stats100b_hw2

Takao Oba

4/10/2022

1

```r
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.32.47 PM.png")
```

1. (a) If T follows a t_7 distribution, find t_o such that $P(|T| < t_o) = 0.9$

(b). Assume that $X_1, \cdots, X_4$ are *i.i.d.* normal random variables $N(0, \sigma^2 = 4)$. Find the constant C such

$$P\left(\frac{X_1}{\sqrt{X_2^2 + X_3^2 + X_4^2}} < C\right) = 0.975$$

a

From the problem, we have the given equation that can be simplified.

$$P(|T| < t_0) = 0.9 = P(-t_0 < T < t_0)$$

Since t-distribution is symmetrical, we can narrow the equation to be

$$1 - 2P(T < -t_0) = 0.9$$

Simplifying the equation, we get that

$$P(T < -t_0) = 0.05$$

We will finally solve this through (this is a t distribution with degree of freedom of 7)

```r
cat("t0 = ", -qt(0.05, 7))
```

```
## t0 =  1.894579
```

b

We can utilize standardization of X through the given information.

$$X \sim N(0, 4)$$

Additionally, through transformation into standard normal, we have

$$\frac{X - \mu}{\sigma} = \frac{X - 0}{\sqrt{4}} = Z \sim N(0, 1)$$

Next, we can reconstruct the X that are given through

$$X = \sigma Z + \mu = \sqrt{4} Z \sim N(0, 4)$$

We will apply the given transformation to the original equation. Additionally, we can simplify the given equaiton

$$P(\frac{X_1}{\sqrt{X_2^2 + X_3^2 + X_4^2}} < C) = P(\frac{\sqrt{4}Z_1}{\sqrt{4Z_2^2 + 4Z_3^2 + 4Z_4^2}} < C) = P(\frac{\sqrt{4}Z_1}{\sqrt{4} \times \sqrt{Z_2^2 + Z_3^2 + Z_4^2}} < C) = P(\frac{Z_1}{\sqrt{Z_2^2 + Z_3^2 + Z_4^2}} < C)$$

We notice that we can transform this to chi squared variables through

$$P(\frac{Z_1}{\frac{1}{\sqrt{3}} \times \sqrt{Z_2^2 + Z_3^2 + Z_4^2}} < \sqrt{3}C) = P(\frac{Z}{\sqrt{\chi_3^2/3}} < \sqrt{3}C) = P(t_3 < \sqrt{3}C) = 0.975$$

```
cat("C = ",qt(0.975,3)/sqrt(3))
```

```
## C =  1.837386
```

2

```
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.32.51 PM.png")
```

2. Let S^2 be the sample variance of n i.i.d random variables from $N(\mu, \sigma^2)$. Suppose n=7, σ^2=9.0 . Find a and b so that
   (a) P(S^2 < b) =0.975 and
   (b) P (a < S^2)=0.975.

a

According to the theorem B from chapter 6.3 gives

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Since n = 7, the degree of freedom is 7-1=6 Additionally, the variance is 9.0

$$P(S^2 < b) = P(\frac{(n-1)S^2}{\sigma^2} < \frac{(n-1)b}{\sigma^2}) = P(\chi_6^2 < \frac{2}{3}b) = 0.975$$

```r
cat("b = ",qchisq(0.975, 6)*3/2)
```

```
## b =  21.67406
```

b

$$P(a < S^2) = P(\frac{(n-1)a}{\sigma^2} < \frac{(n-1)S^2}{\sigma^2}) = P(\frac{2}{3}a < \chi^2_6) = 0.975$$

```r
cat("a = ", qchisq(0.975, 6, lower.tail = F)*3/2)
```

```
## a =  1.856016
```

3

```r
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.32.54 PM.png")
```

3. (a) Assume that $X_1, \cdots, X_5$ are *i.i.d.* normal random variables $N(0, \sigma^2 = 9)$. Find the constant C such

$$P(\frac{X_1^2 + X_2^2 + X_3^2}{X_4^2 + X_5^2} < C) = 0.975$$

(b). Assume that X_1, X_2, X_3 are *i.i.d.* normal random variables $N(0, \sigma^2 = 9)$. Find the constant C such

$$P(\frac{X_2^2 + X_3^2}{X_1^2 + X_2^2 + X_3^2} < C) = 0.975$$

[Hint: are numerator and denominator independent ?]

a

We want to reconstruct X from standard normal of z

$$P(\frac{X_1^2 + X_2^2 + X_3^2}{X_4^2 + X_5^2} < C) = P(\frac{\sigma^2 Z_1^2 + \sigma^2 Z_2^2 + \sigma^2 Z_3^2}{\sigma^2 Z_4^2 + \sigma^2 Z_5^2} < C) = P(\frac{\chi^2_3}{\chi^2_2} < C)$$

We must notice that the ratio of chi squared to f distribution

$$P(\frac{\chi^2_3/3}{\chi^2_2/2} < \frac{2}{3}C) = P(F_{3,2} < \frac{2}{3}C) = 0.975$$

b

The difference between this problem and part a is that there are overlapping X terms in the numerator and the denominator. We will simplify the equation using algebra through the bottom equation.

$$P(\frac{X_2^2+X_3^2}{X_1^2+X_2^2+X_3^2} < C) = P(\frac{X_1^2+X_2^2+X_3^2}{X_2^2+X_3^2} > \frac{1}{C}) = P(1+\frac{X_1^2}{X_2^2+X_3^2} > \frac{1}{C}) = P(\frac{X_1^2}{X_2^2+X_3^2} > \frac{1}{C}-1)$$

Again, we will change x into chi squared.

$$P(\frac{X_1^2}{X_2^2+X_3^2} > \frac{1}{C}-1) = P(\frac{\sigma^2 Z_1^2}{\sigma^2 Z_2^2+\sigma^2 Z_3^2} > \frac{1}{C}-1) = P(\frac{\chi_1^2}{\chi_2^2} > \frac{1}{C}-1)$$

Additionally, noticing that the ratio of the chi square is the F distribution, we can simplify the equation into

$$P(\frac{\chi_2^2/1}{\chi_2^2/2} > \frac{2}{C}-2) = P(F_{1,2} > \frac{2}{C}-2) = 0.975$$

```
cat("C = ", 2/(qf(0.975, 1, 2, lower.tail = F) + 2))
```

```
## C =  0.999375
```

4

```
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.32.58 PM.png")
```

4. The density function of X is $f(x\,|\,\theta) = (1+\theta)x^{\theta},\ 0 \le x \le 1$
where the domain of the parameter is restricted to $\theta > -1$
Suppose we have observed a random sample of size *n* independently generated from this distribution, $x_1, \ldots x_n$.
Find the method of moment estimate θ . For a small dataset of size n= 3, $x_1 = 0.3,\ x_2 = 0.5,\ x_3 = 0.4$, give the estimate.

We intend to find the method of moment estimate of theta. We will first utilize the theorem of the first moment of X and solve the integration.

$$E(X) = \int_0^1 x(1+\theta)x^{\theta}dx = (1+\theta)\int_0^1 x^{1+\theta}dx = \frac{1+\theta}{2+\theta}x^{2+\theta}|_0^1 = \frac{1+\theta}{2+\theta}$$

Additionally, we know that this is a small data set of size n = 3, so we can directly calculate the sample mean

$$\hat{\mu} = \frac{1}{3}(0.3+0.5+0.4) = 0.4$$

Since according to definition, these two values should equal since one is a theoretical expectation and the other is the sample average, we will equate the two values and attempt to solve for the variable theta.

$$0.4 = \frac{1+\theta}{2+\theta}$$

Solving this equation, we get that

$$\hat{\theta} = -\frac{1}{3}$$

This goes along with the domain of the parameter which is given in the problem that the theta should be greater than -1.

