Índice general

Estadística Descriptiva

CAPÍTULO

1 Conceptos previos

1.1 Introducción

En este capítulo hablaremos de la estadística en forma general de sus conceptos básicos y de su utilidad, pero al final pretendemos introducir al estudiante en el manejo de los datos obtenidos, como son: distinguir y clasificar los datos, organizarlos, tabularlos y representarlos gráficamente, para así obtener de ellos una mejor información y realizar así una buena inferencia.

No obstante al realizar todos los procedimientos que enseñaremos en este capítulo debemos ser prudentes al analizar los gráficos obtenidos ya que los mismos datos se pueden representar de diferentes formas y no todas son las más pertinentes, por lo que nuestro objetivo en este capítulo además de construir las bases conceptuales de la estadística desde un punto de vista elemental es enseñarles a realizar los gráficos que mejor representen las variables de interés.

1.2 ¿Qué es la estadística?

A pesar de que la estadística es hoy por hoy una ciencia axiomatizada y por ende una rama de las matemáticas, en sus orígenes era una ciencia puramente experimental, como lo indica la historia de la humanidad, ya que se tiene testimonio de que los egipcios 3050 ac llevaban datos acerca de los esclavos, población, agrimensura y riquezas; los chinos llevaban estadística de la población; la Biblia en el libro de los números se habla de los censos realizados por los levitas, lo mismo que los romanos lo cual lo hacían cada 5 años; en América los Aztecas y los Incas también tabulaban datos.

Pero fue en la edad media cuando Jhonn Graunt realizó un estudio estadístico serio sobre las muertes ocurridas en Londres entre 1603 y 1624 y publicado en 1662, luego Edmond Halley en 1693 realizó la primera tabla de seguros de vida.

Hasta ese momento la estadística aún era una ciencia puramente experimental y fueron los matemáticos Jacobo Bernoulli, Fermat, Pascal, Laplace, Leibniz, Huyghens, Bayes, Tchebitchef y Borel los que la axiomatizaron.

La palabra estadística fue usada por primera vez por Godofredo Achenwall convencido de que los datos estudiados por esta nueva ciencia sería de gran utilidad para los gobernantes.

Como lo indica este pequeño resumen histórico la estadística generalmente se toma como una ciencia que se encarga de presentar informes de encuestas, como nos lo hace pensar también los medios de comunicación masivos, pero la estadística es algo más que eso.

Ella se encarga de establecer los métodos y procedimientos para recolectar, clasificar y resumir datos para luego analizarlos siempre que exista variabilidad e incertidumbre causada intrínsecamente por ellos y después realizar inferencias con el fin de hacer predicciones y tomar decisiones.

Aunque no es fácil definir la estadística y por eso no vamos a tratar de hacerlo aquí por que lo más importante es comprender `que es y como se utiliza` y en el párrafo anterior expresamos lo que es la estadística.

En el siguiente esquema representamos en forma general la estructura de esta ciencia, aunque faltan eslabones estos lo representaremos en el segundo curso.

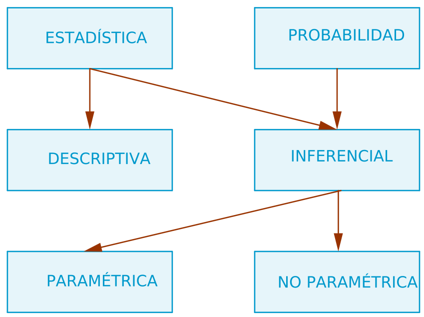


Figura 1.1: Diagrama

Ahora clasificaremos la estadística.

Estadística descriptiva: Describe y analiza los datos utilizando métodos numéricos elementales y gráficos para resumir y presentar la información contenida en ellos.

Estadística inferencial: apoyándose del cálculo de probabilidades y utilizando los datos de una muestra, efectúa estimaciones, decisiones y generaliza sobre un conjunto mayor de datos llamado población.

1.3 Conceptos previos

Definición 1.1 Individuos o elementos:

Individuos o elementos:Son personas u objetos que contienen cierta información de interés.

Definición 1.2 Población:

Es el conjunto de individuos o elementos que poseen la misma información de interés, es decir poseen ciertas propiedades comunes, y lo denotaremos Ω.o S

Definición 1.3 Muestra:

Es un subconjunto representativo de una población.

Definición 1.4 Caracteres o atributos:

Son rasgos o cualidades de los elementos de la población, los cuales pueden ser cualitativos o cuantitativos.

Definición 1.5 Modalidades:

Son diferentes situaciones posibles de un carácter, los cuales deben ser mutuamente excluyentes y exhaustivos, es decir cada elemento posee una y solo una de las modalidades posibles y un carácter debe tener todas las modalidades posibles.

Definición 1.6 Clases:

Conjunto de una o más modalidades en el que se verifica que cada modalidad pertenece a una y solo una de las clases

Definición 1.7 Variabilidad:

Al estudiar unos individuos escogemos uno varios caracteres, pero la medida de le los caracteres presentan desviaciones con respecto a una modalidad cuando cada carácter es analizado con las mismas condiciones, pero con la imposibilidad de controlar esas desviaciones. A cada una de esas desviaciones se le llama variabilidad.

1.3.1. Variables estadísticas

Cuando hablamos de variables estamos en realidad hablando de una función definida:

$$f: \begin{array}{ccc} A & \longrightarrow & B \\ x_i & \longmapsto & X(x_i) \end{array}$$

Donde $A \subseteq \Omega$ y B es un conjunto determinado de modalidades.

Al conjunto B se le denomina dominio de la variable o rango, de acuerdo con el tipo de dominio las variables se clasifican en:

Definición 1.8 Variables cualitativas:

Son aquellas que las modalidades posibles son de tipo nominal o categórico.

Ejemplo: 1.3.1

{si, no}, {hombre, mujer}, {blanco, negro, amarillo,...,etc.}

Definición 1.9 Variables casicuantitativas:

Son variables de tipo nominal, pero se puede establecer un orden entre ellas.

