Problem set 2

Claire Mooney

October 2021

1 Introduction

2 Problem Set 2

3 Question 1

setwd("C:/Users/Student/Desktop/Stats 2 Assignment")
library(ggplot2) library(tidyverse)

4 Question 1 part A

upperclass ¡- c(14,6,7) lowerclass ¡- c(7,7,1) not$_s topped < -c(14, 7) bribed < -c()$

This was done by hand and returned a value of 5.73132381

Assign Observed Frequencies Fo$_1 < -14Fo_2 < -6Fo_3 < -7Fo_4 < -7Fo_5 < -7Fo_6 < -1$

Assign Expected Frequencies Fe$_1 < -((27/42) * 21)Fe_2 < -((27/42) * 13)Fe_3 < -((27/42) * 8)Fe_4 < -((15/42) * 21)Fe_5 < -((15/42) * 13)Fe_6 < -((15/42) * 8)$

Claculate chisq chisq ¡- (((Fo$_1 - Fe_1)^2/Fe_1) + ((Fo_2 - Fe_2)^2/Fe_2) + ((Fo_3 - Fe_3)^2/Fe_3) + ((Fo_4 - Fe_4)^2/Fe_4) + ((Fo_5 - Fe_5)^2/Fe_5) + ((Fo_6 - Fe_6)^2/Fe_6)$

5 Question 1 part B

calculating p value for the test statistic degrees of freedom is (rows-1)(columns-1) (3-1)(2-1)

pchisq(3.801141055, df = 2, lower.tail = F) this returns a p value of 0.1494833 we reject the null hypothesis

6 Question 1 part C roughwork

standardised residuals

roughwork

model ¡ - lm(”upperclass”, ”lowerclass”)

bribe$_c hisq < -chisq.test(mydata(tab$upperclass, mydata$lowerclass))bribe_c hisq$

mydata = read.table(”ExcelProblemSet2.csv”, header=TRUE, sep=’,’) view(mydata) mydata$upperclass < -factor(mydata$upperclass) mydata$lowerclass < -factor(mydata$lowerclass) chisq ¡- chisq.test(table(mydata$upperclass, mydata$lowerclass)) chisq chisq$residualschisq$stdres

7 Question 1 part 3 successful code

mydata = read.table(”ExcelProblemSet2.csv”, header=TRUE, sep=’,’)
view(mydata)
mydata$upperclass < -factor(mydata$upperclass) mydata$lowerclass < -factor(mydata$lowerclass)
chisq ¡- chisq.test(table(mydata$upperclass, mydata$lowerclass)) chisq chisq$residuals$

tab ¡- matrix(c(14, 6, 7, 7, 7, 1), ncol=3, byrow=TRUE)
colnames(tab) ¡- c(’Not Stopped’,’Bribed’,’Warned’) rownames(tab) ¡- c(’Upper Class’, ’Lower Class’) bribes$_d ata < -as.table(tab)bribes_d ata$
$chisq.test < -chisq.test(bribes_d ata)chisq.testchisq.test$residuals chisq.test$stdres$
$notstoppedbribedwarned$
$upperclass$0.1360828 $-$ 0.81537420.8189230$lowerclass$ $-$ 0.18257421.0939393 $-$ 1.0987005

lower class offered more bribes and warned more then was expected

8 Question 1 part d

we were suprised at the amount of people of upper class who were bribing
As the value was the furthest from the expectation small standardised residuals
tell us that the prediction line is a good fit for the data
The residuals can help you detect outliers in your results

9 Question 2

10 question 2 part 1

Null hypothesis that the reservation policy has no affect on irrigation
the alternative hypothesis is that the reservation does have an affect on irrigation

11 Question 2 part 2

economics ¡- read.csv(”https://raw.githubusercontent.com/kosukeimai/qss/master/PREDICTION/women

