Selection Response

By William Beavis, Kendall Lamkey (ISU)

Objectives

- Explain the role of selection on genetic improvement
- Explain all of the components of realized and predicted genetic gain
- Explain why realized genetic gains are always less than predicted genetic gains
- Explain the role of replication in multi-environment tests on predicted and realized genetic gains

Introduction

Consider the underlying theory of selection.

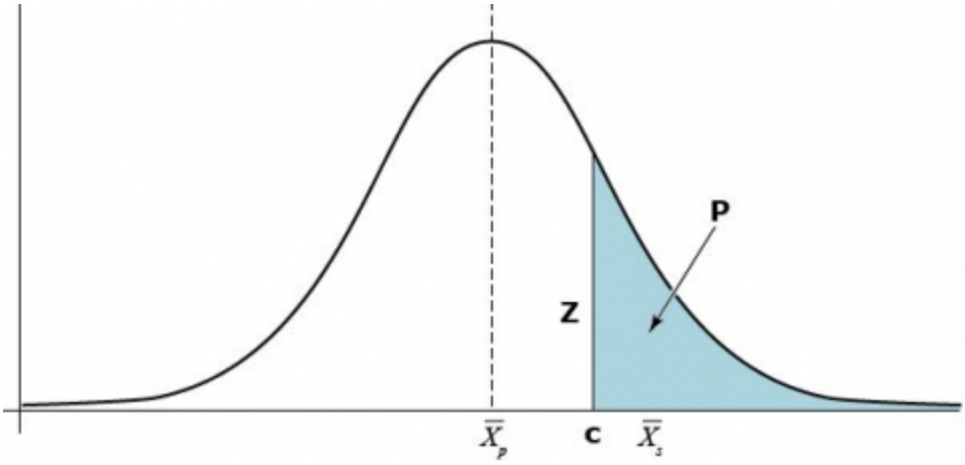

Fig. 1 The normal distribution.

Let $\bar{X}_p$ be the mean phenotypic value of a quantitative trait that is normally distributed in a large random mating population. Also, designate $\bar{X}_s$ as the mean of a selected proportion **P** of this population, where **c** is the truncation point of selection and **Z** is the height of the ordinate at the selection truncation point.

The selection differential is defined as:

$$S = \bar{X}_s - \bar{X}_p$$

If σ^2_p is the phenotypic variance in the population then the standardized selection differential can be written as:

