigma_g^2}$$

or the ratio of genetic to total variance in the population.

The desired probability that the genotypic value will exceed some specified value $g'$ is derived in appendix 2. If $\Phi$ is the standard (cumulative) normal distribution function, and if $x'$ and $g'$ are expressed in standard units as

$$x = \frac{x' - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}} \quad \text{and} \quad g = \frac{g' - \bar{g}}{\sigma_g},$$

then

$$P(G > g' | x) = \Phi\left(\frac{xh - g}{\sqrt{1 - h^2}}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{xh-g}{\sqrt{1-h^2}}} e^{-t^2/2}\, dt. \quad (1)$$

This depends on the variance components $\sigma_e^2$ and $\sigma_g^2$ only through the heritability ratio. A special case of some interest is where $g' = \bar{g}$, and thus $g = 0$, giving $\Phi\left(x\Big/\sqrt{\frac{1 - h^2}{h^2}}\right)$ as the probability that an individual of phenotype $x'$ will be genetically above average; note that when $h^2 = 1/2$ this reduces to $\Phi(x)$ = area to the left of $x$ under the standard normal curve.

For example, with $x = 1.645$ and $h^2 = 1/2$, the probability that the genotypic value of the selected individual exceeds the population mean is $\Phi(1.645) = .95$; if $h^2$ were 1/4 instead of 1/2 then the desired probability would be $\Phi(1.645\sqrt{3}) = \Phi(.950) = .83$, while $h^2 = 3/4$ would give $\Phi(1.645\sqrt{3}) = \Phi(2.849) = .998$. These answers may be interpolated from Figure 1, where the general form of the solution is illustrated by plotting $\Phi([xh - g]/\sqrt{1 - h^2})$ as a function of $g$ with $x$ fixed at the upper 5 percent tail value of 1.645 and with $h^2$ taking the values 0.1 to 0.9 by steps of 0.1.

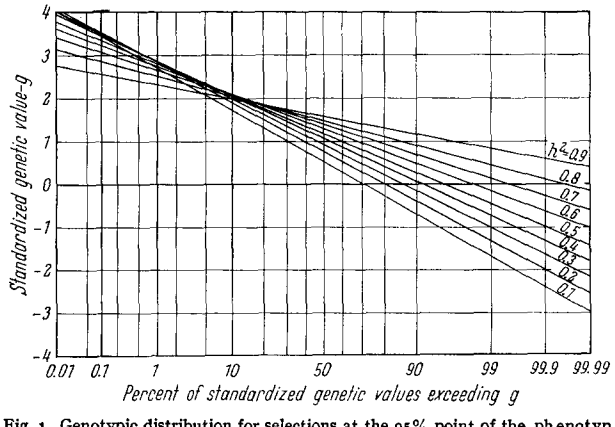


Fig. 1. Genotypic distribution for selections at the 95% point of the phenotyp distribution.

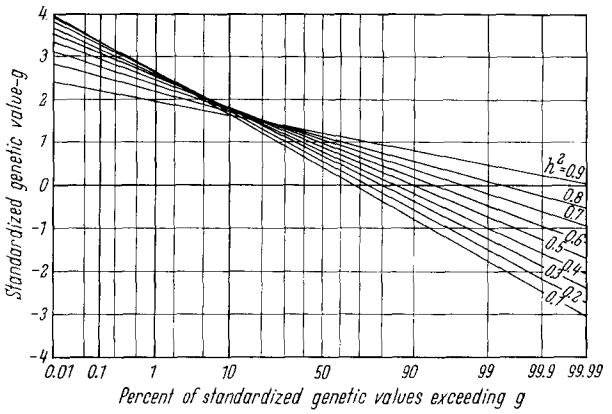


Fig. 2. Genotypic distribution for selections at the 90% point of the phenotypic distribution.

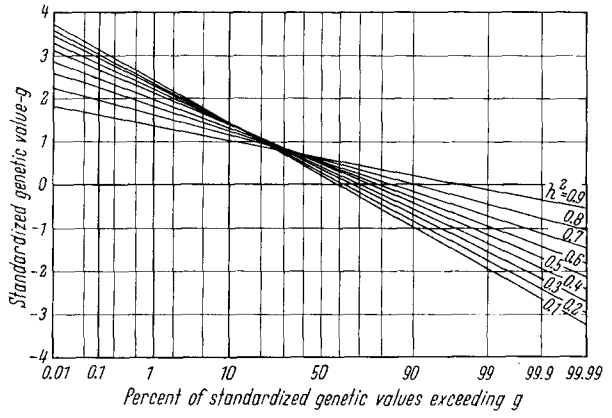


Fig. 3. Genotypic distribution for selections at the 75% point of the phenotypic distribution.

Figures 2 and 3 display the corresponding curves, with $x$ fixed at the upper 10 percent value of 1.282 and the upper quartile value of .674, respectively.

Where selection operates on the entire upper tail of the phenotypic distribution, taking all individuals whose phenotypic value exceeds $x'$, the question of interest becomes the proportion of these selected individuals having genotypic values exceeding $g'$. More precisely, this proportion may be interpreted as the conditional probability that an individual selected at random, and exhibiting a phenotypic value greater than $x'$, will also have a genotypic value greater than $g'$. The solution to this problem is obtained in appendix 3 and again is expressible in integral form as

$$P(G > g' | X > x') = P_{g \cdot x} = \frac{1}{1 - \Phi(x)} \int_x^{\infty} \Phi\left[\frac{hy - g}{\sqrt{1 - h^2}}\right] d\Phi(y) \quad (2)$$

though in this case the integral is not previously tabulated. The only nontrivial case which permits further analytic reduction is $g = 0$ and $h^2 = 1/2$. When half of the total variance is genetic, then the probability that an individual, selected at random from the region $X > x'$, will be genotypically above average is $[1 + \Phi(x)]/2$. (This is established in appendix 3.) Thus, the solution here may be read directly from the cumulative normal tables; for example, the proportion of above-average genotypes in the upper 5 percent tail of the phenotypic distribution is

$$[1 + \Phi(1.645)]/2 = [1 + .95]/2 = .975.$$

A numerical solution in the general case can be obtained only by numerical integration, for which the most convenient approach is to hold $g$ and $h^2$ fixed, and allow $x$ to vary. Graphs of this probability function of $g$, appearing in Figures 4—12, are therefore plotted as functions of $x$, with each curve corresponding to a fixed value of $g$ and $h^2$. For example, from the curve for $P = .05$ (corresponding to $g = 1.645$) and $h^2 = 1/2$, Figure 8, we see that in the upper 25 percent tail of the phenotypic distribution ($x = .674$) the proportion of standardized genetic deviates exceeding $g = 1.645$ is .17. In other words, 17 percent of the phenotypic top-25-percent also belong to the genotypic top-5-percent of the population, when half of the variability is genetic.

A more complete picture of the expected proportions of genetic deviates may be constructed for each value of $h^2$ by a slight extension of the results in Figures 4—12. The case when half of the variance is genetic, $h^2 = 1/2$, serves to illustrate this extension of Figure 8, as summarized in Table 1. The other Figures can be extended in the same fashion.

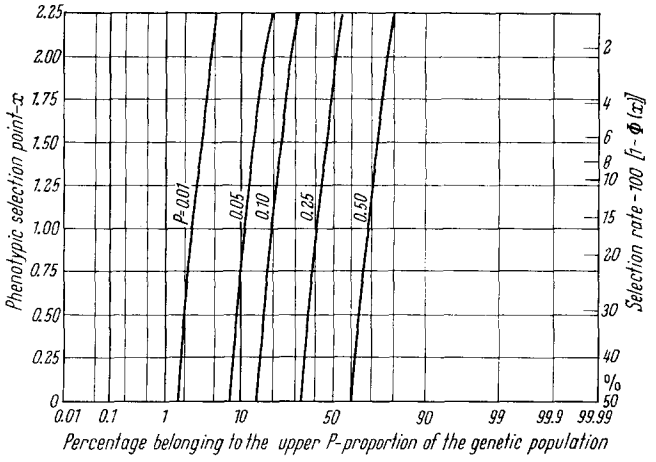


Fig. 4. Relationship between phenotypic truncation point and genetic value when $h^2 = 0.1$.

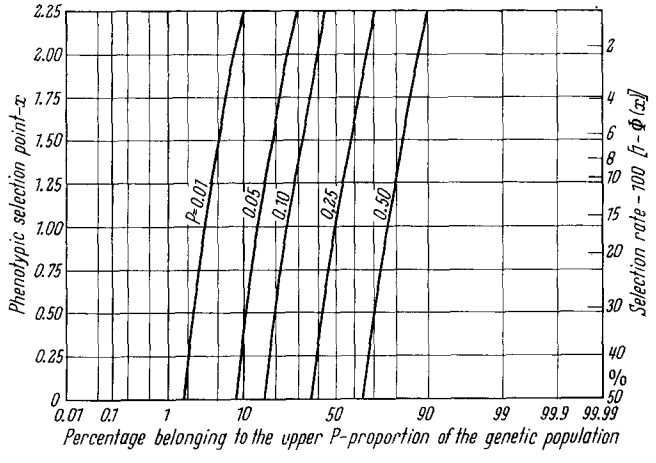


Fig. 5. Relationship between phenotypic truncation point and genetic value when $h^2 = 0.2$.

