Chapter 12: OLS Assumptions and Diagnostic Testing

Exercises

Brian Fogarty

Exercise 1

Using techniques covered in Chapter 12, perform regression diagnostics on `model.1` from Chapter 11's Exercises. As a reminder, using the 2020 Scottish Index of Multiple Deprivation dataset (`simd2020.csv`), this model has `crime_rate` as the outcome variable and `Income_rate`, `Employment_rate`, and `not_participating` as predictors.

a. Convert the three predictors to percentages (i.e., multiply them by 100) and run the regression for `model.1`.
b. Test for functional form violations. If there are violations, try to find solutions.
c. 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?
d. Test for non-normality in the residuals. If non-normality is present, try to find a solution.
e. Test for multicollinearity. If multicollinearity is present, discuss how to solve it.
f. Test for outliers, leverage, and influential data points. If any influential data points are present, discuss how you would handle them. Also, examine one of the highlighted data points in the influential data points plot (i.e., Cook's *d* plot) and examine one of the highlighted outliers in the outliers plot (i.e., Studentized residual plot).

Exercise 2

a. Create a new version of `urban` with value labels. Then run a new regression model on a subsetted version of the `simd` data with `crime_rate` as the outcome variable and `Employment_rate`, `not_participating`, and (the labelled-version of) `urban` as predictors. The data should be subsetted to only keep observations where `crime_rate` is equal to or less than 2,000, `Employment_rate` is equal to or less than 40, and `not_participating` is equal to or less than 30.
b. Test for functional form violations. If there are violations, try to find solutions.
c. 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?
d. Test for non-normality in the residuals. If non-normality is present, try to find a solution.
e. Test for multicollinearity. If multicollinearity is present, discuss how to solve it.
f. Test for outliers, leverage, and influential data points. If any influential data points are present, discuss how you would handle them. Also, examine one of the highlighted data points in the influential data points plot (i.e., Cook's *d* plot) and examine one of the highlighted outliers in the outliers plot (i.e., Studentized residual plot).

For the answers see **Chapter 12 - Answers to Exercises**.