Module 9: One-Way Analysis of Variance (ANOVA)

Jon Starkweather, PhD
jonathan.starkweather@unt.edu
Consultant
Research and **S**tatistical **S**upport

UNT UNIVERSITY OF NORTH TEXAS
Discover the power of ideas.

Introduction to Statistics for the Social Sciences

RSS
Research and Statistical Support

The RSS short courses

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`

RSS
Research and Statistical Support

Theoretical Applicability

- The one-way ANOVA is an extension of the Independent Samples *t* Test.

Theoretical Applicability

- The one-way ANOVA is an extension of the Independent Samples *t* Test.
- Using the one-way ANOVA allows us to analyze more than two groups of one independent variable on one dependent variable.

Theoretical Applicability

- The one-way ANOVA is an extension of the Independent Samples *t* Test.
- Using the one-way ANOVA allows us to analyze more than two groups of one independent variable on one dependent variable.
- The ANOVA family of analysis (all use the *F* distribution) starts with the one-way situation described just above. However, the general principles and many of the equations are applicable with more than one independent variable.

RSS
Research and Statistical Support

Theoretical Applicability

- The one-way ANOVA is an extension of the Independent Samples *t* Test.
- Using the one-way ANOVA allows us to analyze more than two groups of one independent variable on one dependent variable.
- The ANOVA family of analysis (all use the *F* distribution) starts with the one-way situation described just above. However, the general principles and many of the equations are applicable with more than one independent variable.
  - In future modules, we will discuss other analysis of the ANOVA family, such as repeated measures and factorial ANOVA, which features more than one independent variable.

RSS
Research and Statistical Support

Practical Applicability

- 1 IV (with $>$ 2 groups) & 1 interval/ratio DV.

Practical Applicability

- 1 IV (with $>$ 2 groups) & 1 interval/ratio DV.
- The benefit: Can test not only a treatment vs. a control group, but more than one treatment *and* a control group.

Practical Applicability

- 1 IV (with $>$ 2 groups) & 1 interval/ratio DV.
- The benefit: Can test not only a treatment vs. a control group, but more than one treatment *and* a control group.
- Some Examples:

Practical Applicability

- 1 IV (with $>$ 2 groups) & 1 interval/ratio DV.
- The benefit: Can test not only a treatment vs. a control group, but more than one treatment *and* a control group.
- Some Examples:
  - Comparing taste satisfaction ratings of Coke, Pepsi, Mountain Dew, and Dr. Pepper.

RSS
Research and Statistical Support

Practical Applicability

- 1 IV (with $>$ 2 groups) & 1 interval/ratio DV.
- The benefit: Can test not only a treatment vs. a control group, but more than one treatment *and* a control group.
- Some Examples:
  - Comparing taste satisfaction ratings of Coke, Pepsi, Mountain Dew, and Dr. Pepper.
  - Comparing homophobia scores of persons who are Atheist, Christian, Jewish, or Muslim.

RSS
Research and Statistical Support

Practical Applicability

- 1 IV (with $>$ 2 groups) & 1 interval/ratio DV.
- The benefit: Can test not only a treatment vs. a control group, but more than one treatment *and* a control group.
- Some Examples:
  - Comparing taste satisfaction ratings of Coke, Pepsi, Mountain Dew, and Dr. Pepper.
  - Comparing homophobia scores of persons who are Atheist, Christian, Jewish, or Muslim.
  - Comparing the stimulant effects (in terms of alertness) of Cocaine, Crystal Meth., Coffee, Red Bull, Mountain Dew, and Placebo.

RSS
Research and Statistical Support

Assumptions

- As with the Independent Samples *t* Test, if our assumptions are not met, we may have difficulty applying our results to the populations of interest or, we may not have any faith in the actual result of the test (i.e. invalid).

Assumptions

- As with the Independent Samples *t* Test, if our assumptions are not met, we may have difficulty applying our results to the populations of interest or, we may not have any faith in the actual result of the test (i.e. invalid).
- Primary assumptions of one-way ANOVA are exactly as they were for Independent Samples *t* Test.

Assumptions

- As with the Independent Samples *t* Test, if our assumptions are not met, we may have difficulty applying our results to the populations of interest or, we may not have any faith in the actual result of the test (i.e. invalid).
- Primary assumptions of one-way ANOVA are exactly as they were for Independent Samples *t* Test.
  - Normality of the population distributions (i.e. each population should be normally distributed).

RSS
Research and Statistical Support

Assumptions

- As with the Independent Samples *t* Test, if our assumptions are not met, we may have difficulty applying our results to the populations of interest or, we may not have any faith in the actual result of the test (i.e. invalid).
- Primary assumptions of one-way ANOVA are exactly as they were for Independent Samples *t* Test.
  - Normality of the population distributions (i.e. each population should be normally distributed).
  - Homogeneity of Variance of the population distributions (i.e. each population's variance should be similar).

Assumptions

- As with the Independent Samples *t* Test, if our assumptions are not met, we may have difficulty applying our results to the populations of interest or, we may not have any faith in the actual result of the test (i.e. invalid).
- Primary assumptions of one-way ANOVA are exactly as they were for Independent Samples *t* Test.
  - Normality of the population distributions (i.e. each population should be normally distributed).
  - Homogeneity of Variance of the population distributions (i.e. each population's variance should be similar).
  - Independence of Observations; meaning, each score should be independent of the other scores in its population (i.e. no twins in the same group).

RSS
Research and Statistical Support

Assumptions

- As with the Independent Samples *t* Test, if our assumptions are not met, we may have difficulty applying our results to the populations of interest or, we may not have any faith in the actual result of the test (i.e. invalid).
- Primary assumptions of one-way ANOVA are exactly as they were for Independent Samples *t* Test.
  - Normality of the population distributions (i.e. each population should be normally distributed).
  - Homogeneity of Variance of the population distributions (i.e. each population's variance should be similar).
  - Independence of Observations; meaning, each score should be independent of the other scores in its population (i.e. no twins in the same group).
- As with the *t* test situation, we evaluate these assumptions (of the populations) using the samples as representatives.

RSS
Research and Statistical Support

A quick *note* about notation

- Recall from early on, we use subscripts to identify a particular object within a series or collection of objects.

A quick *note* about notation

- Recall from early on, we use subscripts to identify a particular object within a series or collection of objects.
- Here, we will be working extensively with subscripts so it is important to review what the subscripts generally mean.

A quick *note* about notation

- Recall from early on, we use subscripts to identify a particular object within a series or collection of objects.
- Here, we will be working extensively with subscripts so it is important to review what the subscripts generally mean.
  - On the next few slides you will see a great deal of subscript letters, most are explained on the slide in which they appear first.

A quick *note* about notation

- Recall from early on, we use subscripts to identify a particular object within a series or collection of objects.
- Here, we will be working extensively with subscripts so it is important to review what the subscripts generally mean.
  - On the next few slides you will see a great deal of subscript letters, most are explained on the slide in which they appear first.
- However, recall we use i and j as identifiers of row and column respectively among a table of data (or electronic data file) such that: X_{ij}

RSS
Research and Statistical Support

A quick *note* about notation

- Recall from early on, we use subscripts to identify a particular object within a series or collection of objects.
- Here, we will be working extensively with subscripts so it is important to review what the subscripts generally mean.
  - On the next few slides you will see a great deal of subscript letters, most are explained on the slide in which they appear first.
- However, recall we use i and j as identifiers of row and column respectively among a table of data (or electronic data file) such that: X_{ij}
  - For example, we have a score: $X_{23} = 17$ which indicates this score is located at the second row and the third column; or we could say that 17 is the second score of the third variable because, data are generally arranged in a matrix with the variables as column headings and rows representing cases (i.e. participants).

RSS
Research and Statistical Support

ANOVA

- The statistical procedure for testing variation among the means of more than two groups is called Analysis of Variance (ANOVA).

ANOVA

- The statistical procedure for testing variation among the means of more than two groups is called Analysis of Variance (ANOVA).
- NHST with ANOVA is about whether the means of the samples differ more than you would expect if the null hypothesis were true.

ANOVA

- The statistical procedure for testing variation among the means of more than two groups is called Analysis of Variance (ANOVA).
- NHST with ANOVA is about whether the means of the samples differ more than you would expect if the null hypothesis were true.
- This involves analyzing the *variation* **between** the samples (i.e. between the means) and the *variation* **within** each sample.

ANOVA

- The statistical procedure for testing variation among the means of more than two groups is called Analysis of Variance (ANOVA).
- NHST with ANOVA is about whether the means of the samples differ more than you would expect if the null hypothesis were true.
- This involves analyzing the *variation* **between** the samples (i.e. between the means) and the *variation* **within** each sample.
  - We analyze the variance to test if there is a significant difference (somewhere) among the means.

RSS
Research and Statistical Support

ANOVA

- The statistical procedure for testing variation among the means of more than two groups is called Analysis of Variance (ANOVA).
- NHST with ANOVA is about whether the means of the samples differ more than you would expect if the null hypothesis were true.
- This involves analyzing the *variation* **between** the samples (i.e. between the means) and the *variation* **within** each sample.
  - We analyze the variance to test if there is a significant difference (somewhere) among the means.
  - Stated another way, we seek to discover *mean differences* by analyzing the variances.

RSS
Research and Statistical Support

ANOVA

- The statistical procedure for testing variation among the means of more than two groups is called Analysis of Variance (ANOVA).
- NHST with ANOVA is about whether the means of the samples differ more than you would expect if the null hypothesis were true.
- This involves analyzing the *variation* **between** the samples (i.e. between the means) and the *variation* **within** each sample.
  - We analyze the variance to test if there is a significant difference (somewhere) among the means.
  - Stated another way, we seek to discover *mean differences* by analyzing the variances.
- Some authors (textbooks) use 'treatment' instead of 'between' and 'error' instead of 'within'.

RSS
Research and Statistical Support

Variance

- Recall that conceptually, variance is the sums of squared deviations (SOS) divided by the degrees of freedom (df): $S^2 = \sum \left(X - \overline{X}\right)^2 / n - 1 = SOS/df$

Variance

- Recall that conceptually, variance is the sums of squared deviations (SOS) divided by the degrees of freedom (df): $S^2 = \sum \left(X - \overline{X}\right)^2 / n - 1 = SOS/df$
- In ANOVA, we have to calculate the variance for **between** the groups and the variance **within** the groups.

Variance

- Recall that conceptually, variance is the sums of squared deviations (SOS) divided by the degrees of freedom (df):
  $S^2 = \sum \left(X - \overline{X}\right)^2 / n - 1 = SOS/df$
- In ANOVA, we have to calculate the variance for **between** the groups and the variance **within** the groups.
  $MS_b = SOS_b/df_b$

  $MS_w = SOS_w/df_w$

Variance

- Recall that conceptually, variance is the sums of squared deviations (SOS) divided by the degrees of freedom (df):
  $S^2 = \sum \left(X - \overline{X}\right)^2 / n - 1 = SOS/df$
- In ANOVA, we have to calculate the variance for **between** the groups and the variance **within** the groups.
  $MS_b = SOS_b / df_b$

  $MS_w = SOS_w / df_w$
- The ratio of between groups variance to within group variance **is** our F statistic (F_{calc}).

MS_b

- The between groups variance: $MS_b = SOS_b / df_b$

MS_b

- The between groups variance: $MS_b = SOS_b / df_b$
- First, calculate the mean of each group: $\overline{X}_j$

MS_b

- The between groups variance: $MS_b = SOS_b/df_b$
- First, calculate the mean of each group: $\overline{X}_j$
- Then compute the *grand mean*: Sum of the means, divided by the number of means: $\overline{X}_{..} = \sum \overline{X}_j/k$, where k is the number of means (i.e., groups).

MS_b

- The between groups variance: $MS_b = SOS_b / df_b$
- First, calculate the mean of each group: $\overline{X}_j$
- Then compute the *grand mean*: Sum of the means, divided by the number of means: $\overline{X}_{..} = \sum \overline{X}_j / k$, where k is the number of means (i.e., groups).
- Then, $SOS_b = \sum \left[n_j \left(\overline{X}_j - \overline{X}_{..} \right)^2 \right]$ where n_j is the number of scores in the jth group.

MS_b

- The between groups variance: $MS_b = SOS_b / df_b$
- First, calculate the mean of each group: $\overline{X}_j$
- Then compute the *grand mean*: Sum of the means, divided by the number of means: $\overline{X}_{..} = \sum \overline{X}_j / k$, where k is the number of means (i.e., groups).
- Then, $SOS_b = \sum \left[n_j \left(\overline{X}_j - \overline{X}_{..} \right)^2 \right]$ where n_j is the number of scores in the jth group.
  - The sum of, for each group; the number of scores in the group times the group mean minus the grand mean squared.