5

```
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.33.09 PM.png")
```

5. (A) Suppose that X is a discrete random variable with
$P(X=0) = \frac{2}{3}\theta,\ P(X=1) = \frac{1}{3}\theta,\ P(X=2) = 1-\theta$
where the domain of the parameter is $0 < \theta < 1$. Suppose 5 independent observations are obtained from this distribution, yielding a small dataset: 1, 0, 2, 1, 2 .
Find the method of moment estimate of θ .

(B) Suppose.
$P(X=0) = \frac{1}{2}\theta,\ P(X=1) = 1-\theta,\ P(X=2) = \frac{1}{2}\theta$ and the data are 1,0,0, 2, 2
Find the method of moment estimate of θ.
[hint: consider higher order moment if necessary]

a

Unlike the previous problem, this is considering X as a discrete random variable, so we will not be working with integrals.

We will find the theoretical expectation first again.

$$E(X) = 0 \times \frac{2}{3}\theta + 1 \times \frac{1}{3}\theta + 2 \times (1-\theta) = 2 - \frac{5}{3}\theta$$

Moving on, we will find the sample average, which again can be found manually.

$$\hat{\mu} = \frac{1}{5}(1+0+2+1+2) = \frac{6}{5}$$

Again, we will equate the theoretical expectation with the sample average. Doing this will yield the equation

$$\frac{6}{5} = 2 - \frac{5}{3}\theta$$

Solving the equation will give

$$\hat{\theta} = \frac{12}{25}$$

Again, this is in the domain of the parameter which is between 0 and 1 which is stated in the problem.

b

We will find the theoretical expectation first again.

$$E(X) = 0 \times \frac{1}{2}\theta + 1 \times (1-\theta) + 2 \times \frac{1}{2}\theta = 1$$

Moving on, we will find the sample average, which again can be found manually.

$$\hat{\mu} = \frac{1}{5}(1+0+0+2+2) = 1$$

Again, we will equate the theoretical expectation with the sample average. Doing this will yield the equation

$$1 = 1$$

Thus, we can draw the conclusion that all values will be the moment estimate of theta.

6

```r
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.33.15 PM.png")
```

6. An electronic safety component in a delicate device from a company has a lifetime distribution that follows a Gamma distribution with shape and rate parameters α, λ. To improve the life time, a new model is equipped with one additional backup component that can be automatically switched for uninterrupted use, once the current component fails.

Suppose the life time for 20 devices of the new model are reported, $x_1, \cdots, x_{20}$

Find the method of moment estimates for α, λ .

According to the problem, the new model is equipped with on additional backup component in addition to the original device. Both of them follow the gamma distribution. We can add the two gamma distributions and call it a X

Let U be

$$U \sim Gamma(\alpha, \lambda)$$

Let V be

$$V \sim Gamma(\alpha, \lambda)$$

Thus, collectively, we have that according to property,

$$U + V = X \sim Gamma(2\alpha, \lambda)$$

Additionally, according to property of the gamma distribution, we also have that

$$E(X) = \frac{2\alpha}{\lambda}, Var(X) = \frac{2\alpha}{\lambda^2} = E(X^2) - E(X)^2$$

Similarly to the other problems, we can equate the theoretical moments and the empirical moments.

$$\frac{2\alpha}{\lambda} = \frac{1}{20}\sum_{i=1}^{2} x_i = \bar{x}$$

Now, looking at the variance, we have that

$$\frac{2\alpha}{\lambda^2} = \frac{1}{20}\sum_{i=1}^{20} x_i^2 - (\frac{1}{20}\sum_{i=1}^{20} x_i)^2 = \frac{1}{\lambda} \times \frac{2\alpha}{\lambda} = \frac{1}{\lambda} \times \frac{1}{20}\sum_{i=1}^{20} x_i = \hat{\sigma}^2$$

As we have the given values in different variable terms, we have

$$\hat{\lambda} = \frac{\sum_{i=1}^{20} x_i}{\sum_{i=1}^{20} x_i^2 - \frac{1}{20}(\sum_{i=1}^{20} x_i)^2} = \frac{\bar{x}}{\hat{\sigma}^2}$$

Now that we have the moment estimates of lambda, we can find the moment estimates for alpha

$$\hat{\alpha} = \frac{1}{2} \times \lambda \times \frac{1}{20}\sum_{i=1}^{20} x_i = \frac{\frac{1}{40}(\sum_{i=1}^{20} x_i)^2}{\sum_{i=1}^{20} x_i^2 - \frac{1}{20}(\sum_{i=1}^{20} x_i)^2} = \frac{1}{2} \times \frac{\bar{x}^2}{\hat{\sigma}^2}$$

Notice how the values of lambda and alpha are represented as functions of the data x only.

7

```
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.33.19 PM.png")
```

7. *Continued from 6.* Assume that the shape parameter of the Gamma distribution is known to be equal to 1; in other words, the lifetime distribution is exponential with unknown rate parameter λ. It has been suspected that the backup component is of poorer quality than the original component, with a shorter lifetime on the average. But you don't have records on the lifetime of the original component in each device. Is it still possible to estimate the mean lifetime difference from the reported $x_1, \cdots, x_{20}$ only ?

Hint : denote the average lifetime for the original component by μ_1 and the average lifetime for the backup by μ_2. Decompose the lifetime of the device as X=U+V , where U follows exponential distribution with mean μ_1 and V follows exponential distribution with mean μ_2. Use the method of moment to set up two estimation equations and attempt to find a solution for The difference $\mu_1 - \mu_2$.

Let

$$U \sim Exp(\lambda_1)$$

and let

$$V \sim Exp(\lambda_2)$$

Additionally, through algebra, we have that

$$(\mu_1 - \mu_2)^2 = \mu_1^2 - 2\mu_1\mu_2 + \mu_2^2 = (\mu_1 + \mu_2)^2 - 4\mu_1\mu_2$$

Next, we have that

$$(\mu_1 - \mu_2)^2 = (\frac{1}{\lambda_1} + \frac{1}{\lambda_2})^2 - 4\frac{1}{\lambda_1\lambda_2} = (\frac{\lambda_1 + \lambda_2}{\lambda_1\lambda_2})^2 - 4\frac{1}{\lambda_1\lambda_2}$$

If we let X = U+V, we can move on to the moment generating function where

$$M_U(t) = \frac{\lambda_1}{\lambda_1 - t}, M_V(t) = \frac{\lambda_2}{\lambda_2 - t},\ and\ M_X(t) = \frac{\lambda_1}{\lambda_1 - t} \times \frac{\lambda_2}{\lambda_2 - t}$$

Taking the first derivative of this function, we have that

$$M_X'(t) = -\lambda_1\lambda_2 \times (\lambda_1\lambda_2 - \lambda_1 t - \lambda_2 t + t^2)^{-2} \times (-\lambda_1 - \lambda_2 + 2t)$$

Next, we find the second derivative,

$$M_X''(t) = -\lambda_1\lambda_2 \times [-2(\lambda_1\lambda_2 - \lambda_1 t - \lambda_2 t + t^2)^{-3}(-\lambda_1 - \lambda_2 + 2t)^2 + 2(\lambda_1\lambda_2 - \lambda_1 t - \lambda_2 t + t^2)^{-2}]$$

We can then move on to solving the expectation and variance of X

$$E(X) = M_X'(0) = \lambda_1\lambda_2 \times (\lambda_1\lambda_2)^{-2} \times (\lambda_1 + \lambda_2) = \frac{\lambda_1 + \lambda_2}{\lambda_1\lambda_2}$$

We can find the variance through the function

$$Var(X) = E(X^2) - E(X)^2 = (\frac{\lambda_1 + \lambda_2}{\lambda_1\lambda_2})^2 - \frac{2}{\lambda_1\lambda_2}$$

We can again equate the theoretical moments and empirical moments:

$$E(X) = \frac{\lambda_1 + \lambda_2}{\lambda_1\lambda_2} = \bar{x}$$

Moving on to the variance of X, we have that

$$Var(X) = (\frac{\lambda_1 + \lambda_2}{\lambda_1\lambda_2})^2 - \frac{2}{\lambda_1\lambda_2} = \bar(x)^2 - \frac{2}{\lambda_1\lambda_2} = \hat{\sigma}^2$$

working through, we have that

$$\frac{2}{\bar{x}^2 - \hat{\sigma}^2} = \lambda_1\lambda_2$$

Moving on, we will estimate the mean difference using data:

We have that

$$(\mu_1 - \mu_2)^2 = 2\hat{\sigma}^2 - \bar{x}^2$$

We solve for the difference through taking the square root which yields

$$|\mu_1 - \mu_2| = \sqrt{\frac{1}{10}\sum_{i=1}^{20} x_i^2 - 3(\frac{1}{20}\sum_{i=1}^{20} x_i)^2}$$

Again, the final answer is expressed as functions of data x only.

8

```
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.33.23 PM.png")
```

8. Four random numbers, 1.1650, 0.6268, 0.0751, 0.3516, were generated by a computer code of a normal distribution with mean μ_1 and variance σ^2. Five more normal random numbers, 0.3035, 2.6961, 1.0591, 2.7971, 1.2641, were generated by using the same variance σ^2 but a different mean μ_2 . The record about the values of μ_1, μ_2, σ^2 was lost. Please use all available data to find
(a) a 95% confidence interval of $\mu_1 - \mu_2$
(b) A 95% confidence interval of σ^2
(c) A 95% confidence interval of μ_1

a

We can directly use the given data.

$$[(\bar{x}_1 - \bar{x}_2 - t_{1-\frac{\alpha}{2},m+n-2} \times S_p\sqrt{\frac{1}{n} + \frac{1}{m}}, (\bar{x}_1 - \bar{x}_2) + t_{1-\frac{\alpha}{2},m+n-2} \times S_p\sqrt{\frac{1}{n} + \frac{1}{m}}$$

additionally, utilizing the formula to obtain the pooled variance, we have

$$S_p^2 = \frac{(n-1)S_{X_1}^2 + (m-1)S_{X_2}^2}{m+n-2}$$

Utilizing the tool that we have, we will solve for the confidence interval

```
x1 <- c(1.165, 0.6268, 0.0751, 0.3516)
x2 <- c(0.3035, 2.6961, 1.0591, 2.7971, 1.2641)
# nx1 = 4
# nx2 = 5

Sp <- sqrt(((4-1)*var(x1) + (5-1)*var(x2)) / (4 + 5 - 2))
t_score <- qt(1-0.5 / 2, 5 + 4 - 2)
lower <- (mean(x1) - mean(x2) - t_score*Sp*sqrt(1/4 + 1/5))
upper <- (mean(x1) - mean(x2) + t_score*Sp*sqrt(1/4 + 1/5))
cat("The 95 % confidence interval is [", lower, ",", upper, "].")
```

```
## The 95 % confidence interval is [ -1.487063 , -0.6516465 ].
```

b

By definition, the 95 percent confidence interval for the variance is

$$[\frac{(n-1)S_{X_1}^2}{\chi^2_{\alpha/2}}, \frac{(n-1)S_{X_1}^2}{\chi^2_{1-\alpha/2}}]$$

We utilize the five data points to calculate the confidence interval:

```
lower <- (5-1)*var(x1) / qchisq(0.05 / 2, 5-1, lower.tail = F)
upper <- (5-1)*var(x1) / qchisq(1 - 0.05 / 2, 5-1, lower.tail = F)
cat("The 95 percent confidence interval is [", lower, ",", upper,"].")
```

```
## The 95 percent confidence interval is [ 0.07764678 , 1.786142 ].
```

c

Next, we will attempt to solve the confidence interval for mu1. This can be done by looking at the individual values.

Again, by definition, we have that

$$[\bar{x}_1 - t_{1-\frac{\alpha}{2},n-1} \times \frac{S_{X_1}}{\sqrt{n}}, \bar{x}_1 + t_{1-\frac{\alpha}{2},n-1} \times \frac{S_{X_1}}{\sqrt{n}}]$$

Now, we simply will plug values into this above equation

```
t_score <- qt(1-0.05/2, 4-1)
lower <- mean(x1) - t_score*sd(x1) / sqrt(4)
upper <- mean(x1) + t_score*sd(x1) / sqrt(4)
cat("The 95 percent Confidence Interval is [", lower, ",", upper, "].")
```

```
## The 95 percent Confidence Interval is [ -0.1854393 , 1.294689 ].
```

9

```
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.33.32 PM.png")
```

9. A drug is studied for possible improvement of behavior rating of patients suffering from a specific type of mental illness. The study recruited 10 patients who were randomly assigned to a treatment group (patients receiving the drug) and a placebo group (patients receiving a placebo). Two behavioral rating scores were recorded for each subject, a "Before" score (the score at the beginning of the clinical trial) and an "After" score (the score 3 months later). You may assume normal distribution for the scores.
(a) Find a 95% confidence for placebo effect.
(b) Find a 95% confidence interval for the placebo-adjusted net drug improvement effect.
[hint : (a) placebo effect is defined as the parameter = "mean of After " minus "mean of Before" for the placebo group ; (b) the parameter of interest = ("mean of After" minus "mean of Before" for the treatment group) minus the placebo effect defined in (a)]

Placebo group	PL1	PL2	PL3	PL4	PL5
Before	1.9	1.5	1.7	2.4	1.5
After	1.91	1.45	1.54	2.54	1.54

Treatment group	TR1	TR2	TR3	TR4	TR5
Before	1.7	2.6	1.6	1.6	2.3
After	**2.18**	**2.45**	**1.72**	**1.63**	**2.45**

Assume equal variance or assume unequal variance have to write both cases down

a

By definition, we have that

$$[\bar{d}_{placebo} - t_{1-\frac{\alpha}{2},n-1} \times \frac{S_{d_{placebo}}}{\sqrt{n}}, \bar{d}_{placebo} + t_{1-\frac{\alpha}{2},n-1} \times \frac{S_{d_{placebo}}}{\sqrt{n}}]$$

Next, we just have to plug values into the given equation

```
placebo_before <- c(1.9, 1.5, 1.7, 2.4, 1.5)
placebo_after <- c(1.91, 1.45, 1.54, 2.54, 1.54)
difference <- placebo_after - placebo_before
```

```
sddata <- sd(difference)
t_score <- qt(1 - 0.05 / 2, 5 - 1)
lower <- mean(difference) - t_score*sddata / sqrt(5)
upper <- mean(difference) + t_score*sddata / sqrt(5)
cat("The 95 percent confidence interval is [", lower, ",", upper, "].")
```

```
## The 95 percent confidence interval is [ -0.141875 , 0.133875 ].
```

b

We will use pairing for this case

```
treatment_before <- c(1.7, 2.6, 1.6, 1.6, 2.3)
treatment_after <- c(2.18, 2.45, 1.72, 1.63, 2.45)
d_treatment <- treatment_after - treatment_before
d_placebo <- placebo_after - placebo_before
```

Since not explicitly stated, we must test two cases: assume equal variance or assume unequal variance. First, assuming equal variance, we have by definition that

$$[(\bar{d}_{treatment} - \bar{d}_{placebo}) - t_{1-\frac{\alpha}{2},m+n-2} \times S_p\sqrt{\frac{1}{n} + \frac{1}{m}}, (\bar{d}_{treatment} - \bar{d}_{placebo}) + t_{1-\frac{\alpha}{2},m+n-2} \times S_p\sqrt{\frac{1}{n} + \frac{1}{m}}]$$

we will find the pooled variance through definition as well.

