NAME (PRINT): _____
Last/Surname First /Given Name

STUDENT #: _____ **SIGNATURE:** _____

UNIVERSITY OF TORONTO MISSISSAUGA
DECEMBER 2009 FINAL EXAMINATION
STA442H5F
Methods of Applied Statistics
Lawrence Brunner
Duration - 3 hours
Aids: Calculator Model(s): Any calculator is okay

The University of Toronto Mississauga and you, as a student, share a commitment to academic integrity. You are reminded that you may be charged with an academic offence for possessing any unauthorized aids during the writing of an exam, including but not limited to any electronic devices with storage, such as cell phones, pagers, personal digital assistants (PDAs), iPods, and MP3 players. Unauthorized calculators and notes are also not permitted. Do not have any of these items in your possession in the area of your desk. Please turn the electronics off and put all unauthorized aids with your belongings at the front of the room before the examination begins. If any of these items are kept with you during the writing of your exam, you may be charged with an academic offence. A typical penalty may cause you to fail the course.

*Please note, you **CANNOT** petition to **RE-WRITE** an examination once you have begun writing.*

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

Qn. #	Value	Score		Qn. #	Value	Score
1	10			6	16	
2	5			7	10	
3	9			8	12	
4	10			9	18	
5	10					
Total = 100 Points						

10 points

1. In a study of the links between TV violence and aggression, parents of children in a daycare filled out a questionnaire about the TV programs their children watched. Daycare workers recorded the number of attacks and other violent incidents (taking a toy from another child, etc.) for each child. Can this study provide good evidence that violent TV can contribute to violent behaviour? Clearly answer Yes or No, and briefly discuss in terms of concepts from this course.

5 points

2. Make up an original study for which a two-factor multivariate analysis of covariance would be the appropriate tool. Both factors should be between-cases. After briefly describing the study,

   (a) List the independent variable or variables. For each one, say whether it is quantitative or categorical.

   (b) List the dependent variable or variables. For each one, say whether it is quantitative or categorical.

9 points

3. Copy the table below into your examination book. In each cell, write the name of the statistical technique you would recommend. Give the *most elementary test*. For example, any time you could use a matched t-test, you could also use multivariate regression. But "multivariate regression" is wrong. **Don't write your answers on this paper. Write them in the examination book.**

	Dependent Variable		
Independent Variable	Categorical: Two Categories	Categorical: More than Two Categories	Quantitative
Categorical: Two Categories			
Categorical: More than Two Categories			
Quantitative			

continued on page 3

10 points

4. Answer each question below True or False. Write the letters in your examination book, and then write "T" or "F" beside each letter. You must get at least 8 out of 10 correct in order to get credit for this question. **Don't write your answers on this paper. Write them in the examination book.**

   (a) 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.

   (b) The p-value is the probability that the null hypothesis is true.

   (c) We observe $r = -0.70$, $p = .009$. We conclude 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.

   (d) If $p < .05$ we say the results are statistically significant at the .05 level, and we do not have sufficient evidence to conclude that the independent variable and the dependent variable are related in the population.

   (e) 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.

   (f) The p-value is the probability of failing to replicate significant results in a second independent random sample of the same size.

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

   (h) The greater the p-value, the stronger the evidence that the independent and dependent variable are related.

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

   (j) 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.

continued on page 4

10 points

5. The `cars` data file has four variables: length, weight, origin and fuel efficiency in kilometers per litre, for a sample of cars. The three origins are US=1, Japanese=2 and European=3. Presumably these refer to the location of the head office, not to where the car was manufactured.

   (a) Let X_1 = Length, X_2 = Weight and Y = Fuel efficiency. Represent Origin by a collection of dummy variables; you can name them anything you want. *Use cell means coding!* At this point, you need not state how the dummy variables are defined; you'll do that below. For this part of the question, just write $E(Y|\mathbf{X} = \mathbf{x})$.

   (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 $E(Y|\mathbf{X} = \mathbf{x})$. Your expected value formulas must have only numbers, betas (β) and the symbols X_1 and X_2.

   (c) State the null hypothesis you'd test in order to answer this question: "Allowing for weight and length of car, is country of origin a useful predictor of fuel efficiency?" Your answer is a statement about the β quantities in your answer to Question 5a.

