Práctica Dirigida 5

Mg. Sc. J. Eduardo Gamboa U.

Los archivos que se utilizarán en esta práctica dirigida corresponden a la Encuesta Nacional de Hogares (ENAHO), llevada a cabo por el INEI de manera trimestral desde hace más de dos décadas. Esta encuesta está dividida por módulos, cada uno de los cuales aborda una temática en particular.

El archivo **PD5 - datos.xlsx** contiene datos acerca del módulo 1 - cuarto trimestre del 2024, correspondientes solo a hogares de la selva.

Las variables registradas son las siguientes:

- AÑO: Año en el que se realizó la encuesta (2024)
- MES: Mes en el que se realizó la encuesta (10, 11, 12)
- CONGLOME: Conglomerado donde se encuentra la vivienda
- VIVIENDA: Número de vivienda en el conglomerado
- HOGAR: Número de hogar en la vivienda
- DOMINIO: Dominio en el que se encuentra la vivienda (7 = Selva)
- RESULT: Resultado de la encuesta

  1 = Completa

  2 = Incompleta

  3 = Rechazo

  4 = Ausente

  5 = Vivienda desocupada

  6 = No se inició la entrevista

  7 = Otro

- P101: Tipo de vivienda

  1 = Casa independiente

  2 = Departamento en edificio

  3 = Vivienda en quinta

  4 = Vivienda en casa de vecindad(callejón solar o corralón)

  5 = Choza o cabaña)
- P102: Material predominante en las paredes exteriores

  1 = Ladrillo o bloque de cemento

  2 = Piedra o sillar con cal o cemento

  3 = Adobe

  4 = Tapia

  5 = Quincha (caña con barro)

  6 = Piedra con barro

  7 = Madera (pona, tornillo, etc)

  8 = Triplay/calamina/estera

  9 = Otro material
- P103: Material predominante en los pisos

  1 = Parquet o madera pulida

  2 = Láminas asfálticas, vinílicos o similares

  3 = Losetas, terrazos o similares

  4 = Madera (pona, tornillo, etc)

  5 = Cemento

  6 = Tierra

  7 = Otro material

- P103A: Material predominante en los techos

  1 = Concreto armado

  2 = Madera

  3 = Tejas

  4 = Planchas de calamina, fibra de cemento o similares

  5 = Caña o estera con torta de barro o cemento

  6 = Triplay/estera/carrizo

  7 = Paja, hojas de palmera

  8 = Otro material
- P104: Número de habitaciones de la vivienda, sin contar baño ni cocina
- P104A: Número de habitaciones usadas exclusivamente para dormir

Pregunta 1

1. Leer el archivo PD5 - datos.xlsx.

```
library(readxl)
datos1 = read_excel('PD5 - datos.xlsx')
```

2. Ejecutar las siguientes acciones, concatenándolas con pipe:

a. Renombrar la columna RESULT por ENCUESTA

b. Filtrar los registros que correspondan a encuestas completas e incompletas y que la cantidad de habitaciones en la vivienda (sin contar baño ni cocina) sea mayor a 3 o esté perdida.

c. Crear la columna CODIGOH pegando CONGLOME, VIVIENDA y HOGAR.

d. Retirar las columnas CONGLOME, VIVIENDA y HOGAR.

e. Almacenar el resultado en `datos1_ok`.

```
library(dplyr)
datos1 |>
  rename(ENCUESTA = "RESULT") |>
  filter(ENCUESTA %in% 1:2 & (P104 > 3 | is.na(P104))) |>
  mutate(CODIGOH = paste0(CONGLOME,VIVIENDA,HOGAR)) |>
  select(-CONGLOME, -VIVIENDA, -HOGAR) -> datos1_ok
```

