

Module 9.1 Handout: One-way ANOVA Examples

Jon Starkweather, PhD

October 21, 2010

Jon Starkweather, PhD
jonathan.starkweather@unt.edu
Consultant
Research and Statistical Support



<http://www.unt.edu>



<http://www.unt.edu/rss>

The Research and Statistical Support (RSS) office at the University of North Texas hosts a number of “Short Courses”.

A list of them is available at:

<http://www.unt.edu/rss/Instructional.htm>

Contents

1	Introduction to Example	3
2	ANOVA Calculation	3
3	Planned Comparisons	4
3.1	PC1	4
3.2	PC2	5
4	Post-hoc Testing	6
4.1	Tukey's HSD	6
4.2	Games-Howell	7
5	Effect Size	8
5.1	Omega Squared for F	8
5.2	Omega Squared for PCs	8
6	Confidence Intervals	9
6.1	CI for Planned Comparisons	9
6.2	CI for Example Planned Comparisons (PC)	9
6.3	CI for Tukey's HSD Post-hoc (PH) test	10
7	Welch Robust Procedure	11

1 Example Description

A study of the effects of font color on recall of presentation material was conducted. Participants ($n = 12$) were randomly drawn from the UNT undergraduate research pool and randomly assigned to one of three font color groups (Red, Green, & Blue). Participants were then instructed to watch a presentation of 30 words, each word was presented for 45 seconds and the participants were instructed to memorize the words. Participants then completed a 1 hour task unrelated to the experiment. Participants were then instructed to recall as many of the words as possible. The number of words recalled for each participant is below.

Table 1: Font Color Groups' Data

Red Group	Green Group	Blue Group	
25	10	18	
22	12	17	
23	11	19	
21	13	16	Totals
$\sum X = 91$	46	70	207
$n_j = 4$	4	4	$n_t = 12$
$\bar{X}_j = 22.75$	11.50	17.50	$\bar{X}_{..} = 17.25$
$S_j^2 = 2.92$	1.66	1.66	
$S_j = 1.71$	1.29	1.29	
$df_j = 3$	3	3	$df_t = 11$

$\bar{X}_{..}$ is the Grand Mean (mean of the means).

df_t is the total number of individuals (12) minus 1.

2 Standard Calculation of One-way ANOVA

Between Groups (b)

$$SOS_b = \sum n_j (\bar{X}_j - \bar{X}_{..})^2$$

$$SOS_b = 4(22.75 - 17.25)^2 + 4(11.5 - 17.25)^2 + 4(17.5 - 17.25)^2 = 253.50$$

$$df_b = k - 1$$

$$df_b = 3 - 1 = 2$$

$$MS_b = \frac{SOS_b}{df_b} = \frac{253.50}{2} = 126.75$$

$$SOS_w = \sum \sum (X_{ij} - \bar{X}_j)^2$$

$$SOS_w = (25 - 22.75)^2 + (22 - 22.75)^2 + (23 - 22.75)^2 + (21 - 22.75)^2 + \dots$$

$$(10 - 11.50)^2 + (12 - 11.50)^2 + (11 - 11.50)^2 + (13 - 11.50)^2 + \dots$$

$$(18 - 17.50)^2 + (17 - 17.50)^2 + (19 - 17.50)^2 + (16 - 17.50)^2 = 18.75$$

$$df_w = n_t - k$$

$$df_w = 12 - 3 = 9$$

$$MS_w = \frac{SOS_w}{df_w} = \frac{18.75}{9} = 2.083$$

$$F_{calc} = \frac{MS_b}{MS_w} = \frac{126.75}{2.083} = 60.84$$

Sums of Squares total (and degrees of freedom total) will be needed for some effect size measures, but are not strictly necessary for calculation of the ANOVA. However, they do allow you to check your previous calculation because $SOS_b + SOS_w = SOS_t$ and $df_b + df_w = df_t$.

$$SOS_t = \sum (X_{ij} - \bar{X}_{..})^2$$

$$SOS_t = (25 - 17.25)^2 + (22 - 17.25)^2 + (23 - 17.25)^2 + (21 - 17.25)^2 + \dots$$

$$(10 - 17.25)^2 + (12 - 17.25)^2 + (11 - 17.25)^2 + (13 - 17.25)^2 + \dots$$

$$(18 - 17.25)^2 + (17 - 17.25)^2 + (19 - 17.25)^2 + (16 - 17.25)^2 = 272.25$$

$$df_t = n_t - 1$$

$$df_t = 12 - 1 = 11$$

Construct the ANOVA summary table.

