ST.310, Spring 2024

Homework #10

Due on Wednesday, 4/24/24

Genevieve Torkornoo

***Directions:***

* Follow the ***homework format guidelines*** shown on the syllabus.
* Do ***not*** fill in your source code and answers on this problem set. Use the homework template.
* You may work with one partner if you wish.
* Upload a single Word file (saved as ***doc*** or ***docx***) on Moodle by ***11:59 PM*** on the due date.
* Late homework will ***not*** be accepted without any legitimate excuses.

Load the UsingR package and the MASS package. The packages include homework datasets. Show your R command lines and outputs for each question.

**Problem#7.2.** [Page 252]

For the rivers data set, take 1000 random samples of size 10. Compare the mean of the sample means, with the sample mean of the data in rivers. [10 pts]

**Note.** For a random sample of size 10, use the R code, sample(rivers, size=10, replace=TRUE).

data(rivers)

sample\_means <- replicate(1000, mean(sample(rivers, size = 10, replace = TRUE)))

samples\_means <- mean(sample\_means)

sampleMeanRivers <- mean(rivers)

print(paste("Mean of Sample Means:", samples\_means))

print(paste("Sample Mean of Rivers Data:", sampleMeanRivers))

[1] "Mean of Sample Means: 581.4157"

"Sample Mean of Rivers Data: 591.184397163121"

**Problem#8.15.&New** [Page 276]

The homedata data set contains assessed values of homes in Maplewood, NJ for the years 1970 and 2000.

1. Find 95% a confidence interval for the population mean difference of y1970 and y2000, assuming equal vari- ances. [5 pts]

t\_test<- t.test (homedata$y1970, homedata$y2000, paired = FALSE, var.equal = TRUE, conf.level = 0.95)

confidence\_interval <- t\_test$conf.int

print("95% Confidence Interval for Mean Difference:")

print(confidence\_interval)

[1] "95% Confidence Interval for Mean Difference:

[1] -200691.8 -194405.9  
attr(,"conf.level")  
[1] 0.95

1. Find 95% a confidence interval for the population mean difference of y1970 and y2000, assuming unequal variances. [5 pts]

**Note.** Use the t.test function for confidence intervals.

t\_result <- t.test(homedata$y1970, homedata$y2000, paired = FALSE, var.equal = FALSE, conf.level = 0.95)

confidence\_interval <- t\_result$conf.int

print("95% Confidence Interval for Mean Difference Unequal Variances:")

print(confidence\_interval)

[1] "95% Confidence Interval for Mean Difference Unequal Variances:"

[1] -200692.0 -194405.7  
attr(,"conf.level")  
[1] 0.95

1. Which variance assumption is more appropriate for the data? Explain [10 pts]

The confidence interval with the assumed unequal has a slightly bigger range than the assumed equal confidence interval. Therefore, it is more appropriate for the given data.

**Problem#8.16.&New** [Page 276]

The variable weight in the kid.weights data set contains the weights of a random sample of children. Consider a new variable yr5, defined as the weight of 5-year-olds:

yr5 <- subset(kid.weights, subset= 5\*12 <= age & age < 6\*12)

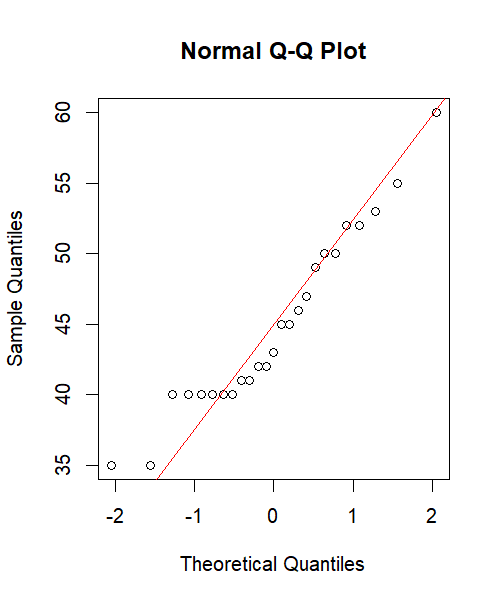
1. Display the normal probability plot for the weight variable of yr5 and check its normality. [10 pts]

