QUEEN'S UNIVERSITY
FACULTY OF ARTS AND SCIENCE

APRIL 2016

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. Allowed calculators include those with blue or gold stickers, the Casio 991, the Sharp EL376S, or other non-programmable calculators. No red-sticker calculators or other aids are allowed.

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. A survey of incomes (measured in thousands) shows that 3 households have income of 20, 5 households have income of 30, and 1 household has income of 87.

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

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

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

2. A pension fund plans to invest its holdings in equal amounts (*i.e.* 50:50) in two investments. The first one has a return r_1 with mean 1 and variance 0. The second one has a return r_2 with mean 4 and variance 4.

(a) Call the portfolio return r_p. Find its mean and variance.

(b) What is the correlation between r_p and r_2?

3. Suppose that within any day the exchange rate between the Hong Kong dollar and the US dollar is continuously and uniformly distributed between 7.7 and 7.9 (measured in Hong Kong dollars).

(a) Find the mean and standard deviation of the exchange rate.

(b) Suppose that there are 256 trading days per year. Find the distribution of the average exchange rate over the year.

4. In Canada the *Labour Force Survey* classifies people into three categories: E (employed), U (unemployed) and N (not in the labour force). It also distinguishes between those under 25 (whom we shall label Y) and those over 25 (whom we shall label O). Suppose that 20% of people are Y and 80% are O. Also $P(E|Y) = 0.5$, $P(E|O) = 0.7$, $P(U|Y) = 0.1$ and $P(U|O) = 0.05$.

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

(b) Find $P(E)$, $P(U)$, and $P(N)$.

(c) A statistician interviews a worker who is unemployed. What is the probability that the worker is young?

5. Suppose that the probability that a species of bird reaches a remote island in any year is 0.02.

(a) What is the probability that such a bird reaches the island one or more times in a 10 year period?

(b) What is the probability that such a bird reaches the island one or more times in a 100 year period?

6. A statistician is testing the hypothesis that the average effect of a training program on wages is zero against the alternative hypothesis that the effect is positive. Suppose that the statistician knows that the population standard deviation of the effect is $\sigma = 2$. There are 36 people in the study. The average effect in the sample is 0.5.

(a) Find the P-value associated with this test.

(b) Suppose that $\alpha = 0.05$. Is the null hypothesis rejected or not?

(c) If the statistician used $\alpha = 0.05$ to conduct the significance test then what would the probability of type II error be if in fact the true value were $\mu = 1$? What would be the power of the test?

7. The news media report that 20% of likely voters support a specific political party. They also report that the margin of error in this opinion poll is 7.84% with 95% confidence.

(a) What must the sample size be?

(b) What would the 95% confidence interval be if you instead used the Wilson (*i.e.* 'plus 4') correction?

8. A sample of 25 people in Indonesia yields an average lifespan of 65 years with a sample standard deviation of 5 years. A sample of 36 people in Vietnam yields an average lifespan of 68 years with a sample standard deviation of 4 years.

(a) Construct the 95% confidence interval for the difference between the population average lifespan in Vietnam and that in Indonesia.

(b) A demographer wishes to test the null hypothesis that average lifespan is the same in the two countries against the alternative that it is higher in Vietnam. Report a range within which the P-value for this test must fall.

9. Suppose that the rate of cell-phone usage in Kenya is 60% based on a survey of 100 people, while the rate in Tanzania is 36% based on a survey of 100 people.

(a) Form a 95% confidence interval for the difference between the two usage 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 Kenya, and report the associated P-value.

10. An econometrics student runs a linear regression to study the relationship between the real wage rate earned by a worker and the number of years since the worker immigrated to Canada using data from 10,000 people. Denote the wage by w and the years since arrival by y. The regression line thus looks like this:

$$w = a + by + e,$$

where e is an error term. The student finds that the estimate of a, labelled $\hat{a}$, is 15 and the estimate of b, labelled $\hat{b}$, is 0.25. The R^2 statistic is 0.8.

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

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

(c) Predict the wage in 2020 for a worker who immigrated in 2000.

(d) Can you suggest a lurking or omitted variable that might matter?

Economics 250
Winter Term 2016
Final Exam Answer Guide

1. (a) The sample mean is 33. The sample standard deviation is 20.82.

(b) The median is 30. This is less than the mean, so the distribution is skewed right.

(c) The CV is $(100) \times (20.82/33) = 63.1\%$.

2. (a) The mean is 2.5. The variance is 1.

(b) The correlation is 1. There is a perfect linear relationship between these two random variables because there is no variation in r_1.

3. (a) The mean is 7.8. The variance is 0.00333 so the standard deviation is 0.0577.

(b) From the CLT the average follows a normal distribution with mean 7.8 and standard deviation $\sigma/\sqrt{n} = 0.0577/16 = 0.0036$.

4. (a) This probability of being Y and N is 0.08 or 8%.

(b) $P(E) = 0.66$, $P(U) = 0.06$, and $P(N) = 0.28$.

(c) $P(Y|U) = 0.02/0.06 = 0.33$.

5. (a) From table C the probability of no arrivals is 0.8171 so the probability of 1 or more is 0.1829.

(b) From the normal approximation the number of arrivals has mean 2 and standard deviation 1.4. So construct $z = (1 - 2)/1.4 = -0.714$. From Table A the area below this is 0.2389 so the area above this is 0.7611. There is a 76.11% chance of at least one arrival in 100 years.

6. (a) The test statistic is $z = (0.5 - 0)/(2/6) = 1.5$.
From Table A the P-value is 1-0.9332 = 0.0668 or 6.68%

(b) If $\alpha = 0.05$ then the null is not rejected since $P > \alpha$.

(c) The critical value for the sample mean would come from:

$$1.645 = \frac{\overline{x}_c - 0}{2/6},$$

so $\overline{x}_c = 0.5483..$ Standardizing this in the new distribution (when $\mu = 1$) gives:

$$z = \frac{0.54833 - 1}{2/6} = -1.355.$$

Averaging the two nearest values in Table A gives 0.0877 to the left of this point in the z distribution. So the probability of Type II error is 8.77% and the power is 91.23%.

7. (a) The margin of error is $0.0784 = 1.96\sqrt{0.16}/\sqrt{n}$ so $n = 100$.

(b) In this case $\tilde{p} = (20 + 2/100 + 4) = 0.2115$. The margin of error (now using $n + 4$) would be 0.0785 so the CI would be (0.133, 0.29).

8. (a) The CI is $3 \pm 2.064 \cdot 1.2018 = 3 \pm 2.480 = (0.52, 5.48)$. Notice that we use the minimum of the two df.

(b) Our t-statistic is $t = (3 - 0)/1.2018 = 2.496$. With $df = 24$ that leaves between 0.01 and 0.05 in one tail so that is the range for the P-value because this is a one-tailed test.

9. (a) The margin of error is 0.134 using $z = 1.96$ so the CI is (0.106, 0.374).

(b) The pooled rate of use is 0.48 and the pooled standard deviation is 0.07. Our test statistic is $z = (0.24 - 0)/0.07 = 3.38$. From Table A the P-value is 1-0.9996 = 0.0004.

10. (a) The intercept is the wage of an immigrant who has just arrived.

(b) Yes, for the R^2 statistic is quite large.

(c) With $y = 20$ the prediction is 20.

(d) Candidates would include education and language skills. Those probably are associated with higher wages controlling for years since immigration.