

Problem Set 3

Applied Stats/Quant Methods 1

Jack Merriman

Question 1

(a)

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

(b)

```
1 model <- lm(prestige ~ income*professional, data = Prestige)
```

Table 1

	<i>Dependent variable:</i>
	prestige
income	0.003*** (0.0005)
professional	37.781*** (4.248)
income:professional	-0.002*** (0.001)
Constant	21.142*** (2.804)
Observations	98
R ²	0.787
Adjusted R ²	0.780
Residual Std. Error	8.012 (df = 94)
F Statistic	115.878*** (df = 3; 94)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

(c)

Where `prestige` = Y , `income` = X_1 , and `professional` = X_2 :

$$Y = 21.142 + 0.003X_1 + 37.781X_2 - 0.002X_1X_2$$

So for blue and white collar jobs:

$$Y = 21.142 + 0.003X_1$$

And for professional jobs:

$$Y = 58.923 + 0.001X_1$$

(d)

An X_1 coefficient of 0.003 suggests that, before accounting for interaction effects, a \$1 increase in income leads to an expected increase of 0.003 in Pineo-Porter prestige score.

(e)

An X_2 coefficient of 37.781 suggests that, holding income constant, the average professional has a Pineo-Porter prestige score 37.781 higher than the average blue or white collar worker.

(f)

Using the equation for professional jobs with $X_1 = 1000$:

$$\hat{Y} = 58.923 + 0.001(1000)$$

$$\hat{Y} = 58.923 + 1$$

$$\hat{Y} = 59.923$$

For professional jobs, a \$1,000 increase in income would lead to an increase in Pineo-Porter prestige score of 1.

(g)

We put \$6,000 into both of the sub-equations from (c):

$$\hat{Y}_0 = 21.142 + 0.003(6000)$$

$$\hat{Y}_0 = 21.142 + 18$$

$$\hat{Y}_0 = 39.142$$

$$\hat{Y}_1 = 58.923 + 0.001(6000)$$

$$\hat{Y}_1 = 58.923 + 6$$

$$\hat{Y}_1 = 64.923$$

$$64.923 - 39.142 = 25.781$$

We can see the effect of making the worker in question an employee would lead to an increase of 25.781 in Pineo-Porter prestige score. If we look at the second lines of the equations above we can see that most of that increase can be attributed to differences in intercept, and that the income of much smaller for the professional equation. This suggests that income could be more influential for the prestige of blue and white collar jobs than professional jobs.

Question 2

(a)

Where β_1 is the linear relationship between the yard signs and a precinct's vote share:

$$H_0 : \beta_1 = 0$$

$$H_A : \beta_1 \neq 0$$

$$t = \frac{\hat{\beta}_1 - \beta_1}{se_{\hat{\beta}_1}}$$

$$t = \frac{0.042 - 0}{0.016}$$

$$t = 2.625$$

$$df = n - k - 1 = 131 - 3 = 128$$

```
1 > 2*pt(2.625, df = 128, lower.tail = FALSE)
2 [1] 0.00972002
```

Our p -value (0.0097) is less than our confidence threshold ($\alpha = 0.05$) therefore we can reject the null hypothesis, suggesting there is a statistically significant linear relationship between the presence of yard signs in a precinct and a precinct's vote share.

(b)

Where β_2 is the linear relationship between yard signs in an adjacent precinct and the original precinct's vote share:

$$H_0 : \beta_2 = 0$$

$$H_A : \beta_2 \neq 0$$

$$t = \frac{\hat{\beta}_2 - \beta_2}{se_{\hat{\beta}_2}}$$

$$t = \frac{0.042 - 0}{0.013}$$

$$t = 3.231$$

$$df = n - k - 1 = 131 - 3 = 128$$

```
3 > 2*pt(0.042/0.013, df = 128, lower.tail = FALSE)
4 [1] 0.00156946
```

Our p -value (0.0016) is less than our confidence threshold ($\alpha = 0.05$) therefore we can reject the null hypothesis, suggesting there is a statistically significant linear relationship between the presence of yard signs in an adjacent precinct and the original precinct's vote share.

(c)

The coefficient for the constant, or intercept, is the mean value of Y for observations in the data where the other two variables are 0. That means, in this model, that Ken Cuccinelli's average vote share in precincts without yard signs that are not adjacent to another precinct with a yard sign was 30.2%.

(d)

We will begin our evaluation by conducting an overall f-test on the model with $\alpha = 0.05$. Our hypotheses as follows:

$$H_0 : \beta_1 = \beta_2 = 0$$

$$H_A : \beta_1 \text{ or } \beta_2 \neq 0$$

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

$$F = \frac{0.094/2}{(1-0.094)/(131-2-1)}$$

$$F = \frac{0.047}{0.906/128}$$

$$F = 6.64$$

```
5 > df((0.094/2)/((1-0.094)/128), 2, 128)
6 [1] 0.001634304
```

With a p-value of 0.0016 below the significance level ($\alpha = 0.05$) we reject H_0 .

Our overall F-test producing a $p < 0.05$ tells us what we have already learned from (a) and (b) which told us that our individual predictors were useful, and therefore our overall model is useful. The presence of negative yard signs against McAuliffe did impact local vote share for his Republican opponent. The R^2 value of 0.094 is quite low, and shows that the yard signs explain around 9% of the variation of the vote share, and that other variables will explain the overwhelming majority of the variation.