

# 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 from our world in data. Focused on countries in 2018 only. First dataset is from <https://ourworldindata.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.

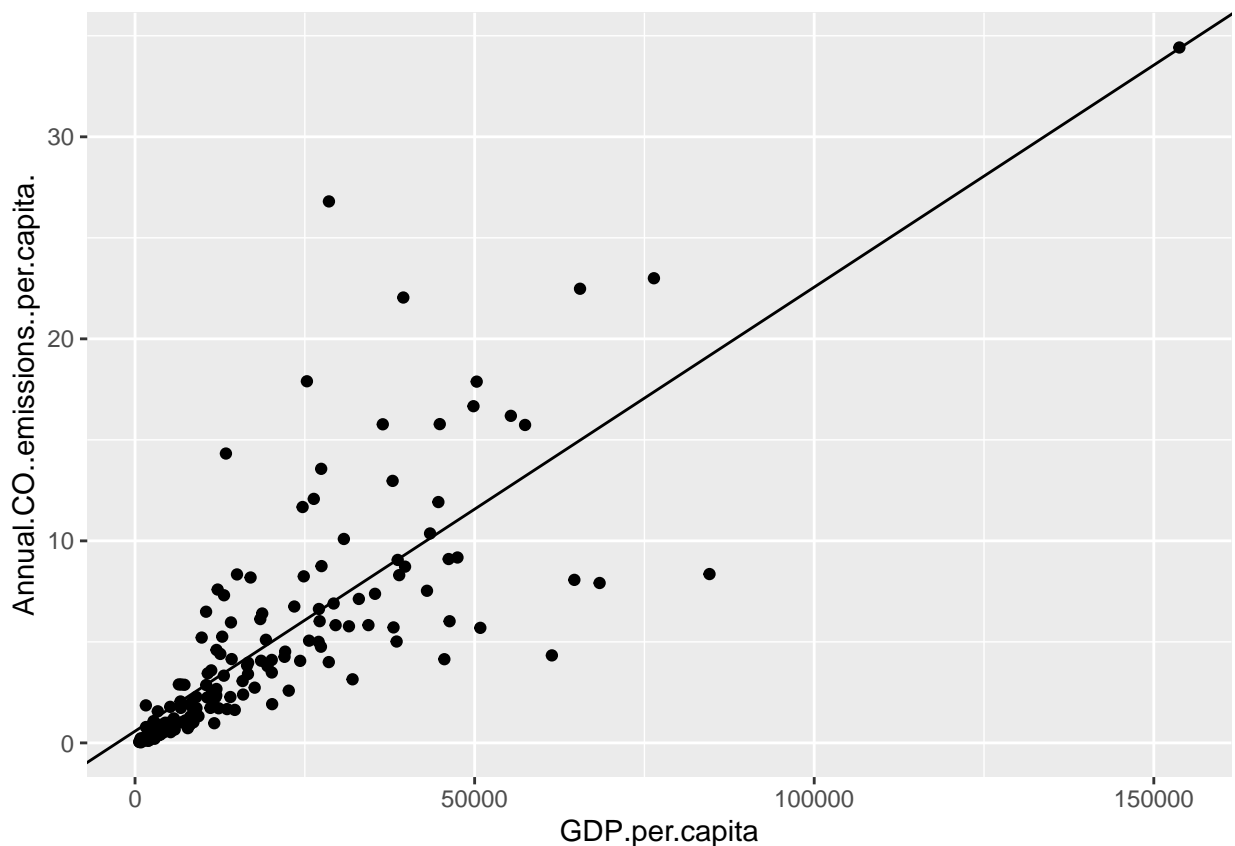
```
lm_a <- lm_robust(Annual.CO..emissions..per.capita. ~ GDP.per.capita, data = full_data)
summary(lm_a)

##
## 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.

$\text{se}(\text{intercept}) : 2.564e-01$ ,  $\text{se}(\text{slope}) : 1.863e-05$

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

(ii)

Add additional variables  $X_2, \dots, X_k$  that may affect  $Y$  and estimate the multiple regression model. Compare the coefficient estimate of  $X_1$  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.c
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 seriously hard data wrangling(keys are not consistent). Making those data can improve the quality of this research.