CONCORDIA UNIVERSITY
Department of Economics

ECON 221/2 SECTION BB
STATISTICAL METHODS I
FALL 2014 – FINAL EXAM
Saturday, December 13, 2:00 pm – 5:00pm

Name: **I.D:** **Section:**

INSTRUCTIONS

- ☐ This is a three-hour exam.
- ☐ This paper is graded out of 120 marks.
- ☐ The examination comprises two sections. Section A contains ten (10) multiple-choice questions worth 2 marks each for a total of 20 marks. Section B contains seven (7) short-answer questions for a total of 100 marks. Candidates should attempt **ALL** questions.
- ☐ Section A: Circle the best answer to each question in Section A. The space next to the choices can be used for calculations but they will not be considered as part of your answer.
- ☐ Section B: Answers to questions in Section B should be written in the spaces provided. You may use the back pages for rough work.
- ☐ Candidates cannot tear pages from the examination paper package and must be returned intact at the end of the examination.
- ☐ Statistical tables are provided.
- ☐ Candidates are allowed to use a non-programmable calculator.
- ☐ Notes or formula crib-sheets are **NOT** allowed. You may use either pen or pencil to provide your answers.
- ☐ Write your name clearly at the top of each page.

FOR EXAMINERS' USE ONLY									
Question	**MC**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**T**
Marks									

SECTION A: TEN (10) MULTIPLE CHOICE QUESTIONS – (TOTAL OF 20 MARKS)

1. Past experience indicates that only 10 percent of new sales staff will last 9 months in the job. If six new staff are hired, what is the probability that at least two of them will still be employed at the end of 9 months?

   A) 0.9841
   B) 0.0984
   C) 0.1143
   D) 0.0159

2. A flock of 18 Michellian sheep contains 7 black ones and 11 white ones. If four Michellian sheep are selected, what is the probability that precisely two are white?

   A) 0.377
   B) 0.522
   C) 0.261
   D) 0.755

3. If $X \sim N(61.7, 27.04)$, what is the value of k such that $P(59 < X < k) = 0.54$?

   A) 65.8
   B) 64.6
   C) 63.7
   D) 66.9

4. Investment A has an expected return of 7.8 percent with a standard deviation of 2 percentage points. Investment B has an expected return of 7.2 percent with a standard deviation of 3.1 percentage points. If the returns on both these investments are normally distributed, which investment is more likely to have a return greater than 10 percent?

   A) Investment A.
   B) Investment B.
   C) Both are equally likely.
   D) Not enough information is provided.

5. Which of the following statements is true regarding the standard error of the mean?

   A) It is less than the population standard deviation.
   B) It decreases as the sample size increases.
   C) It measures the variability of the mean from sample to sample.
   D) All of the above.

6. Banks suffered recently because of high default rates on mortgage and consumer loans. How large a sample is needed to be 98 percent confident that the estimated number of defaults is within \$1000 of the true mean? Assume the population standard deviation is \$10,000.

   A) 567
   B) 233
   C) 240
   D) 543

7. A hypothesis not rejected at the 0.10 level of significance…

   A) Must be rejected at the 0.05 level of significance.
   B) May be rejected at the 0.05 level of significance.
   C) Will not be rejected at the 0.05 level of significance.
   D) Must be rejected at the 0.025 level of significance.

8. The power of a test is the probability of making a(n)…

   A) Correct decision when the null hypothesis is true.
   B) Correct decision when the null hypothesis is false.
   C) Incorrect decision when the null hypothesis is false.
   D) Incorrect decision when the null hypothesis is true.

9. A type II error is committed if we make a(n)…

   A) Correct decision when the null hypothesis is false.
   B) Correct decision when the null hypothesis is true.
   C) Incorrect decision when the null hypothesis is false.
   D) Incorrect decision when the null hypothesis is true.

10. The *p*-value of a test is the…

   A) Smallest α for which the null hypothesis can be rejected.
   B) Largest α for which the null hypothesis can be rejected.
   C) Smallest α for which the null hypothesis cannot be rejected.
   D) Largest α for which the null hypothesis cannot be rejected.

