Name: ______________________
Student No.: ______________________

University of Alberta

Department of Mathematical and Statistical Sciences

Statistics 252 - Midterm Examination Version A (Solution)

Date: March 4, 2010

Instructor: Alireza Simchi

Time: 11:00-12:20

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 first three parts 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 fourth and fifth parts 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	4	
Question 17	6	
	25	

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

A researcher was interested in comparing the resting heart rate of people who exercise regularly and people who do not exercise regularly. Independent simple random samples of 16 people ages 30-40 who do not exercise regularly and 12 people ages 30-40 who do exercise regularly were selected and the resting heart rate of each person was measured. The summary statistics are as follows.

Group Statistics

	Exercise	N	Mean	Std. Deviation	Std. Error Mean
Rate	YES	12	69.6667	2.18812	.63166
	NO	16	72.2500	2.23607	.55902

Is there sufficient evidence to indicate the mean resting heart rate of people who do not exercise regularly is greater than the mean resting heart rate of people who exercise regularly? Questions **1** to **3** are related to this question.

1. What is the standard error of the estimate of $\mu_1 - \mu_2$, where μ_1 is the mean resting heart rate of people who do not exercise regularly and μ_2 is the mean resting heart rate of people who exercise regularly?

a) 0.05 **b)** 0.85 **c)** 1.29 **d)** 1.62 **e)** 2.73

2. What is the absolute value of the test statistic?

a) 0.05 **b)** 1.05 **c)** 2.05 **d)** 3.05 **e)** 4.05

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

a) **less than 0.005**.
b) between 0.005 and 0.01.
c) between 0.01 and 0.05.
d) between 0.05 and 0.10.
e) greater than 0.10

4. In determining a 95% confidence interval for the mean difference for two groups, what is the lower limit of the confidence interval $\mu_1 - \mu_2$, where μ_1 is the mean resting heart rate of people who do not exercise regularly and μ_2 is the mean resting heart rate of people who exercise regularly?

a) 0.84 **b)** 1.53 **c)** 2.08 **d)** 3.66 **e)** 4.32

PART 2

An article reported the results of a planned experiment contrasting five different teaching methods. Forty-five students were randomly allocated, nine to each method. After completing the experimental course, a 1-hour examination was administered. The table below summarizes the scores on a 10-minute retention test that was given 6 weeks later.

Group	Teaching Method	Logo	n	Average	S.D.
1	Lecture	L	9	31.5556	1.6667
2	Programmed text	P	9	29.6667	1.8028
3	Programmed text with lecture	P + L	9	27.4444	1.3333
4	Computer instruction	C	9	32.5556	1.3333
5	Computer instruction with lecture	C + L	9	34.2222	1.9861

The following table is the ANOVA output.

Score	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	247.422	?	?	?	?
Within Groups	?	?	2.706		
Total	?	?			

Suppose we would like to carry out a test to determine if there are any significant differences among the five groups. Questions **5** and **6** are related to this test.

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

a) F(4,40) **b)** F(4, 41) **c)** F(4, 45) **d)** F(5,40) **e)** F(5, 45)

6. In the test for any mean differences, what is the residual sum of squares for reduced model?

a) 108.24 **b)** 110.95 **c)** 355.66 **d)** 358.37 **e)** 369.19

Consider the following two contrasts:

i. The contrast that compares the methods using programmed text as part of the method (groups 2 and 3) with those that do not use programmed text (groups 1, 4 and 5).

ii. The contrast that compares the effect of lecture using programmed text with the effect of lecture using computer instruction.

Questions **7** to **11**are related to this contrast.

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

a) 1.2 **b)** 2.2 **c)** 3.2 **d)** 4.2 **e)** 5.2

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

a) 0.10 **b)** 0.25 **c)** 0.50 **d)** 0.75 **e)** 1.25

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

a) 0.28 **b)** 0.56 **c)** 1.2 **d)** 1.9 **e)** 3.8

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

a) 0.39 **b)** 0.56 **c)** 0.77 **d)** 1.1 **e)** 1.9

11. Suppose we would like to use the Bonferroni method to construct two simultaneous 92% confidence intervals for the above two contrasts. What is (approximately) the critical value for 92% family-wise confidence intervals?

a) 1.7 **b)** 2.0 **c)** 2.1 **d)** 2.4 **e)** 2.7

PART 3

The director of admissions of a small college administered a newly designed entrance test to 20 students selected at random from the new freshman class in a study to determine whether a student's grade point average (GPA) at the end of the freshman year can be predicted from the entrance test score. Using SPSS output in the following, answer questions **12** to **15** based on the model $\mu_Y = \beta_0 + \beta_1 X$, where Y is a student's GPA at the end of the freshman year and X is the entrance test score. (**Hint:** The data for this question are simulated, and you may find unrealistic conclusion. For example, the range of X is from 0 to 6.8. Just answer questions by using provided information.)

