Homework 10

Alizeh Azhar

11/8/2021

Problem 1: OIS Exercise 7.11, page 261 (9 points)

Georgianna claims that in a small city renowned for its music school, the average child takes less than 5 years of piano lessons. We have a random sample of 20 children from the city, with a mean of 4.6 years and a standard deviation of 2.2 years.

(a) Evaluate Georgianna's claim (or that the opposite might be true) using a hypothesis test (4 pts).
Solution
Prepare
The hypothesis to be tested is: H_0 : On average, a child takes less than 5 years of piano lessons VS H_1 : On average, a child takes more than 5 years of piano lessons
Check
The random sample is less than 30 and there are no outliers in the data. We assume that the data come from a nearly normal distribution.

```
#calculate
test_statistic <- (4.6-5)/(2.2/sqrt(20));test_statistic
```

```
## [1] -0.8131156
```

```
d_freedom <- 20-1
```

```
p_value <- pt(test_statistic, df=d_freedom, lower.tail = F);p_value
```

```
## [1] 0.786888
```

Conclusion
The p- value is 0.787 which is greater than 0.05, and we reject the null hypothesis. There is not enough evidence to support the claim that in a small city renowned for its music school, the average child takes less than 5 years of piano lessons.

(b) Construct a 95% confidence interval for the number of years students in this city take piano lessons, and interpret it in the context of the data (2 pts).
Solution

```
#calculate
se <- 2.2/sqrt(20)
t_star <- qt(0.975, df = 19)
MoE <- t_star*se
4.6-MoE
```

```
## [1] 3.570368
```

```
4.6+MoE
```

```
## [1] 5.629632
```

conclusion
We are 95% confident that the true mean of children falls between 3.5703683 and 5.6296317

Problem 2: OIS Exercise 6.26a, page 227 (6 points)

A news article reports that "Americans have differing views on two potentially inconvenient and invasive practices that airports could implement to uncover potential terrorist attacks." This news piece was based on a survey conducted among a random sample of 1,137 aduts nationwide, where one of the questions on the survey was "Some airports are now using 'full-body' digital x-ray machines to electronically screen passengers in airport security lines. Do you think these new x-ray machines should or should not be used at airports?" Below is a summary of responses based on party affiliation

Conduct an appropriate hypothesis test evaluating whether there is a difference in the proportion of Republicans and Democrats who think that the full-body scans should be applied in airports.
Solution
Prepare
H_0 : There is no difference in the proportion of Republicans and Democrats who think that the full-body scans should be applied in airports. H_1 : There is a difference in the proportion of Republicans and Democrats who think that the full-body scans should be applied in airports

```
#Check
(563/687)*687
```

```
## [1] 563
```

```
(1-563/687)*687
```

```
## [1] 124
```

```
# calculate
prop.test(x=c(264,299), n=c(318,369), conf.level = 0.95, correct = F)
```

```
##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  c(264, 299) out of c(318, 369)
## X-squared = 0.45685, df = 1, p-value = 0.4991
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.03758321  0.07736436
## sample estimates:
##    prop 1    prop 2
## 0.8301887 0.8102981
```

Conclusion
The p-value is 0.4991 which is greater than 0.05. We fail to reject the null hypothesis. There is not enough evidence to say that the proportions of republicans and democrats who think that the full-body scans should be applied in airports are different.

Problem 3 (10 points)

