

Problem Set 4

Owen Eglinton

November 18, 2024

Question 1: Economics

- (a) Create a new variable **professional** by recoding the variable **type** so that professionals are coded as 1, and blue and white collar workers are coded as 0 (Hint: **ifelse**).

Answer: I wrote and executed the following code:

```
1 for (i in 1:nrow(Prestige)) {  
2   if (is.na(Prestige$type[i]) | Prestige$type[i] != "prof") {  
3     Prestige$professional[i] <- 0  
4   } else {  
5     Prestige$professional[i] <- 1  
6   }  
7 }
```

This generated the dummy variable with the desired values.

- (b) Run a linear model with **prestige** as an outcome and **income**, **professional**, and the interaction of the two as predictors (Note: this is a continuous \times dummy interaction.)

Answer: I wrote and executed the following code:

```
1 prestige_lm <- lm(prestige ~ income*professional, Prestige)  
2 summary(prestige_lm)
```

This generated the following regression output:

Residuals:

Min	1Q	Median	3Q	Max
-14.854	-5.429	-1.055	4.513	29.762

Coefficients:

```

Estimate Std. Error t value Pr(>|t|)
(Intercept)      20.8035385    2.5387225    8.194 9.74e-13 ***
income            0.0032691    0.0004565    7.161 1.49e-10 ***
professional      38.1200003    4.0797620    9.344 3.22e-15 ***
income:professional -0.0024239    0.0005304   -4.570 1.42e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.018 on 98 degrees of freedom
Multiple R-squared:  0.7893, Adjusted R-squared:  0.7828
F-statistic: 122.3 on 3 and 98 DF,  p-value: < 2.2e-16

```

(c) Write the prediction equation based on the result.

Answer: Based on the above results, the prediction equation can be written as:

$$\hat{Prestige}_i = 20.8 + 0.0032 * Income_i + 38.12 * Professional_i - 0.0024 * Income_i * Professional_i$$

(d) Interpret the coefficient for **income**.

Answer: The correct interpretation is that a unit increase in income is associated with a 0.0032 unit increase in a career's expected prestige.

(e) Interpret the coefficient for **professional**.

Answer: The correct interpretation is that a professional career will have, on average, a prestige that is higher by 38.12 units than a non-professional career with the same income.

- (f) What is the effect of a \$1,000 increase in income on prestige score for professional occupations? In other words, we are interested in the marginal effect of income when the variable `professional` takes the value of 1. Calculate the change in \hat{y} associated with a \$1,000 increase in income based on your answer for (c).

Answer: I wrote and executed the following code:

```
1 prestige_model <- function(income, professional) {  
2   c1 <- prestige_lm$coefficients[1]  
3   c2 <- prestige_lm$coefficients[2]  
4   c3 <- prestige_lm$coefficients[3]  
5   c4 <- prestige_lm$coefficients[4]  
6   c1 + income*c2 + professional*c3 + income*professional*c4  
7 }  
8 prestige_model(1000, 1) - prestige_model(0, 1)
```

This output a value of 0.8452, which is the unit increase in expected prestige associated with a \$1,000 increase in income for professional occupations.

- (g) What is the effect of changing one's occupations from non-professional to professional when her income is \$6,000? We are interested in the marginal effect of professional jobs when the variable `income` takes the value of 6,000. Calculate the change in \hat{y} based on your answer for (c).

Answer: I wrote and executed the following code:

```
1 prestige_model(6000, 1) - prestige_model(6000, 0)
```

This output a value of 23.57631, which is the unit increase in expected prestige associated with a switch from a non-professional to a professional career at a constant income level of \$6,000.

Question 2: Political Science

- (a) Use the results from a linear regression to determine whether having these yard signs in a precinct affects vote share (e.g., conduct a hypothesis test with $\alpha = .05$).

Answer: $H_0: \hat{\beta}_{assigned} = 0$

$H_A: \hat{\beta}_{assigned} \neq 0$

I wrote and executed the following code:

```
1 t_val_1 <- 0.042 / 0.016
2 2*pt(t_val_1, 30-1-1, lower.tail = FALSE)
```

This generated a p-value of 0.01387963. Since this is less than the selected α of 0.05, the null hypothesis is rejected; i.e., we conclude that $\hat{\beta}_{assigned} \neq 0$, and thus that the estimated coefficient is statistically significant.

- (b) Use the results to determine whether being next to precincts with these yard signs affects vote share (e.g., conduct a hypothesis test with $\alpha = .05$).

Answer: $H_0: \hat{\beta}_{adjacent} = 0$

$H_A: \hat{\beta}_{adjacent} \neq 0$

I wrote and executed the following code:

```
1 t_val_2 <- 0.042 / 0.013
2 2*pt(t_val_2, 76-1-1, lower.tail = FALSE)
```

This generated a p-value of 0.001843341. Since this is less than the selected α of 0.05, the null hypothesis is rejected; i.e., we conclude that $\hat{\beta}_{adjacent} \neq 0$, and thus that the estimated coefficient is statistically significant.

- (c) Interpret the coefficient for the constant term substantively.

Answer: The constant term coefficient can be interpreted as showing that, in the 25 precincts that were neither assigned lawn signs nor adjacent to those that were, the average proportion of the vote that went to Ken Cuccinelli was 30.2%, with a standard error of 11%.

- (d) Evaluate the model fit for this regression. What does this tell us about the importance of yard signs versus other factors that are not modeled?

Answer: The R^2 for this regression is 0.094, which tells us that the estimated model accounts for only 9.4% of the variance in vote proportion. Given that this was a RCT, this implies that yard signs are relatively unimportant in determining voter outcome in comparison to factors remaining outside the model.