INDEC

DIRECCION DE CAPACITACION Y
DESARROLLO DE RECURSOS HUMANOS

PROGRAMA DE CAPACITACION EN ESTADISTICA

Directora: Lic. Ana María Edwin

INTRODUCCION a la ESTADISTICA DESCRIPTIVA

MANUAL DE CURSO

Estadístico: Néstor Kvasina

1993

Ml / 338
ej 1

INDICE

UNIDAD 12

UNIDAD I

1 - La Estadística

La palabra **estadística** suele aplicarse generalmente para referir a dos acepciones distintas:

- *Recolección de datos numéricos:* es éste el significado más usualmente asignado a la palabra estadística. La aplicación de la palabra estadística con este sentido es precisa cuando tales datos numéricos están presentados de manera ordenada y sistemática. Una información numérica cualquiera puede no constituir una estadística; para merecer este apelativo, los datos deben ser un conjunto coherente, establecido de forma sistemática según un criterio de ordenación.

- *Ciencia:* en este significado, la Estadística estudia el comportamiento de fenómenos masivos. Como todas las ciencias, busca las características generales de un conjunto y prescinde de las características particulares de cada elemento de dicho conjunto. La cualidad esencial, por lo tanto, es que la Estadística no se detiene a estudiar el comportamiento de un caso aislado; estudia siempre grupos o conjuntos de casos, ya sean grupos de objetos, personas, procesos, etc.

Habitualmente, al examinar un conjunto de datos se aprecia cierta regularidad o estabilidad en su comportamiento, de forma que si el conjunto es lo suficientemente grande, presenta características estables.

El propósito de la Estadística es precisamente hallar esas regularidades en los fenómenos de masa, regularidades que, además de servir para la **descripción** del fenómeno, pueden utilizarse con fines de **predicción**.

Para hallar las regularidades del comportamiento de un conjunto de datos, hay que utilizar los casos particulares, pero estos pierden su individualidad para hacer resaltar tan sólo las características del conjunto del que forman parte. Por ello, puede decirse que la Estadística, como ciencia, no estudia el comportamiento de casos aislados; simplemente, los utiliza como medio para estudiar el comportamiento de un grupo.

Una observación que conviene remarcar, es que la regularidad o estabilidad se obtiene mejor cuanto mayor es el grupo que se observa.

Puede resumirse, sin dar una definición completa, que la Estadística como ciencia estudia los fenómenos de masa para hallar en ellos regularidades en el comportamiento, regularidades que sirven para describir el fenómeno y para efectuar predicciones.

2 - Reseña histórica

Las preocupaciones estadísticas se remontan a la antigüedad, pero el contenido de las mismas ha variado sensiblemente a través del tiempo.

Desde el cuarto milenio a.C. los chinos realizaban censos de población y utilizaban tablas de estadística aplicadas a los problemas agrícolas.

Los egipcios, los griegos y los romanos también realizaron múltiples investigaciones recurriendo a la estadística como herramienta.

Ciertamente, entonces no se conocía la palabra estadística y nadie pensaba en inducir leyes de comportamiento de los datos recogidos con mayor o menor exactitud, pero se conocían las operaciones censales y catastrales que ayudaban a describir situaciones reales.

Las primeras tentativas para sistematizar los conocimientos y hacer de la estadística una ciencia autónoma de otras ciencias, surgen en Alemania en el Siglo XVII, mientras que en Inglaterra se logra un nuevo progreso al superar la fase meramente descriptiva y comenzar a utilizar los datos con fines predictivos.

Más tarde, a partir del estudio de los juegos de azar, el cálculo de las probabilidades se incorpora como un instrumento de análisis extremadamente poderoso para el estudio de fenómenos cuyas causas son demasiado complejas para conocerlas totalmente y poder analizarlas sin su uso.

A partir de comienzos del Siglo XX, la estadística logra su expansión definitiva y es considerada como una ciencia autónoma, desarrollando su aplicación en todas las ramas del saber.

La biología, la meteorología, la investigación agronómica, la demografía, la psicología, la sociología y muchas otras ciencias han sido transformadas mediante el empleo de métodos estadísticos. Esta invasión de la estadística en todos los dominios de la investigación pura o aplicada permite que los métodos estadísticos se desarrollen permanentemente para dar respuesta a los distintos problemas a resolver.

3 - La producción de estadísticas oficiales: El Sistema Estadístico Nacional y el INDEC como organismo rector del Sistema.

