

Medidas de localización o centralidad

Manuel Villalobos Cid

Contents

Introducción	1
Conjunto de datos	1
Medidas de localización	2
Media aritmética o promedio	2
Mediana	3
Moda	4
Deciles, quintiles, cuartiles y percentiles	5
Actividad	6

Introducción

Las **medidas de localización** o **centralidad** caracterizan la distribución de frecuencia de una variable estadística en una población o muestra, permitiendo extraer información sobre su **tendencia central**, y de ciertas posiciones específicas de ella. Existen diversas medidas de localización cada una con diferentes propiedades: **media aritmética**, **mediana**, **moda**, **percentiles**, **deciles**, **cuartiles**, **quintiles**, **entre otros**.

Si bien estas medidas ayudan a caracterizar una muestra, sus resultados no pueden ser generalizables para inferir conclusiones categóricas, requiriendo al menos conocer la dispersión de la variable estadística estudiada (**medidas de dispersión**).

Esta guía se centra en la interpretación de la **medidas de localización**, las que será estudiadas usando un conjunto de datos real proveniente del área hospitalaria.

Conjunto de datos

Las actividades se centran en el **contexto hospitalario**. Los datos (Descargar) a utilizar (**BD_principal_HBLT.csv**) fueron descargados desde la página del **Departamento de Estadística e Informaciones del Ministerio de Salud (DEIS)**. Estos corresponden a múltiples datos codificados de los pacientes egresados durante el año 2017 desde uno de los mayores hospitales públicos de nuestro país: características del establecimiento de salud, datos personales del paciente y de patologías tratadas. Una descripción de cada variable se encuentra en el archivo **descripción_datos.pdf**, y la identificación de los códigos respectivos se encuentran en la carpeta **Diccionarios**.

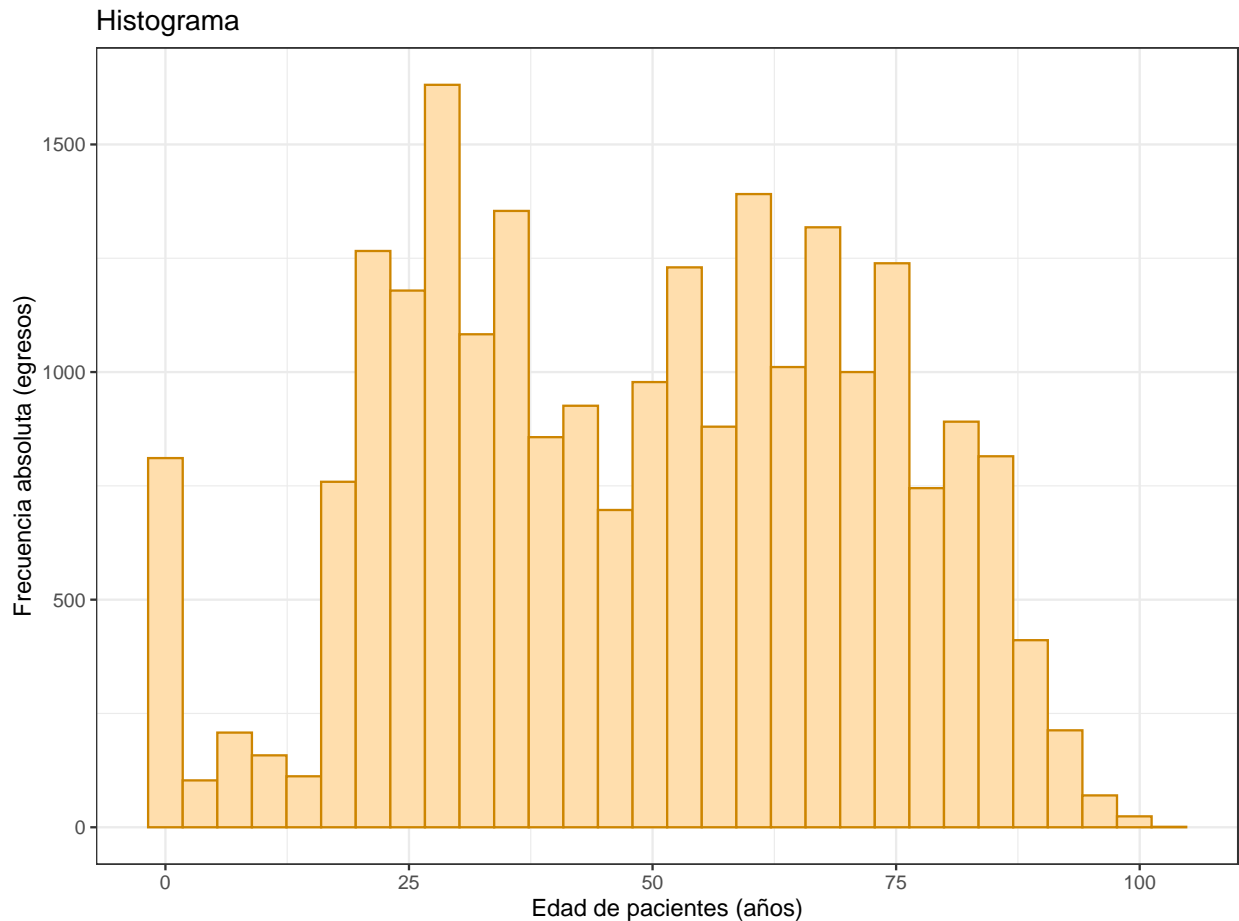
```
datos=read.csv("datos/BD_principal_HBLT.csv",header = T,sep = ",")#Cargar datos
head(datos[,c(3,4,11,12,13,15,16,17)],7) #Mostrar como los datos están almacenados (7 elementos)
```

```
## SEXO EDAD FECHA_EGR SERC_EGR DIAG1 DIAS_ESTAD COND_EGR INTERV_Q
## 1 2 55 08/06/2017 330 M480 2 1 1
## 2 1 56 21/06/2017 321 G122 134 1 1
```

