Hypothesis Testing with One-Way ANOVA

Statistics

Arlo Clark-Foos

Conceptual Refresher

1. Standardized *z* distribution of scores and of means can be represented as percentile rankings.
2. *t* distribution of means, mean differences, and differences between means can all be standardized, allowing us to analyze differences between 2 means
3. Numerator of test statistic is always some difference (between scores, means, mean differences, or differences between means)
4. Denominator represents some measure of variability (or form of standard deviation).

Calculating Refresher

- Test Statistics
  - Numerator = Differences between groups
    - Example: Men are taller than woman
  - Denominator = Variability within groups
    - Example: Not all men/women are the same height

**There is overlap between these distributions.*

$$z = \frac{(M - \mu_M)}{\sigma_M}$$

$$t = \frac{(M - \mu_M)}{s_M}$$

$$t = \frac{\left[(M_X - M_Y) - (\mu_X - \mu_Y)\right]}{s_{Difference}} = \frac{(M_X - M_Y)}{s_{Difference}}$$

"What do you mean, you've outgrown me?"

Analysis of Variance (ANOVA)

TABLE 12-2. Connections Among Distributions

The z distribution is subsumed under the t distributions in certain specific circumstances, and both the z and t distributions are subsumed under the F distributions in certain specific circumstances.

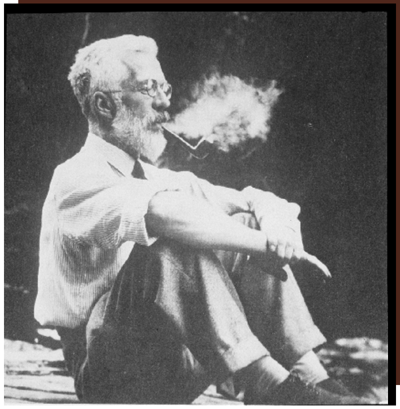
	When Used	Links Among the Distributions
z	One sample; μ and σ are known	Subsumed under the t and F distributions
t	(1) One sample; only μ is known (2) Two samples	Same as z distribution if there is a sample size of ∞ (or just very large)
F	Three or more samples (but can be used with two samples)	Square of z distribution if there are only two samples and a sample size of ∞ (or just very large); square of t distribution if there are only two samples

- *Hypothesis test typically used with one or more nominal IV (with <u>at least</u> 3 groups overall) and an interval DV.*
- *t* Test: Distance between *two distributions*
- *F* ratio: Uses *two* measures of *variability*

F Ratio (Sir Ronald Fisher)

$$F = \frac{\text{between - groups variance}}{\text{within - groups variance}}$$

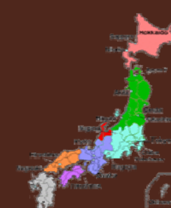
- *Between-Groups Variance: An estimate of the population variance based on the differences among the means of the samples*
- *Within-Groups Variance: An estimate of the population variance based on the differences within each of the three or more sample distributions*



More than two groups

- Example: Speech rates in America, Japan, & Wales

t test?



t test?

t test?

Two Sources of Variance: Between & Within

Problem of Too Many Tests

$$p(A) \text{ AND } p(B) = p(A) \times p(B)$$

$$p(A) \text{ OR } p(B) = p(A) + p(B)$$

- The probability of a Type I error (rejecting the null when the null is true) greatly increases with the number of comparisons.

Fishing Expedition

"If you torture the data long enough,

the numbers will prove anything you want" (Bernstein, 1996)

Problem of Too Many Tests

TABLE 12-1. The Probability of a Type I Error Increases as the Number of Statistical Comparisons Increases

As the number of samples increases, the number of *t* tests necessary to compare every possible pair of means increases at an even greater rate. And with that, the probability of a Type I error quickly becomes far larger than 0.05.

Number of Means	Number of Comparisons	Probability of a Type I Error
2	1	0.05
3	3	0.143
4	6	0.265
5	10	0.401
6	15	0.537
7	21	0.659

Types of ANOVA

- Always preceded by two adjectives
  1. Number of Independent Variables
  2. Experimental Design

- One-Way ANOVA: *Hypothesis test that includes one nominal IV with more than two levels and an interval DV.*

- Within-Groups One -Way ANOVA: *ANOVA where each sample is composed of the same participants (AKA repeated measures ANOVA).*

- Between-Groups One-Way ANOVA: *ANOVA where each sample is composed of different participants.*

Assumptions of ANOVA

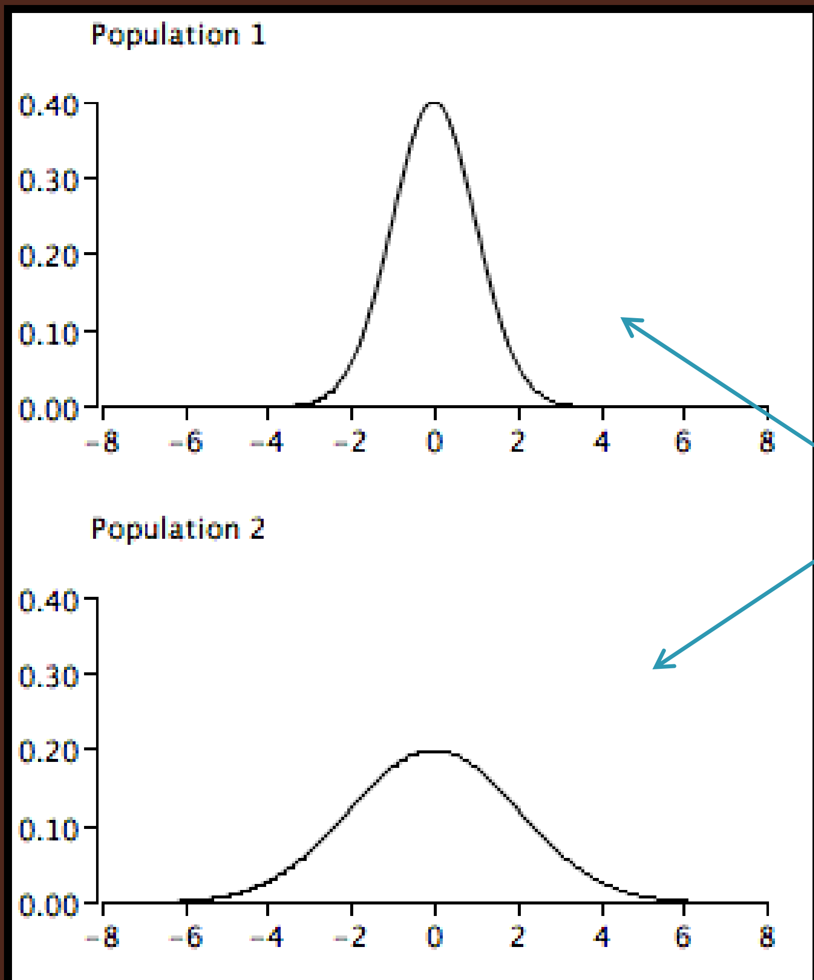
TABLE 10-3. THE ASSUMPTIONS FOR ANOVA: LEARNING AND BENDING THE RULES

We must be aware of the assumptions for ANOVA, and we must be cautious in proceeding with a hypothesis test when our data may not meet all of the assumptions.

