20210 Summer 2025: Problem Set 1 Agustín Esteva
Econometric aesteva@uchicago.edu
Instructor(s): Murilo Ramos Due Date: July 29, 2025

Problem 1

Stock and Watson, Exercises 4.5*, 4.9, 4.12, 5.2, 5.5*, 5.6*, E5.1*

(* indicates exercises that are recommended but not required for submission.)

(a) (4.9)

(i) A regression yields $\hat{\beta}_1 = 0$. Show that $R^2 = 0$.

SOLUTION: It suffices to show that ESS $= 0$. Computing,

$$\begin{aligned}\sum(\hat{y}_i - \bar{y})^2 &= \sum(\hat{\beta}_0 - \bar{y})^2 \\ &= \sum(\hat{\beta}_0 - (\hat{\beta}_0 + \hat{\beta}_1\bar{x}))^2 \\ &= 0\end{aligned}$$

♡

(ii) A regression yields $R^2 = 0$. Does this imply that $\hat{\beta}_1 = 0$?

SOLUTION: Yes, we have that ESS $= 0$ iff

$$\sum(\hat{Y}_i - \bar{Y})^2 = 0 \iff \forall i, \hat{Y}_i = \bar{Y} \iff \forall i,\ \hat{\beta}_0 + \hat{\beta}_1 X_i = \hat{\beta}_0 + \hat{\beta}_1\bar{X}$$

which happens iff $\hat{\beta}_1 = 0$ ♡

(b) (4.12)

(i) Show that $\hat{\text{Corr}}^2_{XY}(X, Y) = R^2$.

SOLUTION: Computing,

$$\begin{aligned}\hat{\text{Corr}}(X,Y)^2 &= \frac{\hat{\text{Cov}}(X,Y)^2}{\hat{\text{Var}}(X)\hat{\text{Var}}(Y)} \\ &= \frac{\frac{1}{n^2}\left(\sum(X_i - \bar{X})(Y_i - \bar{Y})\right)^2}{\frac{1}{n}\sum(X_i - \bar{X})^2\frac{1}{n}\sum(Y_i - \bar{Y})^2} \\ &= \left(\frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}\right)^2\frac{\sum(X_i - \bar{X})^2}{\sum(Y_i - \bar{Y})^2}\end{aligned}$$

$$
\begin{aligned}
&= \frac{\hat{\beta}_1^2 \sum (X_i - \bar{X})^2}{\sum (Y_i - \bar{Y})^2} \\
&= \frac{\sum (\bar{Y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 X_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \\
&= \frac{\sum (\hat{\beta}_0 + \hat{\beta}_1 X_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \\
&= \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \\
&= R^2
\end{aligned}
$$

♡

(ii) Show that R^2 from the regression Y from X is the same ass the R^2 from the regression X from Y.

SOLUTION: From part (i), we have that

$$
\begin{aligned}
R^2_{Y \propto X} &= \frac{\hat{\text{Cov}}(X, Y)^2}{\hat{\text{Var}}(X)\hat{\text{Var}}(Y)} \\
&= \frac{\hat{\text{Cov}}(Y, X)^2}{\hat{\text{Var}}(Y)\hat{\text{Var}}(X)} \\
&= R^2_{X \propto Y}
\end{aligned}
$$

♡

(iii) Show that $\hat{\beta}_1 = r_{XY}(\frac{\sigma_Y}{\sigma_X})$

SOLUTION: From the work in part (i), we see that

$$
r^2 = \frac{\hat{\beta}_1^2 \sum (X_i - \bar{X})^2}{\sum (Y_i - \bar{Y})^2} = \frac{\hat{\beta}_1 \frac{1}{n} \sum (X_i - \bar{X})^2}{\frac{1}{n} \sum (Y_i - \bar{Y})^2} = \hat{\beta}_1^2 \frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2}
$$

Taking roots of both sides and rearranging yields the result. ♡

(c) (5.2) Suppose a researcher, using wage data on 250 randomly selected male workers and 280 female workers, estimates the OLS regression

$$
\widehat{\text{Wage}} = 12.52 + 2.12 \times \text{Male}, \quad R^2 = 0.06, \quad SER = 4.2,
$$

$$
(0.23) \quad (0.36)
$$

where Wage is measured in dollars per hour and Male is a binary variable that is equal to 1 if the person is a male and 0 if the person is a female. Define the wage gender gap as the difference in mean earnings between men and women.

(i) What is the estimated gender gap?

SOLUTION: \$2.12 ♡

(ii) Is the estimated gender gap significantly different from 0? (Compute the p-value for testing the null hypothesis that there is no gender gap.)

