

GENERALIDADES

La *distribución de frecuencias*, es un método utilizado para organizar y resumir información. Bajo este método, los datos recolectados se ordenan y clasifican, indicándonos la frecuencia o sea el número de veces que se repiten.

Se podrá decir también, que nos permite manejar grandes cantidades de información en espacios reducidos, en formas de cuadros o tablas, complementadas con gráficas.

Por *población o universo*, se entiende como un conjunto de medidas para ser aplicadas a una característica cuantitativa, o como el recuento de todas las unidades que presentan una característica común, siendo ésta cualitativa. También se puede definir a la población como un conjunto de elementos o unidades. Lo que se estudia en una *unidad o elemento* son sus *características*.

Cuando se toman todas las unidades o elementos de la población, se habla de una *investigación exhaustiva* o censo. Si sólo se investiga una parte, se le considera como *investigación parcial o muestra*.

La *muestra*, para que sea representativa de la población, requiere que las unidades o elementos sean seleccionadas al *azar*, en tal forma que cada una de ellas tenga la misma posibilidad de ser seleccionada.

Se usan letras mayúsculas o letras del alfabeto griego como símbolos en poblaciones, en cambio, en la muestras se emplean letras minúsculas.

Los caracteres de los *elementos* de una población pueden ser *cualitativos o cuantitativos*. Los datos cualitativos, denominados también *atributos*, son todos aquellos elementos que pueden ser descritos cualitativamente, es decir mediante palabras; son ejemplos de atributos: la clasificación de los alumnos de una universidad por lugar de origen; clasificación de un grupo de personas por ocupación, por cargo, por sexo, etc.

Los caracteres *cuantitativos* denominados *variables*, son todas aquellas características susceptibles de ser expresadas cuantitativamente, es decir, mediante números. Ejemplo: peso, estatura, edad, número de hijos, salarios, etc.

Las variables se dividen en *discretas y continuas*. Es de tener en cuenta que esta clasificación tiene más valor teórico que práctico.

Las *variables discretas* son aquéllas que admiten solamente valores enteros, es decir, no tienen valores intermedios. Ejemplo: el número de hijos por familia será discreta, ya que no se podrá decir que una familia tiene dos hijos y medio; el número de empleados por departamento en una empresa, etc.

Las *variables continuas* son aquéllas que admiten valores fraccionarios, pudiéndose establecer intervalos. Ejemplo: la estatura de una persona que mide un metro con setenta centímetros; que pesa sesenta kilos, una libra y cuatro onzas, etc.

Nota: Se hará un pequeño resumen sobre aquellos términos que más serán utilizados en los capítulos posteriores y que fueron presentados en forma amplia en el anterior capítulo.

ELABORACIÓN DE TABLA DE FRECUENCIAS

VARIABLE DISCRETA

Ante todo es conveniente familiarizarnos con ciertos símbolos que utilizaremos tanto para la variable discreta como para la variable continua.

$\overline{n}$	=	n	=	Tamaño de la muestra.
N	=	N	=	Tamaño de la población o universo.
X_i	=	x_i	=	Identificación para cada valor observado. (minúscula en la muestra)
f_i	=	n_i	=	Frecuencias absolutas.
f_i/n	=	h_i	=	Frecuencias relativas.
F_i	=	N_i	=	Frecuencias absolutas acumuladas.
H_i	=	H_i	=	Frecuencias relativas acumuladas.
X_i	=	y_i	=	Identifica la variable discreta o las marcas de clase en la continua.
$X'_{i-1} - X'_i$	=	$y'_{i-1} - y'_i$	=	Identifica a la variable continua con sus intervalos.
i	=	c	=	Amplitud del intervalo
m	=	m	=	Número de valores de la variable o de intervalos.

Se ha considerado conveniente presentar las dos simbologías más utilizadas en los textos de Estadística, de tal manera que el estudiante pueda consultar o aplicar indistintamente cualquiera de ellas.

