Name: ______________________
Student No.: ______________________

University of Alberta

Department of Mathematical and Statistical Sciences

Statistics 252 - Midterm Examination Version A (Solution)

Date: October 28, 2009

Instructor: Alireza Simchi

Time: 12:00-12:50

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 **four** parts. In the first two parts there are 10 multiple-choice questions. For each multiple-choice question choose the answer that is closest to being correct. Circle one of the letters (a)-(e) corresponding to your chosen answer for each question in the table provided on second page. 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 third and forth 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 **23** points.

Name: ______________________

Signature: ______________________

Component	Worth	Mark
Multiple-choice	10	
Question 11	5	
Question 12	8	
	23	

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**

Name: ______________________________
Student No.:______________________________

PART 1

A psychological experiment was conducted to compare the lengths of response time (in seconds) for two different stimuli. To compare natural person-to-person variability in the responses, both stimuli were applied to each of nine subjects, thus permitting an analysis of the difference between stimuli within each response. Summary statistics foe both stimuli and their difference (Diff = Stimulus I – Stimulus II) were given in the following. Assume all the required assumptions are satisfied.

Descriptive Statistics

	N	Mean	Std. Deviation
StimulusI	9	7.6556	2.30332
StimulusII	9	8.4000	3.08180
Diff	9	-.7444	1.33146
Valid N (listwise)	9		

Is there sufficient evidence to indicate a difference in mean response for the two stimuli? Questions **1** to **3** are related to this question.

Solution: This is a paired design. So, the value of the test statistic is:

$$TS = \frac{\overline{D} - 0}{S_D / \sqrt{n}} = \frac{-0.7444 - 0}{1.33146/\sqrt{9}} = -1.68 .$$

Test statistic has a t-distribution with $df = n - 1 = 9 - 1 = 8$, if H_0 is true. Hence

$p - value = 2P(TS > 1.68) \Rightarrow 2(0.05) < p - value < 2(0.1) \Rightarrow 0.10 < p - value < 0.20$

1. What is the distribution of the test statistic under the null hypothesis?

a) t(8) **b)** t(9) **c)** t(16) **d)** t(17) **e)** t(18)

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

a) 0.24 **b)** 0.58 **c)** 1.35 **d) 1.68** **e)** 2.52

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

- **a)** less than 0.05.
- **b)** between 0.05 and 0.1.
- **c) between 0.1 and 0.2.**
- **d)** between 0.2 and 0.50.
- **e)** greater than 0.50

4. In determining a 95% confidence interval for the mean difference for two groups, what is the upper limit of the confidence interval $\mu_1 - \mu_2$, where μ_1 is the mean of the lengths of response time for stimulus 1 and μ_2 is the mean of lengths of response time for stimulus 2?

Solution: $1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha / 2 = 0.025$, $df = n - 1 = 9 - 1 = 8$. So, $t^* = 2.306$ and the upper limit of the confidence interval of $\mu_1 - \mu_2$ is $\overline{D} + t^* \frac{S_D}{\sqrt{n}} = -0.7444 + 2.306 \frac{1.33146}{\sqrt{9}} = 0.28$

a) 0.28 **b)** 1.43 **c)** 1.77 **d)** 2.00 **e)** 3.46

Name: ______________________________
Student No.:____________________________

PART 2

To assess the relative of four different gasoline blends, researchers conducted an experiment using 24 automobiles of the same type, model, and engine size with six randomly assigned to one of the blends and the mileage per gallon was recorded for each car. Suppose all required assumptions are satisfied. The summary statistics and ANOVA for data are given in the next page. You may use them to answer questions in this part.

Suppose we would like to carry out a test to determine if there is any real difference among average mileage for the four different blends? Questions **5** to **7** are related to this test.

