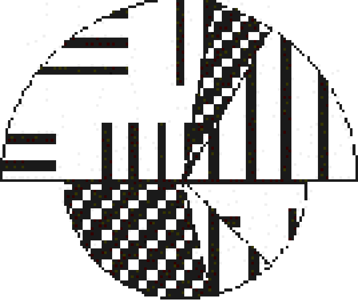
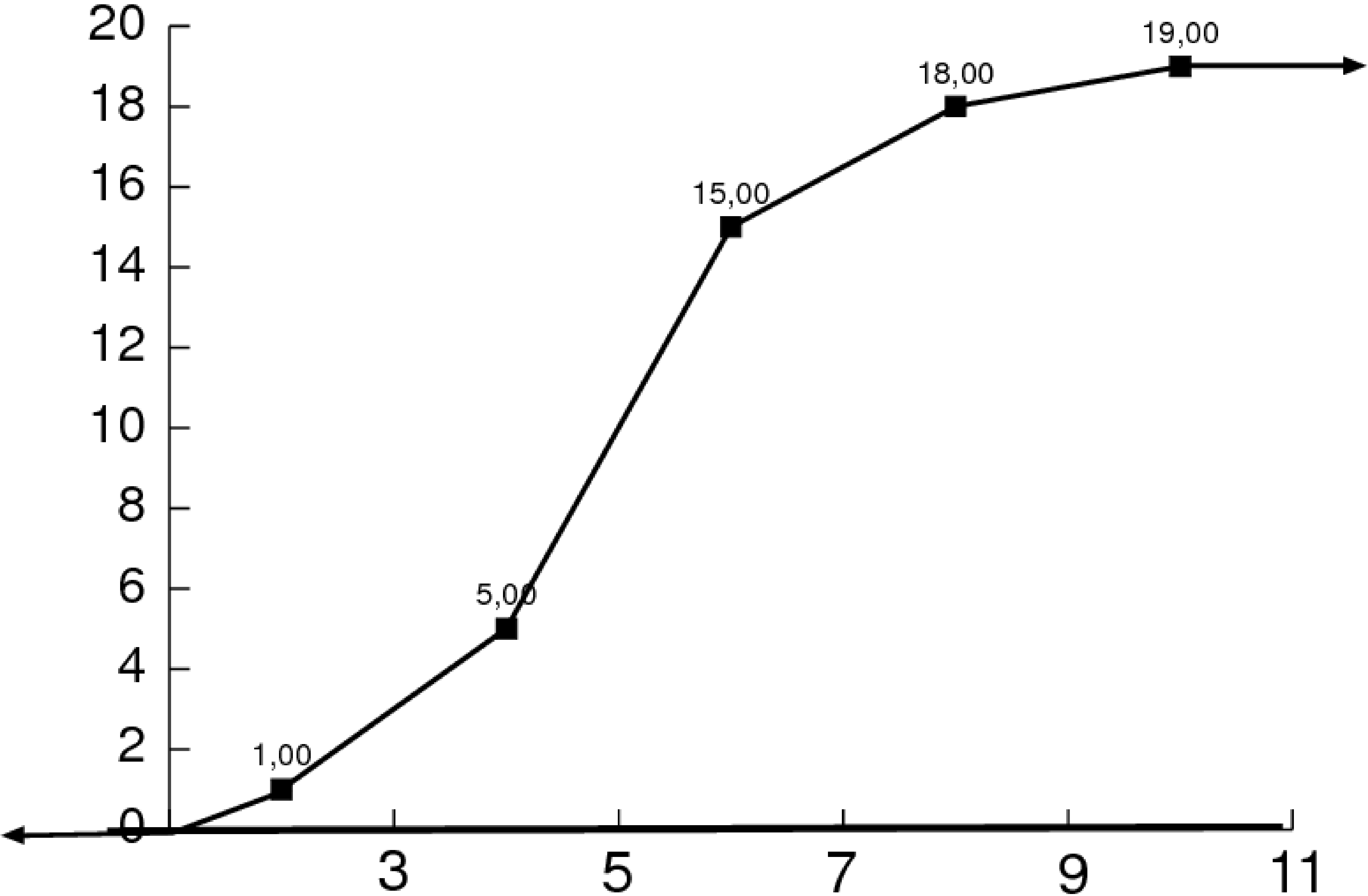
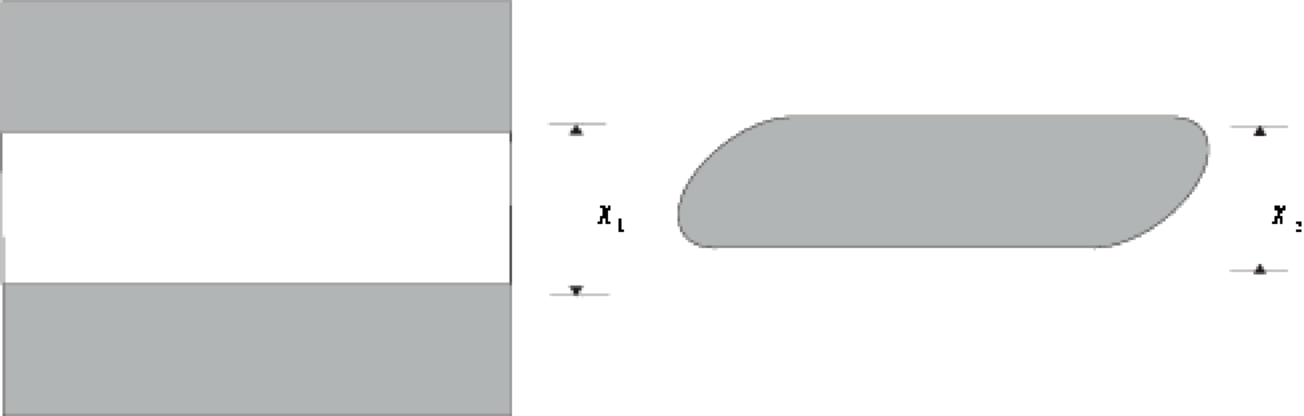
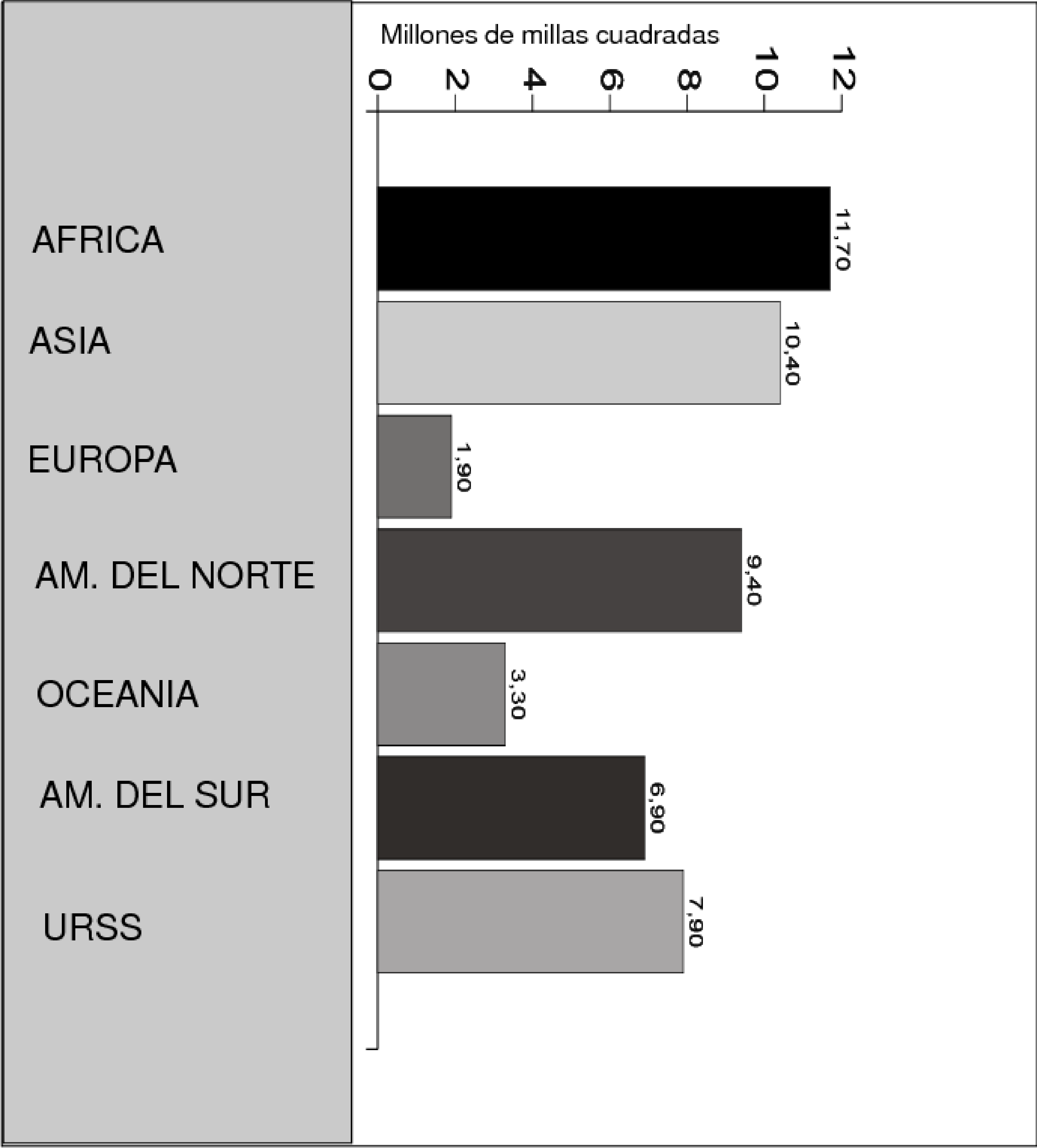


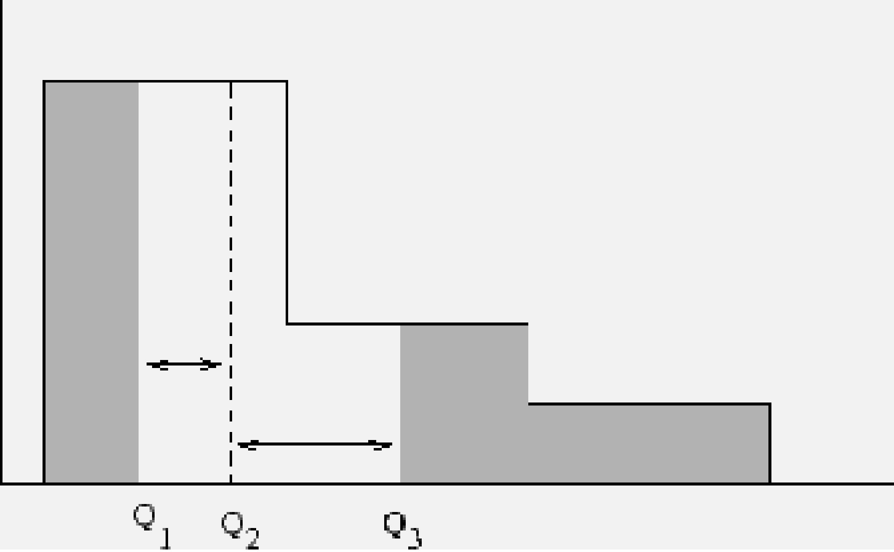


20
18
16
14
12
10
8
6
4
2
0
3
5
7
9
11
1,00
5,00
15,00
18,00
19,00

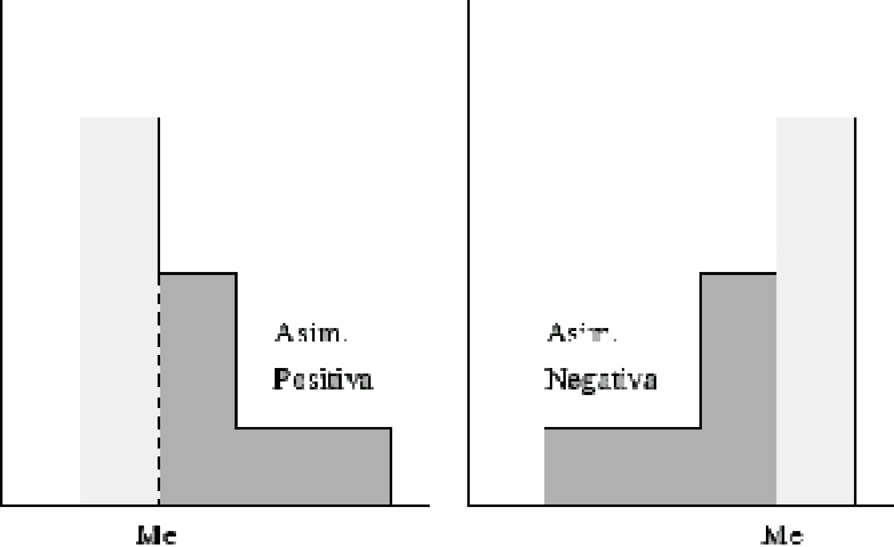


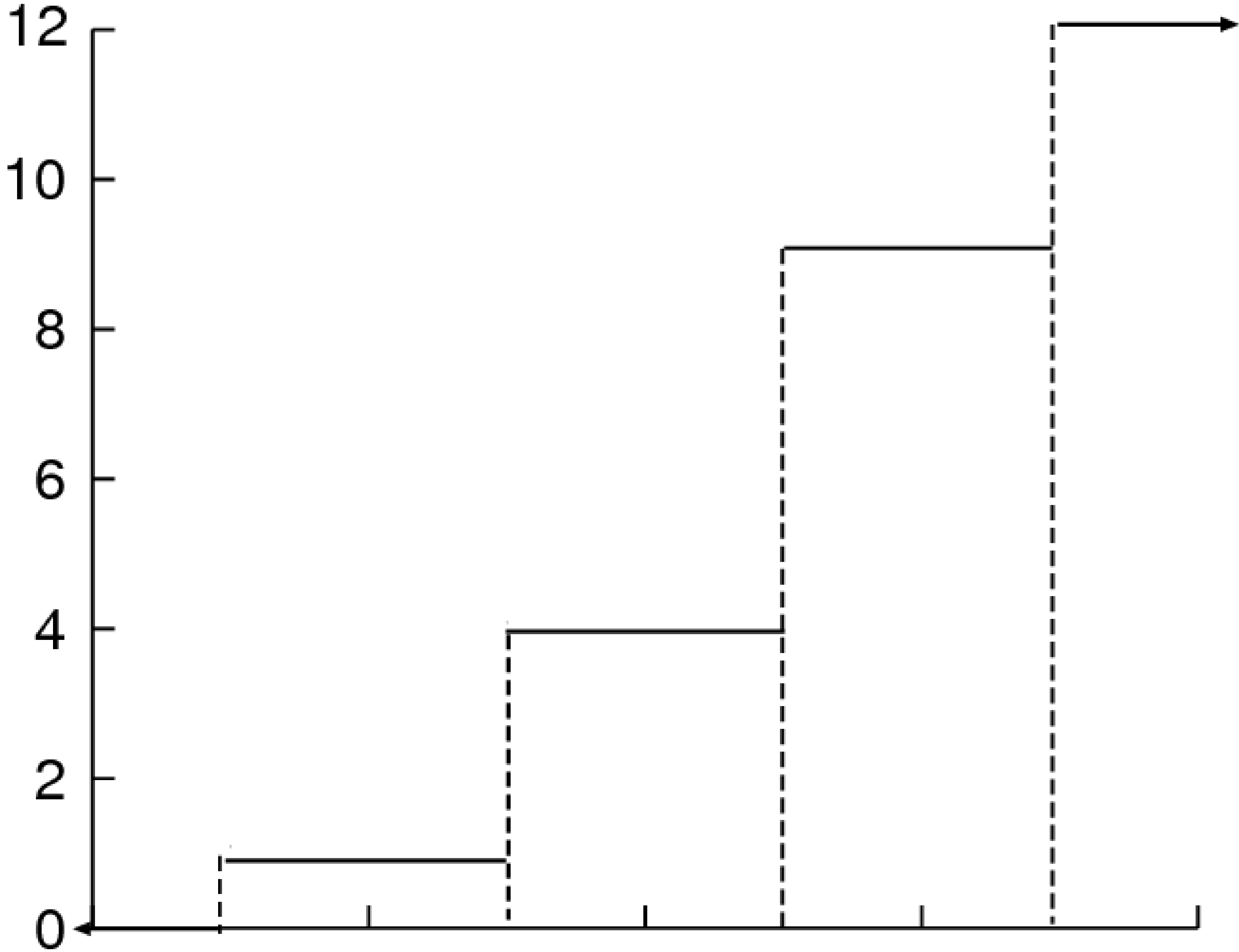
X_1
X_2

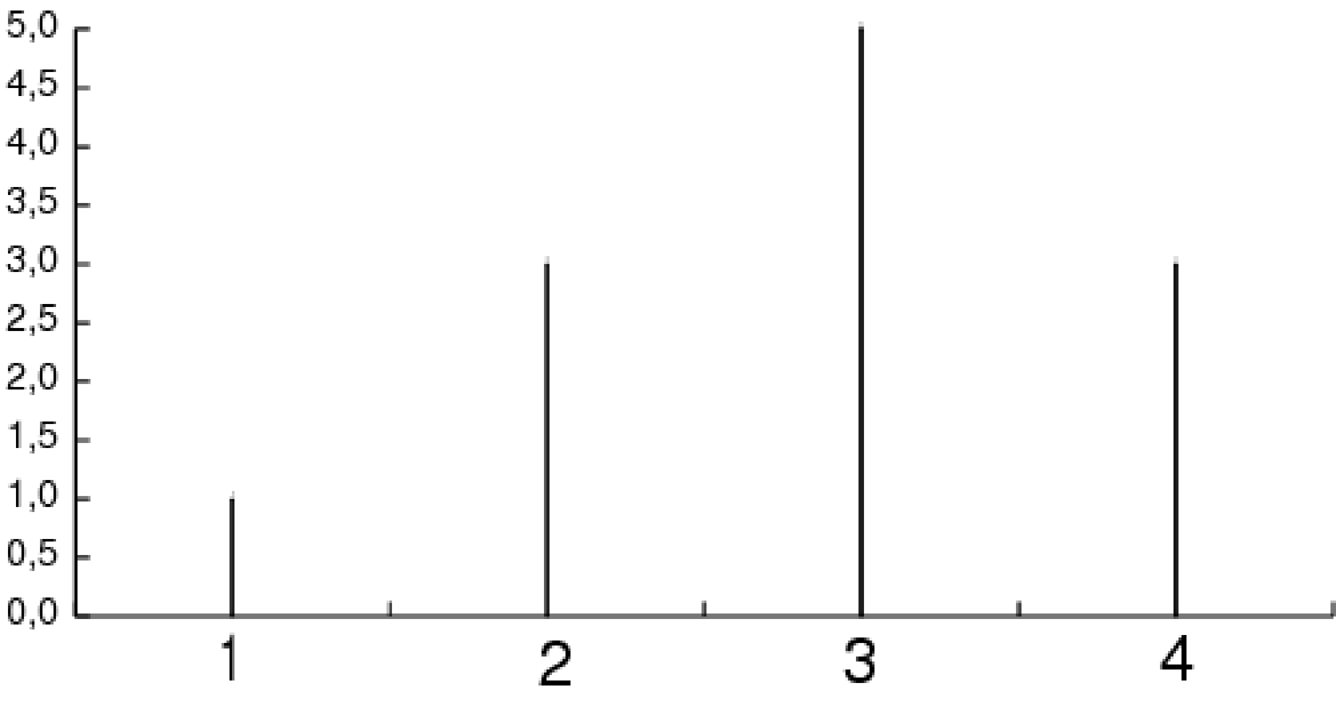

Millones de millas cuadradas
0
2
4
6
8
10
12
AFRICA
11,70
ASIA
10,40
EUROPA
1,90
AM. DEL NORTE
9,40
OCEANIA
3,30
AM. DEL SUR
6,90
URSS
7,90

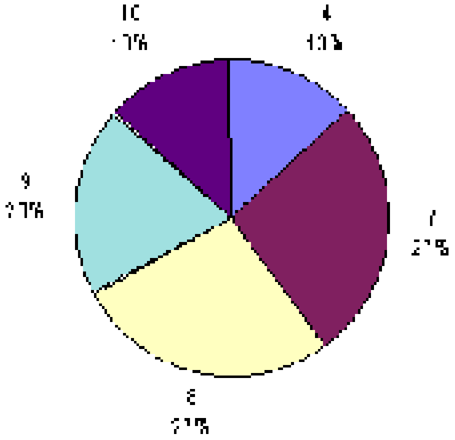


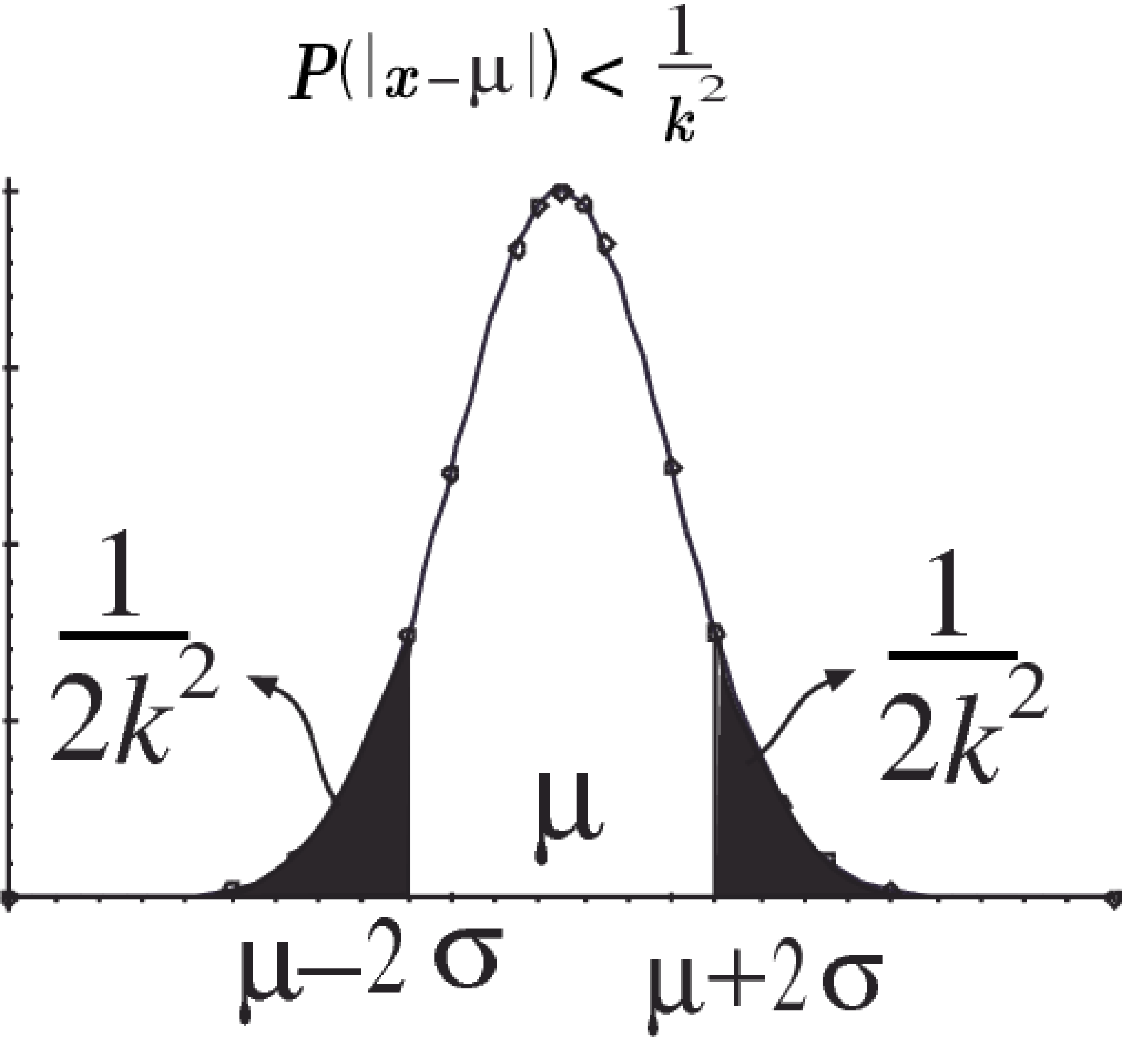
Q_1
Q_2
Q_3


Asim.
Positiva
Me
Asim.
Negativa
Me

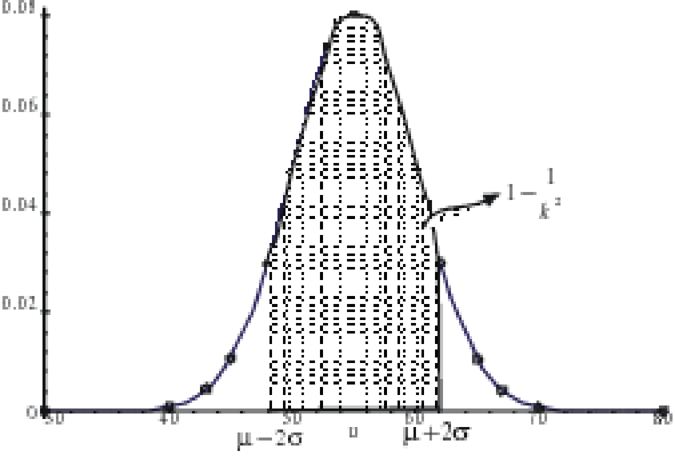

12
10
8
6
4
2
0


5,0
4,5
4,0
3,5
3,0
2,5
2,0
1,5
1,0
0,5
0,0
1
2
3
4

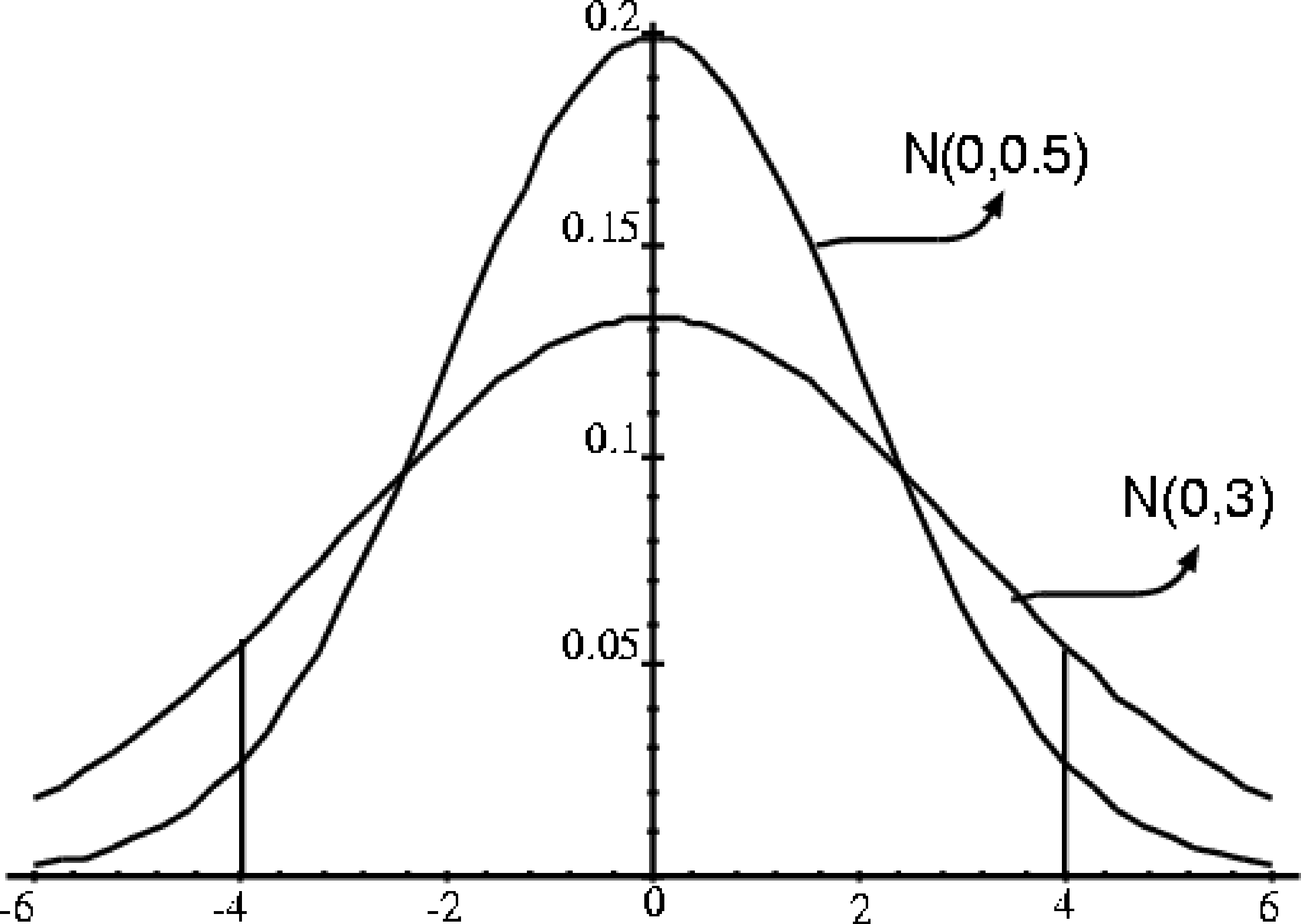

10
4
9

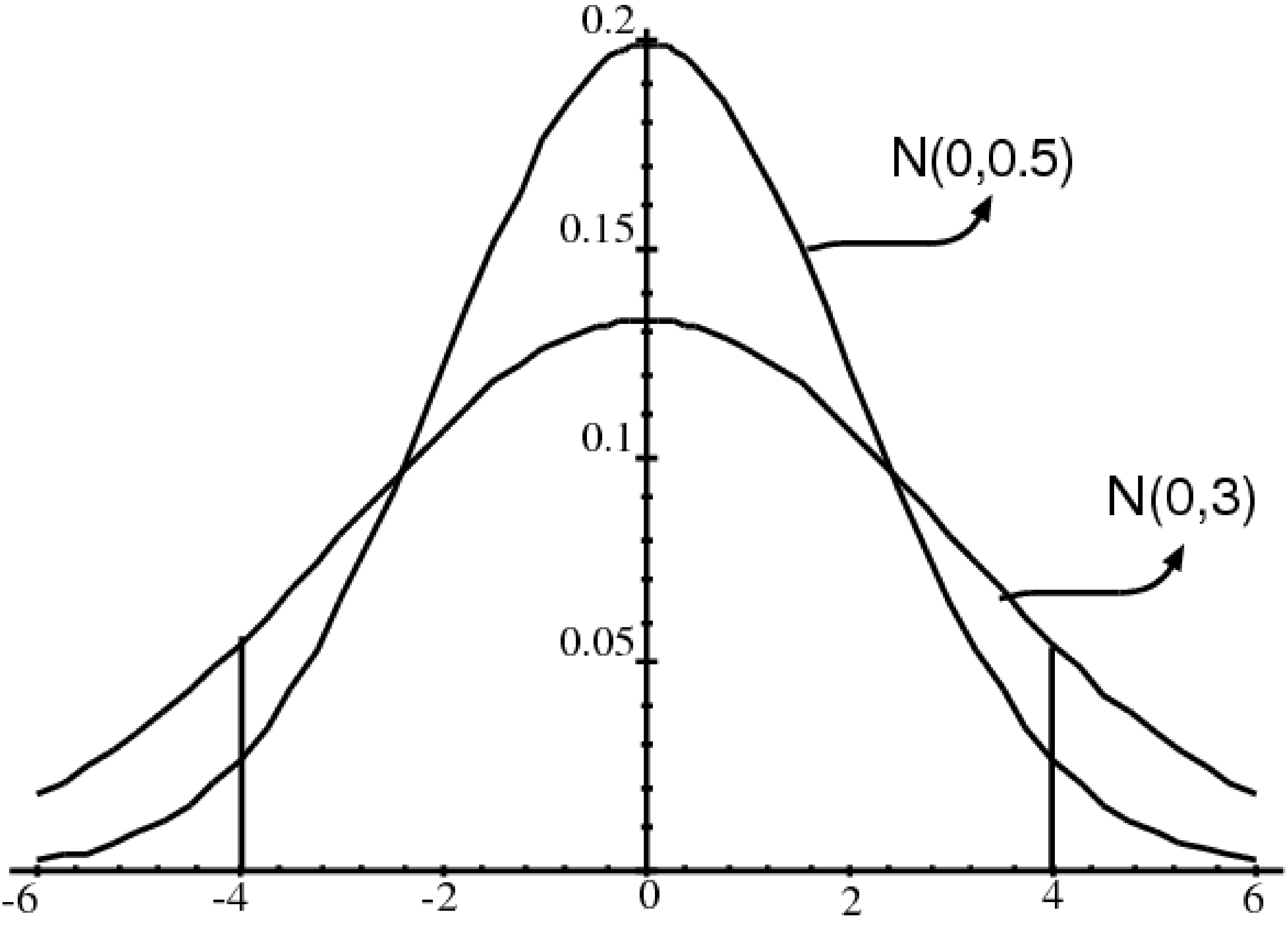


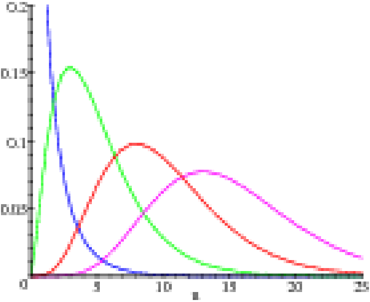
$P(|x-\mu|) < \frac{1}{k^2}$
$\frac{1}{2k^2}$
$\frac{1}{2k^2}$
μ
$\mu-2\sigma$
$\mu+2\sigma$



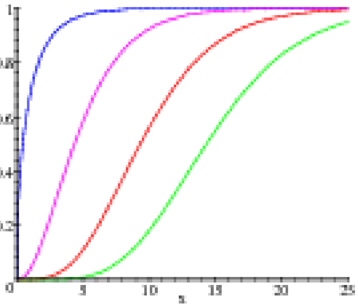
0.08
0.06
0.04
0.02
0
30
40
50
60
70
80
$\mu-2\sigma$
μ
$\mu+2\sigma$
$1-\frac{1}{k^2}$

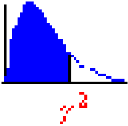

0.2
0.15
0.1
0.05
N(0,0.5)
N(0,3)
-6
-4
-2
0
2
4
6

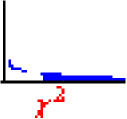

0.2
0.15
0.1
0.05
N(0,0.5)
N(0,3)
-6
-4
-2
0
2
4
6

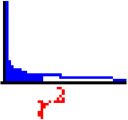


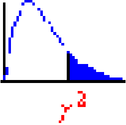
0.2
0.15
0.1
0.05
0
5
10
15
20
25
x

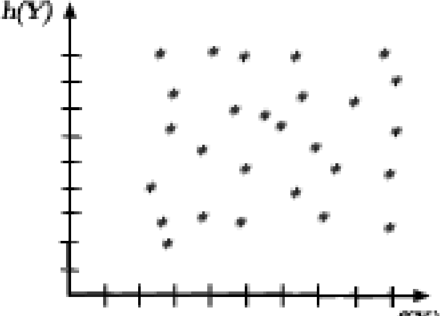

1
0.8
0.6
0.4
0.2
0
5
10
15
20
25
s



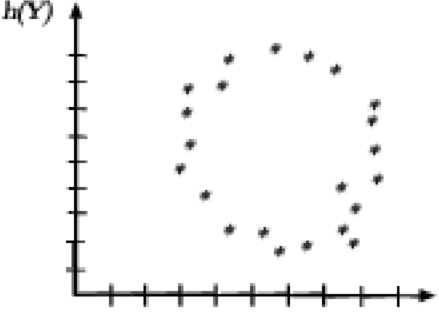




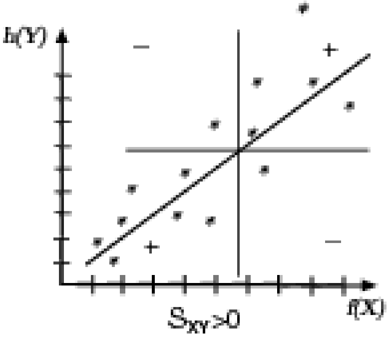

χ^2

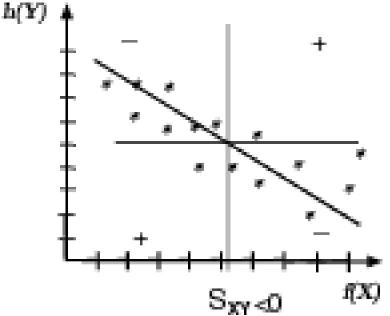


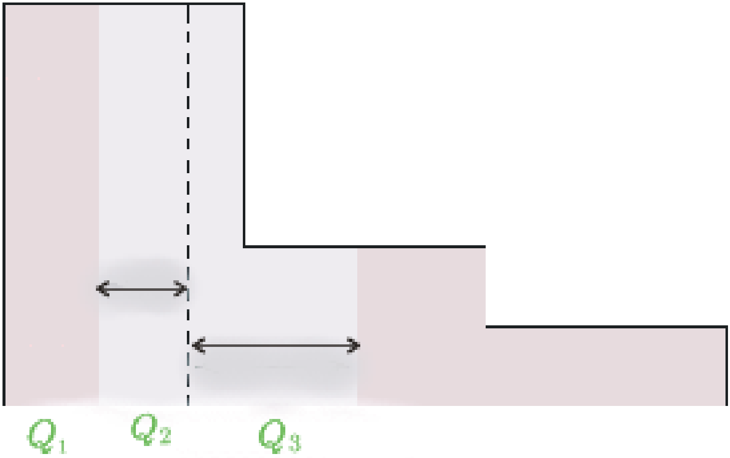
$S_{XY}=0$



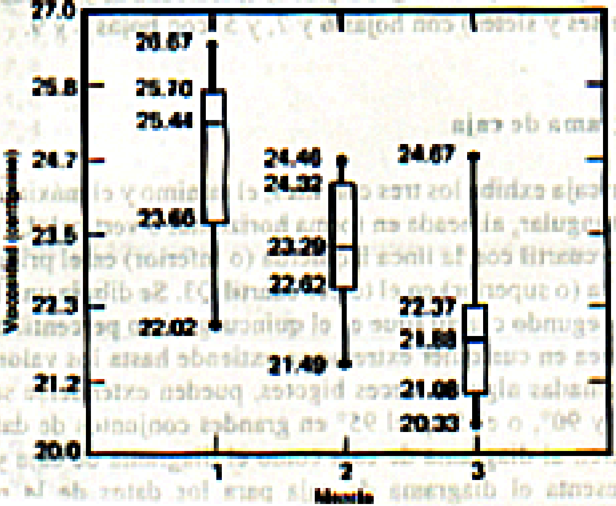
$S_{XY}=0$


h(Y)
−
+
+
−
f(X)
S_XY>0

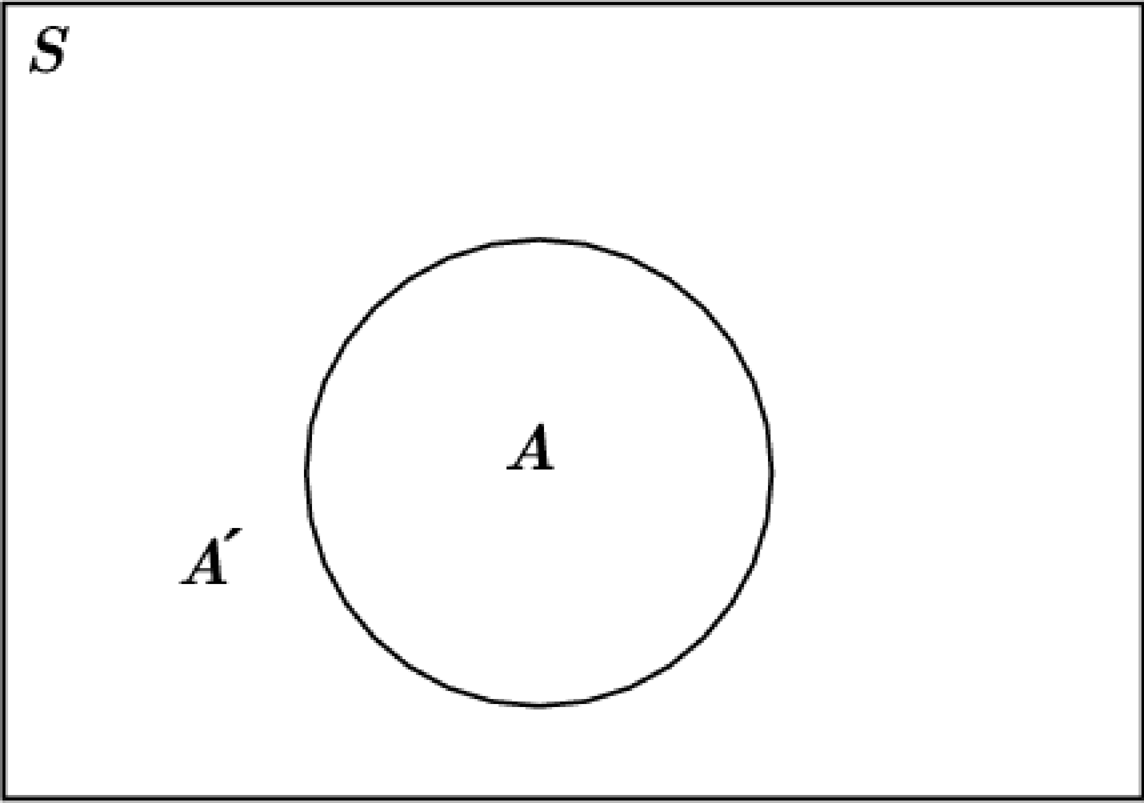

h(Y)
−
+
+
−
f(X)
S_XY<0

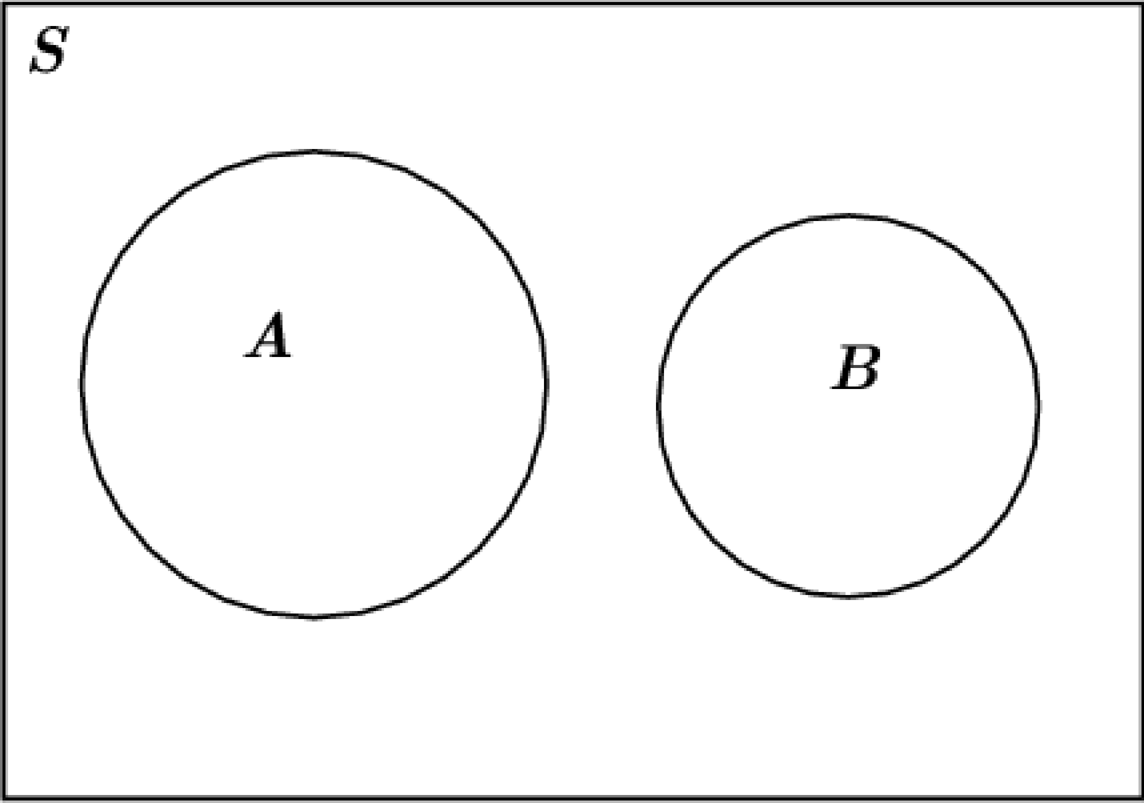


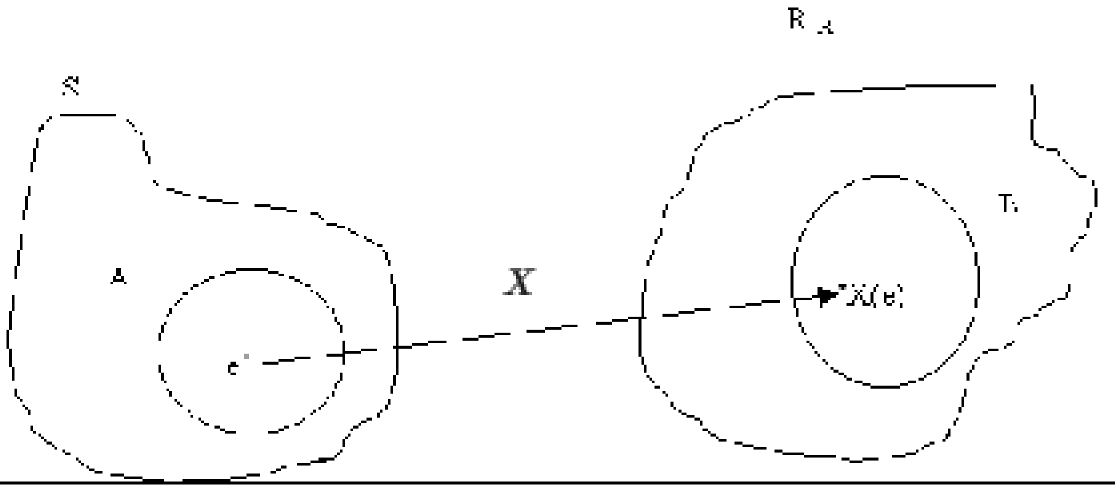
Q_1
Q_2
Q_3



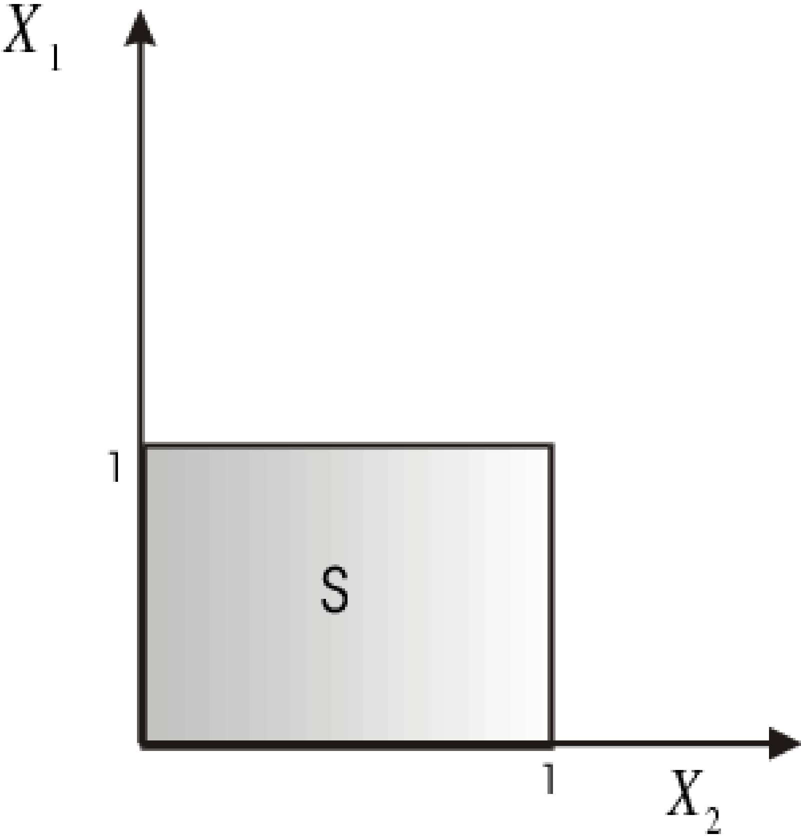
27.0
25.8
24.7
23.5
22.3
21.2
20.0
25.70
25.44
23.66
22.02
24.46
24.32
23.29
22.62
21.49
24.67
22.37
20.33
1
2
3
Mezcla

