STA461: Introduction to ANOVA

Lynn Lin

August 31, 2018

Completely randomized designs

- A design is the rule that determines the assignment of the experimental units to treatments
- A Completely Randomized Design (CRD) is a design in which a fixed number (ν) of treatments are assigned to the experiemtnal units completely at random, subject only to the number of observations to be taken on each treatment
- CRD is the simplest possible design

Model for a CRD

Example: Fertilizer

- Setup: Goal is to compare the effect of 3 kinds of fertilizer: F_1, F_2, and F_3 on the height of plants in a greenhouse
- We have 24 plots for the experiment, and randomly assign 6 plots to F_1, F_2, F_3 and no fertilizer, respectively. I.e., $r_1 = r_2 = r_3 = r_0 = 6$
- The plots with no fertilizer are called the *control group*
- The effects model is

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, i = 0, 1, 2, 3; t = 1, ..., 6$$

$$\epsilon_{it} \sim N(0, \sigma^2), iid$$

- $\mu = 24$ cm; $\sigma^2 = 0.5$ cm; $\tau_0 = -3$ cm; $\tau_1 = 4$ cm; $\tau_2 = 1.5$ cm; $\tau_3 = 3$ cm;

Example: Fertilizer

- Q1: What is the distribution of Y_{11}: the height of the first plant given F_1?
- Q2: What is the distribution of Y_{01}
- Q3: What is the expected difference in height between a plant getting F_3 and a plant in the control group?

Estimation of Parameters

Common approach in data analysis

1. Collect data $\{y_{it}\}$
2. Specify a statistical model for the data: $y_{it} \sim f_y(y|\theta)$
3. Estimate model parameters θ using data $\{y_{it}\}$: Parameter estimates are functions of the data, e.g., $\hat{\sigma}^2 = g(\{y_{it}\})$
4. Conduct hypothesis tests on the estimates, e.g.,

$$H_0 : \tau_1 = \tau_2 \text{ vs. } H_A : \tau_1 \neq \tau_2$$

5. Interpret the results of 3 and 4

Estimation of parameters

- Sometimes we are interested in the estimates of parameters by themselves, e.g., $\hat{\sigma}^2$ tells us something about the magnitude of the random variability around the treatment
- Often we are interested not in a parameter itself, but rather in a function of parameters, e.g., $\hat{\mu} + \hat{\tau}_i$ is the mean response of the ith treatment
- It is possible to obtain many different estimators of a parameter or a function of parameters
- The field of statistics has spent considerable time thinking about "good" approaches to estimating parameters. We often compare estimators by considering their properties as random variables

Example

- Let $Y_t = \mu + \epsilon_t$, $t = 1, ..., 10$, $\epsilon_t \sim N(0, 1)$, iid
- Estimate μ, i.e., we need to find $\hat{\mu}$ which is a function of $Y_1, \ldots, Y_{10}$

Estimation of parameters

- Estimability
- Least squares estimation