Introductory Econometrics

Solutions of Selected Exercises from Tutorial 1

Exercise 1.7 (*Random sample and sample mean.*) The population distribution of the number of teeth (x) has a mean of 20 with a variance of 100. Assume we draw (at random) a sample of 10 people, measure the value of x for each one of them (thus obtaining values $x_1, x_2, \dots, x_{10}$), and then calculate the arithmetic average $\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i$. Due to random sampling, $\bar{x}$ is a random variable.

a) What is the expected value of $\bar{x}$? What is its variance?

Solution: Intuitively, the expectation of the arithmetic average of 10 independent draws should be the same as the expectation of each one of them. Mathematically,

$$\mathsf{E}\bar{x} = \mathsf{E}\left[\tfrac{1}{10}\textstyle\sum_{i=1}^{10} x_i\right] = \tfrac{1}{10}\textstyle\sum_{i=1}^{10} \underbrace{\mathsf{E}x_i}_{=20} = 20.$$

b) (*Law of large numbers.*) Instead of 10 people, we take n now. What happens to $\mathsf{E}\bar{x}$ and $\mathsf{var}\bar{x}$ if we gradually raise n above all limits?

Solution: According to the *Law of large numbers*, the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed. Formally stated, in this case, $\bar{x} \to 20$ almost surely. This means that while $\mathsf{E}\bar{x}$ remains unaltered (i.e., 20), the variance gradually shrinks towards zero ($\mathsf{var}\bar{x} \to 0$).

c) (*Central limit theorem.*) Again, consider a sample of n people, only that now we study the result of

$$y = \sqrt{n}(\bar{x} - 20) = \frac{\sum_{i=1}^{n}(x_i - 20)}{\sqrt{n}}.$$

As n grows, what happens to the distribution of y?

Solution: The *Central limit theorem* (CLT) states that the arithmetic average of n independent draws from a population with mean μ and variance σ^2 will be approximately normally distributed with mean μ and variance $\frac{1}{n}\sigma^2$. More precisely, as n grows,

$$\sqrt{n}(\bar{x} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

where "$\xrightarrow{d}$" denotes convergence in distribution. The strength of the CLT consists in that normality is achieved regardless of the shape of the population's distribution (this is what makes the normal distribution "normal"). In our case, we have

$$\sqrt{n}(\bar{x} - 20) \xrightarrow{d} \mathcal{N}(0, 100).$$

Exercise 1.9 (*Correlation & covariance.*) I leave a and b up to you.

c) If we know that two RVs are negatively correlated, what does it tell us about their *covariance*?

Solution: Their covariance is negative.

d) What are the possible values of a covariance of two RVs?

Solution: Anything between $-\infty$ and ∞.

e) Let x and y be independent. Is it possible that $\mathsf{cov}(x, y) = 0.58$? Why?

Solution: No. If RVs are independent, they are automatically uncorrelated, i.e. $\mathsf{cov}(x, y) = 0$.

f) We know that $\mathsf{cov}(x, y) = 0$. Do x and y have to be independent? (If not, find an example of RVs that are *uncorrelated* despite *not* being *independent*.)

Solution: Let x be an RV with a uniform distribution on the interval $[-1, 1]$ and $y = x^2$. Then x and y are obviously not independent and still they are uncorrelated, i.e. $\mathsf{cov}(x, y) = 0$.

g) What are the possible values of a *correlation coefficient* of two RVs?

Solution: Anything between (and including) -1 and 1.

h) Which of the following can happen:

1) $\mathsf{corr}(x,y) = -1.56$.
2) $\mathsf{corr}(x,y) = 0.28, \quad \mathsf{cov}(x,y) = 0$.
3) $\mathsf{corr}(x,y) = 0.28, \quad \mathsf{cov}(x,y) = -0.5$.
4) $\mathsf{corr}(x,y) = 0.28, \quad \mathsf{cov}(x,y) = 0.5$.

Why? What is the relationship between the covariance and correlation coefficient of two RVs?

Solution: 1,2, and 3 cannot happen; 1 follows directly from *g*, 2 and 3 are easily seen from

$$\mathsf{corr}(x,y) = \frac{\mathsf{cov}(x,y)}{\sqrt{\mathsf{var}x \cdot \mathsf{var}y}}\,,$$

which tells us that cov and corr either have the same sign, or are both zero.

Note: Exercises 1.12, 1.13, and 1.14 focus on *conditional expectations* and *conditional variance*. For a brief treatment of the underlying theory, see Wooldridge's textbook, Appendix B, pages 684–688 (2nd edition).

Exercise 1.12 (*Calculations with conditional expectations.*)

a) Let x and y be *independent* RVs, $\mathsf{E}y = 12.5$. Find $\mathsf{E}[y|x]$.
   Solution: Because x and y are independent, $\mathsf{E}[y|x] = \mathsf{E}y$, i.e. $\mathsf{E}[y|x] = 12.5$.

b) Let x and y be RVs, $\mathsf{E}[y|x] = 2 + 5x$. Find $\mathsf{E}[4y + 3xy + x^2|x]$ and $\mathsf{E}[4y + 3xy + x^2|x = 5]$.
   Solution: We'll be using property CE.2 from Wooldridge, page 686:

$$\begin{aligned}\mathsf{E}[4y + 3xy + x^2|x] &= \mathsf{E}[(4+3x)y + x^2|x] \\ &\overset{\text{CE.2}}{=} (4+3x)\mathsf{E}[y|x] + x^2 \\ &= (4+3x)(2+5x) + x^2.\end{aligned}$$

   Altogether, we have
$$\mathsf{E}[4y + 3xy + x^2|x] = 16x^2 + 26x + 8.$$
   Substituting $x = 5$ into the right-hand side gives us
$$\mathsf{E}[4y + 3xy + x^2|x = 5] = 16\times 5^2 + 26\times 5 + 8 = 538.$$

Exercise 1.13 (*Conditional expectations II.*) Suppose that at a large university, college grade point average, *GPA*, and SAT score, *SAT*, are related by the conditional expectation

$$\mathsf{E}[GPA|SAT] = .70 + .002\,SAT. \tag{1}$$

a) Find the expected *GPA* when $SAT = 800$. Find $\mathsf{E}[GPA|SAT = 1{,}400]$. Comment on the difference.
   Solution: This should be fairly easy:

$$\begin{aligned}\mathsf{E}[GPA|SAT = 800] &= .70 + .002\times 800 &&= 1.67, \\ \mathsf{E}[GPA|SAT = 1{,}400] &= .70 + .002\times 1{,}400 &&= 3.50\,.\end{aligned}$$

b) If the average *SAT* in the university is 1,100, what is the average *GPA*?
   Solution: We'll be using property CE.4 from Wooldridge, page 687 (the so-called *Law of iterated expectations*):

$$\mathsf{E}\,GPA \overset{\text{CE.4}}{=} \mathsf{E}(\mathsf{E}[GPA|SAT]) \overset{(1)}{=} \mathsf{E}(.70 + .002\,SAT) = .70 + .002\,\mathsf{E}SAT = .70 + .002\times 1{,}100.$$

Exercise 1.14 (*Conditional variance.*) Do you think the variance of *wages* varies among groups of people with different levels of *education*? E.g., is there a difference between $\mathsf{var}[wage|educ = 9]$ and $\mathsf{var}[wage|educ = 18]$?

Solution: The obvious answer is *yes*. I leave the explanation up to you. The exercise is only meant as a practice of the "conditional notation".