PS01 Homework

Lucy Kinnear

September 2021

1 Question 1: Education

1. The 90% confidence interval for the average student IQ in school is 93 - 104 (rounded to nearest integer).

y ¡- c(105, 69, 86, 100, 82, 111, 104, 110, 87, 108, 87, 90, 94, 113, 112, 98, 80, 97, 95, 111, 114, 89, 95, 126, 98)

t.test(y, conf.level = 0.9)

One Sample t-test

data: y t = 37.593 df = 24 p-value ¡ 2.2e-16

alternative hypothesis: true mean is not equal to 0

90 percent confidence interval: 93.95993 102.92007

mean of x: 98.44

2. We cannot reject the null hypothesis that the average student IQ in the school is lower than the average IQ in schools across the country.

t.test(y, mu = 100, alternative = "greater")

Null hypothesis: the average student IQ in this school is lower than the average IQ in schools across the country.

I ran a t-test using the sample y, setting the population average to 100. We know IQ is normally distributed, and because we are interested only in one tail of the distribution (the teacher wants to know if her students' average is higher) we insert "alternative = "greater" (i.e. the opposite of our null). We do not need to insert information about the confidence interval because it defaults to 95%.

The t-test returns a p value of 0.7215, meaning we should not be at all surprised to find a mean IQ of 98.44 (the sample mean). The large p value means we cannot reject the null; i.e. the average IQ of the school's pupils does not appear to exceed the average IQ of the country's students.

2 Question 2: Political Economy

2.1 Figure 1 below illustrates the relationships between Y, X1, X2, and X3. X1 and X3 have the strongest correlation/covariation. This means per capita

personal income in state and the number of people per thousand residing in urban areas in state are highly correlated.

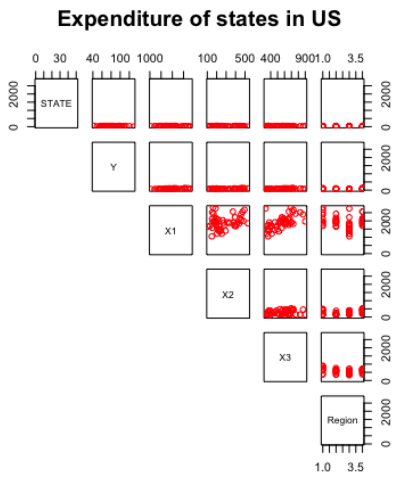

Figure 1: Y, X1, X2, X3 Relationships

2.2 On average, region 4 has the highest per capita expenditure on housing assistance, illustrated with Figures 2 and 3.

2.3 Figure 4 plots the relationship between Y and X1, with different markers for each Region. Personal income and state expenditure are positively correlated, with line of best fit illustrated on the plot.

This is the code I used to attain these results:

```
str␣(expenditure)
lines(expenditure$Y)
lines(expenditure$X1)
lines(expenditure$X2)
lines(expenditure$X3)
plot(expenditure,␣ylim=range(expenditure$Y,␣expenditure$X1,expenditure$X2,expenditure$X3),␣c

plot(expenditure$Region,␣expenditure$Y,
␣␣␣␣␣␣pch␣=␣16,␣frame␣=␣TRUE,
␣␣␣␣␣␣xlab␣=␣"Region",␣ylab␣=␣"Y",␣col␣=␣"red")

scatter.smooth(␣expenditure$X1,␣expenditure$Y,␣xlab␣=␣'Personal␣Income',␣ylab␣=␣'expenditure
expenditure',␣col␣=␣c("blue",␣"red",␣"orange",␣"black"),
pch␣=␣c(0,1,2,3))
```

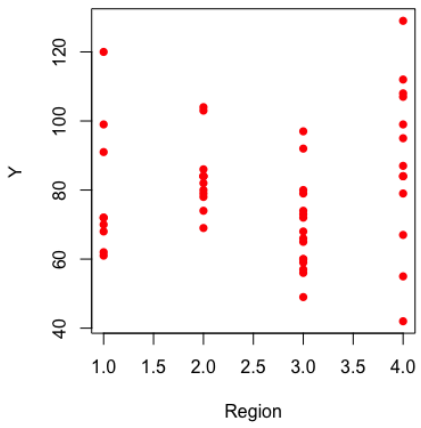

Figure 2: Region expenditure of Y

```
install.packages("ggplot2")
library(ggplot2)
#change␣expenditure␣to␣data␣frame
data=as.data.frame(expenditure[,c(2,6)])
data$Region␣=␣as.factor(data$Region)
mode(data$Region)

ggplot(aes(y␣=␣Y,␣x␣=␣Region,␣fill=Region),␣data␣=␣data)+
geom_boxplot()+ggtitle("Box␣plots␣of␣Expenditure␣by␣Region")
```

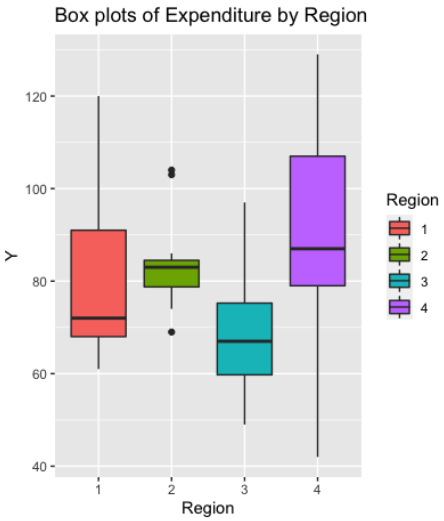

Figure 3: Box Plot of Y Expenditure by Region

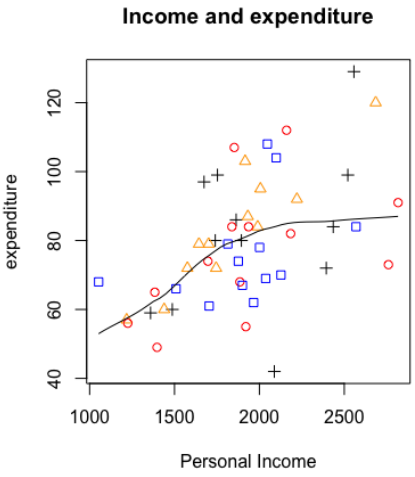

Figure 4: Income and Expenditure