SOLUTION: $H_0 : \beta_1 = 0$, $H_a : \beta_1 \neq 0$. We have that $\hat{\beta}_1 = 2.12$. Statistic is

$$T_{530} = \frac{\hat{\beta}_1 - \beta_1^{H_0}}{\mathrm{SE}(\hat{\beta}_1)} = \frac{2.12}{0.36} = 5.89$$

We approximate with normal because of the large sample size and get a $p-$value of

$$1 - 2\Phi(T) \approx 0$$

We reject the null and hell yeah gender gap yeahhhhhh. ♡

(iii) Construct a 95% confidence interval for the gender gap.

SOLUTION: Pretty sure we use $z_{\frac{\alpha}{2}} = 1.96$. Hence,

$$\mathbb{P}\{|T_n| \leq 1.96\} = 0.95 \implies -1.96 \leq \frac{\hat{\beta}_1 - \beta_1}{\mathrm{SE}(\hat{\beta}_1)} \leq 1.96$$

and so our confidence interval is

$$\left[\hat{\beta}_1 \pm 1.96 \cdot \mathrm{SE}(\hat{\beta}_1)\right] = [1.41, 2.82]$$

♡

(iv) In the sample, what is the mean wage of women? Of men?

SOLUTION: We know that $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ hence,

$$\bar{Y}_W = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_W = \hat{\beta}_0 = \$12.52$$

For men,

$$\bar{Y}_M = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_M = \$14.64$$

♡

(v) Another researcher uses these same data but regresses Wages on Female, a variable that is equal to 1 if the person is female and 0 if the person a male. What are the regression estimates calculated from this regression?

Solution:

$$\widehat{\text{Wage}} = 14.64 + (-2.12) \times \text{Female}, \quad R^2 = 0.06, \quad SER = 4.2.$$

♡

Problem 2

Prove the following result:

$$\mathbb{E}[X(Y-\mathbb{E}[Y])] = \mathbb{E}[(X-\mathbb{E}[X])Y] = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$$

Solution: We just open up parenthesis:

$$\begin{aligned}\mathbb{E}[X(Y-\mathbb{E}[Y])] &= \mathbb{E}[XY - X\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

We see the first equality now:

$$\begin{aligned}\mathbb{E}[(X-\mathbb{E}[X])Y] &= \mathbb{E}[XY - \mathbb{E}[X]Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

For the last equality, we compute

$$\begin{aligned}\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] &= \mathbb{E}[XY - \mathbb{E}[X]Y - \mathbb{E}[Y]X + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

♡

Problem 3

Now let's practice the proof of the expressions for the OLS coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ in the simple linear regression case. Assume that

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

in the population, and that

$$N\hat{\sigma}_x^2 = \sum (X_i - \bar{X})^2 > 0$$

Your data consists of the sequence of observable vectors (X_i, Y_i) for $i = 1, \ldots, N$, collected as an i.i.d. sample from the joint distribution of $(X, Y)N$.

(a) State the minimization problem (in the sample).

SOLUTION:

$$\min_{b_0,b_1} \mathbb{E}[Y - (b_0 + b_1 X)^2] \implies (\hat{\beta}_0, \hat{\beta}_1) = \min_{b_0,b_1} \frac{1}{N} \sum_{n=1}^{N} (Y_i - b_0 - b_1 X_i)^2 =: \min_{(b_0,b_1)\in\mathbb{R}^2} S(b_0, b_1)$$

♡

(b) Derive the two first order conditions (step by step).

SOLUTION: FOC b_0

$$0 = \frac{\partial S(b_0, b_1)}{\partial b_0} = -\frac{2}{N} \sum_{n=1}^{N} (Y_i - b_0 - b_1 X_i)$$

$$0 = \sum_{n=1}^{N} Y_i - N b_0 - b_1 \sum_{n=1}^{N} X_i$$

$$0 = \overline{Y} - b_0 - b_1 \overline{X} \tag{1}$$

FOC b_1

$$0 = \frac{\partial S(b_0, b_1)}{\partial b_1} = -\frac{2}{N} \sum_{n=1}^{N} X_i (Y_i - b_0 - b_1 X_i)$$

$$0 = \overline{XY} - b_0 \overline{X} - b_1 \overline{X^{(2)}} \tag{2}$$

♡

(c) Solve the f.o.c. of $\hat{\beta}_0$ for $\hat{\beta}_0$ as a function of the observables and $\hat{\beta}_1$.

SOLUTION: Clearly, we rearrange (1) to see that

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

♡

(d) Now plug the expression found in (c) into the f.o.c. of $\hat{\beta}_1$ to solve for $\hat{\beta}_1$ as a function of (only) observables.

