Name: _____
Student No.: _____

University of Alberta

Department of Mathematical and Statistical Sciences

Statistics 252 - Midterm Examination Version A (Solution)

Date: November 5, 2010

Instructor: Alireza Simchi

Time: 10:00-10:50

Instructions: (READ ALL INSTRUCTIONS CAREFULLY.)

1. This is a closed book exam. You are permitted to use a non-programmable calculator. Please turn off your cellular phones or pagers.

2. The exam consists of **two** parts. In the first part there are 15 multiple-choice questions. For each multiple-choice question choose the answer that is closest to being correct. Circle one of the letters (a)-(e) **on the second page** corresponding to your chosen answer for each question. All answers will be graded right or wrong (no partial credit) in this part. Each single question is worth 1 point. All numerical answers are rounded. In the second part, there are long-answer problems. **Show all your work to get full credit**. In fact, answers must be accomplished by adequate justification. If you run out of space, use the back of any page for answers as needed. Clearly direct the marker to answers that you provide on the back of a page.

3. This exam has **7** pages including this cover. Please ensure that you have all pages and write your name and your student ID at the top of each page.

4. The statistical tables and formula sheet are provided in a separate booklet.

5. The exam is graded out of a total of **25** points.

Name: _____

Signature: _____

Component	Worth	Mark
Multiple-choice	15	
Question 16	10	
Total	25	

Circle one answer for each question on the following table. Each question is worth 1 mark.

Question	Answer
1	a b c d **e**
2	**a b c d e**
3	a b **c** d e
4	**a b c d e**
5	a **b** c d e
6	**a b c d e**
7	a **b** c d e
8	**a b c d e**
9	a **b** c d e
10	**a b c d e**
11	a b **c** d e
12	**a b c d e**
13	a b **c** d e
14	**a b c d e**
15	**a** b c d e

PART 1

The effect of a new antidepressant drug on reducing the severity of depression was studied in manic-depressive patients at three mental hospitals. In each hospital all such patients were randomly assigned to either a treatment (new drug) or a control (old drug) group with different doses. The results of this experiment are summarized in the following tables; a high mean score indicates more of a lowering in depression level than does a low mean score.

Table 1: The summary statistics of scores for the 6 groups:

Group	Hospital	Drug	n	Dose	Average	S.D.
1	A	New	8	0.5	8.0	1.195
2	A	Old	6	0.5	5.5	1.049
3	B	New	10	0.1	5.0	1.354
4	B	Old	9	0.1	3.0	1.146
5	C	New	4	0.3	6.4	1.317
6	C	Old	5	0.3	4.8	1.643

Table2: The ANOVA table for the scores for the 6 groups:

ANOVA

Score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	?	?	?	?	?.
Within Groups	58.500	?	?	?	
Total	?				

Table 3: The ANOVA table for the comparison of average scores of patients using new drug versus old drug (ignoring hospitals and doses):

ANOVA

Score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	?	1	?	?	?
Within Groups	123.415	?	?		
Total	171.636	?			

Table 4: The ANOVA table for the comparison of average scores of patients between hospitals (ignoring drugs and doses):

ANOVA

Score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	?	2	?	?	?
Within Groups	104.565	?	?		
Total	?	?			

1. Suppose we would like to carry out a test to determine if there are any significant differences among the above six groups. What is extra sum of squares to the nearest integer?

a) 46 **b)** 48 **c)** 65 **d)** 67 **e)** 113

2. Suppose we would like to carry out a t-test to determine if there is a significant difference in the average scores of patients between two drugs (new or old) ignoring hospitals and doses. In fact, we would like to test $H_0 : \mu_{\text{New}} = \mu_{\text{Old}}$. What is (approximately) the value of the test statistic?

a) 3.0 **b)** 3.2 **c)** 3.5 **d)** 3.7 **e)** 4.0

3. Consider a linear contrast in terms of the 6 group means to describe the effect of dose on the average scores of patients. What is (approximately) the estimate of the contrast?

a) 2.9 **b)** 5.6 **c)** 7.2 **d)** 9.4 **e)** 12.5

4. Suppose the researchers conjectured that the new medications caused decrease in average scores of patients and this decrease accumulated as the medication dose was increased. Define a contrast that is a measure of the linear relationship between the dose of the drug and the mean of decrease in scores of patients. What is (approximately) the estimate of the contrast?

a) 0.6 **b)** 1.5 **c)** 2.9 **d)** 3.4 **e)** 8.7

Consider a linear contrast (say γ) that will define the effect of drug (new vs. old) on the average scores of patients. Questions **5** to **8** are related to this contrast.

