Math 244 Lecture Notes

Chapter 26 Day Two: Analysis of Variation (ANOVA)

Overview: Last class, we learned how to compare multiple groups. Today, we will look at terminology more critically and determine how to do ANOVA with technology

Here is the general process from start to finish:

STEP I: State the hypotheses. $H_0 : \mu_1 = ... = \mu_k$ and H_A : At least one average is different.
STEP II: When comparing multiple averages, we use the F-Distribution. Where...

$$\text{Variation Between Groups} = \Sigma \frac{n_i(\bar{x}_i - \bar{X})^2}{k-1}$$

with k the number of groups and $\bar{X}$ the overall average AND

$$\text{Variation Within Groups} = s^2_{pool} = \Sigma \frac{df_i(s_i)^2}{N-k}$$

with N the total number of values.

$$F = \frac{\text{Variation Between Groups}}{\text{Variation Within Groups}}$$

has df's of $k-1$ and $N-k$ in that order.

The assumptions are

(a) Random

(b) Independence/10% Rule

(c) Large n:

(d) Indepedent Groups

(e) Equal Spread:

STEP III: ANOVA is always a right-tailed test.
We find the P-value using "Fcdf(F-score, ∞, $df_{Between}$,df_{Within})"
STEP IV: Conclusion

Example 1. Perform ANOVA for the Excel file located on your desktop.

We've covered the neccessary math at this point, but let's improve our vocabulary.

Variation Between Groups in a medical setting might compare the average duration of various **treatments**.

$$\text{Variation Between Groups} = \frac{\Sigma n_i(\bar{x}_i - \bar{X})^2}{k-1} = \frac{\text{Sum of the Squares of the Treatment}}{df}$$

By dividing the Sum of the Squares of the Treatment (SST), we are finding the Mean for the Squares of the Treatment (MST).

$$\text{Variation Between Groups} = \frac{\text{SST}}{df} = \text{MST}$$

Variation Within Groups is based solely off the Standard **Errors**.

$$\text{Variation Within Groups} = s^2_{pool} = \Sigma\frac{df_i(s_i)^2}{N-k} = \frac{\text{Sum of the Squared standard Errors}}{df}$$

By dividing the Sum of the Squared standard Errors (SSE), we are finding the Mean for the Squares of the Errors (MSE). This is our **pooled** variance.

$$\text{Variation Between Groups} = s^2_{pool} = \frac{\text{SSE}}{df} = \text{MSE}$$

Using the terminology above, we rewrite to get that

$$F = \frac{\text{Variation Between Groups}}{\text{Variation Within Groups}} = \frac{\text{MST}}{\text{MSE}}$$

OBSERVATION: You may have noticed this already, but

$$df_{top} + df_{bottom} = (k-1) + (N-k) = N-1.$$

This will come in handy for the next exercise.

Example 2. Fill in the following ANOVA table:

Source	df	SS	MS	F-Stat	P-Value
Treatment	2	28.22			
Error	24	28.44			
Total	26	56.66			

Verifying our answers in the calculator takes two processes. We first must type in our lists. Then, we must perform the test.

Creating lists: If you have a TI-83, TI-84, hit [Stat]→[Edit] your lists. Type in your dat.

If you have a TI-89, go into [Apps]→[Stat/List Editor]. Then, type in your data.

Checking H-Test with Calculator: In our calculator, we use "ANOVA" (1-way not 2-way) to check our hypothesis test. This is found in either [Stat]→[Tests] on the TI-83/84 OR [Stat/List]→[F6:Tests] on the TI-89. Type in your relevant information and you are good to go!

It will look like "ANOVA(L_1, L_2, L_3)" on the TI-83/84. To get L_1, hit [2nd]→[1]. To get L_2, hit [2nd]→[2] and so on.
In the TI-89, you will have to type in the complete name of the lists. For example "list1".

Example 3. Use your calculator to collect the following information for the warm-up data on tip amounts.

- MST?
- MSE?
- F?
- P-value?
- Pooled Standard Error? s_p?

What do we do if we do notice a difference in averages? Follow it up with 2-sample t-tests. The downside? Error adds up really quickly.

Example 4. If we want to compare three groups, how many pairs will we need to analyze?

Example 5. If we want the total α for all tests to be 0.05, what would each individual α^* need to be?

Bonferroni Correction. Let

$$\alpha^* =$$

when comparing multiple pairs. This will give you a corrected $C^* = 1 - \alpha*$ if you want to create confidence intervals as a follow up.

Example 6. Compare each pair of individuals for the tip data. Which pairs, if any, indicate a difference in average tip amount?

Options, Options, Options: We now have multiple ways of doing problems involving averages.

Example 7. You want to compare two routes to work. Route 1 takes 30 minutes on average with a standard deviation of 15 minutes and $n = 100$. In contrast, Route 2 takes 25 minutes on average with a standard deviation of 10 minutes and $n = 100$ [When both sample sizes are the same size the study is **balanced**]. Compare the averages using ANOVA.

Example 8. You want to compare two routes to work. Route 1 takes 30 minutes on average with a standard deviation of 15 minutes and $n = 100$. In contrast, Route 2 takes 25 minutes on average with a standard deviation of 10 minutes and $n = 100$. Compare the averages using 2-sampletTest. Go ahead and pool.

OBSERVATIONS: