8: One-Way Analysis of Variance (ANOVA)

Learning Objectives For This Lesson

Upon completion of this lesson, you should be able to:

- Logic Behind an Analysis of Variance (ANOVA)
- A Statistical Test for One-Way ANOVA
- Using Minitab to Perform One-Way ANOVA

8.1 - One-Way Analysis of Variance (ANOVA)

One-way ANOVA is used to compare means from at least three groups from one variable. The null hypothesis is that all the population group means are equal versus the alternative that at least one of the population means differs from the others. From a prior stat 200 survey a random sample of 30 students was taken to compare mean GPAs for students who sit in the front, middle and back of the classrooms. What follows is the Minitab output for the one-way ANOVA for this data: [NOTE: For explanations of the shaded pieces, place your mouse over the various acronyms in the row titled "Source"]

Interpreting this output:

1. A one-way analysis is used to compare the populations for one variable or factor. In this instance the one variable is Seating and there are 3 populations, also called group or factor levels being compared: front, middle and back.
2. DF stands for degrees of freedom. - The DF for the variable (e.g. Seating) is found by taking the number of group levels (called k) minus 1 (i.e. $k - 1$). - The DF for Error is found by taking the total sample size, N, minus k (i.e. $N - k$). - The DF for Total is found by $N - 1$.
3. The SS stands for Sum of Squares. The first SS is a measure of the variation in the data between the groups and for the Source lists the variable name (e.g. Seating) used in the analysis. This is sometimes referred to as SSB for "Sum of Squares Between groups". The next value is the sum of squares for the error often called SSE or SSW for "Sum of Squares Within". Lastly, the value for Total is called SST (or sometimes SSTO) for "Sum of Squares Total". These values are additive, meaning SST = SSB + SSW.
4. The test statistic used for ANOVA is the F-statistic and is calculated by taking the Mean Square (MS) for the variable divided by the MS of the error (called Mean Square of the Error or MSE). The F-statistic will always be at least 0, meaning the F-statistic is always nonnegative. This F-statistic is a

ratio of the variability *between* groups compared to the variability *within* the groups. If this ratio is large then the *p*-value is small producing a statistically significant result.(i.e. rejection of the null hypothesis)

5. The *p*-value is the probability of being greater than the F-statistic or simply the area to the right of the F-statistic, with the corresponding degrees of freedom for the group (number of group levels minus 1, or here $3 - 1 = 2$) and error (total sample size minus the number of group levels, or here $30 - 3 = 27$). The F-distribution is skewed to the right (i.e. positively skewed) so there is no symmetrical relationship such as those found with the *Z* or *t* distributions. This *p*-value is used to test the null hypothesis that all the group population means are equal versus the alternative that at least one is not equal. The alternative is not "they are not all equal."
6. The individual 95% confidence intervals provide one-sample t intervals that estimate the mean response for each group level. For example, the interval for Back provides the estimate of the population mean GPA for students who sit in the back. The * indicates the sample mean value (e.g. 3.13). You can inspect these intervals to see if the various intervals overlap. If they overlap then you can conclude that no difference in population means exists for those two groups. If two intervals do not overlap, then you can conclude that a difference in population means exists for those two groups.

8.2 - Hypotheses Statements and Assumptions for One−Way ANOVA

The hypothesis test for analysis of variance for *g* populations:

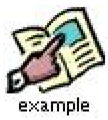
$H_0 : \mu_1 = \mu_2 = ... = \mu_g$
H_a : not all μ_i ($i = 1, ... g$) are equal

Recall that when we compare the means of two populations for independent samples, we use a 2-sample t-test with pooled variance when the population variances can be assumed equal. For more than two populations, the test statistic is the ratio of between group sample variance and the within-group-sample variance. Under the null hypothesis, both quantities estimate the variance of the random error and thus the ratio should be close to 1. If the ratio is large, then we reject the null hypothesis.

Assumptions: To apply or perform a One−Way ANOVA test, certain assumptions (or conditions) need to exist. If any of the conditions are not satisfied, the results from the use of ANOVA techniques may be unreliable. The assumptions are:

1. Each sample is an independent random sample
2. The distribution of the response variable follows a normal distribution
3. The population variances are equal across responses for the group levels. This can be evaluated by using the following *rule of thumb*: if the largest sample variance divided by the smallest sample variance is **not** greater than **two**, then assume that the population variances are equal.