Solution: $df(\text{Between}) = I - 1 = 4 - 1 = 3$, $df(\text{Within}) = n - I = 24 - 4 = 20$. So,

$$MS(\text{Between}) = \frac{SS(\text{Between})}{I-1} = \frac{830.085}{3} = 276.695,$$

$$MS(\text{Within}) = \frac{SS(\text{Within})}{n-I} = \frac{344.272}{20} = 17.2136 \Rightarrow \hat{\sigma} = \sqrt{MS(\text{Within})} = \sqrt{17.2136} = 4.15$$

Hence, the value of the test statistic is:

$$TS = \frac{MS(\text{Between})}{MS(Within)} = \frac{276.695}{17.2136} = 16.1$$

Test statistic has an F-distribution with $df_1 = 3$ and $df_2 = 20$, if H_0 is true.

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

a) F(2,20) **b) F(3, 20)** **c)** F(4, 20) **d)** F(2,21) **e)** F(3, 21)

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

a) 2.61 **b)** 4.05 **c) 4.15** **d)** 16.39 **e)** 17.21

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

a) 12.0 **b) 16.1** **c)** 16.9 **d)** 24.1 **e)** 25.3

Descriptive

Mileage	N	Mean	Std. Deviation	Std. Error
1: Control	6	25.3167	3.78281	1.54433
2: Control + supplement x	6	30.3833	3.78070	1.54346
3: Control + supplement y	6	31.7000	4.59565	1.87617
4: Control + both x and y	6	41.5500	4.37390	1.78564
Total	24	32.2375	7.14556	1.45858

Name: ______________________________
Student No.:____________________________

ANOVA

Mileage

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	830.085				
Within Groups	344.272				
Total					

Consider three contrasts in terms of the 4 group means defined earlier to compare:

i. the overall average mileage effect of supplement X
ii. the overall average mileage effect of supplement Y
iii. the overall average mileage effect of supplement x versus the overall average mileage effect of supplement y.

Suppose we would like to use the Bonferroni method to construct three simultaneous 94% confidence intervals for the above three contrasts. Questions **8** and **10** are related to this method.

Solution: $m = 3, 1 - \alpha_E = 0.94 \Rightarrow \alpha_E = 0.06 \Rightarrow \frac{\alpha_I}{2} = \frac{\alpha_E}{2m} = \frac{0.06}{2(3)} = 0.01, \quad df = df(Within) = 20$.

So, $t^* = 2.528$

Effect of X: $\gamma_1 = \frac{(\mu_2 - \mu_1) + (\mu_4 - \mu_3)}{2} = -0.5\mu_1 + 0.5\mu_2 - 0.5\mu_3 + 0.5\mu_4$

Effect of X: $\gamma_2 = \frac{(\mu_3 - \mu_1) + (\mu_4 - \mu_2)}{2} = -0.5\mu_1 - 0.5\mu_2 + 0.5\mu_3 + 0.5\mu_4$

γ_3= Effect of X versus effect of Y = Effect of X – Effect y = $\gamma_1 - \gamma_2$

$$= (-0.5\mu_1 + 0.5\mu_2 - 0.5\mu_3 + 0.5\mu_4) - (-0.5\mu_1 - 0.5\mu_2 + 0.5\mu_3 + 0.5\mu_4) = \mu_2 - \mu_3$$

$$\hat{\gamma}_3 = \bar{y}_2 - \bar{y}_3 = 30.3833 - 31.7000 = -1.3167$$

$$\hat{\gamma}_2 = -0.5\bar{y}_1 - 0.5\bar{y}_2 + 0.5\bar{y}_3 + 0.5\bar{y}_3$$

$$\Rightarrow S.E.(\hat{\gamma}_2) = \sqrt{MS(Within)}\sqrt{\frac{(-0.5)^2}{n_1} + \frac{(-0.5)^2}{n_2} + \frac{(0.5)^2}{n_3} + \frac{(0.5)^2}{n_4}}$$

$$\Rightarrow S.E.(\hat{\gamma}_2) = \sqrt{17.2136}\sqrt{\frac{1/4}{6} + \frac{1/4}{6} + \frac{1/4}{6} + \frac{1/4}{6}} = 1.7$$