MS_b

- The between groups variance: $MS_b = SOS_b/df_b$
- First, calculate the mean of each group: $\overline{X}_j$
- Then compute the *grand mean*: Sum of the means, divided by the number of means: $\overline{X}_{..} = \sum \overline{X}_j/k$, where k is the number of means (i.e., groups).
- Then, $SOS_b = \sum \left[n_j \left(\overline{X}_j - \overline{X}_{..} \right)^2 \right]$ where n_j is the number of scores in the jth group.
  - The sum of, for each group; the number of scores in the group times the group mean minus the grand mean squared.
- df_b = number of groups minus 1.
  $df_b = k - 1$

MS_w

- MS_w is also called MS_{error} or simply MS_e

MS_w

- MS_w is also called MS_{error} or simply MS_e
- The within groups variance: $MS_w = SOS_w / df_w$

MS_w

- MS_w is also called MS_{error} or simply MS_e
- The within groups variance: $MS_w = SOS_w / df_w$
- First, calculate the $SOS_w = \sum \left(X_{ij} - \overline{X}_j \right)^2$

MS_w

- MS_w is also called MS_{error} or simply MS_e
- The within groups variance: $MS_w = SOS_w / df_w$
- First, calculate the $SOS_w = \sum \left(X_{ij} - \overline{X}_j \right)^2$
- Then calculate the $df_w = n - k$

MS_w

- MS_w is also called MS_{error} or simply MS_e
- The within groups variance: $MS_w = SOS_w / df_w$
- First, calculate the $SOS_w = \sum \left(X_{ij} - \overline{X}_j \right)^2$
- Then calculate the $df_w = n - k$
- Then of course; $MS_w = SOS_w / df_w$

Calculate F

- Simply divide: $MS_b/MS_w = F_{calc}$

Calculate F

- Simply divide: $MS_b/MS_w = F_{calc}$
- Use the df_b and df_w to look up the critical score on the F distribution table for the significance level you are using (typically 0.05).

Calculate F

- Simply divide: $MS_b/MS_w = F_{calc}$
- Use the df_b and df_w to look up the critical score on the F distribution table for the significance level you are using (typically 0.05).
- The 'numerator degrees of freedom' is df_b while the 'denominator degrees of freedom' is df_w

Total *SOS* and *df*

- SOS_t is the sum of the squared deviations of each individual score around the grand mean.

Total *SOS* and *df*

- SOS_t is the sum of the squared deviations of each individual score around the grand mean.
  $SOS_t = \sum \left(X - \overline{X}_{..} \right)^2$

Total *SOS* and *df*

- SOS_t is the sum of the squared deviations of each individual score around the grand mean.
  $SOS_t = \sum \left(X - \overline{X}_{..} \right)^2$
- df_t is the total number of scores minus 1.

Total *SOS* and *df*

- SOS_t is the sum of the squared deviations of each individual score around the grand mean.
  $SOS_t = \sum \left(X - \overline{X}_{..} \right)^2$
- df_t is the total number of scores minus 1.
  $df_t = n - 1$

Total *SOS* and *df*

- SOS_t is the sum of the squared deviations of each individual score around the grand mean.
  $SOS_t = \sum \left(X - \overline{X}_{..} \right)^2$
- df_t is the total number of scores minus 1.
  $df_t = n - 1$
- It is not necessary to calculate the totals, but often when hand calculating they are useful for checking previous calculations since;

Total *SOS* and *df*

- SOS_t is the sum of the squared deviations of each individual score around the grand mean.
  $SOS_t = \sum \left(X - \overline{X}_{..} \right)^2$
- df_t is the total number of scores minus 1.
  $df_t = n - 1$
- It is not necessary to calculate the totals, but often when hand calculating they are useful for checking previous calculations since;
  $SOS_t = SOS_b + SOS_w$ and $df_t = df_b + df_w$

ANOVA Summary Table

Table 1: Example ANOVA Summary Table

Source	SOS	df	MS	F_{calc}
Between	64	2	32	11.59
Within	58	21	2.76	
Total	122	23		

Between and Within should add to Total, if not there was an error in computation.

Recall, $SOS/df = MS$ and $MS_b/MS_w = F_{calc}$

Another way to *think* about ANOVA

- Previously we said: The ratio of between groups variance to within groups variance is our F statistic.
  $F = \frac{MS_b}{MS_w}$

Another way to *think* about ANOVA

- Previously we said: The ratio of between groups variance to within groups variance is our F statistic.
  $F = \frac{MS_b}{MS_w}$
- This could be thought of as: The ratio of differences based on treatment *and chance*, to the differences based on chance alone, or as the ratio of treatment and error to just error.

Another way to *think* about ANOVA

- Previously we said: The ratio of between groups variance to within groups variance is our F statistic.
  $F = \frac{MS_b}{MS_w}$
- This could be thought of as: The ratio of differences based on treatment *and chance*, to the differences based on chance alone, or as the ratio of treatment and error to just error.
  $F = \frac{Treatment\ \ Effects + Error}{Error}$

RSS
Research and Statistical Support

Another way to *think* about ANOVA

- Previously we said: The ratio of between groups variance to within groups variance is our F statistic.
  $F = \frac{MS_b}{MS_w}$
- This could be thought of as: The ratio of differences based on treatment *and chance*, to the differences based on chance alone, or as the ratio of treatment and error to just error.
  $F = \frac{Treatment\ \ Effects + Error}{Error}$
- Simple math principles leave just the treatment effects by 'crossing-out' the error terms.

Another way to *think* about ANOVA

- Previously we said: The ratio of between groups variance to within groups variance is our F statistic.
  $F = \frac{MS_b}{MS_w}$
- This could be thought of as: The ratio of differences based on treatment *and chance*, to the differences based on chance alone, or as the ratio of treatment and error to just error.
  $F = \frac{Treatment\ Effects + Error}{Error}$
- Simple math principles leave just the treatment effects by 'crossing-out' the error terms.
  $F = \frac{Treatment\ differences + Individual\ differences}{Individual\ differences}$

RSS
Research and Statistical Support

Logic behind NHST One-way ANOVA

- If H_0 is true, then all the groups will be very similar; their means will be close to one another.

Logic behind NHST One-way ANOVA

- If H_0 is true, then all the groups will be very similar; their means will be close to one another.
  - Between groups variance (MS_b) will be small.

Logic behind NHST One-way ANOVA

- If H_0 is true, then all the groups will be very similar; their means will be close to one another.
  - Between groups variance (MS_b) will be small.
  - If between groups variance is small, then F will be small as well; and likely not significant ($p < .05$).

Logic behind NHST One-way ANOVA

- If H_0 is true, then all the groups will be very similar; their means will be close to one another.
  - Between groups variance (MS_b) will be small.
  - If between groups variance is small, then F will be small as well; and likely not significant ($p < .05$).
- If H_0 is not true, then all the groups will be very different.

Logic behind NHST One-way ANOVA

- If H_0 is true, then all the groups will be very similar; their means will be close to one another.
  - Between groups variance (MS_b) will be small.
  - If between groups variance is small, then F will be small as well; and likely not significant ($p < .05$).
- If H_0 is not true, then all the groups will be very different.
  - Between groups variance and F will be larger; and likely significant ($p < .05$).

RSS
Research and Statistical Support

Logic behind NHST One-way ANOVA

- If H_0 is true, then all the groups will be very similar; their means will be close to one another.
  - Between groups variance (MS_b) will be small.
  - If between groups variance is small, then F will be small as well; and likely not significant ($p < .05$).
- If H_0 is not true, then all the groups will be very different.
  - Between groups variance and F will be larger; and likely significant ($p < .05$).
- Conversely; if within groups variance is small, then F will be large and if within groups variance is large, F will be small.
- Remember; $F = MS_b/MS_w$

RSS
Research and Statistical Support

Example Study Description

- Study of the effects of font color on recall. We were interested in determining which color font contributes to the best recall of words presented.

Example Study Description

- Study of the effects of font color on recall. We were interested in determining which color font contributes to the best recall of words presented.
- Students (n = 12) were randomly drawn from the UNT undergraduate research pool and randomly assigned to one of three stimuli groups:

Example Study Description

- Study of the effects of font color on recall. We were interested in determining which color font contributes to the best recall of words presented.
- Students (n = 12) were randomly drawn from the UNT undergraduate research pool and randomly assigned to one of three stimuli groups:
  - Red (n = 4), Green (n = 4), and Blue (n = 4).

Example Study Description

- Study of the effects of font color on recall. We were interested in determining which color font contributes to the best recall of words presented.
- Students (n = 12) were randomly drawn from the UNT undergraduate research pool and randomly assigned to one of three stimuli groups:
  - Red (n = 4), Green (n = 4), and Blue (n = 4).
- Participants were then instructed to watch a presentation of 30 words while trying to memorize each word, each presented for 45 seconds.

Example Study Description

- Study of the effects of font color on recall. We were interested in determining which color font contributes to the best recall of words presented.
- Students (n = 12) were randomly drawn from the UNT undergraduate research pool and randomly assigned to one of three stimuli groups:
  - Red (n = 4), Green (n = 4), and Blue (n = 4).
- Participants were then instructed to watch a presentation of 30 words while trying to memorize each word, each presented for 45 seconds.
- Participants were then instructed to recall as many of the words as possible.

RSS
Research and Statistical Support

Raw data

- Three colors are used to present words.

Raw data

- Three colors are used to present words.
- Participants are told to memorize the words.

Raw data

- Three colors are used to present words.
- Participants are told to memorize the words.
- There are 30 words, each presented for 45 seconds.

Raw data

- Three colors are used to present words.
- Participants are told to memorize the words.
- There are 30 words, each presented for 45 seconds.
- Look carefully at the scores.

Raw data

- Three colors are used to present words.
- Participants are told to memorize the words.
- There are 30 words, each presented for 45 seconds.
- Look carefully at the scores.
- What do you notice?

Font Color	Words Recalled
Red	25
Red	22
Red	23
Red	21
Green	10
Green	12
Green	11
Green	13
Blue	18
Blue	17
Blue	19
Blue	16

RSS
Research and Statistical Support

Data and Descriptive Statistics

	Red Group	Green Group	Blue Group	Totals
	25	10	18	
	22	12	17	
	23	11	19	
	21	13	16	
$\sum X =$	91	46	70	207
$n =$	4	4	4	$n_t = 12$
$\overline{X} =$	22.75	11.5	17.5	$\overline{X}_{..} = 17.25$
$S =$	1.71	1.29	1.29	
$S^2 =$	2.92	1.66	1.66	6.24
$df =$	3	3	3	$df_t = 11$

NHST Step 1

- Identify the populations and restate the research question as H_0 and H_1

NHST Step 1

- Identify the populations and restate the research question as H_0 and H_1
- Population 1: UNT undergraduates who are presented words with red font.

NHST Step 1

- Identify the populations and restate the research question as H_0 and H_1
- Population 1: UNT undergraduates who are presented words with red font.
- Population 2: UNT undergraduates who are presented words with green font.

NHST Step 1

- Identify the populations and restate the research question as H_0 and H_1
- Population 1: UNT undergraduates who are presented words with red font.
- Population 2: UNT undergraduates who are presented words with green font.
- Population 3: UNT undergraduates who are presented words with blue font.

NHST Step 1

- Identify the populations and restate the research question as H_0 and H_1
- Population 1: UNT undergraduates who are presented words with red font.
- Population 2: UNT undergraduates who are presented words with green font.
- Population 3: UNT undergraduates who are presented words with blue font.
- Null Hypothesis: $H_0 : \mu_1 = \mu_2 = \mu_3$

NHST Step 1

- Identify the populations and restate the research question as H_0 and H_1
- Population 1: UNT undergraduates who are presented words with red font.
- Population 2: UNT undergraduates who are presented words with green font.
- Population 3: UNT undergraduates who are presented words with blue font.
- Null Hypothesis: $H_0 : \mu_1 = \mu_2 = \mu_3$
- Alternative Hypothesis: $H_1 : \mu_1 \neq \mu_2 \neq \mu_3$

RSS
Research and Statistical Support

NHST Step 1

- Identify the populations and restate the research question as H_0 and H_1
- Population 1: UNT undergraduates who are presented words with red font.
- Population 2: UNT undergraduates who are presented words with green font.
- Population 3: UNT undergraduates who are presented words with blue font.
- Null Hypothesis: $H_0 : \mu_1 = \mu_2 = \mu_3$
- Alternative Hypothesis: $H_1 : \mu_1 \neq \mu_2 \neq \mu_3$
  - Note: The Null Hypothesis could be rejected under several conditions.

NHST Step 1

- Identify the populations and restate the research question as H_0 and H_1
- Population 1: UNT undergraduates who are presented words with red font.
- Population 2: UNT undergraduates who are presented words with green font.
- Population 3: UNT undergraduates who are presented words with blue font.
- Null Hypothesis: $H_0 : \mu_1 = \mu_2 = \mu_3$
- Alternative Hypothesis: $H_1 : \mu_1 \neq \mu_2 \neq \mu_3$
  - Note: The Null Hypothesis could be rejected under several conditions.
  - Any significant difference among the three means.

RSS
Research and Statistical Support

NHST Step 2

- Determine the characteristics of the comparison distribution.

NHST Step 2

- Determine the characteristics of the comparison distribution.
- The comparison distribution is the F distribution with 2 and 9 degrees of freedom:

NHST Step 2

- Determine the characteristics of the comparison distribution.
- The comparison distribution is the F distribution with 2 and 9 degrees of freedom:
  - $df_b = k - 1 = 3 - 1 = 2$

NHST Step 2

- Determine the characteristics of the comparison distribution.
- The comparison distribution is the F distribution with 2 and 9 degrees of freedom:
  - $df_b = k - 1 = 3 - 1 = 2$
  - $df_w = n_t - k = 12 - 3 = 9$

NHST Step 3

- Determine the cutoff sample score on the comparison distribution at which H_0 should be rejected.

NHST Step 3

- Determine the cutoff sample score on the comparison distribution at which H_0 should be rejected.
- Using the F distribution with a significance level of 0.05; and 2 (numerator) and 9 (denominator) degrees of freedom: $F_{crit} = 4.26$

NHST Step 3

- Determine the cutoff sample score on the comparison distribution at which H_0 should be rejected.
- Using the F distribution with a significance level of 0.05; and 2 (numerator) and 9 (denominator) degrees of freedom: $F_{crit} = 4.26$

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

NHST Step 4

- Determine your sample's score on the comparison distribution.

