www.uniremington.edu.co

UNIREMINGTON®
CORPORACIÓN UNIVERSITARIA REMINGTON
RES. 2661 MEN JUNIO 21 DE 1996

ESTADÍSTICA DESCRIPTIVA

TRANSVERSAL

FACULTAD DE CIENCIAS BÁSICAS E INGENIERÍA

Vicerrectoria de Educación a Distancia y virtual

2016

AUTOR

Pablo Emilio Botero Tobón
pbotero@uniremington.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

Jorge Mauricio Sepúlveda Castaño
Decano de la Facultad de Ciencias Básicas e Ingeniería
jsepulveda@uniremington.edu.co

Eduardo Alfredo Castillo Builes
Vicerrector modalidad distancia y virtual
ecastillo@uniremington.edu.co

Francisco Javier Álvarez Gómez
Coordinador CUR-Virtual
falvarez@uniremington.edu.co

GRUPO DE APOYO

Personal de la Unidad CUR-Virtual
EDICIÓN Y MONTAJE

Primera versión. Febrero de 2011.
Segunda versión. Marzo de 2012
Tercera versión. noviembre de 2015

TABLA DE CONTENIDO

Pág.

1 MAPA DE LA ASIGNATURA

ESTADÍSTICA DESCRIPTIVA

PROPÓSITO GENERAL DEL MÓDULO

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.

En todas las profesiones se necesita información, para poder avanzar y cualificar los procesos, por lo tanto, necesita recoger, organizar, analizar y presentar datos con el fin de tomar decisiones que favorezcan este desarrollo profesional; esta herramienta se la proporciona la Estadística Descriptiva, de la cual se realizará un amplio detalle a través de este módulo que le presenta la Corporación Universitaria Remington.

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.

OBJETIVO GENERAL

Al finalizar el módulo el estudiante estará en capacidad de:

Aplicar técnicas de análisis estadístico en la solución de problemas, partiendo de un conjunto de datos y mediciones, para la obtención de conclusiones que permitan la proyección de la estadística hacia la solución de situaciones polémicas en las diferentes áreas.

OBJETIVOS ESPECÍFICOS

UNIDAD 1

Explicar los conceptos generales de la estadística, analizando datos cualitativos y describiendo por medio de tablas, gráficas y medidas los datos ordenados en fila.

UNIDAD 2

Analizar datos cuantitativos agrupados en intervalos de clase.

UNIDAD 3

Definir los conceptos fundamentales de los modelos de representación de procesos propuestos por la estadística y su aplicabilidad.

1.1.1 PROPÓSITO GENERAL DEL MÓDULO:

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.

En todas las profesiones se necesita información, para poder avanzar y cualificar los procesos, por lo tanto, necesita recoger, organizar, analizar y presentar datos con el fin de tomar decisiones que favorezcan este desarrollo profesional; esta herramienta se la proporciona la Estadística Descriptiva, de la cual se realizará un amplio detalle a través de este módulo que le presenta la Corporación Universitaria Remington.

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 alcanzar el propósito del módulo, éste se ha diseñado de forma innovadora: se estudiarán métodos de organización, análisis y presentación de un conjunto de datos asociados a una situación problémica por medio del modelo de representación estadístico y aprenderá a caracterizar un conjunto de datos, a partir de mediciones estadísticas, para obtener conclusiones que sirvan de apoyo en la toma de decisiones.

En la primera parte se definen los conceptos generales que se requieren en estadística, las unidades se han dividido de acuerdo con los tipos de datos, a saber: datos cualitativos, datos cuantitativos ordenados en fila y datos cuantitativos agrupados en intervalos explicando como:

- Se pueden identificar.
- Se pueden recolectar,
- Se pueden organizar,
- Se pueden describir por medio de tablas gráficos y medidas

Una vez determinados estos elementos se puedan obtener conclusiones, que permitan tomar decisiones acertadas y que beneficien los procesos en los cuales se está realizando el estudio estadístico.

El módulo está construido con un lenguaje sencillo y con ejercicios aplicados a la cotidianidad y a situaciones prácticas, coherentes a los planes de estudio de las asignaturas que lo contemplan en su desarrollo académico, con el fin de que, en una forma pedagógica, aprenda y logre los objetivos propuestos en el mismo.

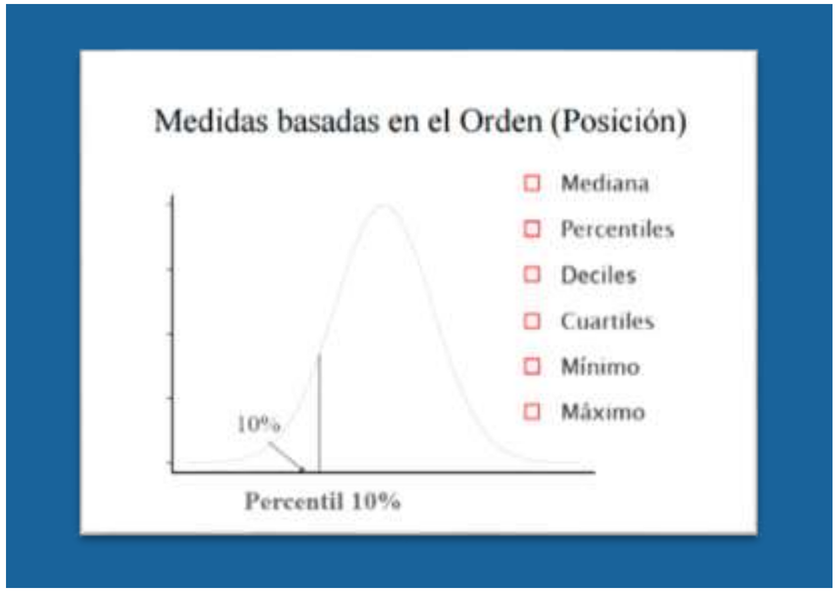
Debido a que el estudiante de educación a distancia, de La Corporación Universitaria Remington, requiere un método de aprendizaje de forma tal que más que un profesor, sea un tutor (orientador y acompañante del proceso) y que el estudiante sea autogenerador de su conocimiento, obviamente con la asesoría del tutor, 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 UNIDAD 1 CONCEPTOS GENERALES Y DATOS CUALITATIVOS

Estadística: conceptos básicos y definiciones:

www.mat.uda.cl/hsalinas/cursos/2010/eyp2/clase1.pdf



2.1.1 RELACIÓN DE CONCEPTOS

2.1.2 OBJETIVO GENERAL

Explicar los conceptos generales de la estadística, analizando datos cualitativos y describiendo por medio de tablas, gráficas y medidas los datos ordenados en fila.

2.1.3 OBJETIVOS ESPECÍFICOS

- Definir los conceptos básicos y fundamentales de la estadística, tales como población, muestra, muestreo, parámetro entre otros y determinarlos cada uno de ellos.
- Aplicar el concepto de redondeo de cifras para el manejo de magnitudes estadísticas, además del manejo adecuado de una ficha técnica utilizada para el manejo de información de una investigación estadística.
- Calcular, dada una tabla de datos estadísticos, la frecuencia, la frecuencia relativa y la frecuencia acumulada.

2.2 TEMA 1 CONCEPTOS GENERALES

A continuación se definen los conceptos fundamentales de la estadística, estos serán utilizados a través de todo el módulo, por lo tanto es de suma importancia que asimile cada una de las definiciones presentadas:

CONCEPTO	DEFINICIÓN
ESTADÍSTICA	Es una ciencia auxiliar y la compilación, organización, resumen, presentación y análisis de datos numéricos cuya función principal es elaborar principios y métodos que nos ayuden a **tomar decisiones frente a la incertidumbre.** Se puede aplicar en todas las áreas de investigación. Técnico, tecnológica, científica. Ejemplo: análisis comparativos de ingresos y egresos. Análisis de mercado para la introducción de un nuevo producto.
ESTADÍSTICA DESCRIPTIVA	Es la presentación, en forma comprensible, de **los datos recolectados**, puede ser en forma tabular, gráfico o numérico.
ESTADÍSTICA INFERENCIAL	"Se puede definir como aquellos **métodos** que hacen posible **la estimación de la población** o **la toma de una decisión referente a una población**, basándose sólo en los resultados de una muestra" (Berenson y Levine, 1996, p.3). Es la rama de la estadística que trata de los procesos y comprensión de la teoría de estimación y prueba de hipótesis.

POBLACIÓN	Es la totalidad de unidades elementales sobre los cuales se desea información. Ejemplos: a) Si el contador de la empresa desea investigar las deudas que ha adquirido la compañía en el último año, la población estaría conformada por todas las cuentas por pagar del último año. b) Si un ingeniero agrónomo desea investigar en una finca cafetera si hay Broca, la población sería todas las matas de café que hay en dicha finca. c) Si un médico veterinario desea investigar si hay gripa aviar en una granja, la población sería todas las aves que hay en la granja.
POBLACIÓN FINITA	Es una población que no es indefinidamente grande o que sólo contiene **un número finito de datos**. Ejemplo: estudiantes de la C.U.R
POBLACIÓN INFINITA	Contiene **un número infinitamente grande** de unidades elementales (datos de la población) Ejemplo: Al lanzar una moneda indefinidamente, el número de caras que se pueden obtener.
CENSO	Estudio de una población. Llamado también **enumeración completa**.
PARÁMETRO	”Es una **medida descriptiva numérica** de una población” (Mendenhall y Sincich, 1997, p.39).

	De los ejemplos de la población: a) Promedio mensual de las cuentas por pagar. ________________________________ b) El 5% de las matas de café tienen Broca. ________________________________ c) En la granja hay 850 pollos.
MUESTRA	Es el elemento **básico** o **parte representativa de una población**. Ejemplo: si deseamos investigar el nivel académico de los estudiantes de educación a distancia de la CUR, **la población** está conformada por **todos los estudiantes de educación a distancia de la CUR** y para extraer la muestra tendríamos que **seleccionar proporciones iguales** de estudiantes de **todos y cada uno** de los sitios donde funciona la educación a distancia de la CUR; si seleccionamos estudiantes de **un solo lugar**, por ejemplo de Puerto Berrio, esta parte **no sería representativa.**
MUESTREO	Es el estudio de la muestra y **la relación entre la población y la muestra** tomada de ella.
MUESTRA AL AZAR	Es aquella que se extrae **con la condición de que posea características de los otros elementos de la población** y **cada elemento de la población tiene la misma probabilidad de ser elegido**. Se denomina también **muestra aleatoria,** recibe este nombre porque para seleccionarla se hace **por sorteo o por números aleatorios.**
ESTADÍSTICA O ESTADÍGRAFO:	“Es una **medida descriptiva numérica** calculada a partir de los **datos de la muestra**” (Mendenhall y Sincich, 1997, p.39), por ejemplo, si en vez de las poblaciones que tenemos de ejemplo, **extraemos muestras**, **los parámetros** se convierten **en estadísticas.**

VARIABLE	Es cualquier **característica** que se analiza de una población, son **los datos que se estudian**; pueden ser cualitativas y cuantitativas.
VARIABLE CUALITATIVA	Es la que se refiere a **atributos o cualidades**, se divide en **categorías** y **no es numérica**. Ejemplos: estado civil (soltero, casado, viudo, unión libre) evaluar un producto (bueno, regular, malo) equipo de fútbol colombiano que más le gusta (Millonarios, Nacional, Medellín, Once Caldas, entre otros).
VARIABLE CUANTITATIVA	Se refiere a **datos numéricos**. Ejemplo: ingresos mensuales, edad, número de nacimientos. Este tipo de variable se divide en dos clases: **a. VARIABLE CUANTITATIVA DISCRETA** Sólo puede tomar **valores enteros**. Ejemplo: número de hijos por familia, número de empleados, número de estudiantes, como podemos apreciar, los ejemplos anteriores se refieren a personas, en estos casos serán **variables cuantitativas discretas**, también podía ser número de goles marcados en un partido, número de carros que pasan por un peaje, número de errores cometidos en una evaluación. **Estas variables se determinan por conteo.** _______________________________________ **b. VARIABLE CUANTITATIVA CONTINUA** Puede asumir **cualquier valor numérico**, es decir, **cualquier número real**. Se encuentra con medición, peso, longitud, tiempo, volumen, velocidad y temperatura. Ejemplos: tiempo que se gasta en la elaboración de una evaluación, las dimensiones de un producto, el peso de un niño al nacer; además todo lo que **se refiera a dinero** como: gastos mensuales, costo de un producto, sueldo de los empleados.

Ejercicios de entrenamiento del tema1 - Conceptos generales

De acuerdo a las definiciones arriba planteadas, en los siguientes cuadros indique el tipo de variable:

1. En el siguiente listado señala con una X, el tipo de variable:

VARIABLE	CUANTITATIVA	CUALITATIVA
Deporte favorito		
Número de estudiantes del curso.		
Comida preferida.		
Estatura de tus compañeros de curso.		
Coeficiente intelectual de tus compañeros de clase.		
Número de alumnos de curso.		
El color de los ojos.		

2. En el siguiente listado señala con una X, el tipo de variable:

VARIABLE	DISCRETA	CONTINUA
Número de acciones vendidas cada día en la Bolsa.		
Temperaturas registradas cada hora en un observatorio.		
Período de duración de un automóvil.		
El diámetro de las ruedas de varios coches.		
Número de hijos de 50 familias.		
Número de alumnos de curso.		
Censo anual de los colombianos.		

3. Clasificar, de acuerdo a su definición, las siguientes proposiciones:

VARIABLE	CUANTITATIVA	CUALITATIVA	DISCRETA	CONTINUA
Número de acciones vendidas cada día en la Bolsa.				
Temperaturas registradas cada hora en un observatorio.				
Período de duración de un automóvil.				
El diámetro de las ruedas de varios coches.				
Número de hijos de 50 familias.				
Número de alumnos de curso.				
Censo anual de los colombianos.				

PISTAS DE APRENDIZAJE

Traer a la memoria:

Recuerde que: Como que como se ha citado anteriormente la **Estadística** trata sobre el recuento, ordenación y clasificación de los datos obtenidos por las observaciones, para poder hacer comparaciones y sacar conclusiones.

2.3 TEMA 2 REDONDEO Y FICHA TÉCNICA

2.3.1.1 REDONDEO

Para el manejo de las magnitudes estadísticas se requiere el redondeo de ellas, por lo tanto es indispensable el manejo de este concepto.

Se entiende por redondear un número, reducir el número de cifras del mismo, consiguiendo un valor parecido, pero que se nos haga más fácil de utilizar en los procesos a desarrollar.

Por ejemplo: si se tiene el número 52, pero se están analizando los múltiplos de número 10, por facilidad aproximaríamos dicho número a 50, que es el múltiplo más próximo, pero si el número fuese 57 lo aproximaríamos a 60 que es el más cercano.

