Problem Set 3

ECON 21020 Spring, 2021

Jake Underland, Groupmates: Ian Bamford, Matthew Chen

2021-04-22

Contents

Question 1

7.

(i)

```r
w_b <- c(8.3, 9.4, 9.0, 10.5, 11.4, 8.75, 10.0, 9.5, 10.8, 12.55, 12.00, 8.65, 7.75, 11.25, 12.65)
w_a <- c(9.25, 9.0, 9.25, 10.0, 12.0, 9.5, 10.25, 9.5, 11.5, 13.1, 11.5, 9.0, 7.75, 11.5, 13.0)
d <- w_a - w_b
df <- data.frame(w_b, w_a, d)
kable(df, "simple")
```

w_b	w_a	d
8.30	9.25	0.95
9.40	9.00	-0.40
9.00	9.25	0.25
10.50	10.00	-0.50
11.40	12.00	0.60
8.75	9.50	0.75
10.00	10.25	0.25
9.50	9.50	0.00

w_b	w_a	d
10.80	11.50	0.70
12.55	13.10	0.55
12.00	11.50	-0.50
8.65	9.00	0.35
7.75	7.75	0.00
11.25	11.50	0.25
12.65	13.00	0.35

```
test <- t.test(d, conf.level = 0.95)
test$conf.int
```

```
## [1] -0.009684686  0.489684686
## attr(,"conf.level")
## [1] 0.95
```

(ii)

$$H_0 : \quad \mu = 0$$
$$H_1 : \quad \mu > 0$$

(iii)

```
x_bar <- mean(d)
s <- sd(d)
n <- length(d)
test_stat <- sqrt(n) * (x_bar - 0) / s
test_stat > qt(0.95, df = 14)
```

```
## [1] TRUE
```

```
test_stat > qt(0.99, df = 14)
```

```
## [1] FALSE
```

```
cat("The test statistic is ", test_stat, "\n",
    "At 5%, this is greater than the 0.95 quantile of the t distribution", "\n", "at df = 14: ",
    qt(0.95, df = 14), " and H_0 is rejected.", "\n",
    "At 1%, this is less than the 0.99 quantile of the t distribution", "\n", "at df = 14: ",
    qt(0.99, df = 14), " and H_1 is rejected.")
```

```
## The test statistic is  2.061595
##  At 5%, this is greater than the 0.95 quantile of the t distribution
##  at df = 14:  1.76131  and H_0 is rejected.
##  At 1%, this is less than the 0.99 quantile of the t distribution
##  at df = 14:  2.624494  and H_1 is rejected.
```

(iv)

```
test <- t.test(d, mu = 0, alternative = "greater")
test$p.value
```

```
## [1] 0.02916138
```

9.

(i)

X follows a Binomial distribution of probability $p = 0.65$, assuming that is the true value of p. Thus, the expected value of X is np.

$$E(X) = np = 200 \cdot 0.65 = 130$$

(ii)

Let x_i denote a Bernoulli trial with probability $p = 0.65$. The variance of X is:

$$\begin{aligned} Var(X) &= Var(\Sigma_{i=1}^{n} x_i) \\ &= \Sigma_{i=1}^{n} Var(x_i) \\ &= np(1-p) \\ &= 200 \cdot 0.65 \cdot 0.35 \\ &= 45.5 \end{aligned}$$

Thus, standard deviation is $\sqrt{45.5} \approx 6.75$

(iii)

We already showed in Problem Set 2 that when $x_1, \dots, x_n$ are iid with $x_i \sim Bernoulli(p)$, then $X_n = \Sigma_{i=1}^{n} x_i \overset{a}{\sim} \mathcal{N}(np, np(1-p))$. Thus, the probability that you would find 115 or fewer yes votes from a random sample of 200 is:

```r
pnorm(115, mean = 130, sd = 6.75, lower.tail=TRUE)
```

```
## [1] 0.01313415
```

(iv)

If what the dictator claims is true, that 65% of the population supported them, then the probability that we would see only 115 yes votes from a random sample of 200 votes is 1.3%. This indicates that the dictator is likely not being transparent about the true rate of support they receive from the public.

