

**Universidad Manuela Beltrán**

**Diplomado en Bioestadística  
modulo 2**

Elaborado por

**Fredy Medina**

**Magister  
(Educación en tecnología)**

Bogotá, Colombia  
Junio 2024.

©2024. - Fredy Medina

Derechos Reservados



# Contenido

<b>1</b>	<b>MÓDULO 2</b>	<b>1</b>
1.1	Variables . . . . .	2
1.2	Variables y escalas de medición . . . . .	3
1.2.1	Exploración de datos EDA . . . . .	6
1.2.2	Variables cuasicuantitativas u ordinales . . . . .	7
1.2.3	Variables cuantitativas o numéricas . . . . .	8
1.2.4	Exploración de datos . . . . .	9
1.3	Medidas de tendencia central . . . . .	11
1.4	Medidas de Variabilidad . . . . .	13
1.4.1	Actividad de comprobación trabajo autonomo . . . . .	16
	<b>Bibliografía</b>	<b>17</b>

# Capítulo 1

## MÓDULO 2

En el ámbito de la bioestadística y la estadística descriptiva, la correcta identificación y análisis de las variables es fundamental para la interpretación de los datos y la toma de decisiones informadas. Este capítulo se enfoca en la clasificación de los distintos tipos de variables, las medidas de tendencia central y las medidas de variabilidad, todos elementos esenciales para resumir y describir conjuntos de datos.

### Tipos de Variables

Las variables estadísticas se dividen en diferentes categorías según sus características y el tipo de datos que representan. Comprender estas categorías nos permite aplicar las técnicas estadísticas adecuadas para analizar la información de manera efectiva.

1. Variables cualitativas: Estas variables describen atributos o categorías sin un orden inherente. Ejemplos incluyen el grupo sanguíneo (A, B, AB, O) y el género (masculino, femenino).

2. Variables cuasi-cuantitativas u ordinales: Aunque estas variables son de tipo nominal, se pueden ordenar. Ejemplos incluyen el grado de recuperación de un paciente (Nada, Poco, Moderado, Bueno, Muy Bueno) y escalas de dolor (1 a 5).

3. Variables cuantitativas o numéricas: Estas variables tienen valores numéricos que permiten operaciones aritméticas. Se dividen en: - Discretas: No admiten valores intermedios. Ejemplo: número de hijos (0, 1, 2, 3, etc.). - Continuas: Admiten valores intermedios. Ejemplo: el peso de un niño al nacer.

### Medidas de Tendencia Central

Las medidas de tendencia central proporcionan un resumen conciso de los datos, facilitando la comprensión de las características principales de los conjuntos de datos.

- Media: Es el promedio de todos los valores. Se calcula sumando todos los valores y dividiendo por el número total de valores. La media es útil para datos cuantitativos y es sensible a valores extremos.

- Mediana: Es el valor central cuando los datos están ordenados. Si hay un número par de observaciones, la mediana es el promedio de los dos valores centrales. La mediana es útil para datos ordinales y no se ve afectada por valores extremos.

- Moda: Es el valor que ocurre con mayor frecuencia en un conjunto de datos. Puede haber más de una moda en un conjunto de datos. La moda es útil para datos cualitativos y ordinales.

#### Medidas de Variabilidad

Las medidas de variabilidad indican cómo se dispersan los datos alrededor de la tendencia central, proporcionando información sobre la variabilidad y consistencia de los datos.

- Rango: Es la diferencia entre el valor máximo y el valor mínimo. Es una medida simple pero sensible a valores extremos.

- Varianza: Es la media de las diferencias al cuadrado entre cada valor y la media del conjunto de datos. La varianza proporciona una medida de la dispersión de los datos en relación con la media.

- Desviación estándar: Es la raíz cuadrada de la varianza. Proporciona una medida de dispersión en las mismas unidades que los datos originales y es útil para interpretar la variabilidad en contextos prácticos.

- Rango intercuartílico (IQR): Es la diferencia entre el tercer cuartil (Q3) y el primer cuartil (Q1). Mide la dispersión de la mitad central de los datos y es menos sensible a valores extremos.

Al final de este capítulo, los lectores tendrán una comprensión clara de los tipos de variables, las medidas de tendencia central y las medidas de variabilidad. Estarán equipados con las habilidades necesarias para aplicar estos conceptos en la práctica de la bioestadística, facilitando la interpretación precisa de los datos biomédicos y mejorando la toma de decisiones informadas en contextos clínicos y de investigación.

## 1.1 Variables

Una variable en estadística es un símbolo que representa una característica, atributo o propiedad que puede tomar diferentes valores. Estos valores pueden variar entre distintos individuos u observaciones y se pueden clasificar en diferentes tipos según sus características. Las variables son fundamentales en el análisis estadístico porque permiten estudiar y entender cómo se comportan ciertos fenómenos y cómo diferentes factores pueden influir en ellos. Un dato es un valor específico que una variable puede tomar. Representa una observación particular que se ha recogido durante la investigación o el estudio estadístico. Los datos son las unidades básicas de la información que se utilizan para el análisis y la toma de decisiones.

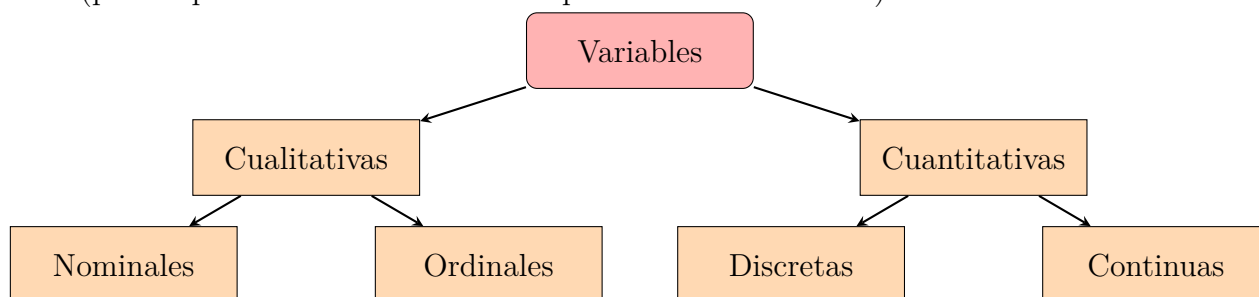
La diferencia fundamental entre una variable y un dato radica en su naturaleza y función: una variable es un concepto abstracto que representa una característica que puede variar entre diferentes observaciones, como "grupo sanguíneo" o "peso al nacer", mientras que un dato es un valor específico que una variable puede tomar, como "A" o "3.2 kg". En resumen, las variables definen qué se está midiendo, y los datos son las observaciones concretas y específicas de esas mediciones. Una variable es cualquier

característica cuyo valor puede cambiar de un objeto a otro en la población. Por su naturaleza podemos considerar variables del tipo:

Univariantes: Son aquellas que en un conjunto de datos univariantes se compone de observaciones realizadas en una sola variable.

Bivariantes: Se tienen datos bivariantes cuando se realizan observaciones en cada unidad experimental de dos variables.

Multivariantes: Estos surgen cuando se realizan observaciones en más de una variable (por lo que bivalente es un caso especial de multivariante).



## 1.2 Variables y escalas de medición

Integer: Objetos con valores enteros.

Ejemplos:

Edad (25, 30, 45)

Número de hijos (0, 2, 3)

Años de experiencia laboral (1, 5, 10)

Convertir a entero: `as.integer()`

Verificar si es entero: `is.integer()`

Numeric: Objetos que admiten números reales (incluyendo  $\pi$  y el número de Euler  $e$ ).

Ejemplos: Altura (1.65, 1.70, 1.75)

Peso (60.5, 70.2, 80.1)

Temperatura (36.6, 37.0, 37.5)

Convertir a numérico: `as.numeric()`

Verificar si es numérico: `is.numeric()`

Character: Objetos nominales.

Ejemplos:

Letras del alfabeto ("A", "B", "C")

Identificaciones ("ID123", "ID456", "ID789")

Nombres ("Ana", "Luis", "Carlos")

Convertir a carácter: `as.character()`

Verificar si es carácter: `is.character()`

Factor: Objetos categóricos con niveles.

Ejemplos:

Sexo ("Masculino", "Femenino")

Raza ("Blanco", "Negro", "Asiático")

Estado civil ("Soltero", "Casado", "Divorciado")

Convertir a factor: `as.factor()`

Verificar si es factor: `is.factor()`

Logical: Objetos con valores lógicos. Solo acepta elementos como TRUE, FALSE y NA (valor faltante).

Ejemplos:

Resultado de una prueba (TRUE, FALSE)

Permiso concedido (TRUE, FALSE)

Valor faltante (NA)

Convertir a lógico: `as.logical()`

Verificar si es lógico: `is.logical()`

Complex: Objetos que corresponden a números complejos.

Incluyen el número imaginario  $i$  (utilizado en R para denotar  $\sqrt{-1}$ ).

Ejemplos:

Números complejos ( $2 + 3i$ ,  $4 - 5i$ )

Verificar si es complejo: `is.complex()`

la siguiente es una tabla de referencia para algunos tipos de variables en R:

Tipo de Variable	Ejemplos	Conversión	Verificación
Integer	Edad (25, 30), Número de hijos	<code>as.integer()</code>	<code>is.integer()</code>
Numeric	Altura (1.65, 1.70), Peso (60.5)	<code>as.numeric()</code>	<code>is.numeric()</code>
Character	Nombres ("Ana"), IDs ("ID123")	<code>as.character()</code>	<code>is.character()</code>
Factor	Sexo ("Masculino"), Raza	<code>as.factor()</code>	<code>is.factor()</code>
Logical	TRUE, FALSE, NA	<code>as.logical()</code>	<code>is.logical()</code>
Complex	Números complejos ( $2 + 3i$ )	N/A	<code>is.complex()</code>

Tabla 1.1: Tipos de variables en R con ejemplos, funciones de conversión y verificación.

En bioestadística, las escalas de medición son fundamentales para la recolección, análisis e interpretación de datos. Las escalas de medición determinan cómo se pueden manipular los datos y qué tipos de análisis son apropiados. Aquí hay una explicación de las escalas de medición aplicadas a la bioestadística, junto con ejemplos:



**Escala Nominal:**

Descripción: La escala nominal clasifica los datos en categorías que no tienen un orden intrínseco. Cada categoría es mutuamente excluyente y colectivamente exhaustiva. Ejemplos en Bioestadística: Grupo sanguíneo: (A, B, AB, O). Sexo: (Masculino, Femenino). Presencia de una enfermedad: (Sí, No).

Análisis posibles: Se pueden realizar pruebas de frecuencias y proporciones, como la prueba de chi-cuadrado para la independencia.

**Escala Ordinal:**

Descripción: La escala ordinal clasifica los datos en categorías con un orden específico, pero las diferencias entre las categorías no son necesariamente uniformes. Ejemplos en Bioestadística: Grado de severidad de una enfermedad: (Leve, Moderada, Severa). Clasificación del dolor: (Sin dolor, Dolor leve, Dolor moderado, Dolor severo). Estadios del cáncer: (I, II, III, IV).

Análisis posibles: Se pueden utilizar medianas, rangos y pruebas no paramétricas como la prueba de Wilcoxon o la prueba de Kruskal-Wallis.

**Escala de Intervalo:**

Descripción: La escala de intervalo no solo clasifica y ordena los datos, sino que también tiene intervalos iguales entre los puntos de datos. No tiene un punto cero absoluto. Ejemplos en Bioestadística: Temperatura corporal en grados Celsius: Las diferencias entre las temperaturas son iguales, pero 0 grados no representa una ausencia de temperatura.

Análisis posibles: Se pueden calcular medias y desviaciones estándar. Sin embargo, no se pueden realizar multiplicaciones o divisiones significativas.

**Escala de Razón:**

Descripción: La escala de razón tiene todas las propiedades de la escala de intervalo y además tiene un punto cero absoluto, que indica la ausencia total de la variable medida.

Ejemplos en Bioestadística: Longitud: (cm, m).

Peso: (kg).

Duración de una enfermedad: (días).

Frecuencia cardíaca: (latidos por minuto).

Análisis posibles: Se pueden realizar todas las operaciones matemáticas, incluyendo promedios, proporciones y tasas. Los análisis estadísticos incluyen regresión lineal y análisis de varianza (ANOVA).

Las escalas de medición en bioestadística son cruciales para determinar los métodos adecuados de análisis. Cada tipo de escala permite diferentes tipos de manipulaciones y análisis de los datos, lo que a su vez afecta las conclusiones que se pueden extraer de los estudios biológicos y médicos. Al comprender y aplicar correctamente las escalas de medición, los investigadores pueden realizar análisis estadísticos precisos y significativos.

Escala de Medición	Tipo de Variable	Ejemplos	Conversión
Nominal	Factor, Character	Factor: <code>sexo &lt;- factor(c("Masculino", "Femenino"))</code> Character: <code>colores &lt;- c("Rojo", "Verde", "Azul")</code>	<code>as.factor() as.character()</code>
Ordinal	Factor ordenado	<code>nivel_educacion &lt;- factor(c("Primaria", "Secundaria", "Terciaria"), levels = c("Primaria", "Secundaria", "Terciaria"), ordered = TRUE)</code>	<code>as.ordered()</code>
Intervalo	Numeric	<code>temperaturas &lt;- c(10, 20, 30)</code>	<code>as.numeric()</code>
Razón	Numeric, Integer	Numeric: <code>altura &lt;- c(1.65, 1.70, 1.75)</code> Integer: <code>edad &lt;- as.integer(c(25, 30, 45))</code>	<code>as.numeric() as.integer()</code>
Lógico	Logical	<code>resultado_prueba &lt;- c(TRUE, FALSE, NA)</code>	<code>as.logical()</code>
Complejo	Complex	<code>num_complejo &lt;- c(2 + 3i, 4 - 5i)</code>	N/A

Tabla 1.2: Tipos de variables en R relacionados con escalas de medición, ejemplos y funciones de conversión.

### 1.2.1 Exploración de datos EDA

La exploración de datos es una fase crucial en el análisis de datos, que implica examinar y entender los datos disponibles antes de realizar análisis más avanzados. Su objetivo es identificar patrones, anomalías, y características generales del conjunto de datos para guiar la selección de técnicas analíticas y asegurar la calidad de los resultados. En un dataset de salud, la exploración de datos puede ayudar a revelar información útil sobre los pacientes, las condiciones de salud, y las variables relevantes.

La exploración de datos (también conocida como análisis exploratorio de datos o EDA, por sus siglas en inglés) incluye:

#### Inspección de Datos:

Visualización Inicial: Examinar un subconjunto de los datos para entender su estructura y contenido. Tipos de Variables: Identificar los tipos de datos (numéricos, categóricos, fechas, etc.).

