Problem Set 4 Answers

KG

11/25/2021

Question 1: Economics

In this question, use the prestige dataset in the car library.

First, run the following commands:

install.packages(car)

library(car)

data(Prestige)

help(Prestige)

We would like to study whether individuals with higher levels of income have more prestigious jobs. Moreover, we would like to study whether professionals have more prestigious jobs than blue and white collar workers.

```
rm(list=ls())
#install.packages("car")
library("car")

## Loading required package: carData

data("Prestige")
help("Prestige")
str(Prestige)

## 'data.frame':    102 obs. of  6 variables:
##  $ education: num  13.1 12.3 12.8 11.4 14.6 ...
##  $ income   : int  12351 25879 9271 8865 8403 11030 8258 14163 11377 11023
...
##  $ women    : num  11.16 4.02 15.7 9.11 11.68 ...
##  $ prestige : num  68.8 69.1 63.4 56.8 73.5 77.6 72.6 78.1 73.1 68.8 ...
##  $ census   : int  1113 1130 1171 1175 2111 2113 2133 2141 2143 2153 ...
##  $ type     : Factor w/ 3 levels "bc","prof","wc": 2 2 2 2 2 2 2 2 2 2 ...

Prestige

##                           education income women prestige census type
## gov.administrators            13.11  12351 11.16     68.8   1113 prof
## general.managers              12.26  25879  4.02     69.1   1130 prof
## accountants                   12.77   9271 15.70     63.4   1171 prof
## purchasing.officers           11.42   8865  9.11     56.8   1175 prof
```

```
## chemists                      14.62   8403 11.68     73.5   2111 prof
## physicists                    15.64  11030  5.13     77.6   2113 prof
## biologists                    15.09   8258 25.65     72.6   2133 prof
## architects                    15.44  14163  2.69     78.1   2141 prof
## civil.engineers               14.52  11377  1.03     73.1   2143 prof
## mining.engineers              14.64  11023  0.94     68.8   2153 prof
## surveyors                     12.39   5902  1.91     62.0   2161 prof
## draughtsmen                   12.30   7059  7.83     60.0   2163 prof
## computer.programers           13.83   8425 15.33     53.8   2183 prof
## economists                    14.44   8049 57.31     62.2   2311 prof
## psychologists                 14.36   7405 48.28     74.9   2315 prof
## social.workers                14.21   6336 54.77     55.1   2331 prof
## lawyers                       15.77  19263  5.13     82.3   2343 prof
## librarians                    14.15   6112 77.10     58.1   2351 prof
## vocational.counsellors        15.22   9593 34.89     58.3   2391 prof
## ministers                     14.50   4686  4.14     72.8   2511 prof
## university.teachers           15.97  12480 19.59     84.6   2711 prof
## primary.school.teachers       13.62   5648 83.78     59.6   2731 prof
## secondary.school.teachers     15.08   8034 46.80     66.1   2733 prof
## physicians                    15.96  25308 10.56     87.2   3111 prof
## veterinarians                 15.94  14558  4.32     66.7   3115 prof
## osteopaths.chiropractors      14.71  17498  6.91     68.4   3117 prof
## nurses                        12.46   4614 96.12     64.7   3131 prof
## nursing.aides                  9.45   3485 76.14     34.9   3135   bc
## physio.therapsts              13.62   5092 82.66     72.1   3137 prof
## pharmacists                   15.21  10432 24.71     69.3   3151 prof
## medical.technicians           12.79   5180 76.04     67.5   3156   wc
## commercial.artists            11.09   6197 21.03     57.2   3314 prof
## radio.tv.announcers           12.71   7562 11.15     57.6   3337   wc
## athletes                      11.44   8206  8.13     54.1   3373 <NA>
## secretaries                   11.59   4036 97.51     46.0   4111   wc
## typists                       11.49   3148 95.97     41.9   4113   wc
## bookkeepers                   11.32   4348 68.24     49.4   4131   wc
## tellers.cashiers              10.64   2448 91.76     42.3   4133   wc
## computer.operators            11.36   4330 75.92     47.7   4143   wc
## shipping.clerks                9.17   4761 11.37     30.9   4153   wc
## file.clerks                   12.09   3016 83.19     32.7   4161   wc
## receptionsts                  11.04   2901 92.86     38.7   4171   wc
## mail.carriers                  9.22   5511  7.62     36.1   4172   wc
## postal.clerks                 10.07   3739 52.27     37.2   4173   wc
## telephone.operators           10.51   3161 96.14     38.1   4175   wc
## collectors                    11.20   4741 47.06     29.4   4191   wc
## claim.adjustors               11.13   5052 56.10     51.1   4192   wc
## travel.clerks                 11.43   6259 39.17     35.7   4193   wc
## office.clerks                 11.00   4075 63.23     35.6   4197   wc
## sales.supervisors              9.84   7482 17.04     41.5   5130   wc
## commercial.travellers         11.13   8780  3.16     40.2   5133   wc
## sales.clerks                  10.05   2594 67.82     26.5   5137   wc
## newsboys                       9.62    918  7.00     14.8   5143 <NA>
## service.station.attendant      9.93   2370  3.69     23.3   5145   bc
```