3. Aplicar la función `is.na` a las 6 primeras filas de datos1_ok. Explicar el resultado

```
datos1_ok |> head() |> is.na()
```

```
       AÑO   MES DOMINIO ENCUESTA  P101  P102  P103 P103A  P104 P104A CODIGOH
[1,] FALSE FALSE   FALSE    FALSE FALSE FALSE FALSE FALSE FALSE FALSE   FALSE
[2,] FALSE FALSE   FALSE    FALSE FALSE FALSE FALSE FALSE FALSE FALSE   FALSE
[3,] FALSE FALSE   FALSE    FALSE FALSE FALSE FALSE FALSE FALSE FALSE   FALSE
[4,] FALSE FALSE   FALSE    FALSE FALSE FALSE FALSE FALSE FALSE FALSE   FALSE
[5,] FALSE FALSE   FALSE    FALSE FALSE FALSE FALSE FALSE FALSE FALSE   FALSE
[6,] FALSE FALSE   FALSE    FALSE FALSE FALSE FALSE FALSE FALSE FALSE   FALSE
```

4. Aplicar la función `na.omit()` al data frame `datos1_ok`. ¿Cuántas filas contiene datos1_ok antes y después de aplicar na.omit()?

```
datos1_ok |> nrow()
```

```
[1] 653
```

```
datos1_ok |> na.omit() |> nrow()
```

```
[1] 633
```

5. Aplicar la función `complete.cases` al data frame `datos1_ok`. Explicar el resultado

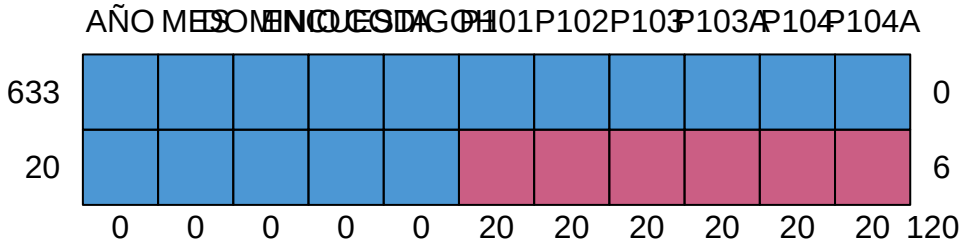
```
datos1_ok |> complete.cases()
```

```
  [1]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [13]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [25]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [37]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [49]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [61]  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
 [73]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [85]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
 [97]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[109]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[121]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[133]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
[145]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
```

```
[157]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[169]  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[181]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE
[193]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE
[205]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[217]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[229]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[241]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[253]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[265]  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[277]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[289]  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
[301] FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[313]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[325]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[337]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[349]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[361]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[373]  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE
[385]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[397]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE
[409]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[421]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[433]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[445]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[457]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[469]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[481]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[493]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[505]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE
[517]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[529]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[541]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[553]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[565]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[577]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
[589]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[601]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[613]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
[625]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[637]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[649]  TRUE  TRUE  TRUE  TRUE  TRUE
```

6. Aplicar la función `mice` al data frame `datos1_ok`. Explicar el resultado

```
library(mice)
datos1_ok |> md.pattern()
```

