Laboratory Exercise Week 11

Ryan Estes | Section 006, 10 pts

11/1/2017

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

library(mosaic)

## 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

library(dplyr)

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"](https://www.ncbi.nlm.nih.gov/pubmed/15531682) 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

1. Compute the differences in the TBBMC between "during lactation" and "postweaning period". Assign this new column into the same data set.
2. Compute summary statistics (mean and standard deviation) on this new column of differences.
3. Compute a 95% confidence interval for the mean difference in TBBMC between "during lactation" and "postweaning period".
4. Based on the computed confidence interval, does the data suggest mean TBBMC is different between "during lactation" and "postweaning period".
5. 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)  
tbbmc\_diff <- TBBMC %>%  
 mutate(d = Postweaning - Lactation)  
  
# ii)  
mean(tbbmc\_diff$d) #105.7

## [1] 105.7

sd(tbbmc\_diff$d) #103.845

## [1] 103.845

# iii)  
t.test(tbbmc\_diff$d, conf.level = 0.95)

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

# iv)  
# Only one value is outside of the confidence interval. Additionally, the p-value of 0.01051 is less than the defined alpha(0.05) which means that we reject the null hypothesis.  
  
# v)  
t.test(TBBMC$Lactation, TBBMC$Postweaning,  
 var.equal = FALSE,  
 conf.level = 0.95)

##   
## 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

# The values vary significantly. The p-value is greater than alpha, the degrees of freedom is double (double amount of values so expected) and the t value is negative now.  
  
# last R code line

1. 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"](https://goo.gl/xjTQM5) gave the accompanying data on both indoor and outdoor concentration (nanograms/cubic meter) for a sample of houses selected from a certain region

airborne <- read.csv("https://www.siue.edu/~jpailde/airborne.csv", header = TRUE)  
airborne

## 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  
## 7 7 0.13 Indoor  
## 8 8 0.14 Indoor  
## 9 9 0.15 Indoor  
## 10 10 0.15 Indoor  
## 11 11 0.17 Indoor  
## 12 12 0.17 Indoor  
## 13 13 0.18 Indoor  
## 14 14 0.18 Indoor  
## 15 15 0.18 Indoor  
## 16 16 0.18 Indoor  
## 17 17 0.19 Indoor  
## 18 18 0.20 Indoor  
## 19 19 0.22 Indoor  
## 20 20 0.22 Indoor  
## 21 21 0.23 Indoor  
## 22 22 0.23 Indoor  
## 23 23 0.25 Indoor  
## 24 24 0.26 Indoor  
## 25 25 0.28 Indoor  
## 26 26 0.28 Indoor  
## 27 27 0.29 Indoor  
## 28 28 0.34 Indoor  
## 29 29 0.39 Indoor  
## 30 30 0.40 Indoor  
## 31 31 0.45 Indoor  
## 32 32 0.54 Indoor  
## 33 33 0.62 Indoor  
## 34 1 0.29 Outdoor  
## 35 2 0.09 Outdoor  
## 36 3 0.47 Outdoor  
## 37 4 0.54 Outdoor  
## 38 5 0.97 Outdoor  
## 39 6 0.35 Outdoor  
## 40 7 0.49 Outdoor  
## 41 8 0.84 Outdoor  
## 42 9 0.86 Outdoor  
## 43 10 0.28 Outdoor  
## 44 11 0.32 Outdoor  
## 45 12 0.32 Outdoor  
## 46 13 1.55 Outdoor  
## 47 14 0.66 Outdoor  
## 48 15 0.29 Outdoor  
## 49 16 0.21 Outdoor  
## 50 17 1.02 Outdoor  
## 51 18 1.59 Outdoor  
## 52 19 0.90 Outdoor  
## 53 20 0.52 Outdoor  
## 54 21 0.12 Outdoor  
## 55 22 0.54 Outdoor  
## 56 23 0.88 Outdoor  
## 57 24 0.49 Outdoor  
## 58 25 1.24 Outdoor  
## 59 26 0.48 Outdoor  
## 60 27 0.27 Outdoor  
## 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

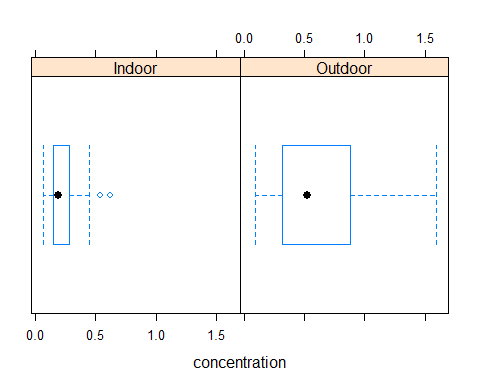
1. Compute the sample mean and sample standard deviation concentration for both indoor and outdoor.
2. Construct boxplots for the concentration for both indoor and outdoor.
3. Based on what you see in parts (i) and (ii), do you suspect the concentration levels for indoor and outdoor are different? Why?
4. Is a paired sample analysis appropriate for this data? Why?
5. 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)  
favstats(~ concentration, data = airborne, group = airborne$Situation)

## airborne$Situation min Q1 median Q3 max mean sd n  
## 1 Indoor 0.07 0.15 0.19 0.28 0.62 0.2309091 0.1284368 33  
## 2 Outdoor 0.09 0.32 0.52 0.88 1.59 0.6369697 0.3923446 33  
## missing  
## 1 0  
## 2 0

# Indoor mean = 0.2309091  
# Indoor sd = 0.1284368  
# outdoor mean = 0.6369697  
# outdoor sd = 0.3923446  
  
# ii)  
bwplot(~ concentration | Situation, data = airborne, fit = "normal")



# iii)  
# Based on these values, the values for the outdoor and indoor concentrations are very different. Indoor values are very scrunched together with low values and variability, while the outdoor values are the exact opposite.  
  
# iv)  
# Due to large variabilities in values, it isn't necessary, however the min values are similar so it wouldn't hurt to calculate a paired sample analysis.  
  
# v)  
t.test(concentration ~ Situation, data = airborne,  
 paired = TRUE,  
 conf.level = 0.95,  
 mu = 0)

## concentration ~ Situation

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

# Conf Int. = -0.5450513 -0.2670700  
# The total mean value, with 95% confidence should fall between these positive values.  
  
# last R code line