$$i = \frac{S}{\sigma_p} = \frac{\bar{X}_S - \bar{X}_p}{\sigma_p}$$

Selection Response

Selection Response - Definition

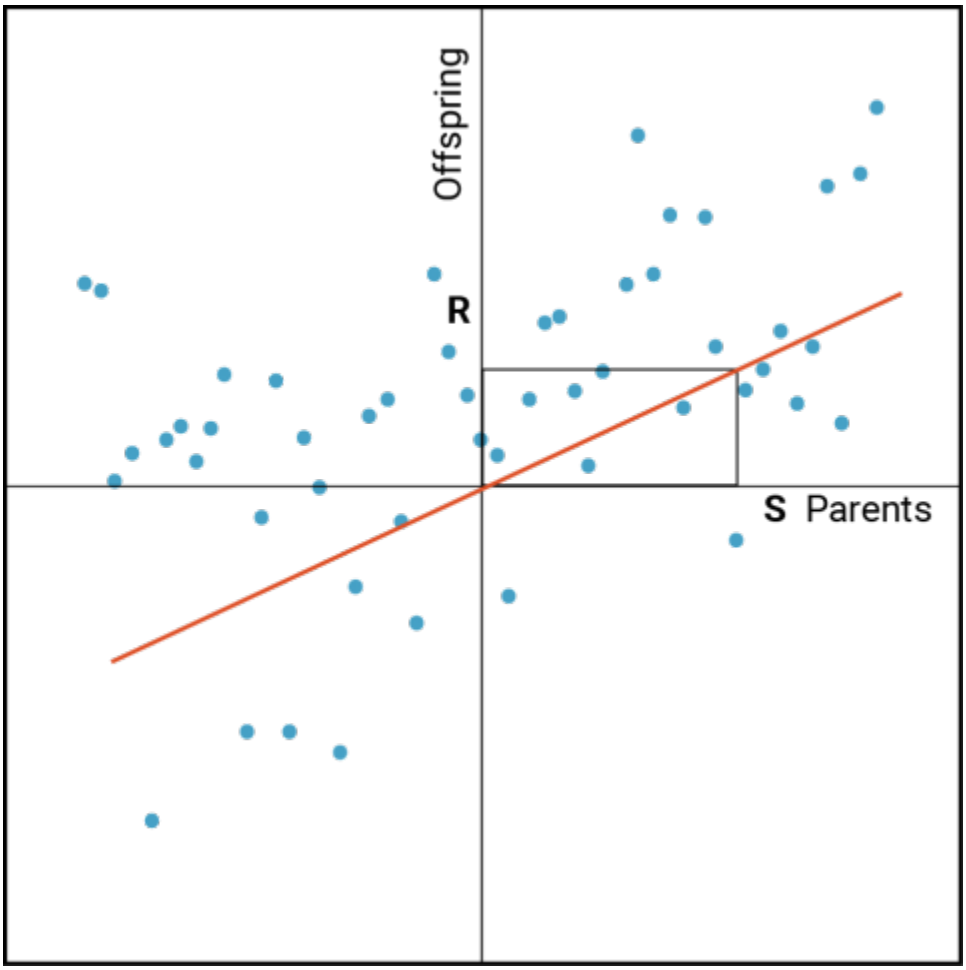

Fig. 2

While $\bar{X}_s$ may be distinctive relative to $\bar{X}_p$, of greater interest are the phenotypes of the progeny derived from crosses among the selected parents $\bar{X}_s$. The predicted response of progeny to selection of their parents can be derived from the relationship between parent and offspring as follows. Designate *R* as the response to selection measured in the offspring (represented as a deviation from the population mean). *S* is the selection differential (represented as a deviation from the population mean) as described on the previous page.

Then the response to selection (R) can be written simply as:

$$R = b_{o\bar{p}}S$$

Equation 1

where $b_{o\bar{p}}$ is the regression coefficient of offspring on the mid-parent value, which can be written as:

$$b_{o\bar{p}} = \frac{Cov(o,\bar{p})}{Var(\bar{p})} = \frac{Cov(o,\bar{p})}{\frac{1}{2}\sigma_P^2}$$

because $Var(\bar{p}) = Var(\frac{P_m + P_f}{2}) = \frac{1}{4}Var(P_m + P_f) = \frac{1}{2}\sigma_P^2$

Also, we can show: $Cov(o,\bar{p}) = \frac{1}{2}\sigma_A^2$

so $b_{O\bar{P}} = \frac{\frac{1}{2}\sigma_A^2}{\frac{1}{2}\sigma_P^2} = \frac{\sigma_A^2}{\sigma_P^2} = h^2$ therefore

$$R = h^2S = ih^2\sigma_p = \Delta G_c$$

Equation 2

which is the selection response or **Genetic Gain**, as Lush defined it in 1940. This equation for ΔG, based on regression of offspring values on mid-parent values, is difficult to apply directly to plant breeding systems because plant breeders typically evaluate hundreds of replicated individuals representing thousands of genotypes grown in replicated plots in dozens to hundreds of environments. Unlike most animal systems, it is possible to replicate progeny genotypes due to the diversity of reproductive biology that is available to plant breeders: clonal propagation, doubled haploids and tolerance to inbreeding through self-pollinations for multiple generations. In the last example, the response units can be several generations removed from the parental (crossing) generation. The type of reproductive biology will affect the details of how we estimate the response to selection, ΔG_c.

Heritability on an Entry-Mean Basis

Recall plant breeders often report heritability from field experiments on an entry-mean basis as

$$\frac{\sigma_g^2}{\sigma_g^2 + (\frac{\sigma_{ge}^2}{e}) + (\frac{\sigma_e^2}{re})}$$

Equation 3

Although Equation 3 is similar to Lush's **broad sense heritability**, it is not exactly the same concept because it can be 'adjusted' by adding replicates and environments to reduce the impact of σ^2_{ge} and σ^2_e on the estimated phenotypic variance.

The problem for plant breeders was that the concept of evaluating individual plants and the performance of their progeny to obtain an estimate of heritability simply is of no practical use for most crops where plot performance is the basis for selection. Hanson attempted to address this by framing the multiple concepts of heritability within the context of genetic gain (1963).

Hanson defined heritability as "the fraction of the selection differential expected to be gained when selection is practiced on a defined reference unit." Given the standard definition for selection response,

$$\Delta G = i\frac{\sigma_g^2}{\sigma_{\bar{y}}^2}\sigma_{\bar{y}} = ih^2\sigma_{\bar{y}} = R$$

we can then solve for h^2 and obtain the following expression:

$$h^2 = \frac{R}{i\sigma_{\bar{y}}}$$

Equation 4

i.e., the standardized response to selection, or realized heritability.

