Multi-trait selection

Prof. Roberto Fritsche-Neto

**rfneto@agcenter.lsu.edu**

Baton Rouge. April 17th. 2023

Genetic correlation

- **The linear relationship between two variables (ρ)**

- **Suggests the relationship's magnitude and direction**

- **The most common estimator is *Pearson* coefficient (r) - normal distribution.**

$$r_{xy} = \frac{Cov(x, y)}{\sqrt{\sigma_x^2 . \sigma_y^2}} \qquad t_{calc} = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

- *Spearman* **– non-normal distribution or rank**

- **Non-Dimensional. and its value does not exceed unity** **$-1 \leq 0 \geq 1$**

- **Significance via *test t***

- *r = 0 means the absence of a linear relationship*
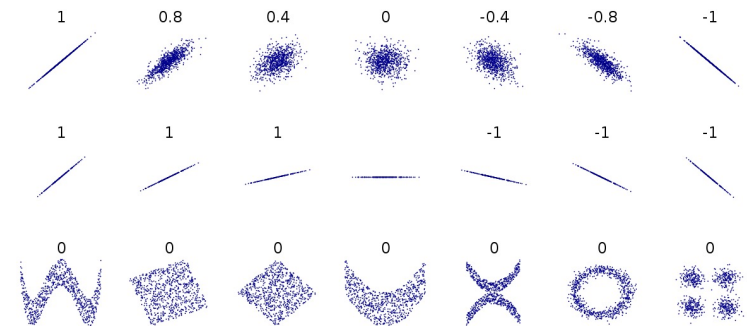
Importance

- **We do not breed for isolated traits**

- **Changes caused by selection in one trait may change another**
- ***Example: Grain yield and plant height in corn***

- **Genetic causes: pleiotropy and linked genes (transient)**

- **Indirect selection of characters that are difficult or costly to evaluate**

- **Distinguish and quantify the degree of genetic and environmental association between characters**

Estimating correlations

- **10 popcorn lines. RCBD 3 rep. GY and EC**

Line	Rep	GY	EC	GY + EC
1	1	249.75	404.48	654.23
1	2	227.56	414.52	642.08
1	3	368.18	490.14	858.32
2	1	615.42	359.91	975.33
2	2	756.11	363.74	1119.85
...	...	...	...	...
10	3	471.44	300.48	771.92

SV	DF	MS		E (MS)	MS$_{x+y}$	E (MS)
		GY	EC		GY + EC	
Block	2	-		-	-	-
Line	9	43978	14527	Ve + **b**Vg	29006	$V_{px} + V_{py} + 2COV_{pxy}$
Residual	18	5630	2515	Ve	7203	$Ve_x + Ve_y + 2COVe_{xy}$
Total	29					
Vg		12783	4004			
Ve		5630	2515		rg = -0.66	
h²		0.87	0.83			

- $E|V(X + Y)| = V_x + V_y + 2COV_{xy}$

- $COV_{Pxy} = (MSTx+y - MSTx - MSTy) / 2 = (29006 - 43978 - 14527) / 2 = -14749$

- $COV_{Exy} = (MSRx+y - MSRx - MSRy) / 2 = (7203 - 5631 - 2515) / 2 = -471$

- $COVg_{xy} = (COV_{Pxy} - COV_{Exy}) / b = -4759$

$$r_g = \frac{COVg_{xy}}{\sqrt{Vg_x . Vg_y}} = \frac{-4759}{\sqrt{12783 . 4004}} = -0,66$$

Cause and effect

- **Correlation is not the same as cause and consequence**
- **Two variables can be highly correlated and not be cause and effect**
- **For example. ice cream consumption x corn growth**
- *Temperature*
- **Alternative – break down into direct and indirect effects**
- *Partial correlations*
- *Canonical*
- *Path analysis*
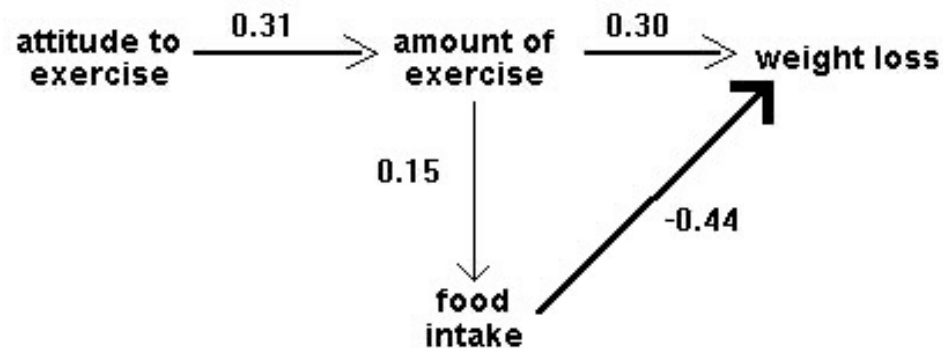
Indirect selection

