- 12-1 Video: https://youtu.be/FploaywVOQo
- 12-2 Video: https://youtu.be/mYsO84WSZcQ

12-1 ONE-WAY ANALYSIS OF VARIANCE

- Analysis of Variance (ANOVA): when an F test is used to test a hypothesis concerning the means of 3 or more populations.
- Assumptions:

1. The populations from which the samples were obtained must be normally or approximately normally distributed.
2. The samples must be independent of one another.
3. The variances of the populations must be equal.
4. The samples must be simple random samples, one from each of the populations.

- Even though we are comparing 3 or more means, variances are used in this F test instead of means.
- 2 different estimates of the population variance are made:

1. Between-group variance estimate: finds the variance of the means.
2. Within-group variance estimate: compute the variance using all the data. Not affected by differences in the means.

- If there is no difference in the means:
  - The between-group and within-group variance estimates will be approximately equal.
  - The F test value will be approximately 1.
  - The null hypothesis will not be rejected.
- If the means differ significantly:
  - The between-group variance will be much larger than the within-group variance.
  - The F test value will be significantly greater than 1,
  - The null hypothesis will be rejected.

Three random samples of times (in minutes) that commuters are stuck in traffic are shown. At the 0.05 level of significance, is there a difference in the mean times among the three cities?

Dallas	59	62	58	63	61
Boston	54	52	55	58	53
Detroit	53	56	54	49	52

1. State the hypotheses:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

H_1: At least one mean is different from the others.

2. Find the critical value:

- d.f.N. = k -1 (k is the number of groups) 3-1=2
- d.f.D. = N – k (N is the sum of the sample sizes)=15-3=12
- Always right tailed. C.V.=3.89
- Table H pgs. 792-796

3. Compute the test value:

a) Find the mean and variance of each sample

$\bar{X}_1 = 60.6$ $\bar{X}_2 = 54.4$ $\bar{X}_3 = 52.8$

$s_1^2 = 4.28$ $s_2^2 = 5.29$ $s_3^2 = 6.71$

b) Find the grand mean:

$$\bar{X}_{GM} = \frac{\sum x}{N} = \frac{839}{15} = 55.93$$

c) Find the between-group variance:

$$s_B^2 = \frac{\sum n_i\left(\bar{X}_i - \bar{X}_{GM}\right)^2}{k-1} = \frac{5(60.6-55.93)^2}{3-1} + \frac{5(54.4-55.93)^2}{3-1} + \frac{5(52.8-55.93)^2}{3-1}$$

$$= 84.87$$

- The numerator is called the sum of the squares between groups. (SS_B) 169.7335
- Also referred to as a mean square (MS_B)

d) Find the within-group variance:

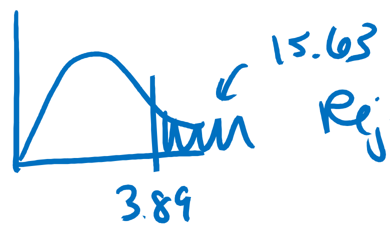
$$s_W^2 = \frac{\sum (n_i - 1)s_i^2}{\sum (n_i - 1)} = \frac{(5-1)(4.28) + (5-1)(5.29) + (5-1)(6.71)}{(5-1) + (5-1) + (5-1)} = 5.43$$

- The numerator is called the sum of squares within groups (sum of squares for the error) (SS_W) 65.12
- Also referred to as a mean square (MS_W)

e) Find the F test value:

$$F = \frac{s_B^2}{s_W^2} = \frac{84.87}{5.43} = 15.63$$

4.) Make the decision:



Reject H_0

5.) Summarize the results: There is enough evidence to support the claim that there is a difference in the mean times among the 3 cities.

- Analysis of Variance (ANOVA) Summary Table

Source	Sum of Squares	d.f.	Mean Square	F
Between	169.7335	K-1 2	84.87	15.63
Within (error)	65.12	N-K 12	5.43	
Total	234.8535	14		

12-2 THE SCHEFFÉ TEST AND THE TUKEY TEST

- If an ANOVA procedure is performed and the null hypothesis is rejected using the F test, this means that at least one of the means is different, but we do not know where the difference lies.

- The Scheffe test and Tukey test are used after an ANOVA procedure to find where the difference among the means is.

SCHEFFÉ TEST

- Compares 2 means at a time using all possible combinations.
- Formula:

$$F_s = \frac{\left(\overline{X}_i - \overline{X}_j\right)^2}{s_W^2\left[\left(1/n_i\right) + \left(1/n_j\right)\right]}$$

$$\overline{X}_i \ \& \ \overline{X}_j = mean \quad of \quad the \quad samples \quad being \quad compared$$

$$n_i \& n_j = respective \quad sample \quad sizes$$

$$s_W^2 = within - group \quad \text{var}\, iance$$

- To find the critical value F′ multiply the critical value for the F test by k-1: $F' = (k-1)(C.V.)$
- There is a significant difference between the 2 means being compared when F_s is greater than F′.

① $\bar{X}_1$ and $\bar{X}_2$

$$F_s = \frac{(60.6 - 54.4)^2}{5.43\left[\frac{1}{5} + \frac{1}{5}\right]} = 17.70$$

② $\bar{X}_1$ and $\bar{X}_3$

$$F_s = \frac{(60.6 - 52.8)^2}{5.43\left[\frac{1}{5} + \frac{1}{5}\right]} = 28.01$$

③ $\bar{X}_2$ and $\bar{X}_3$

$$F_s = \frac{(54.4 - 52.8)^2}{5.43\left[\frac{1}{5} + \frac{1}{5}\right]} = 1.18$$

$$F' = (K-1)(C.V.) = (3-1)(3.89) = 7.78$$

$F_s > F'$

Significant difference between $\bar{X}_1$ and $\bar{X}_2$, $\bar{X}_1$ and $\bar{X}_3$

THE TUKEY TEST

- Used to compare means after an ANOVA procedure has been performed, but sample sizes must be equal.
- Formula: $$q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{s_W^2 / n}}$$
- When the absolute value of q is greater than the critical value for the Tukey test, there is a significant difference between the 2 means being compared.
- To find the critical value use Table N in Appendix C. v= degrees of freedom for within-group variance.

① $\bar{x}_1$ and $\bar{x}_2$

$$q = \frac{(60.6 - 54.4)}{\sqrt{\frac{5.43}{5}}} = 5.95$$

② $\bar{x}_1$ and $\bar{x}_3$

$$q = \frac{(60.6 - 52.8)}{\sqrt{\frac{5.43}{5}}} = 7.48$$

③ $\bar{x}_2$ and $\bar{x}_3$

$$q = \frac{(54.4 - 52.8)}{\sqrt{\frac{5.43}{5}}} = 1.54$$

C.V. = 3.77

$\alpha = 0.05$

use Table N

$K = 3$

$v = N - K = 15 - 3 = 12$

$|q| > C.V$

There is a significant diffrence between $\bar{x}_1$ and $\bar{x}_2$, $\bar{x}_1$ and $\bar{x}_3$

12-1 AND 12-2 PRACTICE

- Pgs.656-658 #2, 7, 8, 10, 11, 14, 15
- Pgs. 664-665 #9-13all
- Technology Pg. 658