Correlated traits: Index selection

Jinliang Yang
Nov. 19th, 2018

The correlation between two traits

$$r_P = r_A h_X h_Y + r_E e_X e_Y$$

- r_P: the phenotypic correlation between two traits X and Y
- r_A: the genetic correlation due to breeding values between X and Y
- r_E: the environmental correlation between X and Y, including non-additive genetic effects
- h^2: heritability
- e^2: $1 - h^2$

Genetic and **environmental correlation** come together to create the phenotypic correlation.

Correlated response to selection

What is the expected BV of Y given a BV of X?

$b_{A_{YX}} = r_A \frac{\sigma_{A_Y}}{\sigma_{A_X}}$ and $R_X = i h_X \sigma_{A_X}$, correlated response of trait Y is:

Correlated response to selection

What is the expected BV of Y given a BV of X?

$b_{A_{YX}} = r_A \frac{\sigma_{A_Y}}{\sigma_{A_X}}$ and $R_X = ih_X \sigma_{A_X}$, correlated response of trait Y is:

$$\begin{aligned} CR_Y &= b_{A_{YX}} R_X \\ &= r_A \frac{\sigma_{A_Y}}{\sigma_{A_X}} ih_X \sigma_{A_X} \\ &= ih_X r_A \sigma_{A_Y} \end{aligned}$$

Correlated response to selection

What is the expected BV of Y given a BV of X?

$b_{A_{YX}} = r_A \frac{\sigma_{A_Y}}{\sigma_{A_X}}$ and $R_X = ih_X\sigma_{A_X}$, correlated response of trait Y is:

$$\begin{aligned} CR_Y &= b_{A_{YX}} R_X \\ &= r_A \frac{\sigma_{A_Y}}{\sigma_{A_X}} ih_X\sigma_{A_X} \\ &= ih_X r_A \sigma_{A_Y} \end{aligned}$$

Because $\sigma_{A_Y} = h_Y\sigma_{P_Y}$, therefore,

$$CR_Y = ih_X h_Y r_A \sigma_{P_Y}$$

Correlated response to selection

What is the expected BV of Y given a BV of X?

$b_{A_{YX}} = r_A \frac{\sigma_{A_Y}}{\sigma_{A_X}}$ and $R_X = ih_X\sigma_{A_X}$, correlated response of trait Y is:

$$
\begin{aligned}
CR_Y &= b_{A_{YX}} R_X \\
&= r_A \frac{\sigma_{A_Y}}{\sigma_{A_X}} ih_X\sigma_{A_X} \\
&= ih_X r_A \sigma_{A_Y}
\end{aligned}
$$

Because $\sigma_{A_Y} = h_Y\sigma_{P_Y}$, therefore,

$$CR_Y = ih_X h_Y r_A \sigma_{P_Y}$$

In the formula:

- $h_X h_Y r_A$ is referred to as the **coheritability**, as it takes the place of the heritability in the direct response equation.

Indirect selection

The trade-off between selection on a correlated trait and direct selection on a trait can also be seen:

$$\frac{CR_Y}{R_Y} = \frac{i_X h_X r_A \sigma_{A_Y}}{i_Y h_Y \sigma_{A_Y}} = \frac{i_X h_X r_A}{i_Y h_Y}$$

Indirect selection

The trade-off between selection on a correlated trait and direct selection on a trait can also be seen:

$$\frac{CR_Y}{R_Y} = \frac{i_X h_X r_A \sigma_{A_Y}}{i_Y h_Y \sigma_{A_Y}} = \frac{i_X h_X r_A}{i_Y h_Y}$$

Assuming seletion intensity is the same

- When $h_X r_A > h_Y$, a correlated response from selection on a secondary trait (X) is greater than response to direct selection on Y .

Indirect selection

The trade-off between selection on a correlated trait and direct selection on a trait can also be seen:

$$\frac{CR_Y}{R_Y} = \frac{i_X h_X r_A \sigma_{A_Y}}{i_Y h_Y \sigma_{A_Y}} = \frac{i_X h_X r_A}{i_Y h_Y}$$

Assuming seletion intensity is the same

- When $h_X r_A > h_Y$, a correlated response from selection on a secondary trait (X) is greater than response to direct selection on Y .
  - Note that this exact same property applies to usefulness of molecular markers for **marker-assisted selection**

Indirect selection

The trade-off between selection on a correlated trait and direct selection on a trait can also be seen:

$$\frac{CR_Y}{R_Y} = \frac{i_X h_X r_A \sigma_{A_Y}}{i_Y h_Y \sigma_{A_Y}} = \frac{i_X h_X r_A}{i_Y h_Y}$$