Descriptive Statistics

	N	Mean	Std. Deviation
Y	20	2.500	.7196
X	20	5.000	.6928

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.434	1	6.434	33.998	.000[a]
	Residual	3.406	18	.189		
	Total	9.840	19			

a. Predictors: (Constant), X

b. Dependent Variable: Y

12. What is the linear correlation between Y and X?

a) 0.52 **b)** 0.65 **c)** 0.77 **d)** 0.81 **e)** 0.96

13. What is the value of the t-test statistic for testing $H_0 : \beta_1 = 0$?

a) 1.8 **b)** 2.8 **c)** 3.8 **d)** 4.8 **e)** 5.8

14. What is the standard error of the estimate of the slope of regression line, if $\hat{\beta}_1 = 0.84$?

a) 0.004 **b)** 0.04 **c)** 0.14 **d)** 1.14 **e)** 2.14

15. Mary Jones obtained a score of 4.7 on the entrance test. Suppose you would like to predict her freshman GPA using a 95% prediction interval. What is the margin of error for the prediction interval ?

a) 0.22 **b)** 0.94 **c)** 1.22 **d)** 1.94 **e)** 2.73

Name: ______________________________
Student No.:____________________________

PART 4

16. (4 Marks in Total) 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.

Define a linear combination (a contrast) in terms of the 6 group means for the following questions. You do not need to find estimates or standard error of estimates.

(a) (2 marks) 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? What is the contrast for the test?

Solution:

The effect of supplement Z when supplement X is not present is $\mu_4 - \mu_1$.

The effect of supplement Z when supplement X is present is $\mu_6 - \mu_2$

Therefore, the contrast is:

$$\gamma = \frac{(\mu_4 - \mu_1) + (\mu_6 - \mu_2)}{2} = -0.5\mu_1 - 0.5\mu_2 + 0.5\mu_4 + 0.5\mu_6$$

(b) (2 marks) 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? What is the contrast for this test?

Solution:

The effect of supplement Z when supplement X is not present is $\mu_4 - \mu_1$.

The effect of supplement Z when supplement X is present is $\mu_6 - \mu_2$

So, the effect of supplement Z does not depend whether or not supplement X is present, if

$\mu_4 - \mu_1 = \mu_6 - \mu_2 \Rightarrow \mu_4 - \mu_1 - \mu_6 + \mu_2 = 0$.

Therefore, the contrast is $\gamma = -\mu_1 + \mu_2 + \mu_4 - \mu_6$.

PART 5

17. (6 Marks in Total) Consider I = 8 means problem. Three models are to be considered; one-mean model, a J-means model, and the 8-mean model. Also 15 observations were collected for each group. Use the given information in the below to help answer the questions:

ANOVA table for 8 means model vs. one mean model

Source of Variation	d.f.	Sum of Squares	Mean-Squares	F-Statistics	P-value
Between (Extra)					
Within (Full)	**112**	**336**	3		
Total (Reduced)		1420			

ANOVA table for J-means model vs. one mean model

Source of Variation	d.f.	Sum of Squares	Mean-Squares	F-Statistics	P-value
Between (Extra)		**980**	245		
Within (Full)	**115**	440			
Total (Reduced)		**1420**			

a) **(2 marks)** What is J?

Solution:

For J-means model, we have:

$$SS(Total) = SS(Between) + SS(Within) \Rightarrow 1420 = SS(Between) + 440$$

(1 mark)

$$\Rightarrow SS(Between) = 1420 - 440 = 980$$

In addition,

$$MS(Between) = \frac{SS(Between)}{J-1} \Rightarrow 245 = \frac{980}{J-1} \Rightarrow J-1 = \frac{980}{245} = 4 \Rightarrow J = 5$$ **(1 mark)**

b) **(4 marks)** Calculate the test statistic and p-value for the test to determine if the 8 mean model fits significantly better than J-mean model. What is your conclusion?

H_0 : 5 - means model is good enough (Reduced model)
H_1 : 8 - means model fits bettert (Full model)

For 8-means model, we have:

$$MS(Within) = \frac{SS(Within)}{n-8} \Rightarrow 3 = \frac{S(Within)}{15(8)-8} \Rightarrow SS(Within) = 3(112) = 336$$

Hence, we have

$SS_{\text{Residual}}(8-means) = 336$, **(0.5 marks)** $SS_{\text{Residual}}(5-means) = 440$ **(0.5 marks)**

, $df_{\text{Residual}}(8-means) = 112$, $df_{\text{Residual}}(8-means) = 115$ **(0.5 marks)**

Extra SS = $SS_{\text{Residual}}(5-means) - SS_{\text{Residual}}(8-mdfeans) = 440 - 336 = 104$
Extra df = $df_{\text{Residual}}(5-means) - df_{\text{Residual}}(8-mdfeans) = 10 - 7 = 3$

Name: ______________________________
Student No.:______________________________

The value of the test-statistic is: $TS = \dfrac{104/3}{336/112} = 11.55$ **(0.5 marks)**

Test statistic has an F-distribution with degrees of freedom of 3 and 7, if H_0 is true. **(0.5 marks)**

α	**0.10**	**0.05**	**0.01**
F^*	**2.139**	**2.696**	**3.984**

0.5 marks for table values.

p-value = P(TS > 11.55) < 0.01. **(0.5 marks)**

Conclusion: with p-value less than 0.01, there is strong evidence to conclude that the 8-means model fits significantly better than 5-mean model. **(0.5 marks)**