NHST Step 4

- Determine your sample's score on the comparison distribution.
  - Compute your sample statistic(s).

NHST Step 4

- Determine your sample's score on the comparison distribution.
  - Compute your sample statistic(s).
- Calculate *SOS*, *df*, *MS*, and F_{calc} for Between groups, Within groups, and Total (as a check of the other calculations) given the information provided above.

Calculate SOS_b and df_b

Sum of the n of each group times the squared deviations of each group mean from the grand mean.

$$SOS_b = \sum \left[n_j \left(\overline{X}_j - \overline{X}_{..} \right)^2 \right]$$

Calculate SOS_b and df_b

Sum of the n of each group times the squared deviations of each group mean from the grand mean.

$$SOS_b = \sum \left[n_j \left(\overline{X}_j - \overline{X}_{..} \right)^2 \right]$$

$$SOS_b = n_1 \left(\overline{X}_1 - \overline{X}_{..} \right)^2 + n_2 \left(\overline{X}_2 - \overline{X}_{..} \right)^2 + n_3 \left(\overline{X}_3 - \overline{X}_{..} \right)^2$$

Calculate SOS_b and df_b

Sum of the n of each group times the squared deviations of each group mean from the grand mean.

$$SOS_b = \sum \left[n_j \left(\overline{X}_j - \overline{X}_{..} \right)^2 \right]$$

$$SOS_b = n_1 \left(\overline{X}_1 - \overline{X}_{..} \right)^2 + n_2 \left(\overline{X}_2 - \overline{X}_{..} \right)^2 + n_3 \left(\overline{X}_3 - \overline{X}_{..} \right)^2$$

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

Calculate SOS_b and df_b

Sum of the n of each group times the squared deviations of each group mean from the grand mean.

$$SOS_b = \sum \left[n_j \left(\overline{X}_j - \overline{X}_{..} \right)^2 \right]$$

$$SOS_b = n_1 \left(\overline{X}_1 - \overline{X}_{..} \right)^2 + n_2 \left(\overline{X}_2 - \overline{X}_{..} \right)^2 + n_3 \left(\overline{X}_3 - \overline{X}_{..} \right)^2$$

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

$$df_b = k - 1 = 3 - 1 = 2$$

Calculate MS_b

Calculation of Mean-Square between, an estimate of the between groups variance:

Calculate MS_b

Calculation of Mean-Square between, an estimate of the between groups variance:

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

$$MS_b = \frac{253.50}{2}$$

$$MS_b = 126.75$$

Calculate SOS_w and df_w

The sum of the squared deviations of each group's score from its group mean.

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

Calculate SOS_w and df_w

The sum of the squared deviations of each group's score from its group mean.

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

$$SOS_w = \left(X_{11} - \overline{X}_1\right)^2 + \left(X_{21} - \overline{X}_1\right)^2 + \left(X_{31} - \overline{X}_1\right)^2 + \left(X_{41} - \overline{X}_1\right)^2 + \left(X_{12} - \overline{X}_2\right)^2 + \ldots + \left(X_{43} - \overline{X}_3\right)^2$$

Calculate SOS_w and df_w

The sum of the squared deviations of each group's score from its group mean.

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

$$SOS_w = \left(X_{11} - \overline{X}_1\right)^2 + \left(X_{21} - \overline{X}_1\right)^2 + \left(X_{31} - \overline{X}_1\right)^2 + \left(X_{41} - \overline{X}_1\right)^2 + \left(X_{12} - \overline{X}_2\right)^2 + \ldots + \left(X_{43} - \overline{X}_3\right)^2$$

$$SOS_w = (25 - 22.75)^2 + (22 - 22.75)^2 + (23 - 22.75)^2 + (21 - 22.75)^2 + (10 - 11.5)^2 + \ldots + (16 - 17.5)^2 = 18.75$$

Calculate SOS_w and df_w

The sum of the squared deviations of each group's score from its group mean.

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

$$SOS_w = \left(X_{11} - \overline{X}_1\right)^2 + \left(X_{21} - \overline{X}_1\right)^2 + \left(X_{31} - \overline{X}_1\right)^2 + \left(X_{41} - \overline{X}_1\right)^2 + \left(X_{12} - \overline{X}_2\right)^2 + \ldots + \left(X_{43} - \overline{X}_3\right)^2$$

$$SOS_w = (25 - 22.75)^2 + (22 - 22.75)^2 + (23 - 22.75)^2 + (21 - 22.75)^2 + (10 - 11.5)^2 + \ldots + (16 - 17.5)^2 = 18.75$$

$$df_w = n_t - k = 12 - 3 = 9$$

Calculate MS_W

Calculation of Mean-Square within (often called Mean-Square error), an estimate of the within groups variance:

Calculate MS_w

Calculation of Mean-Square within (often called Mean-Square error), an estimate of the within groups variance:

$$MS_w = \frac{SOS_w}{df_w}$$

$$MS_w = \frac{18.75}{9}$$

$$MS_w = 2.083$$

Calculate SOS_t and df_t

The sum of the squared deviations of each score from the grand mean.

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

Calculate SOS_t and df_t

The sum of the squared deviations of each score from the grand mean.

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

$$SOS_t = \left(X_{11} - \overline{X}_{..}\right)^2 + \left(X_{21} - \overline{X}_{..}\right)^2 + \left(X_{31} - \overline{X}_{..}\right)^2 + \left(X_{41} - \overline{X}_{..}\right)^2 + \left(X_{12} - \overline{X}_{..}\right)^2 + \ldots + \left(X_{43} - \overline{X}_{..}\right)^2$$

Calculate SOS_t and df_t

The sum of the squared deviations of each score from the grand mean.

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

$$SOS_t = \left(X_{11} - \overline{X}_{..}\right)^2 + \left(X_{21} - \overline{X}_{..}\right)^2 + \left(X_{31} - \overline{X}_{..}\right)^2 + \left(X_{41} - \overline{X}_{..}\right)^2 + \left(X_{12} - \overline{X}_{..}\right)^2 + \ldots + \left(X_{43} - \overline{X}_{..}\right)^2$$

$$SOS_t = (25 - 17.25)^2 + (22 - 17.25)^2 + (23 - 17.25)^2 + (21 - 17.25)^2 + (10 - 17.25)^2 + \ldots (16 - 17.25)^2 = 272.25$$

Calculate SOS_t and df_t

The sum of the squared deviations of each score from the grand mean.

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

$$SOS_t = \left(X_{11} - \overline{X}_{..}\right)^2 + \left(X_{21} - \overline{X}_{..}\right)^2 + \left(X_{31} - \overline{X}_{..}\right)^2 + \left(X_{41} - \overline{X}_{..}\right)^2 + \left(X_{12} - \overline{X}_{..}\right)^2 + \ldots + \left(X_{43} - \overline{X}_{..}\right)^2$$

$$SOS_t = (25 - 17.25)^2 + (22 - 17.25)^2 + (23 - 17.25)^2 + (21 - 17.25)^2 + (10 - 17.25)^2 + \ldots (16 - 17.25)^2 = 272.25$$

$$df_t = n_t - 1 = 12 - 1 = 11$$

Build the ANOVA Summary Table and Calculate *F*

Table 2: NHST Example ANOVA Summary Table

Build the ANOVA Summary Table and Calculate *F*

Table 2: NHST Example ANOVA Summary Table

Source	*SOS*	*df*	*MS*	F_{calc}
Between	253.50	2	126.75	60.84
Within	18.75	9	2.083	
Total	272.25	11		

Build the ANOVA Summary Table and Calculate F

Table 2: NHST Example ANOVA Summary Table

Source	SOS	df	MS	F_{calc}
Between	253.50	2	126.75	60.84
Within	18.75	9	2.083	
Total	272.25	11		

Recall, $SOS/df = MS$ and $MS_b/MS_w = F_{calc}$

Build the ANOVA Summary Table and Calculate F

Table 2: NHST Example ANOVA Summary Table

Source	SOS	df	MS	F_{calc}
Between	253.50	2	126.75	60.84
Within	18.75	9	2.083	
Total	272.25	11		

Recall, $SOS/df = MS$ and $MS_b/MS_w = F_{calc}$

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

Notice that $SOS_b + SOS_w = SOS_t$ and $df_b + df_w = df_t$

Compare and make a decision.

- Decide whether to reject H_0

Compare and make a decision.

- Decide whether to reject H_0
- Because $F_{calc} = 60.84 > 4.26 = F_{crit}$

Compare and make a decision.

- Decide whether to reject H_0
- Because $F_{calc} = 60.84 > 4.26 = F_{crit}$

  We would reject the null hypothesis and conclude that there was a significant difference among the sample means. This study suggests that there is a difference among the three theoretical populations (each represented by a sample here).

Compare and make a decision.

- Decide whether to reject H_0
- Because $F_{calc} = 60.84 > 4.26 = F_{crit}$

  We would reject the null hypothesis and conclude that there was a significant difference among the sample means. This study suggests that there is a difference among the three theoretical populations (each represented by a sample here).
- However, we do not yet know where the significant difference is located, and there may be more than one significant difference among the three groups.

RSS
Research and Statistical Support

What's next?

- The one-way ANOVA only tells us if there was some significant difference among the groups.

What's next?

- The one-way ANOVA only tells us if there was some significant difference among the groups.
- Two possible ways of identifying where the significant difference(-s) lie.

What's next?

- The one-way ANOVA only tells us if there was some significant difference among the groups.
- Two possible ways of identifying where the significant difference(-s) lie.
- Planned Comparisons, which are *planned* prior to data collection (when designing the study).

What's next?

- The one-way ANOVA only tells us if there was some significant difference among the groups.
- Two possible ways of identifying where the significant difference(-s) lie.
- Planned Comparisons, which are *planned* prior to data collection (when designing the study).
  - Used to assess specific hypotheses about specific groups (e.g., Red $>$ Blue; Green $<$ Blue, etc.).

What's next?

- The one-way ANOVA only tells us if there was some significant difference among the groups.
- Two possible ways of identifying where the significant difference(-s) lie.
- Planned Comparisons, which are *planned* prior to data collection (when designing the study).
  - Used to assess specific hypotheses about specific groups (e.g., Red $>$ Blue; Green $<$ Blue, etc.).
- Post-hoc testing, which typically compares all possible pair-wise comparisons of the groups.

What's next?

- The one-way ANOVA only tells us if there was some significant difference among the groups.
- Two possible ways of identifying where the significant difference(-s) lie.
- Planned Comparisons, which are *planned* prior to data collection (when designing the study).
  - Used to assess specific hypotheses about specific groups (e.g., Red $>$ Blue; Green $<$ Blue, etc.).
- Post-hoc testing, which typically compares all possible pair-wise comparisons of the groups.
  - Traditionally, only done when the 'omnibus' F is significant.

RSS
Research and Statistical Support

What's next?

- The one-way ANOVA only tells us if there was some significant difference among the groups.
- Two possible ways of identifying where the significant difference(-s) lie.
- Planned Comparisons, which are *planned* prior to data collection (when designing the study).
  - Used to assess specific hypotheses about specific groups (e.g., Red $>$ Blue; Green $<$ Blue, etc.).
- Post-hoc testing, which typically compares all possible pair-wise comparisons of the groups.
  - Traditionally, only done when the 'omnibus' F is significant.
  - Reality; almost always done to explore pair-wise differences.

RSS
Research and Statistical Support

Planned Comparisons

- Also called: a-priori comparisons or planned contrasts.

Planned Comparisons

- Also called: a-priori comparisons or planned contrasts.
- Used to test specific group differences to evaluate specific hypotheses (stated prior to data collection).

Planned Comparisons

- Also called: a-priori comparisons or planned contrasts.
- Used to test specific group differences to evaluate specific hypotheses (stated prior to data collection).
- Two general types of planned comparisons.

Planned Comparisons

- Also called: a-priori comparisons or planned contrasts.
- Used to test specific group differences to evaluate specific hypotheses (stated prior to data collection).
- Two general types of planned comparisons.
  - Simple Comparisons (we'll stick to these here).

Planned Comparisons

- Also called: a-priori comparisons or planned contrasts.
- Used to test specific group differences to evaluate specific hypotheses (stated prior to data collection).
- Two general types of planned comparisons.
  - Simple Comparisons (we'll stick to these here).
    - Each comparison is used to compare two groups.

Planned Comparisons

- Also called: a-priori comparisons or planned contrasts.
- Used to test specific group differences to evaluate specific hypotheses (stated prior to data collection).
- Two general types of planned comparisons.
  - Simple Comparisons (we'll stick to these here).
    - Each comparison is used to compare two groups.
    - With only two groups $df_b = 1$ always, because $df_b = k - 1$

Planned Comparisons

- Also called: a-priori comparisons or planned contrasts.
- Used to test specific group differences to evaluate specific hypotheses (stated prior to data collection).
- Two general types of planned comparisons.
  - Simple Comparisons (we'll stick to these here).
    - Each comparison is used to compare two groups.
    - With only two groups $df_b = 1$ always, because $df_b = k - 1$
  - Complex Comparisons.

Planned Comparisons

- Also called: a-priori comparisons or planned contrasts.
- Used to test specific group differences to evaluate specific hypotheses (stated prior to data collection).
- Two general types of planned comparisons.
  - Simple Comparisons (we'll stick to these here).
    - Each comparison is used to compare two groups.
    - With only two groups $df_b = 1$ always, because $df_b = k - 1$
  - Complex Comparisons.
    - Each comparison is used to compare two or more groups.