LEARNING THE RULES: ASSUMPTIONS	BREAKING THE RULES: WHEN IT IS OK
Data are selected randomly.	OK if cautious about generalizing.
Population is normally distributed.	Usually OK, especially with large sample sizes.
Variances are equal (homoscedasticity, or homogeneity of variance).	*Same-size samples*: OK if largest variance is less than 5 times the smallest. *Different-size samples*: OK if largest variance is less than twice the smallest.

from 1st edition of textbook

Assumption of Homoscedasticity



- *Homoscedastic populations have the same variance*
- *Heteroscedastic populations have different variances*

BACK to the Six Steps

- Research Question:
  - What influences foreign students to choose an American graduate program? In particular, how important are financial aspects to students in Arts & Sciences, Education, Law, & Business?
- Data Source:
  - Survey of 17 graduate students from foreign countries currently enrolled in universities in the U.S.

Importance Scores					
Arts & Sciences	4	5	4	3	4
Education	4	3	4	4	
Law	3	3	2	3	
Business	4	4	4	3	

1. Identify

- Populations: All foreign graduate students enrolled in __________ programs in the U.S.
- Comparison Distribution: *F* distribution
- Test: One-Way Between-Subjects ANOVA
  - Assumptions:
    - Participants not randomly selected
      - Be careful generalizing results
    - Not clear if population dist. are normal. Data are not skewed.
    - Homoscedasticity
      - We will return to this later during calculations—Don't Forget!

2. Hypotheses

- **Null**: Foreign graduate students in Arts & Sciences, Education, Law, and Business all rate financial factors the same, on average.

$$\mu_1 = \mu_2 = \mu_3 = \mu_4$$

- **Research**: Foreign graduate students in Arts & Sciences, Education, Law, and Business *do not* all rate financial factors the same, on average.

$$\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$$

3. Determine characteristics

- > 2 groups and interval DV: *F* distribution

- *df* for each sample: $N_{Sample} - 1$
  - Arts & Sciences: $df_1 = 5 - 1 = \mathbf{4}$
  - Education: $df_2 = 4 - 1 = \mathbf{3}$
  - Law: $df_3 = 4 - 1 = \mathbf{3}$
  - Business: $df_4 = 4 - 1 = \mathbf{3}$

- $df_{Between}$: $N_{Groups} - 1 = 4 - 1 = 3$
  - Numerator *df*

- df_{Within}: $df_1 + df_2 + df_3 + df_4 = 4 + 3 + 3 + 3 = 13$
  - Denominator *df*

4. Determine Critical Values

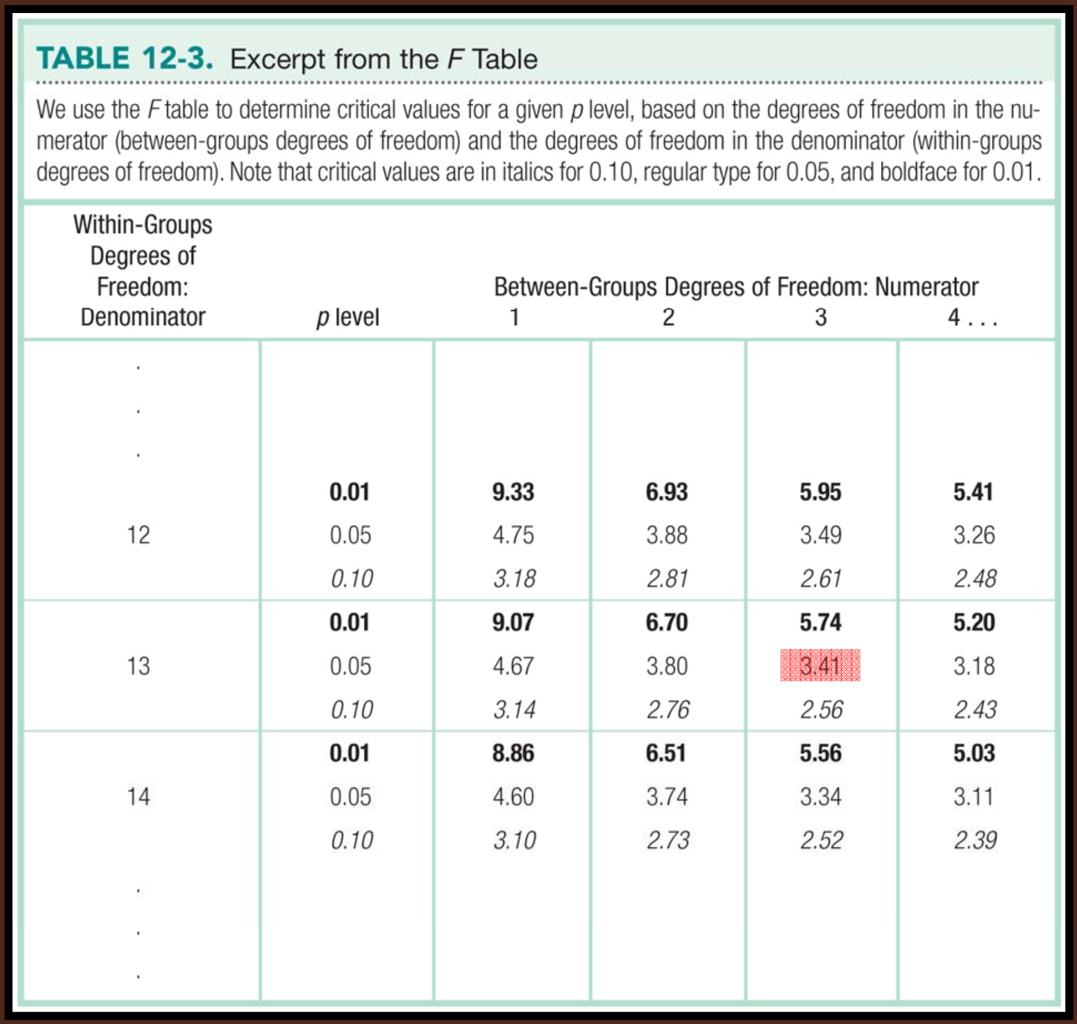


TABLE 12-3. Excerpt from the *F* Table

We use the *F* table to determine critical values for a given *p* level, based on the degrees of freedom in the numerator (between-groups degrees of freedom) and the degrees of freedom in the denominator (within-groups degrees of freedom). Note that critical values are in italics for 0.10, regular type for 0.05, and boldface for 0.01.