- **Relationships may vary in magnitude and meaning**
- **Selection for one character can affect another**
- **Gains based on direct selection**
- **Gains based on indirect selection**
- **When is indirect selection advantageous?**
- *Cost and difficulty of evaluation*

%	20	10	1	0.1
i	1.40	1.76	2.67	3.37

$$GS_d = i.\sigma_{g(x)}.h_x$$

$$GS_i = i.r_g.h_x.\sigma_{g(y)}$$

$$r_g.h_x > h_y$$

SV	DF	MS		E (MS)	MS$_{x+y}$	E (MS)
		GY	EC		GY + EC	
Block	2	-		-	-	-
Line	9	43978	14527	Ve + **b**Vg	29006	V$_{px}$ + V$_{py}$ + 2COV$_{pxy}$
Residual	18	5630	2515	Ve	7203	Ve$_x$ + Ve$_y$ + 2COVe$_{xy}$
Total	29					
Vg		12783	4004			
Ve		5630	2515		rg = -0.66	
h^2		0.87	0.83			

$$GSd = 1.76 \; x \; \sqrt{43978} x \; \sqrt{0.87} = 342$$

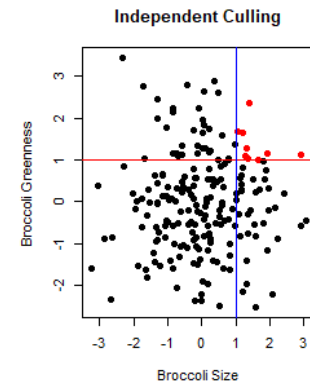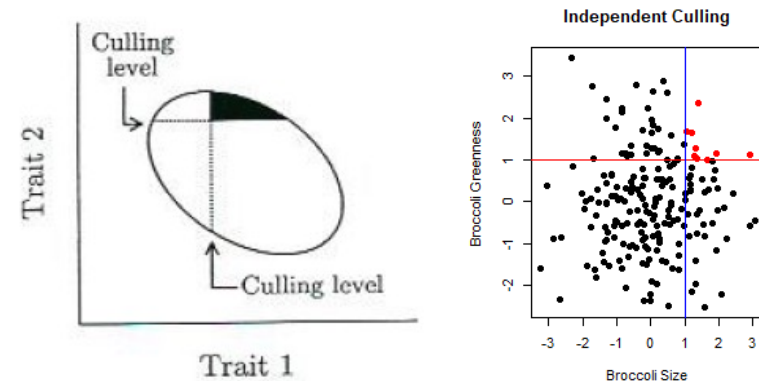$$GSi = 1.76 \; x \; -0.66 \; x \; \sqrt{0.83} \; x \; \sqrt{43978} = -221$$

$$-0.66 \; x \; \sqrt{0.83} > \sqrt{0.88}$$

$$-0.60 > 0.94$$

Multi-trait selection:
Tandem and Independent culling levels
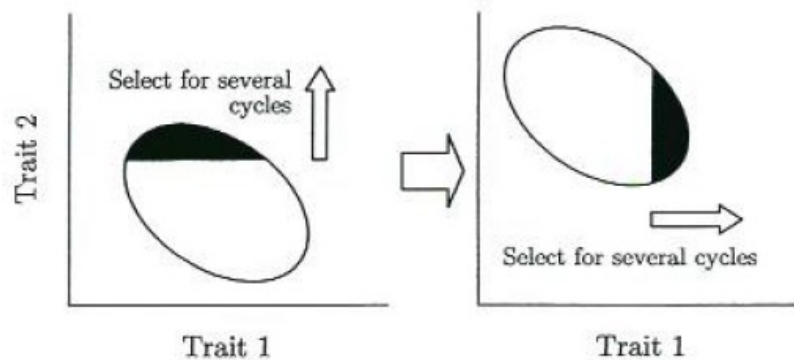
- In the first selection cycle for a trait
- The following cycle is based on another trait
- Sequential selection
- Works well for **independent** traits or positive correlates traits

- Set the minimum performance for each trait
- Selection cannot be too intense for one of the traits
- It tends to be more efficient than a Tandem
- **Easy up to two traits - two-dimensional**

What are selection indices?

