

Taller Aplicado - Grupo 6

Erick Salazar Suarez

Sara Chemas

Jacob Cheney

Karol Suarez

2025-02-25

Instrucciones y librerías en **R** necesarias para los análisis

Se cargan las librerías necesarias

```
library(tidyverse)

## — Attaching core tidyverse packages —————
tidyverse 2.0.0 —
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ forcats   1.0.0      ✓ stringr   1.5.1
## ✓ ggplot2    3.5.1      ✓ tibble    3.2.1
## ✓ lubridate  1.9.4      ✓ tidyr     1.3.1
## ✓ purrr      1.0.4
## — Conflicts —————
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()
## ⓘ Use the conflicted package (<http://conflicted.r-lib.org/>) to force
all conflicts to become errors

library(ggplot2)
library(knitr)
```

Se carga la información en una variable llamada **datos** como se observa a continuación

```
datos <- read.csv("data_colombianos_exterior.csv")
```

1 Introduccion

1.1 Selección de la base de datos

La presente base de datos recopila información sobre ciudadanos colombianos residentes en el exterior. Su propósito es facilitar el análisis de las características demográficas, educativas y socioculturales de esta población, permitiendo identificar tendencias dentro de esta.

1.2 Descripción de la base de datos

País y Ciudad de Residencia: Para conocer la distribución geográfica de los ciudadanos.

Edad y Grupo de Edad: Que permiten entender la estructura etaria de la población migrante.

Área de Conocimiento: Datos útiles para analizar el perfil educativo y profesional.

Estado Civil, Género y Etnia: Muestran datos sociológicos que caracterizan a las personas en el data set.

Fecha de Registro: Fecha en la que se llevo a cabo la toma del registro en cuestion, se toma el año y el mes.

Estatura y Localización Geográfica: Datos que pueden ser utilizados para estudios específicos.

2 Identificación y clasificación de variables

Comando en R

Usamos ***glimpse(data)*** para observar las filas y columnas que tiene el data set, sin embargo se clarifica la informacion en las tablas propuestas mas adelante.

```
glimpse(datos)
## Rows: 1,634,319
## Columns: 17
## $ País                <chr> "AFGANISTAN", "AFGANISTAN", "ALBANIA",
"ALEMANIA..."
## $ Código.ISO.país      <chr> "AFG", "AFG", "ALB", "DEU", "DEU",
"DEU", "DEU",...
## $ Ciudad.de.Residencia <chr> "KABOL/KABUL", "KHOWST/A'IKHEL",
"TIRANE/FACESH"...
## $ Oficina.de.registro  <chr> "C. NUEVA DELHI", "C. NUEVA DELHI", "C.
ROMA", "..."
## $ Grupo.edad           <chr> "ADULTO", "ADULTO", "DESCONOCIDO",
```

```

"ADOLESCENTE"...
## $ Edad..años.      <int> 34, 44, -1, 12, 12, 12, 12, 12, 13, 13,
13, 13, ...
## $ Área.Conocimiento <chr> "NO INDICA", "NO INDICA", "NO INDICA",
"NINGUNA"...
## $ Sub.Area.Conocimiento <chr> "NO INDICA", "NO INDICA", "NO INDICA",
"NINGUNA"...
## $ Nivel.Académico   <chr> "NO INDICA", "BACHILLERATO",
"BACHILLERATO", "PR...
## $ Estado.civil      <chr> "DESCONOCIDO", "CASADO", "DESCONOCIDO",
"SOLTERO...
## $ Género            <chr> "MASCULINO", "FEMENINO", "FEMENINO",
"FEMENINO",...
## $ Etnia.de.la.persona <chr> "SIN ETNIA REGISTRADA", "NINGUNA", "SIN
ETNIA RE...
## $ Estatura..CM.     <int> -1, -1, -1, -1, -1, -1, -1, -1, -1, -1,
-1, -1, ...
## $ Ciudad.de.Nacimiento <chr> "(NO REGISTRA)",
"COLOMBIA/CUNDINAMARCA/BOGOTA",...
## $ Localización      <chr> "(33.93911, 67.709953)", "(33.93911,
67.709953)"...
## $ Fecha.de.Registro  <chr> "2014-09", "2015-09", "2015-10", "2023-
02", "202...
## $ Cantidad.de.personas <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, ...

```

2.1 Identificación

Variable	Descripción
País	Nombre del país de residencia.
Código.ISO.país	Código ISO del país de residencia.
Ciudad.de.Residencia	Ciudad específica donde reside la persona.
Oficina.de.registro	Oficina consular donde se realizó el registro.
Grupo.edad	Clasificación de la persona por grupo etario.
Edad..años.	Edad de la persona en años.
Área.Conocimiento	Área general de conocimiento académico.
Sub.Area.Conocimiento	Subdivisión específica del área de conocimiento.
Nivel.Académico	Nivel educativo alcanzado.
Estado.civil	Estado civil de la persona.
Género	Género de la persona.
Etnia.de.la.persona	Grupo étnico de la persona, si aplica.
Estatura..CM.	Estatura de la persona en centímetros.
Ciudad.de.Nacimiento	Ciudad donde nació la persona.

