

Distribución de frecuencias

Por: Emilio González Olguín

Distribución de frecuencias

Un conjunto de datos puede organizarse de diferentes maneras. La forma como se elija dependerá, entre otras cosas, de la naturaleza de los datos, la cantidad de datos o el aspecto que se desea describir.

Una manera de organizar los datos iniciales es construir una distribución de frecuencias simple. Una distribución de frecuencias simple es la que indica la cantidad de veces que aparece cada número o característica, ordenados desde el menor del conjunto de datos hasta el mayor de ese conjunto. Este tipo de distribución de frecuencias suele usarse cuando los datos se repiten *frecuentemente*.

Por ejemplo, en una escuela secundaria el profesor de Historia desea saber cuántos alumnos obtuvieron calificación de 10, para lo cual realizó la siguiente tabla donde muestra las calificaciones (incluidos los estudiantes que obtuvieron 10).

Calificación	Número de alumnos (frecuencias)
3	2
4	5
5	7
6	8
7	10
8	8
9	6
10	4
TOTAL	50

A la cantidad de veces que un número aparece en el conjunto inicial de datos se le llama **frecuencia**.

Cuando los datos no se repiten (como en el ejemplo anterior), es laborioso e inadecuado construir una distribución de frecuencias simple.

Imagina que al mismo grupo de 50 estudiantes se le organice de acuerdo al peso de cada uno. Aunque es probable que algunos pesen aproximadamente lo mismo, es prácticamente imposible que dos personas pesen exactamente lo mismo, por lo que esta distribución de frecuencias no es la más adecuada para organizar estos datos. Para el caso anterior se prefiere una distribución de **frecuencias con intervalos**.

Tabla de distribución de frecuencias con intervalos

El procedimiento para elaborar la tabla de distribución de frecuencias se explicará al resolver el ejemplo anterior.

Ejemplo

En una escuela preparatoria, el departamento escolar ha detectado problemas con los hábitos alimenticios de los estudiantes. El departamento pretende involucrar a los padres de familia para que en conjunto se atienda el problema, pero con la intención de tener más información al respecto y no sólo basarse en la observación se realizó un proceso de medición de 50 estudiantes. La información sobre los pesos de estos estudiantes se muestra en la siguiente tabla:

67	54	48	64	77
86	70	60	93	60
59	74	68	73	53
49	64	85	81	57
63	56	57	66	72
68	78	83	71	82
104	73	55	64	52
91	59	102	99	50
79	77	58	61	92
85	56	88	70	94

El encargado de analizar la información sobre los pesos de los 50 estudiantes al darse cuenta de que pocos son los pesos que se repiten, decide no utilizar una distribución de frecuencias simple porque la lista de los pesos distintos sería extensa y eso no es lo más conveniente.

Otra forma de organizar los datos es agrupando los pesos similares en intervalos, también denominados clases. La distribución de frecuencia que utiliza intervalos o clases para agrupar datos se conoce como distribución de frecuencia con intervalos.

El procedimiento para organizar los datos en una distribución de frecuencias con intervalos se detalla enseguida (Colegio24hs, 2004):

1.- Encontrar el mayor y el menor de los datos

El primer paso para crear una distribución de frecuencias por intervalos (llamados también clases) es encontrar el número mayor y el número menor. En este ejemplo, el número mayor es 104 y el número menor es 48.

2.- Calcular el rango

A la diferencia entre el dato mayor y el menor se le conoce como rango.

$$\text{Rango} = \text{Número mayor} - \text{Número menor}$$

$$\text{Rango} = 104 - 48 = 56$$

3.- Determinar el número de clases

No existe una regla generalizada para determinar el número de clases, pero una de ellas es:

$$\text{Número de clases} = \sqrt{n}$$

En donde n es el número de observaciones (o datos), por lo que el número de clases es:

$$\text{Número de clases} = \sqrt{50} = 7.07 \approx 7$$

IMPORTANTE

El número de clases tiene que ser un número entero, por lo que si \sqrt{n} da un número con decimales éste se tiene que redondear a un número entero.

Otra recomendación es respecto al número de clases: no debe ser menor a 5 ni mayor a 15, aunque algunos autores manejan hasta 20 clases.

4.- Ancho de clase

Una vez que se tiene el rango y el número de clases, es sencillo determinar el ancho de clase usando la fórmula:

$$\text{Ancho de clase} = \frac{\text{Rango}}{\text{Número de clases}}$$

Sustituyendo valores queda:

$$\text{Ancho de clase} = \frac{56}{7} = 8$$

IMPORTANTE

Para facilitar su manejo, el ancho de clase se puede redondear a un número **ligeramente** mayor. Por ejemplo, si el ancho de clase da 7.33333, se puede redondear a 7.5, ya que éste es más fácil de trabajar.

5.- Determinar los límites de cada clase

Primero se debe obtener el límite inferior de la primera clase, que es igual al dato menor, en este caso es 48.

Debido a que el ancho de clase es de 8, el límite superior de la primera clase se obtiene sumando el ancho de clase con el límite inferior.

Límite superior primera clase = $48 + 8 = 56$.

