



ELEMENTOS DE ESTADÍSTICA DESCRIPTIVA

Recurso compartido de la Web

Orden de datos

- 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
 - Método de tallo y hojas

Ejemplo

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

Peso de 25 estudiantes (en 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 en orden ascendente

- El proceso consiste en ordenarlos de menor a mayor

Peso de 25 estudiantes (en kg)				
42	40	48	51	49
56	44	43	55	52
52	62	44	50	59
63	50	56	55	45
57	66	63	51	58

Peso de 25 estudiantes (en kg)				
40	42	43	44	44
45	48	49	50	50
51	51	52	52	55
55	56	56	57	58
59	62	63	63	66

Método de tallo y hojas

- Si los números de los datos están formados por dos dígitos, se hace una columna con el primer dígito (decenas) y a la derecha de cada uno de ellos se escribe, en fila, sólo el segundo dígito (unidades) de cada uno de los datos que tengan el mismo primer dígito.



□ Datos sin ordenar:

4 2,0,8,9,4,3,4,5
5 1,6,5,2,2,0,9,0,6,5,7,1,8
6 2,3,6,3

□ Datos ordenados:

4 0,2,3,4,4,5,8,9
5 0,0,1,1,2,2,5,5,6,6,7,8,9
6 2,3,3,6

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

Doble tallo

- Una variante de este método es en lugar de dividir en un grupo las decenas, se divide en dos grupos. El primero abarcando los dígitos del 0 al 4 y el segundo del 5 al 9.

- El ejemplo anterior queda:

4	0,2,3,4,4
4	5,8,9
5	0,0,1,1,2,2,
5	5,5,6,6,7,8,9
6	2,3,3
6	6

Caso de variables cualitativas

- El procedimiento es:
 - ▣ Se identifican todos los valores diferentes y se acomodan en columna.
 - ▣ Se agrega una segunda columna en donde se van registrando, mediante una línea vertical, la veces que aparece el valor dado.

Ejemplo

- Considera que la variable de estudio es el color de playera de 25 estudiantes.

Los colores se encuentran en la siguiente tabla:


rosa	azul	blanco	azul	rosa
gris	blanco	café	negro	blanco
rosa	azul	café	blanco	blanco
gris	azul	blanco	rosa	gris
gris	blanco	café	negro	verde

rosa	azul	blanco	azul	rosa
gris	blanco	café	negro	blanco
rosa	azul	café	blanco	blanco
gris	azul	blanco	rosa	gris
gris	blanco	café	negro	verde

Color	Frecuencia
Azul	
Blanco	
Café	
Gris	
Negro	
Rosa	
Verde	

Tabla de Frecuencia de Datos

- Una vez que se tenga ordenados los datos, se acomodan en la “Tabla de distribución de frecuencias o tabla de frecuencias”.
- La tabla es básicamente una tabla de valores x - y , dónde “ x ” representa el dato y “ y ” representa la frecuencia.

- 
- La frecuencia es el número de veces que aparece cada dato.

 - Hay dos clases de tablas de frecuencias:
 - ▣ Para datos NO agrupados.
 - ▣ Para datos agrupados.

Tabla de frecuencias para datos NO agrupados

- Está formada por dos columnas: una para la **variable** “ x_i ” y la otra para su **frecuencia** “ f ”, a esta frecuencia se le llama frecuencia absoluta o frecuencia observada.

Ejemplo

- Tabla de frecuencias de los pesos en kg de 25 alumnos.

Peso de 25 estudiantes (en kg)				
40	42	43	44	44
45	48	49	50	50
51	51	52	52	55
55	56	56	57	58
59	62	63	63	66

xi	f
40	1
42	1
43	1
44	2
45	1
48	1
49	1
50	2
51	2

xi	f
52	2
55	2
56	2
57	1
58	1
59	1
62	1
63	2
66	1
Total	25

Frecuencia relativa y acumulada

- Por lo regular, se agregan dos columnas: la de la frecuencia relativa “fr” y la de la frecuencia acumulada “fa”.
- La frecuencia relativa se obtiene mediante el cociente de la frecuencia y el número total de datos, esto es $fr = f/n$.
- La frecuencia acumulada se obtiene sumando las frecuencias anteriores a las frecuencias de un dato dado.

Ejemplo

xi	f	fr	fa
40	1	0.04	1
42	1	0.04	2
43	1	0.04	3
44	2	0.08	5
45	1	0.04	6
48	1	0.04	7
49	1	0.04	8
50	2	0.08	10
51	2	0.08	12

Annotations:

- Red box: $1/25$ with an arrow pointing to the fr value 0.04 in the first row.
- Red box: $2/25$ with an arrow pointing to the fr value 0.08 in the fourth row.

xi	f	fr	fa
52	2	0.08	14
55	2	0.08	16
56	2	0.08	18
57	1	0.04	19
58	1	0.04	20
59	1	0.04	21
62	1	0.04	22
63	2	0.08	24
66	1	0.04	25
Total	25	1	

Annotations:

- Red box: "Siempre es el número total" with an arrow pointing to the fa value 25 in the last row.
- Red box: "Siempre es 1" with an arrow pointing to the fr value 1 in the Total row.