El redondeo se realiza de acuerdo a **las necesidades** que se tengan de determinados procedimientos, esto es, cual es **la precisión buscada** al desarrollar cierta actividad, cuál es el **nivel de exactitud** requerido para determinar las conclusiones buscadas al desarrollar la actividad.

2.3.1.2 MÉTODOS UTILIZADOS PARA REDONDEAR CIFRAS

En este aparte utilizaremos el método más utilizado para redondear cifras, denominado el **Método Normal,** aunque existen otros métodos diferentes que no analizaremos en este aparte, sin embargo, a continuación, se realizará una actividad con los mismos.

Actividad: Consulte y muestre algunos ejemplos de otros métodos, diferentes al normal.

De acuerdo al **Método Normal, cómo Redondear números:**

1) Se determina cuál es la última cifra que se quiere mantener.

2) Se aumenta en 1 si la cifra siguiente es 5 o más (Redondear arriba).

3) Se deja igual si la siguiente cifra es menor que 5 (redondear abajo)

Nota: Es decir, si la primera cifra que quitamos es 5 o más, entonces aumentamos la última cifra que queda en 1.

2.3.1.3 REDONDEO DE NÚMEROS DECIMALES

Se determina si el redondeo se realizará con décimas, centésimas, milésimas…, esto es, cuantas **cifras decimales** serán tenidas en cuenta, por ejemplo:

Número y condición de redondeo	Redondeo	Razón
3,1416 redondearlo en centésimas	**3,14**	La cifra siguiente (1) **es menor** que 5.
1,2637 redondearlo en décimas	**1,3**	La cifra siguiente (6) **es 5 o más.**
1,2635 redondearlo en tres cifras decimales	**1,264**	La cifra siguiente (5) **es 5 o más.**

2.3.1.4 REDONDEO DE NÚMEROS ENTEROS

Si se quiere redondear a decenas, centenas, entre otras, se tienen que reemplazar las cifras que se quitan por ceros, esto es:

Número y condición de redondeo	Redondeo	Razón
134,9 redondearlo en decenas	**130**	La cifra siguiente (4) **es menor** que 5.
12.690 redondearlo en miles	**13.000**	La cifra siguiente (6) **es 5 o más.**
1,239 redondearlo en unidades	**1**	La cifra siguiente (2) **es menor que 5.**

2.3.1.5 REDONDEO DE CIFRAS SIGNIFICATIVAS

Para redondear **"tantas"** cifras significativas, solo se tiene que contar de izquierda a derecha y realizar el redondeo allí.

> **Nota:** Si el número empieza por ceros, estos no se cuentan, estos ceros indican únicamente lo pequeño que es el número.

Número y condición de redondeo	Redondeo	Razón
1,239 redondearlo a tres cifras significativas	**1,24**	La cifra siguiente (9) **es 5 o más**.
134,9 redondearlo a una cifra significativa	**100**	La cifra siguiente (3) es menor que **5.**
0,0165 redondearlo a dos cifras significativas	**0,017**	La cifra siguiente (5) **es 5 o más.**

2.3.1.6 FICHA TÉCNICA

Se trata de una información que debe tener toda investigación estadística y es un documento en forma de sumario que contiene la descripción de las características de un objeto, material, proceso o programa de manera detallada. Los contenidos varían dependiendo del producto, servicio o entidad descrita, pero en general suele contener datos como el nombre, características físicas, el modo de uso o elaboración, propiedades distintivas y especificaciones técnicas.

La correcta redacción de la ficha técnica es importante para garantizar la satisfacción del consumidor, especialmente en los casos donde la incorrecta utilización de un producto puede resultar en daños personales o materiales o responsabilidades civiles o penales.

Cuando se hace la estadística descriptiva se debe dar una información general sobre la investigación que se hizo. Esta información contiene:

CONCEPTO ESTADÍSTICO	EXPLICACIÓN
1. Población	Se determinan **las unidades elementales** que se investigaron.
2. Muestra	Se define **la muestra** y **el tamaño** en caso de que exista. **Nota:** Cuando la muestra no existe se dice que se hizo **un censo**.
3. Descripción de la variable	Se determina cuál es **la característica de la población** que se va a investigar.
4. Tipo de variable	Se define, claramente, si la variable es **cualitativa** o **cuantitativa**, **discreta** o **continua**.

2.3.2 EJERCICIO DE APRENDIZAJE

Entre las familias de Medellín se eligieron aleatoriamente 5.000 con el fin de investigar el número de hijos por familia.	
Población	Familias de Medellín.
Muestra	5.000 familias de Medellín elegidas al azar.
Descripción de la variable	Número de hijos por familia.
Tipo de variable	Cuantitativa discreta.

Los estudiantes de sistemas de Ciudad Bolívar diseñaron un software administrativo e hicieron la demostración de él ante los empresarios de la misma ciudad, posteriormente, les dijeron que lo evaluaran entre: bueno, regular y deficiente.	
Población	Los empresarios de Ciudad Bolívar
Muestra	No hay, se hizo un censo
Descripción de la variable	Evaluar el software administrativo.
Tipo de variable	Cualitativa.

En la Corporación Universitaria Remington, sede Medellín, se realizó una encuesta entre 300 estudiantes, con el fin de conocer el medio de transporte utilizado en sus desplazamientos hacia la universidad, en metro, en carro particular, en bus de servicio público o en otro medio diferente a los mencionados, para determinar cuál de ellos era el más utilizado y proponer alternativas diferentes de movilización, para favorecer la llegada a tiempo a sus horas de clase	
Población	Estudiantes de la CUR, sede Mdellín
Muestra	300estudiantes
Descripción de la variable	Utilización del servicio de transporte.
Tipo de variable	Cuantitativa discreta.

2.3.3 EJERCICIOS DE ENTRENAMIENTO

Construir tres enunciados y realizar la ficha técnica de cada uno de ellos, como se ilustra en los ejemplos anteriores:

Ejercicio de entrenamiento 1:

Enunciado:	
Población	
Muestra	
Descripción de la variable	
Tipo de variable	

Ejercicio de entrenamiento 2:

Enunciado:	
Población	
Muestra	
Descripción de la variable	
Tipo de variable	

Ejercicio de entrenamiento 3:

Enunciado:	
Población	
Muestra	
Descripción de la variable	
Tipo de variable	

PISTAS DE APRENDIZAJE

Traer a la memoria:

Recuerde que: La correcta redacción de la ficha técnica es importante para garantizar la satisfacción del consumidor, especialmente en los casos donde la incorrecta utilización de un producto puede resultar en daños personales o materiales o responsabilidades civiles o penales.

2.4 TEMA 3 DATOS CUALITATIVOS

Cuando la variable de investigación es de tipo cualitativa (no numérica), después de realizar la ficha técnica, se tienen unos datos cualitativos que contienen tablas y gráficos especiales, por lo que se debe elaborar una tabla de distribución de frecuencias (**distribución de frecuencias:** es la agrupación de datos en categorías mutuamente excluyentes que indican el número de observaciones en cada categoría), que contiene los siguientes elementos:

ELEMENTO	DEFINICIÓN
x_i	Es la **variable de investigación** o **dato** que se investiga.
Frecuencia Absoluta f_{ai} o ni	Es el número de veces que se repite cada clase o categoría. **Al sumar todas las frecuencias absolutas**, se encuentra n o N. N = indica el **tamaño de la población**, se utiliza cuando se hace un censo. n = indica **el tamaño de la muestra**, se utiliza cuando se hace un muestreo.

Frecuencia Relativa f_{ri} o hi	Es la **relación** que existe entre la **frecuencia absoluta** de cada categoría y **el tamaño** de **la muestra** o **población** según el caso. $$f_{ri} = \frac{f_{ai}}{n \text{ o } N}$$ $$hi = \frac{ni}{n \text{ o } N}$$ Debido a que esta relación se da de **una parte al todo**, al sumar todas las frecuencias desde la primera hasta la última categoría, **el resultado es igual a uno.**
Porcentaje %	Cuando se multiplican las **frecuencias relativas** por **cien**, se encuentra el **porcentaje de cada categoría**. **La suma** de todos los porcentajes es **igual a cien**.

PISTAS DE APRENDIZAJE

Traer a la memoria:

Recuerde que: Una distribución de frecuencias: es la agrupación de datos en categorías mutuamente excluyentes que indican el número de observaciones en cada categoría.

2.4.1.1 EJERCICIO DE APRENDIZAJE:

Una empresa productora de software eligió aleatoriamente a un grupo de contadores de Medellín para que evaluaran un paquete de nómina. Los resultados que se encontraron fueron los siguientes:

B	R	R	R	R	R	B	B	B
B	R	M	M	M	R	B	B	B
B	M	R	R	R	R	B	R	R
R	R	B	M	M	M	R	R	R
R	R	B	M	B	R	R	R	R

Dónde:

B: Bueno
R: Regular
M: Malo

Nota: Esto indica que el primer contador dijo que el paquete era bueno, el segundo, regular y así sucesivamente, el último dijo que era regular. Como tenemos 45 resultados este fue el total de contadores encuestados, además como nos dicen que se escogieron aleatoriamente, es una muestra.

1. Se realiza **la ficha técnica:**

Enunciado: Una empresa productora de software eligió aleatoriamente a un grupo de contadores de Medellín para que evaluaran un paquete de nómina.	
Población	Grupo de contadores de Medellín

Muestra	45 contadores escogidos aleatoriamente
Descripción de la variable	Evaluar el paquete de nómina.
Tipo de variable	Cualitativa.

2. Se construye la tabla de Distribución de Frecuencias:

TABLA DE DISTRIBUCIÓN DE FRECUENCIAS			
x_i (*evaluación*)	f_{ai} *Contadores*	f_{ri}	$\%$
Bueno	**13**	$\frac{13}{45} = 0.2888..$	0.289×100 $*28.9\%$
Regular	**24**	$\frac{24}{45} = 0.5333..$	0.533×100 $*53.3\%$
Malo	**8**	$\frac{8}{45} = 0.1777\ldots$	0.178×100 $*17.8\%$
Total	$n = 45$	**1**	**100%**

***Recuerde el concepto de Redondeo y porcentaje:**

- $0.2888.. \cong 0.289 \times 100 = 28.9\%$
- $0.5333 \cong 0.533 \times 100 = 53.3\%$

$$0.1777 \cong 0.178 \times 100 = 17.8$$

Como se realizó **un muestreo**, al sumar las frecuencias absolutas se utiliza $\boldsymbol{n}$.

Para **un censo** se hubiera utilizado $\boldsymbol{N}$.

Interpretación de las Frecuencias

a) Frecuencia Absoluta

El grupo de contadores evaluó el paquete de nómina de la siguiente forma:	
Número de contadores	**Dijeron que**
13	El paquete es **bueno**.
24	El paquete es **regular.**
8	El paquete es **malo**.

b) Frecuencia relativa:

El grupo de contadores evaluó el paquete de nómina de la siguiente forma:	
Número de contadores	**Dijeron que**
28.9%	El paquete es **bueno**.
53.3%	El paquete es **regular.**
17.8%	El paquete es **malo**.

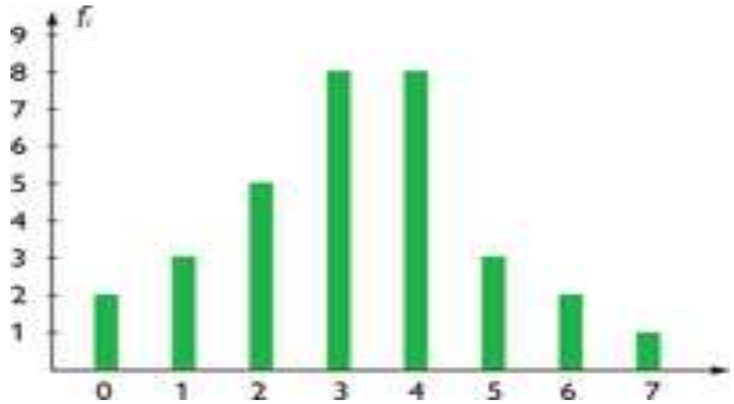
2.4.1.2 GRÁFICOS PARA DATOS CUALITATIVOS

Una vez construida la tabla de frecuencias, estas se pueden representar mediante distintos gráficos el estudio estadístico realizado. Entre los gráficos más utilizado podemos destacar:

a) Diagrama de Barras

Son dos ejes perpendiculares y una barra o un rectángulo para cada uno de los valores de la variable. Por lo general, en el eje horizontal (eje x del plano cartesiano) se colocan los valores de la variable y el otro eje (eje y del plano cartesiano), se gradúa de acuerdo al valor de las frecuencias.

En este diagrama se dibuja una barra o un rectángulo por cada uno de los valores de la variable con una altura igual a la respectiva frecuencia, de la siguiente forma:



Nota: Este diagrama puede ser:

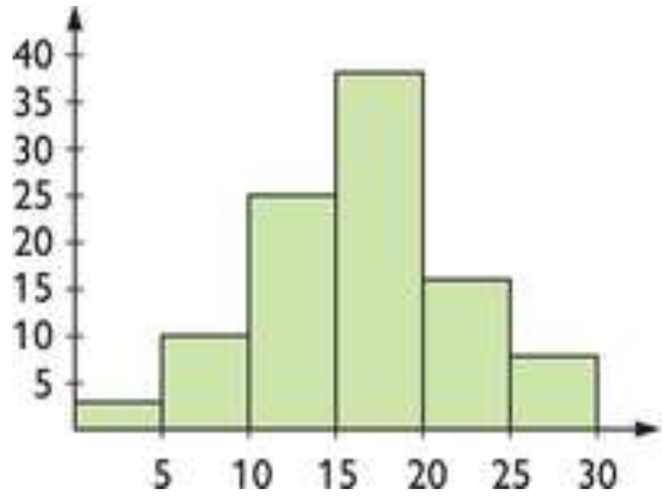
- Diagrama de Barras de horizontales.
- Diagrama de Barras verticales.

*En un ejercicio se visualizará cada uno de ellos.

b) Histograma

Es un caso particular del diagrama de barras para variables continuas; en la representación gráfica los **rectángulos aparecen pegados** y de **igual base** (ver gráfica) si **los intervalos son iguales** (tienen la **misma amplitud**); en caso de que **la amplitud no sea igual** para todos los intervalos, se debe hacer coincidir el **área del rectángulo** con **la frecuencia** de **cada uno de los intervalos** determinados en el problema.

Un histograma muy utilizado es el de la pirámide poblacional.