Previo a la creación del INDEC, la organización del servicio estadístico nacional estaba descentralizada, sin la existencia a nivel nacional de un organismo con el poder suficiente, en el aspecto normativo, que asegurara la coherencia de las tareas estadísticas oficiales.

El Instituto Nacional de Estadística y Censos -INDEC- es el organismo público, de carácter técnico, que unifica la orientación y ejerce la dirección superior de todas las actividades estadísticas oficiales que se realizan en el territorio de la República Argentina. Su creación y funcionamiento está reglamentado por la Ley 17.622/68 y el Decreto 3110/70.

Esta Ley le confiere al Instituto la responsabilidad directa de la organización y dirección de los operativos nacionales de índole censal, así como las grandes encuestas. También le asigna la elaboración de indicadores básicos de orden social y económico y de otras estadísticas básicas.

El INDEC también tiene la responsabilidad de coordinar el Sistema Estadístico Nacional -SEN-, que está integrado por los servicios estadísticos de los organismos nacionales, provinciales y municipales, bajo el principio de centralización normativa y descentralización ejecutiva. Esto significa que el INDEC es responsable del desarrollo metodológico y normativo para la producción de estadísticas oficiales, asegurando la comparabilidad de la información originada en distintas fuentes.

En cada provincia existe una Dirección de Estadística dependiente del gobierno provincial. Dichas Direcciones coordinan las Sistemas Estadísticos Provinciales, e intervienen en la captura, ingreso y procesamiento de información a nivel provincial. Esta es consolidada por el INDEC o por otros servicios nacionales para la obtención de información a nivel nacional.

La Ley 17.622 cita a los integrantes del Sistema Estadístico Nacional -SEN- que se detallan a continuación:

a) El Instituto Nacional de Estadística y Censos

b) Los organismos centrales de estadística que son:

I - Los servicios estadísticos de los Ministerios y Secretarías de Estado.
II - Los servicios estadísticos de los Comandos en Jefe de las Fuerzas Armadas.

III - Los servicios estadísticos de organismos descentralizados de la Administración Nacional.
IV - Los servicios estadísticos de las Empresas del Estado.

c) Los organismos periféricos de estadística que son:

I - Los servicios estadísticos de los gobiernos provinciales.
II - Los servicios estadísticos de los organismos municipales.
III - Los servicios estadísticos de las reparticiones autárquicas y descentralizadas, provinciales y municipales.
IV - Los servicios estadísticos de las Empresas provinciales y municipales.
V - Los servicios estadísticos de los entes interprovinciales".

A su vez, la Ley 17622 además de regular el funcionamiento de la actividad estadística en el ámbito oficial, incluye en su articulado normas estrictas sobre la obligación de tratar con reserva la información individual suministrada por los informantes.

El personal que se desempeña en organismos pertenecientes al Sistema Estadístico Nacional está obligado, cualquiera sea su condición, función y jerarquía, a cumplir con la reserva que impone la Ley Nº 17622 en sus artículos 10 y 13 que señalan:

ART. 10: "Las informaciones que se suministren a los organismos que integran el Sistema Estadístico Nacional, en cumplimiento de la presente Ley, serán estrictamente secretas y sólo se utilizarán con fines estadísticos. Los datos deberán ser suministrados y publicados, exclusivamente en compilaciones de conjunto, de modo que no pueda ser violado el secreto comercial o patrimonial, ni individualizarse las personas o entidades a quienes se refieran.
Quedan exceptuados del secreto estadístico los siguientes datos de registro: nombre y apellido o razón social, domicilio y rama de actividad".

ART. 13: "Todas las personas que por razón de sus cargos o funciones, tomen conocimiento de datos estadísticos o censales, están obligados a guardar sobre ellos absoluta reserva."

La producción de información estadística se realiza a través de distintos métodos -censos, encuestas y registros-, que permiten la confección de estadísticas básicas en relación a diferentes áreas temáticas.

Para producir información estadística es necesario contar con recursos humanos de alta calificación científico-técnica, apoyo logístico de envergadura que incluye el desarrollo de tecnologías informáticas contribuyendo a la captura, procesamiento, obtención de resultados, difusión en tiempo.

UNIDAD 2

1 - Estadística Descriptiva e Inferencial

La Estadística tiene por objeto la recolección, presentación, análisis e interpretación de observaciones o mediciones hechas sobre un conjunto de objetos, personas, procesos, fenómenos, etc.

A través de la cuantificación y ordenamiento de los datos intenta explicar los fenómenos observados, por lo que resulta una herramienta de suma utilidad para la toma de decisiones.

