Quantitative economics (L11QUE-ECON1045)
Quantitative methods (L11QUM–ECON1047)
Mathematical Economics and Statistical Methods (ECON1046)

Answer keys

Alessandro Ruggieri*

April 7, 2020

*Contact: School of Economics, Sir Clive Granger building, University of Nottingham, University Park, Nottingham, NG7 2RD, aruggierimail@gmail.com. The usual disclaimers apply.

Contents

Chapter 1

Descriptive Statistics

Exercise 1 The sample of observations is: $\{14, 28, 40, 13, 25, 27, 20, 29, 49, 66\}$. Since the sample size is $n = 10$, the sample mean is equal to

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{14 + 28 + 40 + 13 + 25 + 27 + 20 + 29 + 49 + 66}{10} = \frac{311}{10} = 31.1$$

To find the median, arrange first the observations in order:

$$13, 14, 20, 25, 27, 28, 29, 40, 49, 66$$

The median is the middle observation. When the number of observations is even, it is the mean of the two middle observations. There are 10 observations in this case, thus the median is 27.5 minutes.

The sample variance is given by

$$s^2 = \frac{\sum_{i=1}^{10}(x_i - \bar{x})^2}{10-1} =$$

$$\frac{(14-31.1)^2+(28-31.1)^2+(40-31.1)^2+(13-31.1)^2+(25-31.1)^2+(27-31.1)^2+(20-31.1)^2+(29-31.1)^2+(49-31.1)^2+(66-31.1)^2}{9}$$

$$= \frac{2428.9}{9} = 269.88$$

The sample standard deviation is given by

$$s = \sqrt{s^2} = \sqrt{269.88} = 16.43$$

The coefficient of variation is given by

$$CV = \frac{s}{\bar{x}} 100 = \frac{16.43}{31.1} 100 = 52.83\%$$

Exercise 2 The sample of prices of certain goods and quantities is the following:

$$\text{Price} = \{10, 15, 20, 25, 30\}$$

$$\text{Quantity} = \{100, 90, 75, 50, 0\}$$

Price and quantity move in opposite directions; there appears to be a negative relation between the two variables. Therefore we anticipate that the covariance will be negative. Define prices as X and quantity as Y . The sample covariance between X and Y is given by

$$\text{COV}[X,Y] = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Notice that

$$\bar{x} = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{100}{5} = 20 \qquad \bar{y} = \frac{\sum_{i=1}^{5} y_i}{5} = \frac{315}{5} = 63$$

Therefore the sample covariance is

$$\text{COV}[X,Y] = \frac{(10-20)(100-63)+(15-20)(90-63)+(20-20)(75-63)+(25-20)(50-63)+(30-20)(0-63)}{4} = \frac{-1200}{4} = -300$$

The sample correlation coefficient is given by:

$$r_{xy} = \frac{\text{COV}[X,Y]}{s_x s_y}$$

Notice that

$$s_x = \frac{\sum_{i=1}^{5}(x_i - \bar{x})^2}{5-1} = \sqrt{\frac{250}{4}} = 7.91 \qquad s_y = \frac{\sum_{i=1}^{5}(y_i - \bar{y})^2}{5-1} = \sqrt{\frac{6380}{4}} = 39.94$$

Therefore, the sample correlation coefficient is

$$r_{xy} = \frac{-300}{7.91(39.94)} = -0.95$$

There is a strong negative linear relationship between price and quantity.

Exercise 3 The sample of observations is:

$$X = \{1.0, 1.0, 1.1, 1.1, 1.1, 1.2, 1.2, 1.4, 1.5, 1.6, 1.6, 2.2, 3.4, 7.6, 10.6, 26.4\}$$

- The sample size n is equal to 16.
- The sample average is equal to:

$$\bar{x} = \frac{\sum_{i=1}^{16} x_i}{16} = \frac{1.0+1.0+1.1+1.1+1.1+1.2+1.2+1.4+1.5+1.6+1.6+2.2+3.4+7.6+10.6+26.4}{16} = \frac{64}{16} = 4$$

- The sample median is equal to:

$$\frac{1.4+1.5}{2} = \frac{2.9}{2} = 1.45$$

- The sample mode is 1.1.

