계량경제학 HW 2차

2020-15709 나승환

Q1

(i) $60.5 + 4.28 \times 2 = 69.06$이므로
$E(Y|X=2) = 69.06$의 예측치이다.

(ii) OLS model에 따른 회귀직선이 $(\bar{X}, \bar{Y})$를 지나므로 $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$이다.
$\bar{Y} = 75$이므로 $\bar{X} = \frac{75 - 60.5}{4.28} \approx 3.3879$
따라서 TestScore 예측치 평균은 75이고, Hours의 평균은 3.3879이다.

(iii) 인과추론을 위한 OLS model의 기본 가정은
$\begin{pmatrix} E(u_i|X_i) = 0 \\ E(X^4) < \infty,\ E(Y^4) < \infty \\ (X_i, Y_i)\text{는 iid} \end{pmatrix}$이다.
해당 가정의 성립 여부를 파악하기에는 정보가 부족하지만, 해당 sample이 꼭 반드시 인과관계의 존재 여부를 판단 가능하다.
한편, $R^2 = 0.2$은 기술적으로 전체 변동 중 회귀직선에 의해 설명되는 부분이 21%라는 의미이고, 각 점추정량은 가정이 만족되는 것에 더해 등분산성까지 존재한다면 BLUE이다.
(가우스-마르코프 정리)

(iv, v) 두 문제는 본질적으로 같으므로 95% asymptotic CI를 우선 구하면 $4.28 \pm 1.96 \cdot 1.645$이므로
β_1의 95% 근사적 CI는 (1.0558, 7.5042)이다.
1이 해당 CI에 포함되지 않으므로, (iv)에서 $H_0: \beta_1 = 1$의 귀무가설을 기각한다.

Q2

(i) X, Y가 각각 기댓값이 μ_X, μ_Y로 존재하는 분포라 하자. (추가 가정)
그러면 $n \to \infty$일 때 $P\left(\left|\frac{\bar{X}_n}{\bar{X}_n} - \frac{\mu_Y}{\mu_X}\right| > \varepsilon\right) = 0$이고, $\frac{\bar{Y}}{\bar{X}}$는 $\frac{\mu_Y}{\mu_X}$로 확률수렴한다. (by LLN)

(ii) (X_i, Y_i)의 iid 가정이 성립하고 X, Y의 2nd moment가 존재하므로 CLT 성립을 위한 전제가 성립한다.
즉 $\sqrt{n}\left(\frac{\bar{Y}}{\mu_X} - \frac{\mu_Y}{\mu_X}\right)$는 $n \to \infty$일 때 $N\left(0, Var\left(\frac{\bar{Y}}{\mu_X}\right)\right) = N\left(0, \frac{\sigma^2}{n\mu_X^2}\right)$으로 분포수렴한다.

Econometrics HW2 coding part

Na SeungChan

Q3

You may continue on your previously collected dataset, or select another cross- sectional data. Let the variables of interest be Y and X1. The main research question is "Does X1 affect Y ? If so, how much?".

```
dataset <- read.csv('co2-emissions-vs-gdp.csv') %>%
  filter(Year == 2018) %>%
  filter(Code != '', Code != 'OWID_WRL')

dataset1 <- read.csv('energy.csv') %>%
  filter(Year == 2018) %>%
  filter(Code != '', Code != 'OWID_WRL')

full_data <- left_join(dataset, dataset1)
```

```
## Joining, by = c("Entity", "Code", "Year")
```

```
glimpse(full_data)
```

```
## Rows: 238
## Columns: 9
## $ Entity                                              <chr> "Afghanistan", "Alb~
## $ Code                                                <chr> "AFG", "ALB", "DZA"~
## $ Year                                                <int> 2018, 2018, 2018, 2~
## $ Annual.CO..emissions..per.capita.                   <dbl> 0.2948759, 1.732364~
## $ GDP.per.capita                                      <dbl> 1934.555, 11104.166~
## $ X417485.annotations                                 <chr> "", "", "", "", "",~
## $ Population..historical.estimates.                   <dbl> 36686788, 2877019, ~
## $ Continent                                           <chr> "", "", "", "", "",~
## $ Primary.energy.consumption.per.capita..kWh.person. <dbl> 1144.532, 14483.855~
```

Data form our world in data. Focused on countries in 2018 only. First dataset is from https://ourworldin data.org/grapher/co2-emissions-vs-gdp , Second dataset is from https://ourworldindata.org/energy

full_data are the merged data of above two.

(i)

Regress Y on X1 (and a constant, of course) and draw the fitted line on the scatter plot. Provide the standard errors for the point estimates. Interpret the result.