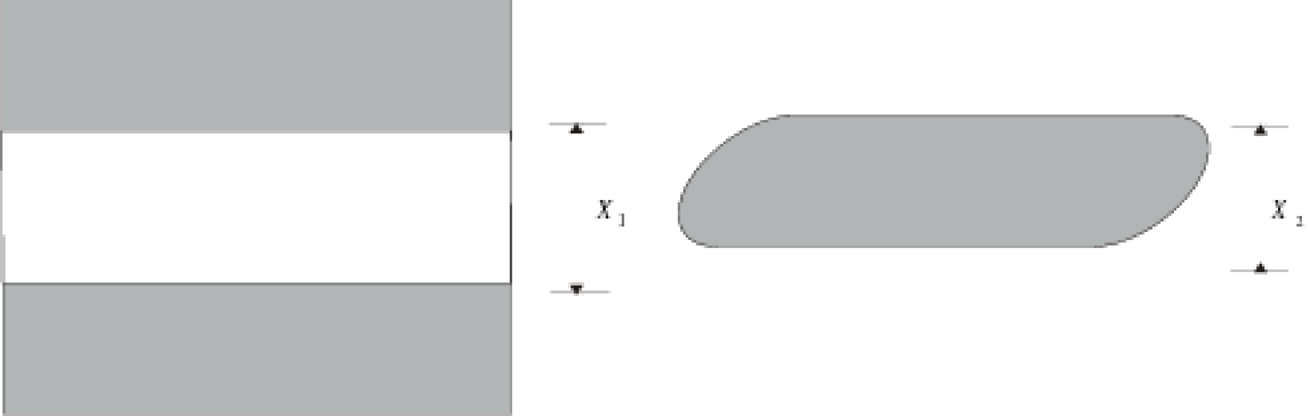

S
A
A′

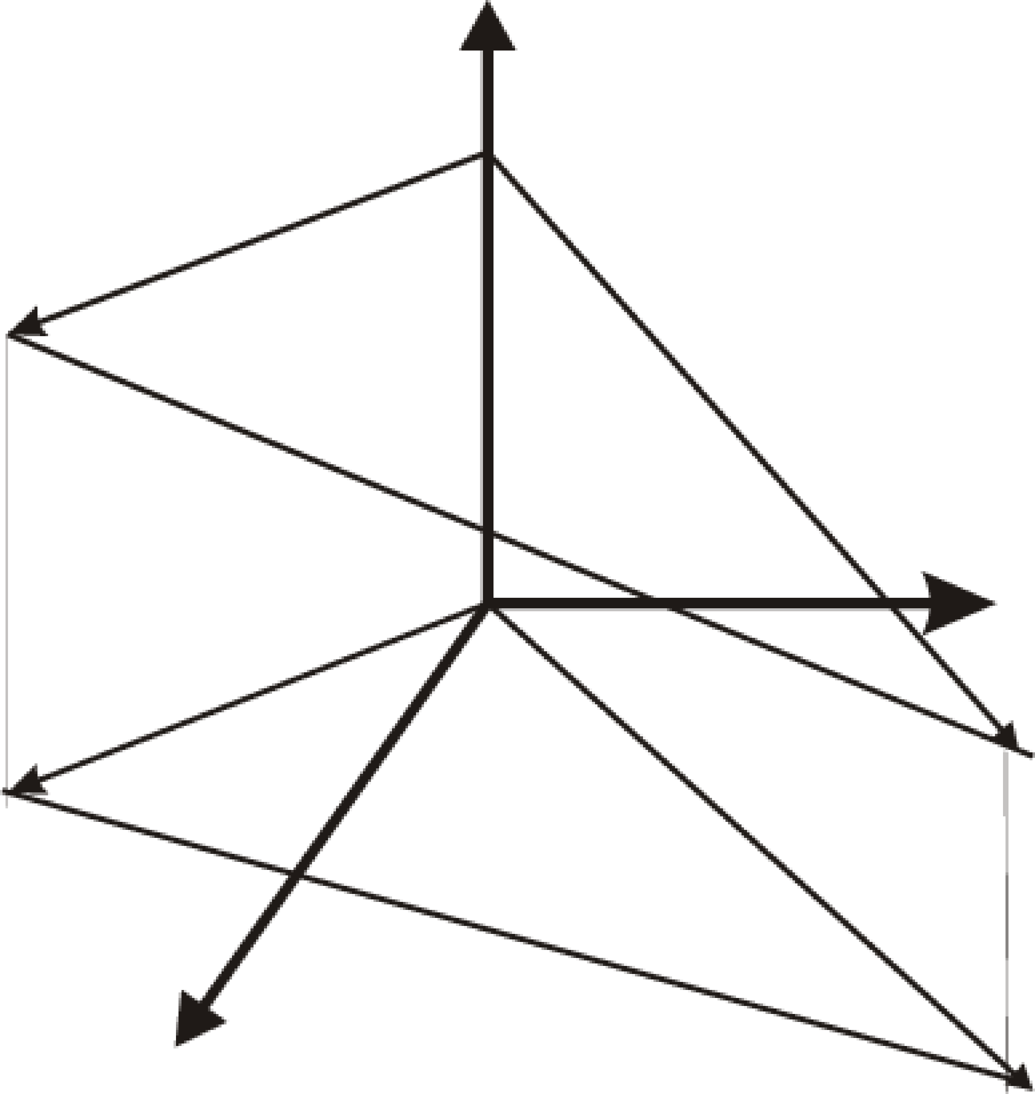

S
A
B

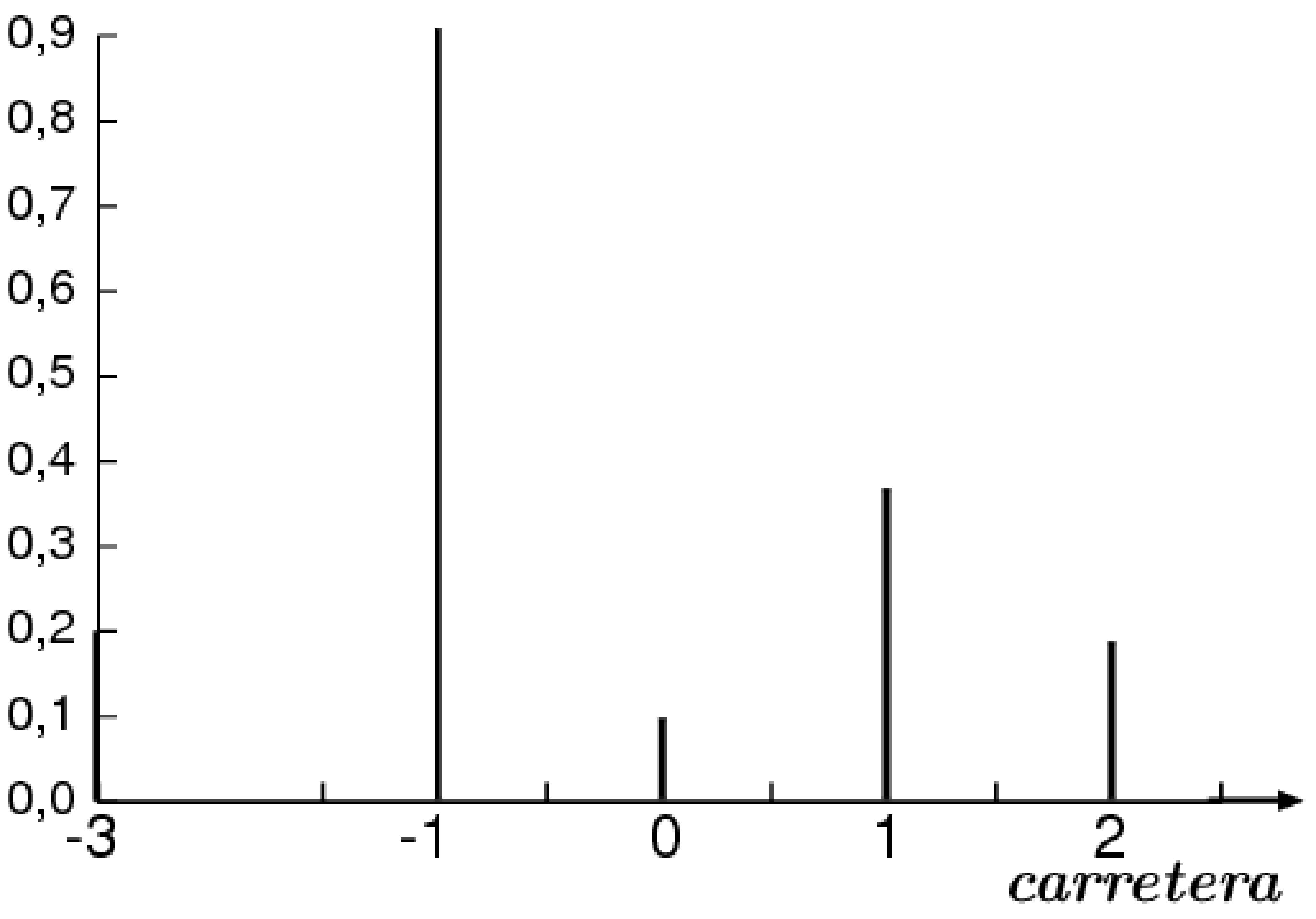


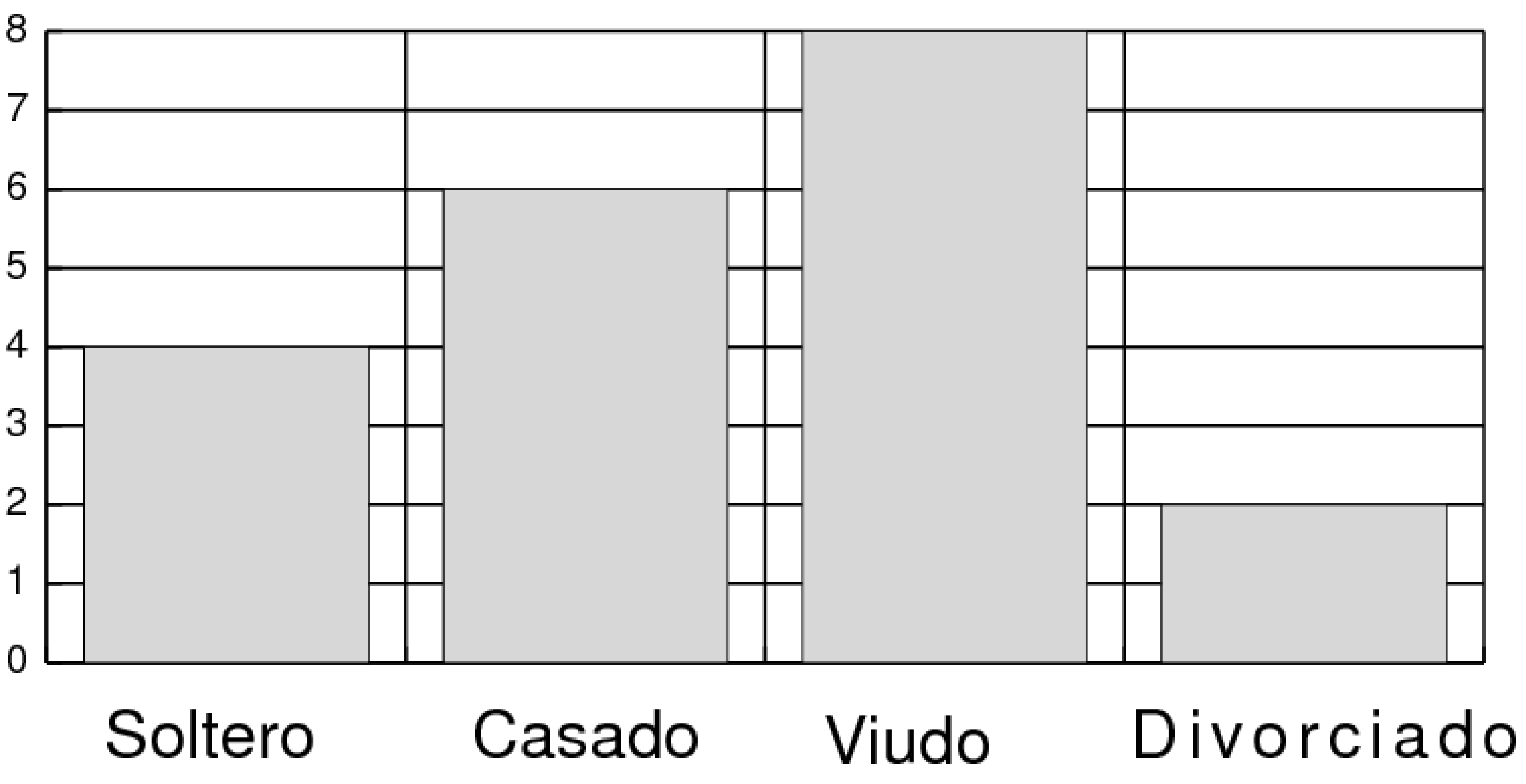
R
S
A
X
e
X(e)

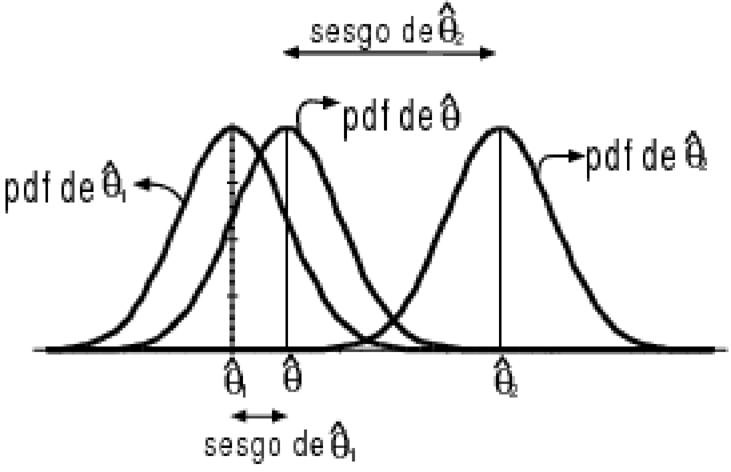

X_1
1
S
1
X_2


X_1
X_2

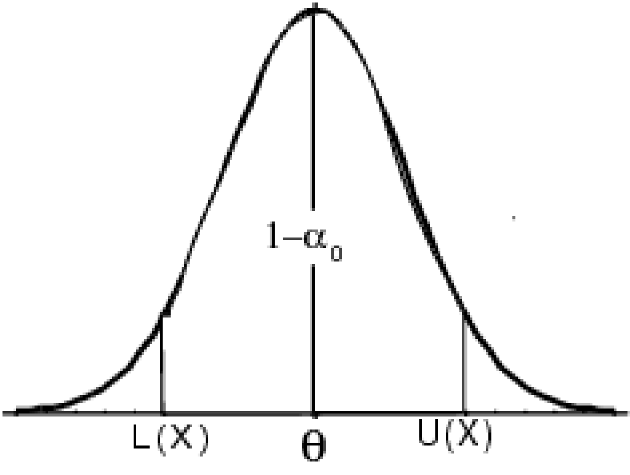



0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-3
-1
0
1
2
carretera

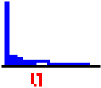

8
7
6
5
4
3
2
1
0
Soltero
Casado
Viudo
Divorciado

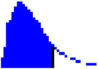


sesgo de $\hat{\theta}_2$
pdf de $\hat{\theta}$
pdf de $\hat{\theta}_1$
pdf de $\hat{\theta}_2$
$\hat{\theta}_1$
$\hat{\theta}$
$\hat{\theta}_2$
sesgo de $\hat{\theta}_1$

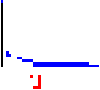


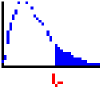
$1-\alpha_0$
L(X)
θ
U(X)

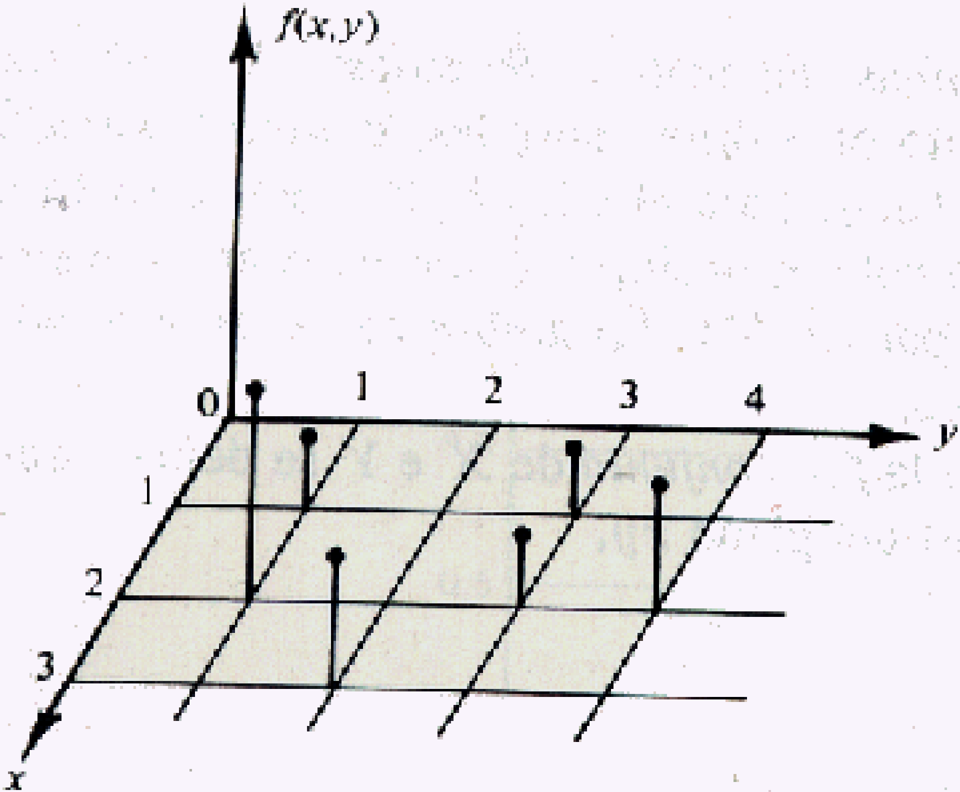


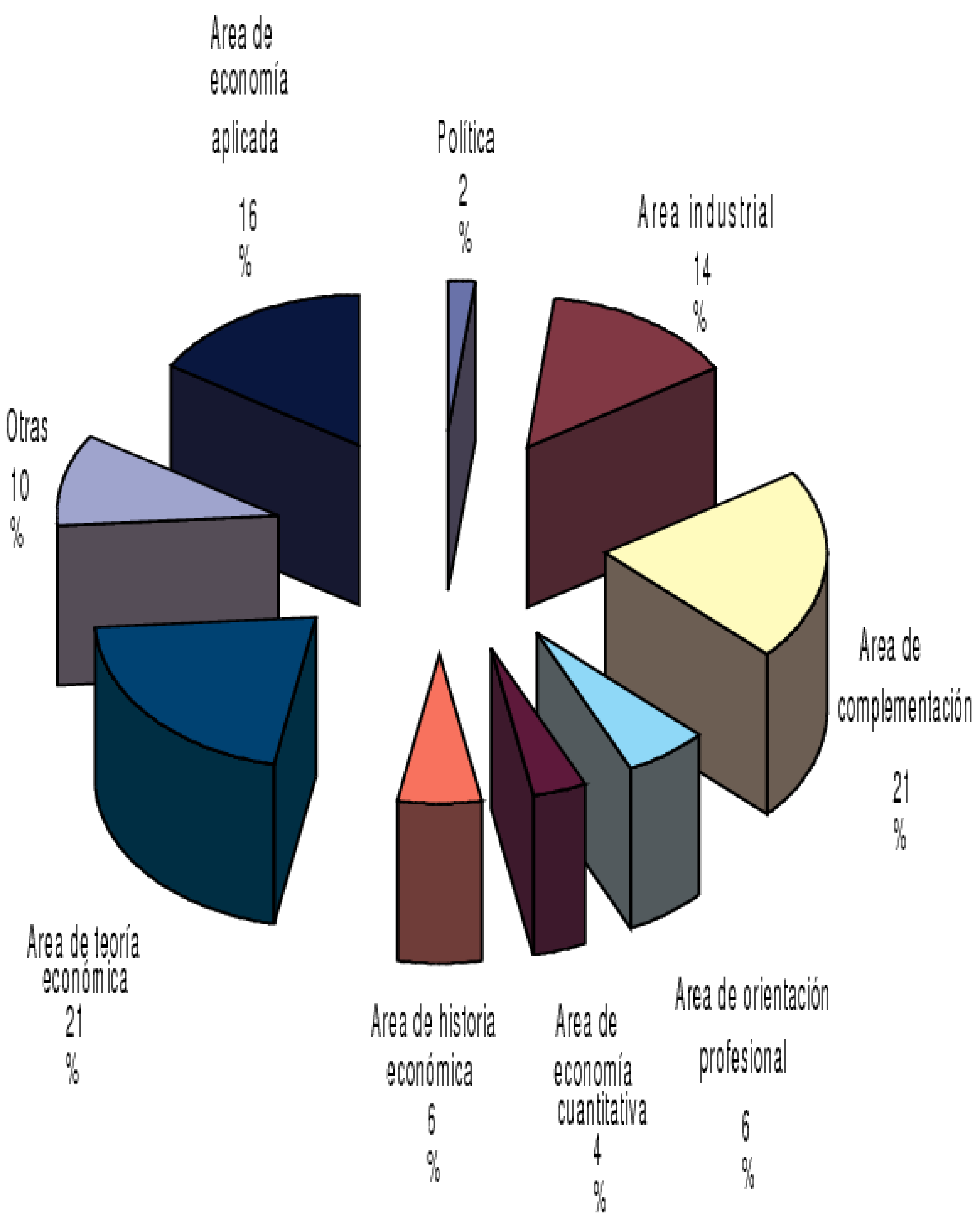


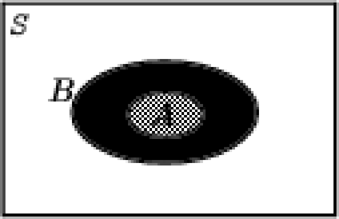
[1]

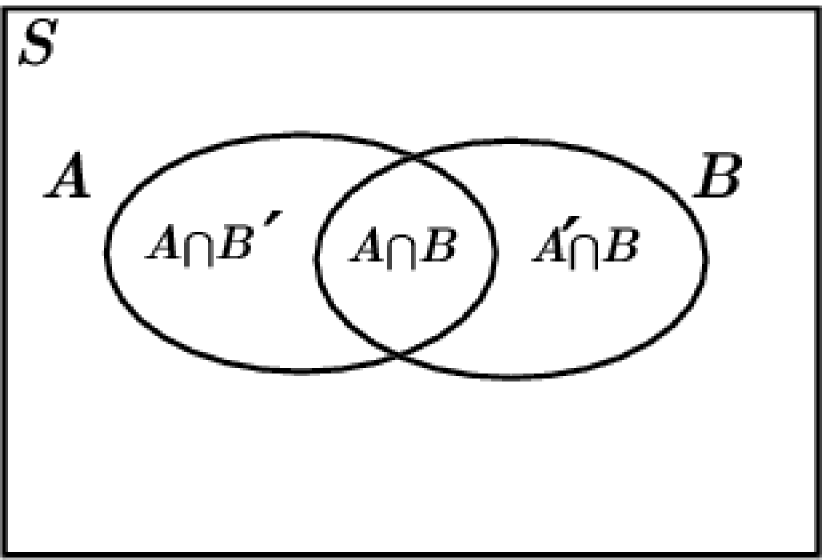


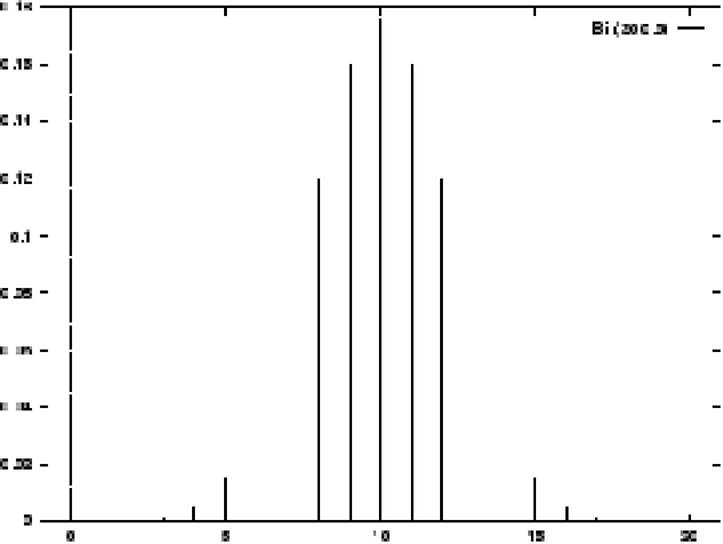



f(x,y)
0
1
2
3
4
y
1
2
3
x

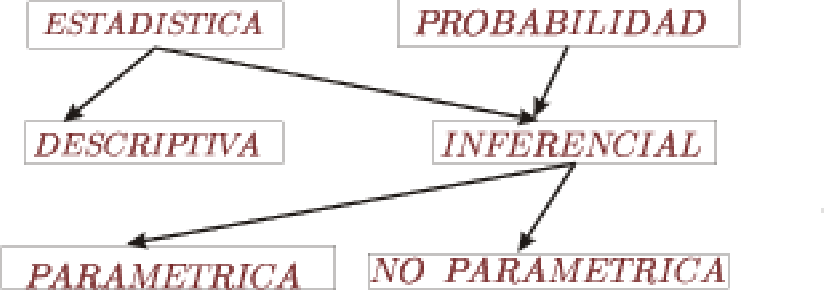

Area de economía aplicada
16 %
Política
2 %
Area industrial
14 %
Otras
10 %
Area de complementación
21 %
Area de teoría económica
21 %
Area de historia económica
6 %
Area de economía cuantitativa
4 %
Area de orientación profesional
6 %

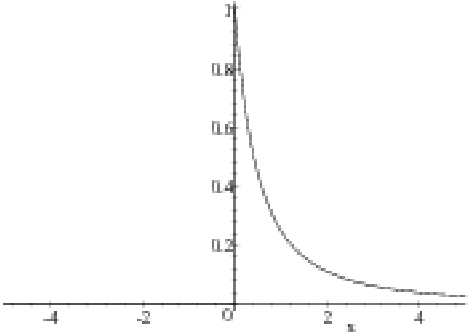

S
B
A


S
A
B
A∩B′
A∩B
A′∩B

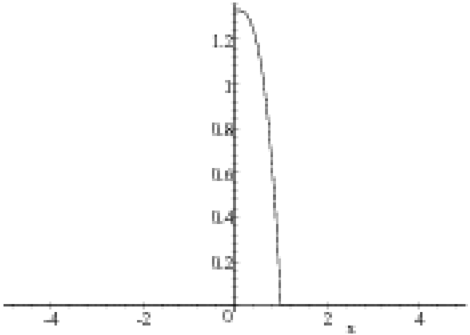


0.1
0
5
10
15
20

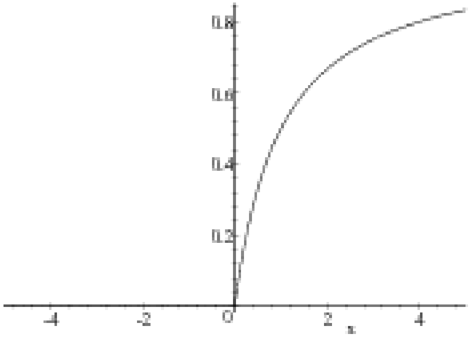

ESTADISTICA
PROBABILIDAD
DESCRIPTIVA
INFERENCIAL
PARAMETRICA
NO PARAMETRICA

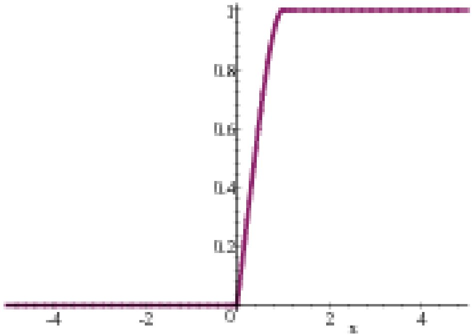


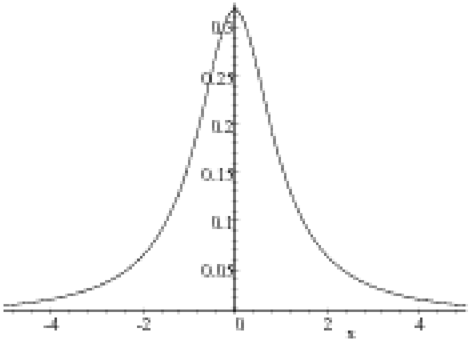
1
0.8
0.6
0.4
0.2
-4
-2
O
2
4
x



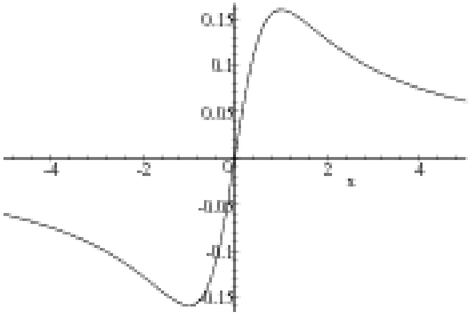
1.2
1
0.8
0.6
0.4
0.2
-4
-2
0
2
4
x

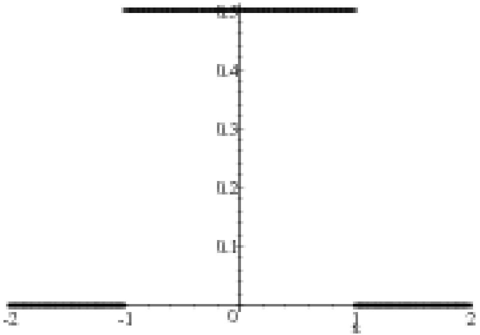

0.8
0.6
0.4
0.2
-4
-2
O
2
4
x


1
0.8
0.6
0.4
0.2
-4
-2
0
2
4
x

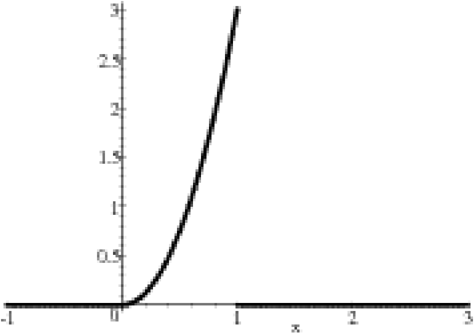


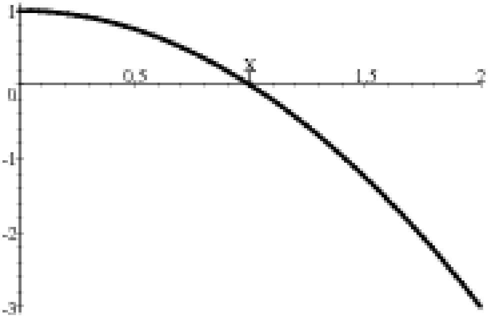
0.3
0.25
0.2
0.15
0.1
0.05
-4
-2
0
2
4
x


0.15
0.1
0.05
-4
-2
0
2
4
x
-0.05
-0.1
-0.15

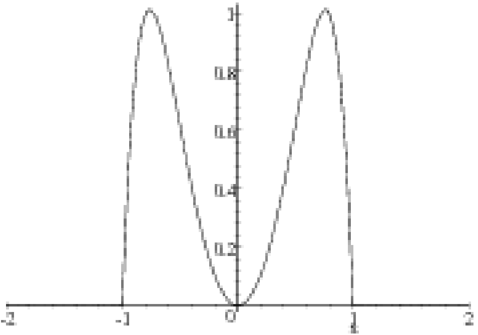


0.4
0.3
0.2
0.1
-2
-1
O
1
2

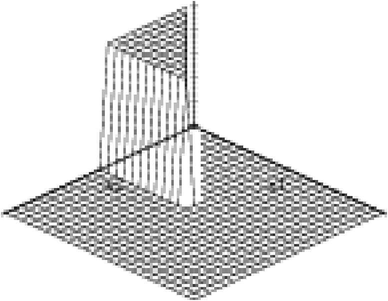

3
2.5
2
1.5
1
0.5
-1
0
1
2
3
x

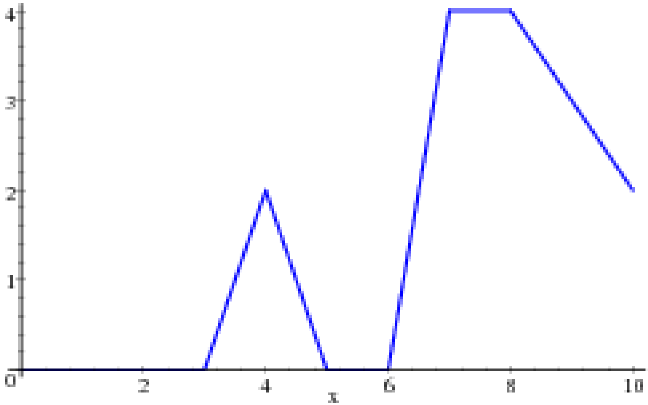


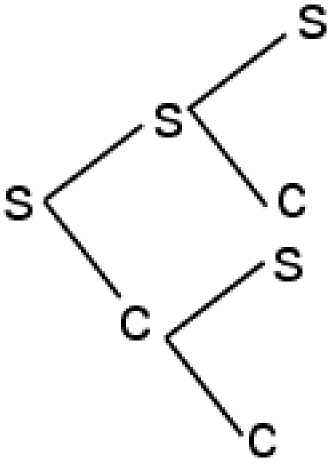
1
0
-1
-2
-3
0.5
x
1.5
2

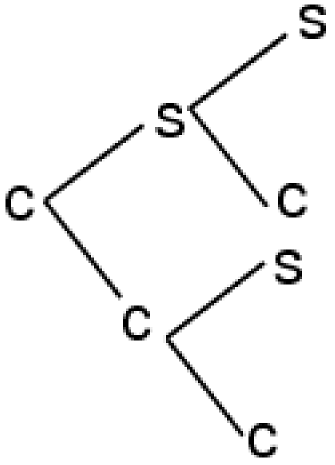


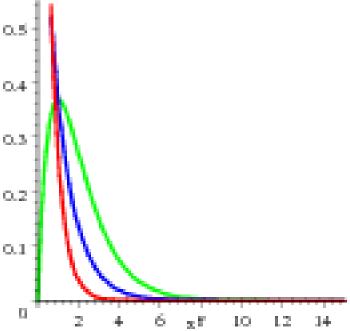
1
0.8
0.6
0.4
0.2
-2
-1
O
1
2



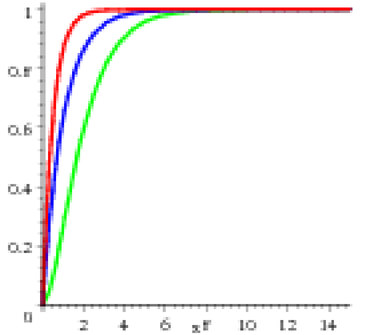

4
3
2
1
0
2
4
6
8
10
x

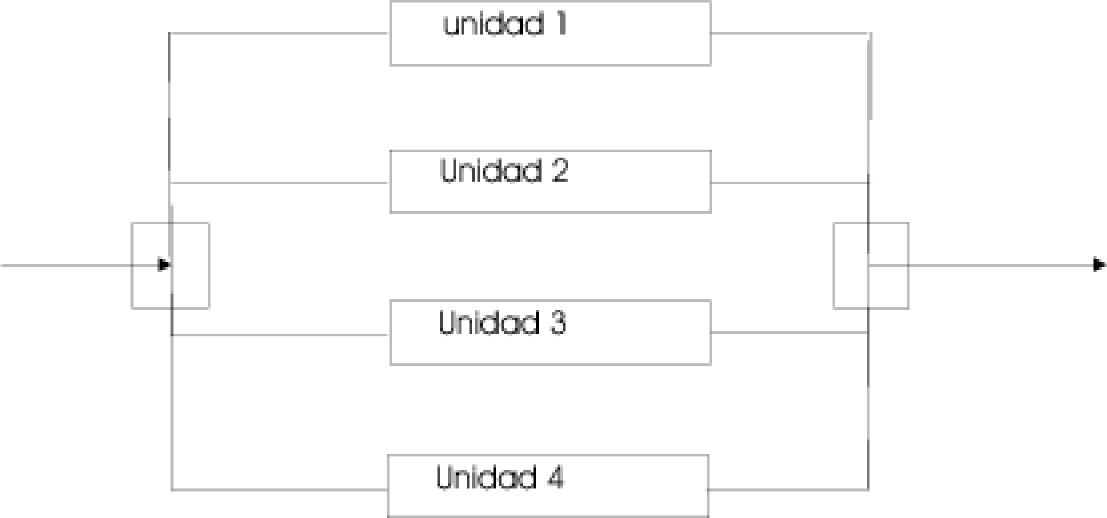


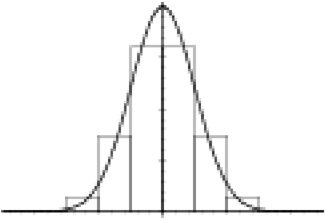


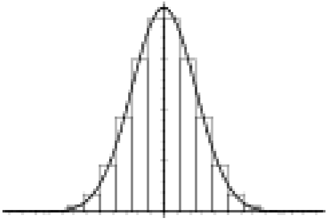


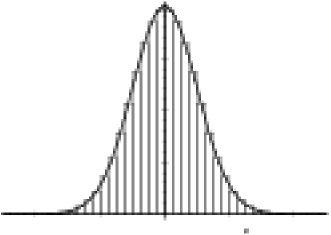
0.5
0.4
0.3
0.2
0.1
0
2
4
6
x
8
10
12
14

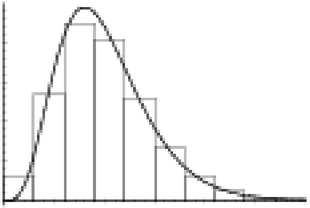

1
0.8
0.6
0.4
0.2
0
2
4
6
x 8
10
12
14

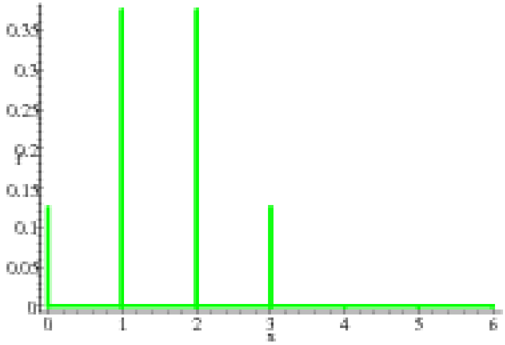

unidad 1
Unidad 2
Unidad 3
Unidad 4



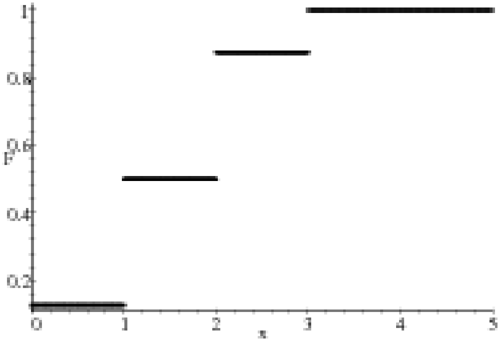




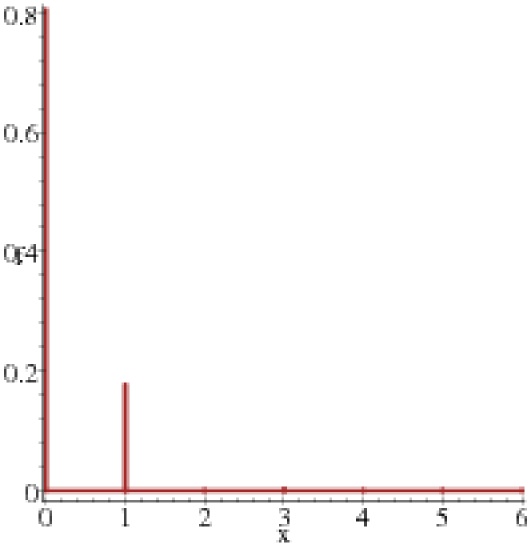




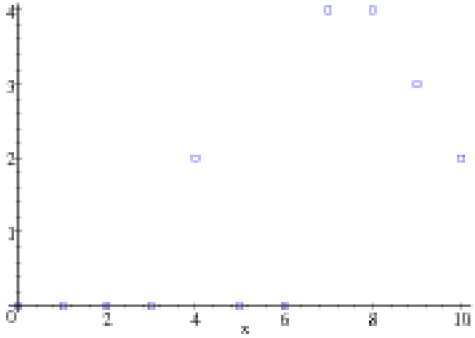
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
1
2
3
4
5
6
x



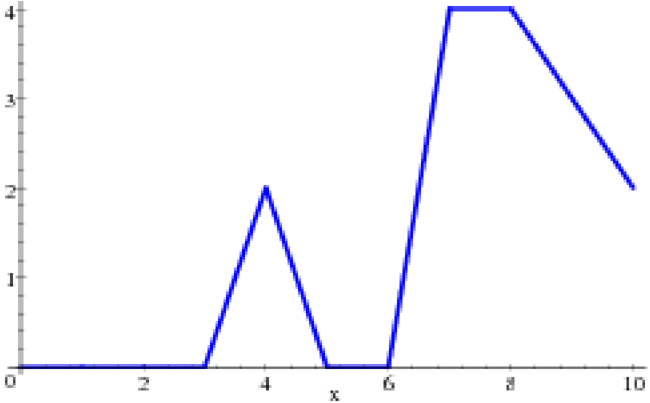
1
0.8
0.6
F
0.4
0.2
0
1
2
3
4
5
x



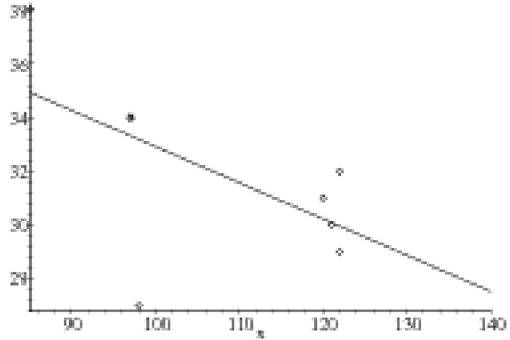
0.8
0.6
0.4
0.2
0
0
1
2
3
4
5
6
x



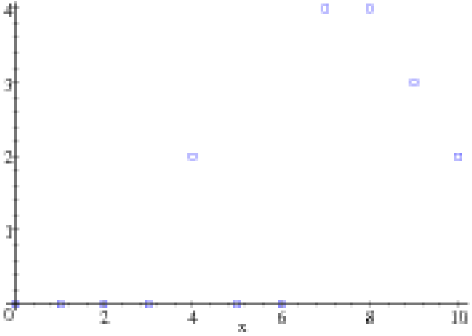
4
3
2
1
0
2
4
6
8
10
x

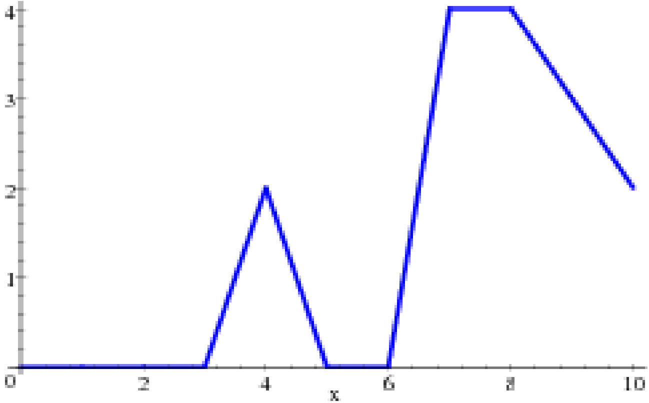


4
3
2
1
0
2
4
6
8
10
x

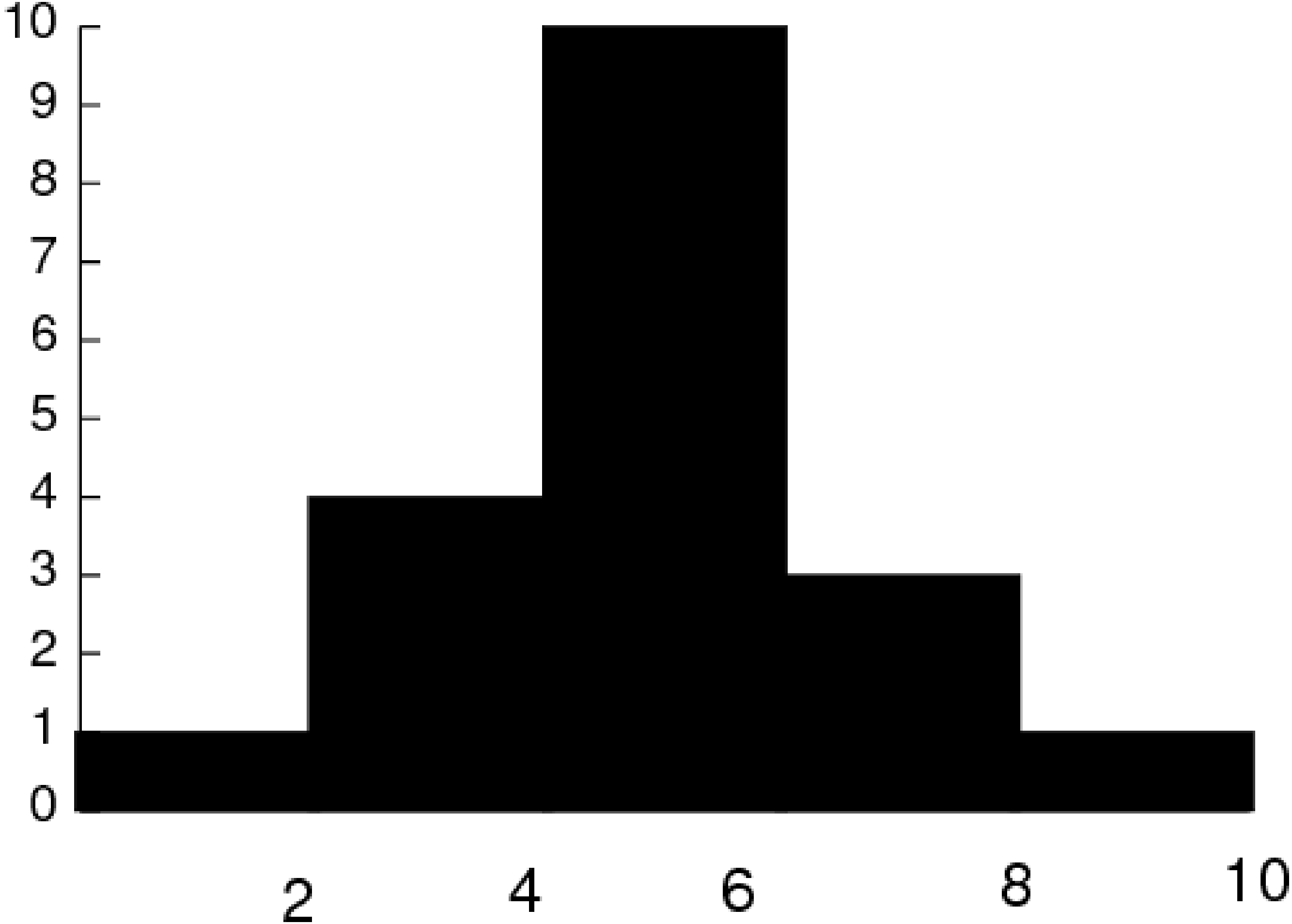


90
100
110
x
120
130
140

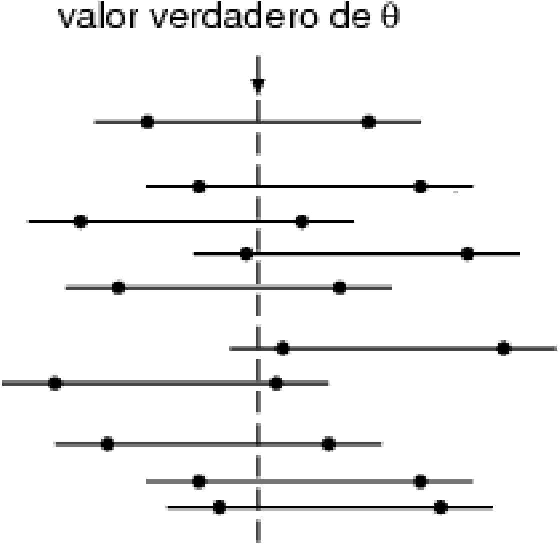

4
3
2
1
0
2
4
6
8
10
x

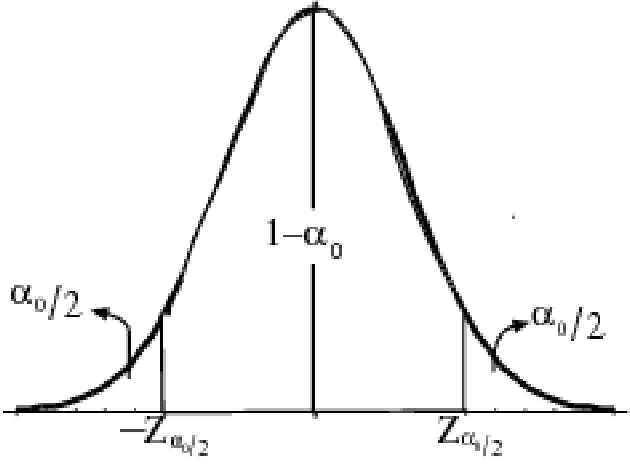


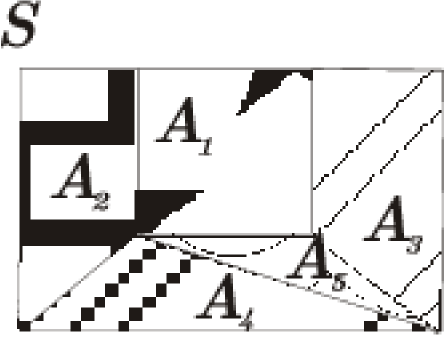
4
3
2
1
0
2
4
6
8
10
x

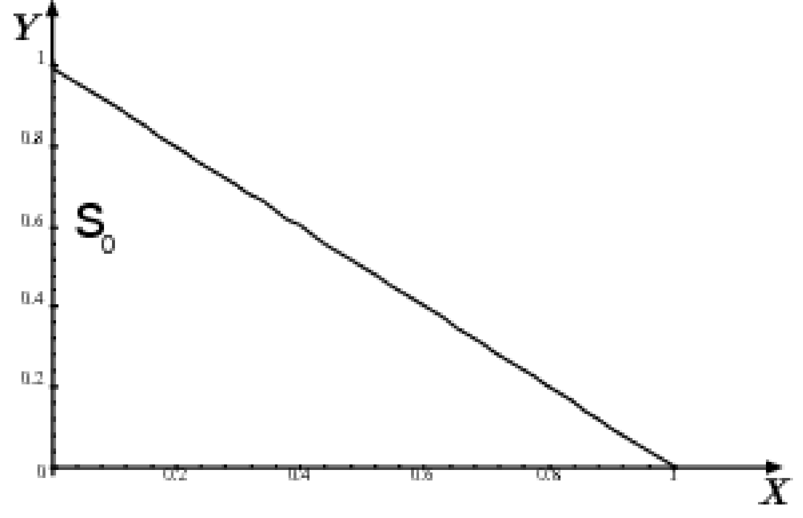


10
9
8
7
6
5
4
3
2
1
0
2
4
6
8
10

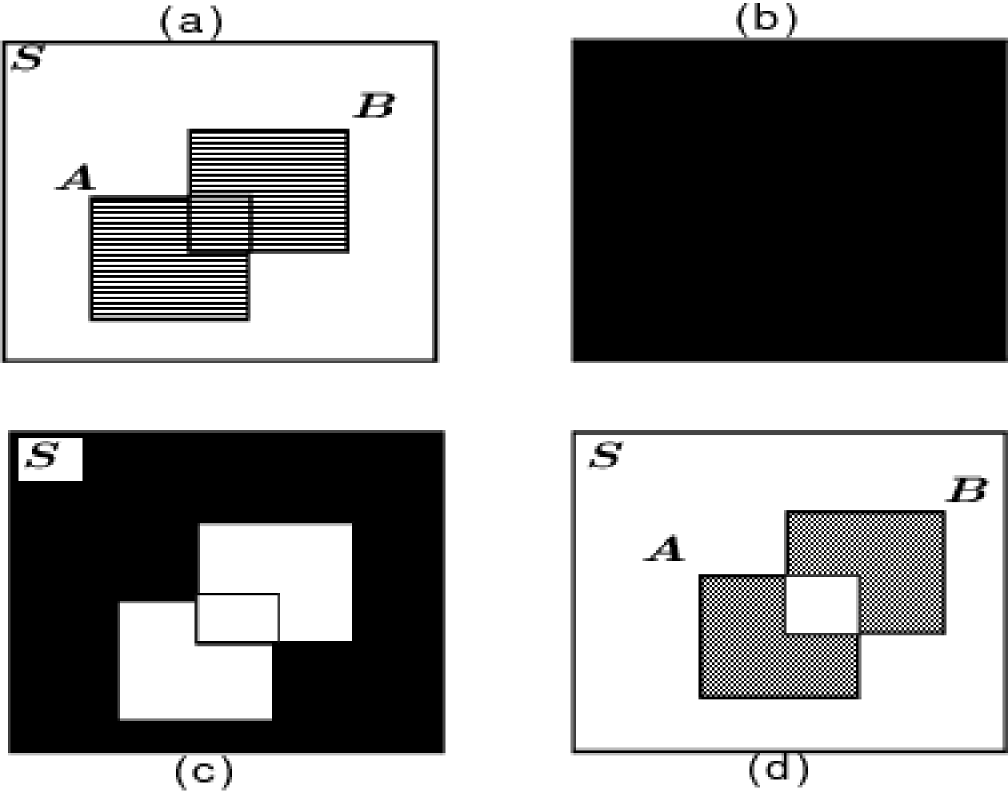

valor verdadero de θ

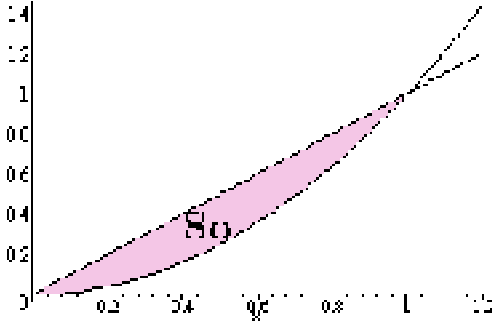

$1-\alpha_0$
$\alpha_0/2$
$\alpha_0/2$
$-Z_{\alpha_0/2}$
$Z_{\alpha_0/2}$


S
A1
A2
A3
A4
A5

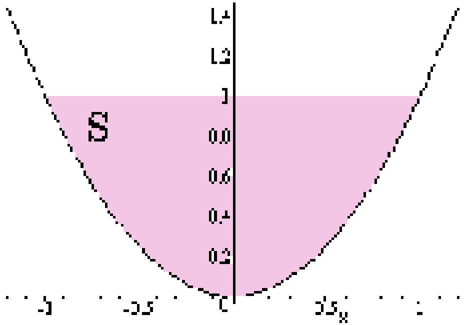


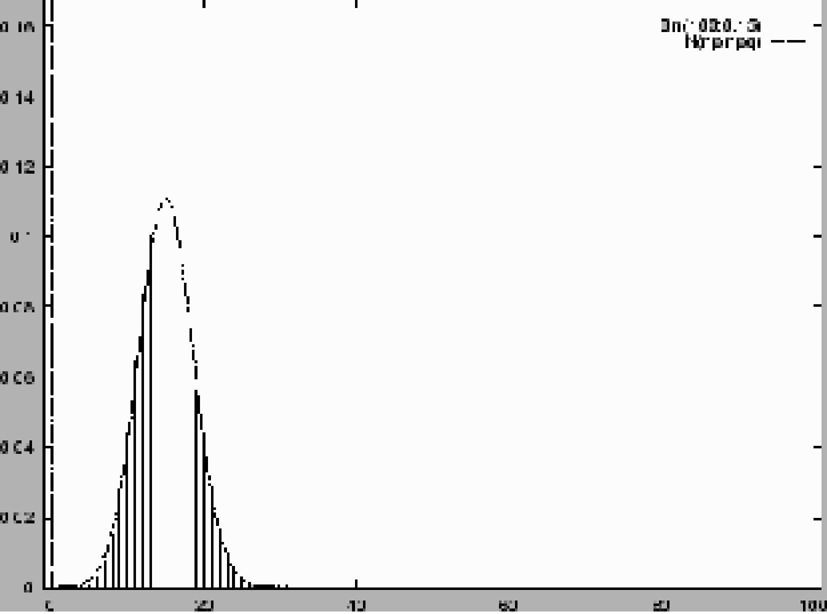
Y
1
0.8
0.6
0.4
0.2
0
S_0
0.2
0.4
0.6
0.8
1
X

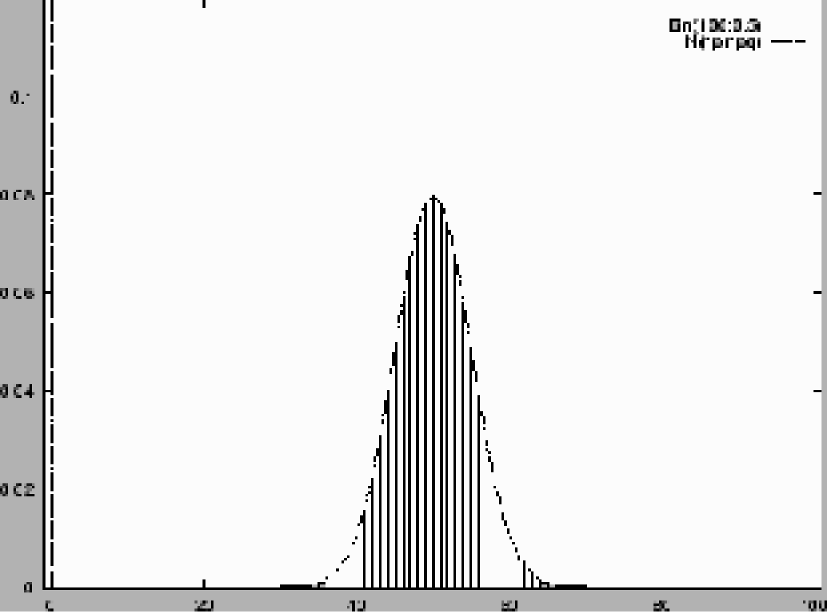

(a)
S
B
A
(b)
S
(c)
S
B
A
(d)

