Tema 03: Organización De Datos Y Distribución De Frecuencias

Organización de los Datos Obtenidos de una Muestra

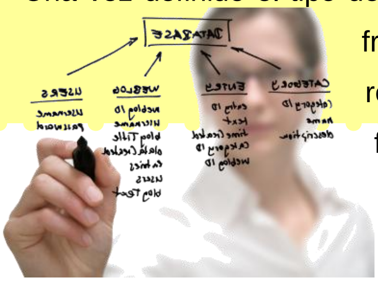
Cuando se han recopilado datos mediante un muestreo o un censo, la primera inquietud que aparece es sobre la manera en la que se puede realizar un análisis descriptivo apropiado con la información recolectada de manera que resulte sencillo entender lo que ocurre en la población de la que se han captado las observaciones.

En este tema se proprocionan algunos procedimientos para la tabulación de datos que conducen a la formación de cuadros o tablas de frecuencias.

Organización De Los Datos Cualitativos

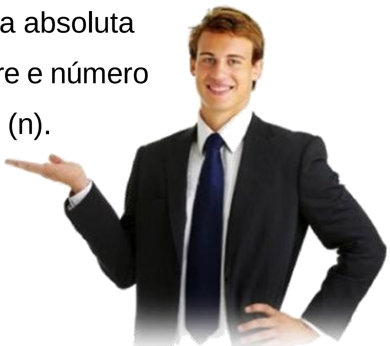
Antes de iniciar el trabajo de organización de datos cualitativos, es necesario determinar si éstos corresponden a variables cualitativas nominales u ordinales. Si los datos son cualitativos nominales, se formar categorías que pueden ser presentadas en cualquier orden: por ejemplo los colores de preferencia de las personas. Si los datos son ordinales, entonces deben estar asociados a algún orden en su presentación.

Una vez definido el tipo de variable, se obtiene mediante un proceso de conteo las frecuencias absolutas (número de veces que se repite cada respuesta), luego las frecuencias relativas (división de cada frecuencia absoluta entre el tamaño de muestra) y/o los porcentajes de cada respuesta (cada frecuencia relativa multiplicada por 100). También se puede encontrar las frecuencias absolutas acumuladas (Fi)



Frecuencia relativa (hi).-

Es la **proporción** del total de observaciones que caen dentro de cada modalidad o valor. Se obtiene dividiendo la frecuencia absoluta (fi) de la modalidad entre e número total de observaciones (n).



Frecuencia acumulada (Fi).-

Para cada clase, valor o modalidad, la frecuencia acumulada equivale la frecuencia absoluta (fi) de la fila sumada a la frecuencia acumulada de la fila anterior. Para la primera fila, la frecuencia acumulada equivale simplemente a la frecuencia absoluta de la misma fila.

Ejemplo 1. Una revista conocida efectuó una encuesta respecto a lo adecuado de la protección policial en la ciudad. Se seleccionó un total de 419 personas. Las respuestas se presentan en la siguiente tabla de frecuencias:

Respuesta	Frecuencia absoluta	Frecuencia relativa	Porcentaje
Si	293	0.6993	69.93
No	80	0.1909	19.09
No sabe/ no responde	46	0.1098	10.98

Modalidad	Frecuencia absoluta	Frecuencia relativa	Porcentaje
Primaria completa Secundaria completa Educac.superior completa	125 70 5	0.625 0.35 0.025	62.5 35 2.5
	200	1	100

Organización De Datos Cuantitativos Discretos

Cuando se tienen datos cuantitativos discretos cuyo número de resultados posibles no es grande, la información puede ser clasificada y presentada directamente sin pérdida de la identidad de la misma. En estos casos, primero se ordenan los posibles valores de la variable según su magnitud, y a continuación se obtienen, mediante un proceso de conteo, las frecuencias absolutas asociadas a cada uno de dichos valores; las frecuencias relativas y porcentuales se obtienen de manera similar a lo descrito para las variables cualitativas.

