



CORPORACIÓN
UNIVERSITARIA
REMINGTON

FACULTAD DE CIENCIAS BÁSICAS E INGENIERÍA
Asignatura transversal
Estadística Descriptiva
Unidad III

CORPORACIÓN UNIVERSITARIA REMINGTON
DIRECCIÓN PEDAGÓGICA

Este material es propiedad de la Corporación Universitaria Remington (CUR), para los estudiantes de la CUR en todo el país.

2011

CRÉDITOS



El módulo de estudio de la asignatura transversal Estadística Descriptiva unidad III es propiedad de la Corporación Universitaria Remington. Las imágenes fueron tomadas de diferentes fuentes que se relacionan en los derechos de autor y las citas en la bibliografía. El contenido del módulo está protegido por las leyes de derechos de autor que rigen al país.

Este material tiene fines educativos y no puede usarse con propósitos económicos o comerciales.

AUTOR

Ángela Quintero Echeverri

Tecnóloga en Sistematización de datos. Estudiante de último semestre de Psicología, Coordinadora académica de la Escuela de Ciencias Básicas e Ingeniería, Profesora de la Corporación Universitaria Remington
rosa.quintero@remington.edu.co

Nota: el autor certificó (de manera verbal o escrita) No haber incurrido en fraude científico, plagio o vicios de autoría; en caso contrario eximió de toda responsabilidad a la Corporación Universitaria Remington, y se declaró como el único responsable.

RESPONSABLES

ESCUELA DE CIENCIAS BÁSICAS E INGENIERÍA

Director Dr. Mauricio Sepúlveda

Director Pedagógico

Octavio Toro Chica

dirpedagogica.director@remington.edu.co

Coordinadora de Medios y Mediaciones

Angélica Ricaurte Avendaño

mediaciones.coordinador01@remington.edu.co

GRUPO DE APOYO

Personal de la Unidad de Medios y Mediaciones

EDICIÓN Y MONTAJE

Primera versión. Febrero de 2011.

Derechos Reservados



Esta obra es publicada bajo la licencia Creative Commons. Reconocimiento-No Comercial-Compartir Igual 2.5 Colombia.

TABLA DE CONTENIDO

1.	INTRODUCCIÓN	4
2.	MAPA DE LA ASIGNATURA	6
3.	DATOS CUANTITATIVOS AGRUPADOS EN INTERVALOS DE CLASE	7
3.1.	Tablas para Datos cuantitativos agrupados en intervalos de clase	8
3.2.	Medidas de tendencia central para datos cuantitativos agrupados en intervalos de clase .	13
3.3.	Medidas de posición relativa para datos cuantitativos agrupados en intervalos de clase...	17
3.4.	Medidas de variabilidad o dispersión para datos cuantitativos	20
3.5.	Glosario	28
3.6.	Fuentes Bibliográficas	29
3.7.	Fuentes Digitales o Electrónicas.....	30

1. INTRODUCCIÓN

La estadística es la ciencia de los datos, por tanto, cuando se aplica el método estadístico, se recolectan, se sintetizan, se organizan, se analizan y se interpretan los datos.

La estadística descriptiva se encarga de describir los datos por medio de tablas, gráficos y medidas; en este módulo se explicará cómo lograrlo. Se pretende que el estudiante aplicando paulatinamente cada paso que se explica, lo logre.

Para que se llegue al objetivo terminal: Estudiar métodos de organización, análisis y presentación de un conjunto de datos asociados a una situación problemática por medio del modelo de representación estadístico caracterizando un conjunto de datos a partir de mediciones estadísticas para la obtención de conclusiones que sirvan de apoyo en la toma de decisiones, se ha diseñado el módulo de una forma especial e innovadora; en la primera parte se definen los conceptos generales que se requieren en estadística y posteriormente se han dividido las unidades de acuerdo con los tipos de datos, a saber:

- ✓ Datos cualitativos.
- ✓ Datos cuantitativos ordenados en fila.
- ✓ Datos cuantitativos agrupados en intervalos.

Nota: En cada tipo de datos, se explica cómo se pueden identificar, recolectar, organizar y describir por medio de tablas gráficos y medidas para finalmente llegar a conclusiones y basado en éstas, tomar las decisiones correspondientes.

