

# Organización y Descripción de Datos

Un dato univariado es aquel que toma sólo un valor de una variable. El objetivo básico de la Estadística Descriptiva es describir un conjunto de datos. Los cuatro conceptos indispensables en la descripción de conjuntos de datos univariados son :

1. Medidas de Tendencia Central
2. Medidas de Dispersión o Variabilidad
3. Medidas de Posición
4. Tipos de Distribución

Una vez recopilados los datos muestrales, se puede utilizar una técnica gráfica que permita ver en dicha representación los patrones de comportamiento de la variable estudiada. Existen varios tipos de gráficas para describir datos. El método que se utilice estará determinado por el tipo de datos y lo que se quiere encontrar.

## 1.1. Gráficas de Barras, Circulares y Tallo-Hoja

La ordenación es el proceso mediante el cual los datos están acomodados de tal manera que se establece un orden (ascendente o descendente) entre ellos. Hay dos métodos comunes: listado en orden ascendente y método de tallo hoja.

Para precisar esta idea, se muestran los siguientes casos:

Considerando que la variable de estudio es el peso de 25 estudiantes. Los pesos se encuentran en la siguiente tabla:

Peso de estudiantes (kg)

40	43	48	51	49
56	44	42	55	52
52	62	44	50	59
63	50	56	55	45
57	66	63	51	58

### Listado Ascendente

Consiste en ordenar un conjunto de datos de menor a mayor. Ordenando los datos presentados anteriormente de forma ascendente, se obtiene

#### Peso de estudiantes (kg)

40	45	51	55	59
42	48	51	56	62
43	49	52	56	63
44	50	52	57	63
44	50	55	58	66

#### Representación Tallo-Hoja

Esta técnica es una combinación de los procedimientos gráficos y de ordenación. El tallo se forma con el (los) primer(os) dígito(s) del dato, mientras que la hoja se forma con los demás dígitos siguientes.

Ejemplo :

Construir una representación tallo- hoja para el siguiente conjunto de 20 calificaciones.

#### Calificaciones

82	74	88	66	58	74	78	84	96	76
62	68	72	92	86	76	52	76	82	78

Su representación tallo- hoja será:

5	8	2
6	6	2 8
7	4 4	8 6 2 6 6 8
8	2 8	4 6 2
9	6	2

## 1.2. Distribuciones de Frecuencias

Frecuencia : Numero de veces que ocurre un valor en una muestra.

Distribución de Frecuencia no Agrupada : Los valores no se combinan para formar grupos, sino que cada valor es un grupo en sí.

Ejemplo :

Utilice el siguiente conjunto de datos, para formar su distribución de frecuencia

3	2	2	3	2
4	4	1	2	2
4	3	2	0	2
2	1	3	3	1

Su distribución de frecuencia es :

$x$	$f$
0	1
1	3
2	8
3	5
4	3

Cuando un conjunto de datos tiene diferentes valores para la variable, se pueden agrupar los valores en un conjunto de clases.

Ejemplo :

Los siguientes datos reflejan las medidas de superficies de casas construidas por la Empresa Roba, en el reparto Mirador de Casa, ubicado en el Departamento de Managua.

27	68	79	91	107
43	71	80	91	108
43	71	81	93	108
44	71	82	94	116
47	73	82	94	120
49	73	84	94	120
50	74	84	96	122
54	75	86	97	123
58	76	88	103	127
65	77	88	106	128

Procedimiento para construir la distribución de frecuencia :

1. Obtener Rango o Recorrido ( $R$ ): Diferencia entre el mayor y menor valor
2. Seleccionar un numero de clase ( $m$ ) : Esto se puede realizar mediante la regla de Sturges, la cual es :

$$m = 1 + 3,3 \log n$$

donde  $n$  es la cantidad de datos analizados. Tambien se puede utlizar la fórmula :  $\sqrt[n]{n}$  para determinar el numero de clase, igualmente se puede usar la fórmula  $2^k$ , donde  $k \in \mathbb{Z}^+$  y se verifica que  $2^k > n$ , donde  $n$  es la cantidad de datos.

3. Amplitud : Esta representado por el cociente  $C = \frac{R}{m}$

Limite inferior de clase : Valor mas pequeño que puede asignarse en cada clase.

Limite superior de clase : Valor mas grande que puede asignarse en cada clase.

Limite de Clases Reales : Se construyen a partir de los limites de clase, restando 0.5 a limite de clase inferior y sumando 0.5 al limite de clase superior. Ademas se trabajan con intervalo de clase semicerrado.

