

University of Toronto at Mississauga
April Examinations 2004
STA 442S
Duration - 3 hours
Aids allowed: Calculator, Printouts
Note: Formula sheet supplied

Name (Please print) _____

Signature _____

Student Number _____

1. In a study of agricultural productivity, small apple farms were randomly assigned to use one of three Pesticides (Type A, B or C) and one of three Fertilizers (Type 1, 2 or 3). The dependent variable was total crop yield in kilograms, and there were two covariates -- number of trees on the farm, and crop yield last year.

a. (3 points) Is this study experimental, or observational, or both? Why?

b. (2 points) Define dummy variables for Pesticide using *effect coding* -- that's the scheme with the minus ones for the last category.

c. (2 points) Define dummy variables for Fertilizer using *effect coding* -- that's the scheme with the minus ones for the last category.

Continued on page 2

1d. (5 points) Now consider a full multiple regression model with an intercept, both covariates, both categorical independent variables, and the interaction between the categorical independent variables -- but no interaction between covariates, and no interaction between covariates and the categorical independent variables. That is, it's a model with parallel regression planes. Write $E[Y]$ for this model. Please put the covariates first.

e. (7 points) For each question in the table below, write the regression coefficients that would be missing from the reduced model. Use the notation of your expression for $E[Y]$ above. Some of the questions have identical answers.

Question	Regression coefficients missing from reduced model
Controlling for the covariates and all other effects, is there a main effect for Pesticide Type?	
Controlling for the covariates and all other effects, is there a main effect for Fertilizer Type?	
For constant values of number of apple trees and crop yield last year, is there any difference among the marginal means for the different types of fertilizer?	
Controlling for the covariates, is there an interaction between the Fertilizer type and Pesticide type?	
For constant values of number of apple trees and crop yield last year, is there any difference among the marginal means for the different types of pesticide?	
For constant values of number of apple trees and crop yield last year, does the effect of fertilizer type depend on type of pesticide?	
Allowing for number of apple trees and crop yield last year, does the effect of pesticide type depend on type of fertilizer?	

Continued on page 3

1f. (4 points) To make this question easier, suppose that there are *no covariates* in the apple study. Now say there were no significant effects except a main effect for fertilizer type. Would it be correct to conclude that differences in fertilizer type **caused** differences in crop yield? Answer Yes or No, and briefly justify your answer. Technical language is okay if it makes your answer shorter.

2. Now please refer to your printout from Assignment 9. That's the analysis of the Math ACT data.

- a. (2 points) What are the independent variables? What is the dependent variable?
- b. (3 points) Is this study experimental, observational, or both? Why?
- c. (3 points) Averaging across different High School math backgrounds, do males and females perform differently on average?
 - i) Is your answer to the question Yes, No, or No Conclusion?
 - ii) Are the results statistically significant at the 0.05 level: Yes or No?
 - iii) What is the value of the test statistic? The answer is a number.
 - iv) What is the p-value? The answer is a number.
 - v) What proportion of the remaining variation is explained by this effect?
- d. (3 points) Averaging across sex, is course work in High School mathematics related to achievement on the Math ACT?
 - i) Is your answer to the question Yes, No, or No Conclusion?
 - ii) Are the results statistically significant at the 0.05 level: Yes or No?
 - iii) What is the value of the test statistic? The answer is a number.
 - iv) What is the p-value? The answer is a number.
 - v) What proportion of the remaining variation is explained by this effect?

Continued on page 4

- 2e. (1 point) In your Bonferroni-corrected tests to follow up the last result, what p-value are you using to declare a follow-up test significant.
- f. (2 points) Give the three F-statistics you are using for the follow-up tests. Your answer to this question is three numbers. They need not be labelled, and order does not matter.
- g. (4 points) In plain, non-statistical language, what do you conclude from these follow-up tests?
- h. (3 points) Does the magnitude of the sex difference in performance depend on upon profile of course work in High School math?
- i) Is your answer to the question Yes, No, or No Conclusion?
 - ii) Are the results statistically significant at the 0.05 level: Yes or No?
 - iii) What is the value of the test statistic? The answer is a number.
 - iv) What is the p-value? The answer is a number.
 - v) What proportion of the remaining variation is explained by this effect?
- i. (4 points) Based on the results of this study, is it proper to conclude that amount of high school course work **causes** differences in performance on the ACT? Briefly justify your answer. Technical language is okay if it makes your answer shorter.

Continued on page 5

3. (3 Points) In the Math ACT study, suppose the main effect for high school math background explained 25% of the remaining *sample* variation after controlling for other effects. Please give the minimum sample size required for significance at the 0.05 level. The answer is a number.
4. (3 Points) In the Math ACT study, suppose the Sex by Math Background interaction explains 7% of the remaining *population* variation after controlling for other effects. What sample size is required for a power of 0.80? As usual, use $\alpha = 0.05$. The answer is a number.
5. (3 Points) In the bank discrimination study, we wanted to test the relationship between sex and starting salary, controlling for Education, Prior Experience, Age and Seniority. If the sex difference explained 10% of the remaining *sample* variation after controlling for the other variables, what sample size would be required to make this effect statistically significant at the 0.05 level? The answer is a number.
6. (3 Points) In the bank discrimination study, suppose we wanted to test Age and Seniority simultaneously, controlling for Sex, Education and Prior Experience. If the variables in question explain 7% of the remaining *population* variation after controlling for other independent variables, what sample size do you need to have a 90% chance of statistical significance at the 0.05 level? The answer is a number.

Continued on page 6

7. (12 Points) In a study of heart attack patients, the variables are Age category (Under 30 vs 30-50 vs Over 50), Family History of Heart Disease (Yes vs No), and Type of Heart Attack. The following symbols will be used for the variables.

A = Age Category

F = Family History of Heart Disease

H = Type of Heart Attack

The table below shows some hierarchical log-linear models that could be fit to the data.

a. [AFH]	b. [A] [FH]	c. [AF] [H]	d. [AH] [F]
e. [A] [F] [H]	f. [AF] [AH]	g. [AF] [FH]	h. [AF] [AH] [FH]

For each question below, write the letter of the reduced model you would test against the saturated model. Write the letter of the model next to the question.

- _____ (Points) Are any of the variables related to one another?
- _____ (Points) Is type of heart attack related to either of the other variables?
- _____ (Points) Controlling for age, is family history of heart disease related to type of heart attack?
- _____ (Points) Controlling for family history of heart disease, is age related to type of heart attack?
- _____ (Points) Does the relationship of family history to type of heart attack depend on age?
- _____ (Points) Does the relationship of age to type of heart attack depend on family history?

8. Now consider the Graduate School Admissions study of Assignment 11.

- a. (3 Points) Is this study experimental, observational, or both? Why?

Continued on page 7

8b. (2 Points) Look at your two-dimensional table of sex by whether or not the person was admitted, and the associated tests. You want to see whether these two variables are related if you ignore what department they applied to. For consistency with the log-linear part of this question, please base your answers on the likelihood ratio test. Are the variables related?

- i) Is your answer to the question Yes, No, or No Conclusion?
- ii) Are the results statistically significant at the 0.05 level: Yes or No?
- iii) What is the value of the test statistic? The answer is a number.
- iv) What is the p-value? The answer is a number.

c. (4 Points) Based just on the preceding analysis, what would you conclude? Use plain, non-statistical language.

d. (4 Points) Do the results of this analysis provide evidence of discrimination in admissions? Answer Yes or No, and say why. Technical language is okay if it makes your answer shorter.

e. (3 Points) Now test the relationship between sex and getting in to graduate school at Berkeley, *controlling for what department the student applied to*.

- i) Are the variables related when you control for Department? Yes, No, or No Conclusion?
- ii) Are the results statistically significant at the 0.05 level: Yes or No?

Continued on page 8

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

iv) What is the p-value? The answer is a number.

f. Now follow up that last result with Bonferroni-corrected tests of the sub-tables.

i) (1 Point) What p-value are you using to determine significance for a sub-table?
Show a little calculation and give a number.

ii) (1 Points) One of the follow-up tests is significant. Give the value of the test
statistic. The answer is a number.

iii) (4 Points) State your conclusion in plain language.

g. Finally, consider the test for three-way association among Major, Sex and Admission.

i) (2 Points) What is the value of the test statistic? The answer is a number.

ii) (4 Points) In plain language, what do you conclude? No credit for repeating the
words from the assignment without demonstrating that you understand.

Total Marks = 100 Points

STA442s04 Formulas

$$\text{SSTO} = \text{SSR} + \text{SSE} \quad R^2 = \frac{\text{SSR}}{\text{SSTO}} \quad R_c^2 = R_F^2 - R_R^2 \quad a = \frac{R_C^2}{1 - R_R^2}$$

There are $p-1$ independent variables in the full model; s of these variables are *not* in the reduced model.

$$F = \frac{(\text{SSR}_F - \text{SSR}_R) / s}{\text{MSE}_F} = \left(\frac{n-p}{s} \right) \left(\frac{a}{1-a} \right) \quad a = \frac{sF}{n-p + sF}$$