**Activity: One Way ANOVA**

Instructor Guide

**Goals:**

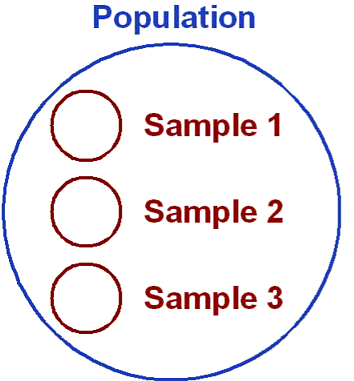
1. Discuss the logic behind a one-way ANOVA.
2. Use Excel to perform a one-way ANOVA.
3. Interpret the results of a one-way ANOVA.

**Required Materials:**

None

**Guiding Question:**

The reason this analysis is called ANOVA rather than multi-group means analysis (or something like that) is because it compares group means by analyzing comparisons of variance estimates. Consider (Draw on board):



1. Why might these means differ?
2. Answer:
   1. Group Membership (i.e., the *treatment effect* or IV).
   2. Differences not due to group membership (i.e., chance or *sampling error*).

**Introduction to the concept:**

**Analysis of variance (ANOVA) is used to test hypotheses about differences between two or more means**. The t-test based on the standard error of the difference between two means can only be used to test differences between two means. When there are more than two means, it is possible to compare each mean with each other mean using t-tests. However, conducting multiple t-tests can lead to severe inflation of the Type I error rate. **Analysis of variance can be used to test differences among several means for significance without increasing the Type I error rate.**

Analysis of variance tests the null hypothesis that all the population means are equal:

H0: µ1 = µ2 = ... =

HA: at least one mean is different

By comparing two estimates of variance (σ²) (σ² is the variance within each of the "a" treatment populations.) One estimate (called the **Mean Square Error Within Groups (MSE)**) is based on the variances **within** the samples. The MSE is an estimate of σ² whether or not the null hypothesis is true. The second estimate (**Mean Square Between** **Groups** (MSG) is based on the variance of the sample means. The MSG is only an estimate of σ² if the null hypothesis is true. If the null hypothesis is false then MSG estimates something larger than σ².

The logic by which analysis of variance tests the null hypothesis is as follows: If the null hypothesis is true, then MSE and MSG should be about the same since they are both estimates of the same quantity (σ²); however, if the null hypothesis is false then MSG can be expected to be larger than MSE since MSG is estimating a quantity larger then σ². Therefore, if MSG is sufficiently larger than MSE, the null hypothesis can be rejected.

**Student Engagement Task #1:**

**ONE WAY ANOVA Class Activity #1 –**

Conditions to run an ANOVA (2-3)

* The samples must be independent
* The populations from which the samples were obtained must be normally or approximately normally distributed
* The variances of the populations must be equal

1. The data are a simple random sample from less than 10% of the population
2. Normal probability plots for each group look good
3. Constant variance is especially important in this example with unequal group sizes

4.

 = (sometimes this is notated with ) = .0852

5. Total Sum of Squares or SS(T) = 0.00991

6. SS Between Samples or measure of the variation between samples (also sometimes called the SS treatment or SS factor

= .00452

7. MS (Between) = SS(B)/(k-1)

k-1 = 5-1 = 4

=.00452/4

=.001113

8. Sum of Squares Within Group, also called SS error is the variability that is assumed to be common to all populations being considered



=.00539

9.

MS(W) = SS(W)

-----------

N-k

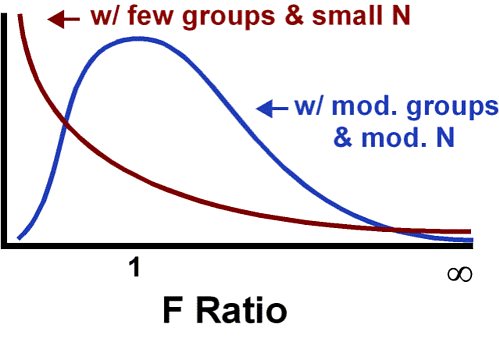
\*\*N is the total number of values in all samples combined.

=.00539/(39-5)

=.0001585

Ask students to “google” an F Distribution Table <http://www.stat.purdue.edu/~jtroisi/STAT350Spring2015/tables/FTable.pdf>

The F-distribution is named after the famous statistician R. A. Fisher. It is also sometimes known as the Fisher F distribution or the Snecedor-Fisher F distribution. F is the ratio of two variances.



**Question: How does the F distribution differ from other distributions you’ve seen?**

1. F is an extended family of distributions, which varies as a function of a pair of degrees of freedom (one for each variance estimate).
2. F is positively skewed.
3. F ratios, like the variance estimates from which they are derived, cannot have a value less than zero.

The F-distribution is most commonly used in Analysis of Variance (ANOVA) and the F test (to determine if two variances are equal). The F-distribution is the ratio of two chi-square distributions, and hence is right skewed. It has a minimum of 0, but no maximum value (all values are positive). The peak of the distribution is not far from 0, as can be seen in the following diagram

Guiding questions: What happens when you change the degrees of freedom?

Run applet with df from above: <http://socr.ucla.edu/htmls/dist/Fisher_Distribution.html>

It is important to note that when referencing the F-distribution the numerator degrees of freedom are always given first, and switching the degrees of freedom changes the distribution (ie. F(10,12) does not equal F(12,10)).

10. F = 0.001113/0.000159

F test statistic = 7.12

The larger the observed variability (between) in the sample means relative to the within group observations, the larger the F will be and the stronger the evidence against the null hypothesis. Because larger values of F represent stronger evidence against the null hypothesis, we use the upper tail of the distribution to computer a p value.

