Multivariate General Linear Models

Lecture 08.1: Multivariate Overview

Lauren Sullivan

Module: Multivariate Models

Readings

Required for class:

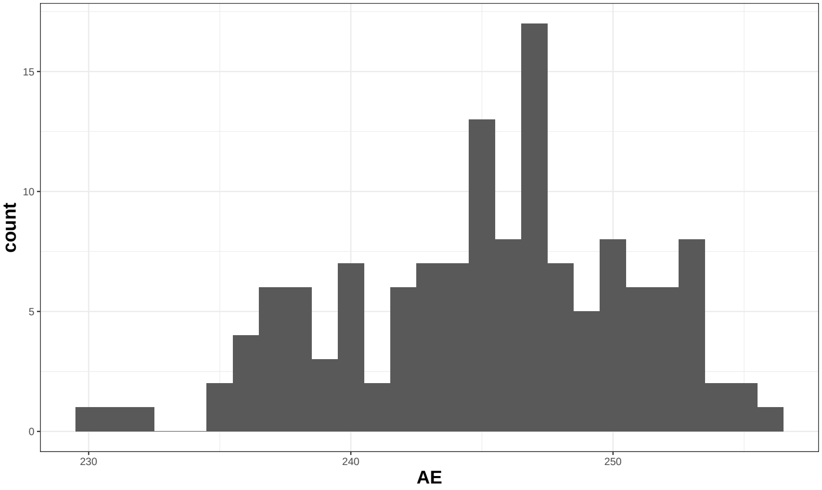
- NA

Optional:

- Healy. (1969) 259 Note: Rao's Paradox Concerning Multivariate Tests of Significance. *Biometrics*.

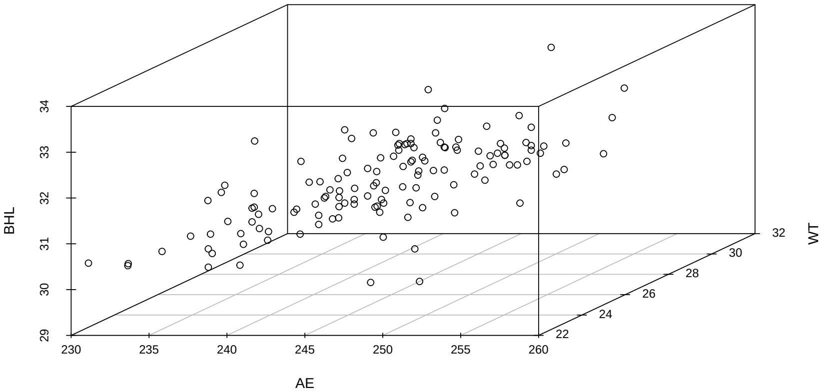
Univariate Analysis

When analyzing univariate data (what we have been doing up till now), we have been learning how to explain the variation in a single dependent variable (Y).



Multivariate Analysis

When analyzing multivariate data, we will attempt to explain variation in more than one dependent variable (multiple Y's).



Why Multivariate?

Multivariate analysis allows you to look at a more complex description of patterns in biology. In many cases, separate univariate analyses miss a covariation signal in the data, so it's important to think about data more fully if possible.

There are several ways to look at multivariate patterns from a matrix of $\mathbf{Y}$'s.

1. Linear models: MANOVA/regression to test patterns
2. Ordination: PCA, nMDS, etc to visualize patterns
3. Permutation tests: PERMANOVA to test patterns

Rao's Paradox and Data Collection

Increasing the number of dimensions of your data (# of Y's) means more information about a process, but for a given number of observations (rows of your dataset) the statistical power decreases. You need to watch out to make sure you don't have too few n's for the number of dependent variables (Y's).

There are a lot of suggestions for the number of observations you need. Make sure you think about this consciously!

- $n = 2 * (\#Y)$
- $n = 4 * (\#Y)$
- $n = (\#Y)^2$
- $n_{group} = 2 * (\#Y)$
- $n_{group} = 4 * (\#Y)$

Multivariate Data

Multivariate data is can be described as:

$$\mathbf{Y} \sim \mathbf{MVN}(\mu, \mathbf{\Sigma})$$

With the Variance-Covariance (VCV) matrix that is described as:

$$\mathbf{\Sigma} = \begin{bmatrix} s_{11} & & \\ s_{21} & s_{22} & \\ s_{31} & s_{32} & s_{33} \end{bmatrix}$$

Correlation Matrix

The VCV can be standardized to be the correlation matrix (last time was ρ but I've replaced with r's). First you pull the σ_i^2 out of the matrix.

$$\boldsymbol{\Sigma} = \sigma_{\mathbf{i}}^{\mathbf{2}} \begin{bmatrix} 1 & & \\ r_{21} & 1 & \\ r_{31} & r_{32} & 1 \end{bmatrix}$$

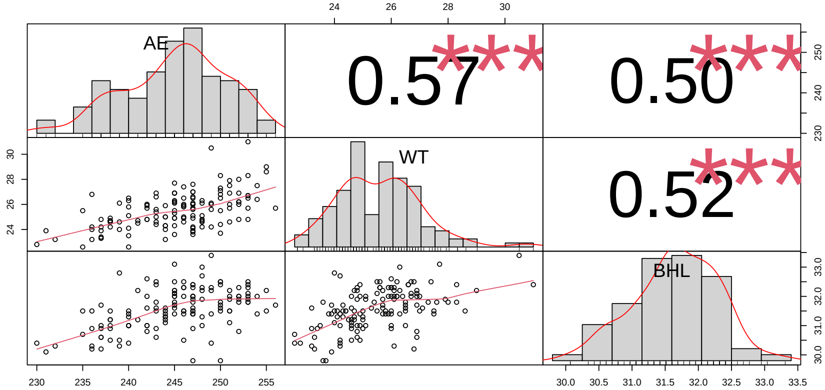
Then you can simply look at the correlation matrix (get rid of σ_i^2, and $\boldsymbol{\Sigma}$ simplifies to $\mathbf{R}$).

$$\mathbf{R} = \begin{bmatrix} 1 & & \\ r_{21} & 1 & \\ r_{31} & r_{32} & 1 \end{bmatrix}$$

Plotting the Correlation Matrix

The correlation matrix can be visualized by plotting the correlation between variables.

```
chart.Correlation(sparrow[,5:7], histogram = TRUE, pch = 19)
```



Another option in tidyverse

The `GGally` library and `ggscatmat()` also makes these figures

```
ggscatmat(sparrow, columns = 5:7, alpha = 1)+
  theme_bw()+
  theme(axis.title = element_text(face="bold", size=16))
```