Problem Set 4

Philip Kruger
18328699
04/12/22

Question 1

Part 1

The following code was used to create an extra column for the professional dummy variable, and remove all rows which have NA as their type. It is unknown if these jobs qualify as professional or not as such they are best removed from the analysis:

```
Prestige$professional <- vector(mode="integer", length=nrow(Prestige)) #adding dummy column

for (i in 1:nrow(Prestige)){
ifelse(Prestige[i,6] == "prof", Prestige[i,7] <- 1, Prestige[i,7] <- 0)
} #filling dummy column

#removing entries with NA for type
Prestige2 <- Prestige

for (i in 1:nrow(Prestige2)){
if (is.na(Prestige2[i,6]) == TRUE){
Prestige2 <- Prestige2[-c(i),]}
}
```

This Yields the following first few lines of the new Prestige2 dataset:

	education	income	women	prestige	census	type	professional
gov.administrators	13.11	12351	11.16	68.8	1113	prof	1
general.managers	12.26	25879	4.02	69.1	1130	prof	1
accountants	12.77	9271	15.70	63.4	1171	prof	1
purchasing.officers	11.42	8865	9.11	56.8	1175	prof	1
chemists	14.62	8403	11.68	73.5	2111	prof	1

Part 2

The following code was used to get the linear regression below:

```
summary(lm(Prestige2$prestige ~ Prestige2$income +Prestige2$professional +
 Prestige2$income:Prestige2$professional)) #getting the coefficients
```

Which resulted in:

```
Residuals:
Min     1Q  Median    3Q     Max
-14.852  -5.332  -1.272   4.658  29.932

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)                              21.1422589  2.8044261   7.539 2.93e-11 ***
Prestige2$income                          0.0031709  0.0004993   6.351 7.55e-09 ***
Prestige2$professional                   37.7812800  4.2482744   8.893 4.14e-14 ***
Prestige2$income:Prestige2$professional -0.0023257  0.0005675  -4.098 8.83e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.012 on 94 degrees of freedom
Multiple R-squared:  0.7872,Adjusted R-squared:  0.7804
F-statistic: 115.9 on 3 and 94 DF,  p-value: < 2.2e-16
```

Part 3
As can be seen from the regression above, the Prediction equation is:

$$21.1422589 + 0.0031709X_1 + 37.78128D_1 - 0.0023257X_1D_1$$

Part 4
The Coefficient for income is 0.0031709. This means that for jobs for which the dummy variable is 0 (non-professional jobs), for a 1 unit increase in income there is on average a 0.0031709 unit increase for prestige. It also has an intercept of 21.1422589 units of prestige.

Part 5
The coefficient for professional is -0.002357, this means that on average the relative increase in prestige for a professional compared to a non-professional per unit income is -0.00237. This is rather significant given that the coefficient for income was 0.0031709 and justifies the use of an interactive rather than an additive model. Taking professional as a regression line the intercept would be $21.1422589 + 37.78128 = 58.9235389$ and the slope would be $0.0031709 - 0.0023257 = 0.0008452$. Meaning that for a professional job with zero income the predicted prestige would be 58.9235389 and for each unit increase in income there would be a 0.0008452 unit increase in prestige.

Part 6
As shown in part 5 a 1 unit increase in income for a professional results in a 0.0008452 unit increase in prestige. Thus a 1000 unit increase in income yields:

$$1000 * 0.0008452 = 0.8452$$

Thus the prestige is increased by 0.8452 units.

Part 6
From the prediction equation found in part 3, the prestige for a professional and non-professional earning 6000 a year is:

$$\text{non-professional: } 21.1422589 + 6000 * 0.0031709 = 40.1676589$$
$$\text{professional: } 58.9235389 + 6000 * 0.0008452 = 63.9947389$$

Thus the change in prestige is 63.9947389 - 40.1676589 = 23.82708.

Question 2

Part 1

We are testing the hypothesis that yard signs in the precinct influences voting preferences. The Hypotheses are as follows:

$$H_0 = \text{the coefficient for areas receiving the treatment was 0.}$$
$$H_a = \text{coefficient is not zero.}$$

The F statistic is gained by dividing the between group variability by the in group variability. In this case it is:

$$0.042/0.016 = 2.625$$

With 29 degrees of freedom this results in a p-value of 0.00504 (F distribution calculator). This is below our α, thus we are able to reject the null hypothesis that the coefficient is zero. Thus we can conclude that yard signs do affect voteshare.

Part 2

We are testing the hypothesis that yard signs in adjacent precincts influence voting preferences. The Hypotheses are as follows:

$$H_0 = \text{the coefficient for areas adjacent to those receiving the treatment was 0.}$$
$$H_a = \text{coefficient is not zero.}$$

The F statistic is gained by dividing the between group variability by the in group variability. In this case it is:

$$0.042/0.013 = 3.323$$

With 75 degrees of freedom this results in a p-value of 0. This is below our α, thus we are able to reject the null hypothesis that the coefficient is zero. Thus we can conclude that yard signs in adjacent precincts do affect voteshare.

Part 3

The coefficient for the constant term is 0.302. This means that for places that did not receive the treatment had an average voteshare of .302. This means that for the precincts which did not have any yard signs and were not adjacent to any lawn signs the average vote share for Ken Cuccinelli was 0.302. When comparing the coefficient for the treatment to this constant we see that on average precincts receiving this treatment had a roughly 14 percent higher voteshare for Cuccinelli.

Part 4

The R^2 of the model was 0.094. This means that there is low to no correlation between the treatment and who one voted for. This indicates that there are other factors, which are not included in the model, which play a larger role in who one votes for than whether your precinct received the treatment.