5. What is (approximately) the estimate of the contrast in absolute value?

a) 1.4 **b)** 2.0 **c)** 3.6 **d)** 4.3 **e)** 6.1

6. What is (approximately) the standard error of the estimate of the contrast?

a) 0.02 **b)** 0.4 **c)** 1.2 **d)** 2.5 **e)** 3.7

7. What is the distribution of the test statistic under the null hypothesis of $H_0 : \gamma = 0$?

a) t(34) **b)** t(36) **c)** t(38) **d)** t(40) **e)** t(42)

8. The range of p-value for $H_0 : \gamma = 0$ versus $H_1 : \gamma \neq 0$ can be described as:

a) less than 0.001 **b)** between 0.001 and 0.01 **c)** between 0.01 and 0.02

d) between 0.02 and 0.05 **e)** greater than 0.05

Suppose we would like to use Bonferroni method to calculate three simultaneous 88% confidence intervals for the difference in average scores of patients using new drug and old drug for each hospital (one for each hospital). Questions **9** and **10** are related to this method.

9. What is (approximately) the critical value for 88% family-wise confidence intervals?

a) 2.0 **b)** 2.1 **c)** 2.4 **d)** 2.7 **e)** 2.9

10. What are (approximately) the margins of error for the three 88% family-wise confidence intervals, sorted in ascending order?

a) 1.1, 1.4, 1.7 **b)** 1.4, 1.7, 2.1 c) 1.7, 1.9, 2.3
d) 2.3, 2.4, 2.7 **e)** 1.7, 2.0, 2.5

Suppose we would like to carry out a single overall test to determine if there is a significant difference in the average scores of patients among different hospitals who have used the same drug (either new or old). Questions **11** to **15** are related to this test.

11. What is the null hypothesis? **(The correct answer is C)**

(a) $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$

(b) $H_0 : \mu_1 = \mu_2 = \mu_3$, $\mu_4 = \mu_5 = \mu_6$

(c) $H_0 : \mu_1 = \mu_3 = \mu_5$, $\mu_2 = \mu_4 = \mu_6$

(d) $H_0 : \mu_1 = \mu_2$, $\mu_3 = \mu_4$, $\mu_5 = \mu_6$

(e) $H_0 : \frac{1}{3}\mu_1 - \frac{1}{3}\mu_2 + \frac{1}{3}\mu_3 - \frac{1}{3}\mu_4 + \frac{1}{3}\mu_5 - \frac{1}{3}\mu_6 = 0$

12. What is the extra degree of freedom?

a) 1 **b)** 2 **c)** 3 **d)** 4 **e)** 5

13. What is the residual sum of squares for reduced model?

a) 58.500 **b)** 104.565 **c)** 123.415 **d)** 171.636 **e)** 214.692

14. What is (approximately) the value of the test statistic?

a) 9.4 **b)** 10.0 **c)** 12.5 **d)** 13.9 **e)** 15.6

15. The range of p-value can be described as:

a) less than 0.01 **b)** between 0.01 and 0.05 **c)** between 0.05 and 0.1

d) between 0.1 and 0.2 **e)** greater than 0.2

PART 2

16. (7 marks in total) Researchers measured the specific activity of the enzyme sucrase extracted from portions of the intestines of 24 patients who underwent an intestinal bypass. After the sections were extracted, they were homogenized and analyzed for enzyme activity. Two different methods can be used to measure the activity of sucrase: the homogenate method and the pellet method. In general, the pellet method is more time-consuming than homogenate method, yet it provides a more accurate measure of sucrase activity. Using SPSS output in the following, answer parts **(a)** to **(d)** based on the model $\mu_Y = \beta_0 + \beta_1 X$, where X is the homogenate method and Y is the pellet method.