Ejemplo. Consideremos la variable número de hijos y tomemos las observaciones de una muestra de 150 familias de zonas marginales de Lima Metropolitana. Los valores obtenidos se pueden agrupar en diferentes valores: 0 hijos, 1 hijo, 2 hijos, 3 hijos o 4 hijos. Para hacer un arreglo de estas observaciones, usaremos una tabla como la siguiente:

Organización De Datos Cuantitativos Continuos

Cuando se tiene información para una variable cuantitativa continua, las observaciones son usualmente diferentes entre sí, lo cual hace que la evaluación descriptiva a través de los valores individuales sea compleja. Para simplificar el análisis, los datos son clasificados de acuerdo con ciertos rangos llamados inte

Número de hijos	Frecuencia absoluta	Frecuencia relativa	Porcentaje
0	2	0.013	1.3
1	15	0.100	10.0
2	40	0.267	26.7
3	55	0.367	36.7
4	38	0.253	25.3
TOTAL	**150**	**1**	**100**

Ejemplo. Tomamos una muestra de 100 niños de 10 años de edad para estudiar su estatura.

Entonces la variable estatura que es cuantitativa continúa se puede presentar en una tabla del siguiente tipo:

INTERVALOS DE CLASE	FRECUENCIA(fi)
[1.0mt.- 1.15mt]	3
[1.15mt. - 1.30mt]	39
[1.30mt.- 1.45mt]	55
[1.45mt.- 1.60mt]	3
TOTAL	**100**

A la organización de las observaciones de una muestra en una tabla para expresar la frecuencia de cada una de sus modalidades o valores se le conoce como **distribución de frecuencias.** En las distribuciones de frecuencia de las variables cuantitativas continuas, también se acostumbra colocar otras columnas además de la frecuencia absoluta (**fi**), estas nos permitirán tener una mayor información sobre los datos y nos facilitarán los cálculos de las medidas descriptivas o estadísticos de la muestra. Estas son la frecuencia relativa (hi), la frecuencia acumulada (Fi y Hi).

La organización de los datos para el caso en que la variable estadística usada tenga muchos valores implica e arreglo de las observaciones en intervalos de clases. El proceso para hallar los intervalos de clase es el siguiente:

Debemos hallar, en primer lugar, en la muestra, el menor valor observado y el mayor valor observado.

El número de intervalos no deberá ser tan pequeño (menor que 5) o tan grande (mayor de 15) que la verdadera naturaleza de la distribución sea imposible de visualizar. La longitud del intervalo de clase deberá ser siempre la misma.

Si la longitud de cada intervalo no fuera exacta, se puede tomar por exceso asegurándonos de este modo que la reunión de todos los intervalos cubrirá a todos los valores observados.

Para construir los intervalos se usa los intervalos cerrados a la izquierda y abiertos a la derecha: [LIi, LSi[, donde LIi,es el límite inferior del intervalo y LSi, es su límite superior.

¿Cómo decidimos cuántos intervalos de clase tomar?

Existen varias reglas que se basan en el tamaño de nuestra población o muestra. Una de las reglas más usadas es la **Regla de Sturges,** regla empírica que funciona bastante bien para grupos de 30 a 300 observaciones.

Esta regla nos dice que el número de intervalos de clase para una muestra de tamaño **n** será **k** si este resulta un número entero o el siguiente número entero a **k,** si **k** resulta un número decimal.

La ecuación para hallar **k** es: **k = 1 + 3.3 * log n,** donde **n** es el tamaño de la muestra.

La **marca de clase (xi),** definida como el punto medio del intervalo de clase, deberá tener de preferencia el mismo número de decimales que los valores observados. **La marca de clase puede considerarse que es un representante de los datos que caen en el intervalo.**

$$Xi = \frac{LIi + LSi}{2}$$

Ejemplo 1. Suponga que los datos que se presentan a continuación corresponden a los valores de la inflación anual durante el año 2008 de un total de 20 ciudades de una región del país. Construir la distribución de frecuencias

8.2 12.8 10.5 9.3 12.7 10.2 9.1 10.7 8.2 12.8 8.5 11.6 8.4 10.1 10.2 13.1 9.8 12.1 13.6 11.7

Solución

1. **R = 13.6 – 8.2 = 5.4**
2. **K = 1 + 3.3 log20 = 1 + 3.3 (1.30.10) = 1 + 4.29 = 5.29 = 5** (redondeo por aproximación)
3. **C = R/k = 5.4 / 5 = 1.08 = 1.1** (redondeo por exceso; los datos tienen un decimal)
4. Los límites de los intervalos se obtienen del siguiente modo:

   LI_1 = 8.2 — **$LS_1 = LI_2$ = 9.3**

   LI_2 = LI1 + c = 8.2 + 1.1= 9.3 — **$LS_2 = LI_3$ = 10.4**

   $LI_3 = LI_2$ + c = 9.3 + 1.1 = 10.4 — **$LS_3 = LI_4$ = 11.5**

   $LI_4 = LI_3$ + c = 10.4 + 1.1 = 11.5 — **$LS_4 = LI_5$ = 12.6**

   $LI_5 = LI_4$ + c = 11.5 + 1.1 = 12.6 — **$LS_5 = LS_4$ + c = 12.6 + 1.1 = 13.7**

5. Las marcas de clase se calculan de la siguiente manera:

   X1 = $\frac{8.2+9.3}{2} = \mathbf{8.75}$; $X2 = \frac{9.3+10.4}{2} = \mathbf{9.85}$ y así sucesivamente

6. Para determinar las frecuencias absolutas se procede como sigue: Se toma la primera observación 8.2 y se busca el intervalo de clase que pertenece, es el 8.2 – 9.3, luego se asigna una tarja en la intersección de la columna de conteo y la fila de ese intervalo. Se toma ahora la otra observación 12.8, la cual pertenece al intervalo 12.6 – 13.7, entonces se asigna una tarja en la intersección de la fila de este nuevo intervalo y la columna de conteo. Así sucesivamente hasta agotar la última observación. Sumando las tarjas se obtiene la frecuencia absoluta de cada clase.
7. Para obtener las frecuencias acumuladas se procede de la siguiente forma:
   F1 = f1 = 5 **F2 = F1 + f2 = 5 + 5 = 10**

Con los resultados anteriores se obtiene el siguiente cuadro de distribución de frecuencias

Intervalos de clase	Marca de clase Xi	Tarjas	Frecuencia absoluta fi	Frecuencia acumulada Fi	Frecuencia relativa hi
8.2 ; 9.3	8.75	/////	5	5	5/20
9.3 ; 10.4	9.85	/////	5	10	5/20
10.4 ; 11.5	10.95	//	2	12	2/20
11.5 ; 12.6	12.05	///	3	15	3/20
12.6 ; 13.7	13.15	/////	5	20	5/20
			20		1

Ejemplo 2. A continuación, se presenta una lista ya ordenada de las observaciones hechas sobre el ingreso de las personas.

53	**57**	**58**	**61**	**61**	**63**	**64**	**66**	**67**	**68**
69	**70**	**71**	**72**	**73**	**74**	**74**	**74**	**74**	**77**
77	**77**	**78**	**78**	**79**	**79**	**79**	**81**	**81**	**81**
82	**82**	**83**	**83**	**84**	**85**	**85**	**86**	**87**	**87**
88	**90**	**90**	**90**	**90**	**92**	**93**	**94**	**96**	**97**

Para estos ingresos, el menor valor de la muestra es 53 dólares y el mayor valor de la muestra es 97 dólares. Luego, el rango de estos valores es: 97 - 53 = 44 dólares

Al aplicar la regla de Sturges con **n= 50**, tendremos: k= 1 + 3.3*(1.69897) = 6.6, lo que equivale a tomar 7 intervalos s)

El tamaño o amplitud de cada intervalo de clase se determina así: **c=R /K= 44 / 7 = 6.29** = **7**. (redondeo por exceso, al entero superior, considerando que los datos son entero. Si los datos tienen decimales el proceso es el mismo).

INTERVALOS	xi	fi	hi	Fi
[53 ; 60]	56.5	3	3/ 50	3
[60 ; 67]	63.5	5	5/ 50	8
[67 ; 74]	70.5	7	7/50	15
[74 ; 81]	77.5	12	12 / 50	27
[81 ; 88]	84.5	13	*13/50*	40
[88 ; 95]	91.5	10	10 / 50	50
TOTAL		50	1	

Siguiendo el mismo proceso utilizado para el ejemplo 1, se tiene la tabla de distribución de frecuencias: