parcial 2 parte 2

Andres Felipe Beltran Rodriguez

12/16/2021

# Punto 1

En un estudio realizado en una finca donde se sembraron árboles de la misma especie 20 en total. Se obtuvieron los siguientes resultados para el diámetro en centímetros de los árboles después de seis meses de sembrados:

## a

Seleccione una muestra aleatoria sistemática de tamaño 8 y estime el parámetro de la media. Construya un intervalo del 99% de confianza para la media, hallar el error de muestreo e intérprete el resultado claramente (tome los 3 primeros aleatorios para el arranque). Detallar la muestra seleccionada.

Antes de empezar cargamos los datos:

tabla <- data.frame(  
 No = c(1:20) ,  
 diametro = c( 18, 12, 15, 13, 15,17,18,12,18,14,13,16,20,17,16,14,12,19,17,15),  
 aleatorio = c(0.268,0.037,0.629,0.349,0.425,0.773,0.062,0.458,0.405,0.131,0.319,0.35,0.561,0.714,0.022,0.895,0.386,0.04,0.149,0.225  
)  
)  
knitr::kable(tabla)

|  |  |  |
| --- | --- | --- |
| No | diametro | aleatorio |
| 1 | 18 | 0.268 |
| 2 | 12 | 0.037 |
| 3 | 15 | 0.629 |
| 4 | 13 | 0.349 |
| 5 | 15 | 0.425 |
| 6 | 17 | 0.773 |
| 7 | 18 | 0.062 |
| 8 | 12 | 0.458 |
| 9 | 18 | 0.405 |
| 10 | 14 | 0.131 |
| 11 | 13 | 0.319 |
| 12 | 16 | 0.350 |
| 13 | 20 | 0.561 |
| 14 | 17 | 0.714 |
| 15 | 16 | 0.022 |
| 16 | 14 | 0.895 |
| 17 | 12 | 0.386 |
| 18 | 19 | 0.040 |
| 19 | 17 | 0.149 |
| 20 | 15 | 0.225 |

Primero, podemos encontrar la relación entre el tamaño de muestra y la población:

N <- 20  
n <- 8  
a <- N/n  
a

## [1] 2.5

Aproximando al número entero inmediatamente mayor:

a <- 3

Para la selección de muestras, podemos crear una matriz:

mat <- matrix(c(1:20,1:4),  
 ncol = a,  
 byrow = T)  
mat

## [,1] [,2] [,3]  
## [1,] 1 2 3  
## [2,] 4 5 6  
## [3,] 7 8 9  
## [4,] 10 11 12  
## [5,] 13 14 15  
## [6,] 16 17 18  
## [7,] 19 20 1  
## [8,] 2 3 4

Teniendo que cada fila de la matriz es un intervalo, se selecciona la posicion que tiene el menor número aleatorio de los primeros tres, en este caso es el 0.037 que corresponde a la segunda posición:

muestraID <- mat[,2]  
muestraID

## [1] 2 5 8 11 14 17 20 3

muestra <- tabla$diametro[muestraID]   
muestra

## [1] 12 15 12 13 17 12 15 15

Podemos entonces calcular la media:

mean(muestra)

## [1] 13.875

la varianza:

var(muestra)

## [1] 3.553571

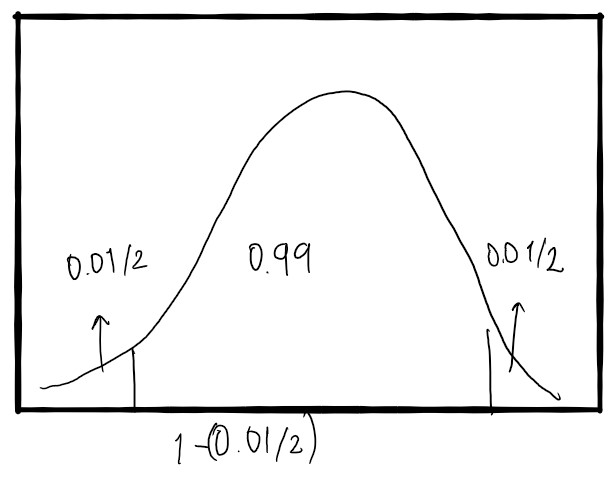
Y el estimador para la varianza de la media:

estimVar <- ((1-(length(muestra)/nrow(tabla)) )/length(muestra))\*var(muestra)  
  
estimVar

## [1] 0.2665179

Dado que la muestra es 8, puede compararse el uso de la distribucion normal estándar y la distribución t para el intervalo de confianza:

### usando normal estándar



qnormal <- qnorm(1-(0.01/2))  
qnormal

## [1] 2.575829

Usando el cuantil y el estimador de varianza podemos calcular el intervalo de confianza:

inferiorNorm <- mean(muestra) - (sqrt(estimVar)\*qnormal)   
inferiorNorm

## [1] 12.54522

SuperiorNorm <- mean(muestra) + (sqrt(estimVar)\*qnormal)   
SuperiorNorm

## [1] 15.20478

### usando t student

qt <- qt(1-(0.01/2), df = length(muestra -1))  
qt

## [1] 3.355387

De nuevo, usando el cuantil y el estimador de la varianza

inferiort <- mean(muestra) - (sqrt(estimVar)\*qt)   
inferiort

## [1] 12.14277

Superiort <- mean(muestra) + (sqrt(estimVar)\*qt)   
Superiort

## [1] 15.60723

### Error de muestreo:

Podemos calcular el error de muestreo de la forma:

sqrt(estimVar)/mean(muestra)\*100

## [1] 3.720747

Al escoger una muestra aleatoria de los diametros de los arboles despues de seis meses de sembrados, se obtiene un valor promedio de 13.87 1.73 cm con un error de muestreo del 3.72 %.

## b

Seleccione una muestra aleatoria simple de tamaño 10 y estime el parámetro del total. Construya un intervalo del 95% de confianza, calcule el porcentaje de error de muestreo e interprete el resultado. Detallar la muestra seleccionada.

Se puede escoger entre el método fan muller y el coordinado negativo. Para este caso utilizamos el coordinado negativo ordenando las muestras en orden ascendiente del número aleatorio asociado a cada fila:

tablaCorNegativo <- tabla[with(tabla, order(aleatorio)),]  
knitr::kable(tablaCorNegativo)

|  |  |  |  |
| --- | --- | --- | --- |
|  | No | diametro | aleatorio |
| 15 | 15 | 16 | 0.022 |
| 2 | 2 | 12 | 0.037 |
| 18 | 18 | 19 | 0.040 |
| 7 | 7 | 18 | 0.062 |
| 10 | 10 | 14 | 0.131 |
| 19 | 19 | 17 | 0.149 |
| 20 | 20 | 15 | 0.225 |
| 1 | 1 | 18 | 0.268 |
| 11 | 11 | 13 | 0.319 |
| 4 | 4 | 13 | 0.349 |
| 12 | 12 | 16 | 0.350 |
| 17 | 17 | 12 | 0.386 |
| 9 | 9 | 18 | 0.405 |
| 5 | 5 | 15 | 0.425 |
| 8 | 8 | 12 | 0.458 |
| 13 | 13 | 20 | 0.561 |
| 3 | 3 | 15 | 0.629 |
| 14 | 14 | 17 | 0.714 |
| 6 | 6 | 17 | 0.773 |
| 16 | 16 | 14 | 0.895 |

Y se seleccionan las primeras 10:

tablaCorNegativo[1:10,]

## No diametro aleatorio  
## 15 15 16 0.022  
## 2 2 12 0.037  
## 18 18 19 0.040  
## 7 7 18 0.062  
## 10 10 14 0.131  
## 19 19 17 0.149  
## 20 20 15 0.225  
## 1 1 18 0.268  
## 11 11 13 0.319  
## 4 4 13 0.349

Podemos entonces calcular el total:

sum(tablaCorNegativo[1:10,2])

## [1] 155

Y su correspondiente estimador:

estimadorTotal <- sum(tablaCorNegativo[1:10,2])/(10/20)  
estimadorTotal

## [1] 310

a su vez la varianza de la muestra:

var(tablaCorNegativo[1:10,2])

## [1] 6.055556

y el estimador de varianza:

estimvarTotal <- 20^2 \*(1 - 10/20)/(10) \* var(tablaCorNegativo[1:10,2])  
estimvarTotal

## [1] 121.1111

Para determinar el intervalo de confianza para el total con un 95 %

ztotal <- 1.96

usando el estimador y el cuantil podemos calcular el intervalo:

inferiortotal <- estimadorTotal - (sqrt(estimvarTotal)\*ztotal)   
inferiortotal

## [1] 288.4301

Superiortotal <- mean(estimadorTotal) + (sqrt(estimvarTotal)\*ztotal)   
Superiortotal

## [1] 331.5699

Con un 95 de confianza se puede reportar que el total de diametros en la población se encientra entre 288 y 332 cm.

# Segundo punto

Un estudiante va todos los días de su trabajo a clase de Estadística en la universidad Nacional. El tiempo promedio para un viaje de ida es 24 minutos, con una varianza de 14.44 minutos. Suponga que la distribución de los tiempos de viaje está distribuida normalmente. (usar dos decimales en los valores estandarizados Z).

## a

Cuál es la probabilidad de que en un viaje tome al menos ½ hora.

media hora son 30 minutos:

Teniendo en cuenta que la variable se distribuye como:

Se puede calcular el cuantil para 30 minutos de la siguiente manera:

z <- (30-24)/sqrt(14.44)  
z

## [1] 1.578947

teniendo el cuantil de

z <- round(z,2)  
z

## [1] 1.58

podemos calcular la probabilidad de que obtener un valor mayor o igual al experimental, es decir, :

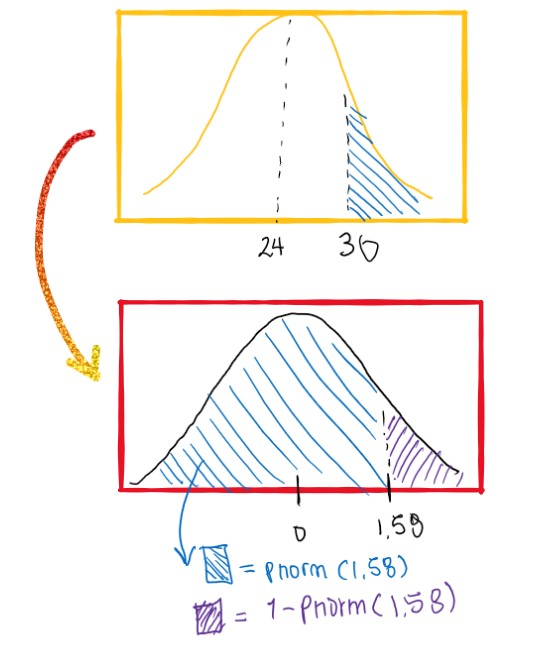
p <- 1-pnorm(1.58)  
p

## [1] 0.05705343

p2 <- 1-pnorm(30, mean = 24, sd = sqrt(14.44))  
p2

## [1] 0.05717406

La probabilidad de que el estudiante se demore en su viaje del trabajo a la universidad un tiempo de por lo menos hora es de

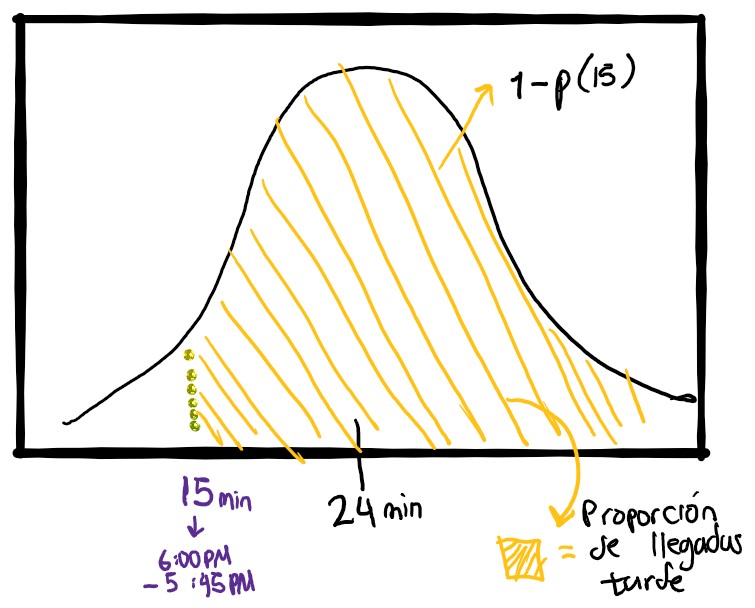


## b

Si la clase inicia a las 6:00 PM y él sale del trabajo para la universidad a las 5:45 PM, ¿Qué porcentaje de las veces llega tarde a la universidad?

Para saber esto, debemos tener en cuenta que el viaje no debe durar más de 15 minutos.

