

Guest Lecture – Response to short-term selection
GN 703 Population and Quantitative Genetics

February 15, 2011

Dr. Fikret Isik
Department of Forestry and Environmental Resources,
North Carolina State University

Source: Summarized from Bruce Walsh's Aug 2003. Nordic Summer Course

Outline

General combining ability (GCA, specific combining ability (SCA) and breeding value (BV)	2
Heritability	5
Estimation of heritability - One-way ANOVA	5
Estimation of heritability – Offspring regressions	7
Short-term vs long-term response.....	9
Response (breeders' equation).....	9
The generalized breeders' equation: Accuracy	11
Truncation selection.....	12
Selection efficiency	15
Index selection (multi-trait) response	16
Change in gene frequency under selection	17
Short-term changes in the variance	19

General combining ability (GCA, specific combining ability (SCA) and breeding value (BV)

When a parent is crossed with a large number of other parents, we measure progeny and estimate a mean of that parent. The deviation of the parent mean (\bar{X}) from the mean of all crosses or population mean (μ) is the general combining ability

$$GCA = \bar{X} - \mu$$

We can use a linear model to define the GCA of a parent as

$$y_i = \mu + GCA_p + e_i$$

Where y_i is the parent mean and e_i is the error term.

Example1: Six parents (A to F) were mated to produce 9 crosses (AD, AE, etc.)
Calculate GCA value of parents (From Dr. Garry Hodge course notes)

Cross mean	A	B	C	half-sib means
D	107	95	122	108
E	90	89	121	100
F	85	86	105	92
half-sib means	94	90	116	grand mean = 100

GCA = mean superiority of all offspring of the parent over the population mean.

$$GCA_A = 94 - 100 = -6$$

$$GCA_B = 90 - 100 = -10$$

$$GCA_D = 108 - 100 = +8 \text{ etc.}$$

Any particular cross would have an expected value, which is the sum of the GCA of its parent. The cross may deviate from the mid-parent GCA value and it is called **specific combining ability (SCA)**.

SCA = difference between mean of the cross and the value expected considering the population mean and the GCAs of the parents

In other words,

$$\text{SCA} = \text{full-sib mean} - (\text{mid-parent value})$$

$$\text{SCA} = (\text{full-sib mean} - \mu) - (\text{GCA}_1 + \text{GCA}_2),$$

$$\text{SCA}_{AD} = 107 - (94+108) / 2 = 5 \quad \text{or} \quad = 7 - (-6 + 8) = 5$$

$$\text{SCA}_{AE} = 90 - (94+100) / 2 = -4 \quad \text{or} \quad = 10 - (-6) - (0) = -4$$

etc..

$$\text{Full-sib mean } Y_{XZ} = \mu + \text{GCA}_X + \text{GCA}_Z + \text{SCA}_{XZ}$$

$$\text{Half-sib mean } Y_X = \mu + \text{GCA}_X$$

full-sib mean	A	B	C	half-sib means
D	100-6+8+5	100-10+8-3	100+16+8-2	100+8 = 108
E	100-6+0-4	100-10+0-1	100+16+0+5	100+0 = 100
F	100-6-9-1	100-10-8+4	100+16-8-3	100-8 = 92
half-sib means	100-6 = 94	100-10 = 90	100+16 = 116	100 (grand mean)

Breeding value is the value of genes transmitted to progeny. In diploids and in absence of epistasis, GCA is the 1/2 of the parental additive genetic value (1/2 A_F) because a parent transmits only 1/2 of its alleles to a progeny. We can define BV in a linear model as the average of parental BVs a_m and a_f , the fixed effects (intercept) and the error (e_i).

$$y_i = \mu + 0.5 (a_m + a_f) + e_i$$

Genetic value is the value of genes to itself. It includes non-additive effects (such as dominance) which cannot be passed on to progeny. The difference between genetic and breeding value is largely dominance deviation (assuming epistatic effects are negligible). We can use a linear model to define genetic values, as average of parental BVs, the Mendelian sampling m_i and the error term.

$$y_{ij} = \mu + 0.5 (a_m + a_f) + m_i + e_{ij}$$

Traditional approach of predicting breeding values is based on phenotype and genetic relationship matrix (**A**) derived from the pedigrees and using a **linear mixed model**.

