In-class exercise - Inference for means and Power Calculations.

EPIB607 - Inferential Statistics[a]

[a]Fall 2019, McGill University

This version was compiled on October 9, 2019

In this exercise you will practice calculating confidence intervals using the t-distribution and the bootstrap.

Sampling distribution | Standard error | Normal distribution | Quantiles | Percentiles | Z-scores

R Code	Value
`qnorm(p = c(0.05, 0.95))`	-1.64, 1.64
`qnorm(p = c(0.025, 0.975))`	-1.96, 1.96
`qnorm(p = c(0.005, 0.995))`	-2.58, 2.58
`qt(p = c(0.025, 0.975), df = 400-1)`	-1.97, 1.97
`qt(p = c(0.025, 0.975), df = 25-1)`	-2.06, 2.06
`qt(p = c(0.025, 0.975), df = 20-1)`	-2.09, 2.09
`qt(p = c(0.025, 0.975), df = 16-1)`	-2.13, 2.13

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. Therfore we expect each of these subjects to gain 56,000/3500=16 pounds (lb). Here are the weights (given in the `weightgain.csv` file) before and after the 8-week period expressed in kilograms (kg):

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

a. Calculate a 95% confidence interval for the mean weight change and give a sentence explaining the meaning of the 95%. State your assumptions.

```
weight <- read.csv("~/git_repositories/EPIB607/exercises/inferencemeans/weightgain.csv")

# Creating new variable for weight change
weight$change <- weight$after-weight$before
weight$change_lb <- weight$change*2.2

# Calculating the mean of weight change and rounding
ybar_change <- round(mean(weight$change),2)

# Calculating the sample standard deviation
ssd_change <- sd(weight$change)
```

```r
# sample size
n <- nrow(weight)

# Calculating a 95% confidence interval version 1
qt_scaled <- function(p, df, mean, sd) {
  mean  + qt(p = p, df = df) * sd
}

(q1_ci95 <- qt_scaled(p = c(0.025, 0.975),
                      df = nrow(weight) - 1,
                      mean = ybar_change,
                      sd = ssd_change / sqrt(n)))
```

```
#  [1] 3.799758 5.660242
```

```r
# Calculating a 95% confidence interval version 2
ybar_change +  qt(p = c(0.025, 0.975), df = n - 1) * ssd_change / sqrt(n)
```

