Problem set 1

Introductory econometrics

Learning goals

- Practice algebraic derivations (without matrix notation)
- Law of iterated expectations
- Expected value and variance calculations
- Deriving the OLS estimator for a simple linear model

Question 1: Practicing algebra

Prove that

$$\sum_{i=1}^{n}(a_i - \bar{a})(b_i - \bar{b}) = \sum_{i=1}^{n}(a_i - \bar{a})b_i,$$

where $\bar{a}$ and $\bar{b}$ denote the sample mean defined as $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Question 2: Law of iterated expectations

This exercise is meant to practice the Law of iterated expectations (L.I.E.). The law states that

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y)).$$

Sometimes you may see it written as $\mathbb{E}(X) = \mathbb{E}_y(\mathbb{E}_x(X|Y))$.

Suppose we have data on $n = 1000$ Austrians. For simplicity assume that the data is not a sample but the whole population of interest[1] and we ask people to tell us how many Schnitzel they eat per month. These are the average quantities regarding Schnitzel that we find:

	Young	Old
Vienna	4	6
other states	0	4

That is, young people from Vienna eat on average 4 Schnitzel per month, young people from other states eat none, etc.

[1] I.e., we do not neeed to worry about the conceptual distinction between sample means and population means.

The frequencies (number of respondents in each category):

	Young	Old
Vienna	300	500
other states	50	150

a) Compute Pr[Vienna|young] and Pr[other state|young].

b) Compute Pr[Vienna|old] and Pr[other state|old].

c) Use the L.I.E. to compute the expected number of Schnitzel eaten, conditional on being young ($\mathbb{E}[S|\text{young}]$).

d) Use the L.I.E. to compute the expected number of Schnitzel eaten, conditional on being old ($\mathbb{E}[S|\text{old}]$).

e) Compute Pr[old], Pr[young].

f) Use the L.I.E. to compute the expected number of Schnitzel eaten $\mathbb{E}[S]$.

Question 3: Means and variance

Let Y_1, Y_2, Y_3, and Y_4 be independent, identically distributed random variables from a population with mean μ and variance σ^2. Let $\bar{Y} = \frac{1}{4}(Y_1 + Y_2 + Y_3 + Y_4)$ denote the average of these four random variables.

a) $\bar{Y}$ is a random variable too. What are the expected value and variance of $\bar{Y}$ in terms of μ and σ^2?

b) Now, consider a different estimator of μ:

$$W = \frac{1}{8}Y_1 + \frac{1}{8}Y_2 + \frac{1}{4}Y_3 + \frac{1}{2}Y_4.$$

This is an example of a weighted average of the Y_i. Show that W is also an unbiased estimator of μ. Find the variance of W.

c) Based on your answers to parts (a) and (b), which estimator of μ do you prefer, $\bar{Y}$ or W.

Question 4: Deriving the Ordinary Least Squares Estimators

Derive the OLS estimator $\hat{\beta}_0$, $\hat{\beta}_1$ for the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + e_i$, through one of these two alternative approaches:

1. Solve the minimization problem:

$$\arg\min_{\beta_0,\beta_1} \sum_{i=1}^{n} e_i$$

2. Replace the population means in these two assumptions by the respective sample mean and solve for β_0 and β_1:

$$(i) \quad \mathbb{E}[e_i] = 0, \qquad (ii) \quad \mathbb{E}[x_i e_i] = 0.$$