- The range is equal to $26.4 - 1 = 25.4$

- The sample variance is equal to:

$$s^2 = \frac{\sum_{i=1}^{16}(x_i - \bar{x})^2}{16-1}$$
$$= \frac{(1.0-4)^2+(1.0-4)^2+(1.1-4)^2+(1.1-4)^2+(1.1-4)^2+(1.2-4)^2+(1.2-4)^2+(1.4-4)^2}{15} + \frac{(1.5-4)^2+(1.6-4)^2+(1.6-4)^2+(2.2-4)^2+(3.4-4)^2+(7.6-4)^2+(10.6-4)^2+(26.4-4)^2}{15}$$
$$= \frac{(-3)^2+(-3)^2+(-2.9)^2+(-2.9)^2+(-2.9)^2+(-2.8)^2+(-2.8)^2+(-2.6)^2}{15} + \frac{(-2.5)^2+(-2.4)^2+(-2.4)^2+(-1.8)^2+(-0.6)^2+(3.6)^2+(6.6)^2+(22.4)^2}{15}$$
$$= \frac{9+9+8.41+8.41+8.41+7.84+7.84+6.76}{15} + \frac{6.25+5.76+5.76+3.24+0.36+12.96+43.56+501.76}{15} = 43.0213$$

- The sample standard deviation is equal to:

$$s = \sqrt{s^2} = \sqrt{43.0213} = 6.559$$

Chapter 2

Set and probability theory

Exercise 1 Define the following two events:

- Event A: enjoys cycling
- Event B: enjoys reading

We want to calculate the number of people in the intersection of A and B ($A \cap B$). 2 out of 20 people enjoy neither cycling nor reading. Therefore 18 people must enjoy either cycling or reading or both and the number of people in the union of A and B ($A \cup B$) is 18. Recall that

$$A + B - (A \cap B) = A \cup B$$

By assumption $A = 15$ and $B = 8$. Then

$$15 + 8 - (A \cap B) = 18 \implies (A \cap B) = 15 + 8 - 18 = 5$$

Exercise 2 Define the following events:

- Event A: the first member is male
- Event B: the second member is male

Both the first and second members being male is the intersection of events A and B. We want to calculate $P(A \cap B)$. We can calculate $P(A)$ as there are six possible members, 4 of whom are male. So the probability of selecting a male is

$$P(A) = \frac{4}{6} = \frac{2}{3}$$

We can calculate the probability that the second member is male, given that the first member was male, i.e. $P(B|A)$. If one male member has been selected, there are 5 candidates left of whom 3 are male and 2 are female.

$$P(B|A) = \frac{3}{5}$$

Using the multiplication rule:

$$P(A \cap B) = P(B|A)P(A) = \frac{3}{5}\frac{2}{3} = 0.40$$

Exercise 3 Define the following events:

- Event A: a customer asks for help. $P(A) = 0.3$.
- Event B: a customer makes a purchase before leaving. $P(B) = 0.2$.

We are told that the intersection of these events is $P(A \cap B) = 0.15$. Using the addition rule:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.2 - 0.15 = 0.35$$

The probability of B occurring conditional on A having already occurred is equal to

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.15}{0.30} = 0.5$$

Are these events mutually exclusive? No, The probability of the intersection is not 0, i.e. $P(A \cap B) = 0$. Are they collectively exhaustive? No. The probability of the union is not 1, i.e. $P(A \cup B) \neq 1$. Are they independent? No. Independence requires: $P(A \cap B) = P(A)P(B)$ but $P(A)P(B) = 0.3(0.2) = 0.06 \neq P(A \cap B)$.

