

Psychology 105, Advanced Research Methods, Homework #4

Available 4/8/2021, due 4/20/2021

Once again, you will continue to work with the Statlab50 data set that you created for Homework #1.

Multiple linear regression

In class on 4/6/2021, we introduced multiple regression and used added variable plots to help us understand what multiple regression is doing. On 4/8/2021, we expanded on these methods with another example that uses the same variables you will use in this homework assignment: Child's Peabody score (CTPEA), mother's age when the child is born (MBAG), and family income in 100s of dollars when the child is 10 years old (FIT). Specifically, we regressed Peabody on mother's age and family income. The steps we took in the 4/8 class will serve as a useful guide as you complete this assignment. Do keep in mind, though, that the results from your sample may be very different from the results we had in class.

Step One:

1. Produce "pretty" scatterplots showing the relationship between Peabody and each predictor.
2. Comment on the linearity in the plots: Do you see evidence that either of the relationships is nonlinear?

Step Two:

1. Forget that you know about multiple regression, and regress Peabody first on mother's age, and then on family income.
2. For each of those regressions, test the null hypothesis about the slope.
3. For each of those regressions, explain what the slope means. (Remember, regression is a model for *conditional mean*.)
4. For each of those regressions, evaluate the assumption of equal error variability across the range of fitted values, and the assumption that errors are normally distributed. This will involve graphics; take the time to make your residuals plots look pretty.

Step Three:

1. Use multiple regression to fit Peabody as a function of mother's age and family income simultaneously.
2. Test the hypotheses about both slopes.
3. Explain what each slope means.
4. Evaluate the assumptions of equal error variability across the range of fitted values and normality of errors. Again, take the time to make your residuals plots look pretty.
5. For each predictor (FIT and MBAG), produce an added variable plot; make it pretty.

6. Confirm that the slopes in the simple regressions implied by the added variable plots match the slopes in your multiple regression.

Step Four:

1. Comment on differences and similarities among the three regressions you have done. Did you reach the same conclusion about the effects of mother's age in the simple regression as in the multiple regression? What about family income?
2. What model would you use if you wanted to understand variation in Peabody scores as a function of these variables? Why?