Laboratory Exercise Week 11

Brian Tipton - STAT-380-001

11/1/23

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 packages `mosaic` and `dplyr` to find numerical and graphical summaries.

```
# install packages if necessary
if (!require(mosaic)) install.packages(`mosaic`)
if (!require(dplyr)) install.packages(`dplyr`)
# load the package in R
library(mosaic) # load the package mosaic to use its functions
library(dplyr) # load the package dplyr to use data management functions
```

1. Lactation promotes a temporary loss of bone mass to provide adequate amounts of calcium for milk reproduction. The paper "Bone Mass is Recovered from Lactation to Postweaning in Adolescent in Adolescent Mothers with Low Calcium Intakes" gave the following data on total body bone mineral content (TBBMC) (g) for a sample both during lactation (L) and in the postweaning period (P).

```
TBBMC <- read.table(header = T, text="
Subject  Lactation  Postweaning
1          1928       2126
2          2549       2885
3          2825       2895
4          1924       1942
5          1628       1750
6          2175       2184
7          2114       2164
8          2621       2626
9          1843       2006
10       2541         2627
")
TBBMC
```

```
##    Subject Lactation Postweaning
## 1        1      1928        2126
## 2        2      2549        2885
```

```
## 3         3      2825        2895
## 4         4      1924        1942
## 5         5      1628        1750
## 6         6      2175        2184
## 7         7      2114        2164
## 8         8      2621        2626
## 9         9      1843        2006
## 10       10      2541        2627
```

i) Compute the differences in the TBBMC between "during lactation" and "postweaning period". Assign this new column into the same data set.

ii) Compute summary statistics (mean and standard deviation) on this new column of differences.

iii) Compute a 95% confidence interval for the mean difference in TBBMC between "during lactation" and "postweaning period".

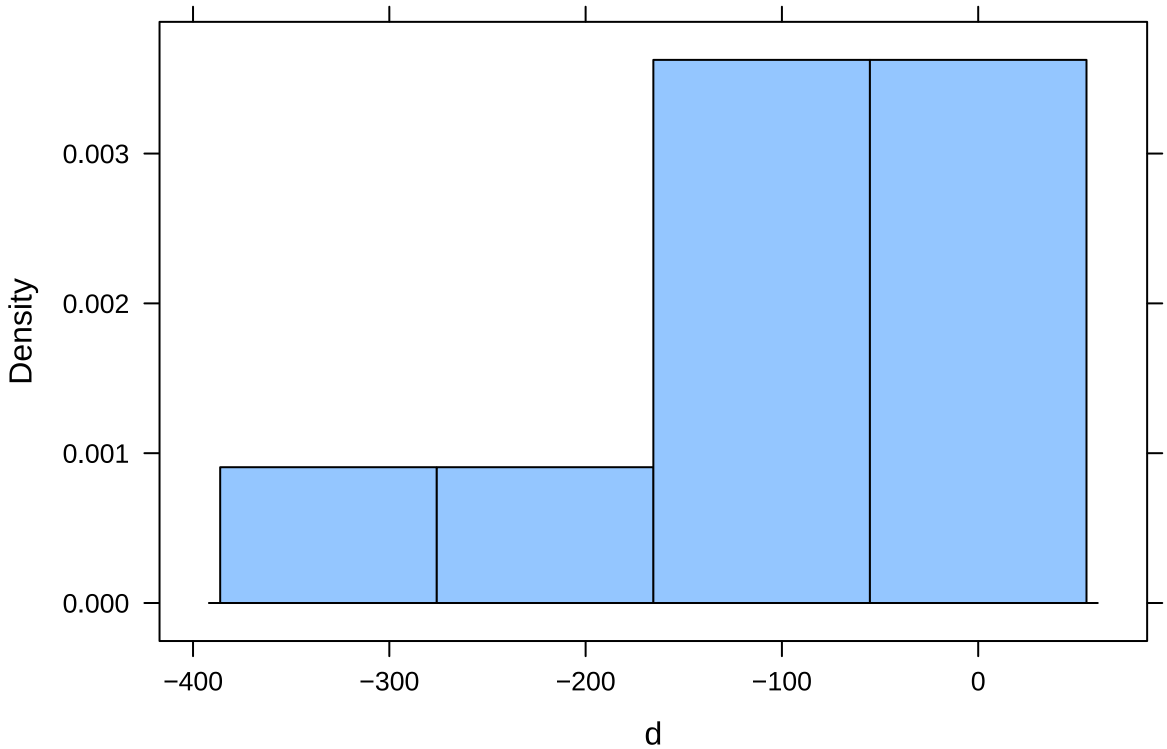
iv) Based on the computed confidence interval, does the data suggest mean TBBMC is different between "during lactation" and "postweaning period".

v) Compute the (incorrect) two-sample t-interval on the data. See Week 10 lesson on how to do this. Does the (incorrect) two-sample t-interval lead to the same conclusion that you obtained in part (iv)? Explain.