Table 2: Summary Table

Source	SOS	df	MS	F
Between	253.50	2	126.750	60.84
Within	18.75	9	2.083	
Total	272.25	11		

Use df_b (as the numerator) and df_w (as the denominator) to look up the critical value; using a significance level of 0.05. We find a critical value (F_{crit}) of 4.26 with $df_b = 2$ and $df_w = 9$ at 0.05.

http://faculty.vassar.edu/lowry/apx_d.html

So, because $F_{calc} = 60.84 > 4.26 = F_{crit}$ we reject the null hypothesis and conclude that the three populations (each represented by a sample group) are not the same. However, the *omnibus* F only tells us whether or not there exists a significant difference among the groups, it does not tell us which of the groups were significantly different from which other group(s). To ascertain the specific difference(-es), we need to do Planned Comparisons **or** Post-hoc testing.

3 Planned Comparisons Example

3.1 Simple Planned Comparison 1

Simple Planned Comparison 1: Red group ($\bar{X} = 22.75$) versus the Blue group ($\bar{X} = 17.50$), with the Grand Mean ($\bar{X}_{..} = 20.125$) for **this comparison**.

Between groups (b):

$$SOS_b = \sum n_j (\bar{X}_j - \bar{X}_{..})^2$$
$$SOS_b = 4(22.75 - 20.125)^2 + 4(17.50 - 20.125)^2 = 55.12$$

df_b for **simple** planned comparisons is always 1 because, **simple** planned comparisons involve only two groups.

$$df_b = k - 1$$
$$df_b = 2 - 1 = 1$$

MS_b for **simple** planned comparisons is always equal to SOS_b of the comparison because, **simple** planned comparisons always have $df_b = 1$.

$$MS_b = \frac{SOS_b}{df_b} = \frac{55.12}{1} = 55.12$$

SOS_w and df_w are the same as was used in the omnibus F , so MS_w is the same as well: 2.083

$$F = \frac{MS_b}{MS_w} = \frac{55.12}{2.083} = 26.46$$

Using $df_b = 1$ and $df_w = 9$ with 0.05 significance, we find $F_{crit} = 5.12$. So, because $F_{calc} = 26.46 > 5.12 = F_{crit}$ we reject the null hypothesis and conclude that the two populations (each represented by a sample group) are not the same. Or, more precisely (after looking at the mean of each group), we could say the Red group ($\bar{X} = 22.75$) recalled significantly more words than the Blue group ($\bar{X} = 17.50$).

3.2 Simple Planned Comparison 2

Simple Planned comparison 2: Blue group ($\bar{X} = 17.50$) versus the Green group ($\bar{X} = 11.50$), with the Grand Mean ($\bar{X}_{..} = 14.50$) for **this comparison**.

Between groups (b):

$$SOS_b = \sum n_j (\bar{X}_j - \bar{X}_{..})^2$$
$$SOS_b = 4(17.50 - 14.50)^2 + 4(11.50 - 14.50)^2 = 72.00$$

df_b for **simple** planned comparisons is always 1 because, **simple** planned comparisons involve only two groups.

$$df_b = k - 1$$
$$df_b = 2 - 1 = 1$$

MS_b for **simple** planned comparisons is always equal to SOS_b of the comparison because, **simple** planned comparisons always have $df_b = 1$.

$$MS_b = \frac{SOS_b}{df_b} = \frac{72.00}{1} = 72.00$$

SOS_w and df_w are the same as was used in the omnibus F , so MS_w is the same as well: 2.083

$$F = \frac{MS_b}{MS_w} = \frac{72.00}{2.083} = 34.57$$

Using $df_b = 1$ and $df_w = 9$ with 0.05 significance, we find $F_{crit} = 5.12$. So, because $F_{calc} = 34.57 > 5.12 = F_{crit}$ we reject the null hypothesis and conclude that the two populations (each represented by a sample group) are not the same. Or, more precisely (after looking at the mean of each group), we could say the Blue group ($\bar{X} = 22.75$) recalled significantly more words than the Green group ($\bar{X} = 17.50$).