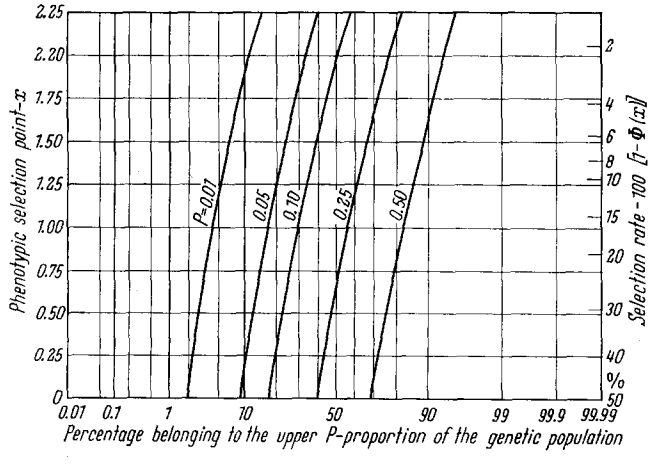


Fig. 6. Relationship between phenotypic truncation point and genetic value when $h^2 = 0.3$.

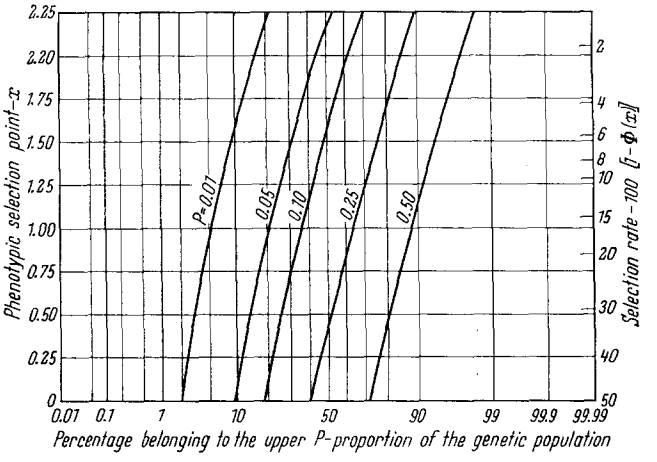


Fig. 7. Relationship between phenotypic truncation point and genetic value when $h^2 = 0.4$.

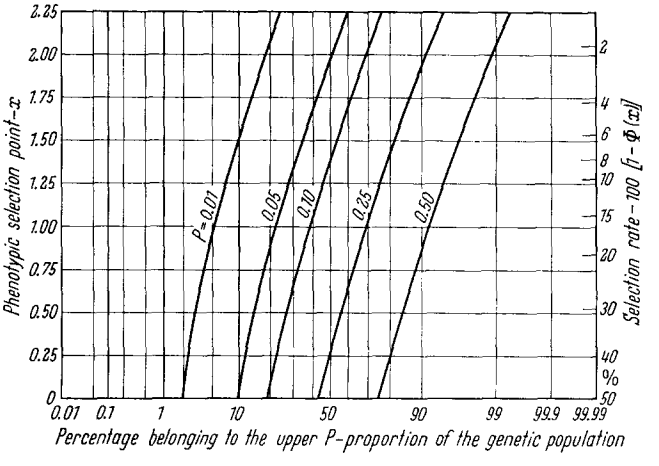


Fig. 8. Relationship between phenotypic truncation point and genetic value when $h^2 = 0.5$.

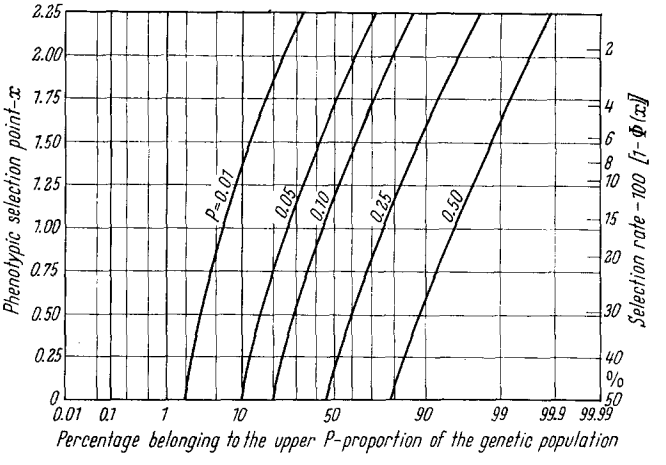


Fig. 9. Relationship between phenotypic truncation point and genetic value when $h^2 = 0.6$.

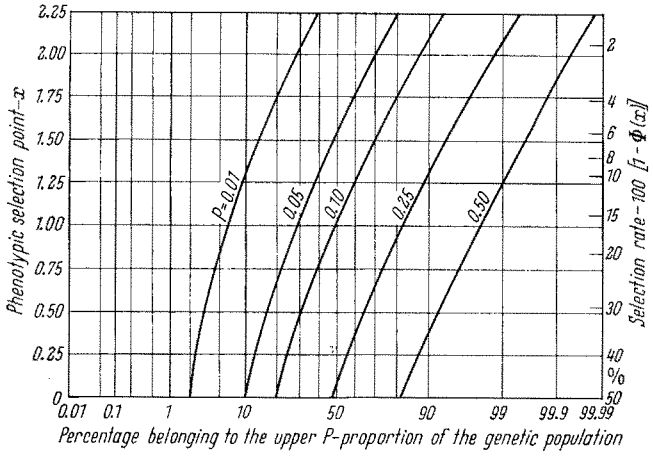


Fig. 10. Relationship between phenotypic truncation point and genetic value when $h^2 = 0.7$.

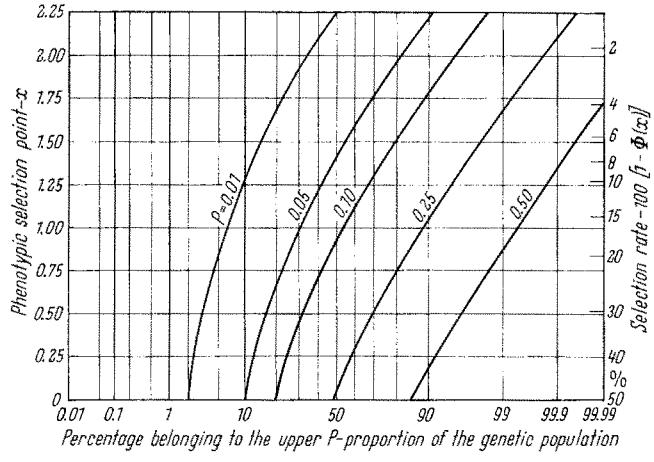


Fig. 11. Relationship between phenotypic truncation point and genetic value when $h^2 = 0.8$.

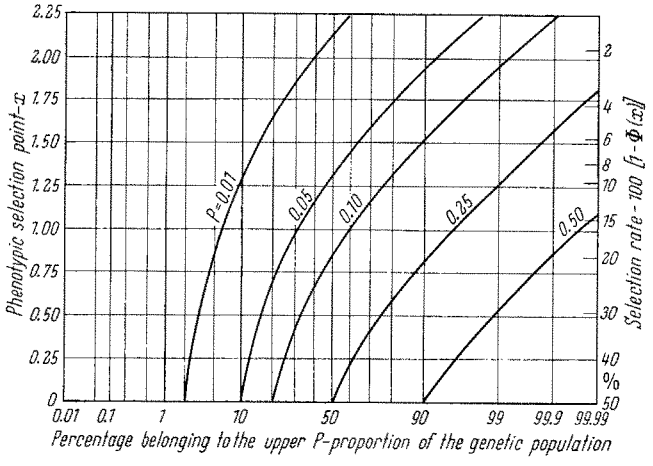


Fig. 12. Relationship between phenotypic truncation point and genetic value when $h^2 = 0.9$.

Table 1. *Percentage of individuals in a phenotypic class which belong to the top P percent of the genotypic distribution when* $h^2 = 1/2$.*

| Phenotypic class interval | Phenotypic frequency (%) | Percentage of the class frequency belonging to the top P percent of the genotypic distribution | | | | |
|---|---|---|---|---|---|---|
| | | P = 50 | P = 25 | P = 10 | P = 5 | P = 1 |
| .25 to .75 | 17.47 | 68.63 | 31.54 | 9.56 | 3.43 | 0.29 |
| .75 to 1.25 | 12.10 | 83.39 | 50.17 | 20.58 | 9.09 | 1.07 |
| 1.25 to 1.75 | 6.55 | 92.67 | 68.85 | 36.79 | 19.85 | 3.51 |
| 1.75 to 2.25 | 2.97 | 97.49 | 83.51 | 55.91 | 35.84 | 9.32 |
| 2.25 to ∞ | 1.22 | 99.18 | 94.26 | 77.05 | 59.84 | 25.41 |

* The development of this table from Figure 8 is discussed in Appendix 4.

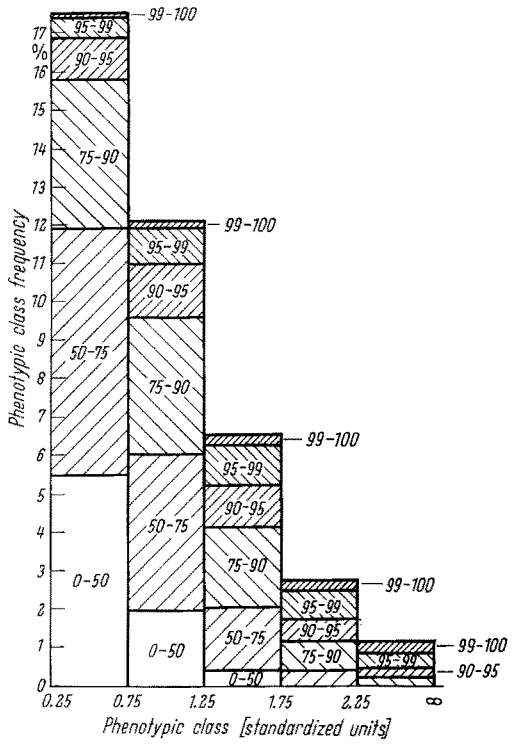


Fig. 13. Distribution of genotypic values within phenotypic classes when $h^2 = 1/2$.

