Introduction to the R Statistical Computing Environment
Linear Models in R: Exercise

John Fox
(McMaster University)
ICPSR

2021

* The data given in the data frame `Burt` in the **carData** package, on the IQs of 27 pairs of identical twins reared apart, were reported by Sir Cyril Burt (1966). (These "data" are wholly fraudulent.) One twin in each pair was raised by his or her biological parents; the other twin was raised in a foster home. In each case, Burt recorded (i.e., made up) the "social class" to which the twins' biological parents belonged. See `?Burt` for more information.

1. Explore the data graphically by plotting `IQbio` (as the response variable) against `IQfoster`, using a different symbol and plotting a separate linear regression line for each social class. *Hint:* You can use the **car** command `scatterplot(IQbio ~ IQfoster | class, data=Burt, smooth=FALSE)` to make this graph.
2. Then regress the IQ of the twins reared by their biological parents (`IQbio`) on the IQ of the twins reared by foster parents (`IQfoster`), dummy variables to represent the three social classes (`class`), and regressors for the interaction between foster-twin IQ and social class. *Suggestion*: You may want to re-order the categories of the factor class so that they are in their natural order rather than in the (default) alphabetic order. Using the estimated regression coefficients, write out the fitted regression equation for each class.
3. Test the interaction between foster-twin IQ and social class. If the interaction proves to be non-significant, test the partial effects of foster-twin IQ and social class on biological-twin IQ. Compute the appropriate Type-II F-tests using the `Anova()` function in the **car** package.
4. Based solely on your statistical analysis of the data, how can you tell with a high level of certainty that the data are "cooked"?
5. Use standard regression diagnostics to examine the adequacy of the regression model fit to the data in part 2.