Question 2

a) Show that in the linear model $y_i = \beta_0 + \beta_1 x_i + u_i$ where the parameters (β_0, β_1) are estimated using ordinary least squares:

$$R^2 = \beta_1^2 \frac{TSS_X}{TSS_Y} = (\hat{\rho}_{XY})^2,$$

where

$$\begin{aligned} \hat{\rho}_{XY} &= \frac{\Sigma_{i=1}^{n}(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sqrt{TSS_X \times TSS_Y}} \\ TSS_X &= \Sigma_{i=1}^{n}(X_i - \bar{X}_n)^2 \\ TSS_Y &= \Sigma_{i=1}^{n}(Y_i - \bar{Y}_n)^2 \end{aligned}$$

ρ_{XY} is the sample correlation coefficient. Interpret your result.

Solution.

First, we know that

$$\begin{aligned}
\hat{\beta_0} &= \bar{Y_n} - \hat{\beta_1}\bar{X}_n \\
\hat{\beta_1} &= \frac{\Sigma(Y_i - \bar{Y}_n)(X_i - \bar{X}_n)}{\Sigma(X_i - \bar{X}_n)^2} \\
\hat{Y}_n &= \hat{\beta_0} + \hat{\beta_1}X_i
\end{aligned}$$

Using these expressions, we derive the following:

$$\begin{aligned}
R^2 &= \frac{ESS}{TSS_Y} \\
&= \frac{\Sigma(\hat{Y}_i - \bar{Y}_n)^2}{\Sigma(Y_i - \bar{Y}_n)^2} \\
&= \frac{\Sigma(\hat{\beta_0} + \hat{\beta_1}X_i - \bar{Y}_n)^2}{\Sigma(Y_i - \bar{Y}_n)^2} \\
&= \frac{\Sigma(\bar{Y_n} - \hat{\beta_1}\bar{X}_n + \hat{\beta_1}X_i - \bar{Y}_n)^2}{\Sigma(Y_i - \bar{Y}_n)^2} \\
&= \frac{\Sigma[\hat{\beta_1}(X_i - \bar{X}_n)]^2}{\Sigma(Y_i - \bar{Y}_n)^2} \\
&= \hat{\beta_1^2}\frac{\Sigma_{i=1}^n(X_i - \bar{X}_n)^2}{\Sigma_{i=1}^n(Y_i - \bar{Y}_n)^2} \\
&= \hat{\beta_1}^2\frac{TSS_X}{TSS_Y}
\end{aligned}$$

Now, by substituting our expression for $\hat{\beta_1}$, we get

$$\begin{aligned}
\hat{\beta_1}^2\frac{TSS_X}{TSS_Y} &= \frac{[\Sigma(Y_i - \bar{Y}_n)(X_i - \bar{X}_n)]^2}{[\Sigma(X_i - \bar{X}_n)^2]^2} \cdot \frac{TSS_X}{TSS_Y} \\
&= \frac{[\Sigma(Y_i - \bar{Y}_n)(X_i - \bar{X}_n)]^2}{TSS_X TSS_Y} \\
&= (\hat{\rho}_{XY})^2
\end{aligned}$$

Since R^2 tells us the fraction of variability in Y explained by X, it is natural to assume that the higher the correlation of the two variables, the higher the explanatory value of X and thus the higher the value of R^2. If the two are perfectly correlated, then Y is a linear function of X, and so naturally there would exist a linear function of X that perfectly explains Y, yielding an R^2 score of 1.

b) Suppose you run the 'reverse' regression $x_i = \gamma_0 + \gamma_1 y_i + \epsilon_i$ and obtain the OLS estimates $\hat{\gamma_0}$, $\hat{\gamma_1}$. When is it true that $\hat{\gamma_1} = 1/\hat{\beta_1}$?

Solution. We can rearrange $\hat{\beta_1}$ such that:

$$\hat{\beta_1} = \frac{\Sigma(Y_i - \bar{Y}_n)(X_i - \bar{X}_n)}{\Sigma(X_i - \bar{X}_n)^2} = \hat{\rho}_{XY}\sqrt{\frac{\hat{Var}(Y)}{\hat{Var}(X)}}$$

Then, if we regress x_i on y_i as $x_i = \gamma_0 + \gamma_1 y_i + \epsilon_i$, we similarly get

$$\hat{\gamma_1} = \frac{\Sigma(Y_i - \bar{Y}_n)(X_i - \bar{X}_n)}{\Sigma(Y_i - \bar{Y}_n)^2} = \hat{\rho}_{XY}\sqrt{\frac{\hat{Var}(X)}{\hat{Var}(Y)}}$$