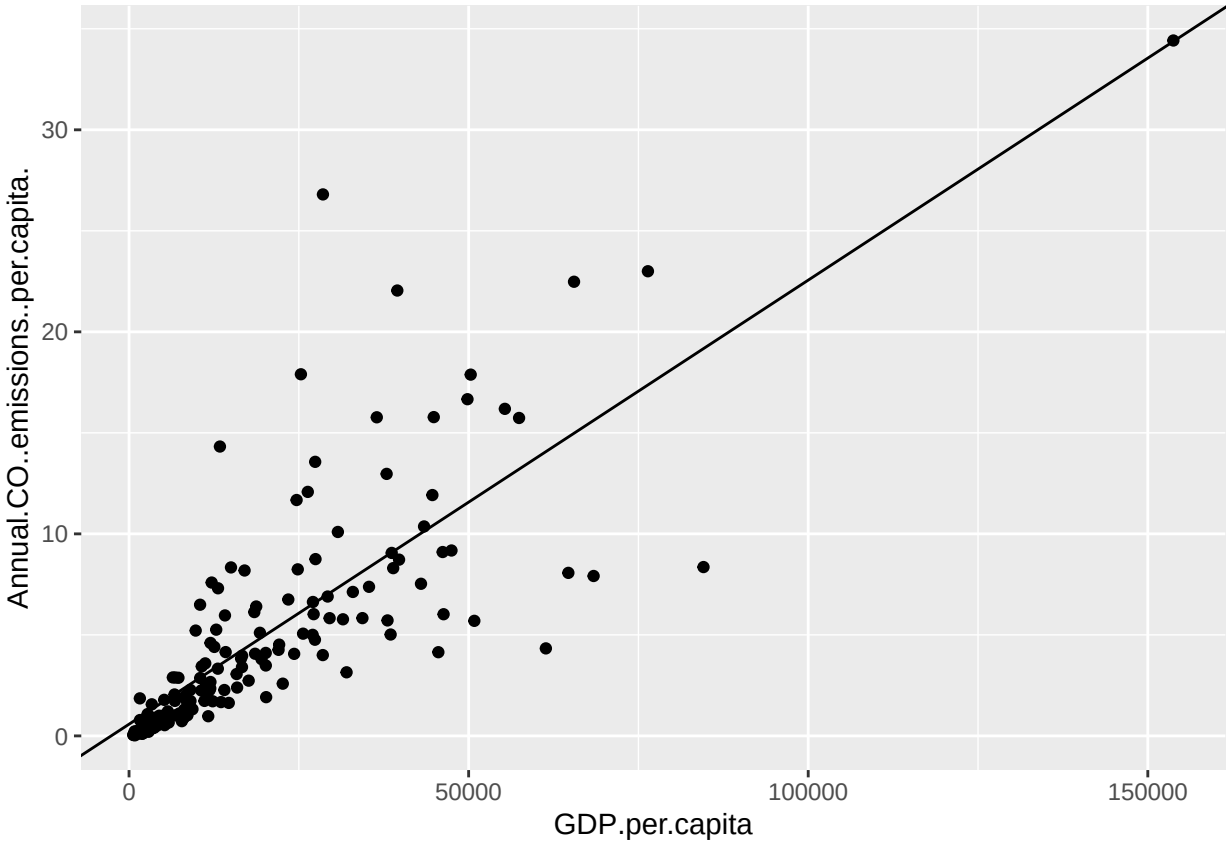
```
lm_a <- lm_robust(Annual.CO..emissions..per.capita. ~ GDP.per.capita, data = full_data)
summary(lm_a)
```

```
##
## Call:
## lm_robust(formula = Annual.CO..emissions..per.capita. ~ GDP.per.capita,
##     data = full_data)
```

```
##
## Standard error type:  HC2
##
## Coefficients:
##                  Estimate Std. Error t value  Pr(>|t|) CI Lower  CI Upper  DF
## (Intercept)    0.5783892  2.564e-01   2.256 2.544e-02 0.072006 1.0847728 162
## GDP.per.capita 0.0002198  1.863e-05  11.798 1.400e-23 0.000183 0.0002565 162
##
## Multiple R-squared:  0.6284 ,    Adjusted R-squared:  0.6262
## F-statistic: 139.2 on 1 and 162 DF,  p-value: < 2.2e-16
```

```
ggplot(data = full_data, aes(x = GDP.per.capita, y = Annual.CO..emissions..per.capita.)) +
  geom_point() +
  geom_abline(intercept = 0.5783892, slope = 0.0002198)
```

```
## Warning: Removed 74 rows containing missing values (geom_point).
```



Research Question is about the causal correlation between the GDP per capita and annual CO2 emission per capita.

se(intercept) : 2.564e-01, se(slope) : 1.863e-05

p-value for slope Pr(>|t|) = <2.2e-16 ***. This means the significance of the regression model.

(ii)

Add additional variables X2, ..., Xk that may affect Y and estimate the multiple regression model. Compare the coefficient estimate of X1 with (i). Note that the main research question is unchanged. The additional variables are supposed to eliminate possible omitted variable bias in the simple regression.

```
lm_b <- lm_robust(Annual.CO..emissions..per.capita. ~ GDP.per.capita + Primary.energy.consumption.per.ca
summary(lm_b)
```

```
##
## Call:
## lm_robust(formula = Annual.CO..emissions..per.capita. ~ GDP.per.capita +
##     Primary.energy.consumption.per.capita..kWh.person., data = full_data)
##
## Standard error type:  HC2
##
## Coefficients:
##                                                  Estimate Std. Error t value
## (Intercept)                                     6.754e-01  2.466e-01   2.739
## GDP.per.capita                                  6.932e-05  5.621e-05   1.233
## Primary.energy.consumption.per.capita..kWh.person. 1.017e-04  4.283e-05   2.375
##                                                 Pr(>|t|)   CI Lower
## (Intercept)                                      0.00686  1.884e-01
## GDP.per.capita                                   0.21928 -4.168e-05
## Primary.energy.consumption.per.capita..kWh.person.  0.01873  1.714e-05
##                                                  CI Upper  DF
## (Intercept)                                     1.1624382 161
## GDP.per.capita                                  0.0001803 161
## Primary.energy.consumption.per.capita..kWh.person. 0.0001863 161
##
## Multiple R-squared:  0.7447 ,    Adjusted R-squared:  0.7416
## F-statistic: 75.44 on 2 and 161 DF,  p-value: < 2.2e-16
```

Now I added X2 as 'energy use per capita.' This can be helpful because some of developed countries may use more non-fossil fuel energy.

p-value for 'energy use per capita' is smaller, and adjusted R^2 is bigger for the second model. But, 'the usage of eco-frendily energy' may be more helpful for this purpose, but making data like that was seriouly hard data wrangling(keys are not consistent). Making those data can improve the quality of this research.