ECON 250 — Introductory Statistics — Final Exam **12 December 2016**

Instructions:

- The exam is 3 hours in length. There are 100 points in total. Allocate your time accordingly.
- Put your name and student number on each answer booklet used.
- You may use a hand calculator: either the standard Queen's Casio 991 series, or an equivalent pre-approved hand calculator.
- Formulas and tables are printed at the end of the question papers. Feel free to detach these from the question pages for easier reference.
- This exam consists of 13 printed pages in all: this cover sheet, 5 question pages, 4 formula pages, and 3 statistic table pages. Please ensure you have all questions/sheets!
- This exam is divided into two sections:
  - Section A (page 2, worth 20 marks) consists of 10 very short questions requiring only a small calculation, value lookup, or a few words.
  - Section B (pages 3–6, worth 80 marks) consists of 9 longer questions with multiple parts. Show your work: part marks cannot be awarded for wrong answers without calculations.
- Answer all questions. The value of each question is shown in the exam.
- Proctors are unable to respond to queries about the interpretation of exam questions. Do your best to answer the exam questions as they are written.
-
- Good luck!

Section A: Very short questions [20 points]

The following questions require only a few words, a lookup of a numerical value, or a small calculation. Each question is worth 2 marks.

1. Assume $X \sim \mathcal{U}[5,7]$. Find $P(X = 6)$.

2. Assume Y follows a t distribution with 27 degrees of freedom. Find $P(Y < -2.052 \;\cup\; Y > 2.052)$.

3. Write down an example of two disjoint random events.

4. Write down an example of two independent random events.

5. If σ is known and you increase the size of the sample by 50%, how does the standard deviation of $\overline{x}$ change?

6. If you have a sample containing heights measured in inches and convert it to centimeters (by multiplying all values by 2.54) what will happen to the sample mean and standard deviation?

7. How many different sequences are there of 9 coin flips show heads exactly 5 times out of the 9 flips?

8. Suppose you calculate a 95% confidence interval and a 99% confidence interval from the same sample. Which interval will be larger?

9. Suppose you run a regression with 4 coefficients $(\beta_1, \ldots, \beta_4)$ estimated using $n = 31$ observations. What is the critical t^* value you would use when constructing a 95% confidence interval for β_3?

10. Suppose you estimate a regression of $income = \beta_1 + \beta_2 age + \beta_4 educ + u$ (where $income$ is a person's annual income in dollars, age is the individual's age, and $educ$ is the number of years of education the individual has completed). Suppose one of the estimates is $\widehat{\beta}_3 = 3010$. Give an economic interpretation of this value.

Section B: Longer questions [80 points]

1. **[9]** Sample values.

a) You collect a sample containing the following data:

X	5	8	7	15	2	9	5

For this sample calculate: the mean, median, mode, first quartile and third quartile.

b) You collect the following sample of 4 observations of two different variables, Y and Z:

Y	1	4	2	9
Z	4	1	7	0

You calculate the means of the two variables to be $\overline{y} = 4$ and $\overline{z} = 3$.

Calculate the variance and standard deviation of each variable.

Calculate the covariance and correlation of the two variables.

2. **[7]** Suppose that United States voters are known to vote for Donald Trump with probability 0.45.

a) Suppose 10 US voters are selected at random. Find the probability that "The Donald" would receive three or fewer of the 10 votes.

b) Find the probability that Donald Trump would get at least 50 votes from 100 randomly selected voters.

3. **[9]** Now suppose, in contrast to question **2.**, that the probability of voting for Donald Trump is unknown, but a recent poll of 1742 randomly selected voters had 831 Trump supporters.

a) Test the hypothesis that Trump receives half of all votes against the alternative that Trump receives less than half of the votes at the $\alpha = 0.03$ level.

b) Suppose that the above sample only contained voters from Florida. A separate poll in Pennsylvania of 1280 voters had 621 Trump supporters.

Test whether there is evidence that the proportion of Trump supporters differs in the two states.

4. **[9]** You are interested in studying the effects of individuals' university major on future income. You randomly select individuals from two groups of graduating students and keep track of them for three years.