Exercise 4 The sets are $A = \{2, 4, 6\}$ and $B = \{4, 5, 6\}$. The answers are:

- complement of A: $\bar{A} = \{1, 3, 5\}$
- complement of B: $\bar{B} = \{1, 2, 3\}$
- intersection of A and B: $A \cap B = \{4, 6\}$
- intersection of $\bar{A}$ and B: $\bar{A} \cap B = \{5\}$
- union of A and B: $A \cup B = \{2, 4, 5, 6\}$
- union of A and $\bar{A}$: $A \cup \bar{A} = \{1, 2, 3, 4, 5, 6\} = S$

Are A and B mutually exclusive? No. The outcomes 4 and 6 are common to both. Are A and B collectively exhaustive? No. The union of A and B does not contain the outcomes 1 or 3.

Chapter 3

Discrete random variables

Exercise 1 Consider the following discrete random variable

$$X = 0 \quad \text{w/ prob} = 0.25$$
$$X = 1 \quad \text{w/ prob} = 0.50$$
$$X = 2 \quad \text{w/ prob} = 0.25$$

Expected value and variance of X are equal to:

$$\text{E}[X] = \sum_{i=1}^{n} X_i P(X_i) = 0(0.25) + 1(0.50) + 2(0.25) = 1$$

$$\text{VAR}[X] = \sum_{i=1}^{n} (X_i - E[X])^2 P(X_i) =$$
$$= \left[(0-1)^2(0.25)\right] + \left[(1-1)^2(0.5)\right] + \left[(2-1)^2(0.25)\right] = 0.25 + 0.25 = 0.5$$

Exercise 2 Consider the joint probability distribution between X and Y

$$P(X = 0, Y = 1) = 0.30$$
$$P(X = 1, Y = 1) = 0.25$$
$$P(X = 0, Y = 2) = 0.20$$
$$P(X = 1, Y = 2) = 0.25$$

The marginal probability distributions for X is equal to:

$$P(X = 0) = P(X = 0, Y = 1) + P(X = 0, Y = 2) = 0.50$$
$$P(X = 1) = P(X = 1, Y = 1) + P(X = 1, Y = 2) = 0.50$$

The marginal probability distributions for Y is equal to:

$$P(Y=1) = P(Y=1, X=0) + P(Y=1, X=1) = 0.55$$
$$P(Y=2) = P(Y=2, X=0) + P(Y=2, X=1) = 0.45$$

The expected value of X and Y are equal to:

$$\mathrm{E}[X] = 0(0.50) + 1(0.50) = 0.50$$
$$\mathrm{E}[Y] = 1(0.55) + 2(0.45) = 1.45$$

The variance of X and Y are equal to:

$$\mathrm{VAR}[X] = (0-0.50)^2(0.50) + (1-0.5)^2(0.50) = 0.25$$
$$\mathrm{VAR}[Y] = (1-1.45)^2(0.55) + (2-1.45)^2(0.45) = 0.2475$$

The probability of having one exam conditional on eating two snacks is:

$$P(X=1|Y=2) = P(X=1, Y=2)/P(Y=2) = \frac{0.25}{0.45} \approx 0.56$$

The covariance of X and Y is equal to

$$\mathrm{COV}[X,Y] = \sum_{j=1}^{n}\sum_{i=1}^{n}(X_j - E[X])(Y_i - E[Y])P(X = X_j, Y = Y_i)$$
$$= (0-E[X])(1-E[Y])P(X=0,Y=1) + (0-E[X])(2-E[Y])P(X=0,Y=2) +$$
$$(1-E[X])(1-E[Y])P(X=1,Y=1) + (1-E[X])(2-E[Y])P(X=1,Y=2)$$
$$= (0-0.50)(1-1.45)0.3 + (1-0.50)(1-1.45)0.25 +$$
$$+(0-0.50)(2-1.45)0.20 + (1-0.50)(2-1.45)0.25 = 0.025$$

Exercise 3 Notice that:

- Number of trials: $n = 6$
- Number of successes: $x = 2$
- Probability of success: $p = 0.6$

