QUEEN'S UNIVERSITY FINAL EXAMINATION
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS

APRIL 2017

ECONOMICS 250
Introduction to Statistics

Instructor: Gregor Smith

Instructions:

The exam is three hours in length.

Do all ten (10) questions.

Be sure to show your calculations and intermediate steps.

Put your student number on each answer booklet.

Formulas and tables are printed at the end of this question paper.

You may use a hand calculator: the Casio 991.

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.

1. Consider the following data: 1, 3, 3, 5, 7, 11.

(a) Find the sample mean and sample standard deviation.

(b) Find the mode and median. Is the distribution symmetric or skewed?

(c) Find the coefficient of variation (expressed as a percentage).

2. Suppose that each day an exchange rate is uniformly and continuously distributed between 1 and 10. It is independent from day to day.

(a) What is the probability that the exchange rate is greater than 4?

(b) What is the probability it is greater than 4 one day *and* greater than 4 the next day?

(c) What is the probability it is greater than 4 one day *or* greater than 4 the next day?

3. Suppose that we classify Canadian residents as university graduates (G) or non-graduates (NG) and that $P(G) = 0.25$. We also classify them as employed (E) or unemployed/not in the labour force (grouped together and denoted U). Suppose that $P(E) = 0.65$ and that $P(E|G) = 0.75$.

(a) What is the probability that someone is in category U given that they are in category NG?

(b) What is the probability that someone is in category G given that they are in category E?

4. Suppose that the interest rate in Canada (labelled r_c) and the interest rate in the US (labelled r_u) can each take only two values: 1 and 3. They are jointly distributed like this:

		r_u = 1	r_u = 3
r_c	1	0.4	0.1
	3	0.1	0.4

(a) Find the mean and variance of r_c.

(b) Find the correlation between r_c and r_u.

[continued over]

(c) Suppose an investor holds a portfolio with a return given by:

$$r = 0.5r_c + 0.5r_u.$$

Find the mean and variance of r.

5. Suppose that in any month the probability of a job-seeker finding a job is 0.4. This event is independent across job-seekers.

(a) If 10 people are looking for jobs this month then what is the probability that 4 or more find jobs?

(b) If 100 people are looking for jobs this month then what is the probability that 46 or more find jobs?

6. Suppose that a sample of $n = 81$ incomes yield a sample average of 40 and a sample standard deviation of 4.

(a) Find a 90% confidence interval for the population average income.

(b) Suppose that you want to test the null hypothesis that the population average income is $\mu = 41$ against the alternative that it is not equal to 41. Find a range within which the P-value for this test must fall.

7. A sample of 9 emerging economies yields an average inflation rate of 4 with a standard deviation of 1. A sample of 16 developed economies yields an average inflation rate of 2 with a sample standard deviation of 0.5.

(a) Construct the 95% confidence interval for the difference between the population average inflation rate in emerging economies and that in developed economies.

(b) A macroeconomist wishes to test the null hypothesis that average inflation rate is the same in the two types of economies against the alternative that it is higher in emerging economies. Report a range within which the P-value for this test must fall.

8. Suppose that researchers study a sample of 100 people and find that 12 have had influenza in the past year. They wish to test the hypothesis that the population proportion of people who have had influenza is $p = 0.10$ against the alternative that it is greater than that.

(a) Suppose that $\alpha = 0.05$. Find the critical value $\hat{p}_c$ for this test. Is the null hypothesis rejected or not?

(b) Suppose that the researchers are using this critical value but, unbeknownst to them, the true, population proportion is 0.13. Find the power of the test.

9. Suppose that the literacy rate in India is 74% based on a survey of 100 people, while the rate in Bangladesh is 62% based on a survey of 100 people.

(a) Form a 95% confidence interval for the difference between the two rates.

(b) Construct a test statistic to test the null hypothesis that the two rates are the same against the alternative hypothesis that the rate is higher in India, and report the associated P-value.

10. Statisticians study the relationship between an indicator of air pollution (labelled y) and GDP per capita, labelled x, across the 50 US states, indexed by i. They estimate this linear regression equation by ordinary least squares:

$$y_i = a + bx_i + \epsilon_i,$$

and find that the estimate of b, labelled $\hat{b}$, is -0.5 with a t-statistic of 2.4. They also find an R^2 statistic of 0.81. They also find $\hat{a} = 60$.

(a) Is there evidence of a statistical relationship between the two variables?

(b) What is the correlation coefficient between x and y?

(c) Predict the value of y for a state with GDP per capita of 40.

Economics 250
Winter Term 2017
Final Exam Answer Guide

1. (a: 2 marks) The sample mean is 5. The sample variance is $64/5 = 12.8$ so the sample standard deviation is 3.577. To avoid mistakes in finding s, remember to make a chart.