8. What is (approximately) the critical value for 94% family-wise confidence intervals?

a) 1.7 **b)** 2.1 **c)** 2.2 **d) 2.5** **e)** 2.8

9. What is (approximately) the estimate of the third contrast?

a) -1.32 **b)** 1.32 **c)** -2.62 **d)** 2.62 **e)** 7.45

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

a) 0.57 **b) 1.7** **c)** 2.4 **d)** 3.4 **e)** 4.8

PART 3

11. (5 marks available) A study was conducted to look at the effects of high school type (public or private) and gender (male or female) on average SAT scores (for the math portion of the SAT's) for students applying to a particular math department. For each high school type 13 observations were collected for males and 13 for females for a total of 26 observations per high school type and 52 observations total.

The four group means are defined as follows:

μ_1 – average SAT score for males that attended a public high school
μ_2 – average SAT score for males that attended a private high school
μ_3 – average SAT score for females that attended a private high school
μ_4 – average SAT score for females that attended a public high school

Consider the three possible models to describe the average response:

1. The one mean model (all means are equal, there are no gender or high school effects)
2. A two mean model with only a high school effect (there are no gender effects)
3. The four mean model (all means are potentially different)

Here is the ANOVA table for the 4-mean vs. the 1-mean model:

Source of Variation	Sum of Squares	d.f.	Mean Square	F-Statistic	p-value
Between					
Within			2400		
Total	497500				

Here is the ANOVA table for the 2-mean vs. the 1-mean model:

Source of Variation	Sum of Squares	d.f.	Mean Square	F-Statistic	p-value
Between	120000				
Within		50			
Total					

Are there any gender effects? Fill in the blanks below for the test for any gender effects. Cleary indicate the models in the null and alternative hypothesis in terms of the parameters defined above. Calculate the test-statistic and clearly indicate how it was calculated. State the distribution of the test statistic under the null hypothesis.

Solution:

$$\begin{cases} H_0 : \mu_1 = \mu_4 \ \ and \ \ \mu_2 = \mu_3 (Two-means-\text{mod}\,el) \\ H_1 : \mu_1 \neq \mu_4 \ \ or \ \ \mu_2 \neq \mu_3 (Four-means-\text{mod}\,el) \end{cases}$$

(One mark for null hypothesis and one mark alternative hypothesis)

$df_{\text{Re}s}(Full) = n - I = 52 - 4 = 48$ **(0.25 marks)**, $SS_{\text{Re}s}(Full) = 48(2400) = 115200$ **(0.5 marks)**

$df_{\text{Re}s}(\text{Re}\,duced) = n - 2 = 52 - 2 = 50$ **(0.25 marks)**

$SS_{\text{Re}s}(\text{Re}\,duced) = SS(Total) - SS(Between) = 497500 - 120000 = 377500$ **(0.5 marks)**

The value of test statistic is:

$$TS = \frac{(377500 - 115200)/(50-48)}{115200/48} = 54.646$$ **(0.5 marks)**

Test statistic has an F-distribution with $df_1 = 2$ and $df_2 = 48$, if H_0 is true. **(1 mark)**

PART 4

12. (8 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:

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) What is the linear correlation between shelf life and storage temperature?

Solution:

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

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)**

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

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) = $4\hat{\beta}_1 \pm t^* S.E.(4\hat{\beta}_1) = 4\hat{\beta}_1 \pm t^* 4S.E.(\hat{\beta}_1)$

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

$1-\alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow \alpha/2 = 0.005$, $df = n-2 = 28-2 = 26$. 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 0.5 marks for correct final answer)

A 99% confidence interval for $4\beta_1$ is (4(-13.90866) , 4(-5.34934)) = (-55.63 , -21.40) **(1 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)**