

Lab 005 - Inference for means - Solutions

EPIB607 - Inferential Statistics^a

^aMcGill University

This version was compiled on October 14, 2021

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
<code>qnorm(p = c(0.05, 0.95))</code>	-1.64, 1.64
<code>qnorm(p = c(0.025, 0.975))</code>	-1.96, 1.96
<code>qnorm(p = c(0.005, 0.995))</code>	-2.58, 2.58
<code>qt(p = c(0.025, 0.975), df = 400-1)</code>	-1.97, 1.97
<code>qt(p = c(0.025, 0.975), df = 25-1)</code>	-2.06, 2.06
<code>qt(p = c(0.025, 0.975), df = 20-1)</code>	-2.09, 2.09
<code>qt(p = c(0.025, 0.975), df = 16-1)</code>	-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. Therefore 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")
```

- 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/inst/labs/005-one-sample-mean/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 <- mean(weight$change))
```

```
# [1] 4.7
```

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

```
# [1] 1.7
```

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

```
# [1] 16
```

```
# 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] 3.8 5.7
```

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

```
# [1] 3.8 5.7
```

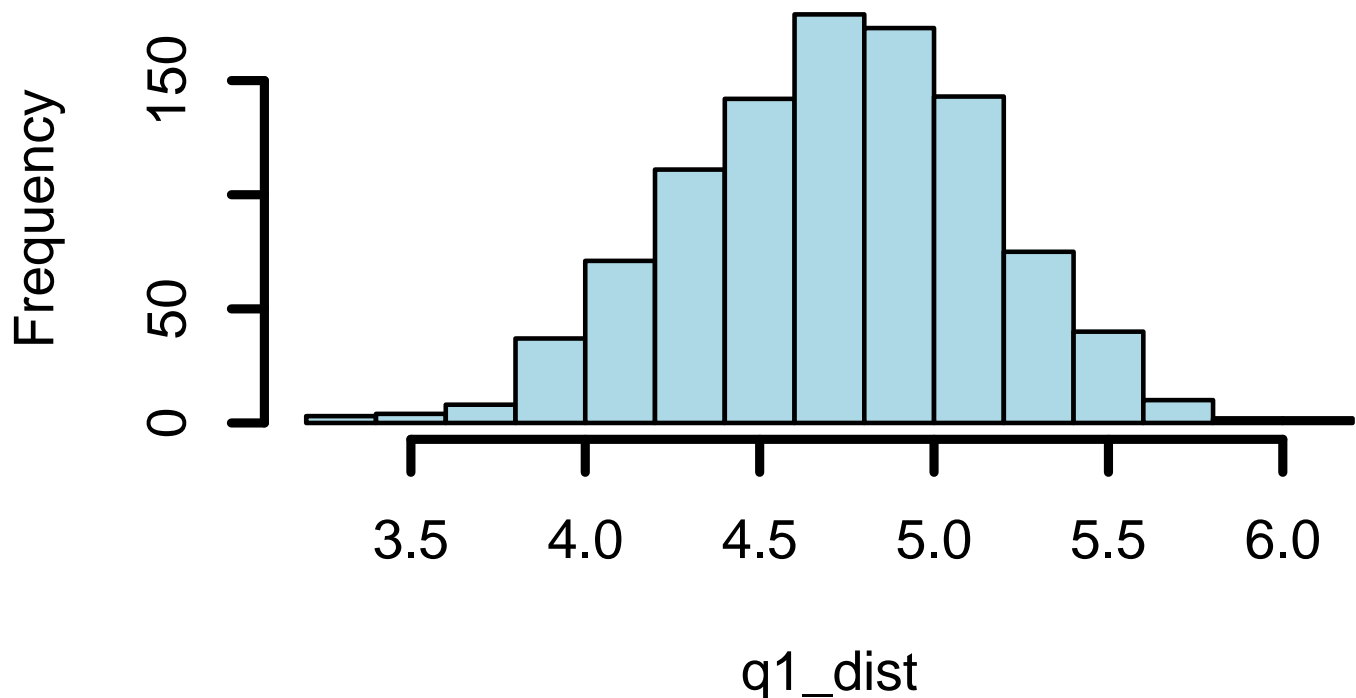
The 95% confidence interval for the mean weight change is 4.73 kg (3.8 kg, 5.66 kg). If the method used in this study were repeated many times, 95% of the time, the interval 3.8 kg and 5.66 kg will cover the true mean weight change. We can also say that we are 95% confident that the population mean weight gain is between (3.8 kg, 5.66 kg). Remember that the our uncertainty is about whether the particular sample we have at hand is one of the successful ones or one of the 5% that fail to produce an interval that captures the true value.

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 because we don't believe the CLT has kicked in.

- 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).

```
q1_dist <- replicate(1000, expr = {
  dplyr::sample_n(weight, size = n, replace = TRUE) %>%
    dplyr::summarize(r = mean(change)) %>%
    dplyr::pull(r)
})
hist(q1_dist, col = "lightblue", lwd = 2)
```

Histogram of q1_dist



```
round(quantile(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% 98% 99%
# 3.3 3.6 3.9 4.0 5.2 5.4 5.5 5.7
```

The 95% Bootstrap interval is given by [3.86, 5.54] kg. Very similar to the interval given in part a). Gives us some more confidence that the CLT has indeed kicked in.

- 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? We convert the Bootstrap CI by simply multiplying the upper and lower limits by 2.2lbs to give:

```
quantile(q1_dist, probs = c(.025, 0.975)) * 2.2
```

```
# 2.5% 98%  
# 8.5 12.2
```

The null hypothesis is that the theory of weight gain is the same as the measured weight gain, $H_0 : \mu = \mu_o = 16$ lbs, and the alternative hypothesis is $H_a : \mu \neq 16$ lbs (two tailed test). I want to test it using the bootstrap method because I don't want to assume that the CLT has kicked in and the sampling distribution is normal. Since the upper limit of the confidence interval is below 16 lbs, there is evidence to suggest that we should reject the null hypothesis, i.e., the actual weight gain might be lower than the theory says.

