Name: _____
Student No.: _____

University of Alberta

Department of Mathematical and Statistical Sciences

Statistics 252 - Midterm Examination Version A Solution

Date: March 7, 2012

Instructor: Alireza Simchi

Time: 12:00 noon - 12:50 p.m.

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 **five** parts. In the parts one to four, 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. Question **16** is a long-answer question. **Show all your work to get full credit**. In fact, answers must have 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.

6. **When referring to "log", I am always referring to the natural log.**

Name: _____
Student No.:_____

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**

Name: _____
Student No.:_____

PART 1

Researchers into the genetic disease sickle cell anemia are interested in how red blood cells adhere to endothelial cells, which form the innermost lining of blood vessels. A set of 14 blood samples are obtained, and each sample is split in half. One half of the blood sample is profused over an endothelial monolayer of type A, and the other half of the blood sample is profused over an endothelial monolayer of type B. The two types of monolayer differ in respect to the simulation conditions of the endothelial cells. After recording data, the researchers would like to carry out a test to determine if (there is any evidence that) the different stimulation conditions affect the adhesion of red blood cells.

1. In the test for any mean differences, what is the distribution of the test statistic under the null hypothesis?

a) t(6) **b)** t(12) **c)** t(13) **d)** t(26) **e)** t(27)

PART 2

Random samples of recent hospital "short" stays for both men and women reveled that a 95% confidence interval for the difference in means, $\mu_{Male} - \mu_{Female}$, was -0.1 to 2.7. Suppose this interval was based on equal variance assumption and sample sizes of 32 men and 30 women. A medical researcher wishes to test whether the average length of "short" hospital stays for men is higher than that for women. Questions **2** to **5** are related to this test.

2. What is the estimate for the common standard deviation of the two groups (males and females)?

a) 0.18 **b)** 0.75 **c)** 1.18 **d)** 2.75 **e)** 3.18

3. What is the value of the test statistic?

a) 0.50 **b)** 0.92 **c)** 1.37 **d)** 1.86 **e)** 2.00

4. The range of p-value can be describe as:

a) less than 0.025
b) between 0.025 and 0.05
c) between 0.05 and 0.10
d) between 0.10 and 0.25
e) greater than 0.25

PART 3

Researchers recorded the yields of corn, in bushels per plot, for five different varieties of corn, A, B, C, D and E. In a controlled greenhouse experiment, the researchers randomly assigned each variety to eight of 40 plots available for the study. The ANOVA table for comparing all five varieties is given in the following:

ANOVA

yeilds

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	?	4	2.123	?	.000
Within Groups	?	?	?		
Total	17.244	39			

Suppose the goal of study was to compare **all** means two at a time (a total of 10 pair-wise comparisons). Therefore, the researcher decided to use Bonferroni method to calculate10 simultaneous 95% confidence interval for the difference in the average of yield for different varieties. Questions **5** and **6** are related to this method.

5. What is (approximately) the critical value for 95% family-wise confidence intervals?

a) 2.0 **b)** 2.1 **c)** 2.4 **d)** 2.7 **e)** 3.0

6. What is (approximately) the margin of error for all 95% family-wise confidence intervals?

a) 0.15 **b)** 0.75 **c)** 1.15 **d)** 1.75 **e)** 2.15

PART 4

To assess the relative of three different supplements on average mileage (mi/gal), researchers conducted an experiment using 24 automobiles of the same type, model, and engine size with four randomly assigned to one of the blends and the mileage per gallon was recorded for each car. The six gasoline blends are described as follow:

Blend	**Description**	
1	C	Control
2	X	Control + Supplement X
3	Y	Control + Supplement Y
4	Z	Control + Supplement Z
5	X+Y	Control + Supplement X + Supplement Y
6	X+Z	Control + Supplement X + Supplement Z.

Note: Because of the chemical make up of supplements Y and Z, they could not be combined in the same blend.

Summary statistics and ANOVA table are given in the following:

Blend	N	Mean	Std. Deviation	Std. Error
1.00	4	24.80	3.59166	1.79583
2.00	4	26.75	3.27058	1.63529
3.00	4	32.675	5.53737	2.76869
4.00	4	33.35	4.36539	2.18270
5.00	4	43.85	3.11181	1.55590
6.00	4	32.00	3.57305	1.78652
Total	24	32.2375	7.14556	1.45858

ANOVA

Mileage

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	887.059	5	?	?	?
Within Groups	?	?	?		
Total	1174.356	?			

Suppose we would like to carry out a test to determine if there is any real difference among average mileage for the six different blends? Questions **7** to **9** are related to this test.

7. In the test for any mean differences, what is the distribution of the test statistic under the null hypothesis?

a) F(5,17) **b)** F(5,18) **c)** F(5, 19) **d)** F(5,20) **e)** F(5,24)

8. In the test for any mean differences, what is the best estimate for the common standard deviation of the six gasoline blends?

a) 3.45 **b)** 3.79 **c)** 3.88 **d)** 3.99 **e)** 4.11

9. In the test for any mean differences, what is the value of the test statistic (approximately)?

a) 10.5 **b)** 11.1 **c)** 11.7 **d)** 12.4 **e)** 14.8

Consider the effects of supplement Z, controlling for the other supplements. Is there an overall average supplement Z effect, all else fixed? That is, does adding supplement Z to gasoline increase mileage, on average? Define a contrast, say γ_1, in terms of the 6 group means for this question. Questions **10** to **12** are related to this contrast.