The first group contains 87 students who obtained a Bachelor's degree with a major in Economics. The mean salary of this group three years after graduating equals \$65,100 with a standard deviation of \$9,000.

The second group has 102 students with Bachelor's degrees with a major in Psychology. Three years after graduation this group has a mean salary of \$61,900 with standard deviation \$15,000.

a) Construct a 95% confidence interval for the difference in mean salaries.

b) Is there evidence at the 95% level that the mean salary is *higher* for Economics graduates than Psychology graduates? Be sure to clearly state the hypotheses and to report the p-value of the test.

5. **[10]** The time to wait for a table at a highly popular local restaurant is uniformly distributed between 10 and 60 minutes (with a mean of 35 minutes and a standard deviation of 14.43 minutes).

a) Find the probability that a person arriving at the restaurant will face: a wait time of at least 20 minutes; a wait time of between 20 and 45 minutes; and a wait time of more than 45 minutes.

b) This customer decides, before arriving, that he will give a tip of \$10 if he waits less than 20 minutes, \$7 if he waits between 20 and 45 minutes, and \$1 if he has to wait more than 45 minutes.

Find the expected value (i.e. the mean) of the tip.

c) If the customer visits the restaurant on 30 different occasions, what is the distribution of the mean time he waits for a table?

6. **[10]** You are interested in testing whether the mean revenue of small businesses in Canada has changed over the past decade. Ten years ago, the mean revenue of small businesses was known to equal 2.50 million dollars per year. The standard deviation of mean revenue of small business is known to equal 2.00 million dollars per year. You conduct a sample of 49 small businesses and calculate a sample mean of 1.89 million dollars.

a) State the null and alternative hypotheses to test whether the mean revenue has changed from its value ten years ago of 2.50 million dollars. Find the p-value of this test. Do you reject the null hypothesis at the $\alpha = .05$ level?

b) For the test in part *a*), calculate the power of the test at the specific alternative $\mu_a = 2.20$ million dollars.

c) What are the probabilities of making a type I and type II error (using the appropriate values given in or calculated for the above parts of the question).

7. **[8]** You conduct a sample of the price of a box of Kollegg's Vegetable Hoops, a popular breakfast cereal, across different stores. Your sample of the prices in 20 randomly selected stores has a mean of 5.78 and a standard deviation of 1.32.

a) Test whether your sample provides evidence at the $\alpha = .05$ level that the mean is different from the producer's ideal retail price of 5.25.

b) Another researcher, working independently, conducted a separate sample of 28 stores which had a mean of 5.99. She reported a 99% confidence interval of $[5.23, 6.75]$. Find the standard deviation of prices in the other researcher's sample.

8. [**10**] An investor is considering buying an investment portfolio consisting of two different stocks and cash. The value of each unit of stock X after one year has mean \$40 and standard deviation \$10. Each unit of stock Y after one year has a mean of \$5 and standard deviation \$1. The monetary value of cash holdings does not change. The investor's portfolio contains 10 units of stock X, 100 units of stock Y, and \$100 of cash.

a) Assuming X and Y are independent, find the mean and standard deviation of the value of the investor's portfolio after one year.

b) X and Y are, in fact, not independent. Rather they have a correlation of 0.5 and covariance of 5.

Find the mean and standard deviation of the value of the portfolio after one year.

c) The investor is considering a second portfolio which is less risky but also has a lower mean value. This portfolio has mean value (after a year) of \$950 and standard deviation of \$100.

Assuming the value of both portfolios are normally distributed, which portfolio is more likely to have a value greater than \$900? Which portfolio is more likely to have a value greater than \$1000?

9. [**8**] Suppose a medical test for diagnosing a rare blood condition that affects 2% of the population returns a positive result 99% of the time for a person *with* the condition and returns a negative result 75% of the time for a person *without* the condition.

a) If a randomly selected person is tested and receives a positive test result, what is the probability that this person actually has the condition?

b) Suppose a person who *does not* have the condition receives a *positive* test result. Is this a Type I or a Type II error?

Suppose a person who *does* have the condition receives a *negative* test result. Is this a Type I or a Type II error?