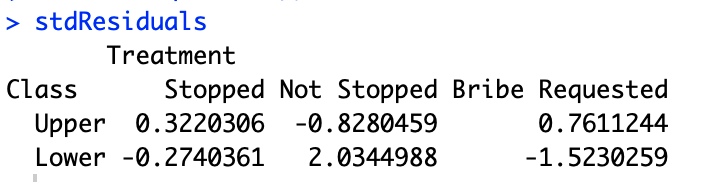
Andrew Celli

PS4

**Problem 1**

1. 3.7912
2. Cannot reject null
3. 
4. It shows us that the largest (absolute and relative) deviations are in the lower class not stopped and bribe requests.

#PROBLEM 1 ---------

FOs = matrix( data = c(14,6,7,7,7,1), nrow = 2, byrow = TRUE)

dimnames(prob1data) <- list(Class=c("Upper", "Lower"),

Treatment=c("Stopped", "Not Stopped", "Bribe Requested"))

prob1data

FEs =

matrix(data = c((27/42)\*21 , (27/42)\*13, (27/42)\*8,

(15/42)\*21, (15/42)\*13, (15/42)\*8), nrow = 2)

X2 = sum((((FOs-FEs)^2)/FEs))

chisq.test(prob1data)

chisq.test(FOs)

pchisq(X2, df = 2, lower.tail = FALSE)

rowprop = matrix(data = c(27/42, 27/42, 27/42, 15/42, 15/42, 15/42), nrow =2, byrow = TRUE)

colprop = matrix(data = c(21/42, 13/42, 8/42, 21/42, 13/42, 8/42), nrow =2, byrow = TRUE)

stdResiduals = matrix(

data = c((FOs - FEs)/(sqrt(FEs\*(1-rowprop)\*(1-colprop))))

, byrow = TRUE, nrow = 2)

dimnames(stdResiduals) <- list(Class=c("Upper", "Lower"),

Treatment=c("Stopped", "Not Stopped", "Bribe Requested"))

stdResiduals

**Problem 2**

**Problem 3**

1. Null : Water quality not effected by reservation policy or Beta = 0; alt Beta != 0
2. model
3. A reserved policy leads to a roughly 9 pt increase in water quality

DataP3 = read.csv(url("https://raw.githubusercontent.com/kosukeimai/qss/master/PREDICTION/women.csv"))

womenData = DataP3$water[DataP3$reserved ==1]

noWomenData = DataP3$water[DataP3$reserved ==0]

mean(noWomenData)

mean(womenData)

model = lm(water ~ reserved, data = DataP3)

summary(model)

model

plot((model))

model

cor(DataP3)

**Problem 4**

3. 1. 37.66 + 30.19x
   2. 27.141176 + 0.002897x
   3. 30.062787 + 22.22977(TYPE) + 0.001473 (INCOME)
4. A 1 increase in income leads to a 0.002897 increase in prestige
5. A 1 increase in type leads to a 30.19 increase in prestige
6. 2.897 (for all); 0.8452 (for profs.)
7. 61.13056 :: 30.062787 + 22.22977 + 0.001473 (6000)

install.packages(car)

library(car)

data(Prestige)

help(Prestige)

?ifelse

Prestige$type <- as.character(Prestige$type)

Prestige$type[Prestige$type == "prof"] <- "b"

Prestige$type[Prestige$type == "b"] <- "1"

Prestige$type[Prestige$type == "wc"] <- "a"

Prestige$type[Prestige$type == "a"] <- "0"

Prestige$type[Prestige$type == "bc"] <- "c"

Prestige$type[Prestige$type == "c"] <- "0"

Prestige$type[is.na(Prestige$type)] <- "na"

Prestige$type[Prestige$type == "na"] <- "0"

Prestige$type <- as.numeric(Prestige$type)

mod1 = lm(Prestige$prestige ~ Prestige$type)

mod2 = lm(Prestige$prestige ~ Prestige$income)

mod3 = lm(Prestige$prestige ~ Prestige$type + Prestige$income)

mod1

mod2

mod3

cor(Prestige)

plot(modpre)

mean(Prestige$prestige[Prestige$type==0])

lm(Prestige$prestige ~ Prestige$type + Prestige$income)

0.002897\*1000

lm(Prestige$prestige[Prestige$type==1]~Prestige$income[Prestige$type==1])

0.0008452\*1000

30.062787 + 22.22977 + 0.001473\*6000

**Problem 5**

* 2. Model R2; R2 adjusted

1. .083; .08
2. .089; .08
3. .244; .23
4. .509; .50
5. .509; .50
6. .418; .41
   1. R squared increases with more data, as the model becomes more accurate. But it is clear that Dean data is most helpful (model 6)

library("faraway")

data("newhamp")

colnames(newhamp)

mod1 = lm(newhamp$pObama ~ newhamp$votesys)

mod2 = lm(newhamp$pObama ~ newhamp$votesys + newhamp$povrate)

mod3 = lm(newhamp$pObama ~ newhamp$votesys + newhamp$povrate + newhamp$pci)

mod4 = lm(newhamp$pObama ~ newhamp$votesys + newhamp$povrate + newhamp$pci + newhamp$Dean)

mod5 = lm(newhamp$pObama ~ newhamp$votesys + newhamp$povrate + newhamp$pci + newhamp$Dean + newhamp$white)

mod6 = lm(newhamp$pObama ~ newhamp$Dean)

summary(mod1)

summary(mod2)

summary(mod3)

summary(mod4)

summary(mod5)

summary(mod6)

plot(newhamp$Dean, newhamp$pObama)

abline(a = 0.22288, b = 0.51053)

**Problem 6**

* 1. mod1
  2. A close up of a map

     Description automatically generated
  3. residualsMod1
  4. 0.579031 + 0.041666(X)

1. 1. mod2
   2. A close up of a map

      Description automatically generated
   3. residualsMod2
   4. 0.507583+ 0.023837 (X)
2. 1. mod3
   2. A close up of a map

      Description automatically generated
   3. 0.441330+ 0.388018 (X)
3. 1. mod4
   2. A close up of a map

      Description automatically generated
   3. 0 + .2569 (x)
4. 1. mod5
   2. 0.4486442 + 0.0355431(difflog) + 0.2568770(presvote)
   3. Presvsote coefficient and the residual2 coefficient. Why you ask? I do not know.
   4. [thinking]

setwd("/Users/andrewcelli/Desktop")

data = read.csv("incumbents\_subset.csv" )

mod1 = lm(data$voteshare ~ data$difflog, data = data)

summary(mod1)

plot( data$difflog, data$voteshare )

abline(mod1, col = "red", )

residualsMod1 = residuals(mod1)

plot(residualsMod1)

#Mod2

mod2 = lm(data$presvote ~ data$difflog)

summary(mod2)

residualsMod2 = residuals(mod2)

plot( data$difflog, data$presvote )

abline(mod2, col = "red")

#Mod3

mod3 = lm(data$voteshare ~ data$presvote)

summary(mod3)

plot( data$presvote, data$voteshar )

abline(mod3, col = "red")

#Mod4

mod4 = lm(residualsMod1~residualsMod2)

summary(mod4)

plot(residualsMod2, residualsMod1)

abline(mod4, col = "red")

#Mod5

mod5 = lm(voteshare ~ difflog + presvote, data =data)

summary(mod5)