- **Genotypic aggregate (H)**
- **Genotypic values weighted by economic values**

$$H = a_1 g_1 + a_2 g_2 + ... + a_n g_n$$

- **Selection index (I) - linear combination of traits**

$$I = b_1 X_1 + b_2 X_2 + ... + b_n X_n$$

- **The weighting coefficients (b's) maximize the correlation between the index and the genotypic aggregate or the desired genetic gain**

Economic weights

- It's challenging to determine the economic values of trait genetic gain

- Let's assume that highly skilled economists have established the economic values (**$ per unit increased**) of some traits

- Making a base index is simple. For each individual. each trait value is multiplied by the weight

- Then the trait values are summed. Selection on the index maximizes profit.

Trait	Example Trait Weight	Example Ind.1 Value	Example Ind.2 Value
Height (cm)	$1 / cm	180	160
Grain Yield (g)	$5 / g	60	40
Disease Resistance (points)	$0.50 / point	1020	1000
Beta Carotene Content (g)	$10 / g	0.002	0.007
Net Merit		(180*1) + (60*5) + (1020*0.50) + (0.002*10) = **990**	(160*1) + (40*5) + (1000*0.50) + (0.007*10) = **860**

Why are selection indices useful?

- Allow making multi-trait decisions **faster** and more **systematically**
- Translates a mental index into data. Helps make complex decisions that are **hard to visualize** at once
- Can prevent "**selecting in a circle**" on negatively correlated traits
- It can help **control trade-offs** among traits if they exist
- Weights are easier to set correctly than optimal levels of independent culling
- Behave as one big trait as long as all traits included are quantitative
- It's not recommended to put non-quantitative traits into a quantitative index
- It's not recommended to put more than four traits into an index

Smith-Hazel index

- **The goal is to adjust economic weights to account for trait heritability and genetic correlations**

- **The adjusted weights are used instead of the raw economic weights**

- **Putting weight on a high heritability trait than a low heritability trait, we can indirectly select the low heritability trait if they are positively correlated**

- **Equivalent to a multivariate BLUP when single phenotypic records are used**

Height (cm)	560	359	321	−0.01		400	389	319	−0.01	1	2.4
Grain Yield (g)	359	1233	115	−0.10		389	900	288	−0.10	5	2.8
Disease Resistance (score)	321	115	22709	−0.24	*	319	288	16000	−0.25	0.5	0.32
Beta Carotene Content (g)	−0.01	−0.10	−0.24	0.00		−0.01	−0.10	−0.25	0.00003	10	−5954

P^{-1}, inverse of phenotypic variance-covariance	*	G, genetic variance-covariance	*	w, economic weights	=	b, Smith-Hazel weights

$$P^{-1}Gw = b$$

Desired Gains Index – Pasek-Baker

- **Another framework to set weights based on desired gains per trait**

- **Desired gain weights also do well for traits with non-linear weights**

- **Economic weights do correspond to an amount of desired gain per trait**

- **In a base index. economic weights also assume how much we value gain per trait is not influenced by the genetic difficulty (heritability. correlations) of obtaining that gain**

- **Whereas desired gains can consider this influence (as would economic weights in a Smith-Hazel index)**

- **The desired gains index uses the inverse of genetic covariance (G^-) and our desired gains (d) in genetic standard deviations as input**

Desired Gains Index – Pasek-Baker

- Imagine that we want make gains of **one genetic standard deviation** in each of four traits.

- How can we set weights to target **one genetic standard deviation** of desired gain per trait?