Within-Groups Degrees of Freedom: Denominator	*p* level	Between-Groups Degrees of Freedom: Numerator 1	2	3	4 . . .
. . .					
	0.01	**9.33**	**6.93**	**5.95**	**5.41**
12	0.05	4.75	3.88	3.49	3.26
	0.10	*3.18*	*2.81*	*2.61*	*2.48*
	0.01	**9.07**	**6.70**	**5.74**	**5.20**
13	0.05	4.67	3.80	3.41	3.18
	0.10	*3.14*	*2.76*	*2.56*	*2.43*
	0.01	**8.86**	**6.51**	**5.56**	**5.03**
14	0.05	4.60	3.74	3.34	3.11
	0.10	*3.10*	*2.73*	*2.52*	*2.39*
. . .					

$p = .05$

$df_{Between} = 3$

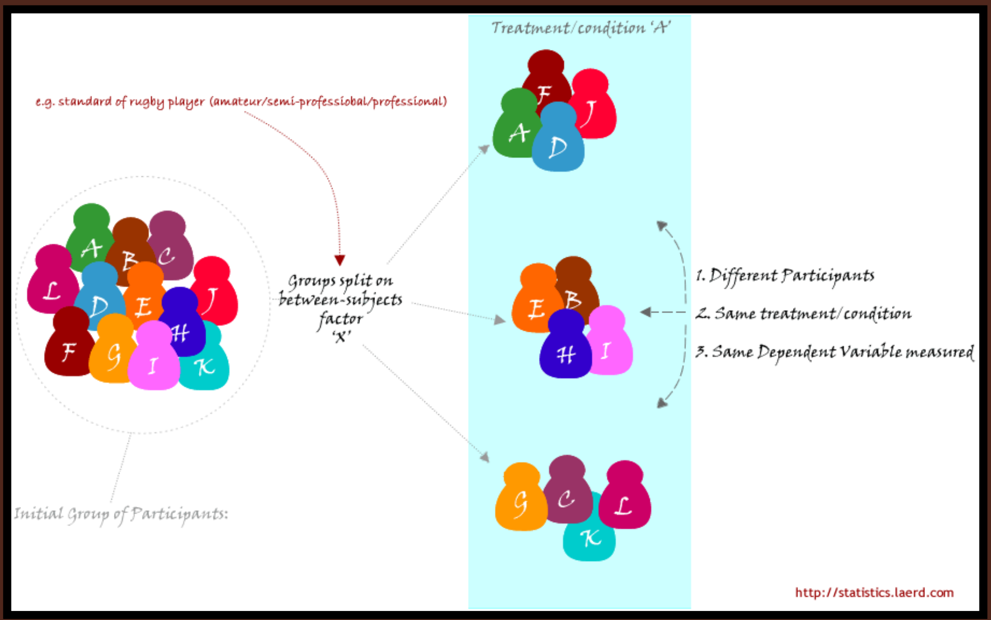
$df_{Within} = 13$

$F_{Critical} = 3.41$

5. Calculate the Test Statistic

- In order to do this, we need 2 measures of variance
  - Between-Groups Variance
  - Within-Groups Variance

  - We will do this shortly…



6. Make a Decision

- If our calculated test statistic exceeds our cutoff, we reject the null hypothesis and can say the following:

"Foreign graduate students studying in the U.S. rate financial factors differently depending on the type of program in which they are enrolled"

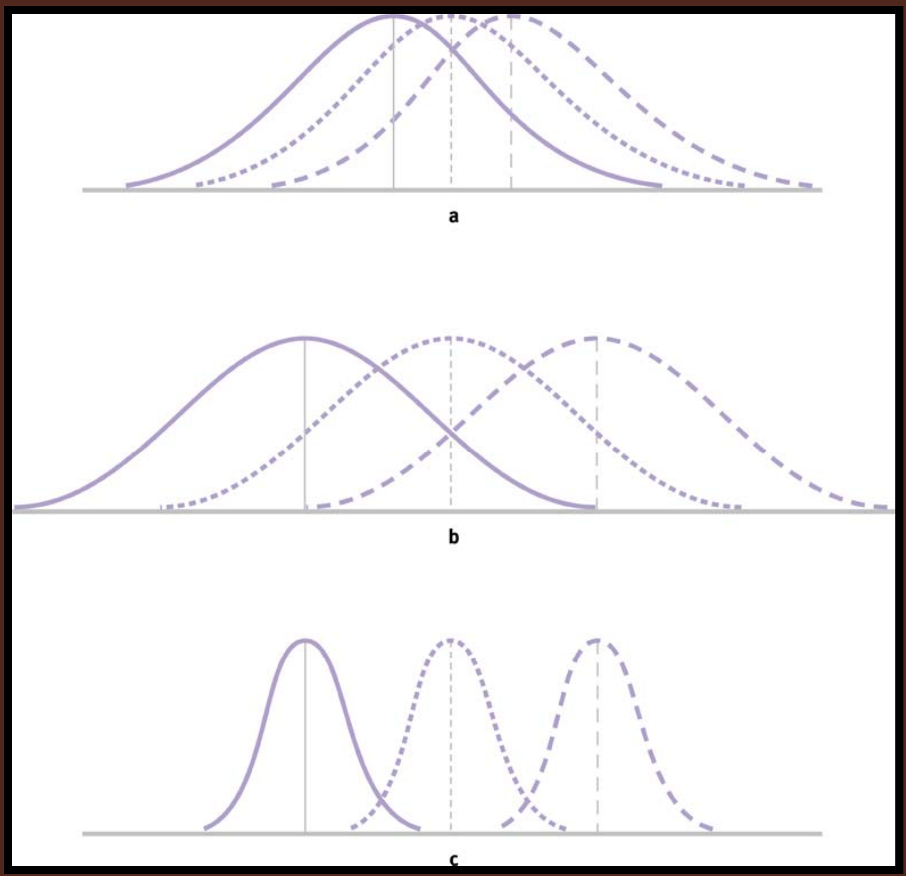
- ANOVA does not tell us where our differences are!
  - We just know that there is a difference somewhere.