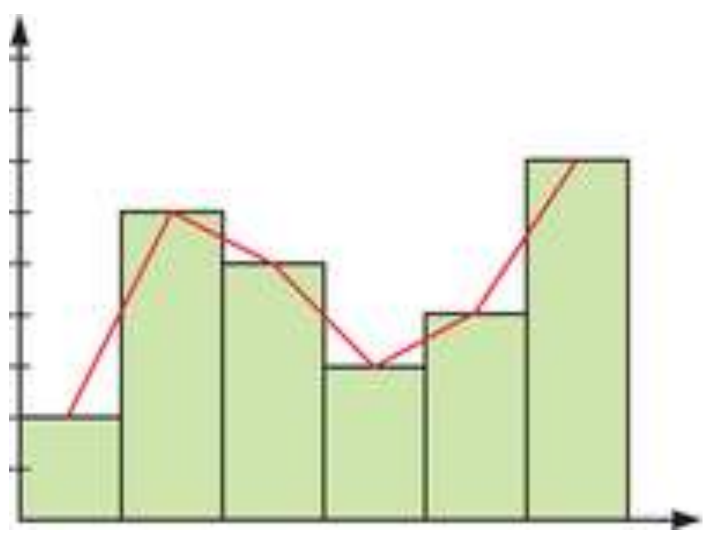


c) Polígono de frecuencias

Son dos ejes perpendiculares, en el eje horizontal se colocan los valores de la variable y en el eje vertical el valor de las frecuencias.

Se determina un par ordenado que tiene como primera coordenada el valor de la variable y como segunda coordenada el valor de la frecuencia, esto es en forma general: $\boldsymbol{P}\ (\boldsymbol{variable}, \boldsymbol{frecuencia})$,, luego se unen estos puntos y se obtiene una **Línea Poligonal** que corresponde a la representación buscada.

***



***

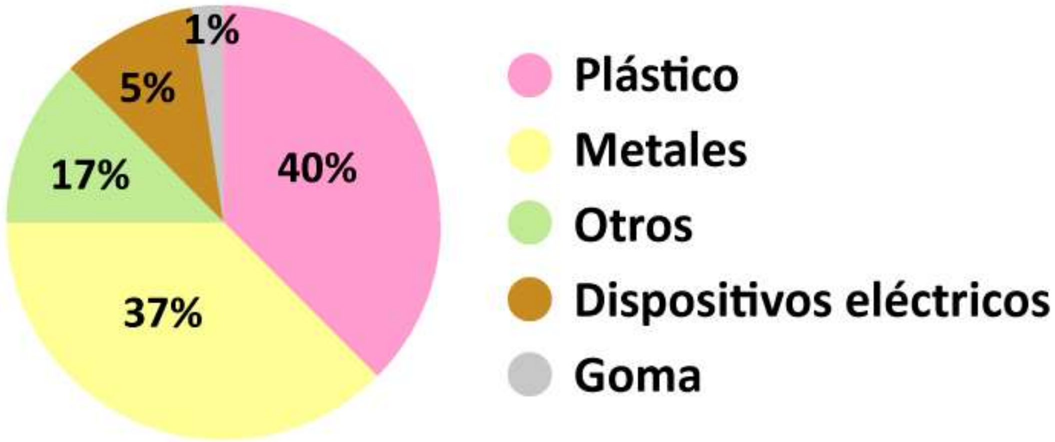
Nótese que la línea poligonal une el centro de cada uno de los lados de los rectángulos, cada uno de esos puntos corresponde a un par ordenado.

d) Diagrama de Sectores o Diagrama Circular

Se divide un círculo en tantos sectores como valores de la variable existan, esto es: si son 2 valores, se divide en dos sectores, si son 3 en tres sectores, si son 4 en cuatro sectores y así sucesivamente.

La amplitud de cada sector (porción de círculo correspondiente) debe ser proporcional a la frecuencia del valor correspondiente.

Es un diagrama que se utiliza para expresar los porcentajes, cada sector del gráfico indica la categoría y es proporcional a su porcentaje.



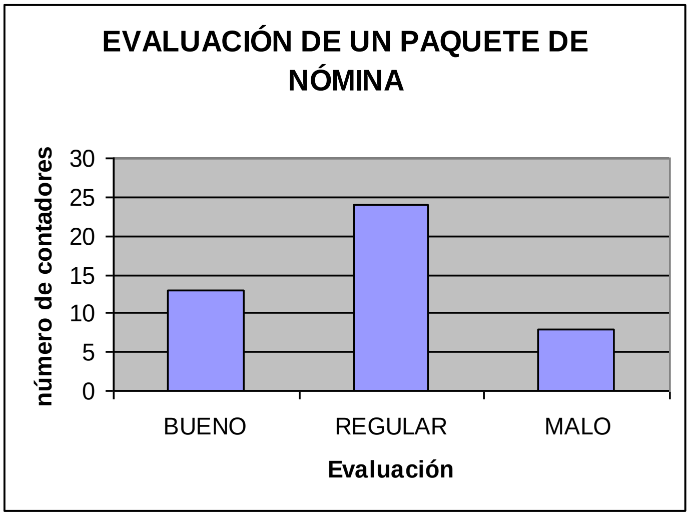
Nota: Todo gráfico estadístico debe tener un título, relacionado con lo que se está investigando.

En **el ejercicio de aprendizaje** realizado, evaluación del paquete de nómina realizado por 45 contadores, la representación gráfica, utilizando cada uno de los diagramas vistos quedaría:

e) Diagrama o gráfico de barras:

- **Diagrama o gráfico de barras verticales:** se hace en el primer cuadrante de un plano cartesiano que tendrá:
- En el eje "x" **las clases o categorías, y**
- En el eje "y" **las frecuencias absolutas.**

- Como información adicional en cada barra puede ir el porcentaje.



Interpretación del gráfico:

El grupo de contadores evaluó el paquete de nómina de la siguiente forma:

- **13** contadores lo evaluaron como bueno
- **24** contadores lo evaluaron como regular
- **8** Contadores lo evaluaron como malo

Nota: Esta interpretación es la misma que la **frecuencia absoluta** porque se refiere a **cantidades**

Diagrama o gráfico de barras horizontales:

- En el eje "x" , las **frecuencias absolutas**, y
- En el eje "y" **las clases** o **categorías.**

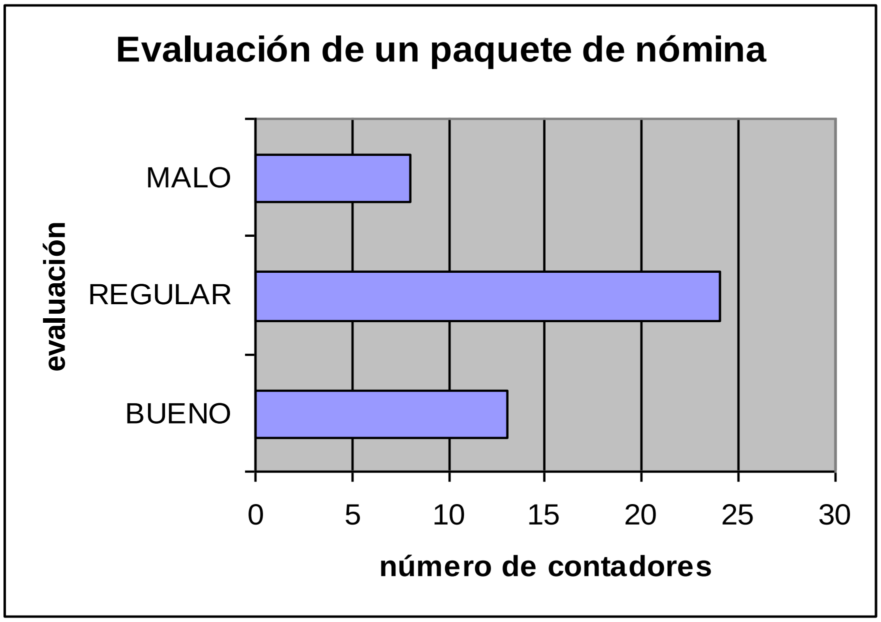
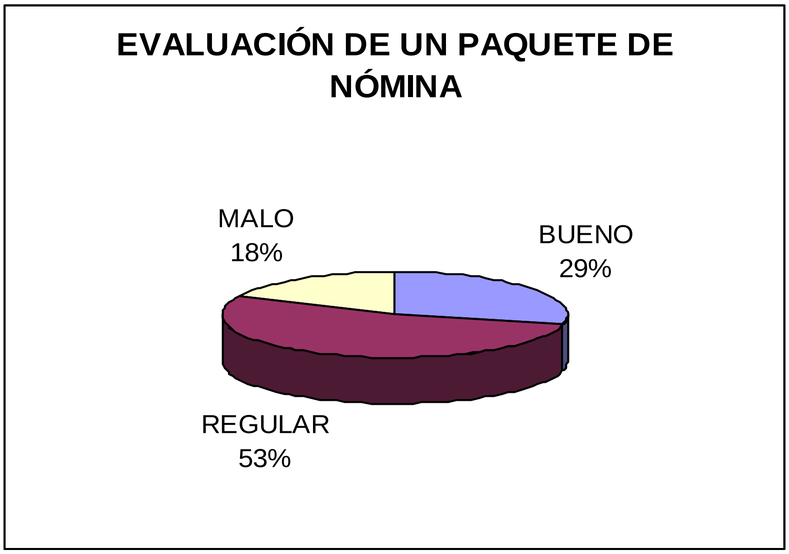


Gráfico de sectores o diagrama circular:

Es un diagrama que se utiliza para expresar **los porcentajes** (tanto por ciento), en el cual, cada uno de los **sectores del gráfico** indican **la categoría**, es proporcional a su porcentaje.



Nota: Siempre se debe seleccionar uno de los gráficos de acuerdo con lo que se quiere mostrar, cantidad o porcentaje.

Después de realizar todo el procedimiento y elaborar la gráfica correspondiente, se deben incluir, como parte del problema y su solución, **las conclusiones** derivadas de **los resultados obtenidos**.

- **CONCLUSIONES:**

De acuerdo a los resultados obtenidos, el paquete no tuvo la acogida y resultados que se esperaban entre los contadores, por lo tanto la empresa productora del software debe reformarlo basándose en las exigencias de los contadores.

2.4.1.3 EJERCICIOS DE ENTRENAMIENTO

1) **Redondeo de cifras**

En los siguientes cuadros realice la actividad indicada:

Nota: Para la solución tenga presente los ejemplos presentados en el desarrollo de la unidad.

Número y condición de redondeo	Redondeo	Justificación
53,38 redondearlo en decenas		
17.750 redondearlo en miles		
8,753 redondearlo en unidades		

2) Redondeo de cifras significativas:

Número y condición de redondeo	Redondeo	Justificación
9,538 redondearlo a tres cifras significativas		
244,53 redondearlo a dos cifras significativas		
0,002389 redondearlo a 3 cifras significativas		

1) DATOS CUALITATIVOS

Para cada uno de los ejercicios planteados a continuación:

- ✓ Elabore la ficha técnica.
- ✓ Elabore la tabla de distribución de frecuencias.
- ✓ Interprete la frecuencia absoluta y la relativa.
- ✓ Elabore el diagrama de barras y el circular.
- ✓ Exprese 2 conclusiones.

a) La compañía "los Dulces" desea lanzar al mercado una nueva chocolatina. Por tal razón realizó una encuesta entre los niños de las escuelas de Medellín para evaluar el producto y seleccionó una muestra al azar con los siguientes resultados (en miles):

x_i	f_{ai}
Bueno	**321**
Regular	**105**
Malo	**195**

b) Entre las amas de casa de Medellín se elaboró una encuesta con el fin de investigar la efectividad de un detergente. Para tal fin se seleccionó una muestra al azar con los siguientes resultados (en miles):

x_i	f_{ai}
Excelente	**225**
Bueno	**104**
Regular	**291**
Malo	**120**

---

c) Se realizó una encuesta entre los visitantes al Éxito de San Antonio en diciembre de 2002 para seleccionar entre 4 artículos: A1, A2, A3, A4 el de mejor calidad. se seleccionó una muestra al azar con los siguientes resultados (en miles):

x_i	f_{ai}
A_1	**25**
A_2	**100**
A_3	**125**
A_4	**7**

---

d) El secretario de gobierno de Medellín ordenó un informe sobre las causas de muertes violentas ocurridas en Medellín durante el último trimestre, para tal fin seleccionó una muestra aleatoria y los resultados fueron los siguientes:

x_i	f_{ai}
Por accidente de tránsito	**65**

Por terrorismo	**90**
Por arma de fuego	**76**
Por otras causas	**29**

---

e) Un investigador judicial realizó un informe sobre el número de reclusos que hay en las cárceles del Área Metropolitana discriminados por sexo, los resultados fueron los siguientes:

x_i	f_{ai}
Masculino	**11.538**
Femenino	**8.983**

---

f) Para cada uno de los siguientes ejercicios, tome los datos de la siguiente tabla:

2	2	2	4	4	4	4	1	1	1	1	1	3	3	1
1	1	3	4	1	1	1	1	3	3	3	4	4	4	4
1	1	1	1	1	2	2	2	4	4	4	1	1	1	1
2	2	3	3	1	1	4	4	1	1	2	3	4	4	1

1) Un grupo de estudiantes de turismo de Corporación Universitaria Remington desean organizar un paseo entre los bachilleres de un colegio de Medellín. Para esto elaboraron un cuestionario donde el bachiller seleccionaba el sitio así: 1.Santa Fe de Antioquia, 2. La Pintada, 3. San Jerónimo, 4.El Peñol. Después de seleccionar una muestra al azar se tiene la siguiente información (ver datos de la tabla)

____

____

____

2) Un grupo de estudiantes de sistemas de Corporación Universitaria Remington elaboró 4 paquetes (software) sobre nómina. Después de hacer la demostración ante un grupo de gerentes, se les entregó un cuestionario para que eligieran el software más conveniente, así: 1.software1, 2.software2, 3.software3, 4.software4 (Ver datos de la tabla)

____
____
____

3) Los estudiantes de Mercadeo de Corporación Universitaria Remington desean montar una microempresa de dulces, para saber con cual producto van a iniciar, encuestaron a una muestra de los niños de un colegio de Medellín para que seleccionaran así: 1.bombón, 2.chocolatina, 3.bocadillo, 4.confites. (Ver datos de la tabla)

____
____
____

4) A los estudiantes del último semestre de Contaduría Pública de Corporación Universitaria Remington se les preguntó en que les gustaría desempeñarse cuando fueran profesionales: 1. contador general, 2. docente, 3.auditor, 4.revisor fiscal. (Ver datos de la tabla)

____
____
____

5) En la institución “Despertar” el psicólogo clasificó a los niños sobre el grado de retardo mental según el DSMIV: 1. leve, 2. moderado, 3. grave 4. Profundo. (Ver datos de la tabla)

____
____
____

---

g) En un cuestionario que el jefe de personal de la empresa “Bienestar” le hizo a sus empleados, una de las preguntas está diseñada así: su estado civil es: 1.soltero, 2.casado, 3.separado, 4.unión libre, 5.viudo. Al recolectar los datos se tienen los siguientes resultados:

1	1	2	1	1	3	1	1	2	1	2	4
2	3	4	4	1	4	5	3	4	3	4	1
1	2	3	2	2	2	2	1	2	4	1	5

---

h) El sociólogo de la empresa "Unidas" realizó un estudio entre los empleados, escogiendo algunos aleatoriamente. Una de las preguntas era sobre el tipo de vivienda, la información recolectada fue la siguiente: 1.propia sin deuda, 2. propia con deuda, 3.arrendada, 4.prestada:

3	3	3	3	1	1	2	3	2	2	3	4	3
3	3	4	4	3	3	2	2	2	2	1	4	2
1	3	3	3	2	2	2	1	4	4	2	3	2
2	2	3	3	3	3	2	2	2	1	1	2	1
4	4	3	3	3	1	3	1	3	3	3	2	1

---

i) Un grupo de contadores públicos y de ingenieros de sistemas de Corporación Universitaria Remington elaboró un software contable; se eligió una muestra aleatoria entre los contadores públicos de Medellín a los cuales le entregaron el paquete y lo evaluaron así:

x_i	f_{ai}
Excelente	400
Bueno	900
Regular	225
Deficiente	15

---

j) A los extranjeros que visitaron la embajada de Colombia en Estados Unidos durante el último trimestre se les preguntó sobre su visita a Colombia y respondieron lo siguiente:

x_i	f_{ai}
Vendrán	257
Nunca vendrán	10
Ya vinieron y volverán	198
Ya vinieron y nunca volverán	5

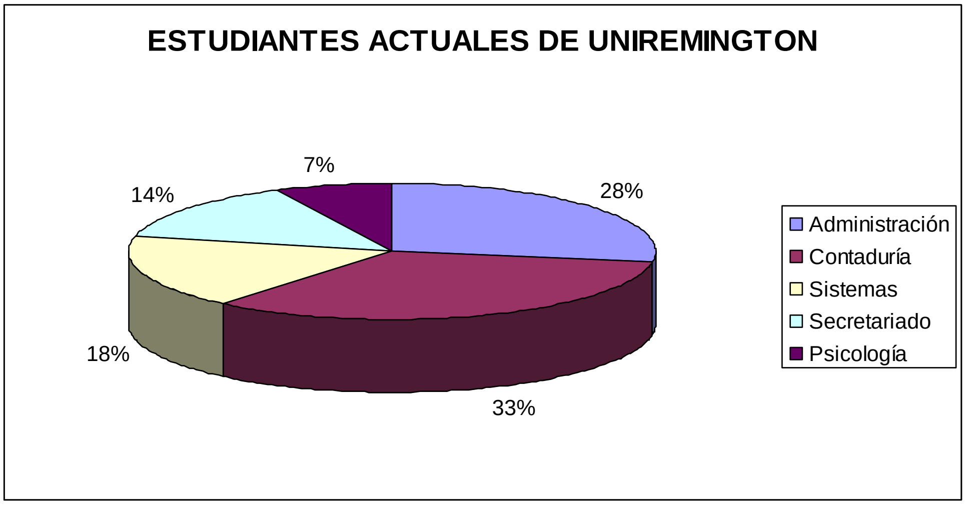
---

k) Se hizo una selección aleatoria entre los estudiantes de Corporación Universitaria Remington y se les preguntó ¿En qué dedican su tiempo libre? y respondieron lo siguiente:

x_i	f_{ai}
Ir acine	159
Hacer deporte	192
Ir a discoteca	98
Leer	105

---

l) Dado el siguiente diagrama de sectores, responda: **Verdadero (V)** o **Falso (F),** según sea el caso, justificando la respuesta



ENUNCIADO	(V)	(F)	JUSTIFICACIÓN
Esta investigación se hizo con datos cuantitativos			
La mayoría de los estudiantes son los de Contaduría			
Se debe promocionar más la Psicología.			
Lo que menos se presenta es que haya estudiantes de Secretariado.			
Este gráfico es el único que se puede elaborar para esta investigación.			
La población de esta investigación son los estudiantes de Corporación Universitaria Remington.			

m) Como ejercicio final realiza una investigación estadística, sobre datos cualitativos en tu medio; ya sea tu lugar de trabajo, tu ciudad o tu familia y realiza todo el proceso: tablas, gráficos y conclusiones o decisiones finales. **(Recuerde los pasos a seguir para desarrollar esta actividad).**

3 UNIDAD 2 DATOS CUANTITATIVOS ORDENADOS EN FILA

Explorando los datos. ¿Cómo resumir la información de tipo cuantitativo? Variables cuantitativas [*Enlace*]

Datos de tipo cuantitativo – Hojamat

www.hojamat.es/estadistica/tema2/documentos/cuantita.pdf

3.1.1 RELACIÓN DE CONCEPTOS

3.1.2 OBJETIVO GENERAL

Describir por medio de tablas, gráficos y medidas los datos ordenados en fila, calculando e interpretando además, las medidas de tendencia central para este tipo de datos.

3.1.3 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.

3.2 TEMA 1 TABLAS PARA DATOS CUANTITATIVOS ORDENADOS EN FILA

- **DATOS CUANTITATIVOS**

Cuando la investigación es de **tipo cuantitativo**, es decir, se tienen **datos numéricos**, estos se pueden organizar de dos formas:

- **Ordenados** en una fila, o
- **Agrupados** en intervalos de clase.

Los datos numéricos, sean pocos o muchos, siempre se organizan en **una Fila en forma ascendente.**

3.2.1 EJERCICIO DE APRENDIZAJE:

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

1) Se elabora la **Ficha Técnica:**

Enunciado: 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:	
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.

2) Se ordenan los datos de **menor a mayor** y se construye la **tabla de distribución de frecuencias**:

x_i	f_{ai}	f_{aai}	f_{ri}	$\%$	f_{rai}	$\%\ acumulado$
1	5	5	$\frac{5}{60}=0{,}083$	$\mathbf{8,3}$	0,083	$\mathbf{8,3}$
2	4	5+4=9	$\frac{4}{60}=0{,}067$	$\mathbf{6,7}$	0,150	$\mathbf{15}$
3	8	9+8=17	$\frac{8}{60}=0{,}133$	$\mathbf{13,3}$	0,283	$\mathbf{28,3}$
4	8	17+8=25	$\frac{8}{60}=0{,}133$	$\mathbf{13,3}$	0,417	$\mathbf{41,7}$
5	21	25+21=46	$\frac{21}{60}=0{,}350$	$\mathbf{35}$	0,767	$\mathbf{76,7}$

6	8	46+8=54	$\frac{8}{60} = 0{,}133$	$\mathbf{13,3}$	0,900	$\mathbf{90,0}$
7	2	54+2=56	$\frac{2}{60} = 0{,}033$	$\mathbf{3,3}$	0,933	$\mathbf{93,3}$
8	4	56+4=60	$\frac{4}{60} = 0{,}067$	$\mathbf{6,7}$	1,000	$\mathbf{100,0}$
	$\boldsymbol{n = 60}$		$\sum = 1$	$\sum = 100$		

3) **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 $\boldsymbol{x_i}$ y que la frecuencia absoluta indica el número de veces que se presenta el dato.

DATO	CARACTERÍSTICA
a. 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**.
b. 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.**
c. 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.**
d. 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.**

Cálculo de datos en la tabla:

- **Frecuencia absoluta acumulada:**

f_{aai} **o** N_i: Se suman las **Frecuencias Absolutas** por cada categoría, esto

es: $f_{aai} = f_{ai} + f_{aa(i-1)}$

Nota: La última **frecuencia absoluta acumulada** es igual a N **o** n

---

- **Frecuencia relativa acumulada:**

f_{rai} **o** H_i: Se suman las **Frecuencias Relativas** por cada categoría, esto

es: $f_{rai} = f_{ri} + f_{ra(i-1)}$, o también, $f_{rai} = \frac{f_{aai}}{N \text{ o } n}$

Nota: La última frecuencia relativa acumulada es igual a uno.

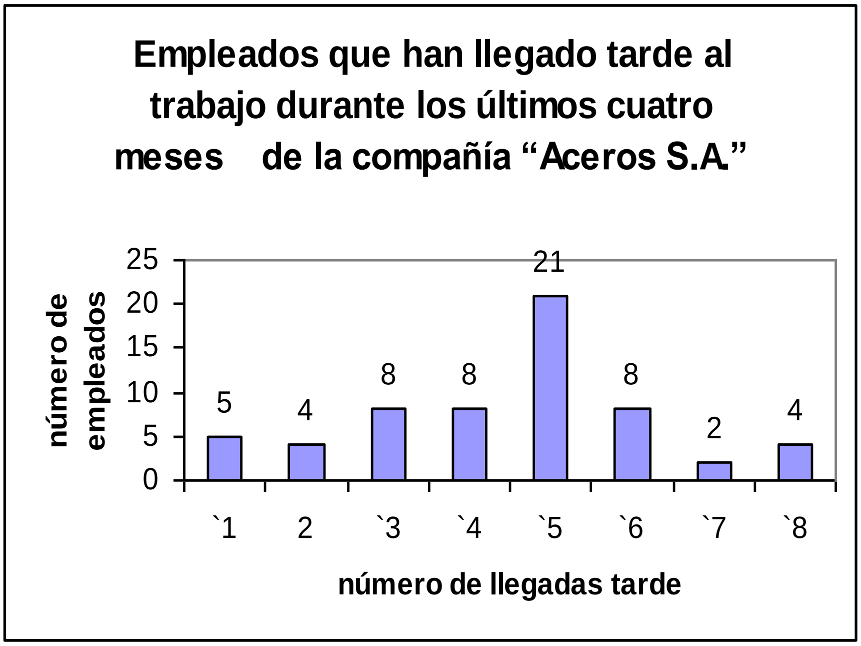
- **Porcentaje acumulado** (% ***acumulado***): Se obtiene multiplicando las **frecuencias relativas acumuladas** por **cien** (**100**).

4) GRÁFICOS PARA DATOS ORDENADOS EN FILA

Se utilizan los mismos diagramas de los datos cualitativos.

Nota: El **diagrama circular** no es muy recomendable utilizarlo 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.



PISTAS DE APRENDIZAJE

Traer a la memoria:

Recuerde que: Los datos numéricos, sean pocos o muchos, siempre se organizan en **una Fila en forma ascendente.**

3.2.2 EJERCICIOS DE ENTRENAMIENTO

Para cada uno de los siguientes ejercicios, realice:

- ✓ La 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

***Recuerde organizar los datos de menor a mayor.

2) En la feria Expo - navidad 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
2500	3

3.3 TEMA 2 MEDIDAS DE TENDENCIA CENTRAL PARA DATOS ORDENADOS EN FILA

En muchas ocasiones nos encontramos con algunos números aislados que no nos proporcionan ninguna información; información que nos permita tomar decisiones respecto a determinado hecho y solo tiene un valor individual, por ejemplo un estudiante del curso de estadística obtuvo en una evaluación un puntaje de $\frac{3.5}{5.0}$, valor que pude ser muy diciente para el estudiante, pero que carece de valor para sacar conclusiones del grupo en un momento determinado.

Para que esta tenga significado se deben tener elementos referentes, generalmente asociados a criterios estadísticos.

Esta referencia está dada por unos elementos estadísticos denominados **Medidas de Tendencia Central,** las cuales nos permiten una interpretación de los datos obtenidos en cualquier medición que se realice.

El propósito de las medidas de tendencia central es:

- Mostrar en qué lugar se ubica el promedio o la nota típica del grupo.
- Sirve como un método para comparar cualquier puntaje, con el puntaje central o típico.
- Sirve como un método para comparar el puntaje obtenido por una misma persona, obtenido en dos o más ocasiones.
- Sirve como método para comparar los resultados medios obtenidos por dos o más grupos.

Estas medidas **centralizan los datos** y dan **información** sobre la parte de la distribución hacia donde se están agrupando los datos

> **Nota:** Las medidas de tendencia central más importantes son: **Media aritmética, La moda, y La mediana.**

- **Definición y cálculo de las Medidas de Tendencia Central**

  - **MEDIA ARITMÉTICA (Promedio Aritmético)**

"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:

SÍMBOLO	UTILIZACIÓN	USABILIDAD
μ	Para un **Censo**	Como **Parámetro**
$\overline{X}$	Para un **Muestreo**	Como **Estadística**

- En el cálculo de **la Media Aritmética** se presentan dos casos:

1) Si los datos no están ordenados:

> **La media aritmética** se obtiene **sumando todos los valores y dividiendo esa suma por el número de valores** de la medición.

$$\mu \ \ o \ \overline{X} = \frac{x_1 + x_2 + \cdots + x_n}{N \ \ o \ \ n}$$

Ejemplo 1: Un estudiante obtuvo las siguientes notas en matemáticas **4.3**, **2.5**, **3.3**, **4.2**, **3.1**, **2.9** y quiere conocer el promedio (Media Aritmética) de las mismas:

El promedio de las notas de matemáticas del alumno es de **3.38.**

Ejemplo 2: A continuación se dan los valores de las boletas para las diferentes tribunas del estadio Atanasio Girardot de Medellín:

- Tribuna alta: $ 100.000

- Occidental numerada: $ 95.000
- Oriental: $ 42.000
- Tribuna sur: $ 23.000
- Tribuna norte: $27.000

Determinar cuál es el valor promedio de la boleta, para el ingreso al mismo.

Solución: Para conocer el promedio del valor de la boleta se aplica el concepto de la media aritmética, esto es:

$$\boldsymbol{\mu \;\; o\, \overline{X} = \frac{x_1+x_2+\cdots+x_n}{N \;\; o \;\; n}}$$, reemplazando valores se tiene:

$$\boldsymbol{\mu \;\; o\, \overline{X} = \frac{\$100.000 + \$95.000 + \$42.000 + +\$23.000 + \$27.000}{5}}$$

$$\boldsymbol{\mu \;\; o\, \overline{X} = \frac{\$287.000}{5} = \$57.400}$$

El valor promedio de la boleta es de $\boldsymbol{\$57.400}$

---

2) Cuando los datos se tienen **ordenados** en **una tabla de distribución de frecuencias**, se aplica la siguiente fórmula para hallar la media aritmética:

$$\boldsymbol{\mu \;\; o\, \overline{X} = \frac{x_1 f_{a_1} + x_2 f_{a_2} + \cdots + x_n f_{a_n}}{N \;\; o \;\; n}}$$

O también se puede expresar como:

$$\mu \ o \ \bar{X} = \frac{\sum_{1}^{n} x_i f_{a_i}}{N \ o \ n}$$

3.3.1 EJERCICIO DE APRENDIZAJE

Tomando una tabla de frecuencias, determinada para una solución cualquiera, se tiene:

***

x_i	f_{ai}	$x_i * f_{ai}$
1	6	6
2	5	10
3	7	21
4	9	36
5	13	65
6	12	72
7	8	56
8	4	32
9	8	72
10	10	100
	$n = 82$	$\sum 470$

$$\bar{X} = \frac{470}{82} = 5,731$$

El promedio (la media aritmética) para el problema anterior es $5,731$

Características 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).