#### Estadísticas Descriptivas:

Medidas de Tendencia Central: Media, mediana, moda. Medidas de Dispersión: Desviación estándar, rango intercuartil. Distribuciones: Histograma, densidad.

#### Identificación de Valores Faltantes:

Frecuencia de Valores Faltantes: Contar cuántos datos faltan en cada variable. Patrones de Datos Faltantes: Determinar si los datos faltantes son aleatorios o siguen algún patrón.

#### Detección de Valores Atípicos:

Visualización: Utilizar diagramas de caja (boxplots) y gráficos de dispersión.

Métodos Estadísticos: Identificar valores que se desvían significativamente de la media.

### **Análisis de Correlación:**

Matriz de Correlación: Analizar la relación entre variables numéricas. Visualización: Gráficos de dispersión, matrices de correlación.

### **Exploración de Relaciones:**

Comparaciones entre Grupos: Analizar diferencias entre subgrupos, por ejemplo, pacientes con diferentes diagnósticos. Tendencias Temporales: Examinar cómo las variables cambian con el tiempo si se tiene una serie temporal.

## **1.2.2 Variables cuasicuantitativas u ordinales**

Las variables cuasicuantitativas, también conocidas como ordinales, son aquellas que, aunque expresan categorías, existe un orden inherente entre ellas. A diferencia de las variables nominales (donde el orden no importa), en las ordinales, un valor es "mayor" o "menor" que otro.

Ejemplos de Variables Ordinales Nivel educativo: Primaria, Secundaria, Preparatoria, Universidad, Posgrado. Grado de satisfacción: Muy insatisfecho, Insatisfecho, Neutral, Satisfecho, Muy satisfecho. Rango de frecuencia: Nunca, Rara vez, A veces, A menudo, Siempre Supongamos que tienes datos sobre la satisfacción de clientes con un servicio, y quieres crear una variable ordinal en R:

```
Vector de datos de satisfacción
satisfaccion <- c("Muy insatisfecho", "Insatisfecho", "Neutral", "Satisfecho", "Muy
satisfecho")
```

```
Convertir a factor ordenado
satisfaccio_ordinal <- factor(satisfaccion, levels = c("Muy insatisfecho", "Insatisfecho", "Neutral", "Satisfecho", "Muy satisfecho"), ordered = TRUE)
```

```
Ver la variable creada:
print(satisfaccion_ordinal)
```

```
Esto nos permite realizar varias operaciones entre las variables como:
comparación entre variables
satisfaccion_ordinal[1] < satisfaccion_ordinal[5] # Retorna TRUE
```

```
crear tablas de frecuencia table(satisfaccion_ordinal)
```

```
También nos permite crear gráficos a partir del vector
barplot(table(satisfaccion_ordinal))
```

Puedes cambiar el orden de las categorías o transformar la variable a un tipo diferente si es necesario:

```
satisfaccion_ordinal <- factor(satisfaccion_ordinal, levels = c("Muy satisfecho", "Satisfecho", "Neutral", "Insatisfecho", "Muy insatisfecho"), ordered = TRUE)
```

En el análisis estadístico, las variables ordinales pueden ser utilizadas en modelos que reconozcan la naturaleza ordinal de los datos, como la regresión logística ordinal. Como también para analizar la asociación entre variables ordinales y otras variables. En el tratamiento de las variables ordinales se debe tener en cuenta que aunque las variables ordinales tienen un orden, no son estrictamente numéricas. Sumar o restar niveles no tiene sentido, La codificación de los niveles es arbitraria, pero el orden es importante. Hablaremos de datos **categoricos** a los datos que pueden tomar solo un conjunto específico y finito de valores que representan un conjunto de posibles categorías.

**1. Binario:** Un caso especial de datos categoricos con solo dos categorías de valores, por ejemplo, 0/1, verdadero Falso. (Sinónimos: dicotómico, lógico, indicador, booleano)

**2. Ordinal:** Datos categoricos que tienen un ordenamiento explícito. (Sinónimo: factor ordenado)

### 1.2.3 Variables cuantitativas o numéricas

Las variables cuantitativas, también conocidas como variables numéricas, son aquellas que pueden medirse y expresarse en términos numéricos. Estas variables permiten realizar operaciones matemáticas y estadísticas, y se dividen en dos tipos principales: variables discretas y variables continuas. A continuación, se presenta una descripción de cada tipo:

**Discretas:** Toman valores enteros y representan cantidades contables. Ejemplo: número de hijos, número de consultas médicas.

**Continuas:** Pueden tomar cualquier valor dentro de un intervalo y representan medidas. Ejemplo: peso, altura, tiempo.

**Intervalo:** No tienen un cero absoluto. Ejemplo: temperatura en grados Celsius.

**Razón:** Tienen un cero absoluto. Ejemplo: edad, peso en kilogramos.

#### Aspectos Clave a Destacar

Escalas de medición: La escala de medición determina las operaciones matemáticas que se pueden realizar con los datos y las pruebas estadísticas apropiadas.

Dicotomización: Tanto las variables nominales como las ordinales pueden dicotomizarse, es decir, reducirse a dos categorías.

Discretización: Las variables continuas pueden discretizarse, pero se pierde información.

Importancia del cero: El cero absoluto es fundamental para determinar si una variable cuantitativa es de intervalo o de razón.

La clasificación de las variables es un primer paso fundamental en cualquier análisis estadístico. Comprender las características de cada tipo de variable permite seleccionar las técnicas estadísticas adecuadas y realizar una interpretación correcta de los resultados.

### 1.2.4 Exploración de datos

La exploración de datos es el siguiente paso fundamental en un análisis estadístico después de identificar las variables porque permite comprender mejor la estructura, las características y las relaciones dentro del conjunto de datos antes de aplicar cualquier método estadístico o modelo más avanzado. Este paso es crucial por varias razones:

1. Comprensión de la Distribución de Datos Visualización de Distribuciones: Ayuda a visualizar cómo se distribuyen las variables cuantitativas (por ejemplo, a través de histogramas o gráficos de densidad). Esto permite identificar si los datos siguen una distribución normal, si están sesgados o si presentan cualquier patrón específico.

2. Identificación de Anomalías

Valores Atípicos (Outliers): La exploración de datos facilita la detección de valores atípicos que podrían distorsionar los resultados del análisis. Los outliers pueden identificarse mediante gráficos como los boxplots o análisis z-score.

Valores Faltantes: Permite identificar patrones de datos faltantes que deben ser abordados antes de proceder con el análisis, ya sea mediante imputación, eliminación de registros incompletos o análisis específico de los valores faltantes.

3. Detección de Relaciones Entre Variables Análisis de Correlación: La exploración inicial permite analizar la relación entre las variables numéricas mediante una matriz de correlación o gráficos de dispersión. Esto ayuda a entender cómo las variables pueden influirse mutuamente. Comparación entre Grupos: En el caso de variables categóricas, se pueden comparar distribuciones o promedios de variables numéricas entre diferentes grupos (por ejemplo, comparando los niveles de glucosa en sangre entre diferentes categorías de diagnóstico).
4. Confirmación de Suposiciones

Verificación de Suposiciones Estadísticas: Muchos métodos estadísticos suponen

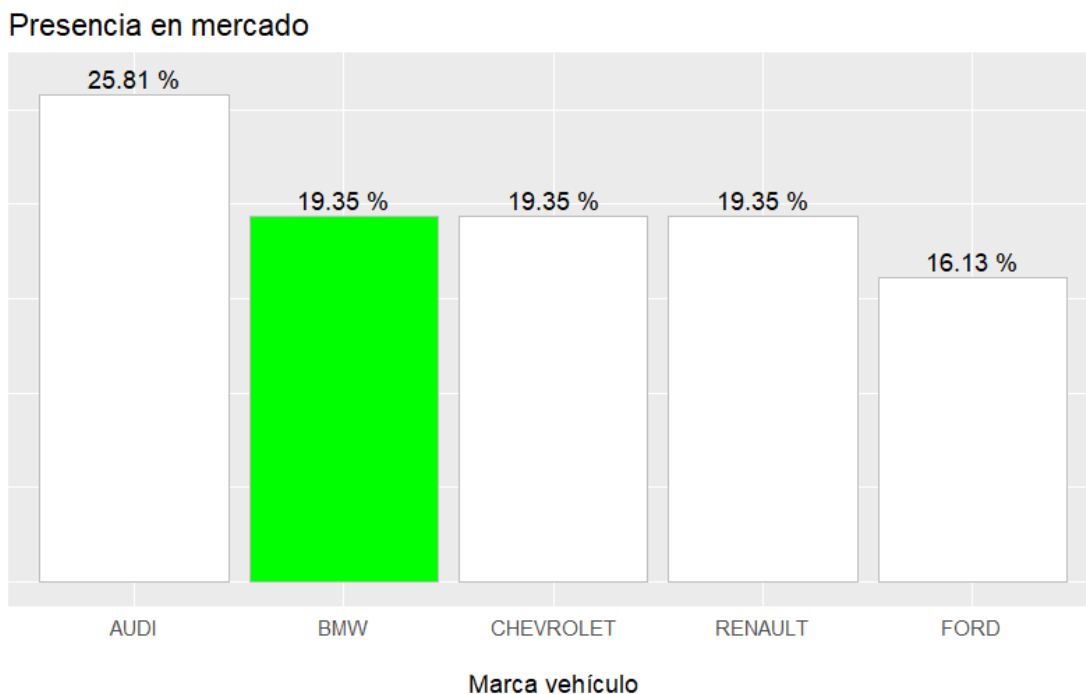


Figura 1.1: Histograma data(cars), agrupadas por marca.

normalidad, homocedasticidad (varianza constante), o independencia de las observaciones. La exploración de datos permite verificar estas suposiciones antes de aplicar técnicas analíticas, como la regresión lineal o ANOVA.

#### 5. Preparación de los Datos

Transformación de Variables: A veces, la exploración revela la necesidad de transformar variables (por ejemplo, aplicando logaritmos a variables con alta asimetría) para cumplir con los supuestos estadísticos o mejorar la interpretabilidad.