SECTION B: SEVEN (7) QUESTIONS – (TOTAL OF 100 MARKS)

Question 1 (16 marks)
A university medical department states that students must pass certain core courses each year before being able to advance to the next year of the five-year programme. Passing core courses in one year is independent of passing core courses in another year. It is estimated that the proportion of students passing their core courses within the expected time is 90 percent at all year-levels.

a. **(4 marks)** Calculate the probability that a new entrant will finish his studies in five years?

Let x be the number of years a student completes successfully. $\Pr(x=5)=0.9^5=0.5905$

b. **(4 marks)** Calculate the probability that two randomly chosen new entrants will finish their studies in five years?

$(\Pr(x=5))^2=0.5905^2=0.3487$

c. **(4 marks)** Assuming that a student has already finished the first two years of core courses, calculate the probability that he will graduate in the next three years?

$\Pr(x=3)=0.9^3=0.7290$

d. **(4 marks)** If the medical department admits 120 new entrants, how many should graduate in five years?

$n\cdot\Pr(x=5)=120\cdot 0.5905=70.86\Rightarrow 70$ *students should graduate in five years.*

Question 2 (21 marks)
A professor claims that the average score on a recent exam was 83 percent. Assume that the test scores are normally distributed. You believe that the average score is less than 83, therefore, you ask some people in class how they did, and you record the following 8 scores:

82 77 85 76 81 91 70 82

a. **(4 marks)** Calculate the sample mean and the sample standard deviation.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{82+77+85+76+81+91+70+82}{8} = 80.5$$

$$s = \sqrt{\frac{\sum x_i^2 - n\bar{x}^2}{n-1}} = \sqrt{\frac{82^2+77^2+85^2+76^2+81^2+91^2+70^2+82^2-8\cdot 80.5^2}{8-1}} = \sqrt{39.\overline{714825}} = 6.3019$$

b. **(2 marks)** State the appropriate null and alternative hypotheses to test the professor's claim.

$H_0: \mu \geq 83$ *and* $H_1: \mu < 83$

c. **(2 marks)** State the critical value (10 percent level of significance) and decision rule for the test.

If $t_{STAT} = \frac{\bar{x}-\mu}{s/\sqrt{n}} < t_{CRIT,df=8-1} = -1.415$ *then reject the null hypothesis in favour of the alternative.*

d. **(3 marks)** Calculate the appropriate test statistic to test the null hypothesis in part (b).

$$t_{STAT} = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{80.5-83}{6.3019/\sqrt{8}} = -1.122$$

e. **(3 marks)** Estimate and interpret the *p*-value of the test statistic calculated in part (d).

Based on the t-tables, a precise figure cannot be calculated. The only thing that can be determined is that the p-value is greater than 10 percent because the test statistic is greater than the critical value. (A precise figure is $\Pr(t < t_7 = -1.122) = 0.1495$*.)*

f. **(3 marks)** Explain whether the null hypothesis should be rejected and what it means regarding the professor's claim.

Since $t_{STAT} \geq t_{CRIT}$*, do not reject the null. Alternately, since* $p > \alpha$*, do not reject the null. In either case, the implication is that there is strong evidence that your belief is inaccurate.*

g. **(4 marks)** Explain if/how you would have tested the hypothesis differently if you knew that the population standard deviation were $\sigma = 6.1$.

Since the population standard deviation is known, a z-statistic incorporating the population standard deviation should be used instead ($z_{STAT} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{80.5-83}{6.1/\sqrt{8}} = -1.159$*). The critical value would be* $z_{CRIT,\alpha=0.1} = -1.2816$*. Despite the changes, the result remains the same.*

Question 3 (12 marks)
Hapolonians love dining out and carefully follow restaurant reviews. Two Hapolonian food critics rated a sample of nine restaurants on a scale from 0 to 100. We want to examine whether the two Hapolonians feel the same way about each restaurant by examining the differences in their *mean rating*.

Restaurant	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Critic A	99	80	72	71	72	84	75	85	91
Critic B	93	67	58	72	67	84	71	78	94

a. **(2 marks)** State the appropriate null and alternative hypotheses.

Let $d = \mu_A - \mu_B$. $H_0 : d = 0$ *and* $H_0 : d \neq 0$.

b. **(2 marks)** State the critical value (5 percent level of significance) and decision rule for the test.