Ejemplo: 1.3.2

$\{1^o, 2^o, 3^o, ...\}$, {alto, medio, bajo}

Definición 1.10 Variables cuantitativas:

Son las que tienen por modalidades un conjunto de elementos dotados de operaciones y estas a su vez se clasifican en discretas y continuas

Definición 1.11 Discretas:

Cuando las variables toman valores puntuales

Ejemplo: 1.3.3

Cuando el rango es subconjunto de los números $I\!N$,

Definición 1.12 Continuas:

si X toma todos los valores en un intervalo

Ejemplo: 1.3.4

si $X \in I\!R$

1.4 Organización de datos

Generalmente los datos se organizan usando tablas y determinando algunas cantidades o valores que definiremos a continuación:

Consideremos una población finita de n elementos o individuos descrita según un carácter o variable X cuyas modalidades han sido agrupadas en un número k de clases, que denotaremos $c_1, c_2, c_3, ..., c_k$ y para cada clase $c_i, i = 1, 2, 3, ..., k$ definimos lo siguiente:

Definición 1.13 Distribución de frecuencias:

Es una función que a cada clase c_i de C (el conjunto de todas las clases) le asigna un valor $n_i \in I\!N_0$, es decir, es una función de la forma

$$\begin{array}{rccc} \mathfrak{F}: & C & \rightarrow & I\!N_0 \\ & c_i & \longmapsto & n_i \end{array}$$

tal que $\mathfrak{F}(c_i) = n_i$

La función de distribución no es única existen cuatro, las cuales definiremos ahora.

Definición 1.14 Frecuencia absoluta:

La frecuencia absoluta de una clase c_i es el número de veces n_i que se observa una modalidad perteneciente a esa clase.

Definición 1.15 Frecuencia relativa:

La frecuencia relativa de una clase c_i es el cociente f_i entre la frecuencia absoluta de la clase c_i y el número total de observaciones de todas las modalidades pertenecientes a todas las clases.

Generalmente esta frecuencia se multiplica por 100 y se da en porcentaje

Definición 1.16 Frecuencia absoluta acumulada:

La frecuencia absoluta acumulada N_i es el número de elementos de la población cuya modalidad es inferior o equivalente a la modalidad c_i es decir,

$$N_i = \sum_{j=1}^{k} n_j$$

[0]La población también puede ser infinita, pero ese caso lo estudiaremos más adelante en el capítulo 3

Definición 1.17 Frecuencia relativa acumulada:

Se denota F_i y es el tanto por uno de los elementos de la población que están en alguna de las clases y que presenta una modalidad inferior o igual a la c_i, es decir

$$F_i = \frac{N_i}{n} = \sum_{j=1}^{k} f_j$$

De éstas definiciones se pueden deducir algunas propiedades evidentes ya que las modalidades son exhaustivas y mutuamente excluyentes

- ♠ $$n = \sum_{j=1}^{k} n_j,$$

- ♠ $$\sum_{j=1}^{k} f_j = 1$$

Definición 1.18 Tabla de frecuencia:

Es una representación de $\mathfrak{F}$ y generalmente una tabla como se muestra en la tabla 1.1
Donde:

- ♠ M: representa modalidades,
- ♠ F.A.: frecuencia absoluta,
- ♠ F.R.: frecuencia relativa,
- ♠ F.A.A: frecuencia absoluta acumulada y
- ♠ F.R.A.: frecuencia relativa acumulada.

M	F.	F.R.	F.A.A.	F.R.A.
C	n_i	f_i	N_i	F_i
c_1	n_1	$f_1=\frac{n_1}{n}$	$N_1=n_1$	$F_1=\frac{N_1}{n}=f_1$
Cc_2	n_2	$f_2=\frac{n_2}{n}$	$N_2=n_1+n_2$	$F_2=f_1+f_2$
c_3	n_3	$f_3=\frac{n_3}{n}$	$N_3=n_1+n_2+n_3$	$F_3=f_1+f_2+f_3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
c_j	n_j	$f_j=\frac{n_j}{n}$	$N_j=\sum_{i=1}^{j} n_i$	$F_j=\sum_{i=1}^{j} f_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
c_k	n_k	$f_k=\frac{n_k}{n}$	$N_k=n$	$F_k=1$
[a]	n	1		

Tabla 1.1: Tabla de Frecuencias

[a] Aunque hemos definido la tabla sólo para $\mathfrak{F}$ en realidad en una tablas de frecuencia se tabulan las otras frecuencias definidas en este apartado

Ejemplo: 1.4.5

Se lanzan cinco monedas 1000 veces. El número de lanzamientos en los que han salido 0,1,2,3,4,5 caras se indican en la Tabla 1.2:

N^o de caras	n_i	f_i	N_i	F_i
0	38			
1	144			
2	342			
3	287			
4	164			
5	25			
Total	1000			

Tabla 1.2: Lanzamiento de monedas

a. Completar la tabla

b. Determinar para que clase F_i es mayor que el 60 %: $F_{\_\_\_} = \_\_\_$

1.4.0.1. Elección de clases

Las clases se pueden seleccionar de diferentes formas, pero siempre hay que seguir los criterios que se ajustan al tipo de variables que estudiamos.

- ♠ Cuando se trata de una variable nominal las clases c_i serán de tipo nominal
- ♠ Cuando la variable es cuantitativa discreta las clases serán valores numéricos del tipo $x_1, x_2, x_3, \cdots, x_k$.
- ♠ Si las variables son cuantitativas continuas las clases se definen mediante intervalos abiertos o semiabiertos, es decir de la forma:

$$(x_{i-1}, x_i)\,, (x_i, x_{i+1})\,, [x_{i-1}, x_i)\,, [x_i, x_{i+1})\,, (x_{i-1}, x_i]\,, (x_i, x_{i+1}]\,.$$

En estos casos las modalidades que contienen una clase son todos los valores numéricos posibles contenidos en el intervalo.

Por convención nosotros de aquí en adelante tomaremos siempre los $(k-1)$ primeros intervalos de la forma $(x_{i-1}, x_i]$ y el último $[x_{k-1}, x_k]$. A cada intervalo lo representaremos $x_{i-1} - x_i = I_i$.

Definición 1.19 Amplitud:

La amplitud de un intervalo se define $a_i = x_i - x_{i-1}$.

Definición 1.20 Marca de clase:

Es un valor $m_i \in I_i$ que representa a la clase y se define

$$m_i = \frac{x_i + x_{i-1}}{2}$$

La marca de clase es una forma abreviada de representar la clase.

m_i se determina de esta forma si las clases son conjuntos acotados.