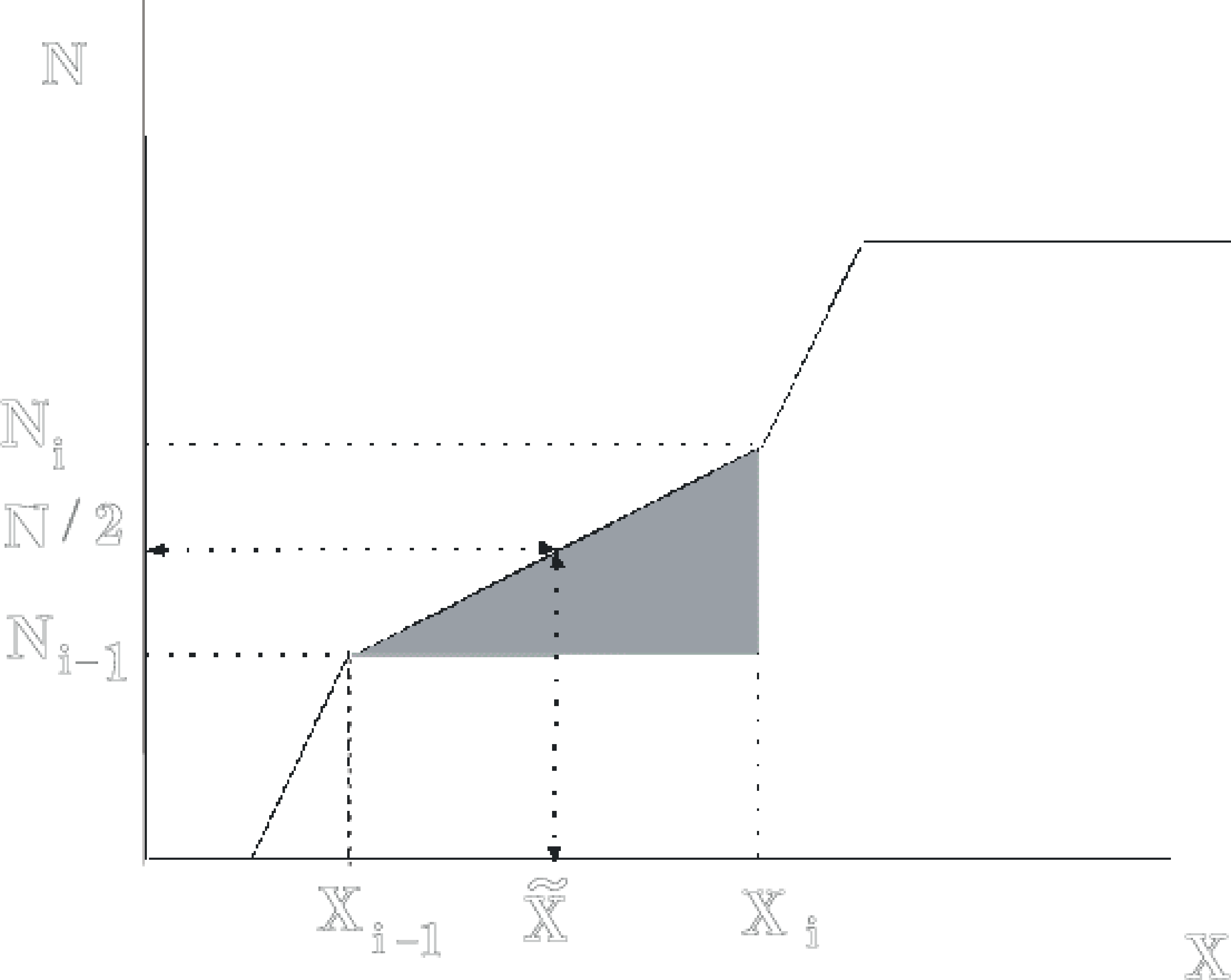


1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.2
0.4
0.6
0.8
1
1.2
S(t)

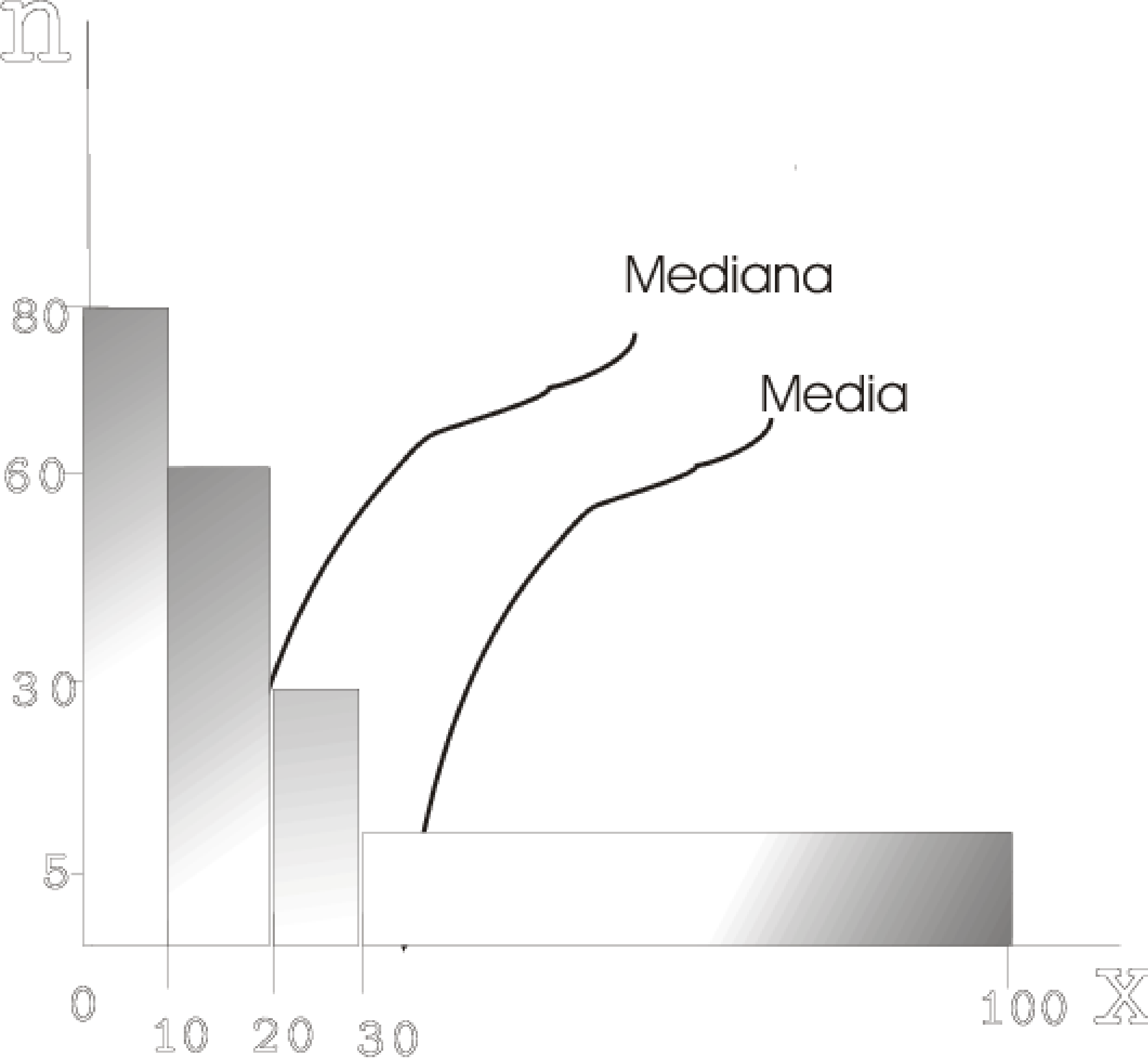

S

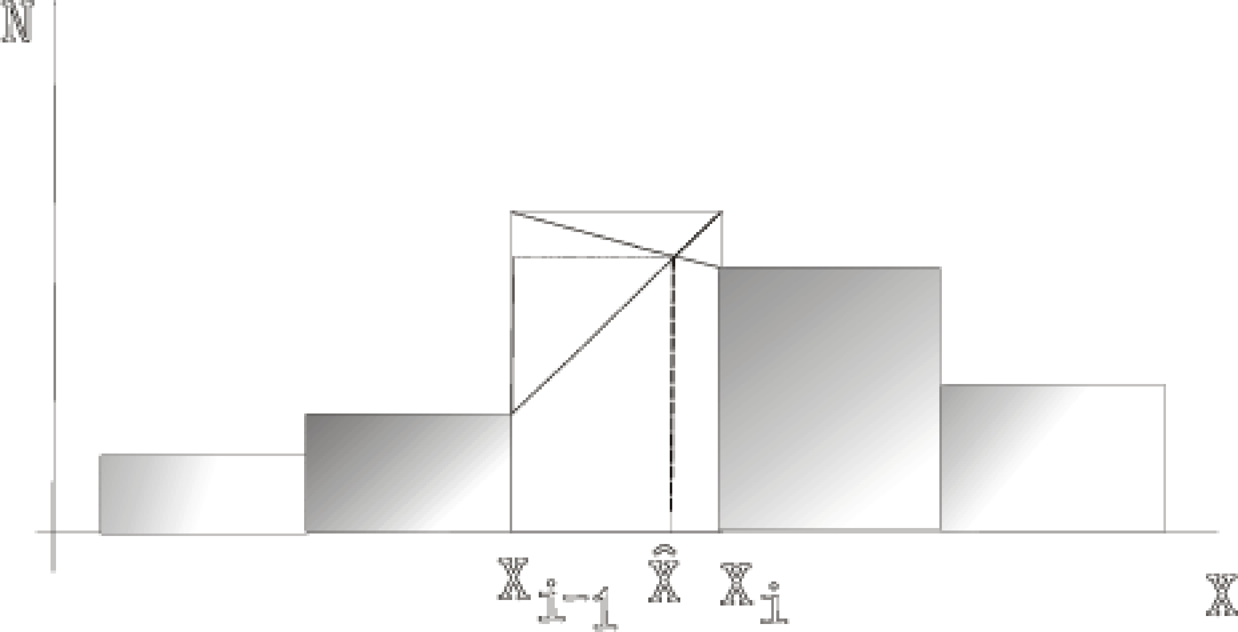


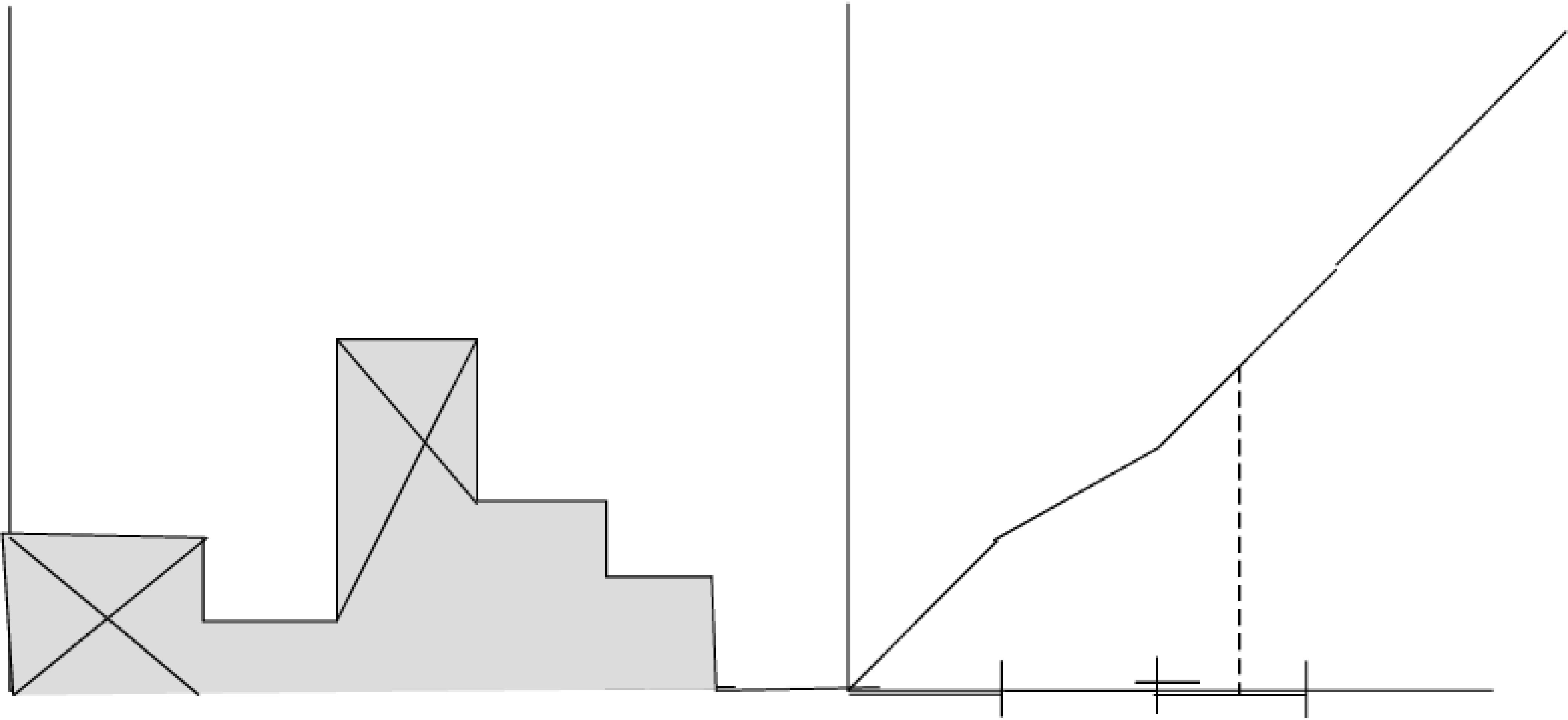


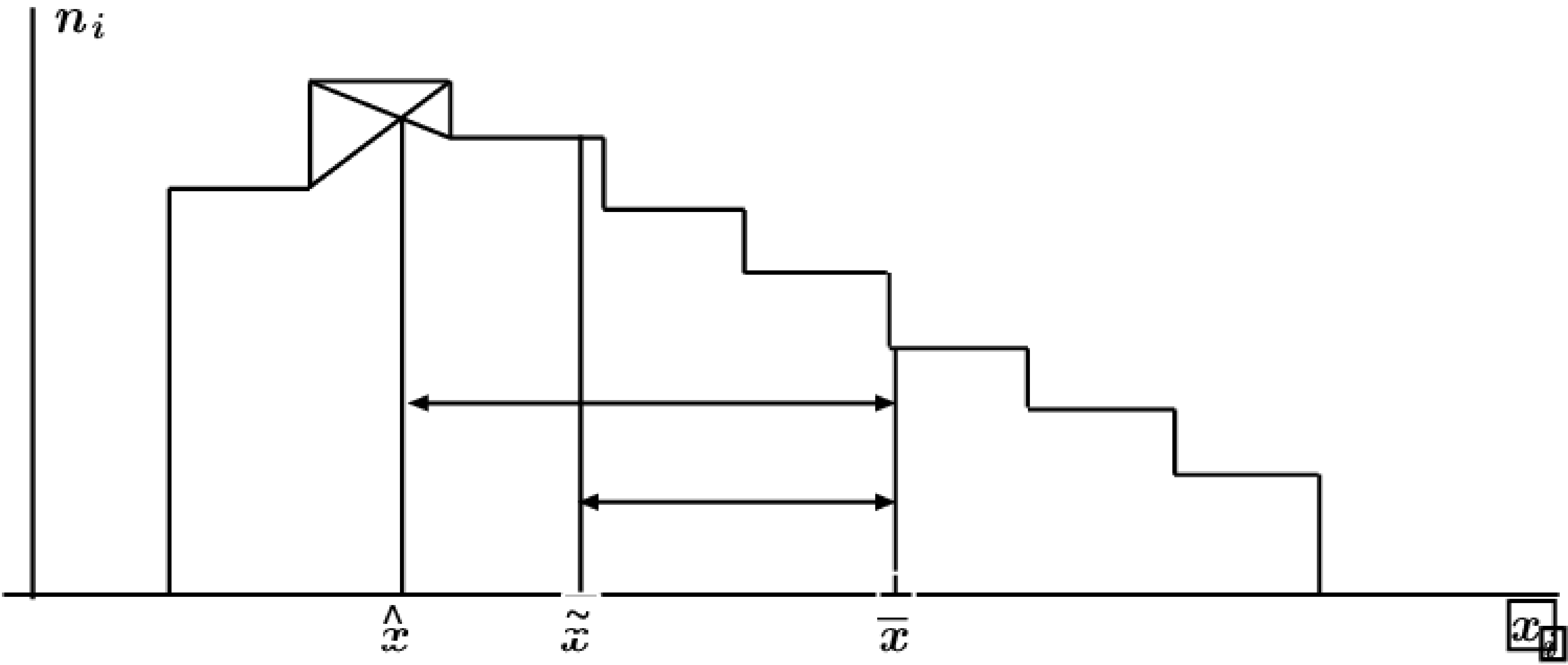


N
N_i
$N/2$
N_{i-1}
X_{i-1}
$\tilde{X}$
X_i
X

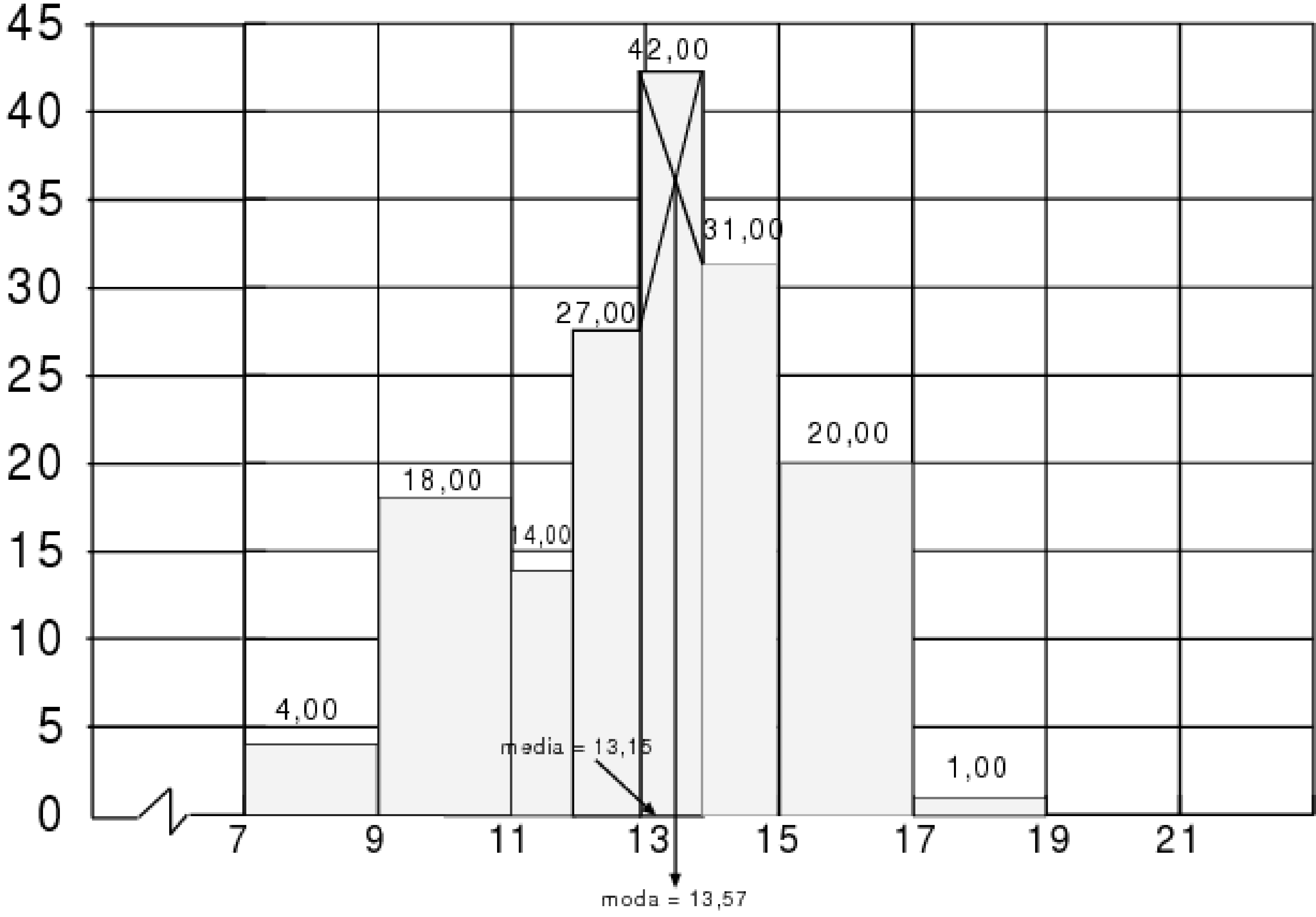

n
80
60
30
5
0
10
20
30
100
X
Mediana
Media


N
X_{i-1}
$\hat{X}$
X_i
X

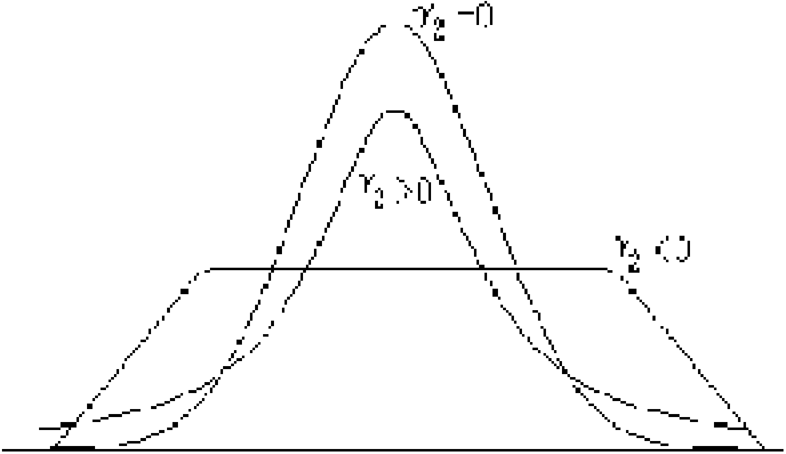


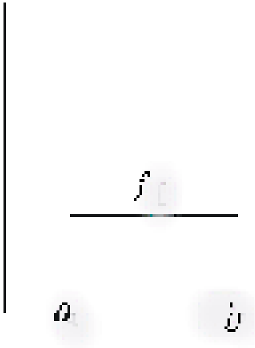


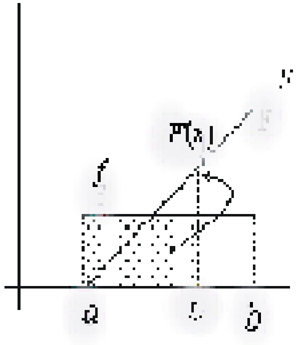
n_i
$\hat{x}$
$\tilde{x}$
$\bar{x}$
x_i

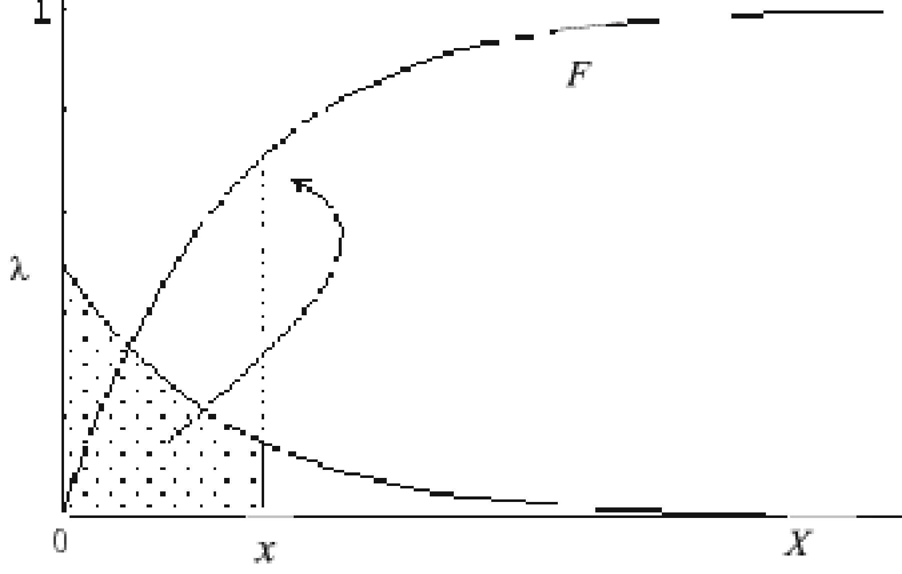


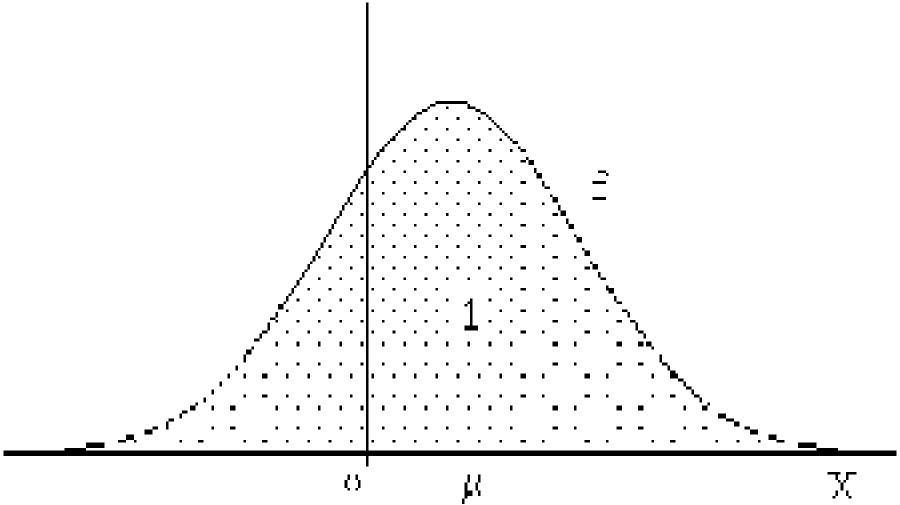
45
40
35
30
25
20
15
10
5
0
7
9
11
13
15
17
19
21
4,00
18,00
14,00
27,00
42,00
31,00
20,00
1,00
media = 13,15
moda = 13,57

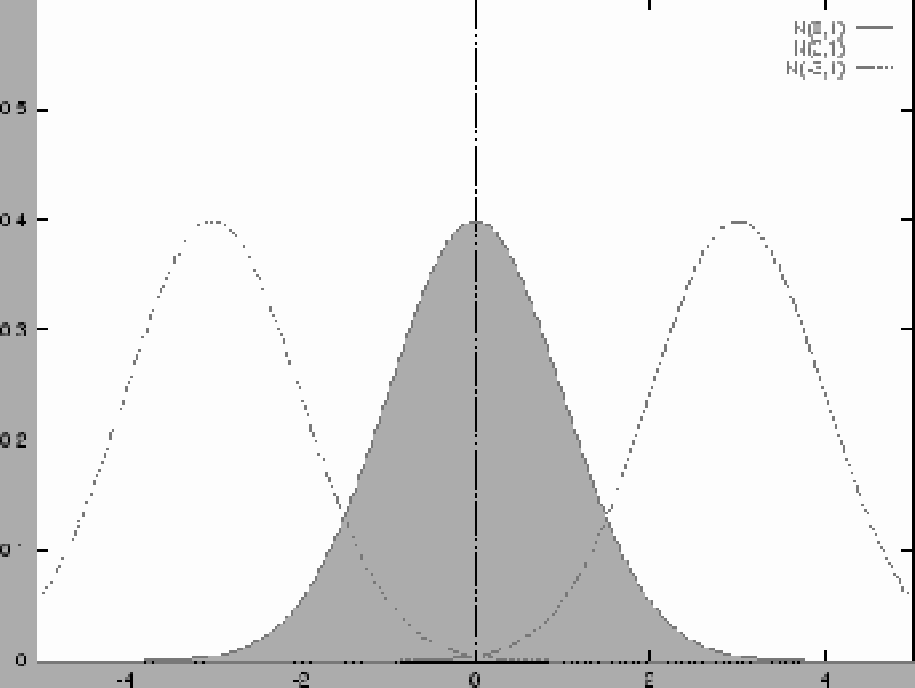

$\gamma_2 = 0$
$\gamma_2 > 0$
$\gamma_2 < 0$

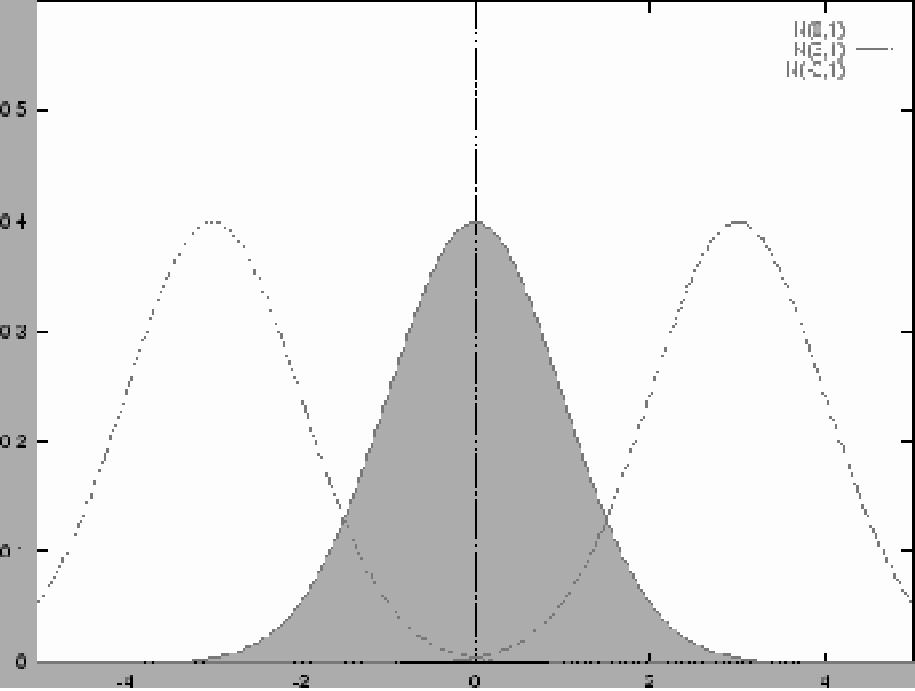

f
a
b

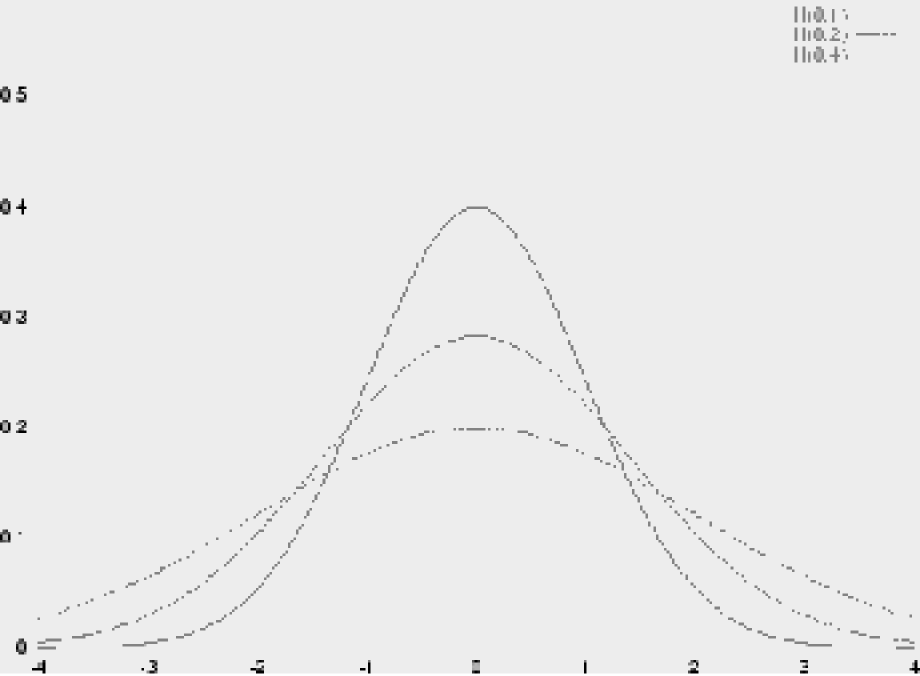

F(x)
f
a
b


1
F
λ
0
x
X

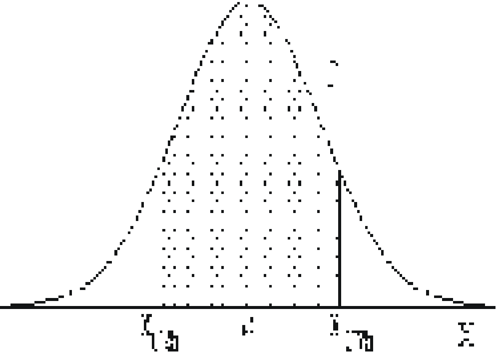

1
0
μ
X

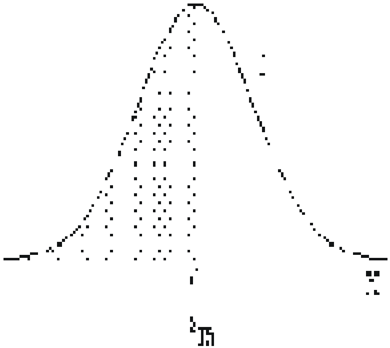

N(0,1)
N(3,1)
N(-3,1)
0.5
0.4
0.3
0.2
0.1
0
-4
-2
0
2
4

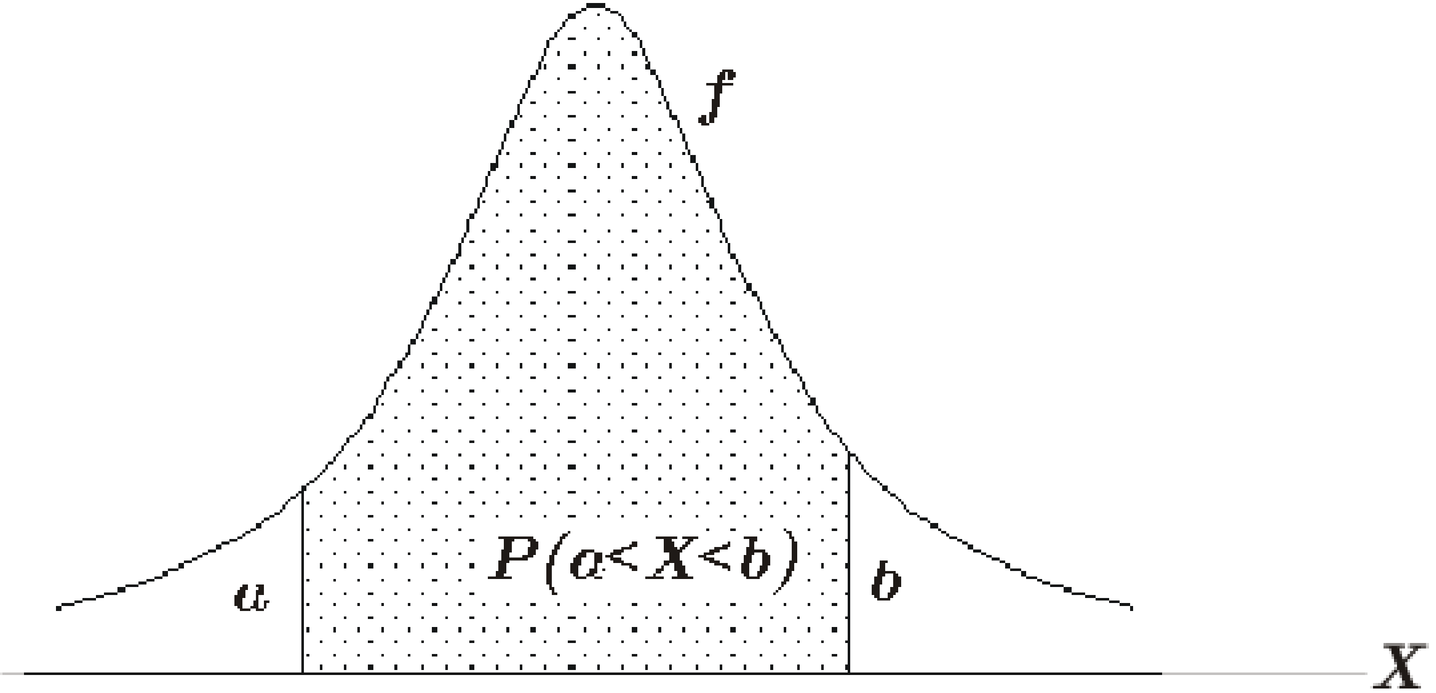

0.5
0.4
0.3
0.2
0.1
0
-4
-2
0
2
4

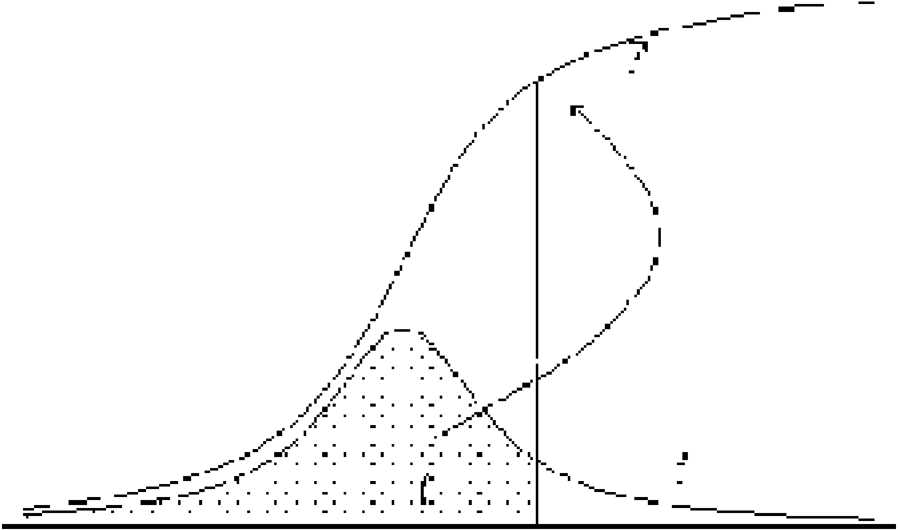


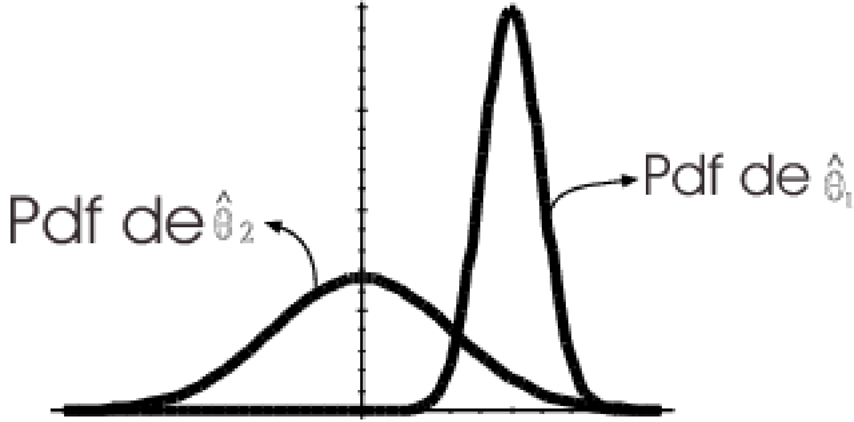
0.5
0.4
0.3
0.2
0
-4
-3
-2
-1
0
1
2
3
4



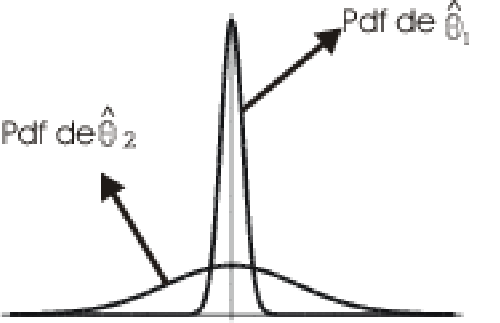


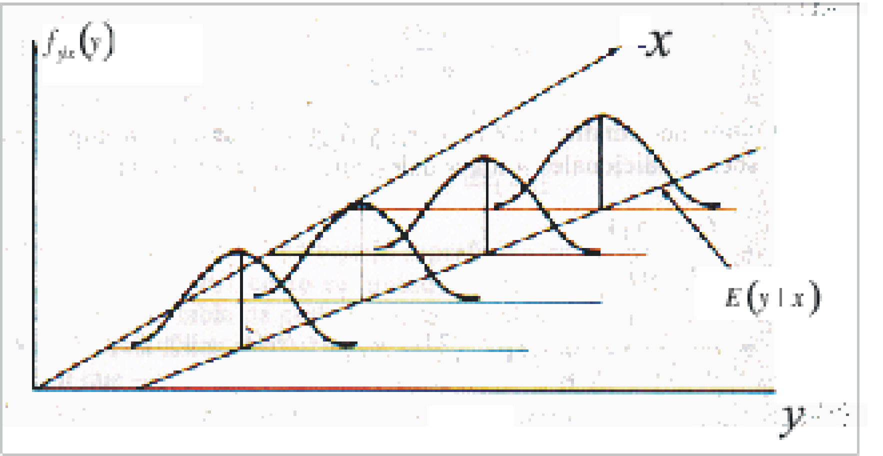

f
P(a<X<b)
a
b
X





Pdf de $\hat{\theta}_2$
Pdf de $\hat{\theta}_1$

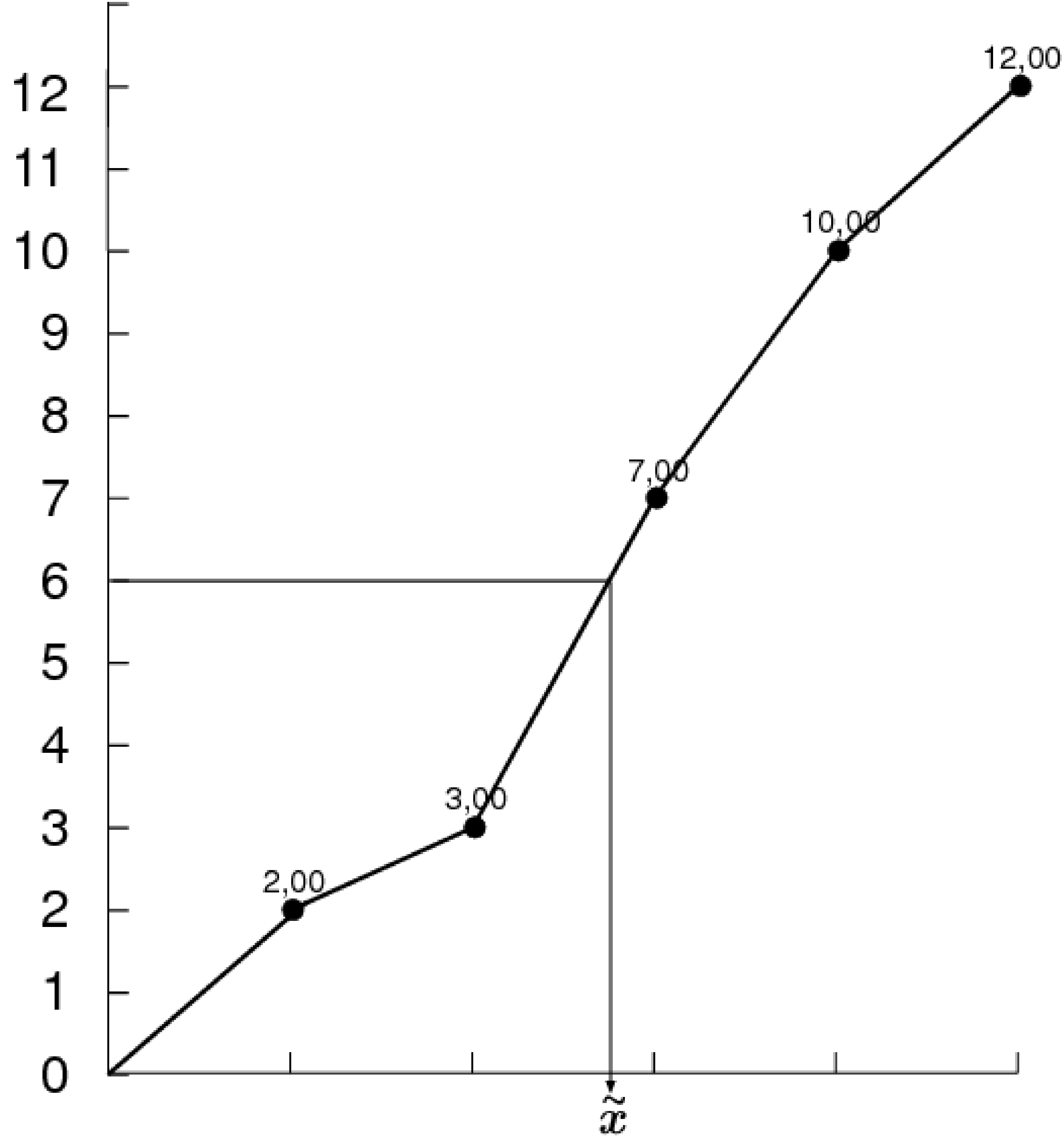

Pdf de $\hat{\theta}_1$
Pdf de $\hat{\theta}_2$

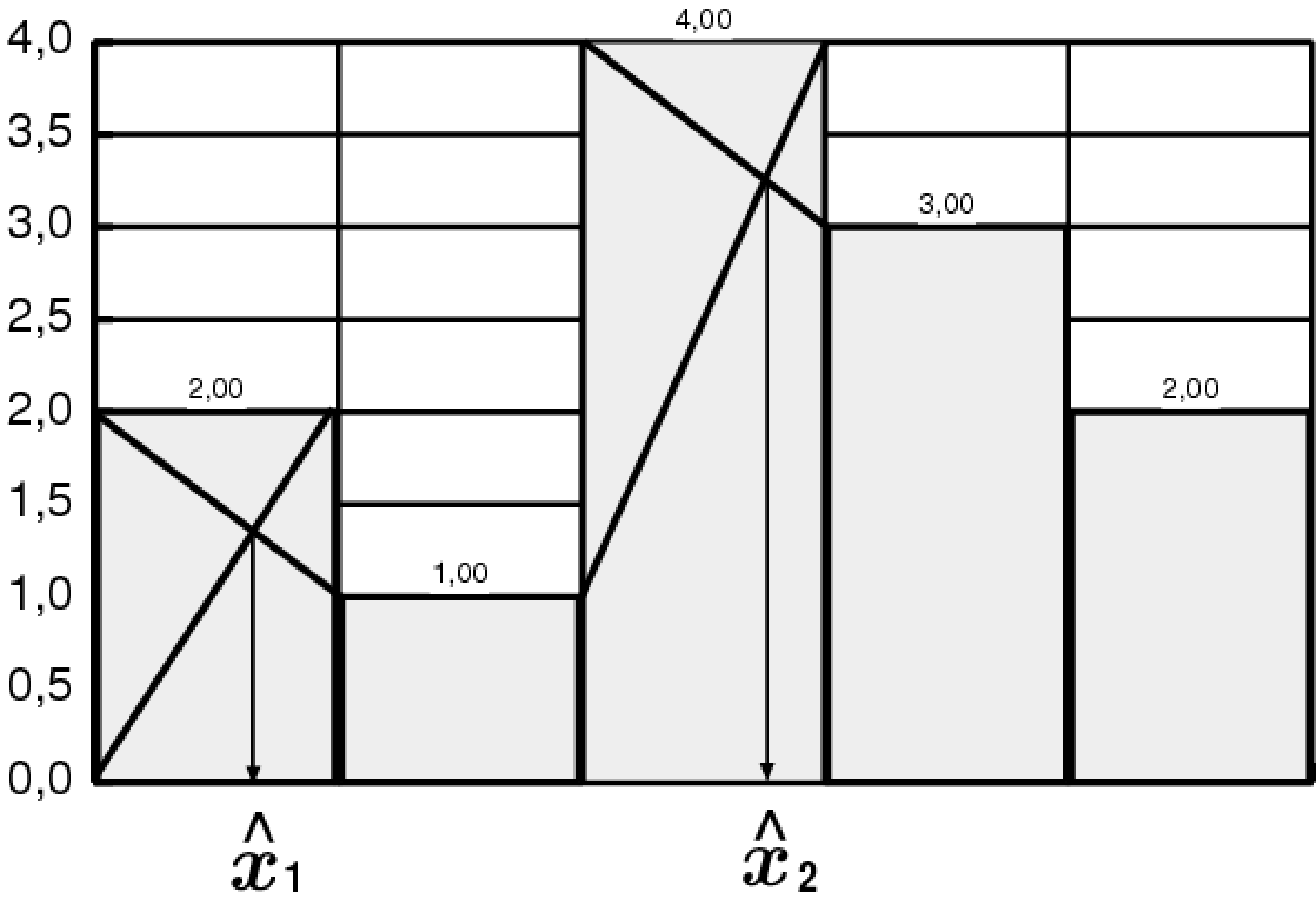


$f_{y|x}(y)$
x
$E(y \mid x)$
y

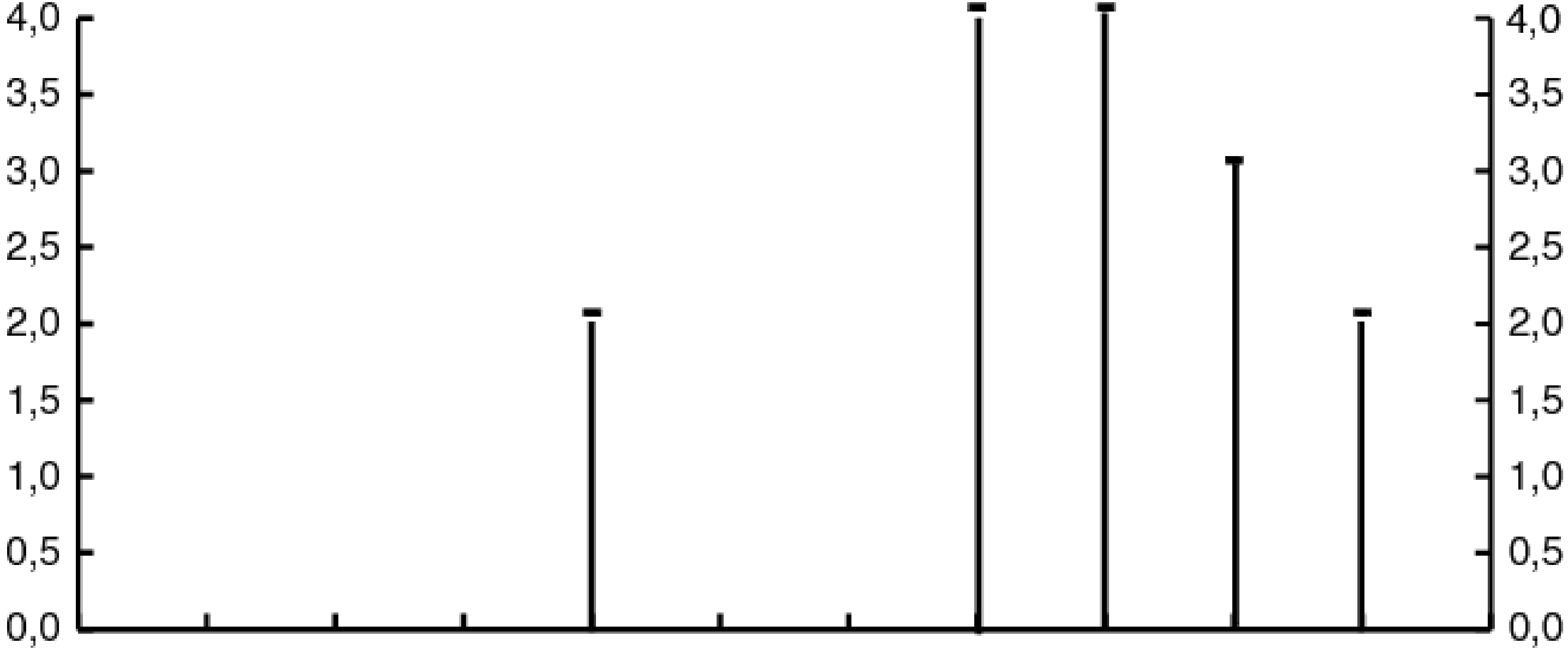
UNIVERSIDAD DEL NORTE

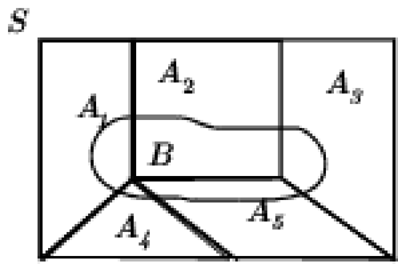
UN


12
11
10
9
8
7
6
5
4
3
2
1
0
2,00
3,00
7,00
10,00
12,00
$\tilde{x}$

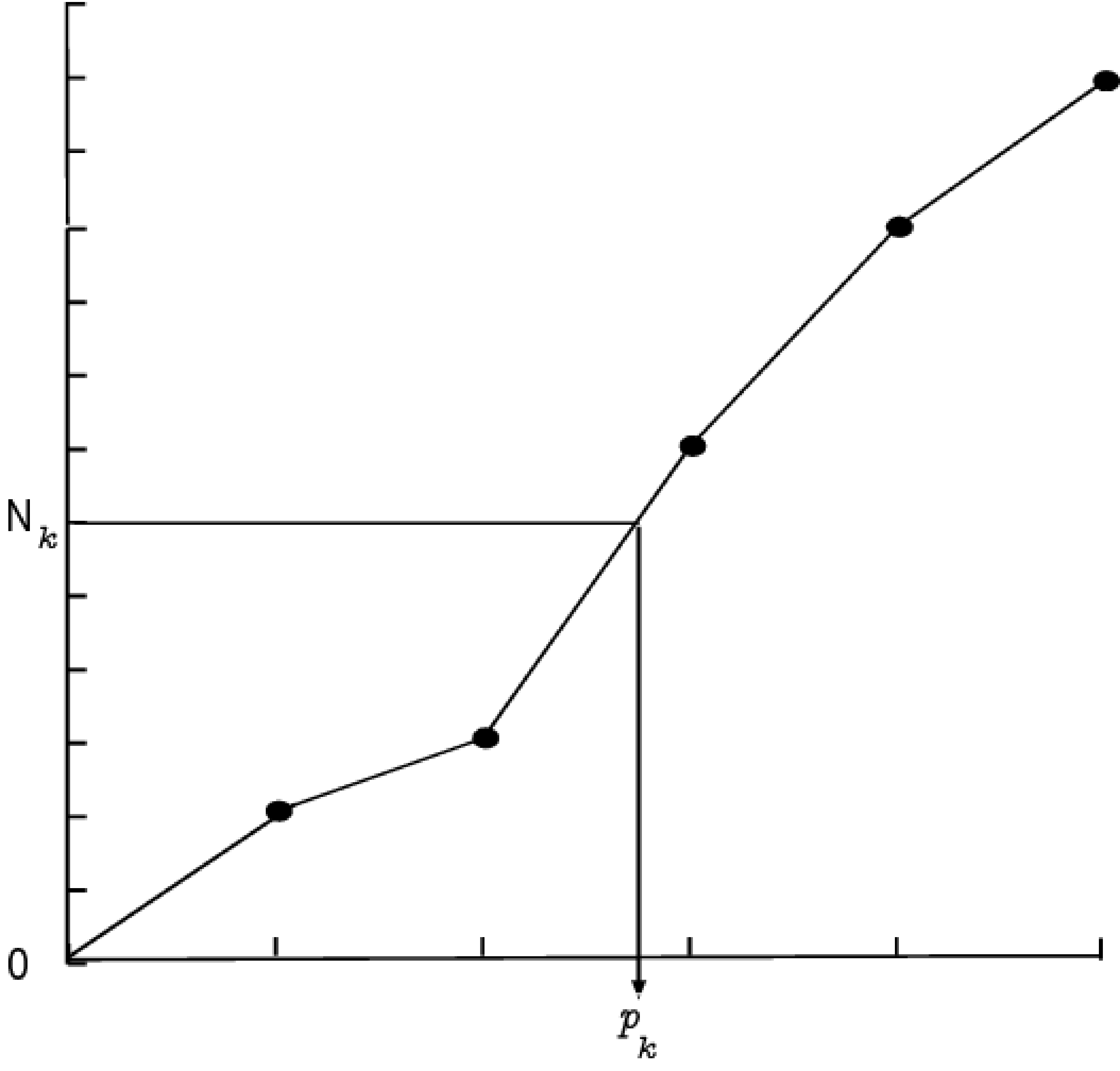


4,0
3,5
3,0
2,5
2,0
1,5
1,0
0,5
0,0
2,00
1,00
4,00
3,00
2,00
$\hat{x}_1$
$\hat{x}_2$


