Economics 250
Introductory Statistics

Exercise 3

Due Wednesday 24 March by noon via onQ

Instructions: No late submissions are possible. Show intermediate steps (such as formulas) in your work both for ease of review and part marks in case of error. Do the exercise on your own and submit only your own work. Plots and graphs must be drawn with software (such as Excel) not by hand. Note: Please submit text in a single .pdf file and not a .doc or multiple photo files. Marks will be deducted otherwise.

1. Suppose that we know that, in a population, spells of unemployment are normally distributed with a mean of 90 days and a standard deviation of 8 days.

(a) What is the probability that a spell lasts more than 94 days?

(b) Researchers collect a random sample of 4 unemployment spells. What is the probability that the average spell lasts more than 94 days?

2. An analyst studies how long people spend at a website. Suppose the analyst knows that the population standard deviation of these times is $\sigma = 4$ seconds. A sample of 16 visits yields an average visit time of 20 seconds.

(a) Suppose the analyst reports a confidence interval of $(19.5, 20.5)$. What is the confidence level?

(b) Find a 66.8% confidence interval for the average visit duration.

(c) Now suppose that a larger study involves 64 visits. Find a 66.8% confidence interval for the average visit duration.

(d) At this larger sample size, what is the confidence level associated with the interval in part (b)?

3. An economist is asked to estimate the average duration of unemployment (measured in months) in a population. From a random sample of 9 spells of unemployment, she finds a sample average duration of 8. Suppose that she knows the population standard deviation of spells is $\sigma = 3$.

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

(b) Test $H_0 : \mu = 7$ against $H_A : \mu > 7$ with $\alpha = 0.05$.

(c) If the true (but unknown) average duration is in fact $\mu = 8$ then what is the power of the test in part (b)?

4. Suppose that a tutoring program raises the grades of 25 high school students by an average of 8 points. The sample standard deviation of the increase in their grades is 4 points.

(a) Find a 95% confidence interval for the population average grade change.

(b) Previous studies suggest the effect of tutoring is to raise grades by an average of 6.9456 points. Test the hypothesis that the average grade change equals 6.9456 points against the alternative that it differs from that value, by reporting the P-value.

5. Suppose that 16 people have an average weight of 70 kg. They then participate in a new exercise program, after which their average weight is 67 kg. The sample standard deviation of the *change* in their weights is 5.631 kg.

(a) Find a 90% confidence interval for the population average change in weight as a result of the program.

(b) Test the null hypothesis that the average effect of the program on weight is zero against the alternative that it is negative (so that people lose weight), and report the P-value.

6. In a sample of 10 people in Japan average life expectancy is 85 years with a sample standard deviation of 3 years. In a sample of 8 people in South Korea the corresponding numbers are 82 years and 2 years.

(a) Find a 95% confidence interval for the population average difference in life expectancy.

(b) Suppose you wish to test the null that the two population life expectancy values are equal against the alternative that they are different. Find a range in which the P-value must fall.

7. Use the data set *OilCAD2021*, where the columns give the value of the Canadian dollar (in US cents) and the world oil price (West texas intermediate, or WTI, in USD) for each month from January 2000 to December 2020.

(a) Use Excel to produce and properly label a time series plot that includes both variables. (Make sure they are labelled so the reader can tell which is which.)

(b) Run a linear regression of the exchange rate (the y_t variable) on the oil price (the x_t variable). Report the the estimated intercept and the estimated slope (the coefficient on the x_t variable). Report the R-square (R^2) statistic.

(c) Run a linear regression of the exchange rate (y_t) on its lagged value (the x-variable, but we'll refer to it as y_{t-1}). (Hint: Create this variable by copying the exchange rate column into a new column, but shifted down one row. You will lose the first observation when you do this so enter the range accordingly.) Report the the estimated intercept and the estimated slope. Report the R-square (R^2) statistic.

(d) An economist writes: "We can explain the movements in the value of the Canadian dollar using the oil price, but we cannot predict its value in advance." Do you agree or disagree and why?

(e) Finally see what happens when you include *both* explanatory variables. This is called multiple regression. In the regression dialogue box simply enter the x range as something like c5...d255 so both variables are included. Report the two estimated coefficients and the R^2. Also report the t-statistics, one for each variable. These test the null that the slope is zero. Is there more evidence that the oil price or the lagged exchange rate statistically explains the current exchange rate?

