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

APRIL 2019

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 5 observations of the value 2, 6 observations of the value 4, and 5 observations of the value 6.

(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 college-or-university graduates (G) or not (NG). We also classify them according to whether they are employed (E), unemployed (U), or not in the labour force (N). Suppose $P(G) = 0.55$. Also, $P(E|G) = 0.74$, $P(U|NG) = 0.08$, $P(E|NG) = 0.6$, and $P(U|G) = 0.06$.

(a) What is the probability that someone is a graduate and not in the labour force?

(b) What is the probability that someone is not a graduate given that they are unemployed?

(c) What is the probability that someone is a graduate 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: 10 and 30. They are jointly distributed like this:

		x	
		10	30
y	10	0.1	0.4
	30	0.4	0.1

(a) Find the mean and standard deviation of x.

(b) Find the correlation between x and y.

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

$$w = x - y.$$

Find the mean and standard deviation of w.

(d) Find the conditional mean of y given $x = 10$.

4. Suppose that a series of independent cancer research projects each has a 5% probability of success.

(a) If there are 15 projects then what is the probability of at least one success? What is the probability of three or more successes?

(b) If there are 150 projects then what is the probability of at least one success? What is the probability of three or more successes?

5. Suppose that a sample of $n = 16$ incomes yields a sample average of 39.659 and a sample standard deviation of 4.

(a) Find a 95% 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 the P-value for this test.

6. A sample of 9 urban residents yields an average lifespan of 80 with a sample standard deviation of 1. A sample of 10 rural residents yields an average lifespan of 79 with a sample standard deviation of 1.2.

(a) Construct the 90% confidence interval for the difference between the population average lifespan in the two areas.

(b) A statistician wishes to test the null hypothesis that average lifespan rate is the same in the areas against the alternative that it is higher for urban residents. Report a range within which the P-value for this test must fall.

7. Suppose that researchers study a sample of 40 people and find that 10 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 0.27 against the alternative that it is less 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 0.24. Find the power of the test.

8. Suppose that the sample unemployment rate in Ontario is 8% based on a survey of 100 people, while the rate in Quebec is 10% 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 Quebec, and report the associated P-value.

9. Statisticians study the relationship between an indicator of high school average grades (labelled y) and hours of home work per day, labelled x, across a sample of 25 students 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 3.0 with a t-statistic of 2.492. They also find an R^2 statistic of 0.70. They also find $\hat{a} = 65$.

(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 $\hat{a}$?

(d) Does the correlation or regression evidence prove that doing more homework raises one's average grade?

Economics 250
Winter Term 2018
Final Exam Answer Guide

1. (a: 2 marks) The sample mean is 4. The median and mode also are 4.

(b: 2 marks) The sample variance is $s^2 = 40/15 = 2.667$ so the sample standard deviation is $s = 1.63$.

(c: 2 marks) Its clear that $Q_1 = 2$ and $Q_2 = 6$ so the IQR is [2,6] with a width of 4.

2. (a: 2 marks) $P(N \cap G) = 0.11$.

(b: 2 marks) $P(NG|U) = P(NG \cap U)/P(U) = 0.036/0.69 = 0.5217$.

(c: 2 marks) $P(G|N) = P(G \cap N)/P(N) = 0.11/0.254 = 0.433$.

3. (a: 2 marks) The mean is 20. The marginal probabilities are 0.5 and 0.5. The standard deviation is 10.

(b: 3 marks) A chart shows the covariance is -60, so the correlation is -0.6 because both standard deviations are 10.

(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 0 and the standard deviation as 17.88.

(d: 2 marks) The conditional probabilities are 0.2 and 0.8 so the conditional mean is $0.2(10) + 0.8(30) = 2 + 24 = 26$.

4. (a: 2 marks) From Table C the probability of 1 or more successes is 1- 4633=0.5367 and the probability of 3 or more successes is 0.0361.

(b: 2 marks) Now $X \sim N(7.5, 2.67)$ so the two z-statistics are -2.43 and -1.68. The probabilities from Table A thus are 0.9925 and 0.9535.

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

$$39.659 \pm 2.131\frac{4}{4} = 39.659 \pm 2.131 = (37.528, 41.79)$$

using the appropriate t-statistic.

(b: 2 marks) Our test statistic is:

$$t = \frac{39.659 - 41}{4/4} = -1.341$$

In Table D, with $df = 15$, the tail area is 0.10 so multiplying by 2 gives a P-value of 0.20.

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

$$80 - 79 \pm 1.86(0.505) = 1 \pm 0.939 = (0.061, 1.939).$$

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

$$\frac{1-0}{0.505} = 1.98,$$

so from the relevant row of Table D the P-value is between 0.025 and 0.05.

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

$$0.25 \pm 1.96(0.0685) = 0.25 \pm 0.1342 = (0.1158, 0.3842)$$

(b: 2 marks) The Wilson value is $\tilde{p} = 12/44 = 0.2727$ (with some rounding) so the CI would be

$$0.2727 \pm 1.96(0.067) = 0.2857 \pm 0.1313 = (0.1414, 0.404)$$

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

$$-1.282 = \frac{\hat{p}_c - 0.27}{\sqrt{0.27(0.73)/40}}$$

(where notice we use 0.27 in the denominator, not the sample value) which yields $\hat{p}_c = 0.18$. Since $\hat{p}_c < \hat{p}$ the null is not rejected.

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

$$z = \frac{0.18 - 0.24}{\sqrt{0.24(0.76)/40}} = -0.06/0.06752 = -0.89.$$

Using -0.89 in Table A gives power as 0.1867.

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

$$0.02 \pm 1.645(0.0404) = 0.02 \pm 0.066 = (-0.046, 0.086)$$

(b: 3 marks) The test statistic involves the pooled estimate which is 0.09. Using that with each sample size gives a standard deviation for the difference of 0.0404, so the test statistic is

$$z = \frac{0.02}{0.0404} = 0.495$$

Averaging the two nearest values in Table A (0.6879 and 0.6915) yields 0.6897 for a P-value of 0.3103.

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

(b: 2 marks) From Table D with $df = 24$ the P-value would be 0.02.

(c: 2 marks) The intercept is an estimate of the average grade for those who do no homework. Notice that this is an estimate: there may be no data point at $x = 0$.

(d: 2 marks) First, with the small probability in part (b) there could still be sampling variability, or type I error, so there could be no population effect. Second, correlation does not prove causation. Perhaps strong students also enjoy homework but would receive high grades even without doing any of it.

Total Marks: 57