Stat 242 L. Katz Midterm xm Fall 2006

1. Brief answer (no formulas needed):
(a) What are the differences between experimental vs. nonexperimental designs? How do interpretations of their results differ?
(b) What is a histogram? What is skewness?
(c) What are standard scores and why are they useful?
(d) What is the standard error of the mean and why is it useful?
(e) What are confidence limits and why are they useful?
(f) Why are data transforms used? What are the ways to interpret anova results when a data transform is used?

2. Concerning the homogeneity of variance assumption in anova:
(a) Describe the assumption (b) Why does the $MS_{S/A}$ depend on the assumption? (c) What criteria (including rule-of-thumb and statistical tests) are used to evaluate the adequacy of the assumption?

3. (a) List and briefly describe the remaining 3 assumptions of anova (besides homogeneity of variance). (b) How sensitive is each assumption to distortion? (c) What criteria are used to evaluate the adequacy of each assumption (including informal criteria and formal statistical tests)?

4. (a) Describe, in words (with or without equations), what SS_{total} and its components, SS_A and $SS_{S/A}$ represent? That is, what different aspects of variability in the data set are encapsulated (i.e., measured) by these three terms? (b) When we divide the latter two terms by their df, we get their respective mean squares. Under what conditions does the ratio of these mean squares produce an F-ratio? (c) What does the alpha level (probability = .05, by convention) mean? (d) What is meant by the *power* of the F-test and what affects power?

5. List 3 measures of effect size that are appropriate for an anova design with more than 2 levels, pointing out what aspect of the results they measure (i.e., what they tell us about the experiment) and how the measures differ from each other. Formulas are not needed but are provided for your assistance on the accompanying formula list.

6. Contrast the logic of null hypothesis testing with the logic of the "probability of replication" (p_{rep}). Comment on importance of the differences between them for the practice of statistical design and analysis.

Formulas: $\sum_j^a \sum_i^n (Y_{ij} - \bar{Y}_{..})^2 \quad n\sum_j^a (\bar{Y}_{.j} - \bar{Y}_{..})^2 \quad \sum_j^a \sum_i^n (Y_{ij} - \bar{Y}_{.j})^2 \quad z = \frac{Y - \bar{Y}}{S}$

$\hat{\sigma}_A^2 = [(a-1)/a][(MS_A - MS_{S/A})/n]$

$\hat{\sigma}_e^2 = MS_{S/A}$

$\hat{f} = \sqrt{\hat{\sigma}_A^2/\hat{\sigma}_e^2}$

$\hat{\omega}^2 = \hat{\sigma}_A^2/(\hat{\sigma}_A^2 + \hat{\sigma}_e^2)$
$= [SS_A - (a-1)MS_{S/A}]/(SS_{tot} + MS_{S/A})$
$= [(a-1)(F-1)]/[(a-1)(F-1) + na]$

$\hat{\lambda} = na\hat{\sigma}_A^2/\hat{\sigma}_e^2$
$= [(a-1)(MS_A - MS_{S/A})]/MS_{S/A}$
$= (a-1)(F-1)$

$\hat{\phi} = \sqrt{n\hat{\sigma}_A^2/\hat{\sigma}_e^2}$
$= \hat{f}\sqrt{n} = \sqrt{\lambda/a}$

$t_{N-1} = \frac{\bar{Y} - \mu}{S/\sqrt{N}}$

$E_S = \frac{(\mu_{A_1} - \mu_{A_2})}{\sigma} = \frac{\bar{Y}_1 - \bar{Y}_2}{S_{pooled}}$

$CI = \bar{Y} \pm 1.96 S_{\bar{Y}} \quad S_{\bar{Y}} = S/\sqrt{n}$

Stat 242 sec 02 Midterm exam Fall, 2004 L. Katz

There are four questions. All of them refer to the data set on the accompanying handout. The data are from an imaginary three-group experiment in which each subject was randomly and independently assigned to one of the three levels of the treatment A. Each of the 15 questions below is worth 7 points (total = 105).

#1. Data distribution assumptions:
#1a. Using the accompanying descriptive statistics and any information you can use from the raw data, determine if the data are suitable to be analyzed by ANOVA by first sketching, by hand, an accurate frequency histogram for each group. #1b. Do the data show more than acceptable amounts of skewness? #1c. What can you say about the normality assumption? #1d. Is the assumption of homogeneity of variance satisfied? Why?