Planned Comparisons

- Also called: a-priori comparisons or planned contrasts.
- Used to test specific group differences to evaluate specific hypotheses (stated prior to data collection).
- Two general types of planned comparisons.
  - Simple Comparisons (we'll stick to these here).
    - Each comparison is used to compare two groups.
    - With only two groups $df_b = 1$ always, because $df_b = k - 1$
  - Complex Comparisons.
    - Each comparison is used to compare two or more groups.
    - e.g., Orthogonal contrasts.

Simple Comparisons

- Suppose, we hypothesized (prior to data collection) that the Red group would recall more words than the Blue group and the Green group would recall less words than the Blue group.

Simple Comparisons

- Suppose, we hypothesized (prior to data collection) that the Red group would recall more words than the Blue group and the Green group would recall less words than the Blue group.
- To do any simple comparison, we need to calculate F for **that comparison**.

Simple Comparisons

- Suppose, we hypothesized (prior to data collection) that the Red group would recall more words than the Blue group and the Green group would recall less words than the Blue group.
- To do any simple comparison, we need to calculate F for **that comparison**.
- We need a new df_b (always = 1) and a new grand mean ($\overline{X}_{..}$) for **each comparison**.

Simple Comparisons

- Suppose, we hypothesized (prior to data collection) that the Red group would recall more words than the Blue group and the Green group would recall less words than the Blue group.
- To do any simple comparison, we need to calculate F for **that comparison**.
- We need a new df_b (always = 1) and a new grand mean ($\overline{X}_{..}$) for **each comparison**.
- But, we use the same SOS_w and df_w which leads to the same $MS_w = 2.083$

Simple Comparisons

- Suppose, we hypothesized (prior to data collection) that the Red group would recall more words than the Blue group and the Green group would recall less words than the Blue group.
- To do any simple comparison, we need to calculate F for **that comparison**.
- We need a new df_b (always = 1) and a new grand mean ($\overline{X}_{..}$) for **each comparison**.
- But, we use the same SOS_w and df_w which leads to the same $MS_w = 2.083$
- Then, for each comparison the usual steps are involved: $F = MS_b/MS_w$

Simple Comparisons

- Suppose, we hypothesized (prior to data collection) that the Red group would recall more words than the Blue group and the Green group would recall less words than the Blue group.
- To do any simple comparison, we need to calculate F for **that comparison**.
- We need a new df_b (always = 1) and a new grand mean ($\overline{X}_{..}$) for **each comparison**.
- But, we use the same SOS_w and df_w which leads to the same $MS_w = 2.083$
- Then, for each comparison the usual steps are involved: $F = MS_b/MS_w$
- You may wish to control for family-wise error rate inflation by dividing the significance level by the number of comparisons.

Simple Comparisons Example

Simple Comparisons Example

- Recall, we hypothesized (prior to data collection) the Red group would recall more words than the Blue group and the Green group would recall less words than the Blue group.

Simple Comparisons Example

- Recall, we hypothesized (prior to data collection) the Red group would recall more words than the Blue group and the Green group would recall less words than the Blue group.
- This necessitates two planned comparisons.

Simple Comparisons Example

- Recall, we hypothesized (prior to data collection) the Red group would recall more words than the Blue group and the Green group would recall less words than the Blue group.
- This necessitates two planned comparisons.
  - One to evaluate $H_{01} : \mu_{red} = \mu_{blue}$ and $H_{11} : \mu_{red} > \mu_{blue}$

Simple Comparisons Example

- Recall, we hypothesized (prior to data collection) the Red group would recall more words than the Blue group and the Green group would recall less words than the Blue group.
- This necessitates two planned comparisons.
  - One to evaluate $H_{01} : \mu_{red} = \mu_{blue}$ and $H_{11} : \mu_{red} > \mu_{blue}$
  - One to evaluate $H_{02} : \mu_{green} = \mu_{blue}$ and $H_{12} : \mu_{green} < \mu_{blue}$

Simple Comparisons Example

- Recall, we hypothesized (prior to data collection) the Red group would recall more words than the Blue group and the Green group would recall less words than the Blue group.
- This necessitates two planned comparisons.
  - One to evaluate $H_{01} : \mu_{red} = \mu_{blue}$ and $H_{11} : \mu_{red} > \mu_{blue}$
  - One to evaluate $H_{02} : \mu_{green} = \mu_{blue}$ and $H_{12} : \mu_{green} < \mu_{blue}$
- Using $df_b = 1$ and $df_w = 9$ and with a significance level of 0.05, we get $F_{crit} = 5.12$ for each comparison.

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

Planned Comparison 1: Red ($\overline{X} = 22.75$) versus Blue ($\overline{X} = 17.50$)

$\overline{X}_{..} = 20.125$ $\quad df_b = k - 1 = 2 - 1 = 1$

Planned Comparison 1: Red ($\overline{X} = 22.75$) versus Blue ($\overline{X} = 17.50$)

$\overline{X}_{..} = 20.125 \qquad df_b = k - 1 = 2 - 1 = 1$

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

Planned Comparison 1: Red ($\overline{X} = 22.75$) versus Blue ($\overline{X} = 17.50$)

$$\overline{X}_{..} = 20.125 \qquad df_b = k - 1 = 2 - 1 = 1$$

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

$$SOS_b = 4\,(22.75 - 20.125)^2 + 4\,(17.50 - 20.125)^2 = 55.12$$

Planned Comparison 1: Red ($\overline{X} = 22.75$) versus Blue ($\overline{X} = 17.50$)

$$\overline{X}_{..} = 20.125 \qquad df_b = k - 1 = 2 - 1 = 1$$

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

$$SOS_b = 4\,(22.75 - 20.125)^2 + 4\,(17.50 - 20.125)^2 = 55.12$$

$$MS_b = SOS_b / df_b = 55.12/1 = 55.12$$

Planned Comparison 1: Red ($\overline{X} = 22.75$) versus Blue ($\overline{X} = 17.50$)

$\overline{X}_{..} = 20.125 \qquad df_b = k - 1 = 2 - 1 = 1$

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

$SOS_b = 4\,(22.75 - 20.125)^2 + 4\,(17.50 - 20.125)^2 = 55.12$

$MS_b = SOS_b / df_b = 55.12/1 = 55.12$

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

Planned Comparison 1: Red ($\overline{X} = 22.75$) versus Blue ($\overline{X} = 17.50$)

$\overline{X}_{..} = 20.125 \qquad df_b = k - 1 = 2 - 1 = 1$

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

$SOS_b = 4\,(22.75 - 20.125)^2 + 4\,(17.50 - 20.125)^2 = 55.12$

$MS_b = SOS_b / df_b = 55.12/1 = 55.12$

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

$F_{calc} = MS_b / MS_w = 55.12/2.083 = 26.46$

Planned Comparison 1: Red ($\overline{X} = 22.75$) versus Blue ($\overline{X} = 17.50$)

$\overline{X}_{..} = 20.125 \qquad df_b = k - 1 = 2 - 1 = 1$

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

$SOS_b = 4\,(22.75 - 20.125)^2 + 4\,(17.50 - 20.125)^2 = 55.12$

$MS_b = SOS_b / df_b = 55.12/1 = 55.12$

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

$F_{calc} = MS_b / MS_w = 55.12/2.083 = 26.46$

So; $F_{calc} = 26.46 > 5.12 = F_{crit}$ we reject the null hypothesis and conclude that the Red group recalled significantly more words than the Blue group.

RSS Research and Statistical Support

Planned Comparison 2: Green ($\overline{X} = 11.50$) versus Blue ($\overline{X} = 17.50$)

$\overline{X}_{..} = 14.50$ $\quad df_b = k - 1 = 2 - 1 = 1$

Planned Comparison 2: Green ($\overline{X} = 11.50$) versus Blue ($\overline{X} = 17.50$)

$\overline{X}_{..} = 14.50 \qquad df_b = k - 1 = 2 - 1 = 1$

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

Planned Comparison 2: Green ($\overline{X} = 11.50$) versus Blue ($\overline{X} = 17.50$)

$$\overline{X}_{..} = 14.50 \qquad df_b = k - 1 = 2 - 1 = 1$$

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

$$SOS_b = 4\,(11.50 - 14.50)^2 + 4\,(11.50 - 14.50)^2 = 72.0$$

Planned Comparison 2: Green ($\overline{X} = 11.50$) versus Blue ($\overline{X} = 17.50$)

$\overline{X}_{..} = 14.50 \qquad df_b = k - 1 = 2 - 1 = 1$

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

$SOS_b = 4\,(11.50 - 14.50)^2 + 4\,(11.50 - 14.50)^2 = 72.0$

$MS_b = SOS_b / df_b = 72.00/1 = 72.00$

Planned Comparison 2: Green ($\overline{X} = 11.50$) versus Blue ($\overline{X} = 17.50$)

$\overline{X}_{..} = 14.50 \qquad df_b = k - 1 = 2 - 1 = 1$

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

$SOS_b = 4\,(11.50 - 14.50)^2 + 4\,(11.50 - 14.50)^2 = 72.0$

$MS_b = SOS_b / df_b = 72.00/1 = 72.00$

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

Planned Comparison 2: Green ($\overline{X} = 11.50$) versus Blue ($\overline{X} = 17.50$)

$\overline{X}_{..} = 14.50 \qquad df_b = k - 1 = 2 - 1 = 1$

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

$SOS_b = 4\,(11.50 - 14.50)^2 + 4\,(11.50 - 14.50)^2 = 72.0$

$MS_b = SOS_b / df_b = 72.00/1 = 72.00$

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

$F_{calc} = MS_b / MS_w = 72.00/2.083 = 34.57$

Planned Comparison 2: Green ($\overline{X} = 11.50$) versus Blue ($\overline{X} = 17.50$)

$\overline{X}_{..} = 14.50 \qquad df_b = k - 1 = 2 - 1 = 1$

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

$SOS_b = 4\,(11.50 - 14.50)^2 + 4\,(11.50 - 14.50)^2 = 72.0$

$MS_b = SOS_b / df_b = 72.00/1 = 72.00$

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

$F_{calc} = MS_b / MS_w = 72.00/2.083 = 34.57$

So; $F_{calc} = 34.57 > 5.12 = F_{crit}$ we reject the null hypothesis and conclude that the Green group recalled significantly fewer words than the Blue group.

Family-wise and Pair-wise Error Rates

- Sometimes called the Bonferroni Procedure (after it's founder), it is used to control family-wise error inflation when multiple tests are done in one study.

Family-wise and Pair-wise Error Rates

- Sometimes called the Bonferroni Procedure (after it's founder), it is used to control family-wise error inflation when multiple tests are done in one study.
- *FW* = desired significance level (or alpha level) divided by the number of comparisons.

Family-wise and Pair-wise Error Rates

- Sometimes called the Bonferroni Procedure (after it's founder), it is used to control family-wise error inflation when multiple tests are done in one study.
- *FW* = desired significance level (or alpha level) divided by the number of comparisons.
- For example, if doing 3 comparisons and wanting a total significance level of 0.05, then:

Family-wise and Pair-wise Error Rates

- Sometimes called the Bonferroni Procedure (after it's founder), it is used to control family-wise error inflation when multiple tests are done in one study.
- FW = desired significance level (or alpha level) divided by the number of comparisons.
- For example, if doing 3 comparisons and wanting a total significance level of 0.05, then:
  - $FW_{.05} = .05/3 = 0.0167$ is used for each comparison.

Family-wise and Pair-wise Error Rates

- Sometimes called the Bonferroni Procedure (after it's founder), it is used to control family-wise error inflation when multiple tests are done in one study.
- FW = desired significance level (or alpha level) divided by the number of comparisons.
- For example, if doing 3 comparisons and wanting a total significance level of 0.05, then:
  - $FW_{.05} = .05/3 = 0.0167$ is used for each comparison.
  - Often necessary to find an F distribution table with exact values, this can be difficult because; most tables are abbreviated for common cutoff points.

Family-wise and Pair-wise Error Rates

- Sometimes called the Bonferroni Procedure (after it's founder), it is used to control family-wise error inflation when multiple tests are done in one study.
- FW = desired significance level (or alpha level) divided by the number of comparisons.
- For example, if doing 3 comparisons and wanting a total significance level of 0.05, then:
  - $FW_{.05} = .05/3 = 0.0167$ is used for each comparison.
  - Often necessary to find an F distribution table with exact values, this can be difficult because; most tables are abbreviated for common cutoff points.
- Should be considered when doing planned comparisons.

RSS
Research and Statistical Support

Family-wise and Pair-wise Error Rates

- Sometimes called the Bonferroni Procedure (after it's founder), it is used to control family-wise error inflation when multiple tests are done in one study.
- *FW* = desired significance level (or alpha level) divided by the number of comparisons.
- For example, if doing 3 comparisons and wanting a total significance level of 0.05, then:
  - $FW_{.05} = .05/3 = 0.0167$ is used for each comparison.
  - Often necessary to find an *F* distribution table with exact values, this can be difficult because; most tables are abbreviated for common cutoff points.
- Should be considered when doing planned comparisons.
- Some Post-hoc tests have embedded controls for *FW* error rates.

RSS
Research and Statistical Support

Post-hoc Testing

- In exploratory research, you would test all possible combinations of groups to determine where the significant differences are located (typically only if the omnibus F is significant).