Thus,

$$
\begin{aligned}
\hat{\gamma}_1 = \hat{\rho}_{XY}\sqrt{\frac{\hat{Var}(X)}{\hat{Var}(Y)}} &= \frac{1}{\hat{\rho}_{XY}}\sqrt{\frac{\hat{Var}(X)}{\hat{Var}(Y)}} = \frac{1}{\hat{\beta}_1} \\
\implies \hat{\rho}_{XY} &= \frac{1}{\hat{\rho}_{XY}} \\
\implies \hat{\rho}_{XY} &= \pm 1
\end{aligned}
$$

Question 3

2.3

(i)

```
# replicating dataframe
student <- c(1:8)
gpa <- c(2.8, 3.4, 3.0, 3.5, 3.6, 3.0, 2.7, 3.7)
act <- c(21, 24, 26, 27, 29, 25, 25, 30)
df2 <- data.frame(student, gpa, act)
```

```
kable(df2, "simple")
```

student	gpa	act
1	2.8	21
2	3.4	24
3	3.0	26
4	3.5	27
5	3.6	29
6	3.0	25
7	2.7	25
8	3.7	30

```
# computing OLS estimate of coefficients
model <- lm(gpa ~ act, data = df2)
model$coefficients
```

```
## (Intercept)         act
##   0.5681319   0.1021978
```

Thus, the relationship can be expressed as

$$\widehat{GPA} = .57 + .10ACT$$

The above equation shows that there is a small but positive correlation between GPA and ACT scores. However, the direction of the causation cannot be established just by looking at a simple linear regression. The intercept shows the expected value of GPA when a student's ACT score is 0. However, in this dataset, it is unrealistic to assume that a student's ACT score can be 0. We can obtain a much more useful intercept by centering the ACT scores at their mean so that the intercept reflects the value of GPA for an average student.

```
# centering ACT scores and computing coefficients
df2_centered <- df2 %>%
  mutate(act = act - mean(act))
model_centered <- lm(gpa ~ act, data = df2_centered)
model_centered$coefficients
```

```
## (Intercept)         act
##   3.2125000   0.1021978
```

The above shows that when ACT takes its mean value, GPA is expected to be 3.2.

When ACT score is increased by 5 points, GPA is expected to increase by the coefficient of ACT in the estimated model times 5.

```
0.1021978 * 5
```

```
## [1] 0.510989
```

(ii)

```
# creating column with fitted values and residuals
df2 <- df2 %>%
  mutate(fitted = 0.5681319 + 0.1021978 * act,
    residuals = gpa - fitted)
```

```
kable(df2, "simple")
```

student	gpa	act	fitted	residuals
1	2.8	21	2.714286	0.0857143
2	3.4	24	3.020879	0.3791209
3	3.0	26	3.225275	-0.2252747
4	3.5	27	3.327472	0.1725275
5	3.6	29	3.531868	0.0681319
6	3.0	25	3.123077	-0.1230769
7	2.7	25	3.123077	-0.4230769
8	3.7	30	3.634066	0.0659341

```
# sum of residuals
sum(df2$residuals)
```

```
## [1] 2e-07
```

(iii)

By substituting $ACT = 20$ into the equation derived in (i), we have

```
gpa_20 = .57 + .10 * 20
cat("Predicted value of GPA when ACT = 20 is:", gpa_20)
```

```
## Predicted value of GPA when ACT = 20 is: 2.57
```

(iv)

```r
# R2 of the OLS model
stargazer(model, type = "latex", header = FALSE)
```

Table 4:

	Dependent variable:
	gpa
act	0.102**
	(0.036)
Constant	0.568
	(0.928)
Observations	8
R^2	0.577
Adjusted R^2	0.507
Residual Std. Error	0.269 (df = 6)
F Statistic	8.199** (df = 1; 6)
Note:	*p<0.1; **p<0.05; ***p<0.01

As we see above, the R^2 value of the regression we ran is 0.577. This indicates that the variation in GPA explained by ACT is around 57.7%.