(b: 2 marks) The mode (the most common value) is 3. The median is 4, found by averaging because there is an even numbr of observations. The median is less than the mean, so the distribution is skewed right. (Remember a long right tail means that a distribution is skewed right).

(c: 2 marks) The CV is $(100) \times (3.577/5) = 71.55\%$.

2. (a: 2 marks) The probability is 6/9 or 0.67. Draw a picture of this uniform distribution to avoid mistakes in calculating this value.

(b: 2 marks) From the multiplication rule the probability of both events is the product: 0.444

(c: 2 marks) From the addition rule the probability of either event is: $0.666+0.666-0.444 = 0.888$. (You will get a slightly different answer if depending on how you round in part (a)).

3. (a: 3 marks) This probability is 0.3833. A foolproof way to answer these questions is to complete the two-way table of joint ('and') probabilities, using the information given in the question. Then the multiplication rule (the link between conditional, joint, and marginal probabilities) can be used.

(b: 3 marks) This probability is 0.28846.

4. (a: 3 marks) The mean is 2. The marginal probabilities are 0.5 and 0.5. The variance is 1.

(b: 3 marks) The covariance is 0.6 so that is the correlation too. Making a chart will help you find the covariance. Then the two standard deviations are 1 so the correlation has the same value.

(c: 3 marks) The mean is 2. The variance is $0.25 + 0.25 + 0.3 = 0.8$. (Look up the formula for the variance of a combination of random variables.)

5. (a: 2 marks) From Table C the probability is $1 - 0.3822 = 0.6178$

(b: 3 marks) We find $x \sim N(40, 4.898)$ so to standardize: $(46 - 40)/4.898 = 1.22$. The probability beyond that is 11.12%.

6. (a: 2 marks) The 90% CI with $df = 80$

$$40 \pm 1.664\frac{4}{9} = 40 \pm 0.7385 = (39.261, 40.739).$$

Remember to use t and not z.

(b: 3 marks) The test statistic is:

$$t = \frac{40 - 41}{4/9} = -2.25.$$

That leaves between 0.02 and 0.01 in one tail so the P-value is between 0.02 and 0.04, because this is a two-tailed test.

7. (a: 3 marks) The 95% CI is

$$2 \pm 2.306(0.356) = 2 \pm 0.8209 = (1.179, 2.821).$$

note that we use the minimum of the two df.

(b: 3 marks) The test statistic is $t = 2/0.356 = 5.618$ with $df = 8$ so the P-value is less than 0.0005. (Remember a large test statistic leads to a small P-value, which is evidence against the null hypothesis.)

8. (a: 3 marks) Under the null we have $\hat{p} \sim N(0.10, 0.03)$. The critical value that leaves 5% in the right tail is

$$\hat{p}_c = 0.10 + 1.645(0.03) = 0.14935.$$

Thus the null is not rejected. Notice that we use 0.01 (the hypothesized value) and not 0.12 (the sample value) to find the distribution under the null.

(b: 4 marks) Under the alternative the standard error now is 0.0336. Standardizing the critical value in the alternative distribution (so using 0.13 to construct the standard deviation) gives:

$$z = \frac{0.14935 - 0.13}{0.0336} = 0.5758.$$

I use 0.58 and find power is 1-0.7190 or 0.281 or 28.1%. Drawing the diagram may help you avoid errors.

9. (a: 3 marks) The standard error of the difference is 0.0656 so the margin of error is 1.96 times that or 0.1285. Thus the 95% CI is 0.12 ± 0.1285 or $(-0.0085, 0.2485)$.

(b: 3 marks) Under the null that the proportions are the same the best estimate of the common proportion is the pooled one. The pooled proportion is 0.68 which then gives a standard deviation of the difference of 0.0659. Thus $z = (0.12 - 0)/0.0659 = 1.82$. From Table A the P-value is 0.0344.

10. (a: 2 marks) Yes, there is a relationship, judged by either R^2 or the t-statistic.

(b: 2 marks) $r = -0.9$ (Note the minus sign.)

(c: 2 marks) $\hat{y} = 60 - 0.5(40) = 60 - 20 = 40$.

Total Marks: 62

Omitted Question

7. Suppose that in Kingston 7 out of 50 people surveyed report having had the flu this winter while in Ottawa 20 out of 200 people report having had the flu.

(a) Find a 95% confidence interval for the difference between the Kingston and Ottawa flu rates.

(b) Test the null hypothesis that the two flu rates are equal, against the alternative that they are not equal, using a significance level of 5%.

7. (a) The difference is 0.14-0.10 or 0.04. The margin of error is 1.96 (0.053) or 0.104 so the 95% confidence interval is (-0.64, 0.144).

(b) We used the pooled proportion, 0.108, to find the standard deviation, which is 0.0491. So our test statistic is

$$z = \frac{0.04 - 0}{0.0491} = 0.81466.$$

The two-sided critical value is 1.96 so we do not reject the null that the rates are equal.