A histogram derived from Table 1 is shown in Figure 13. The proportions of individuals belonging to the various percentiles of the genotypic distribution are indicated for each phenotypic class. For example, the phenotypic class

$$\bar{g} + 1.25\sqrt{\sigma_g^2 + \sigma_e^2} \leq X \leq \bar{g} + 1.75\sqrt{\sigma_g^2 + \sigma_e^2}$$

is represented by a column of height 6.55 percentage points, of which a fraction .9267 (third row of Table 1) was contributed by individuals from the top half of the genotypic distribution. The remaining fraction, .0733, is therefore contributed by individuals belonging to the lower half of the genotypic distribution. Thus, in Figure 13, the first .0733 (6.55) = 0.48 of these 6.55 percentage points are labeled "0 –50". Likewise, a fraction .9267—.6885 = .2382 of these 6.55 percentage points are contributed by individuals belonging to the top half but below the top quarter of the genotypic distribution. The next .2382(6.55) = 1.56 percentage points are therefore labeled "50 to 75" and represent the contribution to this phenotypic class from individuals belonging to the 50%—75% quartile of the genotypic distribution. This partitioning of the 6.55 percentage points is continued according to the columns of Table 1 until we last obtain the .0351(6.55) = 0.23 percentage points contributed by individuals belonging to the top 1 percent of the genotypic distribution (labeled "99 to 100").

## The Effect of Population Size: Before Selection

The preceding results describe some characteristics of an abstract infinite population. The genetic population actually observed and used for selection is of finite size N, representing only a sample from this potential infinite population. Finiteness of the observed population has no effect on the results plotted in Figures 1—3, which are conditional on a single

selected phenotypic value. But the application of Figures 4—12, conditional on a selected upper tail of the phenotypic distribution, requires further explanation.

A selection procedure which takes all individuals in the tail of a distribution may be defined in essentially two different ways, by specifying: (1) the minimum acceptable phenotypic value, or (2) the percentile of the observed phenotypic distribution at which selection begins. When the minimum acceptable phenotype is fixed in advance, the number or percentage of the prospective population of size N which will be selected is a chance variable, unknown in advance of the selection experiment. On the other hand, if the selection rate or percentage is fixed, the minimum phenotypic value which will be accepted is a chance variable and unknown in advance of the experiment. In either case, chance variations in the respective unknown quantities will decrease as N is increased; and, as N approaches infinity, the two procedures become equivalent. In other words, when population size is infinite, the specification that selection will take all individuals with phenotypic values exceeding $x'$ is equivalent to the specification that selection will take a fraction $1\text{-}\Phi(x)$ from the upper tail of the phenotypic distribution. This asymptotic equivalence has been expressed in Figures 4—12 by labeling the ordinate with both the scale of $x$ and $1\text{-}\Phi(x)$.

To plan a finite selection experiment of either type, the potential magnitude of chance fluctuations in the respective unknown quantities must be considered. For example, if a Type I experiment is contemplated with the minimum acceptable phenotype fixed at some preassigned level $x'$, rational planning requires that the population size N chosen be large enough to provide reasonable assurance that at least one of the N phenotypes will exceed $x'$. The preassigned value of $x'$ is in this case presumably based on considerations of the facts revealed in Figures 1—12. It is likewise chosen to provide reasonable assurance that an individual of this phenotype will be of a superior genotype.

Analysis of a Type-I-experiment model shows that the probability of finding at least one phenotype exceeding $x = (X - \bar{g})/\sqrt{\sigma_g^2 + \sigma_e^2}$ in a population of size $N$ is $1\text{-}[\Phi(x)]^N$, and the probability of finding one whose genotype also exceeds some specified value $g = (g' - \bar{g})/\sigma_g$ is $1-[1\text{-}P_{g\cdot x}(1 - \Phi(x))]^N$. Thus, in order to obtain $100(1 - \alpha)$ percent assurance that the Type I experiment will produce at least one selection, population size $N$ must be chosen to satisfy

$$1 - [\Phi(x)]^N = 1 - \alpha \quad \text{or} \quad N = \frac{\log \alpha}{\log \Phi(x)}.$$

In order to provide the same assurance of obtaining at least one selection with a genotypic value exceeding any specified value, g, a larger value of $N$ is required

$$N = \frac{\log \alpha}{\log [1 - P_{g\cdot x}(1 - \Phi(x))]}.$$

For example, if all phenotypes greater than two standard deviations above the mean are to be selected ($x = 2$, $\Phi(x) = .9773$), 90 percent assurance ($\alpha = 0.1$) of obtaining at least one selection requires that $N = \log 0.1/\log (1-.9773) = 101$. If in addition, this selection is to come from the top 5 percent of the genotypic distribution ($P = .05$, $g = 1.645$), $N$ must be chosen as follows:

| $h^2$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|---|---|---|---|---|---|---|---|---|---|---|
| $P_{g\cdot x}$ | 0.1744 | 0.2597 | 0.3427 | 0.4270 | 0.5146 | 0.6071 | 0.7062 | 0.8139 | 0.9293 | 1.0 |
| $N$ | 580 | 390 | 295 | 237 | 196 | 166 | 143 | 124 | 108 | 101 |

These computations demonstrate the influence of $h^2$ on $N$; it is much more dependent on small values of $h^2$ than on large values. The values of $P_{g\cdot x}$ in these computations were obtained from the previously mentioned numerical integration but may be read with 2-digit accuracy from Figures 4—12, respectively. As an illustration, referring to Figure 8 with $h^2 = 0.5$ we find that the curve for $P = .05$ intersects $x = 2$ at $P_{g\cdot x} \doteq .515$, giving

$$N = \frac{\log_{10} 0.1}{\log_{10} [1 - (.515)(.0227)]} = \frac{-1}{\log_{10} 0.9883} = \frac{1}{.00511} = 196.$$

A more detailed characterization of a contemplated Type I experiment is given by the probability of selecting exactly $m$ individuals belonging to the top $100P$ percent of the genotypic distribution, which is the binomial probability

$$\binom{N}{m} P_{g\cdot x}^m [1 - \Phi(x)]^m [1 - P_{g\cdot x}(1 - \Phi(x))]^{N-m}.$$

Thus, for the case illustrated above, with $N = 196$, the probability of obtaining exactly one selection from the genetic top 5 percent is

$$\binom{196}{1}(.515)(.0227)(.9883)^{195} = 2.291338(.1008) = .231.$$

Similarly, the chance of obtaining exactly two selections ($m = 2$) is .265, of three selections is .203, of four selections .116, and so on. The probabilities from $m = 1$ onward add to the previously specified $1 - \alpha = .90$. The expected number of such selections is $NP_{g\cdot x}(1 - \Phi(x))$, or in this case 2.29.

If a Type II experiment is contemplated with a fixed selection rate of $100(1 - \Phi)$ percent, the population size $N$ chosen should be large enough to provide reasonable assurance $(1 - \alpha)$ that the $k = N(1 - \Phi)$ selections are genetically superior, or at least that the best of these selected phenotypes is a genetically superior individual. A characterization of the Type II model, by means of the probability distribution of $m$ (the number of selections exceeding $g$ in genotypic value) is readily accomplished analytically, but is quite cumbersome computationally. A more convenient characterization is afforded by the genotypic distribution associated with either the minimum or maximum selected phenotype, the latter being equivalent to the distribution of $m$ for the special case $k = N(1 - \Phi) = 1$.

The genotypic distribution associated with the $k^{th}$ ranking phenotype in a population of size $N$ is shown in Appendix 5 to be

$$P_{g,k} = 1 - \int_{-\infty}^{\infty} \Phi\left(\frac{g - hx}{\sqrt{1 - h^2}}\right) \times d\left\{\sum_{r=0}^{k-1} \binom{N}{r} [1 - \Phi(x)]^r [\Phi(x)]^{N-r}\right\}, \quad (3)$$

where $P_{g,k}$ denotes the probability that the genotype of the phenotypically $k^{th}$ largest individual will exceed

the population mean by at least an amount $g\sigma_g$. For $k = 1$, or for the largest of $N$ phenotypes, this becomes

$$P_{g,1} = 1 - \int_{-\infty}^{\infty} \Phi\left(\frac{g - h\,x}{\sqrt{1 - h^2}}\right) d\Phi^N(x) .$$

Numerical evaluation of the integral $P_{g,k}$ for the purpose of appraising a contemplated Type II experimental plan is still somewhat tedious. However, numerical integration, employing devices discussed in Appendix 6, was carried out for the case $k = 1$ to obtain a solution to $P_{g,1} = 0.9$ as an equation in $N$. These results are presented in Table 2 as a guide to experiment planning. They indicate the population size required for 90 percent confidence that the best selection will belong to the top 5, 10, 25 or 50 percent of the genetic population.