SOLUTION: We multiply (1) by $\overline{X}$ to find

$$0 = \overline{X}\,\overline{Y} - b_0 \overline{X} - b_1 \overline{X}^2$$

Subtracting this from (2), we find that

$$0 = \overline{XY} - \overline{XY} - b_1(\overline{X^{(2)}} - \overline{X}^2)$$

Thus, we see that

$$b_1 = \frac{\overline{XY} - \overline{X}\,\overline{Y}}{\overline{X^{(2)}} - \overline{X}^2}$$

We see that

$$\frac{\sum_{n=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{n=1}^{N}(X_i - \overline{X})^2} = \frac{\sum_{n=1}^{N} X_i Y_i - \overline{X}\sum Y_i - \overline{Y}\sum X_i + N\overline{X}\,\overline{Y}}{\sum_{n=1}^{N}(X_i - \overline{X})X_i} = b_1$$

as in class. Hence,

$$\hat{\beta}_1 = \frac{\overline{XY} - \overline{X}\,\overline{Y}}{\overline{X^{(2)}} - \overline{X}^2}$$

♡

Hint: Your final expressions should be

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}, \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

Problem 4

Suppose that

$$\text{Col GPA} = \beta_0 + \beta_1 \text{PC} + U$$

where `Col GPA` denotes a student's college GPA and `PC` is a binary variable equal to 1 if the student owns a PC and 0 otherwise.

Define `noPC` as a dummy variable for whether the student does **not** own a PC, with `noPC` = 1 if the student does **not** own a PC and 0 otherwise.

Suppose that:

- There are 34 students who do not own a PC.
- There are 53 students who do own a PC.
- The sample average of Col GPA for those without a PC is 2.5.
- The sample average of Col GPA for those with a PC is 3.5.
- The sample standard deviation of Col GPA for those without a PC is 0.62.
- The sample variance of Col GPA for those with a PC is 0.47.

(a) Prove that the OLS estimators for β_0 and β_1 are $\bar{Y}_0$ and $\bar{Y}_1 - \bar{Y}_0$, respectively. ($\bar{Y}_0$ and $\bar{Y}_1$ are the sample means of the outcome when PC = 0 and PC = 1.)

SOLUTION: Suppose $\hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0$, then

$$\begin{aligned}
\hat{\beta}_0 &= \bar{Y} - (\bar{Y}_1 - \bar{Y}_0)\bar{X} \\
&= \frac{1}{n}\sum Y_i - \Big(\frac{1}{n_1}\sum_{i:X_i=1} Y_i - \frac{1}{n_0}\sum_{i:X_i=0} Y_i\Big)\frac{n_1}{n} \\
&= \frac{1}{n}\sum_{i:X_i=0} Y_i + \frac{(n-n_0)}{n_0}\frac{1}{n}\sum_{i:X_i=0} Y_i \\
&= \frac{1}{n}\sum_{i:X_i=0} Y_i + \frac{1}{n_0}\sum_{i:X_i=0} Y_i - \frac{1}{n}\sum_{i:X_i=0} Y_i \\
&= \bar{Y}_0
\end{aligned}$$

Now to prove $\hat{\beta}_1$, we see that

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum X_i Y_i - \frac{1}{n}\sum X_i \sum Y_i}{\sum (X_i - \bar{X})^2} \\
&= \frac{\sum_{i:X_i=1} Y_i - \frac{n_1}{n} Y_i}{n_1(1-\frac{n_1}{n})^2 + n_0(\frac{n_1}{n})^2} \\
&= \frac{n_1\left(\bar{Y}_1 - \bar{Y}\right)}{\frac{n_0 n_1}{n}}
\end{aligned}$$

$$= \frac{n}{n_0}(\bar{Y}_1 - (\frac{n_1}{n}\bar{Y}_1 + \frac{n_0}{n}\bar{Y}_0))$$
$$= \bar{Y}_1 - \bar{Y}_0$$

♡

(b) Test $H_0 : \beta_1 \leq 0$ versus $H_1 : \beta_1 > 0$ at the 5% significance level. What is the p-value for this hypothesis test? What do you conclude?

SOLUTION: Call

$$V = \frac{1}{n}\sum(X_i - \bar{X})^2$$

We know from class that under the null,

$$T = \frac{\hat{\beta}_1}{\hat{\sigma}} \sim N(0, 1)$$

We know that $n = 83$, $\hat{\sigma}^2 = \frac{\hat{\sigma}_0^2}{n_0} + \frac{\hat{\sigma}_1^2}{n_1}$. Thus,

$$T = 7.042$$

yielding a $p-$value of 0. We reject the null. ♡

Hint: Use the test of difference of means between groups.

(c) Suppose that the researcher instead wishes to do an OLS regression of `Col GPA` on `noPC`. What would be the resulting OLS estimates for that regression?

SOLUTION: It would be

$$\texttt{Col}\hat{\ }\texttt{GPA} = 3.5 + (-1)\texttt{No PC}$$

and thus

$$\hat{\beta}_1 = \bar{Y}_1$$

and

$$\hat{\beta}_0 = \bar{Y}_0 - \bar{Y}_1$$

♡