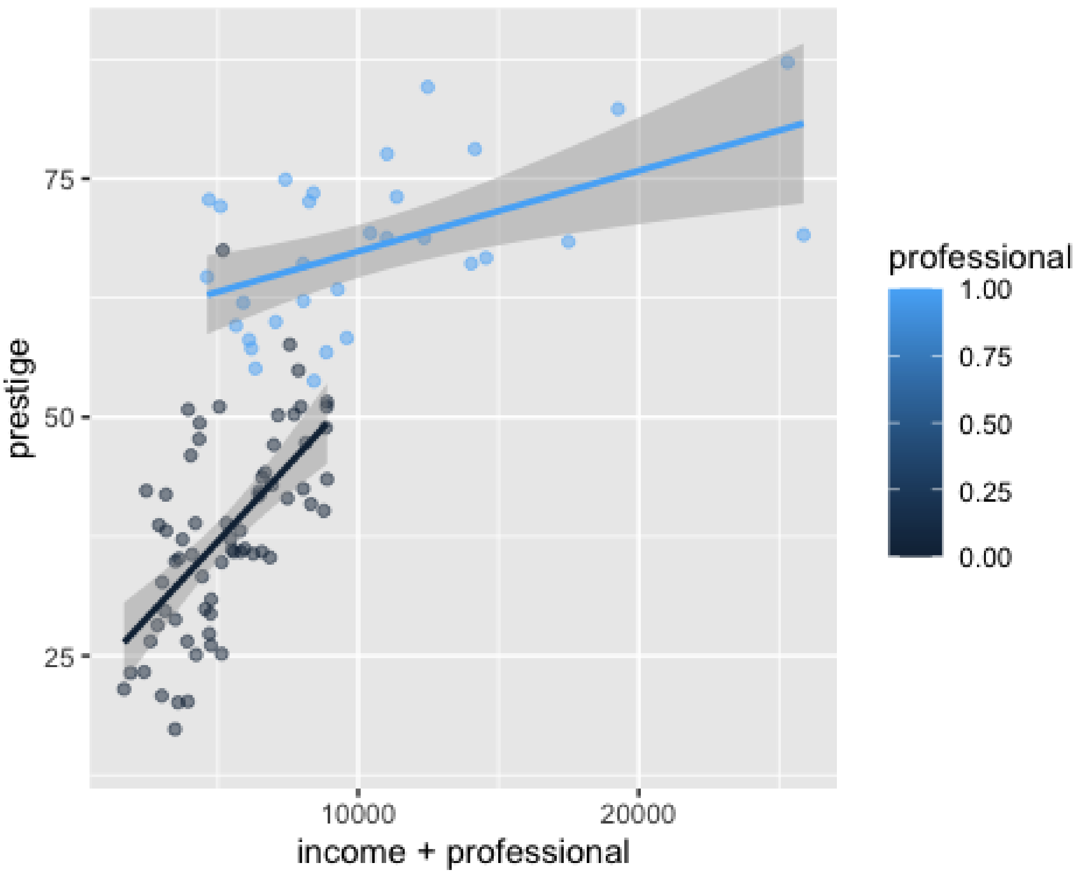
```
## insurance.agents          11.60   8131 13.09     47.3   5171   wc
## real.estate.salesmen      11.09   6992 24.44     47.1   5172   wc
## buyers                    11.03   7956 23.88     51.1   5191   wc
## firefighters               9.47   8895  0.00     43.5   6111   bc
## policemen                 10.93   8891  1.65     51.6   6112   bc
## cooks                      7.74   3116 52.00     29.7   6121   bc
## bartenders                 8.50   3930 15.51     20.2   6123   bc
## funeral.directors         10.57   7869  6.01     54.9   6141   bc
## babysitters                9.46    611 96.53     25.9   6147 <NA>
## launderers                 7.33   3000 69.31     20.8   6162   bc
## janitors                   7.11   3472 33.57     17.3   6191   bc
## elevator.operators         7.58   3582 30.08     20.1   6193   bc
## farmers                    6.84   3643  3.60     44.1   7112 <NA>
## farm.workers               8.60   1656 27.75     21.5   7182   bc
## rotary.well.drillers       8.88   6860  0.00     35.3   7711   bc
## bakers                     7.54   4199 33.30     38.9   8213   bc
## slaughterers.1             7.64   5134 17.26     25.2   8215   bc
## slaughterers.2             7.64   5134 17.26     34.8   8215   bc
## canners                    7.42   1890 72.24     23.2   8221   bc
## textile.weavers            6.69   4443 31.36     33.3   8267   bc
## textile.labourers          6.74   3485 39.48     28.8   8278   bc
## tool.die.makers           10.09   8043  1.50     42.5   8311   bc
## machinists                 8.81   6686  4.28     44.2   8313   bc
## sheet.metal.workers        8.40   6565  2.30     35.9   8333   bc
## welders                    7.92   6477  5.17     41.8   8335   bc
## auto.workers               8.43   5811 13.62     35.9   8513   bc
## aircraft.workers           8.78   6573  5.78     43.7   8515   bc
## electronic.workers         8.76   3942 74.54     50.8   8534   bc
## radio.tv.repairmen        10.29   5449  2.92     37.2   8537   bc
## sewing.mach.operators      6.38   2847 90.67     28.2   8563   bc
## auto.repairmen             8.10   5795  0.81     38.1   8581   bc
## aircraft.repairmen        10.10   7716  0.78     50.3   8582   bc
## railway.sectionmen         6.67   4696  0.00     27.3   8715   bc
## electrical.linemen         9.05   8316  1.34     40.9   8731   bc
## electricians               9.93   7147  0.99     50.2   8733   bc
## construction.foremen       8.24   8880  0.65     51.1   8780   bc
## carpenters                 6.92   5299  0.56     38.9   8781   bc
## masons                     6.60   5959  0.52     36.2   8782   bc
## house.painters             7.81   4549  2.46     29.9   8785   bc
## plumbers                   8.33   6928  0.61     42.9   8791   bc
## construction.labourers     7.52   3910  1.09     26.5   8798   bc
## pilots                    12.27  14032  0.58     66.1   9111 prof
## train.engineers            8.49   8845  0.00     48.9   9131   bc
## bus.drivers                7.58   5562  9.47     35.9   9171   bc
## taxi.drivers               7.93   4224  3.59     25.1   9173   bc
## longshoremen               8.37   4753  0.00     26.1   9313   bc
## typesetters               10.00   6462 13.58     42.2   9511   bc
## bookbinders                8.55   3617 70.87     35.2   9517   bc
```

