

# DATA 300 3 Homework 2 Solution

Aadarsha Gopala Reddy

September 27, 2022

## Contents

|   |   |
|---|---|
| 1. Data Import                                      | 1 |
| 2. Fitting a Bivariate Model                        | 2 |
| 3. Interpreting The Bivariate Model                 | 3 |
| 4. Fitting a Multivariate Model                     | 4 |
| 5. Interpreting the Multivariate Model              | 4 |
| 6. Out-Of-Sample Predictions                        | 4 |
| 7. Subset the Data                                  | 4 |
| 8. Modeling the Decision to Sign the Kyoto Protocol | 4 |
| 9. Out-Of-Sample Predictions for the Kyoto Protocol | 5 |

## 1. Data Import

Load any packages that you will use to complete the assignment. Import carbon\_emissions.csv as an object in R.

library() Loads any packages.

read.csv() Reads the csv file.

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.9
## v tidyr   1.2.0      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
carbon_emissions <- read_csv("carbon_emissions.csv")

## Rows: 8960 Columns: 8
## -- Column specification -----
## Delimiter: ","
```

```
## chr (2): country, iso_code
## dbl (6): year, co2_per_capita, population, gdp_per_capita, energy_per_capita...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
# The unit of observation in this dataset is the country-year.
# country: the name of the country.
# iso_code: a three-letter abbreviation of the country name.
# year: the year of the observation.
# co2_per_capita: an estimate of the amount of carbon dioxide released into the air, measured in tons per person per year.
# population: the population of the country in this year.
# gdp_per_capita: the gross domestic product, a measure of wealth, of the country in this year. Measured in US dollars.
# energy_per_capita: the amount of energy that the country consumed. Measured in kilowatt-hours per person per year.
# kyoto: whether the country signed the Kyoto Protocol in this year. Only measured for 1998.
```

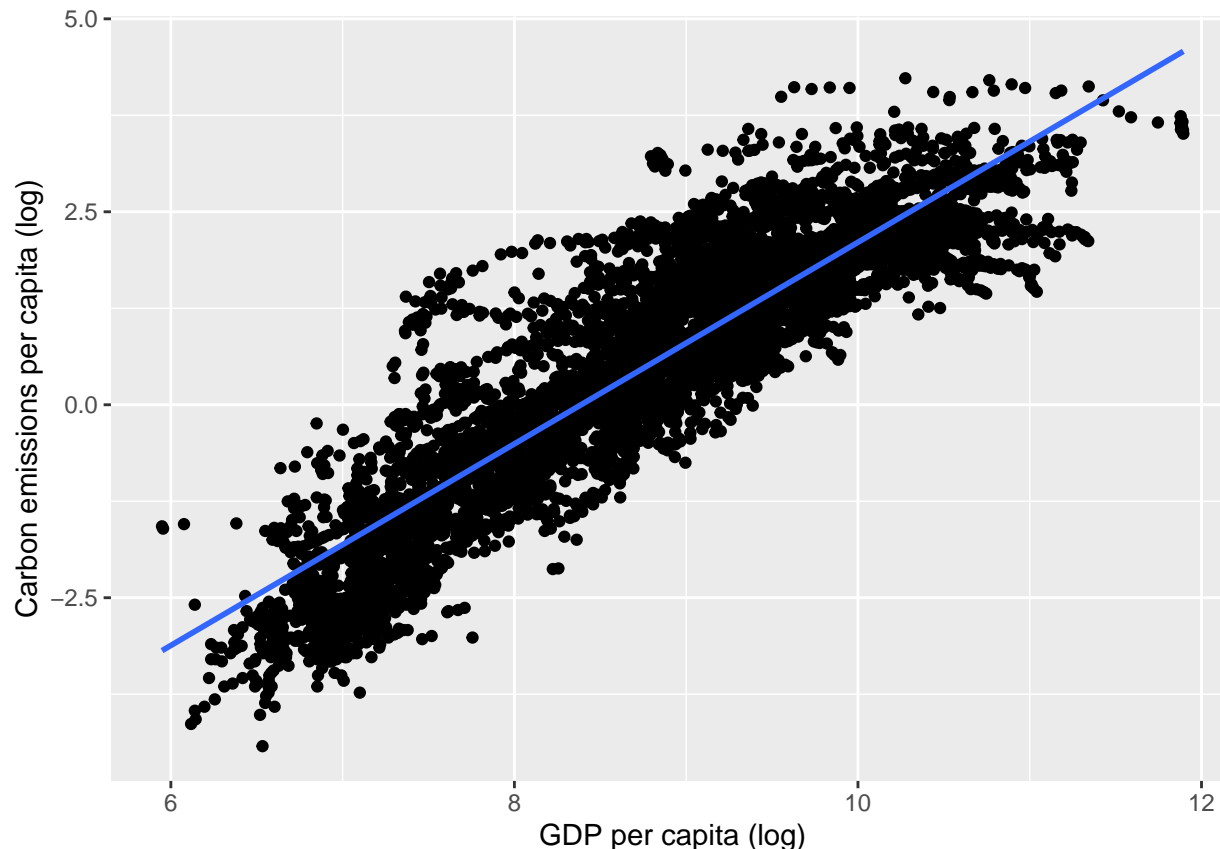
## 2. Fitting a Bivariate Model

Fit a linear model where the outcome is carbon emissions per capita and the predictor is GDP per capita. Apply any data transformations you believe are appropriate. (You might investigate whether data transformations are appropriate by visualizing the data or some other approach.)