Ancho de clase : Diferencia entre un limite inferior de clase y el limite inferior de la siguiente clase ( no es la diferencia entre los limites superior e inferior de la misma clase ).

Fronteras de clases : Son numeros que se localizan enmedio del limite superior de una clase y del limite inferior de la clase siguiente. La diferencia entre las fronteras inferior y superior de una clase se denomina tambien ancho de clase.

Marca de clase ( $x_i$ ) : Punto medio de una clase, que se obtiene como promedio aritmetico del limite inferior y superior de una clase.

Por tanto, su distribución de frecuencia es :

$$\text{Rango : } R = 128 - 27 = 101$$

$$\text{Numero de Clase : } m = \sqrt[1]{50} = 7,0711 \therefore 7$$

$$\text{Amplitud : } C = \frac{R}{m} = \frac{101}{7} = 14,429 \therefore 14$$

Frecuencia Relativa ( $f_r$ ) : La frecuencia relativa se obtiene mediante el cociente de la frecuencia y el numero total de datos. En algunos casos, es concebida como la medida proporcional de la frecuencia de un suceso. El correspondiente histograma se llama histograma de frecuencias relativas ( tambien se utilizan los limites de clases reales ).

Frecuencia Acumulada ( $N_i$ ): La frecuencia acumulada se obtiene sumando las frecuencias anteriores a las frecuencias de un dato dado.

Frecuencia Porcentual ( $f\%$ ) : Se obtiene multiplicando la frecuencia relativa " $f_r$ " por 100.

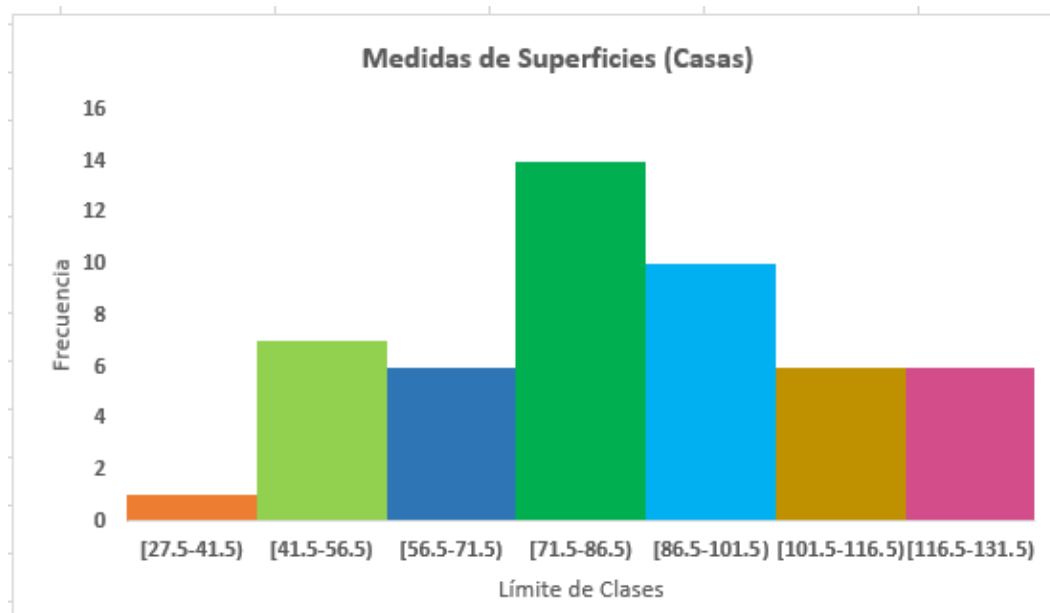
De donde, obtenemos :

<i>Limites de Clases Reales</i>	<i>Límites de Clase</i>	<i>f</i>	<i>Marca de Clase (<math>x_i</math>)</i>	<i><math>f_r</math></i>	<i><math>N_i</math></i>	<i>f %</i>
[27,5 - 41,5)	27 - 41	1	34	0,02	1	2 %
[41,5 - 56,5)	42 - 56	7	49	0,14	8	14 %
[56,5 - 71,5)	57 - 71	6	64	0,12	14	12 %
[71,5 - 86,5)	72 - 86	14	79	0,28	28	28 %
[86,5 - 101,5)	87 - 101	10	94	0,2	38	20 %
[101,5 - 116,5)	102 - 116	6	109	0,12	42	12 %
[116,5 - 131,5)	117 - 131	6	124	0,12	50	12 %
	<i>Total</i>	50		1		100 %

Histograma : Es un tipo particular de grafica de barra que representa a todo un conjunto de datos. Un histograma esta compuesto por:

1. Un titulo que identifica la población de interes
2. Una escala vertical que identifica las frecuencias en la distintas clases
3. Una escala horizontal que identifica a los limites reales de clases.