The Binomial probability formula is:

$$P(X = x) = C_x^n p^x (1-p)^{(n-x)}$$

where

$$C_x^n = \frac{n!}{x!(n-x)!}$$

Since $C_x^n = \frac{6!}{2!4!} = 15$, then

$$P(X = 2) = 15(0.6^2 0.4^4) = 15(0.009) = 0.138$$

Exercise 4 Notice that

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

Since:

$$P(X = 0) = C_0^{12}(0.1^0)(0.9^{12}) = 0.282$$
$$P(X = 1) = C_1^{12}(0.1^1)(0.9^{11}) = 0.377$$
$$P(X = 2) = C_2^{12}(0.1^2)(0.9^{10}) = 0.230$$

Then:

$$P(X \leq 2) = 0.282 + 0.377 + 0.230 = 0.889$$

The probability of a shipment containing 3 or more faulty phones is $1 - P(X \leq 2) = 1 - 0.889 = 0.111$. Let Y be the number of shipments containing 3 or more faulty phones. Y can be modelled with a binomial distribution with $n = 6$ and $p = 0.111$. The probability that at least one shipment will contain 3 or more faulty phones is:

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - C_0^6(0.111^0)(0.889^6) = 1 - 0.494 = 0.506$$

Finally, notice that:

- Expected value: $\mu = np = 6(0.111) = 0.666$
- Variance: $\sigma^2 = np(1 - p) = 6(0.111)(1 - 0.111) = 0.592$
- Standard deviation: $\sigma = \sqrt{(}0.592) = 0.769$

The probability that Y exceeds its mean by more than 2 standard deviations is

$$P(Y \geq \mu + 2\sigma) = P(Y \geq 0.666 + 2(0.769)) = P(Y > 2.205)$$

As Y is discrete, this is equal to $P(Y > 3) = 1 - [P(Y = 0) + P(Y = 1) + P(Y = 2)]$. Since:

$$P(Y = 0) = C_0^6(0.111^0)(0.889^6) = 0.494$$
$$P(Y = 1) = C_1^6(0.111^1)(0.889^5) = 0.370$$
$$P(Y = 2) = C_2^6(0.111^2)(0.889^4) = 0.115$$

then $P(Y > 3) = 1 - 0.494 - 0.370 - 0.115 = 0.021$.

Chapter 4

Continuous random variables

Exercise 1

- To calculate the probability of x we integrate its PDF, $f(x) = \frac{1}{9}x^2$, over the appropriate range of values:

$$P(0 \leq X \leq 1) = \int_0^1 f(x)dx = \int_0^1 \frac{1}{9}x^2 dx = \left[\frac{x^3}{27}\right]_0^1 = \frac{1}{27} - 0 = \frac{1}{27}$$
$$P(0 \leq X \leq 2) = \int_0^2 f(x)dx = \int_0^2 \frac{1}{9}x^2 dx = \left[\frac{x^3}{27}\right]_0^2 = \frac{8}{27} - 0 = \frac{8}{27}$$
$$P(1 \leq X \leq 2) = \int_1^2 f(x)dx = \int_1^2 \frac{1}{9}x^2 dx = \left[\frac{x^3}{27}\right]_1^2 = \frac{8}{27} - \frac{1}{27} = \frac{7}{27}$$

- The expected value of X is:

$$\mathrm{E}[X] = \int_0^3 xf(x)dx \implies \mathrm{E}[X] = \int_0^3 \frac{1}{9}x^3 dx = \left[\frac{x^4}{36}\right]_0^3 = \frac{3^4}{36} - 0 = \frac{9}{4}$$

- The variance of X is:

$$\mathrm{VAR}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$$

Since:

$$\mathrm{E}[X^2] = \int_0^3 x^2 f(x)dx = \int_0^3 \frac{1}{9}x^4 dx = \left[\frac{x^5}{45}\right]_0^3 = \frac{3^5}{45} - 0 = \frac{243}{45} = \frac{27}{5}$$