#2. ANOVA:
#2a. Finish the ANOVA table, using the information that has been provided. #2b. Determine the level of significance, using the F-table on page 665 of the textbook. Write the entire completed ANOVA table in your test booklet. #2c. Write the Expected Mean Squares for A and S/A. Under what condition does the ratio of the EMS qualify as an F-statistic? Demonstrate this using the EMS components.

#3. Contrasts:
#3a. Test all simple contrasts for the data set using the Fisher-Hayter method. Show your calculations (neatly, please).
#3b. Compare the results (in terms of the levels of significance you obtain) to the results of the Tukey, Scheffe, and Bonferroni methods for simple contrasts that are given on the accompanying printout.
#3c. Contrast the average of groups A1 and A2 to A3. Interpret the contrast using the Scheffe method to determine significance.
#3d. Produce a set of *orthogonal* contrasts of these data, beginning with the complex contrast of A1 and A3 compared to A2. Write down all of the orthogonal contrasts in the set. *No* significance test is required (but see the next question); just specify the contrasts. Also, you do *not* need to demonstrate that the contrasts are, in fact, orthogonal (although you may wish to do so for *your own* satisfaction).
#3e. Test only the first contrast (A1 + A3 vs. A2) for significance, using Scheffe's method.

#4. Other statistics:
#4a. The confidence intervals for the treatment means are given on the accompanying handout. How should the three 95% confidence intervals be interpreted? That is, what do they tell us?
#4b. Calculate Cohen's d and interpret this effect size both in practical terms (is it small, medium, or large) and in terms of the theory (i.e., in terms of the relevant population variances).
#4c. Define statistical power. This experiment obviously had adequate power but if it had not, what are the three kinds of changes you could have made to improve power?

Dataset for ANOVA (question #1)

Subject	A	Y
1	1	2
2	1	3
3	1	4
4	1	3
5	1	4
6	1	2
7	1	3
8	1	3
9	2	3
10	2	5
11	2	4
12	2	4
13	2	3
14	2	5
15	2	4
16	2	4
17	3	4
18	3	6
19	3	5
20	3	5
21	3	4
22	3	6
23	3	5
24	3	5

Cohen's $\hat{f} = \sqrt{\frac{(a-1)F_A - 1}{an}}$

f { .10 sm, .25 med, .40 lrg }

$f = \sqrt{\sigma_A^2 / \sigma_e^2}$

$\omega^2 = \frac{\sigma_A^2}{\sigma_e^2 + \sigma_A^2}$

CI:

$P(\bar{Y} - 1.96\,\sigma_{\bar{Y}} \le \mu \le \bar{Y} + 1.96\,\sigma_{\bar{Y}}) = .95$

Formulas

Tukey HSD

$d_{criterion} = q_{\alpha, a, a(n-1)}$

Fisher-Hayter

$d_{criterion} = q_{\alpha, a-1, a(n-1)} \sqrt{\frac{MS_{S/A}}{n}}$

Scheffé

$t = \frac{\hat{\psi}}{S_{\hat{\psi}}} = \frac{\sum_j w_j \bar{Y}_j}{\sqrt{MS_{S/A}\left(\frac{\sum w_j^2}{n_j}\right)}}$

compare t to $S_{criterion}$: $S_{crit} = \sqrt{(a-1)F_{\alpha, a-1, a(n-1)}}$

Bonferroni

$t = \frac{\hat{\psi}}{S_{\hat{\psi}}} = \frac{\sum_j w_j \bar{Y}_j}{\sqrt{MS_{S/A}\left(\frac{\sum w_j^2}{n_j}\right)}}$

t is significant if it is larger than $t_{\alpha_{FWE}}$

where $\alpha_{FWE} = \frac{\alpha_{EC}}{a}$ or $\frac{\alpha_{EC}}{K}$

K = number of contrasts

choice between $\frac{.05}{a}$ and $\frac{.05}{a(a-1)/2}$ depends on how conservative you wish to be.

