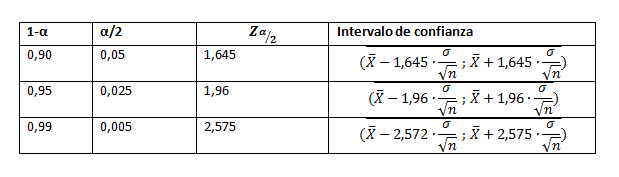
# Estadística Inferencial : Estimación de Parámetros

A tener en cuenta: Tabla de Distribución Normal Estándar - Puntuaciones Típicas Z

Cuando no se conoce la varianza poblacional, se trabaja con la distribución normal estándar (puntuaciones Z). Los intervalos de confianza más usados para la investigación son 90%, 95 % y 99 % cuyas puntuaciones Z son 1.645, 0.025 y 0.005 respectivamente.

# Intervalo de Confianza y Error para una Muestra

Para el cálculo de intervalos de confianza de una media es necesario realizar la siguiente fórmula:



* Es el nivel de confianza que vamos a calcular

Z El valor de distribución normal estándar (puntuaciones Z) que nos brindará el nivel de confianza

 Error estándar de la media y se calcula con la siguiente fórmula:



A continuación realizaremos el cálculo para el error muestral y el intervalo de confianza al 95% para la variable **P25**. Para esto, primero se debe realizar una exploración de la variable inicial.

install.packages ("foreign")

library (foreign)

data2<-read.spss("IOPGENERO.sav", to.data.frame=TRUE, use.value.labels = TRUE)

attach(data2)

install.packages("Hmisc")

library(Hmisc)

**P25.** **¿Cuál es la edad de su actual cónyuge o pareja?**

describe(P25)

P25

n missing unique Info Mean

688 515 62 1 42.85

.05 .10 .25 .50 .75 .90 .95

24 26 32 40 52 63 69

lowest : 17 19 20 21 22

highest: 75 77 78 80 83

Como se puede observar en los resultados, la variable tiene un total de 688 casos válidos y 515 valores perdidos. Esta disgregación se realiza por defecto, pero para estar más seguros de que no están ingresando datos perdidos al análisis realizaremos la siguiente función:

p25r<- P25 #Creamos un nuevo objeto p25r

p25r[P25>83] <- NA # Los datos que sean mayores a 83, valor máximo, serán considerados como NA.

describe(p25r)

p25r

n missing unique Info Mean 05 .10 .25 .50 .75 .90 .95

688 515 6 1 **42.85** 24 26 32 40 52 63 69

.

lowest : 17 19 20 21 22

highest: 75 77 78 80 83

Como podemos observar, en estos resultados no se han adicionado casos NA, pero este paso es necesario ya que nos evita trabajar con posibles valores NA. Ahora iniciaremos el cálculo del error estándar y el intervalo de confianza. En primer lugar, se pide un estadístico de centralidad (media), uno de dispersión (desviación estándar) y el tamaño de la muestra. Luego, se calcula el error estándar, según la fórmula y el valor Z. Por último, se establecen los límites inferior y superior del intervalo.

media <-mean(na.omit(p25r))

#Pedimos la media

desv <-sd(na.omit(p25r))

#La desviación estándar

N<-length(na.omit(p25r))

#El tamaño válido de la muestra

error.est<-desv/sqrt(N)

# Calculamos el error estándar. Recuerden la Fórmula:

**[1] 0.532841**

error <- 1.96\*error.est

#Establecemos la puntuación Z (1.96) para indicar un nivel de confianza de 95%

lim.inf<- media-error

# Límite inferior del intervalo

**[1] 41.80447**

lim.sup <- media+error

# Límite superior del intervalo

**[1] 43.89321**

|  |
| --- |
| **Dato extra**# *Para guardar los datos generados en un solo objeto señale*  *“resultado1 <- data.frame (media, desv, N, error.est, error, lim.inf, lim.sup)"* |

Podemos observar que el intervalo de confianza para la muestra con un intervalo de confianza de 95 % está entre {41.80447 – 43.89321}, lo que nos quiere decir que la edad de su actual cónyuge oscila entre estos límites en el 95% de la población.