## 3	2	77	22/06/2017	110	I219	33	1	2
## 4	2	71	14/03/2017	131	M751	1	1	1
## 5	2	23	14/03/2017	161	O219	2	1	2
## 6	2	63	14/03/2017	162	N908	1	1	2
## 7	2	74	08/02/2017	321	I639	6	1	2

Medidas de localización

Para estudiar las medidas de localización se estudiará el perfil etario de los pacientes egresados en el **Hospital Barros Luco Trudeau (HBLT)** (COD: 113130). Este corresponde a unos de los establecimientos más grandes de nuestro país y posee una **alta complejidad**. Según los datos, durante el año 2017 egresaron **23,361** pacientes. El histograma de la variable edad es el siguiente:



Media aritmética o promedio

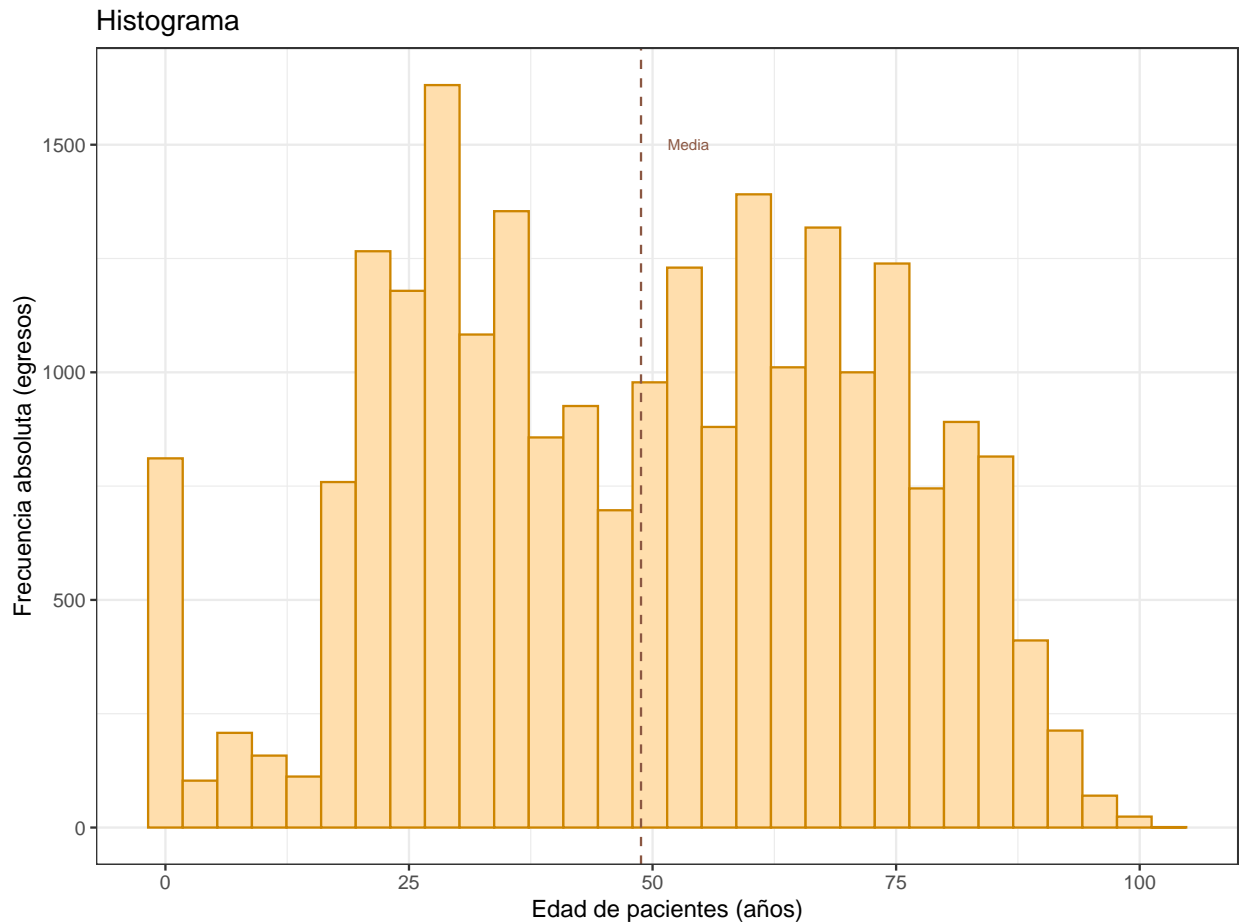
Esta métrica se obtiene sumando de todos los valores de la variable estadística estudiada, dividiéndolos por el número de sumandos. Generalmente la media de una población es representada por la letra μ , mientras que la muestral es caracterizada como \bar{x} . Una de las formas de efectuar su cálculo en R es usando la función `mean()`.

```
media = mean(datos$EDAD)
print(paste("La media de edad de los pacientes es aproximadamente:", round(media), "años"))
```

```
## [1] "La media de edad de los pacientes es aproximadamente: 49 años"
```

Esta función indica que la **media de edad** de los pacientes atendidos es de aproximadamente **49 años**. Este dato también puede ser incorporado al histograma.

```
grafico=grafico+geom_vline(xintercept =media,color="lightsalmon4",
                           linetype="dashed", show.legend = T)
grafico= grafico+annotate("text", x = (media+(media*0.1)), y = 1500, label =
                           "Media",color="lightsalmon4",cex=2.5)
plot(grafico)
```

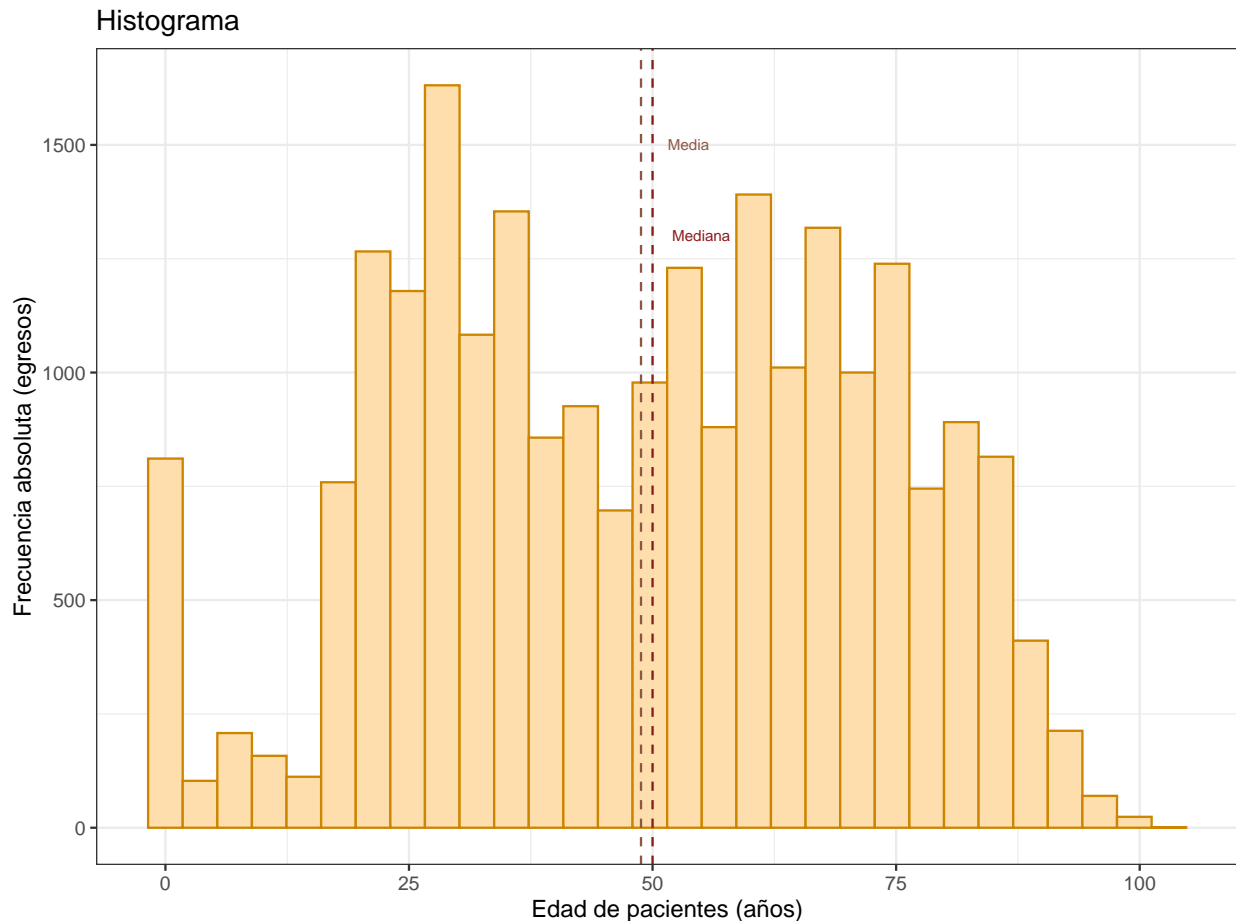


Mediana

La **mediana** es valor medio de los dos valores que se encuentran justo en la mitad de un grupo de datos ordenados. Si se tiene un número par de datos, se deben sumar los dos valores centrales y dividir su resultado entre 2. En cambio, si el número de datos es impar, el dato del centro es la mediana de la muestra. En R la mediana puede ser calculada usando la función **median()**. En el ejemplo estudiado, la mediana (**50 años**) está muy cercana al valor de la media aritmética (**49 años**), sin embargo, esto se debe a las características uniforme de la distribución de frecuencia, y no constituye una regla general.

```
mediana = median(datos$EDAD)
print(paste("La mediana de edad de los pacientes es aproximadamente:",round(mediana),"años"))

## [1] "La mediana de edad de los pacientes es aproximadamente: 50 años"
```



