

1.4. Descripción de variable: distribución de frecuencias y representación gráfica

Para ello debemos tener claro los siguientes conceptos:

- **Frecuencia absoluta (n_i):** Es el número de veces que se repite cada uno de los valores de las variables. La suma de todas las frecuencias absolutas es el total de la muestra (n).
- **Proporción o frecuencia relativa (p_i):** Es el cociente entre la frecuencia absoluta de cada valor y el número de observaciones. $p_i = n_i/n$
- **Porcentaje (P_i):** Es el valor de la frecuencia relativa multiplicado por 100.
- **Frecuencia absoluta acumulada (n_a):** Es el número de veces que se repiten cada uno de los valores o cualquiera de los valores inferiores.
- **Proporción acumulada o frecuencia relativa acumulada (p_a):** Es el cociente entre la frecuencia absoluta acumulada y el total de observaciones. $p_a = n_a/n$
- **Porcentaje acumulado (P_a):** Es el valor de la frecuencia relativa acumulada multiplicado por 100.



1.4. Descripción de variable: distribución de frecuencias y representación gráfica

Para ello debemos tener claro los siguientes conceptos:

- **Frecuencia absoluta (n_i):** Es el número de veces que se repite cada uno de los valores de las variables. La suma de todas las frecuencias absolutas es el total de la muestra (n).
- **Proporción o frecuencia relativa (p_i):** Es el cociente entre la frecuencia absoluta de cada valor y el número de observaciones. $p_i = n_i/n$
- **Porcentaje (P_i):** Es el valor de la frecuencia relativa multiplicado por 100.
- **Frecuencia absoluta acumulada (n_a):** Es el número de veces que se repiten cada uno de los valores o cualquiera de los valores inferiores.
- **Proporción acumulada o frecuencia relativa acumulada (p_a):** Es el cociente entre la frecuencia absoluta acumulada y el total de observaciones. $p_a = n_a/n$
- **Porcentaje acumulado (P_a):** Es el valor de la frecuencia relativa acumulada multiplicado por 100.

20 familias

1 5 4 5 3 3 1 1 2 1
1 4 3 2 4 3 2 2 3 1

x_i	n_i	p_i	P_i	n_a	p_a	P_a
1	6	0,3	30	6	0,3	30
2	4	0,2	20	10	0,5	50
3	5	0,25	25	15	0,75	75
4	3	0,15	15	18	0,9	90
5	2	0,1	10	20	1	100

$$n = 20 \quad 1 \quad 100$$

$$p_i = \frac{6}{20}$$

$$P_a = \frac{n_a}{n} = \frac{18}{20}$$



¡¡¡IMPORTANTE!!!

$n = 20$

x_i	P_i	n_i
1	0,3	6
2	0,2	4
3	0,25	5
4	0,15	3
5	0,1	2

x_i	n_a	n_i
1	6	6
2	10	4
3	15	5
4	18	3
5	20	2

$$P_i = \frac{n_i}{n}$$

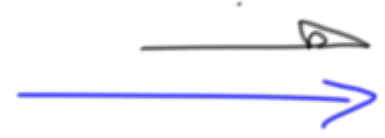
$$0,3 = \frac{n_i}{20}$$

$$0,3 \cdot 20 = n_i$$



¡¡¡IMPORTANTE!!!

x_i	n_i
5	2
4	3
3	0
2	5
1	0



x_i	n_i
1	0
2	5
3	0
5	5
0	2

- **Frecuencia absoluta (n_i):** Es el número de veces que se repite cada uno de los valores de las variables. La suma de todas las frecuencias absolutas es el total de la muestra (n).
- **Proporción o frecuencia relativa (p_i):** Es el cociente entre la frecuencia absoluta de cada valor y el número de observaciones. $p_i = n_i/n$
- **Porcentaje (P_i):** Es el valor de la frecuencia relativa multiplicado por 100.
- **Frecuencia absoluta acumulada (n_a):** Es el número de veces que se repiten cada uno de los valores o cualquiera de los valores inferiores.
- **Proporción acumulada o frecuencia relativa acumulada (p_a):** Es el cociente entre la frecuencia absoluta acumulada y el total de observaciones. $p_a = n_a/n$
- **Porcentaje acumulado (P_a):** Es el valor de la frecuencia relativa acumulada multiplicado por 100.