Post-hoc Testing

- In exploratory research, you would test all possible combinations of groups to determine where the significant differences are located (typically only if the omnibus F is significant).
  - Most researchers always do either planned comparisons (preferred) or post-hoc testing (exploratory), because the omnibus F is not terribly informative on its own.

Post-hoc Testing

- In exploratory research, you would test all possible combinations of groups to determine where the significant differences are located (typically only if the omnibus F is significant).
  - Most researchers always do either planned comparisons (preferred) or post-hoc testing (exploratory), because the omnibus F is not terribly informative on its own.
- There are many types of post-hoc procedures, each has strengths and weaknesses ... but, some *are* better than others.

Post-hoc Testing

- In exploratory research, you would test all possible combinations of groups to determine where the significant differences are located (typically only if the omnibus F is significant).
  - Most researchers always do either planned comparisons (preferred) or post-hoc testing (exploratory), because the omnibus F is not terribly informative on its own.
- There are many types of post-hoc procedures, each has strengths and weaknesses ... but, some *are* better than others.
- Some suggestions:

Post-hoc Testing

- In exploratory research, you would test all possible combinations of groups to determine where the significant differences are located (typically only if the omnibus F is significant).
  - Most researchers always do either planned comparisons (preferred) or post-hoc testing (exploratory), because the omnibus F is not terribly informative on its own.
- There are many types of post-hoc procedures, each has strengths and weaknesses ... but, some *are* better than others.
- Some suggestions:
  - If assumptions met: Tukey's HSD or REGW-Q.

RSS
Research and Statistical Support

Post-hoc Testing

- In exploratory research, you would test all possible combinations of groups to determine where the significant differences are located (typically only if the omnibus F is significant).
  - Most researchers always do either planned comparisons (preferred) or post-hoc testing (exploratory), because the omnibus F is not terribly informative on its own.
- There are many types of post-hoc procedures, each has strengths and weaknesses ... but, some *are* better than others.
- Some suggestions:
  - If assumptions met: Tukey's HSD or REGW-Q.
  - If unequal group sizes: Games-Howell.

RSS
Research and Statistical Support

Post-hoc Testing

- In exploratory research, you would test all possible combinations of groups to determine where the significant differences are located (typically only if the omnibus F is significant).
  - Most researchers always do either planned comparisons (preferred) or post-hoc testing (exploratory), because the omnibus F is not terribly informative on its own.
- There are many types of post-hoc procedures, each has strengths and weaknesses ... but, some *are* better than others.
- Some suggestions:
  - If assumptions met: Tukey's HSD or REGW-Q.
  - If unequal group sizes: Games-Howell.
  - If unequal variances: Games-Howell.

RSS
Research and Statistical Support

More on Post-hoc tests in general

- Most post-hoc tests rely on the t distribution or a modification of it.

More on Post-hoc tests in general

- Most post-hoc tests rely on the t distribution or a modification of it.
  - FYI: $F = t^2$ which means, the simple (two group) comparisons above can be converted with $t = \sqrt{F}$

More on Post-hoc tests in general

- Most post-hoc tests rely on the t distribution or a modification of it.
  - FYI: $F = t^2$ which means, the simple (two group) comparisons above can be converted with $t = \sqrt{F}$
- The modified t distribution is often called the *Studentized t* statistic.

More on Post-hoc tests in general

- Most post-hoc tests rely on the t distribution or a modification of it.
  - FYI: $F = t^2$ which means, the simple (two group) comparisons above can be converted with $t = \sqrt{F}$
- The modified t distribution is often called the *Studentized t* statistic.
  - Symbol for the Studentized t is: q

More on Post-hoc tests in general

- Most post-hoc tests rely on the t distribution or a modification of it.
  - FYI: $F = t^2$ which means, the simple (two group) comparisons above can be converted with $t = \sqrt{F}$
- The modified t distribution is often called the *Studentized t* statistic.
  - Symbol for the Studentized t is: q
  - It is defined by:

More on Post-hoc tests in general

- Most post-hoc tests rely on the t distribution or a modification of it.
  - FYI: $F = t^2$ which means, the simple (two group) comparisons above can be converted with $t = \sqrt{F}$
- The modified t distribution is often called the *Studentized t* statistic.
  - Symbol for the Studentized t is: q
  - It is defined by:

$$q = \frac{\overline{X}_l - \overline{X}_s}{\sqrt{\frac{MS_W}{n_g}}}$$

More on Post-hoc tests in general

- Most post-hoc tests rely on the t distribution or a modification of it.
  - FYI: $F = t^2$ which means, the simple (two group) comparisons above can be converted with $t = \sqrt{F}$
- The modified t distribution is often called the *Studentized t* statistic.
  - Symbol for the Studentized t is: q
  - It is defined by:

$$q = \frac{\overline{X}_l - \overline{X}_s}{\sqrt{\frac{MS_W}{n_g}}}$$

  - where $\overline{X}_l$ refers to the largest mean among the groups, $\overline{X}_s$ the smallest, and n_g is the number of individuals per group.

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

More on q

- When using q we first must rank the means from smallest to largest.

More on q

- When using q we first must rank the means from smallest to largest.
  - For our example: $\{11.50, 17.50, 22.75\}$

More on q

- When using q we first must rank the means from smallest to largest.
  - For our example: $\{11.50, 17.50, 22.75\}$
- There are 3 means in this set, $r = 3$ which is used in looking up values in the q table (the table below uses the familar k instead of r but, most use r).

More on q

- When using q we first must rank the means from smallest to largest.
  - For our example: $\{11.50, 17.50, 22.75\}$
- There are 3 means in this set, $r = 3$ which is used in looking up values in the q table (the table below uses the familar k instead of r but, most use r).

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

More on q

- When using q we first must rank the means from smallest to largest.
  - For our example: $\{11.50, 17.50, 22.75\}$
- There are 3 means in this set, $r = 3$ which is used in looking up values in the q table (the table below uses the familar k instead of r but, most use r).

  http://www.stat.duke.edu/courses/Spring98/sta110c/qtable.html
- Using the formula from above, we can solve for q

More on q

- When using q we first must rank the means from smallest to largest.
  - For our example: $\{11.50, 17.50, 22.75\}$
- There are 3 means in this set, $r = 3$ which is used in looking up values in the q table (the table below uses the familar k instead of r but, most use r).

  http://www.stat.duke.edu/courses/Spring98/sta110c/qtable.html
- Using the formula from above, we can solve for q

$$q = \frac{\overline{X}_l - \overline{X}_s}{\sqrt{\frac{MS_W}{n_g}}} = \frac{22.75 - 11.50}{\sqrt{\frac{2.083}{4}}} = 15.58$$

- So $q_{calc} = 15.58$ and we can use $r = 3$, significance level of 0.05, and $df_w = 9$ to find $q_{crit} = 3.95$ in the q table.

More on q

- When using q we first must rank the means from smallest to largest.
  - For our example: $\{11.50, 17.50, 22.75\}$
- There are 3 means in this set, $r = 3$ which is used in looking up values in the q table (the table below uses the familar k instead of r but, most use r).

  http://www.stat.duke.edu/courses/Spring98/sta110c/qtable.html
- Using the formula from above, we can solve for q

$$q = \frac{\overline{X}_l - \overline{X}_s}{\sqrt{\frac{MS_W}{n_g}}} = \frac{22.75 - 11.50}{\sqrt{\frac{2.083}{4}}} = 15.58$$

- So $q_{calc} = 15.58$ and we can use $r = 3$, significance level of 0.05, and $df_w = 9$ to find $q_{crit} = 3.95$ in the q table.
  - Given $q_{calc} = 15.58 > 3.95 = q_{crit}$, we reject the null hypothesis and conclude there was a significant difference between the largest and smallest means.

RSS
Research and Statistical Support

More on q

- When using q we first must rank the means from smallest to largest.
  - For our example: $\{11.50, 17.50, 22.75\}$
- There are 3 means in this set, $r = 3$ which is used in looking up values in the q table (the table below uses the familar k instead of r but, most use r).

  http://www.stat.duke.edu/courses/Spring98/sta110c/qtable.html
- Using the formula from above, we can solve for q

$$q = \frac{\overline{X}_l - \overline{X}_s}{\sqrt{\frac{MS_W}{n_g}}} = \frac{22.75 - 11.50}{\sqrt{\frac{2.083}{4}}} = 15.58$$

- So $q_{calc} = 15.58$ and we can use $r = 3$, significance level of 0.05, and $df_w = 9$ to find $q_{crit} = 3.95$ in the q table.
  - Given $q_{calc} = 15.58 > 3.95 = q_{crit}$, we reject the null hypothesis and conclude there was a significant difference between the largest and smallest means.
    - Which can replace the omnibus F, but with less statistical power.

RSS Research and Statistical Support

Tukey's Honestly Significant Differences test

- Tukey's HSD is considered one of the original post-hoc tests and offers a fixed family-wise error rate at alpha (significance level).

Tukey's Honestly Significant Differences test

- Tukey's HSD 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 ($q_{crit} = 3.95$) we can calculate the minimum difference between means needed to discover a significant difference.

Tukey's Honestly Significant Differences test

- Tukey's HSD 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 ($q_{crit} = 3.95$) we can calculate the minimum difference between means needed to discover a significant difference.

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

Tukey's Honestly Significant Differences test

- Tukey's HSD 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 ($q_{crit} = 3.95$) we can calculate the minimum difference between means needed to discover a significant difference.

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

- So, if a mean difference ($\overline{X}_i - \overline{X}_j$) is larger than 2.85 we would conclude there is a significant difference between those two means.

Tukey's HSD continued

- 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.

Tukey's HSD continued

- 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.

	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

Tukey's HSD continued

- 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.

	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.

Games-Howell Post-hoc Test

- The Tukey test above assumes equal sample sizes for each group **and** equal variances among the groups.

Games-Howell Post-hoc Test

- 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.

Games-Howell Post-hoc Test

- 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.

Games-Howell Post-hoc Test

- 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.

Games-Howell Post-hoc Test

- 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 $(\overline{X}_i - \overline{X}_j)$ are modified for inclusion of samples sizes and variances.

Games-Howell modification of *df*

- Until now we have consistently used df_w for finding our critical values (e.g., q_{crit}).

Games-Howell modification of *df*

- 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**:

Games-Howell modification of *df*

- 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**:

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

Games-Howell modification of *df*

- 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'

As an example let's consider the Red and Green groups.

- Red (i): $\overline{X} = 22.75, S^2 = 2.92, n = 4$
- Green (j): $\overline{X} = 11.50, S^2 = 1.66, n = 4$

Example df'

As an example let's consider the Red and Green groups.

- Red (i): $\overline{X} = 22.75, S^2 = 2.92, n = 4$
- Green (j): $\overline{X} = 11.50, S^2 = 1.66, n = 4$

$$df' = \frac{\left(\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}\right)^2}{\frac{\left(\frac{s_i^2}{n_i}\right)^2}{n_i - 1} + \frac{\left(\frac{s_j^2}{n_j}\right)^2}{n_j - 1}} = \frac{\left(\frac{2.92}{4} + \frac{1.66}{4}\right)^2}{\frac{\left(\frac{2.92}{4}\right)^2}{4-1} + \frac{\left(\frac{1.66}{4}\right)^2}{4-1}} = 5.5787$$

Example df'

As an example let's consider the Red and Green groups.

- Red (i): $\overline{X} = 22.75, S^2 = 2.92, n = 4$
- Green (j): $\overline{X} = 11.50, S^2 = 1.66, n = 4$

$$df' = \frac{\left(\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}\right)^2}{\frac{\left(\frac{s_i^2}{n_i}\right)^2}{n_i - 1} + \frac{\left(\frac{s_j^2}{n_j}\right)^2}{n_j - 1}} = \frac{\left(\frac{2.92}{4} + \frac{1.66}{4}\right)^2}{\frac{\left(\frac{2.92}{4}\right)^2}{4-1} + \frac{\left(\frac{1.66}{4}\right)^2}{4-1}} = 5.5787$$

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

Example df'

As an example let's consider the Red and Green groups.

- Red (i): $\overline{X} = 22.75, S^2 = 2.92, n = 4$
- Green (j): $\overline{X} = 11.50, S^2 = 1.66, n = 4$

$$df' = \frac{\left(\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}\right)^2}{\frac{\left(\frac{s_i^2}{n_i}\right)^2}{n_i - 1} + \frac{\left(\frac{s_j^2}{n_j}\right)^2}{n_j - 1}} = \frac{\left(\frac{2.92}{4} + \frac{1.66}{4}\right)^2}{\frac{\left(\frac{2.92}{4}\right)^2}{4-1} + \frac{\left(\frac{1.66}{4}\right)^2}{4-1}} = 5.5787$$

- Pay careful attention to the *fact* that S^2 is a **symbol**, not an operation.
  - See the supplemental handout for an example of the complete, step-by-step calculation.

Games-Howell difference between means

Minimum Significant Difference

Games-Howell difference between means

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: $q_{crit} \approx 4.60$

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

Games-Howell difference between means

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: $q_{crit} \approx 4.60$

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

- Where earlier we had:

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

Games-Howell difference between means

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: $q_{crit} \approx 4.60$

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

- Where earlier we had:

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

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

Games-Howell difference between means

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: $q_{crit} \approx 4.60$

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

- Where earlier we had:

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

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

$$\overline{X}_i - \overline{X}_j = q_{crit}\sqrt{\frac{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}}{2}}$$

Games-Howell difference between means

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: $q_{crit} \approx 4.60$

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

- Where earlier we had:

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

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

$$\overline{X}_i - \overline{X}_j = q_{crit}\sqrt{\frac{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}}{2}}$$

- For the current Red vs. Green example:

Games-Howell difference between means

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: $q_{crit} \approx 4.60$

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

- Where earlier we had:

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

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

$$\overline{X}_i - \overline{X}_j = q_{crit}\sqrt{\frac{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}}{2}}$$

- For the current Red vs. Green example:

$$\overline{X}_i - \overline{X}_j = 5.787\sqrt{\frac{\frac{2.92}{4} + \frac{1.66}{4}}{2}} = 4.379$$

Red $\overline{X} = 22.75$ and Green $\overline{X} = 11.50$

- 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.

