

Chapter 11: Linear Regression and Model Building

Exercises

Brian Fogarty

Exercise 1

- Using the `simd2020.csv` dataset, create new versions of `Income_rate`, `Employment_rate`, and `not_participating` that are out of 100 (i.e., multiply them by 100).
- Run a linear regression with `crime_rate` as the outcome variable and the new versions of the variables in 'a.' as predictors. Evaluate the overall model (using R^2 and the F -test) and identify any statistically significant relationships. Additionally, provide an interpretation of any statistically significant coefficient and briefly discuss any significant relationships using plain language.
- Plot the regression coefficients using the `ggcoef_model()` function from the `GGally` package.

Exercise 2

- Re-run the regression model in Exercise 1.b with `urban` added as a predictor (the model now has four predictors). Evaluate the overall model (using R^2 and the F -test), identify whether the statistically significant relationships from Exercise 1.b have changed, and whether `urban` has a statistically significant effect on `crime_rate`. If `urban` is statistically significant provide an interpretation of the coefficient and briefly discuss the result using plain language.
- Plot the regression coefficients using the `ggcoef_model()` function from the `GGally` package.

For the answers see **Chapter 11 - Answers to Exercises**.