Creación de Nuevas Variables: La exploración puede sugerir la creación de nuevas variables derivadas o la combinación de variables existentes que podrían ser más relevantes para el análisis.

#### 6. Generación de Hipótesis

Formulación de Hipótesis: Basado en los patrones observados durante la exploración, los analistas pueden formular hipótesis sobre las relaciones o diferencias significativas en los datos, que luego pueden ser formalmente probadas en etapas posteriores del análisis.

#### 7. Guía para el Análisis Posterior

Selección de Métodos Estadísticos: La exploración de datos proporciona una base sólida para elegir las técnicas estadísticas más adecuadas para el análisis, como la regresión, análisis de componentes principales, o clustering.

Priorización de Variables: Ayuda a identificar cuáles variables son más relevantes

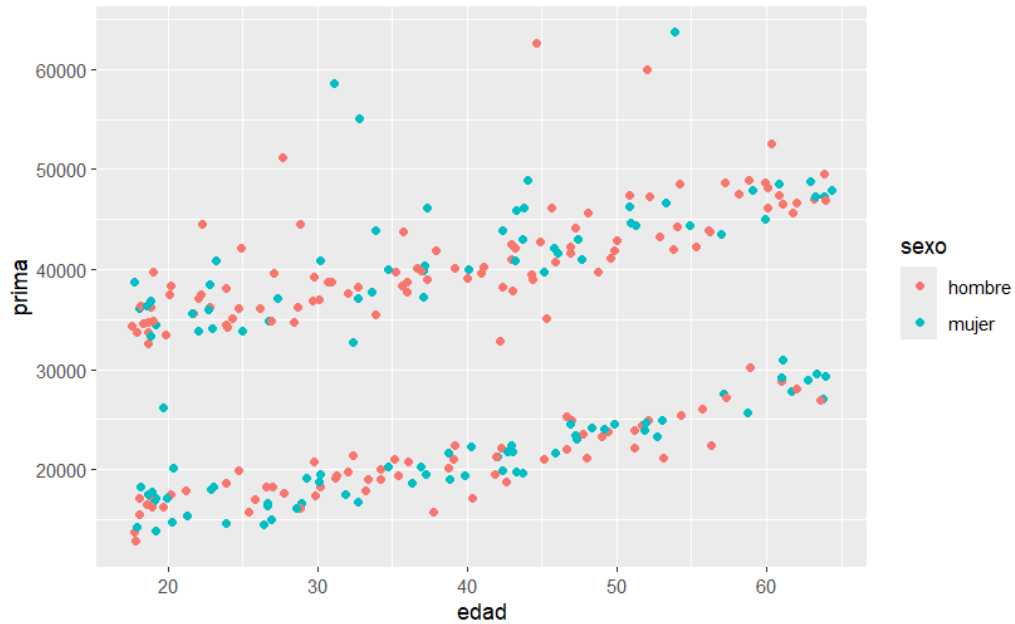


Figura 1.2: Prima de aseguramiento con relación a la edad

para el análisis y cuáles podrían ser redundantes o menos informativas.

Para realizar una exploración de datos en R, es esencial conocer cómo importar, limpiar, y manipular datos, junto con el uso de estadísticas descriptivas y visualizaciones. Debes dominar paquetes clave como *dplyr* para manipulación eficiente, y *ggplot2* para crear gráficos que revelen patrones y outliers. Además, es crucial comprender conceptos básicos de estadística y cómo identificar relaciones entre variables mediante correlaciones y gráficos de dispersión. Estos conocimientos te permitirán preparar los datos adecuadamente y obtener insights valiosos antes de realizar análisis más complejos.

## 1.3 Medidas de tendencia central

Las medidas de tendencia central son estadísticas que describen el punto medio o el centro de una distribución de datos. Estas medidas son fundamentales en estadística ya que permiten resumir grandes conjuntos de datos en un valor representativo, facilitando así la interpretación y comparación entre diferentes grupos de datos.

### 1. Media Aritmética

Definición: La media aritmética es el promedio de un conjunto de valores. Se calcula sumando todos los valores y dividiendo entre el número total de observaciones.

Fórmula:

Ejemplo: Supongamos que tenemos los valores de las edades de cinco personas: 23, 27, 31, 22 y 26 años. La media aritmética sería:

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$$

Interpretación: La media aritmética de 25.8 años indica que, en promedio, las personas en este grupo tienen 25.8 años.

## 2. Mediana

Definición: La mediana es el valor que divide a un conjunto de datos ordenados en dos partes iguales. Es menos sensible a valores extremos o outliers en comparación con la media. Cálculo:

Para un número impar de observaciones: Es el valor en la posición central.

$$\text{Mediana} = x_{(\frac{n+1}{2})}$$

Para un número par de observaciones: Es el promedio de los dos valores centrales.

$$\text{Mediana} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$$

Interpretación: La mediana indica que la mitad de los datos .

## 3. Moda

Definición: La moda es el valor que aparece con mayor frecuencia en un conjunto de datos. Un conjunto de datos puede ser unimodal (una sola moda), bimodal (dos modas) o multimodal (más de dos modas).