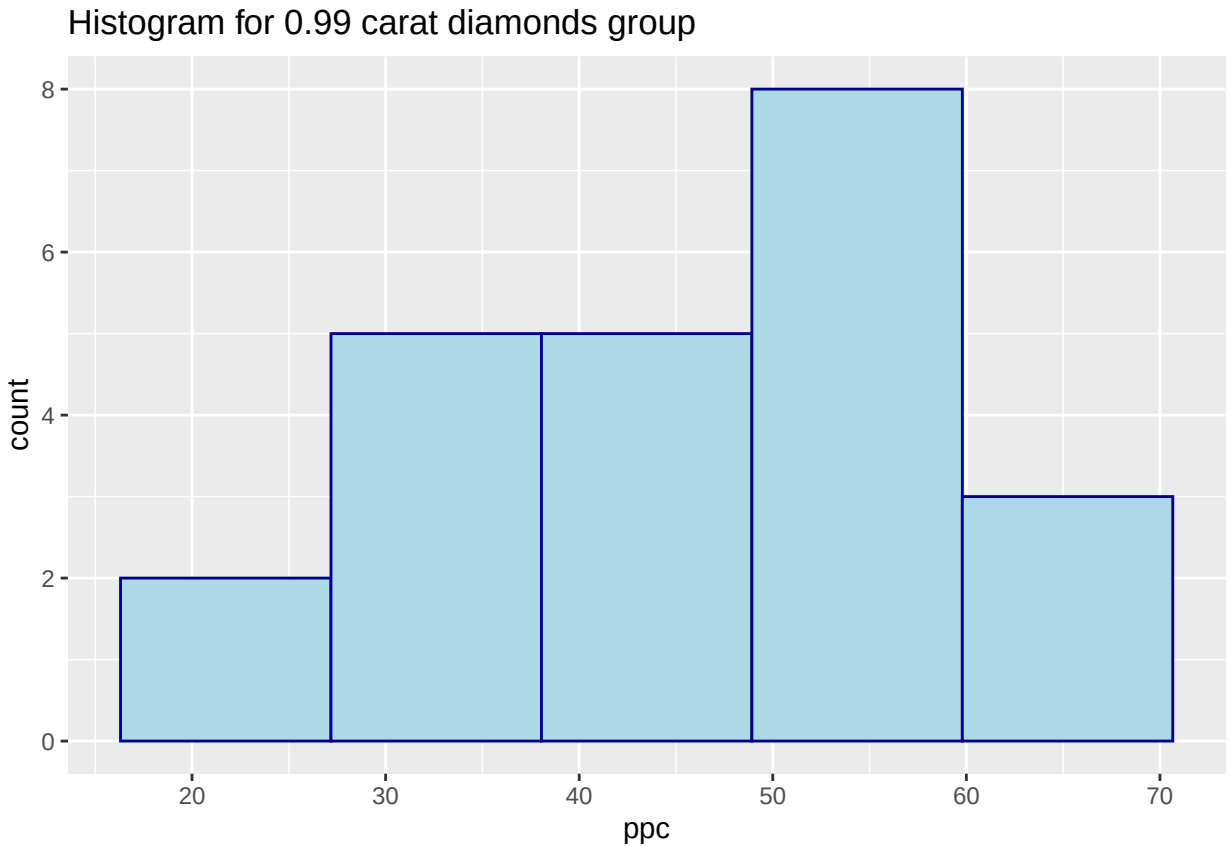
Prices of diamonds are determined by what is known as the 4 C's: cut, clarity, color, and carat weight. The prices of diamonds of up as the carat weight increases, but the increase is not smooth. For example, the difference between the size of a 0.99 carat diamond and a 1 carat diamond is undetectable to the naked human eye, but the price of a 1 carat diamond tends to be much higher than the price of a 0.99 carat diamond. In

this question, we use two random samples of diamonds with carat weights 0.99 and 1. In order to be able to compare equivalent units, we calculate the price per carat, by dividing the price for each diamond by 100 times its weight in carats. That is, for a 0.99 carat diamond, we divide the price by 0.99. For a 1 carat diamond, we divide the price by 100. The variable to analyze is stored in the variable ppc.

