Chapter 5: One-way ANOVA and Randomized Experiments - Part 1

Harsha Perera

Key Topics

- Overview of ANOVA
- The One-Way Randomized Experiment
- Fitting One-Way ANOVA Model

Overview of ANOVA

- Simple Linear Regression and Multiple Linear Regression:
  - Response variable: continuous
  - Predictor variable(s): at least one continuous variable
- In this chapter we will look into models with
  - Continuous response
  - Categorical predictor(s)
- Examples:
  - Weight gain among rats vs. source of protein in diet (beef, cereal or pork)
    - An experiment
  - Difference in average commute time among four cities (Boston, Houston, Minneapolis and Washington)
    - Observational study
- From an overall context we will be investigating whether there is a mean difference in response for each category

Example 5.1 - Fat Rats: A Randomized Experiment

- Objective: to study the effect of three different high-protein diets on weight gain in baby rats
- Response: weight gain (continuous / quantitative)
- Predictor: sources of protein in high-protein diets
  - beef
  - cereal
  - pork
- $n = 30$: 10 on each type of protein diet

Fat rats: a randomized experiment. . .

- Is there a difference in weight gain depending on the different sources of diet?
- If different, how big is the difference?
- Are all groups different or subset of groups different?
- We can answer all theses question by performing an ANOVA

Example 5.2 - Teen Pregnancy and the Civil War

- Response: Pregnancy rate (number of pregnancies per 1000 teenage girls) for each of 50 U.S. states in 2010
- Predictor: role of the state in the U.S. civil war
  - C: confederate state
  - B: border state
  - U: union state
  - O: other state
- Objective: investigate whether teen pregnancy differ depending on the role of the state during civil war (1860)

One-way Randomized Experiment

- Simplest experimental structure for an ANOVA model
- Has a single categorical predictor, and we create the data for this choice of model by randomly assigning units to the categories of that variable
- One-way ANOVA is a regression model when we have one categorical predictor and continuous response

One-way Randomized Experiment - Terminology

- Some terminology in Experimental Designs (example: Fat rats)
  - EXPERIMENTAL UNIT: each rat who get randomly assigned to one of the protein sources
  - EXPLANATORY FACTOR(S): one factor - source of protein
  - LEVELS/TREATMENTS: different levels of the factor (beef, cereal, pork)
  - BALANCED DESIGN: when each level/treatment gets assigned to same number of experimental units (10 rats per level)
  - RESPONSE VARIABLE: measurement of each experimental unit after the treatment applied (weight gain of rats)

- One-way ANOVA can also be used with observation studies as long as the observations are gathered randomly

Fitting One-Way ANOVA model

- For quantitative response Y and single categorical predictor,

$$Y = \mu_i + \in$$

or

$$Y = \mu + \alpha_i + \in$$

- μ: Grand mean
- α_i: treatment effect for i^{th} group mean
- Random error: $\in \sim N(0, \sigma_\in)$
- Mean for the i^{th} group: $\mu_i = \mu + \alpha_i$

Fitting One-Way ANOVA Model

- The goal of ANOVA is to compare the averages of each category

Response = Grand Average + Treatment effect + Residual

- Objective is to answer following questions

1. What is the overall average: Grand mean ($\hat{\mu} = \bar{y}$)
2. How far is each average from overall average: treatment effect ($\hat{\alpha}_i = \bar{y}_i - \bar{y}$)
3. How far is each response from its group average: residual ($y_i - \bar{y}_i$)

ANOVA Table

- For k groups

Source	df	SS	MS	F
Groups	k-1	SSGroups	$\text{MSGroups} = \frac{SSGroups}{k-1}$	$\text{F} = \frac{MSGroups}{MSE}$
Residual	n-k	SSE	$\text{MSE} = \frac{SSE}{n-k}$	
Total	n-1	SSTotal		

Source	df	SS	MS	F
Groups	k-1	$\sum(\bar{y}_i - \bar{y})^2$	$\text{MSGroups} = \frac{SSGroups}{k-1}$	$\text{F} = \frac{MSGroups}{MSE}$
Residual	n-k	$\sum(y_i - \bar{y}_i)^2$	$\text{MSE} = \frac{SSE}{n-k}$	
Total	n-1	$\sum(y_i - \bar{y})^2$		