Variable	Descripción
Localización	Coordenadas geográficas de la residencia.
Fecha.de.Registro	Fecha en que se realizó el registro consular (año-mes).
Cantidad.de.personas	Número de personas registradas en el mismo registro.

2.2 Clasificación

Variable	Clasificación (Tipo y Nivel de Medición)
País	Cualitativa - Nominal
Código.ISO.país	Cualitativa - Nominal
Ciudad.de.Residencia	Cualitativa - Nominal
Oficina.de.registro	Cualitativa - Nominal
Grupo.edad	Cualitativa - Ordinal
Edad..años.	Cuantitativa - Razón
Área.Conocimiento	Cualitativa - Nominal
Sub.Area.Conocimiento	Cualitativa - Nominal
Nivel.Académico	Cualitativa - Ordinal
Estado.civil	Cualitativa - Nominal
Género	Cualitativa - Nominal
Etnia.de.la.persona	Cualitativa - Nominal
Estatura..CM.	Cuantitativa - Razón
Ciudad.de.Nacimiento	Cualitativa - Nominal
Localización	Cualitativa - Nominal
Fecha.de.Registro	Cuantitativa - Intervalo
Cantidad.de.personas	Cuantitativa - Razón

3 Presentación de datos

3.1 Tabla de frecuencia

3.2 Diagrama de barras

3.3 Histograma

4 Medidas de tendencia central

4.1 Calculo de la media

Se filtran valores negativos

```
# Filtrar valores negativos
df_edades <- datos %>%
  filter(Edad..años. >= 0)

df_estaturas <- datos %>%
  filter(Estatura..CM. > 0)
```

Se calcula la media

```
# Cálculo de la media
edad_media <- mean(df_edades$Edad..años., na.rm = TRUE)
estatura_media <- mean(df_estaturas$Estatura..CM., na.rm = TRUE)
personas_media <- mean(df_edades$Cantidad.de.personas, na.rm = TRUE)
```

4.2 Calculo de la mediana

Se calcula la mediana

```
# Cálculo de la mediana
edad_mediana <- median(df_edades$Edad..años., na.rm = TRUE)
estatura_mediana <- median(df_estaturas$Estatura..CM., na.rm = TRUE)
personas_mediana <- median(df_edades$Cantidad.de.personas, na.rm = TRUE)
```

4.3 Calculo de la moda

Se define una funcion y se guardan los resultados

```
# Cálculo de la moda
moda <- function(x) {
  uniq_x <- unique(x)
  uniq_x[which.max(tabulate(match(x, uniq_x)))]
}

edad_moda <- moda(df_edades$Edad..años.)
```

```
estatura_moda <- moda(df_estaturas$Estatura..CM.)
personas_moda <- moda(df_edades$Cantidad.de.personas)
```

4.4 Resultados

Se muestran los resultados obtenidos

```
# Mostrar resultados
kable(data.frame(
  Variable = c("Edad", "Estatura", "Cantidad de Personas"),
  Media = c(edad_media, estatura_media, personas_media),
  Mediana = c(edad_mediana, estatura_mediana, personas_mediana),
  Moda = c(edad_moda, estatura_moda, personas_moda)
))
```

Variable	Media	Mediana	Moda
Edad	43.898256	42	35
Estatura	165.249640	165	160
Cantidad de Personas	1.008273	1	1

Análisis de los resultados

Edad:

La **media** de edad es de aproximadamente 43.9 años, lo que indica que, en promedio, las personas en el conjunto de datos son de mediana edad. La **mediana** es 42 años, lo cual refuerza que la mayoría de las edades están cerca de este valor, mostrando una distribución bastante centrada. La **moda**, que representa el valor más frecuente, es de 35 años, lo que sugiere que esta es la edad más común en el grupo analizado.

Estatura:

La **media** de estatura es de aproximadamente 165.25 cm, y la **mediana** es de 165 cm, lo que muestra una distribución bastante uniforme en torno a esta altura. La **moda** es 160 cm, lo que indica que esta es la estatura más común en el grupo.

Cantidad de Personas:

La **media** es de 1.01, la **mediana** es 1, y la **moda** también es 1, lo que indica que la mayoría de los registros corresponden a individuos que fueron contados de manera individual, sin agrupaciones.

5 Medidas de variabilidad

5.1 Rango

5.2 Varianza

5.3 Desviación Estándar

6 Medidas de posición

6.1 Percentiles

6.2 Cuartiles

6.3 Deciles

7 Aplicación de distribución de probabilidad

8 Aplicación del modelo normal

9 Interpretación de resultados y conclusiones