Red $\overline{X} = 22.75$ and Green $\overline{X} = 11.50$

- 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, $\overline{X}_i - \overline{X}_j = 22.75 - 11.50 = 11.25$ is greater than 4.379; which indicates the Red group recalled significantly more words than the Green group.

Red $\overline{X} = 22.75$ and Green $\overline{X} = 11.50$

- 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, $\overline{X}_i - \overline{X}_j = 22.75 - 11.50 = 11.25$ 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).

Red $\overline{X} = 22.75$ and Green $\overline{X} = 11.50$

- 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, $\overline{X}_i - \overline{X}_j = 22.75 - 11.50 = 11.25$ 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.

RSS
Research and Statistical Support

Red $\overline{X} = 22.75$ and Green $\overline{X} = 11.50$

- 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, $\overline{X}_i - \overline{X}_j = 22.75 - 11.50 = 11.25$ 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.

RSS
Research and Statistical Support

Effect Size Measures for ANOVA

- The ANOVA family of analyses have a variety of effect size measures which can be used.

Effect Size Measures for ANOVA

- The ANOVA family of analyses have a variety of effect size measures which can be used.
- Recall, there are generally two forms of effect size measures;

Effect Size Measures for ANOVA

- The ANOVA family of analyses have a variety of effect size measures which can be used.
- Recall, there are generally two forms of effect size measures;
  - Measures of *shared variance* such as R^2 and $R^2_{adj.}$

Effect Size Measures for ANOVA

- The ANOVA family of analyses have a variety of effect size measures which can be used.
- Recall, there are generally two forms of effect size measures;
  - Measures of *shared variance* such as R^2 and $R^2_{adj.}$
  - Measures of *standardized difference* such as Cohen's d

Effect Size Measures for ANOVA

- The ANOVA family of analyses have a variety of effect size measures which can be used.
- Recall, there are generally two forms of effect size measures;
  - Measures of *shared variance* such as R^2 and $R^2_{adj.}$
  - Measures of *standardized difference* such as Cohen's d
- Well, it turns out that ANOVA is a member of the **General Linear Model** (GLM) and therefore, both types of effect size can be utilized in the ANOVA situation.

Effect Size Measures for ANOVA

- The ANOVA family of analyses have a variety of effect size measures which can be used.
- Recall, there are generally two forms of effect size measures;
  - Measures of *shared variance* such as R^2 and $R^2_{adj.}$
  - Measures of *standardized difference* such as Cohen's *d*
- Well, it turns out that ANOVA is a member of the **General Linear Model** (GLM) and therefore, both types of effect size can be utilized in the ANOVA situation.
  - GLM will be discussed in more detail in future modules (e.g., regression) but, suffice to say; we have specified a *linear model* here, which generally refers to a linear relationship between the variables we are analyzing.

RSS Research and Statistical Support

Available Effect Size Measures

- Cohen's *d* can be computed during individual comparisons (planned comparisons or post-hoc testing).

Available Effect Size Measures

- Cohen's *d* can be computed during individual comparisons (planned comparisons or post-hoc testing).
  - In those situations we are essentially comparing two means (similar to the *t* tests).

Available Effect Size Measures

- Cohen's *d* can be computed during individual comparisons (planned comparisons or post-hoc testing).
  - In those situations we are essentially comparing two means (similar to the *t* tests).
- The most primitive (slightly biased) effect size used with one-way ANOVA is eta-squared (η^2).

Available Effect Size Measures

- Cohen's *d* can be computed during individual comparisons (planned comparisons or post-hoc testing).
  - In those situations we are essentially comparing two means (similar to the *t* tests).
- The most primitive (slightly biased) effect size used with one-way ANOVA is eta-squared (η^2).
- Another effect size for ANOVA is R^2 which we encountered in an earlier module and know it overestimates.

Available Effect Size Measures

- Cohen's *d* can be computed during individual comparisons (planned comparisons or post-hoc testing).
  - In those situations we are essentially comparing two means (similar to the *t* tests).
- The most primitive (slightly biased) effect size used with one-way ANOVA is eta-squared (η^2).
- Another effect size for ANOVA is R^2 which we encountered in an earlier module and know it overestimates.
- Partial eta-squared is a popular effect size for ANOVA; although it is generally reserved for more complex designs than the one-way situation.

Available Effect Size Measures

- Cohen's *d* can be computed during individual comparisons (planned comparisons or post-hoc testing).
  - In those situations we are essentially comparing two means (similar to the *t* tests).
- The most primitive (slightly biased) effect size used with one-way ANOVA is eta-squared (η^2).
- Another effect size for ANOVA is R^2 which we encountered in an earlier module and know it overestimates.
- Partial eta-squared is a popular effect size for ANOVA; although it is generally reserved for more complex designs than the one-way situation.
  - Same as η^2 in the one-way situation.

Available Effect Size Measures

- Cohen's *d* can be computed during individual comparisons (planned comparisons or post-hoc testing).
  - In those situations we are essentially comparing two means (similar to the *t* tests).
- The most primitive (slightly biased) effect size used with one-way ANOVA is eta-squared (η^2).
- Another effect size for ANOVA is R^2 which we encountered in an earlier module and know it overestimates.
- Partial eta-squared is a popular effect size for ANOVA; although it is generally reserved for more complex designs than the one-way situation.
  - Same as η^2 in the one-way situation.
- The **recommended** effect size, for the one-way situation and more complex situations, is Omega squared (ω^2)

Available Effect Size Measures

- Cohen's *d* can be computed during individual comparisons (planned comparisons or post-hoc testing).
  - In those situations we are essentially comparing two means (similar to the *t* tests).
- The most primitive (slightly biased) effect size used with one-way ANOVA is eta-squared (η^2).
- Another effect size for ANOVA is R^2 which we encountered in an earlier module and know it overestimates.
- Partial eta-squared is a popular effect size for ANOVA; although it is generally reserved for more complex designs than the one-way situation.
  - Same as η^2 in the one-way situation.
- The **recommended** effect size, for the one-way situation and more complex situations, is Omega squared (ω^2)
  - Omega squared is virtually unbiased.

Eta squared

- Eta squared (η^2) is very easy to calculate, but as mentioned above it is biased.

Eta squared

- Eta squared (η^2) is very easy to calculate, but as mentioned above it is biased.

Source	*SOS*	*df*	*MS*	*F*
Between	253.50	2	126.75	60.84
Within	18.750	9	2.083	
Total	272.25	11		

Eta squared

- Eta squared (η^2) is very easy to calculate, but as mentioned above it is biased.

Source	*SOS*	*df*	*MS*	*F*
Between	253.50	2	126.75	60.84
Within	18.750	9	2.083	
Total	272.25	11		

$$\eta^2 = \frac{SOS_b}{SOS_t} = \frac{253.50}{272.25} = .9311$$

Eta squared

- Eta squared (η^2) is very easy to calculate, but as mentioned above it is biased.

Source	*SOS*	*df*	*MS*	*F*
Between	253.50	2	126.75	60.84
Within	18.750	9	2.083	
Total	272.25	11		

$$\eta^2 = \frac{SOS_b}{SOS_t} = \frac{253.50}{272.25} = .9311$$

- Which suggests 93.11% of the variance in words recalled (Dependent variable [DV])is attributable to the differences in font color (Indpendent variable [IV]).

RSS
Research and Statistical Support

R^2 for ANOVA

- There are two versions of the R^2 formula for ANOVA, the one below is used in the one-way situation:

R^2 for ANOVA

- There are two versions of the R^2 formula for ANOVA, the one below is used in the one-way situation:

$$R^2 = \frac{(F)(df_b)}{(F)(df_b)+df_w} = \frac{60.84(2)}{60.84(2)+9} = \frac{121.68}{130.68} = .9311$$

R^2 for ANOVA

- There are two versions of the R^2 formula for ANOVA, the one below is used in the one-way situation:

$$R^2 = \frac{(F)(df_b)}{(F)(df_b)+df_w} = \frac{60.84(2)}{60.84(2)+9} = \frac{121.68}{130.68} = .9311$$

- No surprise here, 93.11% of the variance in words recalled (DV)is attributable to the differences in font color (IV).

R^2 for ANOVA

- There are two versions of the R^2 formula for ANOVA, the one below is used in the one-way situation:

$$R^2 = \frac{(F)(df_b)}{(F)(df_b)+df_w} = \frac{60.84(2)}{60.84(2)+9} = \frac{121.68}{130.68} = .9311$$

- No surprise here, 93.11% of the variance in words recalled (DV)is attributable to the differences in font color (IV).
  - This is the same effect size we got for η^2

Omega Squared

- Omega squared (ω^2) is the recommended effect size measure because, it takes into account the number of groups (less biased than the ones above).

Omega Squared

- Omega squared (ω^2) is the recommended effect size measure because, it takes into account the number of groups (less biased than the ones above).

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

Omega Squared

- Omega squared (ω^2) is the recommended effect size measure because, it takes into account the number of groups (less biased than the ones above).

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

- Here we see a less biased, more accurate representation of the relationship between the IV and the DV.

Omega Squared

- Omega squared (ω^2) is the recommended effect size measure because, it takes into account the number of groups (less biased than the ones above).

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

- 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).

Omega Squared

- Omega squared (ω^2) is the recommended effect size measure because, it takes into account the number of groups (less biased than the ones above).

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

- 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 on next slide).

Omega squared for planned comparisons

- For Planned comparison 1 from earlier:

Omega squared for planned comparisons

- For Planned comparison 1 from earlier:

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

Omega squared for planned comparisons

- For Planned comparison 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.

Omega squared for planned comparisons

- For Planned comparison 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.
- For Planned Comparison 2 from earlier:

Omega squared for planned comparisons

- For Planned comparison 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.
- For Planned Comparison 2 from earlier:

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

Omega squared for planned comparisons

- For Planned comparison 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.
- For Planned Comparison 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.

Confidence Intervals

In the ANOVA situation, Confidence Intervals (CIs) are not meaningful, **until** you get down to testing individual pairs of means.

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.

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*).

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.

RSS
Research and Statistical Support

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:

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:

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

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.

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**.

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.

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.

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

Adjustments for square root of a negative

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

Adjustments for square root of a negative

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

- As many of you know, you can not take the square root of a negative number.

Adjustments for square root of a negative

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

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

Adjustments for square root of a negative

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

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

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

RSS
Research and Statistical Support

CI for Example Planned Comparisons (PC)

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

CI for Example Planned Comparisons (PC)

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

$$UL = \sqrt{(+F_{crit}) * (MS_w)} + \overline{X}_i - \overline{X}_j = \sqrt{5.12\,(2.083)} + 5.25 = 8.52$$
$$LL = \sqrt{(-F_{crit}) * (MS_w)} + \overline{X}_i - \overline{X}_j = \sqrt{5.12\,(2.083)} - 5.25 = 1.98$$

CI for Example Planned Comparisons (PC)

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

$$UL = \sqrt{(+F_{crit}) * (MS_w)} + \overline{X}_i - \overline{X}_j = \sqrt{5.12\,(2.083)} + 5.25 = 8.52$$
$$LL = \sqrt{(-F_{crit}) * (MS_w)} + \overline{X}_i - \overline{X}_j = \sqrt{5.12\,(2.083)} - 5.25 = 1.98$$

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 Example Planned Comparisons (PC)

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

$$UL = \sqrt{(+F_{crit}) * (MS_w)} + \overline{X}_i - \overline{X}_j = \sqrt{5.12\,(2.083)} + 5.25 = 8.52$$
$$LL = \sqrt{(-F_{crit}) * (MS_w)} + \overline{X}_i - \overline{X}_j = \sqrt{5.12\,(2.083)} - 5.25 = 1.98$$

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 ($\overline{X} = 11.50$) vs. Blue ($\overline{X} = 17.50$)

CI for Example Planned Comparisons (PC)

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

$$UL = \sqrt{(+F_{crit}) * (MS_w)} + \overline{X}_i - \overline{X}_j = \sqrt{5.12\,(2.083)} + 5.25 = 8.52$$
$$LL = \sqrt{(-F_{crit}) * (MS_w)} + \overline{X}_i - \overline{X}_j = \sqrt{5.12\,(2.083)} - 5.25 = 1.98$$

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 ($\overline{X} = 11.50$) vs. Blue ($\overline{X} = 17.50$)

$$UL = \sqrt{(+F_{crit}) * (MS_w)} + \overline{X}_i - \overline{X}_j = \sqrt{5.12\,(2.083)} + 6.00 = 9.27$$
$$LL = \sqrt{(-F_{crit}) * (MS_w)} + \overline{X}_i - \overline{X}_j = \sqrt{5.12\,(2.083)} - 6.00 = 2.73$$

RSS
Research and Statistical Support

CI for Example Planned Comparisons (PC)

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

$$UL = \sqrt{(+F_{crit}) * (MS_w)} + \overline{X}_i - \overline{X}_j = \sqrt{5.12\,(2.083)} + 5.25 = 8.52$$
$$LL = \sqrt{(-F_{crit}) * (MS_w)} + \overline{X}_i - \overline{X}_j = \sqrt{5.12\,(2.083)} - 5.25 = 1.98$$

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 ($\overline{X} = 11.50$) vs. Blue ($\overline{X} = 17.50$)

$$UL = \sqrt{(+F_{crit}) * (MS_w)} + \overline{X}_i - \overline{X}_j = \sqrt{5.12\,(2.083)} + 6.00 = 9.27$$
$$LL = \sqrt{(-F_{crit}) * (MS_w)} + \overline{X}_i - \overline{X}_j = \sqrt{5.12\,(2.083)} - 6.00 = 2.73$$

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.