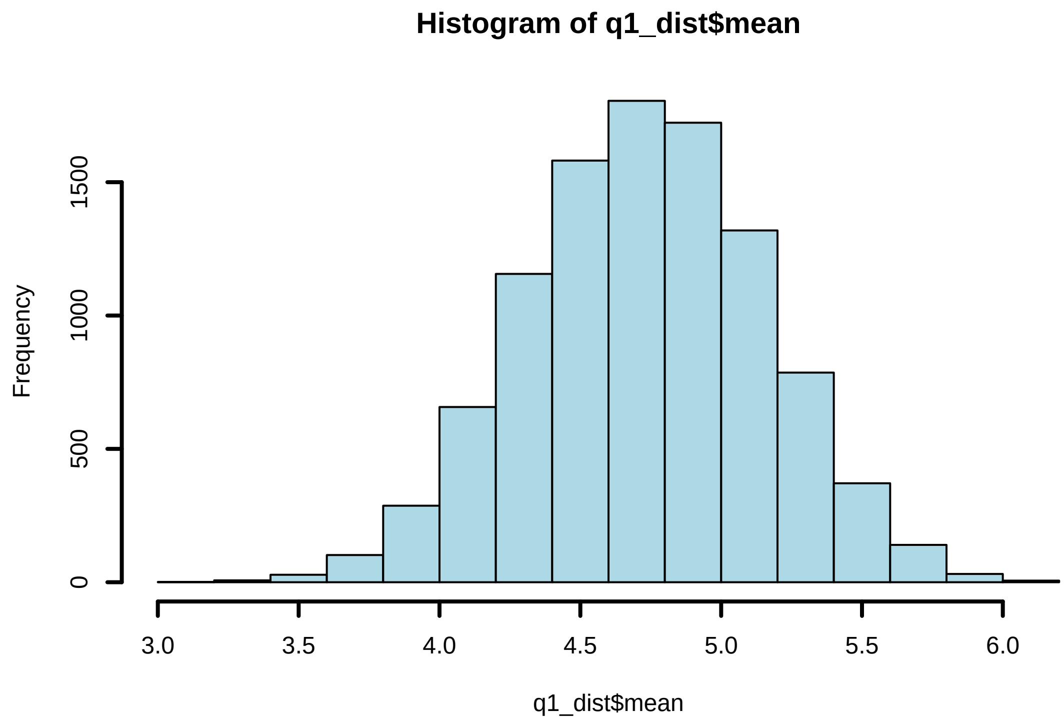
```
#  [1] 3.799758 5.660242
```

You are given that $\bar{y}_{difference} = 4.73$ and $\text{sd}(\bar{y}_{difference}) = 1.75$.

The 95% confidence interval for the mean weight change is 4.73 kg (3.7997582 kg, 5.6602418 kg). If the method used in this study were repeated many times, 95% of the time, the interval 3.7997582 kg and 5.6602418 kg will cover the true mean weight change. As this confidence interval was calculated using the *t* procedure, we are assuming that (1) we can regard our data as a simple random sample (SRS) from the population, (2) we have a representative sample of the population weight change and (3) observations of weight change in the population have a Normal distribution.

b. Calculate a 95% bootstrap confidence interval for the mean weight change and compare it to the one obtained in part (a). Comment on the bootstrap sampling distribution and compare it to the assumptions you made in part (a).

```r
q1_dist <- do(10000) * mean( ~ change, data = resample(weight))
hist(q1_dist$mean, col = "lightblue", lwd = 2)
```




```
round(quantile(~ mean, data = q1_dist, probs = c(0.001, 0.005, .025, 0.05,
                                               0.90, 0.95, 0.975, 0.99)),2)
```

```
#   0.1%  0.5%  2.5%    5%   90%   95% 97.5%   99%
#   3.42  3.67  3.91  4.04  5.28  5.43  5.55  5.69
```

c. 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. State your assumptions and justify your choice of test. Be sure to specify the null and alternative hypotheses. What do you conclude?

2. Attitudes toward school

The Survey of Study Habits and Attitudes (SSHA) is a psychological test that measures the motivation, attitude toward school, and study habits of students. Scores range from 0 to 200. The mean score for U.S. college students is about 115, and the standard deviation is about 30. A teacher who suspects that older students have better attitudes toward school gives the SSHA to 25 students who are at least 30 years of age. Their mean score is $\bar{y} = 132.2$ with a sample standard deviation $s = 28$.

a. The teacher asks you to carry out a formal statistical test for her hypothesis. Perform a test, provide a 95% confidence interval and state your conclusion clearly.
b. What assumptions did you use in part (a). Which of these assumptions is most important to the validity of your conclusion in part (a).

3. Does a full moon affect behavior?

Many people believe that the moon influences the actions of some individuals. A study of dementia patients in nursing homes recorded various types of disruptive behaviors every day for 12 weeks. Days were classified as moon days if they were in a 3-day period centered at the day of the full moon. For each patient, the average number of disruptive behaviors was computed for moon days and for all other days. The hypothesis is that moon days will lead to more disruptive behavior. We look at a data set consisting of observations on 15 dementia patients in nursing homes (available in the `fullmoon.csv` file):

```
fullmoon <- read.csv("fullmoon.csv")
```

```
#     patient moon_days other_days
#  1        1      3.33       0.27
#  2        2      3.67       0.59
#  3        3      2.67       0.32
#  4        4      3.33       0.19
#  5        5      3.33       1.26
#  6        6      3.67       0.11
#  7        7      4.67       0.30
#  8        8      2.67       0.40
#  9        9      6.00       1.59
#  10      10      4.33       0.60
#  11      11      3.33       0.65
#  12      12      0.67       0.69
#  13      13      1.33       1.26
#  14      14      0.33       0.23
#  15      15      2.00       0.38
```

a. Calculate a 95% confidence interval for the mean difference in disruptive behaviors. State the assumptions you used to calculate this interval.
b. Test the hypothesis that moon days will lead to more disruptive behavior. State your assumptions and provide a brief conclusion based on your analysis.
c. Find the minimum value of the mean difference in disruptive behaviors ($\bar{y}$) needed to reject the null hypothesis.
d. What is the probability of detecting an increase of 1.0 aggressive behavior per day during moon days?