Code chunk

```
# start your code

# i) Compute the differences in the TBBMC between "during lactation" and
# "postweaning #period".
# Assign this new column into the same data set.
TBBMC <- TBBMC %>%
  mutate(d = Lactation - Postweaning)
histogram( ~ d, data = TBBMC)
```




```r
# ii) Compute summary statistics (mean and standard deviation) on this new
# column of #differences.
mean( ~ d, data = TBBMC)
```

```
## [1] -105.7
```

```r
sd( ~ d, data = TBBMC)
```

```
## [1] 103.845
```

```r
# iii) Compute a 95\% confidence interval for the mean difference in TBBMC between
# "during lactation" and "postweaning period".
t.test(~ d,
       conf.level = 0.95,
       mu = 0,
       data = TBBMC)
```

```
##
##  One Sample t-test
##
## data:  d
## t = -3.2188, df = 9, p-value = 0.01051
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -179.98625  -31.41375
## sample estimates:
## mean of x
##    -105.7
```

```r
# iv) Based on the computed confidence interval, does the data suggest mean TBBMC
# is different between "during lactation" and "postweaning period".
cat(
  "Since 0 is outside the range so therefore we reject the null hypothesis.\nSo therefore there is a si
)
```

```
## Since 0 is outside the range so therefore we reject the null hypothesis.
## So therefore there is a significant difference between the means
```

```r
# v) Compute the (incorrect)  two-sample t-interval on the data.
# See Week 10 lesson on how to do this.
# Does the (incorrect) two-sample t-interval lead to the same conclusion that you
# obtained in part (iv)? Explain.
t.test(TBBMC$Lactation,
       TBBMC$Postweaning,
       conf.level = 0.95,
       mu = 0)
```

```
##
##  Welch Two Sample t-test
##
## data:  TBBMC$Lactation and TBBMC$Postweaning
## t = -0.58872, df = 17.99, p-value = 0.5634
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -482.9168  271.5168
## sample estimates:
## mean of x mean of y
##    2214.8    2320.5
```

```r
cat(
  "No it does not lead to the same conclusion the P-value is more that alpha (0.56 > 0.05)\nso we do no
)
```

```
## No it does not lead to the same conclusion the P-value is more that alpha (0.56 > 0.05)
## so we do not have sufficient evidence to reject our null.
## Given failure to reject null hypothesis there is not a significant difference.
```

```r
# last R code line
```

2. Hexavalent chromium has been identified as an inhalation carcinogen and an air toxin of concern in a number of different locales. The article "Airborned Hexavalent Chromium in Southwestern Ontario" gave the accompanying data on both indoor and outdoor concentration (nanograms/cubic meter) for a sample of houses selected from a certain region

```r
airborne <- read.csv("https://www.siue.edu/~jpailde/airborne.csv", header = TRUE)
head(airborne) # display first 6 rows
```

```
##   House concentration Situation
## 1     1          0.07    Indoor
## 2     2          0.08    Indoor
## 3     3          0.09    Indoor
## 4     4          0.12    Indoor
## 5     5          0.12    Indoor
## 6     6          0.12    Indoor
```

```r
tail(airborne) # display last 6 rows
```

```
##    House concentration Situation
## 61    28          0.37   Outdoor
## 62    29          1.26   Outdoor
## 63    30          0.70   Outdoor
## 64    31          0.76   Outdoor
## 65    32          0.99   Outdoor
## 66    33          0.36   Outdoor
```

i) Compute the sample mean and sample standard deviation `concentration` for both `indoor` and `outdoor`.

ii) Construct boxplots for the `concentration` for both `indoor` and `outdoor`.

iii) Based on what you see in parts (i) and (ii), do you suspect the `concentration` levels for indoor and outdoor are different? Why?

iv) Is a paired sample analysis appropriate for this data? Why?

v) Calculate a confidence interval for the population mean difference between indoor and outdoor concentrations using a confidence interval of 95%, and interpret the resulting interval.