Ejemplo: Si en un grupo de edades tenemos los valores 22, 22, 23, 26, 26, 26, 27, 31, la moda es 26 años porque aparece más veces que cualquier otro valor. Interpretación: La moda de 26 años indica que la mayoría de las personas en este grupo tienen esa edad.

4. Comparación entre Medidas Media vs. Mediana: La media es afectada por valores extremos, lo que puede hacerla menos representativa en distribuciones asimétricas. La mediana, al no ser sensible a los extremos, es una mejor medida de tendencia central cuando hay outliers o datos muy sesgados. Media vs. Moda: La moda es útil para datos categóricos o cuando se desea conocer el valor más común, pero no proporciona información sobre la distribución general como lo hacen la media o la mediana.

Mediana vs. Moda: La mediana ofrece una visión del centro de la distribución que puede ser más útil en datos numéricos, mientras que la moda es más relevante en contextos donde la frecuencia de un valor específico es de interés.



### 5. Aplicación en Bioestadística

Ejemplo en Salud: En un estudio sobre los niveles de colesterol en pacientes, la media puede proporcionar una idea general de los niveles promedio, mientras que la mediana puede ser más informativa si los datos son asimétricos (por ejemplo, si algunos pacientes tienen niveles extremadamente altos).

Consideración de Outliers: En estudios clínicos, como la medición de presión arterial, la mediana puede ser preferida para evitar que resultados extremos influyan demasiado en la representación de los datos.

## 1.4 Medidas de Variabilidad

Las medidas de variabilidad son estadísticas que describen la dispersión o el grado de extensión de los datos alrededor de la medida de tendencia central. Estas medidas proporcionan información sobre la distribución y la dispersión de los datos, lo que ayuda a comprender la consistencia o variabilidad en el conjunto de datos.

**1. Rango** Definición: El rango es la diferencia entre el valor máximo y el valor mínimo de un conjunto de datos.

Fórmula:  $Rango = Valor\ Máximo - valor\ Mínimo$

Ejemplo: Para el conjunto de datos 3, 7, 8, 5, 12:

Rango =  $12 - 3 = 9$

Rango=12-3=9

Interpretación: El rango de 9 indica la amplitud total de los datos, pero no refleja la distribución interna.

### 2. Varianza

Definición: La varianza mide el promedio de las desviaciones al cuadrado respecto a la media. Proporciona una idea de cuánto se dispersan los datos alrededor de la media.

Fórmula:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

Para una muestra (con corrección de Bessel):

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Ejemplo: Para el conjunto de datos 4, 8, 6, 5, 10, la varianza se calcula primero encontrando la media y luego promediando los cuadrados de las diferencias entre cada dato y la media. Interpretación: Una mayor varianza indica una mayor dispersión de los datos alrededor de la media.

### 3. Desviación Estándar

Definición: La desviación estándar es la raíz cuadrada de la varianza. Es una medida de dispersión que indica cuánto varían los datos en promedio respecto a la media.

Fórmula:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}$$

Para una muestra:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Ejemplo: Usando el conjunto de datos del ejemplo anterior, si la varianza es 8.2, entonces la desviación estándar sería:  $\sigma = \sqrt{8.2} = 2.87$

Interpretación: La desviación estándar proporciona una medida directa de la dispersión en las mismas unidades que los datos originales.

### 4. Coeficiente de Variación

Definición: El coeficiente de variación es una medida de la dispersión relativa que se expresa como un porcentaje de la media. Es útil para comparar la variabilidad entre conjuntos de datos con diferentes unidades o magnitudes.

Fórmula:

$$CV = \frac{\sigma}{\mu} \times 100\%$$

Ejemplo: Si la desviación estándar de un conjunto de datos es 2.87 y la media es 25.8, el coeficiente de variación sería:

$$CV = \frac{2.87}{25.8} \times 100\% = 11.15$$

Interpretación: Un coeficiente de variación del 11.1% indica que la desviación estándar es aproximadamente el 11.1% de la media, proporcionando una visión de la variabilidad relativa.

### 5. Interquartil (IQR)

Definición: El rango intercuartil es la diferencia entre el tercer cuartil (Q3) y el primer cuartil (Q1) y mide la dispersión de la mitad central de los datos. Fórmula:

Ejemplo: Si Q1 es 4 y Q3 es 8 en un conjunto de datos, el IQR sería:

Interpretación: El IQR indica la amplitud de los datos en el rango medio y es resistente a los valores extremos.

