Datos agrupados

**DISTRIBUCIONES DE FRECUENCIA**. Cuando se dispone de gran número de datos, es útil el distribuirlos en clases o categorías y determinar el número de individuos pertenecientes a cada clase, que es la frecuencia de clase. Una ordenación tabular de los datos en clases, reunidas las clases y con las frecuencias correspondientes a cada una, se conoce como una **Distribución de frecuencias o Tabla de frecuencias**.

La tabla de distribución de frecuencias se hace agrupando el conjunto de datos numéricos en clases o intervalos apropiados. Este procedimiento lo explicaremos mediante el desarrollo del siguiente ejercicio.

Los siguientes datos se recopilaron con el fin de determinar la edad de 50 estudiantes del grado 10º de la sección nocturna de la IE María Montessori. Así los datos obtenidos fueron los siguientes:



Organizando este conjunto de datos en forma ascendente y haciendo el correspondiente recuento (número de veces que se repite cada valor) obtenemos el registro indicado a continuación. Observemos que el valor máximo es 60 y el menor es 15.

Edades de 50 estudiantes del grado 10º de IE María montessori sección nocturna.

**Edad Nº de veces que se repite**



En la anterior ordenación la variable X toma muchos valores diferentes y algunos de ellos tienen una frecuencia tan pequeña que no se justifica considerarlos por separado. Además, no podemos visualizar claramente las medidas de tendencia central y su cálculo se dificulta por la cantidad de operaciones que deben realizarse. Por tanto, es necesario agrupar los datos en clases o intervalos.

Para agrupar este conjunto en clases o intervalos de datos debemos seguir los siguientes pasos:

**1er paso: Rango o recorrido**

Calculamos el rango o recorrido que representamos por R y que es la diferencia entre el Xmáx y el Xmín.

R = Xmáx - Xmín

R = 60 – 15 = 45

Un rango de 45 años significa que la diferencia entre la mayor edad y la menor es 45 años.

**2º Paso: Elección del número de intervalos**

Determinamos el número K de clases o intervalos en que vamos a agrupar los datos. No existe una regla única para fijar el K, pero generalmente varía entre 5 y 20 () dependiendo del tamaño n de la muestra; su elección queda al criterio del lector. Elijamos para agrupar nuestra muestra un número K = 7 de clases o intervalos.

**3er Paso: Amplitud de intervalos o clases**

Repartimos el rango en clases o intervalos de la misma longitud o amplitud. Si a representa la amplitud de cada intervalo, entonces:

 

Cuando los datos sean valores enteros de la variable, entonces el cociente  debe ser un número entero. Si no ocurre que  es entero, como en nuestro ejemplo (), debemos aproximar **a** al número entero más próximo por encima, es decir a = 7.

Si los datos de la muestra tienen cifras decimales, entonces debemos tomar una amplitud que tenga el mismo número de cifras decimales.

Así, si en un ejercicio encontramos que , entonces tomamos a = 2,86 si los datos tienen dos cifras decimales; a= 2,9 si los datos tienen una cifra decimal y a=3 si los datos tienen son números enteros.

**4º Paso: Límite de intervalos**

Si K = 7 y a = 7, entonces el rango que vamos a repartir ya no es R = 45 sino 7 x 7 = 49. Este nuevo rango se representa por Ra y se llama rango ampliado.

Si Ra - R es la cantidad en que se amplía el rango, entonces en esta misma cantidad se debe ampliar el Xmáx o disminuir el Xmín ( o ambos) para que se cumpla:

Ra = Xmáx – Xmín.

En nuestro ejemplo: 49 = 64 – 15 ó 49 = 60 – 11 ó 49 = 62 - 13

Si aumentamos el Xmáx en 4, entonces el Xmáx = 64 es el límite superior del último intervalo.

Si al límite inferior Li = 15 del primer intervalo se le adiciona la amplitud a = 7, el resultado Ls = 15 + 7 = 22 es el límite superior del primer intervalo, Así:

Li  Xmín. 15

**Primera clase** Ls =Xmín+ Amplitud

Ls =Li + a

Ls = 15 + 7 = 22

La primera clase está formada por todos los valores de x entre 15 y 22 años.

La segunda clase tiene como límite inferior el límite superior de la primera clase y como límite superior el inferior aumentado en la amplitud.

Segunda clase 22 - 29

Este procedimiento se repite hasta obtener un número (K = 7) de intervalos ya establecido que tiene a 64 como límite superior del último intervalo. (Ver tabla).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Clase | Intervalo | Marca de clase: Xi | Frecuencia absoluta:  fi | Frecuencia absoluta acumulada: Fi | Frecuencia relativa porcentual =% | Frecuencia porcentual acumulada  % | Xi.fi |
| 1º | 15 - 22 | 18,5 | 37 | 37 | 0,74=74% | 74% | 684.5 |
| 2º | 22 - 29 | 25,5 | 5 | 42 | 0,10=10% | 84% | 127.5 |
| 3º | 29 - 36 | 32,5 | 3 | 45 | 0,06=6% | 90% | 97.5 |
| 4º | 36 - 43 | 39,5 | 4 | 49 | 0,08=8% | 98% | 158 |
| 5º | 43 - 50 | 46,5 | 0 | 49 | 0,00=0% | 98% | 0 |
| 6º | 50 - 57 | 53,5 | 0 | 49 | 0,00=0% | 98% | 0 |
| 7º | 57 - 64 | 60,5 | 1 | 50 | 0;02=2% | 100% | 60.5 |
|  |  |  |  |  |  |  |  |
|  | Total |  | n=50 |  | 1,00=100% |  |  |

Cuando el cociente es exacto y no hay necesidad de ampliar el rango, tanto el límite inferior del primer intervalo como el límite superior del último intervalo coinciden con los x mín. y x máx.

**5º paso: Marcas de clase:** Como en cada intervalo podemos considerar infinitos valores reales de la variable x, debemos tomar uno de ellos que nos represente la clase y nos permita hacer gráficas y cálculos(como la media aritmética). A cada uno de estos valores se le llama **marca de clase** y su mejor representante es el punto medio del intervalo (o valor central).

Si xi  representa el punto medio del intervalo i-ésimo, entonces la primera marca de clase es:



Las otras marcas de clase se pueden obtener en forma similar, o sumando a la anterior la amplitud:

 = 18,5 + 7 = 25,5

= 25,5 + 7 = 32,5

= 32,5 + 7 = 39,5



= 53,5 + 7 = 60,5

**6º paso: Tabla de distribución de frecuencias:** Si al elaborar la columna de las frecuencias absolutas, un valor muestral coincide con uno de los límites del intervalo, convenimos en tomar ese valor en aquella clase donde aparece como límite inferior del intervalo. Es decir, son intervalos cerrados –abiertos. Por ejemplo, el valor 22 que aparece como límite superior del primer intervalo pertenece a la segunda clase.

El último intervalo lo tomamos cerrado para que el x máx. y los valores que coinciden con él no queden fuera de la tabla.

En la tabla anterior indicamos las frecuencias absolutas, acumuladas, relativas y acumuladas porcentuales.

Un análisis de la tabla de distribución de frecuencias nos permite afirmar:

37 estudiantes de los grados 10º sección nocturna de la IE María Montessori tienen unan edad entre 15 y 22 años correspondientes al 74% de la muestra.

De los 50 estudiantes 49 son menores de 43 años, lo cual corresponde al 98% de la muestra tomada.

Las edades más frecuentes están entre los 15 y 22 años, por tener esta clase la máxima frecuencia absoluta