SAN JOSÉ STATE UNIVERSITY

College of Social Work
S. W. 242
Spring 2008
Edward Cohen

Week 9
March 21, 2008

- **Note on your calendar: I'll be out of town Thursday April 3 (no office hours)**
- **Let's slow down—nagging questions, unfinished business, general confusion?**
- **Discuss Final Paper (don't slow down too much!)**
- **Prep for ANOVA**
- **Happy Spring Break!**

Concepts you should know:

- When to use ANOVA
- One-way ANOVA
- Two-way ANOVA (or Two-factor ANOVA)
- *F* statistic
- Within-groups variation
- Between-groups variation
- *Post-hoc* tests
- Multiple comparisons

I. When to use ANOVA?

A. Remember that the *t*-test compared means (of ratio or interval level variables) along the attributes of a *dichotomous* nominal variable (e.g. a variable with only two attributes such as male and female), i.e. comparing only two means at a time

B. ANOVA is a distant cousin to the *t*-test. It compares means for categorical (*nominal or ordinal*) variables with *three or more attributes*, i.e. it compares three or more means, for example

1. Are there differences between the mean LCSW licensing scores for MSW graduates of various ethnicities?

2. For children residing in residential treatment does the number of behavioral outbursts vary by level of self esteem (High, Medium, Low)?

3. In an experimental study of treatment for depression, clients are randomized to the following groups: 1) medication only, 2) medication plus individual psychotherapy, and 3) psychotherapy only. Do the clients in the "medication plus psychotherapy" group show more improved outcomes (lower depression scores) than the other two groups?

II. How it works

A. ANOVA is an "Omnibus test"

1. "Omnibus" is an adjective defined as *covering many things or classes*, in this case it makes a global statement about the Null hypothesis.

2. ANOVA only tests the Null hypothesis of "no difference between the means" , **or** "the means of the dependent variables are equal among the groups"– in other words ANOVA tells us whether or not any differences among the means might have been produced due to sampling error

3. The ANOVA statistic and associated *p* value does <u>not</u> prove or disprove your research hypothesis by singling out one of the means as "significantly different than the others" – a statistically significant finding merely says that the Null can be rejected and "the means are different"(or "the means are not equal")

4. The ANOVA statistic is called the "*F* ratio". It has the same function as the *t* statistic and the Chi Square value, and it has its own distribution table (built into SPSS) so it can also be associated with a *p* value. Except the *F* ratio is easier to interpret:

> The *F* ratio reflects the variation of means *between* the groups
> divided by the variation of means *within* the groups.

Why do we need this ratio? Because we need some way to tease out the variation *within* the groups compared to the variation *between* the groups, in order to get a better understanding of the variation *between* the groups. (It's not as simple as just calculating the difference between the means, as the *t*-test does.)

[Technical note: the *F* compares the "sums of squares" (remember the Variance calculation?) for each group with that of the total, or "grand" mean. The ratio is calculated as MEAN SQUARES OF BETWEEN-GROUPS divided by MEAN SQUARES WITHIN GROUPS]

5. Rejecting the Null hypothesis "no difference in means between the groups" is the same process for the other statistical tests we've covered:

(In case you have any room left on your refrigerator door)

> If the *F* ratio falls within the critical region as determined by our alpha, then the actual *p* value will be equal to or less than the alpha. So, if our alpha was pre-determined to be .05, then an actual *p* value $\leq$ .05 from the ANOVA procedure will allow us to reject the Null in favor of the alternative hypothesis.

B. Answering your research question

1. Once you determine that the Null can be rejected, you still have to answer your research question, just like the *t*-test, except you can't just look at the differences

between two means, you have to compare the differences among <u>all</u> (three or more) means. (But if the *F* ratio was not significant, *game over!*)

2. This involves another level of statistical tests called ***post-hoc* tests**, which determine <u>where</u> the significant differences are, i.e.

   a. Group 1 mean compared to group 2 mean,

   b. Group 1 mean compared to group 3 mean, and

   c. Group 2 mean compared to group 3 mean

Three comparisons for three group means.

3. If we had four groups,

Group	Compared to Group
1	2
1	3
1	4
2	3
2	4
3	4

Six comparisons for four groups

4. There are a few different types of *post-hoc* tests (such as the Bonferroni, the Scheffe (pronounced Shef-fay') and the Tukey procedures). We will use one of them in the example and the labs—the Tukey procedure.

5. Why not just do separate *t*-tests for each pair? **(Hint: it has to do with the probability of making a Type I error.)** An alpha=.05 is also the probability of making a Type I error—rejecting the Null when it should <u>not</u> have been rejected. Even if the Null is really true, there's still a 5% probability that significant differences between means will be found purely due to sampling error—at a 5% error rate. You compound this probability of error by adding more comparisons. E.g.: for 20 groups, there are 190 comparisons of means, or 20*(20-1)/2. Five % of 190 is 9.5. So by doing separate *t*-tests for 190 comparisons, you could likely have 9 or 10 of them be significant *purely by chance.*

6. The *post-hoc* procedure adjusts for the inflated risk of making a Type I error, so that the <u>combined</u> probability of falsely rejecting the Null (Type I error) among all the comparisons is no more than your intended alpha (such as .05).

7. The *post-hoc* tests the Null Hypothesis that there are no differences between the two means in each comparison (similar logic as in the *t*-test, except for the probability adjustment stated above).