Context of Heritability

Within the framework of genetic gain, Hanson defined heritability in such a manner as to be consistent with the original concept, while at the same time taking into consideration that it has little meaning unless the selection units (entry means) and response units are defined. Thus, when plant breeders wish to communicate information about heritability they should specify:

1. A reference population of genotypes.
2. A reference population of environments. i.e., the target environments.
3. Selection units
4. Response units

This context emphasizes the purpose for obtaining variance component estimates, usually for the purpose of comparing genetic gains (ΔG) under various possible breeding procedures. The results are used to make decisions about which procedure to employ. Indeed, it is in this context that variance components of heritability are used as “plug in values” (Sprague and Eberhart, 1977) for a six step decision making algorithm that uses ΔG as an arbiter for comparing breeding methods (Fehr, 1994; Chapter 17). Actually, this back of the envelope algorithm is fairly insensitive to the estimated heritability values and there are more effective means of optimizing genetic gain, number of generations and costs.

Holland's Synthesis

A thorough review of heritability and how it should be interpreted to compare ΔG by plant breeders was given by Holland et al, (2003). The review was essentially an update to a review by Nyquist (1991) where the updates were based on computational techniques, REML in particular, for obtaining appropriate estimates of variance components. He indicated that plant breeders have traditionally used the method of moments (covered in later slides) to estimate genotypic and phenotypic correlations between traits on the basis of a multivariate analysis of variance (MANOVA), and pointed out the key drawbacks of using the method that include the possibility of obtaining estimates outside of parameter bounds, reduced estimation efficiency, and ignorance of the estimators' distributional properties when data are missing.

With Hanson's response, the response to selection can be rewritten as:

$$R = \beta_{SR} S$$

where $ß_{SR}$ is the regression coefficient of the response units on the selection units and is equal to:

$$\frac{Cov(R,S)}{Var(S)}$$

Family Structure

Assume our selection and response units are represented by some family structure, say half sibs, or full sibs, or recombinant inbred lines, as examples. Also, recall that we can equate the genotypic variance component, designated as *f* for family relationships, to the genetic covariance of relatives. Thus, the *Cov(R,S) = Var(f)*. Also, note that the *Var(S)* is the phenotypic variance among the entry means. Thus, $ß_{SR}$ is the proportion of variance among family units relative to the phenotypic variance among entry means. We might refer to this as heritability of the family units:

$$\frac{\sigma_f^2}{\sigma_p^2} = h_f^2$$

Equation 5

If the replicated plots consist of half sibs from a random mating population, then the variance component among half sibs on an entry mean basis is equal to the covariance of the half sibs, i.e.,

$Cov(HS) = \frac{1}{4}(1+F)\sigma_A^2$, ignoring epistasis.

Narrow-Sense Heritability of Half-Sibs

Thus, it is possible to utilize the estimated variance components from an ANOVA to estimate a “narrow sense heritability” by simply multiplying this variance component by 4/(1+F) and plugging the value into Equation 5.

$$\frac{\sigma_A^2}{\sigma_G^2 + (\frac{\sigma_{GE}^2}{E}) + (\frac{\sigma_\varepsilon^2}{rE})}$$

Notice that this is not the same as the original narrow sense heritability as defined by Lush (1940), but is a narrow sense heritability for a population of half sibs.

Next consider the numerator in the equation above. For the case of half sibs we’ve learned that

$$\sigma_f^2 = Cov(HS) = \frac{(1+F)}{4}\sigma_A^2 + \left[\frac{(1+F)}{4}\right]^2 \sigma_{AA}^2 + \ldots$$

Equation 6

Covariance Estimation

Again, if the data are not balanced, the variance component will not be estimated correctly unless REML is used. Let's assume that we obtain a 'best' estimate for σ_{HS}; either because our data are balanced or we have used REML. Should we use the previous equation for the Cov(R,S)? To answer this we have to recognize that there is a genetic relationship between selection units and response units, i.e., there is a pedigree relationship or coefficient of coancestry between the selection and response units and Equation 6 does not take this into consideration. In the case where both selection units and response units are half sibs the Cov(R,S)

$$= \frac{(1+F)}{4}\sigma_A^2 + \frac{1}{32}[1+F]^2\sigma_{AA}^2 + \cdots$$

Equation 7

Note that if Equation 7 is used, a slightly biased estimate of heritability will result even if best estimates of variance components are obtained. For other types of progeny the bias in the numerator can be much larger.

