Ejemplos de medidas que resumen la información

1. En un conjunto residencial con 30 condominios, el número de personas que habita en cada uno es:

1	6	3	2	2	4	5	3	1	1
2	5	7	6	2	4	3	4	5	4
6	1	3	1	4	5	4	3	3	2

Calcular lo siguiente:

a. Distribución de frecuencia de los únicos números
b. Medidas de tendencia central: media aritmética, mediana y moda
c. Medidas de dispersión: rango, varianza poblacional y desviación estándar.
d. Medidas de posición y forma

a. Distribución de frecuencia de los únicos números

Los únicos valores son: 1, 2, 3, 4, 5, 6, 7, cuya tabla de frecuencias es

Número de habitantes	1	2	3	4	5	6	7
Frecuencia	5	5	6	6	4	3	1
Frecuencia acumulada	5	10	16	22	26	29	30
Frecuencia relativa porcentual	16.67%	16.67%	20%	20%	13.33%	10%	3.33%

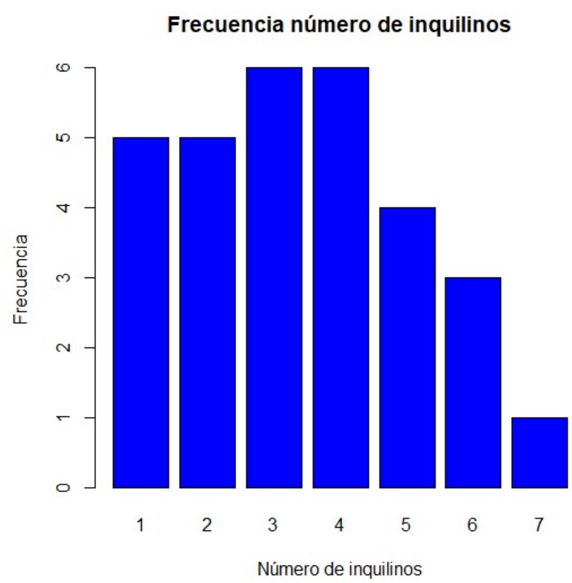


Figura 1. Histograma de frecuencias absolutas

La gráfica de frecuencias permite visualizar la asimetría: en este caso, se observa una mayor concentración de valores entre 1 y 4, lo cual implica que la mayoría de los condominios tienen pocos habitantes.

b. **Medidas de tendencia central**

i. Media aritmética

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1 + 6 + \cdots + 3 + 2}{30} = \frac{102}{30} = 3.4$$

En promedio, en cada condominio viven aproximadamente 3 o 4 personas. Sin embargo, dado que el número de personas debe ser entero, esta media (3.4) indica una tendencia general más que un valor exacto aplicable a este caso.

ii. Mediana

Se colocan los datos de manera ordenada

1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, **3**, **3**, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 7

La mediana es el promedio de los dos números medios (valores resaltados de color amarillo y en negrita).

$$Me = \frac{x_k + x_{k+1}}{2} = \frac{3 + 3}{2} = 3$$

iii. Moda

Con base en la tabla de frecuencias, existen 2 modas: **3 y 4**.

La situación anterior, donde hay dos valores con igual frecuencia máxima, se conoce como una **distribución bimodal**, e indica la presencia de dos subgrupos diferentes en la población.

c. **Medidas de dispersión**

i. Rango

$$rango = x_{max} - x_{min} = 7 - 1 = 6$$

ii. Varianza poblacional

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = \frac{85.2}{30} = 2.84$$

índice	Número inquilinos	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	índice	Número inquilinos	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	1	-2.4	5.76	16	4	0.6	0.36
2	6	2.6	6.76	17	3	-0.4	0.16
3	3	-0.4	0.16	18	4	0.6	0.36
4	2	-1.4	1.96	19	5	1.6	2.56
5	2	-1.4	1.96	20	4	0.6	0.36
6	4	0.6	0.36	21	6	2.6	6.76
7	5	1.6	2.56	22	1	-2.4	5.76
8	3	-0.4	0.16	23	3	-0.4	0.16
9	1	-2.4	5.76	24	1	-2.4	5.76
10	1	-2.4	5.76	25	4	0.6	0.36
11	2	-1.4	1.96	26	5	1.6	2.56
12	5	1.6	2.56	27	4	0.6	0.36
13	7	3.6	12.96	28	3	-0.4	0.16
14	6	2.6	6.76	29	3	-0.4	0.16
15	2	-1.4	1.96	30	2	-1.4	1.96

iii. Varianza muestral

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{85.2}{29} \approx 2.93$$

iv. Desviación estándar

$$s = \sqrt{s^2} = \sqrt{2.93} \approx 1.71$$

La desviación estándar de los datos, ~1.71, indica que, en promedio, el número de habitantes por condominio se desvía ±2, se toma el valor entero más cercano debido a lo que representan los datos analizados, personas respecto a la media (3.4).