10. What is the estimate of the contrast, rounded to the nearest integer?

a) 3 **b)** 4 **c)** 5 **d)** 6 **e)** 7

11. What is the standard error of the estimate of the contrast, rounded to the nearest integer?

a) 1 **b)** 2 **c)** 3 **d)** 4 **e)** 6

12. For testing $H_0 : \gamma_1 = 0$ versus $H_1 : \gamma_1 > 0$, the p-value can be described as:

(a) less than 0.001
(b) between 0.001 and 0.0025
(c) between 0.0025 and 0.01
(d) between 0.01 and 0.05
(e) greater than 0.05

Consider the effects of supplement Z, controlling for the other supplements. Does the effect of supplement Z, depend on whether or not supplement X is present? Define a contrast, say γ_2, in terms of the 6 group means for this question. Questions **13** to **15** are related to this contrast.

13. What is the estimate of the contrast, rounded to the nearest integer?

a) 3 **b)** 4 **c)** 5 **d)** 6 **e)** 7

14. What is the standard error of the estimate of the contrast, rounded to the nearest integer?

a) 1 **b)** 2 **c)** 3 **d)** 4 **e)** 6

15. For testing $H_0 : \gamma_2 = 0$ versus $H_1 : \gamma_2 \neq 0$, the p-value can be described as:

(a) less than 0.001
(b) between 0.001 and 0.01
(c) between 0.01 and 0.05
(d) between 0.05 and 0.1
(e) greater than 0.1

PART 5

16. (10 Marks in Total) Does shelf life of cough syrup depend on storage temperature? Cough syrup manufacturers recommend that after a bottle is unsealed it be kept under cool conditions. They claim the shelf life of the cough syrup is dependent on the temperature at which it is stored. An independent quality control laboratory has obtained data on the shelf life (in days) and storage temperature (in degrees Celsius) for 28 bottles of cough syrup. A regression analysis was proposed to relate the shelf life to storage temperature.

Here is the SPSS output for a SLR analysis of average shelf life on storage temperature.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	79718.558				.000[a]
	Residual	53011.870				
	Total	132730.429				

a. Predictors: (Constant), temp

b. Dependent Variable: life

Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	(Constant)	838.500	34.829		.000
	temp	-9.629	1.540		.000

a. Dependent Variable: life

a) (2 marks) State the value of test-statistic and the null distribution of the test-statistic for the test of any linear significance.

Solution

The value of the Test-Statistic: $TS = \dfrac{\hat{\beta}_1 - 0}{S.E.(\hat{\beta}_1)} = \dfrac{-9.629 - 0}{1.540} = -6.252$ **(One mark)**

Null Distribution: Test statistic has a t-distribution with $df = n - 2 = 28 - 2 = 26$, if H_0 is true. **(One mark)**

b) (2 marks) Find the coefficient of determination. Interpret it

Solution:

$$R^2 = \frac{SS(\text{Re}\,gression)}{SS(Total)} = \frac{79718.558}{132730.429} = 0.600605$$ **(One mark)**

Interpretation: 60.0605% of variation in shelf life is explained by the simple linear regression model (or by temperature). **(One mark)**

OR

Prediction error is 60.0605% smaller when we use regression line to predict shelf life, y, instead of using simple average, $\bar{y}$ to predict y.

c) (2 marks) What is the linear correlation between shelf life and storage temperature?

Solution: For simple linear regression, we have $R^2 = r^2$. In addition, sample correlation and estimate of the slope of the regression line have same sing. Since $\hat{\beta}_1 = -9.629$ is negative, sample correlation is negative too. Hence, $r = -\sqrt{R^2} = -\sqrt{0.600605} = -0.775$
(One mark for square root and one mark for negative sign)

d) (4 marks) Calculate a 99% confidence interval for the effect of increasing storage temperature by 4 degrees on average shelf life. Interpret it.

Solution: Effect of increasing storage temperature by 4 degrees on average shelf life is $4\beta_1$. So, we should find a 99% confidence interval for $4\beta_1$. A confidence interval for $4\beta_1$ is given by

Estimate ± (Critical value) S.E.(Estimate) = $4\hat{\beta}_1 \pm t^* S.E.(4\hat{\beta}_1) = 4\hat{\beta}_1 \pm t^* \left(4S.E.(\hat{\beta}_1)\right)$

We can also find the confidence interval for β_1, and then multiply it by 4.

$1-\alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow \alpha/2 = 0.005$, $df = n-2 = 28-2 = 26$. **(0.5 marks)**
So, $t^* = 2.779$ **(0.5 marks)**

A 99% confidence interval for β_1 is given by

$\hat{\beta}_1 \pm t^* S.E.(\hat{\beta}_1) = -9.629 \pm 2.779(1.540) \Rightarrow (-13.90866, -5.34934)$ **(1.5 marks)**

(One mark for constructing confidence interval and one mark for correct final answer)

A 99% confidence interval for $4\beta_1$ is (4(-13.90866) , 4(-5.34934)) = (-55.63 , -21.40)
(0.5 mark)

Interpretation: It is estimated with 99% confidence that increasing storage temperature by 4 degrees will decrease the average shelf life between 21.40 days and 55.63 days. **(1 mark)**