

# Assignment 5 - Inference for a Population Mean.

## Due October 12, 11:59pm 2018

### EPIB607 - Inferential Statistics<sup>a</sup>

<sup>a</sup>Fall 2018, McGill University

This version was compiled on October 6, 2018

In this assignment you will practice conducting inference for a one sample mean using either the z procedure, t procedure, or the bootstrap. Answers should be given in full sentences (DO NOT just provide the number). All figures should have appropriately labeled axes, titles and captions (if necessary). All graphs and calculations are to be completed in an R Markdown document using the provided template. You are free to choose any function from any package to complete the assignment. Concise answers will be rewarded. Be brief and to the point. Please submit both the compiled HTML report and the source file (.Rmd) to myCourses by October 12, 2018, 11:59pm. Both HTML and .Rmd files should be saved as 'IDnumber\_LastName\_FirstName\_EPIB607\_A5'.

t-test | One sample mean | Bootstrap

### Template

The .Rmd template for Assignment 4 is available [here](#)

### 1. Food intake and weight gain

If we increase our food intake, we generally gain weight. Nutrition scientists can calculate the amount of weight gain that would be associated with a given increase in calories. In one study, 16 nonobese adults, aged 25 to 36 years, were fed 1000 calories per day in excess of the calories needed to maintain a stable body weight. The subjects maintained this diet for 8 weeks, so they consumed a total of 56,000 extra calories. According to theory, 3500 extra calories will translate into a weight gain of 1 pound. Therefore we expect each of these subjects to gain  $56,000/3500=16$  pounds (lb). Here are the weights (given in the `weight_gain.csv` file) before and after the 8-week period expressed in kilograms (kg):

| subject | before | after |
|---------|--------|-------|
| 1       | 55.7   | 61.7  |
| 2       | 54.9   | 58.8  |
| 3       | 59.6   | 66.0  |
| 4       | 62.3   | 66.2  |
| 5       | 74.2   | 79.0  |
| 6       | 75.6   | 82.3  |
| 7       | 70.7   | 74.3  |
| 8       | 53.3   | 59.3  |
| 9       | 73.3   | 79.1  |
| 10      | 63.4   | 66.0  |
| 11      | 68.1   | 73.4  |
| 12      | 73.7   | 76.9  |
| 13      | 91.7   | 93.1  |
| 14      | 55.9   | 63.0  |
| 15      | 61.7   | 68.2  |
| 16      | 57.8   | 60.3  |

```
weight <- read.csv("weight_gain.csv")
```

- Calculate a 95% confidence interval for the mean weight change and give a sentence explaining the meaning of the 95%.
- Convert the units of the mean weight gain and 95% confidence interval to pounds. Note that 1 kilogram is equal to 2.2 pounds.
- Test the null hypothesis that the mean weight gain is 16 lbs. Be sure to specify the null and alternative hypotheses. What do you conclude?

### 2. Deer mice

Deer mice are small rodents native to North America. Their adult body lengths (excluding tail) are known to vary approximately Normally, with mean  $\mu = 86$  millimeters (mm) and standard deviation  $\sigma = 8$  mm. Deer mice are found in diverse habitats and exhibit different adaptations to their environment. A random sample of 14 deer mice in a rich forest habitat gives an average body length of  $\bar{y} = 91.1$  mm. Assume that the standard deviation  $\sigma$  of all deer mice in this area is also 8 mm.

- What is the standard deviation of the mean length  $\bar{y}$ ?

- What critical value do you need to use in order to compute a 95% confidence interval for the mean  $\mu$
- Give a i) 90% confidence interval and ii) 95% confidence interval for the mean body length of all deer mice in the forest habitat.
- Why does the 90% interval have a smaller margin of error?

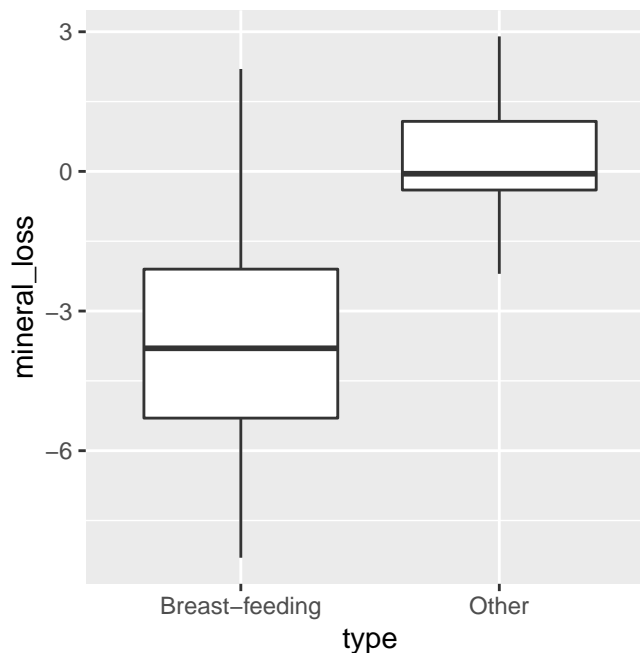
### 3. Does breast-feeding weaken bones?