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$$

A vector of observations (**y**) is modeled as a linear combination of other variables observed along with **y**. In the model;

y is the $n \times 1$ vector of observations,

b is the $p \times 1$ vector of fixed effects (e.g., the intercept, sites and blocks within sites),

X is the design matrix that relates elements of fixed effect vector **b** to **y**,

a is the $q \times 1$ vector of random effects (e.g., family and family by site interaction),

Z is the design matrix that relate elements of **a** and **e** vectors to **y**,

e is the $n \times 1$ the vector of random residuals.

The variance of response variable **y** $\text{Var}(\mathbf{y}) = \mathbf{ZGZ}^T + \mathbf{R}$. The \mathbf{ZGZ}^T accounts for the random effects, while the **R** accounts for residual effects

Heritability

Heritability (h^2) is the percent of variation for a trait explained by the genetic effects. It is the correlation between phenotype and breeding value.

- It ranges between 0 and 1. Zero heritability does not mean that the trait is not genetically determined. It only shows that there is no variation within that specific population.
- It is strictly a population parameter.
- It determines the degree of resemblance between parents and offspring, which in turn determines the response to selection.
- Heritability is a completely unreliable predictor for long-term response, although it is generally a good to excellent predictor of short-term response. Why?

How do we estimate heritability?

There two major ways; family design and parent-offspring regression.

Estimation of heritability - One-way ANOVA

Let's assume we have N full-sib families, each with n offspring. The traditional way to analyze such data is analysis of variance based on the linear model.

$$z_{ij} = \mu + f_i + w_{ij}$$

where z_{ij} is the phenotype of the j th offspring of the i th family, f_i is the effect of the i th family and w_{ij} is the residual error resulting from segregation, dominance, and environmental contributions.

The ANOVA partitions the total phenotypic variance (σ_z^2) into the sum of variances from each of the contributing factors:

$$\sigma_z^2 = \sigma_f^2 + \sigma_w^2$$

The phenotypic covariance between members of the same group equals the variance among groups.
Cov(within) = Var(between).

This identity allow us to relate an estimated variance component (e.g.,the between-family variance σ_f^2) with the casual variance components (e.g., σ_A^2) that are our real interest.
 For example, the variance among family effects equals the covariance between full sibs,

$$\sigma_f^2 = \frac{\sigma_A^2}{2} + \frac{\sigma_D^2}{4} + \sigma_{Ec}^2$$

where Ec is the common (or shared) family environmental effects, A is additive, D is dominance genetic variances.

Example 2: Suppose there are N=10 families, each with n=5 progeny measured. Calculate genetic, phenotypic variances and heritability from ANOVA.

<u>Factor</u>	<u>df</u>	<u>SS</u>	<u>MS</u>	<u>E(MS)</u>
Among families	(10-1)=9	SSf=405	45	$\sigma_w^2 + 5\sigma_f^2$
Within families	10(5-1)=40	SSf=805	20	σ_w^2

SOLUTION:

$$\mathbf{Var(f)} = (MS_f - MS_w) / n = (45 - 20) / 5 = \mathbf{5}$$

$$\mathbf{Var(w)} = MS_w = \mathbf{20}$$

$$\mathbf{Var(z)} = \mathbf{Var(f)} + \mathbf{Var(w)} = 5 + 20 = \mathbf{25}$$

$$\text{Additive variance is } \mathbf{Var(A)} = 2\mathbf{Var(f)} = \mathbf{10}$$

(contaminated by D)

Estimated heritability is $h^2 = \text{Var}(A) / \text{Var}(z) = 10/25 = \mathbf{0.4}$

Estimation of heritability - Offspring regressions

The resemblance between offspring and a parent gives 1/2 of heritability. In some sense the simplest design is the parent-offspring regression. The regression of offspring phenotype (z_{oi}) given the phenotypic value of one of its parents (z_{pi}).

$$z_{oi} = \alpha + bz_{pi} + e_i$$

The slope b (the regression coefficient) can also be written as $b_{y|x}$ to signify that the slope is for the regression of y on x , i.e., the denominator in b is the variance of x .

$$z_{oi} = \alpha + b_{o|p}z_{pi} + e_i$$

The above equation implies that the predicted value by **for y given we know x is**

$$z_{oi} = \mu + b_{o|p}(z_{pi} - \mu) + e_i$$

The alternative formulation follows since the regression passes through the mean of both variables (offspring and parental phenotypes). The expected regression slope $b_{o|p}$ is

$$E(b_{o|p}) = \frac{\sigma(z_o, z_p)}{\sigma^2(z_p)} = \frac{\left(\frac{\sigma_A^2}{2}\right) + \sigma(E_o, E_p)}{\sigma_z^2} = h^2 / 2$$