4,0
3,5
3,0
2,5
2,0
1,5
1,0
0,5
0,0



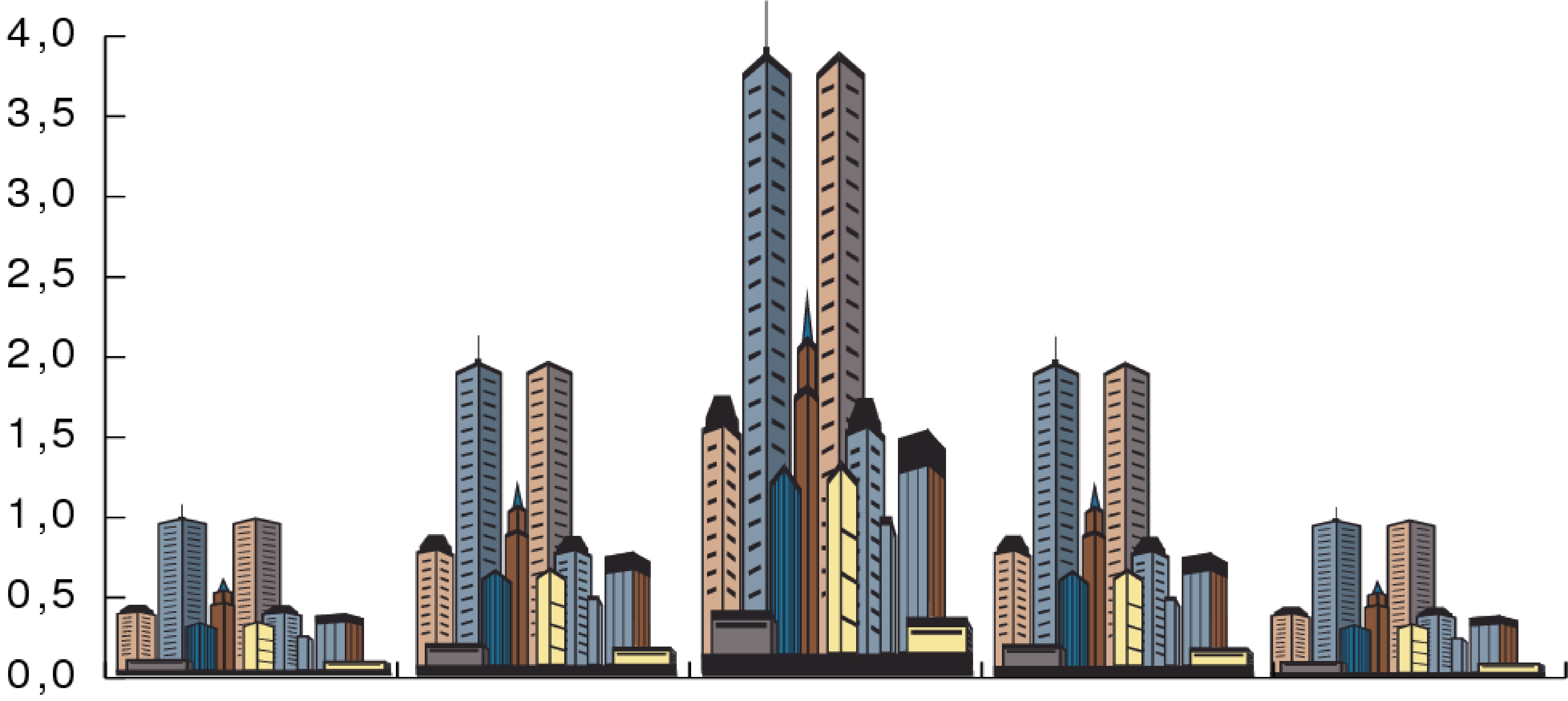
S
A1
A2
A3
B
A4
A5

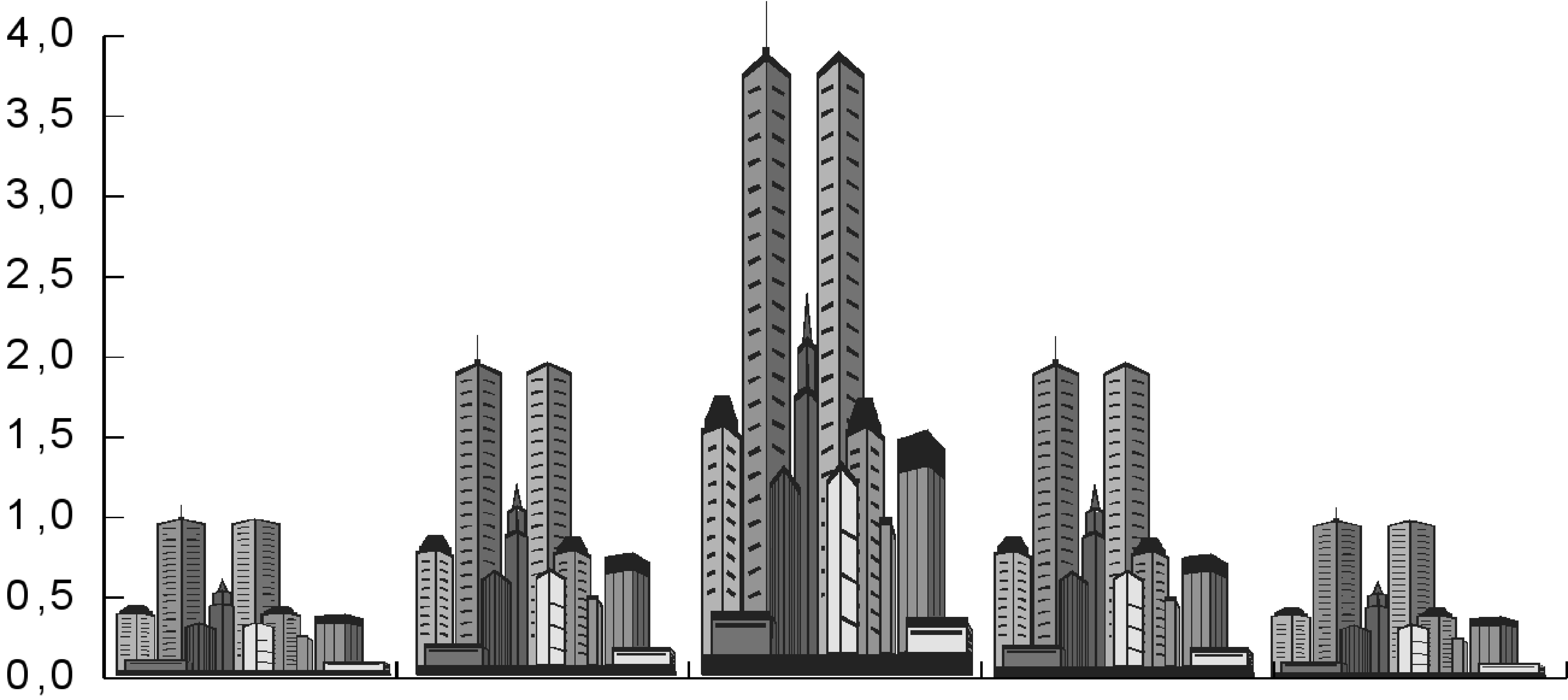


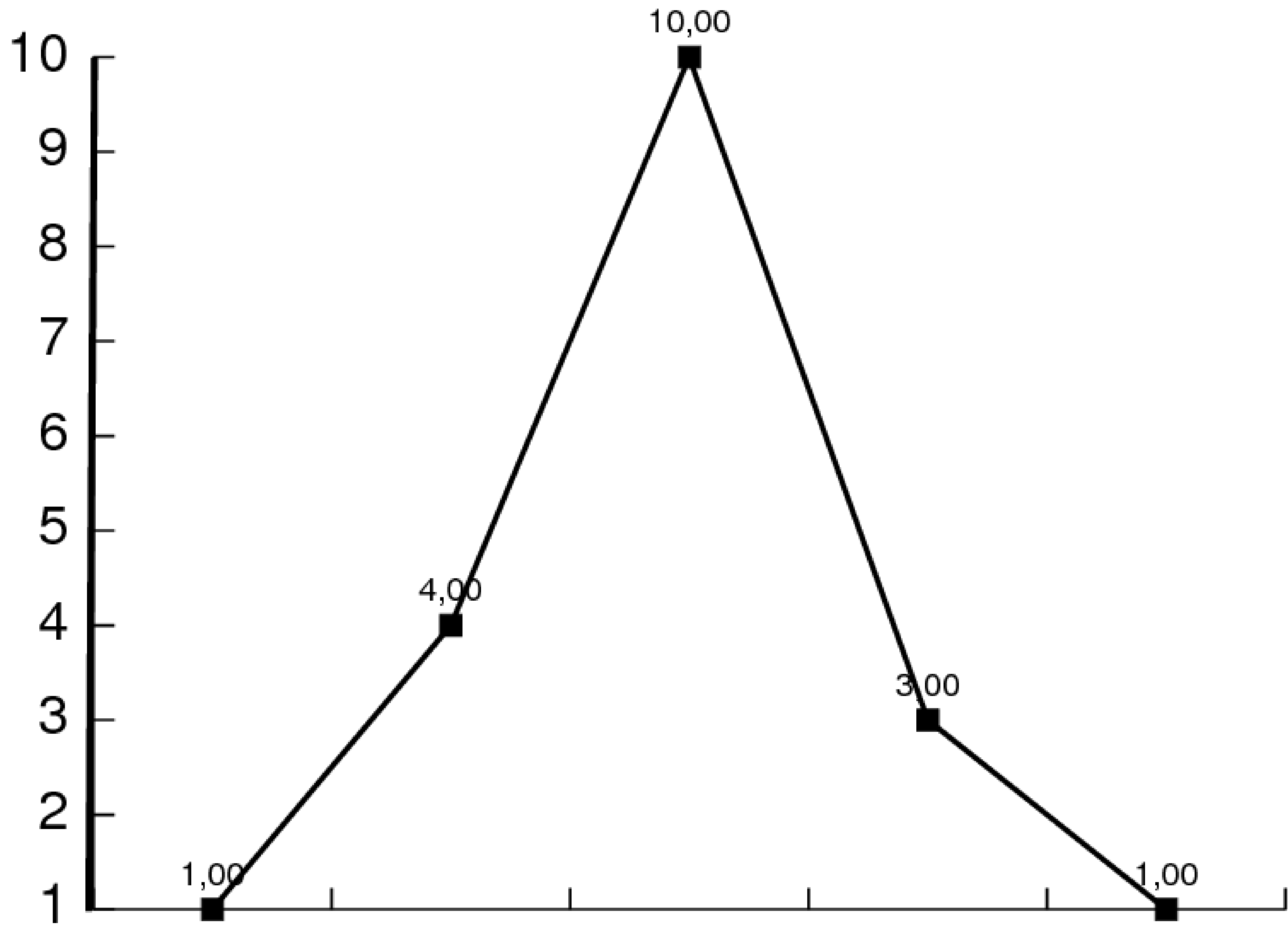
N
k
0
p
k

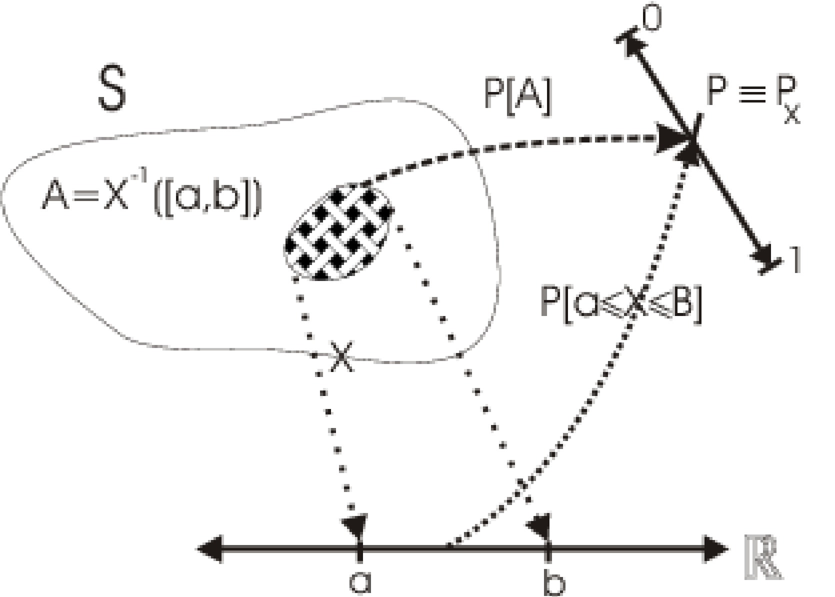
Tablas de Frecuencia para el Peso

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

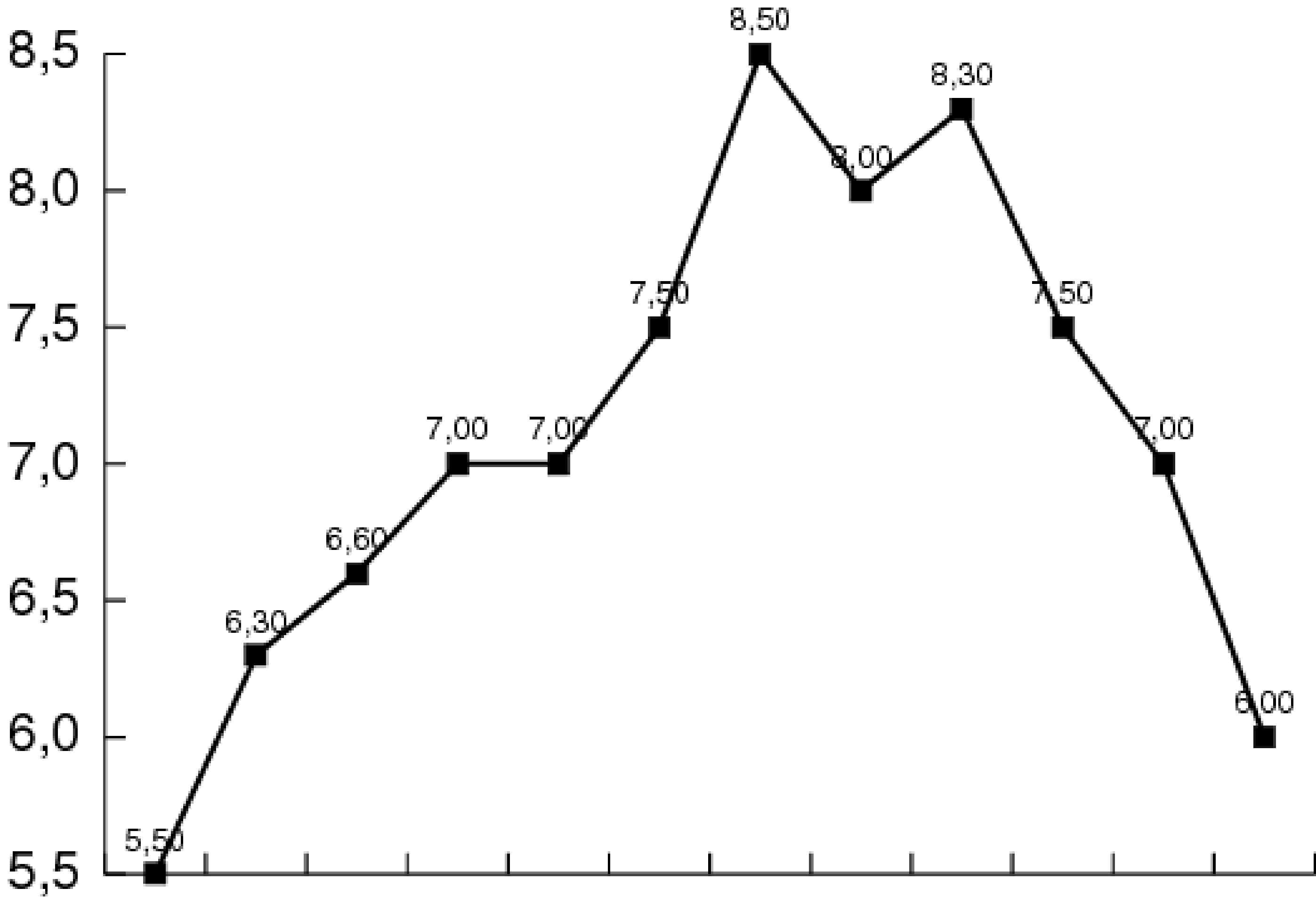

4,0
3,5
3,0
2,5
2,0
1,5
1,0
0,5
0,0


4,0
3,5
3,0
2,5
2,0
1,5
1,0
0,5
0,0

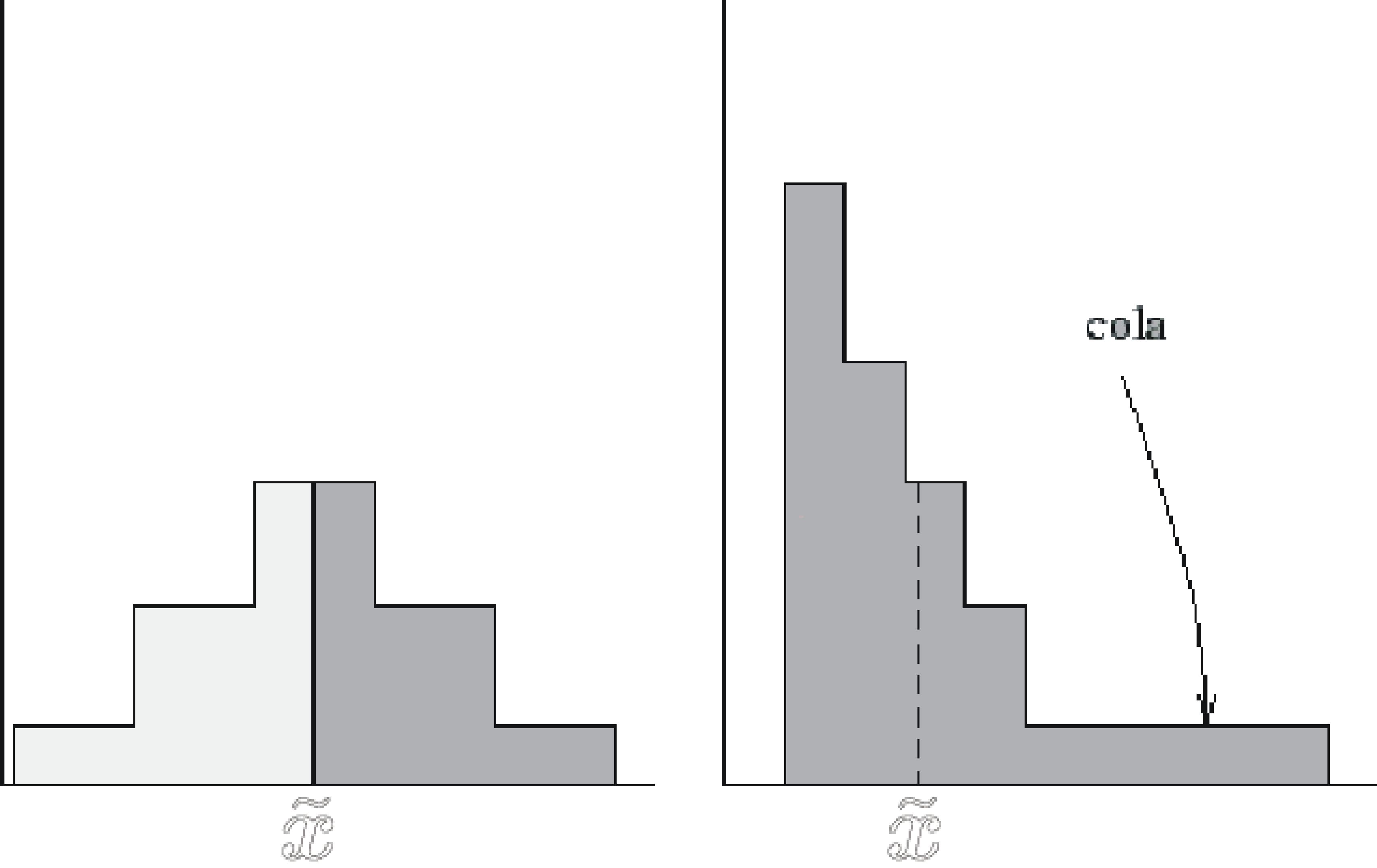

10,00
4,00
3,00
1,00
1,00
10
9
8
7
6
5
4
3
2
1



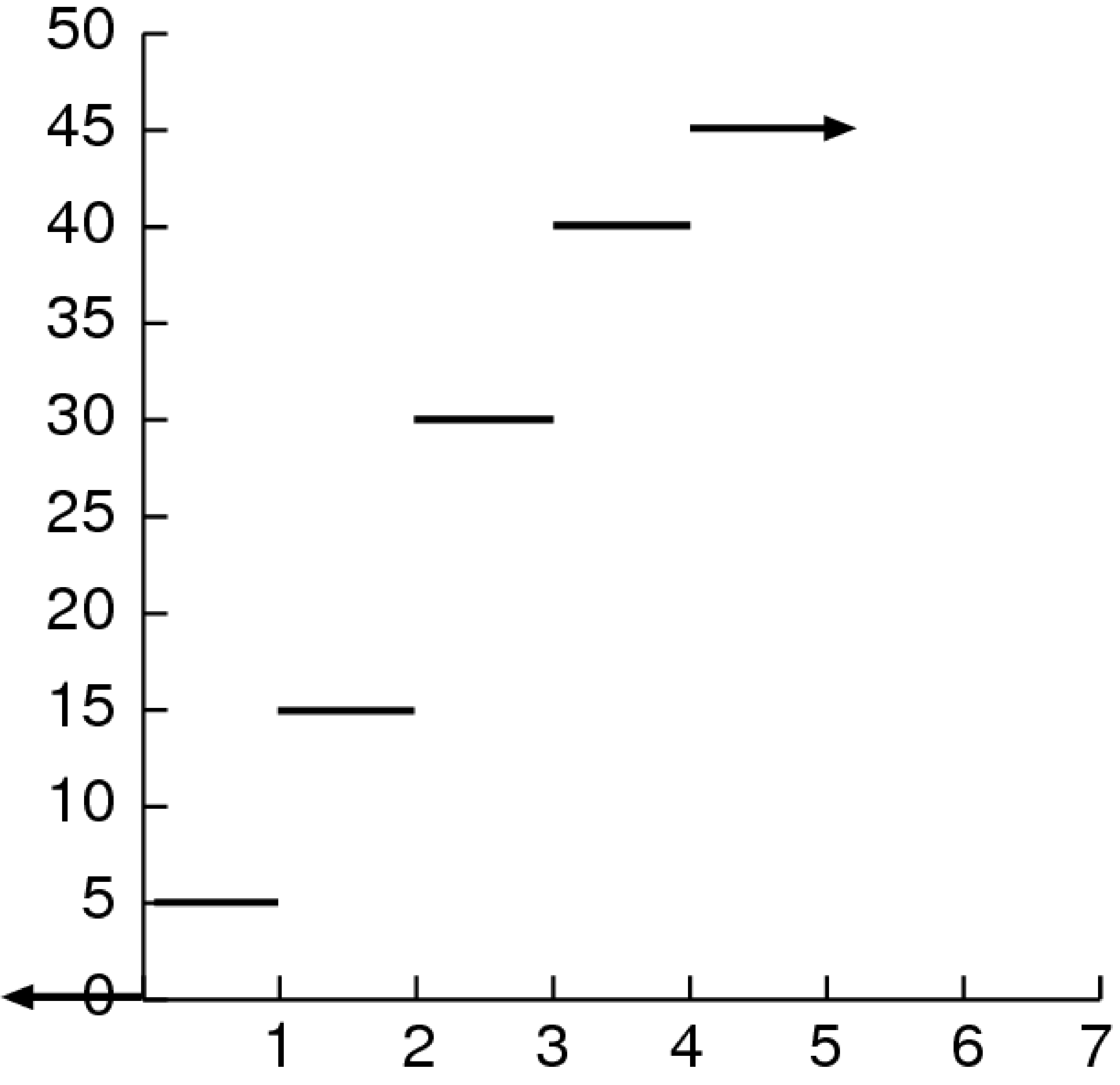
S
P[A]
0
P ≡ P
X
A=X⁻¹([a,b])
1
P[a≤X≤B]
X
ℝ
a
b



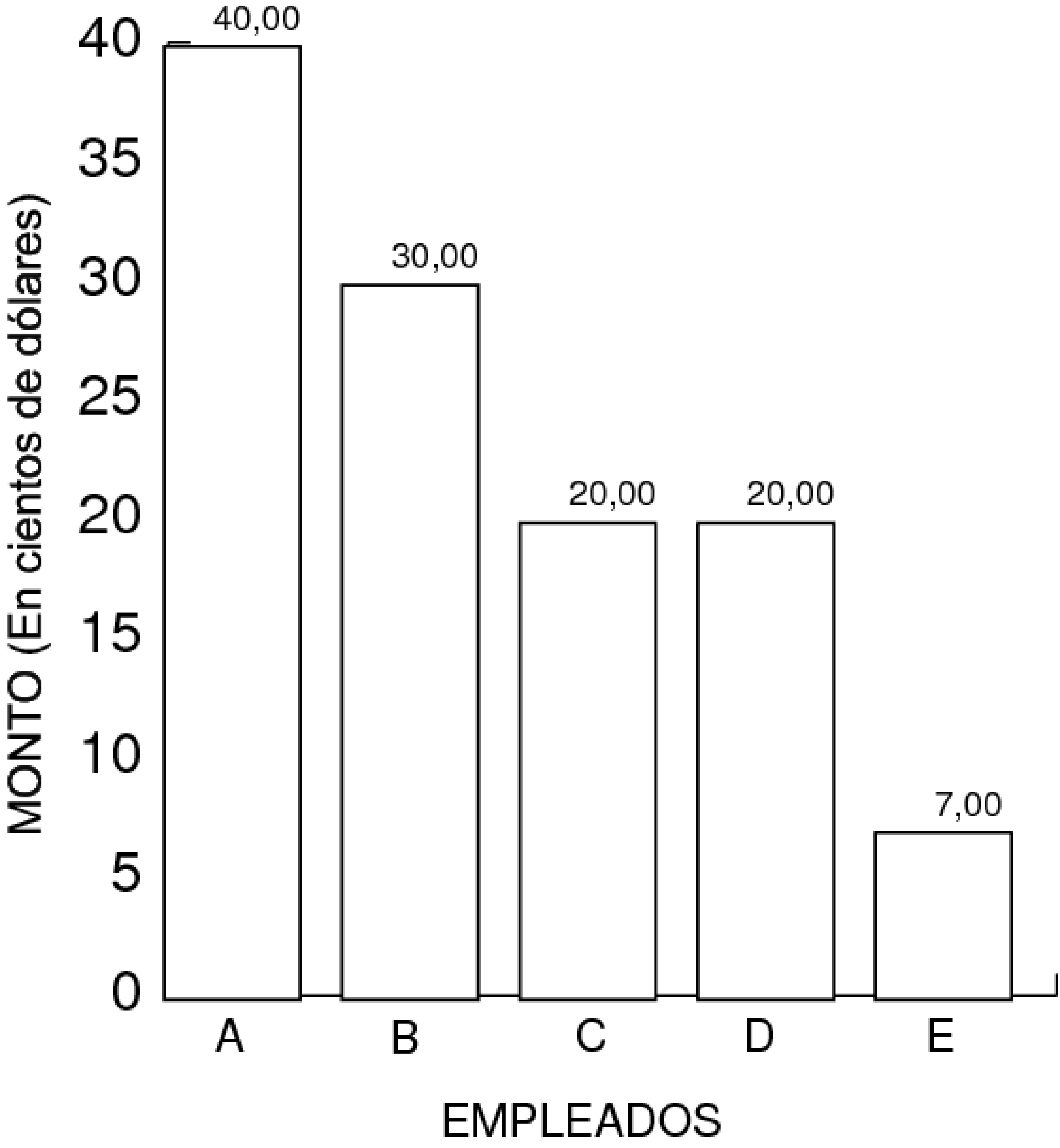
8,5
8,0
7,5
7,0
6,5
6,0
5,5
5,50
6,30
6,60
7,00
7,00
7,50
8,50
8,00
8,30
7,50
7,00
6,00



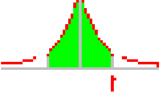
$\tilde{x}$
$\tilde{x}$
cola

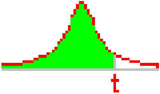


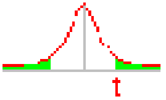
50
45
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7

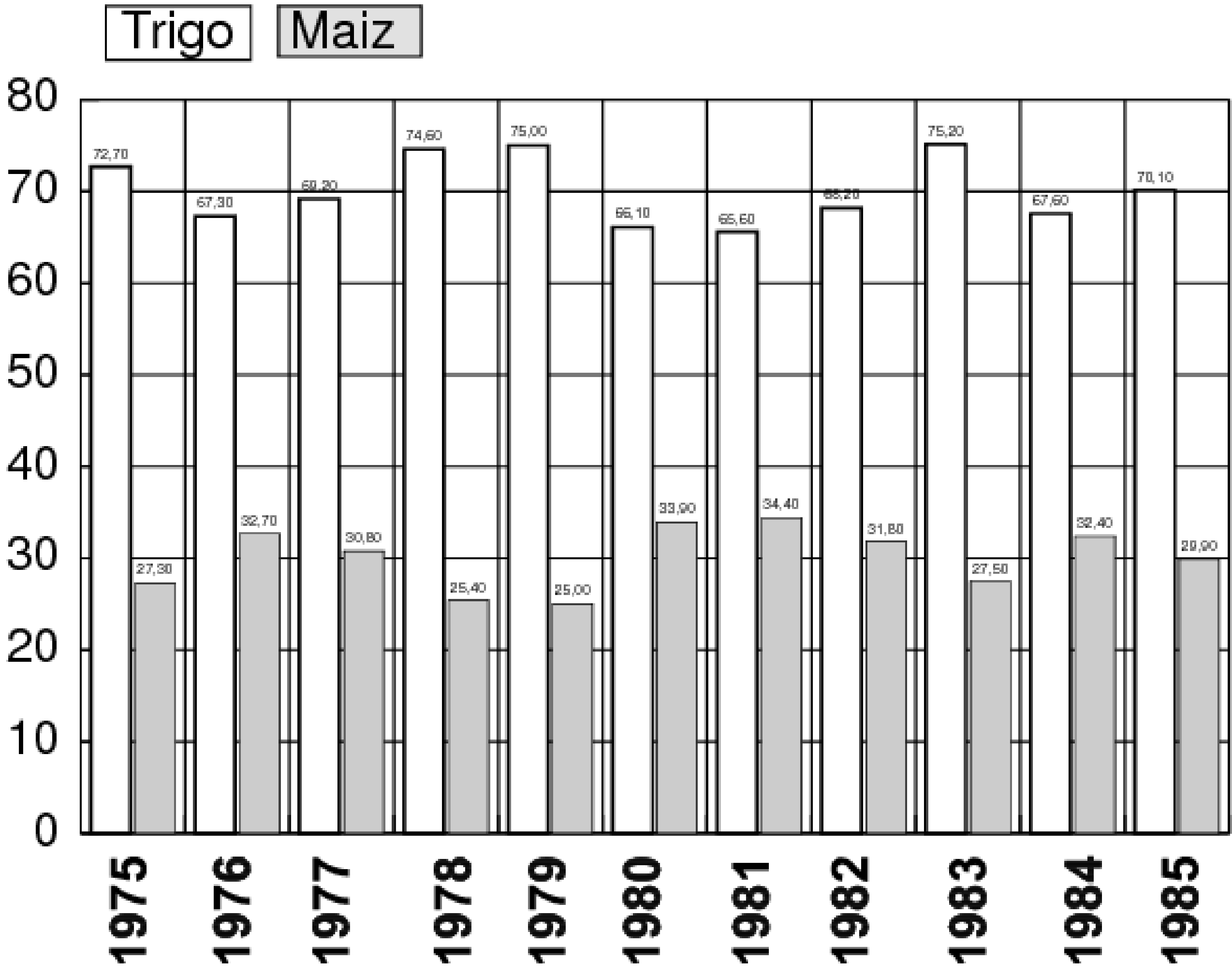


MONTO (En cientos de dólares)
40
35
30
25
20
15
10
5
0
40,00
30,00
20,00
20,00
7,00
A
B
C
D
E
EMPLEADOS

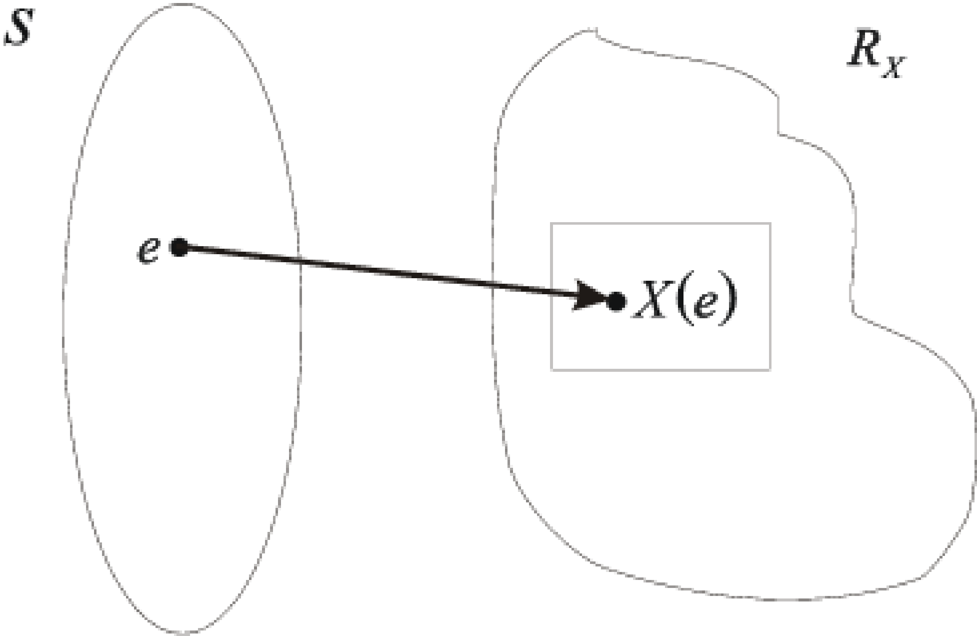


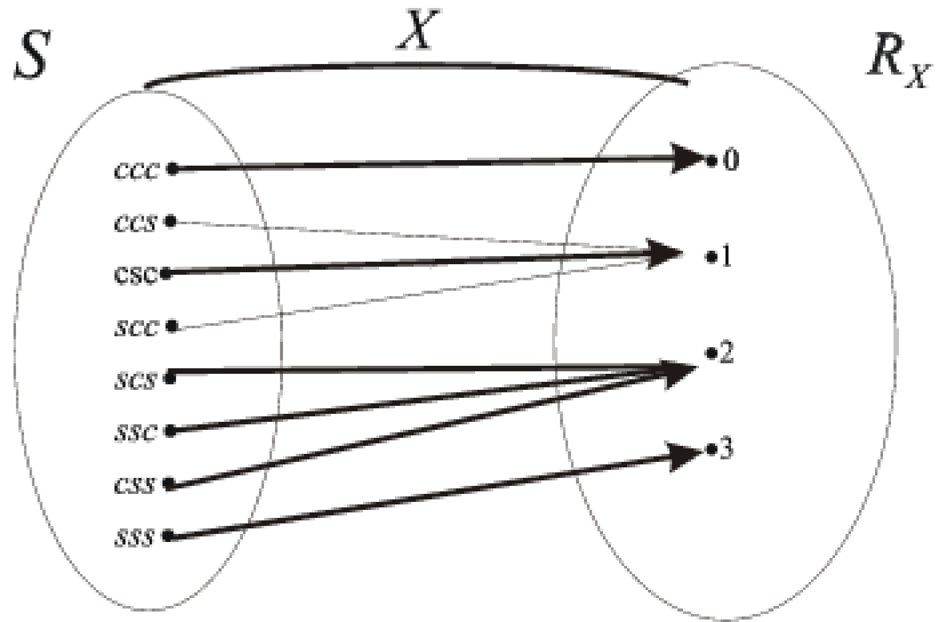

t

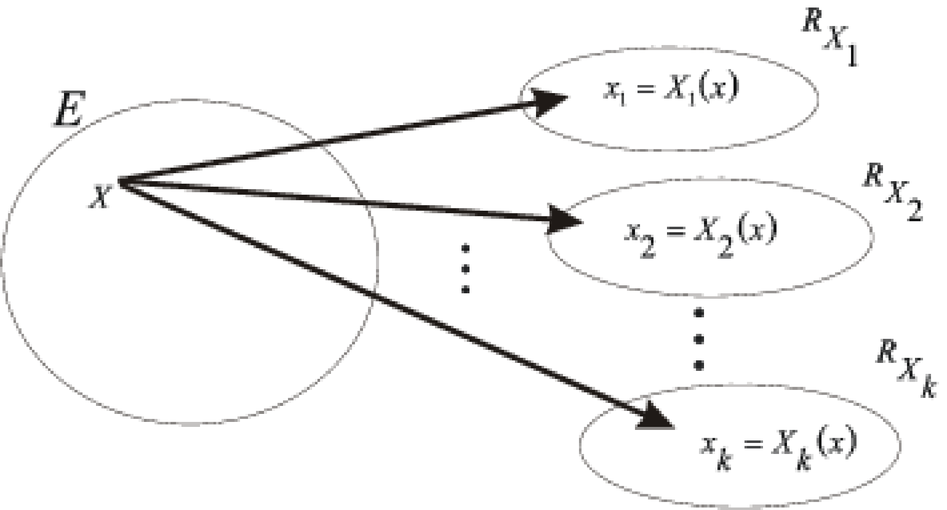

t



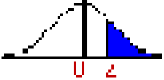
Trigo
Maiz
80
70
60
50
40
30
20
10
0
72,70
27,30
67,30
32,70
69,30
30,80
74,60
25,40
75,00
25,00
66,10
33,90
65,60
34,40
68,20
31,80
75,20
27,50
67,60
32,40
70,10
29,90
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985

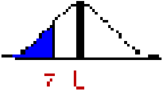

S
R_X
e
$X(e)$

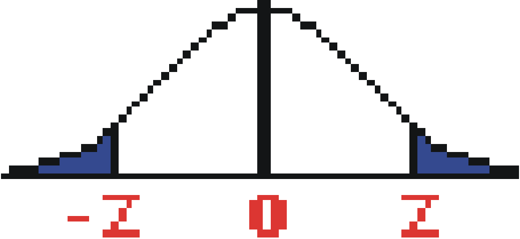

S
X
R_X
CCC
CCS
CSC
SCC
SCS
SSC
CSS
SSS
0
1
2
3

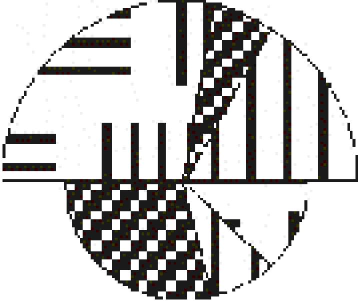


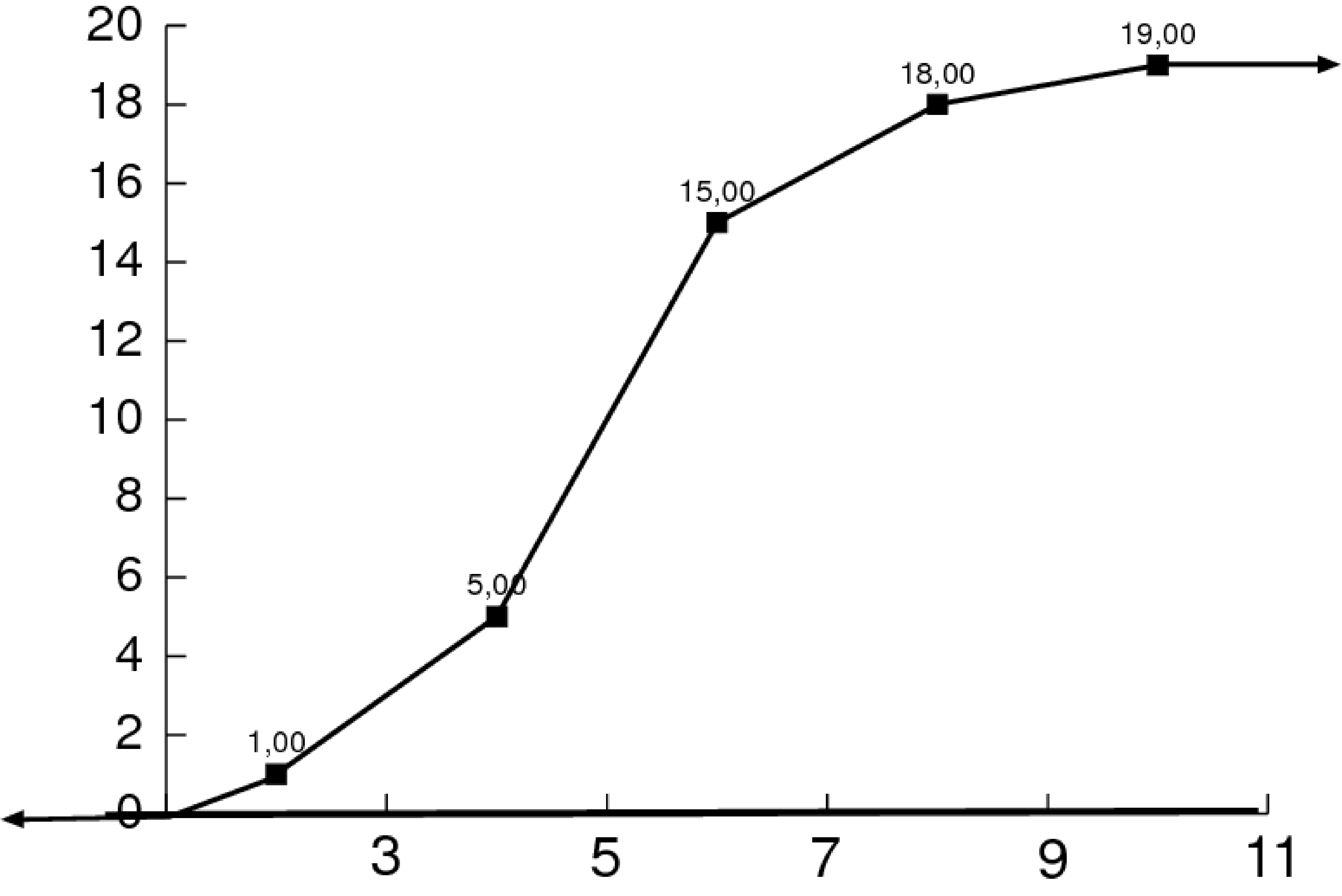
R_{X_1}
$x_1 = X_1(x)$
E
x
R_{X_2}
$x_2 = X_2(x)$
R_{X_k}
$x_k = X_k(x)$




7
L


-Z
0
Z





20
18
16
14
12
10
8
6
4
2
0
3
5
7
9
11
1,00
5,00
15,00
18,00
19,00

1 ANÁLISIS GRÁFICO

El análisis gráfico es un método que nos sirve para determinar la relación matemática entre dos o más variables, (en nuestro caso cantidades físicas) partiendo de una gráfica realizada a partir de datos experimentales tomados al azar.

Este estudio lo vamos presentar en tres partes :
Presentación de la gráfica, escogencia de un modelo matemático y establecimiento de la relación empírica entre las variables.

1.1 Presentación de la gráfica

De acuerdo con el nivel de estas notas sólo presentaremos una introducción del análisis gráfico en dos variables. las cuales representaremos en un plano cartesiano de la siguiente forma:

1. Titulo: El titulo lo colocamos en la parte superior así .
   (nombre de la variable en el eje Y) VS (nombre de la variable en el eje X).

2. En los ejes las letras que representan a las variables y a sus unidades. Se indican las escalas las cuales deben ser múltiplos de 2, 5 o 10

3. Luego se grafican los puntos siguiendo algunos de los siguientes métodos

   (a) Se grafican los puntos y luego se intercala la gráfica entre los puntos, teniendo en cuenta que la gráfica sea suave (es decir que no tenga vértices) y que la misma cantidad de puntos quede a un lado que al otro de la gráfica. por ejemplo

Figura 1

No se deba trazar así:

Figura 2

En las figura 1 observamos que se pueden trazar infinitas gráficas que cumplen las condiciones establecidas.
Lo que nos indica que la gráfica que trazamos no es la gráfica real, ésta sólo nos indica su forma.

(b) Si al tomar los datos podemos fijar cada medida de una variable Por ejemplo: fijamos cada medida x_i de X entonces realizamos varios ensayos para determinar cada valor y_i de la variable Y, obteniendo la media $\bar{y}_i$ para cada medida, luego graficamos para cada medida los puntos $(x_i, \bar{y}_i)$, $(x_i, y_{Máx})$ y (x_i, y_{min}) como indica la figura

Figura 3

Luego con una serie de datos trazamos una gráfica que pase por la mayoría de los puntos medios o que toque todos los segmentos, así:

Figura 4

observemos que los segmentos indican la variabilidad de cada medida, es decir la medida real puede ser cualquier valor sobre el segmento, lo que nos indica que la curva real es una de las infinitas que podemos trazar entre las gráficas punteadas, lo que indica que este método me determina la forma de la curva real.

(c) Si cada medida (x_i, y_i) la obtenemos al realizar varios ensayos, entonces graficamos cada punto $(\bar{x}_i, \bar{y}_i)$ tal que este en la intersección de las diagonales del rectángulo formado por los lados de longitudes $\Delta x_i, \Delta y_i$ como indica la figura 5

Figura 5

Con una serie de datos la curva debe pasar por todos los rectángulos como indica la figura 6.

Figura 6

De la figura observamos lo mismo que en los dos métodos anteriores, que la gráfica que podemos trazar no es única,

1.2 Modelos matemáticos

Conociendo la forma de la curva podemos suponer un modelo matemático, en esta sección presentaremos los cuatro más usados

1. $y = ax^m \;\; a \in \mathbb{R},\; m, a \neq 0,\; m \in \mathbb{Q}^+$

   (a) Se le llama Potencial si tiene la forma

$$m \in \mathbb{Z}^+ \; m \neq 1$$

(b) Radical

$$m \in \mathbb{Q}^+ - \mathbb{Z}^+$$

(c) Inverso

$$m \in \mathbb{Z}^-$$

(d) Lineal

$$m = 1$$

2. Se llama modelo exponencial a $y = ae^{mx}$, $a, m \in I\!R$, $a, m \neq 0$

3. Modelo logarítmico $y = \frac{1}{m} \ln x$ $m \in I\!R$, $m \neq 0$

4. Modelo lineal $y = mx + b$ $m, b \in I\!R$, $m \neq 0$

1.3 Ajuste de la curva

Al observar las gráficas notamos que muchas se parecen y a veces es díficil estar seguro si el modelo que escogemos es adecuado, es más no conocemos todos los modelos. en esta sección explicaremos dos métodos para ayudarnos a escoger un modelo que se aproxime de buena forma a los datos.

1. Linealización :
   Este método consiste en encontrar dos relaciones $h(y)$ y $f(x)$ tales que al graficar $h(y)$ vs $f(x)$ se obtenga una linea recta, es decir si esto sucede el modelo es satisfactorio.
   Por ejemplo si tenemos los datos

$y(cm)$	1	4	9	16	25
$x(s)$	1	2	3	4	5

De acuerdo con la forma el modelo es $y = ax^m$, pero si analizamos los datos ellos nos hacen sospechar que $m = 2$, por lo que podríamos hacer $h(y) = y$ y $f(x) = x^2$ de lo que obtenemos la siguiente tabla

$h(y)$	1	4	9	16	25
$f(x)$	1	4	9	16	25

Como se obtuvo una recta podemos decir que el modelo es aceptable y al determinar la ecuación de la recta nos que al tomar dos puntos cualesquiera

$$a = \frac{y_1 - y_2}{x_1^2 - x_2^2} = \frac{16 - 9}{16 - 9} = 1$$

como la relación matemática es $h(y) = af(x)$, entonces queda $y = x^2$. Este método no es práctico en el sentido de que si no sospechamos nada acerca de la relaciómatemática es casi imposible determinar $h(y)$ y $f(x)$, afortunadamente para los modelos planteados ya existen estas relaciones

(a) Para el modelo $y = ax^m$ son

$$h(y) = \log y, \ f(x) = \log x$$

al obtener la recta

$$\log Y = m \log x + b, a = 10^b$$

podemos encontrar

$$\begin{aligned} m &= \frac{\log y_i - \log y_j}{\log x_i - \log x_j} \\ a &= \frac{y_i}{x_i^m}. \end{aligned}$$

En la práctica debido a los errores al tomar diferentes parejas (x_i, y_i) los valores de m y a no son constantes, por lo que hay que determinarlos varias veces y promediar los resultados, así la relación matemática queda $y = \bar{a}x^{\bar{m}}$

(b) Para el modelo $y = ae^{mx}$ se escogen las relaciones

$$h(y) = \ln y, \;\; f(x) = x$$

y si obtenemos la recta

$$\ln Y = mX + b;\ a = e^{b}$$

podemos utilizar

$$\begin{aligned} m &= \frac{\ln y_i - \ln y_j}{x_i - x_j} \\ a &= \frac{y_i}{e^{mx_i}}. \end{aligned}$$

y calculamos varios valores de m y a para promediarlos y obtener la relación matemática $y = \bar{a}e^{\bar{m}x}$

(c) Si el modelo es $y = \frac{1}{m}\ln x$ las relaciones son

$$h(y) = y,\ f(x) = \ln x$$

si la gráfica $f(x)\ \ vs\ h(y)$ es lineal como se muestra en la figura

podemos usar

$$m = \frac{\ln x_i - \ln x_j}{y_i - y_j}$$

calculamos m varias veces y obtenemos $y = \frac{1}{m}\ln x$.

(d) El modelo lineal es trivial por lo que dejamos de tarea al lector

2. Ajuste lineal

El ajuste lineal es un método estadístico, parecido al anterior, pero usando herramientas diferentes.