> **Nota:** Cuando se tiene **la distribución de frecuencias** se busca **la mayor frecuencia absoluta** y su correspondiente x_i es **la moda**.

En la distribución de frecuencias indicada *** **la Moda** está determinada por el número **5**, que se repite **13** veces.

Tomando los ejemplos anteriores, la moda se interpreta así:

- **Características de la Moda**

1	Existen distribuciones que tienen varias modas, esto es: **Bimodal** cuando tiene **dos modas**, si tiene **más de dos**, se llama **Multimodal**.
2	Pueden existir distribuciones donde **no haya moda**, se da cuando los **valores de la frecuencia absoluta son iguales**, pues indica que todos los datos se **repiten el mismo número de veces.**
3	En el gráfico de **barras verticales** la moda se identifica como **la barra más alta** y su valor se **ubica en el eje X**.

- **La Mediana** (M_e)

Es aquel valor de la serie de datos que se ubica exactamente en la mitad (en el centro) de los mismos.

También se puede decir que es el valor que ocupa el **lugar central** de todos los datos, cuando éstos están **ordenados de menor a mayor**.

- **Características de la Mediana**

1	La **mediana** se representa por M_e
2	No está influido por los **valores extremos**.
3	**No utiliza** en su cálculo **toda la información** de la serie de datos.
4	No toma cada valor por **el número de veces** que se ha repetido.
5	Es el valor que está en todo **el centro** de la distribución.
6	Los datos deben estar **ordenados** para poder ubicarla.
7	La **mediana** se puede **hallar** sólo para **variable cuantitativas**.

- **Cálculo de la Mediana**

1. Para calcular la Mediana se tienen que ordenar los datos de menor a mayor.

2. Si el ordenamiento tiene un **número impar de datos**, la mediana está dada por el dato central de la misma, esto es:

Ejemplo: Se tiene el siguiente grupo de datos ordenados de menor a mayor:

1
3
5
6
7
8
9
11
12
13
14

Como el número de datos es impar (11), la mediana sería entonces el dato del centro, esto es: $M_e = 8$

Quedan, exactamente, **5 valores por encima de 8,** (1,3,5,6,7) y **5 valores por debajo** (9,11,12,13,14)

Si el ordenamiento tiene un **número par de datos**, la mediana está dada por el la **media aritmética** entre los dos datos **centrales**, esto es:

Ejemplo: Se tiene el siguiente grupo de datos ordenados de menor a mayor, calcular la Mediana:

1
3
5
6
7
9
11
12
13
14
17
18

20
21

Quedan, exactamente, **6 valores por encima de 11 y 12,** (1,3,5,6,7,9) y **6 valores por debajo** (13,14,17,18,20,21) , por lo tanto la Mediana sería:

$$M_e = \frac{11+12}{2} = \frac{23}{2} = 11.5$$

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

Si el número de datos $(n \ o \ N)$ **es par**, se calcula $\frac{n+1}{2}$ y se busca la Frecuencia Absoluta Acumulada (f_{aa}), el correspondiente x_i, es **la Mediana (Me)**.

3.3.2 EJEMPLOS

1. **Dada la siguiente tabla ordenada de datos, calcular la Mediana:**

x_i	f_{ai}	f_{aai}
3.8	1	1
4.0	1	2
4.5	3	5
	$N=5$	

$$M_e = \frac{5+1}{2} = 3$$

Corresponde al tercer dato que es 4.5

Si el número de datos (n o N) **es par**, se calcula $\frac{n}{2}$ y su **consecutivo,** se buscan en la Frecuencia Absoluta Acumulada (f_{aa}) y **el promedio** (Media Aritmética) de los correspondientes datos (x_i), es **la Mediana (Me)**.

x_i	f_{ai}	f_{aai}
3.8	2	2
4.0	1	3
4.5	3	6
	$N = 6$	

$M_e = \frac{6}{2} = 3$ que corresponde a **4.0** y su consecutivo que es el 4 y corresponde a 4.5, por lo tanto la Mediana sería:

$$M_e = \frac{4.0 + 4.5}{2} = 4.3$$

Recuerde que son seis datos y ordenados de menor a mayor y con su respectiva frecuencia se tiene:

1	2	3	4	5	6
3.8	3.8	4.0	4.5	4.5	4.5

3.3.3 EJERCICIOS DE ENTRENAMIENTO

Para los siguientes ejercicios calcule las medidas de tendencia central: la Media Aritmética, la Moda y la Mediana

Nota:

- ✓ Recuerde que para solucionar estos ejercicios, se deben ordenar los datos de menor a mayor.
- ✓ Revise los ejercicios de aprendizaje resueltos en el desarrollo del tema.

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**	**1**	**12**	**2**
12	**11**	**9**	**2**	**6**	**4**	**8**	**0**	**7**	**11**
5	**9**	**9**	**3**	**7**	**6**	**90**	**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 Expo - navidad 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	3
200	4
400	10
600	9
800	8
1000	8
1500	7
1600	4
2000	11

3. 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:

x_i *estatura*	f_{ai} (frecuencia)	f_{ri}	%	f_{aai}	f_{rai}	$x_i * f_{ai}$
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**					

4. Termine la tabla y responda verdadero (V) o falso (f), de acuerdo a la distribución anterior, las siguientes preguntas, justificando debidamente su respuesta:

Pregunta	V	F	Justificación
1. El dato más frecuente es **10.**			
2. Lo que **menos** se presenta es que los ancianos midan **1,45.**			
3. Esta distribución es Bimodal			
4. Lo que **más** se presenta es que los ancianos midan **1,90**.			
5. Hay **26** ancianos que miden menos de **1,60**.			
6. El **58%** de los ancianos miden como máximo **1,80.**			
7. El **28%** de los ancianos miden **1,65** o menos.			

En las siguientes preguntas señale la respuesta correcta:

1. El promedio de la estatura entre los ancianos es de:

a) 1,55

b) 1,60

c) 1,638

2. La estatura más frecuente es:

a) 10

b) 1,90

c) 1,60 y 1,65

d) 1,50 1,55 y 1,75

3. El 50% de las estaturas entre los ancianos es de:

a) 1,60

b) 1,70

c) 1,65

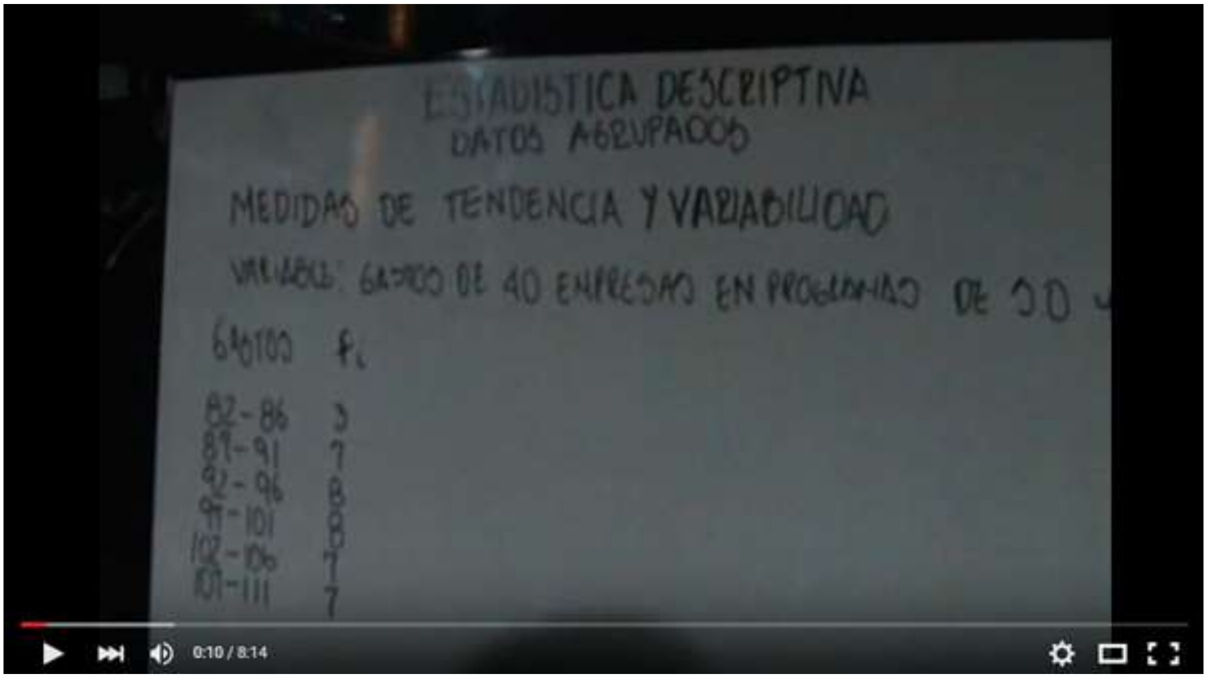
PISTAS DE APRENDIZAJE

Traer a la memoria:

Recuerde que: Las Medidas de Tendencia Central, nos permiten una interpretación de los datos obtenidos en cualquier medición que se realice.

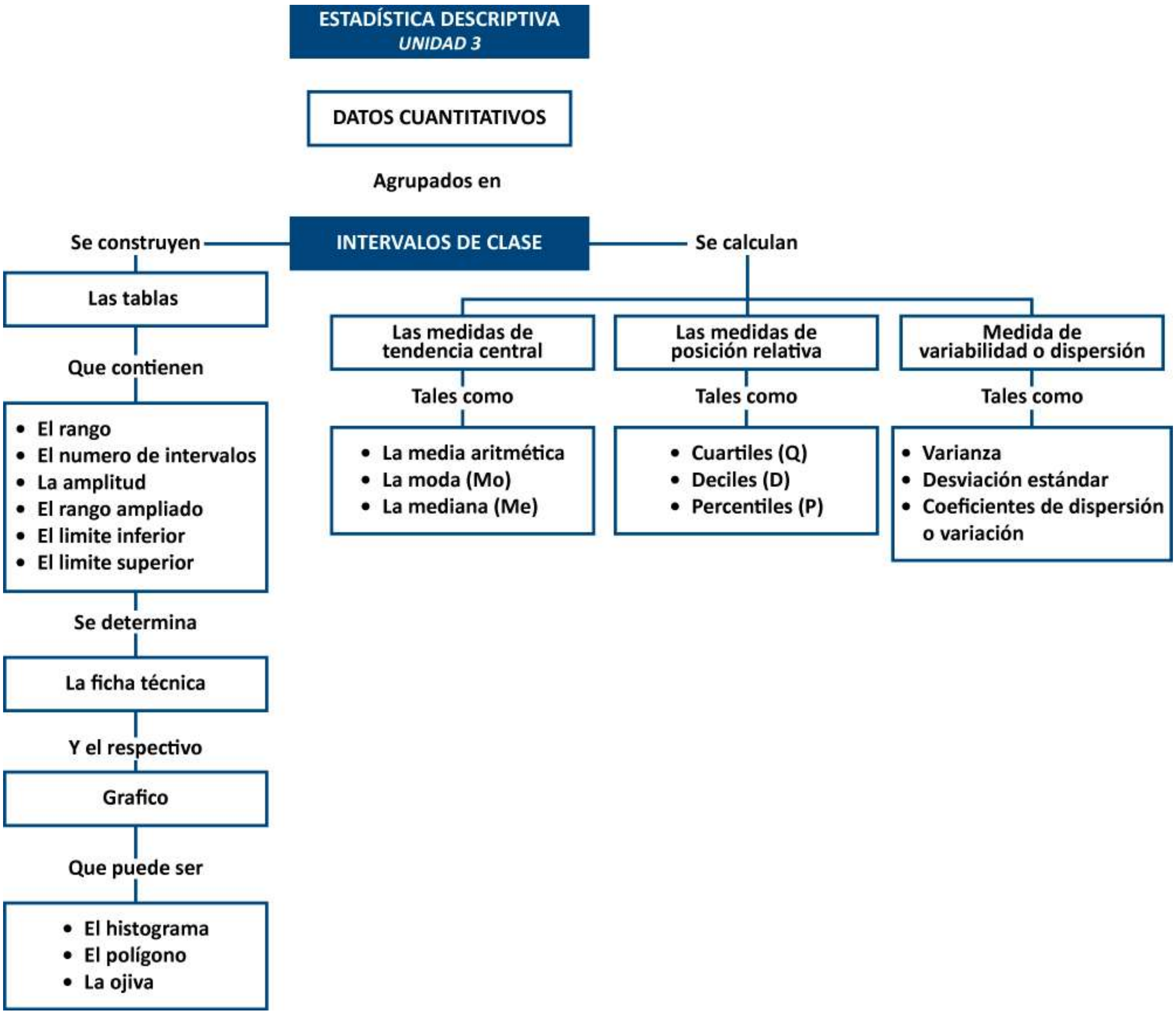
Para encontrar la Media, la Moda y la Mediana, se deben ordenar los datos de menor a mayor.

4 UNIDAD 3 DATOS CUANTITATIVOS AGRUPADOS EN INTERVALOS DE CLASE



Estadística descriptiva - Datos agrupados. Parte 3.1 [Enlace]

4.1.1 RELACIÓN DE CONCEPTOS



4.1.2 OBJETIVO GENERAL

Analizar datos cuantitativos agrupados en intervalos de clase.

4.1.3 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.

4.2 TEMA 1 TABLAS PARA DATOS CUANTITATIVOS AGRUPADOS EN INTERVALOS DE CLASE

Cuando, en un estudio estadístico, se obtienen **muchos datos** en la medición y **la frecuencia** de cada uno de estos datos es **muy baja**, es necesario **agruparlos** por **clase** o por **categoría** y cada una de éstas distribuciones se dispondrá en un **intervalo** determinado.

Para agrupar los datos en intervalos existen varios métodos, entre ellos tenemos, el siguiente:

Cuando los intervalos son de la forma: $\boldsymbol{L_i < x \leq L_s}$

Dónde:

- $\boldsymbol{L_i: es\ el\ Límite\ inferior}$ (No se incluye, por ser un **intervalo abierto** en dicho dato).
- $\boldsymbol{L_s: es\ el\ Límite\ superior}$ (Se incluye, por ser un **intervalo Cerrado** en dicho dato).

