One-Way ANOVA

READING:	5.1 – 5.4
EXERCISES:	CH 5. 27, 37-38, 40
ASSIGNED:	HW 10
PRODUCER:	DR. MARIO

IMG CREDIT: ALEX RIEGERT-WATERS

ANOVA

- Usage of ANOVA
  - Numerical Response Variable (Y)
  - Categorical Predictor Variable (X)

- Questions to Answer:
  - *Do different groups have different means?*
  - *How different are the means across groups?*

- One-Way ANOVA Implies There is One Predictor Variable

ANOVA

- Data: Samples from Different Groups (*K = # of Groups)*

- Summary Statistics

$Group$	n_i	$\bar{y}_i$	s_i
1	n_1	$\bar{y}_1$	s_1
2	n_2	$\bar{y}_2$	s_2
…			
K	n_K	$\bar{y}_K$	s_K
Overall	n	$\bar{y}$	s

- Test Hypotheses

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_K \text{ vs. } H_a: Some\ \mu_i \neq \mu_j$$

Example: Grades on Different Exams

- There are Four Different Exams (1, 2, 3, 4)
- There are Five Students (Barb, Betsy, Bill, Bob, Bud)
- Each Student Takes All Four Exams
- Question: *Is there a significant difference in the average grade among the four different exams?*

Supplement for Lecture 22

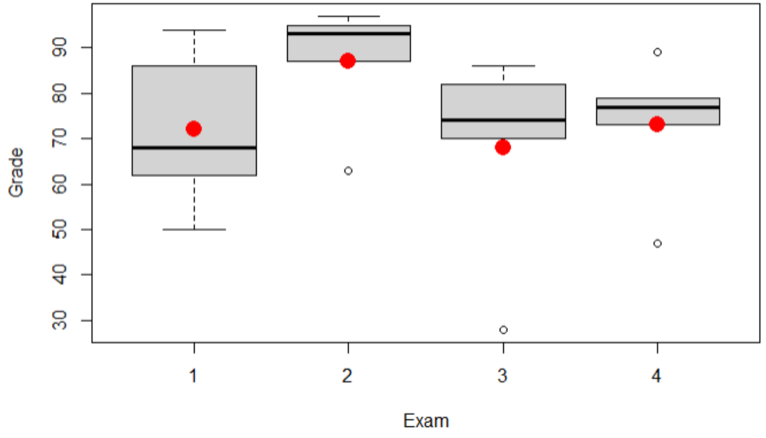
- Inspect the Data

- Calculate Various Summary Statistics
  - Use ***tapply()*** function

- Visual of Relationship
  - Boxplot of Grades by Group
  - Plot Mean Grade by Group

- *Differences due to random chance?*

$Exam$	n_i	$\bar{y}_i$	s_i
1	5	72	17.89
2	5	87	13.93
3	5	68	23.24
4	5	73	15.68
Overall	20	75	18.11



ANOVA

- ANOVA (Means) Model

$$Y = \mu_i + \epsilon$$

Mean for
Group #i

$\epsilon \sim N(0, \sigma_\epsilon)$
Independent and Identically Distributed

- Under $H_0: All\ \mu_i\ Are\ Equal \rightarrow \hat{\mu}_i = \bar{y}$ (overall mean)

- Under $H_a: Some\ \mu_i \neq \mu_j \rightarrow \hat{\mu}_i = \bar{y}_i$ (group mean)

ANOVA

- Making Prediction $\hat{Y}$ under H_0

  $\hat{y} = \bar{y}$ for all groups $\rightarrow residual = y - \bar{y}$

- Making Prediction $\hat{Y}$ under H_0

  $\hat{y} = \bar{y}_i$ for ith group $\rightarrow residual = y - \bar{y}_i$

- Question: *Do we do "significantly" better with separate means?*

  $SSTotal = \sum(y - \bar{y})^2$ vs. $SSE = \sum(y - \bar{y}_i)^2$

ANOVA

- Partitioning Variability

$$SSTotal = SSGroups + \text{SSE}$$

$$\sum(y-\bar{y})^2 = \sum(\bar{y}_i-\bar{y})^2 + \sum(y-\bar{y}_i)^2$$

- ANOVA Table

Source	d.f.	Sum of Squares	Mean Square	F	P-value
Groups	$K-1$	$SSGroups$	$\frac{SSGroups}{K-1}$	$\frac{MSGroups}{MSE}$	$F_{K-1,n-K}$
Residual	$n-K$	SSE	$\frac{SSE}{\text{n}-\text{K}}$		
Total	$n-1$	$SSTotal$			

ANOVA

- ANOVA F-Test
  - $H_0: \mu_1 = \mu_2 = \cdots = \mu_K$
  - H_a: $\mu_i \neq \mu_j$ for some $i \neq j$
  - Small P-value -> Reject H_0 Accept H_a -> Significant Difference Among the Means of K Groups

- Question: *What groups have significantly different means?*

Supplement for Lecture 22

- Visual of Relationship
- Calculate SSTotal, SSE, SSGroups
- ANOVA F-test Using ***aov()*** *function*
- Interpret Results (P-value = 0.395)

Alternate ANOVA

- ANOVA (Effects) Model

$$Y = \mu + \alpha_i + \epsilon$$

Overall Mean

Effect for Group #i

$\epsilon \sim N(0, \sigma_\epsilon)$
Independent and Identically Distributed

- Mean for Group i

$$\mu_i = \mu + \alpha_i$$

- Estimation of Parameters
  - $\hat{\mu} = \bar{y}$
  - $\hat{\alpha}_i = \bar{y}_i - \bar{y}$

Alternate ANOVA

- Before
  - $H_0: \mu_1 = \mu_2 = \cdots = \mu_K$
  - H_a: $\mu_i \neq \mu_j$ for some $i \neq j$
- Now
  - $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$
  - H_a: Some $\alpha_i \neq 0$
- These Hypotheses are Equivalent

Example: Grades on Different Exams

- Recall Data

Exam	n_i	$\bar{y}_i$	s_i
1	5	72	17.89
2	5	87	13.93
3	5	68	23.24
4	5	73	15.68
Overall	20	75	18.11

- Estimate of Grand Mean
  - $\hat{\mu} = \bar{y} = 75$

- Estimates of Effects
  - $\hat{\alpha}_1 = \overline{y_1} - \bar{y} = 72 - 75 = -3$
  - $\hat{\alpha}_2 = \overline{y_2} - \bar{y} = 87 - 75 = 12$
  - $\hat{\alpha}_3 = \overline{y_3} - \bar{y} = 68 - 75 = -7$
  - $\hat{\alpha}_4 = \overline{y_4} - \bar{y} = 73 - 75 = -2$

$\sum \hat{\alpha}_i = 0$

Supplement for Lecture 22

- Plots of Residuals from ANOVA

- Examine Standard Deviations of the Different Exams

Conditions for ANOVA

- **Recall:** $$\epsilon \overset{iid}{\sim} N(0, \sigma_\epsilon)$$
- **Zero Mean:** Always Holds for Sample Residuals
- **Constant Variance:** Compare Standard Deviations of Groups
- **Normality:** Histogram or QQ Plot of Residuals
- **Independence & Randomness:** Details About the Data Collection

Thank You

Make Reasonable Decisions