11. =F.DIST.RT(7.12, 4, 34)

Now show them the technique using the Data Analysis Tookpak. Work through the interpretation.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Anova: Single Factor |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Column 1 | 10 | 0.802 | 0.0802 | 0.000143 |  |  |
| Column 2 | 8 | 0.5984 | 0.0748 | 7.39E-05 |  |  |
| Column 3 | 7 | 0.7241 | 0.103442857 | 0.000263 |  |  |
| Column 4 | 8 | 0.6241 | 0.0780125 | 0.000168 |  |  |
| Column 5 | 6 | 0.5742 | 0.0957 | 0.000168 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.00452 | 4 | 0.001129919 | 7.121019 | 0.000281 | 2.649894 |
| Within Groups | 0.005395 | 34 | 0.000158674 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 0.009915 | 38 |  |  |  |  |

The p value is less than 0.05, indicating the evidence is strong enough to reject the null hypothesis at a significance level of 0.05.

Tukey-Kramer post-hoc analysis

When you reject the null hypothesis in an ANOVA analysis, you might wonder, which of these groups have different means? To answer the question, compare the means of each possible pair of groups. These comparisons can be accomplished using a two sample t test however this may lead to inflation in the Type I error rate. CAUTION: Sometimes an ANOVA will reject the null but no groups will have statistical differences (this does not invalidate the ANOVA conclusion).

To resolve the issue:

1. Bonferroni correction for α

* Which is α\* = α/K where K is the number of comparisons being considered.
* K= k(k-1)/2 or 5(5-1)/2
* K= 10
* .05 /10 = .005
* How does this change the result?

1. Tukeys or Sheffe’s post-hoc analysis

Online app and Tukey’s

[**http://web.mst.edu/~psyworld/tukeyscalculator.htm**](http://web.mst.edu/~psyworld/tukeyscalculator.htm)

and Tukey’s significant/probability table

12. Tukey’s Post Hoc Analysis findings:

|  |  |
| --- | --- |
| Tillamook-Newport | not sig |
| Tillamook-Petersburg | sig |
| Tillamook-Magadan | not sig |
| Tillamook-Tvarminne | not sig |
| Newport-Petersburg | sig |
| Newport-Magadan | not sig |
| Newport-Tvarminne | sig |
| Petersburg-Magadan | sig |
| Petersburg-Tvarminne | not sig |
| Magadan-Tvarminne | not sig |

13. **Student Engagement Task #2**: If time permits, ask students to complete 4.40, p. 215 in the Diez text

14. Lab: One Way ANOVA

**Clicker/WeBwork Questions:**

(1) The null hypothesis for an ANOVA F test is that

A. population means are equal.

B. sample standard deviations are not equal.

C. sample means are equal.

D. sample sizes are equal.

E. population standard deviations are not equal.

(2) In contrast to a chi-square test, an ANOVA F test is most appropriate when the

A. number of samples is two.

B. standard deviations must be estimated.

C. samples sizes are equal.

D. standard deviations are not equal.

E. data are quantitative.

(3) The alternative hypothesis for an ANOVA F test is that

A. all of the population variances are not equal.

B. none of the population means are equal.

C. at least two of the population means are not equal.

D. none of the population variances are equal.

E. all of the population means are not equal.

(4) The P value for an ANOVA F statistic is large if the sample

A. standard deviations are small.

B. sizes are small.

C. sizes are equal.

D. standard deviations are equal.

E. means are about equal.

(5) We anticipate a small P value for an ANOVA F statistic if the box plots for the samples are

A. wide and similarly located.

B. symmetrical.

C. narrow and located differently.

D. wide and have similar medians.

E. identical.

(6) Test the claim that the 5 radiologists have the same mean time to read a scan.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | DF | SS | MS | F | P |
| Radiologist | 4 | 59.08 | 19.76 | 2.37 | .096 |
| Radiology machine | 4 | 93.23 | 46.82 | 5.61 | .005 |
| Error | 16 | 48.33 | 8.35 |  |  |
| Total | 24 | 200.64 |  |  |  |

(a) The F−test statistic is .

(b) The P-value is .

(c) Is there sufficient evidence to warrant the rejection of the claim that the 5 radiologist have the same mean time to read a scan?

A. Yes

B. No

(7) Test the claim that the choice of radiology machine has no effect on the scan read time.

(a) The F− test statistic is .

(b) The P-value is .

(c) Is there sufficient evidence to warrant the rejection of the claim that the choice of radiology machine has no effect on the scan read time?

A. No

B. Yes

(8) The sample data are SAT scores on the verbal and math portions of SAT-I.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | DF | SS | MS | F | P |
| Gender | 1 | 52800 | 52800 | 5 | 0.032 |
| Verbal/Math | 1 | 6043 | 6043 | 0.57 | 0.454 |
| Interaction | 1 | 31618 | 31618 | 3 | 0.092 |
| Error | 36 | 376284 | 10555 |  |  |
| Total | 39 | 466745 |  |  |  |

Using a α=0.05 significance level. Test the claim that SAT scores are not affected by an interaction between gender and test (verbal/math).

(a) The F−test statistic is .

(b) The P-value is .

(c) Does there appear to be a significant effect from the interaction between gender and test?

A. No

B. Yes

2. Test the claim that gender has an effect on SAT scores.

(a) The F−test statistic is .

(b) The P-value is .

(c) Is there sufficient evidence to support the claim that gender has an effect on SAT scores?

A. Yes

B. No

3. Test the claim that the type of test (math/verbal) has an effect on SAT scores.

(a) The F−test statistic is .

(b) The P-value is .

(c) Is there sufficient evidence to support the claim that the type of test has an effect on SAT scores?

A. Yes

B. No