

Author: Brenda Gunderson, Ph.D., 2015

License: Unless otherwise noted, this material is made available under the terms of the Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License: <http://creativecommons.org/licenses/by-nc-sa/3.0/>

The University of Michigan Open.Michigan initiative has reviewed this material in accordance with U.S. Copyright Law and have tried to maximize your ability to use, share, and adapt it. The attribution key provides information about how you may share and adapt this material.

Copyright holders of content included in this material should contact open.michigan@umich.edu with any questions, corrections, or clarification regarding the use of content.

For more information about how to attribute these materials visit: <http://open.umich.edu/education/about/terms-of-use>. Some materials are used with permission from the copyright holders. You may need to obtain new permission to use those materials for other uses. This includes all content from:

Attribution Key

For more information see: <http://open.umich.edu/wiki/AttributionPolicy>

Content the copyright holder, author, or law permits you to use, share and adapt:



Creative Commons Attribution-NonCommercial-Share Alike License



Public Domain – Self Dedicated: Works that a copyright holder has dedicated to the public domain.

Make Your Own Assessment

Content Open.Michigan believes can be used, shared, and adapted because it is ineligible for copyright.



Public Domain – Ineligible. Works that are ineligible for copyright protection in the U.S. (17 USC §102(b)) *laws in your jurisdiction may differ.



Content Open.Michigan has used under a Fair Use determination
Fair Use: Use of works that is determined to be Fair consistent with the U.S. Copyright Act (17 USC § 107)
*laws in your jurisdiction may differ.

Our determination DOES NOT mean that all uses of this third-party content are Fair Uses and we DO NOT guarantee that your use of the content is Fair. To use this content you should conduct your own independent analysis to determine whether or not your use will be Fair.

Stat 250 Gunderson Lecture Notes

10: Analysis of Variance

Data! Data! Data! I can't make bricks without clay! -- *Sherlock Holmes* 

We have already been introduced to the concept of comparing the means of two populations when the data gathered represent independent random samples from normal populations. When the response was quantitative, we learned about two methods, an unpooled method and a pooled method.

We turn to discuss a method that allows us to compare the means of two or more normal populations based on independent random samples when the population variances are assumed to be equal. This method is called "**ANALYSIS OF VARIANCE**" (abbreviated **ANOVA**) and is **an extension of the two independent samples POOLED t-test**.

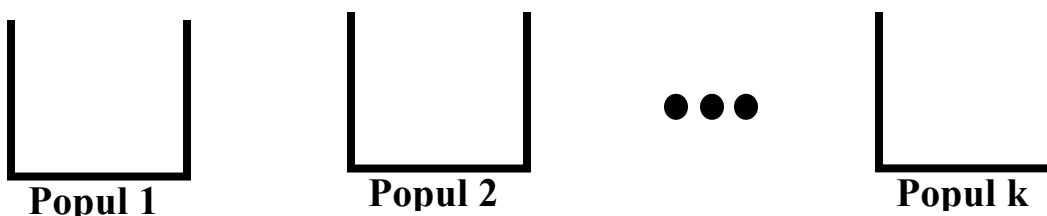
Let's step back for a moment to our two independent samples *t*-test. The purpose of this test was to decide whether or not two population means were equal:

H_0 : _____

The test was based on a *t* statistic that had _____ degrees of freedom.

One-way ANOVA is basically an extension of our two independent samples *t*-test to handling more than 2 populations. One-way ANOVA is a technique for testing whether or not the means of several populations are equal.

Picture:



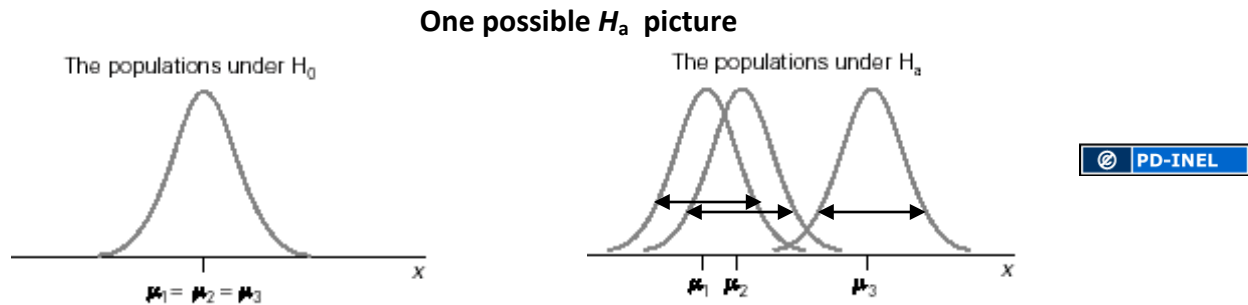
The assumptions are an extension of those for the two independent samples *t*-test to *k* groups.