Economics 250
Exercise 3 Answer Guide

1. (a: 2 marks) Here $x \sim N(90, 8)$ so $z = (94 - 90)/8 = 0.5$. From Table A Prob($z > 0.5$) $= 1 - 0.6915 = .3085$ or 30.85%.

(b: 2 marks) Now $\overline{x} \sim N(90, 8/2)$ so $z = (94 - 90)/4 = 1.0$. The probability thus is $1 - 0.8413 = 0.1587$ or 15.87%. (Notice that an average spell being of long duration is much less likely than an individual spell being of that same long duration.)

2. (a: 2 marks) From table A there is 0.6915 probability below the value 1 so there is 0.3085 in each tail and this is a 38.3% interval.

(b: 2 marks) The value of $z_{\alpha/2} = 0.97$ so the 66.8% interval is (19.03, 20.97).

(c: 2 marks) Now $\sigma/n = 0.5$ so the interval is (19.515, 20.485).

(d: 2 marks) The original margin of error was $0.97 = z_{\alpha/2}4/\sqrt{64}$, so $z_{\alpha/2} = 1.94$ and this now is a 94.76% interval.

3. (a: 2 marks) The 95% confidence interval is:

$$8 \pm 1.645 \cdot \frac{3}{3} = 8 \pm 1.645 = (6.355, 9.645).$$

(b: 2 marks) Leaving 5% in the upper right tail requires $z_c = 1.645$. Our test statistic is:

$$z = \frac{8 - 7}{3/3} = 1,$$

which is less than z_c, so we do not reject the null at the 5% level of significance.

Alternately, you could find the critical value $\overline{x}_c$ and find the sample average is less than that. That gives $\overline{x} = 8.645$.

(c: 2 marks) Under the alternative the distribution is:

$$\overline{x} \sim N(8, 1).$$

Locating the critical value 8.645 in this distribution gives:

$$z = \frac{8.645 - 8}{1} = 0.645.$$

From Table A the area below that point is 0.74055 (averaging the entries) so the area above that point is 1-0.74055 = 0.25945 or 25.945%: That is the power of the test.

4. (a: 2 marks) With df=25 the relevant t statistic is 2.064 so the 95% confidence interval is

$$8 \pm 2.064 \cdot 4/\sqrt{25} = 8 \pm 1.6512 = (6.3488, 9.6512)$$

(b: 2 marks) The test statistic is:

$$t = \frac{8 - 6.9456}{4/5} = 1.318.$$

From Table D with $df = 24$ that leaves 10% in each tail so the P-value is 20%.

5. (a: 2 marks) With $df = 15$ and $\alpha = 0.10$ table D shows $t_{\alpha/2,15} = 1.753$. Thus the 90% confidence interval is:

$$-3 \pm 1.753 \times \frac{5.631}{\sqrt{16}} = -3 \pm 2.468 = (-5.468, -0.532).$$

(b: 2 marks) Our test statistic is:

$$t = \frac{-3 - 0}{1.408} = -2.131$$

The t distribution is symmetric so from table D the P-value is 0.025.

6. (a: 2 marks) The difference in averages is 3. With $df = 7$ the appropriate t statistic is 2.365. So the 95% confidence interval is:

$$3 \pm 2.365\sqrt{\frac{3^2}{10} + \frac{2^2}{8}} = 3 \pm 2.365 \times 1.183 = 3 \pm 2.798 = (0.202, 5.798)$$

(b: 2 marks) Our test statistic is:

$$t = \frac{3}{1.183} = 2.535$$

which we study with 7 df. In table D that is between 0.02 and 0.01 but this is a two-tailed test so the P-value is between 0.02 and 0.04. (That is the chance of finding a test statistic this far from zero in either direction when the null hypothesis is true.)

7. (a: 4 marks) Please see the attached figure.

(b: 4 marks) The estimated line is:

$$y_t = 57.38 + 0.416x_t$$

with $R^2 = 0.83$.

