

Homework 6

Stat 135, summer 2012

Due in class, Monday 6th August

1. Rice (3rd ed.) 14.8
2. Rice 14.10
3. Rice 14.20
4. Rice 14.22
5. In the 2000 Presidential election, who voted for Bush? Who voted for Gore?

The data are on bSpace. The variables into the data set are:

- vote: 0 = Gore, 1 = Bush, NA = other or didn't vote
- female: 0 = male, 1 = female
- age: 1 = 18-29, 2 = 30-44, 3 = 45-64, 4 = 65+
- race: 0 = white, 0.5 = other, 1 = black
- income: 1 = 0th-16th percentile, 2 = 17th-33rd percentile, 3 = 34th-67th percentile, 4 = 68th-95th percentile, 5 = 96th-100th percentile
- educ: 1 = no high school, 2 = high school grad, 3 = some college, 4 = college grad
- party: 1 = strong Democrat, 2 = Democrat, 3 = weak Democrat, 4 = independent, 5 = weak Republican, 6 = Republican, 7 = strong Republican
- ideo: 1 = strong liberal, 7 = strong conservative

Analyses you might consider performing include:

- Histograms or other graphs comparing Bush voters to Gore voters
- Single-variable logistic regressions
- Multivariate logistic regressions
- Multivariate logistic regression with interaction
- Even more advanced models, e.g. the generalized additive model.
- Plots of fitted curves for your final model
- Plots of residuals or deviances

You don't have to do all of these, but you should do most of them.

You should hand in:

- A model that gives a predicted probability of a voter in the 2000 Presidential election voting for Bush or Gore (you can ignore votes for Nader etc.)
- A justification of why you chose that model.
- Graphs or other analyses that show how well the model fits.
- A concise description of how the variables in the data set affect the predicted probability of voting for Bush.
- R code *as an appendix*. You will lose points if you put unnecessary R code in the body of your assignment.