4 Post-hoc Testing

4.1 Tukey's HSD

Tukey's Honestly Significant Differences (HSD) test is considered one of the original post-hoc tests and offers a fixed family-wise error rate at alpha (significance level). Once we have ranked the means {11.50, 17.50, 22.75} and found our critical value ($q_{crit} = 3.95$) using $r = 3$ or $k = 3$ (different symbols which mean the same thing: number of means or groups). (<http://www.stat.duke.edu/courses/Spring98/sta110c/qtable.html>) We can calculate the minimum difference between means needed to discover a significant difference.

$$\bar{X}_i - \bar{X}_j = q_{crit} \sqrt{\frac{MS_w}{n_g}} = 3.95 \sqrt{\frac{2.083}{4}} = 2.85$$

So, if a mean difference ($\bar{X}_i - \bar{X}_j$) is larger than 2.85 we would conclude there is a significant difference between those two means. With our minimum significant difference calculated at 2.85; we can compare the differences among our three means to it, to determine which means differ *significantly* from the others.

Table 3: Pair-wise Comparison Table

	Green	Blue	Red
$q_{r,.05} = 2.85$	11.50	17.50	22.75
Green = 11.50	0	6.00	11.25
Blue = 17.50	–	0	5.25
Red = 22.75	–	–	0

So, we find a significant difference between each pair of means because, each difference is greater than 2.85.

4.2 Games-Howell

The Tukey test above assumes equal sample sizes for each group **and** equal variances among the groups. Often ‘real data’ do not conform to what we would expect and these assumptions are often violated. The Games-Howell test provides a method for when either (or both) of the assumptions are not upheld. Essentially, this test incorporates the samples sizes and variances of each group being compared. The df and the critical difference between means ($\bar{X}_i - \bar{X}_j$) are modified for inclusion of samples sizes and variances. Until now we have consistently used df_w for finding our critical values (e.g., q_{crit}). With the Games-Howell test, we actually calculate df' using each groups’ sample size and variance; such that **for each pair of groups**:

$$df' = \frac{\left(\frac{S_i^2}{n_i} + \frac{S_j^2}{n_j}\right)^2}{\frac{\left(\frac{S_i^2}{n_i}\right)^2}{n_i-1} + \frac{\left(\frac{S_j^2}{n_j}\right)^2}{n_j-1}}$$

where subscript i and subscript j identify descriptive statistics from each group being compared.

Example df'

Red (i): $\bar{X} = 22.75, S^2 = 2.92, n = 4$

Green (j): $\bar{X} = 11.50, S^2 = 1.66, n = 4$

$$df' = \frac{\left(\frac{S_i^2}{n_i} + \frac{S_j^2}{n_j}\right)^2}{\frac{\left(\frac{S_i^2}{n_i}\right)^2}{n_i-1} + \frac{\left(\frac{S_j^2}{n_j}\right)^2}{n_j-1}} = \frac{\left(\frac{2.92}{4} + \frac{1.66}{4}\right)^2}{\frac{\left(\frac{2.92}{4}\right)^2}{4-1} + \frac{\left(\frac{1.66}{4}\right)^2}{4-1}} = \frac{(.73+.415)^2}{\frac{(.73)^2}{3} + \frac{(.415)^2}{3}} = \frac{(1.145)^2}{\frac{.5329}{3} + \frac{.1722}{3}} = \frac{1.311}{.1776+.0574} = \frac{1.311}{.235} =$$

$$df' = 5.5787$$

Pay careful attention to the *fact* that S^2 is a **symbol**, not an operation.

Minimum Significant Difference: So, given $r = 3$, significance level of 0.05, and now $df' = 5.787$ we look to the q table and find: $q_{crit} \approx 4.60$ (its always better to choose the more conservative, larger, critical value).

<http://www.stat.duke.edu/courses/Spring98/sta110c/qtable.html>

Where earlier we had:

$$\bar{X}_i - \bar{X}_j = q_{crit} \sqrt{\frac{MS_w}{n_g}}$$

Now, **for each pair of means**, we have:

$$\bar{X}_i - \bar{X}_j = q_{crit} \sqrt{\frac{\frac{S_i^2}{n_i} + \frac{S_j^2}{n_j}}{2}}$$

