



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

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 II 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 ORDENADOS EN FILA	7
3.1. Tablas para datos cuantitativos ordenados en fila	8
3.2. Medidas de tendencia central para datos ordenados en fila	12
3.3. Glosario	18
3.4. Fuentes Bibliográficas	19
3.5. Fuentes Digitales o Electrónicas	20

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.

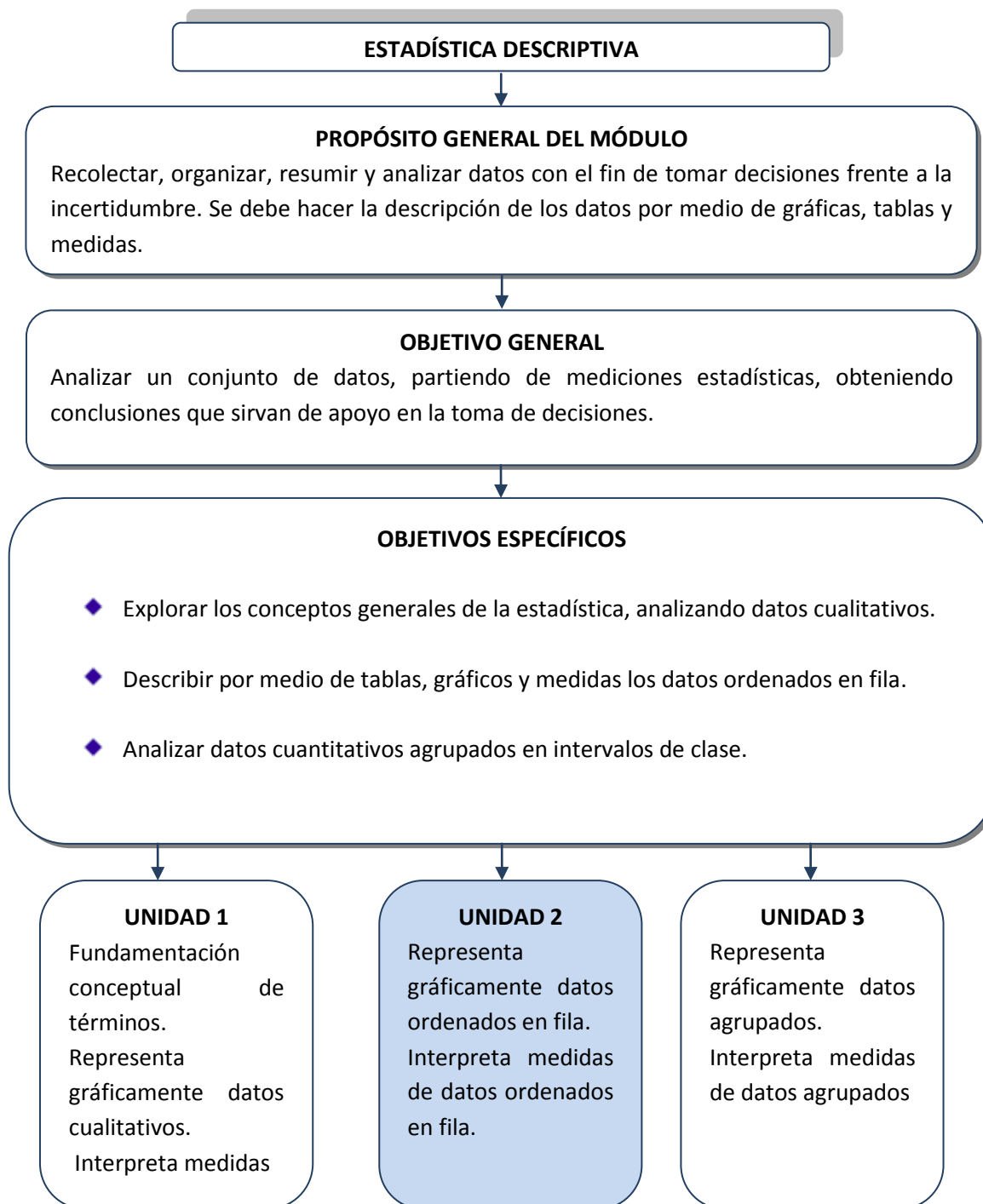
El módulo está construido con un lenguaje sencillo, con ejercicios aplicados a la cotidianidad y a situaciones prácticas inherentes al plan de estudio del estudiante de La Corporación Universitaria Remington de cualquier programa existente en la universidad, con el fin de que, de una forma pedagógica,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

OBJETIVO GENERAL

Describir por medio de tablas, gráficos y medidas los datos ordenados en fila.