Stat 242 Mid-term exam Fall, 2003

#1. *Very short answers.*

a. Distinguish between the omnibus null hypothesis and single-df contrasts.
b. Distinguish between FWE and EC.
c. Distinguish between planned and post hoc contrasts.
d. Distinguish between fixed-effect and random-effect independent variables.
e. Distinguish between central and non-central F distributions.

#2. Below are descriptive statistics and frequency distributions (next page) for a three-group experiment (A1, A2, A3), which has 40 subjects in each of the 3 levels of A. (The data are measurements of the acidity of 40 samples of each of three different hair shampoos). (a) Based on this information, write a brief assessment of the assumptions of normality and homogeneity of variance for these data. (b) Do you think that additional tests are needed? Is a data transform necessary? If so, which tests or transforms?

	A1	A2	A3
Mean	4.918	5.010	5.146
SD	.237	.242	.205
Skewness	-.191	.538	-.486
Kurtosis	.271	-.300	.569

#3. (a) Whether or not you decided in the previous question that it is appropriate to do an ANOVA, do one, nevertheless. The question is: Do the shampoos differ significantly with regard to acidity? Here is some additional information; fill in the blanks. (b) Make a contrast between A1 and A2 and assess its statistical significance. (Use the significance tables in the textbook).

SV	DF	SS	MS	F	p <
A		1.053			
S/A		6.136			

#4. (a) Define *power*. (b) The effect size is one of the three factors that influences the likelihood of your ANOVA being significant at *alpha* = .05. What are the others and what is their effect on significance? (c) Why is it often necessary to calculate the *a priori* power of an experimental design?

#5. (a) Given the formula (attached) for SS_A calculate it for the data below (10 Ss per group):

	A1	A2	A3	A4
Mean A_j :	4.0	4.0	2.0	2.0

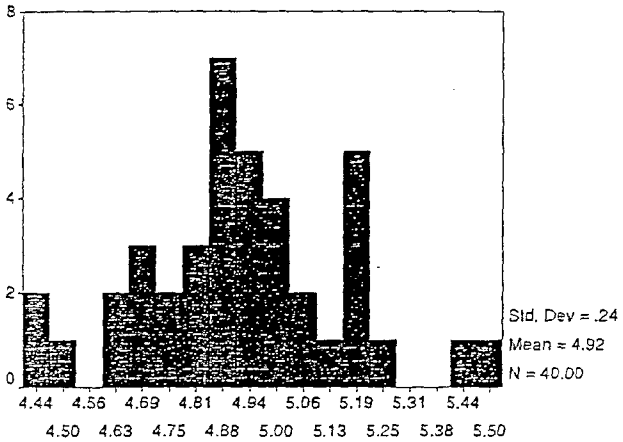
(b) List several ways in which the (central) t- and F-statistics are (1) similar and (2) different.

#6. Brief definitions:

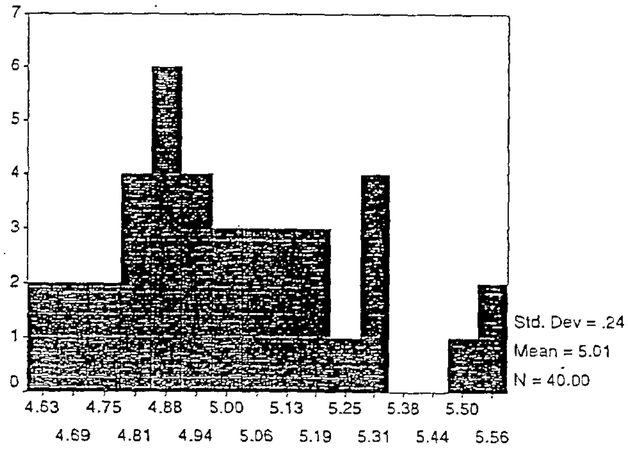
(a) Expected Mean Square
(b) Effect Size
(c) Confidence Interval
(d) Central Limit Theorem
(e) Interquartile Range

#2.