Si la variable con la que estamos trabajando se puede ordenar, podremos calcular todos los conceptos anteriores, en cambio, si es una variable que no se puede ordenar, solamente se calcularán los tres primeros.



VARIABLE DE CARÁCTER CUALITATIVO

En esta tabla tenemos una distribución de frecuencias de carácter cualitativo, en este caso la zona de procedencia de inmigrantes. En ella se resumen de manera organizada los datos correspondientes a 150 sujetos entrevistados.

X	n_i	p_i	P_i
Sudamérica	24	0,16	16
Norte África	60	0,40	40
Centro África	12	0,08	8
Europa del este	18	0,12	12
China	21	0,14	14
Sur Asia	15	0,10	10
	$n=150$	1	100

VARIABLE ORDINAL

En esta tabla nos encontramos con una distribución de frecuencias de una variable ordinal, sobre la preocupación por las cuestiones políticas. En este añadimos tres columnas más, las correspondientes a los valores acumulados.

X	n_i	p_i	P_i	n_a	p_a	P_a
Máxima	12	0,10	10	12	0,10	10
Moderada	48	0,40	40	60	0,50	50
Mínima	42	0,35	35	102	0,85	85
Nula	18	0,15	15	120	1,00	100
	n=120	1	100			

VARIABLE CUANTITATIVA CON POCOS DATOS

En este caso tenemos una distribución de frecuencias de una variable cuantitativa con pocos datos, nos da información sobre la calificación obtenida en el examen de Introducción al Análisis de datos de una muestra de 40 alumnos.

X	n_i	p_i	P_i	n_a	p_a	P_a
1	1	0,025	2,5	1	0,025	2,5
2	1	0,025	2,5	2	0,050	5,0
3	1	0,025	2,5	3	0,075	7,5
4	8	0,200	20	11	0,275	27,5
5	5	0,125	12,5	16	0,400	40,0
6	7	0,175	17,5	23	0,575	57,5
7	8	0,200	20	31	0,775	77,5
8	7	0,175	17,5	38	0,950	95,0
9	1	0,025	2,5	39	0,975	97,5
10	1	0,025	2,5	40	1,000	100
	n=40	1	100			

VARIABLE CUANTITATIVA CON GRAN CANTIDAD DE DATOS

Dentro de las variables cuantitativas podemos encontrarnos con algún caso en el que el número de valores que tome la variable sea muy amplio y será necesario agrupar la variables en intervalos, grupos de valores consecutivos de la variable.

Los intervalos tienen un valor inferior y un valor superior. Estos son los llamados límites aparentes. Tienen la misma unidad de medida que los valores de la variable.

El problema que tenemos con estos límites es que existe una pequeña discontinuidad, ya que el límite superior de un intervalo no coincide con el inferior del siguiente. Por ello, calcularemos los límites exactos.

Además de estos límites, también debemos calcular el punto medio del intervalo, que lo llamaremos marca de clase. Y con ello se harán los posteriores cálculos.

Y por último, otro de los conceptos a tener en cuenta es la amplitud del intervalo, que es la diferencia entre el límite superior exacto y el límite inferior exacto.





$$I < \begin{matrix} 0,5 \\ 0,5 \end{matrix}$$

$$60 - 0,5 / 69 + 0,5$$

VARIABLE CUANTITATIVA CON GRAN CANTIDAD DE DATOS

Dentro de las variables cuantitativas podemos encontrarnos con algún caso en el que el número de valores que tome la variable sea muy amplio y será necesario agrupar la variables en intervalos, grupos de valores consecutivos de la variable.

Los intervalos tienen un valor inferior y un valor superior. Estos son los llamados **límites aparentes**. Tienen la misma unidad de medida que los valores de la variable.

El problema que tenemos con estos límites es que existe una pequeña discontinuidad, ya que el límite superior de un intervalo no coincide con el inferior del siguiente. Por ello, calcularemos los **límites exactos**.

Además de estos límites, también debemos calcular el **punto medio** del intervalo, que lo llamaremos **marca de clase**. Y con ello se harán los posteriores cálculos.