Como ya vimos, la Estadística como ciencia, estudia los fenómenos de masa para hallar en ellos las regularidades del comportamiento colectivo, regularidades que sirven para describir el fenómeno y para efectuar predicciones.

Entonces, podemos decir que la Estadística es la ciencia que estudia las poblaciones o , mejor dicho, el comportamiento de las características de los elementos de una población.

Cuando se observan exhaustivamente todos los elementos de la población, se dispone de todos los datos posibles para el estudio. Con tales datos es posible *describir* exactamente las regularidades, el comportamiento o las características de la población.

La *Estadística descriptiva* es la ciencia dedicada a descubrir las regularidades o características existentes en un conjunto de datos. La estadística descriptiva *obtiene, resume y transforma* datos para poder interpretar la información.

Si la observación de la población no es exhaustiva, sino que se parte de una muestra con la finalidad de conocer mediante ella las características de la población, nos enfrentamos con un proceso de inducción, en virtud del cual se aprovecha la información suministrada por la muestra para conocer, aunque sea aproximadamente, las características de la población.

La *Inferencia estadística* tiene como función *generalizar* los resultados de la muestra para estimar las características de la población.

No obstante, el conjunto de datos muestrales puede describirse o analizarse de la misma forma que una población. Por lo tanto, el conjunto de datos u observaciones de una muestra puede utilizarse en un doble sentido: primero, para describir el propio conjunto de observaciones, y segundo, para inferir o predecir lo que ocurre en la población.

En consecuencia, la fase descriptiva es común a cualquier conjunto de observaciones o datos, ya se refieran éstos a toda la población, a una muestra o incluso a una subpoblación.

La Estadística descriptiva, es la parte más clásica, más conocida y más elemental de la ciencia estadística.

Este curso tiene como propósito la introducción en el conocimiento de las técnicas y métodos de la Estadística descriptiva.

2 - Conceptos básicos

Toda investigación estadística ha de estar referida necesariamente a un conjunto de elementos.

Veremos ahora algunas denominaciones técnicas que se utilizarán como conceptos estadísticos durante el desarrollo del curso.

► ***Unidad de análisis***: es el objeto del cual se desea obtener información. Muchas veces nos referimos a las unidades de análisis con el nombre de ***elementos***. En sentido estadístico, un elemento o unidad de análisis puede ser algo con existencia real, como un automóvil o una casa, o algo más abstracto como la temperatura o un intervalo de tiempo.

► ***Población o Universo***: es el conjunto de unidades de análisis que satisfacen una definición común y en los que interesa analizar una o varias características. Aquí el término población tiene un significado mucho más amplio que el usual, ya que puede referirse a personas, cosas, actos, áreas geográficas e incluso al tiempo.

La población debe estar perfectamente definida en el ***tiempo*** y en el ***espacio***, de modo que ante la presencia de una unidad, se pueda decidir si forma parte o no de la población bajo estudio. Por lo tanto, al definir una población, se debe cuidar que el conjunto de elementos que la integran quede perfectamente delimitado. Si estamos analizando las escuelas primarias, debemos especificar cuáles y cuándo, por ejemplo: escuelas primarias de la Capital Federal, año 1992.

El tamaño de una población viene dado por la cantidad de elementos que la componen. Simbolizaremos esta información con la letra N.

► *Muestra*: es una parte o subconjunto de las unidades de análisis de una población dada, destinado a suministrar información sobre la población. Para que este subconjunto de unidades de análisis sea una muestra de la población bajo estudio, deben reunirse ciertos requisitos en la selección de los elementos.

Las causas por la cual se seleccionan muestras son muchas. Puede ocurrir que la población que se defina tenga tamaño infinito, es decir, no podríamos determinar su tamaño, y en consecuencia no podríamos observar a todos sus elementos. Por ejemplo, si queremos realizar una investigación sobre los peces que habitan el Océano Indico, no tenemos manera de saber exactamente cuál es el tamaño de la población.

En muchas ocasiones, el costo de la observación exhaustiva puede ser muy elevado, el tiempo de recolección de la información muy extenso, o más aún, la observación de los elementos puede ser destructiva. Por ejemplo, si quisiéramos hacer un estudio de la calidad de una partida de fósforos, no podríamos probarlos a todos pues los destruiríamos.

En todos estos casos, la única manera de estudiar la población es obteniendo muestras de ella.

El tamaño de la muestra queda determinado por el número de elementos que la forman y se simboliza con la letra ***n***.