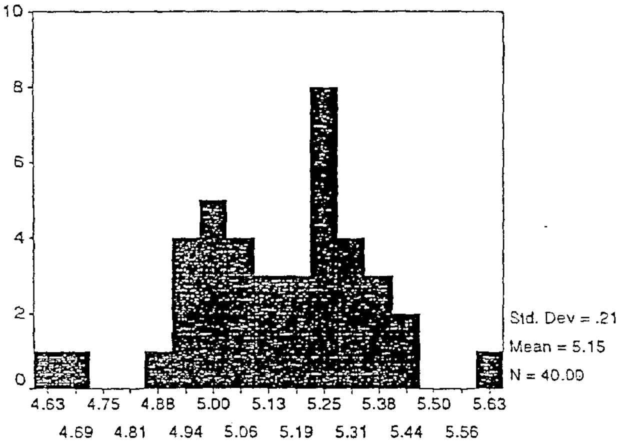
A1



A 2.



A 3.



#5 $$\sum_j^a \sum_i^n (Y_{ij}-\bar{Y}_{..})^2 = \sum_j^a \sum_i^n (Y_{ij}-\bar{Y}_{.j})^2 + n\sum_j^a (\bar{Y}_{.j}-\bar{Y}_{..})^2$$

#3 $$\hat{\psi} = \sum_j w_j \bar{Y}_{.j} \qquad t = \hat{\psi} / S_{\hat{\psi}}$$

$$S_{\hat{\psi}} = S_{pooled}\sqrt{\sum_j \frac{w_j^2}{n_j}}$$

$$S_{pooled} = \sqrt{MS_{S/A}}$$

$$S_{N-1} = \sqrt{\frac{\Sigma (Y-\bar{Y})^2}{N-1}}$$

$$S_{\bar{Y}} = \frac{S_{N-1}}{\sqrt{N}}$$

$$SS_{TOTAL} = SS_{bg} + SS_{wg}$$

$$MS = \frac{SS}{df}$$

$$df_{TOTAL} = N - 1$$

$$F_{max} = \frac{S^2_{largest}}{S^2_{smallest}}$$

$$d = \frac{|\bar{Y}_\ell - \bar{Y}_s|}{\sqrt{MS_{S/A}}}$$

$$\eta^2 = \frac{SS_A}{SS_T}$$

$$F = \frac{n(\Sigma w_j \bar{Y}_j)^2 / \Sigma w_j^2}{MS_{S/A}}$$

$$\sum_{j=1}^{a} w_j = 0$$

$$\alpha_{FW} \leq 1 - (1 - \alpha_{PC})^C \approx C \cdot \alpha_{PC}$$

$$F_S = (a-1) F_{TAB}$$

$$F_T = \frac{q_T^2}{2}$$

$$\bar{d}_T = q_T \sqrt{\frac{MS_{S/A}}{n}}$$

(1) List, define, and give one example of each of the following IV scale types: nominal, ordinal, interval, ratio. For each one, identify and describe briefly a different continuous DV, so that each combination of IV and DV makes a plausible "laboratory" experiment (i.e., an experimental design). How does ANOVA distinguish among these IVs of different scale type?

(2) Consider this experiment. The DV is the number of items remembered from a list of names. The list consists of 100 male first names, which are presented one at a time on a computer screen. After the presentation, each subject writes down as many names as he/she can remember. The subject will receive one reward unit for each name correctly recalled and will lose one unit for each incorrect recall. There are two crossed IVs. One IV is Reward_Amount (the levels, i.e., the units of reward, are $1, $2, and $4). The second IV, crossed with the first, is the Gender of the subject; half are men and half are women. Thirty-one men and 31 women are randomly assigned to each of the 3 levels of Reward_Amount.

Assume that the data look like this: as reward amount increases, people have higher recall scores but the variance of recall scores increases also.

a. First plot the means of the 6 cells on one graph. With regard to Gender and its interaction with Reward_ Amount, you may make the means come out however you like as long as the effect of Reward_ Amount is preserved. Label all parts of the graph.
b. Create histograms of the raw data for the 31 subjects in each of the 6 factorial cells. List the steps you need to take to screen the data before running the ANOVA.
c. Do you recommend any data transforms for your data? If the answer is Yes, state what transform(s) are needed. If the answer is No, state why none are needed.
d. For the same data set, list the sources of variance (SV), the actual numerical df for each SV, and which ratio of mean squares is required for each F-test (i.e., F= MS?/MS?).
e. List two ways to increase the power of this experiment.