III. Example of One-Way ANOVA (from Kirkpatrick & Feeney, Ch. 10)

Consider this dataset of sleep patterns (percentage of time in delta, or deep, sleep) measured over three groups of attachment styles in children: 1=secure, 2=anxious, and 3=avoidant. The study explores the relationship between attachment style and delta sleep, i.e. hypothesizing that children who are insecure in their attachments to parental figures are likely to spend less time in delta sleep compared to their more secure counterparts so that they can be more instinctually attentive to dangers in the environment. The “Delta” variable is the percentage of time in delta, or deep sleep.

Subject	Attachment Style “”attstyle”	Delta Sleep “delta”		Subject	Attachment Style “”attstyle”	Delta Sleep “delta”		Subject	Attachment Style “”attstyle”	Delta Sleep “delta”
1	1	21		11	2	17		21	3	18
2	1	21		12	2	17		22	3	20
3	1	25		13	2	15		23	3	18
4	1	23		14	2	15		24	3	19
5	1	24		15	2	15		25	3	17
6	1	23		16	2	14		26	3	17
7	1	23		17	2	20		27	3	15
8	1	22		18	2	13		28	3	16
9	1	22		19	2	14		29	3	17
10	1	22		20	2	19		30	3	18

Enter the data and set up the variables using the shortened variable names shown in the column headings and the correct levels of measurement.

Once data are entered and variables set up, select the appropriate statistical test and alpha level: **One-way ANOVA, alpha = .05**. **One-way ANOVA** is used when there is only one independent variable. **Two-way or Two-factor ANOVA** involves two independent variables.

a. In SPSS, click **Analyze, Compare Means, One-way ANOVA**
b. Move “attstyle” into the **Factor** box and the “delta” variable in the **Dependent List** box
c. Click **Options** and then select **Descriptive** in the **Statistics** section, and click **Continue**.
d. Click **Post Hoc**, select **Tukey**, and click **Continue**. Then click **OK**.
e. See Output screen (remember to scroll down to bottom for most recent output!)

Eight Steps for Hypothesis Testing:

1. What’s the independent variable and level of measurement? _____________, _____________

2. What’s the dependent variable and level of measurement? _______________, _____________

3. State the Null Hyphothesis (H_O):

4. State the Alternative Hypothesis (H_A):

5. Select appropriate statistical test and alpha level:

6. Review SPSS table of results:

Oneway

Descriptives

Percentage of delta (deep) sleep

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Secure	10	22.6000	1.26491	.40000	21.6951	23.5049	21.00	25.00
Anxious	10	15.9000	2.28279	.72188	14.2670	17.5330	13.00	20.00
Avoidant	10	17.5000	1.43372	.45338	16.4744	18.5256	15.00	20.00
Total	30	18.6667	3.34595	.61088	17.4173	19.9161	13.00	25.00

This table is the report on the means—necessary later for interpreting results!

ANOVA

Percentage of delta (deep) sleep

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	244.867	2	122.433	41.425	.000
Within Groups	79.800	27	2.956		
Total	324.667	29			

This table reports on the *F* ratio. (What's the relationship between the F and the Mean Squares, give or take due to rounding in the complicated formula?)

Post Hoc Tests

Multiple Comparisons

Percentage of delta (deep) sleep

Tukey HSD

(I) Attachment style type	(J) Attachment style type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Secure	Anxious	6.70000*	.76884	.000	4.7937	8.6063
	Avoidant	5.10000*	.76884	.000	3.1937	7.0063
Anxious	Secure	-6.70000*	.76884	.000	-8.6063	-4.7937
	Avoidant	-1.60000	.76884	.113	-3.5063	.3063
Avoidant	Secure	-5.10000*	.76884	.000	-7.0063	-3.1937
	Anxious	1.60000	.76884	.113	-.3063	3.5063

*. The mean difference is significant at the 0.05 level.

This table shows the Multiple Comparison results. How would you interpret it?

Homogeneous Subsets

Percentage of delta (deep) sleep

Tukey HSD

Attachment style type	N	Subset for alpha = 0.05	
		1	2
Anxious	10	15.9000	
Avoidant	10	17.5000	
Secure	10		22.6000
Sig.		.113	1.000

Means for groups in homogeneous subsets are displayed.

This table shows subsets of means that do not differ from one another, confirming the previous table of multiple comparisons. (We can ignore this table for now.)

8. Results:

The *F* value of 41.425 is associated with a *p*-value of less than .001, which is less than the alpha level of .05. Therefore the difference among delta mean scores is significant and we reject the null hypothesis. From the Scheffe *post-hoc* test, we found that the secure group's delta sleep scores are significantly higher than both the avoidant group's and anxious group's mean scores.

Results: [In paper writing format]

From a One-Way Independent-Groups ANOVA, a significant difference was found in mean delta sleep scores among the three attachment groups ($F = 41.425$, $p = .001$). From the Tukey *post-hoc* test, the secure group's mean delta score of 22.6 (sd = 1.26) was significantly greater than both the anxious group's mean score of 15.9 (sd = 2.28) and the avoidant group's mean score of 17.5 (sd = 1.43). There was no significant difference between the anxious and avoidant group delta scores.

9. Explanation and Interpretation of Results: [from a hypothetical literature review]

Children with a secure attachment style benefited from deeper sleep and fewer sleep disturbances than children who had anxious or avoidant attachment styles, indicated by the difference in mean delta scores.

[further explanation would be added depending on additional evidence contained in your research and from your literature review...]

Children with a secure attachment style do not need to monitor their environment as attentively as anxious and avoidant children in order to be aware of their primary caregiver's presence. This attachment style is related to a deeper sleep pattern and better quality rest. In order to promote deeper sleep in children, intervention strategies to increase secure attachment are recommended.