OBJETIVOS ESPECÍFICOS

- ◆ Interpretar los datos: mayor, menor, más frecuente y menos frecuente en un conjunto de datos cuantitativos ordenados en fila.
- ◆ Construir las frecuencias acumuladas e interpretarlas.
- ◆ Realizar el diagrama de barras e interpretarlo.
- ◆ Calcular las medidas de tendencia central e interpretarlas.

Prueba inicial

Se seleccionó una muestra aleatoria entre los estudiantes de Corporación Universitaria Remington y clasificaron de acuerdo con su edad, los siguientes son los datos:

Xi(edad)	fai(nº est)	fri	%	faai	frai	Xi*fai
17	100					
19	140					
23	100					
26	140					
29	100					
35	25					
45	8					

Termine la tabla y responda a las siguientes preguntas, sustentando su respuesta:

1. El dato más frecuente es 100 (F) (V)
2. Lo que menos se presenta es que los estudiantes tengan 35 años (F) (V)
3. Hay 8 estudiantes que tienen la edad mayor (F) (V)

4. El 27% de los estudiantes tiene menos de 26 años (F) (V)
5. Hay 140 estudiantes que tienen como máximo 26 años (F) (V)
6. Hay 340 estudiantes que tienen como máximo 23 años (F) (V)
7. El promedio de la edad entre los estudiantes es de: a) 43,545 b) 25,758 c) 23,548
8. Las edades más frecuentes son: a) 100 y 140 b) 35 y 25 c) 17 y 19 d) 19 y 26
9. El 50% de las edades entre los estudiantes es de: a) 23 b) 19 c) 26
10. El 73,5% de los estudiantes tienen 29 años o menos (F) (V)

3.1. Tablas para datos cuantitativos ordenados en fila

DATOS CUANTITATIVOS

Cuando la investigación es de tipo cuantitativa, es decir que se tienen datos numéricos, los datos se pueden organizar de dos formas: ordenados en una fila o agrupados en intervalos de clase.

DATOS ORDENADOS EN FILA

Cuando los datos numéricos son pocos, o siendo muchos, cada uno representa una alta frecuencia. Los datos se organizan en una fila en forma ascendente.

Ejemplo:

El jefe de personal de la compañía “Aceros S.A.” preocupado por las llegadas tarde de sus empleados, seleccionó una muestra aleatoria entre los empleados que han llegado tarde al trabajo durante los últimos cuatro meses y anotó el número de llegadas tarde de cada uno de dichos empleados y los resultados fueron los siguientes:

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

FICHA TÉCNICA

Población: Empleados que han llegado tarde al trabajo durante los últimos cuatro meses de la compañía "Aceros S.A."

Muestra: 60 empleados escogidos al azar.

Descripción de la variable: Número de llegadas tarde de cada empleado.

Tipo de variable: Cuantitativa discreta.

TABLA DE DISTRIBUCIÓN DE FRECUENCIAS

Xi	fai	fri	%	faai	frai	% Acum
1	5	0,083	8,3	5	0,083	8,3
2	4	0,067	6,7	9	0,150	15,0
3	8	0,133	13,3	17	0,283	28,3
4	8	0,133	13,3	25	0,417	41,7
5	21	0,350	35,0	46	0,767	76,7
6	8	0,133	13,3	54	0,900	90,0
7	2	0,033	3,3	56	0,933	93,3
8	4	0,067	6,7	60	1,000	100,0
	n = 60	$\Sigma = 1$	$\Sigma = 100$			

Interpretación de los datos: mayor, menor, más frecuente y menos frecuente

Como los datos fueron organizados de menor a mayor, entonces siempre el primer dato será el menor y el último dato será el mayor; además recuerde que los datos siempre están en **Xi** y que la frecuencia absoluta indica el número de veces que se presenta el dato.

1. **DATO MENOR:** El menor número de llegadas tarde de los empleados es de **1** y hubo 5 empleados que representan el 8,3% que llegaron tarde al trabajo 1 vez.
2. **DATO MAYOR:** El mayor número de llegadas tarde de los empleados es de **8** y hubo 4 empleados que representan el 6,7% que llegaron tarde al trabajo 8 veces.
3. **DATO MÁS FRECUENTE:** Lo que más se presenta es que los empleados lleguen tarde al trabajo **5 veces**; 21 empleados que representan el 35% llegaron tarde 5 veces.
4. **DATO MENOS FRECUENTE:** Lo que menos se presenta es que los empleados lleguen tarde al trabajo 7 veces; 2 empleados que representan el 3,3% llegaron tarde 7 veces.