Ahora calcularemos el error estándar para una media y el intervalo de confianza para la variable **P26** con un intervalo de confianza de 90 % y 99%.

# Intervalo de Confianza y Error para un parámetro

Para el cálculo de intervalos de confianza para un parámetro será necesario realizar la siguiente fórmula:



* Es el nivel de confianza que vamos a calcular; alfa es el nivel de significancia.

Z El valor de distribución normal estándar (puntuaciones Z) que nos brindará el nivel de confianza

 Error estándar de la proporción y se calcula con la siguiente fórmula:



A continuación realizaremos el cálculo para el error de una proporción y el intervalo de confianza al 95% para la variable **P34.** Primero, se debe explorar la variable inicial.

install.packages ("foreign")

library ("foreign")

data2<-read.spss("IOP-GENERO.sav", to.data.frame=TRUE, use.value.labels = TRUE)

attach (data2)

P34

install.packages("Hmisc") ##paquete necesario para utilizar el comando describe ()

library("Hmisc")

describe(P34)

**P34**

n missing unique

690 513 6

Principalmente yo (114, 17%)

Principalmente mi cÃ³nyuge /pareja (73, 11%)

Algunas veces yo, y otras mi cÃ³nyuge/pareja (141, 20%)

Lo decidimos (o decidÃ­amos) juntos (356, 52%)

Alguna otra persona (4, 1%)

No contesta (2, 0%)

Como se puede observar en los resultados, la variable tiene un total de 690 casos válidos y 513 valores perdidos. A continuación, omitiremos los valores perdidos recodificando la variable:

|  |
| --- |
| Dato# *Para ordenar los datos de frecuencia, porcentaje y porcentaje válido utilice el paquete “descr” y el comando “freq”*  library(descr)  freq(P34, plot=FALSE) |

A continuación, omitiremos los valores perdidos recodificando la variable:

p34r<-na.omit(P34)

Luego seleccionaremos la categoría de la variable con la que trabajaremos y crearemos un objeto que será llamado categoria y que tomará la categoría “Principalmente mi cónyuge/pareja” y su probalididad será comparada con las otras cinco categorías de la variable (“Principalmente yo”, “Algunas veces yo y otras mi cónyuge/pareja”, “Lo decidimos (o decidíamos) juntos”, “Alguna otra persona”, “No contesta”)

categoria<- ifelse (p34r=="Principalmente mi cÃ³nyuge /pareja", 1, 0)

Los valores 1 y 0 significan, la cantidad de personas que seleccionaron la categoría “Principalmente mi cónyuge/pareja” y las que no; por lo tanto, marcaron las otras categorías, respectivamente.

Después de observar los datos mostrados en las tablas, podemos observar que la categoría que analizaremos tiene un total de 73 personas que optaron por esta respuesta, mientras que el número total de respuestas válidas de la muestra es 690. Estos resultados se pueden observar mejor en la tabla siguiente:

prop.table(table(categoria))

categoría

0 1

0.8942029 0.1057971

Realizado todo este procedimiento, se iniciará el cálculo del error estándar y el intervalo de confianza

p<- mean(cat)

p

#Esta es la proporción de personas que respondieron esta opción

media <-mean(na.omit(p34r))

n <- length(categoria)

# Tamaño de la muestra

error.est.p<- sqrt((p\*(1-p))/n)

# Calculamos el error estándar. Recuerden la Fórmula:

# pq = la posibilidad de éxito por la posibilidad de fracaso

# q = 1 – p

# p = f1 / n

[1] 0.01170928

error.p<-1.96\*error.est.p

#Establecemos la puntuación Z (1.96) para indicar un nivel de confianza de 95%

lim.inf.p<-p-error.p

# Límite inferior del intervalo

[1] 0.0828469

lim.sup.p<-p+error.p

# Límite superior del intervalo

[1] 0.1287473

Podemos concluir con la siguiente interpretación: estamos un 95% seguros que para entre el 0,08% al 0,12 % de la población, las decisiones para el fin de semana son tomadas principalmente por el cónyuge/pareja.