

Grant MacEwan College

Stat 151 – Fall 2005

Dr. Karen Buro

Final Exam – December 10th, 2005

Name:	
Student ID:	

Instructions:

- This is an *3 hour* exam.
- Print your name and student id on the heading of the first page.
- This final exam is closed book. Only a calculator, the formula sheet, and provided 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!
- Cheating is lame and may have some unpleasant consequences.

1. In a study the calcium content in wheat from a certain area for four different storage times was investigated.

The data is included in the following table:

Storage Period	Observations						mean	st.dev.
0 month	58.75	57.94	58.91	56.85	55.21	57.30	57.49	1.37
1 month	58.87	56.43	56.51	57.67	59.75	58.48	57.95	1.33
2 month	59.13	60.38	58.01	59.95	59.51	60.34	59.55	0.90
4 month	62.32	58.76	60.03	59.36	59.61	61.95	60.34	1.46

- (a) (6 marks)

Present a side-by-side boxplot for the calcium content in wheat stored 0 month and 4 month.

- (b) (7 marks)

The incomplete ANOVA table is:

Source	df	SS	MS	F	P-value
Period	?	32.14	?	?	0.003
Error	?	32.90	?		
Total	?	?			

Give a complete ANOVA table for the data.

(c) (6 marks)

Is there sufficient evidence to conclude that the mean calcium content is not the same for the different storage times? Test at a significance level of $\alpha = 0.05$.

(d) (6 marks)

Use a 95% confidence intervals to compare the mean calcium content in wheat stored for 0 month with the mean calcium content in wheat stored for 4 month. Comment on your result.

2. The following table gives the size of the living area (in square feet), x , and the selling price, y , of 10 residential properties.

x (sq. ft.)	y (thousand)	
1360	178.5	
1940	275.7	
1750	239.5	
1550	229.8	
1790	195.6	It is $\sum x_i = 17,290$
1750	210.3	$\sum y_i = 2,349.4$
2230	360.5	$\sum x_i y_i = 4,173,646$
1600	205.2	$\sum x_i^2 = 30,477,100$
1450	188.6	$\sum y_i^2 = 578,760$
1870	265.7	

- (a) (4 marks)

Sketch a scatterplot for the two variables (use the back of the paper). Comment on the relationship between x and y .

- (b) (4 marks)

Calculate Pearson Correlation Coefficient for x and y .

What can be learned from this number?

- (c) (4 marks)

Give an estimate for the least squares line for x and y .

(d) (2 marks)

Give an interpretation of the slope in the context of this problem.

(e) (2 marks)

Estimate the selling price of a residential property with a living area of 1500 sq.ft.

3. A survey was conducted to investigate the interest of middle-aged adults in physical fitness programs in Rhode Island, Colorado, California, and Florida. The objective of the investigation was to determine whether adult participation in physical fitness programs varies from one region of the US to another. A random sample of people were interviewed in each state and these data were recorded:

	Rhode Island	Colorado	California	Florida
Participate	46	63	108	121
Do not participate	149	178	192	179

(a) (3 marks)

Give a complete two way table, including marginal counts (row and column totals).

(b) (10 marks)

Do the data indicate a difference in adult participation in physical fitness programs from one state to another?

4. Assume that the level of nitrogen oxide (NOX) in the exhaust of a particular car model, when driven in the city traffic can be modelled by a normal distribution with mean $\mu = 2.1$ grams/km and a standard deviation $\sigma = 0.5$ grams/km.

(a) (3 marks)

If the Environmental Protection Agency (EPA) mandates that a nitrogen oxide level of 2.7 grams/km cannot be exceeded, what proportion of the cars of this model would be in violation of the mandate?

(b) (3 marks)

At most, 25% of the cars of this model exceed what NOX level?

(c) (4 marks)

Obtain the probability that the average NOX level of 16 cars of this model is at least 2.4 grams/km.

5. The cost of automobile insurance is a sore subject in California. The following table gives the 6-month premiums in 2001 for a married male, licensed for 6–8 years, who drives about 15000 miles per year, and who has no accidents or violations:

City	Allstate	21st Century
Long Beach	\$ 1050	\$ 682
Pomona	\$ 984	\$ 638
San Bernadino	\$ 900	\$ 578
Moreno Valley	\$ 964	\$ 524

The following values might help in solving this question: $s_1 = 61.8$, $s_2 = 69.1$, $s_d = 50.9$.

(a) (2 marks)

Why would you expect these pairs of observations to be dependent?

(b) (10 marks)

Do the data provide sufficient evidence to indicate that there is a difference in the average 6-month premiums for Allstate and 21st Century insurance. Test using $\alpha = 0.05$.

(c) (4 marks)

Find a 95% confidence interval for the true difference in the average 6-month premiums for Allstate and 21st Century insurance.

(d) (2marks)

What does it mean in Inferential Statistics to be "95% confident"?

6. A Computer Chess Program is playing 8 games against a human champion. Assume that the two opponents are evenly matched.

(a) (2 marks)

In such a match, how many games do you expect the computer program to win?

(b) (2 marks)

Give the probability that the computer program wins exactly 5 games.

(c) What is the probability that the computer program loses all but one game?

7. (a) (2 marks)

What is the difference between the standard deviation σ and the standard deviation s ?

(b) (2 marks)

The data description for a qualitative variable should include which parts?

(c) (6 marks)

What is Inferential Statistics?

(d) (3 marks)

In ANOVA, what is the difference between the Total Sum Of Squares and the Sum of Squares for Treatments (Groups)?

(e) (3 marks)

Give an example, when we should use a significance level smaller than 0.05. Explain.