Ejemplo 1. Supongamos que se tiene una población constituida por 2.000 cajas y deseamos examinarlas con el fin de determinar el número de piezas o elementos defectuosos que contiene cada caja. Por diferentes razones, se desea que la investigación no sea exhaustiva, es decir, no examinar la totalidad de las 2.000 cajas o universo sino, por el contrario, seleccionar una muestra de tamaño 30, correspondiente a una investigación parcial. $N = 2.000$ $n = 30$

Los resultados obtenidos en esta encuesta, se anotan a continuación. Siendo x_1 , la primera caja examinada y 3, el número de piezas defectuosas encontradas en ella y así sucesivamente:

$x_1 = 3$	$x_2 = 2$	$x_3 = 0$	$x_4 = 2$	$x_5 = 3$	$x_6 = 3$	$x_7 = 1$	$x_8 = 1$	$x_9 = 0$
$x_{10} = 1$	$x_{11} = 3$	$x_{12} = 3$	$x_{13} = 4$	$x_{14} = 4$	$x_{15} = 3$	$x_{16} = 2$	$x_{17} = 4$	$x_{18} = 2$
$x_{19} = 4$	$x_{20} = 2$	$x_{21} = 4$	$x_{22} = 0$	$x_{23} = 0$	$x_{24} = 1$	$x_{25} = 3$	$x_{26} = 4$	$x_{27} = 2$
$x_{28} = 3$	$x_{29} = 1$	$x_{30} = 2$						

En la anterior información o conjunto de datos, observamos que no están ordenados y los denominamos *datos originales* o *no ordenados*. Quizás, dado que la muestra es muy pequeña, podemos darnos cuenta, en forma muy ligera del comportamiento de la característica. Pero si en vez de tener 30 observaciones, fueran 200 o 2.000 es muy poco lo que podemos decir o conocer de la variable; por lo tanto se hace necesario el ordenamiento o la clasificación de esos datos, mediante un proceso que llamaremos *tabulación* o *procesamiento de datos*.

[illegible]

En el caso en que la variable no emplee valores en forma continua, solamente se debe considerar en la distribución aquellos que aparecen en la base de datos, como se puede observar en la Tabla 2.2, donde se omiten los valores 2 4 6 7 y 9.

NÚMERO DE PIEZAS	NÚMERO DE CAJAS	TABULACIÓN	
		A	B
0	4		□
1	5	≡≡	⊠
2	7	≡≡	⊠└
3	8	≡≡	⊠┐
4	6	≡≡	⊠┌
Σ	30	30	30

En el ejemplo 1 de la Tabla 2.1 se puede interpretar, diciendo que en la muestra de 30 cajas encontramos cuatro de ellas sin piezas defectuosas; 5 cajas con una; 7 cajas con dos; 8 cajas con tres; y 6 con cuatro piezas defectuosas. Sin embargo, a veces, es mejor expresarlo en términos porcentuales, diciendo que el 13% de las cajas examinadas no tenían piezas defectuosas, que el 17% tenían una; que el 23%, dos; el 27%, tres; y el 20% con cuatro piezas o unidades defectuosas, como se puede observar en el cuadro siguiente:

NÚMERO DE PIEZAS DEFECTUOSAS
0
1
3
5
8
10

Distribución de frecuencias - variable discreta

y_i	n_i	h_i	N_i	H_i
0	4	0,13	4	0,13
1	5	0,17	9	0,30
2	7	0,23	16	0,53
3	8	0,27	24	0,80
4	6	0,20	30	1,00
Σ	30	1,00	-	-
X_i	f_i	f_i/n	F_i	H_i

y_i	n_i	h_i	N_i	H_i
y_1	n_1	h_1	N_1	H_1
y_2	n_2	h_2	N_2	H_2
y_3	n_3	h_3	N_3	H_3
y_4	n_4	h_4	N_4	H_4
y_5	n_5	h_5	N_5	H_5
Σ	n	1,00	-	-
X_i	f_i	f_i/n	F_i	H_i

