

Grant MacEwan College

Stat 252

Sample questions for Midterm Exam

Definitions – Concepts

1. What information is given through the p-value in a statistical test?
2. Why would we include a block variable in an ANOVA model?
3. What is measured through the Sum of Squares for Treatment in a One-way ANOVA?
4. Give an example, for which you would use a paired t-test.
5. Sketch a line diagram illustrating the means for a response variable in a factorial design, that indicates no main effects, but interaction between two factors.
6. Sketch a QQ-plot that would indicate that the data it is based on is normally distributed / is not normally distributed.
7. When we state: At significance level of 5% the data provide sufficient evidence that the mean is greater than 0.

What is indicated by the significance level of 5%?

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

Applications – Interpretations

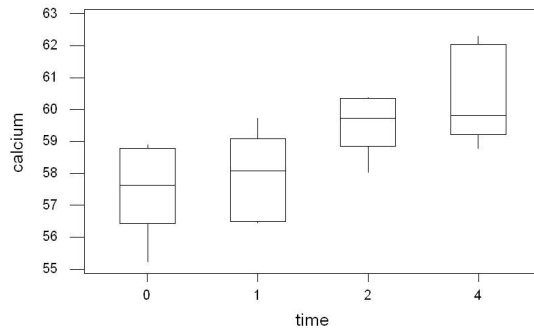
9. In a study the calcium content in wheat from a certain area for four different storage times was investigated.

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)

The following graph illustrates the data in a side by side box plot.



Comment on the graph (center, spread shape).

(b) (6 marks)

Use the box plot and the summary statistics to conclude if you have any concerns, that the assumption for an ANOVA model might be violated. Explain.

(c) (3 marks)

Find the sum of squares for treatments, SST for this data.

(d) (8 marks)

Complete the ANOVA table below:

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

(e) (12 marks)

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

(f) (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. (use the proper contrast)

(g) (8 marks) Assume that the margin of error for the Bonferroni procedure for the mean calcium level after storage times of 0, 1, 2, and 4 month equals $ME = 2.165$.

Why is the margin of error the same for all comparisons?

Use a diagram to illustrate the result of Bonferroni's procedure and interpret the result.

10. 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) (12 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. Conduct a test using $\alpha = 0.05$ (apply the p-value approach).

(c) (4 marks)

Find a 95% confidence interval for the true difference in the average 6-month premiums for Allstate and 21st Century insurance. Interpret the result, does it indicate a difference in the mean cost for insurance from Allstate and 21st Century?.