	N	Mean	Std. Deviation
X	24	52.024	30.791
Y	24	156.175	100.155
Valid N (listwise)	24		

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	173945.330	1	173945.330	67.408	.000[a]
	Residual	56770.432	22	2580.474		
	Total	230715.762	23			

a. Predictors: (Constant), X

b. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.239	20.684		.447	.659
	X	2.824	.344	.868	8.210	.000

a. Dependent Variable: Y

(a) (2 marks) What percentage of variation in Y is explained by the regression model (or X)?

Solution:

$$R^2 = \frac{SS(\text{Re}\,gression)}{SS(Total)} = \frac{173945.330}{230715.762} = 0.7539$$

Therefore, 75.39% of variation in y is explained by the regression model (or X).

(1 mark for calculating R^2 and one mark for interpretation)

(b) (1 mark) Using the ANOVA table, what is the estimated standard error of the model (or the estimated standard deviation of error)? (**Hint:** Just find $\hat{\sigma}$.)

Solution:

$$\hat{\sigma} = \sqrt{MS(\text{Re}\,sidual)} = \sqrt{2580.474} = 50.7984$$ **(1 mark)**

(c) (3 marks) Obtain a 95% confidence interval for β_1. Interpret your confidence interval.

Parameter: β_1

Estimate: $\hat{\beta}_1 = 2.824$ **Standard error of estimate:** $S.E.(\hat{\beta}_1) = 0.344$

$1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$,
$df = n-2 = 24-2 = 22$. **(0.25 marks)**

So , the critical value is $t^*_{22,0.025} = 2.074$ **(0.25 marks)**

A 95% confidence interval for β_1 is:

$$\hat{\beta}_1 \pm t^*_{22,0.025} S.E.(\hat{\beta}_1) = 2.824 \pm 2.074(0.344) = 2.824 \pm 0.713$$ **(1 mark)**

Therefore, a 95% confidence interval for β_1 is: (2.111, 3.537).
(0.5 marks for final correct confidence interval)

Conclusion: It is estimated with 95% confidence that the effect of X on mean of Y is between 2.111 to 3.537. **(1 mark)**

(d) (4 marks) Obtain a 95% confidence interval of the mean sucrase activity measured by pellet methods for patients whose sucrase activity measured by homogenate methods is 50. Interpret your confidence interval.

Parameter: $\mu\{Y \mid x = 50\} = \beta_0 + \beta_1 x = \beta_0 + \beta_1(50)$

Estimate: $\hat{\mu}\{Y \mid x = 50\} = \hat{\beta}_0 + \hat{\beta}_1 x = 9.239 + 2.824(50) = 150.439$ **(0.5 marks)**

Standard error of estimate:

$$S.E.(\hat{\mu}\{Y \mid x = 50\}) = s\sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{(n-1)s_x^2}} = 50.7984\sqrt{\frac{1}{24} + \frac{(50 - 52.024)^2}{(24-1)(30.791^2)}} = 10.3925$$

(1 mark)

$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$,
$df = n - 2 = 24 - 2 = 22$. So , the critical value is $t^*_{22,0.025} = 2.074$ **(0.5 marks)**

A 95% confidence interval for $\mu\{Y \mid x = 50\} = \beta_0 + \beta_1(50)$ is:

$$\hat{\mu}\{Y \mid x = 50\} \pm t^*_{14,0.025} S.E.\left(\hat{\mu}\{Y \mid x = 50\}\right) = 150.439 \pm 2.074(10.3925) \Rightarrow 150.439 \pm 21.554$$

(0.5 marks)

Therefore, a 95% confidence interval for $\mu\{Y \mid x = 50\} = \beta_0 + \beta_1(50)$ is: (128.9, 172.0).

(0.5 marks for final correct confidence interval)

Conclusion: It is estimated with 95% confidence that the mean sucrase activity measured by pellet methods for patients whose sucrase activity measured by homogenate methods is 50 is between 128.9 and 172.0. **(1 mark\)**