(3) Post-hoc tests.
In an experiment in which A, Teaching_Method (A1, A2), is crossed with B, School_Grade (B1, B2), the DV is Reading_Test_Score. Sixteen schoolchildren were tested in each cell. An ANOVA showed that there were A and B main effects (A1<A2 and B1<B2). More importantly, the interaction was significant: $F_{AB}(1, 60) = 10.00$, $MS_{S/AB} = 10.00$, $p < .01$. Here are the means for the four cells:

	A1	A2
B1	2	2
B2	3	8

a. Perform two simple effects post-hoc comparisons for $A_{at\ B1}$ and $A_{at\ B2}$. Test their significance. Use this standard formula for comparisons:

b. Perform a Tukey test to test the significance of the difference between A1B2 and A2B1.

1. (50%) Examine the descriptive statistics and analysis for the data that are given in the handout. The design consists of the independent variable A which has four levels.
   a. On page 1, find the information about the 95% confidence limits. What do they tell us? How are they calculated?
   b. Comment on the frequency distribution of each group.
   c. In general, what transforms of a dv are appropriate when group means and variances are proportional? Do these data warrant such a transform?
   d. What is the homogeneity of variance assumption in ANOVA? Applying Tabachnik & Fidell's criterion, are there sufficient grounds to accept the assumption here?
   e. Find the ANOVA table on page 3. You are given the sums of squares. In your exam book, write the complete ANOVA table, including df, MS, F, and the strongest level of significance that is appropriate (given the table in Tabachnik).
   f. Why is it important to determine a variable's effect size? Produce one measure of effect size for the data given and interpret it.

2. (20%) Given the post hoc tests on page 4 of the handout:
   a. Both Tukey and Sheffe tests are provided for the simple post hoc comparisons we wish to make. Which is the more powerful for this purpose? Demonstrate your answer by comparing the results of the Tukey and Scheffe tests.
   b. What problem is avoided by the use of Tukey and Scheffe (and other) post hoc tests?

3. (20%) Instead of the standard procedure of an omnibus anova plus post hoc tests, you could have performed planned comparisons. Calculate a set of orthogonal comparisons on the data in the handout. After testing the comparisons for significance, show the relationship between the comparison F-ratios and the omnibus F-ratio.

4. (10%) Short answer questions:
   a. Define the power of a statistical test. What design factors affect power (and how do they affect it)?
   b. What is the advantage of equal sample size in all levels of the IV?
   c. In scaling theory, what are empirical and numerical relational systems?
   d. What is the purpose of random assignment of subjects to levels of the IV?
   e. What is the fundamental difference between a so-called experimental design (i.e., a "laboratory" design) and a so-called non-experimental design (i.e., a "field" study).

Stat 242 Fall 00 Midterm exam

1. Perform an analysis of variance on the following data. Children in Grades 3,4,5, and 6 (i.e., A1 – A4) were tested for the amount of time (in seconds) that they needed to read a standard paragraph of text. Is the homogeneity of variance assumption acceptable? Should the null hypothesis be rejected? Present the complete ANOVA table and enough arithmetic work so that you can receive credit even if you make a mistake in your calculations.

	A1	A2	A3	A4
Mean	212.50	193.50	173.25	154.13
s^2	841.71	2022.86	987.93	690.41

n= 8 $SS_A = 15270.844$ $SS_{S/A} = 31800.375$

2. a. Calculate two different measures of the effect size of Grade using (1) omega-squared and (2) the percentage that the A sum of squares contributes to the total data set variability. Briefly evaluate and compare the two measures.

$$\hat{\omega}^2 = \frac{SS_A - (a-1)(MS_{S/A})}{SS_T + MS_{S/A}}$$

b. Graph the means in the data above and describe the apparent relation between Grade and Reading Speed (without a statistical test).

c. Now test the relation with linear and quadratic statistical tests of trend. What additional tests for trend are (theoretically) possible?

3. Are post-hoc tests for differences in the means appropriate? Why? In either case, perform post-hoc tests for a significant difference between A1 and A4 using a. Scheffe's procedure, b. Fisher's procedure, and c. Tukey's procedure. Give the outcome of each test (i.e., is A1 significantly different from A4 and, if so, at what level of significance?). Compare the results of the 3 tests, which is more conservative, which is more sensitive?.

4. For each of the following factorial experiments, describe (without doing the actual f-tests) whether or not the individual main effects and interactions are apparently significant.