Laboratory Exercise Week 9

Ryan Estes | Section 006, 10 pts

10/18/2017

*Directions*:

* Write your R code inside the code chunks after each question.
* Write your answer comments after the # sign.
* To generate the word document output, click the button Knit and wait for the word document to appear.
* RStudio will prompt you (only once) to install the knitr package.
* Submit your completed laboratory exercise using Blackboard's Turnitin feature. Your Turnitin upload link is found on your Blackboard Course shell under the Laboratory folder.

For this exercise, you will need to use the package mosaic to find numerical and graphical summaries.

library(mosaic) # load the package mosaic to use its functions

## Loading required package: dplyr

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

## Loading required package: lattice

## Loading required package: ggformula

## Loading required package: ggplot2

##   
## New to ggformula? Try the tutorials:   
## learnr::run\_tutorial("introduction", package = "ggformula")  
## learnr::run\_tutorial("refining", package = "ggformula")

## Loading required package: mosaicData

## Loading required package: Matrix

##   
## The 'mosaic' package masks several functions from core packages in order to add   
## additional features. The original behavior of these functions should not be affected by this.  
##   
## Note: If you use the Matrix package, be sure to load it BEFORE loading mosaic.

##   
## Attaching package: 'mosaic'

## The following object is masked from 'package:Matrix':  
##   
## mean

## The following objects are masked from 'package:dplyr':  
##   
## count, do, tally

## The following objects are masked from 'package:stats':  
##   
## binom.test, cor, cor.test, cov, fivenum, IQR, median,  
## prop.test, quantile, sd, t.test, var

## The following objects are masked from 'package:base':  
##   
## max, mean, min, prod, range, sample, sum

1. Medical research has shown that repeated interval for extensions beyond 20 degrees increases the risk of wrist and hand injuries. Each of 24 students at Cornell University used a proposed new computer mouse design, and while using the mouse, each student's wrist extension was recorded.

wrist <- data.frame(ID = 1:24,  
 extension = c(27, 28, 24, 26, 27, 25, 25, 24, 24, 24, 25, 28, 22, 25, 24, 28, 27, 26, 31, 25, 28, 27, 27, 25))

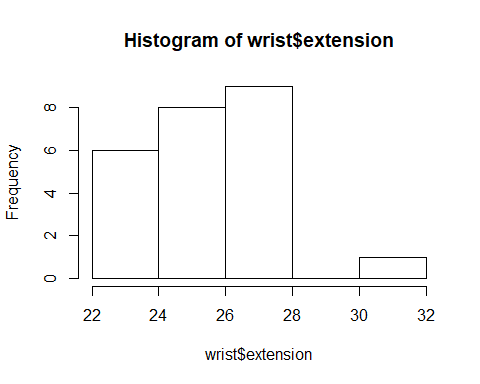
1. Compute the mean and standard deviation of the wrist extensions data above. Describe the sample using these summaries.
2. Create a histogram and QQ-plot of the wrist extensions. Is the normality assumption reasonable?
3. Use the data to estimate the mean wrist extension for people using this new mouse design using a 97% confidence interval. Provide an interpretation of this interval.
4. Use the data to test the hypothesis that the mean wrist extension for people using this new design is greater than 20 degrees. Use alpha = 0.01 (1%) level of significance.

### Code chunk

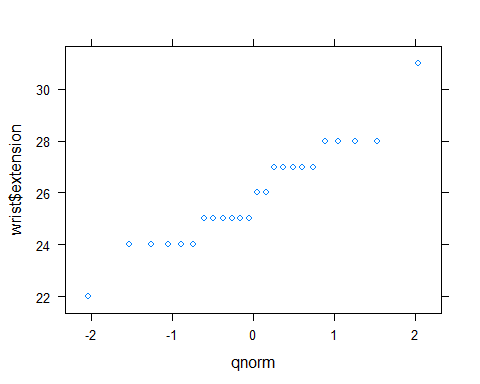
# start your code  
# i)  
favstats(wrist$extension)

## min Q1 median Q3 max mean sd n missing  
## 22 24.75 25.5 27 31 25.91667 1.954185 24 0

#The sample has a mean value of ~25.92 and a sd of ~1.95  
  
# ii)  
hist(wrist$extension)



qqmath(wrist$extension)



#According to the mean, sd, and visual representations, it is safe to assume the normality. The meat of the data seems to be near the median as well with the only real outlier as the 31.  
  
