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

APRIL 2018

ECONOMICS 250
Introduction to Statistics

Instructor: Gregor Smith

Instructions:

The exam is three hours in length.

Do all nine (9) 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. Suppose that a data set consists of 10 observations of the value 1, 12 observations of the value 3, and 10 observations of the value 5.

(a) Find the sample mean, median, and mode.

(b) Find the sample standard deviation.

(c) Find the interquartile range.

2. Suppose that we classify Canadian residents between 15 and 65 as either young (Y, ages 15–24) or old (O, ages 25–65). We also classify them according to whether they are employed (E), unemployed (U), or not in the labour force (N). Suppose $P(Y) = 0.20$. Also, $P(E|Y) = 0.25$, $P(U|Y) = 0.10$, $P(E|O) = 0.7$, and $P(U|O) = 0.2$.

(a) What is the probability that someone between 15 and 65 is not in the labour force?

(b) What is the probability that someone is young given that they are unemployed?

(c) What is the probability that someone is old given that they are not in the labour force?

3. Suppose that two random variables (labelled x and y) can each take only two values: 1 and 3. They are jointly distributed like this:

		x	
		1	3
y	1	0.3	0.2
	3	0.2	0.3

(a) Find the mean and variance of x.

(b) Find the correlation between x and y.

(c) Suppose that a third random variable is given by:

$$w = \frac{x + y}{2}.$$

Find the mean and variance of w.

(d) Find the conditional mean of y given $x = 1$.

4. Suppose that in any month the probability of finding a job is 0.20. This event is independent across job-seekers.

(a) If 10 people are looking for jobs this month then what is the probability that 2 or more find jobs?

(b) If 100 people are looking for jobs this month then what is the probability that 20 or more find jobs?

5. Suppose that a sample of $n = 16$ incomes yields a sample average of 38 and a sample standard deviation of 3.

(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 $\mu = 40$ against the alternative that it is not equal to 40. Find a range for the P-value for this test.

6. A sample of 9 emerging economies yields an average inflation rate of 4 with a sample standard deviation of 1. A sample of 16 developed economies yields an average inflation rate of 3 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.

7. Suppose that researchers study a sample of 50 people and find that 6 say they will vote Green in the next election.

(a) Find a 95% confidence interval for the population proportion who will vote Green.

(b) What would the confidence interval be if the researchers used the Wilson value $\tilde{p}$ instead?

(c) Suppose that an investigator tests the null hypothesis that the population proportion is 0.10 against the alternative that it is greater than that. If $\alpha = 0.10$ then find the critical value $\hat{p}_c$. Using $\hat{p}$ as the sample estimate, would the investigator reject the null?

(d) Suppose that researchers are using this critical value but, unbeknownst to them, the true, population proportion is 0.12. Find the power of the test.

8. Suppose that the sample unemployment rate in Ontario is 8% based on a survey of 200 people, while the rate in Manitoba is 6% based on a survey of 100 people.

(a) Form a 90% confidence interval for the difference between the two population unemployment rates.

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

9. Statisticians study the relationship between an indicator of wages (labelled y) and years of education after high school, labelled x, across a sample of 41 workers 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 2.423. They also find an R^2 statistic of 0.70. They also find $\hat{a} = 20$.

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

(b) If you were testing the null hypothesis that $b = 0$ against the alternative that $b \neq 0$ what would the P-value be?

(c) What is the economic interpretation of $\hat{a}$?

Economics 250
Winter Term 2018
Final Exam Answer Guide

1. (a: 2 marks) The sample mean is 3. The median and mode also are 3.

(b: 2 marks) The sample variance is $s^2 = 80/31 = 2.58$ so the sample standard deviation is $s = 1.61$.

(c: 2 marks) Its clear that $Q_1 = 1$ and $Q_2 = 5$ so the IQR is [1,5] with a width of 4.

2. (a: 2 marks) $P(N) = 0.21$.

(b: 2 marks) $P(Y|U) = P(Y \cap U)/P(U) = 0.02/0.18 = 0.111$.

(c: 2 marks) $P(O|N) = P(O \cap N)/P(N) = 0.381$.

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

(b: 3 marks) A chart shows the covariance is 0.2, so the correlation is also 0.2 because both standard deviations are 1.

(c: 3 marks) We can list the 4 outcomes for w and their probabilities, or simply use the formulas for a linear combination. Those give the mean as 2 and the variance as 0.6.

(d: 2 marks) The conditional probabilities are 0.6 and 0.4 so the conditional mean is 1.8.

4. (a: 2 marks) From Table C the probability of 2 or more successes is 1- 0.2684-0.1074 = 0.6242.

(b: 2 marks) Now $X \sim N(20, 4)$ so the probability is 0.5.

5. (a: 2 marks) The 90% CI is:

$$38 \pm 1.753\frac{3}{4} = 38 \pm 1.31475 = (36.685, 39.315)$$

using the appropriate t-statistic.

(b: 2 marks) Our test statistic is:

$$t = \frac{38 - 40}{3/4} = -2.67$$

In Table D the tail area is between 0.01 and 0.005 so multiplying by 2 gives a P-value between 0.02 and 0.01.

6. (a: 2 marks) Using 8 df the CI is:

$$1 \pm 2.306(0.356) = 1 \pm 0.821 = (0.179, 1.821).$$

(b: 2 marks) Our t-statistic is:

$$\frac{1 - 0}{0.356} = 2.81,$$

so from the relevant row of Table D the P-value is between 0.02 and 0.01.

7. (a: 2 marks) The 95% CI is

$$0.12 \pm 1.96(0.045956) = 0.12 \pm 0.09 = (0.03, 0.21)$$

(b: 2 marks) The Wilson value is $\tilde{p} = 8/54 = 0.148$ (with some rounding) so the CI would be

$$0.148 \pm 1.96(0.0483) = 0.148 \pm 0.09467 = (0.05333, 0.24267)$$

(c: 3 marks) With 10% in one tail, $z = 1.282$ (many students began from an incorrect value here). The critical value solves:

$$1.282 = \frac{\hat{p}_c - 0.10}{\sqrt{0.10(0.90)/50}}$$

(where notice we use 0.10 in the denominator) which yields $\hat{p}_c = 0.154$. Since $\hat{p} < \hat{p}_c$ the null is not rejected.

(d: 3 marks) We standardize the critical value in the alternative distribution:

$$z = \frac{0.154 - 0.12}{\sqrt{0.12(0.88)/50}} = 0.034/0.045956 = 0.7398.$$

Using 0.74 in Table A gives power as 0.2296.

8. (a: 2 marks) The 90% CI is:

$$0.02 \pm 1.645(0.0305) = 0.02 \pm 0.050 = (-0.03, 0.07)$$

(b: 3 marks) The test statistic involves the pooled estimate which is 0.0735. Using that with each sample size gives a standard deviation for the difference of 0.03196, so the test statistic is

$$z = \frac{0.02}{0.03196} = 0.625$$

Averaging the two nearest values in Table A yields a P-value of 0.266.

9. (a: 2 marks) Yes, the R^2 statistic is quite large.

(b: 2 marks) From Table D with $df = 40$ the P-value would be $2 \times 0.01 = 0.02$.

(c: 2 marks) The intercept is an estimate of the wage for those who have just graduated from high school.

Total Marks: 55