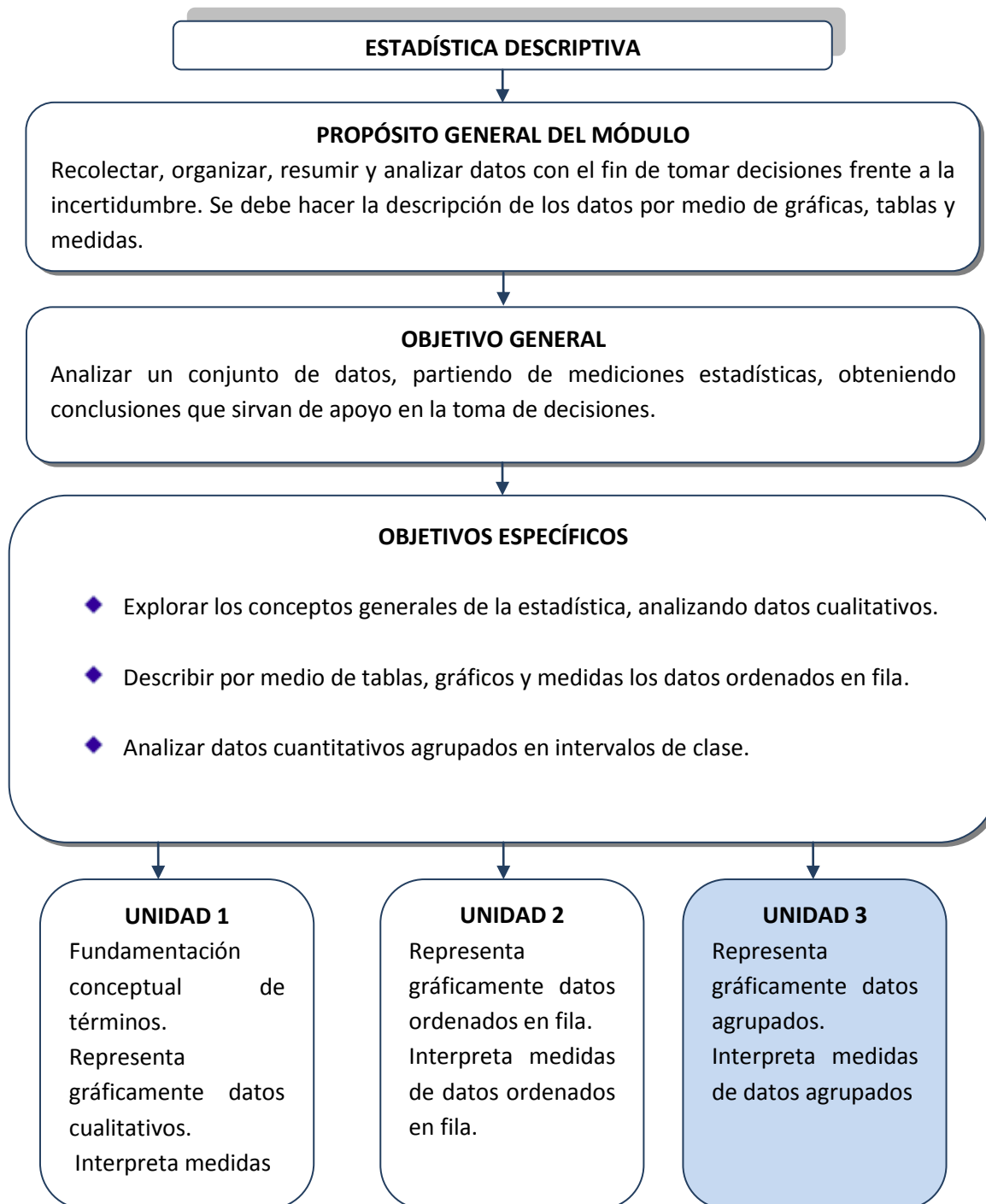
El módulo está construido con un lenguaje sencillo, con ejercicios aplicados a la cotidianidad y a situaciones prácticas inherentes al plan de estudio del estudiante de La Corporación Universitaria Remington de cualquier programa existente en la universidad, con el fin de que, de una forma pedagógica, se aprenda y se logren los objetivos.

La Estadística Descriptiva le proporciona al profesional las herramientas que necesita para recoger, organizar, analizar y presentar datos con el fin de tomar decisiones.

Debido a que el estudiante, de educación a distancia de La Corporación Universitaria Remington, requiere un método de aprendizaje de forma tal que el profesor sea un orientador y que dicho estudiante sea autogenerador de su conocimiento, obviamente con la asesoría del docente, se ha creado este módulo.

Este módulo está diseñado con un lenguaje sencillo y con ejercicios que son aplicados a la cotidianidad del estudiante, a su entorno social y laboral, pues de esta forma podrá realizar investigaciones estadísticas en un futuro, ya sea a corto plazo, en otras asignaturas, o a largo plazo cuando esté realizando su labor como profesional.

2. MAPA DE LA ASIGNATURA



3. DATOS CUANTITATIVOS AGRUPADOS EN INTERVALOS DE CLASE

OBJETIVO GENERAL

Analizar datos cuantitativos agrupados en intervalos de clase.

OBJETIVOS ESPECÍFICOS

- ◆ Organizar una serie de datos numéricos en intervalos.
- ◆ Construir e interpretar las frecuencias acumuladas.
- ◆ Construir gráficos para datos agrupados en intervalos.
- ◆ Calcular, interpretando los resultados, las medidas de tendencia central para datos agrupados en intervalos.
- ◆ Calcular e interpretar las medidas de posición relativa para datos agrupados en intervalos.
- ◆ Calcular e interpretar las medidas de variación o dispersión para datos cuantitativos.

Prueba Inicial

En la compañía “Universal S. A.” se hizo un estudio sobre los sueldos, a continuación se dan los resultados, en miles de \$.

TABLA DE DISTRIBUCIÓN DE FRECUENCIAS									
LI	LS	fai	fri	%	faai	frai	xi	xi . Fai	$(xi - \mu)^2 \cdot Fai$
297	748	15	0.375	37.5	15	0.375	522.5	7837.5	4578429.38
748	1199	10	0.25	25	25	0.625	973.5	9735	102971.756
1199	1650	9	0.225	22.5	34	0.85	1424.5	12820.5	1099509.53
1650	2101	4	0.1	10	38	0.95	1875.5	7502	2563361.1
2101	2552	1	0.025	2.5	39	0.975	2326.5	2326.5	1566314.83
2552	3003	1	0.025	2.5	40	1	2777.5	2777.5	2898591.38
		40	1	100				42999	12809178

De acuerdo con los datos de la tabla, responde lo siguiente:

El promedio de sueldos de la compañía “Universal S.A” es de:

- a) \$ 1.074.975 b) \$ 2.650.000 c) \$ 1.5689.532 d) \$ 3.420.000

El 50%: de sueldos de la compañía “Universal S.A” es de:

- a) \$ 973.500 b) \$ 2.650.000 c) \$ 1.5689.532 d) \$ 3.420.000

El sueldo de la compañía “Universal S.A” más frecuente es:

- a) \$ 973.500 b) \$ 2.650.000 c) \$ 1.5689.532 d) \$ 635.250.

Se puede afirmar que en la compañía “Universal S.A” la mayoría de los sueldos son:

- a) Altos b) Regulares c) Bajos

3.1. Tablas para Datos cuantitativos agrupados en intervalos de clase

Cuando se tienen muchos datos y cada uno tiene una frecuencia baja, es decir que no se repiten o se repiten pocas veces es necesario agrupar estos datos; en cada clase o categoría de la distribución irá un intervalo. Existen varios métodos para agruparlos; uno de ellos es que los intervalos sean de la forma: $LI < X \leq LS$ o sea que se incluye el LS (límite superior) y no se incluye LI (límite inferior) se deben seguir los siguientes pasos:

1. Se calcula el rango

$R = \text{dato mayor} - \text{dato menor}$

2. Se calcula el # de intervalos (k) que tendrá la distribución

$K = 1 + 3,3 \cdot \log n$ donde n es el número de datos

K se redondea a entero

3. Se calcula la amplitud que va a ser la misma en todos los intervalos

$A = R/K$

La amplitud se aproxima al entero siguiente así sea entero o sea se le suma 1 a la amplitud.