Intervalo de clase

- En ocasiones es conveniente acomodar los datos en pequeños grupos de igual tamaño, llamados **intervalos de clase**.
- **El punto medio o marca de clase “xi”**, se obtiene con:

$$\text{Marca de clase} = \frac{\text{Límite inferior} + \text{límite superior}}{2}$$

- **El tamaño del intervalo** se obtiene mediante la diferencia de los límites superior e inferior.

Ejemplo

Límite inferior

Límite superior

$$\frac{\text{Lím inf} + \text{Lim sup}}{2}$$

Intervalo de clase	Punto medio "xi"
38 – 42	40
43 – 47	45
48 – 52	50
53 – 57	55
58 – 62	60
63 – 67	65

Límite verdadero del intervalo

- Frontera de clase o límite verdadero del intervalo:

Intervalo de clase	Punto medio "xi"
37.5 – 42.5	40
40 – 2.5 42.5 – 47.5 40 + 2.5	45
47.5 – 52.5	50
52.5 – 57.5	55
57.5 – 62.5	60
62.5 – 67.5	65

Tabla de intervalos con límites verdaderos

□ Usando símbolos de desigualdad
Está incluido **No está incluido**

Intervalo de clase	Punto medio "xi"
$37.5 \leq x < 42.5$	40
$42.5 \leq x < 47.5$	45
$47.5 \leq x < 52.5$	50
$52.5 \leq x < 57.5$	55
$57.5 \leq x < 62.5$	60
$62.5 \leq x < 67.5$	65

□ Usando paréntesis y corchetes
Está incluido **No está incluido**

Intervalo de clase	Punto medio "xi"
$[37.5, 42.5)$	40
$[42.5, 47.5)$	45
$[47.5, 52.5)$	50
$[52.5, 57.5)$	55
$[57.5, 62.5)$	60
$[62.5, 67.5)$	65

El tamaño del intervalo es de 5

- Si por alguna razón no es fácil decidir el ancho del intervalo y el número de ellos, se pueden utilizar las siguientes fórmulas:
- $K = 1 + 3.3 \log (n)$
 - Donde K = número aproximado de clases
 n = número de datos.
- $\text{Amplitud de los intervalos} = \text{Rango} / K$
 - Donde Rango = diferencia entre el dato mayor y el dato menor.

Ejemplo

- Para el ejemplo de los datos de los pesos de 25 alumnos, el valor de K:

$$K = 1 + 3.3 \log (n) = 1 + 3.3 \log (25) = 5.6.$$

Por lo tanto se requieren aproximadamente **6** intervalos.

- Y la amplitud de los intervalos sería:

$$\text{Amplitud} = \text{Rango} / K = (66 - 40) / 5.6 = 4.64.$$

Aproximadamente **5** unidades es la amplitud de los intervalos.

Tabla de distribución de frecuencias para datos agrupados

- Se elabora con los intervalos de clase, sus puntos medios y las frecuencias correspondientes para cada uno de los intervalos.

Datos sin agrupar

xi	f
40	1
42	1
43	1
44	2
45	1
48	1
49	1
50	2
51	2

52	2
55	2
56	2
57	1
58	1
59	1
62	1
63	2
66	1
Total	25

Datos agrupados

Intervalo de clase	Punto medio "xi"	f
38 – 42	40	2
43 – 47	45	4
48 – 52	50	8
53 – 57	55	5
58 – 62	60	3
63 - 67	65	3
Total		25

- Se agregan las columnas de frecuencia relativa “fr” y frecuencia acumulada “fa”:

Intervalo de clase	Punto medio “xi”	f	fr	Fa
38 – 42	40	2	0.08	2
43 – 47	45	4	0.16	6
48 – 52	50	8	0.32	14
53 – 57	55	5	0.20	19
58 – 62	60	3	0.12	22
63- 68	65	3	0.12	25
Total		25	1	

- Por último se agregan las columnas:
 - ▣ **Frecuencia porcentual**, “ $f\%$ ” ó “ $\%f$ ”, se obtiene multiplicando la frecuencia relativa “ fr ” x 100.
 - ▣ **Frecuencia relativa acumulada** “ fra ”, se obtiene sumando las frecuencias relativas anteriores a un dato dado.
 - ▣ **Frecuencia porcentual acumulada**, “ $f\%a$ ”, se obtiene sumando las frecuencias porcentuales acumuladas a un dato dado.

Tablas de frecuencias absoluta, relativa y acumulada

Intervalo de clase	Punto medio "xi"	f	fr	f%	fa	fra	f%a
38 – 42	40	2	0.08	8	2	0.08	8
43 – 47	45	4	0.16	16	6	0.24	24
48 – 52	50	8	0.32	32	14	0.56	56
53 – 57	55	5	0.20	20	19	0.76	76
58 – 62	60	3	0.12	12	22	0.88	88
63- 68	65	3	0.12	12	25	1	100
Total		25	1	100			

$$0.08 \times 100$$

$$2/25$$

$$0.08 \times 100$$