Table 2. *Population size required for 90 percent certainty that the genotypic value of the phenotypically best individual will fall in a specified upper percentile of the genotypic distribution.*

| Heritability $h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$ | Top 50% | Top 25% | Top 10% | Top 5% |
|---|---|---|---|---|
| 0.1 | 12,137 | $7 \times 10^8$ | $2 \times 10^{14}$ | $10^{18}$ |
| 0.2 | 147 | 31,140 | $3 \times 10^7$ | $5 \times 10^9$ |
| 0.3 | 35 | 1,070 | 84,145 | $2 \times 10^6$ |
| 0.4 | 16 | 193 | 4,650 | $5 \times 10^4$ |
| 0.5 | 9 | 80 | 870 | 5,300 |
| 0.6 | 7 | 34 | 254 | 1,060 |
| 0.7 | 6 | 17 | 109 | 415 |
| 0.8 | 5 | 14 | 58 | 165 |
| 0.9 | 4 | 11 | 34 | 82 |
| 1.0 | 4 | 8 | 22 | 45 |

The population sizes for a Type II experiment, tabulated in the $P = .05$ column of Table 2, can be compared with those computed earlier for a Type I experiment. Earlier computations show that when $h^2 = 1/2$ and $x = 2$ in a Type I experiment, a population size of $N = 196$ is required for 90 percent assurance that *at least one selection* will belong to the top 5 percent of the genotypic distribution. Table 2, on the other hand, indicates that in a Type II experiment, a population size of $N = 5300$ is required for 90 percent assurance that the *phenotypically best selection* will belong to the top 5 percent of the genotypic distribution.

Superficially, these two tabulated results might appear incompatible; however, the conditions to be fulfilled are quite different. In the first, we require only that at least one of the selections be genetically superior, while in the second we require that an identifiable selection (the largest) be genetically superior. Clearly, the latter requirement is much more stringent; the population size necessary to achieve it is correspondingly much greater. For a population size of only 196 in a Type I experiment with $h^2 = 1/2$ and $x = 2$, there is a 0.9 probability that at least one of the selections will have $g > 1.645$; but there is (by numerical integration) less than 0.65 probability that the phenotypically best selection will have $g > 1.645$.

### The Effect of Population Size: After Selection

At the planning stage of a selection experiment the phenotypes to be selected are yet unknown, except in the form of a lower bound in the case of a Type I experiment. Upon completion of the phenotypic selection the genetic values of the chosen individuals are yet unknown. Figures 1—12 are therefore still of some interest to the geneticist at the post-selection stage. Figures 1—3, in particular, provide a description of the genotypic distribution associated with any selected phenotypic value, regardless of the rank of that individual among the selections. If phenotypic values are not actually measured in the selection experiment, but only compared to a standard value $x$ in a Type I experiment (i.e., phenotypes are observed only to be greater or less than $x$), or if they are simply ranked in a Type II experiment, Figures 1—3 do not apply. The use of Figures 4—12 then depends more specifically on the type of information available on the selected phenotypes.

In a Type I experiment where no information is obtained other than the exact number $k$ of phenotypes exceeding the standard value $x$, the computations of Figures 4—12 apply directly for a random one of these $k$ individuals. For all $k$ individuals, a binomial distribution applies. The probability that $m$ of the selected $k$ individuals fall in the top $100P$ percent of the genotypic distribution is

$$\binom{k}{m} P_{g\cdot x}^m (1 - P_{g\cdot x})^{k-m} .$$

For example, if $k = 10$ individuals are found to be phenotypically larger than a preassigned standard value of $x = 1.645$ in a population with heritability $h^2 = 0.4$, the probability that at least 3 of these 10 belong to the top $100P = 5$ percent of the genotypic distribution is

$$1 - \binom{10}{0} P_{g\cdot x}^0 (1 - P_{g\cdot x})^{10} - \binom{10}{1} P_{g\cdot x}^1 (1 - P_{g\cdot x})^9 - \binom{10}{2} P_{g\cdot x}^2 (1 - P_{g\cdot x})^8 .$$

From Figure 7 and the curve $P = .05$ at $x = 1.645$ we find $P_{g\cdot x} = .333$, giving

$$1 - .667^{10} - 10\,(.333)\,(.667)^9 - 45(.333)^2(.667)^8 = .7000 .$$

Notice that knowledge of $k$ will, on the average, decrease the variance of $m$ by a factor of $(1 - P_{g\cdot x})/[1 - P_{g\cdot x} + P_{g\cdot x}\Phi(x)]$, or approximately by $1 - P_{g\cdot x}$.

Additional information concerning the ordering among the $k$ selected phenotypes in a Type I experiment is difficult to analyze numerically, because of integration problems. The genotypic distribution associated with the lowest-ranking selected phenotype, for example, is given by the integral

$$\frac{k}{[1 - \Phi(x)]^k} \int_x^{\infty} \Phi\left(\frac{g - y\,h}{\sqrt{1 - h^2}}\right) [1 - \Phi(y)]^{k-1}\, d\Phi(y)$$

while that for the highest ranking phenotype is

$$\frac{k}{[1 - \Phi(x)]^k} \int_x^{\infty} \Phi\left(\frac{g - y\,h}{\sqrt{1 - h^2}}\right) [\Phi(y) - \Phi(x)]^{k-1}\, d\Phi(y) .$$

Such functions could be integrated numerically by the methods employed in computing Table 1. However, the number of cases to be considered is quite large and requires extensive tables or graphs.

In a Type II experiment, phenotypic ordering, an integral part of the selection process, contributes no additional information beyond that assumed in the planning stage. In this case the number ($k$) selected is fixed in advance; and the minimum selected phenotype $x_k$ is a chance variable corresponding to the selection point $x$ of a Type I experiment. In fact, if $x_k$ were actually measured while the remaining top $k-1$ phenotypes were merely ranked, the information on these $k-1$ individuals would be of a type identical to that obtained on the $k$ selections of a Type I experiment, with $x_k$ now playing the role of $x$.

Usually all selected phenotypes will be measured, and the outcome of the selection experiment will consist of phenotypic observations $x_1 > x_2 > \cdots > x_k$. Figures 1–3 then suffice to describe the probability distribution of genotypic values associated with each of these phenotypes, regardless of which type of experiment was employed.

## Expected Identifiable Numbers of Genetic Deviates

POWERS (1945), POWERS et al. (1958) and DUDLEY and POWERS (1960) introduced the concept of identifiable numbers of genetic deviates in the phenotypic classes of a segregating population. The identifiable numbers of genetic deviates are represented by the differences between these class frequencies and those of a nonsegregating population of the same size, superimposed on the same population mean. As noted by FEDERER, POWERS, and PAYNE (1963), the class frequencies of the segregating normal population exceed those of the corresponding nonsegregating population at a distance of

$$z' = z\sqrt{\sigma_g^2 + \sigma_e^2} = \sigma_e\sqrt{\frac{\log_e(1-h^2)}{-h^2}}$$

or more on either side of the mean. Consequently, in the sense defined by POWERS, the frequency of identifiable genetic deviates is positive in any phenotypic class, beginning at least a distance $z$ away from the mean of an infinite population. In particular, in the entire tail of the distribution, from $z$ to $\infty$, the frequency of identifiable genetic deviates is

$$P^+ = \left[1 - \Phi\left(\frac{z'}{\sqrt{\sigma_g^2 + \sigma_e^2}}\right)\right] - \left[1 - \Phi\left(\frac{z'}{\sigma_e}\right)\right]$$
$$= \Phi\left(\frac{z}{\sqrt{1-h^2}}\right) - \Phi(z) \qquad (4)$$

so that in a segregating population of size $N$ the *expected number of superior identifiable genetic deviates* is defined as $NP^+$.

Unfortunately, in the present context this definition raises certain ambiguities, for in this same phenotypic interval $(z, \infty)$ the proportion of genotypic values exceeding the mean is

$$P_{0\cdot z} = \frac{1}{1-\Phi(z)}\int_z^\infty \Phi\left(\frac{hy}{\sqrt{1-h^2}}\right)d\Phi(y)$$

so that $NP_{0\cdot z}(1-\Phi(z))$ may also be described as the expected number of "superior" genotypes falling in this interval. Confusion may be avoided here by regarding $P^+$ as an index of heritability, comparable to the heritability ratio $h^2$, rather than by attempting to interpret $P^+$ as a probability. In fact, for the normal case, $P^+$ is a monotone (increasing) function of $h^2$ and therefore equivalent to $h^2$ as a heritability index. When the genotypic distribution is not normal then, of course, the two indices $P^+$ and $h^2$ are no longer equivalent and, in general, neither can be regarded as an adequate index of heritability in the sense of uniquely determining the genotypic distribution associated with a given phenotypic distribution. However, if a frequency difference $P^+$ is computed for every phenotypic class interval, all information in the genetic experiment is retained; in this extreme, if trivial form, the $P^+$ index has optimum properties for any distribution model.

## Numerical Illustrations with Sugarbeet Data

Experimental data fulfilling all requirements of the normal model are extremely uncommon, and, in the strictest sense, actually non-existent. Besides the problem of achieving normality and constant environmental variance, by appropriate choices of a scale of measurement, there is also the problem of conducting a practical, completely randomized experiment with a single individual per plot as called for by the simple model considered here. Our data (used here for illustration, and obtained from a population genetic study on sugarbeets conducted in 1960) cannot be rigorously shown to satisfy any of the requirements of the model. Thus the application of the model provides only a crude guide to the true genetic character of the data. Until more general methods of genetic analysis are developed a crude guide is all that can be expected, and may still serve as a valuable tool in the analyses of such data.