For the current Red vs. Green example:

$$\bar{X}_i - \bar{X}_j = 5.787 \sqrt{\frac{\frac{2.92}{4} + \frac{1.66}{4}}{2}} = 5.787 \sqrt{\frac{.73+.415}{2}} = 5.787 \sqrt{\frac{1.145}{2}} = 5.787 \sqrt{.5725} = 5.787 * .7566 =$$

$$\bar{X}_i - \bar{X}_j = 4.379$$

Given our minimum significant difference (*for this pair*) of 4.379, we need an actual mean difference larger than this to conclude a significant difference was present. Clearly, $\bar{X}_i - \bar{X}_j = 22.75 - 11.50 = 11.25$ is greater than 4.379; which indicates the Red group recalled significantly more words than the Green group. Just remember, all that calculating must be done for each group because each group may have a different sample size and more than likely *will* have a different variance (S^2). The ANOVA and associated tests of individual means (planned comparisons or post-hoc tests) are vulnerable to violations of the homogeneity assumption. Therefore, the Games-Howell test is highly recommended.

5 Effect Size Measures

5.1 Omega Squared for F

Omega squared (ω^2) is the recommended effect size measure because, it takes into account the number of groups.

The Omnibus F ANOVA summary table; same as Table 2 above.

Table 4: ANOVA Summary Table

Source	SOS	df	MS	F
Between	253.50	2	126.750	60.84
Within	18.75	9	2.083	
Total	272.25	11		

$$\omega^2 = \frac{SOS_b - (k - 1)MS_w}{SOS_t + MS_w} = \frac{253.5 - (3 - 1)2.083}{272.25 + 2.083} = \frac{249.332}{274.33} = .9089$$

Here we see a less biased, more accurate representation of the relationship between the IV and the DV; 90.89% of the variance in words recalled (DV) is attributable to the differences in font color (IV). Please note; Omega squared should also be calculated on the results of Planned Comparisons if they were performed (examples below).

5.2 Omega Squared for PCs

Recall, in the Post-hoc testing situation or the Simple Planned Comparisons situation, Cohen's d can be used when testing individual pairs of means. However, Omega squared is preferred because it is less biased.

PC 1 from earlier:

$$\omega^2 = \frac{SOS_b - (k - 1)MS_w}{SOS_t + MS_w} = \frac{55.12 - (2 - 1)2.083}{272.25 + 2.083} = \frac{53.037}{274.33} = .1933$$

So, 19.33% of the variance in words recalled is attributable to the difference between Red and Blue font color.

PC 2 from earlier:

$$\omega^2 = \frac{SOS_b - (k - 1)MS_w}{SOS_t + MS_w} = \frac{72 - (2 - 1)2.083}{272.25 + 2.083} = \frac{69.917}{274.33} = .2549$$

So, 25.49% of the variance in words recalled is attributable to the difference between Green and Blue font color.

6 Confidence Intervals

In the ANOVA situation, Confidence Intervals (CIs) are not meaningful, **until** you get down to testing individual pairs of means. At the level of planned comparisons or post-hoc testing, CIs are meaningful. In the situation of testing individual pairs of means, we could review the previous module (*t* tests) and apply those methods to the current testing of individual pairs of means – for calculating a CI (and Cohen’s *d*). Or, we could use current information and our general understanding of the equations for the upper limit (UL) and lower limit (LL) of a CI.

Recall the general equations for CI limits:

$$UL = (+crit) * (SE) + mean$$

$$LL = (-crit) * (SE) + mean$$

6.1 CI for Planned Comparisons

During planned comparison testing, we were interested in the *difference between* pairs of group means and we had all the elements we needed to compute the UL and LL for each comparison. Here we apply those elements to the general form of the UL and LL equations for **each Planned Comparison**. However, we are now in a two group comparison (essentially *t* testing) so given that $F = t^2$ we need to take the square root of F_{crit} and MS_w .

$$UL = \sqrt{(+F_{crit}) * (MS_w)} + \bar{X}_i - \bar{X}_j$$

$$LL = \sqrt{(-F_{crit}) * (MS_w)} + \bar{X}_i - \bar{X}_j$$

As many of you know, you can not take the square root of a negative number. So, recognize that the formulas below are equivalent for what we want to do just above.

$$UL = \sqrt{(F_{crit}) * (MS_w)} + \bar{X}_i - \bar{X}_j$$

$$LL = \sqrt{(F_{crit}) * (MS_w)} - \bar{X}_i - \bar{X}_j$$

6.2 CI for Example Planned Comparisons (PC)

CI for PC 1: Red ($\bar{X} = 22.75$) vs. Blue ($\bar{X} = 17.50$)

$$UL = \sqrt{(+F_{crit}) * (MS_w)} + \bar{X}_i - \bar{X}_j = \sqrt{5.12 (2.083)} + 5.25 = 8.52$$

$$LL = \sqrt{(-F_{crit}) * (MS_w)} + \bar{X}_i - \bar{X}_j = \sqrt{5.12 (2.083)} - 5.25 = 1.98$$

If we drew an infinite number of samples, we would expect 95% of the mean differences between the Red and Blue groups to be between 8.52 and 1.98.