Las frecuencias relativas se obtienen de la siguiente manera:

$$h_1 = \frac{n_1}{n} = \left[\frac{f_1}{n} \right] \quad h_2 = \frac{n_2}{n} = \left[\frac{f_2}{n} \right] \quad h_3 = \frac{n_3}{n} = \left[\frac{f_3}{n} \right] \quad h_4 = \frac{n_4}{n} = \left[\frac{f_4}{n} \right] \quad h_5 = \frac{n_5}{n} = \left[\frac{f_5}{n} \right]$$

$$\left[\frac{f_1}{n} \right] = h_1 = \frac{4}{30} = 0,13 \quad \left[\frac{f_2}{n} \right] = h_2 = \frac{5}{30} = 0,17 \quad \dots \text{ etc.}$$

Las *frecuencias absolutas acumuladas* las utilizaremos más adelante para el cálculo de algunas medidas. Se obtienen así:

$N_1 = n_1$	→	$= N_1 = n_1 = 4$
$N_2 = n_1 + n_2$	→	$= N_1 + n_2 = 4 + 5 = 9$
$N_3 = n_1 + n_2 + n_3$	→	$= N_2 + n_3 = 9 + 7 = 16$
$N_4 = n_1 + n_2 + n_3 + n_4$	→	$= N_3 + n_4 = 16 + 8 = 24$
$N_5 = n_1 + n_2 + n_3 + n_4 + n_5$	→	$= N_4 + n_5 = 24 + 6 = 30$

$F_1 = f_1$	→	$= f_1$
$F_2 = f_1 + f_2$	→	$= F_1 + f_2$
$F_3 = f_1 + f_2 + f_3$	→	$= F_2 + f_3$
$F_4 = f_1 + f_2 + f_3 + f_4$	→	$= F_3 + f_4$
$F_5 = f_1 + f_2 + f_3 + f_4 + f_5$	→	$= F_4 + f_5$

Las frecuencias relativas acumuladas resultan así.

$H_1 = h_1$	→	$= H_1 = h_1 = 0,13$	$= F_1/n$ $= F_2/n$ $= F_3/n$ $= F_4/n$ $= F_5/n$
$H_2 = h_1 + h_2$	→	$= H_1 + h_2 = 0,13 + 0,17 = 0,30$	
$H_3 = h_1 + h_2 + h_3$	→	$= H_2 + h_3 = 0,30 + 0,23 = 0,53$	
$H_4 = h_1 + h_2 + h_3 + h_4$	→	$= H_3 + h_4 = 0,53 + 0,27 = 0,80$	
$H_5 = h_1 + h_2 + h_3 + h_4 + h_5$	→	$= H_4 + h_5 = 0,80 + 0,20 = 1,00$	

Ejemplo 2. Con la información de la tabla No. 2.3 determinar los valores que toma:

Solución

$$y_3 = \left[X_3 \right] = 2 \quad N_2 = \left[F_2 \right] = 9 \quad n_5 = \left[f_5 \right] = 6$$

$$H_3 = \left[\frac{F_3}{n} \right] = 0,53 \quad h_4 = \left[\frac{f_4}{n} \right] = 0,27 \quad x_{16} = 2$$

VARIABLE CONTINUA

Ejemplo 3. Consideremos que se seleccionó una muestra de 30 alumnos, a fin de conocer su peso en kilos; para facilitar el trabajo redondeamos las cifras.