Logic of ANOVA: Quantifying Overlap

$$F = \frac{\text{between - groups variance}}{\text{within - groups variance}}$$

- Whenever differences between sample means are large and differences between scores within each sample are small, the *F* statistic will be large.
  - Remember that large test statistics *indicate* statistically significant results

Logic of ANOVA: Quantifying Overlap



a) Large within-groups variability & small between groups variability
b) Large within-groups variability & large between groups variability
c) Small within-groups variability & small between groups variability.
   - Less Overlap!

Logic of ANOVA: Quantifying Overlap

$$F = \frac{\text{between - groups variance}}{\text{within - groups variance}}$$

- If between-groups = within-groups, $F = 1$
- Null hypothesis predicts $F = 1$
  - No differences between groups
- Within-groups variance based on scores, between-groups variance based on means.
  - Need correction.

Calculating the *F* Statistic: The Source Table

- Source Table: *Presents the important calculations and final results of an ANOVA in a consistent and easy-to-read format.*

TABLE 12-4. The Source Table Organizes Our ANOVA Calculations

A source table helps researchers organize the most important calculations necessary to conduct an ANOVA as well as the final results. The numbers 1–5 in the first row are used in this particular table only to help you understand the format of source tables; they would not be included in an actual source table.

1 Source	2 *SS*	3 *df*	4 *MS*	5 *F*
Between	$SS_{between}$	$df_{between}$	$MS_{between}$	F
Within	SS_{within}	df_{within}	MS_{within}	
Total	SS_{total}	df_{total}		

Calculating the *F* Statistic: The Source Table

TABLE 12-4. The Source Table Organizes Our ANOVA Calculations

A source table helps researchers organize the most important calculations necessary to conduct an ANOVA as well as the final results. The numbers 1–5 in the first row are used in this particular table only to help you understand the format of source tables; they would not be included in an actual source table.

1 Source	2 *SS*	3 *df*	4 *MS*	5 *F*
Between	$SS_{between}$	$df_{between}$	$MS_{between}$	F
Within	SS_{within}	df_{within}	MS_{within}	
Total	SS_{total}	df_{total}		

- Col. 1: The sources of variability
- Col. 5: Value of test statistic, *F* ratio
- Col. 4: Mean Square: arithmetic average of squared deviations
- Col. 3: Degrees of freedom
- Col. 2: Sum of Squares

$$MS_{Between} = \frac{SS_{Between}}{df_{Between}}$$

$$MS_{Within} = \frac{SS_{Within}}{df_{Within}}$$

$$F = \frac{MS_{Between}}{MS_{Within}}$$

Sums of Squared Deviations

TABLE 10-6. CALCULATING THE TOTAL SUM OF SQUARES

The total sum of squares is calculated by subtracting the overall mean, called the grand mean, from every score to create deviations, then squaring the deviations and summing the squared deviations.

SAMPLE	X	$(X - GM)$	$(X - GM)^2$
Arts and sciences	4	0.412	0.170
	5	1.412	1.994
$M_{A\&S} = 4.0$	4	0.412	0.170
	3	−0.588	0.346
	4	0.412	0.170
Education	4	0.412	0.170
	3	−0.588	0.346
$M_{Ed} = 3.75$	4	0.412	0.170
	4	0.412	0.170
Law	3	−0.588	0.346
	3	−0.588	0.346
$M_{Law} = 2.75$	2	−1.588	2.522
	3	−0.588	0.346
Business	4	0.412	0.170
	4	0.412	0.170
$M_{Bus} = 3.75$	4	0.412	0.170
	3	−0.588	0.346
	$GM = 3.588$		$SS_{Total} = \mathbf{8.122}$

from 1st edition of textbook

$$SS_{Total} = \Sigma(X - GM)^2$$

Put all of your scores in one column, with samples denoted in another column.

Grand Mean: *Refers to the mean of all scores in a study, regardless of their sample.*

$$GM = \frac{\Sigma(X)}{N_{Total}}$$

Sums of Squared Deviations

TABLE 10-7. CALCULATING THE WITHIN-GROUPS SUM OF SQUARES

The within-groups sum of squares is calculated by taking each score and subtracting the mean of the sample from which it comes—not the grand mean—to create deviations, then squaring the deviations and summing the squared deviations.

SAMPLE	X	$(X - M)$	$(X - M)^2$
Arts and sciences	4	0	0
	5	1	1
$M_{A\&S} = 4.0$	4	0	0
	3	−1	1
	4	0	0
Education	4	0.25	0.063
	3	−0.75	0.563
$M_{Ed} = 3.75$	4	0.25	0.063
	4	0.25	0.063
Law	3	0.25	0.063
	3	0.25	0.063
$M_{Law} = 2.75$	2	−0.75	0.563
	3	0.25	0.063
Business	4	0.25	0.063
	4	0.25	0.063
$M_{Bus} = 3.75$	4	0.25	0.063
	3	−0.75	0.563
	$GM = 3.588$		$SS_{Within} = \mathbf{4.256}$

from 1st edition of textbook

$$SS_{Within} = \Sigma(X - M)^2$$

Calculate the squared deviation of each score from its own particular sample mean

Sums of Squared Deviations

TABLE 10-8. CALCULATING THE BETWEEN-GROUPS SUM OF SQUARES

The between-groups sum of squares is calculated by subtracting the grand mean from the sample mean for every score to create deviations, then squaring the deviations and summing the squared deviations. The individual scores themselves are not involved in any calculations.

SAMPLE	X	$(M - GM)$	$(M - GM)^2$
Arts and sciences	4	0.412	0.170
	5	0.412	0.170
$M_{A\&S} = 4.0$	4	0.412	0.170
	3	0.412	0.170
	4	0.412	0.170
Education	4	0.162	0.026
	3	0.162	0.026
$M_{Ed} = 3.75$	4	0.162	0.026
	4	0.162	0.026
Law	3	−0.838	0.702
	3	−0.838	0.702
$M_{Law} = 2.75$	2	−0.838	0.702
	3	−0.838	0.702
Business	4	0.162	0.026
	4	0.162	0.026
$M_{Bus} = 3.75$	4	0.162	0.026
	3	0.162	0.026
	$GM = 3.588$		$SS_{Between} = \mathbf{3.866}$

from 1st edition of textbook

$$SS_{Between} = \Sigma(M - GM)^2$$

Calculate the squared deviation of each sample mean from the grand mean.