4. Se calcula el rango ampliado

$Ra = A \cdot K$

5. Se calcula una cantidad que debe ser diferente de cero

$\text{Cant} = \frac{Ra - R}{2}$

6. Se calcula el primer límite inferior que tendrá la distribución.

$LI = \text{dato menor} - \text{cant}$

Ejemplo 1:

Entre los estudiantes de Administración de Corporación Universitaria Remington se hizo una selección aleatoria y se analizó la edad en que iniciaron sus estudios en la universidad, los siguientes son los datos:

25	36	40	45	48	32	17	22	18	20
21	44	46	19	19	26	24	28	29	31
39	36	35	33	32	34	41	42	31	30
20	45	43	23	22	21	38	34	37	43
19	20	23	25	22	25	27	24	19	18
20	21	22	25	24	27	29	19	23	21

Ficha técnica:

POBLACIÓN: Estudiantes de Administración de Corporación Universitaria Remington.

MUESTRA: 60 estudiantes elegidos al azar.

DESCRIPCIÓN DE LA VARIABLE: La edad en que iniciaron sus estudios en la universidad.

TIPO DE VARIABLE: Cuantitativa continua (debido a que pertenece a un intervalo)

Como podemos apreciar, existen muchos datos con baja frecuencia, por tanto hay que agruparlos en intervalos; vamos a aplicar los pasos:

1. Se calcula el rango

$R = \text{dato mayor} - \text{dato menor}$

$$R = 48 - 17 = 31$$

2. Se calcula el # de intervalos (k) que tendrá la distribución

$K = 1 + 3,3 \cdot \log n$ donde n es el número de datos

$$K = 1 + 3,3 \cdot \log 60 = 6,868$$

K se redondea a entero $K = 7$

3. Se calcula la amplitud que va a ser la misma en todos los intervalos

$$A = R/K \quad A = 31/7 = 4,428$$

La amplitud se aproxima al entero siguiente así sea entero o sea se le suma 1 a la amplitud. $A = 5$

4. Se calcula el rango ampliado

$$Ra = A \cdot K = 5 \cdot 7 = 35$$

5. Se calcula una cantidad que debe ser diferente de cero

$$\text{Cant} = Ra - R = 35 - 31 = 4$$

6. Se calcula el primer límite inferior que tendrá la distribución.

$LI = \text{dato menor} - \text{cant}$

$$LI = 17 - 2 = 15$$

Como todos los intervalos tiene la misma amplitud, $LS = LI + A$

Primer límite superior: $LS = 15 + 5 = 20$

Recordemos que vamos a agrupar intervalos de la forma $LI < X \leq LS$ para el primer intervalo contamos los datos que son mayores que 15 y menores o iguales que 20 y en total encontramos 12 datos esta será la primera frecuencia absoluta.

Todo límite superior pasa a ser límite inferior en el siguiente intervalo y para calcular el segundo límite superior aplicamos la fórmula $LS = LI + A$ por tanto el segundo intervalo tendrá como límite inferior 20 y como límite superior $20 + 5 = 25$ y contamos los datos que son mayores que 20 y menores o iguales que 25 y en total encontramos 12 datos esta será la segunda frecuencia absoluta. Así sucesivamente.

TABLA DE DISTRIBUCIÓN DE FRECUENCIAS

LI	LS	fai	fri	%	Xi	faai	frai	% acum
15	20	12	0,2	20	17,5	12	0,2	20
20	25	18	0,3	30	22,5	30	0,50	50
25	30	7	0,117	11,7	27,5	37	0,617	61,7
30	35	8	0,133	13,3	32,5	45	0,75	75
35	40	6	0,1	10	37,5	51	0,85	85
40	45	7	0,117	11,7	42,5	58	0,967	96,7
45	50	2	0,33	3,3	47,5	60	1	100
		N = 60	$\Sigma = 1$	$\Sigma = 100$				

Ejemplo 2:

Los siguientes son los sueldos de los tecnólogos de Medellín de acuerdo con una muestra elegida aleatoriamente (en miles de pesos).

En este caso ya se dan los datos agrupados, por tanto no hay que aplicar los pasos, simplemente se complementa la tabla.

LI	LS	fa
400	600	30
600	800	100
800	1000	55
1000	1200	240
1200	1400	25
1400	1600	20

TABLA DE DISTRIBUCIÓN DE FRECUENCIAS

LI	LS	fai	fri	%	Xi	faai	frai	% acum
400	600	30	0,064	6,4	500	30	0,064	6,4
600	800	100	0,213	21,3	700	130	0,277	27,7
800	1000	55	0,117	11,7	900	185	0,394	39,4
1000	1200	240	0,511	51,1	1100	425	0,904	90,4
1200	1400	25	0,053	5,3	1300	450	0,957	95,7
1400	1600	20	0,043	4,3	1500	470	1,000	100,0
		n = 470	Σ = 1	Σ = 100				

GRÁFICOS PARA DATOS AGRUPADOS

1. **EL HISTOGRAMA:** “Son diagramas de barras verticales en los que se construyen barras rectangulares en los límites de cada clase.” (Berenson, M. L. Y LEVINE, D. M. 1996, p. 70).

Puede ser de frecuencias absolutas o de porcentajes; se realiza colocando en el eje X los límites y en el eje Y las frecuencias absolutas o los porcentajes Según el caso.

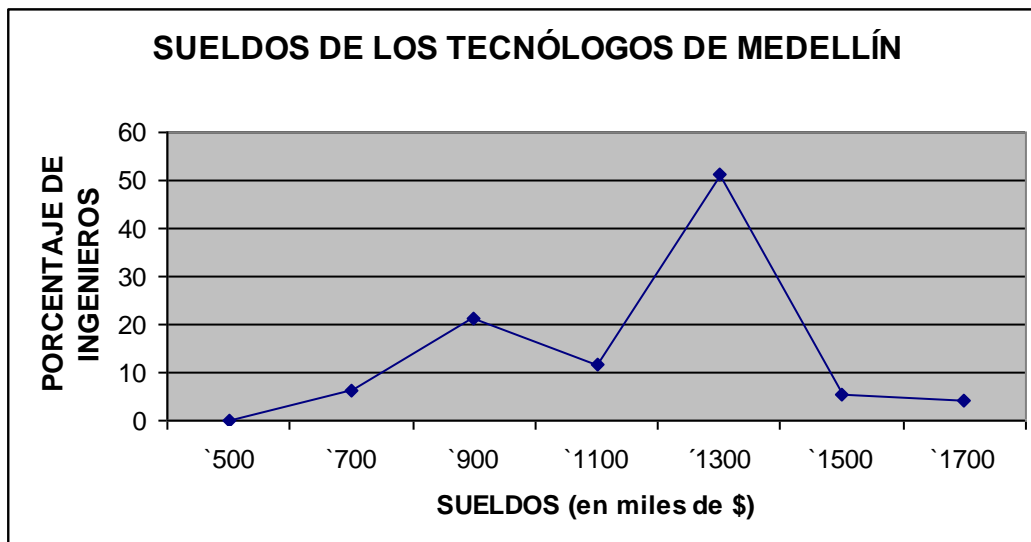
Marcas de clase (Xi):

Es un valor que identifica a cada intervalo “es el punto medio de cada intervalo” (Berenson, M. L. Y LEVINE, D. M., 1996, p. 38).

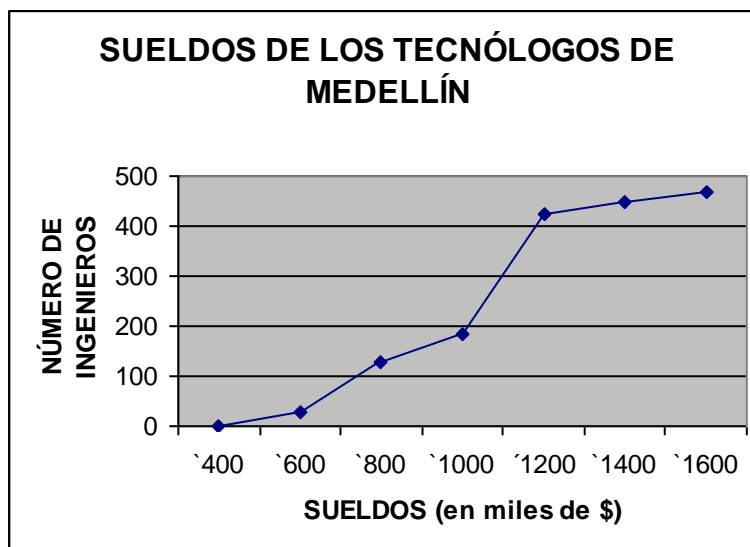
$$xi = \frac{LI + LS}{2}$$

2. **POLÍGONO:** “Se puede obtener uniendo cada punto medio (marca de clase) de los rectángulos del histograma con líneas rectas, teniendo cuidado de agregar al inicio y al final marcas de clase adicionales, con el objeto de asegurar la igualdad del áreas”. (<http://sitios.ingenieria-usac.edu.gt/estadistica/estadistica2/estadisticadescriptiva.html>)

Puede ser de frecuencias absolutas o de porcentaje. Se realiza colocando en el eje X las marcas de clase y en el eje y las frecuencias absolutas o los porcentajes según el caso. El siguiente es el polígono de porcentaje del ejemplo 2



3. **OJIVA:** puede ser de frecuencias absolutas acumuladas o de porcentajes acumulado. Se realiza colocando en el eje X los límites y en el eje Y las frecuencias acumuladas. La siguiente ojiva es de faai del ejemplo 2



Ejercicios tema 1

En cada uno de los siguientes ejercicios realice:

- ◆ La ficha técnica.
- ◆ Interprete el intervalo: mayor, menor, más frecuente y menos frecuente.
- ◆ Elabore el Histograma, el Polígono y la Ojiva.
- ◆ Calcule e interprete: las medidas de tendencia central, las medidas de posición relativas y las medidas de variabilidad.

1. En la compañía “La Delicia” se hizo un estudio sobre los sueldos, a continuación se dan los resultados, en miles de \$.

900	500	450	1900	1200	1250	2500	550	1650	1200
1000	550	950	600	750	1300	850	350	1400	700
300	1100	300	600	1600	1500	1000	1800	900	500
650	2000	450	750	850	600	300	1950	3000	1500

3.2. Medidas de tendencia central para datos cuantitativos agrupados en intervalos de clase

1. Media aritmética:

$$\bar{X} \text{ ó } \mu = \frac{\sum Xi \cdot fai}{n \text{ ó } N}$$

Para el ejemplo 1

LI	LS	fai	Xi	Xi*fai
15	20	12	17,5	210
20	25	18	22,5	405
25	30	7	27,5	192,5
30	35	8	32,5	260
35	40	6	37,5	225
40	45	7	42,5	297,5
45	50	2	47,5	95
		N = 60		Σ = 1685

$$\bar{X} = \frac{1685}{60} = 28,03$$

El promedio de la edad en que iniciaron sus estudios en la universidad los estudiantes de Administración de Corporación Universitaria Remington es de 28,03 años

Para el ejemplo 2

LI	LS	fai	Xi	Xi*fai
400	600	30	500	15000
600	800	100	700	70000
800	1000	55	900	49500
1000	1200	240	1100	264000
1200	1400	25	1300	32500
1400	1600	20	1500	30000
		n = 470		Σ = 461000

$$\bar{X} = \frac{461000}{470} = 980,851$$

El promedio de los sueldos de los tecnólogos de Medellín es de \$980.851

4. Moda:

$$Mo = LI + \left(\frac{d1}{d1 + d2} \right) * A$$

$d1 = fai - fa(i-1) \quad d2 = fai - fa(i+1) \quad A = LSi - Lli$

El intervalo que contiene la Moda es el más frecuente

El intervalo más frecuente es el **rojo** en las tablas, por lo tanto:

Para el ejemplo 1

$$LI = 20 \quad d1 = 18 - 12 = 6 \quad d2 = 18 - 7 = 11 \quad A = 25 - 20 = 5 \text{ reemplazando}$$

$$Mo = 20 + \left(\frac{6}{6 + 11} \right) * 5 = 21,764$$

La edad en que iniciaron sus estudios en la universidad los estudiantes de Administración de Corporación Universitaria Remington más frecuente es de 21,764 años

Para el ejemplo 2

$$LI = 1000 \quad d1 = 240 - 55 = 185 \quad d2 = 240 - 25 = 215 \quad A = 1200 - 1000 = 200$$

Reemplazando en la fórmula:

$$Mo = 1000 + \left(\frac{185}{185 + 215} \right) * 200 = 1092,5$$

Como está dado en miles, se multiplica $1092,5 * 1000 = 1'092.500$

El sueldo de los tecnólogos de Medellín más frecuente es de \$1'092.500

5. Mediana:

$$Me = LI + \left(\frac{\frac{n}{2} - faa(i-1)}{fai} \right) * A$$

Para ubicar el intervalo que contiene la mediana se puede hacer por:

1. $n/2$ y se busca en faa
2. 50 % y se busca en porcentaje acumulado

Para el ejemplo 1

El intervalo que contiene la mediana lo ubicamos $n/2 = 60/2 = 30$ este valor, buscando en la frecuencia absoluta acumulada, está en el 2º intervalo es el mismo de la moda, pero no siempre dan en el mismo, existen ciertas distribuciones que cumplen con esta característica. Si observamos, acá está el 50% en el porcentaje acumulado. Reemplazando en la fórmula:

$$Me = 20 + \left(\frac{30 - 12}{18} \right) * 5 = 25$$

El 50% de la edad en que iniciaron sus estudios en la universidad los estudiantes de Administración de Corporación Universitaria Remington es de 25 años.

Para el ejemplo 2

El intervalo que contiene la mediana lo ubicamos $n/2 = 235$ este valor está en 425 en el mismo de la moda, no siempre dan en el mismo, existen ciertas distribuciones que cumplen con esta característica. Si observamos, acá está el 50% en el porcentaje acumulado. Reemplazando en la fórmula:

$$Me = 100 + \left(\frac{235 - 185}{240} \right) * 200 = 1041,667$$

Como está dado en miles, se multiplica

$$1041,667 * 1000 = 1\,041.667$$

El 50% de los sueldos de los tecnólogos de Medellín es de \$1\,041.667

Ejercicios tema 2

1. A continuación se dan las notas de los estudiantes de sistemas de acuerdo con una muestra elegida al azar:

LI	LS	fai
0	1	15
1	2	15
2	3	20
3	4	28
4	5	22

Calcule e interprete: La Media Aritmética, La Moda y la Mediana de ambos ejercicios.

Calcule e interprete: para el ejercicio 1 Q1, D3, P5 y para el ejercicio 2, Q3, D9 y P55

3.3. Medidas de posición relativa para datos cuantitativos agrupados en intervalos de clase

Estas medidas dividen la distribución en partes iguales, así como La Mediana, por tanto se calculan e interpretan similar a ella. Se tienen las siguientes:

1. **Cuartiles (Q):** dividen la distribución en 4 partes iguales así:

Q1=25%

Q2=50%

Q3=75%

Q4=100%

$$Q = LI + \left(\frac{\#Q * n / 4 - faa(i-1)}{fai} \right) * A$$

Para ubicar el intervalo que contiene Q se puede hacer por

1) $\frac{\#Q * n}{4}$ y se busca en faa

2) el porcentaje respectivo se busca en el porcentaje acumulado

Si nos piden el cuartil 1 del ejemplo 2:

$Q_1 = 1 * 470 / 4 = 117,5$ en la faa está en 130 o el 25% está acá también, en la tabla es el intervalo azul. Reemplazando en la fórmula:

$$Q_1 = 600 + \left(\frac{117,5 - 30}{100} \right) * 200 = 775$$

El 25% de los sueldos de los tecnólogos de Medellín es de \$ 775.000 o menos.

2. **Deciles (D):** “dividen la distribución en 10 partes iguales” (Murray, 1995, p. 66) así:

D1=10%

D2=20%

D10=100%

$$D = LI + \left(\frac{\#D * n / 10 - faa(i-1)}{fai} \right) * A$$

Para ubicar el intervalo que contiene D se puede hacer por

1) $\frac{\#D \cdot n}{10}$ y se busca en faa

2) el porcentaje respectivo se busca en el porcentaje acumulado

Si nos piden el decil 3 del ejemplo 2:

$D_3 = 3 \cdot 470 / 10 = 141$ en la faai está en 130 o el 30% está acá también, en la tabla es el intervalo **marrón**. Reemplazando en la fórmula:

$$D_3 = 800 + \left(\frac{141 - 130}{55} \right) \cdot 200 = 840$$

El 30% de los sueldos de los tecnólogos de Medellín es de \$ 840.000 o menos.

3. Percentiles (P): “dividen la distribución en 100 partes iguales” (Murray, 1995, p. 66), así:

P1=1%

P2=2%

P100=100%

$$P = LI + \left(\frac{\#P \cdot n / 100 - faa(i-1)}{fai} \right) \cdot A$$

Para ubicar el intervalo que contiene P se puede hacer por

1) $\frac{\#P \cdot n}{100}$ y se busca en faa

2) el porcentaje respectivo se busca en el porcentaje acumulado.

Si nos piden el percentil 5 del ejemplo 2:

$P_5 = 5 \cdot 470 / 100 = 23,5$ en la faai está en 30 o el 5% está acá también, en la tabla es el intervalo **fucsia**, el primero. Reemplazando en la fórmula:

$$P_5 = 400 + \left(\frac{23,5 - 0}{30} \right) \cdot 200 = 556,667$$

El 5% de los sueldos de los tecnólogos de Medellín es de \$ 556.667 o menos.

Ejercicios tema 3

Para cada uno de los siguientes ejercicios calcule e interprete las medidas de variabilidad y tome sus propias conclusiones.

