

Inferencia Estadística

Profesor(es): Jarnishs Beltran Ayudante: Pablo Rivera

> Pauta ayudantía Nº8 Primavera 2020

Ejercicio en R

i) Encuentre una estimación puntual de la edad promedio de los estudiantes universitarios con los datos de la muestra de la encuesta.

```
# Presents the MASS package data set survey
library(MASS)

# Save the survey data of student age
ageofsurvey = survey$Age

# Find the point estimate of student age
# As it turned out, not all the student have fill the age section, so we mhave to filter out the missing values
mean(ageofsurvey, na.rm = TRUE)

## [1] 20.37451
```

The result of the coding of the point estimate average above is 20.37451 years

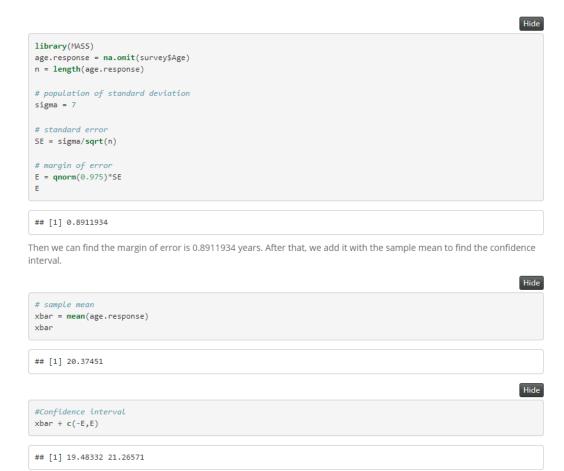
```
point.estimate <-t.test(ageofsurvey, conf.level = 0.95 )
point.estimate$conf.int

## [1] 19.54600 21.20303
## attr(,"conf.level")
## [1] 0.95</pre>
```

The confidence intervals for the average university student age is 19.546 - 21.20303. Hence, the 95% confidence level includes the true population mean which is equal to 20.37451 years.



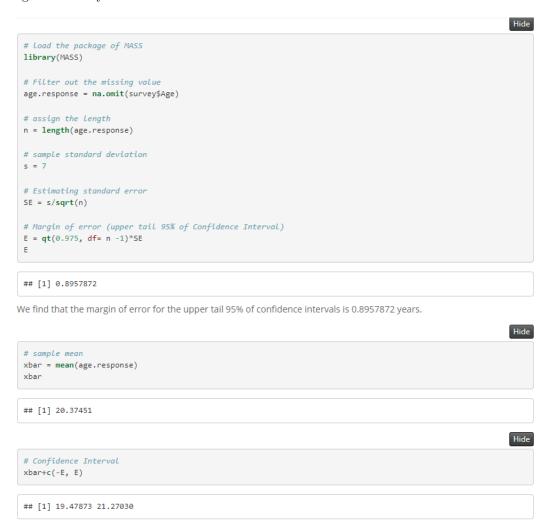
ii) Suponga que la desviación estándard de la población σ de la Edad(\mathbf{Age}) del estudiante en la encuesta de datos es 7. Encuentre el margen de error y la estimación del intervalo a un nivel de confianza del 95 %



The margin of error of the student age by assuming the population standard deviation is 7 at the 95% confidence level is 0.8911934 years. The confidence interval for this case is in between 19.48332 and 21.26571 years.



iii) Sin asumir la desviación estándard de la población σ de la Edad(\mathbf{Age}) del estudiante. Encuentre el margen de error y la estimación del intervalo a un nivel de confianza del 95 %



The result of the margin of error for the student age survey is 0.8957872 years at 95% confidence level and the confidence interval is in between 19.47873 and 21.27030 years.



iv) Mejore la calidad de una encuesta por muestreo aumentando el tamaño de la muestra con una desviación estándard desconocida σ !

```
Hide

zstar = qnorm(0.975)
(zstar^2*(0.5)* (0.5))/ (0.05)^2

## [1] 384.1459
```

So, we got 384.1459 or 384 sample sizes to improve the quality of a sample survey with unknown standard deviation σ !.