En este caso no explicaremos mucho , sólo plantearemos algunas fórmulas e interpretaremos los resultados

Si tenemos dos conjuntos de datos

$$x_1, x_2, x_3, \cdots, x_n$$
$$y_1, y_2, y_3, \cdots, y_n$$

definiremos la

(a) Media aritmética

$$\begin{aligned} \bar{x} &= \frac{\sum_{i=1}^{n} x_i}{n} \\ \bar{y} &= \frac{\sum_{i=1}^{n} y_i}{n} \end{aligned}$$

(b) Varianza

$$\begin{aligned} S_{xx}^2 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \\ S_{yy}^2 &= \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} \end{aligned}$$

(c) Covarianza

$$S_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

(d) Coeficiente de correlación

$$r = \frac{S_{xy}}{\sqrt{S_{xx}^2 S_{yy}^2}}$$

Cuando estudiamos dos variables X y Y, en realidad estas variables estas variables tienen desde el punto de vista geométrico el

comportamiento de dos vectores (rayos con dirección). Y se dice que existe un ajuste lineal si los vectores representados por las desviaciones están alineados o son paralelos.
Para determinar si esto sucede podemos utilizar la trigonometría para determinar el ángulo entre los vectores representados en la figura

Utilizando una rama de las matemáticas llamada algebra lineal se puede comprobar que si θ es el ángulo entre los vectores entonces

$$\cos\theta = r$$

Es decir este coeficiente determina la relación lineal entre las desviaciones de las variables y esta dependencia es perfecta cuando $r = \pm 1$ existe el ajuste lineal en la practica esto no sucede debido a los errores pero decimos que el ajuste es aceptable si $r^2 > 0.95$.
Si el ajuste existe deben existir dos relaciones $h(y)$ y $f(x)$ tal que el modelo

$$h(y) = mf(x) + b$$

debe tener un r adecuado y se puede calcular

$$r = \frac{n\sum_{i,j=1}^{n} f(x_i)\,h(y_j) - \sum_{i=1}^{n} f(x_i)\sum_{j=1}^{n} h(y_i)}{\sqrt{\left[n\sum_{i=1}^{n} f^2(x_i) - \left(\sum_{i=1}^{2} f(x_i)^2\right)\right]}\left[n\sum_{j=1}^{n} h^2(y_j) - \left(\sum_{j=1}^{2} h(y_j)^2\right)\right]}$$

$$\begin{aligned} m &= \frac{n\sum_{i,j=1}^{n} f(x_i)\,h(y_j) - \sum_{i=1}^{n} f(x_i)\sum_{j=1}^{n} h(y_i)}{n\sum_{i=1}^{n} f^2(x_i) - \left(\sum_{i=1}^{2} f(x_i)^2\right)} \\ b &= \frac{\sum_{j=1}^{n} h(y_j) - m\sum_{i=1}^{n} f(x_i)}{n} \end{aligned}$$

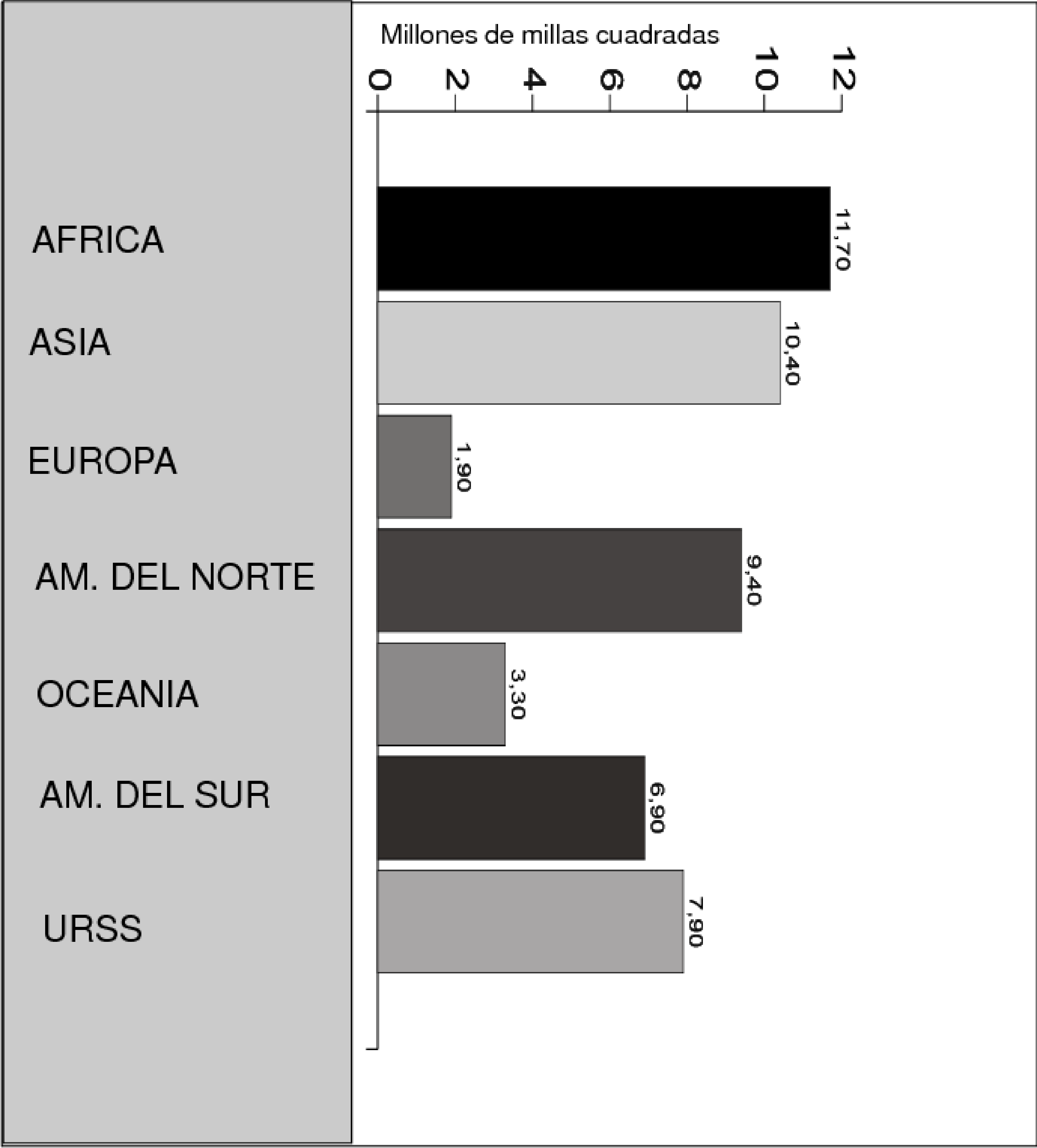
donde a se obtiene

Figure 1: fig a

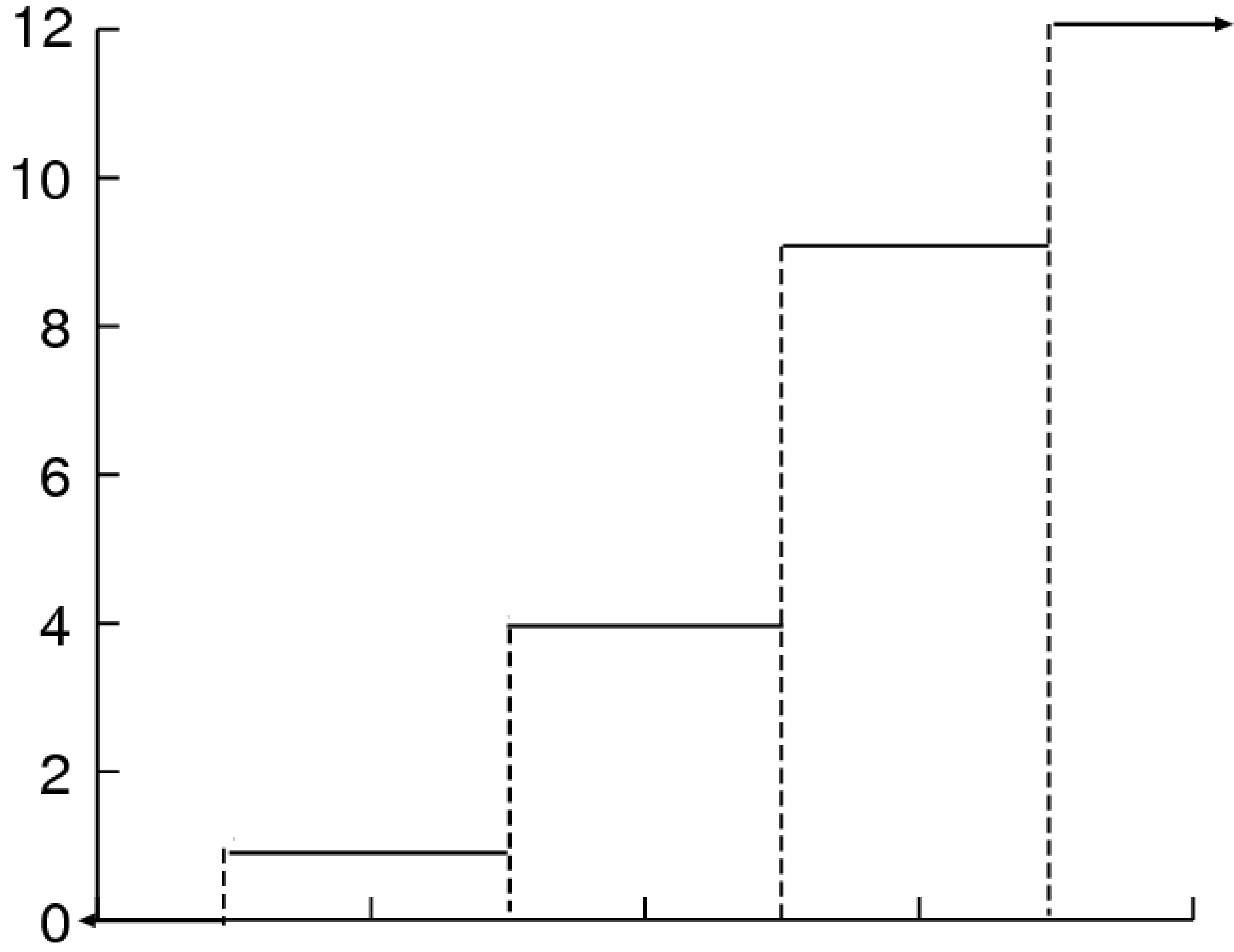
- $a = 10^b$ en el modelo $y = ax^m$
- $a = e^b$ en el modelo $y = ae^{mx}$

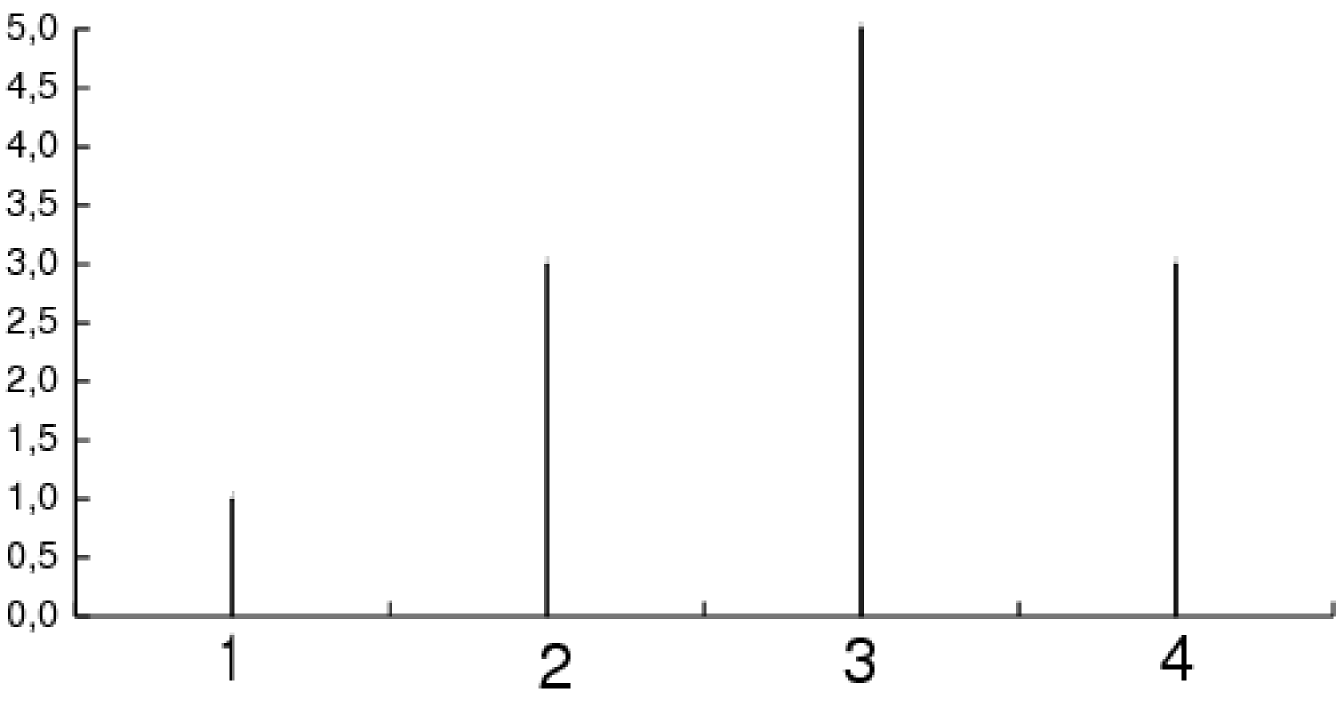
1.4 Interpretación geométrica de la covarianza

Si consideramos una nube de puntos formados por las parejas de los datos concretos de dos variables X e Y (x_i, y_i) el centro de gravedad de esta nube de puntos es $\left(\bar{x}, \bar{y}\right)$, ahora si trasladamos los ejes de tal forma que este punto sea el centro, la nube queda dividida en cuatro cuadrantes los que indica que los puntos que se encuentran en el pimer y tercer cuadrate contribuyen positivamente al valor de la covarianza y los que se encuentran en los otros dos cuadrantes contribuyen negativamente. como lo indica la figura.a y si los puntos se reparten con igual proporción la covarianza será negativa como indica la segunda gráfica de la figura ay no hay relación matemática si la nube de puntos no tiene ninguna tendencia como en la fig b fig b



Millones de millas cuadradas
0
2
4
6
8
10
12
AFRICA
11,70
ASIA
10,40
EUROPA
1,90
AM. DEL NORTE
9,40
OCEANIA
3,30
AM. DEL SUR
6,90
URSS
7,90


12
10
8
6
4
2
0


5,0
4,5
4,0
3,5
3,0
2,5
2,0
1,5
1,0
0,5
0,0
1
2
3
4

INSTITUTO TECNICO DE ARTES Y OFICIOS CARL ROS
SECCION BACHILLERATO
MODULO DE QUIMICA

1. Realizar las siguientes conversiones:

   (a) 53^0F en ^0C
   (b) 36^0C a ^0F
   (c) -17^0C en ^0K
   (d) 140^0F en ^0K
   (e) 71^0F en ^0C y ^0K
   (f) 25^0C ^{0} a F y ^0K
   (g) 280^0C a ^0F
   (h) 0^0C en ^0K y ^0F

2. Reasolver los siguientes ejercicios aplicando las diferentes leyes de los gases

   (a) Un globo lleno de Helio, presenta un volumen de 6,5 litros a una temperatura de 20^0C cuando la presión Es de 1 atmósfera.Si se mantiene la presión constante, qué volumen presentará el globo cuando se desciende la temperatura hasta -15^{0}C?
   (b) Una muestra de 6,5 litros de gas a presión de 1000 torr experimenta un cambio de presión hasta 720 torr, Manteniendo la temperatura constante? Cuál es el nuevo volumen ocupado por el gas?
   (c) Sí 1,9 litros de un gas a ^{0}C se transfiere a un recipiente de 1375 ml a presión constante.Cuál será su temperatura final? Una muestra de un gas ideal ocupa un volumen de 69,3 mililitros a 925 torr y 18^{0}C. Qué volumen. Ocupará el gas a 120^{0}C y 720 torr?
   (d) Si tenemos 5 gr de un gas ideal en un recipiente de 9,6 lts a 0^0C y calentamos el gas a 130^0C ¿Cuál será el nuevo volumen del gas?
   (e) Una muestra de cloro gaseoso ocupa un volumen de 3.8 lts a una presión de 760 ml ¿Cuál será el volumen de la muestra? con una persión de 570 mm de mercurio

Chapter 1

Conceptos previos

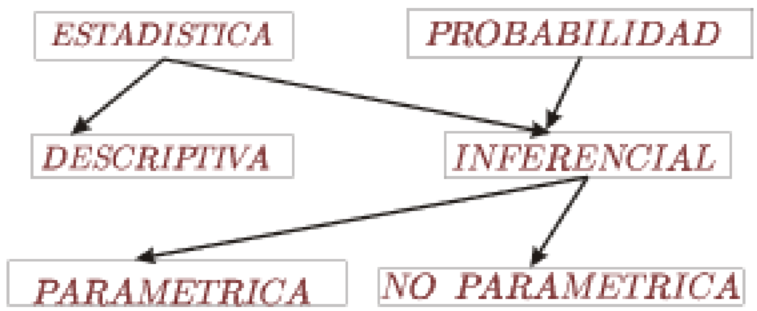
1.1 Introducción

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 numéros se habla de los censos realizados por los levitas, lo mismo que los romanos lo cual lo hacáan 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 historico 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 intrínsicamente por ellos y después ralizar 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.

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

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

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

Muestra: Es un subconjunto representativo de una población.

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

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.

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

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:

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

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

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

{1^o,2^o,3^o,...}, [alto, medio, bajo]

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

Discretas: Cuando las variables toman valores puntuales

Cuando el rango es

Continuas: si X toma todos los valores en un intervalo

si

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:

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.

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

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,

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

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

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

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

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

- $$n = \sum_{j=1}^{k} n_j \; , \; \sum_{j=1}^{k} f_j$$

Distribución de frecuencias: Es una función que a cada clase c_i de C (el conjunto de todas las clases) le asigna un valor es decir si definimos la función

$$\mathfrak{F} : Cac_i \longmapsto n_i$$

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

Tabla de frecuencia: Es una representación de $\mathfrak{F}$ y generalmente es de la

siguiente forma:

M	F.	F.R.	F.A.A.	F.R.A.
C	n_i	f_i	N_i	F_i
c_1	n_1	$f_1=\frac{n_1}{n}$	$N_1=n_1$	$F_1=\frac{N_1}{n}=f_1$
c_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	$f_k=\frac{n_k}{n}$	$N_k=n$	$F_k=1$
	n	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.

[1] 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

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 siguiente tabla:

Nº de caras	n_i	f_i	N_i	F_i
0	38			
1	144			
2	342			
3	287			
4	164			
5	25			
	1000			

a .Completar la tabla

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

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 quí 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 $x_{i-1} - x_i = I_i$.

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

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.

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

Elección de clases para variables continuas

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

- Lo primero es determinar k, entre mayor sea su valor mejor[2], 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[3]

$$\begin{aligned} x_0 &= x_{\min} \\ x_1 &= x_0 + a \\ x_2 &= x_1 + 2a \\ x_k &= x_{\max} + ka \end{aligned}$$

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

$$x_k \geq x_{\max} > x_{\min} \geq x_0$$

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

[3]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.

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

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

Determinar una tabla de frecuencias:

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 fómula $k = 1 + 3.22 * \log_{10} 21 = 5.257\,5 \cong 6$ aunque podríamos escoger también $k = \sqrt{21} \cong 5$.

Ahora hallemos $R = 73 - 40 = 33 \Longrightarrow a_i = a = \frac{33}{6} = 5.5$

$x_0 = x_{\min} = 40$

$x_5 = x_{\max} = 73$

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 $a\prime = 7 \Longrightarrow R\prime = a\prime \cdot 5 = 35$

$d = R\prime - R = 2 \Longrightarrow x_0 = x_{\min} - \frac{d}{2} = 39, \quad x_5 = x_{\max} + \frac{d}{2} = 74$

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 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
--------------------------------------------------------------------------------------------
```

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

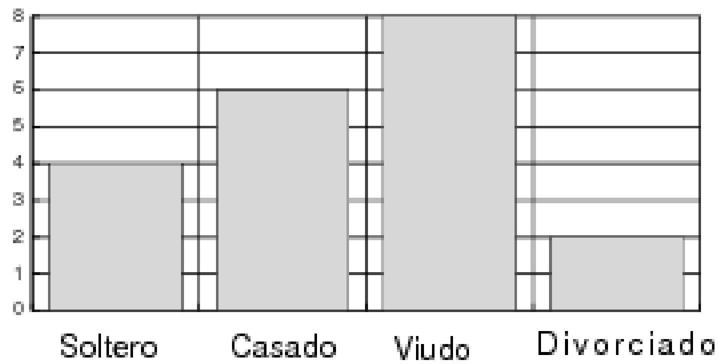


Diagrama de barra para una variable cualitativa

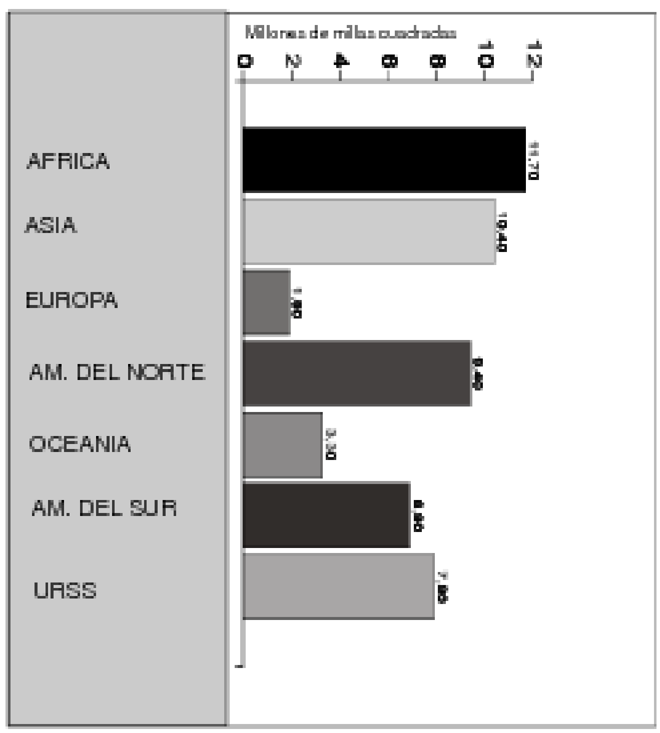
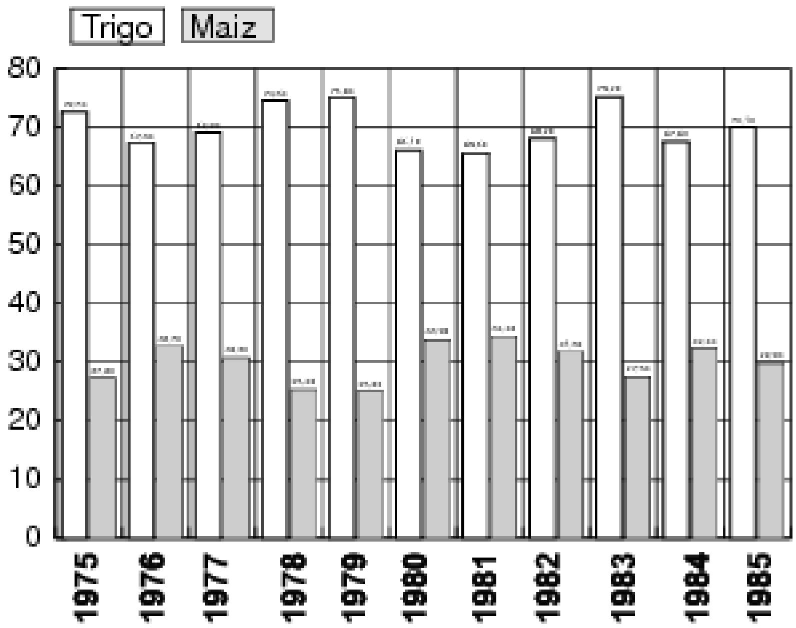


Figure 1.1: AREAS DE LOS CONTINENTES



[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:

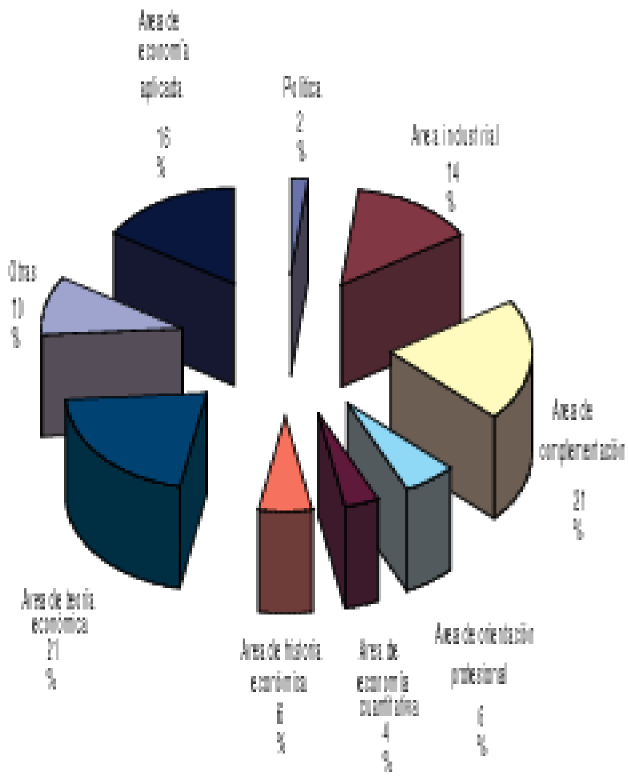
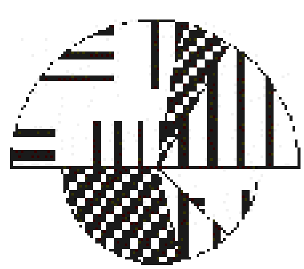


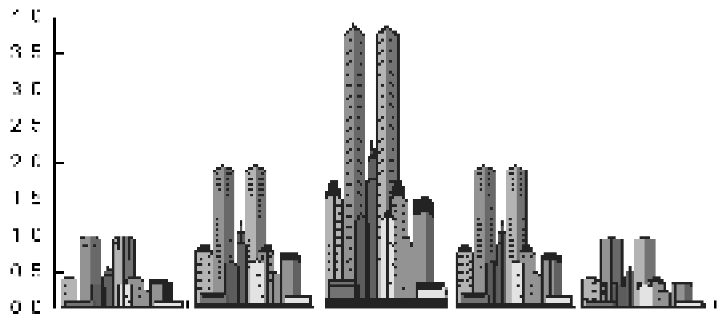
Diagrama circular

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.



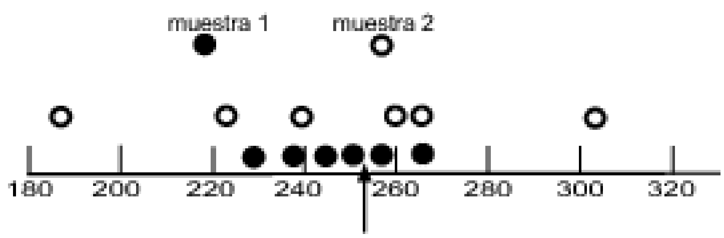
DOS MUESTRAS

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

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 varible 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

La siguiente tabla representa el numéro 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		

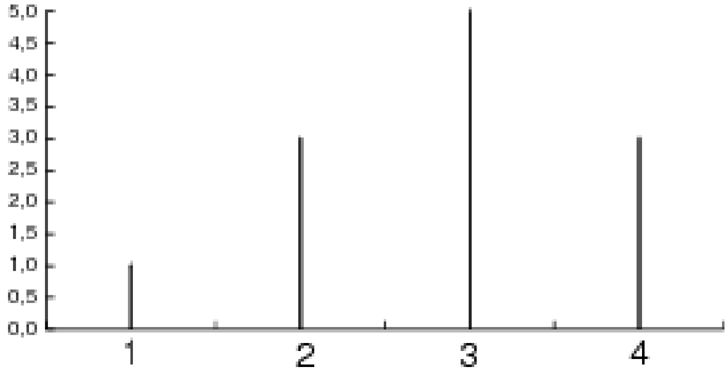
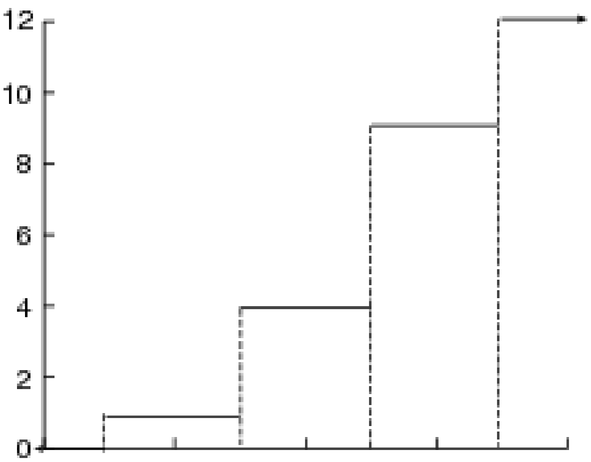


Figure 1.2: FRECUENCIAS ABSOLUTAS



FRECUENCIAS ABSOLUTAS ACUMULADAS

Para las variables continuas también se pueden representar con diagramas circulares

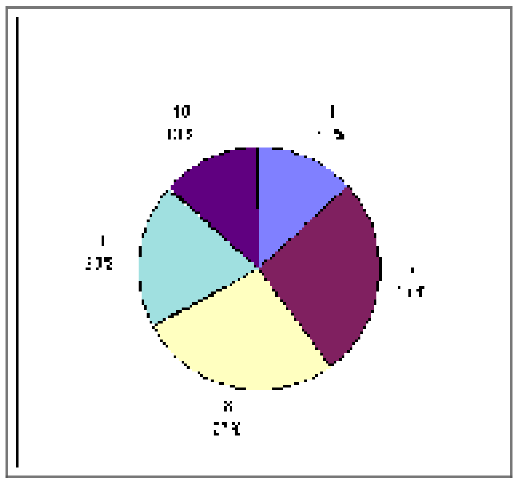
La tabla representa las notas obtenidas por los alumnos de 11^o en un examen de Matemáticas en 1990

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

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	

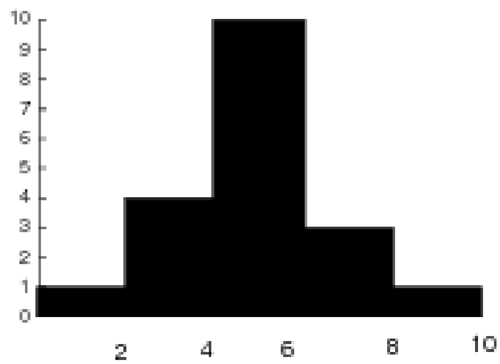
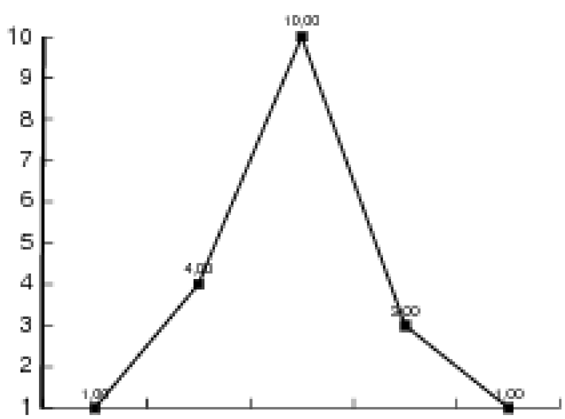


Diagrama Diferencial



Polígono de frecuencias

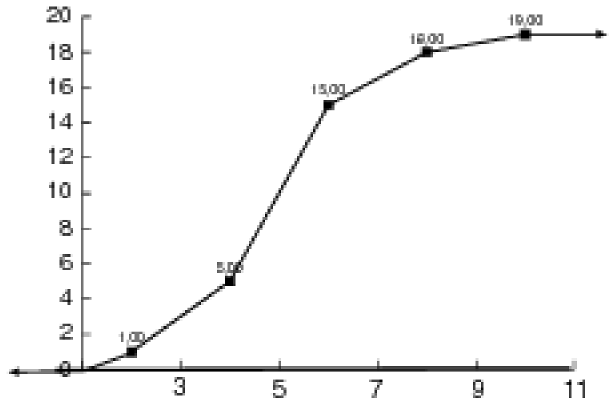


Diagrama acumulado

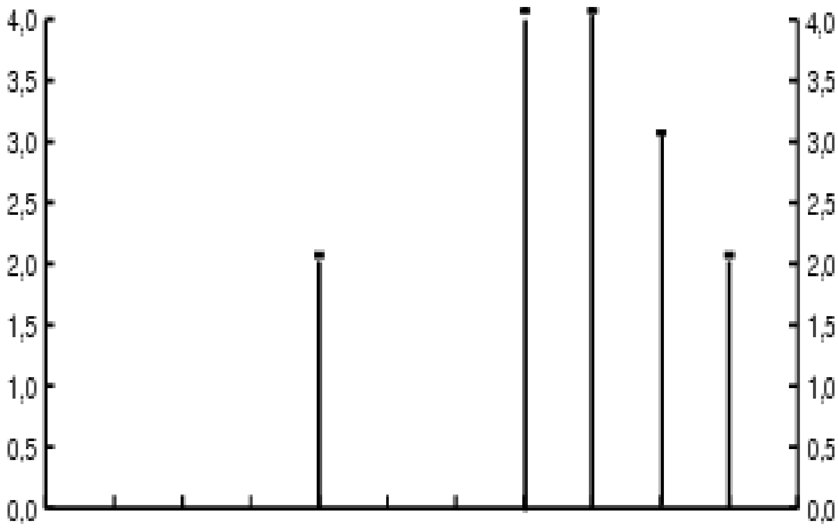
Representar gráficamente los datos de la tabla1

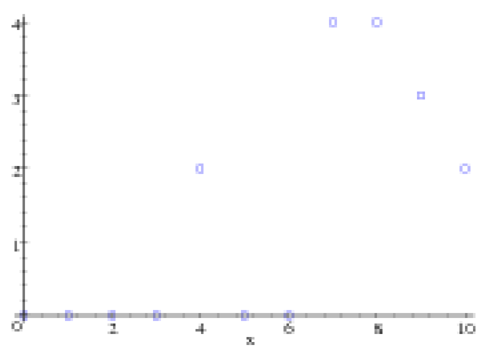
Notas	**Frecuencia**
10	2
9	3
8	4
7	4
6	0
5	0
4	2
3	0
2	0
1	0
0	0

(Tabla 1) Table 1

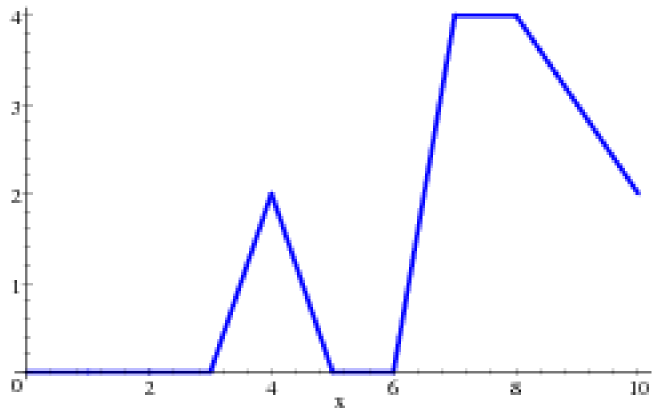
Intervalos	$\mathbf{n}_i$	f_i	$f_i\%$	$\mathbf{N}_i$	$\mathbf{F}_i$
$9-10$	5	$5/15 = .333$	33.3	15	1.000
$7-8$	8	$8/15 = .534$	53.4	10	.667
$5-6$	0	$0/15 = .000$	0.0	2	.133
$3-4$	2	$2/15 = .133$	13.3	2	.133
$0-2$	0	$0/15 = .000$	0.0	0	.000

(Tabla 2)





Test Scores



Test Scores

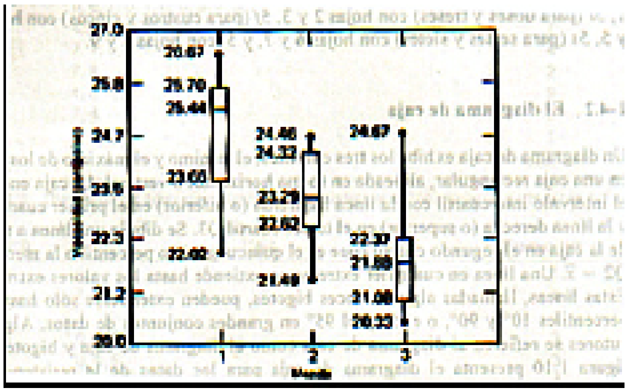
* Realizar una ojiva que represente los datos de la tabla2

- 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.

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

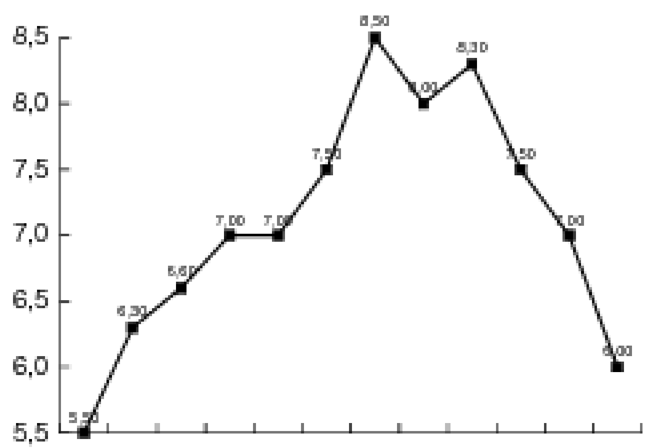
mezcla 1	mezcla 2	mezcla 2
22.02	21.49	20.33
23.83	22.67	21.67
26.67	24.62	24.67
25.38	24.18	22.45
25.48	22.78	22.28
23.50	22.56	21.95
25.90	24.46	20.49
24.98	23.79	21.81



- 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.

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



Serie de tiempo

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 parráfo

(c) Velocidad de un coche

(d) Edad de un trabajador

(e) Numéro de billetes de 20000 que circulan en una ciudad

(f) Numéro de estudiantes matriculados en la Universidad

(g) Caudal del rio,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

i. Intervalos largos e iguales

ii. Intervalos cortos e iguales

iii. Intervalos de diferente rango

(f) Realice un diagrama circular

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

(h) ¿La gáfica 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 parráfo se utilizarón 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

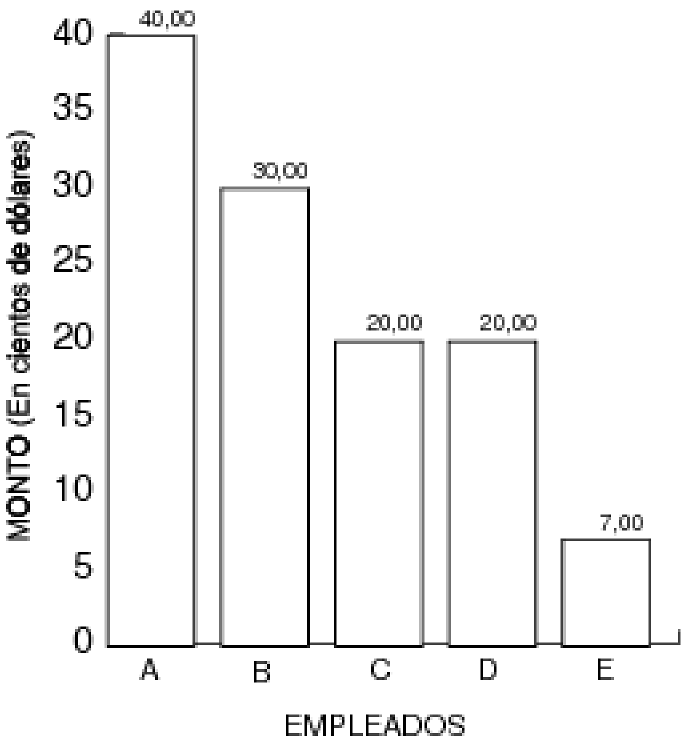
(b) Ordinales

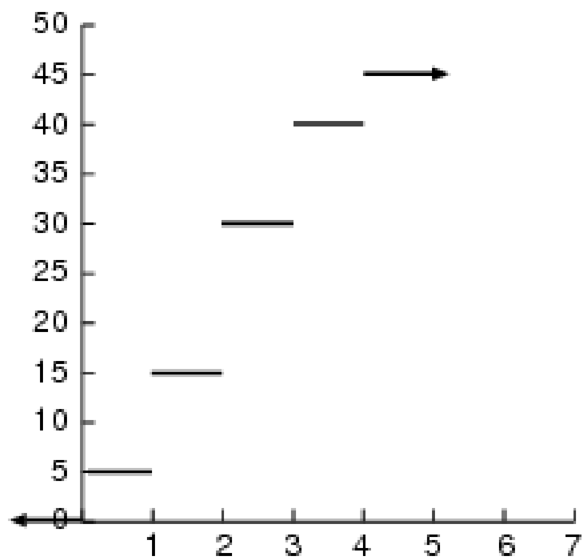
(c) Cuantitativos discretos

(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


40,00
30,00
20,00
20,00
7,00
MONTO (En cientos de dólares)
0
5
10
15
20
25
30
35
40
A
B
C
D
E
EMPLEADOS


0
5
10
15
20
25
30
35
40
45
50
1
2
3
4
5
6
7

Índice general

ESTADÍSTICA PARA INGENIEROS
Notas De Clase

UNIVERSIDAD DEL NORTE

ÁREA DE CIENCIAS BÁSICAS

MATEMÁTICAS Y FÍSICA

ADMINISTRACIÓN DE EMP.
LIC. ANTALCIDES OLIVO

3 de febrero de 2003

Índice general

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 clasifiacar los datos, organizarlos, tabularlos y representarlos graficamente, 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 varibles 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 origenes 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 numéros se habla de los censos realizados por los levitas, lo mismo que los romanos lo cual lo hacian 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 historico 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.

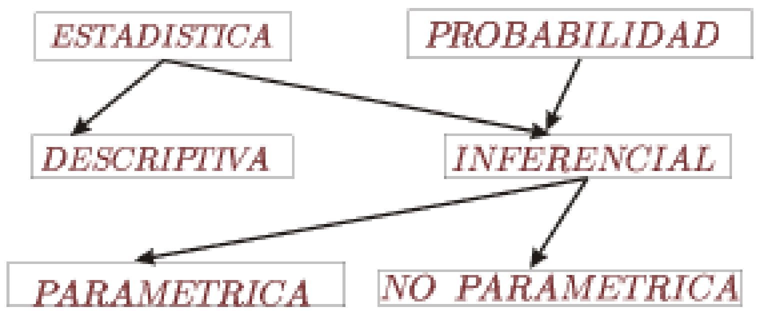
Ella se encarga de establecer los métodos y procedimientos para recolectar, clasificar y resumir datos para luego analizarlos siempre que exista variabilidad e incertidumbre causada intrínsicamente por ellos y después ralizar inferencias con el fin de hacer predicciones y tomar decisiones.

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.

Ahora clasificaremos la estadística.

Concepto 1.1 *Estadística descriptiva: Describe y analiza los datos utilizando métodos numéricos elementales y gráficos para resumir y presentar la información contenida en ellos.*



Concepto 1.2 *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

Concepto 1.3 *Individuos o elementos:Son personas u objetos que contienen cierta información de interés.*

Concepto 1.4 *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*

Concepto 1.5 *Muestra: Es un subconjunto representativo de una población.*

Concepto 1.6 *Caracteres o atributos: Son rasgos o cualidades de los elementos de la población, los cuales pueden ser cualitativos o cuantitativos.*

Concepto 1.7 *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.*

Concepto 1.8 *Clases: Conjunto de una o más modalidades en el que se verifica que cada modalidad pertenece a una y solo una de las clases*

Concepto 1.9 *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 1.1 *Variables cualitativas: Son aquellas que las modalidades posibles son de tipo nominal o categórico.*

Ejemplo 1.1 {*si, no*}, {*hombre, mujer*}, {*blanco, negro, amarillo,...,etc.*}

Definición 1.2 *Variables casicuantitativas: Son variables de tipo nominal, pero se puede establecer un orden entre ellas.*

Ejemplo 1.2 {*1º,2º,3º,...*}, {*alto, medio, bajo*}

Definición 1.3 *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 1.4 *Discretas: Cuando las variables toman valores puntuales*

Ejemplo 1.3 *Cuando el rango es*

Definición 1.5 *Continuas: si X toma todos los valores en un intervalo*

Ejemplo 1.4 *si*

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 1.6 *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 1.7 *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 1.8 *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 1.9 *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

- $$n = \sum_{j=1}^{k} n_j \; , \; \sum_{j=1}^{k} f_j$$

Definición 1.10 *Distribución de frecuencias: Es una función que a cada clase c_i de C (el conjunto de todas las clases) le asigna un valor es decir si definimos la función*

$$\mathfrak{F} : Cac_i \longmapsto n_i$$

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

Definición 1.11 *Tabla de frecuencia: Es una representación de $\mathfrak{F}$*

y generalmente es de la siguiente forma:

M	$F.$	$F.R.$	$F.A.A.$	$F.R.A.$
C	n_i	f_i	N_i	F_i
c_1	n_1	$f_1=\frac{n_1}{n}$	$N_1=n_1$	$F_1=\frac{N_1}{n}=f_1$
c_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	$f_k=\frac{n_k}{n}$	$N_k=n$	$F_k=1$
	n	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.

Ejemplo 1.5 *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 siguiente tabla:*

Nº de caras	n_i	f_i	N_i	F_i
0	*38*			
1	*144*			
2	*342*			
3	*287*			
4	*164*			
5	*25*			
	1000			

a .Completar la tabla

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

[1]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

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 quí 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 1.12 *Amplitud: La amplitud de un intervalo se define* $a_i = x_i - x_{i-1}$.

Definición 1.13 *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.

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

Elección de clases para variables continuas

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

- Lo primero es determinar k, entre mayor sea su valor mejor[2], 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[3]

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

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

[3] 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.

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 $a\prime = [[a]] + 1$

Ejemplo 1.6 *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

Determinar una tabla de frecuencias:

Solución 1.6.1 *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 fómula* $k = 1 + 3{,}22 * \log_{10} 21 = 5.$ $257\,5 \cong 6$ *aunque podríamos escoger también* $k = \sqrt{21} \cong 5.$

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 $a\prime = 7 \Longrightarrow R\prime = a\prime \cdot 5 = 35$

$d = R\prime - R = 2 \Longrightarrow x_0 = x_{\min} - \frac{d}{2} = 39, \quad x_5 = x_{\text{máx}} + \frac{d}{2} = 74$

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 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
--------------------------------------------------------------------------------
```

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

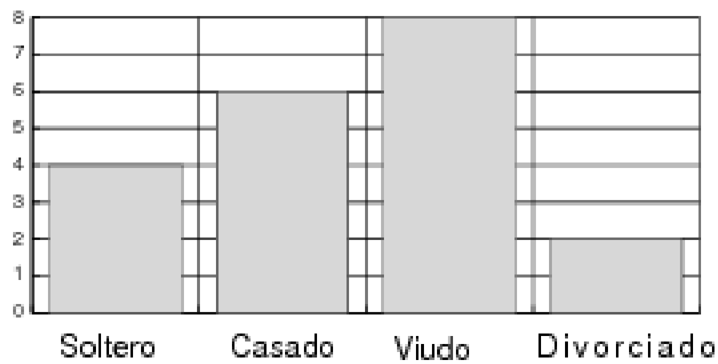


Diagrama de barra para una variable cualitativa

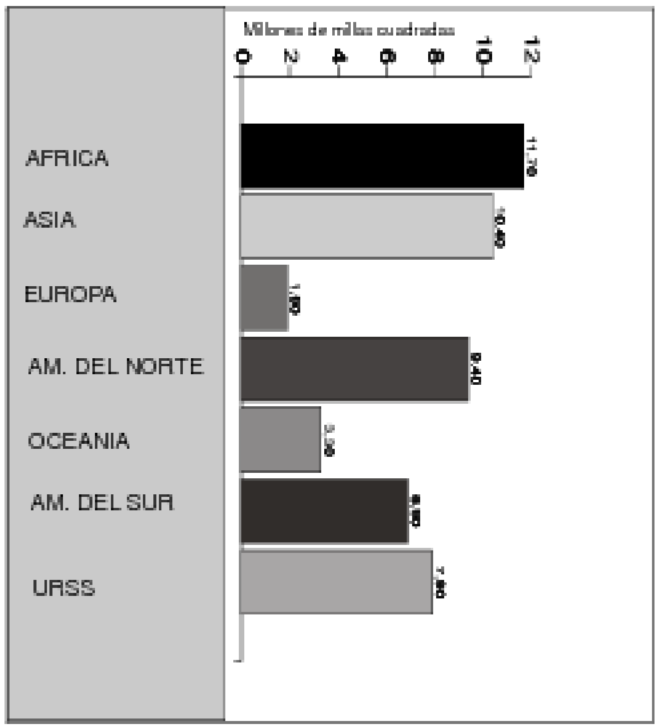
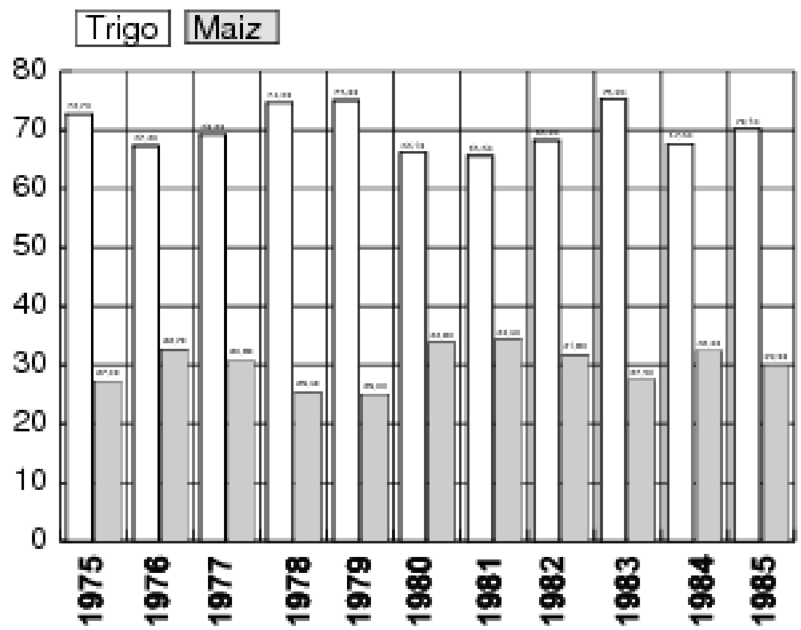


Figura 1.1: AREAS DE LOS CONTINENTES



[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:

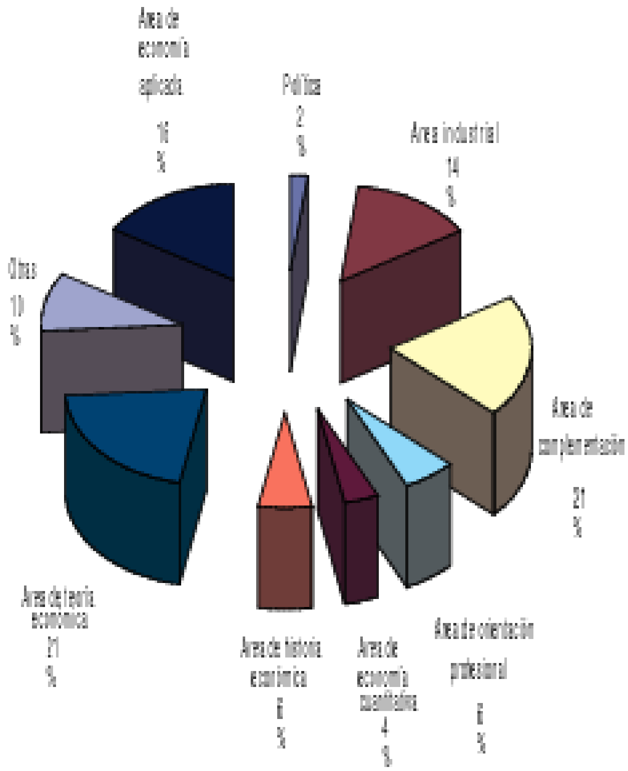
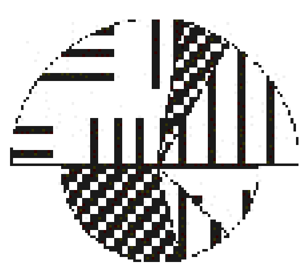


Diagrama circular

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.