Then:

$$\mathrm{VAR}[X] = \frac{27}{5} - \left(\frac{9}{4}\right)^2 = 0.3375$$

Exercise 2 The expected value of X is equal to

$$\mathrm{E}[X] = \int_0^1 x f(x) dx \implies \mathrm{E}[X] = \int_0^1 3x^3 dx = 3\left[\frac{x^4}{4}\right]_0^1 = 3\left[\frac{1}{4} - 0\right] = \frac{3}{4}$$

The variance of X is equal to

$$\mathrm{VAR}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$$

Since:

$$\mathrm{E}[X^2] = \int_0^1 x^2 f(x) dx = \int_0^1 3x^4 dx = 3\left[\frac{x^5}{5}\right]_0^1 = 3\left[\frac{1}{5} - 0\right] = \frac{3}{5}$$

then:

$$\mathrm{VAR}[X] = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = \frac{3}{5} - \frac{9}{16} = 0.0375$$

Exercise 3

- Solution to first question:
  - The following probabilities are obtained from Table 1 of the Statistical Tables booklet:

    $$P(0 \leq Z \leq 1.20) = P(Z \leq 1.20) - P(Z \leq 0) = 0.8849 - 0.5 = 0.3849$$

  - Note that the Statistics Tables only give probabilities for positive Z values, so we cannot find the probability $P(Z < -1.33)$ directly from these tables. But, as the standard normal distribution is symmetric

    $$P(-1.33 \leq Z \leq 0) = P(0 \leq Z \leq 1.33)$$

    which we can calculate as follows

    $$P(0 \leq Z \leq 1.33) = P(Z \leq 1.33) - P(Z \leq 0) = 0.9082 - 0.5 = 0.4082$$

  - $P(Z > 1.33) = 1 - P(Z < 1.33) = 1 - 0.9082 = 0.0918$
  - $P(-0.77 \leq Z \leq 1.68) = P(0 \leq Z \leq 1.68) + P(0 \leq Z \leq 0.77) =$ $(0.9535 - 0.5) + (0.7794 - 0.5) = 0.4535 + 0.2794 = 0.7329$
- Solution to second question: This probability is obtained from Table 1 of the

Statistical Tables booklet. As before, we can write

$$\begin{aligned} P(x \leq Z \leq 1.68) = P(Z \leq 1.68) - P(Z \leq x) &= 0.2 \\ 0.9535 - P(Z \leq x) &= 0.2 \\ P(Z \leq x) = 0.9535 - 0.2 &= 0.7535 \end{aligned}$$

This value (0.7535) is half way between the values 0.7517 and 0.7549 listed in Table 1. The critical values associated are 0.68 and 0.69 Therefore:

$$x = 0.5(0.68) + 0.5(0.69) = 0.685$$

Exercise 4 By assumption $X \sim \mathcal{N}(35000, 4000^2)$. We need to transform the normally distributed variable miles (X) into standard normally distributed random variable Z and use values from Table 1 of the Statistical Tables. We are told that X is normally distributed with a mean of 35000 and a standard deviation of 4000. Therefore:

$$Z = \frac{X - 35000}{4000}$$

is standard normally distributed (it has a mean of 0 and a standard deviation of 1). Applying this standardization technique, for the first question we get:

$$P(35000 < X < 38000) = P\left(\frac{35000 - 35000}{4000} < Z < \frac{38000 - 35000}{4000}\right) = P(0 < Z < 0.75)$$

Therefore:

$$P(0 < Z < 0.75) = P(Z < 0.75) - P(Z < 0) = 0.7734 - 0.5 = 0.2734$$

For the second question, we get:

$$P(X < 32000) = P\left(Z < \frac{32000 - 35000}{4000}\right)$$
$$= P(Z < -0.75) = P(Z > 0.75) = 1 - P(Z < 0.75) = 1 - 0.7734 = 0.2266$$