RSS Research and Statistical Support

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 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 ($\overline{X} = 22.75$) vs. Blue ($\overline{X} = 17.50$)

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 ($\overline{X} = 22.75$) vs. Blue ($\overline{X} = 17.50$)

$$UL = (+q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = +2.85 + 5.25 = 8.10$$
$$LL = (-q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = -2.85 + 5.25 = 2.40$$

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 ($\overline{X} = 22.75$) vs. Blue ($\overline{X} = 17.50$)

$$UL = (+q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = +2.85 + 5.25 = 8.10$$
$$LL = (-q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = -2.85 + 5.25 = 2.40$$

- 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 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 ($\overline{X} = 22.75$) vs. Blue ($\overline{X} = 17.50$)

$$UL = (+q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = +2.85 + 5.25 = 8.10$$
$$LL = (-q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = -2.85 + 5.25 = 2.40$$

- 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.
- Continued on next slide.

CI for Tukey's HSD PH continued

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

CI for Tukey's HSD PH continued

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

$$UL = (+q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = +2.85 + 11.25 = 14.10$$
$$LL = (-q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = -2.85 + 11.25 = 8.40$$

CI for Tukey's HSD PH continued

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

$$UL = (+q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = +2.85 + 11.25 = 14.10$$
$$LL = (-q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = -2.85 + 11.25 = 8.40$$

- 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 Tukey's HSD PH continued

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

$$UL = (+q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = +2.85 + 11.25 = 14.10$$
$$LL = (-q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = -2.85 + 11.25 = 8.40$$

- 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 ($\overline{X} = 17.50$) vs. Green ($\overline{X} = 11.50$)

CI for Tukey's HSD PH continued

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

$$UL = (+q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = +2.85 + 11.25 = 14.10$$
$$LL = (-q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = -2.85 + 11.25 = 8.40$$

- 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 ($\overline{X} = 17.50$) vs. Green ($\overline{X} = 11.50$)

$$UL = (+q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = +2.85 + 6.00 = 8.85$$
$$LL = (-q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = -2.85 + 6.00 = 3.15$$

CI for Tukey's HSD PH continued

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

$$UL = (+q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = +2.85 + 11.25 = 14.10$$
$$LL = (-q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = -2.85 + 11.25 = 8.40$$

- 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 ($\overline{X} = 17.50$) vs. Green ($\overline{X} = 11.50$)

$$UL = (+q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = +2.85 + 6.00 = 8.85$$
$$LL = (-q_{r=3,.05}) + \overline{X}_i - \overline{X}_j = -2.85 + 6.00 = 3.15$$

- 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.

Simple Default graph from some software

Not very informative.

Simple Default graph from some software

Not very informative.

Displays the means
as points

Simple Default graph from some software

Not very informative.

Displays the means as points

Confusing because the points are linked by a line when we are *testing* for differences among them.

Simple Default graph from some software

Not very informative.

Displays the means as points

Confusing because the points are linked by a line when we are *testing* for differences among them.

Also, it shows little of the **variance** in an Analysis of Variance setting.

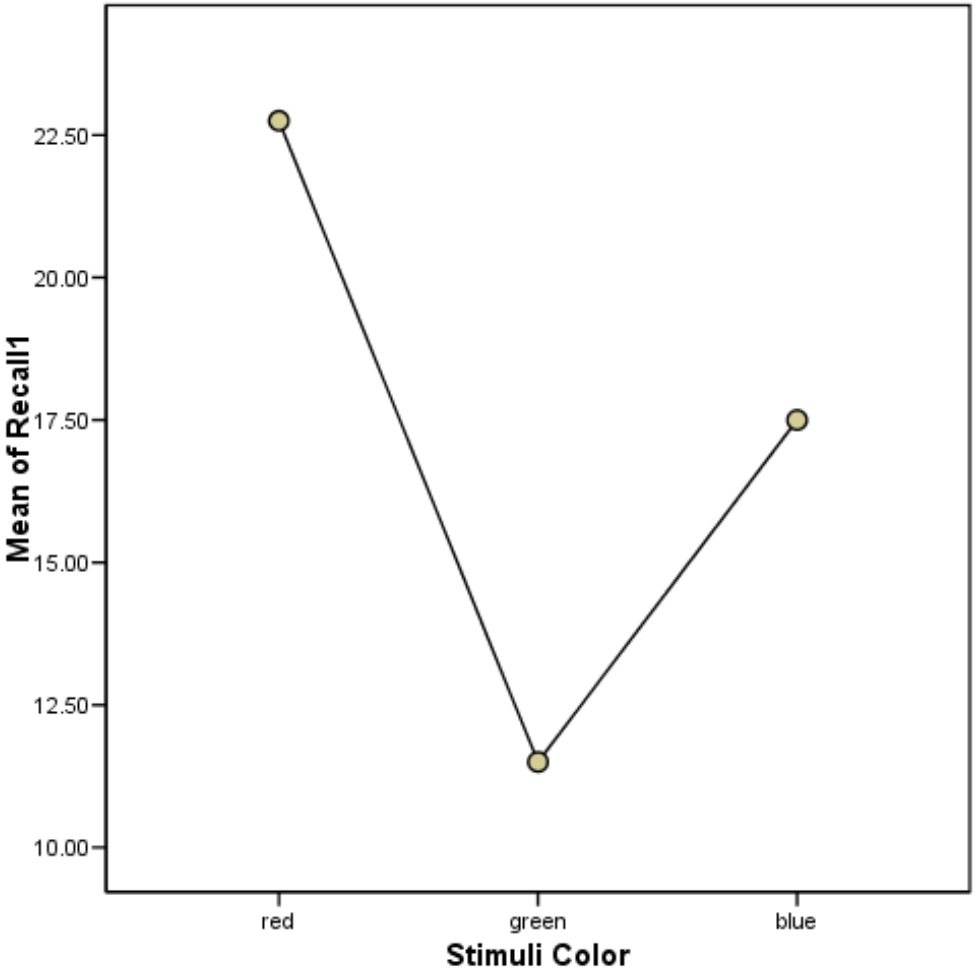
Simple Default graph from some software

Not very informative.

Displays the means as points

Confusing because the points are linked by a line when we are *testing* for differences among them.

Also, it shows little of the **variance** in an Analysis of Variance setting.



Grouped Box plot represents the data better.

Displays the means as solid black lines in the middle of the boxes.

Grouped Box plot represents the data better.

Displays the means as solid black lines in the middle of the boxes.

Between groups variance is represented by the distance between the top whisker and bottom whisker across groups (i.e. top Red, bottom Green).

Grouped Box plot represents the data better.

Displays the means as solid black lines in the middle of the boxes.

Between groups variance is represented by the distance between the top whisker and bottom whisker across groups (i.e. top Red, bottom Green).

Within groups variance is represented by the distance between the boundary of each box's top and bottom whiskers.

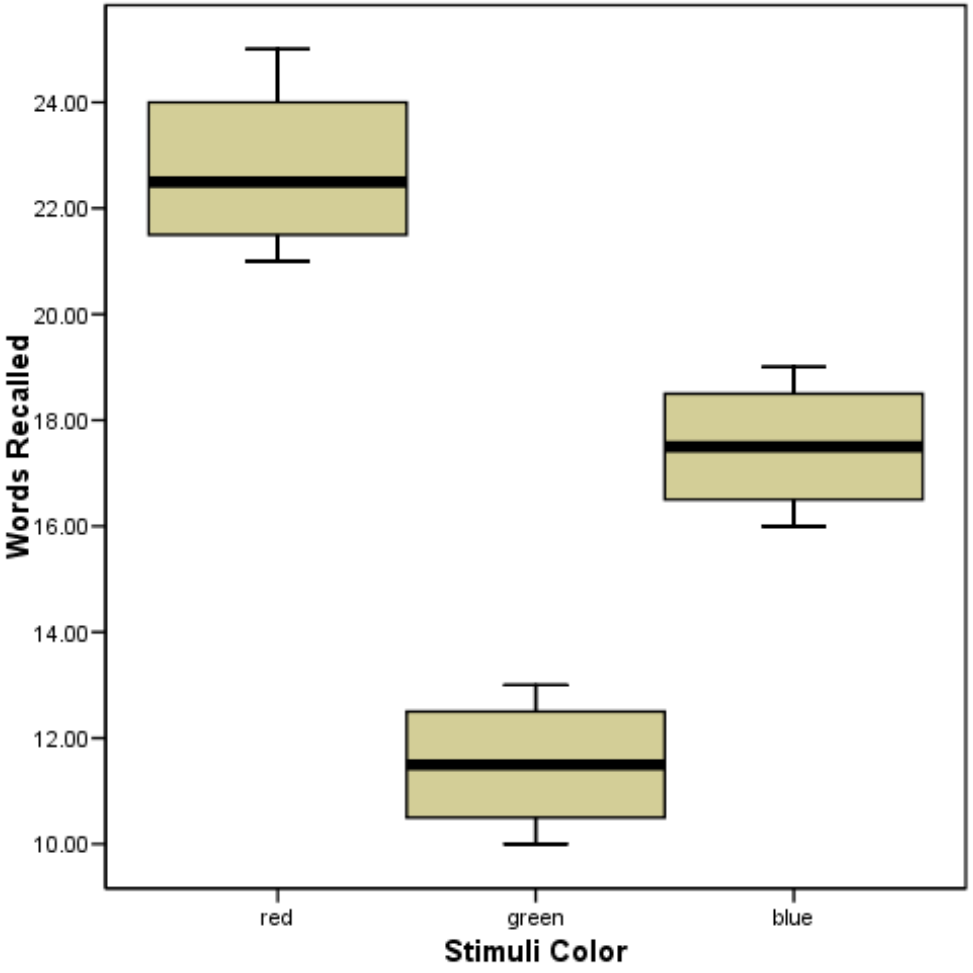
RSS
Research and Statistical Support

Grouped Box plot represents the data better.

Displays the means as solid black lines in the middle of the boxes.

Between groups variance is represented by the distance between the top whisker and bottom whisker across groups (i.e. top Red, bottom Green).

Within groups variance is represented by the distance between the boundary of each box's top and bottom whiskers.



Research and Statistical Support

Quick Review

- Recall, MS_b represents the average between groups variance.

Quick Review

- Recall, MS_b represents the average between groups variance.
- And, MS_w represents the average within groups variance.

Quick Review

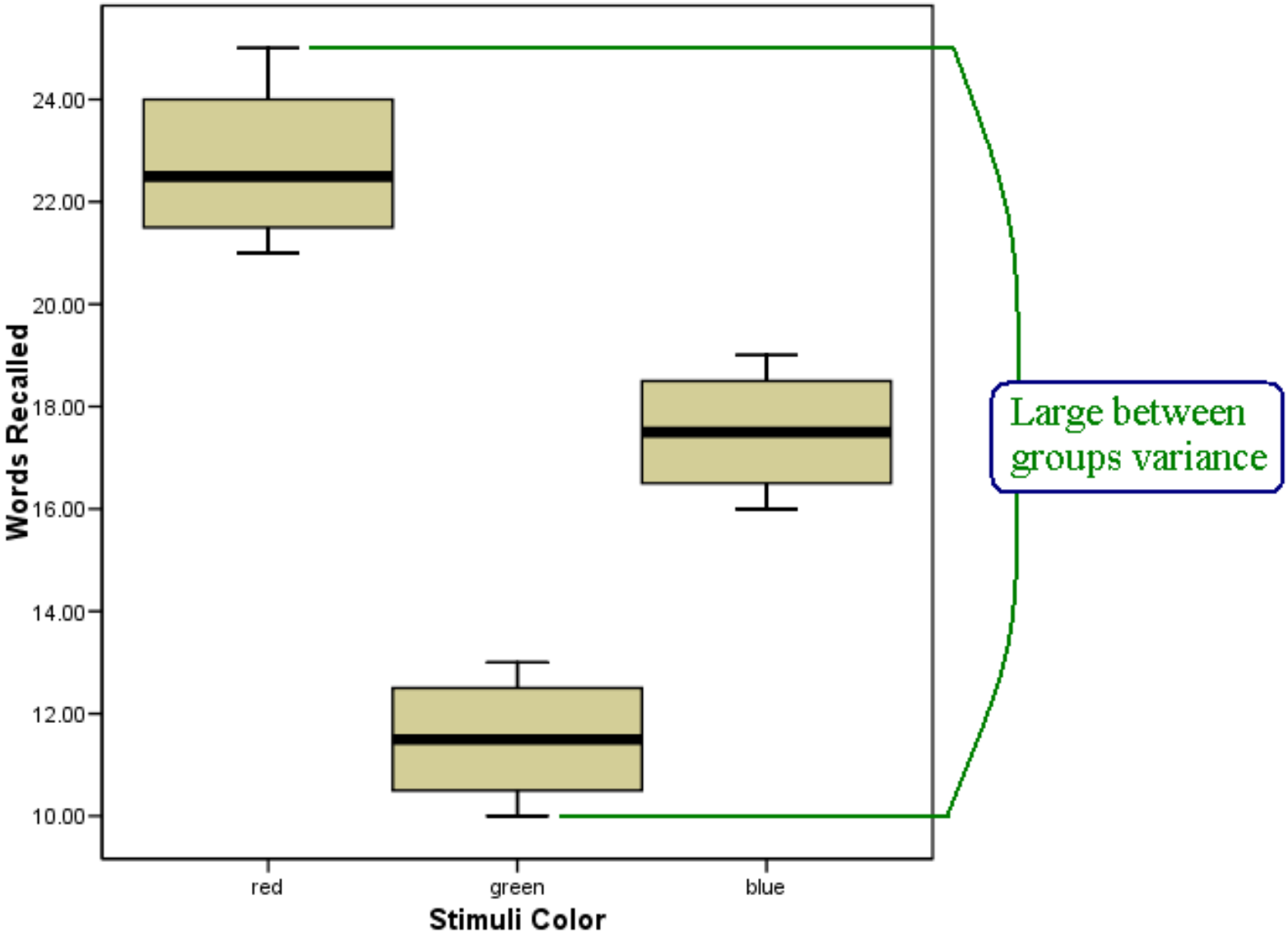
- Recall, MS_b represents the average between groups variance.
- And, MS_w represents the average within groups variance.
- ANOVA analyzes variance to determine if mean differences are significant.