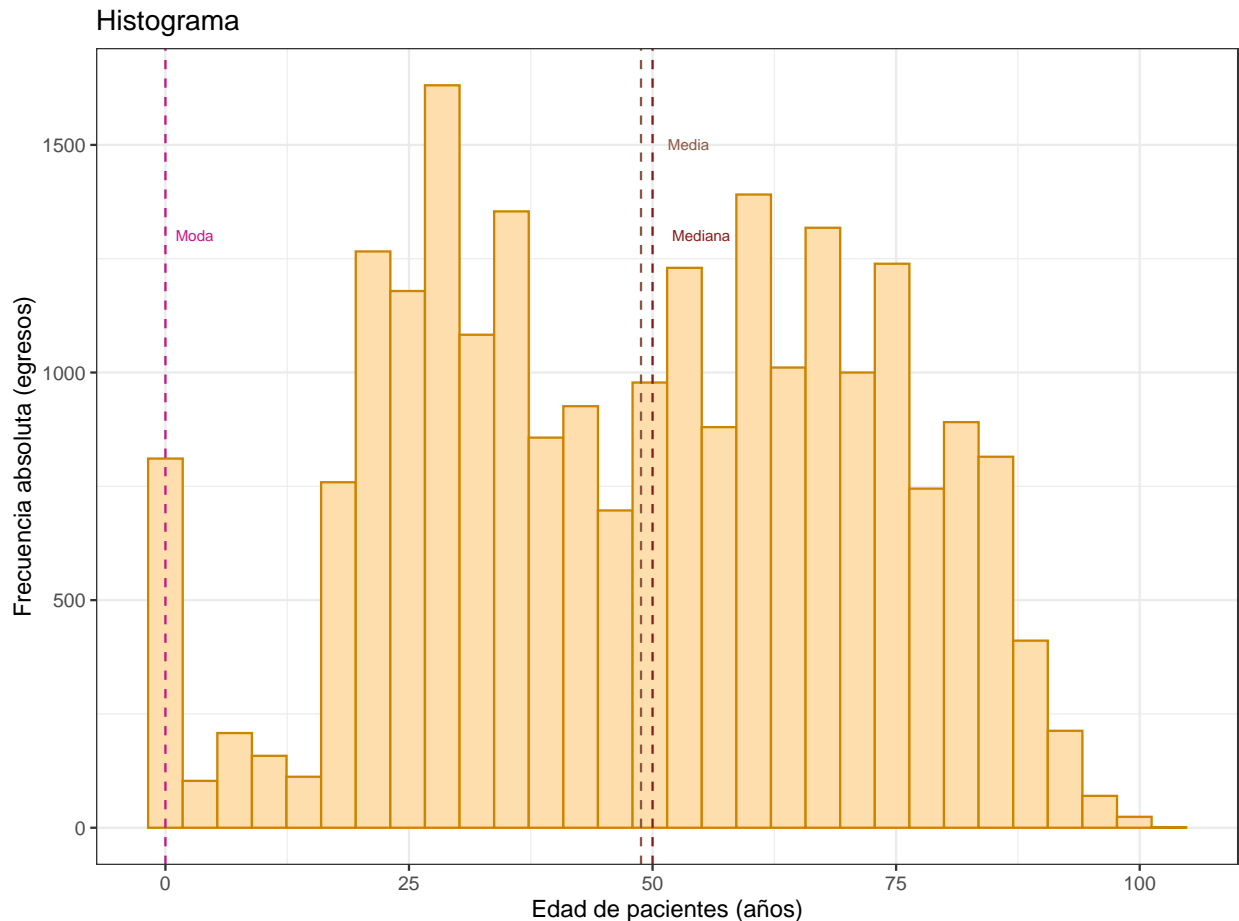
Moda

La **moda** en una agrupación de datos es el valor que más se repite. Cuando sólo existe una moda se habla de una muestra **modal**; si hay dos modas, la muestra es **bimodal** y si hay tres modas o más, la muestra es **multimodal**. Se denomina muestra **amodal** cuando no hay moda. No existe una función particular para calcular esta métrica en R, sin embargo, puede ser calculada empleando un histograma o una tabla de frecuencia.

Para su cálculo se usará la función `table()`. Los resultados indican que la moda es **0** año con **808** pacientes atendidos (3.4% del total de egresos).

```
tabla=table(datos$EDAD) #Crear tabla de frecuencia
moda=names(tabla)[which.max(tabla)] #Buscar elemento más repetido
moda=as.numeric(moda) #Convertir a número
cat(paste("La edad de pacientes que más se repite (moda) es:",round(modas),"años, \n con"
,tabla[which.max(tabla)],"pacientes atendidos\n"))
```

```
## La edad de pacientes que más se repite (moda) es: 0 años,
## con 808 pacientes atendidos
```



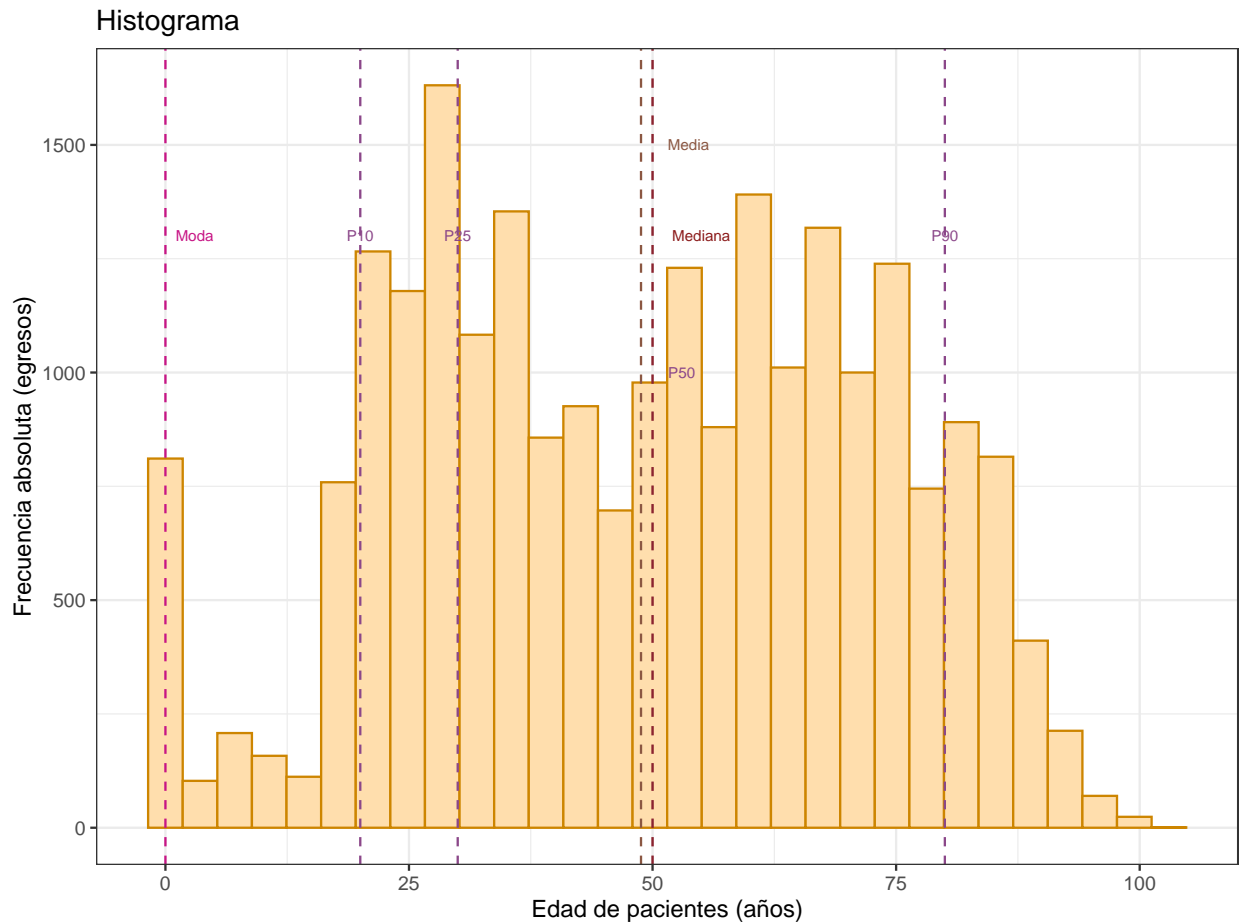
Deciles, quintiles, cuartiles y percentiles

Estas métricas dividen los datos en n partes para encontrar un valor determinado. Por ejemplo, los **deciles** dividen a la población en 10 partes, los **quintiles** dividen la población en 5 partes, los **cuartiles** dividen la población en 4 partes y los **percentiles** en 100. Estas definiciones se pueden combinar, por ejemplo, el Percentil 50 corresponde al segundo cuartil o a la mediana de los datos. Para poblaciones pequeñas R y otras herramientas emplean **interpolación de datos** para el cálculo de percentiles. Los percentiles pueden ser calculados en R usando la función `quantile()`.

Por ejemplo, calcular los percentiles 10, 25 (1 cuartil), 50 (2 cuartil - mediana), y 90. Al hacerlo, los resultados indican que el **10%** de los pacientes atendidos tienen como máximo **20 años**, el **25%** de ellos tiene **30 años**, el **50%** tiene **50 años**, y el **90%** de los pacientes tienen como máximo **80 años**.

```
percentiles=quantile(datos$EDAD, prob = c(0.1,0.25,0.5,0.9))
print(percentiles)
```

```
## 10% 25% 50% 90%
## 20 30 50 80
```



En conclusión, los datos muestran que si bien existe una distribución de frecuencia uniforme (percentiles) sobre los 18 años (con **media** y **mediana** cercana a los **50** años), casi no se atienden pacientes pediátricos (desplazamiento de percentil 10). Esto se debe a que el Hospital Barros Luco Trudeau deriva la mayor parte de ellos al Hospital Dr. Exequiel González Cortés, quién se especializa en su atención. Los pacientes con 0 año de edad corresponden a recién nacidos atendidos cuyas madres fueron ingresadas al establecimiento, y dan cuenta de la **moda** de la población estudiada.

Actividad

- Estudie la distribución de frecuencia de sus datos aplicando métricas de localización.
- Enfóquese en el análisis/interpretación de resultados y documente empleando Markdown.