Practical considerations

- If trait Y is very **expensive** and **difficult** to measure, but trait X is very cheap and easy to measure.
  - e.g. high-throughput phenotyping techonologies

Indirect selection

The trade-off between selection on a correlated trait and direct selection on a trait can also be seen:

$$\frac{CR_Y}{R_Y} = \frac{i_X h_X r_A \sigma_{A_Y}}{i_Y h_Y \sigma_{A_Y}} = \frac{i_X h_X r_A}{i_Y h_Y}$$

Practical considerations

- If trait Y is very **expensive** and **difficult** to measure, but trait X is very cheap and easy to measure.
  - e.g. high-throughput phenotyping techonologies
- Or the desired traits is measurable **in one sex** only, but the secondary traits is measurable in both.
  - e.g. milk yield and body weight in dairy cow

Indirect selection

The trade-off between selection on a correlated trait and direct selection on a trait can also be seen:

$$\frac{CR_Y}{R_Y} = \frac{i_X h_X r_A \sigma_{A_Y}}{i_Y h_Y \sigma_{A_Y}} = \frac{i_X h_X r_A}{i_Y h_Y}$$

Genotype-by-environment interaction

- Performance in different environments can be regarded as **two separate, but correlated traits**.
  - Improvement in one environment by selection in another environment can be predicted by knowing the heritability in each environment and the genetic correlation between them.

Selection on multiple traits

Total economic value is a composite of many traits. How does one **maximize response for many traits** simultaneously?

Selection on multiple traits

Total economic value is a composite of many traits. How does one **maximize response for many traits** simultaneously?

Options for multiple trait selection

- **Tandom selection**:
  - Selection for one trait at a time until that trait is improved to a desired level.
  - After that, selection proceeds for another trait.

Selection on multiple traits

Total economic value is a composite of many traits. How does one **maximize response for many traits** simultaneously?

Options for multiple trait selection

- **Tandom selection**:
    - Selection for one trait at a time until that trait is improved to a desired level.
    - After that, selection proceeds for another trait.
- **Independent culling levels**:
    - Only individuals that meet the minimum standard for each trait are selected.

Selection on multiple traits

Total economic value is a composite of many traits. How does one **maximize response for many traits** simultaneously?

Options for multiple trait selection

- **Tandom selection**:
  - Selection for one trait at a time until that trait is improved to a desired level.
  - After that, selection proceeds for another trait.
- **Independent culling levels**:
  - Only individuals that meet the minimum standard for each trait are selected.
- **Index selection**:
  - Select for multiple traits simultaneously by constructing an index value.
  - Index value is then treated as a single economic trait.

Genetic merit

- Any individual has a genetic value for what we'll call as **genetic merit**, or simply, **merit**.

Genetic merit

- Any individual has a genetic value for what we'll call as **genetic merit**, or simply, **merit**.
    - Merit is the summation of all traits contributing to an individual's worth, fitness, economic value, etc.

Genetic merit

- Any individual has a genetic value for what we'll call as **genetic merit**, or simply, **merit**.
  - Merit is the summation of all traits contributing to an individual's worth, fitness, economic value, etc.

The true value of merit is represented as:

$$\begin{aligned} T &= a_1 G_1 + a_2 G_2 + \ldots + a_m Gm \\ &= \sum_{i=1}^{m} a_i G_i \end{aligned}$$

- G_i is the genetic value for trait i
- a_i is the economic weight placed on trait i
  - The economic weights are set by the breeder according to production needs and value.

Index trait

To accurately predict genentic merit T, we want to combine the values of multiple traits into one value, denoted I.

Index trait

To accurately predict genentic merit T, we want to combine the values of multiple traits into one value, denoted I.

$$
\begin{aligned}
I &= b_1 P_1 + b_2 P_2 + \ldots + b_m Pm \\
&= \sum_{i=1}^{m} b_i P_i
\end{aligned}
$$

- P_i is the phenotypic value for trait i that goes into the index
- b_i is the weighting factor on trait i

Index trait

To accurately predict genentic merit T, we want to combine the values of multiple traits into one value, denoted I.

$$
\begin{aligned}
I &= b_1 P_1 + b_2 P_2 + \ldots + b_m Pm \\
&= \sum_{i=1}^{m} b_i P_i
\end{aligned}
$$

- P_i is the phenotypic value for trait i that goes into the index
- b_i is the weighting factor on trait i

Correlation between T and I

The goal is to **find the values of the b_is** that could **maximize the correlations** between T and I, or r_{TI}.

$$
r_{TI} = \frac{Cov(T, I)}{\sigma_T \sigma_I}
$$

Correlation between T and I

$$r_{TI} = \frac{Cov(T, I)}{\sigma_T \sigma_I}$$

Obtaining the maximum value of the correlation involves taking the derivative and setting to zero.