**Note.** Use the qqnorm and qqline functions. Perform a shapiro.test for the variable.

yr5 <- subset(kid.weights, subset = 5\*12 <= age & age < 6\*12)

qqnorm(yr5$weight)

qqline(yr5$weight, col = "red")



shapiro\_test <- shapiro.test(yr5$weight)

print(shapiro\_test)

Shapiro-Wilk normality test  
  
data: yr5$weight  
W = 0.94157, p-value = 0.1609

1. Find a 90% confidence interval for the population mean of the weight variable of yr5. [5 pts]

tTest<- t.test (yr5$weight , conf.level = 0.90)

confidenceInterval <- tTest$conf.int

print("90% Confidence Interval for Mean Difference:")

print(confidenceInterval)

[1] "90% Confidence Interval for Mean Difference:"

[1] 42.75381 47.08619  
attr(,"conf.level")  
[1] 0.9

**Problem#9.11.&New** [Page 310]

In the babies data set, the variable dht records the father’s height for the sampled cases.

1. Remove the values of 99 from dht, as these indicate missing data. [5 pts]

data(babies)

babiesNew<- subset(babies, dht != 99)

summary (babiesNew$dht)

Min. 1st Qu. Median Mean   
 60.0 68.0 71.0 70.2   
3rd Qu. Max.   
 72.0 78.0

1. Do a significance test of the null hypothesis that the population mean height is 68 inches against an alternative that it is taller. Use *α* = 0*.*05. Interpret the R output. [10 pts]

**Note.** Use the t.test function for the one-sample *t*-test.

hypothesis <- t.test(babiesNew$dht, mu = 68, alternative = "greater")

print(hypothesis)

One Sample t-test  
  
data: babiesNew$dht  
t = 20.796, df = 743, p-value <  
2.2e-16  
alternative hypothesis: true mean is greater than 68  
95 percent confidence interval:  
 70.02973 Inf  
sample estimates:  
mean of x   
 70.2043

The p value is 2.2e-16 which is very small and since it is less than 0.05 we have to reject the null hypothesis meaning their is a relationship between fathers height

**Problem#9.34.&New** [Page 332]

Water-quality researchers wish to measure biomass/chlorophyll ratio for phytoplankton (in milligrams per liter of water). There are two possible tests, one less expensive than the other. To see whether the two tests give the same results, ten water samples were taken and each was measured both ways, providing the data below.

Method measurement

method 1 45.9 57.6 54.9 38.7 35.7 39.2 45.9 43.2 45.4 54.8

method 2 48.2 64.2 56.8 47.2 43.7 45.7 53.0 52.0 45.1 57.5

1. Display the normal probability plots for the two variables, method1 and method2, and check their normality. [10 pts]

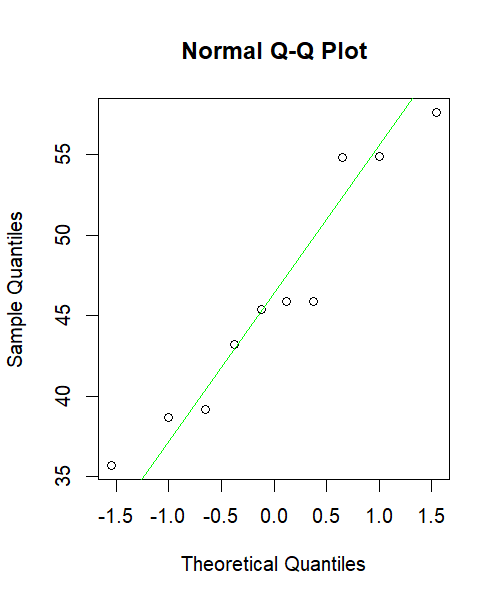
**Note.** Use the qqnorm and qqline functions. Perform a shapiro.test for each variable.

method1 <- c(45.9, 57.6, 54.9, 38.7, 35.7, 39.2, 45.9, 43.2, 45.4, 54.8)

method2 <- c(48.2, 64.2, 56.8, 47.2, 43.7, 45.7, 53.0, 52.0, 45.1, 57.5)

qqnorm(method1)

qqline(method1, col = "green")



shapiro\_test\_method1 <- shapiro.test(method1)

print(shapiro\_test\_method1)