- Each sample is a ... **random sample**
- The *k* random samples are ... **independent**
- For each of population the model for the response is... **a normal distribution**
- The *k* population variances are **equal**

The ANOVA Hypotheses:

H_0 : _____ versus H_a : _____

Notice this alternative does not require all the population means be different from each other.



Question: What call a technique for testing the **equality of the means** “analysis of **VARIANCE**”?

Answer: We are going to **compare two estimators of the common population variance**, σ^2 .

- MS Groups (Mean Square between the Groups):

- MSE (Mean Square Within or due to Error):

These two estimates are used to form the F statistic:

$$F = \frac{\text{Variation among sample means}}{\text{Natural variation within groups}} = \frac{MS \text{ Groups}}{MSE} .$$

If this F ratio is too _____ we would reject the null hypothesis.

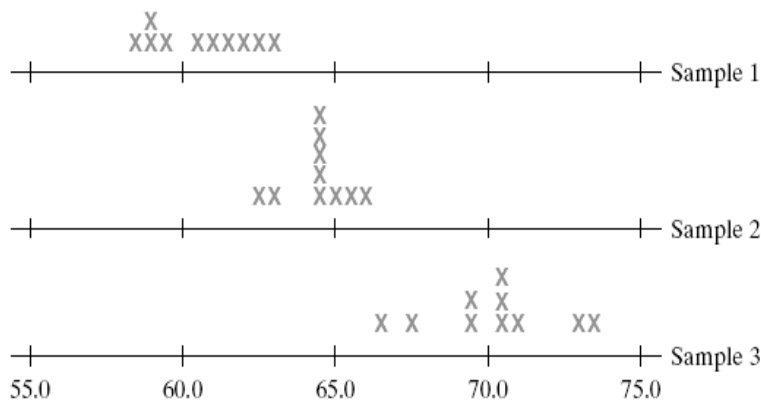
The Logic behind the ANOVA F-Test

Look at the plots below. For each Scenario, we have plotted data obtained by taking independent random samples of size 10 from three populations.

For Scenarios A and B, the three populations each had a normal distribution and the population means were 60, 65, and 70, respectively. So the population means are indeed not all equal.

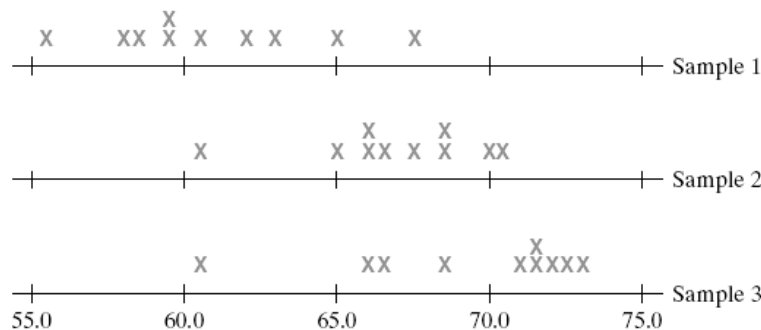
In Scenario A, the population standard deviations were all equal to 1.5. In Scenario B, the population standard deviations were all equal to 3. So in each case the assumption that the populations have equal standard deviations is met.

Scenario A



- Samples from 3 populations whose **means are different**.
- **Variability within** each population is **small**.
- **Difference between sample means more readily seen**.
- **F statistic somewhat big**.

Scenario B



- Samples from 3 populations whose **means are different**.
- **Variability within** each population is **larger**.
- **Difference between sample means not readily seen**.
- **F statistic smaller**.