Code chunk

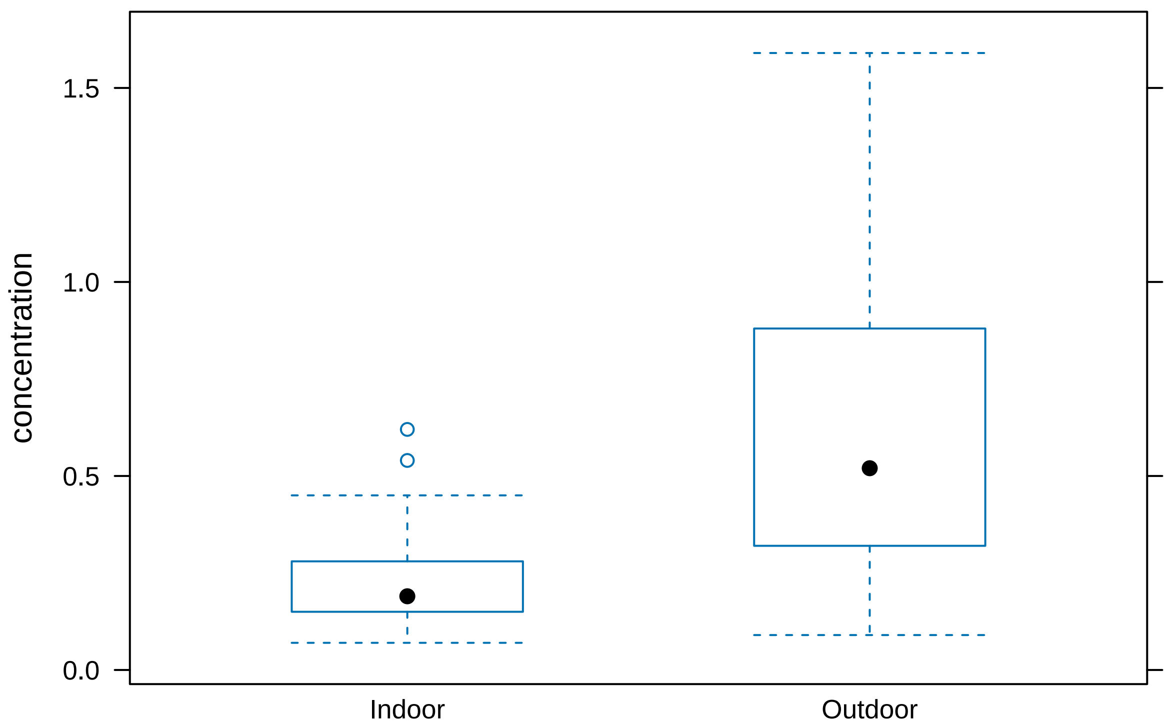
```
# start your code
# i) Compute the sample mean and sample standard deviation `concentration` for both
# `indoor` and `outdoor`.
mean(concentration ~ Situation, data = airborne)
```

```
##    Indoor   Outdoor
## 0.2309091 0.6369697
```

```
sd(concentration ~ Situation, data = airborne)
```

```
##    Indoor   Outdoor
## 0.1284368 0.3923446
```

```
# ii) Construct boxplots for the `concentration` for both `indoor` and `outdoor`.
bwplot(concentration ~ Situation, data = airborne)
```




```
# iii) Based on what you see in parts (i) and (ii), do you suspect the `concentration`
# levels for indoor and outdoor are different?
# Why?
cat("yes there is a difference in the concentration levels.\nThis is because the difference in the mean
```

```
## yes there is a difference in the concentration levels.
## This is because the difference in the mean and median
## between Indoor and Outdoor is signficant.
```

```
# iv) Is a paired sample analysis appropriate for this data?
# Why?
cat("Yes, because the data is related in that they are paired samples using indoor and\noutdoor from the
```

```
## Yes, because the data is related in that they are paired samples using indoor and
## outdoor from the same home.
```

```
# v) Calculate a confidence interval for the population mean difference between indoor and
# outdoor concentrations using a confidence interval of 95\%,
# and interpret the resulting interval.
t.test(
  concentration ~ Situation,
  data = airborne,
  paired = TRUE,
  conf.level = 0.95,
  mu = 0
)
```

```
##
```

```
##  Paired t-test
##
## data:  concentration by Situation
## t = -5.9509, df = 32, p-value = 1.251e-06
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.5450513 -0.2670700
## sample estimates:
## mean difference
##      -0.4060606
```

```r
cat("Since our p value is less than alpha we reject the null hypothesis.\nThere is a significant differe
```

```
## Since our p value is less than alpha we reject the null hypothesis.
## There is a significant difference between the means of the indoor and outdoor samples.
```

```r
# last R code line
```