8.3 - Logic Behind an Analysis of Variance (ANOVA)



We want to see whether the tar contents (in milligrams) for three different brands of cigarettes is different. Lab Precise took 6 samples from each of the three brands and got the following measurements:

Lab Precise's results from their study of tar contents (in milligrams) for three brands of cigarettes.

(Sample 1) Brand A	(Sample 2) Brand B	(Sample 3) Brand C

10.21	11.32	11.60
10.25	11.20	11.90
10.24	11.40	11.80
9.80	10.50	12.30
9.77	10.68	12.20
9.73	10.90	12.20
$\bar{X}_1$ = 10.00	$\bar{X}_2$ = 11.00	$\bar{X}_3$ = 12.00

Lab Sloppy also took 6 samples from each of the three brands and got the following measurements:

Lab Sloppy's results from their study of tar contents (in milligrams) for three brands of cigarettes.

(Sample 1) Brand A	**(Sample 2) Brand B**	**(Sample 3) Brand C**
9.03	9.56	10.45
10.26	13.40	9.64
11.60	10.68	9.59
11.40	11.32	13.40
8.01	10.68	14.50
9.70	10.36	14.42
$\bar{X}_1$ = 10.00	$\bar{X}_2$ = 11.00	$\bar{X}_3$ = 12.00

Data from Lab Precise:

```
 ...       .:
-------+---------+---------+---------+---------+---------Brand A
                 .   .   .     . . .
-------+---------+---------+---------+---------+---------Brand B
                                       .   . .     : .
-------+---------+---------+---------+---------+---------Brand C

     10.00     10.50     11.00     11.50     12.00     12.50
```

Data from Lab Sloppy:

```
.       .     .    .        . .
---+---------+---------+---------+---------+---------+---Brand A
             .     .  :    .                 .
---+---------+---------+---------+---------+---------+---Brand B
             :      .                        .       ..
---+---------+---------+---------+---------+---------+---Brand C

 8.4       9.6      10.8      12.0       13.2     14.4
```

The sample means from the two labs turned out to be the same and thus the differences in the sample means from the two labs are the same. From which data set can you draw more conclusive evidence that the means from the three populations are different? We need to compare the between-sample-variation to the within-sample-variation. Since the between-sample-variation is large compared to the within-sample-variation for data from Lab Precise, we will be more inclined to conclude that the three population means are different using the data from Lab Precise. Since such analysis is based on the analysis of variances for the data set, we call this statistical method the **Analysis of Variance**.

8.4 - Using Software to Perform One-Way ANOVA

The following example involves data that are from a survey sent out each spring to all students registered for STAT200. A total of 1004 students responded. If we assume that this sample represents the PSU-UP undergraduate population, then we can make inferences regarding this population based on the survey results.

Hypotheses Statements

① $H_0 : \mu_1 = \mu_2 = \mu_3$

H_a : not all the mean GPAs are equal

Data files: Course_Survey.MTW or Course_Survey.XLS.

- Using Minitab
- Using SPSS

To perform an Analysis of Variance (ANOVA) test in Minitab:

1. Open Minitab data set
2. Go to Stat > ANOVA > One-Way - Note: if the GPA responses for each sitting area, i.e. three columns one for each sitting area, then we would use One-Way(Unstacked)
3. Select the variable GPA and enter it in the text box for Response
4. Select the variable SitArea and enter it in the text box for Factor
5. Click OK

This should result in the following output:

One-way ANOVA: GPA versus SitArea

```
Source    DF       SS     MS     F      P
SitArea    2    1.329  0.665  3.12  0.045
Error    992  211.417  0.213
Total    994  212.746

S = 0.4617   R-Sq = 0.62%   R-Sq(adj) = 0.42%

                               Individual 95% CIs For Mean Based on
                               Pooled StDev
Level     N    Mean   StDev   ----+---------+---------+---------+-----
Back    182  3.1640  0.4404   (----------*----------)
Front   249  3.2698  0.4795                      (---------*---------)
Middle  564  3.2500  0.4603                      (------*-----)
                              ----+---------+---------+---------+-----
                                3.120     3.180     3.240     3.300

Pooled StDev = 0.4617
```

To the right of the one-way ANOVA table in the Minitab output, under the column headed P, is the *p*-value. In the example, P = 0.045 which is less than our usual alpha level of 0.05, but not less than 0.01 another often used level of significance. This is why setting alpha prior to conducting one"s analysis is important: to avoid potential conflicts of interest after the analysis is completed. The *p*-value is for testing the hypothesis in ① above, the mean GPA from the 3 seating areas are the same vs. not all the means are the same. Because the *p*-value of 0.045 is less than specified significance level of 0.05, we reject H_0. The data provide sufficient evidence to conclude that the mean GPA of students from the three seating locations are not all the same.

