Laboratory Exercise Week 10

Ryan Estes | Section 006, 10 pts

10/25/2017

*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. The experimental data below contains food intake (in Kcal) for 15 men on the day following two nights of only 4 hours of sleep each night and for 15 mean on the day following two nights of 8 hours of sleep each each night. The mean participating in this experiment were randomly assigned to one of the two sleep conditions.

four.hr.grp <- c(3585, 4470, 3068, 5338, 2221, 4791, 4435, 3099, 3187, 3901, 3868, 3869, 4878, 3632, 4518)   
eight.hr.grp <- c(4965, 3918, 1987, 4993, 5220, 3653, 3510, 3338, 4100, 5792, 4547, 3319, 4304, 4057)

1. Compute the mean and standard deviation of the two groups. Comment on what you see.
2. Create a boxplot and histogram (with a fitted normal density curve) for the food intake in the two groups. Is the normal distribution a reasonable assumption for the sodium intake in both classes?
3. Carry out a two-sample t test with alpha = 0.05 to determine if there is a significant difference in mean food intake for the two different sleep conditions.
4. State the results of your t-test. Can you make conclusive statement based on the information in your data? Why?

### Code chunk

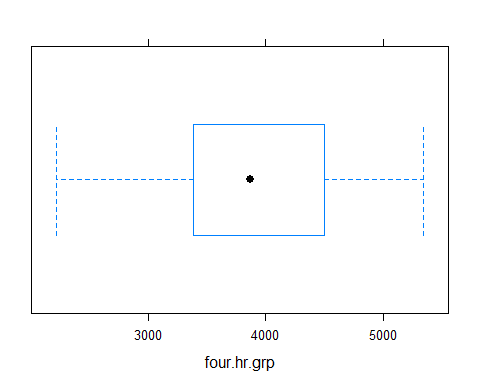
# start your code  
# i)  
favstats(four.hr.grp)

## min Q1 median Q3 max mean sd n missing  
## 2221 3386 3869 4494 5338 3924 829.6681 15 0

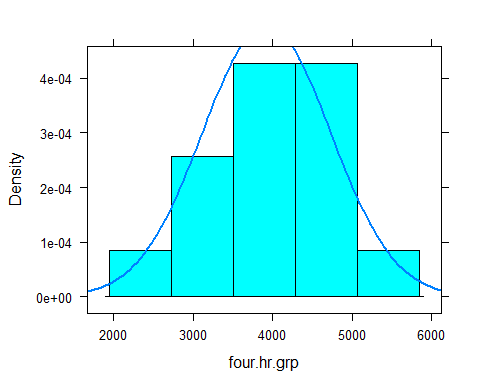
# mean = 3924  
# sd = 829.67  
favstats(eight.hr.grp)

## min Q1 median Q3 max mean sd n missing  
## 1987 3545.75 4078.5 4860.5 5792 4121.643 966.2004 14 0

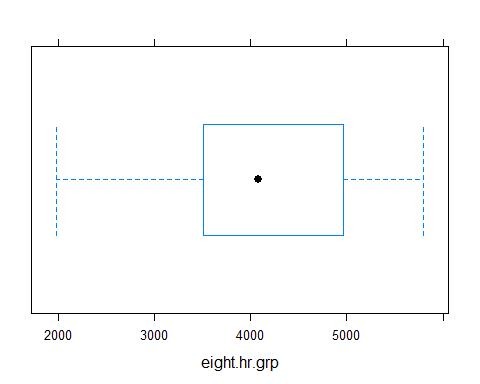
# mean = 4121.64  
# sd = 966.20  
  
# Those who sleep more tend to have a higher sodium intake the next day.  
  
# ii)  
bwplot(four.hr.grp, fit = "normal")



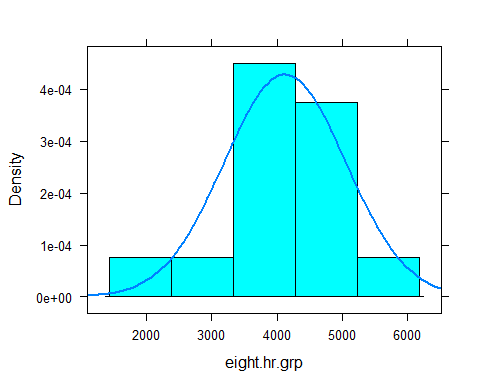
histogram(four.hr.grp, fit = "normal")



bwplot(eight.hr.grp, fit = "normal")



histogram(eight.hr.grp, fit = "normal")



# Both groups are safe assumptions, however it is worth noting that the eight hour group does have a slight left-skewness.  
  
# iii)  
t.test(four.hr.grp, eight.hr.grp,  
 var.equal = FALSE,  
 conf.level = 0.95)

##   
## Welch Two Sample t-test  
##   
## data: four.hr.grp and eight.hr.grp  
## t = -0.58907, df = 25.733, p-value = 0.561  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -887.6561 492.3704  
## sample estimates:  
## mean of x mean of y   
## 3924.000 4121.643

# iv)  
# p-value is significantly larger than the alpha value and the null hypothesis is not rejected. Thus, it is pretty safe to conclude based on this data.   
  