CI for PC 2: Green ($\bar{X} = 11.50$) vs. Blue ($\bar{X} = 17.50$)

$$UL = \sqrt{(+F_{crit}) * (MS_w)} + \bar{X}_i - \bar{X}_j = \sqrt{5.12 (2.083)} + 6.00 = 9.27$$

$$LL = \sqrt{(-F_{crit}) * (MS_w)} + \bar{X}_i - \bar{X}_j = \sqrt{5.12 (2.083)} - 6.00 = 2.73$$

If we drew an infinite number of samples, we would expect 95% of the mean differences between the Green and Blue groups to be between 9.27 and 2.73.

6.3 CI for Tukey's HSD Post-hoc (PH) test

With Tukey's Post-hoc testing, we are using modified t tests, so we can simply use the calculated **minimum difference between means** (2.85) from earlier.

CI for PH 1: Red ($\bar{X} = 22.75$) vs. Blue ($\bar{X} = 17.50$)

$$UL = (+q_{r=3,.05}) + \bar{X}_i - \bar{X}_j = +2.85 + 5.25 = 8.10$$

$$LL = (-q_{r=3,.05}) + \bar{X}_i - \bar{X}_j = -2.85 + 5.25 = 2.40$$

If we drew an infinite number of samples, we would expect 95% of the mean differences between the Red and Blue groups to be between 8.10 and 2.40.

CI for PH 2: Red ($\bar{X} = 22.75$) vs. Green ($\bar{X} = 11.50$)

$$UL = (+q_{r=3,.05}) + \bar{X}_i - \bar{X}_j = +2.85 + 11.25 = 14.10$$

$$LL = (-q_{r=3,.05}) + \bar{X}_i - \bar{X}_j = -2.85 + 11.25 = 8.40$$

If we drew an infinite number of samples, we would expect 95% of the mean differences between the Red and Green groups to be between 14.10 and 8.40.

CI for PH 3: Blue ($\bar{X} = 17.50$) vs. Green ($\bar{X} = 11.50$)

$$UL = (+q_{r=3,.05}) + \bar{X}_i - \bar{X}_j = +2.85 + 6.00 = 8.85$$

$$LL = (-q_{r=3,.05}) + \bar{X}_i - \bar{X}_j = -2.85 + 6.00 = 3.15$$

If we drew an infinite number of samples, we would expect 95% of the mean differences between the Green and Blue groups to be between 8.85 and 3.15.

7 Welch Robust Procedure

As mentioned in the discussion of the Games-Howell post-hoc test, ANOVA can be biased when data are not evenly distributed among the groups and/or the groups have heterogeneous variances (i.e., violated assumptions). Fortunately, there exists a very good alternative when data do not fit the standard ANOVA assumptions: the Welch Robust Procedure. The *Good News*: gain in power and protection against Type I error. The *Bad News*: Requires using different (somewhat cumbersome) calculations; although most of the symbols you already know from previous usage.

Welch Formulas

First, calculate a few new terms w_k and \bar{X}' , where w_k refers to **weight for each group** and \bar{X}' refers to **weighted grand mean**.

$$w_k = \frac{n_k}{S_k^2} \text{ which can be expressed as } w_j = \frac{n_j}{S_j^2}$$

Notice, I have listed above another way of expressing the equation using our standard subscripts (i) and (j), where (j) refers to ‘group’ or column. It is important to note that this equation should be applied to all the groups.

For the current example, the Red group’s weight (w_r) would be the number of individuals in the Red group divided by the variance of the Red group, the Green group’s weight (w_g) would be the number of individuals in the Green group divided by the variance of the Green group, and the Blue group’s weight (w_b) would be the number of individuals in the Blue group divided by the variance of the Blue group.

Notice: All the groups have the same number of participants, which makes calculating Welch a little easier. Also note, the Green and Blue groups have the same variance.

$$w_r = \frac{n_r}{S_r^2} = \frac{4}{2.92} = 1.37$$

$$w_g = \frac{n_g}{S_g^2} = \frac{4}{1.66} = 2.41$$

$$w_b = \frac{n_b}{S_b^2} = \frac{4}{1.66} = 2.41$$

Next, we need to calculate the Weighted Grand Mean (\bar{X}') and the table below helps clarify how to do that.