La escritura de intervalos se indica de la siguiente manera [48,56) lo que significa que se incluye el valor 48 pero no el 56 en este intervalo.

Por lo que la primera clase contendrá valores mayores o iguales a 48 pero menores a 56.

El límite inferior de la segunda clase será igual al límite superior de la primera clase. Su límite superior se obtiene sumando el ancho de clase con el límite inferior.

Límite superior segunda clase = $56 + 8 = 64$.

Para la escritura del intervalo se indica de la siguiente manera [56,64), lo que significa que incluye el valor 56 pero no el 64.

Por lo que la segunda clase contendrá valores mayores o iguales a 56 pero menores a 64. Y así sucesivamente, quedando en este caso los límites de clase de la siguiente forma:

CLASES	
[48	56)
[56	64)
[64	72)
[72	80)
[80	88)
[88	96)
[96	104)

Esta escritura de intervalo cerrado ([) en el límite inferior y abierto()) en el límite superior, suele no colocarse para no recargar la escritura, pero es muy importante considerarla para que se haga un conteo correctamente de los datos.

6.- Construir la distribución de frecuencias

Una vez que se han determinado los límites inferior y superior de cada clase es necesario cuantificar los datos contenidos en cada clase. La tabla 3 muestra una forma de contar cuántos estudiantes tienen un peso dentro de cada intervalo.

Clases		Números dentro de la clase
48	56	48, 49, 50, 52, 53, 54, 55
56	64	56, 56, 57, 57, 58, 59, 59, 60, 60, 61, 63
64	72	64, 64, 64, 66, 67, 68, 68, 70, 70, 71
72	80	72, 73, 73, 74, 77, 77, 78, 79
80	88	81, 82, 83, 85, 85, 86
88	96	88, 91, 92, 93, 94
96	104	99, 102, 104

Notemos que en la tabla en donde se indica “Números dentro de la clase”, aparece el número 56 en la segunda clase, porque está en un intervalo abierto en el primer intervalo y se considera hasta el segundo intervalo por indicarse como cerrado.

IMPORTANTE 

Si el valor mayor es igual al límite superior de la última clase, se incluye en ésta.

Completa el proceso de conteo.

La distribución de frecuencias queda como se muestra a continuación.

Clases		Frecuencia
48	56	7
56	64	11
64	72	10
72	80	8
80	88	6
88	96	5
96	104	3
Total		50

- ¿Cuál es el rango de peso que más se presentó?
- ¿Cuántos son los estudiantes que más peso tienen?
- ¿Cuántos estudiantes se encuentran en el rango menor de peso?

Las respuestas a estas preguntas son:

- El intervalo que más se presentó es de 56 a 64 kilogramos. Once estudiantes del total se encuentran en este rango.
- En la clase de 96 a 104 kilogramos, que es la de mayor peso, se localizaron solamente a tres estudiantes de un total de 50.
- Siete son los estudiantes que se encuentran en el intervalo inferior: de 48 a 56 kilogramos.

Ejemplo

Una empresa maquiladora de mezclilla está interesada en pagar un bono a sus trabajadores para compensar el gasto semanal en pasaje. 45 trabajadores reportaron los siguientes gastos semanales (en pesos).

95	100	165	135	120
110	110	134	112	114
184	96	90	115	135
164	138	146	158	156
154	147	100	140	180
126	125	95	170	170
118	92	137	162	98
162	132	120	136	163
94	118	185	143	182

Usemos nuevamente el procedimiento para organizar los datos en una distribución de frecuencias con intervalos.

1) Encontrar el mayor y el menor de los datos

El dato menor es 90.

El dato mayor es 185.

2) Calcular el rango

$$\text{Rango} = 185 - 90 = 95$$

3) Determinar el número de clases

No existe una regla generalizada para determinar el número de clases, pero una de ellas es:

$$\text{Número de clases} = \sqrt{45} = 6.7 \approx 7$$

Serán 7 clases.

4) Ancho de clase

Una vez que se tiene el rango y el número de clases es sencillo determinar el ancho de clase usando la fórmula:

$$\text{Ancho de clase} = \frac{\text{Rango}}{\text{Número de clases}}$$

Sustituyendo valores queda:

$$\text{Ancho de clase} = \frac{95}{7} = 13.57$$

Se redondea a un número mayor cercano que sea más fácil de utilizar, en este caso se sugiere el 14.

5) Determinar los límites de cada clase

Primero se debe obtener el límite inferior de la primera clase, que es igual al dato menor, que en este caso es 90.

Debido a que el ancho de clase es de 14, el límite superior de la primera clase se obtiene sumando el ancho de clase al límite inferior.

$$\text{Límite superior primera clase} = 90 + 14 = 104$$

Por lo que la primera clase contendrá valores mayores o iguales a 90, pero menores a 104.

El límite inferior de la segunda clase será igual al límite superior de la primera clase. Su límite superior se obtiene sumando el ancho de clase al límite inferior

$$\text{Límite superior segunda clase} = 104 + 14 = 118$$

Por lo que la segunda clase contendrá valores mayores o iguales a 104, pero menores a 118, y así sucesivamente, quedando en este caso los límites de clase de la siguiente forma:

Clases	
90	104
104	118
118	132
132	146
146	160
160	174
174	188

6) Construir la distribución de frecuencias

Una vez que se han determinado los límites inferior y superior de cada clase, es necesario cuantificar los datos que se encuentran entre esos valores. La distribución de frecuencias queda como se muestra a continuación.

Clases		Frecuencias
90	104	9
104	118	5
118	132	6
132	146	9
146	160	5
160	174	7
174	188	4
Total		45

Ahora estás listo para responder preguntas como:

- ¿Cuál es el intervalo de gasto semanal que más se presentó?
- ¿Cuántos trabajadores se encuentran en este rango?
- ¿Cuál es la clase que representa un gasto semanal mayor para los trabajadores?

Respuestas:

- 90-104 y 132-166
- 9
- 174-188

Frecuencia acumulada, frecuencia relativa y frecuencia relativa acumulada

Para contestar las siguientes preguntas:

- ¿Cuántos estudiantes se encuentran en el intervalo de pesos de 48 a 88 kilogramos?
- ¿Qué porcentaje representan este grupo de estudiantes?
- ¿Qué porcentaje de los alumnos pesan entre 88 y 96 kilogramos?

Tenemos que obtener la frecuencia acumulada, frecuencia relativa y la frecuencia relativa acumulada. A continuación se explica cómo hacerlo.

Frecuencia acumulada

Suma a la frecuencia de cada clase las frecuencias de las clases anteriores (en la información del primer ejercicio se muestra cómo determinar la frecuencia acumulada).

Clases		Frecuencia	Frecuencia acumulada
48	56	7	$7+0=7$
56	64	11	$7+11=18$
64	72	10	$18+10=28$
72	80	8	$28+8=36$
80	88	6	$36+6=42$
88	96	5	$42+5=47$

96	104	3	47+3 = 50
Total			50

IMPORTANTE

A la frecuencia acumulada de la primera clase se le suma CERO, ya que no hay otra frecuencia que lo preceda.

La frecuencia acumulada de la última clase debe ser igual al total de datos.

Frecuencia relativa

Muestra la frecuencia de una clase como proporción del total de datos. Por ejemplo, para el caso de los 50 estudiantes, el porcentaje del peso de los estudiantes que se encuentran entre 88 y 96 es:

$$Frecuencia\ relativa = \frac{frecuencia}{Total\ de\ datos} = \frac{5}{50} = 0.10$$

Si el resultado lo multiplicas por 100, lo puedes expresar como porcentaje, que en este caso sería del 10 %.

Frecuencia relativa acumulada

Suma a la proporción de cada clase las proporciones de las clases anteriores.

IMPORTANTE

A la frecuencia relativa acumulada de la primera clase se le suma CERO, ya que no hay otra frecuencia que lo preceda.

La frecuencia relativa acumulada de la última clase debe ser igual a UNO.

Continuando con el primer ejemplo, se mostrará cómo obtener la frecuencia relativa y la frecuencia relativa acumulada:

Clases	Frecuencia	Frecuencia acumulada	Frecuencia relativa	Frecuencia relativa acumulada
--------	------------	----------------------	---------------------	-------------------------------

48	56	7	7	$7/50 = 0.14$	$0.14+0 = 0.14$
56	64	11	18	$11/50 = 0.22$	$0.22 + 0.14 = 0.36$
64	72	10	28	$10/50 = 0.2$	$0.2 + 0.36 = 0.56$
72	80	8	36	$8/50 = 0.16$	$0.16 + 0.56 = 0.72$
80	88	6	42	$6/50 = 0.12$	$0.12 + 0.72 = 0.84$
88	96	5	47	$5/50 = 0.1$	$0.1 + 0.84 = 0.94$
96	104	3	50	$3/50 = 0.06$	$0.06 + 0.94 = 1$
Total		50			

Con esta información adicional ya puedes dar respuesta a las preguntas:

- ¿Cuántos estudiantes se encuentran en el intervalo de pesos de 48 a 88 kilogramos?
- ¿Qué porcentaje representa este grupo de estudiantes?
- ¿Qué porcentaje de los alumnos pesa entre 88 y 96 kilogramos?

Las respuestas a estas preguntas son:

- 42 estudiantes pesan entre 48 y 88 kilogramos.
- El 84% de los estudiantes pesan entre 85 y 95 kilogramos.
- El 10% de los estudiantes pesan entre 88 y 89 kilogramos.

¿Tienes alguna otra interpretación de la información observada?

Con la práctica, tus observaciones serán más concretas y relevantes. ¡Adelante!

Busca distribuciones de frecuencia en los periódicos, en las revistas o Internet, en el lugar menos esperado encontrarás una.



Referencia

Colegio24hs. (2004). *Estadística*. Argentina: Autor.



Bibliografía

Devore, J. L. (2005). *Probabilidad y estadística para ingeniería y ciencias* (6a. ed.). México: Thomson.

Hayter, A. J. (2002). *Probability and Statistics for engineers and scientists* (2nd. ed.). Pacific Grove: Duxbury.