Coeficiente de asimetría de Pearson

A partir de las medidas de tendencia central media y moda, además de la desviación estándar, se puede calcular el coeficiente de asimetría de Pearson:

$$\frac{\bar{x} - moda}{s} = \frac{3.4 - 4}{1.71} = \frac{-0.6}{1.71} = -0.35$$

El coeficiente es negativo, por lo tanto, la distribución de los datos está sesgada a la izquierda (cola larga a la izquierda), esto es, hay más condominios con un número de habitantes menor al promedio. Lo anterior se puede corroborar en la Figura 1.

Un valor negativo de asimetría indica que la cola de la distribución se extiende más hacia los valores bajos, mientras que una asimetría positiva indica una cola más larga hacia valores altos.

Finalmente, existen otros coeficientes de asimetría como el de **Fisher** o de **Bowley**, que se utilizan con mayor precisión en grandes muestras o con software estadístico.

d. Medidas de posición y forma

Las medidas de posición permiten ubicar observaciones dentro del conjunto de datos, facilitando la interpretación de "cuánto es mucho o poco" respecto al grupo completo. Para calcular las siguientes medidas, se usan los datos ordenados.

i. Cuartiles

Q_1	Mediana de la primera mitad de valores	1, 1, 1, 1, 1, 2, 2, **2**, 2, 2, 3, 3, 3, 3, 3
Q_2	Mediana de todos los valores	3
Q_3	Mediana de la segunda mitad de valores	3, 4, 4, 4, 4, 4, 4, **5**, 5, 5, 5, 6, 6, 6, 7

A continuación, se presentan los cuartiles, mínimo y máximo:

$$[x_{min}, Q_1, Me, Q_3, x_{max}] = [1, 2, 3, 5, 7]$$

ii. Deciles

Los deciles son 0.1, 0.2, …, 0.9. El quinto decil coincide con la mediana y con el segundo cuartil, $D_5 = Me = Q_2$. Para calcular la posición de los deciles, se emplea la siguiente fórmula:

$$\frac{k(n+1)}{10}, \quad k = 1, \dots, 9$$

El resultado de la fórmula anterior es un número decimal, por lo que se deben tener en cuenta dos escenarios:

- Si el resultado es un número entero, el decil es el dato que está en esa posición.
- Si el resultado es un número decimal, el valor del decil se ajusta con la siguiente fórmula

$$D = x_i + d(x_{i+1} - x_i)$$

Los cálculos para los deciles 10 y 90 son:

Decil	Cálculos
10	$\frac{1(30+1)}{10} = \frac{31}{10} = 3.1$ Dado que el resultado es un número decimal se aplica la fórmula 2 $D_1 = x_i + d(x_{i+1} - x_i) = 1 + 1(1-1) = 1$
90	$\frac{9(30+1)}{10} = \frac{31}{10} = 27.9$ Dado que el resultado es un número decimal se aplica la fórmula 2 $D_9 = x_{27} + d(x_{28} - x_{27}) = 6 + 9(6-6) = 6$

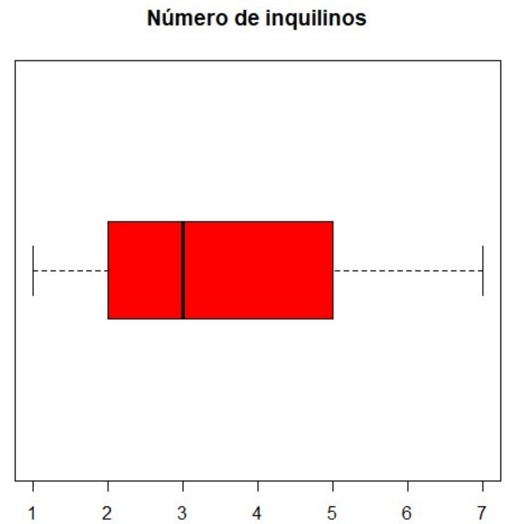
Aquellos condominios que tienen 6 o más habitantes, están en el percentil 90, es decir, el 90% de los condominios tienen 6 habitantes o menos.

iii. Diagrama de caja y bigote

Rango intercuartílico (IQR)

$$IQR = Q_3 - Q_1 = 5 - 2 = 3$$

El IQR se considera el 50% central de los datos, además se emplea para detectar *outliers* en los datos de una variable.