Table 5: Calculation of the Weighted Grand Mean Table

Group (j)	Weight (w_j)	Mean (\bar{X}_j)	Weight times Mean ($w_j * \bar{X}_j$)
Red	1.37	22.75	31.1675
Green	2.41	11.50	27.7150
Blue	2.41	17.50	42.1750
	$\Sigma = 6.19$		$\Sigma = 101.0575$

Next, we need to calculate the Weighted Grand Mean (\bar{X}'), which again we can express the formula in more familiar terms.

$$\bar{X}' = \frac{\sum (w_k \bar{X}_k)}{\sum w_k} = \frac{\sum (w_j \bar{X}_j)}{\sum w_j}$$

For our example we can use the information collected from the table above (Table 5):

$$\bar{X}' = \frac{\sum (w_j \bar{X}_j)}{\sum w_j} = \frac{101.0575}{6.19} = 16.325928 = 16.33$$

Next, calculate the F or rather F'' as it is the correct symbol under Welch's procedure.

$$F'' = \frac{\frac{\sum w_k (\bar{X}_k - \bar{X}')^2}{k-1}}{1 + \left[\frac{2(k-2)}{k^2-1} \right] * \left[\sum \left(\frac{1}{n_k-1} \right) \left(1 - \frac{w_k}{\sum w_k} \right)^2 \right]}$$

Again, it makes the formula less intimidating if we re-express it using the common (j) subscript. We will work downward rather than to the right, because the formula is so long and contains several summations (\sum).

$$F'' = \frac{\frac{\sum w_j (\bar{X}_j - \bar{X}')^2}{k-1}}{1 + \left[\frac{2(k-2)}{k^2-1} \right] * \left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right]}$$

Breaking the formula up into segments or parts, makes it less likely we will make a mistake. So, first focus on the very top line equation (of the equation) which is “the sum of, each group's weight multiplied times each group's mean minus the weighted grand mean, squared...then divide by the number of groups minus one”. All of which sounds a heck-of-a-lot like MS_b if you ask me. Remember; ANOVA (F) is a ratio of a between groups variance estimate (MS_b) divided by a within groups variance estimate (MS_w).

$$\frac{\sum w_j (\bar{X}_j - \bar{X}')^2}{k-1} = \frac{[1.37(22.75-16.33)^2] + [2.41(11.50-16.33)^2] + [2.41(17.50-16.33)^2]}{3-1}$$

$$\frac{\sum w_j (\bar{X}_j - \bar{X}')^2}{k-1} = \frac{[1.37(6.42)^2] + [2.41(-4.83)^2] + [2.41(1.17)^2]}{2}$$

$$\frac{\sum w_j (\bar{X}_j - \bar{X}')^2}{k-1} = \frac{1.37(41.2164) + 2.41(23.3289) + 2.41(1.3689)}{2}$$

$$\frac{\sum w_j (\bar{X}_j - \bar{X}')^2}{k-1} = \frac{56.47 + 56.22 + 3.30}{2}$$

$$\frac{\sum w_j (\bar{X}_j - \bar{X}')^2}{k-1} = \frac{115.99}{2}$$

$$\frac{\sum w_j (\bar{X}_j - \bar{X}')^2}{k-1} = 57.995$$

Next, work on the first half of the bottom line equation (of the equation). Starting with the first (easiest) part, then tackle the summation.

$$1 + \left[\frac{2(k-2)}{k^2-1} \right] * \left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right] = 1 + \left[\frac{2(3-2)}{3^2-1} \right] * \left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right]$$

$$1 + \left[\frac{2(k-2)}{k^2-1} \right] * \left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right] = 1 + \left[\frac{2(1)}{9-1} \right] * \left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right]$$

$$1 + \left[\frac{2(k-2)}{k^2-1} \right] * \left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right] = 1 + \left[\frac{2}{8} \right] * \left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right]$$

$$1 + \left[\frac{2(k-2)}{k^2-1} \right] * \left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right] = 1 + [0.25] * \left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right]$$

Now isolate and work on the summation then come back to put it all together. Remember, $n_j = 4$ for all three groups which will not always be the case.