For males, it is generally expected that the covariance between parent and offspring environmental values is zero and the regression slope is $h^2/2$. Greater precision is obtained when both parents are measured, in which case one can regress offspring phenotypes on the mean phenotypes of their parents (also known as the midparent values). The linear model is now

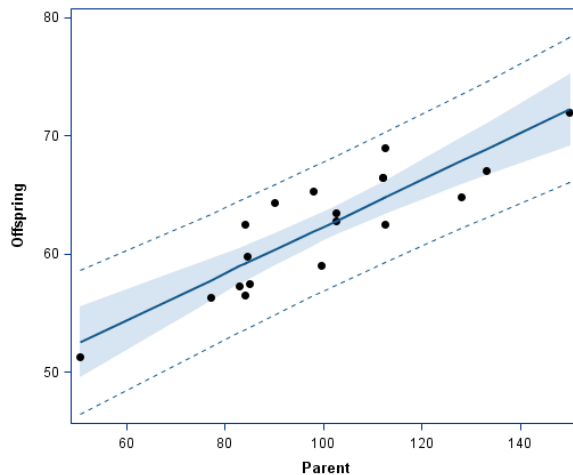
$$z_{oi} = \mu + b_{o|MP} \left(\frac{\sigma(z_{mi} + z_{fi})}{2} - \mu \right) + e_i$$

Correlations and regression slopes are related as follows:

$$r = \text{Cov}(X;Y) / \sqrt{\text{Var}(X)\text{Var}(Y)}$$

$$= \text{Cov}(X;Y) / \text{Var}(X) \sqrt{\text{Var}(X) / \text{Var}(Y)}$$

$$= b_{o|p} z_{pi} / \text{Var}(X) \sqrt{\text{Var}(X) / \text{Var}(Y)} \quad \text{if } \text{Var}(X) = \text{Var}(Y) \text{ then } = b_{o|p} z_{pi} / \text{Var}(X)$$



Example 3: Suppose there are $N=19$ families, each with $n=1$ progeny measured. We regressed height of offspring on height of male parent and obtained the following regression parameters:

Variable	DF	Parameter
Intercept	1	42.57
Parent	1	0.197

Assuming the common environmental effect of parents is negligible, what is the heritability of that population?

SOLUTION: Since $b_{o|p} = h^2/2$, $h^2 = 2 \times 0.197 = 0.394$

Short-term vs long-term response

- If we are only trying to predict a few generations of selection response, knowledge of the base population genetic variances (and in particular the heritability) is usually sufficient to make a satisfactory prediction.

However, as selection proceeds and allele frequencies change significantly, the initial genetic variances essentially lack any predictive power for the long-term response.

The focus in this lecture is on just what predictions we can make for short-term response.

Response (breeders' equation)

Response (R) from short-term selection is the function of heritability (h^2) and the selection differential (S).

$$R = h^2 S \quad (\text{Eq.1})$$

This relationship is often called the breeders' equation. The between-generation change, (the response to selection) R , is the change in means between the population before selection and the population in the next generation.

S is the within-generation change in the mean due to selection as $S = \mu_* - \mu$ where μ is the population mean before selection and μ_* the mean of the parents that reproduce (the population mean after selection). In other words, S is the difference between selected parents and the population as whole (within generation).

Example 4: A tree breeder wants to make selection from a pine population for height growth. The mean of the pine population is $\mu_p = 12\text{m}$. The breeder selects the tallest trees with a mean

of $\mu^* = 18\text{m}$. From the previous studies, the breeder knows that the heritability of height in pine is 0.20.

- 1) What would be the expected response R from this selection and
- 2) What would be the expected mean of the progeny selected?

SOLUTION:

- 1) $R = h^2 S = (\mu^* - \mu) = 0.2*(18-12) = 1.2$ m response from selection.
- 2) The expected mean of the progeny would be $12 + 1.2 = \mathbf{13.2}$ m.

The Selection Intensity i

The selection differential S is not particularly informative when trying to compare the strength of selection on different traits and/or in different populations. A much more useful measure is the **selection intensity i** ,

$$i = S/\sigma_z \quad (\text{Eq.2})$$

which is **the differential expressed as in fractions of phenotypic standard deviations (σ_z)**.