Example A

Estimation: Narrow sense heritability estimated from a half-sib family experiment with data obtained on **individual plants** in **multiple** independent environments.

A. Heritability on an individual plant basis

- selection among individual plants
- 1 Replication in 1 environment
- response measured in outbred progeny

Invalid Equation

Example B

Estimation: Narrow sense heritability estimated from a half-sib family experiment with data obtained on **individual plants** in **multiple** independent environments.

B. Family Heritability on a plot basis (half-sib family, single plot mean values)

- selection among plot means
- 1 Replication in 1 environment
- outbred progeny

Invalid Equation

Computational Considerations

Example C

Estimation: Narrow sense heritability estimated from a half-sib family experiment with data obtained on **individual plants** in **multiple** independent environments.

C. Family Heritability

- selection among half-sib family meansaveraged over environments
- outbred progeny

Invalid Equation

The only way to remove the bias is to include both selection units and response units in the analyses. This is not the same thing as including both groups in the same sets of environments.

Method of Moments

Let's next explore computational nuances of these concepts in the context of plant breeding populations. Consider first the evaluation of half sibs from a random mating population in a replicated Multi-Environment Trial. Let the phenotypic variance of the selection units be designated σ_p^2. From an introductory course in statistics, we were taught that the phenotypic variance on an entry means basis can be obtained directly from Ordinary Least Squares (OLS) ANOVA by equating the estimated Mean Squares (MS) with Expected Mean Squares (EMS). This is also known as the Method of Moments (MoM). Thus an estimate of σ_p^2 =

$$\frac{MS_f}{er} = \sigma_f^2 + \frac{\sigma_{fE}^2}{e} + \frac{\sigma_\epsilon^2}{er}$$

When to Use Method of Moments

It turns out that application of MoM is appropriate only if the data are from a balanced experiment, i.e., the number of genotypes, in this case families or genotypic entries, is the same across reps and environments. Recall that lsmeans are useful for estimates of entry means in the case of unequal replication per environment. Next we need to learn how to obtain estimates of the variance components for unbalanced data sets.

The most obvious problem is that the coefficients of the variance components are not equal to the products of the numbers of reps and environments in the EMS. Addressing this problem is fairly straight-forward (Milliken and Johnson, 1992). A more difficult problem is that the estimates of the variance components themselves are no longer the “best” estimates. The solution, as described by Holland et al (2003) is to obtain Restricted Expected Maximum Likelihood (REML) estimates in a Mixed Model Procedure (MMP).

REML

For example let's consider the case of half sib progeny. Recall that

$$\sigma_f^2 = Cov(HS) = \sigma_{HS} = \frac{(1+4)}{4}\sigma_A^2 + [\frac{(1+4)}{4}]^2\sigma_{AA}^2 + \cdots$$

Equation 8

If the data are not balanced the variance component will not be estimated correctly unless REML is used. Let's assume that we obtain a 'best' estimate for σ_{HS}; either because our data are balanced or we have used REML. Should we use Equation 8 for the *Cov(R,S)*? To answer this we have to recognize that there is a genetic relationship between selection units and response units, i.e., there is a pedigree relationship or coefficient of coancestry between the selection and response units and Equation 8 does not take this into consideration. In the case where both selection units and response units are half sibs the

$$Cov(R,S)(= \frac{(1+F)}{4}\sigma_A^2 + \frac{1}{32}[1+F]^2\sigma_{AA}^2 + \cdots$$

Thus, if Equation 8 is used, a slightly biased estimate of heritability will result even if REML based estimates of variance components are obtained. For other types of progeny the bias in the numerator can be much larger. Thus, the predicted genetic gain that might be used for planning purposes or comparison of possible breeding methods will be overestimated.

Acknowledgements

This module was developed as part of the Bill & Melinda Gates Foundation Contract No. 24576 for Plant Breeding E-Learning in Africa.

Quantitative Genetics Selection Response Author: William Beavis, and Kendall Lamkey (ISU)

Multimedia Developers: Gretchen Anderson, Todd Hartnell, and Andy Rohrback (ISU)

How to cite this module: Beavis, W. and K. Lamkey. 2016. Selection Response. *In* Quantitative Genetics, interactive e-learning courseware. Plant Breeding E-Learning in Africa. Retrieved from **https://pbea.agron.iastate.edu**.