

Grant MacEwan College

Stat 252 – Winter 2008 – 270

Dr. Karen Buro

Final Exam – April 21, 2008

Name:	
Student ID:	

Instructions:

- This is a 3 hour exam.
- Print your name and student id on the heading of this page.
- The exam is closed book, only a calculator can be used during the exam.
- Show your work.
- Cheating is lame and may have some unpleasant consequences.
- When conducting a statistical test state the assumptions, and how you would check if they are met.

Good luck!

1. (a) (5 marks) What is the purpose of inferential statistics?
- (b) (5 marks) In inferential statistics what does it mean to be 95% confident?
- (c) (2 marks) In a multiple linear regression model a categorical variable at 5 levels shall be included. How many dummy variables need to be introduced?
- (d) (2 marks) When giving the adjusted coefficient of determination, what are we adjusting for?
- (e) (4 marks) What is the advantage of including the block factor in a randomized block design?
- (f) (4 marks) Describe a situation where one should consider a multiple comparison.
- (g) (4 marks) Why do we never accept H_0 ?
- (h) (4 marks) In multiple linear regression the predictor variables should not be strongly correlated. Why?
- (i) (4 marks) Give an example when one would use the Wilcoxon Signed-Rank Test.

2. Are mathematics majors better at problem solving than computer science majors?

Random samples from both majors at the UofA were given a problem solving test. Summary statistics for the scores on that test are presented in the table.

Major	Mathematics	Computer science
mean	83.6	72.9
standard deviation	4.3	3.8
sample size	36	36

- (a) (12 marks) Conduct a test at significance level of 0.05 to answer the question stated above.
- (b) (4 marks) Find a 95% confidence interval for the difference in the mean score on this test for math and computer science majors.
- (c) (6 marks) What can be learned from the confidence interval? Explain.
- (d) (3 marks) Which is the corresponding nonparametric test for the test used in part (a)?

3. Nature-Nurture

A 1989 study investigated the effect of heredity and environment on intelligence. From the adoptive register in France, researchers selected samples of adopted children whose biological and adoptive parents came from either the very highest or the very lowest socioeconomic status (SES) categories (based on years of education and occupation).

The 38 selected children were given intelligence quotient (IQ) tests.

Summary statistics for the IQ of the children are

SES of adoptive parents	SES of biological parents	sample size	mean	st. dev.
High	High	10	119.6	12.25
High	Low	10	103.6	12.71
Low	High	8	107.5	11.94
Low	Low	10	92.4	15.41

- (a) (8 marks) Sketch a line graph for the mean IQ, depending on the SES of adoptive and biological parents. Based on the graph should you expect to find main effects for the SES of biological and adoptive parents? Do you expect to find interaction. Explain your answers.
- (b) (4 marks) Give the model for the IQ of the children with factors SES of their adoptive and biological parents.
- (c) (6 marks) Find below an incomplete ANOVA table. Fill in the gaps.

Source	<i>SS</i>	<i>df</i>	$MS=SS/df$	<i>F</i>	P-value
adoptive	1277.388	?	1277.388	?	.011
biologic	2275.788	?	2275.788	13.024	.001
adoptive * biologic	?	?	1.906	.011	.917
Error	5941.200	34	174.741		
Total	9496.282	?			

- (d) (10 marks) Does the data provide sufficient evidence that the SES of biological and adoptive parents interact in their effect on the intelligence of the children?
- (e) (3 marks) Give an estimate for the standard deviation σ in the model.
- (f) (5 marks) A 95% confidence interval for the difference in the mean IQ between children of biological parents with high and low SES is found to be: [6.793, 24.307]
State the result in a sentence and interpret this result.
- (g) (2 marks) Which contrast was used to obtain that confidence interval?
- (h) (4 marks) A 95% confidence interval for the difference in the mean IQ between children with adoptive parents with high and low SES is found to be: [2.893, 20.407]
Based on the confidence intervals: Which SES seems to have a greater impact on the mean IQ of children, of biological or adoptive parents? Explain.
- (i) (5 marks) What are the model assumptions and how could one check if they are met?
- (j) (3 marks) Would it be meaningful to conduct a multiple comparison for the influence of the SES of adoptive parents on the mean IQ of adopted children? Why or why not?
4. In a study to determine a person's yearly income 10 years after high school(y), it was found that the two biggest predictors are the number of math courses taken (x_1) and number of hours worked per week during a person's senior year of high school (x_2).

The estimated regression function based on a sample of size 65 is

$$\hat{y} = 6000 + 4540x_1 + 1290x_2$$

- (a) (4 marks) Give the multiple linear regression model relating the average annual salary with the two predictor variables (no interaction).
- (b) (6 marks) Interpret the intercept and slopes within the context of the question.
- (c) (5 marks) $R_a^2 = 0.5450$. Interpret the value. What does it tell about the model utility?

- (d) (12 marks) Test if the slope for x_1 is positive. Assume that the value of the test statistic is $t_0 = 3.14$.
- (e) (5 marks) What are the assumptions for the multiple linear regression model, and how would you check, if they are met.
- (f) (4 marks) A 95% prediction interval for the yearly income of a person who took 8 math classes and worked 20 hours per week during her/his senior year in high school was found to be $[55100; 81140]$. Interpret this result.