A continuación se explican otros datos en la tabla:

Frecuencia absoluta acumulada (faai o Ni): Es sumar las frecuencias absolutas por cada categoría:

$$faai = fai + faa_{(i-1)}$$

La última frecuencia absoluta acumulada es igual a N ó n.

Frecuencia relativa acumulada (frai o Hi): Es sumar las frecuencias relativas por cada categoría:

$$frai = fri + fra_{(i-1)} \quad \text{o también} \quad frai = \frac{faai}{N \text{ ó } n}$$

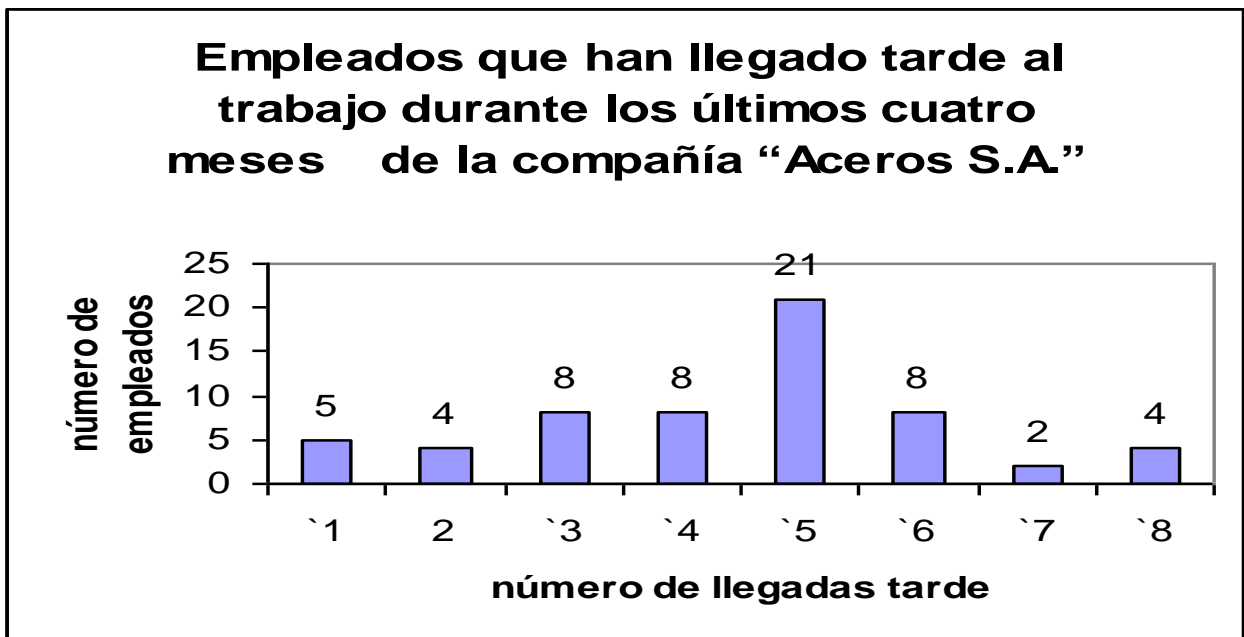
La última frecuencia relativa acumulada es igual a uno.

Porcentaje acumulado: Se obtiene multiplicando las frecuencias relativas acumuladas por cien.

GRÁFICOS PARA DATOS ORDENADOS EN FILA

Se utilizan los mismos diagramas de los datos cualitativos, aunque el diagrama circular no es muy usual porque se presta a confusiones pues queda el valor y el porcentaje en las categorías.

A continuación se tiene el diagrama de barras verticales de la anterior investigación.



Ejercicios tema 1

En cada uno de los siguientes ejercicios, realice lo siguiente:

- ◆ Ficha técnica
- ◆ Interprete dato: mayor, menor, más frecuente y menos frecuente.
- ◆ Elabore el diagrama de barras e interprételo.

1. La profesora de estadística realizó una prueba a un grupo de alumnos que constaba de 12 puntos. Eligió una muestra al azar y los resultados del número de respuestas acertadas fueron los siguientes:

12	10	10	8	4	3	0	2	1	12
12	11	9	2	6	4	8	0	7	11
5	9	9	3	7	6	9	0	8	8
4	10	9	7	8	7	9	1	7	8
6	10	4	8	11	10	7	12	9	8

2. En la feria Exponavidad del año pasado se realizó un estudio sobre lo que los visitantes tenían disponible para gastar en regalos navideños. Se seleccionó una muestra al azar con los siguientes resultados: (el dinero disponible está dado en miles)

X_i	f_{ai}
50	10
200	25
400	15
600	10
800	8
1000	8
1500	2
2000	3

3.2. Medidas de tendencia central para datos ordenados en fila

Son medidas que centralizan los datos y dan información sobre la parte de la distribución hacia donde se están agrupando los datos. Las más importantes son: Media aritmética, la moda y la mediana.