Correlation between T and I

$$r_{TI} = \frac{Cov(T,I)}{\sigma_T \sigma_I}$$

Obtaining the maximum value of the correlation involves taking the derivative and setting to zero.

Finally, we get:

$$\begin{bmatrix} Var(G_1) & Cov(G_1,G_2) & \cdots & Cov(G_m,G_1) \\ Cov(G_1,G_2) & Var(G_2) & \cdots & Cov(G_m,G_2) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(G_1,G_m) & Cov(G_2,G_m) & \cdots & Var(G_m) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} =$$

$$\begin{bmatrix} Var(P_1) & Cov(P_1,P_2) & \cdots & Cov(P_m,P_1) \\ Cov(P_1,P_2) & Var(P_2) & \cdots & Cov(P_m,P_2) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(P_1,P_m) & Cov(P_2,P_m) & \cdots & Var(P_m) \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Matrix Notation

$$\mathbf{Ga} = \mathbf{Pb}$$

- $\mathbf{P}$ is the phenotypic variance-covariance matrix
- $\mathbf{G}$ is the genetic variance-covariance matrix
- $\mathbf{a}$ is the vector of **known** economic weights
- $\mathbf{b}$ is the vector of **unknow** weights, or the weights to be applied to the phenotypic values of the different traits composing the index, I.

Matrix Notation

$$\mathbf{G}\mathbf{a} = \mathbf{P}\mathbf{b}$$

- $\mathbf{P}$ is the phenotypic variance-covariance matrix
- $\mathbf{G}$ is the genetic variance-covariance matrix
- $\mathbf{a}$ is the vector of **known** economic weights
- $\mathbf{b}$ is the vector of **unknow** weights, or the weights to be applied to the phenotypic values of the different traits composing the index, I.

In matrix algebra, taking the inverse of a matrix and multiplying is equivalent to division in scalar algebra.

$$\begin{aligned}
\mathbf{G}\mathbf{a} &= \mathbf{P}\mathbf{b} \\
\mathbf{P}^{-1}\mathbf{G}\mathbf{a} &= \mathbf{P}^{-1}\mathbf{P}\mathbf{b} \\
\mathbf{P}^{-1}\mathbf{G}\mathbf{a} &= \mathbf{b}
\end{aligned}$$

Optimum index

The vector of weights, which is called the **Smith-Hazel index**, or **optimum index**:

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{a}$$

Optimum index

The vector of weights, which is called the **Smith-Hazel index**, or **optimum index**:

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{a}$$

- The vector of $\mathbf{b}$ can be estimated if we know the **phenotypic and genetic variances** and the **phenotypic and genetic covariances**

Optimum index

The vector of weights, which is called the **Smith-Hazel index**, or **optimum index**:

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{a}$$

- The vector of $\mathbf{b}$ can be estimated if we know the **phenotypic and genetic variances** and the **phenotypic and genetic covariances**
- This assumes that we have good information on the economic weights, which can be hard to determine.

Optimum index

The vector of weights, which is called the **Smith-Hazel index**, or **optimum index**:

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{a}$$

- The vector of $\mathbf{b}$ can be estimated if we know the **phenotypic and genetic variances** and the **phenotypic and genetic covariances**
- This assumes that we have good information on the economic weights, which can be hard to determine.
- Then the resulting vector of weights will give a selection index that maximized genetic gain in T, the true genetic marit.

Optimum index

The vector of weights, which is called the **Smith-Hazel index**, or **optimum index**:

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{a}$$

- The vector of $\mathbf{b}$ can be estimated if we know the **phenotypic and genetic variances** and the **phenotypic and genetic covariances**
- This assumes that we have good information on the economic weights, which can be hard to determine.
- Then the resulting vector of weights will give a selection index that maximized genetic gain in T, the true genetic marit.
- The drawback of this index is that the genetic variance and covariances are often estimated with large amounts error, which may reduce the correlation between I and T.

A problem

This litter size of mice could be increased:

- 1) by selection of females for their little size
- 2) by selection of both parents for body weight

Which would be the better of these two simple procedures, given the following parameters?

- h^2 of litter size: 0.22
- h^2 of body weight: 0.35
- Genetic correlation r_A: 0.43
- Proportion of selection:
  - Females: 25% (i_f =1.271)
  - Males: 10% (i_m =1.755)