13 Lecture Notes 14: One-Way Analysis of Variance (ANOVA)

"You know it's real when you are who you think you are." - Aubrey Drake Graham

13.1 Theory Behind One-Way (ANOVA)

The purpose of ANOVA is to determine if the means are the same between more than 2 groups[1]. For example:

- GPA of Engineering Majors, Biology Majors and English Majors
- 4 different strengths of Lipitor
- Different types of trees

ANOVA is a generalization of the Two Independent Sample t test,the only difference is that we have more than 2 groups.

The way we determine if the different strengths of Lipitor reduces cholesterol are the same, for example, is to analyze two types of variances:

- Variance *between* the groups
- Variance *within* the group

At the end of the day, we want to compare the variability between the groups to the variability within groups.


---
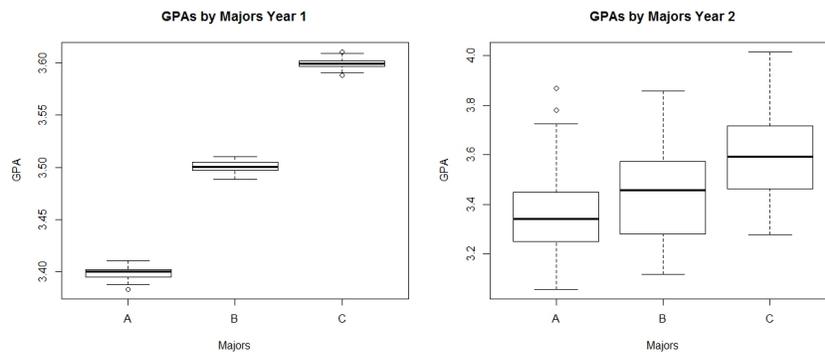
[1]In the book factor/treatment is used for groups.

Assumptions:

1. The groups are independent simple random samples

2. The populations are normally distributed

3. The population variances are equal $\sigma_1 = \sigma_2 = \sigma_3 = \cdots = \sigma_k$

Example 1: Process for Hypothesis Testing for this class (example): Head injury data from crash test dummies used in the 3 types of cars (Model Q, Model F, P100I) has been collected. There is 40 of each type of car in study. The measurement of head injury are in hic, which denotes a standard head injury criterion. Use a 0.05 significance level to test the null hypothesis that the types of cars, have the same mean.

1. Identify and State the Statistical Question: **Is there a statistical difference between types of cars.**
   - Determine the variable(s) of interest: **quantitative**
   - Identify and state the hypotheses (Null and Alternative Hypotheses) based on the question at hand: $H_0 : \mu_1 = \mu_2 = \mu_3$ and H_1 **: at least 1 mean is different from the others**

2. Identify and state level of significance α (the probability of rejecting the H_0 when H_0 is true) $\boldsymbol{\alpha = 0.05}$

F distribution
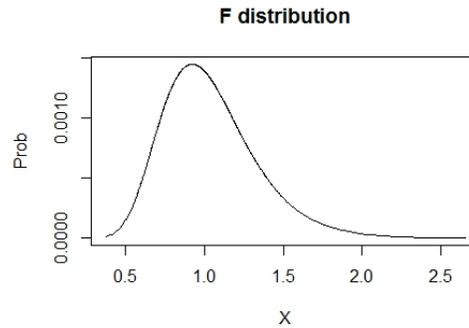
Draw and state the rejection region.

(a) J is the number of groups
(b) n_1 is the number of Model Q cars
(c) n_2 is the number of Model F cars
(d) n_3 is the number of P100I cars
(e) n is the sample size

(a) $J = 3$ number of groups
(b) $n_1 = 40$ Model Q cars
(c) $n_2 = 40$ Model F cars
(d) $n_3 = 40$ P100I cars
(e) $n = 120$ cars

Really IMPORTANT:
   - Degrees of Freedom: $\boldsymbol{df_{num} = J - 1 = 2, df_{den} = n - J = 127}$:
   - α: $\boldsymbol{\alpha = 0.05}$
   - Critical Value $= 3.0738$ (F table, Appendix A, given to you on test)

3. Perform Statistical Test and Interpret Results Fun begins:

Type of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Test Statistic (TS)
Between	SSB	$df_b = J - 1$	MSB	TS
Within	SSW	$df_w = n - J$	MSW	
Total	SST	$df_t = n - 1$		