Para utilizar este método se debe realizar el siguiente procedimiento:

1. Se calcula el **Rango** ($\boldsymbol{R}$), esto es:

$$\boldsymbol{R = Dato\ Mayor - Dato\ Menor}$$

2. Se determina el **número de intervalos** (K) que tendrá la distribución, mediante la fórmula:

$$K = 1 + 3.3 \times \log n$$

Dónde:

- $\boldsymbol{K}$: $\boldsymbol{N\acute{u}mero\ de\ intervalos}$
- $\boldsymbol{n}$: $\boldsymbol{N\acute{u}mero\ de\ datos}$

Nota: Una vez obtenido **K,** se debe realizar el redondeo correspondiente en **números enteros.**

3. Se calcula la **Amplitud** ($\boldsymbol{A}$**:** es **la misma** para todos los intervalos) de cada uno de los intervalos mediante la ecuación:

$$A = \frac{R}{K}$$

Dónde:

- $\boldsymbol{R}$: $\boldsymbol{Rango}$
- $\boldsymbol{K}$: $\boldsymbol{N\acute{u}mero\ de\ intervalos}$

Nota: La Amplitud (A) **se aproxima al entero siguiente** (así sea entero), esto es se le suma 1 a la **Amplitud.**

4. Se calcula el **Rango Ampliado** (R_a):

$$R_a = A * K$$

Dónde:

- A: *Amplitud*
- K: *Número de intervalos*

5. Se calcula **una cantidad** que debe ser **diferente de cero**, de la siguiente manera:

$$* \, Cantidad = \frac{R_a - R}{2}$$

Dónde:

- R_a: *Rango Ampliado*
- R: *Rango*

6. Se calcula el **primer Límite Inferior** de al Distribución., de la siguiente forma:

$$L_i = Dato\ Menor - * \, Cantidad$$

4.2.1 EJERCICIOS DE APRENDIZAJE

Ejercicio 1: Entre los estudiantes de Administración de **UNIREMINGTON** 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

a. Se elabora la **Ficha técnica**:

CONCEPTO	CARACTERÍSTICA
Población	Estudiantes de Administración de **UNIREMINGTON**.
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** (Por pertenecer a **un intervalo**).

Nota: En la tabla determinada en la recolección de datos, existen muchos de ellos con **baja frecuencia**, por lo tanto para sacar conclusiones al respecto se deben **agrupar en intervalos.**

b. PROCEDIMIENTO

Se realiza de acuerdo a los pasos indicados en la definición de conceptos, esto es:

1. Se calcula el **Rango** $(\boldsymbol{R})$:

$$\boldsymbol{R = Dato\ Mayor - Dato\ Menor}$$

Reemplazando, se tiene:

$$\boldsymbol{R = 48 - 17 = 31}$$

$$\boldsymbol{R = 31}$$

---

2. Se calcula el **número de intervalos** $(\mathbf{K})$, que tendrá la Distribución:

$$K = 1 + 3.3 \times \log n$$

Reemplazando, se tiene:

$$\boldsymbol{K = 1 + 3.3 \times \log n \rightarrow K = 1 + 3.3 \times \log 60 \rightarrow}$$

$\boldsymbol{K} =$ **6,868,** utilizando el concepto de redondeo (ver unidad 1) a **números enteros**, se tiene que: $\boldsymbol{K = 7.}$

---

3. Se calcula la **Amplitud** $(\boldsymbol{A})$**,** (es la misma para todos los intervalos):

$$A = \frac{R}{K}$$

Reemplazando, se tiene:

$$A = \frac{31}{7} = 4{,}428$$

Por norma (determinada en la definición del concepto) se aproxima al número entero siguiente, por lo tanto:

$$\boldsymbol{A = 5}$$

Nota: Si el número obtenido hubiese sido un **número entero** (resultado de una **división exacta**), se aproxima **al entero siguiente** sumando **1.**

---

4. Se calcula el **Rango Ampliado** $(\boldsymbol{R_a})$:

$$\boldsymbol{R_a = A * K}$$

Reemplazando, se tiene:

$$\boldsymbol{R_a = A * K \rightarrow R_a = 5 * 7 \rightarrow R_a = 35}$$

---

5. Se calcula **una cantidad** que debe ser **diferente de cero**, de la siguiente manera:

$$* \boldsymbol{Cantidad = \frac{R_a - R}{2}}$$

Reemplazando, se tiene:

$$Cantidad = \frac{R_a - R}{2} \rightarrow Cantidad = \frac{35 - 31}{2} \rightarrow Cantidad = * 2$$

---

6. Se calcula el **primer Límite Inferior** de la Distribución., de la siguiente forma:

$$L_i = Dato\ Menor - * \ Cantidad$$

Reemplazando, se tiene:

$$L_i = 17 - * \ 2 \rightarrow L_i = 15$$

- Para determinar el **primer límite superior**, se tiene que todos los intervalos tiene **la misma amplitud**, por lo tanto:

$Primer\ Límite\ Superior$: $L_s = L_i + A \rightarrow L_s = 15 + 5 \rightarrow L_s = 20$

Como se van a agrupar intervalos de la forma: $L_i < X \leq L_s$, entonces:

Para el primer intervalo, se tiene entonces que: los datos **son mayores que 15** y **menores o iguales que 20**: en total se encontraron **12 datos** con esta característica, siendo esta la **primera Frecuencia Absoluta**.

Nota: Todo límite superior, es el límite inferior del siguiente intervalo.

Para calcular el **segundo límite superior** se aplica la fórmula $L_s = L_i + A$, se tiene entonces que el segundo intervalo tiene como **límite inferior 20** y como **límite superior**, $L_s = 20 + 5 = 25$, se encontraron **12 datos** que cumplen esta característica siendo esta será la **segunda frecuencia absoluta**.

Nota: Este procedimiento se repite tantas veces sea necesario, hasta que queden incluidos todos los datos del problema.

---

7. La tabla de Distribución de Frecuencias queda de la siguiente manera:

TABLA DE DISTRIBUCIÓN DE FRECUENCIAS

L_i	L_s	f_{ai}	f_{ri}	$\%$	X_i	f_{aai}	f_{rai}	$\%$ acumulado
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
		$N = 60$	$\sum = 1$	$\sum = 100$				

Nota: Los intervalos, de acuerdo a $L_i < X \leq L_s$, también se pueden expresar de la siguiente Forma:

L_i	L_s	$L_i < X \leq L_s$
15	20	$(15, 20]$
20	25	$(20, 25]$
25	30	$(25, 30]$
30	35	$(30, 35]$
35	40	$(35, 40]$
40	45	$(40, 45]$
45	50	$(45, 50]$

Donde:

$$(: \textit{indica que es abierto en diccho extremo}.$$

$$]: \textit{indica que es cerrado en dicho extremo}$$

Ejercicio 2: Los siguientes son los sueldos de los tecnólogos de Medellín de acuerdo con una muestra elegida aleatoriamente (en miles de pesos).

Nota: En este caso ya están los datos agrupados, por tanto no hay que aplicar los pasos para calcular los intervalos, simplemente se complementa la tabla.

L_i	L_s	$L_i < X \leq L_{s,}$	f_a
400	600	$(400, 600]$	30
600	800	$(600, 800]$	100
800	1000	$(800, 1000]$	55
1000	1200	$(1000, 1200]$	240
1200	1400	$(1200, 1400]$	25
1400	1600	$(1400, 1600]$	20

TABLA DE DISTRIBUCIÓN DE FRECUENCIAS

L_i	L_s	f_{ai}	f_{ri}	%	X_i	f_{aai}	f_{rai}	% *acumulado*
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
		$n = 470$	$\sum = 1$	$\sum = 100$				

- **GRÁFICOS PARA DATOS AGRUPADOS**

> **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).

Este **Histograma** puede ser de: **frecuencias absolutas** o de **porcentajes**.

Se grafica tomando:

- En el eje $\boldsymbol{X}$ los **límites** de los **intervalos**, y
- En el eje $\boldsymbol{y}$ las **frecuencias absolutas** o los **porcentajes** según sea el caso.

- **Marcas de clase** $(\boldsymbol{X_i})$**:**

Es un valor que identifica a cada intervalo "es el punto medio de cada intervalo", está dada por la siguiente ecuación:

$$\boldsymbol{x_i = \frac{L_i + L_s}{2}}$$

(Berenson, M. L. Y LEVINE, D. M., 1996, p. 38).

-

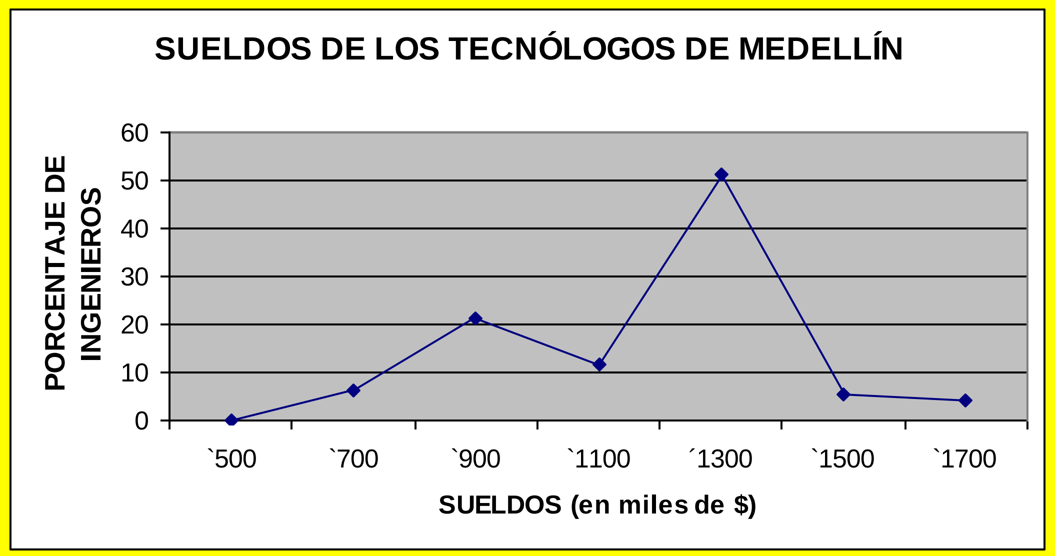
POLÍGONO: "Se puede obtener uniendo cada punto medio (marca de clase) de los rectángulos del histograma con líneas rectas, teniendo cuidado de agregar al inicio y al final marcas de clase adicionales, con el objeto de asegurar la igualdad del áreas".
http://sitios.ingenieriausac.edu.gt/estadistica/estadistica2/estadisticadescriptiva.html

Este **Polígono** puede ser de: **frecuencias absolutas** o de **porcentajes**.

Se grafica tomando:

- En el eje X las **Marcas de Clase**, y
- En el eje y las **frecuencias absolutas** o los **porcentajes** según sea el caso.

El siguiente gráfico corresponde al **polígono de porcentaje** del **ejemplo 2**



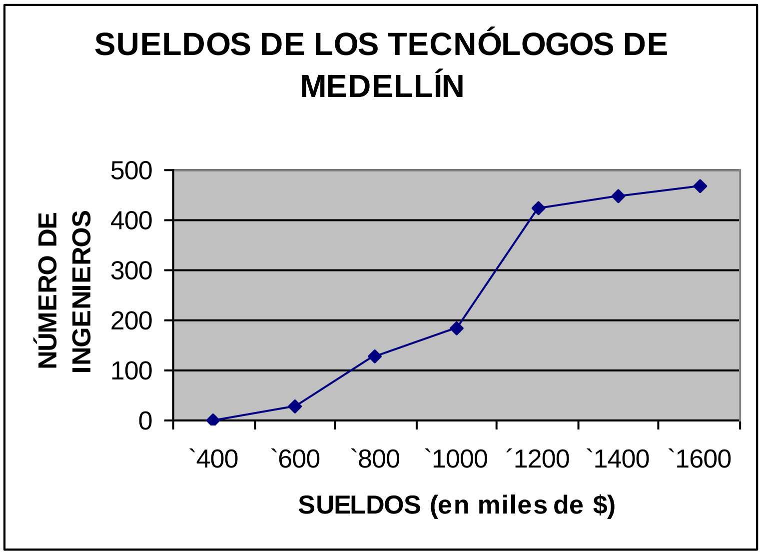
OJIVA: "Es el polígono de frecuencias acumuladas, es decir, que en ella se permite ver cuántas observaciones se encuentran por encima o por debajo de ciertos valores, en lugar de solo exhibir los números asignados a cada intervalo".
Tomado de: Ojiva (estadística)- Wikipedia, la enciclopedia libre es.wikipedia.org/wiki/Ojiva_(estadística

Esta **Ojiva** puede ser de: **frecuencias absolutas acumuladas** o de **porcentajes acumulados**.

Se grafica tomando:

- En el eje X los límites de los intervalos, y
- En el eje y las **frecuencias absolutas acumuladas.**

El siguiente gráfico corresponde al f_{aai} del **ejemplo 2:**



4.2.2 EJERCICIOS DE ENTRENAMIENTO

Para 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

PISTAS DE APRENDIZAJE

Traer a la memoria:

Recuerde que: Cuando, en un estudio estadístico, se obtienen **muchos datos** en la medición y **la frecuencia** de cada uno de estos datos es **muy baja,** es necesario **agruparlos** por **clase** o por **categoría** y cada una de éstas distribuciones se dispondrá en un **intervalo** determinado.

4.3 TEMA 2 CÁLCULO DE LAS MEDIDAS DE TENDENCIA CENTRAL PARA DATOS CUANTITATIVOS AGRUPADOS EN INTERVALOS DE CLASE

1. **Media Aritmética:**

$$\bar{X} \ o \ \mu = \frac{\sum X_i * f_{ai}}{n \ o \ N}$$

Para el ejemplo 1

L_i	L_s	f_{ai}	x_i	$x_i * f_{ai}$
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
		$N = 60$		$\sum 1685$

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

El promedio (la media aritmética) de la edad en que iniciaron sus estudios en la universidad los estudiantes de Administración de UNIREMINGTON es de **28,03 años.Para el ejemplo 2**

L_i	L_s	f_{ai}	x_i	$x_i * f_{ai}$
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
		$N = 470$		$\sum 461.000$

$$\bar{X}\ o\ \mu = \frac{461.000}{470} \rightarrow \bar{X}\ o\ \mu = 980{,}851$$

El promedio (la media aritmética) de los sueldos de los tecnólogos de Medellín es de **$980.851**

2. La moda (M_o)

$$M_o = L_i + \left(\frac{d_1}{d_1 + d_2}\right) * A$$

Dónde:

- $d_1 = f_{ai} - f_{a(i-1)}$
- $d_2 = f_{ai} - f_{a(i+1)}$
- $A = L_{si} - L_{li}$

Por lo tanto el intervalo que contiene la Moda es el más frecuente.