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.)**

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

(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 × dummy interaction.)**

```
reg1= lm(data = Prestige,  prestige ~ income+professional)
reg1

##
## Call:
## lm(formula = prestige ~ income + professional, data = Prestige)
##
## Coefficients:
##  (Intercept)        income  professional
##    30.618334      0.001371     22.757000

library(ggplot2)
ggplot(Prestige, aes(income+professional, prestige, group = professional, col
our = professional)) +
  geom_point(alpha = 0.5, aes(colour = professional)) +
  geom_smooth(method = "lm", aes(colour = professional))

## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 4 rows containing non-finite values (stat_smooth).

## Warning: Removed 4 rows containing missing values (geom_point).
```



Created regression, visualised data in ggplot. Added in colour for profession to allow visualisation of professional (1=blue) and non-professional (0=grey) and how it interacts with prestige.

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

```
reg1$coefficients

##  (Intercept)       income professional
## 30.618333810  0.001370625 22.756999857
```

${0} = 30.618333810 $ ${1} = 0.001370625 $ $_{2} = 22.756999857 $ $x_1 = income\ x_2 = professional\ y = prestige$

$ = beta_{0} + beta_{1}x_{1} +beta_{2}x_{2} $ $\hat{y} = 30.618333810 + 0.001370625\timesx_{1} + 22.756999857\timesx_{2}$

(d) Interpret the coefficient for income.

For every one point increase in the Pineo-Porter prestige score, the average income of incumbents increases by $0.001371

(e) Interpret the coefficient for professional.

This is a highly positive regression coefficient. Since professional jobs are coded as one this indicates there is a strong positive relationship between professional jobs and job prestige. If the coefficient was negative it would indicate that jobs that are coded as 0 (non-professional jobs) were the stronger relationship with prestige.

(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 yˆ associated with a $1,000 increase in income based on your answer for (c).