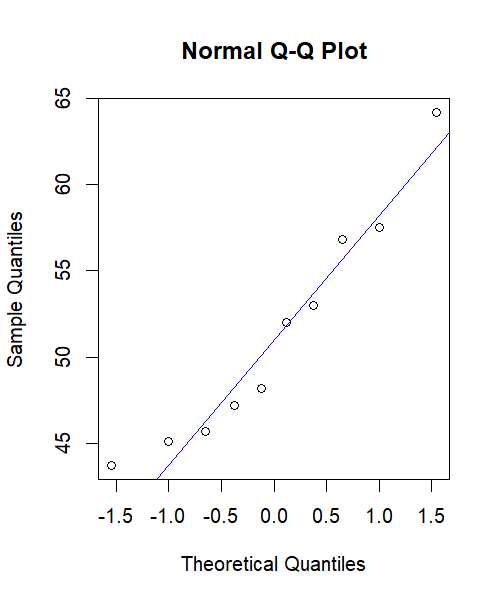
Shapiro-Wilk normality test  
  
data: method1  
W = 0.92317, p-value = 0.3841

qqnorm(method2)

qqline(method2, col = "blue")

shapiro\_test\_method2 <- shapiro.test(method2)

print(shapiro\_test\_method2)



Shapiro-Wilk normality test  
  
data: method2  
W = 0.92805, p-value = 0.4289

1. Do a *t*-test to see if there is a difference in the means of the measured amounts. Use *α* = 0*.*05. Interpret the R output. [10 pts]

**Note.** Use the t.test function for the two-sample *t*-test.

twoSample <- t.test(method1, method2, paired = FALSE, alternative = "two.sided")

print(twoSample)

Welch Two Sample t-test  
  
data: method1 and method2  
t = -1.65, df = 17.722, p-value  
= 0.1166  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 -11.851292 1.431292  
sample estimates:  
mean of x mean of y   
 46.13 51.34

The P value is 0.1166, greater than 0.05, meaning we fail to reject the null hypothesis. This means that there is insufficient evidence to prove that there is a difference in the means of measure amount.

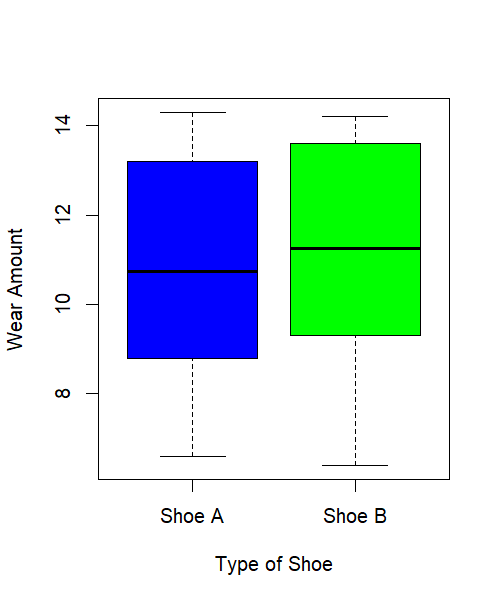
1. If you assume equal variances or unequal variances, explain why. [5 pts]

I assume it was an equal variance because I assumed that the line would line up with the plots.

**Problem#9.35.** [Page 332]

The shoes (from the MASS package) data set contains a famous data set on shoe wear. Ten boys wore two different shoes each, then measurements were taken on shoe wear. The wear amounts are stored in variables A and B. First draw comparative boxplots of the data, then compare the mean wear for the two types of shoes using the paired-sample *t*-test. Use *α* = 0*.*05. Interpret the R output. [15 pts]

boxplot(shoes$A, shoes$B, names = c("Shoe A", "Shoe B"), xlab = "Type of Shoe", ylab = "Wear Amount", col = c("blue", "green"))



t\_result <- t.test(shoes$A, shoes$B, paired = TRUE, alternative = "two.sided")

print(t\_result)

Paired t-test  
  
data: shoes$A and shoes$B  
t = -3.3489, df = 9, p-value =  
0.008539  
alternative hypothesis: true mean difference is not equal to 0  
95 percent confidence interval:  
 -0.6869539 -0.1330461  
sample estimates:  
mean difference   
 -0.41