MEDIA ARITMÉTICA

“Es la medida más conocida, más fácil de calcular y con la que siempre estamos más familiarizados” (Martínez Bencardino, C. 2004, p. 74) Es el promedio de los datos y se representa:

μ para un censo es decir, como parámetro

\bar{X} para un muestreo es decir, como estadística.

Para calcular la media aritmética: se debe tener en cuenta lo siguiente:

1. **si los datos no están ordenados:** “Se obtiene sumando todos los valores y dividiendo por la cantidad de valores” (Anderson, D., Sweeney, D. y Williams, T., 1999, p.65).

$$\mu \text{ ó } \bar{X} = \frac{x_1 + x_2 + \dots + x_n}{N \text{ ó } n}$$

Ejemplo 1: a continuación se tienen las notas de seguimiento en estadística de un estudiante 4,5 3,8 4,5 4,0 4,5

$$\mu = \frac{4,5 + 3,8 + 4,5 + 4 + 4,5}{5} = 4,26$$

El promedio de las notas de estadística del alumno es de 4,26

Ejemplo 2: a continuación se dan los sueldos de los empleados del almacén “El Reloj” de acuerdo con una muestra aleatoria: 590, 590, 880, 1000, 590, 590, 1000 (en miles)

$$\bar{X} = \frac{590 + 590 + 880 + 1000 + 590 + 590 + 1000}{7} = 748,571$$

El promedio de los sueldos de los empleados del almacén el administrador es de \$748571

2. cuando se tienen ordenados en una tabla de distribución de frecuencias para hallar la media aritmética se aplica la siguiente fórmula:

$$\mu \text{ ó } \bar{x} = \frac{x_1 \cdot fa_1 + x_2 \cdot fa_2 + \dots + x_n \cdot fa_n}{N \text{ ó } n}$$

Tomando el ejemplo de la tabla de frecuencias:

Xi	fai	Xi*fai
1	5	5
2	4	8
3	8	24
4	8	32
5	21	105
6	8	48
7	2	14
8	4	32
	n = 60	Σ = 268

$$\bar{X} = \frac{268}{60} = 4,467$$

El promedio de las llegadas tarde de los empleados de la compañía “Aceros S.A.” es de 4,467 (si se analiza, llevándolo a la realidad, es que en promedio llegan 4 veces tarde al trabajo)

Propiedades de la media aritmética:

1. Todo conjunto de datos cuantitativos tiene una media aritmética.
2. Cuando se calcula la media aritmética se incluyen todos los datos.
3. La media aritmética de un conjunto de datos es única es decir, sólo existe una media aritmética.
4. Cuando se tienen valores extremos (es decir, muy altos o muy bajos) la media aritmética no puede tomarse como representativa de los datos.

LA MODA (Mo)

“Es el valor que ocurre con mayor frecuencia; es decir, el valor más frecuente” (Spiegel, M. R., 1995).

Cuando se tiene la distribución de frecuencias se busca la mayor frecuencia absoluta y su correspondiente “Xi” es la moda.

Tomando los ejemplos anteriores, la moda se interpreta así:

Ejemplo 1: La nota de seguimiento en estadística del estudiante más frecuente es 4,5

Ejemplo 2: El sueldo de los empleados del almacén “El Reloj” más frecuente es \$ 590000

Ejemplo 3: Las llegadas tarde de los empleados de la compañía “Aceros S.A.” Más frecuente es de 5

Existen distribuciones que tienen varias modas por lo tanto se denomina **bimodal** cuando tiene dos modas, si tiene más de dos, se llama **multimodal**.

Pueden existir distribuciones donde no haya moda, si todos los valores de la frecuencia absoluta son iguales, no hay moda, pues indica que todos los datos se repiten el mismo número de veces. En el gráfico de barras verticales la moda se identifica como barra más alta y su valor se ubica en el eje X

LA MEDIANA (Me)

“Es el valor de la serie de datos que se sitúa justamente en el centro de la muestra (un 50% de valores son inferiores y otro 50% son superiores).