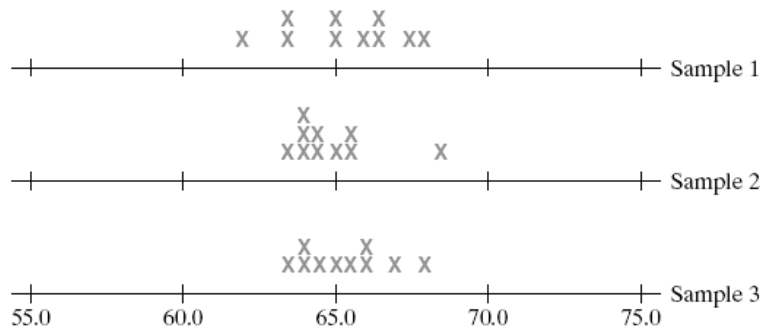
All images

Which of the above two scenarios do you think would provide more evidence that at least one of the population means is different from the others? Scenario A or Scenario B?

Below is a final set of plots for three independent random samples of size 10 each taken from a population with a normal model with a population mean of 65 and population standard deviation of 1.5. So in Scenario C, the population means are indeed all equal—that is, the null hypothesis tested in one-way ANOVA is true. Notice that, although the population means were all equal, there is still some of variation between the sample means.

Also in Scenario C there is still some natural variation within the samples, making the slight variation between the sample means hardly noticeable. The data in Scenario C do not provide evidence that the population means are different.

Scenario C



- Samples from 3 populations whose **means are all equal**.
- Still **some variability within** each population.
- **Very little differences between the sample means**.
- **F statistic very small**.

The F-statistic will be sensitive to differences between the sample means. The larger the **variation between the sample means**, the larger the value of the F-statistic and larger values of the F-statistic provide more support for rejecting the null hypothesis. The variation between the sample means was greatest for Scenarios A and B compared to Scenario C. The **natural variation within the samples** was greatest for Scenario B compared to Scenarios A and C. The F-statistic is the ratio of these two measures of variation:

$$F = \frac{\text{Variation among sample means}}{\text{Natural variation within groups}}$$

So which scenario would you expect to result in the largest value of the F-statistic? Provided below are the values of the F-statistic for the test of equality of the population means.

Scenario	Value of F Statistic	p-value
A (H_a is true)	F = 80.4	0.0000
B (H_a is true)	F = 16.4	0.01
C (H_0 is true)	F = 0.17	0.84

Note the value of the F-statistic is smallest and the p-value the largest when the null hypothesis is true (Scenario C). For Scenarios A and B, the population means are different, but the smaller population standard deviation in Scenario A accentuates the differences by producing a larger F-ratio and an extremely small p-value. **A larger F-statistic value (and thus smaller p-value) corresponds to more evidence that the population means are not all equal.**

Computing the F Test Statistic

We will see how to get MS Groups and MSE and perform the F test. These two mean squares will be a sum of squares (SS) divided by a corresponding degrees of freedom (DF).

The data can be generically represented below, where $X_{ij} = j^{\text{th}}$ observation from the i^{th} population. However we really don't have to worry too much about these subscripts, as we will go through the steps using words!

Data from Population 1	Data from Population 2	...	Data from Population k
X_{11}	X_{21}		X_{k1}
X_{12}	X_{22}		X_{k2}
\vdots	\vdots		\vdots
X_{1n_1}	X_{2n_2}		X_{kn_k}

The details leading to the F statistic are presented in six steps, ending with an ANOVA table.

Step 1: Calculate the mean and variance for each sample: \bar{x}_i, s_i^2

Step 2: Calculate the overall sample mean
(using all $N = n_1 + n_2 + \dots + n_k$ observations): \bar{x}

Step 3: Calculate the sum of squares between groups:

$$\text{SS Groups} = \sum_{\text{groups}} n_i (\bar{x}_i - \bar{x})^2$$

Step 4: Calculate the sum of squares within groups (due to error):

$$\text{SSE} = \sum_{\text{groups}} (n_i - 1) s_i^2$$

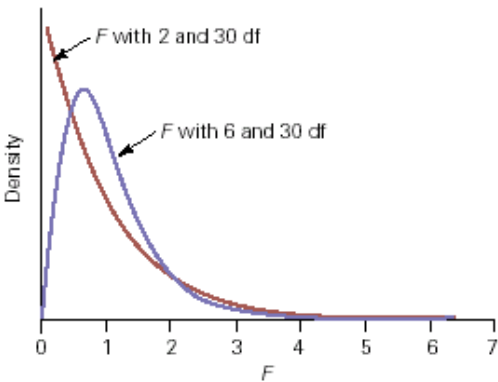
Step 5: OPTIONAL: Calculate the total sum of squares:

$$\text{SS Total} = \sum_{\text{values}} (x_{ij} - \bar{x})^2$$

Step 6: Fill in the ANOVA table:

Source	DF	Sum of Squares	Mean Square	F
Groups	k-1	SS Groups		
Error (Within)	N-k	SSE		
Total	N-1	SS Total		

If H_0 is true, then the F statistic, $F = \frac{MS \text{ Groups}}{MSE}$, has an $F(k - 1, N - k)$ distribution. Below are a few pictures of some F distributions.



From Utts, Jessica M. and Robert F. Heckard. *Mind on Statistics, Fourth Edition*. 2012. Used with permission.

Table A.4 provides percentiles of an F distribution. However, standard computer output also provides the exact p -value and completed ANOVA table. We will rely on R output to provide the p -value, but you should know how the ANOVA table is constructed and be able to sketch a picture of the p -value for an F -test.

Stat 250 Formula Card Summary of ANOVA

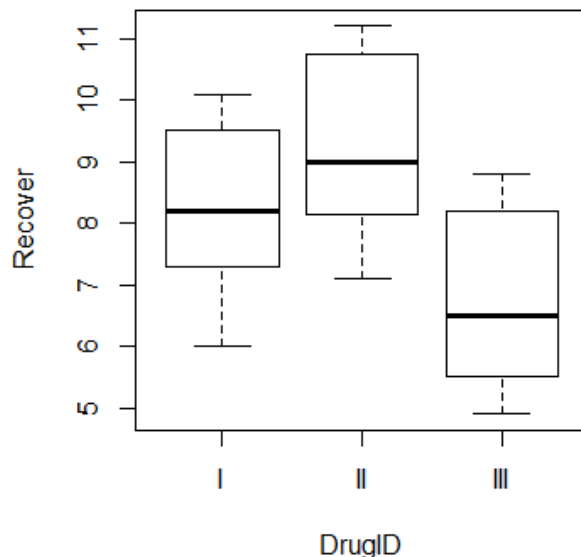
One-Way ANOVA				
$SS \text{ Groups} = SSG = \sum_{\text{groups}} n_i (\bar{x}_i - \bar{x})^2$	$MS \text{ Groups} = MSG = \frac{SSG}{k - 1}$	ANOVA Table		
$SS \text{ Error} = SSE = \sum_{\text{groups}} (n_i - 1) s_i^2$	$MS \text{ Error} = MSE = s_p^2 = \frac{SSE}{N - k}$	Source	SS	DF
$SS \text{ Total} = SSTO = \sum_{\text{values}} (x_{ij} - \bar{x})^2$	$F = \frac{MS \text{ Groups}}{MS \text{ Error}}$	MS	F	
Confidence Interval	$\bar{x}_i \pm t^* \frac{s_p}{\sqrt{n_i}}$	Groups	SS Groups	$k - 1$
	$df = N - k$	Error	SS Error	$N - k$
		Total	SSTO	$N - 1$
Under H_0 , the F statistic follows an $F(k - 1, N - k)$ distribution.				

Try It! Comparing 3 Drugs

We wish to compare three drugs for treating some disease. A quantitative response (time to cure in days) is measured such that a smaller value indicates a more favorable response.

A total of $N = 19$ patients are randomly assigned to one of the three drug (treatment) groups. The data are provided below:

Drug 1	Drug 2	Drug 3
7.3	7.1	5.8
8.2	10.6	6.5
10.1	11.2	8.8
6.0	9.0	4.9
9.5	8.5	7.9
	10.9	8.5
	7.8	5.2



Recall the assumptions for performing an F -test. Think about how you would check them.

- Each sample is a ... **random sample**
- The k random samples are ... **independent**
- For each of population the model for the response is... **a normal distribution**
- The k population variances are **equal**.