- $b = G^{-1} d$
  to target desired gains. d. in genetic standard deviations

- given the inverse of the genetic variance-covariances in our set. **G**$^{-1}$. we can set weights b
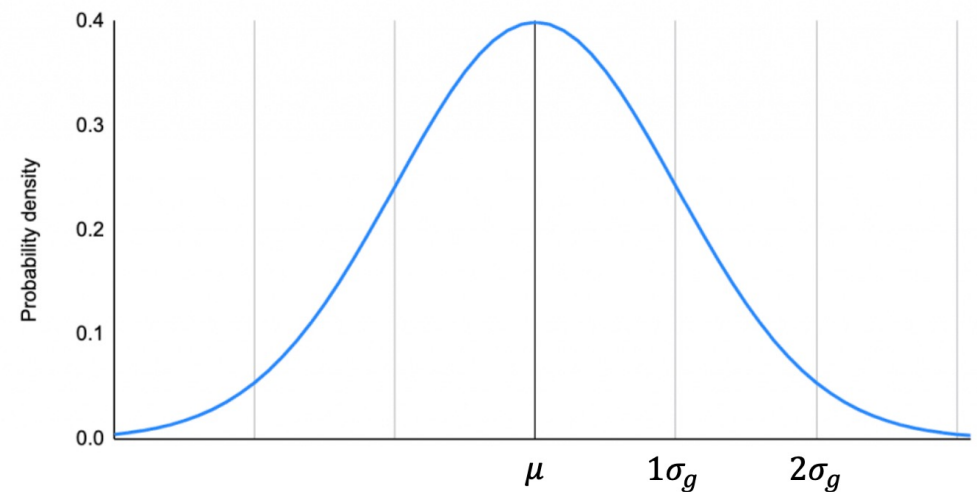

$$\begin{bmatrix} 1 \\ 5 \\ 100 \\ 30 \end{bmatrix} \quad \begin{bmatrix} 1.70 \\ 6.58 \\ 107.07 \\ 33.87 \end{bmatrix} \quad \begin{bmatrix} 2.4 \\ 8.16 \\ 114.14 \\ 37.74 \end{bmatrix}$$
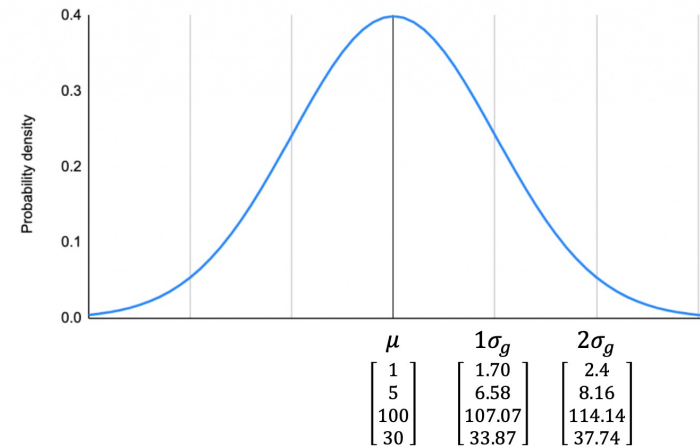
	Genetic Mean (μ)	Genetic Variance (σ_g^2)	Genetic Standard Deviation (σ_g)	Heritability
Trait A	1	0.5	0.70	1
Trait B	5	2.5	1.58	1
Trait C	100	50	7.07	1
Trait D	30	15	3.87	1

Trait A						
Trait A	$\begin{bmatrix} 1.40 \\ 0.63 \\ 0.14 \\ 0.26 \end{bmatrix}$	$=$	$\begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 2.5 & 0 & 0 \\ 0 & 0 & 50 & 0 \\ 0 & 0 & 0 & 15 \end{bmatrix}^{-}$		$\begin{bmatrix} 0.70 \\ 1.58 \\ 7.07 \\ 3.87 \end{bmatrix}$

weights (*b*)   inverse of genetic variance-covariances, **G**$^{-1}$ desired gains (*d*)

Scaled Desired Gains Index – Pasek-Baker

- How can we set weights to target **one genetic standard deviation** of desired gain per scaled trait?