F distribution
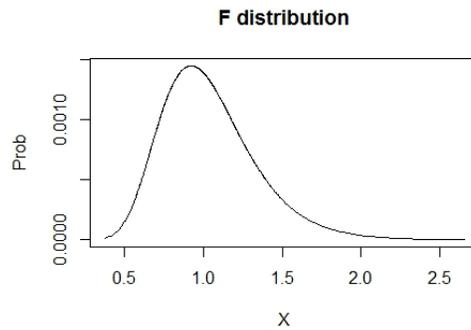
Sums of Squares Between:
$$SSB = \text{Nope} \tag{43}$$

Sums of Squares Within:
$$SSW = \text{Nope} \tag{44}$$

Sums of Squares Total:
$$SST = SSB + SSW \tag{45}$$

Mean Square Between:
$$MSB = \frac{SSB}{df_b} \tag{46}$$

Mean Squares Within:
$$MSW = \frac{SSW}{df_w} \tag{47}$$

```
            Df Sum Sq Mean Sq F value Pr(>F)
   Group      2 0.8052  0.4026   14894 <2e-16 ***
   Residuals 117 0.0032  0.0000
   ---
   Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Test Statistic: **14894**
- p-value: $\mathbf{< 2e - 16 ***}$ (You would not have to do this)

4. State the sample, null hypothesis, test that was used, and conclusion with non-statistical terms
   - $n = 120$ cars
   - All the types cars hic have the same means
   - Analysis of Variance
   - Reject the claim that All the types cars hic have the same means

What is this test not telling us?

Example 2: There is data from the amounts of nicotine (mg per cigarette) in the small size cigarettes (42 cigarettes sampled), king size cigarettes (25 cigarettes sampled), the 100 mm menthol cigarettes (52 cigarettes sampled), and the 100 mm nonmenthol cigarettes (37 cigarettes sampled). The king size cigarettes are nonfiltered, nonmenthol, and non-light. Use a 0.05 significance level to test the claim that the three categories of cigarettes yield the same mean amount of nicotine.

1. Identify and State the Statistical Question:
   - Determine the variable(s) of interest:
   - Identify and state the hypotheses (Null and Alternative Hypotheses) based on the question at hand:

2. Identify and state level of significance α (the probability of rejecting the H_0 when H_0 is true):

F distribution
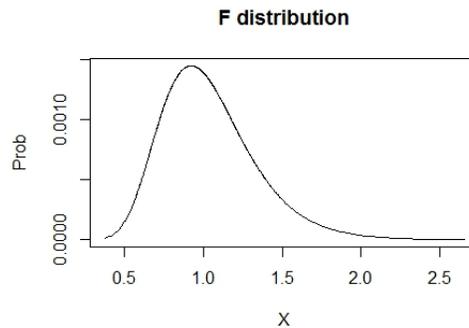
Draw and state the rejection region.

(a) J is the number of groups
(b) n_1 is the number of small size cigarettes
(c) n_2 is the number of king size cigarettes
(d) n_3 is the number of 100 mm menthol cigarettes
(e) n_4 is the number of 100 mm nonmenthol cigarettes
(f) n is the sample size

(a) $J =$ groups
(b) $n_1 =$ small size cigarettes
(c) $n_2 =$ king size cigarettes
(d) $n_3 =$ 100 mm menthol cigarettes
(e) $n_4 =$ 100 mm nonmenthol cigarettes
(f) $n =$ is the sample size

Really IMPORTANT:
   - Degrees of Freedom: $df_{num} = J - 1 =$, $df_{den} = n - J =$:
   - α: $\alpha =$
   - Critical Value = 2.6441 (F table, Appendix A, given to you on test)

124

3. Perform Statistical Test and Interpret Results

```
          Df Sum Sq Mean Sq F value    Pr(>F)
Group     A.    409    C.       E.    0.202
Residuals B.  13297    D.
```

Type of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Test Statistic (TS)
Between	$SSB =$	$df_b =$	$MSB =$	$TS =$
Within	$SSW =$	$df_w =$	$MSW =$	
Total	$SST =$	$df_t =$		

- Test Statistic:
- p-value:

4. State the sample, null hypothesis, test that was used, and conclusion with non-statistical terms
   -
   -
   -
   -