C2

```r
data("ceosal2")
```

(i)

```r
cat(" Average salary is:", mean(ceosal2$salary), "thousand dollars \n",
    "Average tenure is:", mean(ceosal2$ceoten), "years")
```

```
##  Average salary is: 865.8644 thousand dollars
##  Average tenure is: 7.954802 years
```

(ii)

```r
# filtering data to first year CEOs
ceo_first <- ceosal2 %>%
  filter(ceoten == 0)
cat("number of first year CEOs:", nrow(ceo_first))
```

```
## number of first year CEOs: 5
```

```r
# Longest tenure as CEO
max(ceosal2$ceoten)
```

```
## [1] 37
```

(iii)

```r
# estimating the OLS regression
model2 <- lm(log(salary) ~ ceoten, data = ceosal2)
stargazer(model2, header = FALSE, type = "latex")
```

Table 5:

	Dependent variable:
	log(salary)
ceoten	0.010
	(0.006)
Constant	6.505^{***}
	(0.068)
Observations	177
R^2	0.013
Adjusted R^2	0.008
Residual Std. Error	0.604 (df = 175)
F Statistic	2.334 (df = 1; 175)
Note:	*p<0.1; **p<0.05; ***p<0.01

The approximate predicted percentage increase in salary given one more year as a CEO is 1%.

C3

```r
# loading data
data(sleep75)
```

(i)

```r
# Estimate model
sleepmodel <- lm(sleep ~ totwrk, data=sleep75)
stargazer(sleepmodel, type = "latex", header = FALSE)
```

```r
# Results found in Table 6
```

$$sleep = 3,586.4 - .151\, totwrk$$

Where observations = 706, $R^2 = 0.103$.
The intercept in this equation shows how much sleep in a week a person would get if they worked 0 minutes that week.

(ii)

Since 2 hours is 120 minutes, the estimated effect on *sleep* is

```r
120 * -0.151
```

```
## [1] -18.12
```

Table 6:

	Dependent variable:
	sleep
totwrk	-0.151^{***}
	(0.017)
Constant	$3{,}586.377^{***}$
	(38.912)
Observations	706
R^2	0.103
Adjusted R^2	0.102
Residual Std. Error	421.136 (df = 704)
F Statistic	81.090^{***} (df = 1; 704)
Note:	*p<0.1; **p<0.05; ***p<0.01

so 18 minutes less sleep that week. This is a pretty sizable drop. If we consider 2 hours of overtime per day, which is not atypical, then that would result in almost 2 hours less sleep throughout the week, 18 minutes per night, which is large if you're a person who highly values sleep, like me.

Question 4

Suppose you have an iid sample of observations $\{y_i, x_i\}_{i=1}^n$, where y and x are random variables.

a) You wish to find the best predictor of y given x using only the functions $\{f : f(x) = bx \text{ for some } b \in \mathbb{R}\}$. You write the model

$$y_i = \beta x_i + u_i$$

where β is the optimal choice of b in terms of mean squared prediction error:

$$MSE(b) = E[(y - bx)^2].$$

State the minimization problem to be solved and show that

$$\beta = \frac{E(xy)}{E(x^2)}.$$

Argue that $E(xu) = 0$ as a consequence of selecting β in this manner. What is the ordinary least squares estimator of β?

Solution. The minimization problem:

$$\min_b E[(y - bx)^2]$$

FOCs:

$$\begin{aligned}
\{\beta\} : \ & \frac{d}{d\beta} E[(y - \beta x)^2] = 0 \\
\implies & -2E[x(y - \beta x)] = 0 \\
\implies & -2E[xy] + 2\beta E[x^2] = 0 \\
\implies & \beta = \frac{E[xy]}{E[x^2]}
\end{aligned}$$

Since $u \equiv y - \beta x$, $y = \beta x + u$. Plug this into the FOC to yield

$$
\begin{aligned}
& -2E[x(\beta x + u)] + 2\beta E[x^2] = 0 \\
\implies & -2E[\beta x^2 + ux] + 2\beta E[x^2] = 0 \\
\implies & -2\beta E[x^2] + E[ux] + 2\beta E[x^2] = 0 \\
\implies & E[ux] = 0
\end{aligned}
$$

The OLS estimator of β is the solution to the analogous problem:

$$\min_b \frac{1}{n}\Sigma(y_i - bx_i)^2$$

FOCs:

$$
\begin{aligned}
\{\hat{\beta}\}: \quad & \frac{d}{d\hat{\beta}}\frac{1}{n}\Sigma(y_i - \hat{\beta}x_i)^2 = 0 \\
\implies & -2\frac{1}{n}\Sigma x_i(y_i - \hat{\beta}x_i) = 0 \\
\implies & -2\frac{1}{n}\Sigma[x_i y_i] + 2\hat{\beta}\frac{1}{n}\Sigma[x_i^2] = 0 \\
\implies & \hat{\beta} = \frac{\Sigma[x_i y_i]}{\Sigma[x_i^2]}
\end{aligned}
$$

b) Now suppose you are willing to assume that $E(u|x) = 0$.

i. Is $E(u|x) = 0$ stronger or weaker than assuming$E(ux) = 0$? Is βx still the best predictor of y in the class of functions

$$F^* = \{f : f(x) = bx \text{ for some } b \in \mathbb{R}\}$$

under this new assumption?