- Traits are scaled first

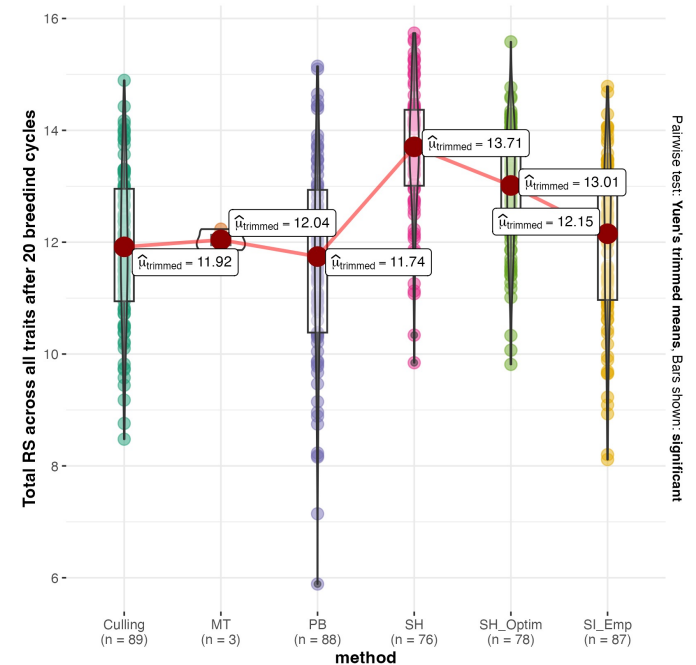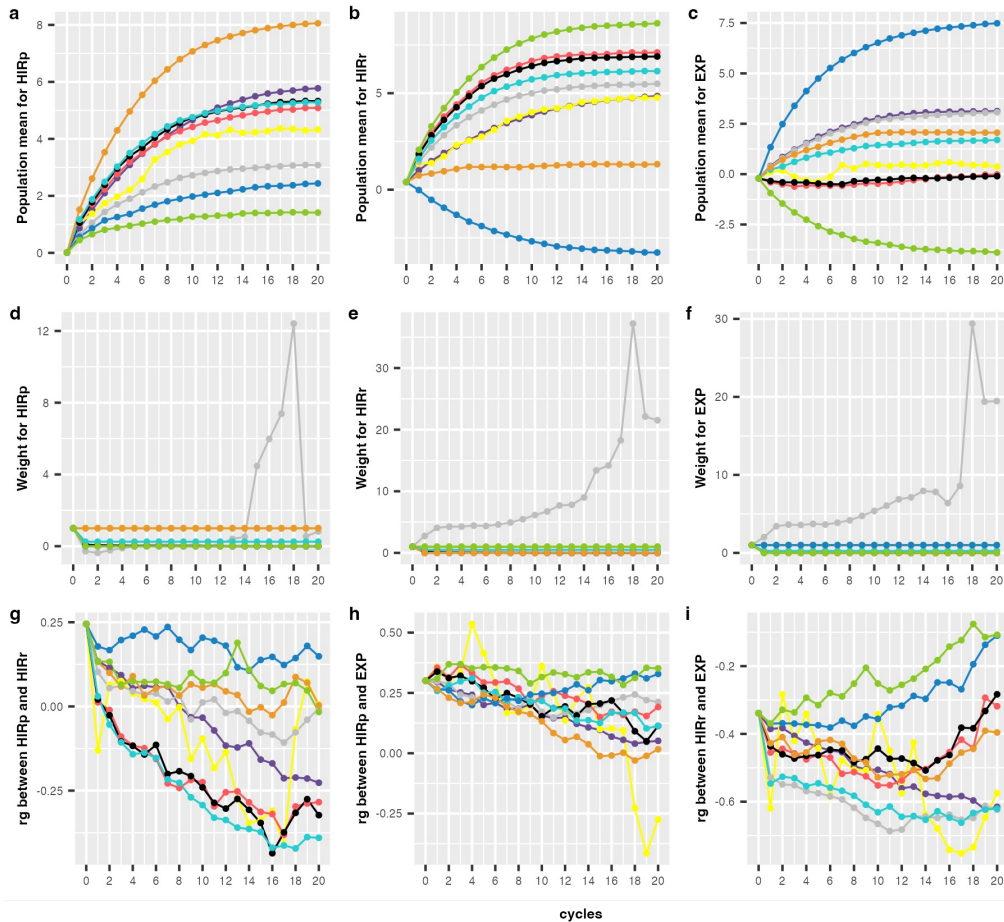- Subtract the mean from each observation and divide by the standard deviation

- $b = G^{-1}\,d$

Long-term

- **Weights must be re-estimated every cycle**
- **Do not attempt to interpret weights**. When taking into account genetic (co)variance and heritability. their meaning is difficult to interpret
- Their purpose is to **drive selection** in the direction set by the breeder.