Unidad 1: Conceptos generales y Análisis exploratorio de datos

Mg. J. Eduardo Gamboa U.

Table of contents

Carga de paquetes	2
Lectura de datos	2
Medidas de tendencia central	3
Media	3
Mediana	4
Moda	5
Medidas de posición	6
Medidas de dispersión	8
Rango	8
Rango intercuartil	8
Varianza	8
Desviación estándar	8
Varianza	8
Coeficiente de variabilidad	8
Medidas de asimetría	8
Tablas de frecuencia	8
Gráficas	8

Carga de paquetes

```
library(readr)
library(dplyr)
library(modeest)
library(sjstats)
library(cleaner)
library(DescTools)
```

Lectura de datos

Se empleará el archivo Salud.csv, el cual recopila datos de pacientes en torno a las siguientes cuatro variables:

- Edad (en años)
- Tiempo semanal de ejercicios (en minutos)
- Índice de Masa Corporal
- Presión sistólica (en mmHg)

```
datos <- read_csv('Salud.csv')

Rows: 100 Columns: 4
-- Column specification ------
Delimiter: ","
dbl (4): Edad, Minutos_ejercicio, IMC, Presion_sistolica

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.

datos |> head(5)
```

 2
 42
 142
 20.1
 142

 3
 61
 58
 20.2
 139

```
      4
      41
      25
      17.8
      120

      5
      40
      46
      17.4
      133
```

```
datos |> tail(3)
```

```
# A tibble: 3 x 4
  Edad Minutos_ejercicio
                             IMC Presion_sistolica
  <dbl>
                    <dbl> <dbl>
                                             <dbl>
1
     44
                       176 20.4
                                                111
2
     37
                        8 20.1
                                                131
                       98 15.7
     44
                                                128
```

Medidas de tendencia central

Media

Ejemplo 1

Interpretar la media aritmética de la edad

```
datos |> summarize(Media = mean(Edad))
```

```
# A tibble: 1 x 1
   Media
   <dbl>
1 46.0
```

La edad promedio de los pacientes es de 46 años.

Ejemplo 2

Interpretar la presión sistólica media de los pacientes mayores de 50 años.

```
datos |> filter(Edad > 50) |> summarize(Media = mean(Edad))
```

```
# A tibble: 1 x 1
   Media
   <dbl>
1 59.9
```

La presión sistólica promedio de los pacientes mayores de 50 años es de 59.9 mmHg.

Mediana

Ejemplo 3

Interpretar la mediana del IMC

Al menos la mitad de las personas tiene un IMC menor o igual a 19.2.

Ejemplo 4

Interpretar la mediana de la presión sistólica para las personas que son sedentarias (menos de 30 minutos de ejercicios a la semana) y las que no lo son.

```
datos |>
  mutate(Sedentario = ifelse(Minutos_ejercicio<30, "Sí", "No")) -> datos

datos |>
  group_by(Sedentario) |>
  summarize(Medianas = median(Presion_sistolica))
```

Al menos la mitad de las personas sedentarias presenta una presión sistólica de como máximo 138 mmHg (¡elevada!). Por otro lado, al menos el 50% de las personas que no son sedentarias tiene una presión sistólica menor o igual a 121 mmHg (casi en el rango normal).

Moda

Ejemplo 5

Interpretar la moda de la presión sistólica

```
datos |>
   summarize(Moda = mfv(Presion_sistolica))

# A tibble: 1 x 1
   Moda
   <dbl>
1 121
```

La presión sistólica más frecuente es de 121 mHg.

Ejemplo 6

Interpretar la moda de la edad

```
datos |>
  reframe(Moda = mfv(Edad))

# A tibble: 2 x 1
    Moda
  <dbl>
1    36
2    59
```

Las edades más frecuentes de los pacientes son 36 y 59 años.

Ejemplo 7

Intepretar la moda del tiempo semanal de ejercicio de los pacientes sedentarios

```
datos |>
  filter(Sedentario == "Sí") |>
  reframe(Moda = mfv(Minutos_ejercicio))
```

```
# A tibble: 1 x 1
    Moda
    <dbl>
1    15
```

El tiempo de ejercicios más frecuente entre los pacientes sedentarios es de 12 minutos.

Medidas de posición

Ejemplo 8

Interpretar el percentil 41 de la edad

```
datos |>
   summarize(P41 = quantile(Edad, 0.41))

# A tibble: 1 x 1
    P41
   <dbl>
1 42
```

Al menos el 41% de los pacientes tiene 42 años de edad o menos.

Ejemplo 9

Interpretar los percentiles 12 y 74 de los tiempos semanales de ejercicio de las personas no sedentarias

Al menos el 12% de los pacientes no sedentarios realiza como máximo 61.4 minutos de ejercicio a la semana, mientras que al menos el 74% realiza hasta 241 minutos semanales de actividad física.

Ejemplo 10

Interpretar los cuartiles del IMC de las personas adultas mayores (60 años a más)

```
datos |>
  filter(Edad >= 60) |>
  reframe(Cuartiles = quantile(IMC, c(0.25,0.50,0.75)))

# A tibble: 3 x 1
  Cuartiles
```

Al menos el 25% de los pacientes tiene un IMC igual o inferior a 19.4, mientras que como máximo el 50% tiene un IMC igual o inferior a 20. Además, hasta el 75% de los pacientes presenta un IMC igual o inferior a 20.9.

Ejemplo 11

¿Cuál es el tiempo máximo de ejercicio semanal que realiza un paciente joven (menor de 30 años) para estar dentro del 20% que menos ejercicio realiza?

```
datos |>
  filter(Edad < 30) |>
  summarize(P20 = quantile(Minutos_ejercicio, 0.20))

# A tibble: 1 x 1
    P20
    <dbl>
1 58
```

58 minutos semanales es el tiempo máximo de ejercicio que realiza un paciente joven (menor de 30años) para estar dentro del 20% que menos ejercicio realiza.

Rango Rango intercuartil Varianza Desviación estándar Varianza Coeficiente de variabilidad Medidas de asimetría Tablas de frecuencia

Gráficas

Medidas de dispersión