Postwork sesión 8

Un centro de salud nutricional está interesado en analizar estadística y probabilísticamente los patrones de gasto en alimentos saludables y no saludables en los hogares mexicanos con base en su nivel socioeconómico, en si el hogar tiene recursos financieros extra al ingreso y en si presenta o no inseguridad alimentaria. Además, está interesado en un modelo que le permita identificar los determinantes socioeconómicos de la inseguridad alimentaria.

La base de datos es un extracto de la Encuesta Nacional de Salud y Nutrición (2012) levantada por el Instituto Nacional de Salud Pública en México. La mayoría de las personas afirman que los hogares con menor nivel socioeconómico tienden a gastar más en productos no saludables que las personas con mayores niveles socioeconómicos y que esto, entre otros determinantes, lleva a que un hogar presente cierta inseguridad alimentaria.

La base de datos contiene las siguientes variables:

- nse5f (nivel socieconómico del hogar): 1 "Bajo", 2 "Medio bajo", 3 "Medio", 4 "Medio alto", 5 "Alto"
- area (zona geográfica): 0 "Zona urbana", 1 "Zona rural"
- numpeho (número de personas en el hogar)
- refin (recursos financieros distintos al ingreso laboral): 0 "no", 1 "sí"
- edadjef (edad del jefe/a de familia)
- sexoje (sexo del jefe/a de familia): 0 "Hombre", 1 "Mujer"
- añosedu (años de educación del jefe de familia)
- ln_als (logarítmo natural del gasto en alimentos saludables)
- ln_alns (logarítmo natural del gasto en alimentos no saludables)
- IA (inseguridad alimentaria en el hogar): 0 "No presenta IA", 1 "Presenta IA"

df <- read.csv("https://raw.githubusercontent.com/beduExpert/Programacion-R-Santander-2022/main/Sesion-

- 1) Plantea el problema del caso
- 2) Realiza un análisis descriptivo de la información
- 3) Calcula probabilidades que nos permitan entender el problema en México
- 4) Plantea hipótesis estadísticas y concluye sobre ellas para entender el problema en México
- 5) Estima un modelo de regresión, lineal o logístico, para identificiar los determinantes de la inseguridad alimentaria en México
- 6) Escribe tu análisis en un archivo README.md y tu código en un script de R y publica ambos en un repositorio de Github.

NOTA: Todo tu planteamiento deberá estár correctamente desarrollado y deberás analizar e interpretar todos tus resultados para poder dar una conclusión final al problema planteado.

1) Planteamiento del problema del caso

Objetivos

- Analizar patrones de gasto en alimentos saludables y no saludables en familias mexicanas con base en:
 - el nivel socioeconómico
 - disponibilidad de recursos financieros adicionales al ingreso
 - presencia de inseguridad alimentaria
- Elaborar un modelo que prediga la posibilidad de presentar inseguridad alimentaria (IA) con base en las variables disponibles en el extracto de la Encuesta Nacional de Salud y Nutrición 2012 (ENSANUT

2) Análisis descriptivo de la información

Obtención de los datos:

```
url <- "https://raw.githubusercontent.com/beduExpert/Programacion-R-Santander-2022/main/Sesion-08/Postw
datos <- read.csv(url, encoding = "UTF-8")
head(datos)</pre>
```

```
##
     nse5f area numpeho refin edadjef sexojef añosedu IA
                                                            ln_als ln_alns
## 1
         5
              0
                      4
                            0
                                   43
                                            0
                                                    24 0 5.393628
                      5
                                                    24 0 7.024649
## 2
         5
              0
                            1
                                   NA
                                            NA
## 3
         5
              0
                      4
                            0
                                   46
                                             0
                                                    24 0 6.767343 4.605170
                                                    24 0 3.401197 4.094345
## 4
         5
             1
                      1
                            0
                                   54
                                             0
## 5
         5
                      2
                                   39
                                                    24 0 6.115892 5.480639
              0
                            1
                                             0
## 6
         5
              0
                      5
                            1
                                    NA
                                            NA
                                                    24 0 7.514800 5.598422
```

Para obtener un resumen estadístico de los datos, se convierten en factores las variables discretas.

```
datos$nse5f <- factor(datos$nse5f,</pre>
                       levels = 1:5,
                       labels = c("Bajo", "Medio bajo", "Medio", "Medio alto", "Alto"),
                       ordered = TRUE)
datos$area <- factor(datos$area,
                      levels = 0:1,
                      labels = c("Zona urbana", "Zona rural"))
datos$refin <- factor(datos$refin,</pre>
                       levels = 0:1.
                       labels = c("no", "sí"))
datos$sexojef <- factor(datos$sexojef,</pre>
                        levels = 0:1,
                        labels = c("Hombre", "Mujer"))
datos$IA <- factor(datos$IA,
                    levels = 0:1,
                    labels = c("No presenta IA", "Presenta IA"))
```