$$S_p^2 = \frac{(n-1)S^2_{d_{treatment}} + (m-1)S^2_{d_{placebo}}}{m+n-2}$$

Through plugging in the numbers again, we have that

```
Sp <- sqrt(((5-1)*var(d_treatment) + (5-1)*var(d_placebo)) / (5+5-2))
t_score <- qt(1 - 0.05 / 2, 5 + 5 - 2)
lower <- (mean(d_treatment) - mean(d_placebo)) - t_score*Sp*sqrt(1/5 + 1/5)
upper <- (mean(d_treatment) - mean(d_placebo)) + t_score*Sp*sqrt(1/5 + 1/5)
cat("The 95 percent confidence interval is [", lower, ",", upper, "]")
```

```
## The 95 percent confidence interval is [ -0.1332483 , 0.3932483 ]
```

Now, we take the case of assuming unequal variance

$$[(\bar{d}_{treatment} - \bar{d}_{placebo}) - t_{1-\frac{\alpha}{2},df} \times \sqrt{\frac{S^2_{d_{after}}}{n} + \frac{S^2_{d_{before}}}{m}}, (\bar{d}_{treatment} - \bar{d}_{placebo}) + t_{1-\frac{\alpha}{2},df} \times \sqrt{\frac{S^2_{d_{after}}}{n} + \frac{S^2_{d_{before}}}{m}}]$$

We have from definition that the degrees of freedom is

$$df = \frac{(\frac{S^2_{d_{after}}}{n} + \frac{S^2_{d_{before}}}{m})^2}{(\frac{S^2_{d_{after}}}{n})^2/(n-1) + (\frac{S^2_{d_{before}}}{m})^2/(m-1)}$$

```r
ESE <- sqrt(var(d_treatment) / 5 + var(d_placebo) / 5)
df <- (var(d_treatment) / 5 + var(d_placebo) / 5)^2 / ((var(d_treatment) / 5) ^2 / 4 + (var(d_placebo)/

t_score <- qt(1-0.05/2, df)
lower <- (mean(d_treatment) - mean(d_placebo)) - t_score*ESE
upper <- (mean(d_treatment) - mean(d_placebo)) + t_score*ESE
cat("the 95 percent confidence interval is [", lower, ",", upper, "]" )
```

```
## the 95 percent confidence interval is [ -0.1520471 , 0.4120471 ]
```

10

```r
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.33.37 PM.png")
```

10.
(a) Three hundred students reported 95% confidence intervals of population mean based on the data each student collected independently . Approximately 15 of the 300 answers will not contain the true population mean. True or False ?
(b) A 90% confidence interval for μ, the average number of children per household in a city, is found to be [0.87, 1.15]. The probability is 0.90 for $0.87 \le \mu \le 1.15$. True or False ?
(c) A 90% confidence interval for μ, the average number of children per household in a city, is found to be [0.87, 1.15]. Then because 1 is the only integer between 0.87 and 1.15,
A student infers that about 90 percents of families in the city have only one child. Is the student's claim true or false ?
(d) The confidence interval [0.87, 1.15] was based on a sample of size 100. Next year a similar survey will take place. What sample size would you recommend if the city hopes to shorten the width of the confidence interval from 0.28 (= 1.15-0.87) to about 0.10 ?

a

This is a TRUE statement

b

This is FALSE. We cannot necessarily say that the probability is 0.9 to interpret a Confidence interval. We must say that we are 90 % confident that the true average number of children per household in a study is between 0.87 and 1.15.

c

This is FALSE. Although the number of children can only be realistically represented by integers, we are taking means and variance and the conclusion that the student is making does not make sense.

d

According to the problem, we intend to find a sample size such that

$$(\bar{x} + Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n'}}) - (\bar{x} - Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n'}}) = 0.1$$

Since we also know that

$$2 \times Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{100}} = 0.28$$

We can derive to

$$\frac{\sqrt{100}}{\sqrt{n'}} = \frac{0,1}{0,28}$$

Thus, we get that the approximation of n is

$$n' \approx 784$$