For details of the design of the experiment, a description of the populations studied, and the adjustment of the frequency distributions to eliminate variation due to replications and populations, see POWERS et al. (1964). Except for predicted values, the data in Table 3 were taken from this article. They determined sucrose percentage of the sugarbeet roots, transformed the data to the logarithmic scale, and made the adjustments mentioned above. The resulting frequency distributions are thus freed of both replication and population mean effects, and hence are subject only to within-plot sources of genetic and environmental variation. Intra-plot correlations (positive before adjustment) are negative in the adjusted observations. Their effects are not taken into account in the following analysis. The frequency distributions (see POWERS et al. (1958) and POWERS et al. (1964)), present a skew appearance which indicates that the environmental distribution depicted by the non-segregating entry may change shape with each genotype.

First, the correspondence between the observed frequency difference beyond the points of intersection of the segregating and the non-segregating frequency distributions are considered. The method of identifying these points of intersection is given by POWERS et al. (1958). The predicted frequency differences, based on the observed within-plot heritability ratio, and the normal theory formula for $NP^+$, are shown in Table 3 for each of the segregating populations. For example, with a heritability ratio of $h^2 = .60569$ and $N = 450$, the predicted frequency difference to the right of

$$z = \sqrt{\frac{.39431\log_e(.39431)}{-.60569}} = .77836$$

is

$$NP^{+} = 450\left[\Phi\left(\frac{.77836}{\sqrt{.39431}}\right) - \Phi\,(.77836)\right]$$
$$= 450\,[.8924 - .7808] = 50\,.$$

Examination of Table 3 reveals that this prediction corresponds fairly closely with observation for most entries, though a few major discrepancies reduce the correlation between observed and predicted to 0.69. A somewhat surprising side result is the high correlation of .99976 between $h^2$ and $NP^{+}$ in this table, indicating that despite its analytically complicated form, $NP^{+}$ is a nearly linear function of $h^2$. In the light of these clear, as well as suspected violations of the model, the degree of fit to the model predictions are somewhat surprising. They lend some credence to imponderable assumptions.

Table 3. *Comparison of observed and predicted identifiable numbers of genetic deviates. Adjusted 1960 log percentage sucrose data, N = 450.*[1] *Predictions based on equation 4.*

| Population | Heritability ratio $h^2$ | Identifiable numbers of genetic deviates | | | |
|---|---|---|---|---|---|
| | | Superior | | Total | |
| | | observed | predicted | observed | predicted |
| CMS X 4W-34 | .60569 | 45 | 50 | 87 | 100 |
| CMS X A54-1 | .62211 | 52 | 52 | 97 | 104 |
| A54-1 | .63228 | 60 | 53 | 108 | 106 |
| CMS X 4W-34 $S_2$ | .48490 | 42 | 35 | 75 | 70 |
| 4W-34 $S_2$ | .59884 | 45 | 49 | 77 | 97 |
| CMS X 4W-34 AR | .52778 | 44 | 40 | 82 | 81 |
| 4W-34 AR | .41712 | 25 | 29 | 52 | 58 |
| 52-430 X 54-520 | .41019 | 32 | 29 | 64 | 57 |
| 54-520 X 52-305 | .35738 | 34 | 24 | 65 | 48 |
| A56-3 | .51308 | 56 | 39 | 94 | 78 |
| 54-520 | .67032 | 44 | 59 | 83 | 118 |
| Totals | | 479 | 459 | 884 | 917 |
| | | $r = .69$ | | $r = .75$ | |

With exception of the predicted values the data are taken from POWERS, REMMENGA, and URQUHART (1964).

The predictions embodied in Table 3 were based upon the heuristically derived Equation (4) for $NP^{+}$. However, this prediction contains no information about the genetic properties of the selected individuals. Yet individuals from the top 10 percent of the phenotypic distribution should have desirable genetic properties. Table 4 was constructed to illustrate this point using again data from POWERS et al. (1964). The number of observations above $x'$ is the observed tail number. When the truncation point is divided by an estimate of $\sqrt{\sigma_e^2 + \sigma_g^2}$, the standardized lower limit results. Since $\sigma_g^2$ is different for each population, the standardized limit varies from population to population.

The a posteriori probability of a superior genetic deviate in the tail is computed by interpolation from Figures 4—12. For an example, consider the first population in Table 4: in the region $X > \bar{g} + 1.2667\sqrt{\sigma_g^2 + \sigma_e^2}$ of a normal segregating population with $h^2 = .60569$ the proportion of genotypes falling in the first quartile of the (normal) genotypic distribution is given by linear interpolation between the points $P = .25$, $x = 1.2667$ in Figures 9 and 10. For $h^2 = .6$ in Figure 9 at $x = 1.2667$ and $P = .25$ we find $P_{g\cdot x} = .825$ and for $h^2 = .7$ in Figure 10, $P_{g\cdot x} = .887$. Interpolation to $h = .60569$ then gives

$$P_{g\cdot x} = .0569(.887 - .825) + .825 = .828$$

and since 32 individuals instead of the predicted 46 belonged to this tail of the observed (standardized) phenotypic distribution, the predicted number of genotypes is

$$32P_{g\cdot x} = 32(.828) = 26\,.$$

A direct measure of departure from normality is present in Table 4, where the observed frequencies in the upper tails of the distributions ($X > x'$) are compared to predicted frequencies $\{N(1 - \Phi(x))\}$. Here the skewness of the phenotypic distributions is apparent; the predicted upper tail frequencies almost always exceed the observed. Consequently the predicted numbers should be scaled downward in some manner as, for example, by multiplying them by the ratio of the observed to predicted numbers, i.e.

$$26\left(\frac{32}{46}\right) = 18\,.$$

Table 5 contains a similar illustration based on selections from the upper quarter of the phenotypic distribution and the upper half of the genotypic distribution.

## Truncation Selection for Two Traits

Graphs of the integral $P_{g\cdot x}$, shown in Figures 4—12, also have application in selecting on two positively correlated traits $X_1$ and $X_2$ for improvement in a third trait $Y$. If truncation selection is practiced by retaining all individuals whose phenotype with respect to $X_1$ and $X_2$ satisfies the relation

$$\frac{X_1 - \mu_{x_1}}{\sigma_{x_1}} > x_1 \qquad \frac{X_2 - \mu_{x_2}}{\sigma_{x_2}} > x_2$$

Table 4. *Normal theory predicted proportions of individuals from the top quarter of the genotypic distribution. Data from* POWERS *et al. (1964).*

| Population | Heritability ratio $h^2$ | Standardized lower limit $x$ | Normal tail | | Observed tail number | Predicted genotypes from the top quarter of the genotypic distribution | | |
|---|---|---|---|---|---|---|---|---|
| | | | area $p = 1-\Phi(x)$ | number $\pm 2\sigma$ $Np \pm 2\sqrt{Np(1-p)}$ | | proportion | number | adjusted |
| CMS X 4W-34 | 0.6057 | 1.2667 | .1026 | 46 ± 13 | 32 | .828 | 26 | 18 |
| CMS X A54-1 | 0.6221 | 1.2402 | .1075 | 48 ± 13 | 30 | .822 | 25 | 16 |
| A54-1 | 0.6323 | 1.2232 | .1106 | 50 ± 13 | 32 | .834 | 27 | 17 |
| CMS X 4W-34 $S_2$ | 0.4849 | 1.4478 | .0738 | 33 ± 11 | 23 | .793 | 18 | 13 |
| 4W-34 $S_2$ | 0.5988 | 1.2776 | .1007 | 45 ± 13 | 34 | .827 | 28 | 21 |
| CMS X 4W-34 AS | 0.5278 | 1.3861 | .0829 | 37 ± 12 | 31 | .807 | 25 | 21 |
| 4W-34 AS | 0.4171 | 1.5401 | .0618 | 28 ± 10 | 20 | .766 | 15 | 11 |
| 52-430 X 54-520 | 0.4102 | 1.5493 | .0607 | 27 ± 10 | 19 | .763 | 14 | 10 |
| 52-430 X 52-307 | 0.0000 | 2.0172 | .0218 | 10 ± 6 | 7 | | | |
| 54-520 X 52-305 | 0.3574 | 1.6171 | .0529 | 24 ± 9 | 19 | .735 | 14 | 11 |
| A56-3 | 0.5131 | 1.4077 | .0796 | 36 ± 11 | 24 | .803 | 19 | 13 |
| 54-520 | 0.6703 | 1.1583 | .1234 | 56 ± 14 | 35 | .841 | 29 | 18 |