Sums of Squared Deviations

TABLE 10-9. THE THREE SUMS OF SQUARES OF ANOVA

The calculations in ANOVA are built on the foundation we learned in Chapter 2, sums of squared deviations. We calculate three types of sums of squares, one for between-groups variance, one for within-groups variance, and one for total variance. Once we have our three sums of squares, most of the remaining calculations involve simple division.

SUM OF SQUARES	TO CALCULATE THE DEVIATIONS, SUBTRACT THE . . .	FORMULA
Between-groups	grand mean from the sample mean (for each score)	$SS_{Between} = \Sigma(M - GM)^2$
Within-groups	sample mean from each score	$SS_{Within} = \Sigma(X - M)^2$
Total	grand mean from each score	$SS_{Total} = \Sigma(X - GM)^2$

TABLE 10-10. A SOURCE TABLE WITH FORMULAS

This table summarizes the formulas for calculating an *F* statistic.

SOURCE	*SS*	*df*	*MS*	*F*
Between	$\Sigma(M - GM)^2$	$N_{Groups} - 1$	$\frac{SS_{Between}}{df_{Between}}$	$\frac{MS_{Between}}{MS_{Within}}$
Within	$\Sigma(X - M)^2$	$df_1 + df_2 + \ldots + df_{Last}$	$\frac{SS_{Within}}{df_{Within}}$	
Total	$\Sigma(X - GM)^2$	$N_{Total} - 1$		

[Expanded formula: $df_{Within} = (N_1 - 1) + (N_2 - 1) + \ldots + (N_{Last} - 1)$]

from 1st edition of textbook

Source Table for our Example

TABLE 10-11. A COMPLETED SOURCE TABLE

Once we've calculated the sums of squares and the degrees of freedom, the rest is just simple division. We use the first two columns of numbers to calculate the variances and the *F* statistic. We divide the between-groups sum of squares and within-groups sum of squares by their associated degrees of freedom to get the between-groups variance and within-groups variance. Then we divide between-groups variance by within-groups variance to get the *F* statistic, 3.94.

SOURCE	*SS*	*df*	*MS*	*F*
Between	3.866	3	1.289	**3.94**
Within	4.256	13	0.327	
Total	8.122	16		

from 1st edition of textbook

What is our decision?

TABLE 10-12. CALCULATING SAMPLE VARIANCES

We calculate the variances of the samples by dividing each sum of squares by the sample size minus 1 to check one of the assumptions of ANOVA. For unequal sample sizes, as we have here, we want our largest variance (0.500 in this case) to be no more than twice our smallest (0.251 in this case). Two times 0.251 is 0.502, and so we meet this assumption.

SAMPLE	ARTS AND SCIENCES	EDUCATION	LAW	BUSINESS
Squared deviations:	0	0.063	0.063	0.063
	1	0.563	0.063	0.063
	0	0.063	0.563	0.063
	1	0.063	0.063	0.563
	0			
Sum of squares:	2	0.752	0.752	0.752
$N - 1$	4	3	3	3
Variance	**0.500**	**0.251**	**0.251**	**0.251**

from 1st edition of textbook

- Back to Step 1.
  - ***Homoscedasticity***

- Because the largest variance (.500) is not more than twice (unequal sample sizes) the smallest variance (.251) then we have met this assumption.

What is our decision?

- Step 6. Make a decision

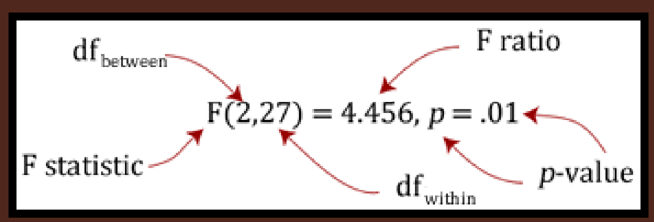
$$F = 3.94 \;\; > \;\; F_{crit} = 3.41$$

  - We can reject the null hypothesis. There is (are) a difference somewhere.
  - Where?
  - post-hoc test: *Statistical procedure frequently carried out after we reject the null hypothesis in an ANOVA; it allows us to make multiple comparisons among several means.*
    - post-hoc: Latin for "after this"
    - Examples: Tukey's HSD, Scheffe, Dunnet, Duncan, Bonferroni…

Reporting ANOVA in APA Style

1. Italic letter *F*: F
2. Open parenthesis : $F($
3. Between Groups df then comma: $F(df_{Between},$
4. Within Groups df: $F(df_{Between}, df_{Within})$
5. Close parentheses, equal sign: $F(df_{Between}, df_{Within}) =$
6. *F* Statistic then comma: $F(df_{Between}, df_{Within}) = 1.23,$
7. Lower case, italic letter p: $F(df_{Between}, df_{Within}) = 1.23, p$
8. Significant, less than .05: $F(df_{Between}, df_{Within}) = 1.23, p < .05$
   - OR non significant: $F(df_{Between}, df_{Within}) = 1.23, p > .05$
   - OR exact p value: $F(df_{Between}, df_{Within}) = 1.23, p = .02$

Another example:



Between-Subjects One Way ANOVA

Example: Memory for Emotional Stimuli

Between-Subjects One Way ANOVA: Memory for Emotional Stimuli

Do you have differences in memory for emotional vs. neutral events?

- Do others have the same differences or is it something unique to *you*?
- Let's find out…

- Research Question: Will people asked to study pure lists of either positive, negative, or neutral pictures have differences in recall of those pure lists?

- Research Design: We asked 17 participants study one single list of either 30 positive, 30 negative, or 30 neutral pictures (from IAPS). Following a brief delay all participants were asked to recall as many of the 30 studied photos as they could. These data are on the following slide.

Between-Subjects One Way ANOVA: Memory for Emotional Stimuli

Already Stated: N_{Total} = 17, one IV with 3 levels (Emotion) is between-sub.

Below are the proportion of pictures on their studied lists that each participant successfully recalled (100% = perfect memory):



0.69	0.59	.64
0.84	0.64	.73
0.93	0.62	.51
0.91	0.71	.68
0.89	0.50	.61
0.90	0.60	
***M* = .86**	***M* = .61**	***M* = .634**

Between-Subjects One Way ANOVA: Memory for Emotional Stimuli

Already Stated/Calculated		
$N_{Total} = 17$		
$N_{Neg} = 6$	$N_{Neut} = 6$	$N_{Pos} = 5$
$df_{Neg} = 5$	$df_{Neut} = 5$	$df_{Pos} = 4$
$df_{Between} = 2$		
$df_{Within} = 14$		
$M_{Neg} = .86$	$M_{Neut} = .61$	$M_{Pos} = .634$

- Six Steps to Hypothesis Testing…again!
  1. Population: All memories for negative, neutral, and positive events.