No presentan el problema de estar influido por los valores extremos, pero en cambio no utiliza en su cálculo toda la información de la serie de datos (no pondera cada valor por el número de veces que se ha repetido).” (<http://www.aulafacil.com/CursoEstadistica/CursoEstadistica.htm>)

Es el valor que está en todo el centro de la distribución. Para ubicarla los datos deben estar ordenados. Tomando los ejemplos anteriores, la mediana se interpreta así:

Ejemplo 1:

3,8 4,0 4,5 4,5 4,5

El 50% de las notas de seguimiento en estadística del estudiante es 4,5

Ejemplo 2:

590, 590, 590, 590, 880, 1000, 1000

El 50% de los sueldos de los empleados del almacén “El Reloj” es \$590000

Cuando se tiene un número de datos par se toman los dos valores centrales y se promedian.

Ejemplo:

3,8 3,8 4,0 4,5 4,5 4,5 $Me = \frac{40 + 4,5}{2} = 4,3$

Cuando se tiene la distribución de frecuencias para ubicar la mediana se procede así:

1. Si el número de datos (N ó n) es impar se calcula $\frac{n}{2}$, se busca en la faa (frecuencia absoluta acumulada) y su correspondiente dato (Xi) es la Me.
2. Si el número de datos (N ó n) es impar se calcula $\frac{n+1}{2}$, se busca en la faa (frecuencia absoluta acumulada) y su correspondiente dato (Xi) es la Me.

xi	fai	faai
3.8	1	1
4.0	1	2
4.5	3	5
N=5		

$Me = \frac{5+1}{2} = 3$ Corresponde al tercer dato que es 4.5

3. Si el número de datos (N o n) es par se calcula $\frac{n}{2}$ y su consecutivo, se buscan en la faa (frecuencia absoluta acumulada) y el promedio de sus correspondientes datos (Xi) es la Me.

Xi	fai	faai
3,8	2	2
4,0	1	3
4,5	3	6
N = 6		

$Me = \frac{6}{2} = 3$ Y su consecutivo es 4 $Me = \frac{4,0 + 4,5}{2} = 4,3$

En el ejemplo de la compañía “Aceros S.A.” $N/2 = 60/2 = 30$ y su consecutivo 31 como ambos están incluidos en 46, la Mediana sería 5

El 50% de las llegadas tarde de los empleados de la compañía “Aceros S.A.” es de 5

Ejercicios tema 2

1. La profesora de estadística realizó una prueba a un grupo de alumnos que constaba de 12 puntos. Eligió una muestra al azar y los resultados del número de respuestas acertadas fueron los siguientes:

12 10 10 8 4 3 0 2 1 12
12 11 9 2 6 4 8 0 7 11
5 9 9 3 7 6 9 0 8 8
4 10 9 7 8 7 9 1 7 8
6 10 4 8 11 10 7 12 9 8

2. En la feria Exponavidad del año pasado se realizó un estudio sobre lo que los visitantes tenían disponible para gastar en regalos navideños. Se seleccionó una muestra al azar con los siguientes resultados: (el dinero disponible está dado en miles)

Xi	fai
50	10
200	25
400	15
600	10
800	8
1000	8
1500	2
2000	3

Calcule las medidas de tendencia central de los anteriores ejercicios.

Actividad final de la unidad 2

Se seleccionó una muestra aleatoria entre los habitantes de la tercera edad que viven en el Área Metropolitana y se clasificaron de acuerdo con su estatura, los siguientes son los datos:

Xi(estat)	fai(nºancia)	fri	%	faai	frai	Xi*fai
1,45	6					
1,50	10					
1,55	10					
1,60	18					
1,65	18					
1,70	12					
1,75	10					
1,80	7					
1,85	3					
1,90	1					

Termine la tabla y responda a las siguientes preguntas, sustentando su respuesta:

- 1 El dato más frecuente es 10 (F) (V)
- 2 Lo que menos se presenta es que los ancianos midan 1,45 (F) (V)
- 3 Esta distribución es Bimodal (F) (V)
- 4 Lo que más se presenta es que los ancianos midan 1,90 (F) (V)
- 5 Hay 26 ancianos que miden menos de 1,60 (F) (V)
- 6 El 58% de los ancianos miden como máximo 1,80 (F) (V)
- 7 El promedio de la estatura entre los ancianos es de: a) 1,55 b) 1,60 c) 1,638
- 8 La estatura más frecuente es: a) 10 b) 1,90 c) 1,60 y 1,65 d) 1,50 1,55 y 1,75
- 9 El 50% de las estaturas entre los ancianos es de: a) 1,60 b) 1,70 c) 1,65
- 10 El 28% de los ancianos miden 1,65 o menos (F) (V)

Actividad:

Realiza una investigación estadística, sobre datos cuantitativos en fila 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.

3.3. 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.

3.4. 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.

3.5. 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>