Table 5. *Normal theory predicted proportions of individuals from the top half of the genotypic distribution. Data from* POWERS *et al. (1964).*

| Population | Heritability ratio $h^2$ | Standardized lower limit $x$ | Normal tail area $p = 1-\Phi(x)$ | Normal tail number $\pm 2\sigma$ $Np \pm 2\sqrt{Np(1-p)}$ | Observed tail number | Predicted genotype from the top half of the genotypic distribution proportion | number | adjusted |
|---|---|---|---|---|---|---|---|---|
| CMS X 4W-34 | 0.6057 | .6000 | .2742 | 123 ± 19 | 112 | .902 | 101 | 92 |
| CMS X A54-1 | 0.6221 | .5875 | .2785 | 125 ± 19 | 119 | .908 | 108 | 103 |
| A54-1 | 0.6323 | .5794 | .2812 | 127 ± 19 | 127 | .910 | 116 | 116 |
| CMS X 4W-34 $S_2$ | 0.4849 | .6858 | .2464 | 111 ± 18 | 109 | .870 | 95 | 93 |
| 4W-34 $S_2$ | 0.5988 | .6052 | .2725 | 123 ± 19 | 112 | .902 | 101 | 92 |
| CMS X 4W-34 AS | 0.5278 | .6566 | .2557 | 115 ± 18 | 111 | .883 | 98 | 95 |
| 4W-34 AS | 0.4171 | .7295 | .2328 | 105 ± 18 | 92 | .848 | 78 | 68 |
| 52-430 X 54-520 | 0.4102 | .7338 | .2317 | 104 ± 18 | 99 | .846 | 84 | 80 |
| 52-430 X 52-307 | 0.0000 | .9555 | .1697 | 76 ± 16 | 67 | | | |
| 54-520 X 52-305 | 0.3574 | .7660 | .2218 | 99 ± 18 | 101 | .826 | 83 | 85 |
| A56-3 | 0.5131 | .6668 | .2524 | 114 ± 18 | 123 | .879 | 108 | 117 |
| 54-520 | 0.6703 | .5487 | .2917 | 131 ± 19 | 111 | .920 | 102 | 86 |

then the mean value of $Y$ in the selected fraction of the population is shown in Appendix 7 to be

$$\mu_y + \frac{\sigma_y}{[1-\Phi(x_1)]\,P_{x_2\cdot x_1}}\left\{\varrho_{yx_1}\varphi(x_1)\,\Phi\left(\frac{x_1\varrho_{x_1x_2}-x_2}{\sqrt{1-\varrho^2_{x_1x_2}}}\right) + \varrho_{yx_2}\varphi(x_2)\,\Phi\left(\frac{x_2\varrho_{x_1x_2}-x_1}{\sqrt{1-\varrho^2_{x_1x_2}}}\right)\right\}. \quad (5)$$

The symbol $\varrho$ in this expression denotes a simple correlation coefficient; $\varphi$ denotes the standard univariate normal density function and

$$[1-\Phi(x_1)]\,P_{x_2\cdot x_1} = \int_{x_1}^{\infty}\varphi(z)\,\Phi\left(\frac{\varrho_{x_1x_2}z-x_2}{\sqrt{1-\varrho^2_{x_1x_2}}}\right)dz\,.$$

These formulae are based upon the assumption that $X_1$, $X_2$ and $Y$ follow a trivariate normal distribution.

If equal selection rates are applied to both traits so that $x_1 = x_2 = x$ (say), then the expression for mean gain in $Y$ reduces to

$$\text{Gain} = \frac{\sigma_y}{[1-\Phi(x)]\,P_{x\cdot x}} \times \varphi(x)\left[1-\Phi\left(x\sqrt{\frac{1-\varrho_{x_1x_2}}{1+\varrho_{x_1x_2}}}\right)\right](\varrho_{yx_1}+\varrho_{yx_2})\,.$$

Maximum gain is obtained when the two selected traits are independent of one another, $\varrho_{x_1x_2} = 0$,

$$\text{Maximum gain} = \frac{\varphi(x)}{1-\Phi(x)}(\varrho_{yx_1}+\varrho_{yx_2})\,\sigma_y\,,$$

while a perfect positive correlation, $\varrho_{x_1x_2} = 1$, would result in just half this maximum gain; and a correlation of $\varrho_{x_1x_2} = -1$ would result in no gain whatsoever in the mean value of $Y$. A selection rate of 25 percent on each trait ($x = .674$), for example, could produce a maximum gain of

$$\frac{\varphi(.674)}{1-\Phi(.674)}(\varrho_{yx_1}+\varrho_{yx_2})\,\sigma_y = 1.272\,(\varrho_{yx_1}+\varrho_{yx_2})\,\sigma_y$$

when the selected traits are independent. If $X_1$ and $X_2$ were highly (positively) correlated, say $\varrho^2_{x_1x_2} = .9$, the corresponding gain would be reduced to

$$\frac{1-\Phi\left(.674\sqrt{\frac{1-.9487}{1+.9487}}\right)}{P_{.674\cdot.674}}\,1.272\,(\varrho_{yx_1}+\varrho_{yx_2})\,\sigma_y = \frac{.4564}{.83}(1.272)\,(\varrho_{yx_1}+\varrho_{yx_2})\,\sigma_y$$

or 55 percent of the maximum possible gain. Figure 12 was employed in this calculation to obtain

$$P_{.674\cdot.674} = .83$$

as the abscissa of the curve $P = 0.25$ at a selection rate of 25 percent.

In many practical situations where this model might apply, the two phenotypic traits being selected are negatively correlated. The calculation of numerical predictions in the form outlined here would then require extension of our Figures 4—12 to include negative correlations. A case in point arises in sugar-beet selection when

$$X_1 = G_1 + E_1 = \text{log weight per root}$$
$$X_2 = G_2 + E_2 = \text{log percent sucrose}$$

and

$$Y = G_1 + G_2\,.$$

Selections on weight per root are commonly made in the field and brought into the laboratory for selection on percent sucrose, the aim being to improve the genetic value with respect to sucrose yield.

## Appendix 1: Conditional Distribution of G given X

If $X = G + E$ where $G$ and $E$ are independent normal chance variables with means $\bar{g}$ and 0 and variances $\sigma^2_g$ and $\sigma^2_e$, respectively, then the conditional distribution of $G$ for fixed $X$ is normal with

$$\text{ave}\,(G\,|x) = \text{ave}\,(G) + \frac{\text{cov}(X,G)}{\text{var}(X)}[x' - \text{ave}\,(x)] = \bar{g} + \frac{\sigma^2_g}{\sigma^2_g+\sigma^2_e}(x'-\bar{g})$$

$$\text{var}\,(G\,|x) = \text{var}\,(G)\left[1-\frac{[\text{cov}\,(X,G)]^2}{\text{var}(X)\,\text{var}(G)}\right] = \sigma^2_g\left(1-\frac{\sigma^2_g}{\sigma^2_g+\sigma^2_e}\right)$$

## Appendix 2: Derivation of Equation (1)

For

$$h^2 = \frac{\sigma^2_e}{\sigma^2_e+\sigma^2_g}, \qquad x = \frac{x'-\bar{g}}{\sqrt{\sigma^2_g+\sigma^2_e}}, \qquad g = \frac{g'-\bar{g}}{\sigma_g},$$

the conditional probability that $G$ will exceed $g'$ is

$$P\,(G>g'\,|x) = P\left[\frac{G-\text{ave}\,(G\,|x)}{\sqrt{\text{var}\,(G\,|x)}} > \frac{g'-\text{ave}\,(G\,|x)}{\sqrt{\text{var}\,(G\,|x)}}\right] = 1-\Phi\left[\frac{g-hx}{\sqrt{1-h^2}}\right] = \Phi\left[\frac{hx-g}{\sqrt{1-h^2}}\right]$$

where $\Phi$ is the standard (cumulative) normal distribution function. If $g = 0$ and $h^2 = 1/2$, this becomes merely $\Phi(x)$.

## Appendix 3: Derivation of Equation (2)

$$P_{g\cdot x} = P\,(G > g' \,|X > x') = \frac{P(G > g',\, X > x')}{P\,(X > x')}$$

$$= \frac{1}{1 - F_X(x')} \int_{x'}^{\infty} \int_{g'}^{\infty} f_{G|X}\,(u\,|v)\, f_X(v)\, \mathrm{du}\, \mathrm{dv}$$

$$= \frac{1}{1 - F_X(x')} \int_{x'}^{\infty} [1 - F_{G|X}\,(g'\,|v)]\, dF_X\,(v)$$

where

$$F_X(v) = \Phi\left(\frac{v - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}}\right)$$

$$1 - F_{G|X}\,(g'\,|v) = 1 - \Phi\left(\frac{g' - \bar{g} - h^2\,(v - \bar{g})}{\sigma_g \sqrt{1 - h^2}}\right)$$

$$= \Phi\left(\frac{h^2\,(v - \bar{g}) - (g' - \bar{g})}{\sigma_g \sqrt{1 - h^2}}\right).$$

With the change of variable

$$y = \frac{v - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}}$$

then, in terms of the standardized variables $x$ and $g$ defined in Appendix 2,

$$P_{g\cdot x} = \frac{1}{1 - \Phi(x)} \int_{x}^{\infty} \Phi\left(\frac{h\,y - g}{\sqrt{1 - h^2}}\right) d\Phi(y)\,.$$

This integral is undefined when $h^2 = 1$. However, this means that $X = G$ with probability one, i.e. the genotype is observable. Thus,

$$P_{g\cdot x} = P\,(G > g' \mid X > x') = \begin{cases} 1 & g' \leq x' \\ \dfrac{1 - \Phi(g)}{1 - \Phi(x)} & g' > x' \end{cases}$$

For the special case $g = 0$ and $h^2 = 1/2$, the integral reduces to

$$P_{0\cdot x} = \frac{1}{1 - \Phi(x)} \int_{x}^{\infty} \Phi(y)\, d\Phi(y)$$

$$= \frac{1}{1 - \Phi(x)} \left[\frac{1}{2}\, \Phi^2(y)\right]_x^{\infty}$$

$$= \frac{1 - \Phi^2(x)}{2\,(1 - \Phi(x))} = \frac{1}{2}\,(1 + \Phi(x))\,.$$

## Appendix 4: Development of Table 1

For a set of phenotypic intervals $(x_1, x_2), \ldots, (x_m, +\infty)$ on the standard scale

$$x_i = \frac{X_i - g}{\sqrt{\sigma_g^2 + \sigma_e^2}}$$

the proportion $P_g(x_i, x_{i+1})$ of genotypic values exceeding the population mean by an amount $g\sigma_g$ can be computed from the formula

$$P_g\,(x_i, x_{i+1}) = \frac{P_{g\cdot x_i}\,[1 - \Phi(x_i)] - P_{g\cdot x_{i+1}}\,[1 - \Phi(x_{i+1})]}{[1 - \Phi(x_i)] - [1 - \Phi(x_{i+1})]}$$

Referring to Figure 8 for the case $h^2 = 0.5$, we illustrate this procedure with intervals of length $x_{i+1} - x_i = 0.5$ between $x_1 = .25$ and $x_m = x_5 = 2.25$. Taking the interval (1.25, 1.75), for example, we find (either by visual interpolation on the right-hand scale of Figure 8 or directly from tables of the standard cumulative normal distribution):

$$1 - \Phi\,(1.25) = .1056 \qquad 1 - \Phi\,(1.75) = .0401$$
$$[1 - \Phi\,(1.25)] - [1 - \Phi\,(1.75)] = .0655\,.$$

Then, to find the expected proportion of above-average genotypes in this interval we refer to the $P = 0.50$ (or $g = 0$) curve in Figure 8; at $x_i = 1.25$ this gives

$$P_{g\cdot x_i} = P_{0\cdot 1.25} = .947$$

and at $x_{i+1} = 1.75$

$$P_{g\cdot x_{i+1}} = P_{0\cdot 1.75} = .980$$

so that

$$P_g\,(x_i, x_{i+1}) = P_0\,(1.25, 1.75)$$
$$= \frac{.947(.1056) - .980(.0401)}{.0655}$$
$$= \frac{.0607}{.0655}$$
$$= .9267$$

Thus, when $h^2 = 1/2$, 6.55 percent of the population falls in the phenotypic interval

$$1.25 < \frac{X - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}} < 1.75$$

and 92.67 percent of the individuals falling in this interval are genotypically above average.

Similarly, from the curve for $P = 0.25$ ($g = .6745$) in Figure 8 we find

$$P_{g\cdot x_i} = P_{.6745\cdot 1.25} = .756 \quad P_{g\cdot x_{i+1}} = P_{.6745\cdot 1.75} = .867$$

so that

$$P_g\,(x_i, x_{i+1}) = P_{.6745}\,(1.25, 1.75)$$
$$= \frac{.756(.1056) - .867(.0401)}{.0655}$$
$$= \frac{.0451}{.0655}$$
$$= .6885\,.$$

Thus, 68.85 percent of the individuals in this interval belong to the top quartile of the genotypic distribution. The remaining computations for this and other intervals, performed in the same manner, are shown in Table 1.

Tables similar to Table 1, but for other values of $h^2$ require the use of figures other than Figure 8. For example, Figure 5 would be used in place of Figure 8 to get a table for $h^2 = 0.2$, but it would be used exactly as Figure 8 was used above.

## Appendix 5: Derivation of Equation (3)

$$P_{g,k} = P(G_k > g') = \int_{-\infty}^{\infty} \int_{g'}^{\infty} f_{G_k|X_k}\,(u\,|v)\, f_{X_k}\,(v)\, \mathrm{du}\, \mathrm{dv}$$

$$= \int_{-\infty}^{\infty} [1 - F_{G_k|X_k}\,(g'\,|v)]\, dF_{X_k}\,(v)$$

$$= 1 - \int_{-\infty}^{\infty} F_{G_k|X_k}\,(g'\,|v)\, (dF_{X_k}\,(v)$$

where

$$F_{X_k}\,(v)$$

$$= \sum_{r=0}^{k-1} \binom{N}{r} \left[1 - \Phi\left(\frac{v - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}}\right)\right]^r \left[\Phi\left(\frac{v - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}}\right)\right]^{N-r}$$

and

$$F_{G_k|X_k}\,(g'\,|x) = \Phi\left(\frac{g' - \bar{g} - h^2\,(v - \bar{g})}{\sigma_g \sqrt{1 - h^2}}\right).$$

With the changes of variable

$$x = \frac{v - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}} \quad \text{and} \quad g = \frac{g' - \bar{g}}{\sigma_g \sqrt{1 - h^2}},$$

$$P_{g,k} = 1 - \int_{-\infty}^{\infty} \Phi\left(\frac{g - h\,x}{\sqrt{1 - h^2}}\right) \times d\left\{\sum_{r=0}^{k-1} \binom{N}{r} [1 - \Phi(x)]^r [\Phi(x)]^{N-r}\right\}.$$

## Appendix 6: Numerical Integration Leading to Table 2

Arbitrarily close bounds on the integral in Equation (3) may be computed as

$$\sum_{i=0}^{m} \Phi\left(\frac{x_i - g\,h}{\sqrt{1 - h^2}}\right)[F_k(x_{i+1}) - F_k(x_i)] < P_{g,k} < \sum_{i=0}^{m} \Phi\left(\frac{x_{i+1} - g\,h}{\sqrt{1 - h^2}}\right)[F_k(x_{i+1}) - F_k(x_i)]$$

where $-\infty = x_0 < x_1 < \ldots < x_m < x_{m+1} = +\infty$ is any judiciously chosen monotone sequence and

$$F_k(x) = \sum_{r=0}^{k-1} \binom{N}{r} [1 - \Phi(x)]^r [\Phi(x)]^{N-r}.$$

Furthermore, when $k$ is small and $N$ is large, $F_k(x)$ remains extremely small until $\Phi(x)$ gets near unity, and then $F_k(x)$ is closely approximated by

$$F_k(x) \doteq e^{-N[1-\Phi(x)]} \sum_{r=0}^{k-1} \frac{[N(1 - \Phi(x))]^r}{r!}.$$

In particular, when $k = 1$,

$$F_1(x) = \Phi^N(x) \doteq e^{-N[1-\Phi(x)]}$$

remains less than .005 until $\Phi(x)$ attains the value

$$\Phi(x_1) = 1 - \frac{5.29832}{N}$$

and increases to .995 at

$$\Phi(x_m) = 1 - \frac{.00501}{N}$$

thus indicating a judicious range for the sequence $x_1 < \cdots < x_m$.

Certain special cases do exist where the integral $P_{g,k}$ can be evaluated explicitly, the most interesting being the case $h^2 = 1/2$ and $g = 0$. When half of the phenotypic variability in the (infinite) population is of genetic origin, the probability that the $k^{th}$ ranking phenotype in a population of size $N$ will be genetically above average is

$$P_{0,k} = 1 - k\binom{N}{k} \int_{-\infty}^{\infty} [\Phi(x)]^k [1 - \Phi(x)]^{N-k}\, d\Phi(x).$$

With the change of variable $u = \Phi(x)$, then

$$P_{0,k} = 1 - k\binom{N}{k} \int_{-\infty}^{\infty} u^k (1 - u)^{N-k}\, du$$

$$= 1 - \frac{k\,N!}{k!\,(N-k)!} \cdot \frac{k!\,(N-k)!}{(N+1)!} = 1 - \frac{k}{N+1}.$$

From this result it follows that the expected number of above average genotypes among the best $k$ out of $N$ phenotypes is

$$\sum_{j=1}^{k} P_{0,j} = k\left[1 - \frac{k+1}{2\,(N+1)}\right].$$

A second, but rather trivial, special case arises when $h^2 = 1$, giving

$$P_{g,k} = 1 - F_k(g)$$

which is readily evaluated from tables of the standard normal distribution. While this limiting case is of no particular genetic interest, it does provide a convenient bound or useful check against numerical integration at other values of $h$.

## Appendix 7: Justification of Equation 5

$$\begin{bmatrix} W \\ Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} \frac{Y - \mu_y}{\sigma_y} \\ \frac{X_1 - \mu_{x_1}}{\sigma_{x_1}} \\ \frac{X_2 - \mu_{x_2}}{\sigma_{x_2}} \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \varrho_{yx_1} & \varrho_{yx_2} \\ \varrho_{yx_1} & 1 & \varrho_{x_1x_2} \\ \varrho_{yx_2} & \varrho_{x_1x_2} & 1 \end{bmatrix}\right]$$

For simplicity of notation, write $\varrho_{x_1x_2}$ as $\varrho$, $\varrho_{yx_1}$ as $\varrho_1$ and $\varrho_{yx_2}$ as $\varrho_2$.

$$E\left(Y \left| \frac{X_1 - \mu_{x_1}}{\sigma_{x_1}} > x_1, \frac{X_2 - \mu_{x_2}}{\sigma_{x_2}} > x_2\right.\right) = \mu_y + \sigma_y E(W \mid Z_1 > x_1,\ Z_2 > x_2). \quad \text{(A-1)}$$

$$E(W \mid Z_1 > x_1, Z_2 > x_2) = \frac{1}{P(Z_1 > x_1, Z_2 > x_2)} \int_{x_1}^{\infty}\int_{x_2}^{\infty}\int_{-\infty}^{\infty} w f_{W,Z_1,Z_2}(w,u,v)\, dw\, dv\, du$$

$$= \frac{1}{P(Z_1 > x_1, Z_2 > x_2)} \int_{x_1}^{\infty}\int_{x_2}^{\infty} E(W \mid Z_1 = u, Z_2 = v) \times f_{z_1,z_2}(u,v)\, dv\, du \quad \text{(A-2)}$$

where

$$P(Z_1 > x_1, Z_2 > x_2) = P(Z_2 > x_2 \mid Z_1 > x_1)\, P(Z_1 > x_1) = P_{x_2 \cdot x_1}(1 - \Phi(x_1)), \quad \text{(A-3)}$$

$$E(W \mid Z_1 = u, Z_2 = v) = \frac{\varrho_1 (u - \varrho\, v) + \varrho_2 (v - \varrho\, u)}{(1 - \varrho^2)} \quad \text{(A-4)}$$

and

$$f_{z_1,z_2}(u,v) = \frac{1}{2\pi\sqrt{1 - \varrho^2}} \exp\left[-\frac{1}{2\,(1 - \varrho^2)}(u - \varrho\, v)^2 - \frac{1}{2} v^2\right]. \quad \text{(A-5)}$$