1. Las siguientes son las notas de estadística dos grupos de Corporación Universitaria Remington, de una muestra elegida al azar:

grupo 1		grupo 2	
Xi (nota)	fai(nº estud)	Xi(nota)	fai(nº estud)
1	5	1	7
2	10	2	6
3	8	3	8
4	3	4	5
5	10	5	6

El entrenador de un equipo de atletismo está evaluando a tres estudiantes para poderlos incluir en su equipo. Se hicieron competir a estos 3 atletas en 5 carreras de 500 mts. Y se obtuvieron los siguientes resultados (en segundos):

Atleta A	61.3	62.1	61.7	62.9	63.2
Atleta B	62.8	63	62.5	61.9	60.7
Atleta C	61.9	63.7	62.9	61.9	61.5

3.4. Medidas de variabilidad o dispersión para datos cuantitativos

Para entender en que consiste la variabilidad, veamos el siguiente ejemplo:

A continuación se tiene el número de unidades producidas por hora durante un día por dos operarios:

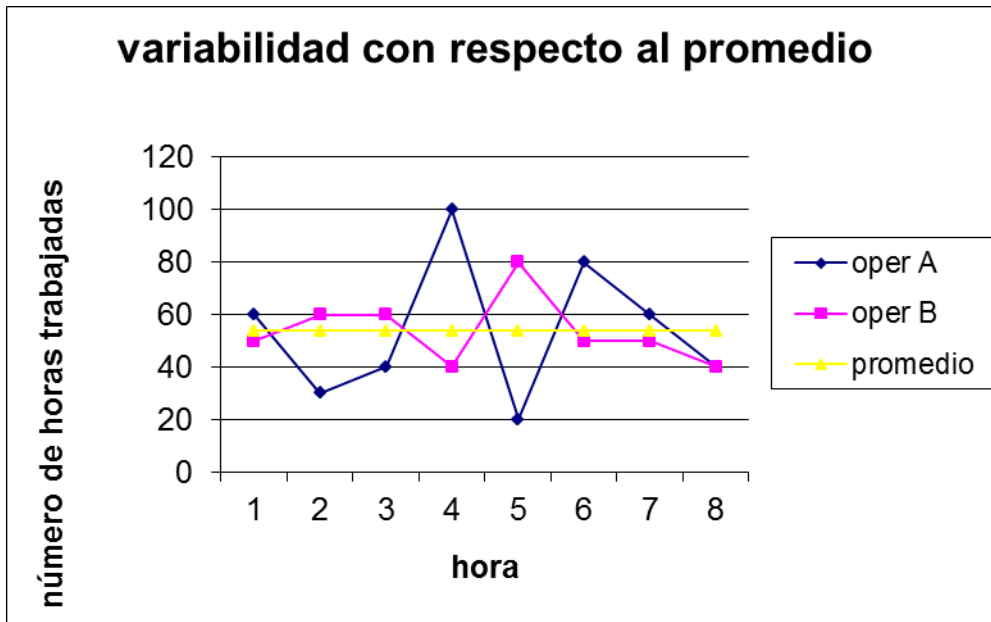
oper A	60	30	40	100	20	80	60	40
Oper B	50	60	60	40	80	50	50	40

Si vamos a buscar el promedio del número de unidades producidas por hora en el día de cada operario, sumamos las unidades y dividimos por las 8 horas así:

$$\mu = \frac{60+30+40+100+20+80+60+40}{8} = \frac{430}{8} = 53.75 \quad \text{para el operario A}$$

$$\mu = \frac{50+60+60+40+80+50+50+40}{8} = \frac{430}{8} = 53.75 \quad \text{Para el operario B}$$

Como podemos apreciar ambos operarios tienen el mismo promedio. Pero veamos algo adicional, vamos a trazar un diagrama de líneas para cada operario:



Como podemos ver, el operario A presenta mayor variación con respecto al promedio que el operario B. Las medidas que vamos a ver a continuación tienen dos aplicaciones:

1. Para analizar cómo varía un conjunto de datos con relación a su propio promedio.
2. Para comparar la variabilidad de dos o más conjuntos de datos entre sí.

Ahora veamos la primera de estas medidas:

1. VARIANZA

Si los datos no están organizados en una tabla de frecuencias, la varianza se calcula así:

Como parámetro, es decir, si los datos se toman de una población:

$$\sigma^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + \dots + (x_n - \mu)^2}{N} = \frac{\sum (x_i - \mu)^2}{N}$$

Como estadística, es decir, si los datos se toman de una muestra:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1} = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

Veamos el ejemplo que tenemos sobre los operarios:

OPERARIO A			OPERARIO B	
X_i	$(X_i - \mu)^2$		X_i	$(X_i - \mu)^2$
60	39,0625		50	14,0625
30	564,0625		60	39,0625
40	189,0625		60	39,0625
100	2139,0625		40	189,0625
20	1139,0625		80	689,0625
80	689,0625		50	14,0625
60	39,0625		50	14,0625
40	189,0625		40	189,0625
$\Sigma=430$	$\Sigma=4987,5$		$\Sigma=430$	$\Sigma=1187,5$

$$\mu_A = \frac{\sum Xi}{N} = \frac{430}{8} = 53.75$$

$$\mu_B = \frac{\sum Xi}{N} = \frac{430}{8} = 53.75$$

$$\sigma_A^2 = \frac{\sum (Xi - \mu)^2}{N} = \frac{4987.5}{8} = 623.438$$

$$\sigma_B^2 = \frac{\sum (Xi - \mu)^2}{N} = \frac{1187.5}{8} = 148.438$$

Pero si los datos están organizados en una tabla de frecuencias, la varianza se calcula

Como parámetro, es decir, si los datos se toman de una población:

$$\sigma^2 = \frac{(x_1 - \mu)^2 \cdot fa_1 + (x_2 - \mu)^2 \cdot fa_2 + (x_3 - \mu)^2 \cdot fa_3 + \dots + (x_n - \mu)^2 \cdot fa_n}{N} = \frac{\sum (x_i - \mu)^2 \cdot fa_i}{N}$$

Como estadística, es decir, si los datos se toman de una muestra:

$$s^2 = \frac{(x_1 - \bar{x})^2 \cdot fa_1 + (x_2 - \bar{x})^2 \cdot fa_2 + (x_3 - \bar{x})^2 \cdot fa_3 + \dots + (x_n - \bar{x})^2 \cdot fa_n}{n - 1} = \frac{\sum (x_i - \bar{x})^2 \cdot fa_i}{n - 1}$$

