

# CAPITULO UNO

## Concepto de Estadística

Es una ciencia que trata sobre la recopilación, organización, presentación, análisis e interpretación de los datos numéricos, con el fin de tomar una decisión más efectiva.

## Objetivo:

- Estimar las propiedades de una población a partir del conocimiento de las propiedades de sólo una muestra de ella (Inferencia).

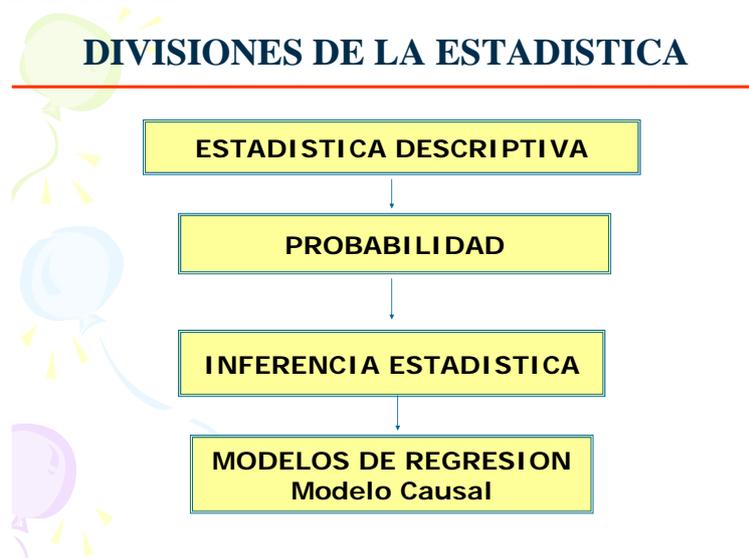
## Diferencia entre estadísticas y estadística

A un dato numérico o valor aislado se le denomina dato estadístico. A un conjunto de datos numéricos presentados para caracterizar a una población se le denomina estadística. El estudio general de las estadísticas se define como la ciencia estadística

## Uso de la estadística

Las técnicas estadísticas se utilizan de manera amplia en mercadotecnia, contabilidad, control de calidad y en otras actividades como: estudios de consumidores, análisis de resultados en deportes, administración de instituciones, en la toma de decisiones, etc. Igualmente se usa en el área social, y los aplican los analistas de investigación, los partidos políticos, los directivos de las empresas, los consumidores, etc.

## Divisiones de la estadística



La estadística se divide en:

1. **La estadística descriptiva**, comprende la tabulación, representación y descripción de una serie de datos que pueden ser cuantitativas, como medidas de estatura y peso o cualitativas como sexo y tipo de personalidad.

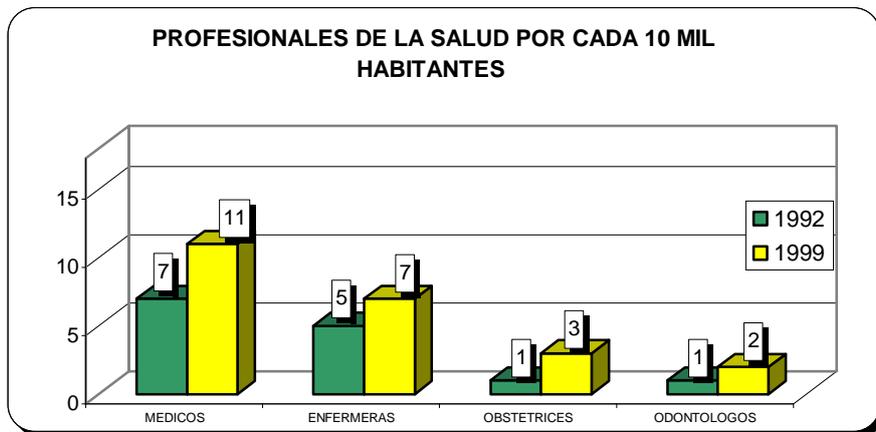
2. **La estadística inferencial**, son procedimientos estadísticos que sirven para deducir o inferir algo acerca de un conjunto de datos numéricos (población), seleccionando un grupo menor de ellos (muestra).

**Tipos de variables:**

Los tipos de variables de la estadística son:

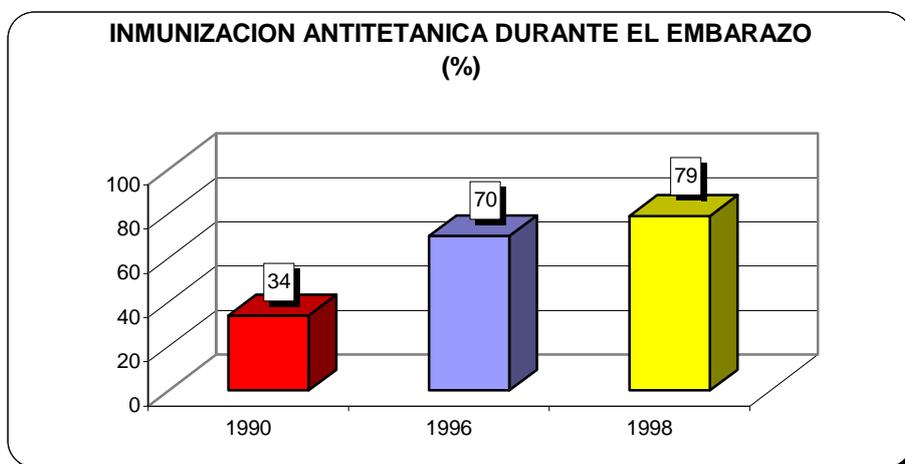
**Variable cualitativa**, es cuando la variable en estudio no es numérica. Por ejemplo el género, la religión, las profesiones, atenciones por tipo de consulta etc.

**Variable cuantitativa**, es cuando la variable estudiada es expresable numéricamente. Por ejemplo, el ingreso, el peso, la talla, edad, el valor de las ventas, o compras, etc. Estas a su vez pueden ser discretas o continuas.



**Variable discreta**, es decir, que su medida no puede dar a lugar a cualquier valor. Se llama así si su rango o conjunto de valores sólo puede ser finito o infinito enumerable. Normalmente puede tomar como valor un entero. Por ejemplo el número de médicos, enfermeras, obstetricas por departamento.

**Variable continua**, es decir, que dentro de determinado rango, su medición puede dar a lugar a cualquier valor. Se le identifica como un valor en el campo de los números reales.



### **Nivel de medición**

Los niveles de medición pueden ser de los tipos siguientes:

**Medición nominal**, es cuando se clasifican los objetos y las clases se nominan.

**Ejemplo:** el origen racial, color de los ojos, los diagnósticos clínicos, el tipo de alumbrado, los departamentos, provincias o distritos de un país, las enfermedades,.. etc.

**Medición ordinal**, son los tamaños relativos de los números asignados a los objetos, quienes reflejan las cuantías de los atributos que los objetos poseen. Por ejemplo, el orden de mérito de los estudiantes de análisis matemático, ordenación en un rasgo tipo de personalidad.

**Medición por intervalo**, es una unidad de medida que permite que los objetos no sólo puedan ser ordenados, sino también asignados a ciertos números, de tal manera que unas diferencias iguales entre los números asignados a esos objetos reflejen diferencias idénticas en las cuantías de los atributos que se han medido. Por ejemplo, el día, las temperaturas termométricas etc. El cero no tiene un significado especial.

**Medición de razón o cociente**, son los números asignados a los objetos que tienen todas las propiedades de la escala intercalar, pero el cero significa la ausencia de algo. Por ejemplo, la edad que se asigna a un individuo, el peso que se le asigna a un escolar, etc.

### **EL MÉTODO ESTADÍSTICO.**

Se encarga de la recopilación, organización, procesamiento, análisis e interpretación de la información. Para la aplicación del método estadístico se requiere tener los objetivos bien definidos, para que el diseño de la investigación sea el más adecuado.

La **recopilación** puede referirse a estudios o trabajos ya realizados; sólo en caso que no existiera se aplica esta etapa. Las fuentes de datos son:

Las Encuestas y los registros administrativos. Las Encuestas a su vez pueden ser por censo que es una recopilación a todas las unidades de la población, o en base a un muestreo a una porción de ella.

La **organización** de la información se efectúa para facilitar el procesamiento, es una forma de ordenar la información que debe responder a las necesidades de los usuarios. En el campo económico se aplican clasificaciones internacionales para fines de comparabilidad.

El **procesamiento** comprende a su vez la crítica y codificación de la data y su ingreso automatizado.

El **análisis** es referido a la posibilidad de inferir los resultados de la muestra a la población. Ello se realiza construyendo un intervalo donde se supone está el verdadero valor del parámetro con un nivel de confianza.

La **interpretación** de los resultados se aplica cuando ya se tiene la seguridad que la información es confiable.

## CAPITULO DOS

### DISTRIBUCIONES DE FRECUENCIAS Y REPRESENTACIONES GRAFICAS

#### Objetivo

La distribución de frecuencias es la primera forma de sintetizar la información proveniente de las investigaciones estadísticas. Tiene por objetivo presentar la distribución de las variables en estudio a fin de poder aplicar posteriormente el método más adecuado para estimar los promedios y su dispersión absoluta o relativa.

#### Elaboración de una distribución de frecuencias

Es un método estadístico muy útil para organizar un conjunto de observaciones en forma significativa. Además indica el número de veces que ocurre cada valor o dato en cada clase.

Los pasos para elaborar una distribución de frecuencia son los siguientes:

1. Determinación del rango.
2. Selección de los intervalos de clase.
3. Determinar los límites de las clases.
4. Efectuar la tabulación.

#### Los gráficos

Por muy expresiva que haya tenido la intención de hacerla su autor, una tabla descriptiva o una descripción sintética exigen una cierta educación por parte del que la utiliza.

La consulta de los números resulta árida, y ciertas veces, por otro lado perfectamente normales, tiende a interpretarlos erróneamente.

Por ello la representación gráfica, cuyo objeto es representar las distribuciones estadísticas por magnitudes o figuras geométricas, constituyendo los gráficos, es un gran recurso para las descripciones.

El gráfico es el instrumento ideal para la comparación de distribuciones mediante su superposición en la misma hoja o en hojas transparentes distintas.

#### DISTRIBUCION DE FRECUENCIAS DE VARIABLES DISCRETAS CON POCA VARIABILIDAD (TIPO I)

Cuando el rango de variación es muy pequeño, la distribución de frecuencias de la variable discreta se hace en función de cada uno de los datos.

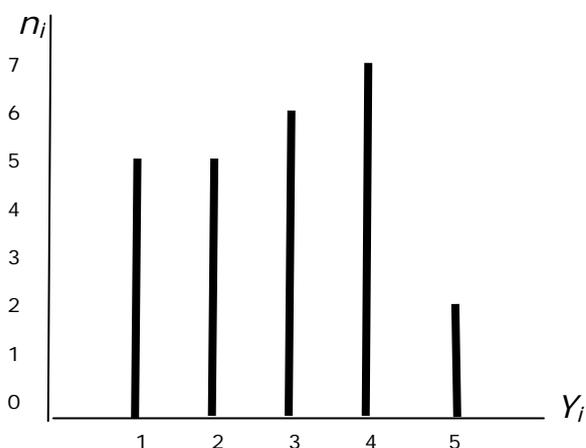
Ejemplo: En base a información del número de hijos que tienen cada una de las 25 madres de una maternidad se ha elaborado la siguiente distribución de frecuencias.

Nº de hijos	Frec. Abs.	Frec. Relativa	Frec. Abs. Acum.	Frec. Rel. Acum.
	$n_i$	$h_i$	$N_i$	$H_i$
0	0	0		
1	5	0,20	5	0,20
2	5	0,20	10	0,40
3	6	0,24	16	0,64
4	7	0,28	23	0,92
5	2	0,08	25	1,00
<b>TOTAL</b>	<b>25</b>	<b>1</b>		

### GRAFICOS DE DISTRIBUCION DE FRECUENCIAS DE VARIABLE DISCRETA: FORMA I

Para el gráfico de este tipo de variables; así como el de variables continuas; se utiliza el primer cuadrante de un sistema de coordenadas en el cual en el eje horizontal (eje de abscisas) se consideran los valores que toma la variable y en el eje vertical (eje de las ordenadas) cualquiera de las frecuencias.

#### Gráfico de Frecuencias Absolutas ( $n_i$ )



#### Gráfico de Frecuencias Absolutas y Relativas

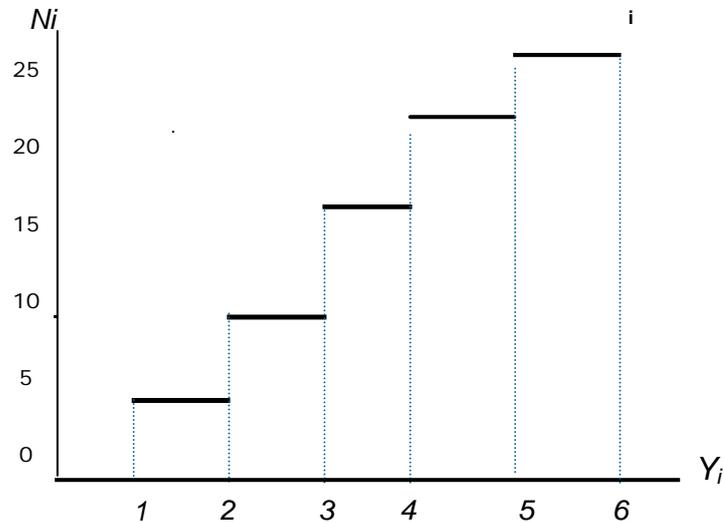
El gráfico se obtiene de trazar perpendiculares al eje horizontal, desde el punto de intersección de cada valor de la variable con su respectiva frecuencia absoluta o frecuencia relativa, según sea el caso.

#### Gráfico de Frecuencias Absolutas y Relativas (Acumuladas)

En este tipo de gráfico, se levanta sobre cada uno de los valores de la variable un segmento vertical discontinuo de longitud equivalente a la frecuencia acumulada respectiva. Luego se

dibujan diversos tramos horizontales correspondientes a los intervalos dentro de los cuales no pueden existir observaciones, dada la índole de la variable.

### Gráfico de Frecuencias Absolutas Acumuladas ( $N_i$ )



### ELABORACIÓN DE TABLAS DE FRECUENCIA DE VARIABLE DISCRETA: FORMA II

Cuando existen abundante información y/o la dispersión es alta. En estas condiciones, la presentación de cuadros de variables discretas se puede realizar con intervalos de clase cerrados.

Ejemplo aplicativo: se ha realizado una encuesta sobre el número de conexiones domiciliarias que se han ejecutado en el mes de noviembre de 2001; en 40 distritos de la provincia de Lima, los resultados son los siguientes:

42	58	79	86	98	120	134	120
59	62	85	89	75	110	104	78
84	96	90	75	120	130	122	95
82	94	108	79	105	115	102	80
56	78	84	66	69	78	84	98

Pasos a seguir para elaborar la tabla de frecuencias

- 1.- Valor máximo 134  
Valor mínimo 42
- 2.- Valor máximo - valor mínimo  $134 - 42 + 1 = 93$
- 3.- Determinar el número de intervalos de clase:

### Conceptos previos

**Clase.-** Es cada una de las agrupaciones en las cuales se han dividido el conjunto de la información ( $y_i$ ). Las mismas se han seguido un ordenamiento sucesivo de los valores, menor a mayor o viceversa.

**Amplitud de clase.** - Es la diferencia entre el límite superior y el límite inferior de cada intervalo de clase. Cuando sea posible debe formarse un cuadro de frecuencias con intervalos de igual amplitud; aunque algunas veces, por el tipo de información se deben hacer intervalos de clase distinta amplitud.

**Marca de clase.** - Es el punto medio de cada intervalo de clase y se obtiene de la semisuma de los límites inferior y superior:

$$MC = \frac{L_i + L_s}{2}$$

La marca de clase, se utiliza como una aproximación al promedio real, en cada intervalo de clase; su valor representa a cada agrupación.

### SUGERENCIAS

- Los intervalos de clase utilizados en una distribución de frecuencia deben tener la misma amplitud, porque los intervalos de clase de diferentes rangos (o amplitud) ofrecen problemas al representarse en una gráfica.
  - Evite tener menos de cinco clases o más de quince clases.
  - Evite tener clases traslapadas (superpuestas).
- 4.- Obtención de la amplitud de cada intervalo

$$A = \frac{R}{NI} \quad \text{donde } R = \text{Rango o recorrido}$$

NI = Número de intervalos

Ejemplo

Con del ejemplo aplicativo

$$A = \frac{93}{5} = 18.5 = 19$$

- 5.- Nuevo rango o recorrido, se modifica cuando existe un decimal aunque sea pequeño; se redondea la amplitud del intervalo hacia el número inmediato superior.

$$R^1 = NI \times A$$

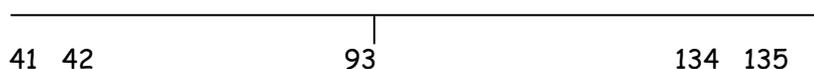
$$R^1 = 5 \times 19 = 95$$

- 6.- Obtención de la diferencia

$$d = R^1 - R$$

$$d = 95 - 93 = 2$$

- 7.- Prorrrateo gráfico de la diferencia



Se coloca una unidad en cada extremo.

Generalmente se reparte equitativamente en los extremos las unidades de diferencia.

Observación: El número de intervalos de clase también se puede obtener mediante la propuesta de Sturges =  $1 + 3\text{Ln}(n)$ , donde  $n$  es el número de observaciones.

#### 8.- Elaboración de la tabla de frecuencias con intervalos cerrados

Intervalos de clase	Marca de clase	Nº de Distritos Tabulación	$n_i$	$N_i$	$h_i$	$H_i$
41 - 59	50	////	4	4	0.1	0.1
60 - 78	69	<del>///</del> ///	8	12	0.2	0.3
79 - 97	88	<del>///</del> <del>///</del> ////	14	26	0.35	0.65
98 - 116	107	<del>///</del> ///	8	34	0.2	0.85
117-135	126	<del>///</del> /	6	40	0.15	1.00

$n = 40$

La marca de clase (M.C.) se obtiene de la semisuma de cada intervalo de clase.

$$\frac{41 + 59}{2} = 50$$

Una vez obtenida la primera M.C las siguientes también se pueden obtener al sumarlas con las amplitudes (19 en este caso)

$$50 + 19 = 69 + 19 = 88 + 19 = 107 + 19 = 126$$

$$\frac{60 + 78}{2} = 69$$

$$\frac{79 + 97}{2} = 88$$

$$\frac{98 + 116}{2} = 107$$

$$\frac{117 + 135}{2} = 126$$

#### Intervalo cerrado

$[41 - 59]$  se incluye ambos valores de los extremos

#### Intervalo semiabierto por la izquierda

$(41 - 59]$  se incluye solo el valor de la derecha

#### Intervalo semiabierto por la derecha

$[41 - 59)$  se incluye solo el valor de la izquierda

#### Intervalo abierto

$(41 - 59)$  no se incluye ninguno de los valores consignados

### Interpretación de la tabla de frecuencia

$n_5$  = En 6 distritos el número de conexiones domiciliarias varía entre 117 y 135

$N_2$  = En 12 distritos el número de conexiones domiciliarias es a lo más 78

$h_4$  = De los 40 distritos en el 20%; el número de conexiones domiciliarias varía entre 117 y 135

$H_3$  = En el 65% de los distritos el número de conexiones domiciliarias varía entre 41 y 97.

¿Cuál es el número de distritos cuyo número de conexiones domiciliarias varía entre 60 a 97?

La respuesta sería: la suma de  $n_2 + n_3 = 8+14=22$  distritos

### Distribución de frecuencias relativas

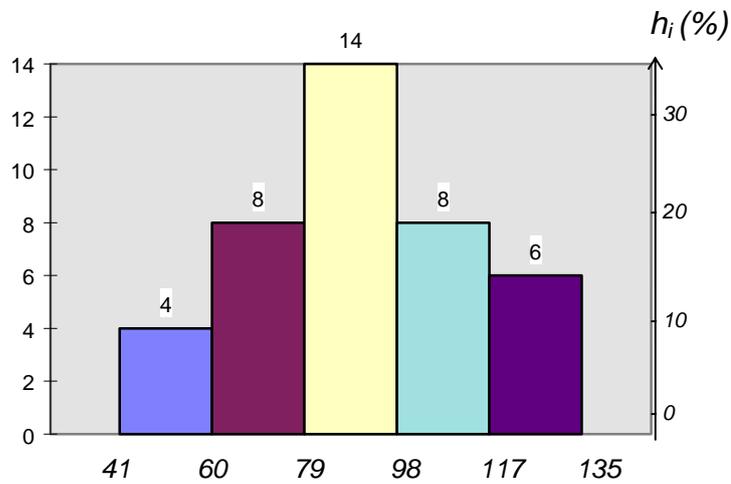
Para convertir una distribución de frecuencias absolutas a una distribución de frecuencias relativas, cada una de las frecuencias de clase se divide entre el número total de observaciones.

Ingresos mensuales	Frecuencia absoluta	Frecuencia relativa
300-400	3	0,025
400-500	7	0,058
500-600	11	0,092
600-700	21	0,175
700-800	41	0,342
800-900	24	0,200
900-1000	9	0,075
1000-1100	4	0,033
Total	120	

### GRÁFICAS DE DISTRIBUCIONES DE FRECUENCIA DE VARIABLE DISCRETA DE TIPO II

En el caso de variables discretas agrupadas en intervalos, la representación gráfica de las frecuencias absolutas y relativas, se hace mediante rectángulos alargados, cuya longitud será proporcional a la magnitud representada, y la amplitud de la base será de acuerdo a los intervalos de clase, dejando un espacio de una unidad entre cada intervalo, para caracterizar la discontinuidad de la variable.

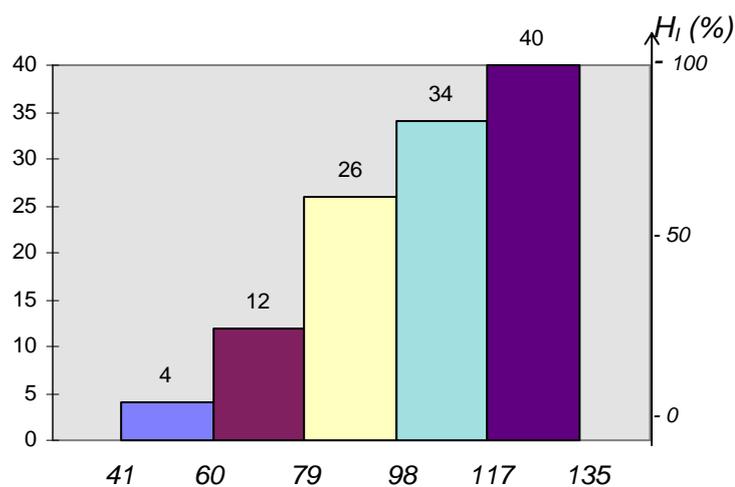
### Gráfico de frecuencias absolutas y relativas



Cuando las amplitudes de los intervalos son mayores, el espacio al quedar vacío es ínfimo, haciéndose los rectángulos continuos; convirtiéndose de esta forma los gráficos en histogramas.

En el gráfico de las frecuencias acumuladas se levanta sobre los límites de cada intervalo un segmento vertical, de longitud equivalente a la frecuencia acumulada respectiva. Luego se dibujan diversos tramos horizontales correspondientes a la amplitud de cada intervalo, alargándola en una unidad; como en esta no existe nunca observaciones, el gráfico está sin adulteraciones.

### Gráfico de frecuencias absolutas y relativas acumuladas



**Ejemplo:**

A continuación daremos los ingresos mensuales en nuevos soles, de 120 trabajadores de una empresa ubicada en Lima, en el mes de enero del año 1999.

**Ingresos mensuales (nuevos soles)**

585	604	791	639	653	736	539	660	769	925
666	709	975	702	872	766	610	448	750	836
736	701	521	690	411	779	559	767	755	880
913	655	713	644	697	773	516	645	348	402
720	711	665	702	359	729	725	728	1026	839
560	510	702	721	797	981	632	894	751	834
676	670	730	912	726	569	796	491	991	546
714	802	850	619	663	795	571	713	775	457
735	892	809	716	779	448	831	796	776	806
625	710	1081	687	771	816	784	611	986	857
475	770	817	819	825	804	820	870	770	1094
876	824	989	320	868	611	895	594	1046	915

Se revisa el cuadro, con el fin de hallar la renta (nuevos soles) más baja (S/.320), posteriormente se ordenan los valores desde el menor hasta el mayor.

**Arreglo ordenado de los 120 ingresos**

320	521	611	666	711	735	773	804	839	913
348	539	619	670	713	736	775	806	850	915
359	546	625	676	713	736	776	809	857	925
402	559	632	687	714	750	779	816	868	975
411	560	639	690	716	751	779	817	870	981
448	569	644	697	720	755	784	819	872	986
448	571	645	701	721	766	791	820	876	989
457	585	653	702	725	767	795	824	880	991
475	594	655	702	726	769	796	825	892	1026
491	604	660	704	728	770	796	831	894	1046
510	610	663	709	729	770	797	834	895	1081
516	611	665	710	730	771	802	836	912	1094

Luego:

1. Establecer un conjunto de agrupaciones, llamado también clase. Para ello se debe seguir los siguientes pasos:
  - Se determina el valor máximo, 1094 nuevos soles, y el valor mínimo, 320 nuevos soles.
  - Se resta el valor mínimo del valor máximo:  
Valor máximo - Valor mínimo = 1094 - 320 = 774
  - Determinar el número de intervalos de clase (8)

- Luego debemos obtener la amplitud (A) de cada intervalo

$$A = \frac{R}{NI} = \frac{774}{8} = 96$$

R = Rango o recorrido  
NI = Número de intervalos

En este caso, por comodidad de trabajo, se ha redondeado la amplitud a 100, y por la misma razón el primer intervalo empezará en 300.

El primer intervalo terminará en:  $300 + A = 300 + 100 = 400$

Entonces la primera clase será: 300 - 400

Una clase puede contener todos los ingresos. Por ejemplo la primera comprende los que tienen ingresos desde S./ 300 hasta menos de S./ 400. Es decir es cerrada a la izquierda y abierta a la derecha (semiabierto a la derecha). Las siguientes clases se hallan de la misma forma.

Ingresos mensuales
300-400
400-500
500-600
600-700
700-800
800-900
900-1000
1000-1100

2. Llevar la cuenta de los valores en las clases. La práctica común es utilizar una marca de cuenta (/) para señalar un valor. Por ejemplo el ingreso de S./ 594 de la esquina superior izquierda del cuadro que contiene los datos originales se marca para la cuenta en la clase de S./ 500-S./600 y así sucesivamente. Al terminar, las marcas deben quedar como sigue:

Ingreso Mensual	Tarjas
300 -400	///
400 -500	//// //
500 -600	//// //// /
600 -700	//// //// //// //// /
700 -800	//// //// //// //// //// //// //// /
800 -900	//// //// //// //// ////
900 -1000	//// ////
1000-1100	////

3. Contar el número de marcas en cada clase. En el ejemplo de ingresos mensuales:

Ingresos Mensuales	Frecuencia Absoluta
300-400	3
400-500	7
500-600	11
600-700	21
700-800	41
800-900	24
900-1000	9

1000-1100	4
Total	120

**Limite de clases**

Los límites son tales que el límite superior verdadero de una clase es igual al límite inferior verdadero de la siguiente.

**Puntos medios**, se determina localizando la mitad entre los límites de clase declarados o los límites de clase verdaderos. Se evalúa sumando los límites inferior y superior, y dividiendo el total entre dos.

**Intervalo de clase**

Es el símbolo que caracteriza a cada clase y consta de un límite inferior ( $L_i$ ) y de un límite superior ( $L_s$ ). Las mismas que hacen siguiendo un ordenamiento sucesivo de los valores, de menor a mayor o viceversa.

Existen tres tipos de intervalos de clase de acuerdo con los tamaños de las clases en una distribución de frecuencia:

1. clases de igual tamaño
2. clases de desigual tamaño
3. clases abiertas.

**Distribución de frecuencias relativas**

Para convertir una distribución de frecuencias absolutas a una distribución de frecuencias relativas, cada una de las frecuencias absolutas se divide entre el número de observaciones.

**Ejercicios**

**1.- ¿Qué es una ordenación o arreglo ordenador?**

La ordenación o arreglo ordenador es un proceso en la cual se reorganizan los valores desde el menor hasta el mayor.

**2.- ¿Que entiende por distribución de frecuencias?**

Una distribución de frecuencias es una agrupación de datos o valores numéricos dentro de categorías mutuamente excluyentes.

**3.- Una distribución de frecuencias debe ser mutuamente excluyente.**

**Explique.**

Efectivamente una distribución de frecuencias tiene que ser mutuamente excluyente ya que una persona o dato no puede estar en más de una categoría al mismo tiempo.

**4.- El director del programa de reconocimiento de una universidad tiene 16 solicitudes para admisión próxima. Las puntuaciones especiales de los solicitantes son.**

27	27	27	28	27	25	25	28
26	28	26	28	31	30	26	26

Tales puntuaciones han de ser organizadas en una distribución de frecuencias.

- a.- ¿Cuántas clases recomendaría?
- b.- ¿Qué intervalo de clases sugeriría usted?
- c.- ¿Qué límite inferior recomendaría para la primera clase?
- d.- Organice las puntuaciones en una distribución de frecuencias.
- e.- Comente acerca de la forma de la distribución

**Solución:**

- a.- El número de clases recomendadas serían de 4.
- b.- Aplicando la fórmula de la amplitud de clase, tenemos:

$$\text{Amplitud} = \frac{\text{Valor más alto} - \text{valor más bajo}}{\text{Número de clases}}$$

$$\text{Amplitud} = \frac{31 - 25}{4} = 1.5$$

Por lo tanto la amplitud de clase sería 1.5, para mayor comodidad es recomendable redondear a un valor más próximo, es decir a 2.0. Los intervalos de clases serían los siguientes:

- 24 - 25
- 26 - 27
- 28 - 29
- 30 - 31

- c.- El límite inferior recomendado es 24 como se puede observar en la respuesta b del ejercicio.
- d.- Organizando las puntuaciones de la frecuencia quedaría:

Puntuaciones de los solicitantes	Frecuencia Absoluta
24 - 25	2
26 - 27	8
28 - 29	4
30 - 31	2
Total	16

- 5.- El Director de Operaciones de una institución de salud está interesado en el número de atenciones que se solicitan en el área de enfermedades cardiovasculares en 51 establecimientos de salud. El número de atenciones que se registraron fueron las siguientes:

5 3 3 1 4 4 5 6 4 2 6 6 6 7 1

1 14 1 2 4 4 4 5 6 3 5 3 4 5 6  
 8 4 7 6 5 9 11 3 12 4 7 6 5 15 1  
 1 10 8 9 2 12

- a.- Iniciando con 1 como el límite inferior de la primera clase y utilizando el intervalo de clase de 2, organice los datos en una distribución de frecuencias.
- b.- Describa la distribución. ¿Dónde tienden los datos a acumularse?
- c.- Encuentre la distribución de frecuencias relativas.

**Solución**

a.- Organizando los datos en una distribución de frecuencia quedaría así:

Número de atenciones	Frecuencia Absoluta
1 - 3	14
4 - 6	24
7 - 9	7
10 - 12	4
13 - 15	2
Total	51

- b.- En este caso la mayor concentración está en el intervalo 4 - 6.
- c.- Aplicando la fórmula para hallar las frecuencias relativas, se obtiene:

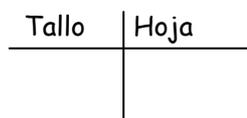
Número de Atenciones	Frecuencia Absoluta	frecuencia relativas
1 - 3	14	0.275
4 - 6	24	0.471
7 - 9	7	0.137
10 - 12	4	0.078
13 - 15	2	0.039
Total	51	1

**REPRESENTACIONES DE TALLO Y HOJA**

El objetivo de una representación tallo y hoja es organizar datos no agrupados en forma significativa.

El diagrama de hojas y tallos de TUKEY es un procedimiento semi - gráfico (tabular y gráfico) de presentar la información para datos cuantitativos, especialmente útil cuando el número total de observaciones es pequeño (menor de 50). Los principios básicos para construirlo son:

- a. Redondear los datos a dos de tres cifras significativas, expresándolos en unidades convenientes.
- b. Disponerlos en una tabla con dos columnas separadas por una línea como sigue:
  - Para datos con dos dígitos, escribir a la izquierda de la línea los dígitos de las decenas, que forman el tallo, y a la derecha las unidades que serán las ramas. Por ejemplo 94 se escribe:



9      4

- Para datos con tres dígitos el tallo estará formado por los dígitos de las centenas y decenas, que se escribirán a la izquierda, separados de las unidades. Por ejemplo, 237 se escribirá:

Tallo	Hoja
23	7

- a. Cada tallo define una clase, y se escribe sólo una vez. El número de "hojas" representa la frecuencia de dicha clase, que se ubica en una tercera columna del diagrama.

**Ejemplo:**

Los siguientes datos representan la longitud en cm. de 16 camarones de un criadero: 114; 125; 114; 124; 143; 152; 133; 113; 178; 127; 135; 161; 126; 134; 147; 132.

Los datos se adecuan al caso b, por lo tanto el diagrama será el siguiente:

Tallos	Hojas	Frecuencia
11	443	3
12	5476	4
13	3542	4
14	37	2
15	2	1
16	1	1
17	8	1

## REPRESENTACIÓN GRÁFICA DE UNA DISTRIBUCIÓN DE FRECUENCIAS

### HISTOGRAMA

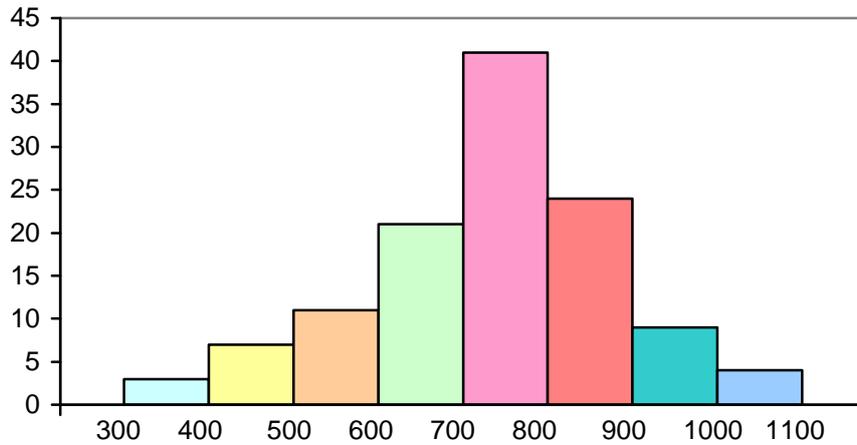
El histograma describe una distribución de frecuencias utilizando una gráfica de barras en la que la altura de cada barra es proporcional a la frecuencia de la clase que representa.

**Ejemplo:**

Tomando como referencia de los datos sobre los ingresos mensuales (nuevos soles) de los trabajadores de la empresa x. Siendo sus frecuencias:

Ingresos mensuales	Frecuencia Absoluta
300-400	3
400-500	7
500-600	11
600-700	21
700-800	41
800-900	24
900-1000	9
1000-1100	4
Total	120

### HISTOGRAMA SOBRE LOS INGRESOS MENSUALES (NUEVOS SOLES)

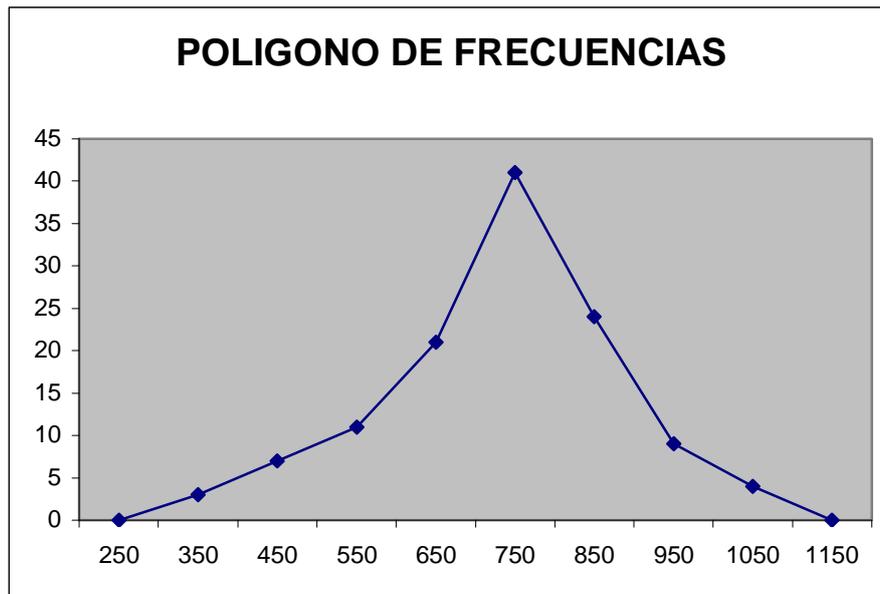


### POLÍGONO DE FRECUENCIAS

El polígono de frecuencia consiste en una línea poligonal formado por segmentos de recta que unen los puntos determinados por la intersección de la vertical del punto medio de clase, y la horizontal de la frecuencia de clase.

El polígono de frecuencia está estrechamente relacionado con un histograma. Tanto el histograma como el polígono de frecuencias permiten obtener una imagen rápida de las principales características de los datos (máximos, mínimos, punto de concentración, y otros) y el objetivo de estas dos representaciones es similar.

Ingresos mensuales	Marca de Clase	Frecuencia Absoluta
300-400	350	3
400-500	450	7
500-600	550	11
600-700	650	21
700-800	750	41
800-900	850	24
900-1000	950	9
1000-1100	1050	4



#### **POLÍGONO DE FRECUENCIAS ACUMULADAS**

Un polígono de frecuencias se utiliza cuando se desea determinar cuántas observaciones se encuentran por encima o por de abajo de ciertos valores.

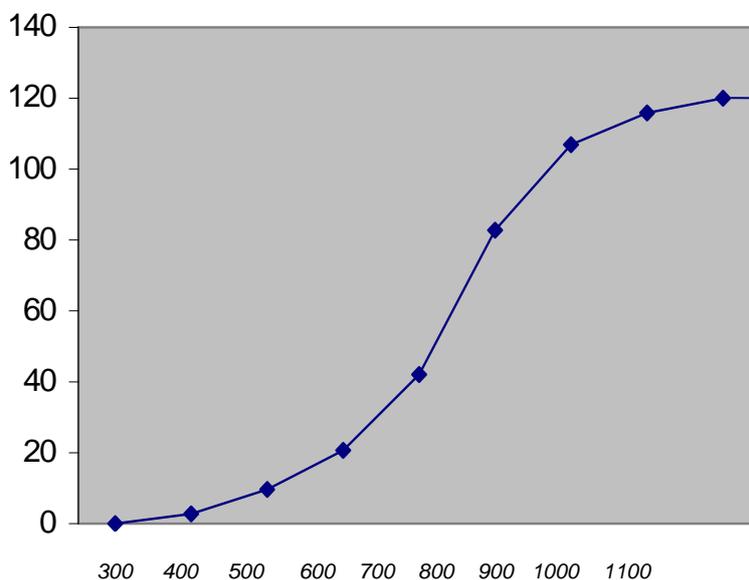
Las frecuencias acumuladas de los ingresos mensuales son los siguientes

Ingresos mensuales	Marca de Clase	$N_i$	$N'_i$
300-400	350	3	120
400-500	450	10	117
500-600	550	21	106
600-700	650	42	85
700-800	750	83	44
800-900	850	107	20
900-1000	950	116	11
1000-1100	1050	120	5

#### **Polígono de frecuencias acumuladas "menos de"**

Una distribución de frecuencias acumuladas "menos de", se traza iniciando con la clase más baja y laborando hacia adelante, sumando las frecuencias hasta llegar a la clase más alta. Del ejemplo anterior:

### Polígono de frecuencias acumuladas "menos de"

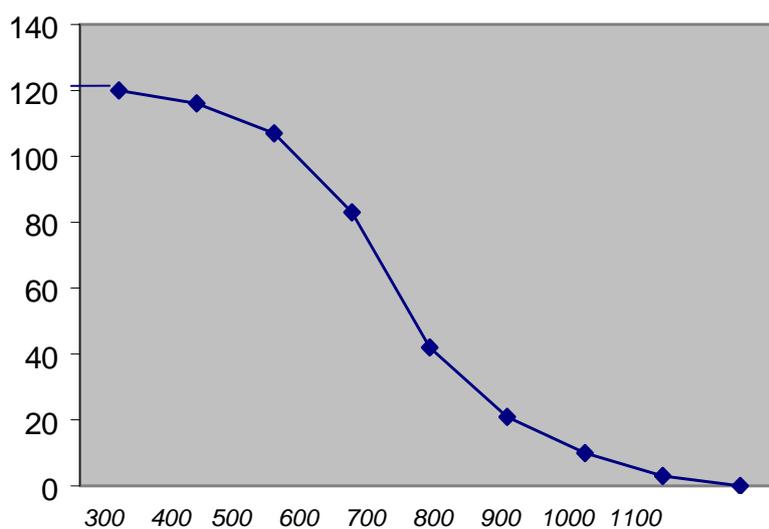


### Polígono de frecuencias acumuladas "más de"

Una distribución de frecuencias acumuladas "más de", se traza iniciando con la clase más alta y laborando hacia atrás, sumando las frecuencias hasta llegar a la clase más baja.

De los datos del ejemplo anterior, se obtiene:

### Polígono de frecuencias acumuladas "más de"

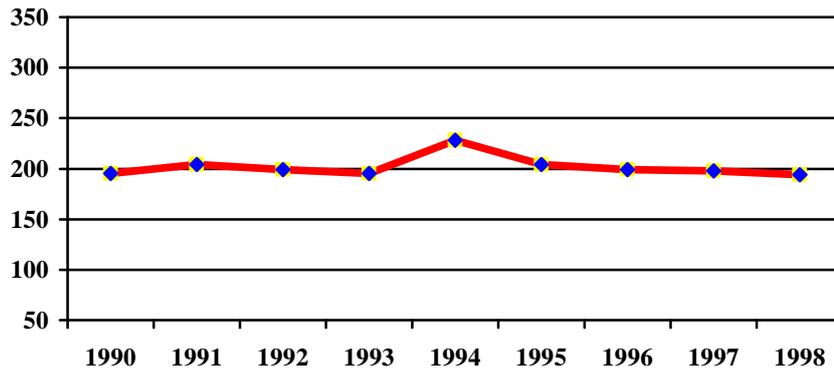


## REPRESENTACIÓN GRÁFICA DE LOS DATOS

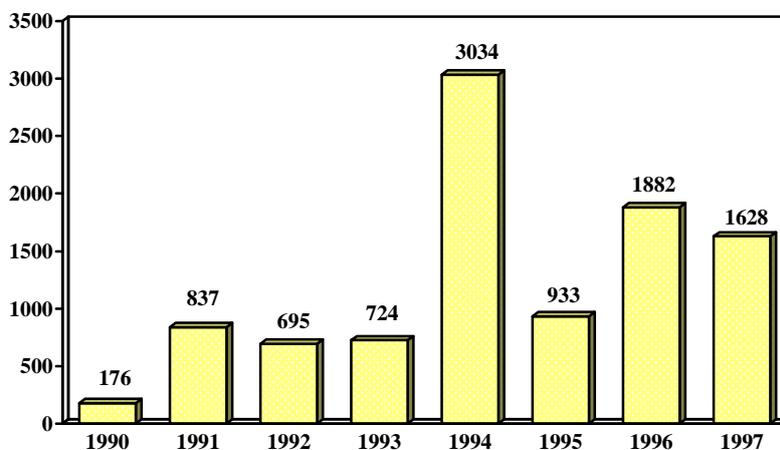
### Gráficas simples de línea y de barras

Las gráficas simples de líneas son ideales para representar tendencias de ventas, importaciones y otras series de valores durante un cierto período.

### SALARIOS REALES DEL SECTOR PRIVADO DE LIMA METROPOLITANA: 1990 - 98 - Índice Base Agosto 1990 = 100



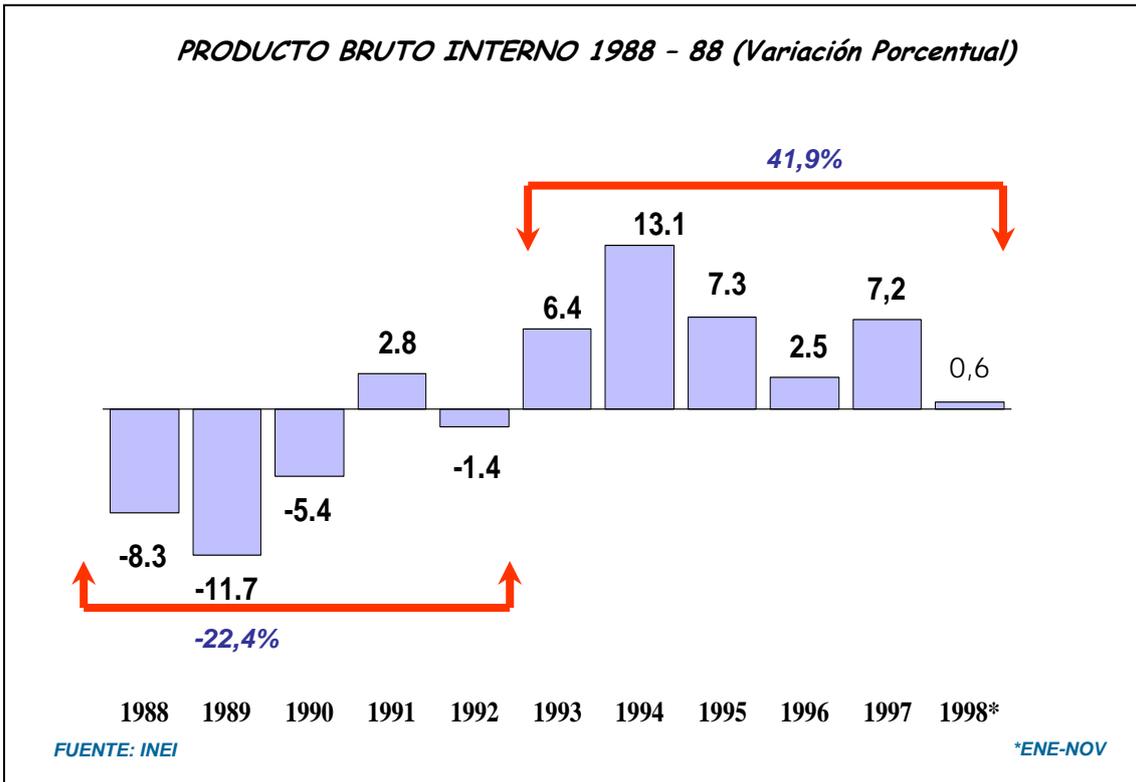
### BALANZA DE PAGOS: 1990-97 (millones de \$)



### Gráficas de barras seleccionadas

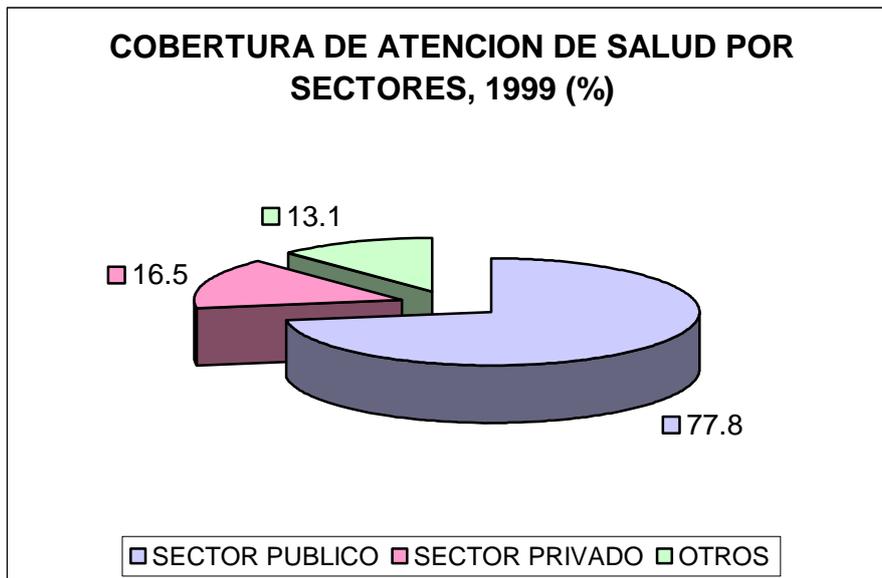
#### Gráfica de barra bidireccional

Una gráfica bidireccional se puede utilizar para mostrar pérdidas y ganancias, actividades por encima y por debajo de lo normal, y cambios porcentuales de un período a otro.



**Gráfica de sectores (o circular)**

Básicamente este tipo de gráfica sirve para representar una distribución de frecuencias relativas y obviamente consiste en trazar un círculo, y hay 360 grados en el mismo.



**Ejercicio:**

Se tiene información de los diversos indicadores que miden la Salud Materna de la población, para los departamentos del Perú del año 2000. Se pide elaborar la tabla de frecuencias y sus representaciones gráficas para el caso de atención de parto (3).

El alumno desarrollará el mismo análisis para las demás variables

Sea el siguiente cuadro:

**PERÚ: INDICADORES DE LA SALUD MATERNA, NIVEL DEPARTAMENTAL, 2000.**

DEPARTAMENTOS	SALUD MATERNA					
	Control Prenatal (%)	Vacuna contra el tétanos neonatal (%)	Atención del parto (%)	Parto institucional (%)	Atención postnatal (%)	Mujeres con alto riesgo reproductivo (%)
	(1)	(2)	(3)	(4)	(5)	(6)
PERU	83.8	80.3	59.3	57.9	32.3	61.8
AMAZONAS	59.1	78.0	28.3	27.3	12.5	64.4
ANCASH	78.4	77.4	42.0	38.4	27.0	71.8
APURIMAC	89.8	83.1	55.2	49.3	59.8	64.6
AREQUIPA	89.2	74.1	77.6	79.4	55.4	59.3
AYACUCHO	80.9	83.8	55.6	47.2	33.4	70.3
CAJAMARCA	67.6	72.9	27.1	22.2	18.0	64.6
CUSCO	95.1	77.9	47.7	39.4	43.6	67.0
HUANCAVELICA	78.1	75.9	21.0	19.6	32.7	77.0
HUANUCO	65.6	81.4	27.8	28.3	17.5	70.2
ICA	95.7	87.3	94.2	92.5	45.9	52.5
JUNIN	78.2	80.8	54.0	46.0	35.4	66.2
LA LIBERTAD	77.0	73.1	51.5	51.6	28.4	62.9
LAMBAYEQUE	81.2	80.9	61.2	57.4	41.8	58.6
LIMA 1/	94.8	83.2	91.1	89.5	54.2	57.9
LORETO	62.8	85.3	38.2	39.6	9.2	65.4
MADRE DE DIOS	87.7	80.5	75.7	74.9	37.4	61.8
MOQUEGUA	94.7	71.5	86.6	83.6	68.0	52.1
PASCO	72.8	78.7	56.7	50.7	21.9	64.9
PIURA	83.1	81.9	66.2	63.4	26.9	57.0
PUNO	86.8	77.1	27.8	20.5	46.2	72.4
SAN MARTIN	83.3	86.4	48.4	45.8	30.8	55.5
TACNA	95.0	64.3	87.2	82.2	76.0	56.2
TUMBES	92.7	81.4	88.4	85.7	42.5	51.1
UCAYALI	72.9	83.0	45.8	46.3	27.2	61.0
LIMA METROPOLITANA	96.4	83.2	94.3	93.5	61.5	57.7

1/ El departamento de Lima incluye la Provincia Constitucional del Callao

(3-4) Para nacimientos en los últimos 5 años que precedieron a la encuesta.

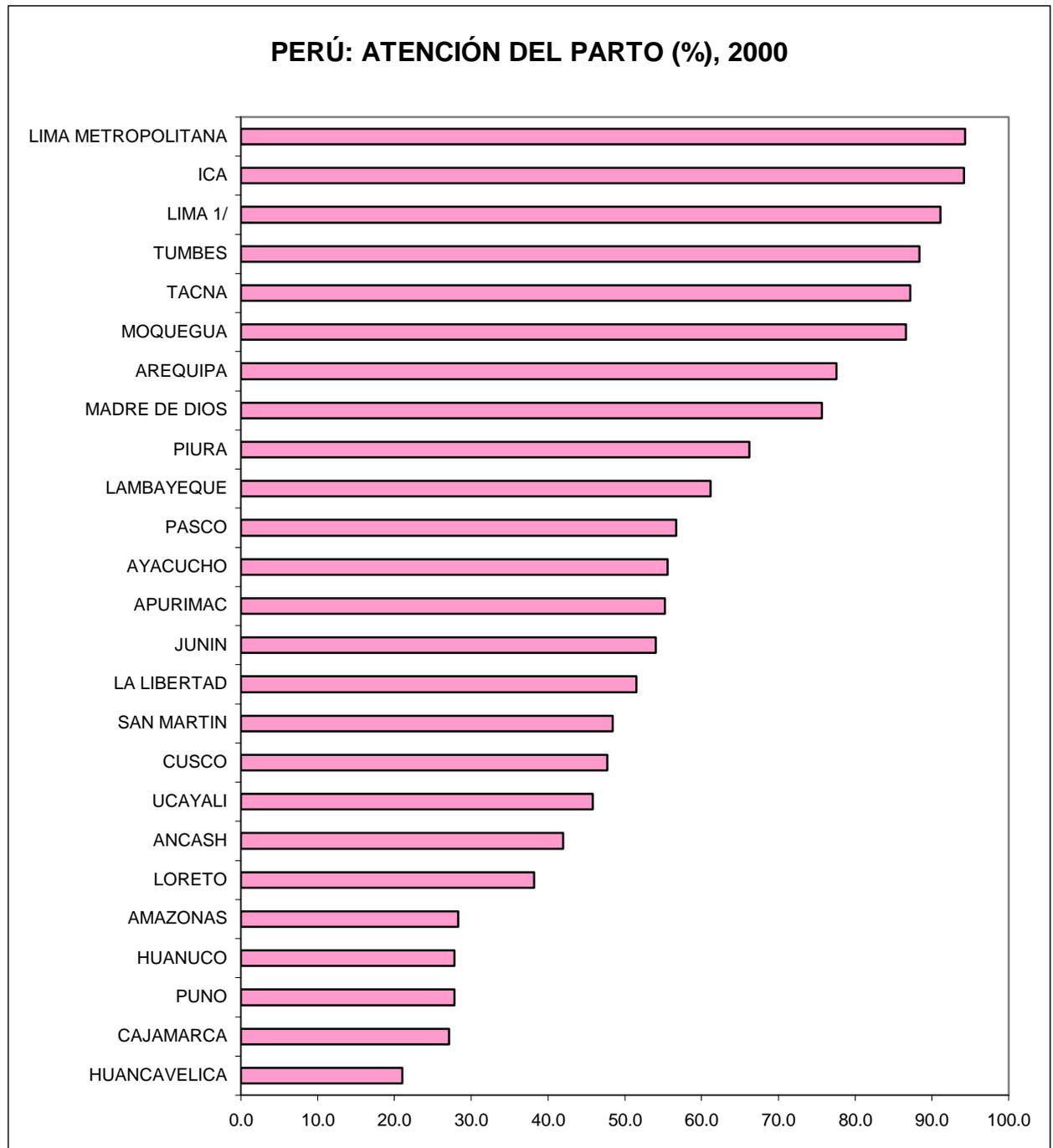
(1, 2 y 5) Para los últimos nacimientos de los 5 años anteriores a la encuesta.

(1, 3 y 5) Cuidado prenatal, atención durante el parto y postnatal recibidos por médico, obstetrix o enfermera

(4) Nacimientos en establecimientos de salud.

(6) Una mujer es considerada en grupo de alto riesgo si su embarazo es demasiado pronto (menos de 24 meses después del nacimiento Previo) cuando es demasiado joven (menor de 18 años), demasiado mayor (más de 35 años) o tiene demasiados hijos (3 o más nacimientos).

Una forma de representación, es la gráfica de barras, la cual describe el porcentaje de atenciones por cada departamento.



El siguiente paso a seguir es sintetizar la información obtenida para esa variable, es por ello que se elaborará la tabla de frecuencias.

Primero debemos verificar que tipo de variable vamos analizar, ello con el fin de construir su respectiva tabla de frecuencias, en este caso se trata de una variable continua.

1. Determinar la amplitud de cada intervalo

$$A = \frac{\text{Rango}}{\text{N}^\circ \text{ de intervalos}}$$

En este caso, el rango es 73.3 y el número de intervalos que escogeremos es de 5.

La amplitud será:  $A = \frac{73.3}{5} = 14.7 = 15$

La amplitud lo redondearemos a un número entero para que la construcción de la tabla sea más sencilla. El límite inferior del primer intervalo será de 20.

Obtenemos los siguientes intervalos:

[20,35) Se incluye sólo el valor de la izquierda  
[35,50)  
[50,65)  
[65,80)  
[80,95)

2. La marca de clase para cada intervalo será:

$$\frac{20+35}{2} = 27.5$$

$$\frac{35+50}{2} = 42.5$$

$$\frac{50+65}{2} = 57.5$$

$$\frac{65+80}{2} = 72.5$$

$$\frac{80+95}{2} = 87.5$$

3. En tercer lugar distribuiremos las frecuencias en sus respectivos intervalos.

Por ejemplo:

Para el año 2000 existieron 5 departamentos que tuvieron un porcentaje de atenciones de parto mayores e iguales a 20% y menores a un 35%. Entonces la frecuencia para este primer intervalo será de 5.

4. Luego de haber determinado las frecuencias absolutas se determinará las frecuencias absolutas acumuladas de cada intervalo.

Por ejemplo:

En el año 2000, 16 departamentos tuvieron a lo más un porcentaje de atención del 65%. Entonces para el tercer intervalo la frecuencia acumulada es de 16.

5. La frecuencia relativa y la frecuencia relativa acumulada se obtienen simplemente de la división de la frecuencia absoluta y la frecuencia absoluta acumulada entre el número de departamentos, respectivamente. (En este caso son 25 por que se ha considerado a Lima y callao como una observación más). Por ejemplo:

$$h_2 = \frac{5}{25} = 0.2$$

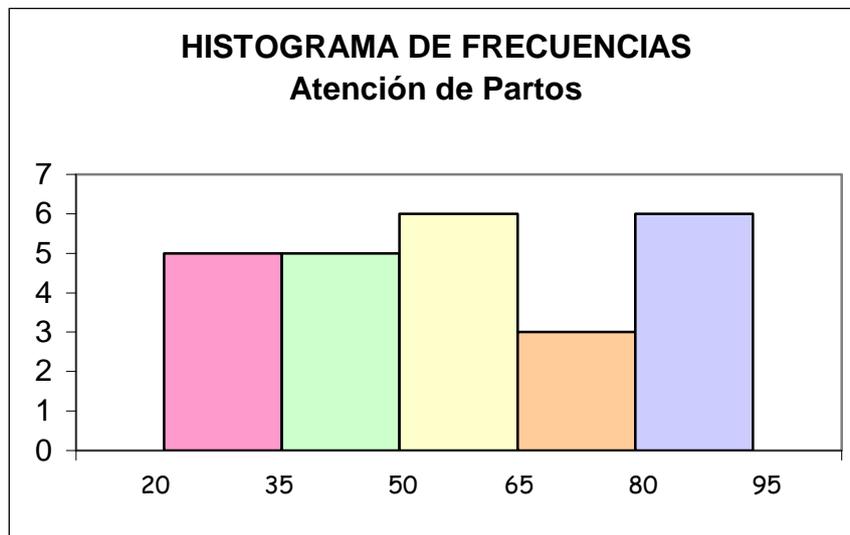
$$H_2 = \frac{10}{25} = 0.4$$

La tabla de frecuencias con intervalos semiabiertos por la derecha sería el siguiente:

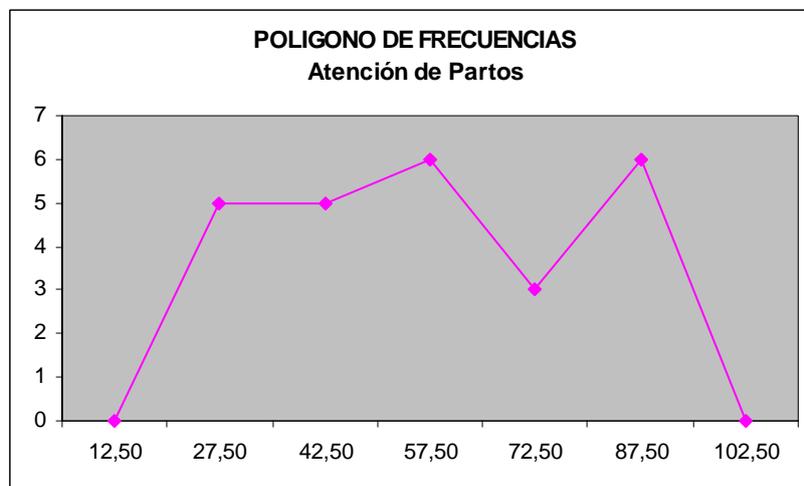
Intervalos de clase [ a, b )	Marca de clase	Nº de Departam. Tabulación	$n_i$	$N_i$	$h_i$	$H_i$	$N'_i$
20 35	27.5	///\	5	5	0.20	0.20	25
35 50	42.5	///\	5	10	0.20	0.40	20
50 65	57.5	///\ /	6	16	0.24	0.64	15
65 80	72.5	///	3	19	0.12	0.76	9
80 95	87.5	///\ /	6	25	0.24	1.00	6

n = 25

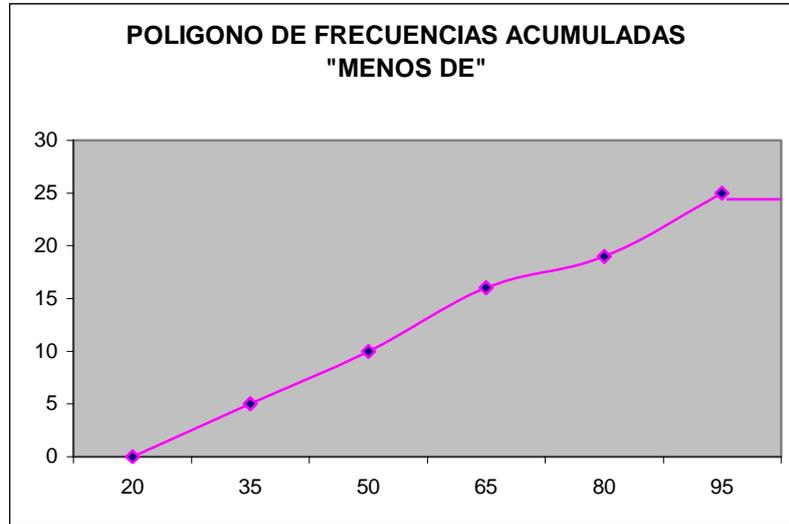
Tomando los datos de la tabla anterior, de la distribución en los diversos intervalos de las atenciones de partos de los departamentos del Perú, elaboramos el siguiente histograma de frecuencias:



Del mismo modo elaboramos el Polígono de frecuencias y el polígono de frecuencias acumuladas "menos de" y "mas de".



### Numero Acumulado De Departamentos Según Porcentaje De Atención En El Parto



### Numero Acumulado De Departamentos Según Porcentaje De Atención En El Parto