There are different equivalent expressions of breeders' equations:

$$R = h^2 S = \frac{\sigma_A^2}{\sigma_z^2} S \quad (\text{Eq.3})$$

$$R = \sigma_A \frac{\sigma_A}{\sigma_z} \frac{S}{\sigma_z} = \sigma_A \frac{\sigma_A}{\sigma_z} i = h \sigma_A i \quad (\text{Eq.4})$$

The response becomes the function of additive genetic standard deviation (σ_A), square root of heritability (h) and the selection intensity i . Alternatively,

$$R = h^2 \frac{\sigma_z}{\sigma_z} i = h^2 \sigma_z i \quad (\text{Eq.5})$$

Example 5: Given the following parameters, calculate the response R and additive genetic variance? $h^2 = 0.5$, $\sigma_z = 4$, $i = 1$

SOLUTION:

$$\text{Since } R = h^2 \sigma_z i = 0.5 \times 4 \times 1 = 2$$

$$2 = h^2 \sigma_z i = \frac{\sigma_A^2}{\sigma_z^2} \sigma_z i = \frac{\sigma_A^2}{\sigma_z} i$$

$$\sigma_A^2 = \sigma_z i = 2 \times 4 = 8$$

- The breeders' equation can be valid in predicting response over several generations but not longer.
- The most important point is the changes in gene frequencies from one generation to other.
- As gene frequencies changes (due to selection) genetic variances and heritability also change.

The generalized breeders' equation: Accuracy

We can extent the breeders' equation to apply to much more general selection schemes beyond simply choosing an individual solely on the basis of its phenotype. First note that

$$h^2 \sigma_z = \left(\frac{\sigma_A^2}{\sigma_z^2} \right) \sigma_z = \left(\frac{\sigma_A \sigma_A}{\sigma_z} \right) = \left(\frac{\sigma_A}{\sigma_z} \right) \sigma_A = h \sigma_A$$

Hence we can write equation 5 as

$$R = h \sigma_A i$$

Recall that h is simply the correlation between an individual's breeding (A) and phenotypic (P) values, $h = r(PA)$. This correlation quantifies the ability to predict the breeding value of an individual from some measure (here that individual's phenotype) and is called the accuracy of the selection scheme used to choose parents.

We can thus express the breeders' equation in terms of the accuracy of selection as

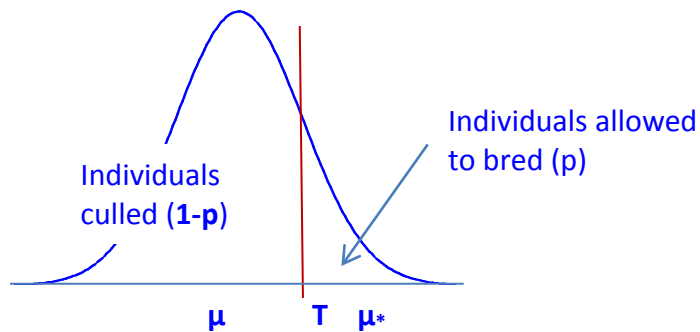
$$R = i r(PA) \sigma_A$$

Hence, the breeders' equation can be considered as the following product:

Response = (Intensity)*(Accuracy in Predicting Breeding Value)*(Usable Variance)

Truncation selection

- Truncation selection is described as percent of population saved
- Only the largest/tallest or smallest individuals are allowed to breed to form the next population.
- Very common in plant and animal breeding programs



The mean of selected tail is μ_* , the mean of whole populations is μ . T is the truncation point; p is the percent of population selected. We can easily calculate selection intensity as $S = \mu_* - \mu$.

Correlated response (CR) and indirect selection

The mean of trait X in a breeding population can change in two ways:

1) As a direct response to selection on X (R_x), which we already covered above

$$R_x = i_x h_x \sigma_{Ax} = i_x \left(\frac{\sigma_{Ax}}{\sigma_{Px}} \right) \sigma_{Ax} = i_x \left(\frac{\sigma_{Ax}^2}{\sigma_{Px}} \right)$$

2) As a correlated, indirect, response to direct selection on Y (CR_x)

When X and Y are genetically correlated, selection on X will result in change in Y too. Such a change in the unselected trait (Y) is correlated response.

