

Some helpful notes for plant breeding related topics

From Lynch and Walsh 1998

Difference between IBD and IBS

- all measures of relatedness are based upon the concept of **identity by descent (IBD)**. Genes that are identical by descent are direct descendants of a specific gene carried in some ancestral individual.
- Two genes that have identical nucleotide sequences but have descended from different copies in the reference population are identical by state but not by descent.
- Genes that are identical by descent are necessarily identical by state, barring mutation
- Individuals that contain pairs of alleles that are IBD are said to be **inbred**.

Intro to Matrix Algebra and Linear Models

Standard regression model

- Depending on the causal connections between two variables, x and y , their true relationship may be linear or non-linear. However, regardless of the true pattern of association, a linear model can always serve as a first approximation

$$y = \alpha + \beta x + e$$

Where:

α is the y-intercept

β is the slope of the line (or the regression coefficient)

e is the residual error

Letting:

$$\hat{y} = \alpha + \beta x$$

- be the value of y predicted by the model, then the residual error is the deviation between the observed and predicted y value, i.e., $e = y - \hat{y}$.
- When info on x is used to predict y , x is referred to as the predictor or independent variable and y as the response or dependent.
- The objective of linear regression analysis is to estimate the model parameters, α and β , that give the "best-fit" for the joint distribution of x and y . The true parameters α and β are only obtainable if the entire population is sampled

Standard multiple regression formula

$$y = \alpha + \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_n z_n + e$$

Where:

y is the response variable

z_i is the predictor (or explanatory) variable used to predict the value of the response variable

y, z_1, \dots, z_n are observed measures

$\alpha, \beta_1, \dots, \beta_n$ are constants to be measured

e is the residual error --> is the deviation between the observed and fitted value of y

- Similar to a simple linear regression ($y = \alpha + \beta x + e$), except that y is now a function of n predictor variables, rather than one
- No assumption on the relationship between y and z_1, \dots, z_n , it simply gives the best linear approximation between observations

What's bookmarked?

- pg 49, additive genetic variance (ch. covariance, regression, and correlation)
- pg 444 detected marker effects (mapping QTLs: inbred line crossing)
- pg 454 marker approximations, likelihood ratio (mapping QTLs)
- pg 784 maximum likelihood estimates of variance components in the general mixed model (variance component estimation)