Lo que pregunta el enunciado, es la probabilidad de obtener un tiempo de viaje mayor a 15 minutos:



Podemos entonces calcular la probabilidad:

p15 <- 1-pnorm(15, mean = 24, sd = sqrt(14.44))  
p15

## [1] 0.9910679

p15\*100

## [1] 99.10679

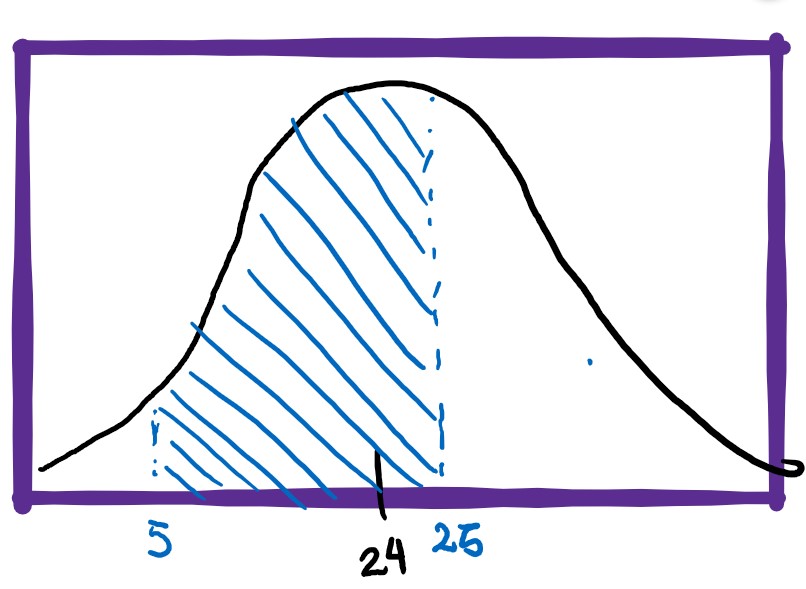
Dado que la clase inicia a las 6:00 PM y el estudiante sale de trabajar a las 5:45 PM, el porcentaje de veces que llega tarde a la clase es de .

## c

Si sale de su trabajo a las 6:15 PM y el profesor hace quiz de 6:20 PM a 6:40 PM. Cuál es la probabilidad de que el estudiante pierda la oportunidad de presentar el quiz ?

Podemos calcular el tiempo de viaje en cada limite de tiempo del quiz:

Entonces debemos hallar:



Para hallar el area entre los límites 5 y 25 podemos:

1. hallar el area de menos infinito a 25
2. hallar el aria de menos infinito a 5 3 . restarle el area de menos infinito a 5 al area de menos infinito a 25

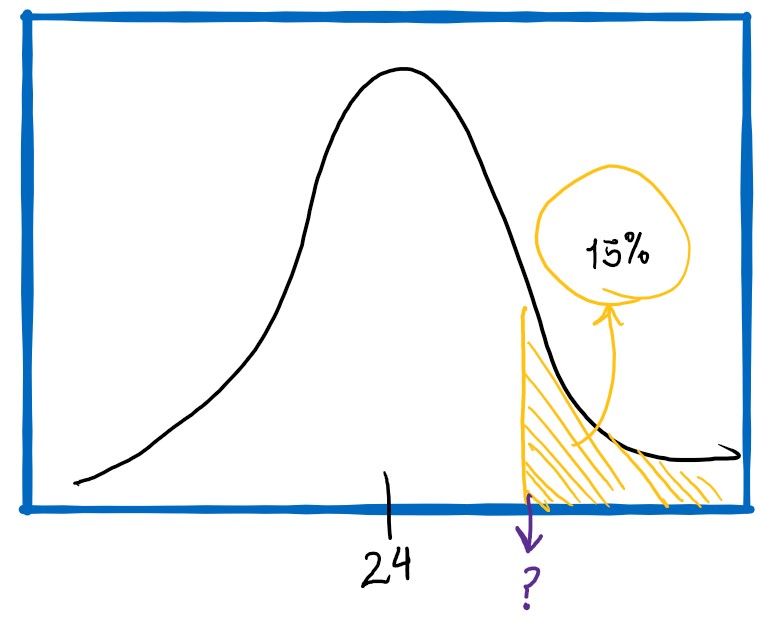
pquiz <- pnorm(25, mean = 24 , sd = sqrt(14.44)) - pnorm(5,mean = 24 , sd = sqrt(14.44))  
pquiz

## [1] 0.6037853

La probabilidad de que el estudiante puerda la oportunidad de presentar el quiz agendado de 6:20 PM a 6:40 PM es de 0.60.

1. Encuentre la longitud de tiempo por arriba de la cual encontramos el 15% de los viajes más lentos.

Los viajes más lentos, son aquellos que duran más tiempo. podemos entonces calcular el area de todo excepto lo resaltado:



tLentos <- qnorm(1-0.15, mean = 24, sd = sqrt(14.44))  
tLentos

## [1] 27.93845

El 15% de los viajes más lentos, tienen un tiempo de viaje de por lo menos 27.9 minutos.

# Tercer punto