Y, enseguida, el resumen estadístico.

```
nrow(datos)
```

```
## [1] 40809
```

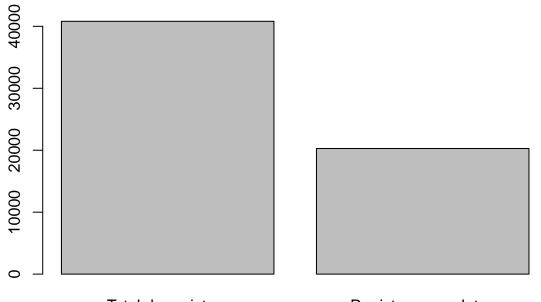
summary(datos)

```
##
          nse5f
                              area
                                            numpeho
                                                          refin
                     Zona urbana:26591
                                         Min. : 1.000
                                                          no:33046
  Bajo
             :8858
## Medio bajo:8560
                     Zona rural :14218
                                         1st Qu.: 3.000
                                                          sí: 7763
                                         Median: 4.000
              :8323
## Medio alto:7903
                                         Mean
                                               : 3.941
                                         3rd Qu.: 5.000
##
  Alto
             :7165
##
                                         Max.
                                                :19.000
##
##
      edadjef
                    sexojef
                                    añosedu
                                                              ΙA
                                                 No presenta IA:10781
##
  Min. : 18
                  Hombre:26957
                                 Min. : 0.00
##
   1st Qu.: 37
                  Mujer: 8861
                                 1st Qu.: 9.00
                                                 Presenta IA
                                                               :30028
## Median: 47
                  NA's : 4991
                                 Median: 9.00
## Mean : 49
                                       :10.36
                                 Mean
```

```
##
    3rd Qu.: 60
                                     3rd Qu.:12.00
##
    Max.
            :111
                                     Max.
                                            :24.00
            :5017
##
    NA's
##
        ln_als
                          ln_alns
##
    Min.
            :0.6931
                      Min.
                              :0.000
##
    1st Qu.:5.7038
                      1st Qu.:3.401
##
    Median :6.1633
                      Median :4.025
            :6.0665
                              :4.125
##
    Mean
                      Mean
##
    3rd Qu.:6.5511
                      3rd Qu.:4.868
##
    Max.
            :8.9699
                      Max.
                              :8.403
##
    NA's
            :787
                      NA's
                              :17504
sum(complete.cases(datos))
```

[1] 20280

La base de datos tiene 40,809 registros, de los cuales 20,280, poco menos de la mitad, tienen información completa para todos los campos.

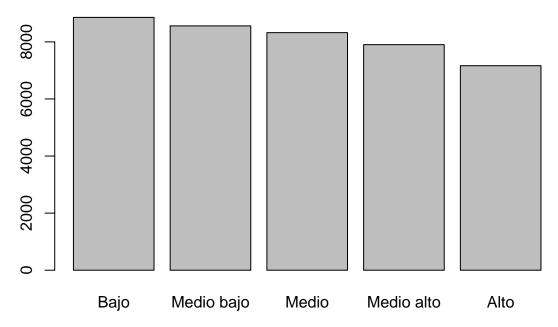


Total de registros

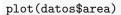
Registros completos

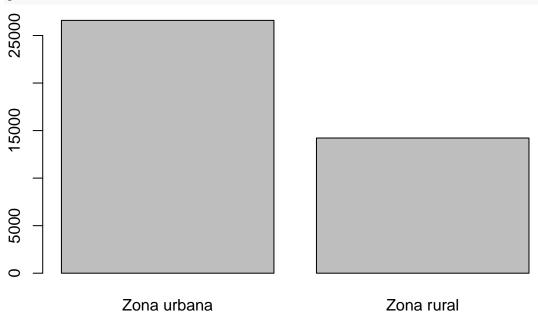
Con una mezcla más o menos homogénea de niveles socioeconómicos, con una tendencia ligeramente decreciente conforme aumenta el nivel socioeconómico, con entre 7,000 y 9,000 hogares por nivel.

plot(datos\$nse5f)



La muestra es mayoritariamente de hogares urbanos, aproximadamente en dos terceras partes.

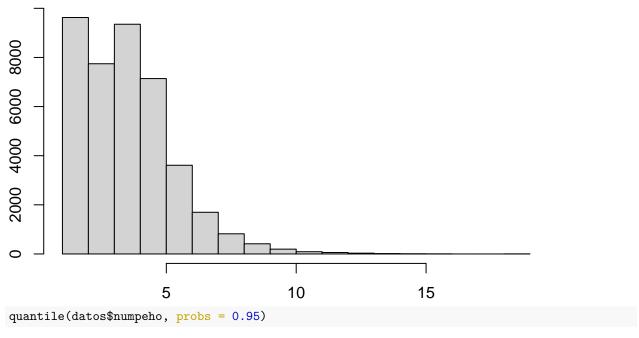




El número de personas por hogar varía entre 1 y 19, con una media de 3.9 y una mediana de 4. El 75% de los hogares es habitado por 5 personas o menos y el 50% por entre 3 y 5 personas. Asimismo, el 95% de los hogares tiene 7 o menos habitantes, siendo la excepción los valores mayores.

```
hist(datos$numpeho,
    main = "Número de personas en el hogar",
    xlab = "",
ylab = "")
```

Número de personas en el hogar



95%

7