$x_1 = 74$	$x_2 = 67$	$x_3 = 94$	$x_4 = 70$	$x_5 = 69$	$x_6 = 61$	$x_7 = 71$	$x_8 = 79$	$x_9 = 47$
$x_{10} = 85$	$x_{11} = 82$	$x_{12} = 55$	$x_{13} = 65$	$x_{14} = 88$	$x_{15} = 52$	$x_{16} = 58$	$x_{17} = 76$	$x_{18} = 57$
$x_{19} = 72$	$x_{20} = 66$	$x_{21} = 48$	$x_{22} = 56$	$x_{23} = 63$	$x_{24} = 71$	$x_{25} = 60$	$x_{26} = 64$	$x_{27} = 68$
$x_{28} = 83$	$x_{29} = 74$	$x_{30} = 92$						

El primer paso a seguir consiste en determinar el valor máximo y el mínimo que toma la variable. En este ejemplo tenemos: $x_{\max} = 94$ y $x_{\min} = 47$. La diferencia entre el valor máximo y el mínimo se denomina recorrido. En el presente caso el recorrido es de 47.

$$x_{\max} - x_{\min} = \text{Recorrido o rango}$$

$$94 - 47 = 47$$

Introducimos dos nuevos símbolos que son: $m = \text{número de intervalos}$ y

$$C = \text{amplitud del intervalo} = [i]$$

El valor de m se puede determinar de varias maneras:

- Un número arbitrario que sea mayor o igual a cinco, y menor o igual a 16

$$5 \leq m \leq 16 \quad \text{ó} \quad 5 \leq m \leq 18$$

- Métodos muy utilizados en clase, consiste en la aplicación de la fórmula Sturges

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

- De acuerdo a la clasificación que requiera la variable o la forma en que se ha venido presentando la información.

- Algunos utilizan un procedimiento, no muy recomendado, calculando m mediante la fórmula $m = \sqrt{\log n}$

En la práctica, se determina m atendiendo a varios factores, tales como: finalidad del estudio, grado de variabilidad de los datos, necesidad de efectuar comparaciones con otros estudios que obligue a mantener los intervalos y su número. Sin embargo, se recomienda que el número de intervalos, hasta donde sea posible, no sea menor de 5 ni mayor de 16. Algunos lo admiten hasta 18.

En este ejercicio se tendrá que $m = 1 + 3,3 \log 30 = 5,29 \cong 6$, lo aproximamos a 6, pero se hubiera podido considerar que $m = 5$. La aproximación no tiene importancia alguna en estos casos. Esta fórmula recibe el nombre de Sturges, su descubridor.

En cuanto la amplitud $[C]$ o $[i]$ que debe tomar cada intervalo de la distribución, dependerá del criterio establecido para presentar la información. Puede variar el valor de la amplitud para cada intervalo, sin embargo en clase y con fines de explicación es mejor trabajar con una amplitud constante, para ello se tendrá que:

$$i = C = \frac{\text{Rango}}{m}$$

Reemplazando se tiene que:

$C = 47 \div 6 = 7,83$. Siempre aproximamos al número inmediatamente superior, por pequeña que sea la fracción. Si C hubiese resultado 7,06 o 7,10, se aproxima a 8. Si usted no aproxima al número superior por pequeña que sea la fracción al calcular amplitud, es muy posible que le queden valores observados por fuera del rango.

Para nuestro ejercicio fijamos m igual a 6, con lo cual podemos determinar el valor de $C = [i]$

$$C = \frac{X_{\text{máx}} - X_{\text{mín}}}{m} \quad C = \frac{97 - 47}{6} = \frac{47}{6} = 7,83 \quad C = [i] = 7,83 \cong 8$$

Siempre que el resultado de C sea un valor entero con alguna fracción, por pequeña que ésta sea, debe ser aproximada siempre al valor inmediatamente superior; según esto, el valor de C será de 8, con el cual el recorrido se aumenta en una unidad, puesto que $6 \times 8 = 48$. La unidad de incremento se le puede restar al límite inferior o aumentarse al límite superior. En este caso se lo restamos a 47, quedando el límite inferior en 46, a partir de este valor comenzamos a agregar el valor de C , para formar los distintos intervalos.

