Michael Rimmey

Professor Jeffrey Ziegler

Quantitative Political Methods

18 April 2019

Problem Set 4

Question 1

**Null Hypothesis**: Officers not more likely to solicit bribe based on social class.

**Alternative Hypothesis**: Officers not more likely to solicit bribe based on social class.

1 a)

For question 1, we must conduct a **chi-squared test** by hand (or in R). To do this, we first subtract the mean from each column (mean of 7, columns of 14, 7, 7, 7, 6, 1). We then square each sum and get 49, 0, 0, 0, -1, and -36. Add those numbers together, and we get an **X2** of 12.

Code:

1. policedata = matrix(c(14, 6, 7, 7, 7, 1), nrow = 2, byrow = TRUE)
2. policedata
3. policedatatExpected = matrix(c(((27/42)\*21), ((27/42)\*13), ((27/42)\*8), ((15/42)\*21), ((15/42)\*13), ((15/42)\*8)), nrow = 2, byrow = TRUE)
4. Residualpolicedata = policedata - policedatatExpected
5. Residualpolicedata
6. chisquareteststat = sum((Residualpolicedata^2)/policedatatExpected)
7. chisquareteststat
8. chisq.test(policedata)
9. StandResPolice = Residualpolicedata/sqrt(policedatatExpected)
10. StandResPolice

1 b)

To get the **P-value**, we simply plug these numbers into R with the pchisq command.

pchisq(12, 5, lower.tail = FALSE)

The P-value we get from doing this is 0.03478778. Therefore, we can reject the null hypothesis since this value is less than 0.05.

1 c)

Standardized Residuals:

Upper Class: 0.136, -0,185, 0.819

Lower Class: -0.183, 1.094, -1.099

1 d)

All values fall within frequency of what can be expected (-2 to 2).

Question 2

**Null Hypothesis**: Yard signs in a given precinct do not affect vote share.

**Alternative Hypothesis:** Yard signs in a given precinct affect vote share.

2 a)

Null Hypothesis cannot be rejected. This is because the R-squared value is too low.

2 b)

Being in an adjacent precinct does not affect the respondents in any significant way.

2 c)

The small impact being made seems to imply that the coefficient should be larger.

2 d)

The no relation-correlation model would suit this problem well.

Question 3

For question 3, we must conduct a hypothesis test. We must first remember the 5 steps for hypothesis tests: Make assumptions about the data, state the null and alternative hypotheses, calculate a test statistic, calculate a P-value, and draw a conclusion based on that P-value.

3 a)

**Null Hypothesis:** Women do not promote different policies than men.

**Alternative Hypothesis:** Women promote different policies than men.

3 b)

1. femalepols = read.csv(url("https://raw.githubusercontent.com/kosukeimai/qss/master/PREDICTION/women.csv"))
2. femalepols
3. reserved = femalepols$reserved
4. water = femalepols$water
5. regressionfemalepols = lm(water ~ reserved, data=femalepols)
6. summary(regressionfemalepols)
7. summary(femalepols)

The resulting P-value allows us to reject the null hypothesis.

3 c)

With more women in power, approximately 9.25 more water facility projects were taken.

Question 4

1. install.packages(car)
2. library(car)
3. data(Prestige)
4. help(Prestige)
5. View(Prestige)
6. Prestige$professional = NA
7. Prestige$professional = Prestige$type
8. Prestige$professional = as.character(Prestige$professional)
9. Prestige$professional[ Prestige$professional=="prof" ] <- "1"
10. Prestige$professional[ Prestige$professional=="bc" ] <- "0"
11. Prestige$professional[ Prestige$professional=="wc" ] <- "0"
12. dummyint1 = "Prestige$income\*Prestige$professional"
13. lm4B = lm(Prestige$prestige ~ Prestige$income + Prestige$professional + (Prestige$income : Prestige$professional))
14. summary(lm4B)
15. scatterplot(Prestige$prestige ~ Prestige$income + Prestige$professional)

4 a)

(See Code)

4 b)

(See Code)

4 c)

Y = 21.14 + 0.0032 X-income + 37.78 X-professional - 0.0023 X-relations

4 d)

Prestige increases only after income is increased by a great deal.

4 e)

Interestingly, “profession” has a much greater impact on Prestige than “income.” Respondents who transitioned from Non-professional to professional jobs experienced an impactful rise in Prestige, whereas with “income,” the rise in prestige was much slower.

