

**Econ 108   FALL 2022**  
**Problem Set 3**

**This problem set is due at 7pm, Friday, October 21, 2022.**

1. In this exercise we will use the spam.csv data and the spam.R code from the textbook.
  - (a) Understand and run the spam.R code.
  - (b) Run a logit function of  $p(y = 1|x)$  using only two features/covariates as  $x$ : char\_dollar and word\_george. How does the in-sample  $R^2$  of this simplified model compare to the in-sample  $R^2$  of the full regression that uses all features (as in page 52 of the paper copy of the textbook, or page 44 of the lecture slides)?
  - (c) In the above regression, how does having the word "george" in an email change the odds of the email being a spam? Is the change in the odds statistically significant?
  - (d) Randomly split the sample into a leave-out subsample with 1000 emails and a remaining training subsample. Estimate the above two feature logit model using the training subsample, and compute its  $R^2$  on the leave-out subsample. How does the out-of-sample  $R^2$  compare to the in-sample  $R^2$ ?
  - (e) Run a logit function of  $p(y = 1|x)$  using every feature in the original data set except char\_dollar and word\_george. Compare the in-sample  $R^2$  and the out-of-sample  $R^2$ . (Hint: You might find the following command useful.)

```
email[,-which((names(email) %in% c("word_george","char_dollar")))]
```

2. In this exercise we will use the oj.csv data and the oj.R code from the textbook.
  - (a) Rerun the price elasticity regression from the previous problem set where log price is interacted with brand.
  - (b) Plot the residuals of this regression against log price. Does the plot suggest possible conditional heteroscedasticity of the error term?
  - (c) Draw a box plot of the residuals of this regression against the brand factors. Does the plot suggest possible conditional heteroscedasticity of the error term?
  - (d) Compute the HC-robust standard errors of the regression coefficients using the AER package. Are the robust standard errors visibly different from the non-robust standard errors?