Solution.

It is stronger. Let us assume $E(u|x) = 0$. Then, from the law of iterated expectations,

$$E(ux) = E[E[ux|x]] = E[xE[u|x]] = E[x \cdot 0] = 0$$

Thus, $E(u|x) = 0 \implies E(ux) = 0$. Now, from this assumption we know that

$$E(y|x) = E(\beta x + u|x) = \beta x + E[u|x] = \beta x$$

Thus, the problem

$$\min_\beta E[(y - E(y|x))^2] = \min_\beta E[(y - \beta x)^2]$$

Since we know $E(y|x)$ is the best predictor of y given x, βx is still the best predictor of y.

ii. Show that β can be represented in the following ways:

$$\beta = \frac{E(xy)}{E(x^2)} \text{ and } \frac{E(y)}{E(x)}$$

Solution. Note that $u = y - \beta x$. From the iterated law of expectation,

$$E[u] = E[E[u|x]] = E[0] = 0$$

Then,

$$
\begin{aligned}
& E[y - \beta x] = 0 \\
\implies & E[y] - \beta E[x] = 0 \\
\implies & \beta = \frac{E(y)}{E(x)}
\end{aligned}
$$

Furthermore, we also know from i. that $E[ux] = 0$. Thus,

$$
\begin{aligned}
E[x(y - \beta x)] &= 0 \\
\implies E[xy] - \beta E[x^2] &= 0 \\
\implies \beta &= \frac{E(xy)}{E(x^2)}
\end{aligned}
$$

iii. Use your answer to part b)ii and the sample analogue principle to construct two estimators of β. Are they unbiased? Justify your answer.

Solution.

Using the sample analogue principle,

$$
\begin{aligned}
\hat{\beta} &= \frac{\frac{1}{n}\Sigma x_i y_i}{\frac{1}{n}\Sigma x_i^2} = \frac{\Sigma x_i y_i}{\Sigma x_i^2} \\
E[\hat{\beta}|x_1, \cdots, x_n] &= E[\frac{\Sigma x_i y_i}{\Sigma x_i^2}|x_1, \cdots, x_n] \\
&= \frac{\Sigma x_i E[y_i|x_1, \cdots, x_n]}{\Sigma x_i^2} \\
&= \frac{\Sigma x_i E[\beta x_i + u_i|x_1, \cdots, x_n]}{\Sigma x_i^2} \\
&= \frac{\beta \Sigma x_i^2 + \Sigma x_i E[u_i|x_1, \cdots, x_n]}{\Sigma x_i^2} \\
&= \frac{\beta \Sigma x_i^2}{\Sigma x_i^2} \\
&= \beta \\
&\implies E[\hat{\beta}] = E[E[\hat{\beta}|x_1, \cdots, x_n]] = \beta
\end{aligned}
$$

Thus, this first estimator is unbiased.

$$
\begin{aligned}
\hat{\beta} &= \frac{\frac{1}{n}\Sigma y_i}{\frac{1}{n}\Sigma x_i} = \frac{\Sigma y_i}{\Sigma x_i} \\
E[\hat{\beta}|x_1, \cdots, x_n] &= E[\frac{\Sigma y_i}{\Sigma x_i}|x_1, \cdots, x_n] \\
&= \frac{\Sigma E[\beta x_i + u_i|x_1, \cdots, x_n]}{\Sigma x_i} \\
&= \frac{\beta \Sigma x_i + \Sigma E[u_i|x_1, \cdots, x_n]}{\Sigma x_i} \\
&= \frac{\beta \Sigma x_i}{\Sigma x_i} \\
&= \beta \\
&\implies E[\hat{\beta}] = E[E[\hat{\beta}|x_1, \cdots, x_n]] = \beta
\end{aligned}
$$

Thus, the second estimator is also unbiased.

Question 5

Suppose you have a sample of observations $\{y_i, x_i\}_{i=1}^n$, where y and x are random variables. You write the model

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

where $E(u) = E(xu) = 0$ (i.e. $\beta_0 + \beta_1 x$ is the best linear predictor of y given x under square loss). Suppose you know that $\beta_0 = 2$. Derive the ordinary least squares estimator of β_1.