1.4.0.2. Elección de clases para variables continuas

Cuando tenemos una muestra es importante escoger en una forma adecuada las clases y el número de clases k y para ello indicaremos los siguientes pasos:

- ♠ Lo primero es determinar k, entre mayor sea su valor mejor[1], pero de todas formas hay que acotarlo por que la idea es reducir el número de datos en la muestra.

$$k = \left\{ \begin{array}{cc} \sqrt{n} & \text{si } n \text{ es muy grande} \\ 1 + 3{,}22 \log n & \text{en otro caso} \end{array} \right\}$$

n se considera grande si $n \geq 40$

- ♠ Como segundo paso se determina el rango $R = x_k - x_0$
- ♠ Determinado el rango de la muestra podemos hallar la amplitud de cada intervalo que generalmente la tomamos constante:

$$a = \frac{R}{k}, \quad a_i = a \quad \forall i = 1, 2, 3, ..., k$$

- ♠ Ahora determinaremos los intervalos[2]

$$\begin{aligned} x_0 &= x_{\text{mín}} \\ x_1 &= x_0 + a \\ x_2 &= x_1 + 2a \\ x_k &= x_{\text{máx}} + ka \end{aligned}$$

Como se puede observar es posible que el valor de a no sea un número fácil de manejar entonces en estos casos como

$$x_k \geq x_{\text{máx}} > x_{\text{mín}} \geq x_0$$

entonces se varían los extremos simétricamente y a se aproxima al mayor entero es decir $a\prime = [[a]] + 1$

[1] Se aconseja que se escojan entre 5 y 20 clases dependiendo del número de datos

[2] Se aconseja que las marcas de clases coincidan con un gran número de datos y que los datos no sean extremos de las clases, para que los cálculos posteriores queden mejor aproximados.

Ejemplo: 1.4.6

Queremos observar el peso de las personas en una población y se toma una muestra de 21 individuos, los cuales están tabulados en la siguiente tabla

Peso en Kg						
58	42	51	54	40	40	49
56	58	57	59	63	58	66
70	73	71	69	70	68	64

Construir la tabla de frecuencias:

Solución

Lo primero que hay que identificar es la variable y en este caso vemos que la variable es de tipo cuantitativa continua por lo que ahora hay que determinar los intervalos y su longitud para que la perdida de información no sea significativa entonces sea X la variable peso, utilizaremos la fórmula $k = 1 + 3{,}22 * \log_{10} 21 = 5.\,257\,5 \approxeq 6$ aunque podríamos escoger también $k = \sqrt{21} \approxeq 5$.
Ahora hallemos $R = 73 - 40 = 33 \Longrightarrow a_i = a = \frac{33}{6} = 5.\,5$
$x_0 = x_{\text{mín}} = 40$
$x_5 = x_{\text{máx}} = 73$

i	Intervalos	m_i	n_i	f_i	N_i	F_i
1	40-45,5					
2	45,5-51					
3						
4						
5	62-67.5					≈ 1
					21	≈ 1
			21	≈ 1		

Existe otra posibilidad de construir la tabla y es escogiendo por ejemplo $a\prime = 7 \Longrightarrow R\prime = a\prime \cdot 5 = 35$
$d = R\prime - R = 2 \Longrightarrow x_0 = x_{\text{mín}} - \frac{d}{2} = 39, \quad x_5 = x_{\text{máx}} + \frac{d}{2} = 74$

i	Intervalos	m_i	n_i	f_i	N_i	F_i
1	39-46					
2	46-53					
3						
4						
5	67-74					
6					21	≈ 1
			21	≈ 1		

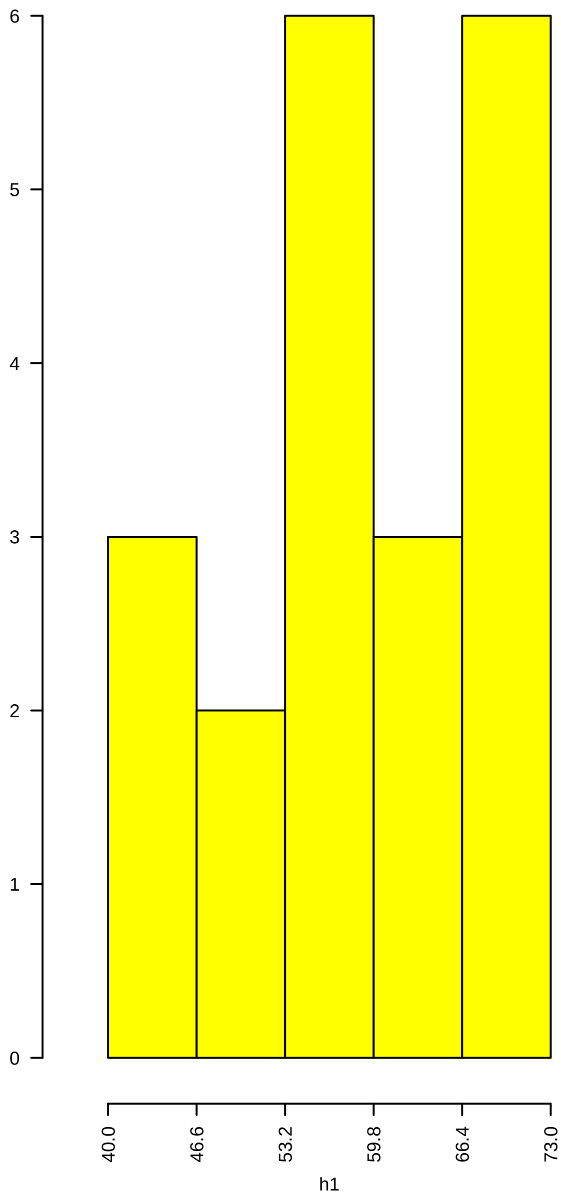
Utilizando un software estadístico ***Statgraphics*** ®, obtenemos la siguiente tabla

```
Tablas de Frecuencia para el Peso

------------------------------------------------------------------------------------
       Límite    Límite   Marca      Absoluta   Relativa   Acumulada  Acumulada.Rel.
Clases Inferior Superior  de Clase   Frecuencia  Frecuencia  Frecuencia  Frecuencia
------------------------------------------------------------------------------------

   Antes de        38,0                0        0,0000        0
   1       38,0   44,6667   41,3333    3        0,1429        3        0,1429
   2    44,6667   51,3333      48,0    2        0,0952        5        0,2381
   3    51,3333      58,0   54,6667    6        0,2857       11        0,5238
   4       58,0   64,6667   61,3333    3        0,1429       14        0,6667
   5    64,6667   71,3333      68,0    6        0,2857       20        0,9524
   6    71,3333      78,0   74,6667    1        0,0476       21        1,0000
Después de 78,0                        0        0,0000       21        1,000
------------------------------------------------------------------------------------
```

Podemos usar R para resolver el ejercicio, en este caso cargamos el paquete *Agricolae* como se muestra a continuación

```r
library(agricolae) # cargar el paquete agricole a la memoria
peso<-c(58,42,51,54,40,40,49,56,58,57,59,63,58,66,70,73,71,69,70,68,64)
par(mfrow=c(1,2),mar=c(4,3,0,1),cex=0.6)
h1<- graph.freq(peso,col="yellow",frequency=1,las=2,xlab="h1")
round(table.freq(h1), 2)
```

```
##   Lower Upper Main Frequency Percentage CF   CPF
## 1  40.0  46.6 43.3         3       14.3  3  14.3
## 2  46.6  53.2 49.9         2        9.5  5  23.8
## 3  53.2  59.8 56.5         7       33.3 12  57.1
## 4  59.8  66.4 63.1         3       14.3 15  71.4
## 5  66.4  73.0 69.7         6       28.6 21 100.0
```



1.5 Representaciones gráficas y diagramas

En la sección anterior hemos visto que al organizar los datos en una tabla se reduce la información, pero con ello podemos analizarlos de manera más sistemática y de esa manera podemos concentrarnos en los más importante, pero aún así a veces no es fácil observar todo lo que queremos, sobre todo si la persona interesada en los resultados no le interesa la estadística y sabe muy poco de ella, por lo que una representación gráfica simplifica aún más los datos.

1.5.1. Gráficos para variables cualitativas

i. Diagramas de barras

Se establece una especie de plano cartesiano representando las modalidades en el eje de ordenadas y las frecuencias absolutas o relativas en el eje de las abscisas, con este gráfico si se comparan varias poblaciones entre sí es conveniente utilizar las frecuencias relativas

Diagrama de barra para una variable cualitativa

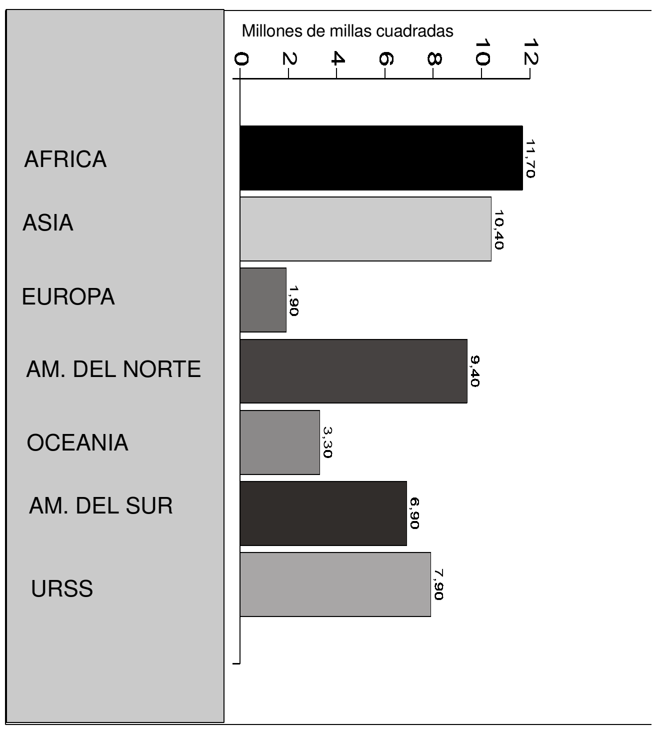
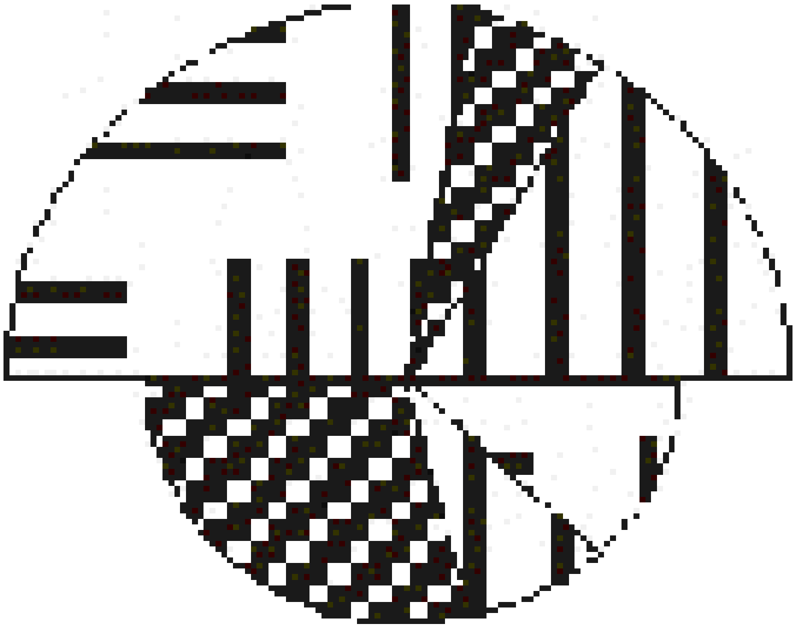


Figura 1.2: AREAS DE LOS CONTINENTES



ii. Diagramas circulares o sectores:

En estos diagramas se toma un círculo o cilindro y se dividen en tantos sectores como clases existan de modo que cada sector es proporcional a su frecuencia absoluta o acumulada, como se indica en los siguientes diagramas:

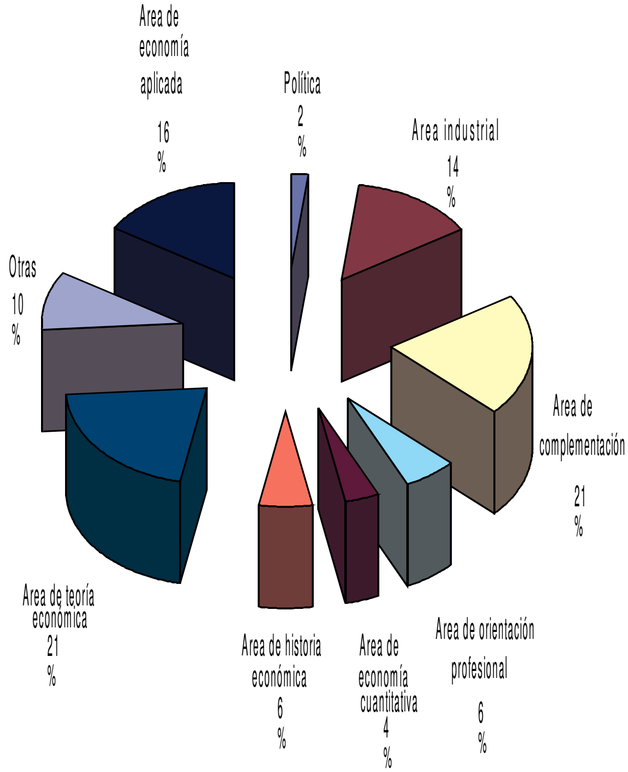
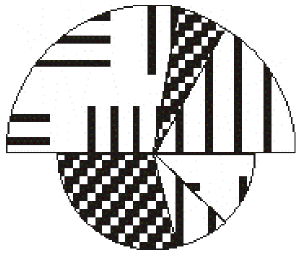


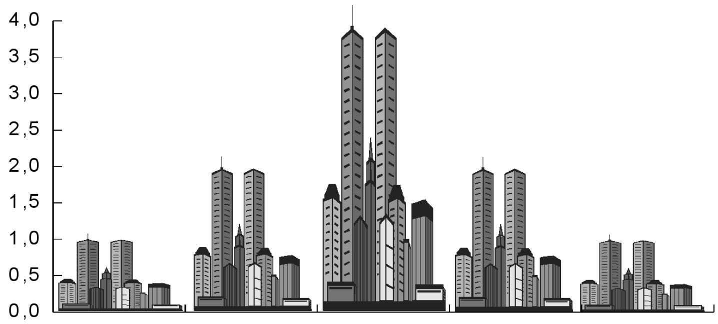
Diagrama circular

Con estos diagramas también se pueden comparar dos poblaciones utilizando dos semicírculos de radios r_1 y r_2 tal que $r_2 > r_1$ y cumplan la relación: $r_2 = r_1\sqrt{\frac{n_2}{n_1}}$ donde las n representan el tamaño de las poblaciones.



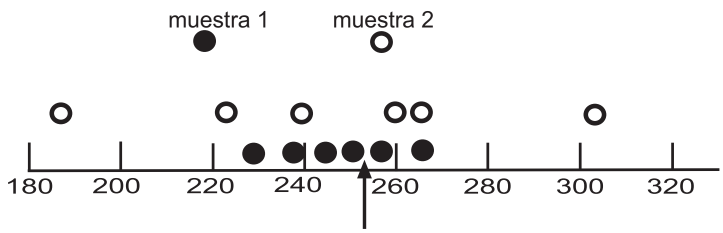
DOS MUESTRAS

iii. Pictogramas: Cuando se expresan dibujos alusivos al tema en estudio estos dibujos se hacen de tal forma que se utilizan diferentes escalas para representar la frecuencia absoluta o relativa.



1.5.2. Gráficos para variables cuantitativas

i. Diagrama de puntos: En este diagrama se coloca la frecuencia absoluta o relativa de una modalidad en una recta numérica y nos sirve para analizar dos o más modalidades cuando el número de datos es pequeño, con este gráfico analizamos fácilmente la tendencia y la variabilidad de la muestra. lo mismo que características poco usuales.



ii. Diagrama de tallo y hoja: Cuando el conjunto de datos

$$x_1, x_2, x_3, \cdots, x_n.x_1, x_2, x_3, \cdots, x_n.$$

es grande cada x_i, $\quad i = 1, 2, 3, ..., n$ tiene más de dos dígitos, entonces se dividen los x_i en dos partes . Un tallo formado por los primeros dígitos. Una hoja formada por el resto de dígitos

Ejemplo: 1.5.7

En un examen de clasificación para seleccionar alumnos que pueden ver directamente cálculo en el primer semestre en la facultad de ingeniería se obtuvieron los siguientes resultados:

$$\begin{bmatrix} 95 & 95 & 100 & 100 & 100 & 100 & 100 & 105 & 105 & 105 \\ 110 & 110 & 110 & 110 & 110 & 110 & 110 & 110 & 110 & 115 \\ 115 & 115 & 115 & 115 & 115 & 115 & 115 & 115 & 115 & 115 \\ 120 & 120 & 120 & 120 & 120 & 120 & 120 & 120 & 125 & 125 \\ 125 & 125 & 130 & 130 & 130 & 130 & 135 & 135 & 140 & 140 \end{bmatrix}$$

09	5	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	0	0	0	0	0	5	5	5	-	-	-	-	-	-	-	-	-	-	-	-
11	0	0	0	0	0	0	0	0	0	5	5	5	5	5	5	5	5	5	5	5
12	0	0	0	0	0	0	0	0	5	5	5	5	-	-	-	-	-	-	-	-
13	0	0	0	0	5	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

iii. Diagramas diferenciales: Son aquellos en los que se representan gráficamente frecuencias absolutas y relativas

iv. Diagramas integrales: Son los diagramas en los que se representan gráficamente el número de elementos que presentan una modalidad inferior o igual a una modalidad dada y se generan a tráves de las frecuencias acumuladas

Debido a que existen dos tipos de variables cuantitativas entonces debemos clasificar estos dos tipos de diagramas de acuerdo con el tipo de variable cuantitativa en estudio.

Gráficos para variables discreta

Al representar gráficamente la frecuencia absoluta o relativa de una variable discreta usamos los diagrama de barras, pero a diferencia de los diagrama de barra de las variables cualitativas las barras aquí se presentan con líneas delgadas, para indicar así la naturaleza de la variable. En el caso de los diagramas integrales tienen la forma del gráfico de una función escalonada

Ejemplo: 1.5.8

La siguiente tabla representa el número de hijos que tenían 12 familias encuestadas de un caserío cerca a Baranoa:

x_i	n_i	f_i	N_i
1	1		
2	3		
3	5		
4	3		
	12		

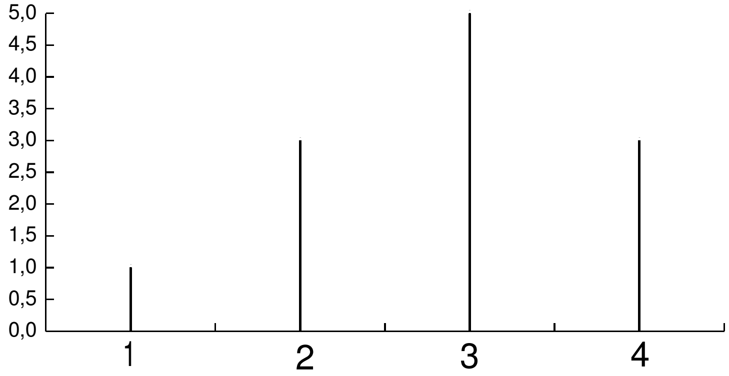
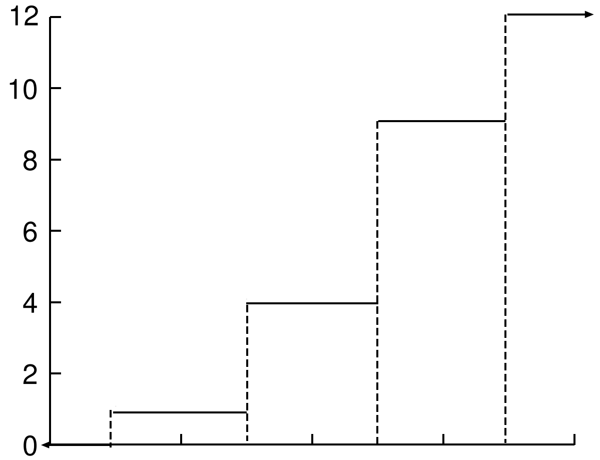


Figura 1.3: FRECUENCIAS ABSOLUTAS



FRECUENCIAS ABSOLUTAS ACUMULADAS

Para las variables continuas también se pueden representar con diagramas circulares

Ejemplo: 1.5.9

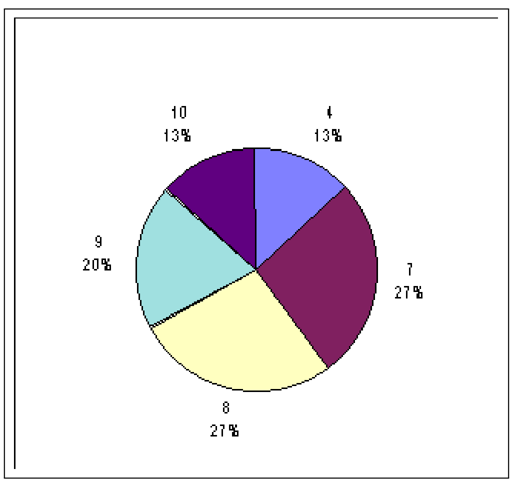
La tabla representa las notas obtenidas por los alumnos de 11en un examen de Matemáticas en 1990

Notas	Frecuencia	frecuencia Relativa
10	2	$2/15 = ,133$
9	3	$3/15 = ,200$
8	4	$4/15 = ,267$
7	4	$4/15 = ,267$
4	2	$2/15 = ,133$

(Tabla 1)

notas	$\mathbf{n}_i$	$\mathbf{f}_i$	f_i %	$\mathbf{N}_i$	$\mathbf{F}_i$
10	2	$2/15 = ,133$	13,3	15	1,000
9	3	$3/15 = ,200$	20,0	13	,867
8	4	$4/15 = ,267$	26,7	10	,667
7	4	$4/15 = ,267$	26,7	6	,400
4	2	$2/15 = ,133$	13,3	2	,133

(Tabla 2)



Gráficos para variables continuas

Para las variables continuas existen dos tipos de gráficos:

- ♠ Los histogramas: Los cuales se construyen representando sobre cada intervalo, un rectángulo que tiene la longitud del segmento como base y la altura debe ser un valor proporcional a el valor de la frecuencia para ese intervalo.
- ♠ Polígono de frecuencias: Este se elabora uniendo los puntos que corresponden a las imágenes de las marcas de clase.

En el caso de los diagramas integrales a estos polígonos se les llama ojiva y se obtienen uniendo las abscisas a partir de los extremos de los intervalos en los que se han organizado los datos

Ejemplo: 1.5.10

Representar gráficamente la información que aparece en la siguiente tabla

Intervalos	m_i	n_i	N_i
0-2	1	1	
2-4	3	4	
4-6	5	10	
6-8	7	3	
8-10	9	1	

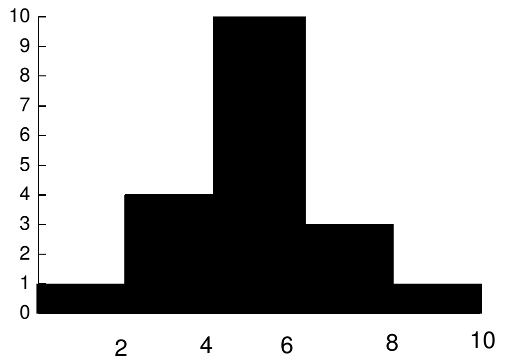
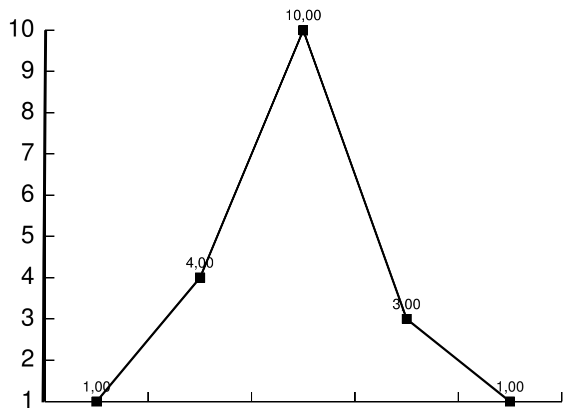


