Name _____

Student Number _____

University of Toronto in Mississauga

December Examinations 2005

STA 442F

Duration - 3 hours

Aids allowed: Calculator. Printouts will be supplied.

$$F = \left(\frac{n-p}{s}\right)\left(\frac{a}{1-a}\right) \qquad a = \frac{R_F^2 - R_R^2}{1-R_R^2} = \frac{sF}{n-p+sF}$$

1. (8 Points) Make up an original study that would employ a one-way multivariate analysis of covariance with one quantitative covariate and two dependent variables. Give a paragraph briefly describing the study, followed by list of the variables; classify each variable as independent or dependent, and quantitative or categorical.

Continued on page 2

2. (8 Points) Answer each question below True or False. Write "T" or "F" on the line. You must get at least 9 out of 10 correct in order to get credit for this question.

___ In an observational study, a statistically significant relationship between the independent variable and the dependent variable can provide some evidence of a causal relationship if the study is well controlled.

___ If a subject (case) provides data for more than one value of an independent variable, we call that independent variable a *within-subjects* variable.

___ We observe $r = -0.70$, $p = .009$. We conclude that that high values of X tend to go with low values of Y and low values of X tend to go with high values of Y.

___ If $p < .05$ we say the results are statistically significant at the .05 level, and we conclude that the independent variable and the dependent variable are unrelated.

___ In a study attempting to predict income from education and race, there is a significant interaction between education and race. This means that income and race are related.

___ When you add another independent variable in multiple regression, R^2 cannot go down.

___ We observe $r = 0.50$, $p = .002$. This means that 50% of the variation in the dependent variable is explained by a linear relationship with the independent variable.

___ An experimental study is one in which cases are randomly assigned to the different values of an independent variable.

___ When a relationship between the independent variable and the dependent variable is *not* statistically significant, we conclude there is no relationship between the two variables in the population.

___ A multivariate analysis is one with multiple independent variables.

Continued on page 3

3. (7 Points) Recall the **cars data**, in which you want to test for differences in fuel efficiency as a function of country (actually, region) of origin, while controlling for weight (X_1) and length (X_2) of vehicle. You need to set up a regression model using *cell means coding*; there will be no β_0 in the model.

   (a) List all the independent variables in your model. Say how the dummy variables are defined.

   (b) Make a table with three rows, one for each country. Make a column for each dummy variable, and show the value of the dummy variable for each country. Make one more column, a wider one. In that column, give the expected value of Y. Your expected value formulas must have only numbers, betas (β) and the symbols X_1 and X_2.

Continued on page 4

4. Now please look at the printout for the cars data, and answer these questions:

   (a) (5 Points) We want to test whether fuel efficiency is related to *any* of weight, length and country of origin. The null hypothesis is that fuel efficiency is not related to any of these variables.

      i. What is the value of the test statistic? The answer is a single number.

      ii. What is the p-value? The answer is a number or possibly a range of numbers, if $p < .0001$.

      iii. Are the results statistically significant at the 0.05 level? Answer Yes or No.

   (b) (3 Points) For any country of origin, and holding weight constant at a fixed level, is length positively related to fuel efficiency, or is it negatively related? How do you know? Your answer to this question should be the words "Positively Related" or "Negatively related," and a single number — a parameter estimate.

   (c) (5 Points) Controlling for country of origin, we want to test whether fuel efficiency is related to weight, length, or both.

      i. What is the value of the test statistic? The answer is a single number.

      ii. What is the p-value? The answer is a number or possibly a range of numbers, if $p < .0001$.

      iii. Are the results statistically significant at the 0.05 level? Answer Yes or No.

Continued on page 5

(d) (6 Points) Controlling for weight and length, we want to test whether fuel efficiency is related to country of origin.

i. What is the value of the test statistic? The answer is a single number.

ii. What is the p-value? The answer is a number or possibly a range of numbers, if $p < .0001$.

iii. Are the results statistically significant at the 0.05 level? Answer Yes or No.

iv. After allowing for the reduced model, what proportion of the remaining variation is explained by this effect?

(e) (14 Points) Now we will look at *one* of the pairwise comparisons.

i. We want to test the difference between cars of U.S. versus Japanese origin, controlling for weight and length.

A. What is the value of the test statistic? The answer is a single number.

B. What is the p-value? The answer is a number or possibly a range of numbers, if $p < .0001$.

C. Are the results statistically significant at the 0.05 level? Answer Yes or No.

D. Are the results still statistically significant with a Bonferroni correction for the fact that we are making three pairwise comparisons? Answer Yes or No.

E. After allowing for the reduced model, what proportion of the remaining variation is explained by this effect?

Continued on page 6

ii. What is the predicted fuel efficiency in Kilometers per litre for a Japanese car of average weight and length? By average, I mean set weight and length to their sample mean values. The answer to this question is a single number. Please "show your work" by writing down the expression you are trying to calculate, and then doing it with your calculator.

iii. What is the predicted fuel efficiency in Kilometers per litre for a U.S. car of average weight and length? By average, I mean set weight and length to their sample mean values. The answer to this question is a single number. Please "show your work" by writing down the expression you are trying to calculate, and then doing it with your calculator.

iv. In plain, non-statistical language, what do you conclude about the relative fuel efficiency of U.S. and Japanese cars *once you allow for length and weight*? The answer is one short sentence; you have more room than you need.

Continued on page 7

5. (8 Points) Consider the Titanic data. For just adults, we want to know if survival was related to sex when one allows for class.

   (a) Give the value of the test statistic. Your answer is a single number.

   (b) The test is significant beyond any doubt. What degrees of freedom would you use to calculate the p-value? Your answer is a single number.

   (c) Are the Bonferroni-corrected tests on the sub-tables all significant at the 0.05 level? Answer Yes or No. If the answer is No, specify which ones are not significant.

   (d) In plain, non-statistical language, what do you conclude from this analysis?

6. (5 Points) In the analysis of the **shoe data**, I followed up a significant main effect for period (Before, During and After the advertising campaign) with Bonferroni-corrected pairwise comparisons of marginal means. In plain, non-statistical language, what do you conclude from the pairwise comparisons? *Marks will be deducted if you use even a single statistical term.*

Continued on page 8

7. (8 Points) Make up an original study with one categorical within-subjects independent variable, one quantitative between-subjects independent variable, and one categorical dependent variable. Give a paragraph briefly describing the study, followed by list of the variables; classify each variable as independent or dependent, and categorical or quantitative. State whether each independent variable is between-subjects or within-subjects.

Continued on page 9

8. (5 Points) For the **exercise tolerance data**, please focus upon the following question: Does the relationship of Smoking to exercise tolerance depend on Fat level? Answer the question and describe the results in plain, non-statistical language. Your answer should be guided by the results of a statistical test, but you must not mention that test explicitly.

9. (5 Points) For the **farm data**, we want to know whether the effect of irrigation method depends upon the type of fertilizer.

   (a) What is the value of the test statistic? The answer is a single number.

   (b) What is the p-value? The answer is a number or possibly a range of numbers, if $p < .0001$.

   (c) Are the results statistically significant at the 0.05 level? Answer Yes or No.

   (d) In plain, non-statistical language, what do you conclude?

Continued on page 10

10. (8 Points) Give an example of a true experimental study with one independent variable, and the conclusions of the study are invalid because of a confounding variable. Hint: consider a drug study without a double blind.

11. (5 Points) In the **wine study**, 6 judges rated 4 wines on a scale from 0 to 40, presented in random order and with a double blind. Based on the Bonferroni-corrected tests for pairwise differences between means, sort the wines into groups such that the members of each group of wines are *not* significantly different, and arrange the groups in order from highest to lowest quality. Here is an example of what I mean (of course this answer is incorrect):

    "Wine 3 is rated highest, followed by Wines 1 and 4, followed by Wine 2."

Total marks = 100 points