We can see below the ANOVA table, another table that provides the sample sizes, sample means, and sample standard deviations of the 3 samples. Beneath that table, the pooled StDev is given and this is an estimate for the common standard deviation of the 3 populations. Since the largest SD is not more than twice the smallest SD we can assume, using the rule of thumb, that the equal variance assumption is satisfied. With the sample sizes of each seating location level exceeding 30 we can assume normality by the Central Limit Theorem. Please note, however, that equal sample sizes can be very helpful in the design of ANOVA studies. Having equal sample sizes helps to protect against violations to the equal variance assumption.

Finally, the lower right side of the Minitab output gives individual 95% confidence intervals for the population means of the 3 seating locations. This example provides an excellent illustration as to why 2-sample *t*-tests can be misleading when more than two groups are being compared. As we see from the confidence interval output all of the intervals overlap. This indicates that none of the means would differ if we conducted a series of 2-sample tests of means. However, the ANOVA test does result in at least one mean being significantly different from the other.

8.5 - Multiple Comparisons

If you have a rejection of the null hypothesis the next logical step is to determine which group difference or differences led to the rejection. This is accomplished through a procedure known as **multiple comparisons** or **simultaneous** confidence intervals. Several comparison methods exist, but we are only going to introduce one: the Tukey method. This method is preferred since it is designed to produce an overall level of confidence that is very close to the level desired (e.g. 90% or 95%), and the intervals produced are narrower than another frequently applied method called the Bonferroni method. However, the Tukey formula is rather complex, but fortunately is readily available in most statistical software packages

Returning to the Open Course Survey, (Course_Survey.MTW or Course_Survey.XLS), data, repeat the steps for conducting the ANOVA test, but at the window interface click the 'Options' button (note that newer versions of Minitab may have a button titled 'Comparisons') and check the box for Tukey. For now, leave the error rate level at 5. This error rate as we will see in the output relates to the overall simultaneous rate of confidence. For an error rate of 5 we get 95% simultaneous rate; for an error rate of 2 we get 98%; and so on. What is meant by this *simultaneous level of confidence* is that we are this confident that ***all*** intervals in the entire set contain the true parameter family and not just any one of the intervals are correct.

```
Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons among Levels of SitArea

Individual confidence level = 98.05%


SitArea = Back subtracted from:

SitArea    Lower   Center   Upper  ---+---------+---------+---------+------
Front     0.0005   0.1058  0.2112            (------------*------------)
Middle   -0.0061   0.0860  0.1782           (-----------*----------)
                                   ---+---------+---------+---------+------
                                   -0.080     0.000     0.080     0.160


SitArea = Front subtracted from:

SitArea    Lower    Center   Upper  ---+---------+---------+---------+------
Middle   -0.1020   -0.0198  0.0624  (----------*---------)
                                    ---+---------+---------+---------+------
                                    -0.080     0.000     0.080     0.160
```

From the output, we see that we have three multiple comparisons: Back to Front, Back to Middle, and Front to Middle. We have only 3 because, with 3 levels for our factor of Sitting Area there are only 3 possible combinations of two−way comparisons. The Tukey level of confidence says that we are 95% confident that these three intervals contain the true difference between the factor levels being compared.

Alternatively, the output states that for any one interval we are 98.05% confident that the interval is correct. In other words, for example, we are 98.05% confident that the true mean GPA difference between students sitting in the Back compared to students sitting in the front is between 0.0005 to 0.2112.

We can interpret these confidence intervals as we would any confidence intervals for checking if a difference exists, i.e. does the interval contain the value of zero. If not, then there is a significant difference in the response between the two levels.

In this example we see that only the interval comparing Back to Front does not contain zero with a confidence interval of 0.0005 to 0.2112, while the remaining two intervals do include the value of zero. Notice that this interval is close to including zero and helps to explain why we are rejecting at the 0.05 level that not all of the means are equal, but with a p-value that is fairly close to alpha.