     Comparison Distribution: *F* distribution

     Test: One-Way Between-Subjects ANOVA
     - Assumptions:
       - Participants were randomly selected from subject pool
       - Not clear if population dist. are normal. Data are not skewed.
       - Homoscedasticity

Between-Subjects One Way ANOVA: Memory for Emotional Stimuli

2. Hypotheses

Null: On average, memories for negative, neutral, and positive pictures will not differ.

$$\mu_{Neg} = \mu_{Neut} = \mu_{Pos}$$

Research: On average, memories for negative, neutral, and positive pictures will be different.

$$\mu_{Neg} \neq \mu_{Neut} \neq \mu_{Pos}$$

Between-Subjects One Way ANOVA: Memory for Emotional Stimuli

3. Determine characteristics
   - > 2 groups and interval DV: *F* distribution



0.69	0.59	.64
0.84	0.64	.73
0.93	0.62	.51
0.91	0.71	.68
0.89	0.50	.61
0.90	0.60	
M = .86	M = .61	M = .634
s^2 = .00784	s^2 = .00472	s^2 = .00683

Between-Subjects One Way ANOVA: Memory for Emotional Stimuli

Digression: Test for Homoscedasticity

Rule

If sample sizes differ across conditions, largest variance must not be more than twice (2x) the smallest variance

0.69	0.59	.64
0.84	0.64	.73
0.93	0.62	.51
0.91	0.71	.68
0.89	0.50	.61
0.90	0.60	
M = .86	M = .61	M = .634
s^2 = .00784	s^2 = .00472	s^2 = .00683
.00784	.0047 * 2 =.00944	
.00784 < .00944 so this assumption is met.		

Between-Subjects One Way ANOVA: Memory for Emotional Stimuli

4. Determine critical values

WITHIN-GROUPS df	SIGNIFICANCE (p) LEVEL	1	2	3	4	5	6
		BETWEEN-GROUPS DEGREES OF FREEDOM					
12	.01	9.33	6.93	5.95	5.41	5.07	4.82
	.05	4.75	3.89	3.49	3.26	3.11	3.00
	.10	3.18	2.81	2.61	2.48	2.40	2.33
13	.01	9.07	6.70	5.74	5.21	4.86	4.62
	.05	4.67	3.81	3.41	3.18	3.03	2.92
	.10	3.14	2.76	2.56	2.43	2.35	2.28
14	.01	8.86	6.52	5.56	5.04	4.70	4.46
	.05	4.60	3.74	3.34	3.11	2.96	2.85
	.10	3.10	2.73	2.52	2.40	2.31	2.24
15	.01	8.68	6.36	5.42	4.89	4.56	4.32
	.05	4.54	3.68	3.29	3.06	2.90	2.79
	.10	3.07	2.70	2.49	2.36	2.27	2.21
16	.01	8.53	6.23	5.29	4.77	4.44	4.20
	.05	4.49	3.63	3.24	3.01	2.85	2.74
	.10	3.05	2.67	2.46	2.33	2.24	2.18
17	.01	8.40	6.11	5.19	4.67	4.34	4.10
	.05	4.45	3.59	3.20	2.97	2.81	2.70
	.10	3.03	2.65	2.44	2.31	2.22	2.15
18	.01	8.29	6.01	5.09	4.58	4.25	4.02
	.05	4.41	3.56	3.16	2.93	2.77	2.66
	.10	3.01	2.62	2.42	2.29	2.20	2.13
19	.01	8.19	5.93	5.01	4.50	4.17	3.94
	.05	4.38	3.52	3.13	2.90	2.74	2.63

Already Stated/Calculated		
$N_{Total} = 17$		
$N_{Neg} = 6$	$N_{Neut} = 6$	$N_{Pos} = 5$
$df_{Neg} = 5$	$df_{Neut} = 5$	$df_{Pos} = 4$
$df_{Between} = 2$		
$df_{Within} = 14$		
$M_{Neg} = .86$	$M_{Neut} = .61$	$M_{Pos} = .634$
$s^2 = .00784$	$s^2 = .00472$	$s^2 = .00683$

$F_{crit} = 3.74$

Between-Subjects One Way ANOVA: Memory for Emotional Stimuli

$$GM = \frac{\Sigma(X)}{N_{Total}}$$

5. Calculate a test statistic…

Source	*SS*	*df*	*MS*	*F*
Between		2		
Within		14		
Total		16		

$$SS_{Within} = \Sigma\left(X - M\right)^2$$

$$SS_{Between} = \Sigma\left(M - GM\right)^2$$

$$SS_{Total} = \Sigma\left(X - GM\right)^2$$

Between-Subjects One Way ANOVA: Memory for Emotional Stimuli

5. Calculate a test statistic…

$$SS_{Total} = \Sigma(X - GM)^2$$

$$GM = \frac{\Sigma(X)}{N_{Total}}$$

$GM = .7053$

X	$(X - GM)$	$(X - GM)^2$
0.69	-0.02	0.0002
0.84	0.135	0.0181
0.93	0.225	0.0505
0.91	0.205	0.0419
0.89	0.185	0.0341
0.90	0.195	0.0379
0.59	-0.12	0.0133
0.64	-0.07	0.0043
0.62	-0.09	0.0073
0.71	0.005	0.0
0.50	-0.21	0.0421
0.60	-0.11	0.0111
0.64	-0.07	0.0043
0.73	0.025	0.0006
0.51	-0.2	0.0381
0.68	-0.03	0.0006
0.61	-0.1	0.0091

$SS_{Total} = .3135$

Between-Subjects One Way ANOVA: Memory for Emotional Stimuli

5. Calculate a test statistic…

$$SS_{Within} = \Sigma(X - M)^2$$

X	$(X - M)$	$(X - M)^2$
0.69	-0.17	0.0289
0.84	-0.02	0.0004
0.93	0.07	0.0049
0.91	0.05	0.0025
0.89	0.03	0.0009
0.90	0.04	0.0016
0.59	-0.02	0.0004
0.64	0.03	0.0009
0.62	0.01	0.0001
0.71	0.1	0.01
0.50	-0.11	0.0121
0.60	-0.01	0.0001
0.64	0.006	0
0.73	0.096	0.0092
0.51	-0.124	0.0154
0.68	0.046	0.0021
0.61	-0.024	0.0006

$M_{Neg} = .86$

$M_{Neut} = .61$

$M_{Pos} = .634$

$SS_{Within} = .0901$

Between-Subjects One Way ANOVA: Memory for Emotional Stimuli

5. Calculate a test statistic…

$$SS_{Between} = \Sigma(M - GM)^2$$

$GM = .7053$

X	M	$(M - GM)$	$(M - GM)^2$
0.69	0.86	0.155	0.024
0.84	0.86	0.155	0.024
0.93	0.86	0.155	0.024
0.91	0.86	0.155	0.024
0.89	0.86	0.155	0.024
0.90	0.86	0.155	0.024
0.59	0.61	-0.1	0.009
0.64	0.61	-0.1	0.009
0.62	0.61	-0.1	0.009
0.71	0.61	-0.1	0.009
0.50	0.61	-0.1	0.009
0.60	0.61	-0.1	0.009
0.64	0.634	-0.07	0.005
0.73	0.634	-0.07	0.005
0.51	0.634	-0.07	0.005
0.68	0.634	-0.07	0.005
0.61	0.634	-0.07	0.005

$SS_{Between} = .223$

Between-Subjects One Way ANOVA: Memory for Emotional Stimuli

5. Calculate a test statistic…

Source	*SS*	*df*	*MS*	*F*
Between	.223	2	.1115	17.969
Within	.0901	14	.0064	
Total	~.3135	16		

$$MS_{Between} = \frac{SS_{Between}}{df_{Between}}$$

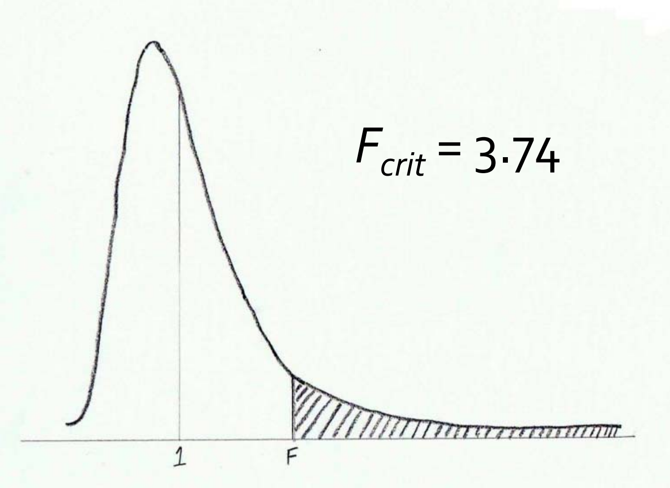
$$MS_{Within} = \frac{SS_{Within}}{df_{Within}}$$

$$F = \frac{MS_{Between}}{MS_{Within}}$$

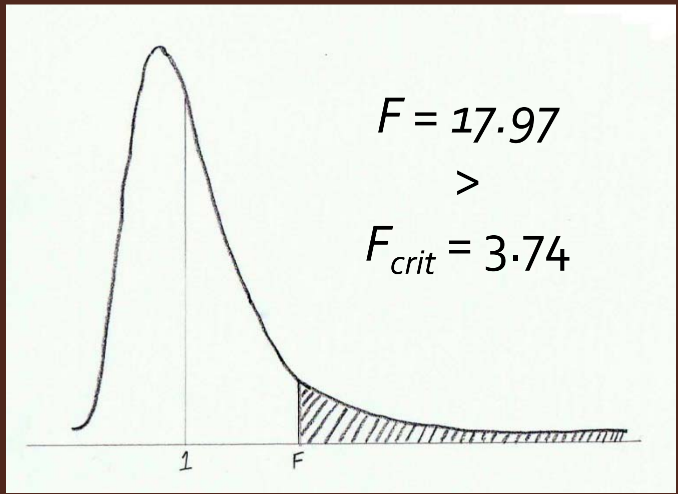
Between-Subjects One Way ANOVA: Memory for Emotional Stimuli

6. Make a decision

Source	*SS*	*df*	*MS*	*F*
Between	.223	2	.1115	17.969
Within	.0901	14	.0064	
Total	~.3135	16		



Between-Subjects One Way ANOVA: Memory for Emotional Stimuli



6. Make a decision

Recall of negative, neutral, and positive pictures was different, $F(2, 14) = 19.97$, $p < .05$.

But which pictures were remembered best? Worst?

A Priori & Post-Hoc Tests

Hindsight is 20-20

- Although your data may suggest a new relationship, and thus new analyses…
- Theory should guide research and thus comparisons should be decided on before you conduct your experiment.

With hindsight, maybe Colin didn't pick the best time to cover up his acne with a balaclava!

Planned & A Priori Comparisons

- Based on literature review
  - Theoretical

- Planned comparisons
  - *A test that is conducted when there are multiple groups of scores, but specific comparisons have been specified prior to data collection.*
    - A Priori Comparisons

Planned & A Priori Comparisons

- If you have planned comparisons…
  - Just run *t* tests
    - Subjective Decision about *p* value
      - $p = .05$?
      - $p = .01$?
      - Bonferroni Correction?

Post-Hoc: Tukey HSD

- Tukey Honestly Significant Difference
  - *Determines differences between means in terms of standard error*
    - 'Honest' because we adjust for making multiple comparisons
    - The HSD is compared to a critical value
  - Overview
    1. Calculate differences between a pair of means
    2. Divide this difference by the standard error

    * Basically this is a variant of a *t* test *

    Oh no, that means the six steps again…sort of.

Tukey HSD

$$HSD = \frac{\left(M_1 - M_2\right)}{s_M}$$

$$t = \frac{\left(M_1 - M_2\right)}{s_{Difference}}$$

- For Tukey HSD, standard error is calculated differently depending on whether your sample sizes are equal or not.

Tukey HSD

- Equal Sample Sizes

$$s_M = \sqrt{\frac{MS_{Within}}{N}}$$

$N = $ Sample size within each group

- Unequal Sample Sizes

$$s_M = \sqrt{\frac{MS_{Within}}{N'}}$$

$$N' = \frac{N_{Groups}}{\sum\left(\frac{1}{N}\right)}$$

Tukey HSD

- Determine Critical Value from Table
- Make a Decision
- Let's go back to our memory for emotional pictures example…

Tukey HSD: Example

- Memory for Emotional Pictures Example:

  Between-Subjects One Way ANOVA
  - Decision: Recall of negative, neutral, and positive pictures was different, $F(2, 14) = 19.97$, $p < .05$..
  - Where are our differences?
  - Let's get our q_{crit} first…

Tukey HSD: Example

TABLE B.5 THE q STATISTIC (TUKEY *HSD* TEST)

WITHIN-GROUPS df	SIGNIFICANCE (p) LEVEL	k = NUMBER OF TREATMENTS (LEVELS): 2	3	4	5	6	7	8	9	10	11	12
5	.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32
	.01	**5.70**	**6.98**	**7.80**	**8.42**	**8.91**	**9.32**	**9.67**	**9.97**	**10.24**	**10.48**	**10.70**
6	.05	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65	6.79
	.01	**5.24**	**6.33**	**7.03**	**7.56**	**7.97**	**8.32**	**8.61**	**8.87**	**9.10**	**9.30**	**9.48**
7	.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43
	.01	**4.95**	**5.92**	**6.54**	**7.01**	**7.37**	**7.68**	**7.94**	**8.17**	**8.37**	**8.55**	**8.71**
8	.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18
	.01	**4.75**	**5.64**	**6.20**	**6.62**	**6.96**	**7.24**	**7.47**	**7.68**	**7.86**	**8.03**	**8.18**
9	.05	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87	5.98
	.01	**4.60**	**5.43**	**5.96**	**6.35**	**6.66**	**6.91**	**7.13**	**7.33**	**7.49**	**7.65**	**7.78**
10	.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83
	.01	**4.48**	**5.27**	**5.77**	**6.14**	**6.43**	**6.67**	**6.87**	**7.05**	**7.21**	**7.36**	**7.49**
11	.05	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71
	.01	**4.39**	**5.15**	**5.62**	**5.97**	**6.25**	**6.48**	**6.67**	**6.84**	**6.99**	**7.13**	**7.25**
12	.05	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51	5.61
	.01	**4.32**	**5.05**	**5.50**	**5.84**	**6.10**	**6.32**	**6.51**	**6.67**	**6.81**	**6.94**	**7.06**
13	.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43	5.53
	.01	**4.26**	**4.96**	**5.40**	**5.73**	**5.98**	**6.19**	**6.37**	**6.53**	**6.67**	**6.79**	**6.90**
14	.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46
	.01	**4.21**	**4.89**	**5.32**	**5.63**	**5.88**	**6.08**	**6.26**	**6.41**	**6.54**	**6.66**	**6.77**
15	.05	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40
	.01	**4.17**	**4.84**	**5.25**	**5.56**	**5.80**	**5.99**	**6.16**	**6.31**	**6.44**	**6.55**	**6.66**

Already Stated/Calculated		
$N_{Total} = 17$		
$N_{Neg} = 6$	$N_{Neut} = 6$	$N_{Pos} = 5$
$df_{Neg} = 5$	$df_{Neut} = 5$	$df_{Pos} = 4$
$df_{Between} = 2$ $(k = 3)$		
$df_{Within} = 14$		
$M_{Neg} = .86$	$M_{Neut} = .61$	$M_{Pos} = .634$

$q_{crit} = \pm 3.70$

Tukey HSD: Example

0.69	0.59	.64
0.84	0.64	.73
0.93	0.62	.51
0.91	0.71	.68
0.89	0.50	.61
0.90	0.60	

Already Stated/Calculated		
$N_{Total} = 17$		
$N_{Neg} = 6$	$N_{Neut} = 6$	$N_{Pos} = 5$
$df_{Neg} = 5$	$df_{Neut} = 5$	$df_{Pos} = 4$
$df_{Between} = 2$ $(k = 3)$		
$df_{Within} = 14$		
$M_{Neg} = .86$	$M_{Neut} = .61$	$M_{Pos} = .634$

$q_{crit} = \pm 3.70$

Source	*SS*	*df*	*MS*	*F*
Between	.223	2	.1115	17.969
Within	.0901	14	.0064	
Total	~.3135	16		

Tukey HSD: Example

- Standard Error: Unequal Sample Sizes

$$N' = \frac{N_{Groups}}{\sum\left(\frac{1}{N}\right)}$$

$$N' = \frac{3}{\frac{1}{6}+\frac{1}{6}+\frac{1}{5}} = \frac{3}{.533} = 5.625$$

$$s_M = \sqrt{\frac{MS_{Within}}{N'}}$$

$$s_M = \sqrt{\frac{.0064}{5.625}} = \sqrt{.0011378} = 0.034$$

Tukey HSD: Example

- Negative (*M*=0.86) vs. Neutral (*M*=0.61)

$$HSD = \frac{(M_1 - M_2)}{s_M} = \frac{(.86 - .61)}{.034} = 7.35$$

- Negative (*M*=0.86) vs. Positive (*M*=0.634)

$$HSD = \frac{(M_1 - M_2)}{s_M} = \frac{(.86 - .634)}{.034} = 6.65$$

- Neutral (*M*=0.61) vs. Positive (*M*=0.634)

$$HSD = \frac{(M_1 - M_2)}{s_M} = \frac{(.61 - .634)}{.034} = -0.71$$

Tukey HSD: Example

- Make a Decision
  - Post hoc comparisons using the Tukey HSD test revealed that negative pictures were better remembered (M = .86) than either positive (M = .634) or neutral (M = .61) pictures, with no differences between the latter two.

Bonferonni Correction

An alternative post-hoc strategy

Bonferroni Correction

Fishing Expedition

- Remember the problem of too many tests?
  - Inflates the risk of a Type I error.
    - False positives
  - Is there a way to address that without a new test?
    - We've hinted at it already…

Bonferroni Correction

TABLE 12-14. The Bonferroni Test: Few Groups, Many Comparisons

Even with a few means, we must make many comparisons to account for every possible difference. Because we run the risk of incorrectly rejecting the null hypothesis just by chance if we run so many tests, it is a wise idea to use a more conservative procedure, such as the Bonferroni test, when comparing means. The Bonferroni test requires that we divide an overall p level, such as 0.05, by the number of comparisons we will make.

Number of Means	Number of Comparisons	Bonferroni p Level (overall $p = 0.05$)
2	1	0.05
3	3	0.017
4	6	0.008
5	10	0.005
6	15	0.003
7	21	0.002

Summary

- Between-Subjects One Way ANOVA
  - Two Sources of Variance
    - New *Sums of Squares*
    - New *df*
  - Homoscedasticity
  - The problem of too many tests
  - Source Table
- Post-Hoc tests
  - Tukey's HSD
  - Bonferroni
  - LSD
  - etc.