# iii)  
t.test(wrist$extension, mu = mean(wrist$extension), conf.level = 0.97)

##   
## One Sample t-test  
##   
## data: wrist$extension  
## t = 0, df = 23, p-value = 1  
## alternative hypothesis: true mean is not equal to 25.91667  
## 97 percent confidence interval:  
## 24.99393 26.83941  
## sample estimates:  
## mean of x   
## 25.91667

#With a 97% confidence interval, the range of values are from a minimum of ~24.99(basically 25) and a max of ~26.84  
  
# iv)  
# 0.01 alpha = 0.99 conf.level  
t.test(wrist$extension, mu = mean(wrist$extension), conf.level = 0.99)

##   
## One Sample t-test  
##   
## data: wrist$extension  
## t = 0, df = 23, p-value = 1  
## alternative hypothesis: true mean is not equal to 25.91667  
## 99 percent confidence interval:  
## 24.79683 27.03650  
## sample estimates:  
## mean of x   
## 25.91667

# last R code line

1. Recall the Going Wireless data first mentioned Week 2 of this class. The article Going Wireless (AARP Bulletin, June 2009) reported the estimated percentage of house- holds with only wireless phone service (no land line) for the 50 U.S. states and the District of Columbia. In the accompanying data table, each state was also classified into one of three geographical regionsâWest (W), Middle states (M), and East (E). Consider only the variable Wireless in this data.

wireless.data <- read.csv("https://goo.gl/72BKSf", header = TRUE)  
str(wireless.data)

## 'data.frame': 51 obs. of 3 variables:  
## $ Wireless: num 13.9 11.7 18.9 22.6 9 16.7 5.6 5.7 20 16.8 ...  
## $ Region : Factor w/ 3 levels "E","M","W": 2 3 3 2 3 3 1 1 1 1 ...  
## $ State : Factor w/ 51 levels "AK","AL","AR",..: 2 1 4 3 5 7 6 9 8 10 ...

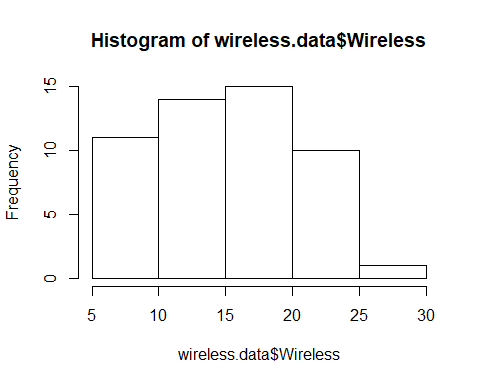
1. Compute the mean and standard deviation of the wireless data above. Describe the sample using these summaries.
2. Create a histogram and QQ-plot of the wireless data. Is the normality assumption reasonable?
3. Use the data to estimate the mean wireless percentage per state using a 90% confidence interval. Provide an interpretation of this interval.
4. Use the data to test the hypothesis that the mean wireless percentage per state is less than 17. Use alpha = 0.05 (5%) level of significance.

### Code chunk

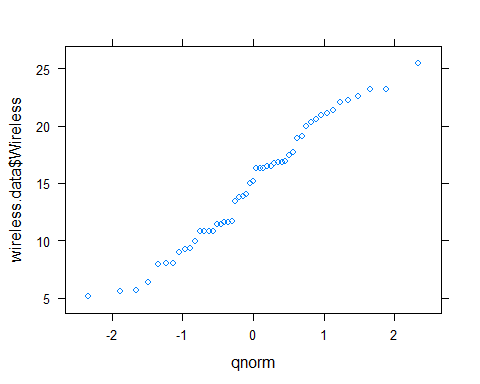
# start your code  
# i)  
favstats(wireless.data$Wireless)

## min Q1 median Q3 max mean sd n missing  
## 5.1 10.8 15.2 19 25.5 14.81569 5.344469 51 0

#The mean percentage of households in the USA with only cellular phones are ~14.82 with a sd of 5.34. This means that the amount varies quite a bit from state to state.  
  
# ii)  
hist(wireless.data$Wireless)



qqmath(wireless.data$Wireless)



#The normality assumption is reasonable.  
  
# iii)  
t.test(wireless.data$Wireless, mu = mean(wireless.data$Wireless), conf.level = 0.90)

##   
## One Sample t-test  
##   
## data: wireless.data$Wireless  
## t = 0, df = 50, p-value = 1  
## alternative hypothesis: true mean is not equal to 14.81569  
## 90 percent confidence interval:  
## 13.56148 16.06989  
## sample estimates:  
## mean of x   
## 14.81569

# The interval is between ~13.56 and ~16.07. The mean is near the Lower Limit meaning most of the data is skewed left.  
  
# iv)  
t.test(wireless.data$Wireless, mu = mean(wireless.data$Wireless), conf.level = 0.95)

##   
## One Sample t-test  
##   
## data: wireless.data$Wireless  
## t = 0, df = 50, p-value = 1  
## alternative hypothesis: true mean is not equal to 14.81569  
## 95 percent confidence interval:  
## 13.31253 16.31884  
## sample estimates:  
## mean of x   
## 14.81569

# last R code line