4 f)

Difference in Prestige: **3.2**

4 g)

Difference in Prestige: **37.78**

Question 5

1. library("faraway")
2. data("newhamp")
3. colnames(newhamp)
4. help(lm)
5. votesysonly = lm(newhamp$pObama ~ newhamp$votesys)
6. summary(votesysonly)
7. votesyspovrate = lm(newhamp$pObama ~ newhamp$votesys + newhamp$povrate)
8. VotPovPci = lm(newhamp$pObama ~ newhamp$votesys + newhamp$povrate + newhamp$pci)
9. VPPD = lm(newhamp$pObama ~ newhamp$votesys + newhamp$povrate + newhamp$pci + newhamp$Dean)
10. VPPDW = lm(newhamp$pObama ~ newhamp$votesys + newhamp$povrate + newhamp$pci + newhamp$Dean + newhamp$white)
11. DeanOnly = lm(newhamp$pObama ~ newhamp$Dean)
12. summary(votesysonly)
13. summary(votesyspovrate)
14. summary(VotPovPci)
15. summary(VPPD)
16. summary(VPPDW)
17. summary(DeanOnly)

5 a)

(See Code)

5 b)

Model 1: R-squared = 0.083; R-adjusted = 0.080

Model 2: R-squared = 0.0897; R-adjusted = 0.083

Model 3: R-squared = 0.244; R-adjusted = 0.236

Model 4: R-squared = 0.509; R-adjusted = 0.502

Model 5: R-squared = 0.509; R-adjusted = 0.500

Model 6: R-squared = 0.418; R-adjusted = 0.416

5 c)

Highest R-squared values: Models 4 and 5 (tie).

Question 6

6 a)

1. setwd("GitHub/QPMspring2019")
2. read.csv("problemSets/incumbents\_subset.csv")
3. Incumbents <- read.csv("problemSets/incumbents\_subset.csv")
4. colnames(Incumbents)

7. scatterplot(Incumbents$voteshare ~ Incumbents$difflog)
8. ModelOne <- lm(voteshare ~ difflog, data=Incumbents)
9. abline(ModelOne, col="Blue", lwd=2)
10. summary(ModelOne)
11. ResidualOne = (residuals(ModelOne))
13. plot(residuals(ModelOne) ~ fitted(ModelOne), data=Incumbents)
14. abline(h=0)
15. summary(ResidualOne)
17. coef(ModelOne)
18. abline(a=coef(ModelOne)[1], b=coef(ModelOne)[2],
19. lwd=2, col="Blue")

Prediction Equation: ybar = **0.579031 + 0.041666** X-difflog

For question 6 a), I ran a regression of the two variables, “voteshare” and “difflog” after assigning the dataset to the term “Incumbents.”

6 b)

1. plot(Incumbents$voteshare, Incumbents$difflog)
2. ModelOne <- lm(voteshare ~ difflog, data=Incumbents)
3. abline(ModelOne, col="Blue", lwd=2)
4. summary(ModelOne)
5. ResidualTwo = (residuals(ModelTwo))
7. residuals <- plot(residuals(ModelTwo) ~ fitted(ModelTwo), data=Incumbents)
8. abline(h=0)
9. summary(ResidualTwo)
11. coef(ModelTwo)
12. abline(a=coef(ModelTwo)[1], b=coef(ModelTwo)[2],
13. lwd=2, col="Blue")

ybar = **0.507583 + 0.023837** X-difflog

6 c)

1. scatterplot(Incumbents$presvote ~ Incumbents$voteshare,
2. main="Model Three", xlab="Presidential Vote", ylab="Vote Share")
3. ModelThree <- (lm(presvote ~ voteshare, data=Incumbents))
4. abline(ModelThree, col="Red", lwd=2)
5. summary(ModelThree)
7. ResidualThree = (residuals(ModelThree))
9. residuals <- plot(residuals(ModelThree) ~ fitted(ModelThree), data=Incumbents)
10. abline(h=0)
11. summary(ResidualThree)
13. coef(ModelThree)
14. abline(a=coef(ModelThree)[1], b=coef(ModelThree)[2],
15. lwd=2, col="Blue")

ybar = **0.20363 + 0.53042** X-voteshare

6 d)

1. scatterplot(ResidualOne ~ ResidualTwo,
2. main="Residual Comparison", xlab="Residual Two", ylab="Residual One")
3. ResidualComparison <- (lm(ResidualOne ~ ResidualTwo, data=Incumbents))
4. abline(ModelThree, col="Red", lwd=2)
5. summary(ModelThree)

ybar = **4.860e18 + 2.569e-01** X-Residual Two

6 e)

1. Incumbents$difflogpresvote = Incumbents$difflog + Incumbents$presvote
3. scatterplot(Incumbents$voteshare ~ Incumbents$difflogpresvote,
4. main="Model Five", xlab="Diff Log and Presidential Vote", ylab="Vote Share")
5. ModelFive <- (lm(presvote ~ voteshare, data=Incumbents))
6. abline(ModelFive, col="Red", lwd=2)
7. summary(ModelFive)

ybar = **0.5554993 + 0.0419046** X-difflogpresvote

The “max” residuals are exactly the same in d and e.

**COLLABORATED WITH LUKE EHRENSTROM AND SINCLAIR BOWMAN**