

Homework 2

SPEA-V 507

Due on March 17, 2021

The data associated with this homework is `HW2Data.zip`. Each question is worth 20 points. For each question, the R output (including graphs) must be included in the document you upload to Canvas. Make sure, the interpret and explain the output.

1. Consider the data in `indyhomes`. A real estate agent once mentioned that homes built in the eighties are of lower quality and thus, are cheaper. One possibility to asses this claim is with an ANOVA model. Add three new dummy variables to the data frame: `eighties` (equal to 1 if the home was built in any year between 1980 to 1989), `northwest` (equal to 1 if the home is in the 46268 ZIP code), and `single story` (equal to 1 if the home has one story). Estimate the following ANOVA model:

$$price = \beta_0 + \beta_1 \cdot d_{80s} + \beta_2 \cdot d_{northwest} + \beta_3 \cdot d_{single}$$

What do you conclude? Next, add square footage to the model. What has changed in terms of coefficients? How do you interpret that change?

2. Using the data in `wdi`, estimate the following equation for the year 2018 and report the results:

$$fertrate = \beta_0 + \beta_1 \cdot gdp + \beta_2 \cdot litrate$$

The variable *fertrate* represents fertility rate (birth per woman), *gdp* represents the GDP per capita in in real terms, and *litrate* is the literacy rate.

3. Use the data in `sleep.csv` to estimate the following sleep equation:

$$sleep = \beta_0 + \beta_1 \cdot totwrk + \beta_2 \cdot educ + \beta_3 \cdot age + \beta_4 \cdot yngkid + \beta_5 \cdot male$$

Estimate the parameters of the model using OLS and interpret the coefficients. Use the Breusch-Pagan-Test for heteroscedasticity. What is the null hypothesis? Do you reject the null hypothesis at the 5% and 10% level? Obtain the heteroscedastic robust standard errors. What changes from the OLS results? What do you conclude?

4. The data `indyhomes` contains home values of two ZIP codes in Indianapolis. The model estimates the home value (dependent variable) based on a set of independent variables. The variables *levels* and *garage* refers to the number of stories and garage spots, respectively. The remaining variables are self-explanatory.

- (a) Create a dummy variable called *northwest* for the 46268 ZIP code.
- (b) Report and interpret the results of the following regression equation:

$$\ln(\text{price}) = \beta_0 + \beta_1 \cdot \ln(\text{sqft}) + \beta_2 \cdot \text{northwest} + \beta_3 \cdot \ln(\text{lot}) + \beta_4 \cdot \text{bed} \\ + \beta_5 \cdot \text{garage} + \beta_6 \cdot \text{levels} + \beta_7 \cdot \text{northwest} \cdot \text{levels}$$

- (c) What is the expected home value of a house in the 46228 ZIP code area with the following characteristics: 1900 sqft, 0.65 acres lot, 3 bedrooms, 3 bathrooms, 2 garage spots, and 2 story.
 - (d) Conduct a Breusch-Pagan-Godfrey test for heteroscedasticity. What do you conclude?
 - (e) Estimate the above model with heteroscedasticity-consistent (HC) standard errors. What changes compared to the model from Part b?
5. Use the command `subset()` on the WDI data and to select the variables `fertrate`, `gdp`, `litrte`, `lifeexp`, and `mortrate` for the year 2015. Estimate the following model:

$$\text{fertrate} = \beta_0 + \beta_1 \cdot \text{gdp} + \beta_2 \cdot \text{litrte} + \beta_3 \cdot \text{lifeexp} + \beta_4 \cdot \text{mortrate}$$

Interpret the results. What do you conclude in terms of statistical significance and the value of R^2 ? Use the function `vif` from the package `car`. What do you think causes the issue of multicollinearity in this case? Adjust your model to avoid multicollinearity.