Y por último, otro de los conceptos a tener en cuenta es la **amplitud del intervalo**, que es la diferencia entre el límite superior exacto y el límite inferior exacto.

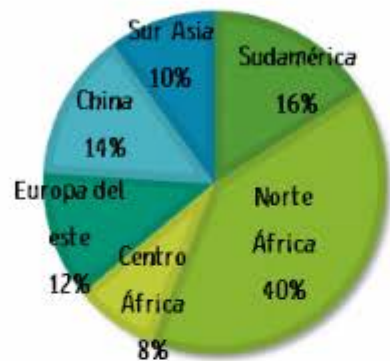
L.A	n _i	L.E	P. medio
60 - 69	4	59,5 - 69,5	64,5
70 - 79	20	69,5 - 79,5	74,5
80 - 89	40	79,5 - 89,5	84,5
90 - 99	100	89,5 - 99,5	94,5
100 - 109	250	99,5 - 109,5	...
110 - 119	115
120 - 129	25
130 - 139	10

$$P. medio : \frac{LS + LI}{2} = \frac{59,5 + 69,5}{2}$$



REPRESENTACIÓN GRÁFICA

DIAGRAMA DE SECTORES



HISTOGRAMA

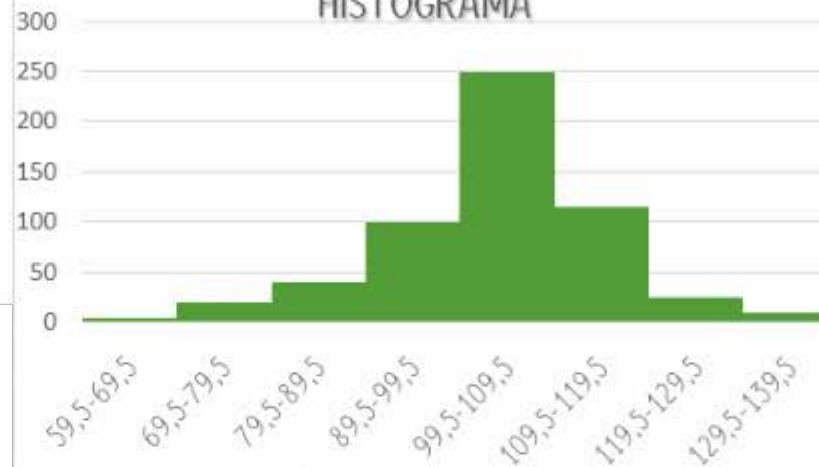


DIAGRAMA DE BARRAS

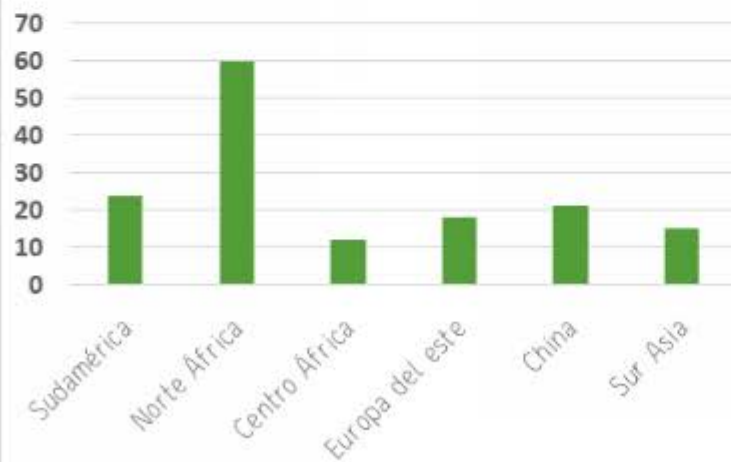


DIAGRAMA DE LÍNEAS



1.5. Tendencia central, variabilidad y forma de una variable: aproximación gráfica

En los temas siguientes se calcularán una serie de índices estadísticos que los utilizaremos para medir:

- La tendencia central: lugar donde se centra una distribución particular en la escala de valores.
- La variabilidad: Se refiere al grado de concentración de los valores entre sí o con respecto al valor central. Será homogénea si los valores están cercanos al promedio y será heterogénea si los valores se dispersan mucho.
- La forma: Se analizará la asimetría y la curtosis.