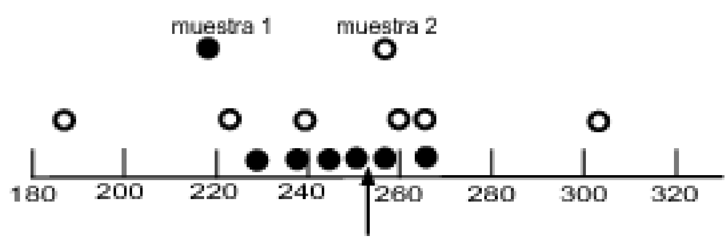


DOS MUESTRAS

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 1.7 *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 varible 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 1.8 *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		

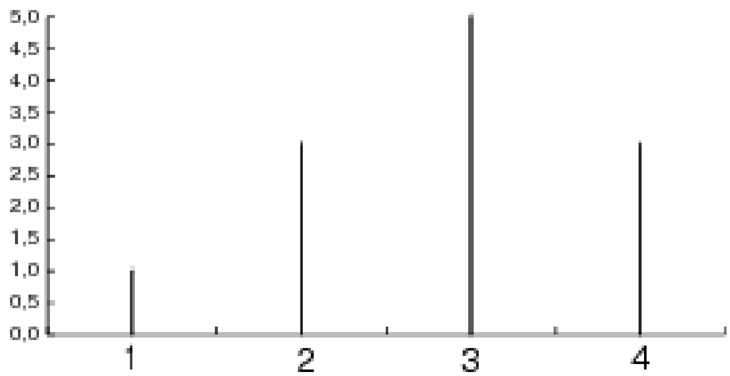
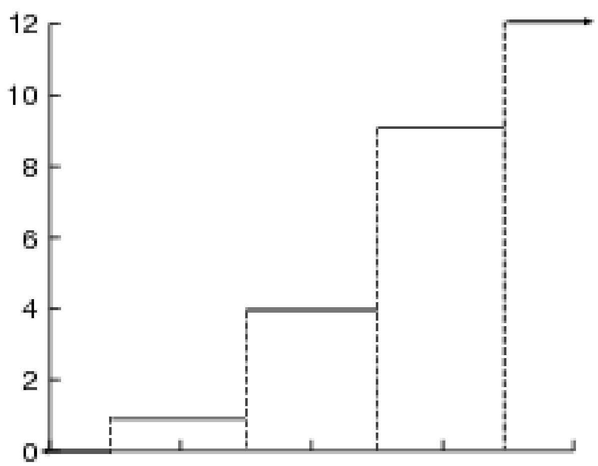


Figura 1.2: FRECUENCIAS ABSOLUTAS



FRECUENCIAS ABSOLUTAS ACUMULADAS

Para las variables continuas también se pueden representar con diagramas circulares

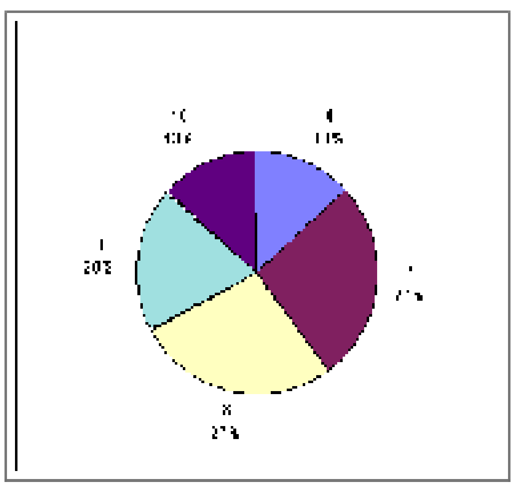
Ejemplo 1.9 *La tabla representa las notas obtenidas por los alumnos de 11º en un examen de Matemáticas en 1990*

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	$\mathbf{n}_i$	$\boldsymbol{f_i}$	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: Éste 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 1.10 *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	

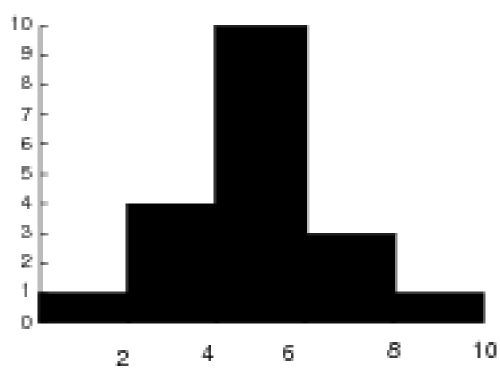
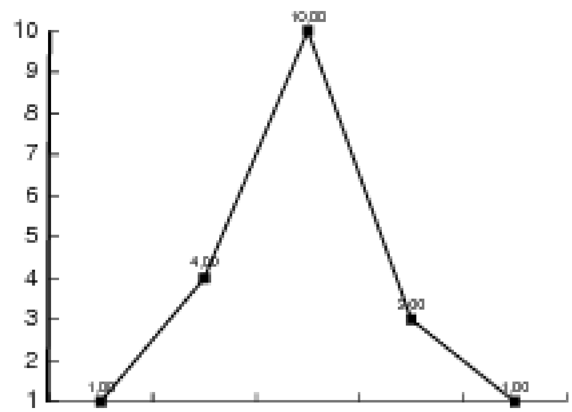


Diagrama Diferencial



Polígono de frecuencias

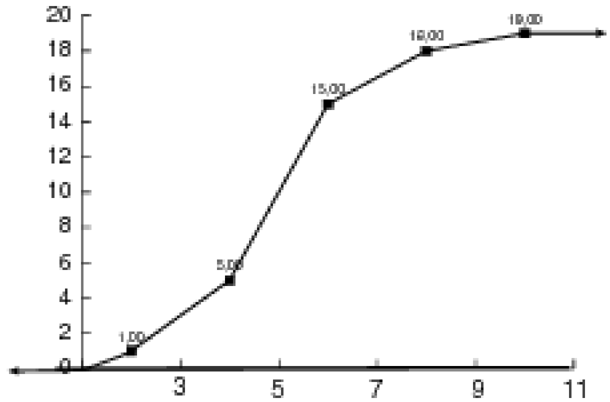


Diagrama acumulado

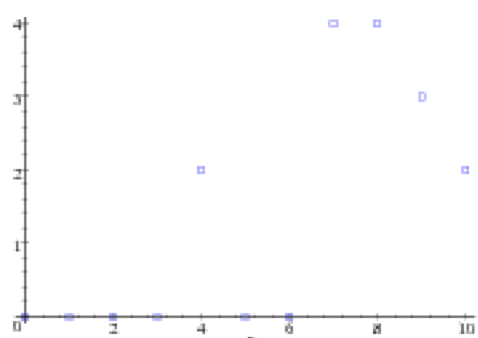
Ejemplo 1.11 *Representar gráficamente los datos de la tabla1*

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

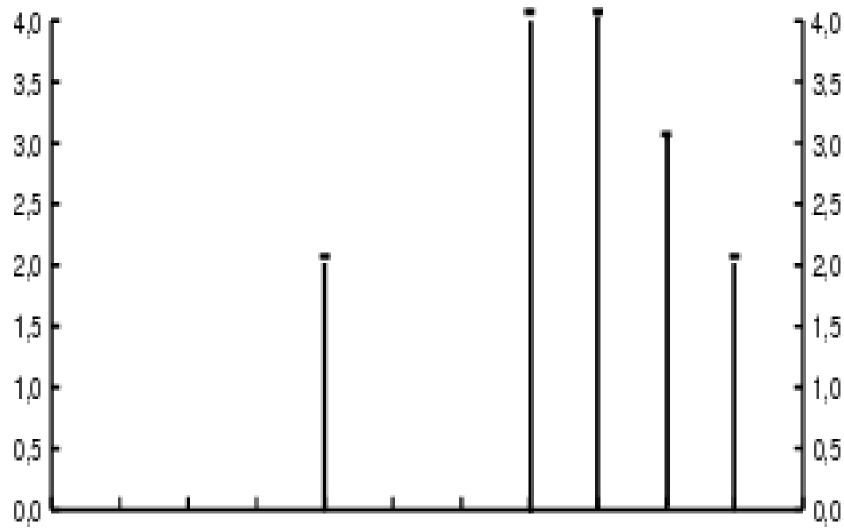
(Tabla 1)

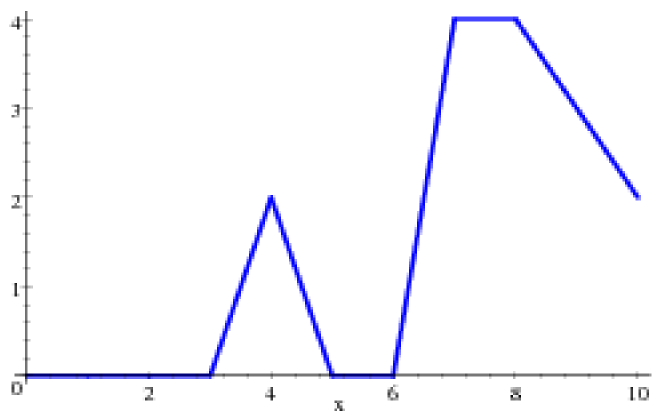
Intervalos	$\mathbf{n}_i$	f_i	$f_i\,\%$	$\mathbf{N}_i$	$\mathbf{F}_i$
$9-10$	5	$5/15=,333$	33,3	15	1,000
$7-8$	8	$8/15=,534$	53,4	10	,667
$5-6$	0	$0/15=,000$	0,0	2	,133
$3-4$	2	$2/15=,133$	13,3	2	,133
$0-2$	0	$0/15=,000$	0,0	0	,000

(Tabla 2)



Test Scores





Test Scores

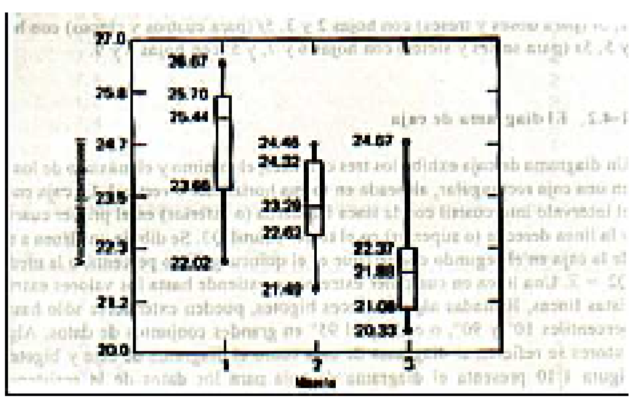
* Realizar una ojiva que represente los datos de la tabla2

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

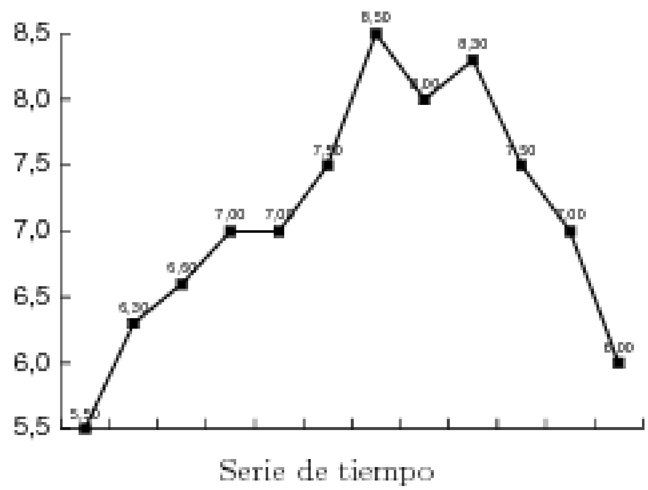
mezcla 1	mezcla 2	mezcla 2
22.02	21.49	20.33
23.83	22.67	21.67
26.67	24.62	24.67
25.38	24.18	22.45
25.48	22.78	22.28
23.50	22.56	21.95
25.90	24.46	20.49
24.98	23.79	21.81



- 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 1.13 *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*



Tarea 1 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 parráfo*

c) *Velocidad de un coche*

d) *Edad de un trabajador*

e) *Numéro de billetes de 20000 que circulan en una ciudad*

f) *Numéro de estudiantes matriculados en la Universidad*

g) *Caudal del rio, 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 gáfica 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 parráfo se utilizarón 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*

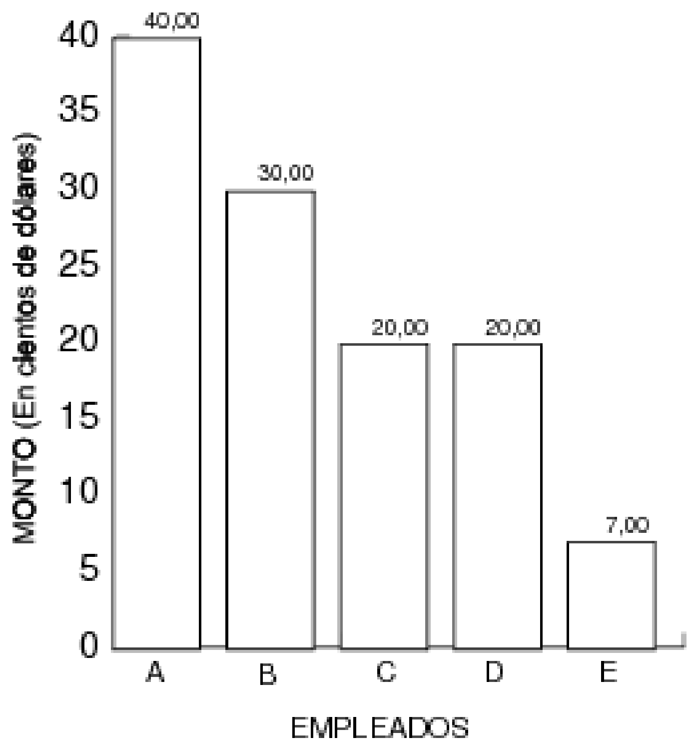
b) *Ordinales*

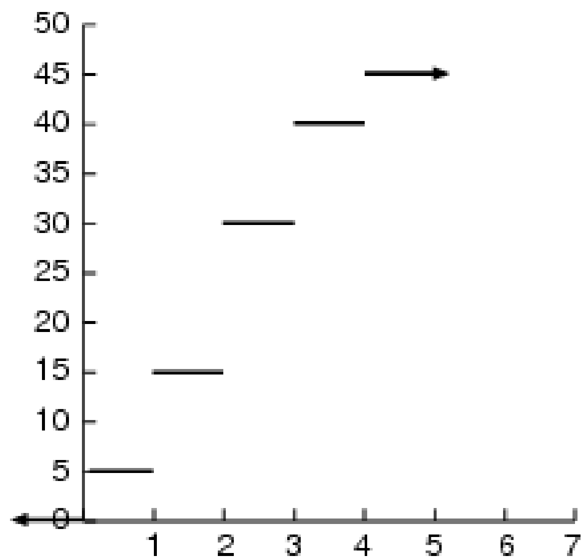
c) *Cuantitativos discretos*

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*


40,00
30,00
20,00
20,00
7,00
MONTO (En cientos de dólares)
0
5
10
15
20
25
30
35
40
A
B
C
D
E
EMPLEADOS


0
5
10
15
20
25
30
35
40
45
50
1
2
3
4
5
6
7

Ω. S

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

$A \subseteq \Omega$ B

B

□ □ □

$\mathbb{N}$ $\mathbb{Z}$ $\mathbb{Q}$

X

$X \in \mathbb{R}$

n

X

k

$c_1, c_2, c_3, ..., c_k$ $c_i, i = 1, 2, 3, ..., k$

c_i n_i

c_i f_i

c_i

N_i

c_i

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

F_i

c_i

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

- $$n = \sum_{j=1}^{k} n_j \; , \; \sum_{j=1}^{k} f_j$$

c_i C

$n_i \in I\!N_0$

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

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

$\mathfrak{F}$

	i	i	i	i
1	1	1 $\frac{n_1}{n}$	1 1	1 $\frac{N_1}{n}$ 1
2	2	2 $\frac{n_2}{n}$	2 1 2	2 1 2
3	3	3 $\frac{n_3}{n}$	3 1 2 3	3 1 2 3
j	j	j $\frac{n_j}{n}$	j $\sum_{i=1}^{j}$ i	j $\sum_{i=1}^{j}$ i
k	k	k $\frac{n_k}{n}$	k	k

a	n_i	f_i	N_i	F_i

F_i F $=$

8

- c_i

- $x_1, x_2, x_3, \cdots, x_k.$

-

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

$(k-1)$ $(x_{i-1}, x_i]$ $[x_{k-1}, x_k].$ $x_{i-1} - x_i = I_i$

$a_i = x_i - x_{i-1}.$

$m_i \in I_i$

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

m_i

k

- k

$$k = \left\{ \begin{array}{ll} \sqrt{n} & n \\ 1 + 3{,}22 \log n & \end{array} \right\}$$

n $n \geq 40$

- $R = x_k - x_0$
-

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

-

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

a

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

a

$a\prime = [[a]] + 1$

X

$k = 1 + 3{,}22 * \log_{10} 21 = 5.$

$257\,5 \cong 6$ $k = \sqrt{21} \cong 5$

$$R = 73 - 40 = 33 \Longrightarrow a_i = a = \tfrac{33}{6} = 5.\,5$$

$x_0 = x_{\text{mín}} = 40$

$x_5 = x_{\text{máx}} = 73$

i		m_i	n_i	f_i	N_i	F_i
						≈ 1
						≈ 1
				≈ 1		

$$a\prime = 7 \Longrightarrow R\prime = a\prime \cdot 5 = 35$$
$$d = R\prime - R = 2 \Longrightarrow x_0 = x_{\min} - \frac{d}{2} = 39, \quad x_5 = x_{\max} + \frac{d}{2} = 74$$

i		m_i	n_i	f_i	N_i	F_i
						≈ 1
				≈ 1		

r_1 r_2 $r_2 > r_1$

$r_2 = r_1\sqrt{\frac{n_2}{n_1}}$ n

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

$x_i, \quad i = 1, 2, 3, ..., n$

x_i

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

x_i	n_i	f_i	N_i

10	2	$2/15 = ,133$
9	3	$3/15 = ,200$
8	4	$4/15 = ,267$
7	4	$4/15 = ,267$
4	2	$2/15 = ,133$

	$\mathbf{n}_i$	$_i$	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

	m_i	n_i	N_i

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

	$\mathbf{n}_i$	f_i	f_i	$\mathbf{N}_i$	$_i$
$9-10$	5	$5/15 =$,333	33,3	15	1,000
$7-8$	8	$8/15 =$,534	53,4	10	,667
$5-6$	0	$0/15 =$,000	0,0	2	,133
$3-4$	2	$2/15 =$,133	13,3	2	,133
$0-2$	0	$0/15 =$,000	0,0	0	,000

$n = 128$

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		

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

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

X

ESTADÍSTICA PARA INGENIEROS
Notas De Clase

UNIVERSIDAD DEL NORTE

ÁREA DE CIENCIAS BÁSICAS

MATEMÁTICAS Y FÍSICA

ADMINISTRACIÓN DE EMP.
LIC. ANTALCIDES OLIVO

28 de agosto de 2002

Índice general

Capítulo 2

Medidas descriptivas

2.1. Introducción

En el capítulo anterior analizamos los los datos de una muestra utilizando diagramas o gráficos con lo que nos dimos cuenta que al tratar de realizar inferencias a partir de ellos, los gráficos no estaban bien definidos ya que para un mismo grupo de datos podemos realizar gráficos diferentes, por los que debemos definir expresiones bien definidas para realizar mejores inferencias.

Estas expresiones
son cantidades matemáticas con las que podemos obtener conclusiones validando las inferencias, estas expresiones matemáticas que representan propiedades importantes de la muestra se llaman medidas descriptivas.

2.2. Medidas de tendencia central

2.2.1. Medias

Media aritmética

La media aritmética de una muestra es la suma de cada uno de los valores posibles multiplicado por su frecuencia, es decir.

Si la siguiente tabla representa la tabla de frecuencia de la muestra

M	n_i	f_i
x_1	n_1	f_1
$\vdots$	$\vdots$	$\vdots$
x_k	n_k	f_k

La media es el valor :

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i n_i}{n} \tag{2.1}$$

y si los datos no están ordenados entonces

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i}{n} \tag{2.2}$$

Nota 2.1 *En la definición de media se consideró que la variable de interés X es discreta, pero si la variable X no es discreta sino continua. En la fórmula se reemplaza cada valor x_i por la marca de clase correspondiente es decir*

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i n_i}{n} \tag{2.3}$$

Este proceso hace que la media aritmética difiera de la media obtenida según (2.1), es decir habrá una perdida de precisión que será mayor en cuanto mayor sea la diferencia entre las marcas de clase y los valores reales, o sea entre mayor sea la longitud a_i de los intervalos

Desviaciones

$$\sum_{i=1}^{k} (x_i - \overline{x})\, n_i \approx 0 \qquad \textit{Dispersión total} \tag{2.4}$$

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 \gtrapprox 0 \quad \textit{Error cuadrático} \tag{2.5}$$

$$\sum_{i=1}^{n} |x_i - \bar{x}| \gtrapprox 0 \quad \textit{Error absoluto} \tag{2.6}$$

Estas fórmulas determinan la variabilidad entre los datos.

Ejemplo 2.1 *Obtener la media y las desviaciones con respecto a la media*

$N°de\,caras$	n_i	$x_i n_i$	$x_i - \bar{x}$
0	38	0	
1	144	144	
2	342	684	
3	287	861	
4	164	656	
5	25	125	
	1000		

Ejemplo 2.2 *Determinar la media aritmética del ejemplo 1.10*

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

Proposición 2.1 *Dados r grupos con $n_1, n_2, n_3, \cdots, n_r$ observaciones, si $\bar{x}_1$, $\bar{x}_2, \bar{x}_3, \cdots, \bar{x}_r$ son las respectivas medias, entonces la media de las $n = n_1 + n_2 + n_3 + \cdots + n_r$ observaciones es:*

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i \bar{x}_i}{n} \tag{2.7}$$

Desventajas de la media La media es una medida muy usada en estadística, pero a pesar de eso posee ciertas desventajas

- La media aritmética es muy sensible a los valores extremos, es decir si una medida se aleja mucho de las otras hará que la media se aproxime mucho a ella
- No se recomienda usar cuando los datos se desplazan hacia los extremos
- En el caso de variables continuas depende de los intervalos de clase
- en le caso de variables discretas el valor puede no ser un valor de la muestra.

Otras medias

Media geométrica [1]
$\bar{x}_g$ es la media de los logaritmos de las medidas

$$\lg \bar{x}_g = \frac{\sum_{i=1}^{k} \lg \bar{x}_i}{n}$$

es decir

$$\bar{x}_g = \sqrt[n]{x_1.x_2.x_3 \cdots x_3}$$

y en el caso de datos agrupados tenemos

$$\bar{x}_g = \sqrt[n]{x_1^{n_1}.x_2^{n_2}.x_3^{n_3} \cdots x_3^{n_k}}$$

Media armónica

$\bar{x}_a$ se define como el recíproco de la media aritmética de los recíprocos :

$$\frac{1}{\bar{x}_a} = \frac{\sum_{i=1}^{n} \frac{1}{x_i}}{n}$$

[1]Se aconseja usar cuando la diferencia entre los extremos es muy grande comparada con la diferencia entre los otros datos o cuando la diferencia entre un dato y otro siempre aumenta.

Media cuadrática

[2]

$\bar{x}_c$ es la raíz cuadrada de la media aritmética de los cuadrados

$$\bar{x}_c = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

2.2.2. La mediana

Sea X una variable discreta cuyas observaciones han sido ordenadas de mayor de mayor a menor, entonces se le llama mediana $\widetilde{x}$ al primer valor de la variable que deja por debajo de si el 50% de las observaciones es decir si n es el número de observaciones, la mediana será la observación $\left[\left|\frac{n}{2}\right|\right] + 1$

Definición 2.1 *Sea $x_{(1)}$,$x_{(2)}$, $x_{(3)}$, $\cdots$, $x_{(n)}$ las observaciones de una muestra para una variable X donde $x_{(1)}$ representa la observación más pequeña, $x_{(2)}$ la observación que le sigue en valor y así sucesivamente $x_{(n)}$ denota la observación de mayor valor, entonces la mediana se define*

$$\widetilde{x} = \begin{cases} x_{(|n+1|/2)} & \textit{si } n \textit{ impar} \\ \frac{x_{(n/2)} + x_{(|n/2|+1)}}{2} & \textit{si } n \textit{ par} \end{cases}$$

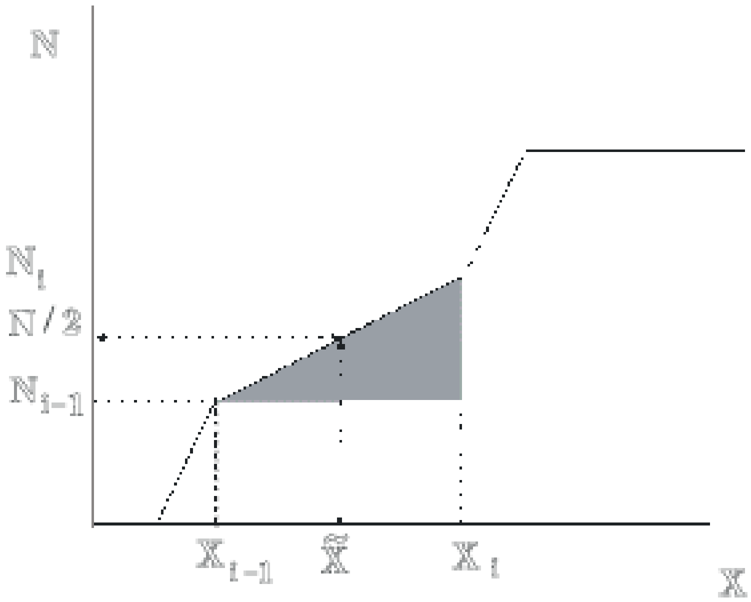
En el caso de variables continuas, las clases vienen dadas por intervalos como se indicó en el capítulo anterior por tal razón para determinar la mediana se escoge el intervalo donde se encuentra el valor para el cual están debajo de él la mitad de los datos. Entonces a partir de ese intervalo se observan las frecuencias absolutas acumuladas y se aplica la siguiente fórmula

$$\widetilde{x} = x_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i} a_i$$

de aquí se puede deducir que $\widetilde{x}$ el "punto" que divide al histograma en dos partes de áreas iguales

Gráficamente la moda para variables continuas se determina de acuerdo con la siguiente gráfica

[2] Es usada muy a menudo cuando se estudian experimentos científicos



Propiedades y desventajas de la mediana

1. Tiene la ventaja de no ser afectada por los valores extremos y por eso se aconseja para distribuciones para las cuales los datos no se concentran en el centro

2. Es fácil de calcular

3. En el caso de variables discretas el valor de la mediana es un valor de la variable

4. El mayor defecto es que las propiedades matemáticas son muy complicadas y esto hace que muy poco se use para realizar inferencias

5. Es función de los intervalos escogidos en el caso de variables continuas

Ejemplo 2.3 *Sea X una variable discreta que ha presentado las siguientes observaciones*

$X \rightsquigarrow 2, 5, 7, 9, 12 \Rightarrow \widetilde{x} = 7$ y $\overline{x} = 7$

Ahora si cambiamos la ultima observación, no se afectará la mediana pero si la media

$X \rightsquigarrow 2, 5, 7, 9, 125 \Longrightarrow \widetilde{x} = 7$ y $\overline{x} = 29,6$

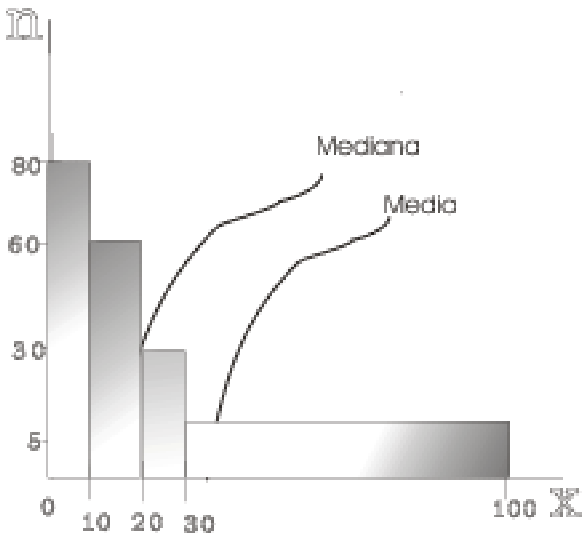
Además la media no es valor de la variable

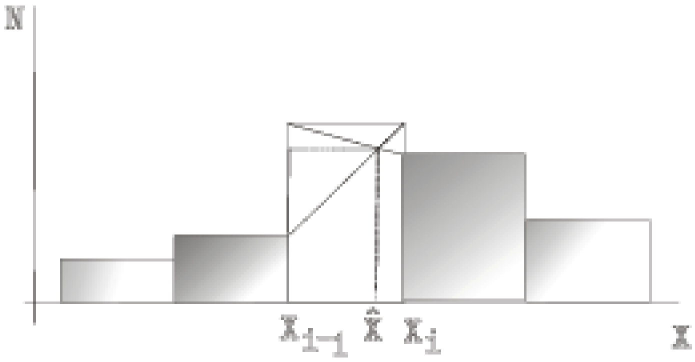
Ejemplo 2.4 *Obtener la media y la mediana de la tabla frecuencias siguientes*

$x_{i-1} - x_i$	m_i	n_i	a_i	$x_i n_i$	N_i
0-10		80			
10-20		60			
20-30		30			
30-100		5			
100-500		5			

$\overline{x} = \frac{\quad}{\quad} \qquad \widetilde{x} = \qquad + \frac{\qquad}{\qquad} \qquad =$

- Realizar el histograma de frecuencia y comparar la media y la mediana





La moda

Llamaremos moda $\widehat{x}$ a cualquier máximo relativo de la distribución de frecuencias, es decir, cualquier valor de la variable que posea una frecuencia mayor que su anterior y posterior valor.

En el caso de variables continuas es más correcto hablar de intervalos modales. Luego de determinar el intervalo de clase o intervalo modal,que es aquel para el cual la distribución de frecuencia posee un máximo relativo, se determina la moda utilizando la siguiente fórmula

$$\widehat{x} = x_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} a_i$$

La cual se obtiene geométricamente como se indica en la fig. 2.3

Propiedades de la moda

La moda posee la siguientes propiedades

- Es muy fácil de calcular
- Puede no ser única
- Es función de los intervalos de su amplitud,número y límites

Relación entre la media, la moda y la mediana

En el caso de distribuciones unimodales, la mediana está con frecuencia comprendida entre la media y la moda (incluso más cerca de la media).

En distribuciones que presentan cierta inclinación, es más aconsejable el uso de la mediana. Sin embargo en estudios relacionados con propósitos estadísticos y de inferencia suele ser más apta la media.

Veamos un ejemplo de cálculo de estas tres magnitudes.

Ejemplo 2.5 *Consideramos una tabla estadística relativa a una variable continua, de la que nos dan los intervalos, las marcas de clase m_i, y las frecuencias absolutas, n_i, para determinar las tres medidas de tendencia central básicas*

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

.

La mediana es el valor que deja debajo de sí el 50 % de los datos, en la tabla en la columna de las frecuencias absolutas acumuladas observamos que sexto dato se encuentra en la tercera clase, es decir

$i = 3$

$(x_{i-1}, x_i] = (4, 6]$ intervalo donde se encuentra la mediana

$\widetilde{x} = \widetilde{x} = x_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i} a_i = 4 + \frac{\frac{12}{2} - 3}{4} \cdot 2 = 5,5$

para calcular la moda,debemos encontrar el o los intervalos modales, buscando los máximos relativos en las columnas de las frecuencias absolutas n_i.

Encontramos dos intervalos con un máximo relativo,

para $i = 1$ cuya frecuencia absoluta es 2

para $i = 3$ cuya frecuencia absoluta es 4.

Por tanto la moda en cada caso se calcula

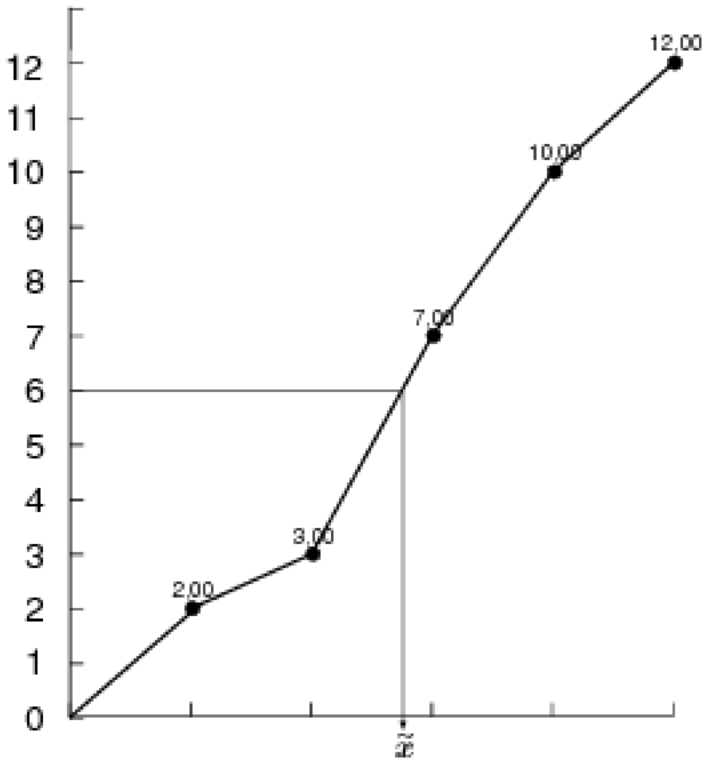
cuando $i = 1 \Rightarrow$

$\widehat{x} = x_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} a_i = 0 + \frac{2-0}{(2-0)+(2-1)} \cdot 2 = 1,\widehat{3}$

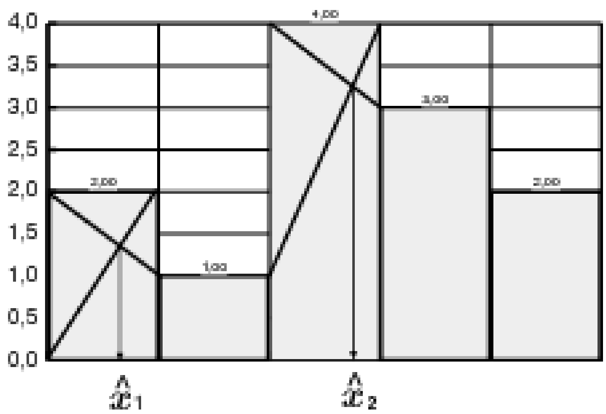
cuando $i = 3 \Rightarrow$

$\widehat{x} = x_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} a_i = 4 + \frac{4-1}{(4-1)+(4-3)} \cdot 5 = 5,5$

Ahora lo resolveremos geométricamente



$\tilde{x} = 5,5$



$\hat{x}_1 = 1,33$; $\hat{x}_2 = 5,5$

La media es $\bar{x} = 5.\hat{3}$

2.3. Medidas de posición

A veces es importante obtener los valores de la variable que dividen la población en cuatro,diez o cien partes iguales,usualmente llamados cuartiles deciles y percentiles respectivamente.

El procedimiento es similar al utilizado para determinar la mediana, como lo indicaremos ahora.

Percentil

Para una variable discreta, se define el percentil de orden k, como la observación que deja por debajo de si el $k\,\%$ de la población es decir

$$N_k = n\frac{k}{100}, \text{ si } n \text{ es impar}$$
$$\text{es decir } p_k = x_{[|(n+1)\frac{k}{100}|]}$$

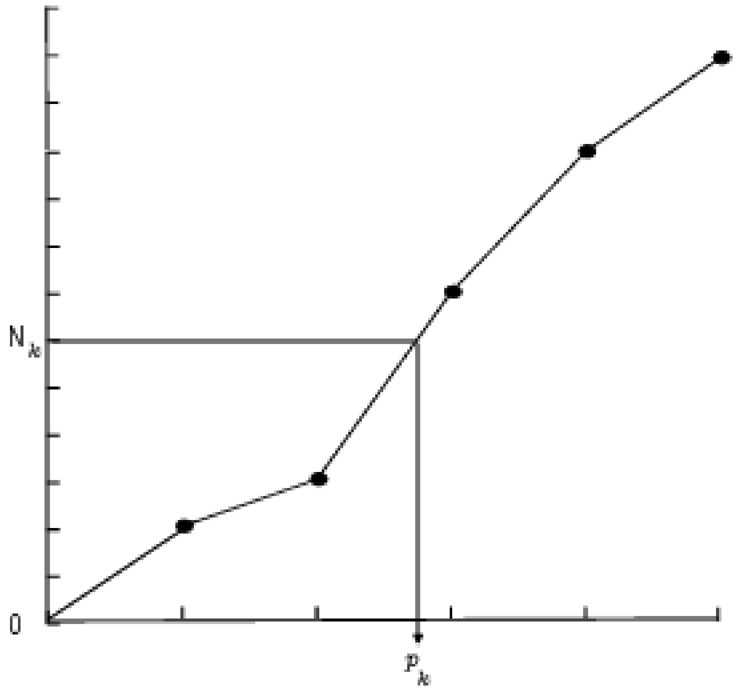
donde en sub índice $[|(n+1)\frac{k}{100}|]$ indica que es la posición k a la que le corresponde ese valor de la frecuencia absoluta acumulada. Si n es par

$$p_k = \frac{x_{[|n\frac{k}{100}|]} + x_{[|n\frac{k}{100}|]+1}}{2}$$

En el caso de variables continuas se busca el intervalo donde se encuentra p_k, es decir se busca el valor que deja por debajo de si el $k\,\%$ de las observaciones y se determina el intervalo $(x_{i-1}, x_i]$ donde se encuentra y se utiliza la relación

$$p_k = x_{i-1} + \frac{n\frac{k}{100} - N_{i-1}}{n_i} \cdot a_i$$

para determinarlo geométricamente observe la figura



.

Quartiles

Los cuartiles son tres y se definen:

- $Q_1 = p_{25}$
- $Q_2 = p_{50}$
- $Q_3 = p_{75}$

2.3.1. Deciles

De manera análoga se definen los deciles
los deciles son los valores que dividen las observaciones en 10 grupos de igual tamaño es decir son el conjunto
$D_1, D_2, D_3, \cdots, D_{10}$ y se definen
$D_i = p_{10.i} \qquad i = 1, 2, 3, \cdots, 10$

Ejemplo 2.6 *1. Dada la siguiente tabla de distribución de frecuencia Determine los tres cuartiles*

x_i	n_i	N_i
0	*14*	
1	*10*	
2		
3	*26*	
4	*20*	
5	*15*	*100*
	100	

Solución
Primer cuartil

$$\frac{n}{4} = 25 \Rightarrow N_i > \frac{n}{4} = 39 \Rightarrow Q_1 = 2$$

Segundo cuartil

$$\frac{2n}{4} = 50 \Rightarrow N_i > \frac{2n}{4} = 65 \Rightarrow Q_2 = 3$$

Tercer cuartil

$$\frac{3n}{4} = 75 \Rightarrow N_i > \frac{3n}{4} = 85 \Rightarrow Q_3 = 4$$

Ejemplo 2.7 *Calcular los cuartiles en las siguiente distribución de una variable continua*

x_{i-1}	n_i	N_i
0-1	*10*	
1-2	*12*	
2-3		
3-4	*10*	
4-5	*7*	*51*
	51	

primer cuartil

$$\frac{n}{4} = 12,75 \Rightarrow 22 = N_i > \frac{n}{4} \Rightarrow Q_1 \in (1,2]$$
$$Q_1 = x_{i-1} + \frac{n/4 - N_{i-1}}{n_i} \cdot a_i = 1 + \frac{12,75 - 10}{12} \cdot 1 = 1,23$$

segundo cuartil

$$\frac{2n}{4} = 25,5 \Rightarrow 34 = N_i > \frac{2n}{4} \Rightarrow Q_2 \in (2,3]$$
$$Q_2 = x_{i-1} + \frac{2n/4 - N_{i-1}}{n_i} \cdot a_i = 2 + \frac{25,5 - 22}{12} \cdot 1 = 2,29$$

Tercer cuartil

$$\frac{3n}{4} = 38,5 \Rightarrow 44 = N_i > \frac{3n}{4} \Rightarrow Q_3 \in (3,4]$$
$$Q_3 = x_{i-1} + \frac{3n/4 - N_{i-1}}{n_i} \cdot a_i = 3 + \frac{38,25 - 34}{10} \cdot 1 = 3,45$$

Ejemplo 2.8 *La distribución de una variable tiene por polígono acumulativo de frecuencias relativas.Si el número total de observaciones es de 50*

Intervalo	m_i	F_i	a_i
0-5	*2,5*	*0,2*	*5*
5-7	*6*	*0,7*	*2*
7-12	*9,5*	*0,8*	*5*
12-15	*13,5*	*1*	*3*

1. *Elaborar una tabla estadística con los siguientes elementos : intervalos de clase, marcas de clase, frecuencia absoluta, frecuencia absoluta acumulada, frecuencia relativa y frecuencia relativa acumulada.*

2. *Cuantas observaciones tuvieron un valor inferior a 10, cuantas inferior a 8 y cuantas fueron superior a 11.*

3. *Calcule las modas.*

4. *Determine los cuartiles*

1. En la siguiente tabla se proporciona la información pedida y algunos cálculos auxiliares que nos permitirán responder otras preguntas.

Intervalo	m_i	n_i	N_i	f_i	F_i	a_i
0-5	2,5	10			0,2	5
5-7	6	25			0,7	2
7-12	9,5	5			0,8	5
12-15	13,5	10			1	3

2. Calcularemos ahora el número de observaciones que se han pedido

- Hasta el segundo intervalo van 35 datos ahora analizaremos cuantos datos tienen un valor menor que diez en el tercer intervalo

$$\begin{array}{lcc} 7\,a\,12 & \longrightarrow & 5 \\ 7\,a\,10 & \longrightarrow & x \end{array} \Longleftrightarrow \begin{array}{ccc} 5 & \longrightarrow & 5 \\ 3 & \longrightarrow & x \end{array} \Longrightarrow x = 3$$

  Por tanto el número de observaciones que tomaron un valor inferior a 10 es 38

- De igual forma que en el paso anterior hasta el segundo intervalo van 35 observaciones y en el tercer intervalo tenemos que

$$\begin{array}{lcc} 7\,a\,12 & \longrightarrow & 5 \\ 7\,a\,8 & \longrightarrow & x \end{array} \Longleftrightarrow \begin{array}{ccc} 5 & \longrightarrow & 5 \\ 1 & \longrightarrow & x \end{array} \Longrightarrow x = 1$$

  Por tanto el número de observaciones que tomaron un valor menor que 8 es 36

- En el cuarto intervalo hay 10 observaciones mayores que 11, por tanto hay que determinar cuantas mayores que 11 hay en el tercer intervalo

$$\begin{array}{lcl} 7\,a\,12 & \longrightarrow & 5 \\ 7\,a\,11 & \longrightarrow & x \end{array} \Longleftrightarrow \begin{array}{lcl} 5 & \longrightarrow & 5 \\ 4 & \longrightarrow & x \end{array} \Longrightarrow x = 4$$

  Entonces hay cuatro observaciones entre 7 y 11 por tanto mayor que 11 hay 1 es decir hay 11 observaciones mayores que 11

3. Tenemos dos modas

- Cuando $i = 2$

$$\widehat{x_1} = x_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} a_i$$
$$\widehat{x_1} = 5 + \frac{25 - 10}{(25 - 10) + (25 - 5)} \cdot 2 = 5,86$$

- Cuando $i = 4$

$$\widehat{x_2} = x_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} a_i$$
$$\widehat{x_2} = 12 + \frac{10 - 5}{(10 - 5) + (10 - 0)} \cdot 7 = 14,\widehat{3}$$

4 Los Cuartiles son

$$Q_1 = x_{i-1} + \frac{n/4 - N_{i-1}}{n_i} \cdot a_i = 5 + \frac{12,5 - 10}{25} \cdot 2 = 5,2$$
$$Q_2 = x_{i-1} + \frac{2n/4 - N_{i-1}}{n_i} \cdot a_i = 5 + \frac{25 - 10}{25} \cdot 2 = 6,2$$
$$Q_3 = x_{i-1} + \frac{3n/4 - N_{i-1}}{n_i} \cdot a_i = 7 + \frac{37,5 - 35}{5} \cdot 5 = 9,5$$

2.4. Medidas de variabilidad o dispersión

Las medias de posición o medidas de tendencia central nos indican donde se sitúa un grupo de medidas. Los de variabilidad o dispersión nos indican si esas medidas o valores están próximas entre sí o si por el contrario están dispersas.

Una medida razonable da la variabilidad podría ser la amplitud o rango, el cual se obtiene restando el valor o medida mínima de la medida máxima de un conjunto de observaciones

Esta forma de determinar la variabilidad es fácil de calcular y tiene las mismas unidades de las medidas, pero tiene algunos inconvenientes

o No utiliza todas las observaciones si no dos de ellas

o Es afectada por las observaciones extremas

o El rango aumenta con el número de medidas, o bien queda igual. En ningún caso disminuye

Por eso a continuación determinaremos otras formas de medir la dispersión de una variable

2.4.1. Desviación media

Se define la desviación media D_m como la media de las diferencias en valor absoluto de los valores de la variable a la media, es decir, si tenemos un conjuntos de n observaciones $x_1, x_2, x_3, \cdots, x_n$ entonces

$$D_m = \frac{1}{n}\sum_{i=1}^{n}|x_i - \overline{x}|$$

Si los datos están agrupados en una tabla estadística es más sencillo usar la relación

$$D_m = \frac{1}{n}\sum_{i=1}^{k}|x_i - \overline{x}| \cdot n_i$$

Como se observa, la desviación media guarda las mismas dimensiones que las observaciones. La suma de los valores absolutos es relativamente fácil de calcular, pero esta simplicidad tiene un inconveniente: Desde el punto de vista geométrico, la distancia que induce la desviación media en el espacio de observaciones no es natural (no permite definir ángulos entre dos conjuntos de observaciones). Esto hace que sea muy engorroso trabajar con ella a la hora de hacer inferencias.

2.4.2. Varianza y desviación típica

Como forma de medir la dispersión de los datos hemos descartado:

- $\sum_{i=1}^{n}(x_i - \overline{x})$, pues sabemos que esa suma vale 0, ya que las desviaciones con respecto a la media se compensan al haber términos en esa suma que son de signos distintos.

- Para tener el mismo signo al sumar las desviaciones con respecto a la media podemos realizar la suma con valores absolutos. Esto nos lleva a la D_m, pero como hemos

mencionado, tiene poco interés por las dificultades que presenta.

- Si las desviaciones con respecto a la media las consideramos al cuadrado, $(x_i - \overline{x})^2$, de nuevo obtenemos que todos los sumandos tienen el mismo signo (positivo). Esta es además la forma de medir la dispersión de los datos de forma que sus propiedades matemáticas son más fáciles de utilizar. Vamos a definir entonces dos medidas que serán fundamentales en el resto del curso:

La varianza y la desviación típica.

La varianza, S_n^2, se define como la media de las diferencias cuadráticas de n puntuaciones con respecto a su media aritmética, es decir

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Para datos agrupados en tablas, usando las notaciones establecidas en el capítulo anterior, la varianza se puede escribir como

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x})^2 \cdot n_i$$

Una fórmula equivalente para el cálculo de la varianza está basada en lo siguiente:

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 n_i - \overline{x}^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} \left(x_i^2 - 2x_i\overline{x} + \overline{x}^2\right)$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\overline{x}\frac{\sum_{i=1}^{n} x_i}{n} + \frac{1}{n}n\overline{x}^2$$

Con lo cual se tiene

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2$$

Si los datos están agrupados en tablas, es evidente que

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 n_i - \overline{x}^2$$

La varianza no tiene la misma magnitud que las observaciones (ej. si las observaciones se miden en metros, la varianza lo hace en $metros^2$). Si queremos que la medida de dispersión sea de la misma dimensionalidad que las observaciones bastará con tomar su raíz cuadrada. Por ello se define la desviación típica, S_n, como

$$S_n = \sqrt{S_n^2}$$

Ejemplo 2.9 *Calcular la varianza y desviación típica de las siguientes cantidades medidas en metros:*

3,3,4,4,5

Solución 2.9.1 *Para calcular dichas medidas de dispersión es necesario calcular previamente el valor con respecto al cual vamos a medir las diferencias. Éste es la media:*

$$\overline{x} = 3,8\, metros$$

La varianza es:

$$\begin{aligned} S_n^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 n_i - \overline{x}^2 \\ &= \frac{1}{5}\left(9+9+16+16+25\right) - 3,8^2 = 0,56m^2 \end{aligned}$$

siendo la desviación típica su raíz cuadrada:

$$S_n = \sqrt{0,56m^2} = 0,748m$$

Las siguientes propiedades de la varianza (respectivamente, desviación típica) son importantes a la hora de hacer un cambio de origen y escala a una variable. En primer lugar, la varianza (resp. Desviación típica) no se ve afectada si al conjunto de valores de la variable se le añade una constante. Si además cada observación es multiplicada por otra constante, en este caso la varianza cambia en relación al cuadrado de la constante (resp. La desviación típica cambia en relación al valor absoluto de la constante). Esto queda precisado en la siguiente proposición:

Proposición 2.2 *Si $Y = aX + b$ entonces $S_{n,Y}^2 = a^2 S_{n,X}^2$*

Demostración. Para cada observación x_i de $X, i = 1, 2, \cdots, n$, tenemos una observación de Y que es por definición $y_i = ax_i + b$, es fácil demostrar que $\overline{y} = a\overline{x} + b$. Por tanto, la varianza de Y es

$$\begin{aligned} S_{n,Y}^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[\left(ax_i + b\right) - \left(a\overline{x} + b\right)\right]^2 \\ &= \frac{1}{n}\sum_{i=n}^{n} a^2\left(x_i - \overline{x}^2\right) \\ &= a^2 S_{n,X}^2 \end{aligned}$$

■

Nota 2.2 *Las consecuencias del anterior resultado eran de esperar: Si los resultados de una medida son trasladados una cantidad b, la dispersión de los mismos no aumenta. Si estos mismos datos se multiplican por una cantidad $a < 1$, el resultado tenderá a concentrarse alrededor de su media (menor varianza). Si por el contrario $a > 1$ habrá mayor dispersión.*

Otra propiedad fundamental de la varianza es la siguiente:

Proposición 2.3 *Dados r grupos, cada uno de ellos formado por n_i observaciones de media $\overline{x_i}$ y de varianza S_i^2. Entonces la varianza,*
S^2, del conjunto de todas las $n = n_1, n_2, n_3, \cdots, n_r$ observaciones vale

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{r} n_i S_{n_i}^2 + \frac{1}{n}\sum_{i=1}^{r} n_i \left(\overline{x_i} - \overline{x}\right)^2$$

Proposición 2.4 *Además de las propiedades que hemos demostrado sobre la varianza (y por tanto sobre la desviación típica), será conveniente tener siempre en mente otras que enunciamos a continuación:*

- Ambas son sensibles a la variación de cada una de las puntuaciones, es decir, si una puntuación cambia, cambia con ella la varianza. La razón es que si miramos su definición, la varianza es función de cada una de las puntuaciones.

- Si se calculan a través de los datos agrupados en una tabla, dependen de los intervalos elegidos. Es decir, cometemos cierto error en el cálculo de la varianza cuando los datos han sido resumidos en una tabla estadística mediante intervalos, en lugar de haber sido calculados directamente como datos no agrupados. Este error no será importante si la elección del número de intervalos, amplitud y límites de los mismos ha sido adecuada.

- La desviación típica tiene la propiedad de que en el intervalo

$$(\overline{x} - 2S_n, \overline{x} + 2S_n) = \overline{x} \pm 2S_n$$

se encuentra, al menos, el 75 % de las observaciones (véase más adelante el teorema de Thebycheff). Incluso si

tenemos muchos datos y estos provienen de una distribución normal (se definirá este concepto más adelante), podremos llegar al 95 %

- No es recomendable el uso de ellas, cuando tampoco lo sea el de la media como medida de tendencia central.

Grados de libertad

Los grados de libertad de un medida calculado sobre n datos se refieren al número de cantidades independientes que se necesitan en su cálculo, menos el número de restricciones que ligan a las observaciones y la medida.

Ilustrándolo con un ejemplo. Consideraremos una serie de valores de una variable,

$X \rightsquigarrow 2, 5, 7, 9, 12$

que han sido tomados de forma independiente.

Su media es $\overline{x} = 7$ y se ha calculado a partir de las $n = 5$ observaciones independientes x_i, que están ligadas a la media por la relación:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i$$

Luego el número de grados de libertad de la media es $n - 1 = 4$.

$$\frac{(x_i - \overline{x})^2}{n}$$

Sin embargo esas cantidades no son totalmente independientes, pues están ligadas por una restricción:

$$\sum_{i=1}^{n}\left(x_i - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)\right) = 0$$

El número de grados de libertad de la medida es el número de observaciones de la variable menos el número de restricciones que se verifican, así que en este caso, los grados de libertad de la varianza sobre los $n = 5$ datos son también $n - 1 = 4$.

Un principio general de la teoría matemática nos dice que si pretendemos calcular de modo aproximado la varianza de una población a partir de la varianza de una muestra suya, se tiene que el error cometido es generalmente más pequeño, si en vez de considerar como estimación de la varianza de la población, a la varianza muestral

$$S_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

consideramos lo que se denomina cuasivarianza muestral, S^2 que se calcula como la anterior, pero cambiando el denominador por el número de grados de libertad, $n - 1$:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{nS_n^2}{n-1}$$

Sobre este punto insistiremos más adelante, ya que es fundamental en estadística inferencial.

Tipificación

Se conoce por tipificación al proceso de restar la media y dividir por su desviación típica a una variable X. De este modo se obtiene una nueva variable

$$Z = \frac{X - \overline{x}}{S}$$

de media $\overline{z} = 0$ y desviación típica, $S_Z = 1$. que denominamos variable tipificada.

Esta nueva variable carece de unidades y permite hacer comparables dos medidas que en un principio no lo son, por aludir a conceptos diferentes. Así por ejemplo nos podemos preguntar si un elefante es más grueso que una hormiga determinada, cada uno en relación a su población. También es aplicable al caso en que se quieran comparar individuos semejantes de poblaciones diferentes. Por ejemplo si deseamos comparar el nivel académico de dos estudiantes de diferentes Universidades para la concesión de una beca de estudios, en principio sería injusto concederla directamente al que posea una nota media más elevada, ya que la dificultad para conseguir una buena calificación puede ser mucho mayor en un centro que en el otro, lo que limita las posibilidades de uno de los estudiante y favorece al otro. En este caso, lo más correcto es comparar las calificaciones de ambos estudiantes, pero tipificadas cada una de ellas por las medias y desviaciones típicas respectivas de las notas de los alumnos de cada Universidad.

Coeficiente de variación

Hemos visto que las medidas de centralización y dispersión nos dan información sobre una muestra. Nos podemos preguntar si tiene sentido usar estas magnitudes para comparar dos poblaciones. Por ejemplo, si nos piden comparar la dispersión de los pesos de las poblaciones de elefantes de dos circos diferentes, S nos dará información útil.

¿Pero qué ocurre si lo que comparamos es la altura de unos elefantes con respecto a su peso? Tanto la media $\overline{x}$, como la desviación típica, S, se expresan en las mismas unidades que la variable. Por ejemplo, en la variable altura podemos usar como unidad de longitud el *metro* y en la variable peso, el kilogramo. Comparar una desviación (con respecto a la media) medida en *metros* con otra en *kilogramos* no tiene ningún sentido.

El problema no deriva sólo de que una de las medidas sea de longitud y la otra sea de masa. El mismo problema se plantea si medimos cierta cantidad, por ejemplo la masa, de dos poblaciones, pero con distintas unidades. Este es el caso en que comparamos el peso en *toneladas* de una población de 100 elefantes con el correspondiente en *miligramos* de una población de 50 hormigas.

El problema no se resuelve tomando las mismas escalas para ambas poblaciones. Por ejemplo, se nos puede ocurrir medir a las hormigas con las mismas unidades que los elefantes (toneladas). Si la ingeniería genética no nos sorprende con alguna barbaridad, lo lógico es que la dispersión de la variable peso de las hormigas sea prácticamente nula (¡Aunque haya algunas que sean 1.000 veces mayores que otras!)

En los dos primeros casos mencionados anteriormente, el problema viene de la dimensionalidad de las variables, y en el tercero de la diferencia enorme entre las medias de ambas poblaciones. El coeficiente de variación es lo que nos permite evitar estos problemas, pues elimina la dimensionalidad de las variables y tiene en cuenta la proporción existente entre medias y desviación típica. Se define del siguiente modo:

$$Cv = \frac{S_X}{\overline{x}}$$

Basta dar una rápida mirada a la definición del coeficiente de variación, para ver que las siguientes consideraciones deben ser tenidas en cuenta:

∘ Sólo se debe calcular para variables con todos los valores positivos. Todo índice de variabilidad es esencialmente no negativo. Las observaciones pueden ser positivas o nulas, pero su variabilidad debe ser siempre positiva. De ahí que sólo debemos trabajar con variables positivas, para la que tenemos con seguridad que $\overline{x} > 0$.

∘ No es invariante ante cambios de origen. Es decir, si a los resultados de una medida le sumamos una cantidad positiva, $b > 0$, para tener $Y = X + b$, entonces, $C_{vY} < C_{vX}$ ya que la desviación típica no es sensible ante cambios de origen, pero si la media. Lo contrario ocurre si restamos ($b < 0$).

$$C_{vY} = \frac{S_Y}{\overline{y}} = \frac{S_X}{\overline{x} + b} < \frac{S_X}{\overline{x}} = C_{vX}$$

∘ Es invariante a cambios de escala. Si multiplicamos X por una constante a, para obtener,$Y = aX$, entonces

Es importante destacar que los coeficientes de variación sirven para comparar las variabilidades de dos conjuntos de valores (muestras o poblaciones), mientras que si deseamos comparar a dos individuos de cada uno de esos conjuntos, es necesario usar los valores tipificados.

Dada la distribución de edades (medidas en años) en un colectivo de 100 personas, obtener:

1.

La variable tipificada Z.

2.

Valores de la media y varianza de Z.

3.

Coeficiente de variación de Z.

Horas trabajadas

Horas trabajadas	*N° de empleados*
$0-4$	47
$4-10$	32
$10-20$	17
$20-40$	4
	100

Para calcular la variable tipificada

$$Z = \frac{X-\overline{x}}{S_X}$$

partimos de los datos del enunciado. Será necesario calcular en primer lugar la media y desviación típica de la variable original ($X = años$).