$$\left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right] = \left(\frac{1}{4-1} \right) \left(1 - \frac{1.37}{6.19} \right)^2 + \left(\frac{1}{4-1} \right) \left(1 - \frac{2.41}{6.19} \right)^2 + \left(\frac{1}{4-1} \right) \left(1 - \frac{2.41}{6.19} \right)^2$$

$$\left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right] = \left(\frac{1}{3} \right) (1 - .221)^2 + \left(\frac{1}{3} \right) (1 - .389)^2 + \left(\frac{1}{3} \right) (1 - .389)^2$$

$$\left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right] = .333 (.779)^2 + .333 (.611)^2 + .333 (.611)^2$$

$$\left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right] = .333 (.6068) + .333 (.3733) + .333 (.3733)$$

$$\left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right] = .2021 + .1243 + .1243$$

$$\left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right] = .4507$$

Now we can put the elements back together into a relatively easy equation.

$$F'' = \frac{\frac{\sum w_j (\bar{X}_j - \bar{X}')^2}{k-1}}{1 + \left[\frac{2(k-2)}{k^2-1} \right] * \left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right]} = \frac{57.99}{1 + [0.25] * [0.4507]}$$

$$F'' = \frac{\frac{\sum w_j (\bar{X}_j - \bar{X}')^2}{k-1}}{1 + \left[\frac{2(k-2)}{k^2-1} \right] * \left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right]} = \frac{57.99}{1 + .112675}$$

$$F'' = \frac{\frac{\sum w_j (\bar{X}_j - \bar{X}')^2}{k-1}}{1 + \left[\frac{2(k-2)}{k^2-1} \right] * \left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right]} = \frac{57.99}{1.112675}$$

$$F'' = \frac{\frac{\sum w_j (\bar{X}_j - \bar{X}')^2}{k-1}}{1 + \left[\frac{2(k-2)}{k^2-1} \right] * \left[\sum \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2 \right]} = 52.117644$$

$$F'' = 52.12$$

And the giant falls...

Next, calculate the df'

$$df' = \frac{k^2 - 1}{3 \sum \left(\frac{1}{n_k - 1} \right) \left(1 - \frac{w_k}{\sum w_k} \right)^2}$$

Again, we can make this formula more readable using the familiar (j) subscript.

$$df' = \frac{k^2 - 1}{3 \sum \left(\frac{1}{n_j - 1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2}$$

Much of what is below is similar to what was done above (you will see some familiar numbers).

$$df' = \frac{3^2 - 1}{3 \sum \left(\frac{1}{n_j - 1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2}$$

$$df' = \frac{9 - 1}{3 \sum \left(\frac{1}{n_j - 1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2}$$

$$df' = \frac{8}{3 \sum \left(\frac{1}{n_j - 1} \right) \left(1 - \frac{w_j}{\sum w_j} \right)^2}$$

$$df' = \frac{8}{3 \left[\left(\frac{1}{4-1} \right) \left(1 - \frac{1.37}{6.19} \right)^2 + \left(\frac{1}{4-1} \right) \left(1 - \frac{2.41}{6.19} \right)^2 + \left(\frac{1}{4-1} \right) \left(1 - \frac{2.41}{6.19} \right)^2 \right]}$$

$$df' = \frac{8}{3 \left[\left(\frac{1}{3} \right) (1 - .221)^2 + \left(\frac{1}{3} \right) (1 - .389)^2 + \left(\frac{1}{3} \right) (1 - .389)^2 \right]}$$

$$df' = \frac{8}{3 [.333(.779)^2 + .333(.611)^2 + .333(.611)^2]}$$

$$df' = \frac{8}{3 [.333(.6068) + .333(.3733) + .333(.3733)]}$$

$$df' = \frac{8}{3 [.2021 + .1243 + .1243]}$$

$$df' = \frac{8}{3 [.4507]}$$

$$df' = \frac{8}{1.3521}$$

$$df' = 5.9167221$$

$$df' = 5.92$$

Now use $df' = 6$ and $k - 1 = df_b = 3 - 1 = 2$ to find the critical value in the standard F table using df_b as the numerator and df' as the denominator, with a significance level of 0.05. All of which gives us $F_{crit} = 5.14$

http://faculty.vassar.edu/lowry/apx_d.html

So, finally; we see that $F''_{calc} = 52.12 > 5.14 = F''_{crit}$ and therefore, we can reject the null hypothesis and conclude there was a significant difference among the three font color groups in terms of the number of words recalled.