Solution.

The minimization problem to be solved is

$$\min_{b_1} \Sigma(y_i - 2 - b_1 x_i)^2$$

FOCs:

$$\begin{aligned}
\{\hat{\beta}_1\}: \ & \frac{d}{d\hat{\beta}_1}\Sigma(y_i - 2 - \hat{\beta}_1 x_i)^2 = 0 \\
\implies & -2\Sigma x_i(y_i - 2 - \hat{\beta}_1 x_i) = 0 \\
\implies & -2\Sigma[x_i y_i] + 4\Sigma x_i + 2\hat{\beta}_1 \Sigma[x_i^2] = 0 \\
\implies & \hat{\beta}_1 = \frac{\Sigma x_i(y_i - 2)}{\Sigma x_i^2}
\end{aligned}$$

Question 6

a) Derive the ordinary least squares estimators of β_0 and β_1 in the model

$$y_i = \beta_0 + \beta_1 x_i + u_i.$$

Solution.

The minimization problem is:

$$\min_{b_0, b_1} \frac{1}{n}\Sigma(y_i - b_0 - b_1 x_i)^2$$

FOCs:

$$\begin{aligned}
\{\hat{\beta}_0\}: \ & \frac{d}{d\hat{\beta}_0}\Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 0 \\
\implies & -2 \cdot \frac{1}{n}\Sigma y_i + 2\hat{\beta}_0 + 2\hat{\beta}_1 \frac{1}{n}\Sigma x_i = 0 \\
\implies & \hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_n \ldots \text{ where } \bar{y}_n = \frac{1}{n}\Sigma y_i,\ \bar{x}_n = \frac{1}{n}\Sigma x_i
\end{aligned}$$

$$\begin{aligned}
\{\hat{\beta}_1\}: \ & \frac{d}{d\hat{\beta}_1}\Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 0 \\
\implies & -2 \cdot \frac{1}{n}\Sigma x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\
\implies & -2\frac{1}{n}\Sigma x_i y_i + 2\hat{\beta}_0 \frac{1}{n}\Sigma x_i + 2\hat{\beta}_1 \frac{1}{n}\Sigma x_i^2 = 0 \\
\implies & -\frac{1}{n}\Sigma x_i y_i + \hat{\beta}_0 \bar{x}_n + \hat{\beta}_1 \frac{1}{n}\Sigma x_i^2 = 0 \\
\implies & -\frac{1}{n}\Sigma x_i y_i + \bar{x}_n \bar{y}_n - \hat{\beta}_1 \bar{x}_n^2 + \hat{\beta}_1 \frac{1}{n}\Sigma x_i^2 = 0 \\
\implies & \hat{\beta}_1 = \frac{\frac{1}{n}\Sigma x_i y_i - \bar{x}_n \bar{y}_n}{\frac{1}{n}\Sigma x_i^2 - \bar{x}_n^2} = \frac{\Sigma x_i(y_i - \bar{y}_n)}{\Sigma x_i(x_i - \bar{x}_n)}
\end{aligned}$$

b) You obtain the residuals $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$. If you now regress $\hat{u}_i$ on a constant and x_i, what will be the ordinary least squares estimates of the intercept and slope?

Solution.

We formulate the minimization problem to estimate the OLS estimates of the intercept and slope:

$$\min_{b_0,b_1} \frac{1}{n}\Sigma(\hat{u}_i - b_0 - b_1 x_i)^2$$

Since we know that $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$, we plug this into the above and rearrange

$$\begin{aligned}
&\min_{b_0,b_1} \frac{1}{n}\Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - b_0 - b_1 x_i)^2 \\
=&\min_{b_0,b_1} \frac{1}{n}\Sigma(y_i - \underbrace{(b_0 + \hat{\beta}_0)}_{B_0} - \underbrace{(b_1 + \hat{\beta}_1)}_{B_1} x_i)^2
\end{aligned}$$

This problem is exactly the same as the one we solved in a), and we already know that the best estimates of B_0 and B_1 are $\hat{\beta}_0$ and $\hat{\beta}_1$. Thus, the best estimators of b_0, b_1 in the above problem are those that satisfy $B_0 = b_0 + \hat{\beta}_0 = \hat{\beta}_0$ and $B_1 = b_1 + \hat{\beta}_1 = \hat{\beta}_1$, ergo, both 0. Hence, the intercept and slope of this regression will be 0.