Exercise 5 Let X be a continuous random variable defined in $[0, 2]$ with pdf $f(x) = 0.5$. By definition, the cdf of X is

$$P(X \leq z) = \int_0^z f(x)dx$$

Therefore:

$$P(X \leq z) = \int_0^z 0.5dx = 0.5[x]_0^z = 0.5z$$

The probability that X takes values between 0.5 and 1.5 is equal to

$$P(0.5 \leq X \leq 1.5) = P(X \leq 1.5) - P(X \leq 0.5) = 0.5(1.5) - 0.5(0.5) = 0.50$$

Exercise 6 By assumption, $X \sim \mathcal{N}(1000000, 30000^2)$. The probability that the portfolio value is between 970000 and 1060000 is equal to

$$P(970000 < X < 1060000) = P\left(\frac{970000 - 1000000}{30000} < Z < \frac{1060000 - 1000000}{30000}\right) =$$
$$P(-1 < Z < 2) = P(Z < 2) - P(Z < -1) = 0.97725 - 0.15866 = 0.81859$$

Chapter 5

Estimators

Exercise 1 We want to compute $P(\bar{x} < 245)$. We are not told that the volume of the bottles is normally distributed, but $n \geq 30$ so the Central Limit Theorem applies here, and we can assume that the sample mean of bottle volume is normally distributed with mean μ and variance $\frac{\sigma^2}{n}$. Therefore:

$$\bar{x} \sim \mathcal{N}\left(250, \frac{400}{30}\right)$$

We need to transform this normal distribution into a standard normal:

$$P(\bar{x} < 245) = P\left(Z < \frac{245 - 250}{\sqrt{\frac{400}{30}}}\right) =$$
$$P(Z < -1.37) = P(Z > 1.37) = 1 - P(Z < 1.37) = 1 - 0.9147 = 0.0853$$

Exercise 2 To compute the 95% confidence interval for the population mean, with unknown standard deviation, we need the formula

$$P\left(\bar{x} - t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

We know that $n = 16$, $\bar{x} = 20.4$, $s = 6.4$, and $\alpha = 0.05$. The relevant critical value $t_{n-1,\frac{\alpha}{2}} = t_{15,0.025} = 2.131$. Therefore

$$P\left(20.4 - 2.131\frac{6.4}{4} \leq \mu \leq 20.4 + 2.131\frac{6.4}{4}\right) = 0.95$$
$$P\left(16.9904 \leq \mu \leq 23.8096\right) = 0.95$$

Exercise 3 We know that $n = 400$, $\bar{x} = 250$, $s = 64$, and $\alpha = 0.05$. The relevant critical value $t_{n-1,\frac{\alpha}{2}} = t_{399,0.025} = 1.96$. Note that $t_{399,0.025} = z_{0.025}$, i.e. given that n is large, the standard normal distribution provide a good approximation of the t-student distribution. Hence, we could use the critical values from normal distribution rather than the t-values. Therefore:

$$P\left(250 - 1.96\frac{64}{\sqrt{400}} \leq \mu \leq 250 + 1.96\frac{64}{\sqrt{400}}\right) = 0.95$$
$$P\left(243.728 \leq \mu \leq 256.272\right) = 0.95$$

Chapter 6

Hypothesis testing

Exercise 1 The null and alternative hypotheses are given by

$$\mathrm{H}_0 : \mu = 1500 \quad \text{versus} \quad \mathrm{H}_0 : \mu < 1500$$

We are not told that earnings are normally distributed, but as n is large ($n > 30$), the central limit theorem applies and we can assume that the sample mean $\bar{x}$ is normally distributed. The test statistic is given by

$$T = \frac{\bar{x} - 1500}{\frac{s}{\sqrt{n}}} = \frac{1262 - 1500}{\frac{432}{\sqrt{150}}} = -6.747$$

The relevant critical value for a one-sided hypothesis testing with $n = 150$ observations and one parameter to estimate at 5% significance level is $-t_{n-1,\alpha} = -t_{149,0.05} = -1.655$. We reject the null hypothesis if $T < \mathrm{cv}_\alpha$. Here, $T = -6.747 < -1.655$ thus we reject the null hypothesis. The sample provides evidence that the population mean earnings are less than 1500GBP.

