Economics 250
Introductory Statistics

Exercise 2
Due Friday 7 March 2025

Instructions: 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. Write in pen (or type) and not in pencil. Work can be sumbitted on paper in class or in a single file via OnQ.

1. Suppose that a financial asset pays a dividend of $4 sometimes and $8 at other times. Let us study the dividends in two successive years, denoted d_1 and d_2. These two discrete random variables are jointly distributed like this:

		d_1	
		4	8
d_2	4	0.4	0.1
	8	0.1	0.4

(a) Find the marginal distribution for d_2. Then find the mean and standard deviation using that distribution. (Sometimes these are called the unconditional mean and standard deviation.)

(b) Suppose that you observe that $d_1 = 4$. Find the conditional distribution for d_2. Then find the conditional mean and standard deviation.

(c) Are the two random variables independent? Find the covariance and the correlation.

2. Suppose that 10% of athletes in a competition are cheating by taking banned stimulants. A test is available that yields evidence of cheating (*i.e.* a positive test) 8% of the time for athletes who are not cheating and 70% of the time for athletes who are cheating.

(a) What is the probability of a randomly selected athlete testing positive?

(b) If an athlete tests positive what is the probability they were cheating?

3. Suppose that an after-school tutoring program succeeds in raising any student's mathematics test score with a probability of 0.7.

(a) If 10 students participate in the program then what is the probability that at least 6 of them (*i.e.* 6 or more) experience a test score increase?

(b) What is the probability that only 1, 2 or 3 of them experience a test score increase?

(c) If 100 students participate then what is the probability that at least 60 of them experience a wage increase? (Hint: Think of this sample size as large and do not use the binomial formula.)

(d) For large n what is the distribution of the success rate or proportion, defined as $\hat{p} = x/n$?

4. Statistics are central to managing an investment portfolio. Imagine that an investor is choosing how to divide each dollar between two investments, labelled A and B. Investment A has a return with mean 2 and standard deviation 1 and so does investment B. The correlation between the two returns is 0.5. The portfolio invests a fraction ω in investment A and a fraction $1-\omega$ in investment B so that the portfolio's return is:

$$r_p = \omega r_A + (1-\omega) r_B.$$

(a) Find formulas for the mean of the portfolio return, labelled μ_p, and the variance of the portfolio return, labelled σ_p^2, as functions of ω. Do the mean and variance of the portfolio return depend on the value of ω?

(b) In finance, risk somestimes is measured using the variance. Find the value of ω that minimizes the variance of r_p. (Hint: Use your mathematics skills to minimize the function you derived.)

5. Suppose that we know that, in a population, spells of unemployment are normally distributed with a mean of 144 days and a standard deviation of 12 days.

(a) What is the probability that a spell lasts more than 150 days?

(b) Researchers collect a random sample of 4 unemployment spells. What is the probability that the average spell lasts more than 150 days?

6. Use the data set *Fisher* from the Projects web page for this course. For years from 1962 to 2024 it shows Canadian 10-year bond interest rates and CPI inflation rates, both in percentage points for each month.

(a) Use Excel to produce and properly label a scatter plot of the 10-year interest rate (on the vertical axis) plotted against the inflation rate (on the horizontal axis).

Next run a linear regression like this:

$$y = a + bx,$$

In Excel, go to the 'data' tab then to 'data analysis" and 'regression'.

(b) Report the coefficient on the x-variable (*i.e.* the inflation rate) and its standard error. Also report the R-square (R^2) statistic.

(c) Sections 2.3–2.4 of the textbook tell us how to interpret these statistics. Based on these statistics is there much evidence of a relationship between this interest rate rate and the inflation rate, in these time series data?

(d) If the inflation rate is 2% then what will the predicted interest rate rate be, given this historical relationship?

7. Use the data set *OilCAD*, 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 2024.

(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.) 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?

Economics 250
Exercise 2 Answer Guide

1. (a: 2 marks) The marginal distribution for d_2 has probability 0.5 on 4 and 0.5 on 8. Thus the mean is 6. The variance is 4. The standard deviation is 2.

(b: 3 marks) The conditional distribution of d_2, given $d_1 = 4$, has weight 0.8 on the value 4 and 0.2 on the value 8. Thus the conditional mean is 4.8. The conditional variance uses the conditional mean and the conditional probabilities, which gives 2.56 so the standard deviation is 1.6. (Notice, by the way, that the new information leads the researcher to revise down the mean and the standard deviation.)

(c: 4 marks) The two random variables are not independent. Our work in parts (a) and (b) shows that the conditional probabilities differ from the marginal ones. Another way to show they are not independent is to find the correlation. The covariance is 2.4 so the correlation is 0.6.

2. (a: 3 marks) Here $P(C) = 0.10$, $P(Pos|C) = 0.70$, and $P(Pos|Not\,C) = 0.08$. So,

$$\begin{aligned} P(Pos) &= P(Pos|C)P(C) + P(Pos|Not\,C)P(Not\,C) \\ &= 0.70(0.10) + 0.08(0.90) = 0.07 + 0.072 = 0.142, \end{aligned}$$

or 14.2%.

(b: 3 marks) Now

$$P(C|Pos) = \frac{P(C \cap Pos)}{P(Pos)} = \frac{0.07}{0.142} = 0.493$$

or 49.3%.

3. (a: 3 marks) This is a binomial problem with $n = 10$ and $p = 0.7$. (We can use Table C remembering that the probability of failure is thus 0.3 and we want to count 4 or fewer failures.) From Table C the answer is 0.8497.

(b: 3 marks) Reading up the column in Table C gives 0.0105.

(c: 3 marks) We use the normal approximation, so that

$$x \sim N(70, 4.58).$$

That means that:

$$z = \frac{60 - 70}{4.58} = -2.18.$$

So from Table A the probability above this value is 1-0.0146=0.9854 or 98.54%.

(d: 3 marks) The distribution for large n is:

$$\hat{p} \sim N(0.70, \frac{0.458}{\sqrt{n}}).$$

4. (a: 3 marks) The mean is:

$$\mu_p = \omega(2) + (1 - \omega)2 = 2\omega + 2 - 2\omega = 2.$$

The variance is:

$$\sigma_p^2 = \omega^2(1^2) + (1 - \omega)^2(1^2) + 2\omega(1 - \omega)(0.5)(1)(1) = \omega^2 + 1 - \omega$$

The mean does not depend on ω, but the variance does.

(b: 3 marks) Now

$$\sigma_p^2 = \omega^2 + 1 - \omega$$

Differentiate with respect to ω and set the derivative to zero:

$$\frac{d\sigma_p^2}{d\omega} = 2\omega - 1 = 0,$$

which gives

$$\omega = 0.5.$$

Even though the two investments have the same mean and variance, the portfolio variance can be reduced by exactly diversifying as long as the correlation is not 1. (Incidentally, the second derivative is positive, so this is indeed a minimum.)

5. (a: 2 marks) Here $x \sim N(144, 12)$ so $z = (150 - 144)/12 = 0.5$. From Table A Prob$(z > 0.5) = 1 - 0.6915 = .3085$ or 30.85%.

(b: 2 marks) Now $\overline{x} \sim N(144, 12/2)$ so $z = (150 - 144)/6 = 1.0$. 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.)

6. (a: 3 marks) Please see the attached figure.

(b: 3 marks) The slope is 0.711 with a standard error of 0.0299. The R^2 statistic is 0.43.

(c: 3 marks) Yes, there is a moderately strong, positive relationship, as seen from the value of R^2.

(c: 3 marks) The estimated intercept is 3.753 so the predicted value of the interest rate is 3.753 + 0.711 (2) = 5.175%.

7. (a: 3 marks) Please see the attached figure.

(b: 3 marks) The estimated line is:

$$y_t = 59.70 + 0.343 x_t$$

with $R^2 = 0.60$.

(c: 3 marks) The estimated line now is:

$$y_t = 0.805 + 0.990 y_{t-1},$$

with $R^2 = 0.98$.

(d: 3 marks) I disagree. It is true that the first regression has fairly high R^2, 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. (Incidentally, we'll later see we do not need to choose between these but can combine them, using a method called multiple regression.)