Grant MacEwan University

Stat 161 – Fall 2015

Dr. Karen Buro

Final Exam Sample Questions

Instructions:

- This final exam is closed book. Only a calculator, the formula sheet, and provided tables are permitted.
- 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 were 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) (10 marks)
   Is there sufficient evidence to conclude that the mean calcium content is not the same after 0 and 4 month? Test at a significance level of $\alpha = 0.05$.

   (c) (8 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. Give a complete interpretation of your result.

   (d) (4 marks)
   Are the results from (b) and (c) consistent?

   (e) (10 marks)
   Do the data provide sufficient evidence that the mean contents of calcium after four month of storage exceeds 60?

   (f) (8 marks)
   Find a 97% confidence interval for the mean calcium content in wheat after 4 month of storage. Interpret.

   (g) (6 marks)
   Could this data be used to learn about the relationship between storage time and calcium content in wheat using a linear regression model(explain)? If yes, would the slope be positive or negative, explain.

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
1750	210.3
2230	360.5
1600	205.2
1450	188.6
1870	265.7

It is
$\sum x_i = 17,290$
$\sum y_i = 2,349.4$
$\sum x_i y_i = 4,173,646$
$\sum x_i^2 = 30,477,100$
$\sum y_i^2 = 578,760$
$s_x = 254.45, s_y = 54.56, s_{xy} = 12392.6$

(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)
State the linear regression mode relating the variables under investigation.

(c) (4 marks)
Calculate Pearson Correlation Coefficient for x and y.
What can be learned from this number?

(d) (4 marks)
Give an estimate for the least squares line for x and y.

(e) (2 marks)
Give an interpretation of intercept and slope in the context of this problem.

(f) (8 marks)
Give a confidence interval for the slope of the linear regression function and interpret, use that the standard error $s_{b_1} = 0.03417$. (If you could not find b_1 you may use $b_1 = 0.2$)

(g) (3 marks)
Is it appropriate to use the result obtained through this data to estimate the selling price of a residential property with living area of 2500 sq.ft.? Explain.

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) (3 marks)
When randomly choosing one person from the sample, what is the probability the person was from Colorado and participated in a fitness program?

(c) (3 marks)
When randomly choosing one person from the sample, what is the probability the person participated in a fitness program?

(d) (3 marks)
When randomly choosing one person from the sample, what is the probability the person participated in a fitness program given the person is from Colorado?

(e) (3 marks)
When randomly choosing one person from the sample, what is the probability the person participated in a fitness program?

(e) (3 marks)
Are the events being from Colorado and participating in a fitness program independent from each other? Explain.

(e) (3 marks)
Are the events being from Colorado and participating in a fitness program mutually exclusive? Explain.

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 of $\sigma = 0.5$ 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)
25% of the cars of this model exceed which 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.

(d) (4 marks)
How large a sample from cars of this model should be taken if one wants to be 95% confident that the true mean NOX level is estimated within a margin of error of 0.25 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.

6. (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 inclose which parts?

(c) (6 marks)
What is the purpose of Inferential Statistics?

(d) (20 marks)
What is the subject of Statistics?

(e) (4 marks)
What does it mean in Inferential Statistics to be "95% confident"?

(f) (6 marks)
What is measured by the P-value of a test?

(g) (6 marks)
When rejecting a null hypothesis at significance level of 5%, what can we say about the likelihood that this decision is wrong?

(h) (6 marks)
When NOT rejecting a null hypothesis at significance level of 5%, what can we say about the likelihood that this decision is wrong?

(i) (3 marks)
Sketch a right skewed histogram.