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

- **Para el ejemplo 1,** se tiene que:

  - $L_l = 20$
  - $d_1 = 18 - 12 = 6$
  - $d_2 = 18 - 7 = 11$
  - $A = 25 - 20 = 5$

Reemplazando en la ecuación de la **Moda**, se tiene:

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

La edad en que iniciaron sus estudios en la universidad los estudiantes de Administración de UNIREMINGTON más frecuente es de **21,764 años.**

- **Para el ejemplo 2,** se tiene que:

- $L_l = 1000$
- $d_1 = 240 - 55 = 185$
- $d_2 = 240 - 25 = 215$
- $A = 1200 - 1000 = 200$

Reemplazando en la ecuación de la **Moda**, se tiene:

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

Como está dado en miles, se multiplica por 1000, esto es:

$$M_o = 1092,5 \times 1000 = 1.092.500$$

El sueldo de los tecnólogos de Medellín más frecuente es de **$1`092.500**

3. **Mediana (M_e):**

Esta dada por la ecuación:

$$M_e = L_i + \frac{\frac{n}{2} - f_{aa(i-1)}}{f_{ai}} * A$$

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

1. $\frac{n}{2}$ y se busca en f_{aa}
2. 50% y se busca en **porcentaje acumulado**.

- **Para el ejemplo 1**

El intervalo que contiene la mediana lo ubicamos en $\frac{n}{2}=\frac{60}{2}=30$ 60/2 = 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, se tiene que:

$$M_e = L_i + \frac{\frac{n}{2} - f_{aa(i-1)}}{f_{ai}} * A \rightarrow M_e = 20 + \left(\frac{30-12}{18}\right) * 5$$

$$M_e = 25$$

El 50% de la edad en que iniciaron sus estudios en la universidad los estudiantes de Administración de UNIREMINGTON es de **25 años**.

- **Para el ejemplo 2**

El intervalo que contiene la mediana lo ubicamos $\frac{n}{2} = 235,$ 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, se tiene que:

$$M_e = L_i + \frac{\frac{n}{2} - f_{aa(i-1)}}{f_{ai}} * A \rightarrow M_e = 100 + \left(\frac{235-185}{240}\right) * 200$$

$$M_e = 1041,667$$

Como está dado en miles, se multiplica por 1000, esto es:

$$M_e = 1041,667 \times 1000 = 1.041.667$$

El 50% de los sueldos de los tecnólogos de Medellín es de $\$\ 1.041.667$

4.3.1 EJERCICIO DE APRENDIZAJE

A continuación se dan las notas de los estudiantes de sistemas de acuerdo con una muestra elegida al azar, calcule e interprete: La Media Aritmética, La Moda y la Mediana

L_i	L_s	f_{ai}
0	**1**	**15**
1	**2**	**15**
2	**3**	**20**
3	**4**	**28**
4	**5**	**22**

PISTAS DE APRENDIZAJE

Traer a la memoria:

Recuerde que: El intervalo que contiene la Moda es el más frecuente.

4.4 TEMA 3 MEDIDAS DE POSICIÓN RELATIVA PARA DATOS CUANTITATIVOS AGRUPADOS EN INTERVALOS DE CLASE

Las medidas de posición facilitan información sobre la serie de datos que se están analizando.

Una vez definidos los conceptos básicos en el estudio de una distribución de frecuencias de una variable, se estudiarán las distintas formas de resumir dichas distribuciones mediante medidas de posición (o de centralización), teniendo presente el error cometido en el resumen mediante las correspondientes medidas de dispersión.

Se trata de encontrar unas medidas que sinteticen las distribuciones de frecuencias. En vez de manejar todos los datos sobre las variables, tarea que puede ser pesada, se puede caracterizar su distribución de frecuencias mediante algunos valores numéricos, eligiendo como resumen de los datos un valor central alrededor del cual se encuentran distribuidos los valores de la variable.

Tomado de:http://www.monografias.com/trabajos14/medidasposicion/medidasposicion.shtml#ixzz3DQMs8KcO

Estas medidas **dividen la distribución** en **partes iguales**, así como La Mediana, por lo tanto se calculan e interpretan similar a ella. Entre estas medidas de posición relativa se tienen:

1. $\boldsymbol{Cuartiles\ (Q)}$: Dividen la distribución en **4 partes iguales**, de la siguiente manera:

$$Q_1 = 25\%$$

$$Q_2 = 50\%$$

$$Q_1 = 75\%$$

$$Q_1 = 100\%$$

Se calcula con la siguiente ecuación:

$$Q = L_i + \left(\frac{\#Q \times \frac{n}{4} - f_{aa(i-1)}}{f_{ai}}\right) * A$$

Nota: Para ubicar el intervalo que contiene Q se utiliza:

1. $\frac{\#Q * n}{4}$ y se busca en f_{aa}

2. El **porcentaje respectivo** se busca en **el porcentaje acumulado**.

4.4.1 EJERCICIO DE APRENDIZAJE:

Si nos piden **el cuartil 1** del **ejemplo 2**, quedaría:

$Q_1 = \frac{1 * 470}{4} \rightarrow Q_1 = 117,5,$ en la f_{aai} está en 130 o el 25% está acá también (en la tabla es el intervalo azul); Reemplazando en la fórmula, se tiene:

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

Como está dado en miles de pesos: $775 \times 1000 = 775.000$

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

1. ***Deciles*** (D): "Dividen la distribución en **10 partes iguales**" (Murray, 1995, p. 66).

La división se dará de la siguiente forma:

$$D_1 = 10\%$$

$$D_2 = 20\%$$

$$D_3 = 30\%$$

$$D_4 = 40\%$$

$$\vdots$$

$$D_{10} = 100\%$$

Se calcula con la siguiente ecuación:

$$D = L_i + \left(\frac{\# D * \frac{n}{10} - f_{aa(i-1)}}{f_{ai}}\right) * A$$

Nota: Para ubicar el intervalo que contiene D se utiliza:

1. $\frac{\#D * n}{10}$ y se busca en f_{aa}

2. El **porcentaje respectivo** se busca en **el porcentaje acumulado.**

4.4.2 EJERCICIO DE APRENDIZAJE:

Determinar el **Decil 3** del ejemplo 2:

Se tiene que:

$D_3 = 3 * \frac{470}{10} \rightarrow D_3 = 141,$ en la f_{aai} está en 185 o el 30% está acá también (En la tabla es el intervalo de **color marrón**).

Reemplazando en la ecuación se tiene:

$$D_3 = L_i + \left(\frac{\# D * \frac{n}{10} - f_{aa(i-1)}}{f_{ai}}\right) * A = 800 + \left(\frac{141 - 130}{55}\right) * 200 \rightarrow$$

$$D_3 = 840$$

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

1. $Percentiles\ (P)$: "**Dividen** la distribución en **100 partes iguales**"

(Murray, 1995, p. 66).

La división se dará de la siguiente forma:

$$P_1 = 1\%$$

$$P_2 = 2\%$$

$$P_3 = 3\%$$

$$P_4 = 4\%$$

$$\vdots$$

$$P_{100} = 100\%$$

Se calcula con la siguiente ecuación:

$$P = L_i + \left(\frac{\#\, P * \frac{n}{100} - f_{aa(i-1)}}{f_{ai}}\right) * A$$

Nota: Para ubicar el intervalo que contiene D se utiliza:

1. $\frac{\#P*n}{100}$ y se busca en f_{aa}

2. El **porcentaje respectivo** se busca en **el porcentaje acumulado.**

4.4.3 EJERCICIO DE APRENDIZAJE

Determinar **el percentil 5** del ejemplo 2:

Se tiene que:

$P_5 = 5 * \frac{470}{100} \rightarrow P_5 = 23,5$ En la f_{aai} está en 30 *o el* 5% está acá también (En la tabla es el intervalo de **color Fucsia)**.

Reemplazando en la ecuación se tiene:

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

Como está dado en miles de pesos:

$$P_5 = 556,667 * 1000 \rightarrow P_5 = 556.667$$

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

4.4.4 EJERCICIOS DE ENTRENAMIENTO

En cada uno de los siguientes ejercicios calcule e interprete **las medidas de variabilidad** y escriba sus propias conclusiones:

1. Las siguientes son las notas de dos grupos de estadística de UNIREMINGTON, de una muestra tomada al azar:

GRUPO I	
x_i $(nota)$	f_{ai} $(\#\ estudiantes)$
1	4
2	11
3	10
4	5
5	12
NP: $no\ presentaron\ la\ prueba$	4

GRUPO II	
x_i $(nota)$	f_{ai} $(\#\ estudiantes)$
1	17
2	16
3	2
4	3
5	4
NP: $no\ presentaron\ la\ prueba$	6

2. 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	***TIEMPO (en segundos)***					
1	**61.3**	**62.1**	**6.17**	**62.9**	**63.2**	**61.9**
2	**62.8**	**63**	**62.5**	**61.9**	**60.7**	**61.0**
3	**61.9**	**63.7**	**62.9**	**61.9**	**61.5**	**62.0**

PISTAS DE APRENDIZAJE

Traer a la memoria:

Recuerde que: Una vez definidos los conceptos básicos en el estudio de una distribución de frecuencias de una variable, se estudiarán las distintas formas de resumir dichas distribuciones mediante medidas de posición (o de centralización), teniendo presente el error cometido en el resumen mediante las correspondientes medidas de dispersión.

4.5 TEMA 4 MEDIDAS DE VARIABILIDAD O DISPERSIÓN PARA DATOS CUANTITATIVOS

“Una medida de **dispersión** o **variabilidad** nos determina el grado de **acercamiento** o **distanciamiento** de los valores de una distribución frente a su **promedio de localización**, sobre la base de que:

Entre **más grande** sea el **grado de variación, menor uniformidad** tendrán **los datos** (sinónimo de **heterogeneidad**) y por lo tanto **menor representatividad** o **confiabilidad del promedio** de **tendencia central** o **localización** por haber sido obtenido de **datos dispersos**.

Por el contrario, si este valor **es pequeño** (respecto a la unidad de medida) entonces, hay **una gran uniformidad** entre **los datos**. Cuando **es cero** quiere decir que todos los **datos son iguales**.

Hay básicamente **dos tipos de medidas de dispersión**: **Medidas Absolutas** y **Medidas Relativas.**

Las **medidas absolutas** se caracterizan por ser **números concretos**, es decir, valores expresados en las **mismas unidades** de la variable en estudio y que por lo tanto **no permiten comparaciones** o **análisis** respecto a **la mayor** o **menor** dispersión de **series** expresadas en **diferentes unidades**. Estas medidas son: **la varianza**, **la desviación estándar** y el **rango intercuartilico**.

Las **medidas relativas** de **dispersión** son **valores abstractos**, es decir, **medidas adimensionales** y por lo tanto **no expresadas en ninguna unidad específica**, obviando así el inconveniente señalado para las medidas absolutas. La principal medida es el coeficiente de variación".

Tomado de: Introducción a medidas de variabilidad - Probabilidad y
www.virtual.unal.edu.co/cursos/ciencias/2001065/.../cont_129_29.htm

Para tener una mejor comprensión de estos conceptos, veamos el siguiente ejemplo:

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

OPERARIO	PRODUCCIÓN/HORA DURANTE UN DÍA							
A	60	30	40	100	20	80	60	40
B	50	60	60	40	80	50	50	40

Para calcular el promedio (la media aritmética) de las unidades producidas por hora en el día de cada operario, se aplica la fórmula correspondiente a la media aritmética, esto es:

- **Operario A**:

$$\mu_A = \frac{60 + 30 + 40 + 100 + 20 + 80 + 60 + 40}{8} \rightarrow \mu_A = \frac{430}{8}$$

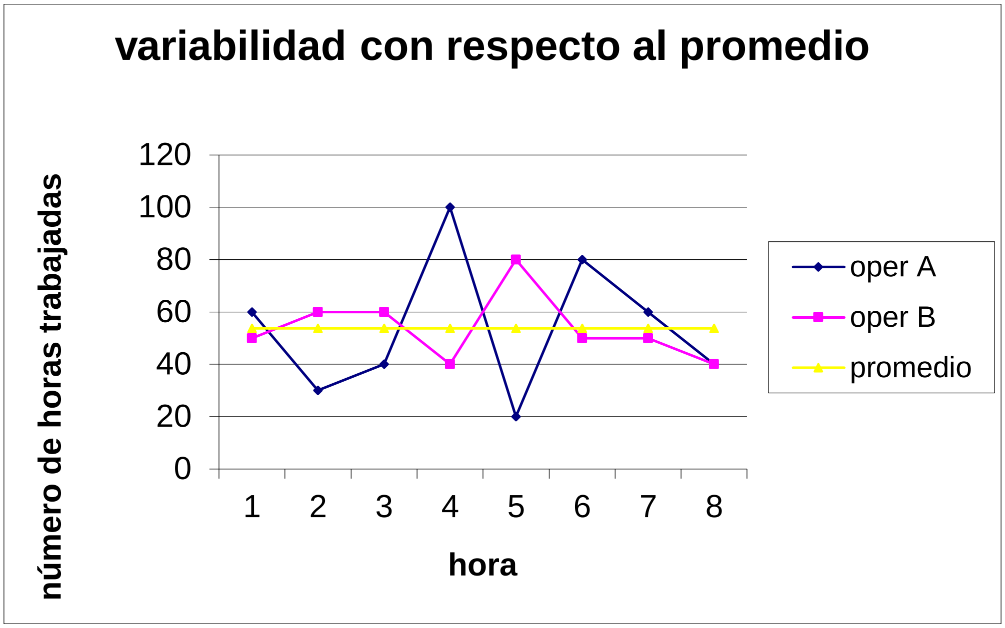
$$\mu_A = 53.75\ unidades$$

- **Operario B**:

$$\mu_B = \frac{50 + 60 + 60 + 40 + 80 + 50 + 50 + 40}{8} \rightarrow \mu_B = \frac{430}{8}$$

$$\mu_B = 53.75\ unidades$$

De acuerdo a los resultados obtenidos, ambos operarios tienen **el mismo promedio**. Pero analicemos algo adicional, se traza un diagrama de líneas para cada operario:



Analizando el gráfico obtenido se tiene que: el **operario A** presenta **mayor variación** con respecto **al promedio** que **el operario B**.