Quick Review

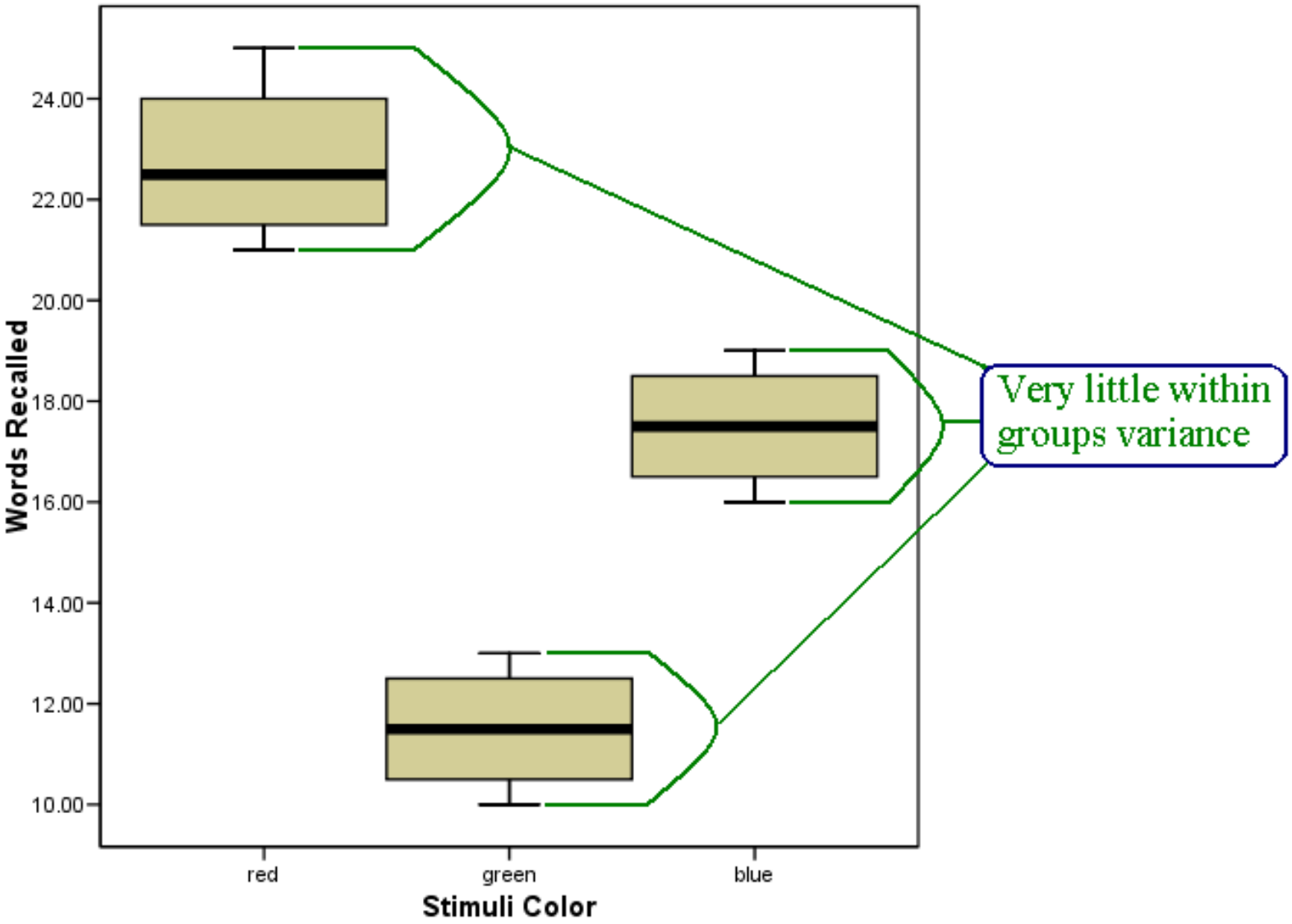
- Recall, MS_b represents the average between groups variance.
- And, MS_w represents the average within groups variance.
- ANOVA analyzes variance to determine if mean differences are significant.

$$F = \frac{MS_b}{MS_w}$$

Box plots show between groups variance



Box plots show within groups variance



Power

- As mentioned in previous modules, post-hoc power (calculated after conducting the analysis) is meaningless.

Power

- As mentioned in previous modules, post-hoc power (calculated after conducting the analysis) is meaningless.
  - Power is a direct function of sample size.

Power

- As mentioned in previous modules, post-hoc power (calculated after conducting the analysis) is meaningless.
  - Power is a direct function of sample size.
- The best way to use power is to specify it and the effect size you would like to attain, and use those to calculate the number of subjects/participants needed to achieve those goals.

Power

- As mentioned in previous modules, post-hoc power (calculated after conducting the analysis) is meaningless.
  - Power is a direct function of sample size.
- The best way to use power is to specify it and the effect size you would like to attain, and use those to calculate the number of subjects/participants needed to achieve those goals.
  - This is called a-priori power calculation.

Power

- As mentioned in previous modules, post-hoc power (calculated after conducting the analysis) is meaningless.
  - Power is a direct function of sample size.
- The best way to use power is to specify it and the effect size you would like to attain, and use those to calculate the number of subjects/participants needed to achieve those goals.
  - This is called a-priori power calculation.
- A-priori power calculations (i.e. calculating the necessary sample size) should always be done when planning a study.

RSS
Research and Statistical Support

Power

- As mentioned in previous modules, post-hoc power (calculated after conducting the analysis) is meaningless.
  - Power is a direct function of sample size.
- The best way to use power is to specify it and the effect size you would like to attain, and use those to calculate the number of subjects/participants needed to achieve those goals.
  - This is called a-priori power calculation.
- A-priori power calculations (i.e. calculating the necessary sample size) should always be done when planning a study.
  - A-priori calculations should be done **prior** to data collection.

RSS
Research and Statistical Support

G-power 3

- The best way to calculate a-priori power/sample size is to use G-power 3, which is available for free at the link below.

G-power 3

- The best way to calculate a-priori power/sample size is to use G-power 3, which is available for free at the link below.

  http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/

G-power 3

- The best way to calculate a-priori power/sample size is to use G-power 3, which is available for free at the link below.

  http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/

- At the G-power 3 web site, you will find links to the User Manual for G-power 3 which are very easy to follow and guild you through the process of carrying out the necessary calculations **for a variety of analyses**.

RSS
Research and Statistical Support

G-power 3

- The best way to calculate a-priori power/sample size is to use G-power 3, which is available for free at the link below.

  http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/

- At the G-power 3 web site, you will find links to the User Manual for G-power 3 which are very easy to follow and guild you through the process of carrying out the necessary calculations **for a variety of analyses**.
- Also at the G-power site, a Literature section is available which contains a references list related to power and sample size calculation.

RSS
Research and Statistical Support

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).

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.

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.

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.
  - *Good News*: Gain in power and protection against Type I error.

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.
  - *Good News*: Gain in power and protection against Type I error.
  - *Bad News*: Requires using different (somewhat cumbersome) calculations; although most of the symbols you already know from previous usage.

RSS
Research and Statistical Support

Welch Formulas

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

Welch Formulas

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

$$w_k = \frac{n_k}{S_k^2}$$

Welch Formulas

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

$$w_k = \frac{n_k}{S_k^2}$$

$$\overline{X}'_{.} = \frac{\sum\left(w_k \overline{X}_k\right)}{\sum w_k}$$

Welch Formulas

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

$$w_k = \frac{n_k}{S_k^2}$$

$$\overline{X}'_{.} = \frac{\sum\left(w_k \overline{X}_k\right)}{\sum w_k}$$

- For example, the Red group's weight would be the number of individuals in the Red group divided by the variance of the Red group:

Welch Formulas

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

$$w_k = \frac{n_k}{S_k^2}$$

$$\overline{X}'_{.} = \frac{\sum \left(w_k \overline{X}_k \right)}{\sum w_k}$$

- For example, the Red group's weight would be the number of individuals in the Red group divided by the variance of the Red group:

$$w_{red} = \frac{n_{red}}{S_{red}^2} = \frac{4}{2.92} = 1.37$$

RSS
Research and Statistical Support

Welch Robust Procedure continued

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

Welch Robust Procedure continued

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

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

Welch Robust Procedure continued

- Next, calculate the df'

Welch Robust Procedure continued

- Next, calculate the df'

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

Welch Robust Procedure continued

- Next, calculate the df'

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

- Now use df' and $k-1 = df_b$ to find the critical value in the standard F table using df_b as the numerator and df' as the denominator.

Welch Robust Procedure continued

- Next, calculate the df'

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

- Now use df' and $k-1 = df_b$ to find the critical value in the standard F table using df_b as the numerator and df' as the denominator.
- See the Supplemental Handout for a complete example of the Welch Robust Procedure.

Summary of Module 9

Module 9 covered the following topics:

- Introduction to One-way ANOVA

Summary of Module 9

Module 9 covered the following topics:

- Introduction to One-way ANOVA
- One-Way ANOVA Hypothesis Testing

Summary of Module 9

Module 9 covered the following topics:

- Introduction to One-way ANOVA
- One-Way ANOVA Hypothesis Testing
  - Planned Comparisons

Summary of Module 9

Module 9 covered the following topics:

- Introduction to One-way ANOVA
- One-Way ANOVA Hypothesis Testing
  - Planned Comparisons
  - Post-hoc Testing

RSS
Research and Statistical Support

Summary of Module 9

Module 9 covered the following topics:

- Introduction to One-way ANOVA
- One-Way ANOVA Hypothesis Testing
  - Planned Comparisons
  - Post-hoc Testing
  - Effect Size

Summary of Module 9

Module 9 covered the following topics:

- Introduction to One-way ANOVA
- One-Way ANOVA Hypothesis Testing
  - Planned Comparisons
  - Post-hoc Testing
  - Effect Size
  - Confidence Intervals

Summary of Module 9

Module 9 covered the following topics:

- Introduction to One-way ANOVA
- One-Way ANOVA Hypothesis Testing
  - Planned Comparisons
  - Post-hoc Testing
  - Effect Size
  - Confidence Intervals
- Graphing

RSS
Research and Statistical Support

Summary of Module 9

Module 9 covered the following topics:

- Introduction to One-way ANOVA
- One-Way ANOVA Hypothesis Testing
  - Planned Comparisons
  - Post-hoc Testing
  - Effect Size
  - Confidence Intervals
- Graphing
- Additional Considerations

RSS
Research and Statistical Support

Summary of Module 9

Module 9 covered the following topics:

- Introduction to One-way ANOVA
- One-Way ANOVA Hypothesis Testing
  - Planned Comparisons
  - Post-hoc Testing
  - Effect Size
  - Confidence Intervals
- Graphing
- Additional Considerations
  - Power

RSS
Research and Statistical Support

Summary of Module 9

Module 9 covered the following topics:

- Introduction to One-way ANOVA
- One-Way ANOVA Hypothesis Testing
  - Planned Comparisons
  - Post-hoc Testing
  - Effect Size
  - Confidence Intervals
- Graphing
- Additional Considerations
  - Power
  - Welch Robust Procedure

RSS
Research and Statistical Support

Summary of Module 9

Module 9 covered the following topics:

- Introduction to One-way ANOVA
- One-Way ANOVA Hypothesis Testing
  - Planned Comparisons
  - Post-hoc Testing
  - Effect Size
  - Confidence Intervals
- Graphing
- Additional Considerations
  - Power
  - Welch Robust Procedure

A firm understanding of the topics covered here and previously will be necessary for understanding future topics.

RSS
Research and Statistical Support

This concludes Module 9

Next time Module 10.

- Next time we'll begin covering Regression.
- Until next time; have a nice day.

These slides initially created on: October 15, 2010
These slides last updated on: October 21, 2010

- The bottom date shown is the date this Adobe.pdf file was created; LaTeX[1] has a command for automatically inserting the date of a document's creation.

[1] This document was created in LaTeX using the Beamer package