Midterm Exam
Psychology 407
Spring 2008

Name: ______________________________

Whenever possible, use fractions.

Questions I through III are all based on the following storyline and data:

A certain university was engaged in teaching Peace Corps volunteers the foreign languages they would need during their tours of duty. An experiment (very small scale) was carried out to determine the effectiveness of the different teaching methods that were used. All students were first given a language-aptitude test, which provided the measures that we will label by the letter X.

The experiment carried out was to evaluate two methods of teaching the foreign languages and to determine the value of language laboratory sessions. The two teaching methods were: (1) formal classroom meetings with lectures, and (2) no formal classroom meetings — only conversation periods held in a congenial atmosphere. In addition, half of the students being taught by each teaching method spent three hours a day in the language laboratory wing the tape-recording equipment. The other half of the students in each method group never entered the language lab. subjects were randomly assigned to the four conditions, and the total instructional time was the same for all individuals. Each volunteer was independently rated (by an expert) on a ten-point scale (1 to 10) for their language ability after the completion of the course. These scores are represented by the letter Y.

The data that resulted from this experiment are as follows:

	classroom		conversation	
lab	X	Y	X	Y
	62	5	46	5
	75	7	53	4
	41	3	57	3
	88	8	49	7
	72	7	62	6
no lab	84	2	58	9
	91	3	72	10
	68	1	61	8
	77	1	65	8
	85	3	59	10

To aid in the computations that you might find yourself in need of completing, the following table given the various sums and (raw) sums of squares for the Y variable — within each of the four cells, within each of the rows and columns, and overall. Sums are given first; sums of squares second.

30; 196	25; 135	55; 331
10; 24	45; 409	55; 433
40; 220	70; 544	110; 764

Suppose the data are treated in the form of a one-way analysis-of-variance layout with 4 groups: group 1 – lab/classroom; group 2 – lab/conversation; group 3 – no lab/classroom; group 4 – no lab/conversation.

a) Complete the analysis-of-variance source table – label all terms appropriately. (Form the test statistic, and indicate what degrees-of-freedom you would use in assessing significance.)

b) Suppose I represent the vector of observations $\mathbf{Y}$ in the form $\mathbf{X}\ \boldsymbol{\beta} + \boldsymbol{\epsilon}$, where the 20 observations in $\mathbf{Y}$ are ordered in the usual way – i.e., those in group 1, 2, 3, and then 4. Define $\boldsymbol{\beta}\prime = [\mu.\ \tau_1\ \ldots\ \tau_{r-1}]$, where $\mu.$ is based on the weighted average. What is the design matrix $\mathbf{X}$?

c) Considering the model in (b) to be the full model, what is the design matrix $\mathbf{X}$ and what is $\boldsymbol{\beta}$ for specifying a restricted model under $\mathrm{H}_o : \tau_1 = \cdots = \tau_r = 0$?

d) What is the sum of squares for the reduced model in (c)? For the full-model in (b)?

e) Show how the F-ratio in (a) is formed from the information calculated in (d)?

II. Still treating the data in the form of a one-way analysis-of-variance layout, the following three pages give the output from several SYSTAT analyses. The questions below should be answered in relation to those analyses.

a) If I were concerned with the regression of Y on X for each of the four groups separately, give the slopes and intercepts (for raw scores):

	intercept	slope
group 1:		
group 2:		
group 3:		
group 4:		

b) Indicate how we would test the hypothesis that the four within-group regression slopes are equal. Give the F-ratio and the degrees of freedom that you would use.

c) Assuming that the assumption of within-group regression slopes is reasonable, carry out a test of the treatment effects. Give the F-ratio and the degrees of freedom that you would use.

d) If all the 20 observations on Y and X were considered and the group structure ignored, what is the correlation of Y and X? Test it for significance and give the *exact* two-tailed p-value.

d) Provide the “adjusted” means on Y for each of the four groups (assuming a common within-group regression slope).

III. Suppose the data are treated in the form of a two-way analysis-of-variance layout with the two crossed factors of lab/no lab and of classroom/conversation.

a) Complete the analysis-of-variance source table — label all terms appropriately. (Form the test statistics, and indicate what degrees-of-freedom you would use in assessing significance.)

b) Suppose I represent the vector of observations $\mathbf{Y}$ in the form $\mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where the 20 observations in $\mathbf{Y}$ are ordered in the same way as in I(b). Define an appropriate vector $\boldsymbol{\beta}$ containing the main effect and interaction parameters, and give the design matrix $\mathbf{X}$ that would result.

c) Estimate the specific effects of the lab/no lab factor *within* the two levels of the classroom/converstion factor.

IV. Completion:

(a) In single factor analysis-of-variance, a(n) ____________________ factor is one where each level is described by a numerical value on a scale.

(b) In single factor analysis-of-variance, if the treatments are considered randomly selected from a larger population of treatments, one uses a particular model for the analysis. This model goes by a variety of names, three of which are ____________________, ____________________, and ____________________.