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.

2.1. Using the t procedure. The null hypothesis is that older student mean SSHA score equals all student mean SSHA scores, $H_0 : \mu = \mu_0 = 115$. The alternate hypothesis is that older student mean SSHA score is greater than 115, $H_a : \mu > 115$ (ie: one-tail test). The sample size is on the smaller side (ie: below 30), so I chose to use a one sample t-test because I don't trust the sample sd to be a good estimate of the population sd. A two tailed 95% CI is [120.64, 143.76]. Based on this, I reject the null hypothesis, i.e. , our data provides evidence that there might be a difference in SSHA scores between older students and the general population of students.

2.2. Using the z procedure.

```
qnorm(p = c(0.025, 0.975), mean = 132.2, sd = 30/sqrt(25))
```

```
# [1] 120 144
```

```
# alternatively  
132.2 + qnorm(p = c(0.025, 0.975)) * 30/sqrt(25)
```

```
# [1] 120 144
```

Assuming that the standard deviation of the all U.S. college students of 30 is accurate and taken as sigma, the 95% confidence interval for the population mean SSHA score is [120.44, 143.96].

- 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).

We are assuming that this is a simple random sample of older students. If using the t -distribution, we are assuming that the standard deviation of the population is not a good estimate of the standard deviation of our sample. The most important assumption we've made is that the CLT has kicked in and therefore the sampling distribution is normal.

If using z procedure \rightarrow that the standard deviation of the all U.S. college students of 30 is accurate and taken as sigma, and that the sample size is enough that the CLT has kicked in.

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.

```
(moon.mean <- mean(moon.diff))
```

```
# [1] 2.4
```

```
(moon.sdev <- sd(moon.diff))
```

```
# [1] 1.5
```

```
(moon.n <- length(moon.diff))
```

```
# [1] 15
```

```
qt_scaled(p=c(0.025, 0.975), df = moon.n - 1, mean = moon.mean, sd = moon.sdev/sqrt(moon.n))
```

```
# [1] 1.6 3.2
```

```
# alternatively
moon.mean + qt(p = c(0.025, 0.975), df = moon.n - 1) * moon.sdev/sqrt(moon.n)
```

```
# [1] 1.6 3.2
```

Assuming a simple random sample and that our sample is representative of the population, and that the difference in disruptive events is normally distributed in the population (OR you can say that you believe the Central Limit Theorem (CLT) has kicked in), I calculated a 95% confidence interval for the population mean difference of [1.62, 3.24] disruptive events on full moon days when compared to other days. This was done using the t procedure due to the unknown sigma and small sample size of $n = 15$.

- 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.

```
#t-statistic
(t_statistic <- (moon.mean - 0) / (moon.sdev/sqrt(moon.n)))
```

```
# [1] 6.5
```

```
# p-value
pt(q = t_statistic, df = moon.n-1, lower.tail = F)
```

```
# [1] 7.6e-06
```

$H_0 : \mu = 0$ (i.e. that there is no difference in disruptive behaviours on full moon nights when compared to other nights) and $H_a : \mu > 0$ (i.e. that there are more disruptive behaviours on full moon nights when compared to other nights).

I calculate a one-sided (to the right) p-value using the t statistic (because unknown sigma and small sample size of $n = 15$) of 7.59×10^{-6} . Assuming the null hypothesis is true, there is very small probability that the observed mean difference of 2.43 came from the null distribution. We could also say that since $\mu = 0$ is not contained within our 95% confidence interval, this provides evidence against the null hypothesis. This is based on the assumptions that the sample distribution here is enough that the CLT has kicked in as well.

- c. Find the minimum value of the mean difference in disruptive behaviors (\bar{y}) needed to reject the null hypothesis.

We first need to figure out for what values we can reject the null. Since this is a one-sided alternative we want and $\alpha = 0.05$ in the right tail. The cutoff is then given by:

```
(tscore.null <- qt(p = 0.95, df = 15-1))
```

```
# [1] 1.8
```

Then we solve for \bar{y} in the t -statistic formula:

$$t_{\text{statistic}} = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$
$$1.76 = \frac{\bar{y} - 0}{1.46/\sqrt{15}}$$

The solution to the above equation is:

```
tscore.null*(moon.sdev/sqrt(moon.n))
```

```
# [1] 0.66
```

Assuming H_0 is true, under the null distribution the minimum value in mean difference in disruptive behaviours needed to reject the null hypothesis at an $\alpha = 0.05$ level is 0.66 disruptive behaviours.

d. What is the probability of detecting an increase of 1.0 aggressive behavior per day during moon days?

$$H_0 : \mu = 0 \quad H_A : \mu > 1$$

This is a standard power calculation. In the first step, we calculate the difference in mean disruptive behaviours needed to reject the null hypothesis, which we calculated in part c) to be 0.66. The corresponding t value under the alternative hypothesis distribution is given by

$$\begin{aligned} t_{\text{statistic}} &= \frac{\bar{y} - \mu_A}{s/\sqrt{n}} \\ &= \frac{0.664 - 1}{1.46/\sqrt{15}} \\ &= -0.890 \end{aligned}$$

We then calculate the probability of observing this value or more under the alternative hypothesis distribution:

```
pt(q = -0.890, df = moon.n - 1, lower.tail = FALSE)
```

```
# [1] 0.81
```

Therefore, the power to detect an increase of 1.0 aggressive behavior per day during moon days is 80.57%.