

## 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 socioeconómico 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** (logaritmo natural del gasto en alimentos saludables)
- **ln\_alns** (logaritmo 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-8")
```

- 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 identificar 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 estar 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)



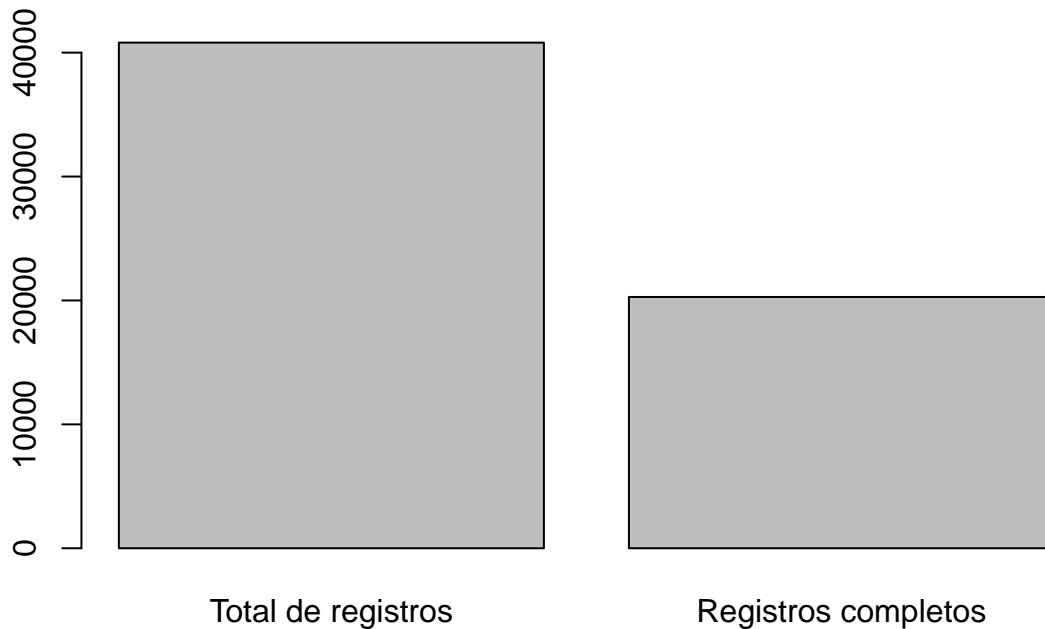
```
## 3rd Qu.: 60          3rd Qu.:12.00
## Max.    :111        Max.    :24.00
## NA's    :5017
##      ln_als      ln_alns
## Min.    :0.6931   Min.    :0.000
## 1st Qu.:5.7038   1st Qu.:3.401
## Median :6.1633   Median :4.025
## Mean    :6.0665   Mean    :4.125
## 3rd Qu.:6.5511   3rd Qu.:4.868
##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

```
sum(complete.cases(datos))
```

```
## [1] 20280
```

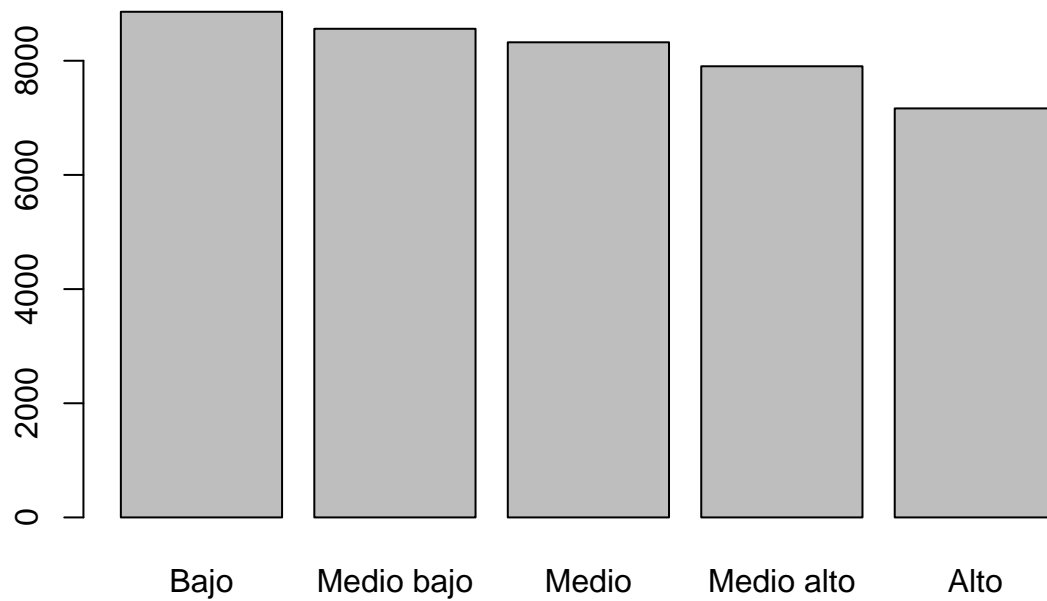
La base de datos tiene 40,809 registros, de los cuales 20,280, poco menos de la mitad, tienen información completa para todos los campos.

```
barplot(c(nrow(datos), sum(complete.cases(datos))),
        names.arg = c("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.

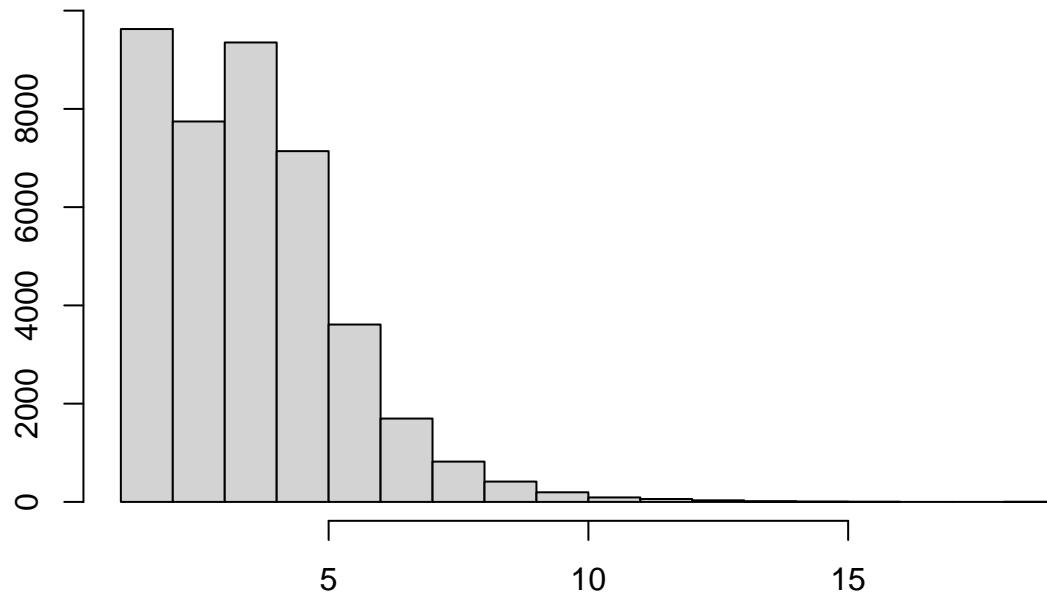
```
plot(datos$area)
```



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



```
quantile(datos$numpeho, probs = 0.95)
```

```
## 95%
```

```
## 7
```