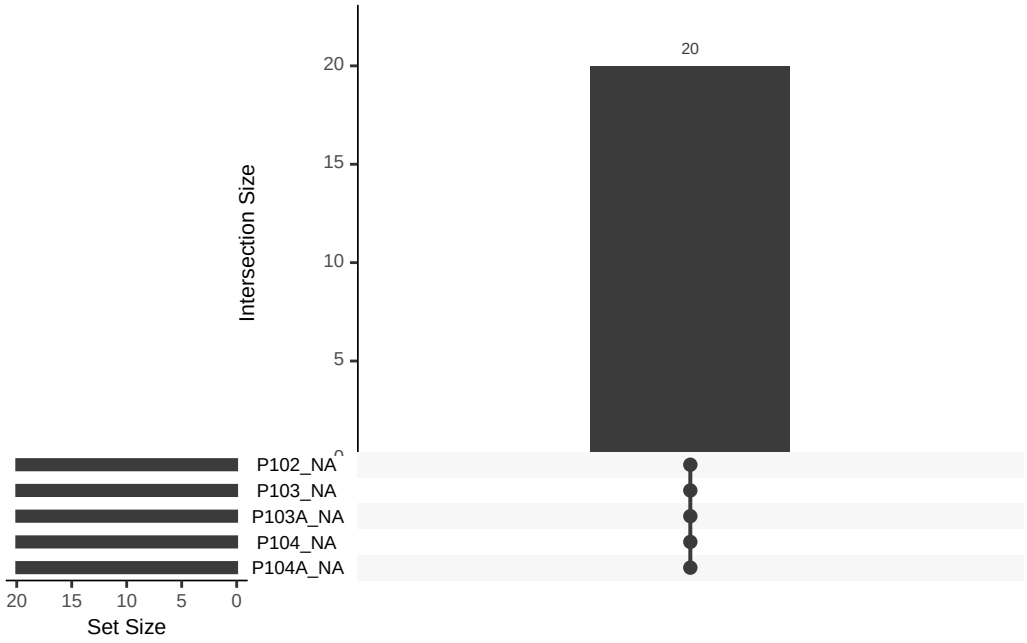



```
    AÑO MES DOMINIO ENCUESTA CODIGOH P101 P102 P103 P103A P104 P104A
633   1   1       1        1       1    1    1    1     1    1     1   0
20    1   1       1        1       1    0    0    0     0    0     0   6
      0   0       0        0       0   20   20   20    20   20    20 120
```

7

. Aplicar la función `gg_miss_upset` al data frame `datos1_ok`. Explicar el resultado

```
library(naniar)
datos1_ok |> gg_miss_upset()
```



8. Extraer la columa MES usando la función `select` y almacenar el resultado en `mes1`. Realizar lo mismo, pero usando `pull` y almacenar en `mes2` ¿Qué diferencia existe entre `mes1` y `mes2`?

```
datos1_ok |> select(MES) -> mes1
datos1_ok |> pull(MES) -> mes2
```

9. Almacenar el data frame `datos1_ok` en el archivo **datos_ej1.xlsx**

```
  library(writexl)
datos1_ok |> write_xlsx("datos_ej1.xlsx")
```

10. A partir de `datos1_ok` realizar las siguientes acciones, concatenándolas con pipe:

a. Crear una nueva variable que se llame `Paredes`, la cual corresponderá a "Ladrillo o cemento" si P102 = 1, "Adobe" si P102 = 3, "Madera" si P102 = 7, y "Otros" para los demás valores de P102, considerando que los NA deben mantenerse como NA.

b. Crear una nueva variable P104B, la cual resultará de la diferencia de P104 - P104A.

c. Seleccionar los registros que correspondan a paredes de ladrillo o cemento o adobe, y que el material predominante en los pisos sea tierra.

d. Retirar las columnas P102 y P103.

e. Ordenar de mayor a menor según la variable "Número de habitaciones usadas exclusivamente para dormir".

f. Almacenar el data frame resultante en `datos2_ok`.

```
datos1_ok |>
  mutate(Paredes = case_when(P102 == 1 ~ "Ladrillo o cemento",
                             P102 == 3 ~ "Adobe",
                             P102 == 7 ~ "Madera",
                             is.na(P102) ~ NA,
                             P102 %in% c(2,4,5,6,8,9) ~ "Otros"),
         P104B   = P104 - P104A) |>
  filter(Paredes %in% c("Ladrillo o cemento", "Adobe") & P103 == 5) |>
  select(-P102,-P103) |>
  arrange(-P104A) -> datos2_ok
```

11. Almacenar el data frame `datos2_ok` en el archivo **datos_ej2.csv**

```
datos2_ok |> write.csv('datos_ej2.csv',row.names = FALSE)
```