ANOVA – Midterm Review
24 November 2007

1. **Null hypothesis tested by one-way ANOVA:** Two or more population means are equal

    * Asking: are observed differences between sample means due to actual differences in the population means
        - H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$
        - H_A: H_0 is false

2. **Assumptions needed for ANOVA:**

    * random, independent sampling from each population
    * normal population distributions
    * equal variances within each population

3. **ANOVA compares** two estimates of the population variance

    * one is the pooled variance of scores within groups
    * the other is based on the observed variance between group means

MS_{WG} = Pooled Estimate of $\sigma_y^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)} = \dfrac{\sum (y_{ij} - \bar{y}_j)^2}{\sum (n_j - 1)}$

* From Introduction to One-Way ANOVA handout: Formula 2 = Formula 3 (pp. 22 & 23)

4. **Variance of the sampling distribution of the mean** (variance of all possible sample means of independent samples of size n drawn from some population) = $\sigma_{\bar{y}}^2 = \dfrac{\sigma_y^2}{n}$

    * Are trying to estimate the population variance based on the observed variance of sample means = MS_{BG} = Between Group Estimate of $\sigma_y^2 = \sum \dfrac{n_j(\bar{y}_j - \bar{y}_{..})^2}{(k-1)}$

5. **F ratio tests:** how likely it is our two estimates of the same population variance would differ so widely if our assumptions are valid

6. Source Table According to Howell (p. 305):

	SS	df	MS	F	p
Between Groups	$\sum n_j\left(\overline{Y_j} - \overline{Y..}\right)^2$	k – 1	SS_{BG}/df_{BG}	MS_{BG}/MS_{WG}	
Within Groups	$\sum_{ij}\left(Y_{ij} - \overline{Y_j}\right)^2$	k(n – 1)	SS_{WG}/df_{WG}		
Total	$\sum_{ij}\left(Y_{ij} - \overline{Y..}\right)^2$	N – 1			

7. Transformation of Data:

* transformations are used when the assumptions for a statistical test are not met (e.g., homogeneity of variance)
* first plot a histogram to see if the distribution suffers from skew
* SPSS will compute skew and kurtosis (values > 1 should be investigated; outliers cause positive kurtosis, negative kurtosis is generally not a problem)
* The three most common transformations to reduce positive skew (from least intensive to most intensive): square root, logarithm, & reciprocal
* negative skew can be reduced in the three ways listed above but the data must first be reflected
    - to reverse a variable, Yr = M + 1 – Y
        - ~ M = maximum score on the original Y variable
        - ~ Yr = the reversed variable
    - for example, a 1 to 7 Likert scale would be reversed with Yr = 8 – Y
    - all values must be greater than zero when you apply transformations
    - after reversal & transformation, you may wish to reverse again
    - note that reciprocal transformation reverses the scale
* the transformation brings in the positive tail of the distribution and spreads out the lower end of the distribution
* start with the least intensive transformation & progress to the most intensive as needed
* transformed variables must be interpreted as such when drawing conclusions (e.g., "salary" becomes "the log of salary" in write-ups)
* ANOVA is quite robust with respect to violations of normality, and homogeneity of variance, especially if the samples are large and the group sample sizes are equal

8. A priori vs. post-hoc (a posteriori) tests: controlling family-wise error – the more comparisons we make between two means, the more we inflate alpha error. There are several tests that can be used to control this phenomenon.

9. A priori contrasts:

* Chosen before the data are collected
* Based on theoretical predictions

10. Contrasts or comparisons:

* Orthogonal contrasts
  - Completely partition SS_{BG}
  - Occur when members of a set of contrasts are independent of one another
* Each contrast has df = 1
* $SS_{contrast} = MS_{contrast} = L^2 \Big/ \sum\left(\frac{a_j^2}{n_j}\right)$
* Where L is the numerical value of the contrast – found by multiplying each coefficient in the linear contrast with the mean for its respective group and summing these
* $F = MS_{contrast}/MS_{WG}$

11. Bonferroni Test:

* Any set of a priori comparisons
* Recalculates alpha to control for inflated error over multiple comparisons
* Alpha is split by the # of tests
* Critical value = $\alpha' = \alpha/c$, where c = # of comparisons you are going to do

12. Dunn-Sidak Test:

* Variation on the Bonferroni procedure
* Minor differences between Bonferroni and Dunn-Sidak, but Dunn-Sidak yields more power
* Critical value = $\alpha' = 1 - (1 - \alpha)^{1/c}$

13. Holm:

* Adjusts the denominator of the Bonferroni test
* More liberal and more powerful than Bonferroni
* Order tests in terms of "p" values – largest (rank 1) to smallest (rank c)
  - c = the number of tests (selected a priori)
* Successively "loosens" the requirement (critical value) you compare against as you go down the ordered values
* Critical value = $\alpha' = \alpha/\text{rank}$

14. A posteriori

* Planned after the data are collected
* Usually some examination of data – experimenter notes which means are farther apart than others and bases further analysis on these observations of the data