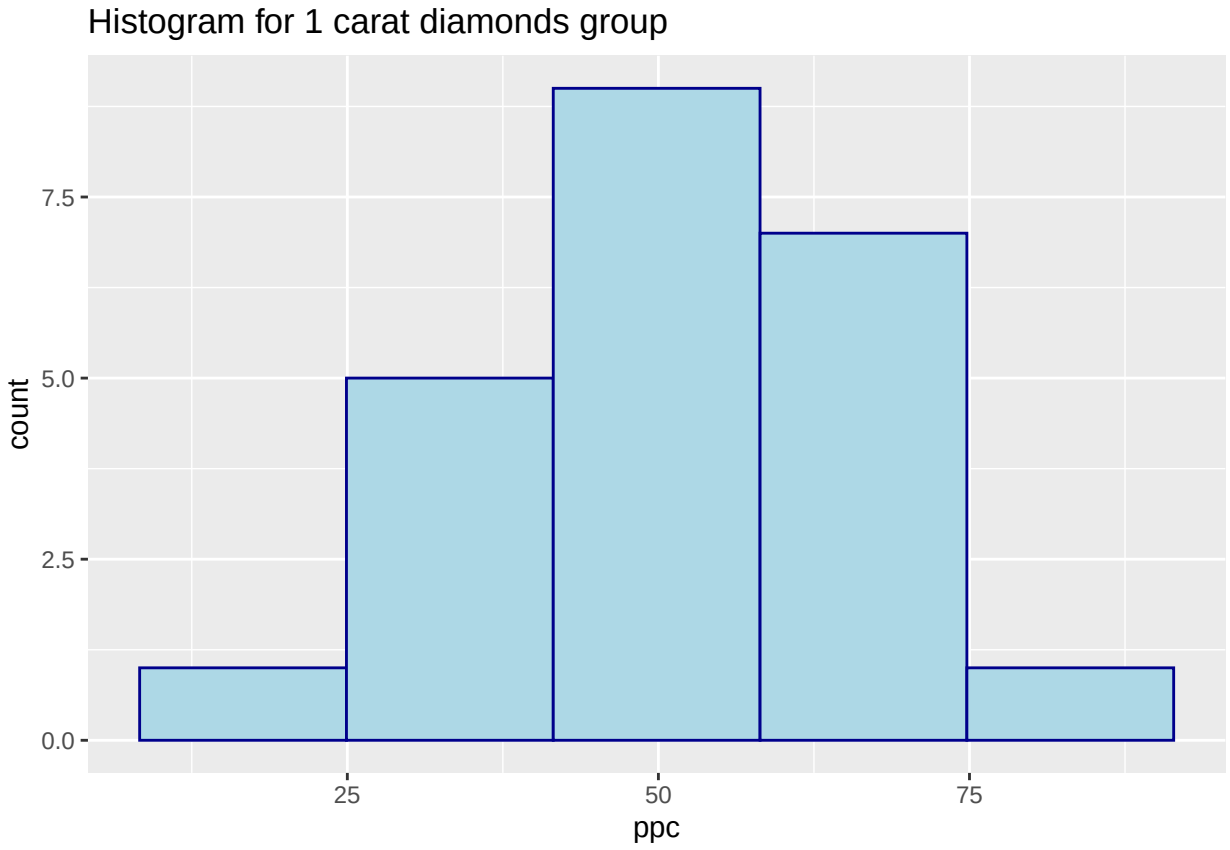
```
diamonds <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/diamonds.csv")

diamonds <- diamonds[c(386, 424, 1276, 3171, 4308, 6025, 6168, 6217, 6226, 6460, 6558,6955, 7006, 7816,
16294, 16550, 16899, 20147, 46680, 47803, 48260),]

diamonds$ppc <- as.numeric(diamonds$price/(100*diamonds$carat))
```

(a) Create two histograms of the price per carat, one for the group of 0.99 carat diamonds and one for thegroup of 1 carat diamonds. The histograms should have five bins (4 pts).

```
library(tidyverse)
par(mfrow= c(1,2))
data1 <- diamonds %>% filter(carat== 0.99)
data2 <- diamonds %>% filter(carat==1)
ggplot(data = data1, aes(x=ppc))+geom_histogram(bins = 5, color="darkblue", fill= "lightblue")+
  labs(title = "Histogram for 0.99 carat diamonds group")
```




```
ggplot(data = data2, aes(x=ppc))+geom_histogram(bins = 5, color="darkblue", fill= "lightblue")+
  labs(title = "Histogram for 1 carat diamonds group")
```



(b) Comment on the two histograms (2 pts).
The histograms for both the 1 carat diamonds and 0.99 carat diamonds groups are nearly symmetric. We can conclude that the normality assumption is met in both cases.

(c) Conduct an appropriate hypothesis test evaluating whether there is a difference in price per carat for 0.99 carat diamonds and 1 carat diamonds. Use your results from parts (a) and (b) to check the conditions (4 pts)
Solution

prepare
The hypothesis to be tested is: H_0 : there no a difference in price per carat for 0.99 carat diamonds and 1 carat diamonds VS H_1 : there is a difference in price per carat for 0.99 carat diamonds and 1 carat diamonds

```
mean(data1$ppc)
```

```
## [1] 44.50681
```

```
sd(data1$ppc)
```

```
## [1] 13.32311
```

```
mean(data2$ppc)
```

```
## [1] 51.41261
```

```
sd(data2$ppc)
```

```
## [1] 13.67794
```

check

From the histograms obtained, we can conclude that the data is normally distributed

```
#calculate

t.test(data= diamonds, log(ppc)~ carat, conf.level =0.99)
```

```
##
##  Welch Two Sample t-test
##
## data:  log(ppc) by carat
## t = -1.6707, df = 42.427, p-value = 0.1021
## alternative hypothesis: true difference in means between group 0.99 and group 1 is not equal to 0
## 99 percent confidence interval:
##  -0.41876992  0.09837956
## sample estimates:
## mean in group 0.99    mean in group 1
##           3.742282           3.902478
```

Conclusion

The p- value is 0.1021, which is greater than 1%. We fail to reject the null hypothesis. There is not enough evidence to say that there is a difference in price per carat for 0.99 carat diamonds and 1 carat diamonds.