(c: 4 marks) The estimated line now is:

$$y_t = 0.995 + 0.988y_{t-1},$$

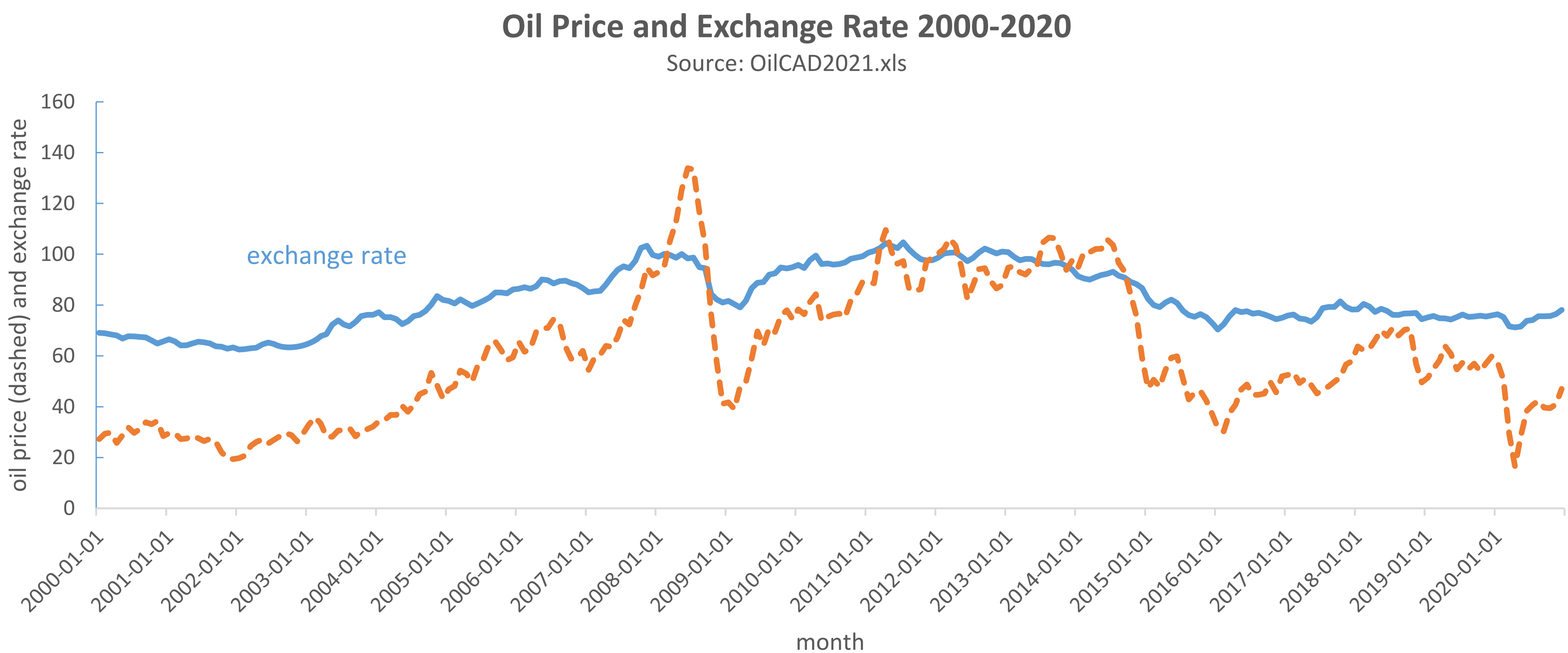
with $R^2 = 0.98$.

(d: 4 marks) I disagree. It is true that the first regression has high explanatory power, so there is a strong, positive association. But the R^2 statistic is even higher in the second regression. That means we can predict the exchange rate more accurately using last month's value than using this month's oil price.

(e: 4 marks) The multiple regression gives:

$$y_t = 2.67 + 0.16x_t + 0.956y_{t-1}$$

as the fitted line. The R^2 value is 0.981. The two t-statistics are 1.66 and 45.33. (That means there is much more evidence against the null of a zero effect for the lagged exchange rate than there is for the oil price, in other words more evidence that the lagged exchange rate matters.)


Oil Price and Exchange Rate 2000-2020
Source: OilCAD2021.xls
oil price (dashed) and exchange rate
exchange rate
160
140
120
100
80
60
40
20
0
2000-01-01
2001-01-01
2002-01-01
2003-01-01
2004-01-01
2005-01-01
2006-01-01
2007-01-01
2008-01-01
2009-01-01
2010-01-01
2011-01-01
2012-01-01
2013-01-01
2014-01-01
2015-01-01
2016-01-01
2017-01-01
2018-01-01
2019-01-01
2020-01-01
month