15. Fisher's LSD (Least Significant Difference) Method:

* Requires a significant F from the overall ANOVA
* Used for testing pairwise comparisons
* Not recommended by Howell when comparing more than 3 means
* Also, can use "q" table with r = 2
* $t = \dfrac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{MS_{WG}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$
* LSD = $\overline{Y}_1 - \overline{Y}_2 = t_{(df,\alpha)}\left(\sqrt{MS_{WG}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}\right)$

16. Tukey's HSD (Honestly Significant Difference) Method:

* Controls family-wise error
* Used for testing all possible pairwise comparisons
* HSD = $q_{(r,df,\alpha)}\left(\sqrt{\dfrac{MS_{WG}}{2}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}\right)$
* Where:
  - r = number of groups
  - df = df_{WG}
  - q = from table (Howell, p. 679)
  - α = .05 or .01

17. Studentized Range Statistic (q):

* Compromise between LSD and HSD tests
* Rank means from smallest to largest
* Consider the number of "steps" between means to be compared

18. Ryan Procedure (REGWQ):

* Use for testing pairwise comparisons
* More powerful than Tukey

19. Scheffe Test:

* Most conservative post hoc test
* Tests complex contrasts: all possible sets of contrasts
* Do not use for pairwise tests
* Do not use for a priori tests
* Do not use the Scheffe test in SPSS for one-way ANOVA
* Instead, use t^2 ($t_{observed}$ from your contrast) to calculate F
* Compare calculated F to Scheffe F_s
* $F_s = (k-1)F_{k-1, N-k, \alpha}$
* Where:
  - N = total # of subjects
  - k = the # of groups

20. Dunnett:

* Use when comparing one group mean (control group) against each other group mean (one at a time)
* Critical value = $\overline{Y_j} - \overline{Y}_{control} = t_{d(df, \alpha, k)}\left(\sqrt{\frac{2MS_{WG}}{n_j}}\right)$
* Where:
  - k = number of groups
  - df = df_{WG}
  - t_d = from table
  - $\alpha = .05$

Trend Analysis - When to use:

* when you are interested in looking at whether there is some polynomial trend in the data **and** you have:
  - several levels of treatment
  - the levels differ on a quantifiable scale
  - the intervals between levels are specified and (ideally) equal, or close to being equal
* may be helpful to look at the plot and to consider the relative size of the different trend components
* remember to look at what the variables are, and whether it makes sense for there to be a linear trend (i.e., it wouldn't make sense to look for a polynomial trend if the treatment levels are nominal and don't lend themselves to some meaningful ordering)
* also, is there theoretical justification for a polynomial trend?

Simple to complex trends:

* in general, we seek the simplest model that is appropriate for our data
* from simple to complex: linear, quadratic, cubic, etc.
* the # of mutually orthogonal trend contrasts that can be performed equals k – 1, where k = the # of groups or different treatment levels

Interpreting SPSS Output: (Taken from HW#2):

ANOVA

Score Reading Test

			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)		1213.200	3	404.400	8.153	.002
	Linear Term	Contrast	547.560	1	547.560	11.040	.004
		Deviation	665.640	2	332.820	6.710	.008
	Quadratic Term	Contrast	369.800	1	369.800	7.456	.015
		Deviation	295.840	1	295.840	5.965	.027
	Cubic Term	Contrast	295.840	1	295.840	5.965	.027
Within Groups			793.600	16	49.600		
Total			2006.800	19			

Above we see that there is a significant linear trend ($F_{1,16} = 11.04$, $p < 0.01$). Interestingly, we also see that the data significantly deviates from being linear as well ($F_{2,16} = 6.71$, $p < 0.01$). Though this deviation is significant, we find that neither the quadratic, nor the cubic terms are significant at alpha = .01. But we see that they were close to being significant (i.e., p = .015 and .027 for quadratic and cubic, respectively), which may explain the significant deviation from linearity.

Although there is a significant linear trend, this may not be the best model to use in describing the data. If you have strong theoretical justification for a linear model, then this might be appropriate. If not, a more appropriate model might be just that groups C and D combined are significantly higher on reading score than groups A and B combined (see multiple pairwise comparisons, e.g., REGW-Q). **The point here is that your interpretation should be grounded to some extent in theory and what makes sense.**

ANOVA – Notes: Homework #2
24 November 2007

Write-Up
* read homework directions thoroughly – for example, this homework (#2) states "attach and interpret your output, and present a separate summary…"
* include appropriate stats in your summarized results
* it's helpful to include tables from SPSS in your summarized results, but also include output
* include what test you chose & why, findings/results (with stats) – interpret for your audience using "plain" language
* description of results should also include the size of effects & the direction of effects
* spell-check & proofread

SPSS
* attach separate SPSS output with annotation
* include syntax
* formulas when appropriate (don't need on repeated tables)

Other
* Hypothesis testing – when you reject the null hypothesis, it does NOT mean that you have proven the alternate hypothesis is true
* When starting statistical analysis, best to get "close" to the data (descriptives, graphs) and to check for violation of any assumptions that apply to the tests you plan to run