If $t_{STAT} = \left|\frac{\bar{d} - d}{s_d / \sqrt{n}}\right| > \left|t_{CRIT, df=9-1}\right| = 2.306$ *then reject the null hypothesis in favour of the alternative.*

c. **(6 marks)** Calculate the appropriate test statistic to test the null hypothesis in part (a).

$$t_{STAT} = \frac{\bar{d} - d}{s_d / \sqrt{n}} = \frac{5 - 0}{5.874 / \sqrt{9}} = 2.554$$

d. **(2 marks)** Explain whether the null hypothesis should be rejected and what it means regarding the reviews of the Hapolonian food critics.

Since the t-statistic exceeds the critical value, then reject the null hypothesis in favour of the alternative and conclude that the Hapolonian food critics do not feel the same way about each restaurant.

Question 4 (17 marks)
Tootie birds migrate from Flin Flon, Manitoba every year: some to Edmonton, some to Montréal. Because Tooties look very similar, it is hard to tell the precise proportions that go where, but experts believe that at least 25 percent go to Edmonton. Bird watchers in Flin Flon recently saw that 63 out of a random sample of 300 Tooties were heading for Edmonton.

a. **(2 marks)** State the appropriate null and alternative hypotheses.

$$H_0 : p_E \geq 0.25 \text{ and } H_1 : p_E < 0.25$$

b. **(2 marks)** State the critical value (5 percent level of significance) and decision rule for the test.

If $z_{STAT} = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}} < z_{CRIT,0.05} = -1.645$ *then reject the null hypothesis in favour of the alternative.*

c. **(3 marks)** Calculate the appropriate test statistic to test the null hypothesis in part (a).

$$z_{STAT} = \frac{0.21 - 0.25}{\sqrt{\frac{0.25(1-0.25)}{300}}} = -1.600$$

d. **(2 marks)** Explain whether the null hypothesis should be rejected and whether it supports the experts' claim.

Since the z-statistic is less than the critical value, then do not reject the null hypothesis and conclude that the proportion of Tootie birds that migrate to Edmonton is at least 25 percent.

e. **(3 marks)** Construct a 95 percent confidence interval for the true population proportion.

$$\hat{p} - z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \Rightarrow$$

$$0.21 - 1.96\sqrt{\frac{0.21(1-0.21)}{300}} < p < 0.21 + 1.96\sqrt{\frac{0.21(1-0.21)}{300}} \Rightarrow$$

$$0.1639 < p < 0.2561$$

f. **(3 marks)** Explain whether the interval in part (e) supports your decision in part (d).

The confidence interval contains the hypothesised population proportion. In repeated sampling, this proportion should appear in the confidence interval 95 percent of the time. Therefore, there is insufficient evidence to reject the expert's claim. This supports the decision in part (d).

g. **(2 marks)** How many Tooties need to be observed to ensure that the margin of error is no more than three percentage points.

$$n = \frac{0.25 z_{0.025}^2}{ME^2} = \frac{0.25 \cdot 1.96^2}{0.03^2} = 1067.\dot{1} \Rightarrow 1068$$

Question 5 (5 marks)

In Excel, explain how you would illustrate the proposition that the sample mean, $\bar{Z}$, of a random sample of size 10, drawn from a standard normal distribution, satisfies $\Pr\left(\left|\bar{Z}\right| > \frac{1.96}{\sqrt{10}}\right) = 0.95$.

Generate an independent draw from N(0, 1) in each of the first 10 columns of the first row. In column 11, calculate the absolute value of the mean of the 10 columns. Repeat this over a "large" number of rows (say, 100,000). Count the number of times the value in the final column is bigger than $\frac{1.96}{\sqrt{10}}$. *The number should be close to 5,000 (ie, 5 percent of the "large" number).*

Question 6 (16 marks)

Mean monthly profits at McDonald's restaurants are normally distributed with a population standard deviation of $10 thousand. In 2013, mean monthly profits were $150 thousand. A sample of 25 McDonald's restaurants for the first half of this year produced mean monthly profits of $148 thousand. An investor wonders if mean monthly profits have fallen below $150 thousand.

a. **(2 marks)** State the appropriate null and alternative hypotheses.

