Chapter 15: Putting It Altogether

Exercises

Brian Fogarty

Exercise 1

a. Read-in the `kenya_wvs.xlsx` dataset (or whatever name you used to save the Kenya data from the WVS website) and subset the data to only keep `responsibility`, `ideology`, `age`, `sex`, and `edu`. (Hint: follow the code in Chapter 15, but only include these five variables.)

b. Write an empirical hypothesis with `responsibility` as the outcome variable and each of the other variables as predictors. (You will write four hypotheses.) Also, write one generic null hypothesis.

c. Run a linear regression with `responsibility` as the outcome variable and the four other variables as predictors. Evaluate the overall model (using R^2 and the F-test) and identify any statistically significant relationships.

d. Plot the regression coefficients using the `ggcoef_model()` function from the `GGally` package.

Exercise 2

Perform the following diagnostics on `model.1`.

a. Test for functional form violations. If there are violations, try to find solutions.
b. Test for heteroscedasticity. If heteroscedasticity is present, re-run the regression using robust standard errors. Are any predictors that were statistically significant now not significant?
c. Test for non-normality in the residuals. If non-normality is present, discuss how to solve it.
d. Test for multicollinearity. If multicollinearity is present, discuss how to solve it.
e. Test for outliers, leverage, and influential data points. If any influential data points are present, discuss how you would handle them.

Exercise 3

Provide an interpretation of any statistically significant coefficient and discuss any significant relationships using plain language.

For the answers see **Chapter 15 - Answers to Exercises**.