β_0 = 30.618333810

β_1= 0.001370625 $

β_2= 22.756999857 $

$x_1 = 1000 \;\; x_2 = 1 \;\; y = prestige$

$\hat{y}$= β_0 + $\beta_1 x_1$ +β_2 x_2

$\hat{y}$= 30.618333810 + 0.001370625*x_1} + 22.756999857*x_2

$$\hat{y} = 30.618333810 + 0.001370625 * 1000 + 22.756999857 * 1$$

$\hat{y}$= 30.618333810 + 0.001370625*1000 + 22.756999857*1

$\hat{y}$= 54.74596

$$\hat{y}$$

(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 yˆ based on your answer for (c).

professional

$\hat{y}$= 30.618333810 + 0.001370625* x_1 + 22.756999857*x_2

$$\hat{y} = 30.618333810 + 0.001370625 * 6000 + 22.756999857 * 1$$

```
30.618333810 + 0.001370625*6000 + 22.756999857*1

## [1] 61.59908
```

$\hat{y}$ = 30.618333810 + 0.001370625*6000 + 22.756999857*1

$\hat{y}$ = 61.59908

non-professional

$\hat{y}$= β_0 + $\beta_1 x_1$ +β_2 x_2

$\hat{y}$ = 30.618333810 + 0.001370625* x_1 + 22.756999857*x_2

$$\hat{y} = 30.618333810 + 0.001370625 * 6000 + 22.756999857 * 0$$

$\hat{y}$= 30.618333810 + 0.001370625 *6000 + 22.756999857*1 $

```
30.618333810 + 0.001370625*6000 + 22.756999857*0

## [1] 38.84208
```

$\hat{y}$= 38.84208 $

Difference between professional and non-professional prestige:

```
61.59908 - 38.84208

## [1] 22.757
```

$\hat{y}$ change = 22.757

Question 2: Political Science

Researchers are interested in learning the effect of all of those yard signs on voting preferences.1 Working with a campaign in Fairfax County, Virginia, 131 precincts were randomly divided into a treatment and control group. In 30 precincts, signs were posted around the precinct that read, "For Sale: Terry McAuliffe. Don't Sellout Virgina on November 5."

Below is the result of a regression with two variables and a constant. The dependent variable is the proportion of the vote that went to McAuliff's opponent Ken Cuccinelli. The first variable indicates whether a precinct was randomly assigned to have the sign against McAuliffe posted. The second variable indicates a precinct that was adjacent to a precinct in the treatment group (since people in those precincts might be exposed to the signs).

Impact of lawn signs on vote share Precinct assigned lawn signs (n=30) | 0.042 (0.016) Precinct adjacent to lawn signs (n=76) | 0.042 (0.013) Constant | 0.302 (0.011) Notes: R2=0.094, N=131

(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 α = .05).

H0: yard signs in a precinct do not affect vote share

HA: yard signs in a precinct do affect vote share

Get t-statistic so can get p-value.

Divide coefficient by standard error.

```
tstat2a<- 0.042/0.016
```

tstat2a = 2.625

Find p value pt(q, df, lower.tail=TRUE) q = t-score df = degrees of freedom (n-2) Multiply by 2 (2*pt) for two tailed

```
pval2a <- 2*pt(tstat2a, (30-2), lower.tail=TRUE)
```

pval2a = 1.98612

alpha = 0.05

The p-value is considerably higher than alpha, so we cannot reject the null hypothesis. The results if this study do not allow us to infer that yard signs in a precinct affect vote share.

(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 α = .05).

H0: Being next to precincts with these signs does not affect vote share

HA: Being next to precincts with these signs affects vote share

Adjacent to lawn sign (n = 76) coefficient (0.042) SE (0.013)

```
tstat2b<- 0.042/0.013
```

tstat2b = 3.230769

```
pval2b <- 2*pt(tstat2b, (76-2), lower.tail=TRUE)
```

pval2b = 1.998157

alpha = 0.05

The p-value is considerably higher than alpha, so we cannot reject the null hypothesis. The results if this study do not allow us to infer that being adjacent to precincts with yard signs affects vote share.

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

The constant coefficient indicates what the proportion of the vote that went to McAuliffe would be if there were no lawn signs in the precinct. It represents the intercept of the slope of the regression. It is likely a meaningful coefficient as it should have been derived from the control group in the experiment.

A constant regression coefficient of 0.302 indicates that within the control group (without signs) the proportion of the vote that goes to McAuliffe is 0.302 units of the metric they are using to measure the proportion of votes.

(Maybe it means 30.2% of the vote would already go to McAuliffe, but I am reluctant to assume that without having seen the actual data set.)

(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?

R squared = 0.094 R squared is a measure of how well the line fits the data points. A small difference in the predicted values and the observation values would result in a better-fitting line and an R squared statistic which is closer to 100%.

An r squared statistic of 0.094, 9.4% indicates that the line does not fit well and that the model (signage) explains only a small amount of the variation around McAuliffe's voteshare in these precincts.