Diagrama Diferencial



Polígono de frecuencias

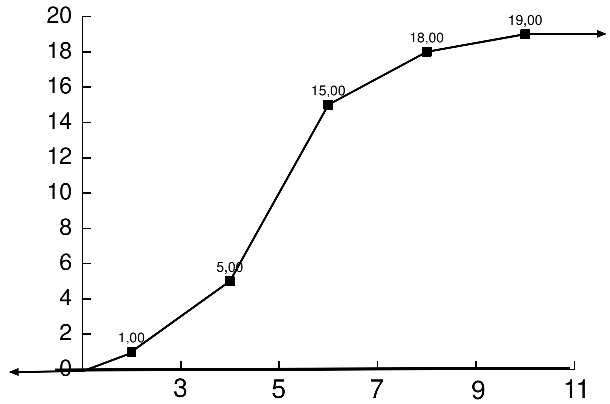


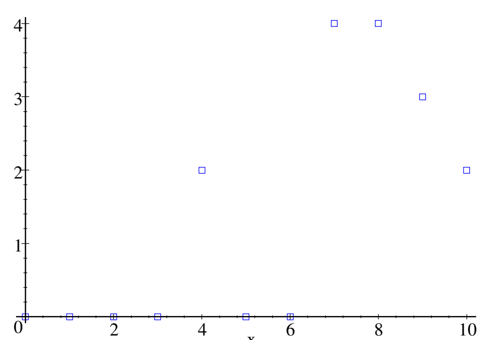
Diagrama acumulado

Ejemplo: 1.5.11

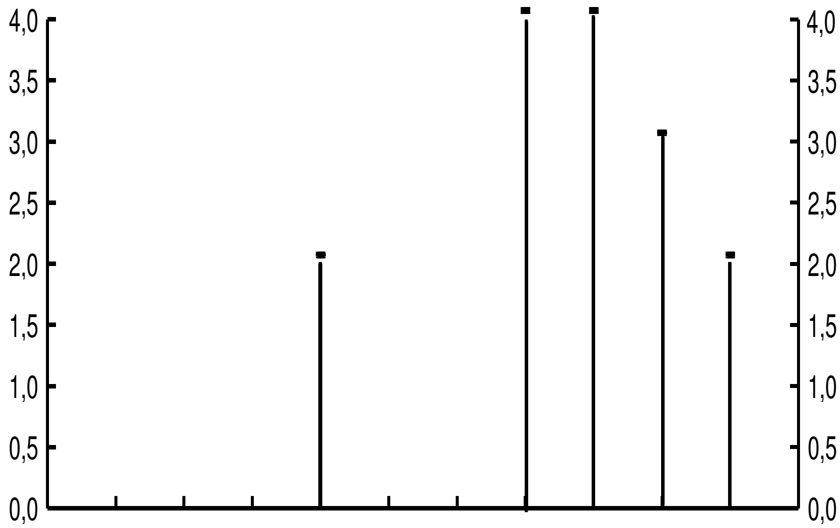
Representar gráficamente los datos de la tabla1

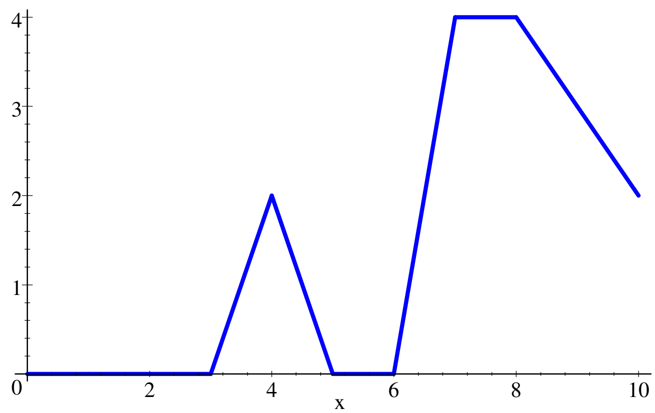
Notas	Frecuencia
10	2
9	3
8	4
7	4
6	0
5	0
4	2
3	0
2	0
1	0
0	0

(Tabla 1)



Test Scores





Test Scores

* Realizar una ojiva que represente los datos de la tabla2

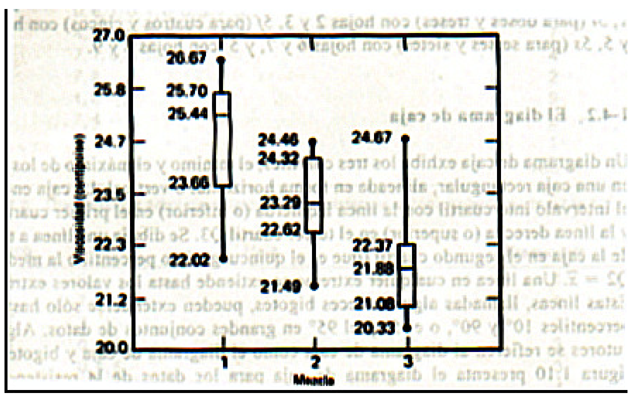
♠ Diagrama de caja: El diagrama de caja es una representación visual que describe varias características importantes tales coma son: la dispersión, la variabilidad, la simetría e identifica los datos que se alejan de manera poco usual del resto de los datos.

En un diagrama de caja se concentran el 50 % de los datos en un rectángulo, ubicando los valores para los cuales la frecuencia relativa acumulada son el 25 % y el 75 % respectivamente en líneas horizontales o verticales formando los lados paralelos del rectángulo, también se utiliza una línea que pasa por el centro del rectángulo para representar el dato que le corresponde a la frecuencia relativa acumulada del 50 %. De los dos lados del rectángulo antes mencionados se extiende una línea perpendicular hasta los extremos más alejados, los valores que se encuentran en esta región se llaman valores atípicos o extremos.