State the hypotheses to be tested:

H_0 : _____ H_a : _____

Note: We would use a computer or calculator to work at least the basic summaries in steps 1 and 2, and likely to create the entire ANOVA table for us. Let's be sure we understand where the values are coming from and how to interpret the final results.

Drug 1	Drug 2	Drug 3
7.3	7.1	5.8
8.2	10.6	6.5
10.1	11.2	8.8
6.0	9.0	4.9
9.5	8.5	7.9
	10.9	8.5
	7.8	5.2

Step 1: Calculate the mean and variance for each sample:

$$\bar{x}_1 = \quad \quad \quad s_1^2 =$$

$$\bar{x}_2 = \quad \quad \quad s_2^2 =$$

$$\bar{x}_3 = \quad \quad \quad s_3^2 =$$

Step 2: Calculate the overall sample mean (based on all $N = n_1 + n_2 + \dots + n_k$ observations):

$$\bar{x} =$$

Step 3: Calculate the sum of squares between groups:

$$SS \text{ Groups} = \sum_{groups} n_i (\bar{x}_i - \bar{x})^2$$

Step 4: Calculate the sum of squares within groups (due to error):

$$SSE = \sum_{groups} (n_i - 1) s_i^2$$

Step 5: OPTIONAL: Calculate the total sum of squares: *No Thank You!*

Step 6: Fill in the ANOVA table:

Source	Sum of Squares	DF	Mean Square	F
Groups				
Error (Within)				
Total				

Here are the results from R:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
DrugID	2	21.98	10.991	4.188	0.0345 *
Residuals	16	41.99	2.624		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

One of the assumptions in ANOVA is that the population standard deviations are all equal. Using the data, give an estimate of this common population standard deviation.

Give the observed test statistic value.

What is the distribution of the test statistic if the three drugs are equally effective in terms of the mean response?

What is the corresponding p -value for assessing if the three drugs are equally effective in terms of the mean response?

At the 5% level, what is your conclusion?

We Rejected H_0 in ANOVA: What is next? Multiple Comparisons

The term **multiple comparisons** is used when two or more comparisons are made to examine the specific pattern of differences among means. The most commonly analyzed set of multiple comparisons is the set of all pairwise comparisons among population means. In our previous Drug example, the possible pairwise comparisons are: Drug 1 with Drug 2, Drug 1 with Drug 3, and Drug 2 with Drug 3. To compare the pair of means we could ...

- Compute a **confidence interval** for the difference between the two population means and see if 0 falls in the interval or not.
- Perform a **test of hypotheses** to assess if the two population means differ significantly.

When many statistical tests are done there is an increased risk of making at least one type I error (erroneously rejecting a null hypothesis). Consequently, several procedures have been developed to **control the overall family type I error rate** or the **overall family confidence level** when inferences for a set (family) of multiple comparisons are done.

Tukey's procedure is one such procedure for the family of pairwise comparisons. If the family error rate is not a concern, Fisher's procedure is used.

Try It! Comparing 3 Drugs

In the comparison of the three drugs, we rejected the null hypothesis at the 5% significance level. We follow with a multiple comparison procedure to determine which group means are significantly different from each other.

R gives family-wise confidence interval comparisons using Tukey's method and a family confidence level of 95%.

```
95% family-wise confidence level

Linear Hypotheses:
      Estimate 1wr      upr
II - I == 0    1.0800 -1.3670  3.5270
III - I == 0   -1.4200 -3.8670  1.0270
III - II == 0  -2.5000 -4.7338 -0.2662

      I      II      III
"ab"  "b"   "a"
```

- Use the above output to report about the three pairwise comparisons:
Does the confidence interval for comparing Drug I and II contain 0? _____
Does the confidence interval for comparing Drug I and III contain 0? _____
Does the confidence interval for comparing Drug II and III contain 0? _____
- State your conclusions regarding the differences between the mean response for the three drug groups based on the Tukey family-wise comparison method.

We can conclude that the population mean responses differ for ...

but do not differ for ...

Individual Confidence Intervals for the Population Means

Sometimes it is helpful to examine a confidence interval for the mean for each population. Since in ANOVA we assume the population standard deviations are all equal, the estimate of that common population standard deviation $s_p = \sqrt{MSE}$ is used in forming the individual confidence intervals. The degrees of freedom used to find the t^* multiplier will be those associated with the estimated standard deviation, namely $N - k$. The formula for the individual confidence intervals is provided below.