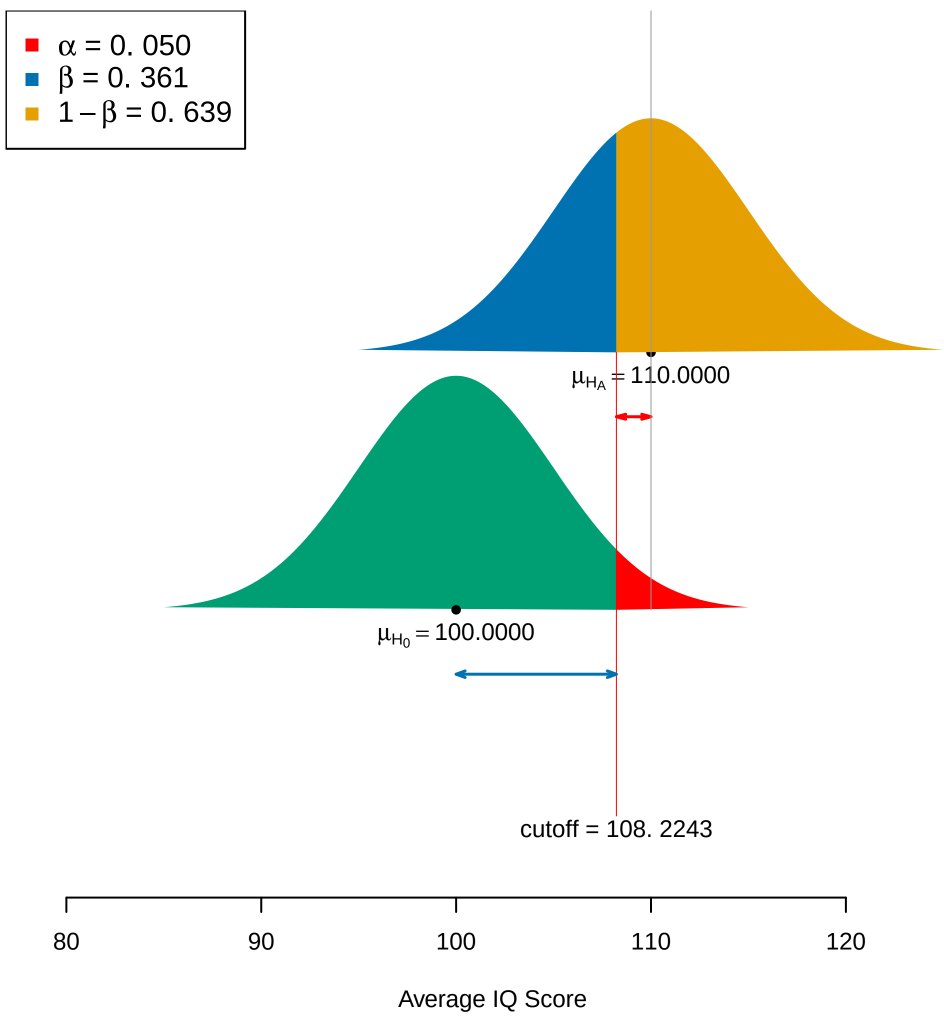
4. Lake Wobegon

It is claimed that the children of Lake Wobegon are above average. Take a simple random sample of 9 children from Lake Wobegon, and measure their IQ to obtain a sample mean of 112.8. IQ scores are scaled to be Normally distributed with mean 100 and standard deviation 15.

a) Does this sample provide evidence to reject the null hypothesis of no difference between children of Lake Wobegon and the general population?