Exercise 2

- The test statistic T has a sampling distribution. It is therefore possible that we do not always make the right decision from our hypothesis test. In fact, there are two possible types of error we can make. A Type I error occurs if we reject the null hypothesis H_0 when it is true (and therefore shouldn't be rejected). A Type II error occurs if we fail to reject H_0 when it is false (and therefore should be rejected).

- Power is the probability of rejecting the null hypothesis, H_0, given that it is false, i.e. the probability of correctly rejecting the null. This probability is equal to $P(T > \mathrm{cv}_\alpha|\mathrm{H}_1)$ or $P(T < -\mathrm{cv}_\alpha|\mathrm{H}_1)$ for a one sided test and to $P(|T| > \mathrm{cv}_{\frac{\alpha}{2}}|\mathrm{H}_1)$ for a two sided test. If as n increases, the power approaches 1, we call the test consistent. This is a desirable property for a test to possess.

Exercise 3 The null and alternative hypotheses are given by

$$\text{H}_0 : \mu = 50 \quad \text{versus} \quad \text{H}_1 : \mu \neq 50$$

The test statistic is

$$T = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{49.3 - 50}{\frac{1.25}{\sqrt{20}}} = \frac{-0.7}{0.280} = -2.5$$

With $n = 20$ observations and one parameter to estimate, the critical value for a two-sided test at 5% is given by

$$t_{\frac{0.05}{2},19} = t_{0.025,19} = 2.093$$

We reject the null hypothesis if $|T| > \text{cv}_{\frac{\alpha}{2},n-1}$. Since $|T| = 2.5 > 2.093$, we reject the null hypothesis. We have evidence to suggest the population mean number of matches in a box is NOT equal to 50.

Exercise 4 The null and alternative hypotheses are given by

$$\text{H}_0 : \mu = 50 \quad \text{versus} \quad \text{H}_1 : \mu > 50$$

The test statistic is

$$T = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{50.4 - 50}{\frac{1.25}{\sqrt{20}}} = \frac{0.4}{0.280} = 1.429$$

With $n = 20$ observations and one parameter to estimate, the critical value for a one-sided test at 5% is given by

$$t_{0.05,19} = 1.729$$

We reject the null hypothesis if $T > \text{cv}_{\alpha}$. Since $T = 1.429 < 1.729$, we fail to reject the null hypothesis. Evidence suggests that the population mean number of matches in a box is not larger than 50.

Exercise 5 The null and alternative hypotheses are given by

$$\text{H}_0 : \mu = 55 \quad \text{versus} \quad \text{H}_1 : \mu < 55$$

The test statistic is

$$T = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{52 - 50}{\frac{1.5}{\sqrt{16}}} = \frac{-3}{0.375} = -8$$

With $n = 16$ observations and one parameter to estimate, the critical value for a one-sided test at 5% is given by

$$t_{0.05,15} = -1.753$$

We reject the null hypothesis if $T < -\text{cv}_\alpha$. Since $T = -8 < -1.753$, we reject the null hypothesis. There is evidence to suggest that the population mean number of matches in a box is less than 55.

Exercise 6 The null and alternative hypotheses are given by

$$\text{H}_0 : \mu_x - \mu_y = 0 \quad \text{versus} \quad \text{H}_1 : \mu_x - \mu_y \neq 0$$

The test statistic is

$$T = \frac{\bar{x} - \bar{y} - \mu_0}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} = \frac{420 - 408 - 0}{\sqrt{\frac{12.5^2}{30} + \frac{18^2}{30}}} = \frac{12}{\sqrt{16.008}} = 2.999$$

Using the conservative approach that degrees of freedom v are equal to the smaller of $n_x - 1$ or $n_y - 1$, i.e. $v = 29$, the critical value for a two-sided test at 5% is given by

$$t_{\frac{\alpha}{2},v} = t_{\frac{0.05}{2},29} = 2.045$$

We reject the null hypothesis if $|T| > \text{cv}_\alpha$. Since $|T| = 2.999 > 2.045$ we reject the null hypothesis that the number of cars produced in the two factories have the same population mean. There is evidence to suggest that the population mean number of cars produced is different for the different factories.