- **Medidas de Dispersión o Variabilidad**

Las medidas que vamos a ver a continuación tienen dos aplicaciones:

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

1. La Varianza

Se define como:

> "La varianza es la media aritmética del cuadrado de las desviaciones respecto a la media de una distribución estadística".
> La varianza se representa por: σ^2
> *Tomada de: Varianza - Vitutor www.vitutor.com/estadistica/descriptiva/a_15.html*

La varianza se calcula de la siguiente manera:

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

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

$$\sigma^2 = \frac{(x_1-\mu)^2+(x_2-\mu)^2+(x_3-\mu)^2+\cdots+(x_n-\mu)^2}{N} \rightarrow$$

$$\sigma^2 = \frac{\sum(x_i-\mu)^2}{N}$$

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

$$S^2 = \frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+(x_3-\overline{x})^2+\cdots+(x_n-\overline{x})^2}{n-1} \rightarrow$$

$$\sigma^2 = \frac{\sum(x_i-\overline{x})^2}{n-1}$$

4.5.1 EJERCICIO DE APRENDIZAJE

En el ejemplo que se tiene sobre los operarios:

Recuerde que: $\mu_A = 53.75 = \mu_B = 53.75$

OPERARIO A	
x_i	$(x_i-\mu_A)^2$
60	$(60-53.75)^2 = (6.25)^2 = 39{,}0625$
30	$(30-53.75)^2 = (-23{,}75)^2 = 564{,}0625$
40	$(40-53.75)^2 = (13{,}75)^2 = 189{,}0625$

100	$(100-53.75)^2=(46.25)^2=2.139,0625$
20	$(20-53.75)^2=(-33.75)^2=1139,0625$
80	$(80-53.75)^2=(26.25)^2=689,0625$
60	$(60-53.75)^2=(6.25)^2=39,0625$
40	$(40-53.75)^2=(13,75)^2=189,0625$
$\sum=430$	$\sum=4987,5$

$$\mu_A=\frac{\sum x_i}{N}\rightarrow\mu_A=\frac{430}{8}\rightarrow\mu_A=53,75$$

$$\sigma^2(A)=\frac{\sum(x_i-\mu)^2}{N}\rightarrow\sigma^2(A)=\frac{4987,5}{8}\rightarrow\sigma^2(A)=623,438$$

OPERARIO B	
x_i	$(x_i-\mu_B)^2$
50	$(50-53.75)^2=(-3,75)^2=14,0625$
60	$(60-53.75)^2=(6,25)^2=39,0625$
60	$(60-53.75)^2=(6,25)^2=39,0625$

40	$(40-53.75)^2=(-13,75)^2=189,0625$
80	$(80-53.75)^2=(26,25)^2=689,0625$
50	$(50-53.75)^2=(-3,75)^2=14,0625$
50	$(50-53.75)^2=(-3,75)^2=14,0625$
40	$(40-53.75)^2=(-13,75)^2=189,0625$
$\sum=430$	$\sum=1187,5$

$$\mu_B=\frac{\sum x_i}{N}\rightarrow \mu_B=\frac{430}{8}\rightarrow \mu_B=53,75$$

$$\sigma^2(B)=\frac{\sum(x_i-\mu)^2}{N}\rightarrow \sigma^2(B)=\frac{1187,5}{8}\rightarrow \sigma^2(A)=148,438$$

b. si los datos **están organizados** en una tabla de frecuencias, la **varianza** se calcula:

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

$$\sigma^2=\frac{(x_1-\mu)^2f_{a_1}+(x_2-\mu)^2f_{a_2}+(x_3-\mu)^2f_{a_3}+\cdots+(x_n-\mu)^2f_{a_n}}{N}$$

$$\sigma^2=\frac{\sum(x_i-\mu)^2f_{a_i}}{N}$$

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

$$\sigma^2 = \frac{(x_1-\overline{x})^2 f_{a_1} + (x_2-\overline{x})^2 f_{a_2} + (x_3-\overline{x})^2 f_{a_3} + \cdots + (x_n-\overline{x})^2 f_{a_n}}{n-1}$$

$$\sigma^2 = \frac{\sum (x_i-\overline{x})^2 f_{a_i}}{n-1}$$

4.5.2 EJERCICIO DE APRENDIZAJE

En el caso de los operarios, se tiene:

- **OPERARIO A**

x_i	f_{a_i}	$x_i * f_{a_i}$	$(x_i-\mu)^2 f_{a_i}$
20	1	$20 * 1 = 20$	$(20-53,75)^2 * 1 = (-33,75)^2 * 1 = 1139,0625$
30	1	$30 * 1 = 30$	$(30-53,75)^2 * 1 = (-23,75)^2 * 1 = 564,0625$
40	2	$40 * 2 = 80$	$(40-53,75)^2 * 2 = (-13,75)^2 * 2 = 378,125$
60	2	$60 * 2 = 120$	$(60-53,75)^2 * 2 = (6,25)^2 * 2 = 78,125$
80	1	$80 * 1 = 80$	$(80-53,75)^2 * 1 = (26,25)^2 * 1 = 689,0625$
100	1	$100 * 1 = 100$	$(100-53,75)^2 * 1 = (46,25)^2 * 1 = 2139,0625$
	$N = 8$	$\sum = 430$	$\sum = 4987,5$

$$\mu_A = \frac{\sum x_i * f_{a_i}}{N} = \frac{430}{8} = 53,75$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2 f_{a_i}}{N} = \frac{4987,5}{8} = 623,438$$

- **OPERARIO B**

x_i	f_{a_i}	$x_i * f_{a_i}$	$(x_i - \mu)^2 f_{a_i}$
40	2	$40 * 2 = 80$	$(40 - 53,75)^2 * 1 = (-13,75)^2 * 2 = 378,125$
50	3	$50 * 3 = 150$	$(50 - 53,75)^2 * 1 = (-3,75)^2 * 3 = 42,1875$
60	2	$60 * 2 = 120$	$(60 - 53,75)^2 * 2 = (6,25)^2 * 2 = 78,125$
80	1	$80 * 1 = 80$	$(80 - 53,75)^2 * 2 = (26,25)^2 * 1 = 689,0625$
	$N = 8$	$\sum = 430$	$\sum = 1187,5$

$$\mu_B = \frac{\sum x_i * f_{a_i}}{N} = \frac{430}{8} = 53,75$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2 f_{a_i}}{N} = \frac{1187,5}{8} = 148,438$$

- **Procedimiento para calcular la Varianza**

A continuación se da el paso a paso seguido para obtener el valor de **la varianza**:

Cálculo de la Varianza
1. Se calcula **el promedio** o **Media Aritmética** $\mu \; o \; \bar{x}$.
2. Se calcula la **Desviación** de cada dato con respecto **al promedio** y este resultado se **eleva al cuadrado**: $(x_i - \mu)^2 \; o \; (x_i - \bar{x})^2$
3. Cada una de las **Desviaciones** anteriores se multiplican por la **frecuencia absoluta**, en caso de que exista: $(x_i - \mu)^2 * f_{a_i} \; o \; (x_i - \bar{x})^2 * f_{a_i}$
4. Se **suman**, para cada caso, los anteriores resultados.
5. Esta sumatoria **se divide** por el número de datos, $N \; o \; n - 1$, según el caso.

> **Nota:** Como se puede observar, **la varianza** da en **unidades al cuadrado,** esto resulta ilógico y **no se puede interpretar**, por esta razón se creó la siguiente medida de variación: **La Desviación Estándar**

2. DESVIACIÓN ESTÁNDAR

También se conoce como **Desviación Típica**, se define como:

"**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)

La varianza se calcula de la siguiente manera:

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

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

$$\sigma = \sqrt{\frac{(x_1-\mu)^2 f_{a_1} + (x_2-\mu)^2 f_{a_2} + (x_3-\mu)^2 f_{a_3} + \cdots + (x_n-\mu)^2 f_{a_n}}{N}}$$

2. Como **estadística**, es decir, si **los datos** se toman de **una muestra.**

$$S = \sqrt{\frac{(x_1-\bar{x})^2 f_{a_1} + (x_2-\bar{x})^2 f_{a_2} + (x_3-\bar{x})^2 f_{a_3} + \cdots + (x_n-\bar{x})^2 f_{a_n}}{n-1}}$$

Para el ejercicio de aprendizaje, con los operarios A y B, la desviación estándar quedaría:

- $\boldsymbol{OPERARIO\ A}$: $\boldsymbol{\sigma = \sqrt{623,438} = 24,969}$

Esto es: 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.**

- $\boldsymbol{OPERARIO\ B}$: $\boldsymbol{\sigma = \sqrt{148,438} = 12,184}$

Esto es: 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**.

> **Nota: 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.

4.5.3 EJERCICIO DE APRENDIZAJE

- 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**; para evaluar este tipo de resultados se tiene un elemento denominado **Coeficiente de Dispersión o de variación**, cuya conceptualización es la siguiente:

- **COEFICIENTE DE DISPERSIÓN O DE VARIACIÓN (C. V)**

Es 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, se calcula de la siguiente forma:

- $C.V = \frac{\sigma}{\bar{x}} * 100\ (para\ una\ población)$
- $C.V = \frac{S}{\bar{x}} * 100\ (para\ una\ muestra)$

Para el ejercicio anterior:

- $C.V_C = \frac{5}{40} * 100 = 12.5\%$
- $C.V_D = \frac{15}{160} * 100 = 9.4\%$

Se concluye, entonces, que el **operario D** presenta **menor variabilidad**.

Para el ejemplo de los operarios A y B:

- $C.V_A = \frac{24{,}969}{53{,}75} * 100 = 46{,}454\ \%$
- $C.V_B = \frac{12{,}184}{53{,}75} * 100 = 22{,}668\ \%$

4.5.4 EJERCICIO DE APRENDIZAJE

Para un conjunto de datos agrupados:

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

L_i	L_s	f_{a_i}	x_i	$x_i * f_{a_i}$	$(x_i-\mu)^2 * f_{a_i}$
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
		$N = 81$		$\sum 2810.5$	$\sum 7172.84$

$\mu = \frac{\sum x_i * f_{a_i}}{N} = \frac{2810.5}{81} = 34,698$ Promedio de edad de los empleados.

$\sigma^2 = \frac{\sum (x_i-\mu)^2 f_{a_i}}{N} = \frac{7172,84}{81} = 88,554$, no se interpreta.

$$\sigma = \sqrt{\sigma^2} = \sqrt{88,554} = 9,41$$

$C.V = \frac{9,41}{34,698} * 100 = 27,127\%$ Porcentaje de variación.

Las edades de los empleados de la empresa tienen una **Desviación Promedio** de **9,41 años**, con relación al promedio que es de **34,689 años**.

4.5.5 EJERCICIOS DE ENTRENAMIENTO

1. 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	88	88	123

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

2. 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		
L_I	L_S	F_{a_i}
0	**1**	**13**
1	**2**	**40**
2	**3**	**10**
3	**4**	**24**
4	**5**	**3**

CENTRO PSICOLÓGICO B		
L_I	L_S	F_{a_i}
0	**1**	**33**
1	**2**	**10**
2	**3**	**10**
3	**4**	**40**
4	**5**	**7**

3. Seleccione la respuesta correcta:

a. Los ingresos más frecuentes del Centro psicológico A están en el intervalo:

   1. De 0 a 1
   2. De 1 a 2
   3. De 3 a 4

b. Los ingresos menos frecuentes del Centro psicológico A están en el intervalo:

   1. De 2 a 3
   2. De 4 a 5
   3. De 2 a 3

c. Los ingresos más frecuentes del Centro psicológico B están en el intervalo:

   1. De 4 a 5
   2. De 1 a 2
   3. De 3 a 4

d. Los ingresos menos frecuentes del Centro psicológico B están dos intervalos:

   1. Verdadero
   2. Falso

e. Los ingresos menores de los dos Centros psicológicos están en el intervalo:

   1. De 0 a 1
   2. De 1 a 2
   3. De 3 a 4

f. Los ingresos mayores de los dos Centros psicológicos están en el intervalo:

g.

   1. De 4 a 5
   2. De 1 a 2
   3. De 3 a 4

h. Se puede afirmar que el Centro psicológico B tuvo el mayor promedio de ingresos:

1. Verdadero
2. Falso

PISTAS DE APRENDIZAJE

Traer a la memoria:

Recuerde que: Hay básicamente **dos tipos de medidas de dispersión: Medidas Absolutas y Medidas Relativas.**

Las **medidas absolutas** se caracterizan por ser **números concretos,** es decir, valores expresados en las **mismas unidades** de la variable en estudio y que por lo tanto **no permiten comparaciones o análisis** respecto a **la mayor o menor** dispersión de **series** expresadas en **diferentes unidades.** Estas medidas son: **la varianza, la desviación estándar y el rango intercuartílico.**

Las **medidas relativas** de **dispersión** son **valores abstractos,** es decir, **medidas adimensionales** y por lo tanto **no expresadas en ninguna unidad específica,** obviando así el inconveniente señalado para las medidas absolutas. La principal medida es el coeficiente de variación".

5 PISTAS DE APRENDIZAJE

Recuerde que: Como que como se ha citado anteriormente la **Estadística** trata sobre el recuento, ordenación y clasificación de los datos obtenidos por las observaciones, para poder hacer comparaciones y sacar conclusiones

Recuerde que: La correcta redacción de la ficha técnica es importante para garantizar la satisfacción del consumidor, especialmente en los casos donde la incorrecta utilización de un producto puede resultar en daños personales o materiales o responsabilidades civiles o penales.

Recuerde que: Una distribución de frecuencias: es la agrupación de datos en categorías mutuamente excluyentes que indican el número de observaciones en cada categoría.

Recuerde que: Los datos numéricos, sean pocos o muchos, siempre se organizan en **una Fila en forma ascendente.**

Recuerde que: Las **Medidas de Tendencia Central,** nos permiten una interpretación de los datos obtenidos en cualquier medición que se realice.

Para encontrar la Media, la Moda y la Mediana, se deben ordenar los datos de menor a mayor.

Recuerde que: Cuando, en un estudio estadístico, se obtienen **muchos datos** en la medición y **la frecuencia** de cada uno de estos datos es **muy baja**, es necesario **agruparlos** por **clase** o por **categoría** y cada una de éstas distribuciones se dispondrá en un **intervalo** determinado.

Recuerde que: El intervalo que contiene la Moda es el más frecuente.

Recuerde que: Una vez definidos los conceptos básicos en el estudio de una distribución de frecuencias de una variable, se estudiarán las distintas formas de resumir dichas distribuciones mediante medidas de posición (o de centralización), teniendo presente el error cometido en el resumen mediante las correspondientes medidas de dispersión.

Recuerde que: Hay básicamente **dos tipos de medidas de dispersión**: **Medidas Absolutas** y **Medidas Relativas.**

Las **medidas absolutas** se caracterizan por ser **números concretos**, es decir, valores expresados en las **mismas unidades** de la variable en estudio y que por lo tanto **no permiten comparaciones** o **análisis** respecto a **la mayor** o **menor** dispersión de **series** expresadas en **diferentes unidades**. Estas medidas son: **la varianza**, **la desviación estándar** y el **rango intercuartilico**.

Las **medidas relativas** de **dispersión** son **valores abstractos**, es decir, **medidas adimensionales** y por lo tanto **no expresadas en ninguna unidad específica**, obviando así el inconveniente señalado para las medidas absolutas. La principal medida es el coeficiente de variación".

Tomado de: Introducción a medidas de variabilidad - Probabilidad y ... www.virtual.unal.edu.co/cursos/ciencias/2001065/.../cont_129_29.html/

6 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 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.

7 BIBLIOGRAFÍA

- 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.

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 21de agosto de 2007. Extraído el 24 de octubre de 2009 de http://sitios.ingenieria-usac.edu.gt/estadistica/estadistica2/estadisticadescriptiva.html