Ejemplo: 1.5.12

La siguiente tabla muestra los datos sobre la viscosidad de tres mezclas

mezcla 1	mezcla 2	mezcla 2
22.02	21.49	20.33
23.83	22.67	21.67
26.67	24.62	24.67
25.38	24.18	22.45
25.48	22.78	22.28
23.50	22.56	21.95
25.90	24.46	20.49
24.98	23.79	21.81

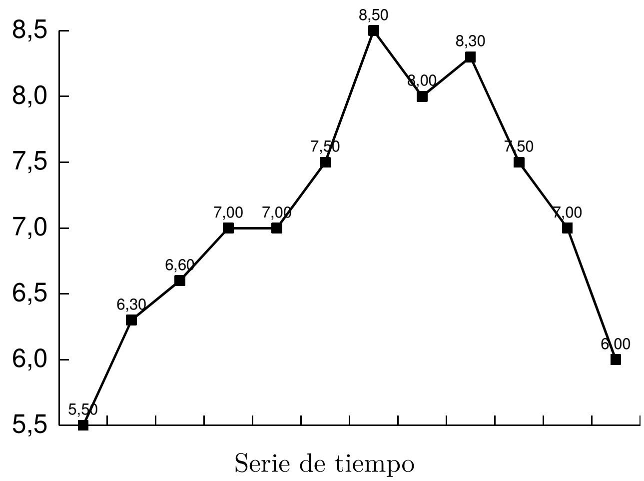


♠ Gráficas de serie de tiempo: Estas gráficas de series de tiempo se utilizan cuando nos interesa averiguar si el momento en que se tomaron afecta su variabilidad.

Ejemplo: 1.5.13

El número de piezas (en miles) existentes en el almacén de una determinada fábrica el último día de cada mes del año 1981, viene dado por la tabla

Meses	Ene.	Feb.	Mar.	Abr.	Ma..	Jun.	Jul.	Ag.	Sep.	Oct.	Nov.	Dic.
Piezas	5,5	6,3	6,6	7	7	7,5	8,5	8	8,3	7,5	7	6



1.6 Problemas

1. Decir si los datos que representan los siguientes item son continuos o,discretos o cualitativos

   a) Color de los ojos

   b) Letras usadas en un párrafo

   c) Velocidad de un coche

   d) Edad de un trabajador

   e) Número de billetes de 20000 que circulan en una ciudad

   f) Número de estudiantes matriculados en la Universidad

   g) Caudal del río,medido en diferentes días del año

2. El conjunto de datos adjunto está formado de observaciones del gasto de agua en regaderas (L/min) para una muestra de $n = 128$ casas de un sector exclusivo de Barranquilla

4,6	12,3	7,1	4,0	9,2	6,7	6,9	11,5	5,1	3,8
11,2	10,5	14,3	8,0	8,8	6,8	5,1	5,6	9,6	7,5
7,5	6,2	5,8	2,3	3,4	10,4	9,8	6,6	3,7	6,4
6,0	8,3	6,5	7,6	9,3	9,2	7,3	5,0	6,3	13,8
6,2	5,4	4,8	7,5	6,0	6,9	10,8	7,5	6,6	5,5
3,3	7,6	3,9	11,9	2,2	15,0	7,2	6,1	15,3	18,9
7,2	5,4	5,5	4,3	9,0	12,7	11,5	7,4	5,0	3,5
8,2	8,4	7,3	10,3	11,9	6,0	5,6	9,5	9,3	10,4
9,7	5,1	6,7	10,2	6,2	8,4	7,0	4,8	5,6	10,5
14,6	10,8	15,5	7,5	6,4	3,4	5,5	6,6	5,9	15,0
9,6	7,8	7,0	6,9	4,1	3,5	11,9	3,7	5,7	6,8
11,3	9,3	9,6	10,4	9,3	6,9	9,8	9,1	10,6	4,5
6,2	8,3	3,2	4,9	5,0	6,0	8,2	6,3		

   a) Dibuje un histograma de frecuencias absolutas en el eje vertical

   b) Dibuje un diagrama acumulado

   c) Construya una representación de tallo y hoja

   d) Dibuje una gráfica de series de tiempo

   e) Tome tres opciones diferentes de intervalos de clase para el conjunto de datos

      1) Intervalos largos e iguales

      2) Intervalos cortos e iguales

      3) Intervalos de diferente rango

   f) Realice un diagrama circular

   g) Compare las gráficas y determine cual de ellas representa mejor los datos

   h) ¿La gráfica está centrada o sesgada?

   i) ¿Existen Puntos inusuales?, es decir muy alejados de la mayoría

3. Un artículo sobre semilla de maní (Sept de 1990) reportó los siguientes resultados

Cremoso	56	44	62	36	39	53	50	65	56
	68	41	30	40	50	56	30	22	40
Crujiente	62	53	75	42	47	40	34	62	52
	50	34	42	36	75	80	47	56	62

   a) Construya un diagrama de tallo y hojas para las muestras

   b) Construya un histograma para los muestras

   c) Realice un diagrama de puntos en ambos casos

   d) Realice sendos diagramas de cajas

   e) Compare las dos muestras

f) Obtuvo los mismos resultados al comparar

g) ¿Que decisión tomaría usted? Explique porqué.

4. Construya un diagrama de barras para los datos representados en la siguiente tabla.

X	n
480	75
498	56
516	42
534	30
552	21
570	15
588	11
606	6
624	2
720	1

Donde X representa el sueldo en miles de pesos que reciben los empleados de una empresa industrial de acuerdo con la labor que desempeñan

5. Considere la muestra de calificaciones de un grupo de primaria

a	c	d	b	c	c	c	d	f	f
d	f	a	d	c	b	c	d	d	b

Construya un

a) Diagrama de barras

b) Diagrama circular

6. En un párrafo se utilizaron las siguientes letras

a	b	c	e	h	t	r	o	i	u	w	x
20	15	12	15	3	10	4	20	12	10	1	1

a) Diagrama de barras

b) Diagrama circula

7. ¿Qué clase de gráficas son apropiadas para los datos

a) Nominales

b) Ordinales

c) Cuantitativos discretos

d) Cuantitativos continuos

8. Use los datos de la gráfica 1.3 para realizar una tabla de frecuencias

9. De acuerdo con los datos de la gráfica 1.4 realice una tabla de frecuencias

aaaaa

Teorema 1.1 aaaaa

aaaaaaa