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

Hide

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

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

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(**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 %

Hide

```
library(MASS)
age.response = na.omit(survey$Age)
n = length(age.response)

# population of standard deviation
sigma = 7

# standard error
SE = sigma/sqrt(n)

# margin of error
E = qnorm(0.975)*SE
E
```

```
## [1] 0.8911934
```

Then we can find the margin of error is 0.8911934 years. After that, we add it with the sample mean to find the confidence interval.

Hide

```
# sample mean
xbar = mean(age.response)
xbar
```

```
## [1] 20.37451
```

Hide

```
#Confidence interval
xbar + c(-E,E)
```

```
## [1] 19.48332 21.26571
```

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(**Age**) del estudiante. Encuentre el margen de error y la estimación del intervalo a un nivel de confianza del 95 %

Hide

```
# load the package of MASS
library(MASS)

# Filter out the missing value
age.response = na.omit(survey$Age)

# assign the length
n = length(age.response)

# sample standard deviation
s = 7

# Estimating standard error
SE = s/sqrt(n)

# Margin of error (upper tail 95% of Confidence Interval)
E = qt(0.975, df= n -1)*SE
E
```

```
## [1] 0.8957872
```

We find that the margin of error for the upper tail 95% of confidence intervals is 0.8957872 years.

Hide

```
# sample mean
xbar = mean(age.response)
xbar
```

```
## [1] 20.37451
```

Hide

```
# Confidence Interval
xbar+c(-E, E)
```

```
## [1] 19.47873 21.27030
```

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 estimaciòn 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

Hide

```
library(MASS)
gender.response = na.omit(survey$Sex)
n = length(gender.response)
k = sum(gender.response == "Male")
k
```

```
## [1] 118
```

Hide

```
pbar  = k/n;pbar
```

```
## [1] 0.5
```

The number of male student is 118. The proportion of the male student is 0.5.

Now, we want to find the sample size to achieve 5% margin of error for the male student `survey` at 95% confidence level

Hide

```
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 = ",")

# 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
E
```

```
## [1] 0.1920964
```

Hide

```
xbar <- mean(avghour.response)
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
SE=sigma/sqrt(n)

# Margin of error
E= qnorm(0.975)*SE
E
```

```
## [1] 0.06340892
```

Hide

```
xbar <- mean(age.respon)
xbar
```

```
## [1] 29.75445
```

Hide

```
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")
k
```

```
## [1] 3313
```

Hide

```
#Standard Deviation
sigma = sd(female.response)

#standard error of the mean
SE=sigma/sqrt(n)

# Margin of error
E= qnorm(0.975)*SE
E
```

```
## [1] 0.01080662
```

Hide

```
xbar <- mean(female.response)
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")
k
```

```
## [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
E
```

```
## [1] 0.01092388
```

Hide

```
xbar <- mean(bachelor.response)
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