View(economics) lm1 ¡- lm(economics*irrigation economics*reserved)
lm1
summary(lm1)
p value is 0.7422 and therefore is bigger than 0.5 p value we fail to reject the null hypothesis
reservation policy has a negative impact on irrigation system

12 Question 2 part 3

p value is 0.7422 and therefore is bigger than 0.5 p values therefore we fail to reject the null hypothesis therefore reservation policy has a negative impact on irrigation system

13 Question 3

library(ggplot2) library(tidyverse)
data(”fruit”) data(”fruitflies”)

FruitFlies ¡- read.csv(”https://www.zoology.ubc.ca/ bio501/R/data/fruitflies.csv”) view(FruitFlies)

dat¡-read.csv(”http://stat2.org/datasets/FruitFlies.csv”) attach(dat)

install.packages(”tidyverse”) library(ggplot2)

all dollar signs are removed for the purpose of latex

14 question 3 part 1

plot(fruitflies-longevity.days)

plot(fruitflies-Longevity, main = ”Scatter Plot of Two variables”, xlab = ”Predictor Variable on X axis”, ylab = ”Target Variable on y axis”)

15 Question 3 part 2

plot(fruitflies-Longevity, fruitflies-Thorax, main = "Scatter Plot of Two variables", xlab = "Predictor Variable on X axis", ylab = "Target Variable on y axis")
There is a weak linear relationship between the two

ggplot(aes(Longevity, Thorax), data = fruitflies) + geom$_{p}oint()$

cor(fruitflies-Longevity, fruitflies-Thorax)
the correlation coeffcient between the two variables is 0.6364835

16 Question 3 part 3

lm(fruitflies-Longevity fruitflies$Thorax$)

lm2 ¡- lm(fruitflies-Longevity fruitflies-Thorax)
it return an intercept values for fruitflies-Thorax of -61.05 and 144.33
this gives the intercept and the slope which shows us a steep slope in the data however not an extremetly steep slope

summary(lm2)
plot(fruitflies-Longevity, fruitflies-Thorax)
abline(lm(fruitflies-Longevity fruitflies-Thorax), col = "red")
the slope shows a positive correlation between longevity and thorax however the data is somewhat unevenly distributed

17 Question 3 part 4

ggplot(aes(Longevity, Thorax), data = fruitflies) + geompoint()+ geomsmooth()

A strong linear relationship is when the observations fluctuate Tightly around the fitted line
There is a linear relationship between the two but observations do not consistently gather around line so it is not significant

18 Question 3 part 5

class(lm2) predict(lm (fruitflies-Longevity fruitflies-Thorax), interval = "confidence", level = 90) confint(lm (fruitflies-Longevity fruitflies-Thorax), interval

= ”confidence”, level = 90)

fit ¡- (lm (fruitflies-Longevity   fruitflies-Thorax, fruitflies))
summary(fit)
lm (formula = fruitflies-Longevity   fruitflies-Thorax, data = fruitflies)
confint(fit,Longevity,level=90)

I was unable to get the data to successfully run the code so I was unable to find the confidence interval - I have shown some above different attempts of code I tried
In terms of formula for the confidence interval I would calculated confidence = slope
and the margin of error. The margin of error = critical value and standard error

19 Question 3 part 6

having been unable to calculate the confidence interval meant I was unsure how to proceed
I understand the predict function will give you the coefficient of the intercept and the slope
you use the predict function when you have new data what y variables would be given new x values
This would have been the code I probably would have written if I could run it

class(fruitflies-Longevity) individual ¡- data.frame(lifespan = c(runif(fruitflies-Thorax = 0.8)))
predict(fruitflies-Longevity, newdata = individual)

average ¡- data.frame(lifespan = c(runif(fruitflies-Thorax= 0.8)))
predict(lm(fruitflies-Longevity   fruitflies-Thorax), newdata = average, se.fit = T))

20 Question 3 part 7

this is the code I would have used to plot

plot(new$_df-percollege, predict(fruitflies-Thorax, newdata = individual)$)