► *Variable:* **es la cualidad o cantidad medible que se estudia de las unidades de análisis y que varían de una unidad a otra. Por ejemplo: edad, ingreso de un individuo, sexo, cantidad de lluvia caída, etc.**

► *Nivel de medición:* las variables pueden ser medidas con mayor o menor grado de precisión según la escala de medida utilizada para su observación. Podemos distinguir los siguientes niveles de medición de una variable:

Nominal: sólo permite clasificar a las unidades de análisis en categorías. Por ejemplo, sexo -varón y mujer-.

Ordinal: además de clasificar a los elementos en distintas categorías, permite establecer una relación de orden de las mismas. Por ejemplo, clase social -baja, media y alta-.

Intervalar: permite clasificar, ordenar y medir la distancia entre las diferentes categorías. Por ejemplo, edad -18, 25, 14, 76 años-.

Las variables se clasifican en dos grupos de acuerdo al nivel de medición utilizado para su observación.

► *Variables cualitativas:* a las variables medidas en escala nominal u ordinal se las conoce con el nombre de variables cualitativas, ya que la característica que miden de la unidad de análisis es una cualidad.

► *Variables cuantitativas:* a las variables medidas en escala intervalar se las llama variables cuantitativas, puesto que lo que miden es una cantidad. A su vez, a éstas se las clasifica en:

Variables discretas: cuando sólo pueden asumir valores sobre los números enteros. Por ejemplo, cantidad de alumnos.

Variables continuas: cuando pueden asumir cualquier valor sobre los números reales. Por ejemplo, peso de las personas.

Entre los niveles de medición -nominal, ordinal e intervalar- existe un orden jerárquico. Cada nivel posee las propiedades del nivel inferior. Esto significa que una variable que permite ser clasificada en el nivel intervalar puede asumir categorías que la clasifiquen en cualquiera de los otros dos niveles.

Por ejemplo:

VARIABLE	CATEGORIA	NIVEL DE MEDICION
temperatura	2º, 3º, 4º	intervalar continua
temperatura	baja, mediana, alta	ordinal
temperatura	frío, calor	nominal

► *Dato u Observación:* es el valor que toma la variable para cada unidad de análisis, que se obtiene mediante algún método de captación.

En el siguiente cuadro se presentan algunos ejemplos de los conceptos vistos anteriormente:

UNIDAD DE ANALISIS	POBLACION	VARIABLE	NIVEL DE MEDICION	TIPO DE VARIABLE	DATO
Hab. de la Rep. Arg.	Población total de la Rep. Arg. en 1990	Sexo	Nominal	Cualitat.	Femenino
Idem	Idem	Ingreso	Interv.	Cuant. continua	$ 100
Hogar	Hogares de la Cap. Fed. en 1985	Tamaño del hogar	Interv.	Cuant. Discreta	4 pers.
Idem	Idem	Educación del Jefe	Ordinal	Cualitat.	Secund.

Ejercicio 2.1

Indicar cuál es la población o universo y la unidad de análisis en los siguientes casos:
a) Censo Nacional de Población y Vivienda de 1991.
b) Un estudio socioeconómico de las familias que viven en la Capital Federal.
c) Un relevamiento de equipamiento en los hospitales de todo el país.
d) Un estudio de consumo de electricidad en las industrias del Gran Buenos Aires.
e) Un estudio de preferencias deportivas en alumnos de escuelas primarias de Rosario.

Resolución del ejercicio 2.1
a) b) c)

d)

e)

Ejercicio 2.2

Defina categorías apropiadas y establezca el nivel de medición correspondiente para cada una de las siguientes variables:

a) Sexo
b) Nacionalidad
c) Estado Civil
d) Ingreso de las personas
e) Nivel de educación
f) Grado de participación política
g) Cantidad de hijos
h) Cantidad de obreros en establecimientos industriales
i) Grupo étnico de pertenencia
j) Productividad de las empresas líderes
k) Período de duración de ciertos tubos de luz.

Resolución del ejercicio 2.2
a) b) c) d) e) f)

g)

h)

i)

j)

k)

Ejercicio 2.3

Enumere 2 ejemplos de variables cuyos niveles de medición sean:
a) Nominales, pero no ordinales ni intervalares
b) Ordinales, pero no intervalares
c) Intervalares

Resolución del ejercicio 2.3
a)
b)
c)

Ejercicio 2.4

Defina una variable intervalar e indique como transformaría esa variable en ordinal y luego en nominal.

Resolución del ejercicio 2.4

UNIDAD 3

LA ESTADISTICA EN UNA INVESTIGACION

1 - Etapas

En cualquier investigación se pueden definir claramente dos etapas de trabajo en las que la estadística cumple un rol fundamental.

