PASPE 2024: R básico

Tarea 2

Rossana Torres Alvarez

July 23, 2024

## Problema 1. Explorar la base Antropometria

1. Muestra el numero de columnas
2. Muestra las primeras 15 observaciones
3. Estima los cuartiles de talla
4. Estima la talla minima y la talla maxima

# Llamo a la libreria haven porque se que Antropometria tiene un formato dta  
library(haven)  
  
# Uso la funcion read\_dta y pego la direccion local donde tengo la base  
Antropometria <- read\_dta("~/Documents/GitHub/Rbasico/files/Antropometria.dta")  
  
# NOTA: para deshacer vectores del tipo haven\_labelled  
# aplico la funcion as.numeric  
# ESTO SOLO ES RECOMENDABLE EN VARIABLES QUE  
# VERDADERAMENTE SEAN NUMERICAS COMO EN ESTE CASO  
  
Antropometria$talla\_unlabelled <- as.numeric(Antropometria$talla)  
  
# Uso la funcion ncol para observar el numero de columnas  
ncol(Antropometria)

## [1] 65

# Uso la funcion head para enseñar las primeras 15 observaciones  
head(Antropometria, 15)

## # A tibble: 15 × 65  
## folio intp entidad desc\_ent munici desc\_mun locali desc\_loc manzana ageb   
## <chr> <dbl> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>  
## 1 010001 1 01 01 AGUASC… 001 001 AGU… 0001 0001 AG… "" 0322   
## 2 010001 3 01 01 AGUASC… 001 001 AGU… 0001 0001 AG… "" 0322   
## 3 010001 5 01 01 AGUASC… 001 001 AGU… 0001 0001 AG… "" 0322   
## 4 010008 1 01 01 AGUASC… 001 001 AGU… 0001 0001 AG… "" 0322   
## 5 010008 3 01 01 AGUASC… 001 001 AGU… 0001 0001 AG… "" 0322   
## 6 010008 4 01 01 AGUASC… 001 001 AGU… 0001 0001 AG… "" 0322   
## 7 010008 5 01 01 AGUASC… 001 001 AGU… 0001 0001 AG… "" 0322   
## 8 010009 1 01 01 AGUASC… 001 001 AGU… 0001 0001 AG… "" 0322   
## 9 010011 2 01 01 AGUASC… 001 001 AGU… 0001 0001 AG… "" 0322   
## 10 010011 4 01 01 AGUASC… 001 001 AGU… 0001 0001 AG… "" 0322   
## 11 010011 5 01 01 AGUASC… 001 001 AGU… 0001 0001 AG… "" 0322   
## 12 010011 6 01 01 AGUASC… 001 001 AGU… 0001 0001 AG… "" 0322   
## 13 010015 1 01 01 AGUASC… 001 001 AGU… 0001 0001 AG… "" 0322   
## 14 010015 5 01 01 AGUASC… 001 001 AGU… 0001 0001 AG… "" 0322   
## 15 010015 6 01 01 AGUASC… 001 001 AGU… 0001 0001 AG… "" 0322   
## # ℹ 55 more variables: maq <chr>, sexo <dbl>, edad\_i <dbl>, edad <dbl>,  
## # meses <dbl>, intsel <dbl>, intsel2 <dbl>, peso <dbl+lbl>, peso2 <dbl+lbl>,  
## # ropa <dbl+lbl>, rpeso <dbl+lbl>, talla <dbl+lbl>, talla2 <dbl+lbl>,  
## # rtalla <dbl+lbl>, emb <dbl+lbl>, temb <dbl+lbl>, cintura <dbl+lbl>,  
## # cintura2 <dbl+lbl>, rcintura <dbl+lbl>, cadera <dbl+lbl>,  
## # cadera2 <dbl+lbl>, rcadera <dbl+lbl>, ppeso <dbl+lbl>, amputa <dbl+lbl>,  
## # time1 <chr>, time2 <chr>, panto <dbl+lbl>, rpanto <dbl+lbl>, …

# Opcion 1 para estimar cuantiles   
# Uso el operador $ para seleccionar la columna talla y despues aplico  
# la funcion quantile. NOTA utilizo la opcion na.rm = TRUE en caso de que tenga observaciones con NA  
  
quantile(Antropometria$talla\_unlabelled, na.rm = TRUE)

## 0% 25% 50% 75% 100%   
## 46.0 131.0 152.2 162.9 222.2

# Opcion 1 para estimar el maximo y el minimo:  
#Uso la funcion max y min para estimar el maximo y minimo de talla  
max(Antropometria$talla\_unlabelled, na.rm = TRUE)

## [1] 222.2

min(Antropometria$talla\_unlabelled, na.rm = TRUE)

## [1] 46

# Opcion 2 uso la funcion summary para obtener un resumen de las estadisticas  
summary(Antropometria$talla)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 46.0 131.0 152.2 146.7 162.9 222.2 544

# Mismo resultado si uso mi nueva variable sin etiquetas  
summary(Antropometria$talla\_unlabelled)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 46.0 131.0 152.2 146.7 162.9 222.2 544

## Problema 2. Manejo de datos

1. Crea una nueva variable llamada id, que junte las variables folio e intp
2. Usa la funcion unique() para contar el numero de ids unicos.
3. Selecciona las variables sexo, edad, peso y talla
4. Usando base R
5. Usando dplyr
6. Crea una variable llamada imc y estima el indice de masa corporal (kg/(mt)^2)
7. Usando base R
8. Usando dplyr
9. Crea una variable con las categorias de imc de la OMS. Apoyate utilizando los criterios usados en la tarea 1.
10. Usando base R
11. Usando dplyr
12. Guarda un dataframe solo para mujeres y uno solo para hombres, llamalos mujeres\_df y hombres\_df.

# 1. Uso el operador $ y la funcion paste0 para juntar las variables folio e intp  
Antropometria$id <- paste0(Antropometria$folio, Antropometria$intp)  
  
# 2. Uso la funcion unique() y length() para contar el numero de ids unicos.  
length(unique(Antropometria$id))

## [1] 85291

# 3. Selecciono las variables sexo, edad, peso y talla. NOTA: Usare la funcion head para que la consola solo imprima las primeras observaciones como demostracion de la respuesta  
  
# a) Usando base R  
head(Antropometria[, c("sexo", "edad", "peso", "talla")], 3)

## # A tibble: 3 × 4  
## sexo edad peso talla   
## <dbl> <dbl> <dbl+lbl> <dbl+lbl>  
## 1 1 42 75 166.   
## 2 1 19 64 170.   
## 3 1 9 28.4 134.

# b) Usando dplyr  
# LLAMO A LA LIBRERIA TIDYVERSE AL INICIO DE MI SCRIPT  
library(tidyverse)  
head(select(Antropometria, "sexo", "edad", "peso", "talla"), 3)

## # A tibble: 3 × 4  
## sexo edad peso talla   
## <dbl> <dbl> <dbl+lbl> <dbl+lbl>  
## 1 1 42 75 166.   
## 2 1 19 64 170.   
## 3 1 9 28.4 134.

# 4. Creo una variable llamada imc y estima el indice de masa corporal  
# (kg/(mt)^2). NOTA: Usare la funcion head para que la consola solo imprima las primeras observaciones como demostracion de la respuesta  
  
# a) Usando base R  
Antropometria$imc <- Antropometria$peso/(Antropometria$talla/100)^2  
head(Antropometria$imc,3)

## [1] 27.05408 22.24991 15.91129

# b) Usando dplyr  
  
# Sin pipes  
Antropometria <- mutate(Antropometria, imc = peso/(talla/100)^2)  
  
# Con pipes (equivalente a lo de arriba)  
  
Antropometria <- Antropometria %>%   
 mutate( imc = peso/(talla/100)^2)  
  
head(Antropometria$imc,3)

## [1] 27.05408 22.24991 15.91129

# 5. Creo una variable con las categorias de imc de la OMS utilizando los criterios usados en la tarea 1.  
  
# a) Usando base R  
Antropometria$imc\_cat <- rep(NA, nrow(Antropometria))  
Antropometria$imc\_cat[Antropometria$imc < 18] <- "Bajo peso"  
Antropometria$imc\_cat[Antropometria$imc >= 18 & Antropometria$imc <= 24.9] <- "Peso normal"  
Antropometria$imc\_cat[Antropometria$imc > 24.9 & Antropometria$imc <= 29.9] <- "Sobrepeso"  
Antropometria$imc\_cat[Antropometria$imc > 29.9 & Antropometria$imc <= 34.5] <- "Obesidad grado I"  
Antropometria$imc\_cat[Antropometria$imc > 34.5 & Antropometria$imc <= 39.9] <- "Obesidad grado II"  
Antropometria$imc\_cat[Antropometria$imc > 39.9] <- "Obesidad grado III"  
  
# Verifico usando funcion table y la opcion exclude = NULL para ver observaciones que no cayeron en ninguna categoria (es decir NAs)  
  
table(Antropometria$imc\_cat, exclude = NULL)

##   
## Bajo peso Obesidad grado I Obesidad grado II Obesidad grado III   
## 21523 9509 3755 6245   
## Peso normal Sobrepeso <NA>   
## 25529 18186 544

# b) Usando dplyr  
  
Antropometria <- Antropometria %>%   
 mutate(imc\_cat = case\_when(imc < 18 ~ "Bajo peso",  
 imc >= 18 & imc <= 24.9 ~ "Peso normal",  
 imc> 24.9 & imc <= 29.9 ~ "Sobrepeso",  
 imc > 29.9 & imc <= 34.5 ~ "Obesidad grado I",  
 imc > 34.5 & imc <= 39.9 ~ "Obesidad grado II",  
 imc > 39.9 ~ "Obesidad grado III"))  
  
# Verifico  
  
table(Antropometria$imc\_cat, exclude = NULL)

##   
## Bajo peso Obesidad grado I Obesidad grado II Obesidad grado III   
## 21523 9509 3755 6245   
## Peso normal Sobrepeso <NA>   
## 25529 18186 544

# Extra, ver quienes no tienen categoria  
  
sinimc <- filter(Antropometria, is.na(imc))  
  
sinimc[, c("peso","talla","imc", "imc\_cat")]

## # A tibble: 544 × 4  
## peso talla imc imc\_cat  
## <dbl+lbl> <dbl+lbl> <dbl> <chr>   
## 1 NA NA NA <NA>   
## 2 NA NA NA <NA>   
## 3 NA NA NA <NA>   
## 4 NA NA NA <NA>   
## 5 NA NA NA <NA>   
## 6 NA NA NA <NA>   
## 7 NA NA NA <NA>   
## 8 NA NA NA <NA>   
## 9 NA NA NA <NA>   
## 10 NA NA NA <NA>   
## # ℹ 534 more rows

# Concluyo que son los individuos sin datos completos  
  
# 6. Guarda un dataframe solo para mujeres y uno solo para hombres, llamalos mujeres\_df y hombres\_df.  
  
mujeres\_df <- filter(Antropometria, sexo == 2)  
  
# Verifico  
table(mujeres\_df$sexo, exclude = NULL)

##   
## 2   
## 45057

hombres\_df <- filter(Antropometria, sexo == 1)  
  
# Verifico  
table(hombres\_df$sexo, exclude = NULL)

##   
## 1   
## 40234