$H_0 : \mu \geq 150$ *and* $H_1 : \mu < 150$

b. **(2 marks)** State the critical value (10 percent level of significance) and decision rule for the test.

If $z_{STAT} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z_{CRIT,0.05} = -1.28$ *(alternately,* $\bar{x} < \mu - z_{CRIT,0.05} \cdot \frac{\sigma}{\sqrt{n}} = 147.44$*), then reject the null hypothesis in favour of the alternative.*

c. **(2 marks)** Calculate the appropriate test statistic to test the null hypothesis in part (a).

$$z_{STAT} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{148 - 150}{10/\sqrt{25}} = -1.00$$

d. **(2 marks)** Explain whether the null hypothesis should be rejected and whether it confirms the investor's fear.

Since the z-statistic is less than the critical value, then do not reject the null hypothesis and conclude that the investor's fear is unjustified.

e. **(4 marks)** What level of significance will generate a conclusion opposite to the one in part (d).

$p(z < -1) = 0.1587$

f. **(4 marks)** If this year's mean monthly profits have fallen to $145 thousand, calculate the probability of rejecting the (false) null hypothesis in part (a).

$$\Pr\left(\bar{x} < 147.55 \middle| \mu = 145\right) = \Pr\left(z < \frac{147.44 - 145}{10/\sqrt{25}} = 1.22\right) = 0.8888$$

Question 7 (13 marks)
The joint probability distribution of variables *X* and *Y* is shown in the table below, where *X* is the number of tennis racquets and *Y* is the number of golf clubs sold daily in a small sports store.

	Y = 1	*Y = 2*
X = 1	*0.42*	*0.24*
X = 2	*0.19*	*0.15*

a. **(2 marks)** Calculate the marginal probability distributions of *X* and *Y*.

$$\Pr(X=x)=p(x,1)+p(x,2) \ \ \textit{and} \ \ \Pr(Y=y)=p(1,y)+p(2,y)$$

X	*X = 1*	*X = 2*	*Y*	*Y = 1*	*Y = 2*
$p(X=x)$	0.66	0.34	$p(Y=y)$	0.61	0.39

b. **(3 marks)** Calculate $P(Y|X=2)$ (ie, the conditional probability distribution of *Y* given *X* = 2).

$$\Pr(Y=1|X=2)=\frac{\Pr(Y=1\cap X=2)}{\Pr(X=2)}=\frac{0.19}{0.34}=0.5588,$$

$$\Pr(Y=2|X=2)=\frac{\Pr(Y=2\cap X=2)}{\Pr(X=2)}=\frac{0.15}{0.34}=0.4412$$

c. **(4 marks)** Explain, using probabilities, whether the number of tennis racquets and golf clubs sold are statistically independent.

Close, but not quite. If racquets and clubs were independent variables, then $\Pr(x,y)=\Pr(x)\cdot\Pr(y)$ *and the matrix would look like:*

	Y = 1	*Y = 2*
X = 1	$0.66\cdot 0.61=0.4026$	$0.66\cdot 0.39=0.2574$
X = 2	$0.34\cdot 0.61=0.2074$	$0.34\cdot 0.39=0.1326$

d. **(4 marks)** Calculate and interpret the correlation coefficient of *X* and *Y*.

$$\overline{x}=0.66\cdot 1+0.34\cdot 2=1.34, \ \overline{y}=0.61\cdot 1+0.39\cdot 2=1.39$$
$$\mathrm{var}(x)=0.66\cdot 1^2+0.34\cdot 2^2-1.34^2=0.2244, \ \mathrm{var}(y)=0.61\cdot 1^2+0.39\cdot 2^2-1.39^2=0.2379$$
$$\mathrm{cov}(x,y)=0.42\cdot 1\cdot 1+0.24\cdot 1\cdot 2+0.19\cdot 2\cdot 1+0.15\cdot 2\cdot 2-1.34\cdot 1.39=0.0174$$

$$\rho=\frac{\mathrm{cov}(x,y)}{\sqrt{\mathrm{var}(x)\cdot\mathrm{var}(y)}}=\frac{0.0174}{\sqrt{0.2244\cdot 0.2379}}=0.0753$$