La primera es la de ***Planeamiento***, que comprende:

- *# Planteo del problema a estudiar*
- *# Definición de conceptos, variables y categorías*
- *# Operacionalización de los conceptos*
- *# Selección del procedimiento de recolección de datos (muestra, censo, registro administrativo)*
- *# Definición de los instrumentos de captación (cuestionarios, entrevistas, observación directa, etc.)*
- *# Preparación del plan de tabulación y codificación*
- *# Diseño de los programas de procesamiento*
- *# Pruebas experimentales*

La segunda etapa comprende la ***recolección, elaboración y sistematización***, y el ***análisis e interpretación*** de los datos.

Recolección: es la fase de la investigación en que se reúnen o recolectan los datos.

Los datos pueden ser primarios, cuando son elaborados por el propio organismo que realiza la investigación con instrumentos diseñados para ese fin específico, o secundarios, cuando han sido elaborados por otros organismos.

Elaboración y sistematización: consiste en procesar la información obtenida en la fase de recolección con el fin de presentarla adecuadamente para su estudio y comparación.

Comprende la organización y presentación de los datos.

La ***organización*** de los datos consiste en:

- *# Crítica y corrección: es la comprobación de la calidad de los datos - errores u omisiones -.*
- *# Clasificación: en general, los datos se ordenan y clasifican en función de una base geográfica, cronológica, cuantitativa o cualitativa.*
- *# Codificación: consiste en asignar un símbolo numérico específico a cada categoría de las variables en estudio. En algunos casos los datos ya se encuentran en términos numéricos, por ejemplo, edad. Cuando las variables no son numéricas, se hace necesario codificarlas.*
- *# Tabulación: en general los datos se presentan resumidos en tablas o cuadros resumen. La cantidad de información a tabular, la disponibilidad de personal y recursos financieros, y la complejidad de la tabulación son factores que se tienen en cuenta para seleccionar el procedimiento de tabulación.*

La ***presentación*** de los datos puede realizarse en forma de texto, en cuadros o gráficos.

Análisis e interpretación de la información: puede consistir en:

- el análisis simple empleando métodos básicos de la estadística descriptiva;
- la realización de predicciones con utilización de conceptos de la estadística inferencial;
- la medición de los cambios de una variable a través del tiempo;
- el estudio de relaciones entre distintas variables.

2 - Métodos de relevamiento

La forma de obtener la información original de las unidades de análisis que componen el universo por investigar puede ser efectuada a través de un registro administrativo o de una investigación estadística, y en este último caso, la información puede ser obtenida por medio de un censo o una muestra.

2.1 - Censo

Es un método estadístico en el cual la información se obtiene de la totalidad de los elementos de información que componen la población o universo por investigar.

Debe cumplir las condiciones de **universalidad** -censar a todos los elementos de la población o universo- y **simultaneidad** - realizarse en un momento determinado-. Un censo es equivalente a una fotografía de la población en un momento dado.

Dado que los censos se realizan en el marco de una investigación estadística, la información se obtiene tal como se necesita para cumplimentar sus fines estadísticos, y ésta constituye la principal diferencia que tiene este procedimiento de recolección respecto del registro administrativo.

El término censo no sólo se aplica a aquellos relevamientos que comprenden todas las unidades de todo el territorio nacional y que se realizan con una frecuencia de recolección quinquenal o decenal, como es el caso de los censos de población, económicos, agropecuarios, etc., sino también a todo relevamiento, cualquiera sea su cobertura geográfica, número de unidades de información, o frecuencia de su recolección, siempre que incluya todas las unidades que componen el universo que se investiga.

2.2 - Muestra

Es un procedimiento estadístico que permite estudiar el universo de interés con base en la información que se obtiene de una parte de los elementos que componen dicho universo.

La muestra debe ser **representativa** de la población de la que proviene.

Al igual que en el caso de los censos, mediante este procedimiento de recolección la información se obtiene tal como se necesita para los fines estadísticos.

Su uso ha ido en rápido aumento, en la medida en que las instituciones productoras de información disponen de personal capacitado para efectuar su organización, diseño y análisis, debido a su menor costo y a que en determinadas circunstancias la información resulta más exacta debido a que los errores ajenos al muestreo (errores en la recolección y en el procesamiento) pueden ser reducidos a través de una mejor capacitación de los empadronadores y la utilización de métodos de captación de información más objetivos.

La muestra no cumple con la condición de universalidad y puede o no ser simultánea.

2.3 - *Registro administrativo*

Es un procedimiento de recolección por el cual un servicio administrativo obtiene información para sus propios fines.