Confidence Interval	$\bar{x}_i \pm t^* \frac{s_p}{\sqrt{n_i}}$	$df = N - k$
----------------------------	--	--------------

Try It! Comparing 3 Drugs

We were comparing $k = 3$ groups based on a total of $N = 19$ observations. The pooled standard deviation for the comparison of the three drugs data set is $s_p = 1.62$. The sample means and sample sizes were:

Drug 1:	Sample mean = 8.22	Sample size = 5
Drug 2:	Sample mean = 9.30	Sample size = 7
Drug 3:	Sample mean = 6.80	Sample size = 7

The degrees of freedom for the t^* multiplier is $N - k =$ _____.

From the table of t^* multipliers (Table A.2) with confidence level = 0.95

and the above degrees of freedom we have $t^* =$ _____

Drug 3 was descriptively the best. Compute a 95% confidence interval for the population mean time to cure for all subjects taking Drug 3.

Try It! Memory Experiment

In a memory experiment, three groups of subjects were given a list of words to try to remember. The length of the list for the first group was 10 words (short list), whereas for the second group it was 20 words (medium list) and for the third group 40 words (long list). The percentage of words recalled for each subject was recorded. The sample mean percentage of words recalled was 68.3% for the short list, 48% for the medium list, and 39.2% for the long list. A one-way ANOVA was used to assess whether the *length of the list* had a significant effect on the *percentage of words recalled*.

	df	SS	Mean Square	F	Sig.
List Length	2	2668.8			.0003
Residuals			84.6		
Total	16	3852.9			

- Some values in the ANOVA table are missing. Complete the above table.
- State the null and alternative hypotheses that the above F statistic is testing.

H_0 : _____ vs. H_a : _____

- Suppose the necessary assumptions hold. Using a 5% significance level, does it appear that the average percentage of words recalled is the same for the three different lengths of lists? Explain.
- Family-wise comparisons were performed using Tukey's method.

95% family-wise confidence level			
Linear Hypotheses:			
	Estimate	lwr	upr
medium - short == 0	-20.33	-39.61	-1.06
long - short == 0	-29.17	-47.54	-10.79
long - medium == 0	-8.83	-28.11	10.44
short	medium	long	
	"a"	"a"	

Use the results and circle the pairs that are significantly different at a 5% level.

short versus medium short versus long medium versus long

- Give a 99% confidence interval for the population mean percentage of words recalled for the long list group. Recall that the sample mean based on the 6 subjects in the long list group was 39.2 percent.

What if some conditions do not hold?

You probably won't be surprised to learn that the necessary conditions for using an analysis of variance F -test don't hold for all data sets. There are methods that can be used when one or both of the assumptions about equal population standard deviations and normal distributions are violated.

When the observed **data are skewed**, or when extreme outliers are present, it usually is better to **analyze the median rather than the mean**. One test for comparing medians is the **Kruskal-Wallis Test**. It is based on a comparison of the relative rankings (sizes) of the data in the observed samples, and for this reason is called a rank test. The term **nonparametric test** also is used to describe this test because there are no assumptions made about a specific distribution for the population of measurements. Another nonparametric test used to compare population medians is **Mood's Median Test**.

Two-Way ANOVA

So far we have focused on the **one-way ANOVA** procedure. The "*one-way*" referred to having only one explanatory variable (or factor) and one quantitative response variable.

Two-way ANOVA examines the effect of two explanatory variables (or factors) on the mean response. The researcher is interested in the *individual effect* of each explanatory variable on the mean response and also in the *combined effect* of the two explanatory variables on the mean response. The individual effect of each factor on the response is called a **main effect**. If one of the factors does not have an effect on the response, we say there is no main effect due to that factor.

Besides assessing the main effects of each factor on the response, an interesting feature in two-way analyses is the possibility of interaction between the two factors. We say **there is interaction between two factors if the effect of one factor on the mean response depends on the specific level of the other factor**. The interpretation of the factor main effects can be more difficult when interaction is present.

Additional Notes

A place to ... jot down questions you may have and ask during office hours, take a few extra notes, write out an extra problem or summary completed in lecture, create your own summary about these concepts.