En la variable continua. Se designa con $y'_{i-1} - y'_i$ ó $X'_{i-1} - X'_i$, la columna correspondiente a los distintos valores que toma la variable. Al primer valor y'_{i-1} ó X'_{i-1} se denomina límite inferior del intervalo y y'_i ó X'_i al límite superior; la diferencia entre estos límites corresponde al valor de la amplitud del intervalo C ó i . Es preferible con fines de trabajo, que la amplitud sea constante, sin embargo, en la práctica no es muy frecuente que esto se presente.

La tabla de frecuencias para este ejercicio nos quedará de la siguiente forma:

Tabla 2.4

Distribución de frecuencias - variable continua

$y'_{i-1} - y'_i$	n_i	h_i	N_i	H_i	y_i	$y'_{i-1} - y'_i$	n_i
46,1 - 54	3	0,10	3	0,10	50	$y'_0 - y'_1$	y_1
54,1 - 62	6	0,20	9	0,30	58	$y'_1 - y'_2$	y_2
62,1 - 70	8	0,27	17	0,57	66	$y'_2 - y'_3$	y_3
70,1 - 78	6	0,20	23	0,77	74	$y'_3 - y'_4$	y_4
78,1 - 86	4	0,13	27	0,90	82	$y'_4 - y'_5$	y_5
86,1 - 94	3	0,10	30	1,00	90	$y'_5 - y'_6$	y_6
Σ	30	1,00			-	Σ	-
$X'_{i-1} - X'_i$	f_i	f_i/n_i	F_i	H_i	X_i	$X'_{i-1} - X'_i$	f_i

Es de anotar que a cada límite inferior del intervalo se le ha agregado 0,1, lo que permite una fácil tabulación y así se sabe a cuál intervalo corresponde el valor de x_p obtenido. **El 0,1 se omitirá para el cálculo de las marcas de clase y de los diferentes estadígrafos o medidas que se requiera calcular.**

El cuadro de tabulación con el cual se elaboró la tabla de frecuencias, se estableció de la siguiente forma:

Tabla 2.5

Tabulaciones de datos - variable continua

$y'_{i-1} - y'_i$	n_i	TABULACIÓN	
		A	B
46,1 – 54	3	□	III
54,1 – 62	6	☒ I	HHI
62,1 – 70	8	☒ □	HH III
70,1 – 78	6	☒ I	HHI
78,1 – 86	4	□	III
86,1 – 94	3	□	III
Σ	30	30	30

Las *marcas de clase* se pueden obtener de 3 maneras:

a) Se promedia cada intervalo, de la siguiente forma:

$$\begin{aligned}
 [X_1] &= y_1 = \frac{y'_0 + y'_1}{2} = \frac{46 + 54}{2} = 50 = \left[\frac{X'_0 + X'_1}{2} \right] \\
 [X_2] &= y_2 = \frac{y'_1 + y'_2}{2} = \frac{54 + 62}{2} = 58 = \left[\frac{X'_1 + X'_2}{2} \right] \\
 [X_3] &= y_3 = \frac{y'_2 + y'_3}{2} = \frac{62 + 70}{2} = 66 = \left[\frac{X'_2 + X'_3}{2} \right] \\
 [X_4] &= y_4 = \frac{y'_3 + y'_4}{2} = \frac{70 + 78}{2} = 74 = \left[\frac{X'_3 + X'_4}{2} \right] \dots, \text{ etc.}
 \end{aligned}$$

b) Se obtiene la primera marca de clase por el método anterior y si la amplitud C es constante, se le suma a la primera marca de clase obtenida y así sucesivamente.

$$y_1 = \frac{y'_0 + y'_1}{2} = \frac{46 + 54}{2} = 50 = \left[\frac{X'_0 + X'_1}{2} \right] \quad [X_2] = X_2 = 50 + 8 = 58 \quad [X_3] = X_3 = 58 + 8 = 66$$

c) Se divide la amplitud de cada intervalo por dos y se le suma al límite inferior del intervalo o se le resta al límite superior del intervalo.

$$\left[\frac{i}{2} \right] = \frac{C}{2} = \frac{8}{2} = 4$$

$$\begin{aligned}
 y'_0 + \frac{C}{2} &= y_1 & [X_1] &= y_1 = 46 + 4 = 50 & [X_2] &= y_2 = 54 + 4 = 58, \text{ etc.} \\
 y'_1 - \frac{C}{2} &= y_1 & [X_1] &= y_1 = 54 - 4 = 50 & [X_2] &= y_2 = 62 - 4 = 58, \text{ etc.}
 \end{aligned}$$

Nota: una distribución es simétrica cuando las frecuencias absolutas o relativas, equidistantes a un punto central son iguales, además, la gráfica que la representa tiene forma acampanada o convexa como se observa con los datos que se presentan a continuación:

Tabla 2.6

A continuación se presenta dos ejemplos de Distribuciones simétricas:

a)	$y'_{i-1} - y'_i$	n_i	h_i	b)	y_i	n_i	h_i
	46,1 – 54	2	0,05		2	2	0,10
	54,1 – 62	6	0,15		4	5	0,25
	62,1 – 70	12	0,30		6	6	0,30
	70,1 – 78	12	0,30		8	5	0,25
	78,1 – 86	6	0,15		10	2	0,10
	86,1 – 94	2	0,05		Σ	20	1,00
	Σ	40	1,00		X_i	f_i	f_i/n
	$X'_{i-1} - X'_i$	f_i	f_i/n				

Propiedades de las frecuencias

- Las frecuencias absolutas son siempre valores enteros.
- La suma de las frecuencias absolutas es igual a n .

$$\sum_{i=1}^n n = n$$

$$\sum_{i=1}^5 n = n = n_1 + n_2 + n_3 + n_4 + n_5 = n$$

$$\sum_{i=1}^5 n_i = 30 = 4 + 5 + 7 + 8 + 6 = 30$$

$$\sum_{i=1}^n f_i = n$$

$$\sum_{i=1}^5 f_i = f_1 + f_2 + f_3 + f_4 + f_5 = n$$

$$\sum_{i=1}^5 f_i = 30$$

- Las frecuencias relativas son siempre valores fraccionarios.

$$0 < h_i < 1$$

$$0 < f_i/n < 1$$

- La suma de las frecuencias relativas es igual a 1

$$\sum h_i = 1$$

$$\sum_{i=1}^5 h_i = h_1 + h_2 + h_3 + h_4 + h_5 = 1$$

$$\sum_{i=1}^5 h_i = 0,13 + 0,17 + 0,23 + 0,27 + 0,20 = 1,00$$

$$\sum f_i/n = 1$$

$$\sum_{i=1}^5 f_i/n = \frac{f_1}{n} + \frac{f_2}{n} + \frac{f_3}{n} + \frac{f_4}{n} + \frac{f_5}{n} = 1$$

- El último valor de las frecuencias absolutas acumuladas es igual a n .

$$N_m = n \quad N_5 = n \quad 30 = 30 \quad \left[F_m = n \right] \quad \left[F_5 = n = 30 \right]$$

- El último valor correspondiente a las frecuencias relativas acumuladas debe ser igual a 1.

$$H_m = 1 \quad H_5 = 1 \quad \left[F_m / n = 1 \right] \quad \left[F_6 / n = 1 \right]$$

Ejemplo 3.

- a. Indique dos ejemplos de distribuciones unidimensionales, dos de bidimensionales y dos de multidimensionales, relacionadas con la economía del país. b. Señale cuatro distribuciones de variable discreta, tres de variable continua y tres de atributos.

Solución:

- | | |
|------------------------------------|---|
| a. Unidimensional
(univariante) | 1. Producción |
| | 2. No. de explotaciones agrícolas |
| | 3. Ingresos |
| Bidimensional | 1. Producción e inversión (bivariante) |
| | 2. No. de explotaciones agrícolas y hectáreas cultivadas. |
| | 3. Ingresos y gastos. |
| Pluridimensional
(polivariante) | 1. Producción; inversión; horas laborales |
| | 2. No. de explotaciones agrícolas, hectáreas cultivadas número de obreros |
| | 3. Ingreso, gasto, No. de familias |
| b. Variable discreta | 1. No. de hijos por familia |
| | 2. No. de hermanos por alumno. |
| | 3. No. de billetes por numeración |
| | 4. No. de piezas defectuosas. |
| Variable continua | 1. Ingresos |
| | 2. Peso en Kg. |
| | 3. Longitud. |
| c. Atributos | 1. Exportaciones por puertos colombianos. |
| | 2. Número de alumnos por Universidad en Caracas. |
| | 3. Consumo de cigarrillos por marcas. |

Ejemplo 4.

En los ejemplos siguientes señalar con una **A** las series constituidas por datos cualitativos (atributos) y con una **V** los datos cuantitativos (variables).

Solución:

- | | |
|---|-----|
| a. Distribución de alumnos por mes de nacimiento | (A) |
| b. Distribución de alumnos por nacionalidad | (A) |
| c. Distribución de profesionales por estatura y peso. | (V) |
| d. Distribución de oficiales por sueldo. | (V) |
| e. Distribución de accidentes por causa. | (A) |

Ejemplo 5.

Señale con **C** las series de variable continua y con **D** las de variable discreta.

Solución:

- a. Distribución de obreros por salarios (C)
 b. Distribución de fallecimientos por edades (C)
 c. Distribución de alumnos por número de hermanos (D)
 d. Distribución de alumnos por estatura. (C)

Ejemplo 6.

Leer el siguiente texto: “Una vez recolectados los datos en forma ordenada, es necesario presentarlos en forma tal que se facilite su comprensión y su posterior análisis y utilización. Para ello se ordenan en cuadros numéricos y luego se representan en gráficos, para variable discreta mediante diagramas de frecuencias, tanto para absolutas o relativas”.

- a. Formar una tabla de frecuencias absolutas (n_i), según el número de letras que componen cada palabra (rr y ll, debe considerarse una sola letra).
 b. Considerando el número de letras por palabra. ¿Qué valor tienen x_4 x_{27} x_{12} x_{34} x_2 ?
 c. En la tabla de frecuencias del punto a. ¿Qué valores poseen y_2 y_4 y_7 ?
 d. Análogamente ¿Qué valores poseen n_2 n_4 n_7 ?
 e. ¿Qué valores poseen h_2 h_4 h_7 ?
 f. Calcular las frecuencias relativas acumuladas (H_i) y las frecuencias absolutas acumuladas (N_i).
 g. Puede ocurrir que $h_4 > h_7$ (en general). ¿Qué $H_4 < H_7$?

Solución:

x_1 = una = 3 letras

x_2 = vez = 3 letras, etc.

a)

$x_1 = 3$	$x_{11} = 12$	$x_{21} = 2$	$x_{31} = 7$	$x_{41} = 8$
$x_2 = 3$	$x_{12} = 2$	$x_{22} = 9$	$x_{32} = 9$	$x_{42} = 8$
$x_3 = 12$	$x_{13} = 5$	$x_{23} = 8$	$x_{33} = 1$	$x_{43} = 9$
$x_4 = 3$	$x_{14} = 3$	$x_{24} = 1$	$x_{34} = 5$	$x_{44} = 2$
$x_5 = 5$	$x_{15} = 3$	$x_{25} = 11$	$x_{35} = 2$	$x_{45} = 11$
$x_6 = 2$	$x_{16} = 2$	$x_{26} = 4$	$x_{36} = 11$	$x_{46} = 5$
$x_7 = 5$	$x_{17} = 8$	$x_{27} = 3$	$x_{37} = 2$	$x_{47} = 4$
$x_8 = 8$	$x_{18} = 2$	$x_{28} = 2$	$x_{38} = 8$	$x_{48} = 9$
$x_9 = 2$	$x_{19} = 11$	$x_{29} = 7$	$x_{39} = 4$	$x_{49} = 1$
$x_{10} = 9$	$x_{20} = 1$	$x_{30} = 2$	$x_{40} = 8$	$x_{50} = 9$

