

# Instrucción 2 - Proyecto Final XS1130

Diego Alberto Vega Viquez

2024-11-09

## Carga datos

Primero debemos cargar los datos

```
ACTUALIDADES2022 <- as.data.frame(  
  read_dta("https://raw.githubusercontent.com/Diego-Vega-Viquez/Proyecto_Final_XS1130/main/d
```

## Selección de variables

```
datos <- ACTUALIDADES2022 %>% select(SbjNum, Date, Duration,  
  matches("^RC[0-79]\\D?"),  
  matches("^A\\d"),  
  matches("^CS[0-79]"),  
  edad,  
  edad_imputada, edu, edu_imputada,  
  factor, etnia)
```

## Limpieza de datos

```
datos_limpios <- datos  
  
# Elimina las variables CS13  
datos_limpios <- datos_limpios %>% select(-c(CS13,CS13_NOMBRE))
```

```

# Elimina a las personas ecuestadas que no saben quien es Rodrigo Chavez
datos_limpios <- datos_limpios %>% filter(datos_limpios$RC1 != 6)

# Convierte todos los 9 en NA de la parte de Chavez (RC)
for (j in c(1:7, 9:11)) {
  variable <- paste0("RC",as.character(j))
  datos_limpios <- datos_limpios %>% mutate(across(all_of(variable),
                                                    ~ ifelse(. == 9, NA, .)))
}

# Convierte todos los 9 y 999 en NA de la parte de Sociodemográficas (CS)
for (j in c(1:7, 9:10)) {
  variable <- paste0("CS",as.character(j))
  datos_limpios <- datos_limpios %>% mutate(across(all_of(variable),
                                                    ~ ifelse(. == 9 | . == 999,
                                                            NA, .)))
}

```

Lo que se hizo anteriormente corresponde a la asignación de NA a los valores correspondientes en la encuesta y se eliminan algunas unidades estadísticas que son las que no participaron en el módulo con el que se está trabajando en el presente trabajo

## Formateo de variables

```

datos_formateados <- datos_limpios
datos_formateados$Date <- as.Date(datos_formateados$Date)
datos_formateados$Duration <- as.difftime(datos_formateados$Duration)
# Datos de Chavez

for (j in c(1:7, 9:11)) {
  variable <- paste0("RC", as.character(j))
  datos_formateados[[variable]] <- as.factor(datos_formateados[[variable]])
}

# Datos Sociodemográficos

datos_formateados$CS1 <- as.factor(datos_formateados$CS1)
datos_formateados$CS2 <- as.numeric(datos_formateados$CS2)
datos_formateados$CS3 <- as.factor(datos_formateados$CS3)

```

```

datos_formateados$CS4 <- as.factor(datos_formateados$CS4)
datos_formateados$CS5 <- as.factor(datos_formateados$CS5)
datos_formateados$CS6 <- as.factor(datos_formateados$CS6)
datos_formateados$CS7 <- as.factor(datos_formateados$CS7)
datos_formateados$CS9 <- as.factor(datos_formateados$CS9)
datos_formateados$CS10 <- as.numeric(datos_formateados$CS10)
datos_formateados$CS11A <- as.numeric(datos_formateados$CS11A)
datos_formateados <- datos_formateados %>%
  mutate(CS11A = ifelse(CS11A == 99, NA, CS11A))
datos_formateados$CS11B <- as.numeric(datos_formateados$CS11B)
datos_formateados <- datos_formateados %>%
  mutate(CS11B = ifelse(CS11B == 99, NA, CS11B))
datos_formateados$CS11C <- as.numeric(datos_formateados$CS11C)
datos_formateados <- datos_formateados %>%
  mutate(CS11C = ifelse(CS11C == 99, NA, CS11C))
datos_formateados$CS12P <- as.factor(datos_formateados$CS12P)
datos_formateados <- datos_formateados %>%
  mutate(CS12P = na_if(CS12P, "9. NS/NR")) %>% droplevels()
datos_formateados$CS12C <- as.factor(datos_formateados$CS12C)
datos_formateados <- datos_formateados %>%
  mutate(CS12C = na_if(CS12C, "9. NS/NR")) %>%
  mutate(CS12C = na_if(CS12C, "9.NS/NR")) %>% droplevels()
datos_formateados$CS12D <- as.factor(datos_formateados$CS12D)
datos_formateados <- datos_formateados %>%
  mutate(CS12D = na_if(CS12D, "9. NS/NR")) %>%
  mutate(CS12D = na_if(CS12D, "9.NS/NR")) %>%
  mutate(CS12D = na_if(CS12D, "-1")) %>% droplevels()

# La CS13 no tiene nada que ver con el presente trabajo

datos_formateados$A1 <- as.numeric(datos_formateados$A1)
datos_formateados$A3 <- as.numeric(datos_formateados$A3)
datos_formateados$A6 <- as.numeric(datos_formateados$A6)
datos_formateados$A7 <- as.factor(datos_formateados$A7)
datos_formateados$edad <- as.factor(datos_formateados$edad)
datos_formateados$edad_imputada <- as.factor(datos_formateados$edad_imputada)
datos_formateados$edu <- as.factor(datos_formateados$edu)
datos_formateados$edu_imputada <- as.factor(datos_formateados$edu_imputada)
datos_formateados$factor <- as.numeric(datos_formateados$factor)
datos_formateados$etnia <- as.factor(datos_formateados$etnia)

```

## Selección de la Muestra

```
set.seed(123)
muestra <- datos_formateados[sample(500, replace = F),]
```

Para el presente trabajo se realizará la semilla `set.seed(123)`, la versión de R que se utilizará corresponde a

```
[1] "R version 4.3.3 (2024-02-29)"
```

## Análisis Descriptivo de los Datos

```
# Muestra cómo se ven los datos

glimpse(muestra)

# Muestra cuantas variables hay de cada tipo de dato (cuali & cuanti)
tipos <- c()
for (word in colnames(muestra)) {
  tipos <- c(tipos, class(muestra[[word]]))
}
summary(as.factor(tipos))
```

## Estimaciones puntuales

### Proporciones

Recuerde que el error estándar para proporciones viene dado por

$$\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Nos basaremos en la pregunta

*En general, ¿diría usted que está haciendo un trabajo muy malo, malo, bueno o muy bueno?*

Cual tiene código RC1, esta variable corresponde a una cualitativa ordinal en con 5 niveles, se utilizarán todas las respuestas que puedan se consideradas como “ni malo ni regular”, por tanto tomaremos las respuestas “BUENO” y “MUY BUENO”.

### Proporción de hombres que califican positivamente Adm. Chaves Robles

Buscaremos estimar la proporción poblacional de hombres que califican positivamente la administración Chaves Robles ( $p_1$ ) con un nivel de confianza del 98%, de la siguiente forma:

```
# Número de hombres que califican positivamente la adm. Chaves Robles muestral
n1 <- muestra %>% filter(as.numeric(RC1)>3 & CS1 == 1) %>% nrow()
# Tamaño de la población
N1 <- muestra %>% nrow()

# Proporción muestral
p1 <- n1 / N1
# Complemento muestral
q1 <- 1-p1

# Nivel de confianza
confianza1 <- 98/100
# Nivel de significancia (alpha)
alpha1 <- 1-confianza1
# Valor crítico / estadístico Z
z1 <- abs(ifelse(n1>=30,qnorm(alpha1/2),qt(alpha1/2)))
# Error estandar
error1 <- sqrt(p1*q1 / n1)

# Intervalo de Confianza (IC)
lim_inf1 <- p1 - z1*error1
lim_sup1 <- p1 + z1*error1

round(lim_inf1*100,2) %>% paste0("%") %>% print();
```

```
[1] "30.02%"
```

```
round(lim_sup1*100,2) %>% paste0("%") %>% print()
```

```
[1] "46.38%"
```

De esta forma se tiene que

$$30.02\% \leq p_1 \leq 46.38\%$$

### Proporción de mujeres que califican positivamente Adm. Chaves Robles