Esta información puede ser utilizada con fines estadísticos y se obtiene tal como está disponible para los fines administrativos. Los fines administrativos no siempre coinciden totalmente con los fines estadísticos.

Por ejemplo, para un estudio sobre determinada enfermedad se puede recurrir a los registros disponibles en hospitales, sanatorios, etc. Estos registros habrán sido diseñados para dar respuesta a ciertos requerimientos administrativos y seguramente la información que contienen no coincidirá exactamente con los requerimientos estadísticos.

Ejercicio 3.1

Una empresa de productos lácteos realizó una investigación de mercado con la finalidad de ampliar las ventas de sus distintos tipos de leche, en el ámbito de la Capital Federal. Para tal fin, fue necesario conocer el comportamiento de los consumidores con respecto a los propios productos y a los de la competencia. Se averiguó qué tipo de leche se consumía preferentemente: común, especial o en polvo. En cada caso se preguntó la marca preferida: A, B, C o D. También interesó saber el precio y la cantidad que se

vendía de las distintas marcas de leche, en sus diversos tipos.
a) Determine cuál fue el objetivo de la investigación.
b) Establezca cuál fue la población estudiada y la unidad de análisis correspondiente.
c) Determine las variables analizadas, la escala de medida en que dichas variables fueron medidas y las categorías asumidas por cada una de ellas.
d) Establezca si analizaría a la totalidad de la población o sólo a una parte de ella. Justifique su respuesta.

Resolución del ejercicio 3.1
a)
b)
c)
d)

Ejercicio 3.2

Se quiere llevar a cabo un estudio sobre preferencias de candidatos para una futura elección presidencial.
a) Qué cobertura geográfica tendría el estudio y cuál sería la población a estudiar?
b) Indique qué método de relevamiento utilizaría para recolectar la información y a través de qué medios. Justifique su respuesta.

Resolución del ejercicio 3.2
a)   b)

Ejercicio 3.3

En el Censo Nacional de Población y Vivienda de 1991 se utilizaron para el relevamiento de los datos dos tipos de Cédulas Censales. La Cédula "A" o ampliada y la Cédula "B" o básica. La Cédula "B" contenía un número menor de preguntas que coincidían con una parte de las preguntas de la Cédula "A". Las preguntas de la Cédula "B" se realizaron a todas las personas de todas las viviendas, mientras que las de la Cédula "A" se aplicaron sólo a una parte de la población.
Qué puede decir acerca del tipo de relevamiento de este Censo?

Resolución del ejercicio 3.3

UNIDAD 4

PRESENTACION DE DATOS

Existen tres formas para presentar los datos de cualquier estudio estadístico una vez que han sido organizados y procesados:

 # TEXTO

 # CUADRO

 # GRAFICO

Cada una de estas formas presenta ventajas y desventajas, de acuerdo al tipo de variables en análisis, y al objetivo que se persiga con la presentación.

Veremos ahora detalladamente cada una de estas formas de presentación de los datos.

1 - Texto

Es una combinación de cifras y texto. Esta forma de presentación permite llamar la atención sobre comparaciones de importancia y resaltar ciertas cifras, pero sólo puede utilizarse cuando los datos a presentar son pocos.

Ejemplo:

LA POBLACION MUNDIAL

Según los datos censales de 1980, la población mundial estaba constituída por 4.432.147.000 personas.
El continente con mayor población era Asia, donde vivían 2.578.610.000 personas, es decir, el 62% de los habitantes del mundo. El continente que poseía el menor número de habitantes era Oceanía, con sólo 23.820.000 personas, lo que representaba el 0,005% de la población mundial.
Si comparamos el total de habitantes de Oceanía con el de nuestro país, descubrimos que, en este caso, la población de un solo país es mayor a la de todo un continente.
La población argentina era de 27.947.446, lo que representaba el 4,6% de la población americana.

2 - Cuadros

Este tipo de presentación de la información permite volcar un gran número de datos en forma resumida haciendo fácil y clara su lectura. Es más breve, puesto que los encabezados de las columnas y los títulos de las filas evitan repetir explicaciones y fundamentalmente facilita las comparaciones de los datos.

Generalmente un cuadro completo debe incluir las siguientes partes:

► TITULO: debe enunciar qué datos están incluídos en el cuerpo del cuadro, el lugar y el momento de referencia.

► ENCABEZADO: en general el encabezado del cuadro explicita las variables seleccionadas y las categorías correspondientes.
En los cuadros de doble entrada, usualmente se ubica en la parte superior la variable "independiente" y a la izquierda la variable "dependiente". Esta última es la variable que se trata de explicar y se llama variable dependiente porque se supone que los valores que toma dependen de los valores que presentan otras variables, llamadas independientes o explicativas.