(c) Given the usual assumptions on the error terms in Model I analysis-of-variance, the observations within the i^{th} group come from what type of distribution (be explicit and include parameter information to characterize the distribution): ____________________.

(d) In interpreting one-way analysis-of-variance as a multiple regression model, the fitted value ($\hat{\mathrm{Y}}_{ij}$) for Y_{ij} (person j, group i) can be given very simply as ____________________.

(e) Tukey's method of multiple comparisons is based on the ____________________ distribution.

(f) In a single factor analysis-of-variance where the levels of the factor are labeled numerically and a particular polynomial response function is fit using these labels to generate an independent variable, the adequacy (lack-of-fit) of the model can be evaluated because a pure error term exists in the form of ____________________.

(g) Within a Model II analysis-of-variance, the proportion of variance accounted for by the treatments is typically referred to as the ____________________.

(h) When two experimental factors are present and each category or level of one factor occurs with each level of the other, the two factors are said to be ______________________.

(i) If treatment effects exist in Model I analysis of variance, then mean-square error is a(n) ______________________ estimate of σ^2.

(j) When there is no interaction, effects are said to be ______________________, because the effect of a combination of treatments is the sum of the effects of the treatments involved.

(k) An interaction that results in "crossed lines" in graphing the means irrespective of what factor is treated along the horizontal axis is called a(n) ______________________ interaction.

(l) Given J independent sample means, there can be no more than ______________________ comparisons, each comparison being independent of both the grand mean and of each other.

(m) An experimental arrangement is said to be ______________________, when within each treatment combination, there are at least two independent observations made under identical circumstances.

(n) A multivariate alternative to a repeated measures analysis based on Model III would use a statistic extending the usual paired t-test to more than one difference. This statistic is called ______________________ T^2 statistics.

(o) If a one-way analysis-of-variance model (fixed effects) is represented as $\mathrm{Y}_{ij} = \mu_i + \epsilon_{ij}$, it is termed a ______________________ linear model.

(p) A priori comparisons among means are also known as __________________ comparisons.

(q) In testing equality of variances in a one-way layout, Box developed a more robust modification of a test originally due to __________________.

(r) The reliability coefficient in classical test theory can be defined in terms of the __________________ using Model II analysis-of-variance.

(s) Multiple comparisons among the levels of the fixed factor in a Model III analysis, are carried out using __________________ as a replacement for MSE.

V. True or False

________________ (a) When there is no interaction, effects are said to be additive.

________________ (b) If all pairwise comparisons are of interest in a single factor analysis-of-variance layout and we wish to control the overall significance level, the Bonferroni method is generally superior to the Tukey method in the sense of leading to narrower confidence intervals.

________________ (c) If observations are proportions, variance-stabilizing transformations are generally defined through the use of the reciprocal transformation.

________________ (d) Scheffe's test for the equality of variances is generally less robust to nonnormality than is Bartlett's test.

________________ (e) In a single-factor analysis-of-variance based on Model II, the observations are all assumed to be independent.

________________ (f) In a one-way ANOVA a(n) ________________ is defined by a difference between a particular cell mean and a grand mean.

________________ (g) A p-value provided by a Geisser-Greenhouse test will never be larger than that provided by a Huynh-Feldt test.

________________ (h) There is no need to correct tests involving only between-subject factors with strategies such as those suggested by Huynh-Feldt or Geisser-Greenhouse.

____________________ (i) The variance of a sample comparison based on adjusted means in analysis-of-covariance is identical to the variance of the same sample comparison based on unadjusted means.

____________________ (j) In analysis-of-covariance, the degrees-of-freedom for the adjusted mean square for treatments is the same as the degrees of freedom for the unadjusted mean square for treatments.

____________________ (k) If all sample sizes are equal in a single-factor analysis-of-variance layout, the significance testing strategy is robust to a lack of independence among the error terms.

____________________ (l) In general, two comparisons are orthogonal if the sum of the pairwise products of their defining coefficients is zero.

____________________ (m) Given $r - 1$ mutually orthogonal comparisons in a one-way analysis-of-variance layout with r groups, the sum of the $r-1$ sum of squares for the comparisons is always equal to the sum of squares for treatments.

____________________ (n) Among all alternative hypotheses in a one-way fixed effects analysis-of-variance context with r groups for which some pair of means differ by a value T, the largest noncentrality parameter is generated for the alternative in which $r - 2$ means are equal and the other two differ by T.

____________________ (o) The F-test for equality among the means in Model III analysis-of-variance (which assumes that the error terms are independent) continues to be valid under the same assumptions plus a relaxation that allows the error terms to be correlated as long as these correlations are no greater than .50.

VI. Given a two-way ANOVA source table of the following form:

Source	df	SS
A	6	60
B	5	120
AxB	30	120
Error	126	756
Total	167	1056

What are the appropriate F-ratios and degrees-of-freedom in the following chart:

Test	A,B fixed	df	A,B random	df	A random; B fixed	df
A						
B						
AxB						

VII. suppose I have a one-way analysis-of-variance layout with 5 cells and equal n's. Construct a set of orthogonal comparisons that would "use up" all of the SS Between.