The response to selection of trait X is (by definition) the mean breeding value of selected individuals.

Thus, the change in trait Y in response to selection on X is the regression of breeding value of Y on the breeding value of X (Falconer and MacKay page 317).

The slope (b) of the regression is;

$$b_{Ax|Ay} = \text{cov}_A / \sigma_{Ax}^2$$

Recall that the correlation between X and Y equals $r = \text{cov}(X;Y) / \sigma_{Ax}\sigma_{Ay}$. Rearranging, we can express the covariance as $\text{cov}(X; Y) = r \sigma_{Ax}\sigma_{Ay}$

$$= (r_A \sigma_{Ax} \sigma_{Ay}) / \sigma_{Ax}^2$$

$$b = r_A \left(\frac{\sigma_{Ay}}{\sigma_{Ax}} \right)$$

The regression of breeding values of Y on breeding values of X;

$$Y = b_{Ax|Ay} X, = r_A \left(\frac{\sigma_{Ay}}{\sigma_{Ax}} \right) X$$

$$CR_y = b R_x$$

$$= b (i_x h_x \sigma_{Ax}) = r_A \left(\frac{\sigma_{Ay}}{\sigma_{Ax}} \right) (i_x h_x \sigma_{Ax})$$

$$= i_x h_x r_A \sigma_{Ay} \quad (\text{substituting } \sigma_{Ay} \text{ with } h_y \sigma_{Py} \text{ gives})$$

$$= i_x h_x r_A h_y \sigma_{Py}$$

The $h_x h_y r_A$ is the co-heritability of trait X and Y. Where, i_x is the selection intensity.

Selection efficiency

If traits X and Y are genetically correlated, and if trait X is difficult, expensive and time consuming to measure, then we may make selection on Y to improve the mean response in X. Sometimes indirect selection of X could be more efficiently than direct selection.

$$R_x = i_x h_x \sigma_{Ax} \quad (\text{R to direct selection})$$
$$CR_x = i_y h_y r_A \sigma_{Ax}, \quad (\text{Correlated R to indirect selection})$$

$$E = CR_x / R_x$$
$$= i_y h_y r_A \sigma_{Ax} / i_x h_x \sigma_{Ax} \quad (\text{remember that } h_x = \sigma_{Ax} / \sigma_{Px})$$
$$= i_y h_y r_A / i_x h_x \quad (\text{assuming } i_y = i_x)$$
$$= r_A h_y / h_x$$

Selection efficiency E can be greater than 1 if $h_y > h_x$ and if r_A is high.

Example 6: Wood density of trees was measured indirectly using a drilling tool called the Resistograph (Y). The actual of wood density was also measured (X). The objective of the research was to develop indirect wood density assessment of trees in tree improvement programs. Researchers estimated the following genetic parameters:

$$Y = h^2_{(\text{resistograph})} = 0.29, \quad X = h^2_{(\text{density})} = 0.61, \quad r_A = 0.95,$$

Assuming the same selection intensity for two methods, calculate efficiency of indirect selection.

$$\text{SOLUTION: } E = 0.95 * \sqrt{0.29} / \sqrt{0.61} = 0.66$$

Index selection (multi-trait) response

Selection response for one trait can be written in matrix format

$$\begin{aligned} R &= i h^2 \sigma_P = h^2 S \\ &= (\sigma_A^2 / \sigma_P^2) S \\ &= \sigma_A^2 \sigma_P^{-1} S \end{aligned}$$

Where, S is the selection differential. In multi trait selection there are multi genetic and phenotypic variances. Suppose there are n traits, their selection differentials S would be a vector. Their genetic variances and covariances would be a matrix. Similarly their phenotypic variances would be a matrix. For two characters;

$$\mathbf{S} = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}$$

Likewise, define the phenotypic and additive genetic covariance matrices, P and G , respectively, as matrices whose element in the i th row and j column is the covariance (phenotype or additive genetic) between traits i and j .

$$\mathbf{G} = \begin{bmatrix} \sigma_{Ax}^2 & \text{COV}_{Axy} \\ \text{COV}_{Axy} & \sigma_{Ay}^2 \end{bmatrix}$$
$$\mathbf{P} = \begin{bmatrix} \sigma_{Px}^2 & \text{COV}_{Pxy} \\ \text{COV}_{Pxy} & \sigma_{Py}^2 \end{bmatrix}$$

Note that the diagonal elements are the variances.