El boxplot permite identificar posibles valores atípicos si algún valor está por debajo de $Q_1 - 1.5 \cdot IQR$ o por encima de $Q_3 + 1.5 \cdot IQR$.

En el gráfico anterior no se identifican *outliers* para el número de habitantes.

Media armónica

La media armónica se emplea cuando los datos representan tasas, relaciones o relaciones reciprocas. Ejemplos de este tipo de datos se encuentran principalmente en finanzas, física e ingeniería.

2. Una empresa realizó la misma inversión a 4 productos, pero sus ganancias son diferentes:

Producto	Inversión	Ganancia	Proporción
A	$ 5000	$ 500	10
B	$ 5000	$ 1000	5
C	$ 5000	$ 2500	2
D	$ 5000	$ 4000	1.25

$$H = \frac{n}{\sum_{i=1}^{n} \frac{n_i}{x_i}} = \frac{4}{\frac{1}{10} + \frac{1}{5} + \frac{1}{2} + \frac{1}{1.25}} = \frac{4}{\frac{16}{10}} = 2.5$$

La media aritmética es:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{18.25}{4} = 4.56$$

Para validar que media es más adecuada, se calcula la cartera real:

- Inversión total: $ 20000
- Total de ganancias: $ 8000
- Cartera real: 20000/8000 = 2.5

El valor de la media armónica es igual al valor de la cartera real, al asignar menos peso al valor extremo 10. Además, la media armónica proporciona una estimación más precisa de la relación entre inversión y ganancia cuando se distribuyen fondos de forma uniforme en opciones con rendimientos muy dispares.

¿Cuándo usar la media armónica?

- Cuando los cálculos se basan en tasas
- Es menos sensible a valores atípicos (*outliers*)
- Brinda promedios precisos para relaciones reciprocas

Desventajas

- No es adecuada para un conjunto de datos que tienen 0
- Es sensible a valores pequeños
- No es intuitivo para audiencias no técnicas

Cuando se promedian **tasas** o **razones** sobre un denominador constante (por ejemplo, misma distancia o inversión), la media armónica es preferible a la media aritmética.

Ejercicios.

a. Calcular la media armónica para un automóvil que recorre 50 km con una velocidad de 60 km/h y los siguientes 50 km a una velocidad de 120 km/h. ¿Cuál es la velocidad promedio?
b. Un ciclista recorre 3 tramos iguales de 10 km a velocidades de 15 km/h, 20 km/h y 30 km/h, ¿cuál es la velocidad promedio del ciclista?

Media geométrica

La media geométrica es la n-raíz del producto de n números, los cuales pueden ser porcentajes o tasas de crecimiento. Está medida permite comparar valores que cambian con el tiempo, tal como el crecimiento demográfico o el rendimiento de inversiones. Su fórmula es

$$\bar{x}_G = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

3. Una persona contrato un fondo de inversión con una inversión inicial de $100,000. Las ganancias durante los primeros 4 años son:

Año	Monto ($)	Factor de crecimiento
0	100,000.00	
1	100,500.00	1.005
2	115,500.00	1.149
3	136,290.00	1.18
4	156.733.50	1.15

$$\bar{x}_G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[4]{(1.005 \cdot 1.149 \cdot 1.18 \cdot 1.15)} = \sqrt[4]{1.5669} \approx 1.1188$$

El valor 1.1188 indica que, en promedio, la inversión creció un **11.88% anual**, considerando el efecto compuesto de los 4 años. Así, el crecimiento de la inversión en 4 años se puede calcular de manera aproximada, como:

$$100{,}000 \cdot 1.1188^4 = 156{,}678.65$$

La media geométrica es útil cuando las tasas **se acumulan de forma multiplicativa**, como en intereses compuestos, inflación acumulada, crecimiento poblacional o retornos anuales sobre fondos.

Ejercicios

a) Calcular la media geométrica de los índices de inflación de 5 años, los cuales fueron 1.02, 1.05, 1.01, 1.04 y 1.03, ¿cuál es la inflación promedio anual?
b) Una ciudad tuvo crecimientos anuales del 5%, 3%, 4% y 2% en los últimos 4 años. Calcular su tasa promedio de crecimiento.

Referencias

1. https://www.geeksforgeeks.org/r-bar-charts/
2. https://medium.com/swlh/visualizing-the-geometric-and-harmonic-means-e8b9c5a818ae