Problem Set 4

Luke Duggan, 16316834

November 25, 2021

Question 1

1. The R code to create this dummy variable is:

```
Prestige$professional <- ifelse(Prestige$type=="prof", 1, 0)
```

2. The R code to run this linear model is:

```
results <- lm(prestige ~ income + professional + income*professional, data = Prestige)
```

3. Rounding to three decimal places, the prediction equation is:

$$\text{prestige} = 21.142 + (\text{income})*(0.003) + (\text{professional})*(37.781) - (\text{income*professional})*(0.002)$$

4. Usually, the coefficient on `income` would just be the marginal effect of an increase in `income` on `prestige`. However, the presence of the interaction term slightly complicates matters.

   The coefficient on `income` is the effect of `income` on `prestige` *when the interaction variable is zero.* Accordingly, in this case, the coefficient gives the marginal effect of `income` on `prestige` for *non*-professionals. For professionals, the estimated marginal effect is equal to this coefficient *plus or minus* the estimated coefficient on the interaction term.

5. The coefficient on `professional` is the difference in the constant or y-intercept term for *professionals* compared to non-professionals, i.e., those for whom the dummy is coded as 0. The intuition is straightforward: for non-professionals, the dummy is 0,

and the constant term is just 21.142. For professionals, we must add 37.781 to this constant term because the dummy takes the value 1.

6. When `professional`=1, the constant term 58.924 (this isn't the sum of the rounded coefficients given here, but a rounding of the estimated coefficients given by R). Then, factoring out `income` from the two other terms gives a slope coefficient of 0.0008452, around 0.001. Finally, plugging `income` = 1000 into this equation gives a prestige value of 59.7682. Comparing this to the constant term, we see that a $1,000 increase income leads to around about a 0.84 increase in `prestige`.

7. For this question, we plug in `income` = 6000. Then, we evaluate `prestige` first when `professional`=1 and then when `professional`=0. Subtracting the two answers gives the effect on `prestige` of switching from nonprofessional to professional at this level of income.

   Working with the rounded figures (so there may be a little rounding error), the answer is about 88.923 - 39.142 = 49.781. Accordingly, at an income of $6,000, the effect of switching from nonprofessional to professional is about 49 units of the outcome variable, `prestige`.

Question 2

1. We want to test if the estimated coefficient on the dummy variable for being assigned a lawn sign is significantly different from zero.

   We do this using a familiar test statistic following Student's t-distribution. In this case, the test statistic is the estimated coefficient, 0.042, divided by the standard error of the estimate, which is 0.016. This follows Student's t-distribution with n - k - 1 = 30 - 2 - 1 = 27 degrees of freedom.

   We can use R to calculate both the value of the test-statistic and the critical value at the 0.05 level of significance.

   ```
   test_statistic <- (0.042)/(0.016)
   test_statistic

   qt(0.95, 27)
   ```

   The value of the test statistic is 2.625; the critical value at this level of significance is a little over 1.7. Thus, the test statistic is well in excess of the critical value, and we reject the null hypothesis that the coefficient is zero.

2. The method is as before, and the estimated coefficient happens to be the same. However, this time the standard error is 0.013, and the degrees of freedom are 76 - 2 - 1 = 73. Changing the R code appropriately, we see that the value of the test statistic is about 3.23, and the critical value is around 1.67. Thus, we can also reject the null of a zero coefficient in this case.

3. In this case, the coefficient on the constant term is the "baseline" or "reference" value relative to the dummy variables; that is, in a precinct not having the lawn signs or being next to a precinct having the lawn signs, the estimated share of the vote for Cuccinelli is around 30%. The coefficients on the dummies express how the presence of the lawn signs (either in a precinct or in a neighbouring precinct) adds to or subtracts from this baseline value (in this case, the vote share is estimated to be increasing in both cases).

4. The coefficient of determinaton (the R^2 value) is 0.094. This means that the model explains around 9.5% of the variation in the data. This is not particularly large: around 90% of the variation in the data is due to factors that have not been modelled. Nevertheless, it might be enough to be of relevance in deciding a particularly close election.