► CUERPO DEL CUADRO: contiene los datos clasificados de acuerdo a las variables del encabezado.

► NOTA DE PIE: son utilizadas para realizar aclaraciones sobre partes incluídas en el cuadro, por ejemplo: "cifras redondeadas", "no se posee dato", etc.

► FUENTE: da cuenta de la fuente de la cual fueron extraídos los datos, en el caso de ser secundarios.

Recomendaciones generales para la construcción de cuadros:

* *No deben ser largos. Si hay varios hechos a presentar será preferible construir varios cuadros más sencillos.*

* *Ordenar las categorías de cada variable de una manera lógica para facilitar su interpretación y análisis. Las categorías pueden ser ordenadas:*

  - *geográficamente, en orden alfabético o de acuerdo a la importancia de ciertas áreas.*
  - *cronológicamente, del más antiguo al más reciente.*
  - *cuantitativamente, en forma ascendente o descendente.*
  - *cualitativamente, por orden de importancia.*

Se debe utilizar el mismo ordenamiento lógico para todas las variables del cuadro. Por ejemplo en un cuadro de dos variables cuantitativas, usar el orden ascendente para las dos.

Veamos como ejemplo el siguiente cuadro:

POBLACION ARGENTINA
CLASIFICADA POR JURISDICCION SEGUN CENSOS
En miles de habitantes

Jurisdicción	Censos						
	1869	1895	1914	1947	1960	1970	1980
Capital Federal	187	663	1577	2983	2967	2972	2923
Buenos Aires	308	922	2066	4272	6766	8775	10865
Catamarca	80	90	101	147	168	172	208
Córdoba	211	351	735	1498	1754	2060	2408
Corrientes	129	240	347	525	533	564	661
Chaco	-	10	46	431	543	567	701
Chubut	-	4	23	92	142	190	263
Entre Ríos	134	292	425	787	805	812	908
Formosa	-	5	19	114	179	234	296
Jujuy	40	50	78	167	241	302	410
La Pampa	-	26	101	169	159	172	208
La Rioja	49	70	80	111	128	136	164
Mendoza	65	116	278	588	824	973	1196
Misiones	-	33	54	246	361	443	589
Neuquén	-	15	29	87	110	155	244
Río Negro	-	9	42	134	193	263	383
Salta	89	118	142	291	413	510	663
San Juan	60	84	119	261	352	384	466
San Luis	53	81	116	166	174	183	214
Santa Cruz	-	1	10	43	53	84	115
Santa Fé	89	397	900	1703	1885	2136	2466
S. del Estero	133	162	262	479	477	495	595
Tucumán	109	216	333	593	774	766	973
T. del Fuego	-	-	3	5	8	16	27
Total	1737	3955	7885	15894	20011	23364	27947

NOTA: "-" no existía la jurisdicción
FUENTE: INDEC

3 - Gráficos

La representación gráfica de los datos contenidos en un estudio estadístico tiene como finalidad ofrecer una visión de conjunto del fenómeno sometido a investigación, más rápidamente perceptible que la observación directa de los datos numéricos.

De aquí que las representaciones gráficas sean un medio eficaz para el análisis de las estadísticas, ya que las magnitudes y las regularidades se aprecian y recuerdan con más facilidad cuando se examinan gráficamente.

Hay que advertir, sin embargo, que la representación gráfica no es más que un medio auxiliar de la investigación estadística, pues ésta es fundamentalmente numérica.

Las representaciones gráficas pueden hacerse utilizando un sistema geométrico de representación, en cuyo caso gozan de rigurosidad y precisión, o bien pueden utilizarse símbolos alusivos al tema en estudio (por ejemplo, casas, árboles, figuras humanas, etc).

Mediante este último sistema de representación no se persigue una rigurosa exactitud, sino lograr efectos impresionistas en quien está leyendo la información.

Aunque un cuadro estadístico encierra o proporciona toda la información disponible, es útil traducir esta información mediante un gráfico para realizar una síntesis visual. La función de un gráfico es la presentación ilustrativa de los datos, para ayudar a ver la ***forma y figura*** de una situación compleja y efectuar comparaciones.

Existe una gran variedad de gráficos. Su elección depende de la variable en estudio y de las características que se quieran resaltar.

Para la construcción de gráficos no hay reglas únicas.

Siempre se debe tener presente que un gráfico permite una interpretación más rápida de la información, pero esta información es menos precisa que la que puede proporcionar un cuadro.