Chapter 7

Linear regression

Exercise 1 The y variable is weight and the x variable is height. The OLS estimator of β_1 and β_0 are:

$$\hat{\beta}_1 = \frac{\text{COV}[x,y]}{\text{VAR}[x]} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Notice that:

$$\begin{aligned} \bar{x} &= \frac{\sum_{i=1} n x_i}{n} = \frac{892}{5} = 178.4 \\ \text{VAR}[x] &= \frac{\sum_{i=1} n (x_i - \bar{x})^2}{n-1} = \frac{411.2}{4} = 102.8 \\ \bar{y} &= \frac{\sum_{i=1} n y_i}{n} = \frac{373}{5} = 74.6 \\ \text{COV}[x,y] &= \frac{\sum_{i=1} n (x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{300.8}{4} = 75.2 \end{aligned}$$

Therefore:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\text{COV}[x,y]}{\text{VAR}[x]} = \frac{75.2}{102.8} = 0.732 \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} = 74.6 - 0.732(178.4) = -55.99 \end{aligned}$$

Exercise 2 The OLS estimates of β_1 and β_0 are equal to:

$$\begin{aligned} \hat{\beta}_1 &= \frac{25.2}{44.5} = 0.566 \\ \hat{\beta}_0 &= 1980 - 0.566(1280) = 1255.52 \end{aligned}$$

Exercise 3 Recall that the coefficient of determination R^2 is equal to coefficient of correlation squared, i.e.

$$R^2 = (r_{xy})^2 = \left(\frac{\text{COV}[x,y]}{s_x s_y} \right)^2$$

The coefficient of correlation is equal to the covariance between x and y divided by the product of standard deviations, i.e.

$$r_{xy} = \frac{\text{COV}[x, y]}{s_x s_y}$$

Therefore, we have

$$r_{xy} = \frac{25.2}{\sqrt{44.5}\sqrt{29.7}} = 0.693$$

and

$$R^2 = 0.693^2 = 0.480$$

i.e. 48% of the variation in the data is explained by the explanatory variable.

Exercise 4

- The hypothesis tested is:

  $$\text{H}_0 : \beta_1 = 0 \quad \text{versus} \quad \text{H}_1 : \beta_1 \neq 0$$

  Since $n = 30$, and the parameters to estimate are two, the relevant critical value at 5% significance level for this test is $\text{cv}_\alpha = t_{\frac{\alpha}{2}, n-2} = t_{0.025,28} = 2.048$. The test statistic is

  $$T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{\text{VAR}}[\beta_1]}} = \frac{0.2 - 0}{0.08} = 2.5$$

  We reject H_0 if $|T| > \text{cv}_\alpha$. Since $2.5 > 2.048$ we reject the null hypothesis at a 5% level of significance.

- The hypothesis tested is:

  $$\text{H}_0 : \beta_1 = 0.3 \quad \text{versus} \quad \text{H}_1 : \beta_1 < 0.3$$

  Since $n = 30$, and the parameters to estimate are two, the relevant critical value at 5% significance level for this test is $\text{cv}_\alpha = t_{\alpha, n-2} = t_{0.05,28} = 1.701$. The test statistic is

  $$T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{\text{VAR}}[\beta_1]}} = \frac{0.2 - 0.3}{0.08} = -1.25$$

  We reject H_0 if $T < -\text{cv}_\alpha$. Since $-1.25 < -1.701$ we fail to reject the null hypothesis at a 5% level of significance.