y_i	n_i	h_i	N_i	H_i
1	4	0,08	4	0,08
2	11	0,22	15	0,30
3	6	0,12	21	0,42
4	3	0,06	24	0,48
5	5	0,10	29	0,58
7	2	0,04	31	0,62
8	7	0,14	38	0,76
9	6	0,12	44	0,88
11	4	0,08	48	0,96
12	2	0,04	50	1,00
Σ	50	1,00	-	-
X_i	f_i	f_i/n	F_i	N_i/n

- b) $x_4 = 3$ $x_{27} = 3$ $x_{12} = 2$ $x_{34} = 5$ $x_2 = 3$

- c) $y_2 = 2 = X_2$ $y_4 = 4 = X_4$ $y_7 = 8 = X_7$

- d) $n_2 = 11 = f_2$ $n_4 = 3 = f_4$ $n_7 = 7 = f_7$

- e) $h_2 = 0,22$ $h_4 = 0,06$ $h_7 = 0,14$

$$f_2/n = 0,22$$

$$f_4/n = 0,06$$

$$f_7/n = 0,14$$

- f) En general puede ocurrir que $h_4 > h_7$ ó $h_7 > h_4$
 Pero siempre se tendrá $H_4 < H_7$ y nunca se dará que $H_4 > H_7$



EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL**.

1. En un grupo de familias considerando el número de hijos, se han obtenido los siguientes valores:

2	0	2	4	4	6	6	4	6	7	4	4	7	4	2	0	4	6	7	7
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Construir una tabla con base en los anteriores datos. Calcular las columnas correspondientes a: n_i , h_i , N_i y H_i .

2. Dé contestación a los siguientes puntos, señalando con una X si es cierto o falso.

	Cierto	Falso
a. $H_5 = 0,36$ $N_4 = 30$ $n_5 = 6$ $n = 50$	()	()
b. Para calcular las marcas de clase, se suma el límite inferior al superior del intervalo y se divide entre dos.	()	()
c. El número de accidentes según sus causas, es una variable discreta.	()	()
d. Si $H_6 = 0,7$ y $H_4 = 0,3$ un 40% de los valores de la variable es menor que Y_6 y mayor que Y_4 .	()	()
e. $H_5 = 1,10$ $H_4 = 0,80$ $h_5 = 0,30$	()	()
f. La suma de las frecuencias absolutas es igual a 1.	()	()

3. Reconstruya la siguiente distribución simétrica.

y_i	n_i	H_i
10	6	0,12
20	-	0,32
30	-	-
40	-	-
50	-	-
X_i	f_i	F_i/n

4. Se realiza un estudio en la ciudad de Guayaquil a 150 familias de clase media, para conocer el tipo de aceite o manteca usados en la cocina. Los resultados son los siguientes: Maíz, 14 hogares; Soya, 65 hogares; Ajonjolí, 21 hogares; compran aceite al detal sin especificar tipo, 17 hogares; manteca de cerdo, 21 hogares; grasas de origen vegetal, 6 hogares; Oliva, 13 hogares.
- ¿Cuál es la población?
 - ¿Cuál es el tamaño de la muestra?
 - ¿Qué carácter tiene la población?
 - ¿Cuál es la característica?
 - ¿Cuántas clases de características tiene la distribución?
 - Construya una distribución de frecuencias en el mismo orden indicado.
 - ¿Cuál es la 5ª clase?
 - ¿Cómo se explica que la suma de frecuencias sea superior al número de hogares?