

536 Homework 1

Adrian Dobra `adobra@uw.edu`

This homework concerns the Labor Force Participation data which is available on the course website. A version of this dataset is also discussed in Chapters 3 and 4 of Long. You are asked to analyze this dataset by following these steps:

1. Determine which variables should be excluded from the analysis up front. Examples include “hours”, “wage”, “lwage.” These are variables that should not be seen as explanatory for the binary response “inlf”.
2. Determine if you need to take any transformation of the remaining explanatory variables. For example, “age” should be used as “log(age)”, etc.
3. Determine an initial logistic regression M_0 using a heuristic argument of your choice. You could include the explanatory variables with the highest absolute correlation with the response. You can come up with another idea if you like.
4. Fit the model M_0 and make sure there are no numerical errors when the MLEs are calculated. Give a formula for the fitted model and discuss its validity in term of standardized residuals, fitted values, etc. Produce relevant plots.
5. Use the function `step` to improve M_0 in terms of AIC. Let M be the final model returned by `step`.
6. Consider all the models that are obtained by deleting one variable from M . Perform likelihood ratio tests to see whether any variables should be removed from M .
7. Compute the BIC scores of the same models. Which model would you pick with respect to BIC?
8. Now calculate the Brier scores of the same models. Which model would you pick now? What is the error rate of your preferred model?
9. Draw conclusions about what you learned from your analysis. Give a formula for your final model and interpret the regression coefficients (an interpretation in term of log-odds is fine). Discuss the statistical significance of each variable selected in the model.

It is critical to remember that there is no perfect answer when it comes to data analysis. What matters is your line of reasoning. Try to follow all these steps, but do not be too verbose in your write-up. The longest homeworks are not necessarily the best. However, you need to include all the relevant details.