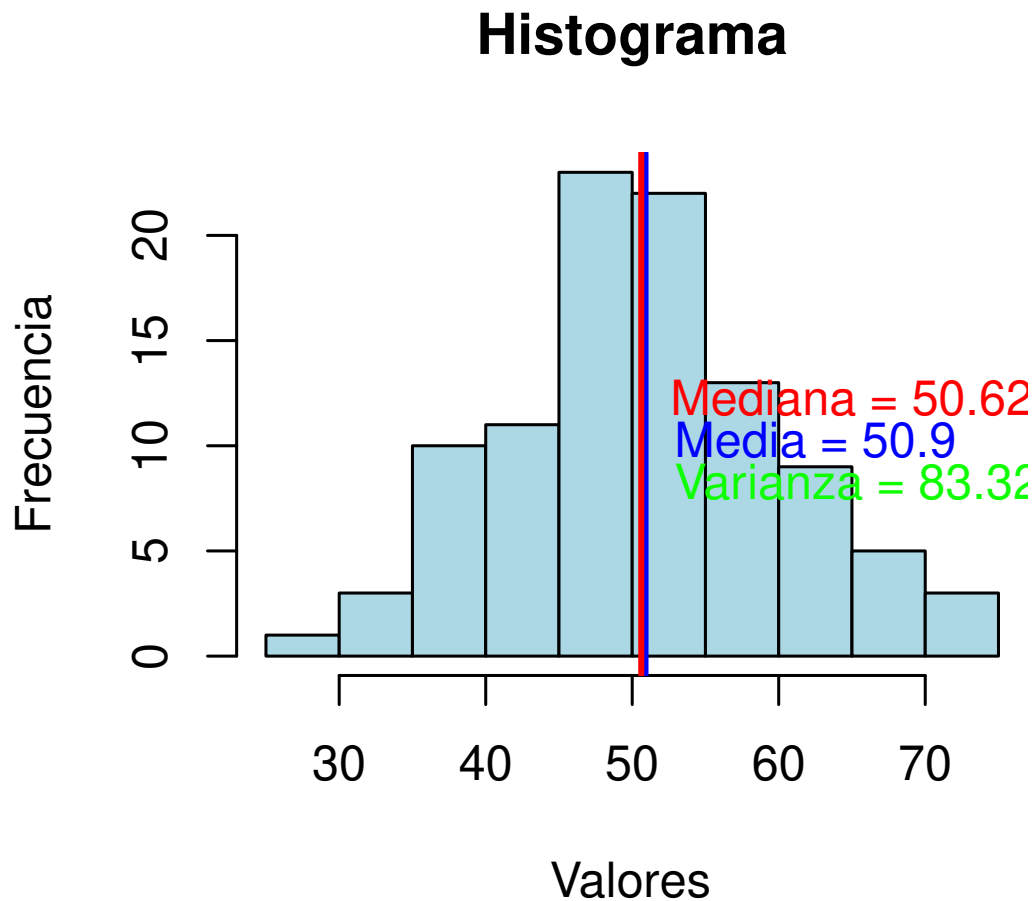


Figura 1.3: Histograma con Medidas de Tendencia Central y Variabilidad ejemplo

### Conclusión

Las medidas de variabilidad proporcionan una visión más completa de la dispersión de los datos en torno a la medida de tendencia central. Entender estas medidas es crucial para interpretar la consistencia, la amplitud y la distribución de los datos en un conjunto de datos, lo cual es esencial para la toma de decisiones y el análisis en diversas áreas, incluida la bioestadística.

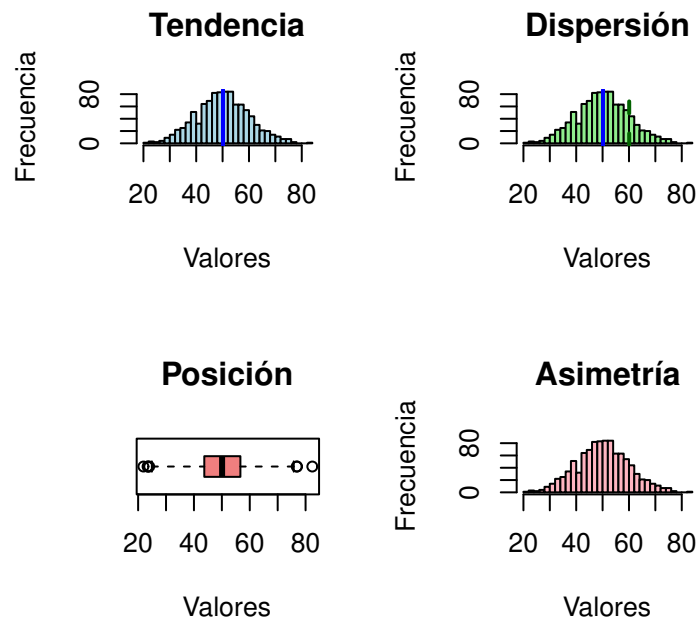


Figura 1.4: Histograma con Medidas de Tendencia Central y Variabilidad ejemplo

### 1.4.1 Actividad de comprobación trabajo autonomo

- Para la aplicación de los contenidos presentados en el presente documento utilizaremos el R Markdown es un archivo **.Rmd**, siga las instrucciones generado nuevos **Chunks** para su desarrollo.

# Bibliografía

- [1] Rius Diaz, F., & Baron Lopez, F. J. (2005). Bioestadística. Thomson-Paraninfo..
- [2] Martagón, V. L. (2014). Bioestadística. Editorial El Manual Moderno.