Intervalos	m_i	n_i	$m_i n_i$	$m_i^2 n_i$
$0-4$	2	47	94	188
$4-10$	7	32	224	1568
$10-20$	15	17	255	3825
$20-40$	30	4	120	
Total			693	9181

$$\overline{x} = \frac{693}{100} = 6,93\, años$$
$$S_X^2 = \frac{9181}{100} - 6,93^2 = 43,78\, (años)^2$$
$$S_X = \sqrt{43,78} = 6,6\, años$$

A partir de estos valores podremos calcular los valores tipificados para las marcas de clase de cada intervalo y construir su distribución de frecuencias:

$$z_1 = \frac{2 - 6,93}{6,6} = -0,745$$
$$z_2 = \frac{7 - 6,93}{6,6} = 0,011$$
$$z_2 = \frac{15 - 6,93}{6,6} = 1,22$$
$$z_2 = \frac{30 - 6,93}{6,6} = 3,486$$

z_i	n_i	$z_i n_i$	$z_i^2 n_i$
$-0,745$	47	$-35,015$	26,086
0,011	32	0,352	0,004
1,220	17	20,720	25,303
3,486	4	13,944	48,608
	$n = 100$	0,021	100,002

$$\overline{z} = \frac{0,021}{100} \approx 0$$
$$S_Z^2 = \frac{100,02}{100} - 0^2 \approx 1$$
$$S_X = \sqrt{1} = 1$$

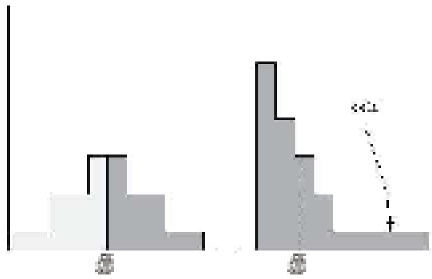
A pesar de que no se debe calcular el coeficiente de variación sobre variables que presenten valores negativos (y Z los presenta), lo calculamos con objeto de ilustrar el porqué:

$$C_v = \frac{S_Z}{\overline{z}} = \frac{1}{0} = \infty$$

Es decir, el coeficiente de variación no debe usarse nunca con variables tipificadas.

2.4.3. Asimetría y apuntamiento

Sabemos cómo calcular valores alrededor de los cuales se distribuyen las observaciones de una variable sobre una muestra y



sabemos cómo calcular la dispersión que ofrecen los mismos con respecto al valor de central. Nos proponemos dar un paso más allá en el análisis de la variable. En primer lugar, nos vamos a plantear el saber si los datos se distribuyen de forma simétrica con respecto a un valor central, o si bien la gráfica que representa la distribución de frecuencias es de una forma diferente del lado derecho que del lado izquierdo.

Si la simetría ha sido determinada, podemos preguntarnos si la curva es más o menos apuntada (larga y estrecha). Este apuntamiento habrá que medirlo comparado a cierta distribución de frecuencias que consideramos normal (no por casualidad es éste el nombre que recibe la distribución de referencia).

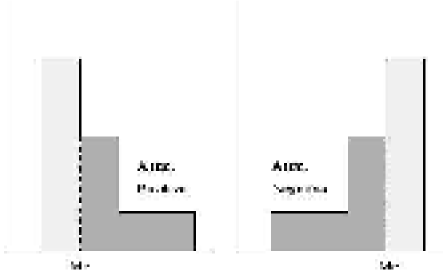
Estas ideas son las que vamos a desarrollar en lo que resta del capítulo.¨

Índices de asimetría o sesgo

Para saber si una distribución de frecuencias es simétrica, hay que precisar con respecto a qué. Un buen candidato es la mediana, ya que para variables continuas, divide al histograma de frecuencias en dos partes de igual área. Podemos basarnos en ella para, de forma natural, decir que una distribución de frecuencias es simétrica si el lado derecho de la gráfica (a partir de la mediana) es la imagen por un espejo del lado izquierdo(figura).

Cuando la variable es discreta, decimos que es simétrica, si lo es con respecto a la media.

2.4.4.



- Se podría pensar que definir la simetría usando la mediana para variables continuas y usando la media para variables discretas es una elección arbitraria. En realidad esto no es así, pues si una variable es continua, coinciden ambos criterios de simetría (con respecto a la media y a la mediana). Es más, se tiene que media y mediana coinciden para distribuciones continuas simétricas. Por otro lado,
- En el caso de variables discretas, la distribución es simétrica si el lado derecho del diagrama se obtiene por imagen especular desde la media. En este caso coincide la media con la mediana si el número de observaciones es impar.
- Si la variable es continua simétrica y unimodal, coinciden la media, la mediana y la moda.

Dentro de los tipos de asimetría posible, vamos a destacar los dos fundamentales (2.4.4):

Asimetría positiva:

Si las frecuencias más altas se encuentran en el lado izquierdo de la media, mientras que en derecho hay frecuencias más pequeñas (cola).

Asimetría negativa:

Cuando la cola está en el lado izquierdo.

Cuando realizamos un estudio descriptivo es altamente improbable que la distribución de frecuencias sea totalmente simétrica.

En la práctica diremos que la distribución de frecuencias es simétrica si lo es de un modo aproximado. Por otro lado, aún observando cuidadosamente la gráfica, podemos no ver claro de qué lado están las frecuencias más altas. Conviene definir entonces unos estadísticos que ayuden a interpretar la asimetría, a los que llamaremos índices de asimetría, y que denotaremos mediante A_s. Vamos a definir a continuación algunos de los índices de asimetría más usuales como son el índice basado en los tres cuartiles, el momento de tercer orden y la distancia entre la moda y la media o la media y la mediana.

Índice basado en los tres cuartiles (Yule-Bowley)

Si una distribución es simétrica, es claro que deben haber tantas observaciones entre la que deja por debajo de sí las tres cuartas partes de la distribución y la mediana, como entre la mediana y la que deja por debajo de sí un cuarto de todas las observaciones. De forma abreviada esto es,

$$Q_3 - Q_2 = Q_2 - Q_1$$

Una pista para saber si una distribución de frecuencias tiene una asimetría positiva, es observando la figura):

$$Q_3 - Q_2 > Q_2 - Q_1$$

Por analogía, si es asimétrica negativa, se tendrá

$$Q_3 - Q_2 < Q_2 - Q_1$$

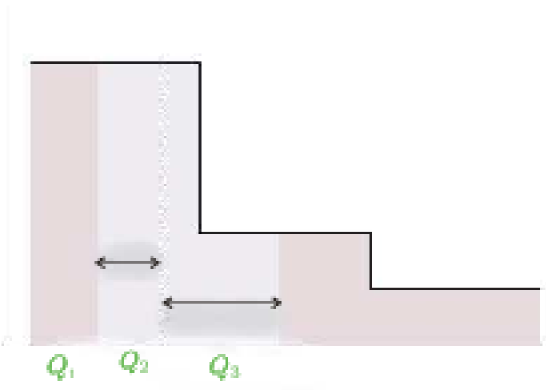
Para quitar dimensionalidad al problema, utilizamos como índice de asimetría la cantidad, llamada índice intercuartílico

$$A_s = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} \tag{2.8}$$

Es claro que

$$1 \geq \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1) + (Q_2 - Q_1)} \geq -1$$

El número obtenido, A_s, es invariante ante cambios de origen de referencia y de escala.



2.4.5. Índice basado en el momento central de tercer orden

Sea X una variable cuantitativa y Llamamos momento de orden p a:

$$\mu_p = \frac{1}{n}\sum_{i=1}^{n} x_i^p$$

Se denomina momento central de orden p a la cantidad

$$m_p = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^p$$

Si los datos están agrupados en una tabla, m_p admite otra expresión equivalente:

$$m_p = \frac{1}{n}\sum_{i=1}^{n} n_i(x_i - \overline{x})^p$$

si $p = 1$ se tiene que $m_1 = 0$.

El momento de orden 2 es la varianza muestral:

$$m_2 = S^2$$

Es sencillo comprobar que los momentos de orden p impar, son siempre nulos en el caso de variables simétricas, ya que para cada i

que esté a un lado de la media, con,$(x_i - \overline{x}) < 0$, le corresponde una observación j del otro lado de la media tal que. $(x_i - \overline{x}) = (x_j - \overline{x})$ Elevando cada una de esas cantidades a p impar, y sumando se tiene que $m_p = 0$ si la distribución es simétrica

Si la distribución fuese asimétrica positiva, las cantidades,$(x_i - \overline{x})^p$ con, $p \geq 3$ impar positivas estarían muy aumentadas al elevarse a p. Esta propiedad nos indica que un índice de asimetría posible consiste en tomar $p = 3$y definir

$$A_s = a_3 = \frac{m_3}{m_{2\sqrt{m_2}}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{S^3} \tag{2.9}$$

que para datos organizados en una tabla sería

$$a_3 = \frac{\frac{1}{n}\sum_{i=1}^{n} n_i (x_i - \overline{x})^3}{S^3} \tag{2.10}$$

Apoyándonos en este índice, diremos que hay asimetría positiva si $a_3 > 0$, y que la asimetría es negativa si $a_3 < 0$.

Hemos dividido m_3 por el cubo de para S que a_3 sea un número abstracto sin dimensiones, independiente de la variabilidad de la variable. Por otro lado, la cantidad A_s definida por la relación (2.8) no es la misma que la definida en (2.9). Simplemente las notamos A_s para simbolizar que es un índice de asimetría.

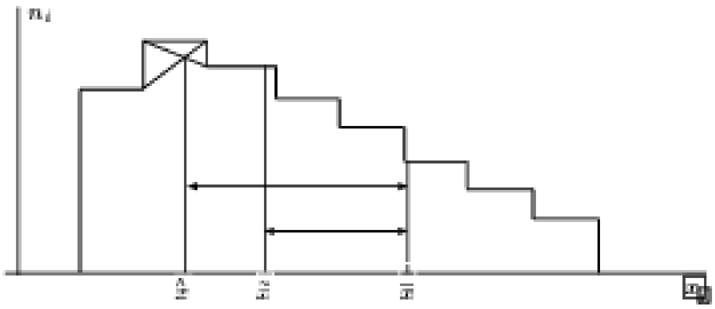
2.9.2.6 Otros índices de asimetría Basándonos en que si una distribución de frecuencias es simétrica y unimodal, entonces la media, la mediana y la moda coinciden, podemos definir otras medidas de asimetría, como son:

$$A_s = \frac{\overline{x} - \widehat{x}}{S} \tag{2.11}$$

o bien,

$$A_s = \frac{3(\overline{x} - \widetilde{x})}{S} \tag{2.12}$$

Diremos que hay asimetría positiva si $A_s < 0$ y negativa si $A_s > 0$ (véase la figura).



Ejemplo 2.10 *Las edades de un grupo de personas se reflejan en la tabla siguiente:*

Intervalos	n_i
7-9	*4*
9-11	*18*
11-12	*14*
12-13	*27*
13-14	*42*
14-15	*31*
15-17	*20*
17-19	*1*

Determinar la variabilidad de la edad mediante los estadísticos varianza, desviación típica, coeficiente de variación y rango intercuartílico. Estudie la simetría de la variable.

Solución 2.10.1 *En primer lugar realizamos los cálculos necesa-*

rios a partir de la tabla de frecuencias:

$Intervalos$	n_i	m_i	N_i	$m_i x_i$	$m_i^2 n_i$
$7-9$					
$9-11$					
$11-12$					
$12-13$					
$13-14$					
$14-15$					
$15-17$					
$17-19$					

La media es $\overline{x} = 13,15$ *años. La varianza la calculamos a partir de la columna de* $m_i^2 n_i$ *como sigue:*

$$S^2 = 27742,25/157 - 13,15^2 = 3,78\,años^2$$
$$S = \sqrt{3,78} = 1,94\,años$$

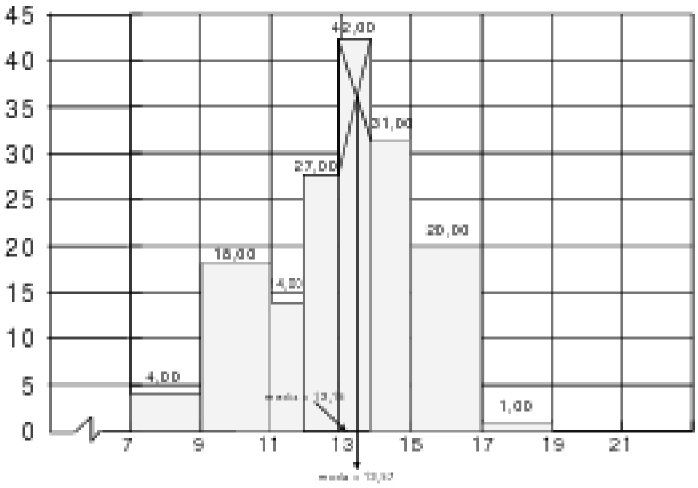
El coeficiente de variación no posee unidades y es:

$$C_v = \frac{1,94}{13,15} = 0,15$$
$$15\,\% \;\; de\,variabilidad$$

En lo que concierne a la simetría podemos utilizar el coeficiente de asimetría de Yule-Bowley, para el cual es preciso el cálculo de los cuartiles:

$$Q_1 = 12 + \frac{39,25-36}{27} \times 1 = 12,12$$
$$Q_2 = 13 + \frac{78,5-63}{42} \times 1 = 13,37$$
$$Q_3 = 14 + \frac{117,75-105}{31} \times 1 = 14,41$$

Lo que nos dice que aproximadamente en un rango de $Q_3 - Q_2 = 2,29$ *años se encuentra el 50 % central del total de observaciones*



Además:

$$A_s = \frac{(14,41 - 13,37) - (13,37 - 12,12)}{14,41 - 12,12} = -0,09$$

Este resultado nos indica que existe una ligera asimetría a la izquierda (negativa). Un resultado similar se obtiene si observamos que la distribución de frecuencias es unimodal, siendo la moda:

$$\widehat{x} = 13 + \frac{42 - 27}{(42 - 27) - (42 - 31)} \times 1 = 13,57$$

en cuyo caso podemos usar como medida del sesgo:

$$A_s = \frac{\overline{x} - \widehat{x}}{S} = -0,21$$

Medidas de apuntamiento o curtosis

Se define el coeficiente de aplastamiento de Fisher como:

$$\gamma = \frac{m_4}{\sigma^4} - 3 \qquad (2.13)$$

donde m_4 es el momento empírico de cuarto orden. Es éste un coeficiente adimensional, invariante ante cambios de escala y de origen. Sirve para medir si una distribución de frecuencias es muy apuntada o no. Para decir si la distribución es larga y estrecha, hay que tener un patrón de referencia. El patrón de referencia es la distribución normal o gaussiana para la que se tiene

$$\gamma = 0$$

De este modo, atendiendo a, γ se clasifican las distribuciones de frecuencias en

- Leptocúrtica:Cuando,

  $\gamma > 0$ o sea, si la distribución de frecuencias es más apuntada que la normal;

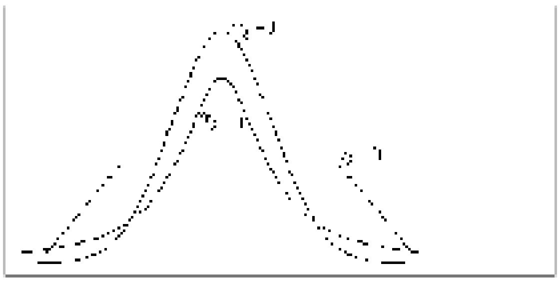
- Mesocúrtica:Cuando,

  $\gamma = 0$ es decir, cuando la distribución de frecuencias es tan apuntada como la normal;

- Platicúrtica:Cuando,

  $\gamma < 0$ o sea, si la distribución de frecuencias es menos apuntada que la normal;

  Estadísticos de tendencia central o posición nos indican donde se sitúa un grupo de puntuaciones. Los de variabilidad o dispersión nos indican si esas puntuaciones o valores están próximas entre sí o si por el contrario están o muy dispersas.



Tarea 1

1. a) *Calcule en los ejercicios de la tarea del capítulo 1 las siguientes medidas*
    1) *Tendencia central*
    2) *Posición*
    3) *Variabilidad*
    4) *Asimetría*
    5) *Apuntamiento*

   b) *Determine cuales medias representan mejor la muestra*

   c) *Interprete desde un punto de vista estadístico los resultados.*

2. *En un experimento de aprendizaje se registró el tiempo que los alumnos emplearon al resolver un problema de matemáticas* x_i : *en el primer intento,* y_i : *en el segundo intento.*

	i	1	2	3	4	5	6
Individuo	x_i	23,7	36,9	25,5	30,2	28,0	34,8
	y_i	15,0	25,4	21,0	22,3	25,2	28,8

   a) *Calcule la diferencia* $d = x_i - y_i$ *en cada caso y luego determine las medidas de*
    1) *Tendencia central*
    2) *Posición*
    3) *Variabilidad*
    4) *Asimetría*
    5) *Apuntamiento*

   b) *Determine lo mismo para* x_i *y* y_i

   c) *Compare*
    1) $\bar{x} - \bar{y}$ *con* $\bar{d}$
    2) $S_x^2 - S_y^2$ *con* S_d^2
    3) *Que puede concluir de la comparación*

   d) *Determine* $\sum_{i=1}^{6} \frac{(x_i - c)^2}{N}$ *donde c representa una medida de tendencia central diferente de la media aritmética*

e) *Comparando el resultado del inciso b. con los resultados del inciso d. para que valor de c se obtuvo un valor mínimo*

f) *Minimice la función*

$$f(c) = \sum_{i=1}^{n} \frac{(x_i - c)^2}{N}$$

¿Está de acuerdo este resultado con el del inciso anterior.?

g) *Si*

$$z = \frac{x_i - \bar{x}}{S}$$

es. Calcule $\bar{z}$ *y* S_z^2 *¿Qué puede concluir?*

h) *Realice los incisos del d. al g. para las variables* y, d *y* z

i) *Realice histogramas para las variables:* x, y, d *y* z*.* *Compare los resultados con los histogramas*

3. *calcule la media y la varianza a partir de los datos tabulados en la siguiente tabla*

Clases	n_i
$10 \leq x < 20$	121
$20 \leq x < 30$	165
$30 \leq x < 40$	184
$40 \leq x < 50$	173
$50 \leq x < 60$	142
$60 \leq x < 70$	120
$70 \leq x < 80$	118
$80 \leq x < 90$	110
$90 \leq x < 100$	90

4. *Calcule las medidas de asimetría y de apuntamiento para los*

datos de la siguiente muestra

Clases	n_i
$600 \leq x < 650$	*41*
$650 \leq x < 700$	*46*
$700 \leq x < 750$	*50*
$750 \leq x < 800$	*52*
$800 \leq x < 850$	*60*
$850 \leq x < 900$	*64*
$900 \leq x < 950$	*65*
$950 \leq x < 1000$	*66*
$1000 \leq x < 1050$	*70*

5. *Si el ingreso medio de 20 trabajadores es de 40000 dólares, ¿Cuál es su ingreso total?*

6. *Si la estatura media de una muestra de 25 jugadores de basketball es 6,9 pies, ¿cuál es la suma de la estatura de los 23 jugadores?*

7. *En un esfuerzo de reducir su consumo de café, un trabajador de oficina registra los números siguientes de tazas de café consumidas durante un periodo de 20 días*

4	*5*	*3*	*6*	*7*	*1*	*2*	*3*	*0*	*5*
6	*5*	*8*	*4*	*0*	*2*	*3*	*7*	*5*	*6*

¿Qué medida de tendencia central le servirá mejor a su propósito? ¿Explique por qué? ¿ Cuál es el valor numérico?

8. *En un investigación realizada por la secretaria de un médico para averiguar los tiempos de espera en minutos de los pacientes que acuden con el doctor, una muestra de pacientes de un día arrojó los resultados:*

35	*25*	*35*	*50*	*25*	*55*	*30*	*50*	*35*	*35*
5	*5*	*35*	*60*	*30*	*30*	*25*	*55*	*30*	*20*
60	*25*	*25*	*40*	*80*	*20*	*20*	*5*	*5*	*10*

a) *Describa un tiempo típico de espera usando la media.*

b) *Describa un tiempo típico de espera usando la mediana.*

c) *¿Cuál medida, media, moda o mediana, se considera usted que es más representativa del conjunto de datos? Explique.*

d) *Determine los tres cuartiles*

e) *Determine los deciles*

f) *Realice un diagrama de caja*

g) *Determine los índices de asimetría y de apuntamiento*

h) *Realice una inferencia sustentada en los resultados.*

9. *Se escogió una muestra de 705 conductores de autobús y se registró en la tabla siguiente el número de accidentes de tránsito que tuvieron durante cuatro años.*

Número de accidentes	*Frecuencia*
0	144
1	157
2	158
3	115
4	78
5	44
6	21
7	7
8	6
10	3
11	1

a) *¿Cuál es la moda?*

b) *Calcule la media*

c) *Determine la mediana, analítica y geométricamente*

d) *Determine el rango medio*

e) *Calcule el sesgo*

f) *?Cuál medida de tendencia central usaría para determinar el valor central? Explique*

g) *Calcule los cuartiles*

h) *calcule los deciles*

i) *Realice un diagrama de caja*

10. *Si 20 puntajes tienen una media de 15 y 30 puntajes , una media de 30, ¿cuál es la media del grupo total de 50 puntajes?*

11. *La media armónica que a menudo se utiliza para promediar velocidades desarrolladas en distancias iguales. A continuación se presentan se darán las velocidades (en millas por hora) de un automóvil que viaja a velocidad constante cada 20 millas*

    30 60 40 60 30 20 25

    ¿Cuál es la rapidez promedio en el viaje de 140 millas?

12. *Con los datos del ejercicio anterior determine:*

    a) *La media aritmética*

    b) *La media geométrica*

    c) *La media cuadrática*

    d) *Escoja una y explique por qué*

13. *Si la desviación estándar de un conjunto de datos es 0, ¿qué puede afirmarse de dicho conjunto?*

14. *¿Será posible que el rango y la varianza sean iguales ? explique*

15. *¿Puede ser la varianza negativa?*

16. *Suponga que una muestra tiene una media de 25 y una desviación estándar de 3.3 Aproxime un intervalo que contenga el 90 % de los datos*

17. *¿Qué efecto tiene sobre la desviación el tamaño de la muestra?*

Índice general

Capítulo 3

Teoría de probabilidades

3.1. Introducción

Si el único propósito del investigador es describir los resultados de un experimento concreto, los métodos descriptivos analizados en los capítulos anteriores pueden considerarse suficientes. No obstante, si lo que se pretende es utilizar la información obtenida para extraer conclusiones generales sobre todos aquellos objetos del tipo de los que han sido estudiados, entonces estos métodos constituyen sólo el principio del análisis, y debe recurrirse a métodos de inferencia estadística, los cuales implican el uso de una de las ramas de la matemática, llamada teoría de la probabilidad.

Analizaremos en este párrafo la noción de probabilidad y la terminología subyacente a esta área de las matemáticas, ya que la probabilidad constituye por sí misma un concepto básico que refleja su relación con la faceta del mundo exterior que pretende estudiar: los fenómenos aleatorios, los cuales obedecen unas ciertas reglas de comportamiento.

El concepto de probabilidad es importante cuando se estudian procesos físicos, biológicos, químicos, sociales que generan observaciones que no es fácil o factible predecir con exactitud, pero se puede determinar la frecuencia relativa con que ocurren con cierta precisión, si realizamos un gran número de observaciones.

Los eventos que poseen esta propiedad se denominan eventos aleatorios estocásticos y la frecuencia relativa con que ocurren es una interpretación intuitiva de la probabilidad, pero no nos proporciona una definición exacta de ella.

En su definición y en sus aplicaciones nos dedicaremos en este capítulo.

3.2. Experimentos y sucesos aleatorios

Definición 3.1 *Diremos que un experimento es aleatorio si se verifican las siguientes condiciones,*

1. *Se puede repetir indefinidamente, siempre con las mismas condiciones.*
2. *Antes de realizarlo no se puede predecir el resultado.*
3. *El resultado "e" que se obtiene pertenece a un conjunto de resultados posibles conocido previamente el cual llamaremos espacio muestral y lo denotaremos con la letra Ω o S. Los elementos del espacio muestral se denominan sucesos elementales.*

Es decir si $e_1, e_2 \in S \Longrightarrow e_1, e_2$ son sucesos elementales o puntos muestrales.En otras palabras: Un suceso es elemental si su ocurrencia o no ocurrencia no está relacionada con ningún otro suceso.

Cualquier subconjunto A de S se denomina suceso o evento aleatorio.

El espacio muestral puede ser de dos tipos:

- Discreto si está formado por un conjunto finito o numerable de resultados.
- Continuo si está compuesto por un conjunto no numerable de elementos.

Definición 3.2 (Suceso determinista) *Se denomina experimento determinista a el experimento que al realizarlo varias veces con las mismas condiciones iniciales obtenemos siempre el mismo resultado*

Ejemplo 3.1 *Si dejamos caer dos cuerpos de diferentes masas y a la misma altura en el vacío, los dos cuerpos caerán al mismo tiempo.*

Cuando en un experimento no se puede predecir el resultado final, decimos que el experimento es aleatorio

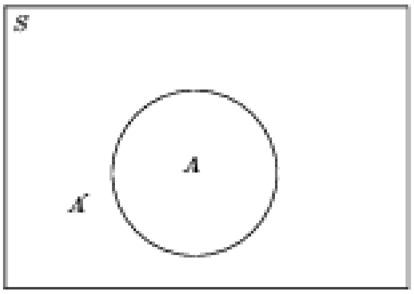
Ejemplo 3.2 *En una ruleta legal no se sabe que número va a salir*

3.2.1. Tipos de eventos

1. Evento seguro es aquel que siempre se verifica después del experimento aleatorio, es decir S.

2. Evento imposible es aquel que nunca se verifica como resultado del experimento aleatorio y como un evento debe ser un subconjunto de S el suceso imposible es el conjunto vacío ϕ.

3. Evento complementario. Se denomina suceso complementario de un suceso A al que se verifica si no se verifica A es decir El complemento de A con respecto a S es el conjunto de todos los elementos de S que no están en A. Denotaremos el complemento del conjunto A como A'



Eventos complementarios

Ejemplo 3.3 *Si realizamos el experimento aleatorio de lanzar un dado al aire. Determinar el conjunto de eventos posibles*

Solución 3.3.1 *En este caso tenemos el espacio muestral es*
$S = \{1, 2, 3, 4, 5, 6\}$
y los sucesos elementales son
$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$

$$\textit{El conjunto de eventos posibles es } \mathcal{A} = \begin{cases} \phi \text{ suceso imposible} \\ S \text{ suceso seguro} \\ \{1,2,3\} \\ \{4,5\} \\ \{2,4,6\} = \{1,2,3\}' \\ \vdots \end{cases}$$

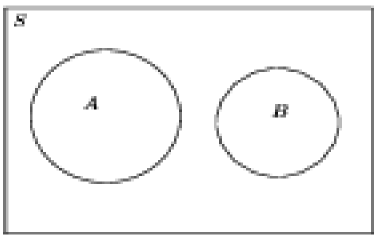
3.3. Operaciones básicas con eventos aleatorios

Como los eventos son en realidad conjuntos podemos aplicarles el álgebra de conjuntos tratada en su curso de álgebra elemental visto en I semestre.

1. Intersección. La intersección de dos eventos A y $B \subset S$ que se representa $A \cap B$, es el evento que contiene a todos los sucesos elementales comunes de A y B, es decir

$$A \cap B = \{e \in S : e \in A \wedge e \in B\}$$

2. Dos eventos son mutuamente excluyentes o disjuntos si $A \cap B = \phi$, esto es si A y $B \subset S$ no tienen elementos comunes.



Eventos disjuntos

3. La unión. Dados dos sucesos aleatorios A y B, se denomina suceso unión de A y $B \subset S$ al evento formado por todos los sucesos elementales que pertenecen a A o bien a B, es decir

$$A \cup B = \{e \in S : e \in A \vee e \in B\}$$

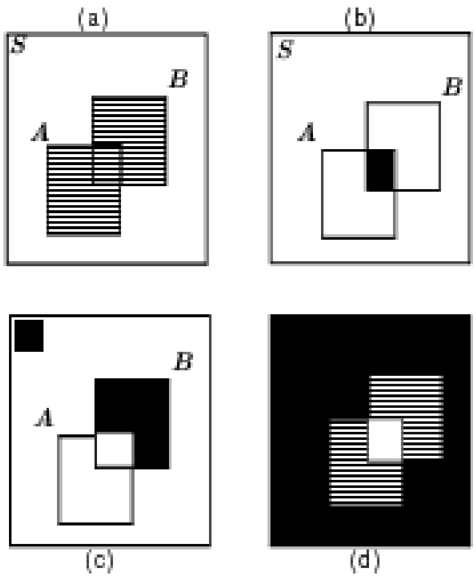
4. Dados dos sucesos aleatorios $A, B \subset S$, se llama diferencia de A y B, y se representa $A - B$ o $A \backslash B$, al suceso aleatorio formado por todos los sucesos elementales que pertenecen a A, pero no a B, es decir

$$A \backslash B = \{e \in S : e \in A \wedge e \notin B\} = A \cap B'$$

   Donde a B' se le llama complemento de B y se define $S \backslash B$

5. Diferencia simétrica. Si $A, B \subset S$ se denomina evento diferencia simétrica de A y B y se denota por $A \triangle B$ al evento formado por todos los sucesos elementales que pertenecen a A y no a B, y los que están en B y no en A es decir,

$$A \triangle B = \{e \in S : e \in (A \cup B) \wedge e \notin (A \cap B)\}$$



Dados dos eventos $A, B \subset S$ se presenta en: (a) $A \cup B$; en (b) $A \cap B$; en (c) $A - B$; en (d) $A \triangle B$.

Hay algunas propiedades importantes que valen la pena recordar

- Leyes de Morgan

i $(A \cup B)' = A' \cap B'$

ii $(A \cap B)' = A' \cup B'$

- $A \triangle B = (A \backslash B) \cup (B \backslash A) = (A \cup B) \backslash (A \cap B)$
- $A \cap A' = \phi$
- $A \cap B' = A \backslash B$
- $A' = S \backslash A$
- $A \cup A' = S$

Definición 3.3 *Sea $A_1, A_2, A_3, \cdots A_n$ una serie de eventos se tiene,*
◦ *El evento formado por todos los sucesos comunes y no comunes de la serie es,*

$$A_1 \cup A_2 \cup A_3 \cdots \cup A_n = \cup_{i=1}^{n} A_i$$

◦ *El evento formado por los sucesos comunes entre los eventos de la serie es el evento,*

$$A_1 \cap A_2 \cap A_3 \cdots \cap A_n = \cap_{i=1}^{n} A_i$$

Definición 3.4 *Sea A un conjunto finito o enumerable, Se le denomina $n(A)$, al cardinal de A, es decir el número de elementos que posee el conjunto A.*

3.4. Probabilidad

3.4.1. Probabilidad estocástica

En los experimentos aleatorios observamos que si conservamos las mismas condiciones iniciales al aumentar el números de ensayos la frecuencia relativa con la que ocurre el suceso A es,

$$f_n(A) = \frac{número\, de\, veces\, que\, ocurre\, A}{n}$$

y tiende a converger hacia un valor, el cual se denomina probabilidad de A

$$P(A) = \lim_{n \to \infty} f_n(A)$$

Esta noción de probabilidad no se puede aplicar en la práctica ya que:

- Se requiere realizar un número infinito de veces el experimento para calcular la probabilidad
- Los experimentos aleatorios a veces no pueden ser realizados en la práctica por ejemplo. Si se quiere saber cuantas personas mueren al fumigar con un tóxico

3.4.2. Probabilidad de laplace

Si un experimento cualquiera se puede repetir obteniendo un número finito de resultados posibles y no existe ninguna razón para pensar que un resultado tiene privilegios sobre otro, se calcula la probabilidad del suceso A según la regla de Laplace

$$P(A) = \frac{número\, de\, casos\, favorables\, para\, A}{número\, de\, casos\, posibles} = \frac{n(A)}{n(S)}$$

Ejemplo 3.4 *Calcular la probabilidad de que al lanzar un dado se obtenga un número par*

Solución 3.4.1 *El espacio muestral es* $S = \{1, 2, 3, 4, 5, 6\}$ *llamaremos* A *al suceso que da como resultado un número par el lanzamiento del dado,* $A = \{2, 4, 6\}$*, si suponemos que el dado es legal, es decir ninguna cara tiene privilegio sobre otra para salir, entonces*

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

3.4.3. Definición axiomática de la probabilidad

Como en toda rama de las matemáticas debemos establecer una serie de axiomas y definiciones básicas

Definición 3.5 *Sea* $\mathcal{A}$ *una clase no vacía formada por subconjuntos de* S*, diremos que la clase* $\mathcal{A}$ *es un* $\sigma -$ *álgebra de eventos si los eventos complementarios de cada elemento de* $\mathcal{A}$ *está también en* $\mathcal{A}$ *y además las uniones numerables de eventos de* $\mathcal{A}$ *es elemento de* $\mathcal{A}$ *es decir.*
Se dice que una clase $\mathcal{A}$ *no vacía de conjuntos* S *de es una* $\sigma -$ *álgebra si y sólo si cumple las siguientes propiedades*

$$S \in \mathcal{A}$$
$$\forall A \in \mathcal{A} \Longrightarrow A' \in \mathcal{A}$$
$$\forall A_1, A_2, A_3, \cdots \in \mathcal{A} \Longrightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$$

Definición 3.6 ***Definición axiomática de probabilidad***
Dado un espacio muestral S*, y una* $\sigma -$ *álgebra de sucesos* $\mathcal{A}$ *sobre él, diremos que* P *es una probabilidad sobre* $\mathcal{A}$ *si cumple las siguientes propiedades*

A$_1$ *La probabilidad es una función definida sobre* $\mathcal{A}$*, que toma solo valores positivos comprendidos entre 0 y 1, es decir*

$$\begin{array}{rcl} P : \mathcal{A} & \rightarrow & [0,1] \subset IR \\ A \in \mathcal{A} & \mapsto & 1 \geq P(A) \geq 0 \end{array}$$

$\mathbf{A}_2$ *La probabilidad del suceso seguro es 1*

$$P(S) = 1$$

$\mathbf{A}_3$ *Para cualquier sucesión infinita $A_1, A_2, A_3, \cdots$ de sucesos disjuntos de $\mathcal{A}$ se tiene que la probabilidad de el evento $\bigcup_{i=1}^{\infty} A_i$ es la serie infinita de las probabilidades, es decir*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Ejemplo 3.5 *Cuando el conjunto S formado por todos los posibles resultados de un experimento es finito o infinito enumerable podemos considerar que la clase $\mathcal{A}$ es el conjunto denominado **partes de** S*

$$\mathbb{P}(S) = \{A | A \subseteq S\}$$

el cual es un $\sigma -$ álgebra.

Ejemplo 3.6 *Consideremos el experimento de lanzamiento de un dado, entonces*

$$S = \{1,2,3,4,5,6\}$$
$$\mathbb{P}(S) = \{\phi, S, \{1\}, \{2\}, \cdots, \{6\}, \{1,2\}, \{1,3\}, ...\}$$

Es decir $\mathbb{P}(S)$ tiene $2^6 = 64$ elementos

Ejemplo 3.7 *Cuando S es infinito no enumerable por ejemplo queremos realizar el experimento de esperar el tiempo necesario para que el polietileno se desintegre naturalmente, observamos que $S = I\!R+$ y consideramos $\mathcal{A}$ como el conjunto de intervalos abiertos o cerrados, y sus uniones finitas, es decir*

$$\mathcal{A} = \{\phi, I\!R^+, [2,2], (2,3), (4,5] \cup [8,+\infty), ...\}$$

De acuerdo con el nivel de estas notas no trabajaremos con $\sigma -$ *álgebras* diferentes a las indicadas en los ejemplos 3.6 y 3.7[1], por lo que de aquí en adelante suponemos que estas son las clases $\mathcal{A}$

[1] El lector interesado en $\sigma - \acute{a}\lg ebras$ puede remitirse a Real Analysis probability de Robert B. Ash

Teorema 3.1 $P(\phi) = 0$

Demostración. consideremos la sucesión infinita de eventos $A_1, A_2, A_3, \cdots \in \mathcal{A}$ los cuales son todos vacíos y además disjuntos ya que $\phi \cap \phi = \phi$, de acuerdo con esto tenemos que $\bigcup_{i=1}^{\infty} A_i = \phi$ al aplicar A_3 tenemos

$$P(\phi) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

y el único real que cumple esta propiedad es el cero por lo que $P(\phi) = 0$ ■

Teorema 3.2 *Para cualquier sucesión finita de n eventos disjuntos $A_1, A_2, A_3, \cdots A_n \in \mathcal{A}$*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

Demostración. Consideremos la sucesión infinita de eventos $A_1, A_2, A_3, \cdots \in \mathcal{A}$ en los cuales hay $A_1, A_2, A_3, \cdots A_n$ eventos disjuntos y los $A_{n+1}, A_{n+2}, A_{n+3}, \cdots$ son vacíos. Entonces $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{n} A_i$. Por A_3 tenemos

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) \\ &= \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(A_i) \\ &\sum_{i=1}^{n} P(A_i) + 0 \\ &\sum_{i=1}^{n} P(A_i) \end{aligned}$$

Teorema 3.3 *para cualquier suceso $A \in \mathcal{A}$ se tiene que $P(A') = 1 - P(A)$*

■

Demostración. Como $A \in \mathcal{A} \Longrightarrow A' \in \mathcal{A}$ y además $A \cup A' = S$ y $P(S) = 1$ tenemos por el teorema 4 que

$$P(A \cup A') = P(A) + P(A') = P(S)$$

pero de acuerdo con A_2 $P(S) = 1$, entonces

$$\begin{aligned} P(A) + P(A') &= 1 \\ P(A') &= 1 - P(A) \end{aligned}$$

■

Teorema 3.4 *Si $A, B \in \mathcal{A}$ y $A \subset B$, entonces $P(B) \geq P(A)$*

Demostración. Como se ilustra en la fig 3.1 El evento B lo podemos escribir como $B = A \cup (B \cap A')$ entonces

$$\begin{aligned} P(B) &= P(A \cup (B \cap A')) \\ &= P(A) + P(B \cap A') \end{aligned}$$

y del A_1 $P(B \cap A') \geq 0$ de lo que se deduce $P(B) \geq P(A)$ ■

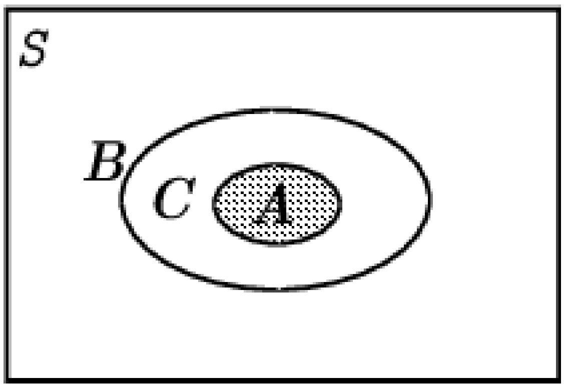


Figura 3.1 $B = A \cup (B \cap A')$ donde
$C = (B \cap A')$

Teorema 3.5 *Para dos sucesos $A, B \subset \mathcal{A}$ cualesquiera*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Como se ilustra en la fig 3.2

$$A \cup B = (A \cap B') \cup (A \cap B) \cup (A' \cap B)$$

como los eventos $(A' \cap B), (A \cap B')$ y $(A \cap B)$ son disjuntos podemos aplicar el teorema 3.1

$$\begin{aligned} P(A \cup B) = & \\ & P(A \cap B') + P(A \cap B) + P(A' \cap B) \end{aligned}$$

pero en la gráfica se observa que

$$\begin{gathered} A = (A \cap B') \cup (A \cap B) \\ y \\ B = (A \cap B) \cup (A' \cap B) \end{gathered}$$

al aplicar el teorema 5 queda

$$\begin{aligned} P(A) &= P(A \cap B') + P(A \cap B) \\ P(B) &= P(A \cap B) + P(A' \cap B) \end{aligned}$$

al despejar y sustituir obtenemos lo que queremos

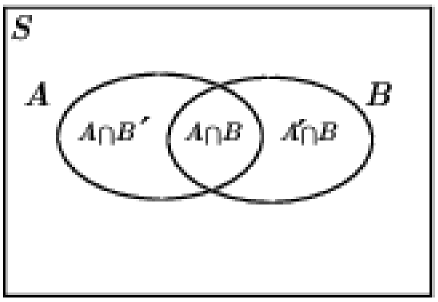


Figura 3.2 Partición de $A \cup B$

Teorema 3.6 *Sean $A_1, A_2, A_3, \cdots A_n \in \mathcal{A}$ una sucesión finita de eventos se tiene*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j=2}^{n} P(A_i \cap A_j) + \\ + \sum_{i<j<r=3}^{n} P(A_i \cap A_j \cap A_r) + \\ + \cdots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

Ejemplo 3.8 *Un estudiante de clase puede ser hombre o mujer. Si la probabilidad de que un hombre sea seleccionado es 0.3 ¿Cuál es la probabilidad de que sea seleccionada una mujer?*

Solución 3.8.1 *Sea A el evento de que se seleccione un hombre de la clase y sea B el evento de seleccionar una mujer, entonces es obvio que $S = A \cup B$, y además $A \cap B = \phi$, aplicando el teorema 7*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

y de los teorema 4 y A_2 obtenemos

$$\begin{aligned} 1 &= P(A) + P(B) \\ 1 &= 0,3 + P(B) \\ P(B) &= 0,7 \end{aligned}$$

Ejemplo 3.9 *Se selecciona una bola de una urna que contiene bolas rojas, blancas, azules, amarillas y verdes. Si la probabilidad de seleccionar una bola roja es $\frac{1}{5}$ y la de seleccionar una blanca es $\frac{2}{5}$ ¿Cuál es la probabilidad de seleccionar una bola azul, amarilla o blanca?*

Solución 3.9.1 *Sean $A_1, A_2, A_3, A_4, A_5, A_6$ los eventos de seleccionar de la urna una bola roja, blanca, azul, amarilla y verde respectivamente. Nos dicen que $P(A_1) = \frac{1}{5}$ y $P(A_2) = \frac{2}{5}$, entonces*

$$P(A_3 \cup A_4 \cup A_5 \cup A_6) = P(B) = ?$$

Como los eventos son disjuntos

$$S = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6$$

y por A_2 y el th.3

$$1 = P(A_1) + P(A_2) + P(B)$$
$$P(B) = 1 - \frac{1}{5} - \frac{2}{5} = \frac{2}{5}$$

Ejemplo 3.10 *Si la probabilidad de que un estudiante A pierda un examen de estadística es de 0,5, la probabilidad de que un estudiante B pierda el mismo examen es de 0,2 y la probabilidad de que ambos pierdan el examen es de 0,1*

1. *¿Cuál es la probabilidad de que al menos uno de estos estudiantes gane el examen?*

2. *¿Cuál es la probabilidad de de que ninguno de los dos pierda el examen?*

Solución 3.10.1 *Sea A el evento que el estudiante A pierda el examen y sea B el evento de que el estudiante B pierda el examen de estadística*

Nos piden la probabilidad de que uno de los dos gane el examen pero no los dos, es decir si C es este evento

$$P(C) = P(A') + P(B') - P(A' \cap B')$$
$$P(A') = 1 - P(A) = 1 - 0,5 = 0,5$$
$$P(B') = 1 - P(B) = 1 - 0,2 = 0,8$$
$$A' \cap B' = (A \cup B)'$$
$$P(A' \cap B') = 1 - P(A \cup B)$$

Pero

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,6$$
$$P(A' \cap B') = 0,4$$

entonces

$$P(C) = 0,5 + 0,8 - 0,4 = 0,9$$

$P(A' \cap B') = 0,4$

3.5. Técnica para la enumeración de puntos muestrales

Cuando S o cualquiera de sus subconjuntos tiene muchos eventos elementales describirlo por extensión para determinar los casos favorables y casos posibles se hace engorroso por lo que en esta sección utilizaremos el análisis combinatorio para determinar los casos favorables y posibles de una manera más simple.

3.5.1. Diagrama de árbol

En experimentos simples es muy útil utilizar un método llamado diagrama de árbol el cual explicaré con un ejemplo.

Ejemplo 3.11 *Si se lanza una moneda legal tres veces ¿Cuál es la probabilidad de que en la primera y la ultima lanzada salga cara?.*

Solución 3.11.1 *En el diagrama observamos que para el primer lanzamiento hay dos posibilidades sello (S) y cara (C) en el segundo lanzamiento para cada posibilidad anterior hay dos nuevas posibilidades y en el tercer se da lo mismo por lo que al final hay 8 casos posibles y dos casos favorables de acuerdo con Laplace, entonces si A es el evento que sale cara en el primer y último lanzamiento*

$$P(A) = \frac{2}{8} = \frac{1}{4}$$

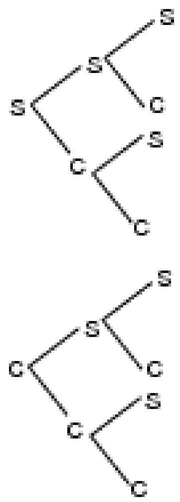


Diagrama de árbol

Teorema 3.7 *Considérese un experimento que tiene las dos características siguientes*

- *El experimento se realiza en dos partes*
- *La primera parte del experimento tiene m resultados posibles $x_1, x_2, x_3, \cdots, x_m$ independientemente del resultado x_i obtenido la segunda parte del experimento tiene n resultados posibles $y_1, y_2, y_3, \cdots, y_n$.*
  Cada resultado del espacio muestral S del experimento será por tanto, un par de la forma (x_i, y_j) es decir

$$S = \{(x_i, y_j) \,|i = 1, 2, 3, \cdots m; j = 1, 2, 3 \cdots n\}$$
$$n\,(S) = mn$$

Este teorema puede generalizarse de la siguiente forma

Si un experimento puede realizarse con las siguientes características

- El experimento se realiza en k partes
- la primera parte puede realizarse con n_1 resultados posibles $x_{11}, x_{12}, x_{13}, \cdots, x_{1n_1}$ e independientemente del resultado x_{1i}

se pueden realizar la segunda parte con n_2 resultados posibles $x_{21}, x_{22}, x_{23}, \cdots, x_{2n_2}$ y así sucesivamente cada parte del experimento puede realizarse de n_l formas, $l = 1, 2, 3, \cdots k$ obteniéndose un espacio muestral de la forma

$$S = \{(x_{1i}, x_{2j}, x_{3r}, \cdots, x_{km})\}$$
$$n(S) = n_1 \cdot n_2 \cdot n_3 \cdot \cdots \cdot n_k$$

Ahora enunciamos este teorema de una manera diferente.

Teorema 3.8 (Fundamental) . *Si una acción puede efectuarse de una de p maneras diferentes, y si después de que esta acción ha sido efectuada de una de esas maneras, una segunda acción puede efectuarse de una de q maneras diferentes, entonces el número total de maneras diferentes en que las acciones pueden efectuarse siguiendo el orden mencionado es pq.*

Corolario 3.1 *Si una acción puede efectuarse de p maneras diferentes, y una segunda acción puede efectuarse de q maneras diferentes, y una tercera acción puede efectuarse de r maneras diferentes, y así sucesivamente, entonces el número total de maneras diferentes en que pueden realizarse todas estas acciones en el orden mencionado es pqr...*

Corolario 3.2 *Si r acciones pueden efectuarse sucesivamente de p maneras diferentes cada una, entonces el número total de maneras diferentes en que pueden efectuarse las r acciones sucesivamente es* p^r.

3.6. Permutaciones

Definición 3.7 *Una permutación es un arreglo de objetos distintos de tal manera que una permutación difiere de otra si el orden del arreglo o su contenido difieren*

Conviene observar que el orden es una característica de especial importancia en una permutación. Cuando cambiamos el orden de los elementos de este arreglo, se dice que permutamos dichos elementos

3.6.1. Muestreo sin reemplazo

Consideremos un experimento en el cual se selecciona un objeto de n objetos distintos, y luego se selecciona un segundo objeto de los $n-1$ objetos restantes, y así sucesivamente hasta seleccionar el último objeto Este proceso se llama muestreo sin reemplazo de acuerdo con el teorema anterior los n objetos se pueden seleccionar de $n(n-1)(n-2)\cdots 3\cdot 2\cdot 1 = n!$ formas diferentes

Ahora si no se escogen todos los objetos, si no k objetos, los k objetos se pueden seleccionar

$$
\begin{aligned}
p_{n,k} &= n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!} \quad r \leq n \\
0! &= 1 \\
1! &= 1 \\
p_{n,n} &= n(n-1)(n-2)\cdots 1 = n!
\end{aligned}
$$

Ejemplo 3.12 *Supongamos que se necesitan dos representantes del grupo 02 de estadística I-ad para asistir a un congreso*

Solución 3.12.1 *Como el grupo 02 de estadística tiene 40 alumnos y se necesitan dos, eso quiere decir que que escogen 2 de 40*

$$p_{40,2} = \frac{40!}{(40-2)!} = 1560$$

Ejemplo 3.13 *Se necesita colocar 7 libros en un estante ¿De cuantas formas posibles se pueden colocar?*

Solución 3.13.1 *Como se escogerán 7 de 7 entonces*

$$p_{7,7} = 7! = 5040$$

3.6.2. Muestreo con reemplazo

Si suponemos que tenemos una urna con n objetos numerados del 1 al n y se selecciona un objeto de la urna y se anota su número, y luego se coloca nuevamente en la urna, luego se selecciona otro objeto el cual pude ser el primero y así sucesivamente se pueden seleccionar tantos objetos como se quieran. Este proceso se denomina muestreo con reemplazo.

Si queremos realizar un total de k selecciones distintas, se nos presentan dos posibilidades

- Si $k > n$, es imposible ya que hay n números posibles y no están repetidos.
- Si $n \geq k$ Entonces existen n^k posibles formas de escoger los k objetos

Ejemplo 3.14 *Si se colocan al azar 12 bolas en 20 urnas, ¿Cuál es la probabilidad de que ninguna urna contenga más de una bola*

Solución 3.14.1 *Como hay 20 urnas y 12 bolas entonces podemos escoger 12 de 20 posibilidades de colocar las bolas, pero hay 20^{12}, posibles formas de escoger una urna con una bola*

$$p = \frac{20!}{8! * 20^{12}} = \frac{235\,702\,467}{16\,000\,000\,000}$$

Coeficiente multinomial El número de formas en las que se pueden asignar n objetos distintos de k grupos diferentes que contienen $n_1, n_2, n_3, \cdots, n_k$ objetos respectivamente es

$$N = \frac{n!}{n_1!n_2!n_3!\cdots n_k!} = \left(\begin{array}{c} n \\ n_1 n_2 n_3 \cdots n_k \end{array} \right)$$

$$\text{donde } n = \sum_{i=1}^{k} n_i$$

3.6.3. Combinación

Una combinación es un arreglo de objetos distintos donde una combinación difiere de otra si difiere el contenido del arreglo. Si nos interesa determinar el número de combinaciones cuando en n objetos distintos deben seleccionarse r a la vez entonces

$$C_{n,r} = \frac{p_{n,r}}{r!} = \left(\begin{array}{c} n \\ r \end{array} \right) \quad r \leq n$$

ya que el numerador es el número de permutaciones al escoger r objetos de n posibles, pero hay que descontar los casos en que el orden determina para la combinación el mismo elemento, que es exactamente $r!$.

Propiedades

- $\begin{pmatrix} n \\ 0 \end{pmatrix} = \begin{pmatrix} n \\ n \end{pmatrix} = 1$
- $\begin{pmatrix} n \\ k \end{pmatrix} = \begin{pmatrix} n \\ n-k \end{pmatrix}$
- $\begin{pmatrix} n \\ k \end{pmatrix} + \begin{pmatrix} n \\ k-1 \end{pmatrix} = \begin{pmatrix} n+1 \\ k \end{pmatrix}$
- $C_{n,r} = C_{n-1,r-1} + C_{n-1,r}$
- El número de maneras en que mn objetos diferentes pueden dividirse en m grupos de n objetos cada uno, en donde el orden de los objetos en cada grupo no se toma en consideración, es
  $\frac{(mn)!}{(n!)^m}$, considerando el orden en que se forman los grupos,
  $\frac{(mn)!}{(n!)^m m!}$, sin considerar el orden en que se forman los grupos.

Ejemplo 3.15 *Calcular el número de maneras distintas en que 15 libros diferentes pueden dividirse en tres grupos de 9, 4 y 2 libros respectivamente.*

Solución 3.15.1 *En este caso*

$$N_m = \frac{15!}{9!4!2!} = 75075.$$

Ejemplo 3.16 *Se tiene una baraja*[2] *de 52 cartas diferentes.*

[2]Muchos problemas de probabilidad están relacionados con cartas, y aunque algunos estudiantes están relacionados con este juego, lo describiremos brevemente para que los problemas sean entendidos con claridad.

Una baraja ordinaria tiene 52 cartas divididas en cuatro grupos o palos con

Encontrar

1. El número de maneras en que pueden repartirse las cuatro manos de 13 cartas a cuatro jugadores de bridge
2. El número de maneras en que las 52 cartas pueden dividirse en cuatro grupos de 13 cartas cada uno.

1. En el juego de bridge cada distribución diferente de las manos entre los jugadores constituye una división diferente. Por tanto, en este caso, los grupos aparecen permutados, y por lo tanto, el número de maneras diferentes es
$$\frac{52!}{(13!)^4} = 53\,644\,737\,765\,488\,792\,839\,237\,440\,000$$
2. En este caso, no importa el orden de los grupos, así el número de maneras es
$$\frac{52!}{(13!)^4\,4!} = 2235\,197\,406\,895\,366\,368\,301\,560\,000$$

Ejemplo 3.17 *Una moneda se tira 10 veces. Calcular la probabilidad de que aparezcan exactamente 7 caras*

Solución 3.17.1 *Ya que la moneda tiene dos formas diferentes de aparecer en cada tiro, en 10 sería 2^{10} formas. Y de las 10 caras vamos a seleccionar 7 caras, por lo que serían $C_{10,7}$ formas diferentes, por lo que la probabilidad buscada es*

$$P = \frac{\binom{10}{7}}{2^{10}} = \frac{15}{128}$$

trece cartas cada uno. Los nombres de los palos y sus colores son los siguientes: Bastos(negros), diamantes(rojos), corazones(rojos) y espadas(negras.

