Grant MacEwan College

Stat 252 – Fall 2008 – Section 100
Midterm Exam – October 28, 2008

Name:	
Student ID:	

Instructions:

- This is a 50 minute exam.
- Print your name and student id on the heading of this page.
- This midterm exam is closed book. Only a calculator, the formula sheet, and statistical tables are permitted.
- Write your answers **legibly** in the space below the question or the back of the paper. No other sheets are accepted.
- Show your work.
- No conversations, please.
- Cheating is lame and may have some unpleasant consequences.

Good luck!

Definitions – Concepts

1. (3 marks) What is the purpose of Inferential Statistics?

2. (1 mark) ANOVA stands for .....?

3. (3 marks) What is the difference between experiment wise error rate and comparison wise error rate?

4. (3 marks) What is reflected by the value of the p-value in a statistical test?

5. (4 marks) State the model for the 2-way ANOVA model (factorial design).

6. (3 marks) Why do we never "accept H_0"?

7. (3 marks) Sketch a line graph that would indicate a main effect of one factor and an interaction effect of the two factors on the mean of a response variable.

Applications – Interpretations

8. In a study by the Fraser Institute prescription drug prices in different provinces were compared. Some of their findings were on the the cost of a 40 mg prescription drug cost. The results are summarized in the following table

province	sample size	mean cost ($)	standard deviation ($)
Ontario	41	34.82	1.96
Manitoba	36	32.36	1.60

(a) (1 marks) Give a point estimate for the difference in the mean price for this drug in Ontario and Manitoba.

(b) (12 marks) Test at significance level of 0.05, if the data provide sufficient evidence that the mean price for the drug is higher in Ontario than in Manitoba.

(c) (4 marks) Obtain a 95% confidence interval for the difference in the mean price for the drug in Ontario and Manitoba.

(d) (3 marks) What does it mean to be 95% confident?

(e) (4 marks) Are the findings in part (b) and (c) consistent. Explain.

9. In order to prove that the selling price of a real estate object depends on the number of bathrooms a real estate agent randomly chose houses on the market from three different locations. The prices found are given in the following table in hundreds of $.

	Location			
Bathrooms	Greater Sudbury	Kingston	Windsor	mean
1	1797	2025	1578	1800
2	692	2199	1890	1594
3	5299	1725	2439	3154
4	3590	3290	7499	4793
mean	2844	2310	3351	2835

(a) (4 marks) Identify the design of this experiment, give the response variable and factor(s) used.

(b) (3 marks) Find the sum of squares for bathrooms. (Use SS-treatment= 20,000,000 for subsequent questions if you can not solve this part.)

(c) (8 marks) An incomplete ANOVA table for the data above is:

ANOVA Table

Source	*df*	*SS*	*MS*	*F*
bathr	?	?	?	?
location	?	2,170,999.5	?	?
Error	?	17,379,107.8	?	
Total	?	?		

Complete the ANOVA table.

(d) (2 marks) Give an estimate for the standard deviation σ in the model?

(e) (12 marks) Conduct an appropriate test at significance level of 5% to check if the data provide sufficient evidence that the number of bathrooms influences the mean price of houses.

(f) (3 marks) Does the test result indicate that you should follow up with a multiple comparison? Explain.

(g) (6 marks) A Bonferroni multiple comparison at experiment wise error rate of $\alpha = 0.05$ for the mean price of houses was conducted depending on the number of bathrooms.

i. (2 marks) What is the comparison wise error rate in this multiple comparison?

ii. (5 marks) Assume that the margin of error for all pairwise comparisons is 3000. Obtain a diagram illustrating the result of the multiple comparison. Interpret your result.

(h) (4 marks) In order to compare the mean prices of houses with 4 bathrooms with the mean price of all other houses, what contrast should be used.

(i) (3 marks) Assume the 95% confidence interval for the contrast is [-825, 2175]. Interpret the result.

(j) A QQplot for the residuals is displayed below. Does it indicate that the assumption that the data is normally distributed might be violated? Explain.