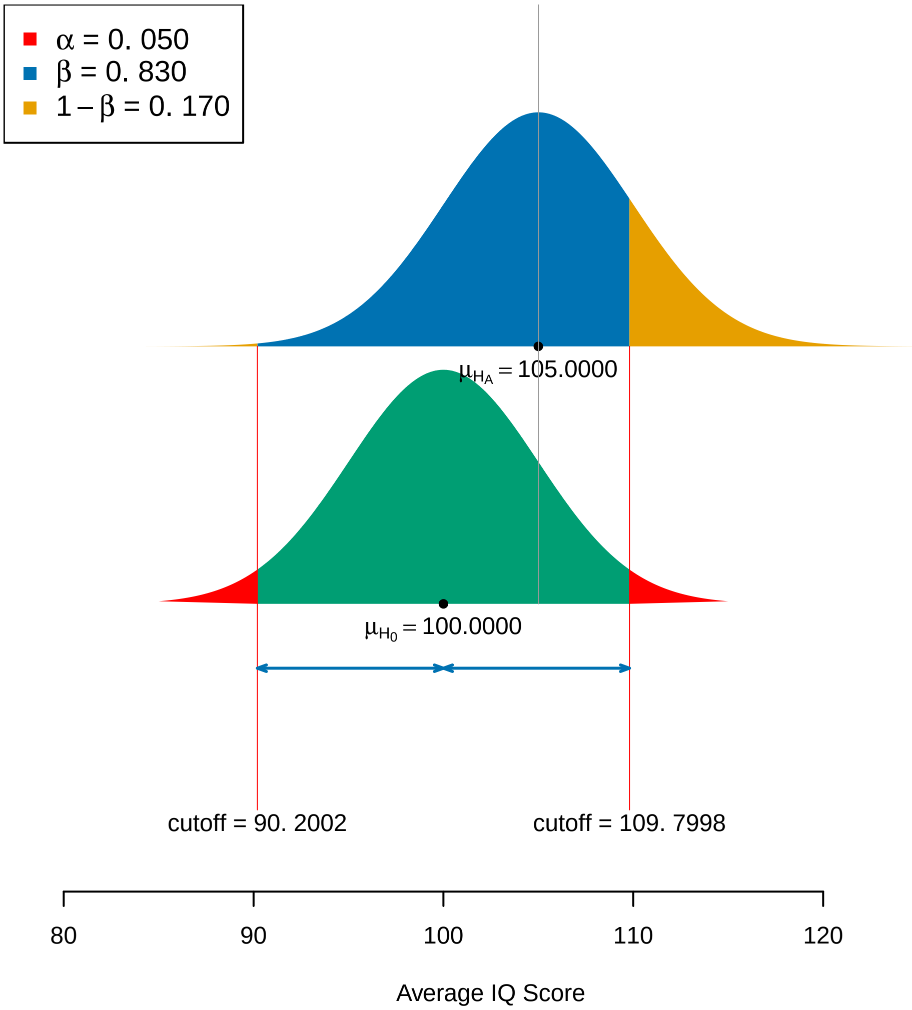
b) Suppose you hope to use a one-sided test to show that the children from Lake Wobegon are at least 10 points higher than average on the IQ test. What power do you have to detect this with the sample of 9 children if using a 0.05-level test?

```
source("https://raw.githubusercontent.com/sahirbhatnagar/EPIB607/master/code/plot_null_alt.R")
power_plot(n = 9, s = 15, mu0 = 100, mha = 110,
           cutoff = 100 + qnorm(0.95) * 15 / sqrt(9),
           alternative = "greater", xlab = "Average IQ Score")
```



c) If you hoped to use a **two-sided** test to show that the children from Lake Wobegon are at least 5 points higher than average on the IQ test, what power do you have with the sample size of 9 and a 0.05-level test?

```r
power_plot(n = 9, s = 15, mu0 = 100, mha = 105,
           cutoff = 100 + qnorm(c(0.025, 0.975)) * 15 / sqrt(9),
           alternative = "equal", xlab = "Average IQ Score")
```



5. Bias in step counters

Following the study by Case et al., JAMA, 2015, suppose we wished to assess, via a formal statistical test, whether (at an *population*, rather than an individual, level) a step-counting device or app is unbiased (H_0) or under-counts (H_1). Suppose we will do so the way Case et al. did, but measuring n persons just once each. We observe the device count when the true count on the treadmill reaches 500.

a. Using a planned sample size of $n = 25$, and $\sigma = 60$ steps as a pre-study best-guess as to the s that might be observed in them, calculate the critical value at $\alpha = 0.01$.
b. Now imagine that the mean would not be the null 500, but $\mu = 470$. Calculate the probability that the mean in the sample of 25 will be less than this critical value. Use the same s for the alternative that you used for the null. What is this probability called?
c. Determine the sample size required for 80% power using a 1% level of significance. Plot the null and alternative distributions in a diagram using the `plot_power` function.