

Sesión 3: Datos Agrupados

Módulo 1: Estadística Descriptiva

Magdalena Cornejo

Datos agrupados

- En la práctica, la mayor parte de los conjuntos de datos contienen muchas observaciones, por lo que cuando la cantidad de información es muy grande resulta conveniente reducirla agrupando las observaciones en intervalos o **tablas de frecuencia**.
- Muchas veces, no solo se representan los datos en forma agrupada para reducir la cantidad de información sino también porque hay variables que son sensibles a la no-respuesta.
- Es decir, es más probable que una persona en una encuesta conteste cierta información cuando se le pide que se ubique en un rango, que cuando se le pide el valor exacto (ejemplo: nivel de ingreso).

Ejemplo

La siguiente tabla tiene datos de ingreso mensual en US\$ de 3200 individuos en cierta ciudad:

Ingresos	Nro de individuos
(150-450]	250
(450-750]	800
(750-1050]	1250
(1050-1350]	700
(1350-1650]	180
(1650-1950]	16
(1950-2250]	3
(2250-2550]	1
Suma	3200

El paréntesis excluye el valor y el corchete incluye el valor.

Frecuencias

Definición

Se llama **frecuencia** al número de observaciones de cada intervalo.

Definición

La **frecuencia acumulada** es el número total de observaciones que hay en ese intervalo y en los anteriores.

Ingresos	Frecuencia	Frec. Acumulada
(150-450]	250	250
(450-750]	800	1050
(750-1050]	1250	2300
(1050-1350]	700	3000
(1350-1650]	180	3180
(1650-1950]	16	3196
(1950-2250]	3	3199
(2250-2550]	1	3200
Suma	3200	

Frecuencia relativa

Definición

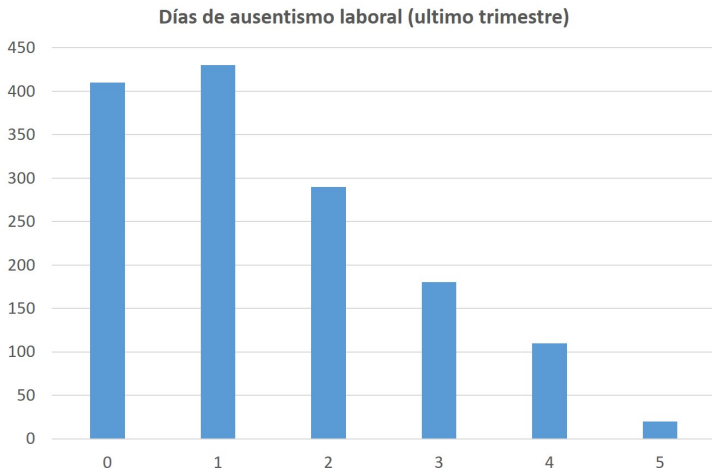
La proporción de observaciones de cada intervalo o clase se denomina **frecuencia relativa**.

Días de ausentismo laboral en el último trimestre:

Nro de días	Nro de empleados	%
0	410	28.5
1	430	29.9
2	290	20.1
3	180	12.5
4	110	7.6
5	20	1.4
Total	1440	100.0

Histograma

El **histograma** nos permite hacernos una idea visual rápida y adecuada de la proporción de observaciones que se encuentran dentro de un determinado intervalo.



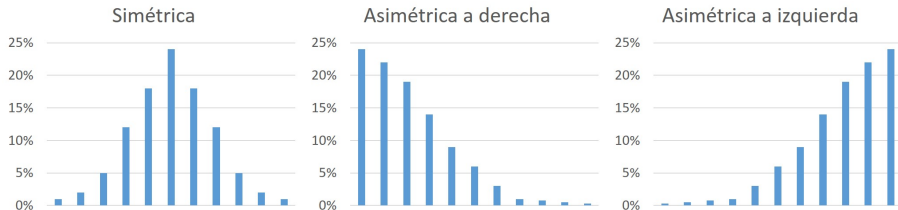
Histograma

Un histograma sirve para:

- Tener una primera vista de los datos, cómo se distribuyen.
- Detectar casos extremos.
- Detectar problemas con los datos.
- Ver qué es lo que sucede más frecuentemente.

Asimetría

La distribución de los datos puede ser **simétrica** o **asimétrica**.



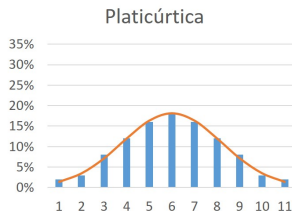
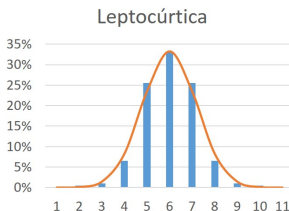
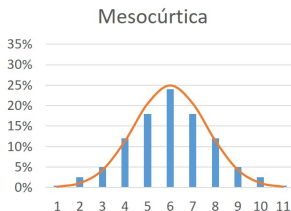
Cuando la distribución es:

- **simétrica:** $\text{Media} = \text{Mediana} = \text{Moda}$
- **asimétrica a derecha (positiva):** $\text{Media} > \text{Mediana} > \text{Moda}$
- **asimétrica a izquierda (negativa):** $\text{Media} < \text{Mediana} < \text{Moda}$

Curtosis

Definición

La **curtosis** mide el grado de concentración que presentan los valores en la región central de la distribución.



Usando como referencia la distribución normal, la distribución puede ser:

- Platicúrtica
- Leptocúrtica
- Mesocúrtica (la distribución normal)

Aplicación en Excel

- Bajar de **Yahoo Finance** (finance.yahoo.com) el precio de cierre ajustado de Intel entre el 4 de enero de 2016 y el 22 de julio de 2016.
- Compute el retorno diario de Intel a través de la siguiente fórmula:

$$R_t = \frac{(P_t - P_{t-1})}{P_{t-1}} \times 100 = \left(\frac{P_t}{P_{t-1}} - 1 \right) \times 100$$

- Utilice el complemento de Excel “Herramienta para análisis” obtenga los principales estadísticos descriptivos.
- Genere una tabla de frecuencias y grafique el histograma.
- ¿Qué observa?