OPERARIO A

Xi	fai	Xi*fai	(Xi - μ) ² *fai
20	1	20	1139,0625
30	1	30	564,0625
40	2	80	378,125
60	2	120	78,125
80	1	80	689,0625
100	1	100	2139,0625
	N = 8	Σ = 430	Σ = 4987.5

$$\mu = \frac{\sum Xi \cdot fai}{N} = \frac{430}{8} = 53.75$$

$$\sigma^2 = \frac{\sum (Xi - \mu)^2 \cdot fai}{N} = \frac{4987.5}{8} = 623.438$$

OPERARIO B

X_i	f_{ai}	$X_i * f_{ai}$	$(X_i - \mu)^2 * f_{ai}$
40	2	80	378,125
50	3	150	42,1875
60	2	120	78,125
80	1	80	689,0625
	$N = 8$	$\Sigma = 430$	$\Sigma = 1187.5$

$$\mu = \frac{\sum X_i * f_{ai}}{N} = \frac{430}{8} = 53.75$$

$$\sigma^2 = \frac{\sum (X_i - \mu)^2 * f_{ai}}{N} = \frac{1187.5}{8} = 148.438$$

Pasos para calcular la varianza:

1. Se busca el promedio: μ ó \bar{x}
2. Se calcula la desviación de cada dato con respecto al promedio y este resultado se eleva al cuadrado: $(x_i - \mu)^2$ ó $(x_i - \bar{x})^2$
3. Cada uno de estos se multiplican por la frecuencia absoluta, en caso de que exista: $(x_i - \mu)^2 \cdot f_{ai}$ ó $(x_i - \bar{x})^2 \cdot f_{ai}$
4. Se suman los anteriores resultados.
5. Esta sumatoria se divide por el número de datos. N o n-1, según el caso.

Como podemos observar, la varianza da en unidades al cuadrado lo cual resulta ilógico y no se puede interpretar, por esta razón se creó la siguiente medida de variación:

2. DESVIACIÓN ESTÁNDAR

También conocida como desviación típica, “es la raíz cuadrada de la varianza. Expresa la dispersión de la distribución y se expresa en las mismas unidades de medida de la variable. La desviación típica es la medida de dispersión más utilizada en estadística”.

(<http://www.fisterra.com/mbe/investiga/10descriptiva/10descriptiva.asp#introduccion>)

Como parámetro, es decir, si los datos se toman de una población:

$$\sigma = \sqrt{\frac{(x_1 - \mu)^2 \cdot fa_1 + (x_2 - \mu)^2 \cdot fa_2 + (x_3 - \mu)^2 \cdot fa_3 + \dots + (x_n - \mu)^2 \cdot fa_n}{N}}$$

Como estadística, es decir, si los datos se toman de una muestra:

$$S = \sqrt{\frac{(x_1 - \bar{x})^2 \cdot fa_1 + (x_2 - \bar{x})^2 \cdot fa_2 + (x_3 - \bar{x})^2 \cdot fa_3 + \dots + (x_n - \bar{x})^2 \cdot fa_n}{n - 1}}$$

Para el ejemplo que tenemos de los operarios, la desviación estándar se calcula así:

Operario A $\sigma = \sqrt{623.438} = 24.969$

El número de unidades producidas por hora durante día del operario A tiene una desviación promedio de 24.964 unidades con relación a su promedio 53.75

Operario B $\sigma = \sqrt{148.438} = 12.184$

El número de unidades producidas por hora durante día del operario B tiene una desviación promedio de 12.184 unidades con relación a su promedio 53.75

La desviación estándar es útil siempre y cuando se estén comparando conjuntos de datos que tengan promedios similares y que sean las mismas unidades de investigación como en el ejemplo de los dos operarios, pero cuando los promedios y las unidades de investigación son diferentes, no basta con esta medida; por ejemplo si se está comparando el número de unidades producidas por hora de un operario y el sueldo de otro operario, no se podrían comparar las variaciones de estos.

Veamos un ejemplo:

Un operario C produce en promedio 40 unidades por hora, con una desviación estándar de 5.

Otro operario D produce en promedio 160 unidades por hora, con una desviación estándar de 15. A simple vista parece ser que el operario D tiene 3 veces más variabilidad que el operario C; pero debe tenerse en cuenta que el operario D produce unidades en promedio 4 veces más que el operario C; así que para estos casos se tiene la siguiente medida:

1. COEFICIENTE DE DISPERSIÓN O DE VARIACIÓN

También conocida como variación relativa, puesto que muestra en qué porcentaje está variando un conjunto de datos; es decir, que no está expresado en las unidades de investigación.

$$C.V = \frac{\sigma}{\bar{x}} * 100 \text{ para una población} \quad C.V = \frac{s}{\bar{x}} * 100 \text{ para una muestra}$$

En ejemplo anterior:

$$C.V_c = \frac{5}{160} * 100 = 3.125\% \quad C.V_D = \frac{15}{160} * 100 = 9.375\%$$

Así podemos ver que el operario D presenta menor variabilidad.

Para el ejemplo que tenemos de los operarios A y B:

$$C.V_A = \frac{24.969}{53.75} * 100 = 46.454\% \quad C.V_B = \frac{12.184}{53.75} * 100 = 22.668\%$$

Realicemos un ejemplo con un conjunto de datos agrupados:

A continuación se dan las edades de los empleados de una empresa:

LI	LS	fai	Xi	Xi*fai	(Xi-μ) ² * fai
18	23	7	20,5	143,5	1411,08243
23	28	20	25,5	510	1692,06408
28	33	15	30,5	457,5	264,34806
33	38	7	35,5	248,5	4,502428
38	43	4	40,5	162	134,652816
43	48	25	45,5	1137,5	2917,0801

48	53	3	50,5	151,5	749,109612
		N =81		Σ=2810.5	Σ=7172.84

$$\mu = \frac{\sum Xi * fai}{N} = \frac{2810.5}{81} = 34.698$$

Este es el promedio de las edades de los empleados

$$\sigma^2 = \frac{(Xi - \mu)^2 * fai}{N} = \frac{7172.84}{81} = 88.554$$

No se interpreta

$$\sigma = \sqrt{\sigma^2} = \sqrt{88.554} = 9.41$$

$$C.V = \frac{9.41}{34.689} * 100 = 27.127\%$$

Este es el porcentaje de variación

Las edades de los empleados de la empresa tienen una desviación promedio de 9.41 años, con relación a su promedio 34.689 años.