```
carbon_emissions <- carbon_emissions %>%
  mutate(co2_per_capita = log(co2_per_capita),
         gdp_per_capita = log(gdp_per_capita))

#plot a chart
ggplot(carbon_emissions, aes(x = gdp_per_capita, y = co2_per_capita)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(x = "GDP per capita (log)", y = "Carbon emissions per capita (log)")

## `geom_smooth()` using formula 'y ~ x'
## Warning: Removed 2612 rows containing non-finite values (stat_smooth).
## Warning: Removed 2612 rows containing missing values (geom_point).
```



### 3. Interpreting The Bivariate Model

Interpret the output from the bivariate model.

- Is the relationship between GDP and carbon emissions positive or negative? What is the magnitude of the association? (How much should we expect carbon emissions to change when GDP increases?)
- Is the coefficient estimate statistically significant at the .05 level?
- How well does the model fit? (You might check metrics like the R-squared value or use visualization to answer this question.)

```
bivariate_model <- lm(co2_per_capita ~ gdp_per_capita, data = carbon_emissions)
summary(bivariate_model)
```

```
##
## Call:
## lm(formula = co2_per_capita ~ gdp_per_capita, data = carbon_emissions)
##
## Residuals:
```

| ## | Min      | 1Q       | Median   | 3Q      | Max     |
|----|----------|----------|----------|---------|---------|
| ## | -2.19079 | -0.49833 | -0.09689 | 0.40296 | 2.77028 |

```
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -10.953202   0.071300  -153.6   <2e-16 ***
## gdp_per_capita  1.305839   0.008013   163.0   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7511 on 6346 degrees of freedom
## (2612 observations deleted due to missingness)
## Multiple R-squared:  0.8071, Adjusted R-squared:  0.8071
## F-statistic: 2.656e+04 on 1 and 6346 DF,  p-value: < 2.2e-16
```

## 4. Fitting a Multivariate Model

Fit a linear model where the outcome is carbon emissions per capita. This time, include GDP per capita and at least one other predictor variable of your choosing. Again, apply any data transformations that you believe are appropriate.

Justify your choice of predictor variables. Why are the variables you chose appropriate for the model?

## 5. Interpreting the Multivariate Model

Interpret the output from the multivariate model.

- Are the relationships between the predictors positive or negative?
- What is the magnitude of the association between each predictor and carbon emissions?
- Which of the coefficient estimates are statistically significant at the .05 level?
- How well does the model fit? (You might check metrics like the R-squared value or use visualization to answer this question.)

## 6. Out-Of-Sample Predictions

Using the multivariate model from the previous question, calculate the predicted values of carbon emissions per capita across the range of GDP per capita. Set all other predictor variables to their mean value for this prediction.

You can generate predictions using manual calculations or by using R's `predict()` function. Give the at least the predicted values of carbon emission per capita for the maximum, minimum, and mean GDP observed in the dataset.

## 7. Subset the Data

Today, almost every country in the world has signed an international pledge to lower carbon emissions. This wasn't always the case. In 1998, a group of world leaders were the first to sign the Kyoto Protocol, an international agreement to limit carbon emissions and combat climate change.

Which countries were most likely to sign the Kyoto Protocol when it was introduced? We will build a model to answer this question. First, subset the dataset to only observations from the year 1998.

## 8. Modeling the Decision to Sign the Kyoto Protocol

The variable `kyoto` is a binary variable that records whether a country signed the Kyoto Protocol in the year 1998. Countries with a value of 1 signed the Kyoto Protocol in that year, and countries with a value of 0 did

not.

Build a model where the outcome is `kyoto`. Justify your choice of model and your choice of predictor variable(s).

## 9. Out-Of-Sample Predictions for the Kyoto Protocol

Choose one predictor variable from the model in the previous question. Calculate the predicted probability that a country signed the Kyoto protocol across the range of that variable. (If there are other variables in your model, set them to their mean value.)

Give at least the predicted probabilities of signing the treaty for the minimum, maximum, and mean values of the variable you chose. (Hint: if using R's `predict()` function, you may find it helpful to check the help page for `predict.glm`.)