   (d) Suppose you want to test for a difference in fuel efficiency between Japanese and European cars, controlling for weight and length. State your null hypothesis in terms of β quantities.

16 points

6. Steel is made by heating iron and adding some carbon. A steel company conducted an experiment in which knife blades were manufactured using two different amounts of carbon (Low and High), and three different temperatures (Low, Medium and High). Of course even the Low temperature was very hot. A sample of knife blades was manufactured at each combination of carbon and temperature levels, and then the breaking strength of each blade was measured by a specially designed machine. The dependent variable is breaking strength.

   (a) The table below shows all six treatment combinations. Please copy the table into your examination book, and make columns giving the coefficients of the contrast or contrasts you would use to test for main effects of Temperature. **Don't write your answer on this paper. Write it in the examination book.**

Low Carbon, Low Temperature	
Low Carbon, Medium Temperature	
Low Carbon, High Temperature	
High Carbon, Low Temperature	
High Carbon, Medium Temperature	
High Carbon, High Temperature	

continued on page 5

(b) The table below shows all six treatment combinations. Again, copy the table into your examination book, and make columns giving the coefficients of the contrast or contrasts you would use to test the Temperature by Carbon Level interaction. **Don't write your answer on this paper. Write it in the examination book.**

Low Carbon, Low Temperature	
Low Carbon, Medium Temperature	
Low Carbon, High Temperature	
High Carbon, Low Temperature	
High Carbon, Medium Temperature	
High Carbon, High Temperature	

(c) One last time, here is the table showing the six treatment combinations. Again, copy the table into your examination book. Now make columns showing how you would set up dummy variables for both independent variables, using *effect coding*. **Don't write your answer on this paper. Write it in the examination book.**

Low Carbon, Low Temperature	
Low Carbon, Medium Temperature	
Low Carbon, High Temperature	
High Carbon, Low Temperature	
High Carbon, Medium Temperature	
High Carbon, High Temperature	

(d) Write $E(Y|\mathbf{X} = \mathbf{x})$ for the regression model, using the names from your table above. Include the interactions!

(e) Using the β values from your answer to the preceding question, state the null hypothesis you'd use to test whether the effect of carbon level on breaking strength depends on the temperature. **Don't write your answer on this paper. Write it in the examination book.**

continued on page 6

10 points

7. Please refer to the printout for the **Bird Keeping Study**. The main question is this: Controlling for all the other independent variables, is bird keeping associated with lung cancer?

   (a) Allowing for other potential risk factors, the estimated odds of cancer are ____ times as great for a bird keeper. Your answer is a single number from the printout.

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

   (c) What is the p-value? The answer is a single number from the printout.

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

   (e) In plain, non-technical language, what do you conclude, if anything?

12 points

8. Refer to the printout for the **Eating Study**.

   (a) Please follow this strategy. First, look at the multivariate tests. For each test, be guided by the usual 0.05 criterion for statistical significance. Then, if a multivariate test is significant, look at the corresponding univariate tests, with a Bonferroni correction for the number of corresponding univariate tests. (You probably know what I mean by "corresponding," but just to make it clearer, consider the multivariate test for the covariate. The corresponding univariate tests are the univariate tests for the covariate, one for each dependent variable.)

   Following this strategy, what if anything do you conclude from the Eating study? Use plain, non-statistical language. Do *not* mention any numbers.

   (b) Once you control for reported hunger and all other effects in the model, what proportion of the remaining variation in food eaten is explained by whether the subjects were friends? Your answer is a single number.

   Don't write your answers on this paper. Write them in the examination book.

continued on page 7

18 points

9. Finally, please refer to the printout for the **Dichotic Listening Study**. The two independent variables are Presentation (Signal in left ear and noise in right, Signal in right ear and noise in left, and Both signal and noise in both ears) and Noise Type (Hiphop music, classical music or Radio commercials).

   (a) Classify each of the independent variables as within-cases or between-cases.

   (b) What is the multivariate output from "Test 1" telling us? Use simple, non-statistical language.

   (c) Consider the test for main effect of Presentation.

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

      ii. What is the p-value? The answer is a single number from the printout.

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

      iv. In plain, non-technical language, what do you conclude, if anything? Base your conclusions on Bonferroni-protected pairwise comparisons of the marginal means for Presentation. **Don't write your answer on this paper. Write it in the examination book.**

Total Marks = 100 points