When equations (A-4) and (A-5) are substituted into (A-2), a sum of two integrals results. If we ignore the factor $P(Z_1 > x_1, Z_2 > x_2)$ for the moment, the first of these integrals is

$$\frac{\varrho_1}{(1 - \varrho^2)\, 2\pi\sqrt{1 - \varrho^2}} \int_{x_2}^{\infty} \exp\left[-\frac{1}{2} v^2\right] \int_{x_1}^{\infty} (u - \varrho\, v) \times \exp\left[-\frac{1}{2\,(1 - \varrho^2)}(u - \varrho\, v)^2\right] du\, dv$$

$$= \frac{\varrho_1}{2\pi\sqrt{1 - \varrho^2}} \int_{x_2}^{\infty} \exp\left[-\frac{1}{2} v^2\right] \times \exp\left[-\frac{1}{2\,(1 - \varrho^2)}(x_1 - \varrho\, v)^2\right] dv$$

$$= \varrho_1 \left[\frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} x_1^2\right]\right] \frac{1}{\sqrt{2\pi}\sqrt{1 - \varrho^2}} \times \int_{x_2}^{\infty} \exp\left[-\frac{1}{2\,(1 - \varrho^2)}(v - \varrho\, x_1)^2\right] dv$$

$$= \varrho_1 \varphi(x_1)\left[1 - \Phi\left(\frac{x_2 - \varrho\, x_1}{\sqrt{1 - \varrho^2}}\right)\right] = \varrho_1 \varphi(x_1)\, \Phi\left(\frac{\varrho\, x_1 - x_2}{\sqrt{1 - \varrho^2}}\right).$$

The second integral can be treated in the same fashion. Substituting these results and Equation (A-3) into (A-2) and thence into (A-1),

$$E\left(Y \left| \frac{X_1 - \mu_{x_1}}{\sigma_{x_1}} > x_1, \frac{X_2 - \mu_{x_2}}{\sigma_{x_2}} > x_2 \right.\right) = \mu_y + \frac{\sigma_y}{[1 - \Phi(x_1)]\, P_{x_2 \cdot x_1}} \left[\varrho_1 \varphi(x_1)\, \Phi\left(\frac{\varrho\, x_1 - x_2}{\sqrt{1 - \varrho^2}}\right) + \varrho_2 \varphi(x_2)\, \Phi\left(\frac{\varrho\, x_2 - x_1}{\sqrt{1 - \varrho^2}}\right)\right]$$

as shown by COCHRAN (1950).

## Zusammenfassung

Wenn genetische und umweltbedingte Effekte auf einen quantitativen Phänotyp $X = G + E$ von einander unabhängig und normal verteilt sind, dann entspricht die Wahrscheinlichkeitsverteilung des genetischen Werts $G$ unter Individuen mit fixiertem phänotypischem Wert $X$ gleichfalls einer Normalverteilung. Das Mittel dieser a posteriori-Verteilung der genetischen Werte ist $\bar{g} + h^2 (X - \bar{g})$ und die Varianz ist $\sigma_g^2 (1 - h^2)$; wobei $\bar{g}$ das a priori-Mittel von $X$, $h^2$ der Heritabilitätskoeffizient und $\sigma_g^2$ die genetische Varianz sind. Für jeden fixierten Wert von $h^2$ und $\sigma_g^2$ kann daher die a posteriori-Wahrscheinlichkeit, daß der genetische Wert $G$ das Populationsmittel in Verbindung mit einem gegebenen Phänotyp $X$ um einen bestimmten Wert übersteigt, direkt aus den Tabellen einer standardisierten Normalverteilung abgelesen werden. Der erwartete Anteil dieser überlegenen, genetisch bedingten Abweichung unter Individuen, deren phänotypischer Wert einen vorgegebenen Konstantwert übersteigt, kann ebenfalls numerisch errechnet werden. Er wird im vorliegenden Fall graphisch dargestellt. Wenn eine phänotypische Selektion zur Auswahl der besten Phänotypen aus N Individuen erfolgt, sollte N groß genug sein, um mit hoher Wahrscheinlichkeit eine überlegene, genetisch bedingte Abweichung zu erhalten. Die wirksamen Charakteristiken dieses Typs der Selektion werden in tabellierter Form wiedergegeben, die gleichfalls auf numerischer Integration beruht.

## Literature

1. COCHRAN, W. G.: Improvement by means of selection. Proc. Second Berkeley Symposium on Mathematical Statistics and Probability, pp. 449-470 (1950). — 2. DUDLEY, J. W., and LEROY POWERS: Population genetic studies on sodium and potassium in sugar beets (*Beta vulgaris* L.). J. Amer. Soc. Sugar Beet Tech. **11**, 97—127 (1960). — 3. EISENHART, C.: The assumptions underlying the analysis of variance. Biometrics **3**, 1—21 (1947). — 4. FEDERER, W. T., LEROY POWERS, and M. G. PAYNE: Studies on statistical procedures applied to chemical genetic data from sugar beets. Colorado Agr. Exp. Sta. Tech. Bul. 77. 39 pp. (1963). — 5. POWERS, LEROY: Strawberry breeding studies involving crosses between the cultivated varieties (× *Fragaria ananassa*) and the native Rocky Mountain strawberry (*F. ovalis*). J. Agr. Res. **70**, 95—122 (1945). — 6. POWERS, LEROY, D. W. ROBERTSON, and A. G. CLARK: Estimation by the partitioning method of the numbers and proportions of genetic deviates in certain classes of frequency distributions. J. Amer. Soc. Sugar Beet Tech. **9**, 677-696 (1958). — 7. POWERS, LEROY, E. E. REMMENGA, and N. S. URQUHART: The partitioning method of genetic analysis applied to a study of weight per root and percentage sucrose in sugarbeets. Colorado Agr. Exp. Sta. Techn. Bul. 84. 23 pp. (1964).

# Buchbesprechungen / Book Reviews

**Frey, Kenneth J. (Editor): Plant Breeding. A Symposium held at Iowa State University.** Ames, Iowa: Iowa State University Press 1966. $ 6,95.

The papers presented at symposia inevitably vary in relevance to theme and in quality. This symposium on Plant Breeding was no exception, although all of the topics are relevant to the practicing plant breeder. Eleven reports, with accompanying discussions, are presented. The first, by D. C. SMITH, provides a broad review of the development of the profession. Chapters two through nine are devoted to more detailed reviews of major research areas of the science. They include: Plant Explorations and Biosystematics, J. R. HARLAN; Influence of Morphology and Sterility on Breeding Methodology, D. N. DUVICK; Cytogenetics and Plant Improvement, C. R. BURNHAM; Mutation Breeding, W. C. GREGORY; The Role of Pest Resistance in New Varieties, J. C. WALKER; Significance of Experimental Design in Plant Breeding, E. L. LE CLERG; Quantitative Genetics in Plant Improvement, G. F. SPRAGUE; and, Genetic Control of Physiological Processes, A. M. SRB. Emphasis in this section is on review of past performance as it relates to our present state of knowledge. In this regard E. L. LE CLERG provides a particularly lucid presentation of the historical development of concepts, principles and techniques currently utilized in the design of plant breeding experiments. The reviews of DUVICK and BURNHAM are also straightforward.

W. C. GREGORY and G. F. SPRAGUE have placed more emphasis on the future. Furthermore these reports are sufficiently specific to invite critical thought. GREGORY argues from theory of FISHER and experimental results of his own and others, for increased use of mutation breeding. Using FISHER's model of population adaptation, he attacks the argument that mutations are generally deleterious in effect. GREGORY's logic train, if I follow it correctly, is as follows: 1) experimental evidence demonstrates that the phenotypic effects of induced mutations are distributed exponentially, the most frequent class of effects being very low in magnitude, the least frequent being very large in magnitude; 2) according to FISHER's adaptation model, as the magnitude of phenotypic effects of a mutant approaches zero, the probability the effect improves adaptability approaches one half; 3) given 1) and 2) then mutagenesis followed by screening of mutants with large phenotypic effects should result in increased genetic variance without a reduction in the mean. It seems inevitable this hypothesis will stimulate controversy. Hopefully it will stimulate clear cut experimental tests.

G. F. SPRAGUE reviews the most commonly used techniques for estimating genetic variance components in agronomic crops. But his discussion becomes intriguing when he turns to breeding methods. Rather than pushing for new breeding methodology, he suggests a second look at the method of mass selection. SPRAGUE argues that additive genetic variance plays a more important role in crop improvement than it has received credit for in the recent past. He supports this argument with considerable evidence. Of course the comparative efficiencies of breeding methods can be determined in a rather straightforward way — given good estimates of appropriate variance components. The catch here is in obtaining good estimates.

The final two chapters, Plant Breeding — Prospects for the Future by G. W. BURTON and International Needs in Plant Breeding by A. H. MOSEMAN wind up the symposium with a grand view of future problems and strategies for meeting these problems.

It appears that the symposium was designed for the exchange of information and ideas primarily among professional breeders. The intent of publishing these proceedings must therefore be to disseminate this information and these ideas more broadly. The book will be of more interest to breeders of field crops than to breeders of horticultural or vegetable species.

*P. E. Hansche, Davis*