Ejercicios tema 4

El gerente de una empresa comercializadora tiene el record de las ventas de sus 3 vendedores durante los últimos 5 meses (en millones de pesos)

Mónica	88	68	90	101	89
Alex	77	89	90	87	78
Sandra	104	88	118	88	123

Calcule: Varianza, Desviación Estándar y coeficiente de Variación o Dispersión.

Actividad final unidad 3

Los siguientes son los ingresos semanales (en millones de pesos) de 2 centros de atención psicológica durante los últimos 2 años, de acuerdo con una muestra aleatoria:

Centro psicológico A			Centro psicológico B		
Li	Ls	fai	Li	Ls	fai
0	1	13	0	1	33
1	2	40	1	2	10
2	3	10	2	3	10
3	4	24	3	4	40
4	5	3	4	5	7

Seleccione la respuesta correcta:

- Los ingresos más frecuentes del Centro psicológico A están en el intervalo:
a) De 0 a 1 **b)** De 1 a 2 **c)** De 3 a 4
- Los ingresos menos frecuentes del Centro psicológico A están en el intervalo:
a) De 2 a 3 **b)** De 4 a 5 **c)** De 2 a 3
- Los ingresos más frecuentes del Centro psicológico B están en el intervalo:
a) De 4 a 5 **b)** De 1 a 2 **c)** De 3 a 4
- Los ingresos menos frecuentes del Centro psicológico B están dos intervalos:
a) Falso **b)** verdadero
- Los ingresos menores de los dos Centros psicológicos están en el intervalo:
a) De 0 a 1 **b)** De 1 a 2 **c)** De 3 a 4
- Los ingresos mayores de los dos Centros psicológicos están en el intervalo:
a) De 4 a 5 **b)** De 1 a 2 **c)** De 3 a 4
- Se puede afirmar que el Centro psicológico B tuvo el mayor promedio de ingresos:
a) Falso **b)** verdadero

Actividad

Realiza una investigación estadística, sobre datos cuantitativos agrupados, en tu medio; ya sea tu lugar de trabajo, tu ciudad o tu familia y realiza todo el proceso: tablas, gráficos, medidas y conclusiones o decisiones finales.

3.5. Glosario

Dato más frecuente: Es el dato que más se repite; es decir la moda. Se identifica como el que tiene la frecuencia absoluta más alta.

Frecuencias: Indica en forma numérica (absoluta) o en forma porcentual (relativa) las veces que se presenta cada dato.

Inferencia: Es la generalización que se obtiene, partiendo de una o varias muestras, sobre la población.

Dispersión: Indica cómo se dispersan o varían los datos en la distribución; existen varias medidas para analizar dicha dispersión; las más utilizadas son las que varían con relación al promedio.

Histograma: Es un gráfico de barras continuas y puede ser de frecuencias absolutas o frecuencias relativas.

Ojiva: Muestra gráficamente el comportamiento numérico o porcentual de los datos en la forma: “menor o igual que el dato”

Diagrama de barras: Es el que más se aplica en datos cuantitativos ordenados en fila.

Frecuencias acumuladas: Las frecuencias absolutas y las relativas, se acumulan por cada clase y se utilizan para hacer interpretaciones de los datos como: mayor o igual, menor, menor o igual.

Interpretación de datos: mayor, dato menor, dato más frecuente, dato menos frecuente. Consiste en el análisis de los datos con el fin de analizar el comportamiento de ellos y concluir.

3.6. Fuentes Bibliográficas

Anderson, D., Sweeney, D. & Williams, T. (1999). Estadística para Administración y Economía. (7ª edición). México: Internacional Thomson Editores.

Berenson, M. L. & LEVINE, D. M. (1996). Estadística básica en Administración. (6ª edición). México: Prentice-Hall.

Cáceres Hernández, J. (2009). Conceptos básicos de Estadística para ciencias sociales. Madrid: Delta Publicaciones.

Espejo, M. (2003). Estadística descriptiva y probabilística. Cádiz: Universidad de Cádiz.

Martínez Bencardino, C. (2004). Estadística y muestreo. (11ª edición). Bogotá: Ecoe ediciones.

Mendenhall, W. & Sincich, T. (1997). Probabilidad y Estadística para Ingeniería y administración. (4ª edición). México: Prentice-Hall.

Pérez López, J. (2007) Muestreo estadístico. Madrid: Prentice-Hall.

Ross, Sheldon, M. (2005). Introducción a la Estadística. Barcelona: Reverte.

Spiegel, M. R. (1995). Estadística. (2ª edición). Madrid: McGraw-Hill.

Berenson, M. L. & Levine, D. M. (1996). Estadística básica en Administración. (6ª edición). México: Prentice-Hall.

Anderson, D., Sweeney, D. & Williams, T. (1999). Estadística para Administración y Economía. (7ª edición). México: Internacional Thomson Editores.

Spiegel, M. R. (1995). Estadística. (2ª edición). Madrid: McGraw-Hill.

3.7. Fuentes Digitales o Electrónicas

Compas3 Comercio Electrónico. Introducción a la estadística descriptiva [Versión electrónica]. Madrid, España, 2000. Extraído el 10 de octubre de 2009 de:
<http://www.aulafacil.com/CursoEstadistica/CursoEstadistica.htm>

María Da Silva Ramis. Definición y Aplicaciones de la estadística descriptiva [Versión electrónica]. Extraído el 27 de octubre de 2009 de:

<http://www.monografias.com/trabajos10/esta/esta.shtml?monosearc>

Pita Fernández, S. Estadística descriptiva de los datos Uso de la estadística y la epidemiología en atención primaria. En: Gil VF, Merino J, Orozco D, Quirce F.

Manual de metodología de trabajo en atención primaria. Universidad de Alicante. Madrid, Jarpyo Editores, S.A. 1997; 115-161. Actualizado 06/03/2001. Extraído el 27 de octubre de 2009 de:
<http://www.fisterra.com/mbe/investiga/10descriptiva/10descriptiva.asp#introduccion>

Universidad de San Carlos. Estadística descriptiva: Conceptos básicos [Versión electrónica]. Guatemala, actualizado el 21 de agosto de 2007. Extraído el 24 de octubre de 2009 de
<http://sitios.ingenieria-usac.edu.gt/estadistica/estadistica2/estadisticadescriptiva.html>