Cada palo consiste de 9 cartas numeradas del 2 al 10 inclusive y 4 cartas más llamadas as, rey, reina y sota (ordenadas en valor descendente). La expresión de que una carta se escoge o saca al azar, significa que la carta se toma de una baraja bien mezclada del modo que todas las cartas tengan igual oportunidad de ser escogida.

Ejemplo 3.18 *Si se sacan 3 cartas al azar de una baraja de 52 cartas, calcular la probabilidad de que sean as, rey y reina.*

Solución 3.18.1 *Se pueden seleccionar 3 cartas al azar entre 52, por lo que hay $C_{52,3}$ formas diferentes. Y como hay 4palos y en cada palo hay un as, un rey y una reina, entonces resulta que estas barajas pueden obtenerse de $4 \times 4 \times 4$ formas diferentes .Por lo que la probabilidad buscada es*

$$\frac{4^3}{\binom{52}{3}} = \frac{16}{5525} = 2.8959 \times 10^{-3}$$

Ejemplo 3.19 *De una bolsa que que contiene 4 bolas blancas, 2 negras y 3 rojas, se sacan 5 al azar. Calcular la probabilidad de que 2 sean blancas 1 negra y 2 rojas.*

Solución 3.19.1 *Del total de $4 + 2 + 3 = 9$ bolas se pueden seleccionar 5 bolas en $C_{9,5}$ formas diferentes. Ahora entre las 4 bolas blancas 2 de ellas pueden seleccionarse $C_{4,2}$, entre las 2 blancas $C_{2,1}$ y entre las 3 rojas $C_{3,2}$ formas por lo que el total de casos favorables es $C_{4,2}C_{2,1}C_{3,2}$, así*

$$P = \frac{\binom{4}{2}\binom{2}{1}\binom{3}{2}}{\binom{9}{5}} = \frac{2}{7}$$

3.7. Probabilidad condicionada e independencia de eventos

Definición 3.8 *Sean $A, B \in \mathcal{A}$ y sea B un evento de probabilidad no nula para el evento A, llamamos probabilidad condicionada de A a B a la cantidad que representamos $P(A|B)$ y que definimos*

$$P(A|B) = \frac{P(AB)}{P(B)},$$

la cantidad $P(A|B)$ se lee la probabilidad de A dada la ocurrencia de B

Ejemplo 3.20 *Se lanza al aire un dado ¿Cuál es la probabilidad de que salga el número 4? si sabemos que el resultado ha sido par*

Solución 3.20.1 *Sea* $A = \{4\}$, *entonces* $P(A) = \frac{1}{6}$,
$B = \{2, 4, 6\}$, *entonces* $P(B) = \frac{3}{6} = \frac{1}{2}$
por tanto

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Definición 3.9 *Sean* $A, B \in \mathcal{A}$ *dos eventos de probabilidad no nula se dice que son independientes si y solo si*

$$P(AB) = P(A)P(B)$$

Ejemplo 3.21 *Para cierta población de empleados, los porcentajes de quienes aprueban un examen de aptitud para un trabajo, especificados según el sexo, se muestran en la tabla. Es decir todas las personas que presentan el examen el 24 % cae en la categoría de hombre aprobado, el 16 % en la categoría de de hombre reprobado, y así sucesivamente. Se selecciona al azar un empleado de esta población. Sea A el evento de que el empleado aprueba el examen y H el evento de que se selecciona un hombre ¿Son independientes los eventos A y H ?*

	sexo		
Resultado	*Mujer(M)*	*Hombre(H)*	*Total*
Aprueba (A)	*24*	*36*	*60*
Reprueba(A′)	*16*	*24*	*40*
Total	*40*	*60*	*100*

Solución 3.21.1 *Determinemos*

$$\begin{aligned} P(A|H) &= \\ &= \_\_\_\_\_\_ \\ &= \_\_\_\_\_\_ \\ &= \_\_\_\_\_\_\_ \end{aligned}$$

3.7.1. Reglas multiplicativas

Teorema 3.9 *Sea $A_1, A_2, A_3, \cdots A_n \in \mathcal{A}$ una sucesión de eventos aleatorios, entonces*

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)\,P(A_2|A_1)\,P(A_3|A_1A_2)\cdots P\left(A_n|\bigcap_{i=1}^{n-1} A_i\right)$$

Definición 3.10 *Se dice que una colección de eventos independientes $A_1, A_2, A_3, \cdots A_n \in \mathcal{A}$ es un evento exhaustivo y excluyente de sucesos, si se verifican las siguientes condiciones*

$$\bigcup_{i=1}^{n} A_i = S$$
$$A_i \cap A_j = \phi \qquad \forall i \neq j$$

generalmente estas dos condiciones son representadas por una sola

$$\biguplus_{i=1}^{n} A_i = S$$

Figura 3.1: A_1, A_2, A_3, A_4 y A_5 forman un sistema exhaustivo y excluyente de sucesos

Teorema 3.10 *sea $A_1, A_2, A_3, \cdots A_n \in \mathcal{A}$ un sistema exhaustivo y excluyente de sucesos. Entonces*

$$\forall B \subset \mathcal{A}, \Longrightarrow P(B) = \sum_{i=1}^{n} P(B|A_i)\,P(A_i)$$

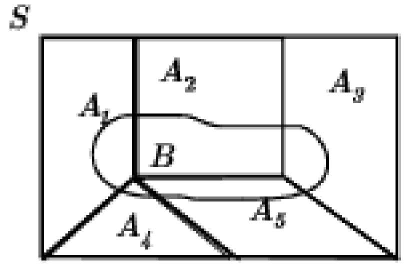


Figura 3.2: Si A_1, A_2, A_3, A_4 forman un sistema exhaustivo y excluyente de sucesos , podemos calcular la probabilidad de B a partir de las cantidades $P(B \cap A_i)$

Demostración. Como $B \subset S$ podemos decir

$$\begin{aligned} P(B) &= P\left(B \cap \left(\bigcup_{i=1}^{n} A_i\right)\right) \\ &= P\left(\bigcup_{i=1}^{n} (B \cap A_i)\right) \\ &= \sum_{i=1}^{n} P(B \cap A_i) \\ &= \sum_{i=1}^{n} P(B|A_i)\, P(A_i) \end{aligned}$$

Ejemplo 3.22 *Se tienen dos urnas, y cada una de ellas contiene un número diferente de bolas blancas y rojas:*

-
    - *primera urna U_1: 3 bolas blancas y 2 rojas*
    - *Segunda urna U_2 4 bolas blancas y 2 rojas*

Se realiza el siguiente experimento :

Se tira una moneda al aire y si sale cara se elige una bola de la primera urna y si sale sello se escoge una bola de la segunda urna.

¿Cuál es la probabilidad de que salga una bola blanca?

Solución 3.22.1 *La situación que tenemos la podemos esquematizar de la siguiente manera*
Si B: es el evento de sacar una bola blanca y R : el evento de sacar una bola roja, entonces

$$\begin{aligned} P(U_1) &= \frac{1}{2} \\ P(B|U_1) &= \frac{3}{5} \\ P(U_2) &= \frac{1}{2} \\ P(B|U_2) &= \frac{4}{6} \end{aligned}$$

Como U_1 y U_2 forman un sistema exhaustivo y excluyente de sucesos aplicando el teorema de la probabilidad total podemos afirmar que,

$$\begin{aligned} P(B) &= P(B|U_1)\,P(U_1) + P(B|U_2)\,P(U_2) \\ &= \frac{3}{5}\cdot\frac{1}{2} + \frac{4}{6}\cdot\frac{1}{2} = \frac{19}{30} \end{aligned}$$

Teorema 3.11 *(Bayes) Sea $A_1, A_2, A_3, \cdots A_n \in \mathcal{A}$ un sistema exhaustivo y excluyente de eventos. Sea $B \subset \mathcal{A}$ un suceso del que conocemos todas las cantidades*

$$P(B|A_i)\,, i = 1, 2, 3, \cdots, n$$

a las que denominamos verosimilitudes, entones se verifica

$$\forall j = 1, 2, 3, \cdots, n, \qquad P(A_j|B) = \frac{P(B|A_j)\,P(A_j)}{\sum_{i=1}^{n} P(B|A_i)\,P(A_i)}$$

Demostración. *Es consecuencia de la definición de probabilidad condicionada en términos de la intersección, y del teorema de la probabilidad total*

$$\begin{aligned} P(A_j|B) &= \frac{P(A_j \cap B)}{P(B)} \\ &= \frac{P(B|A_j)\,P(A_j)}{\sum_{i=1}^{n} P(B|A_i)\,P(A_i)} \end{aligned}$$

■

Ejemplo 3.23 *Se tienen tres urnas. Cada una de ellas contiene un número diferente de bolas blancas y rojas. La primera urna tiene 3 bolas blancas y 2 rojas, la segunda 4 bolas blancas y 2 rojas y la tercera 3 bolas rojas.*
Se realiza el siguiente experimento:
Alguién elije al azar y con la misma probabilidad una de las tres urnas, saca una bola.
Si el resultado del experimento ha sido sacar una bola blanca ¿Cuál es la probabilidad de que provenga de la primera urna? Calcular lo mismo para las otras dos urnas.

Solución 3.23.1 *Si B: es el evento de sacar una bola blanca y R: el evento de sacar una bola roja, entonces como*
U_1 : *3 bolas blancas y 2 rojas*
U_2 : *4 bolas blancas y 2 rojas*
U_3 : *3 bolas rojas*
tenemos

$$
\begin{aligned}
P(U_1) &= \frac{1}{3} \\
P(U_2) &= \frac{1}{3} \\
P(U_3) &= \frac{1}{3} \\
P(B|U_1) &= \frac{3}{5} \\
P(B|U_2) &= \frac{4}{6} \\
P(B|U_3) &= 0
\end{aligned}
$$

En este caso U_1, U_2, U_3 *forman un sistema incompatible y excluyente de eventos, por lo que es posible aplicar el teorema de Bayes*

$$
\begin{aligned}
P(U_1|B) &= \frac{P(B|U_1)\,P(U_1)}{P(B|U_1)\,P(U_1) + P(B|U_2)\,P(U_2) + P(B|U_3)\,P(U_3)} \\
&= \frac{\left(\frac{3}{5}\right)\left(\frac{1}{3}\right)}{\left(\frac{3}{5}\right)\left(\frac{1}{3}\right) + \left(\frac{4}{6}\right)\left(\frac{1}{3}\right) + 0\left(\frac{1}{3}\right)} \\
&= \frac{\mathbf{9}}{19}
\end{aligned}
$$

Para los otros dos casos resulta de manera equivalente

$$\begin{aligned} P(U_2|B) &= \frac{P(B|U_2)P(U_2)}{P(B|U_1)P(U_1)+P(B|U_2)P(U_2)+P(B|U_3)P(U_3)} \\ &= \frac{10}{19} \end{aligned}$$

$$\begin{aligned} P(U_3|B) &= \frac{P(B|U_3)P(U_3)}{P(B|U_1)P(U_1)+P(B|U_2)P(U_2)+P(B|U_3)P(U_3)} \\ &= 0 \end{aligned}$$

Ejemplo 3.24 *En la segunda guerra mundial, uno de los primeros intentos de investigación de operaciones en la Gran Bretaña se orientaba en establecer patrones de búsqueda de submarinos desde vuelos de escuadrones o mediante un sólo avión. Por algún tiempo, la tendencia fue concentrar los vuelos en la costa, pues se pensaba que le mayor número de avistamientos que ocurrían ahí. El grupo de investigación registró 1000 vuelos de un solo avión con los siguientes resultados(Los datos no son reales)*

	En la playa	*Fuera de la costa*	*Total*
Observación	*80*	*20*	*100*
No observación	*820*	*80*	*900*
Total de salidas	*900*	*100*	*1000*

sea

S_1 *Hubo un avistamiento*
S_2 *No hubo avistamiento*
B_1 *Salida a la costa*
B_2 *Salida en altamar*

Vemos de inmediato que

$$\begin{aligned} P(S_1|B_1) &= \frac{80}{900} = \frac{4}{45} = 8.8889 \times 10^{-2} \\ P(S_2|B_2) &= \frac{20}{100} = 0.2 \end{aligned}$$

lo cual indica una estrategia de búsqueda es contraria a la primera práctica.

Ejemplo 3.25 *Supóngase que se va a seleccionar una muestra aleatoria de tamaño 2 de un lotes de tamaño 100, y que se sabe que 98 de los 100 artículos se encuentran en buen estado. La muestra se toma de manera tal que el primer artículo se observa y se regresa antes de seleccionar el segundo artículo*

Solución 3.25.1 . *Si aceptamos que*
A : El primer artículo observado está en buen estado
B : El segundo artículo observado está en buen estado
Y si deseamos determinar la probabilidad de que ambos artículos estén en buen estado, entonces

$$P(A \cap B) = P(A)\,P(B) = \left(\frac{98}{100}\right)\left(\frac{98}{100}\right) = 0.960\,4$$

Si se selecciona el artículo sin reemplazo de modo que que el primer artículo no se regresa antes de seleccionar el segundo, entonces

$$P(A \cap B) = P(A)\,P(B|A) = \frac{98}{100} \cdot \frac{97}{99} = 0{,}9602$$

Los resultados son muy parecidos por lo que generalmente suponemos que los eventos son independientes cuando la fracción de muestreo (tamaño de la muestra/tamaño de la población) es menor que 0.1

Ejemplo 3.26 *Tres industrias suministran microprocesadores a un fabricante de equipos de telemetría. Todos se elaboran supuestamente con las mismas especificaciones. No obstante, el fabricante ha probado durante varios años los microprocesadores, y los registros indican la siguiente información*

Industria	*Fracción de defectos*	*Fracción suministrada por*
1	*0.02*	*0.15*
2	*0.01*	*0.80*
3	*0.03*	*0.05*

EL fabricante ha interrumpido las pruebas por causa de los costos involucrados, y puede ser razonable suponer que la proporción defectuosa y la mezcla de inventarios son las mismas que durante el periodo en el que se efectuaron los registros. El director de manufactura selecciona un microprocesador al azar, lo lleva al departamento de

pruebas y descubre que está defectuoso. Determinar la probabilidad que el artículo proviene de la industria 3
Sea A el evento de que el artículo es defectuoso y B_i $i = 1, 2, 3$ es el evento que el artículo proviene de la empresa i

$$\begin{aligned} P(B_3|A) &= \frac{P(A|B_2)\,P(B_2)}{P(A|B_1)\,P(B_1) + P(A|B_2)\,P(B_2) + P(A|B_3)\,P(B_3)} \\ &= \frac{(0{,}05)\,(0{,}03)}{(0{,}15)\,(0{,}02) + (0{,}8)\,(0{,}01) + (0{,}05)\,(0{,}03)} = 0.12 \end{aligned}$$

Ejemplo 3.27 *Una caja contiene 8 bolas rojas, 3 bolas blancas y nueve azules. Si se sacan 3 bolas al azar, determinar la probabilidad de que*

a. *Las 3 sea rojas*

b. *Las tres sean blancas*

c. *2 sean rojas*

d. *Al menos una blanca*

e. *Sea una de cada color*

f. *Salgan en el orden roja,blanca y azul*

Solución 3.27.1 *Denotaremos por R_i $i = 1, 2, 3$ los eventos que la primera, la segunda y la tercera bola sean rojas respectivamente. Denotaremos por B_i $i = 1, 2, 3$ los eventos que la primera, la segunda y la tercera bola sean blancas respectivamente, y denotaremos por A_i $i = 1, 2, 3$ los eventos que la primera, la segunda y la tercera bola sean rojas respectivamente, entonces*

a.

$$P\left(\bigcap_{i=1}^{3} R_i\right) = \frac{\begin{pmatrix} 8 \\ 3 \end{pmatrix}}{\begin{pmatrix} 20 \\ 3 \end{pmatrix}} = \frac{14}{285}$$

b.

$$P\left(\bigcap_{i=1}^{3} B_i\right) = \frac{\binom{3}{3}}{\binom{20}{3}} = \frac{1}{1140}$$

c.

$$P\left(\bigcap_{i=1}^{2} R_i \cap B\right) = \frac{\binom{8}{2}}{\binom{20}{3}} = \frac{}{}$$

d.

$$P\,(\text{ninguna es blanca}) = \frac{\binom{17}{3}}{\binom{20}{3}} = \frac{}{}$$

e.

$$P\,(\text{sacar una de cada color}) = \frac{\binom{8}{1}\binom{3}{1}\binom{9}{1}}{\binom{20}{3}} = \frac{}{}$$

f.

$$\begin{aligned} & P\,(\text{bolas en el orden R, B,A}) \\ & = \frac{1}{3!} P\,(\text{una de cada colore}) = \frac{3}{95} \end{aligned}$$

Ejemplo 3.28 *De una baraja de 52 naipes bien mezclada se sacan 5 naipes. Hallar la probabilidad de que 3 sean de un palo y 2 de otro*

Solución 3.28.1

$$\begin{aligned} P\,(\textit{3 de cualquier figura y 2 de otra}) = & \\ & = \frac{4\binom{13}{3} \cdot 3\binom{13}{2}}{\binom{52}{5}} \\ & = 0.103 \end{aligned}$$

Ejemplo 3.29 *Supóngase que se lanzan 12 dados. Se determina la probabilidad p de que cada uno de los seis números distintos aparezca dos veces*

Solución 3.29.1 *tenemos que S es el conjunto formado por la sucesión de 12-tupla donde la i-ésimo valor de la sucesión es el resultado i-ésimo lanzamiento, de lo que se deduce que $n(S) = 6^{12}$ donde todos los 6^{12} resultados posibles son igualmente probables, ahora podemos considerar los seis valores posibles que se puede dar en cada dado como seis casillas las cuales se le pueden asignar solo dos posibilidades. Por lo que al utilizar el coeficiente multinomial para determinar los casos favorables tenemos*

$$\begin{aligned} n &= 12 \\ k &= 6 \\ n_1 &= n_2 = \cdots = n_6 = 2 \\ N &= \frac{12!}{(2!)(2!)\cdots(2!)} = \frac{12!}{(2!)^6} \end{aligned}$$

Por lo que la probabilidad buscada es
$P = \frac{12!}{2^6 6^{12}} = \frac{1925}{559872} = 3.4383 \times 10^{-3}$

Ejemplo 3.30 *Una baraja de 52 cartas contiene 13 corazones. Supóngase que se barajan las cartas y se distribuyen entre cuatro jugadores A,B,C y D de tal manera que a cada jugador le correspondan 13 cartas, Se determinará la probabilidad P de que cada jugador reciba 6, 4, 2 y 1 corazones respectivamente*

Solución 3.30.1 *El número de combinaciones distintas posibles de las 13 posiciones ocupadas en la baraja por los corazones es $\binom{52}{13}$. Si el jugador A recibe 6 corazones, hay $\binom{13}{6}$ combinaciones posibles de las 6 posiciones que ocupan estos corazones entre las 13 cartas que recibe el jugador A. De la misma manera podemos establecer las combinaciones posibles de cada jugador entre las 13 cartas que recibe cada uno como $\binom{13}{4}$, $\binom{13}{2}$ y $\binom{13}{1}$ respectivamente para B,C y D. Por tanto como los eventos A,B,C y D son independientes*

$$P(ABCD) = \frac{\binom{13}{6}\binom{13}{4}\binom{13}{2}\binom{13}{1}}{\binom{52}{13}} = 1.9592 \times 10^{-3}$$

Ejemplo 3.31 *Supóngase que una moneda equilibrada va a ser lanzada diez veces y se desea determinar*

a. *la probabilidad P de obtener exactamente 3 caras*

b. *La probabilidad P de obtener a lo sumo tres caras*

Solución 3.31.1

a. El número total de posibles combinaciones distintas de 10 caras y sellos es 2^{10} y se puede suponer que todas estas combinaciones son igualmente probables y el número de casos favorables es $\binom{10}{3}$ por lo que

$$p = \frac{\binom{10}{3}}{2^{10}} = 0.117\,19$$

Solución 3.31.2 **b.** *Puesto que lo que me piden es la union de los eventos A_i : obtener 0,1,2 y 3 caras $i = 0, 1, 2, 3,$ y como estos eventos son disjuntos tenemos que*

$$P\left(\bigcup_{i=0}^{3} A_i\right) = \frac{\binom{10}{0} + \binom{10}{1} + \binom{10}{2} + \binom{10}{3}}{2^{10}} = 0.171\,88$$

Ejemplo 3.32 *Supóngase que en una clase hay 15 hombres y 30 mujeres, y que se van a seleccionar al azar 10 estudiantes para una tarea especial. se determinará la probabilidad P de seleccionar exactamente 3 hombres*

Solución 3.32.1 *el número de combinaciones distintas que se pueden formar con 45 estudiantes para obtener una muestra de 10 es $\binom{45}{10}$, y como el muestreo es al azar y sin reemplazo entonces todas estas combinaciones son igualmente posibles por lo que hay que determinar el números de combinaciones distintas que se pueden seleccionar con 3 hombres y 7 mujeres, entonces con los hombres se pueden realizar $\binom{15}{3}$ y con las mujeres $\binom{30}{7}$ por lo que*

$$P = \frac{\binom{15}{3}\binom{30}{7}}{\binom{45}{10}} = \frac{3958\,500}{13\,633\,279} = 0.290\,36$$

Ejemplo 3.33 *Supóngase que se baraja un naipe de 52 cartas que contiene 4 ases y que las cartas se reparten entre cuatro jugadores, de forma que cada una reciba 13 cartas. se determinará la probabilidad de que cada jugador reciba un as.*

Solución 3.33.1 *El número de combinaciones diferentes es* $\binom{52}{4}$*. Y podemos suponer que todas estas combinaciones son igualmente probables. si cada jugador recibe un as entonces debe haber un as entre las 13 cartas que recibe por ejemplo el primer jugador y un as recibirá cada uno de los jugadores de las 13 cartas que recibirán, es decir hay 13 posiciones posibles para el as que recibe cada jugador por lo que los casos favorables será* 13^4*, entonces*

$$P = \frac{13^4}{\binom{52}{4}} = \frac{2197}{20\,825} = 0.105\,5$$

Ejemplo 3.34 *Supóngase que una máquina produce un artículo defectuoso con probabilidad* $p = 0{,}4$ *y produce un artículo no defectuoso con una probabilidad* $q = 0,6$*. Supóngase además que se seleccionan aleatoriamente para su control seis de los artículos producidos por la máquina y que los resultados de control son independientes para estos seis artículos. Se determinará la probabilidad de que exactamente dos de los seis artículos sean defectuosos*

Solución 3.34.1 *Si consideramos los eventos* D_i *el evento del que el* $i-$*ésimo artículo sea defectuoso y* N_j *el evento de que el* $i-$*ésimo artículo sea no defectuoso para,* $i = 1, 2, 3, 4, 5, 6$*, como los eventos son independientes entonces por ejemplo*

$$P\left(N_1 N_3 D_2 D_4 D_5 D_6\right) = qqpppp$$

como esta es una de las posibles combinaciones hay que determinar las otras que no es más que $\binom{6}{2}$ *combinaciones distintas y por el th 3.13 tenemos*

$$P = \binom{6}{2}(0{,}4)^2(0{,}6)^4 = 0.311\,04$$

Ejemplo 3.35 *considérese una máquina que produce un artículo defectuoso con una probabilidad $p = 0{,}4$ y uno no defectuoso con probabilidad $q = 0{,}6$. Supóngase que el control se realiza seleccionando artículos al azar y de uno en uno hasta obtener exactamente cinco artículos defectuosos. Se determinará la probabilidad P de que deban ser seleccionados 20 artículos para obtener 5 defectuosos*

Solución 3.35.1 *El quinto artículo defectuoso será el $n-$ésimo controlado si, y sólo si, hay exactamente cuatro defectuosos entre los primeros $n-1$ artículos y el $n-$ésimo es defectuoso entonces por el th 3.13 tenemos que*

$$\begin{aligned} P &= \binom{n-1}{4} p^5 q^{n-5} = \binom{19}{4} (0{,}4)^5 (0{,}6)^{15} \\ &= 1.8662 \times 10^{-2} \end{aligned}$$

Ejemplo 3.36 *Supóngase que se van a extraer dos bolas aleatoriamente y sin reemplazo, de una urna que contiene 8 bolas rojas y 10 bolas azules. Se determinará la probabilidad P de obtener la primera bola roja y la segunda azul*

Solución 3.36.1 *Sea A el evento de que la primera bola sea roja y B de que la segunda bola sea azul, entonces $P(A) = \frac{8}{18} = 0.44444$, además si el suceso A ha ocurrido, entonces se ha obtenido una bola roja de la urna en la primera extracción por lo que la probabilidad de obtener una bola azul en la segunda extracción es*

$$P(B|A) = \frac{10}{17} = 0.58824$$

entonces

$$P(AB) = (0.44444)(0.58824) = 0.26144$$

Ejemplo 3.37 *Para la fabricación de un gran lote de artículos similares se utilizaron tres máquinas M_1, M_2 y M_3. Supóngase que el 20 % de los artículos fueron fabricados por la máquina M_1, el 30 % por la máquina M_2 y el 50 % por la máquina M_3 Suponemos además que el 1 % de los artículos fabricados por la máquina M_1*

son defectuosos y así respectivamente 2 % y 3 % de los artículos fabricados por la máquinas M_2 y M_3. Se quiere seleccionar al azar uno de los artículos del lote que resultan defectuosos. Determine la probabilidad de que el artículo haya sido fabricado por la máquina M_2.

Solución 3.37.1 *Sea A_i el evento de que el artículo haya sido fabricado por la máquina $M_i, i = 1, 2, 3,$ y sea B el suceso de que el artículo seleccionado sea defectuoso, de acuerdo con esto hay que determinar $P(A_2|B)$, entonces*

$$\begin{aligned} P(A_1) &= 0{,}2 \\ P(A_2) &= 0{,}3 \\ P(A_3) &= 0{,}5 \\ P(B|A_1) &= 0{,}01 \\ P(B|A_2) &= 0{,}02 \\ P(B|A_3) &= 0{,}03 \end{aligned}$$

aplicando el teorema de Bayes

$$P(A_2|B) = \frac{P(A_2)\,P(B|A_2)}{\sum_{i=1}^{3} P(A_i)\,P(B|A_i)} = 0{,}26$$

Tarea 1

1. *Tres clases diferentes tienen 20, 18 y 25 estudiantes, respectivamente, y cada estudiante pertenece a una sola clase. Si se forma un equipo con un estudiante de cada una de estas tres clases, ¿de cuántas maneras distintas se pueden seleccionar los miembros del equipo?*

2. *¿De cuántas maneras distintas se pueden ordenar cinco letras a, b, c, d y e?*

3. *Si un hombre tiene seis camisas distintas y cuatro pares distintos de pantalones, ¿de cuántas formas distintas se puede vestir combinando esas prendas?*

4. *Si se lanzan cuatro dados, ¿cuál es la probabilidad de que los cuatro números que aparecen sean distintos?*

5. *Si se lanzan seis dados, ¿cuál es la probabilidad de que cada uno de los seis números posibles aparezcan exactamente una vez?*

6. *Si se colocan al azar 12 bolas en 20 urnas, ¿cuál es la probabilidad de que ninguna urna contenga más de una bola?*

7. *El ascensor de un edificio empieza a subir con cinco personas y para en siete pisos. Si la probabilidad de que cualquier pasajero salga del ascensor en un piso concreto es igual para todos los pisos y los pasajeros salen independientemente unos de otros, ¿cuál es la probabilidad de que no haya dos pasajeros que salgan en el mismo piso?*

8. *Si k personas se sientan aleatoriamente en una fila de n asientos ($n > k$), ¿cuál es la probabilidad de que ocupen k asientos contiguos en la fila?*

9. *Sik personas se sientan aleatoriamente en n sillas dispuestas en círculo ($n > k$), ¿cuál es la probabilidad de que ocupen k sillas contiguas del círculo?*

10. *Si n personas se sientan aleatoriamente en una fila de $2n$ asientos, ¿cuál es la probabilidad de que no haya dos personas sentadas en asientos contiguos?*

11. *Una caja contiene 24 bombillas, de las cuales 2 son defectuosas. Sí una persona selecciona 10 bombillas al azar, sin reemplazo, ¿cuál es la probabilidad de seleccionar las 2 bombillas defectuosas?*

12. *Supóngase que se ha de seleccionar un comité de 12 personas aleatoriamente escogidas entre un grupo de 100. Determínese la probabilidad de que dos personas concretas, A y B, sean seleccionadas.*

13. *Supóngase que 35 personas se dividen aleatoriamente en dos equipos de manera que uno de los equipos consta de 10 personas y el otro de 25. ¿Cuál es la probabilidad de que dos personas concretas, A y B, estén en el mismo equipo?*

14. *Una caja contiene 24 bombillas, de las cuales 4 son defectuosas. Si una persona selecciona aleatoriamente 10 bombillas de la caja, y una segunda persona toma entonces las 14 bombillas restantes, ¿cuál es la probabilidad de que la misma persona seleccione las 4 bombillas defectuosas?*

15. *Demuéstrese que, para cualquier entero positivo n y k ($n > k$),*

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$$

16. *Demuéstrese que*

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = 2^n$$

17. *El Senado de Estados Unidos está constituido por dos senadores de cada uno de los 50 estados,*

    a) *Si se selecciona aleatoriamente un comité de 8 senadores, ¿cuál es la probabilidad de que contenga al menos uno de los dos senadores de un estado concreto?*

    b) *¿Cuál es la probabilidad de que un grupo de 50 senadores seleccionados aleatoriamente contenga un senador de cada estado?*

18. *Supóngase que$A$ es un suceso tal que $P(A) = 0$ y que B es cualquier otro suceso. Demuéstrese que A yB son sucesos independientes.*

19. *Supóngase que una persona lanza tres veces dos dados equilibrados. Determínese la probabilidad de que en cada uno de los tres lanzamientos la suma de los dos números que aparecen sea 7.*

20. *Supóngase que la probabilidad de que el sistema de control utilizado en una nave espacial no funcione en un vuelo concreto*

es 0.001. Supóngase además que la nave también tiene instalado un segundo sistema de control idéntico, pero completamente independiente del primero, que toma el control cuando el primero falla. Determínese la probabilidad de que en un vuelo concreto la nave espacial esté bajo control, ya sea del sistema original o del sistema duplicado.

21. *Supóngase que una lotería consta de 10 000 boletos y que otra lotería consta de 5000. Si una persona compra 100 boletos de cada lotería, ¿cuál es la probabilidad de que gane al menos un primer premio?*

22. *Dos estudiantes A y B están inscritos en un curso. Si el estudiante A asiste a las clases el 80 % de las veces y el estudiante B el 60 %, y si las ausencias de los dos estudiantes son independientes, ¿cuál es la probabilidad de que al menos uno de los dos estudiantes esté en clase un día concreto?*

23. *Si se lanzan tres dados equilibrados, ¿cuál es la probabilidad de que los tres números que aparecen sean iguales?*

24. *Considérese un experimento en el cual se lanza una moneda equilibrada hasta que aparece una cara por primera vez. Si este experimento se repite tres veces, ¿cuál es la probabilidad de que se necesite exactamente el mismo número de lanzamientos para cada una de las tres repeticiones?*

25. *Supóngase que A, B y C son tres sucesos independientes tales que $P(A) = \frac{1}{4}$, $P(B) = \frac{1}{3}$, y $P(C) = \frac{1}{2}$.*

    a) *Determínese la probabilidad de que ninguno de estos tres sucesos ocurra*

    b) *Determínese la probabilidad de que ocurra exactamente uno de estos tres sucesos.*

26. *Supóngase que la probabilidad de que una partícula emitida por un material radiactivo penetre en cierto campo es 0.01. Si se emiten diez partículas,*

a) *¿cuál es la probabilidad de que exactamente una de ellas penetre en el campo?*

b) *Si se emiten diez partículas, ¿cuál es la probabilidad de que al menos una de ellas penetre en el campo?*

c) *¿Cuántas partículas tienen que ser emitidas para que la probabilidad de que al menos una partícula penetre en el campo sea al menos 0.8?*

27. *En la Serie Mundial de Béisbol, dos equipos A y B juegan una serie de partidos uno contra otro y el primer equipo que gana un total de cuatro partidos es el ganador de la Serie Mundial. Si la probabilidad de que el equipo A gane un partido contra el equipo B es $\frac{1}{3}$, ¿cuál es la probabilidad de que el equipo A gane la Serie Mundial?*

28. *15. El Senado de Estados Unidos está constituido por dos senadores de cada uno de los 50 estados, (a) Si se selecciona aleatoriamente un comité de 8 senadores, ¿cuál es la probabilidad de que contenga al menos uno de los dos senadores de un estado concreto? (b) ¿Cuál es la probabilidad de que un grupo de 50 senadores seleccionados aleatoriamente contenga un senador de cada estado?*

29. *La probabilidad de que una persona nade es de 0.45 y la probabilidad de que una persona cace es de 0.58. Si la probabilidad de que una persona cace sabiendo que también caza es 0.21, encuentre la probabilidad de que:*

    a) *Cace y nade*

    b) *Cace si también nada*

    c) *Cace y no nade*

    d) *Cace o nade*

30. *La tabla adjunta muestra las frecuencias relativas para el daltonismo en hombres y mujeres, donde H representa hombres,*

M mujeres, D daltónico y ND no daltónico

	H	*M*
D	*0.042*	*0.007*
ND	*0.485*	*0.466*

Si se escoge a una persona al azar, use la tabla para determinar las probabilidades siguientes:

a) $P(H)$

b) $P(H \cap D)$

c) $P(D)$

d) $P(H \cap ND)$

31. *Si* $P(E) = 0{,}2$ *y* $P(F) = 0{,}3$. *Responda si puede ser cierto en cada una de las preguntas dadas y si es posible plantee un ejemplo.*

    a) ¿$P(E \cup F) = 0{,}5$?

    b) ¿$P(E \cup F) = 0{,}7$?

    c) ¿$P(E \cup F) = 0{,}4$?

    d) ¿$P(E \cap F) = 0{,}2$?

    e) ¿$P(E \cap F) = 0{,}3$?

    f) ¿$P(E \cap F) = 0{,}1$?

    g) ¿$P(E \cap F) = 0{,}4$?

32. *Si cuatro hombres y cuatro mujeres se colocan en fila, ¿Cuál es la probabilidad de que un arreglo aleatorio de los ocho individuos tenga*

    a) *hombres y mujeres alternados?*

    b) *A los hombres todos juntos?*

33. *De un conteo de tarjetas numeradas dei 1 al 10000. ¿Cuál es la probabilidad de que el número que te toque sea divisible exactamente por 5?*

34. *Un especialista en alergias alega que el 50 % de sus pacientes sufre de alergia ¿Cuál es la probabilidad de que*

    a) *de que 3 de sus siguientes cuatro pacientes sufran de alergia?*

    b) *Ninguno de los cuatro pacientes sufran de alergias?*

35. *De una caja que contiene 6 pelotas negras y 4 verdes, se sacan 3 en sucesión, reemplazándose cada una en la caja antes de extraer la siguiente. ¿Cuál es la probabilidad de que*

    a) *las 3 sen del mismo color*

    b) *sea al menos una de cada color*

36. *Un embarque de 12 televisores contiene 3 defectuosos. ¿En cuantas formas puede un hotel comprar 5 y recibir al menos 2 de los defectuosos?*

37. *En una cierta ciudad, 40 % de los votantes son Liberales y el 60 % son conservadores; 70 % de los republicanos y el 80 % de los demócratas están a favor de una de una consulta popular. al seleccionar al azar un votante de la ciudad, ¿Cuál es la probabilidad de qué está a favor de la consulta ?*

38. *¿Cuántas manos de bridge que contengan 4 espadas, 6 diamantes, 1 de bastos y 2 de corazones son posibles?*

39. *Una empresa industrial grande utiliza 3 hoteles locales para proporcionar alojamiento a sus clientes durante la noche. De pasadas experiencias se sabe que al 20 % de ellos se les asigna habitación en el hotel de Santa Marta, el 50 % en Cartagena y al 30 % en Tolú. Si existe una falla en el servicio de plomería en el 5 % de la habitaciones en Santa Marta , en el 4 % en Cartagena y del 8 % en Tolú. ¿Cuál es la probabilidad de que*

    a) *a un cliente se le asigne un cuarto con problemas en plomería ?*

b) *a una persona con un problema en plomería se asigne el hotel en Cartagena?*

40. *Un espacio muestral de 200 adultos se clasifica de acuerdo con su sexo y nivel de educación*

Educación	*Hombre*	*Mujer*
Primaria	*38*	*45*
Secundaria	*28*	*50*
Bachillerato	*22*	*17*

Si se selecciona aleatoriamente a una persona de ese grupo, encuentre la probabilidad de que

a) *no tenga grado de profesional dado de que sea mujer*

b) *sea hombre dado que tiene educación superior*

41. *Se lanzan un par de dados, si se sabe que uno de ellos resulta en un 4, ¿cuál es la probabilidad de que*

a) *el otro caiga en 6*

b) *el total de ambos sea 9*

42. *Supongamos que se selecciona al azar un individuo de la población de todos los adultos hombres que en los Estados Unidos. Sea a el evento en que el individuo seleccionado tenga una estatura de más de 6 pies, y B el evento de que el individuo seleccionado sea un jugador profesional de baloncesto. ¿Cuál considera que sea mayor $P(B|A)$ o $P(A|B)$?. ¿Por qué?*

43. *Un circuito flexible se selecciona al azar de una corrida de producción de 1000 circuitos. Los defectos de manufactura se clasifican en tres diferentes tipos, denominados A. B y C. Los defectos de tipo A ocurren el 2 por ciento de las veces, los del tipo B, el 1 por ciento, y los de tipo C, el 1.5 por ciento. Además, se sabe que el 0.5 por ciento tienen los defectos de tipo A y B, el 0.6 por ciento, los defectos B y C, y el 0.4 por ciento presenta los defectos B y C, en tanto que el 0.2 por*

ciento tiene los tres defectos. ¿Cuál es la probabilidad de que el circuito flexible seleccionado tenga al menos uno de los tres tipos de defectos?

44. *En un laboratorio de factores humanos, se miden tiempos de reacción, por ejemplo, el tiempo que transcurre desde el instante en que se despliega un número de posición en un tablero digital hasta que el sujeto presiona un botón localizado en la posición indicada. Participan dos sujetos, midiéndose el tiempo en segundos para cada individuo* (t_1, t_2). *¿Cuál es el espacio muestral para este experimento? Presente los siguientes eventos como subconjuntos y márquelos sobre un diagrama:*

$$\left(\frac{t_1+t_2}{2}\right) \leq 0{,}15, \ \text{máx}\,(t_1,t_2) \leq 0{,}15, \ |t_1 - t_2| < 0{,}6.$$

45. *Durante un periodo de 24 horas se entrará a un procesamiento computarizado. En un tiempo X y se saldrá en tiempo Y ≥ X. Considérense X y Y medido en horas en la linea del tiempo con el inicio del periodo de 24 horas como el origen. El experimento consiste en observar X y Y, (X, Y).*

    a) *Describa el espacio muestral S.*

    b) *Dibuje los siguientes eventos en el plano X, Y.*

        1) *El tiempo de utilización es una hora o menos*

        2) *El acceso es antes de* t_1 *y la salida después de* t_2, *donde* $0 \leq t_1 \leq t_2 \leq 24$

        3) *El tiempo de utilización es menor que el 20 por ciento del periodo.*

46. *Se prueban diodos de un lote uno a la vez y se marcan ya sea como defectuosos o como no defectuosos. Esto continúa hasta encontrar dos artículos defectuosos o cuando se han probado cinco artículos. Describa el espacio muestral para este experimento.*

47. *Un conjunto tiene cuatro elementos* $A = \{a, b, c, d\}$. *Describa el conjunto partes de* A .

48. *Describa el espacio muestral para cada uno de los siguientes experimentos:*

    a) *Un lote de 120 tapas de baterías para celdas de marcapasos contiene varias defectuosas debido a un problema con el material de barrera que se aplica en el sistema de alimentación. Se seleccionan tres tapas al azar (sin reemplazo) y se inspeccionan con cuidado siguiendo una reducción.*

    b) *Una paleta de 10 piezas fundidas contiene una unidad defectuosa y nueve en buen estado. Se seleccionan cuatro piezas al azar (sin reemplazo) y se inspeccionan.*

49. *El gerente de producción de cierta compañía está interesado en probar un producto terminado, que se encuentra disponible en lotes de tamaño 50. A él le gustaría volver a elaborar un lote si tiene la completa seguridad de que el 10 por ciento de los artículos son defectuosos. Decide seleccionar una muestra al azar de 10 artículos sin reemplazo y volver a producir el lote si éste contiene uno o más artículos defectuosos. ¿Este procedimiento parece razonable?*

50. *Una firma de transporte tienen un contrato para enviar una carga de mercancías de la ciudad W a la ciudad Z. No hay rutas directas que enlacen W con Z, pero hay seis carreteras de W a X y cinco de X a Z. ¿Cuántas rutas en total deben considerarse?*

51. *Un estado tienen un millón de vehículos registrados y está considerando emplear placas de licencia con seis símbolos en los que los primeros tres sean letras y los últimos tres, dígitos. ¿Es éste esquema factible?*

52. *El gerente de una pequeña planta desea determinar el número de maneras en que puede asignar trabajadores al primer turno. Cuenta con 15 hombres que pueden servir como operadores del equipo de producción, 8 que pueden desempeñarse como personal de mantenimiento y 4 que pueden ser supervisores. Si el*

turno requiere 6 operadores, 2 trabajadores de mantenimiento, y 1 supervisor, ¿de cuántas maneras puede integrarse el primer turno?

53. *Un lote de producción tiene 100 unidades de las cuales 20 se sabe que están defectuosas. Una muestra aleatoria de 4 unidades se selecciona sin reemplazo. ¿Cuál es la probabilidad de que la muestra no contenga más de 2 unidades defectuosas?*

54. *En la inspección de lotes de mercancías que están por recibirse, se emplea la siguiente regla de inspección en lotes que contienen 300 unidades; se selecciona una muestra al azar de 10 artículos. Si no hay más que un artículo defectuoso en la muestra, se acepta el lote. De otro modo se regresa al vendedor. Si la fracción defectuosa en el lote original es p', determinar la probabilidad de aceptar el lote como una función de p'*

55. *En una planta de plásticos, 12 tubos vacían diferentes químicos en un tanque de mezcla. Cada tubo tienen una válvula de cinco posiciones que mide el flujo dentro del tanque. Un día, mientras se experimenta con diferentes mezclas, se obtiene una solución que emite un gas venenoso, no habiéndose registrado los valores en las válvulas. ¿Cuál es la probabilidad de obtener esta misma solución cuando se experimenta de nuevo de manera aleatoria?*

56. *Ocho hombres y ocho mujeres con las mismas habilidades solicitan dos empleos. Debido a que los dos nuevos empleados deben trabajar estrechamente, sus personalidades deben ser compatibles. Para lograr esto, el administrador de personal ha aplicado una prueba y debe comparar las calificaciones para cada posibilidad. ¿Cuántas comparaciones debe efectuar el administrador?*

57. *En forma casual, un químico combinó dos sustancias de laboratorios que produjeron un producto conveniente. Desafortunadamente, su asistente no registró los nombres de los ingredientes. Hay cuarenta sustancias disponibles en el laboratorio.*

Si las dos en cuestión deben encontrarse mediante experimentos sucesivos de ensayo y error, ¿cuál es el número máximo de pruebas que pueden realizarse?

58. *Un prisionero político será enviado a Siberia o a los Urales. Las probabilidades de que lo envíen a estos dos lugares son 0.6 y 0.4, respectivamente. Se sabe además que si un residente de Siberia se elige al azar hay una probabilidad de 0.5 de que lleve un abrigo de piel, en tanto que la probabilidad para lo mismos es de 0.7 en el caso de un residente de los Urales. Al llegar al exilio, la primera persona que ve el prisionero no lleva un abrigo de piel. ¿Cuál es la probabilidad de que esté en Siberia?*

59. *Se diseña un dispositivo de frenado para evitar que un automóvil patine en el que incluye un sistema electrónico e hidráulico. El sistema completo puede descomponerse en tres subsistemas en serie que operan de manera independiente: un sistema electrónico, un sistema hidráulico y un accionador mecánico. En un frenado particular, las confiabilidades de estas unidades son aproximadamente 0.995, 0.993 y 0.994, respectivamente. Estime la confiabilidad del sistema.*

60. *Dos bolas se extraen de una urna que contiene m bolas numeradas del 1 a m. Se conserva la primera bola si tiene el número 1, y se regresa en caso contrario. ¿Cuál es la probabilidad de que la segunda bola extraída tenga el número 2?*

61. *Se eligen dos dígitos al azar de los dígitos del 1 al 9 y la selección es sin reemplazo (el mismo dígito no puede escogerse en ambas selecciones). Si la suma de los dígitos es par, encuentre la probabilidad de que ambos dígitos sean impares.*

62. *En cierta universidad 20 por ciento de los hombre y 1 por ciento de las mujeres miden más de dos metros de altura. Asimismo, 40 por ciento de los estudiantes son mujeres. Si se selecciona un estudiante al azar y se observa que mide más de dos metros de altura, ¿Cuál es la probabilidad de que sea mujer?*

63. *En un centro de maquinaria hay cuatro máquinas automáticas que producen tornillos. Un análisis de los registros de inspección anteriores produce los siguientes datos:*

Máquina	*Porcentaje de producción*	*Porcentaje de defectos producidos*
1	*15*	*4*
2	*30*	*3*
3	*20*	*5*
4	*35*	*2*

Las máquinas 2 y 4 son más nuevas y se les ha asignado más producción que a las máquinas 1 y 3. Suponga que la combinación de inventarios refleja los porcentajes de producción indicados.

a) *Si se elige un tornillo al azar del inventario, ¿cuál es la probabilidad de que esté defectuoso?*

b) *Si se elige un tornillo y se encuentra que está defectuoso, ¿cuál es la probabilidad de que se haya producido en la máquina 3?*

64. *Supóngase que $A \subset B$. Demuéstrese que $B' \subset A'$.*

65. *Para tres sucesos cualesquiera A, B y C, demuéstrese que*

$$A(B \cup C) = (A \cap B) \cup (A \cap C).$$

66. *Para dos sucesos cualesquiera A y B, demuéstrese que*

$$(A \cup B)' = A' \cap B' \text{ y } (A \cap B)' = A' \cup B'.$$

67. *4. Para cualquier conjunto de sucesos $A_i (i \in I)$, demuéstrese que*

a) $\left(\bigcup_{i \in I} A_i\right)' = \bigcap_{i \in I} A_i'$

b) $\left(\bigcap_{i \in I} A_i\right)' = \bigcup_{i \in I} A_i'$

68. *5. Supóngase que se selecciona una carta de una baraja de veinte cartas que contiene diez cartas rojas numeradas del 1 al 10 y diez cartas azules numeradas del 1 al 10. Sea A el suceso de seleccionar una carta con un número par, sea B el suceso de seleccionar una carta azul y sea C el suceso de seleccionar una carta con un número menor que 5. Descríbanse el espacio muestral S y cada uno de los siguientes sucesos en palabras y en subconjuntos de S:*

   a) $A \cap B \cap C$.
   b) $B \cap C'$.
   c) $A \cup B \cup C$.
   d) $A \cap (B \cup C)$.
   e) $A' \cap B' \cap C'$.

69. *Un estudiante seleccionado de una clase puede ser chico o chica. Si la probabilidad de que un chico sea seleccionado es 0.3, ¿cuál es la probabilidad de que sea seleccionada una chica?*

70. *Se selecciona una bola de una urna que contiene bolas rojas, blancas, azules, amarillas y verdes. Si la probabilidad de seleccionar una bola roja es $\frac{1}{5}$ y la de seleccionar una blanca es $\frac{2}{5}$ ¿cuál es la probabilidad de seleccionar una bola azul, amarilla o verde?*

71. *Si la probabilidad de que un estudiante A suspenda un cierto examen de estadística es 0.5, la probabilidad de que un estudiante B suspenda el examen es 0.2 y la probabilidad de que ambos estudiantes A y B suspendan el examen es 0.1,*

   a) *¿cuál es la probabilidad de que al menos uno de estos dos estudiantes suspenda el examen?*
   b) *¿cuál es la probabilidad de que ni el estudiante A ni el B suspendan el examen?*
   c) *¿cuál es la probabilidad de que exactamente uno de los dos estudiantes suspenda el examen?*

72. *Considérense dos sucesos A y B tales que $P(A) = \frac{1}{3}$ y $P(B) = \frac{1}{2}$. Determínese el valor de $P(B \cap A')$ para cada una de las siguientes condiciones:*

    a) *A y B son disjuntos*

    b) *$A \subset B$*

    c) *$P(A \cap B) = \frac{1}{8}$*

73. *Si el 50 % de las familias de cierta ciudad están suscritas al periódico matinal, el 65 % de las familias al periódico vespertino y el 85 % al menos a uno de los dos periódicos, ¿ cuál es la proporción de familias que están suscritas a los dos periódicos?*

74. *Considérense dos sucesos A y B con $P(A) = 0{,}4$ y $P(B) = 0{,}7$. Determínense los posibles valores máximo y mínimo de $P(A \cap B)$ y las condiciones en las cuales se consigue cada uno de estos valores.*

75. *Demuéstrese que para dos sucesos A y B cualesquiera, la probabilidad de que exactamente uno de los dos sucesos ocurra está dada por la expresión*

$$P(A) + P(B) - 2P(A \cap B).$$

76. *Se selecciona un punto (x, y) del cuadrado S que contiene todos los puntos (x, y) tales que $0 \leq x \leq 1$ y $0 \leq y \leq 1$. Supóngase que la probabilidad de que el punto seleccionado pertenezca a cualquier subconjunto específico de S es igual al área de ese subconjunto. Determínese la probabilidad de cada uno de los siguientes subconjuntos:*

    a) *el subconjunto de puntos tales que $(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 > \frac{1}{4}$*

    b) *el subconjunto de puntos tales que $\frac{1}{2} < x + y < \frac{3}{2}$*

    c) *el subconjunto de puntos tales que $y < 1 - x^2$*

    d) *el subconjunto de puntos tales que $x = y$.*

77. *Sea* $A_1, A_2, A_3, \cdots$ *una serie numerable de eventos y sea* $B_1, B_2, B_3, \cdots$ *otra serie numerable de eventos tal que*

$$B_1 = A_1,\ B_2 = A_1' \cap A_2,\ B_3 = A_1' \cap A_2' \cap A_3...$$

Demuéstrese que

a) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(B_i)$

b) $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(B_i)$ *para* $n = 1, 2, 3, ...$

78. *Sea* $A_1, A_2, A_3, \cdots A_n$ *una serie cualquiera de eventos, demostrar que*

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

79. *Supóngase que una urna contiene una carta azul y cuatro cartas rojas. A, B, C y D. Supóngase también que dos de estas cinco cartas se extraen al azar sin reemplazo*

a) *Si se sabe que se ha extraído la carta A, ¿cuál es la probabilidad de que ambas cartas sean rojas?*

b) *Si se sabe que se ha extraído una carta roja, ¿cuál es la probabilidad de que ambas cartas sean rojas?*

80. *La probabilidad de que cualquier niño de una familia determinada tenga ojos azules es 1/4, y esta característica es heredada por cada niño de la familia independientemente de los demás. Si hay cinco niños en la familia y se sabe que al menos uno de estos niños tiene ojos azules,*

a) *¿cuál es la probabilidad de que al menos tres de los niños tengan ojos azules?*

b) *Considérese la familia de cinco nios*

   1) *Si se sabe que el niño más pequeño de la familia tiene los ojos azules, ¿cuál es la probabilidad de que al menos tres de los niños tengan ojos azules?*

2) *Explíquese por qué la respuesta es diferente de la respuesta de la parte (a)*

81. *Considérese la siguiente versión del juego de dados: El jugador lanza dos dados. SÍ la suma en el primer lanzamiento es 7 u 11, el jugador gana el juego. Si la suma en el primer lanzamiento es 2, 3 ó 12, el jugador pierde. Sin embargo, si la suma en el primer lanzamiento es 4, 5, 6, 8, 9 ó 10, entonces se lanzan los dos dados una y otra vez hasta que la suma sea 7, lío el valor original. Si el valor original se obtiene por segunda vez antes de obtener 7 u 11, entonces el jugador gana. Si se obtiene un total de 7 o de 11 antes de obtener el valor original por segunda vez, entonces el jugador pierde. Determínese la probabilidad de que el jugador gane este juego.*

82. *En una ciudad determinada, el 30 % de las personas son conservadores, el 50 % son liberales y el 20 % son independientes. Los registros muestran que en unas elecciones concretas, votaron el 65 % de los conservadores, el 82 % de los liberales y el 50 % de los independientes. Si se selecciona al azar una persona de la ciudad y se sabe que no votó en las elecciones pasadas, ¿ cuál es la probabilidad de que sea un liberal?*

83. *Supóngase que cuando una máquina está correctamente ajustada, el 50 % de los artículos que produce son de alta calidad y el otro 50 % son de calidad media. Supóngase, sin embargo, que la máquina está mal ajustada durante el 10 % del tiempo y que, en estas condiciones, el 25 % de los artículos producidos por ella son de alta calidad y el 75 % de los artículos son de calidad media.*

   a) *Supóngase que cinco artículos producidos por la máquina en un tiempo determinado son seleccionados al azar e inspeccionados. Si cuatro de estos artículos son de alta calidad y uno es de calidad media, ¿cuál es la probabilidad de que la máquina estuviera correctamente ajustada durante ese tiempo?*

b) *Supóngase que se selecciona un artículo adicional, que fue producido por la máquina al mismo tiempo que los otros cinco, y resulta ser de calidad media. ¿Cuál es la nueva probabilidad final de que la máquina estuviera correctamente ajustada?*

84. *Supóngase que una caja contiene cinco monedas y que la probabilidad de obtener cara en un lanzamiento es distinta para cada moneda. Sea* p_i *la probabilidad de obtener cara al lanzar la* $i-$ *ésima moneda* $(i = 1, ..., 5)$ *y Supóngase que* $p_1 = 0$, $p_1 = \frac{1}{4}$, $p_3 = 1/2$, $p_4 = 3/4$ *y* $p_5 = 1$. *Si se selecciona una moneda de la caja al azar y que al lanzarla una vez se obtiene una cara. ¿Cuál es la probabilidad final de que se haya seleccionado la* $i-$ *ésima moneda* $(i = 1, ..., 5)$*?*