

Week 8
March 13, 2009

- Review of exam
- Introduction to ANOVA
- Lab 2a, 2b (continue next week if necessary)

Concepts you should know:

- When to use ANOVA
- One-way ANOVA
- Omnibus test
- *F* statistic
- Within-groups variation
- Between-groups variation
- *Post-hoc* tests
- Multiple comparisons

I. Review—"Name that Statistical Test"

A. Is there an association between receiving CBT and feeling less anxious ("Yes" vs "No")?

IV: Treatment Level of Measurement: Categorical (e.g. receiving vs. not receiving CBT)

DV: Anxiety Level of Measurement: Categorical

Statistical Test: Chi-squared test of association

B. Does CBT improve anxiety symptoms (based on a ratio level anxiety scale) more so than those receiving psychodynamic psychotherapy?

IV: Treatment group
(i.e. CBT vs. psychodynamic psychotherapy) Level of Measurement: Categorical (nominal)

DV: Anxiety Level of Measurement: Continuous

Statistical Test: Independent groups *t*-test

If ...	Independent variable is:	And dependent variable is:	Statistical test is:
	Categorical (nominal or ordinal)	Categorical (nominal or ordinal)	<u>Chi square</u>
	Categorical (nominal—dichotomous, two independent groups)	Ratio or interval	<u>Independent groups t-test</u>
	Categorical (two dependent groups)	Ratio or interval	<u>Dependent groups (or paired samples) t-test</u>
New!	Categorical (nominal or ordinal) more than two attributes	Ratio or interval	One-way ANOVA

II. When to use ANOVA?

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

D. ANOVA is a distant cousin to the independent groups *t*-test. ANOVA is a *parametric test*—requires data that are normally distributed in the population. 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 test 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) individual psychotherapy only. Do the clients in the “medication plus psychotherapy” group show more improved outcomes (a larger rate of change in depression scores) than the other two groups?

***Identify the independent and dependent variables in each of these scenarios.

E. Types of ANOVA

1. One-way ANOVA only has one nominal independent variable with three or more groups (**this is the one we’ll focus on in this class, as in the examples above**)
2. Two-way ANOVA has two nominal independent variables (and these can have two or

more attributes)

3. MANOVA—multivariate ANOVA—has two *dependent* variables.

III. 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 not 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 of two groups, 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]

Within groups vs. between groups

Child_ID	Self esteem level	Number behavioral outbursts
1	Low	7
2	Low	4
3	Low	5
4	Medium	6
5	Medium	6
6	Medium	2
7	High	2
8	High	4
9	High	1

Diagram illustrating the grouping of data for ANOVA analysis. The data is grouped into three self-esteem levels: Low (Child IDs 1, 2, 3), Medium (Child IDs 4, 5, 6), and High (Child IDs 7, 8, 9). Within each self-esteem level, the data is further grouped into 'Within Groups' (indicated by red text and brackets) and 'Between Groups' (indicated by red text and a large bracket). The 'Between Groups' label is positioned to the right of the 'Within Groups' labels, indicating the comparison between the three self-esteem levels.

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. If the F ratio is not significant, *game over!* There is no evidence to reject the Null, hence there is no evidence that *any* of the means are different.
2. But, if it is significant and 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 all (three or more) means.
3. This involves another level of statistical tests called *post-hoc tests*, which determine where 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.

- 4. If we had four groups, then there would be six comparisons for four groups:

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

- 5. There are a few different types of *post-hoc* tests (such as the Bonferroni (pronounced Bohn-fur-own'-ee), the Scheffe (pronounced Shef-fay') and the Tukey (pronounced Too'-key) procedures. We will use one of them in the example and the labs—the Tukey procedure.
- 6. 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 not 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 (i.e. the probability of *avoiding* a Type I error decreases with each additional comparison).
- 7. The *post-hoc* procedure adjusts for the inflated risk of making a Type I error, so that the combined probability of falsely rejecting the Null (Type I error) among all the comparisons is no more than your intended alpha (such as .05).
- 8. 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).

IV. Example of One-Way ANOVA

Consider this dataset of sleep patterns 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 (or deep 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 sleep” 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

Source of dataset: Kirkpatrick, L. A. & Feeney, B. C. (2007). A simple guide to SPSS for Windows for version 14.0. Belmont, CA: Thomson/Wadsworth.

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.

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

Eight Steps for Hypothesis Testing:

- What's the independent variable and level of measurement? Attachment style; categorical
- What's the dependent variable and level of measurement? Percent delta sleep; continuous
- State the Null Hypothesis (H_0):
The three attachment style groups do not differ in percentage of delta sleep.
- State the Alternative Hypothesis (H_A) and identify whether one-tailed or two-tailed:
a) The three groups differ in percentage of delta sleep (two-tailed).
b) The "secure" attachment group will have a higher percentage of delta sleep (one-tailed).
- Select appropriate statistical test and alpha level:
One-way ANOVA (or, One-way Independent groups ANOVA, or just ANOVA). $\alpha \leq .05$

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?**

*****How would you interpret the *F* ratio here, i.e. what's the relationship between the variation between the groups and variation within the groups?**

Post Hoc Tests

Multiple Comparisons

Percentage of delta (deep) sleep

Tukey HSD

(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval		
					Lower Bound	Upper Bound	
Attachment style type	Attachment style type						
	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.)

9. Results:

The F value of 41.425 is associated with a p -value of less than .001 (remember—when SPSS says $p = .000$, SPSS stops at three decimal places), which is less than the alpha level of .05. Therefore the difference among the three groups' delta mean scores is significant and we reject the null hypothesis. From the Tukey *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.

10. 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. Future research should focus on whether this relationship is also true across ethnicity groups. Also, since children in foster care are thought to experience post-traumatic stress disorder and difficulties attaching to the new environment, research should include comparisons between foster and non-foster children.