v) Suponga que ud. no tiene una estimación de proporción planificada, encuentre el tamaño de la muestra necesario para conseguir un margen de error del $5\,\%$ para las encuentas de estudiantes hombres con un nivel de confianza del $95\,\%$

First we need to find out the number of male students. We can find it using sum function and dividing it by n to find the male student proportion in this sample survey

```
zstar = qnorm(0.975)

p=0.5

#Margin or error

E = 0.05

zstar^2*p*(1-p)/E^2

## [1] 384.1459
```

The, we get that we need 384.1459 or 384 sample size to achieve 5% margin of error for the male student survey at 95% confidence level.



vi) Realice un análisis de intervalos de confianza en este data set de 2004 que incluye datos sobre ingresos por hora promedio, estado civil, sexo y edad de miles de personas.

First we read the csv of cps04 before we find the confidence interval of the average hourly earnings, marital status, gender, and age for thousands of people.

```
Hide
cps04 <-read.csv("cps04.csv", header = T, sep = ",")</pre>
# Average hourly earnings
avghour.response = na.omit(cps04$ahe)
n = length(avghour.response)
# Standard Deviation
sigma = sd(avghour.response)
# Standard error of the mean
SE = sigma /sqrt(n)
# Margin of error
E = qnorm(0.975)*SE
## [1] 0.1920964
                                                                                                              Hide
xbar <- mean(avghour.response)</pre>
xbar
## [1] 16.7712
                                                                                                              Hide
xbar +c(-E,E)
## [1] 16.57911 16.96330
```

From the code above, we can know that the margin of error of average hourly earnings is 0.1920964. xbar (sample mean) is 6.7712 while the confidence interval is inbetween 16.57911 and 16.96330.



```
Hide
# Age
age.respon = na.omit(cps04$age)
n = length(age.respon)
#Standard Deviation
sigma = sd(age.respon)
#standard error of the mean
{\sf SE=sigma/sqrt(n)}
# Margin of error
E= qnorm(0.975)*SE
## [1] 0.06340892
                                                                                                              Hide
xbar <- mean(age.respon)</pre>
xbar
## [1] 29.75445
xbar+c(-E,E)
## [1] 29.69104 29.81785
```

From the code above, we can know that the margin of error of age is 0.06340892. xbar (sample mean) is 29.75445 while the confidence interval is inbetween 29.69104 and 29.81785 years.



```
Hide
# Female
female.response = na.omit(cps04$female)
n = length(female.response)
k = sum(female.response == "1")
## [1] 3313
#Standard Deviation
sigma = sd(female.response)
#standard error of the mean
SE=sigma/sqrt(n)
# Margin of error
E= qnorm(0.975)*SE
## [1] 0.01080662
                                                                                                                    Hide
xbar <- mean(female.response)</pre>
xbar
## [1] 0.414851
                                                                                                                    Hide
xbar+c(-E,E)
## [1] 0.4040444 0.4256576
```

From the code above, we can know that the total of the female is 3313 and the margin of error of age is 0.01080662. xbar (sample mean) is 0.414851 while the confidence interval is inbetween 0.4040444 and 0.4256576. From this interval we know that, there are more male than female participants



```
# Bachelor
bachelor.response = na.omit(cps04$bachelor)
n = length(bachelor.response)
k = sum(bachelor.response == "1")
## [1] 3640
                                                                                                            Hide
#Standard Deviation
sigma = sd(bachelor.response)
#standard error of the mean
SE=sigma/sqrt(n)
# Margin of error
E= qnorm(0.975)*SE
Е
## [1] 0.01092388
                                                                                                            Hide
xbar <- mean(bachelor.response)</pre>
xbar
## [1] 0.4557976
                                                                                                            Hide
xbar+c(-E,E)
## [1] 0.4448738 0.4667215
```

From the code above, we can know that the total of the bachelor is 3460 and the margin of error of age is 0.01092388. xbar (sample mean) is 0.4557976 while the confidence interval is inbetween 0.4448738 and 0.4667215. From this interval we know that, there are more not bachelor than bachelor participants