# last R code line

1. Instructors in two nutrition education programs have their SNAP-Ed students keep diaries of what they eat for a week, and then calculate the daily sodium intake in milligrams. Since the classes have received different nutrition education programs, they want to see if the mean sodium intake is the same for both classes.

The data are displayed in the table below:

sodium <- read.table(header = T, text ="  
Instructor Student Sodium  
'Brendon Small' a 1200  
'Brendon Small' b 1400  
'Brendon Small' c 1350  
'Brendon Small' d 950  
'Brendon Small' e 1400  
'Brendon Small' f 1150  
'Brendon Small' g 1300  
'Brendon Small' h 1325  
'Brendon Small' i 1425  
'Brendon Small' j 1500  
'Brendon Small' k 1250  
'Brendon Small' l 1150  
'Brendon Small' m 950  
'Brendon Small' n 1150  
'Brendon Small' o 1600  
'Brendon Small' p 1300  
'Brendon Small' q 1050  
'Brendon Small' r 1300  
'Brendon Small' s 1700  
'Brendon Small' t 1300  
'Coach McGuirk' u 1100  
'Coach McGuirk' v 1200  
'Coach McGuirk' w 1250  
'Coach McGuirk' x 1050  
'Coach McGuirk' y 1200  
'Coach McGuirk' z 1250  
'Coach McGuirk' aa 1350  
'Coach McGuirk' ab 1350  
'Coach McGuirk' ac 1325  
'Coach McGuirk' ad 1525  
'Coach McGuirk' ae 1225  
'Coach McGuirk' af 1125  
'Coach McGuirk' ag 1000  
'Coach McGuirk' ah 1125  
'Coach McGuirk' ai 1400  
'Coach McGuirk' aj 1200  
'Coach McGuirk' ak 1150  
'Coach McGuirk' al 1400  
'Coach McGuirk' am 1500  
'Coach McGuirk' an 1200  
")

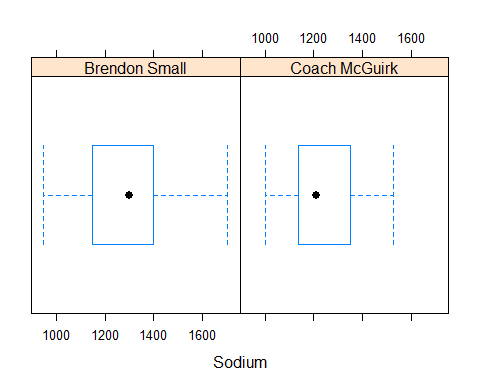
1. Compare the sodium intake between the two classes. Use the favstats function. Comment on what you see.
2. Create a boxplot and histogram (with a fitted normal density curve) for the sodium intake in the two classes. Is the normal distribution a reasonable assumption for the sodium intake in both classes?
3. Carry a two-sample t-test with alpha = 0.05 to determine if there is a significant difference in mean sodium intake for the two different classes.
4. State the results of your t-test. Can you make conclusive statement based on the information in your data? Why?

### Code chunk

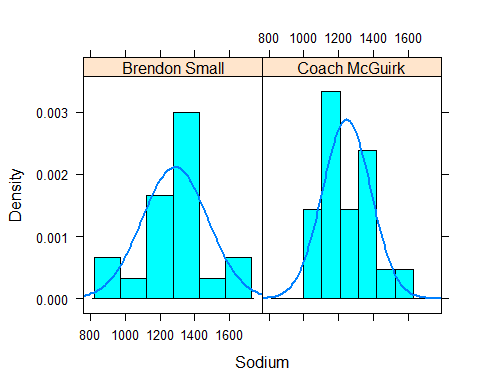
# start your code  
# i)  
favstats(~Sodium, data = sodium, group = sodium$Instructor)

## sodium$Instructor min Q1 median Q3 max mean sd n  
## 1 Brendon Small 950 1150.00 1300.0 1400 1700 1287.50 193.7341 20  
## 2 Coach McGuirk 1000 1143.75 1212.5 1350 1525 1246.25 142.4123 20  
## missing  
## 1 0  
## 2 0

# The class instructed by Brendon Small has a higher mean, lower min, higher max, higher sd, and higher median, which means that this class fluctuates heavily more than Coach McGuirk's class.  
  
#ii)  
bwplot(~Sodium | Instructor, data = sodium, fit = "normal")



histogram(~Sodium | Instructor, data = sodium, fit = "normal")



#From the data, it safe to assume the sdoium intake for Brendon Small's class. However, it is unsafe to assume for Coach McGuirk's class.  
  
# iii)  
Brendon <- sodium$Sodium[1:20]  
Coach <- sodium$Sodium[21:40]  
  
t.test(Brendon, Coach,  
 var.equal = FALSE,  
 conf.level = 0.95)

##   
## Welch Two Sample t-test  
##   
## data: Brendon and Coach  
## t = 0.76722, df = 34.893, p-value = 0.4481  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -67.91132 150.41132  
## sample estimates:  
## mean of x mean of y   
## 1287.50 1246.25

# iv)  
# p-value is significantly larger than the alpha value and the null hypothesis is not rejected. Thus, it is pretty safe to conclude based on this data.   
  
# last R code line