Breast-feeding mothers secrete calcium into their milk. Some of the calcium may come from their bones, so mothers may lose bone mineral content. Researchers compared 47 breast-feeding women with 22 women of similar age who were neither pregnant nor lactating. They measured the percent change in the mineral content of the women's spines over three months. A negative value of `mineral_loss` indicates a loss in mineral content. The data can be read into R and a boxplot comparing the two groups is shown in the Figure below:

```
boneloss <- read.csv("https://github.com/sahirbhatnagar/EPIB607/raw/master/data/boneloss.csv")
head(boneloss)
```

```
#   type mineral_loss
# 1 Other           2.4
# 2 Other           0.0
# 3 Other           0.9
# 4 Other          -0.2
# 5 Other           1.0
# 6 Other           1.7
```

```
library(mosaic)
ggformula::gf_boxplot(mineral_loss ~ type, data = boneloss)
```



- Give a 95% confidence interval for the mean mineral loss for each group.
- Based on these confidence intervals, are you convinced that the data show distinctly greater bone mineral loss among the breast-feeding women?
- (BONUS) Let  $\bar{y}_1$  be the mean mineral loss for breast-feeding women and  $\bar{y}_2$  be the mean mineral loss for the other group. Construct a 95% bootstrap confidence interval for the difference  $\bar{y}_2 - \bar{y}_1$  based on  $B = 10000$  bootstrap samples. Interpret this confidence interval and compare it to the ones obtained in b. *Hint*: split the data in two data.frames by `type`. For each data.frame, create  $B$  resamples and calculate the means for each. Plot the histogram of the difference in means, which gives you the bootstrap distribution for the difference. From this you can use the `quantile` function to calculate the 95% CI.

### 4. How deep is the ocean?

This question is based on the [in-class Exercise](#) on sampling distributions. Refer to the [slides on Bootstrap confidence intervals](#) for R code on how to calculate bootstrap confidence intervals. For your sample of  $n = 20$  of depths of the ocean, calculate the

- 95% Confidence interval using the  $\pm$  formula
- 95% Confidence interval using the qnorm function
- 95% Confidence interval using  $B = 10000$  bootstrap samples
- Plot all three confidence intervals on the same plot and comment on the difference/similarities between the 3 intervals. You may use the compare\_CI function provided below to produce the plot. This takes as input, the sample mean (ybar), and the CIs calculated from a,b,c in the form of a numeric vector of length 2 into the arguments PM, QNORM and BOOT, respectively.

```
compare_CI <- function(ybar, PM, QNORM, BOOT,
                      col = c("#E41A1C", "#377EB8", "#4DAF4A")) {

  dt <- data.frame(type = c("plus_minus", "qnorm", "bootstrap"),
                  ybar = rep(ybar, 3),
                  low = c(PM[1], QNORM[1], BOOT[1]),
                  up = c(PM[2], QNORM[2], BOOT[2])
  )

  plot(dt$ybar, 1:nrow(dt), pch = 20, ylim = c(0, 5),
       xlim = range(pretty(c(dt$low, dt$up))),
       xlab = "Depth of ocean (m)", ylab = "Confidence Interval Type",
       las = 1, cex.axis = 0.8, cex = 3)

  abline(v = 37, lty = 2, col = "black", lwd = 2)
  segments(x0 = dt$low, x1 = dt$up,
          y0 = 1:nrow(dt), lend = 1,
          col = col, lwd = 4)

  legend("topleft",
        legend = c(eval(substitute( expression(paste(mu," = ",37))))),
                  sprintf("plus/minus CI: [%.f, %.f]",PM[1], PM[2]),
                  sprintf("qnorm CI: [%.f, %.f]",QNORM[1], QNORM[2]),
                  sprintf("bootstrap CI: [%.f, %.f]",BOOT[1], BOOT[2])),
        lty = c(1, 1,1,1),
        col = c("black",col), lwd = 4)
}

# example of how to use the function:
compare_CI(ybar = 36, PM = c(28, 40), QNORM = c(25,40), BOOT = c(31, 38))
```