

## Problems: Logit and Probit Models

1. Using the data on the 1988 Chilean plebiscite (the data frame `Chile` in the `car` package – see `?Chile`), perform a logistic regression of voting intention (yes or no) on `region`, `population`, `sex`, `age`, `education`, and `income`. Then add the variable `statusquo` (support for the status-quo) to the regression equation. Fit a similar probit model to the data and compare the results with the logit model.

Because there are several levels to the factor `vote`, you will have to create the binary response variable for these regressions; here is one way to do so:

```
Chile$vote1 <- factor(with(Chile, ifelse(vote=="Y", "yes",  
                                       ifelse(vote=="N", "no", NA))))
```

You will also probably want to order the levels of education naturally rather than alphabetically, e.g., by the command

```
Chile$education <- factor(Chile$education, levels=c("P", "S", "PS"))
```

2. Continuing with the Chilean plebiscite data, now include undecided voters, by creating a factor or ordered factor with levels `N`, `U`, and `Y`. One way to do this is as follows:

```
Chile$vote2 <- Chile$vote  
Chile$vote2[Chile$vote2 == "A"] <- NA # eliminate abstainers  
Chile$vote2 <- ordered(Chile$vote2) # could skip this step
```

Redo the analysis in Problem 1 for this new response variable, employing both proportional-odds models and multinomial logit models. Compare the results from the two kinds of models.

*Note:* You will have trouble fitting the proportional-odds model with the `polr()` function without first transforming `income` and `population`; you can use `log2(income)` and `log2(population)` in the regression equations.