Where **G** is the additive genetic variance-covariance matrix of two traits, **P** is the phenotypic variance and covariance matrix. The response of selection for multi traits becomes;

$$\mathbf{R} = \mathbf{G} \mathbf{P}^{-1} \mathbf{S}$$

This equation is often referred to as the **multidimensional breeders' equation**.

Change in gene frequency under selection

With artificial selection we allow individuals with favorable alleles to breed. The frequency of alleles affecting the trait will change because of the change in gene frequencies, which in turn changes genetic variance and this heritability.

How quickly does selection change the frequency of alleles at loci contributing to a trait under selection? Consider a diallelic locus, with alleles A1 and A2, whose genotypes have the following relative fitnesses:

Genotype	A1A1	A1A2	A2A2
Fitness	1	1+s	1 + 2s

This is an example of **additive fitness**. If q represents the frequency of allele A2 before selection, then the change in the frequency of q after selection is given by

$$\Delta q = \frac{sq(1-q)}{1+2sq} \cong sq(1-q) \quad \text{when } |2sq| \ll 1$$

Under these fitnesses, the change in the frequency of the favorable allele is proportional to s .

In finite populations, genetic drift can overpower the effects of selection. In particular, when $4N_e|s| \ll 1$ the fate of an allele is largely determined by gene drift, rather than selection. In such cases, favorable alleles can easily be lost by drift.

Now consider a locus contributing to a character z under selection. Suppose the genotypes at this locus make the following contribution to the character:

Genotype	A1A1	A1A2	A2A2
Contribution	0	a	2a

For a trait with phenotypic variation σ_z^2 under selection intensity i , this induces additive fitnesses on these genotypes, with $s \cong \frac{a}{\sigma_z} i$

Hence, the change in allele frequency depends on both the strength of selection i and the relative contribution a/σ_z of the character to the overall trait value.

As expected, loci with larger contributions are under stronger selection than loci with minor contributions and hence have faster allele frequency changes. Thus, many favorable QTL alleles can be lost by drift if either their effects (a/σ_z^2), the strength of selection on the character (i), or the effective population size (N_e) are sufficiently small.

More generally, if the locus shows dominance towards the character, the fitnesses become

Genotype	A1A1	A1A2	A2A2
Contribution	0	a(1 + k)	2a
Induced fitness	1	1 + s(1 + h)	1+2s

where for the induced fitnesses $s = a/\sigma_z^2$ (as above) and $h = k$.

Short-term changes in the variance

Selection has two routes by which to change the genetic variances, and hence the heritability and selection response. First, it can change the frequencies at individual alleles. When the contribution to a trait from any locus is very small, these selection-induced changes in allele frequencies over a few generations are also very small. However, selection also creates correlations between alleles at different loci (**linkage disequilibrium**), and this can result in an immediate change in the variance.

Consider the within-generation change in the variance, $\delta\sigma_z^2 = \sigma_{z^*}^2 - \sigma_z^2$, where $\sigma_{z^*}^2$ is variance after selection. Using regression arguments similar to those leading to the breeders' equation, the expected response in the variance to a single generation of selection is

$$d = \sigma_o^2 - \sigma_p^2 = \frac{h^4}{2} \delta\sigma_z^2$$

Where σ_o^2 is the variance in the offspring, σ_p^2 is the phenotypic variance in the unselected population. This is the variance response analog to the response in mean (the breeders' equation), with $h^4/2$ replacing h^2 and $\delta\sigma_z^2$ replacing S . In many situations (such as truncation selection), we can write **expected change in variance**

$$\sigma_{z^*}^2 = (1-k) \sigma_z^2$$

so that the result of selection is a proportional change in the variance.

It turns out that all the change in the variance is due to a change in the additive genetic variance, so that if V_a denotes the additive variance before selection, then after one generation of selection

$$V_A(1) = V_a + d; \quad V_P(1) = V_A(1) + V_D + V_E = V_P + d$$

where V_P is the phenotypic variance in the base (pre-selection) population. The heritability thus becomes

$$h^2 = \frac{V_A(1)}{V_P(1)} = \frac{V_a + d}{V_P + d}$$

Truncation selection reduces the variance ($\delta\sigma_z^2 < 0$), which results in reduced additive genetic variance and heritability in the next generation, slowing response. This reduction in variance due to selection creating linkage disequilibrium is referred to as the **Bulmer effect**.