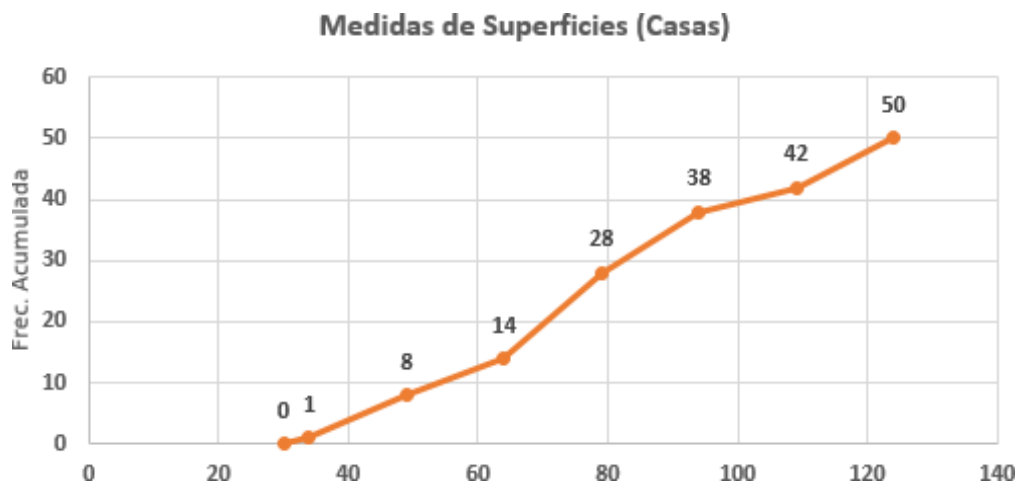
El grafico de histograma para el conjunto de datos que reflejan las medidas de superficies de casas construidas por la Empresa Roba, en el reparto Mirador de Casa, ubicado en el Departamento de Managua, es el siguiente:



Ojiva : Es una grafica de una distribución de frecuencia ( o de frecuencias relativas ) acumuladas. Una ojiva tiene los siguientes componentes :

1. Un titulo que identifica la población de interes.
2. Una escala vertical que identifica las frecuencias acumuladas o las frecuencias relativas acumuladas.
3. Una escala horizontal que identifica la marca de clase de cada clase creada.

Nota : Toda ojiva comienza con una frecuencia relativa igual a cero asociada a la frontera inferior de la primera clase y termina con una frecuencia relativa del 100 % asociada a la frontera superior de la ultima clase.



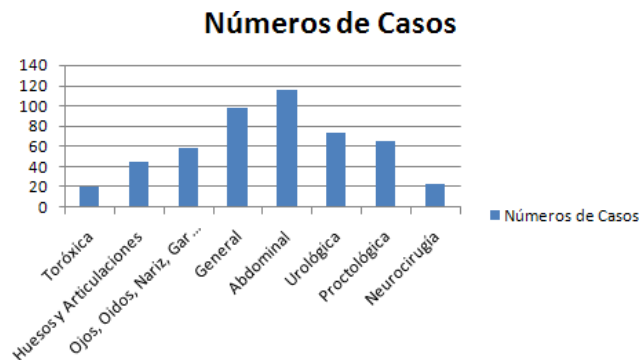
Gra.fico de Barra : Se realiza de modo similar al Histograma, sólo que las barras se construyen separadas unas de otras

Ejemplo:

La siguiente tabla muestra las operaciones realizadas en el hospital Roberto Calderón de la Ciudad de Managua, durante el primer mes del año de 1999.

Tipos de Operación	Numeros de Casos	$f_r$	$f_r * 360^0$
Torácica	20	0,04	14 <sup>0</sup>
Huesos y Articulaciones	45	0,09	33 <sup>0</sup>
Ojos, Oídos, Nariz, Garganta	58	0,12	42 <sup>0</sup>
General	98	0,20	71 <sup>0</sup>
Abdominal	115	0,23	83 <sup>0</sup>
Urológica	74	0,15	53 <sup>0</sup>
Proctológica	65	0,13	47 <sup>0</sup>
Neurocirugía	23	0,05	17 <sup>0</sup>
Total	498	1	360 <sup>0</sup>

0



Gra.fico Circular o de Pastel: Sólo se representan datos de frecuencias relativas o frecuencias porcentuales.

Además se debe dividir el área del círculo de manera proporcional a las frecuencias obtenidas.