Un gráfico completo debe reunir las siguientes partes:

- TITULO: debe describir correctamente el contenido de la gráfica.

- DIAGRAMA: al igual que el cuerpo de un cuadro representa los datos mostrados en la gráfica.

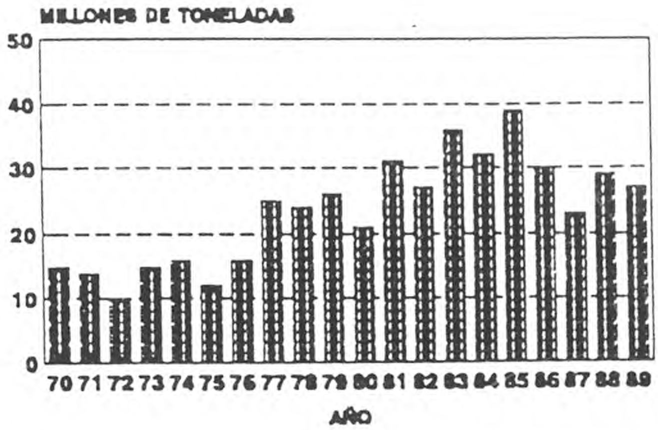
- VARIABLE: en todo gráfico debe indicarse la variable que se está representando.

▸ *ESCALA:* todo gráfico debe contener una escala de medida que dé cuenta de las magnitudes que se están representando.

▸ *FUENTE:* debe ser colocada al pie de la gráfica e indicar cuál fue el origen de los datos representados a través de la gráfica.

Veamos con un ejemplo, las distintas partes que componen un gráfico:

EXPORTACIONES ARGENTINAS
1970-1989



INDEC

De acuerdo a la clasificación de los datos -cronológica, geográfica, cualitativa, cuantitativa, etc- se usan distintos tipos de gráficos:

 # GRAFICO DE LINEA

 # GRAFICO DE BARRA

 # GRAFICO DE SECTORES

 # PICTOGRAMAS

 # MAPAS ESTADISTICOS

A continuación veremos las características de cada uno de estos tipos de gráficos.

3.1 - Gráfico de línea

Es un gráfico donde se muestra la variación de una variable a través del tiempo, haciendo énfasis en los cambios y tendencias.

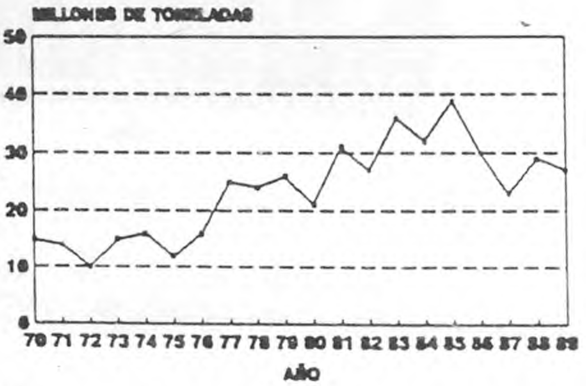
En el eje de abscisas se representa la variable temporal y en el de ordenadas la variable en estudio. El gráfico es una línea quebrada o "poligonal", donde los puntos de quiebre indican los datos correspondientes a los distintos períodos de tiempo.

En un mismo gráfico pueden representarse diferentes líneas, lo que permite establecer comparaciones.

Los gráficos de línea son útiles cualquiera sea la escala de medida utilizada en la medición de la variable.

Veamos un ejemplo:

EXPORTACIONES ARGENTINAS
1970-1989



INDEC

3.2 - Gráfico de barras

Es un gráfico compuesto por un grupo de barras rectangulares donde cada barra representa un valor de la variable. Se utiliza para enfatizar las diferencias entre valores.

Para su construcción hay que tener en cuenta las siguientes pautas:

- *todas las barras deben tener el mismo ancho.*
- *la altura de cada barra muestra los datos representados.*
- *las barras no deben ser excesivamente cortas y anchas, ni largas y angostas.*
- *el espacio entre barra y barra debe ser menor que el ancho de la barra.*
- *se puede utilizar para datos cronológicos, cuantitativos, geográficos y cualitativos.*

* *Cuando los datos son geográficos o cualitativos se usan barras horizontales.*
* *Cuando los datos son cronológicos o cuantitativos se usan barras verticales.*

Los gráficos de barras son efectivos cuando es poca la cantidad de datos a representar. Si la variable que se está representando toma muchos valores distintos es aconsejable usar gráficos de línea.

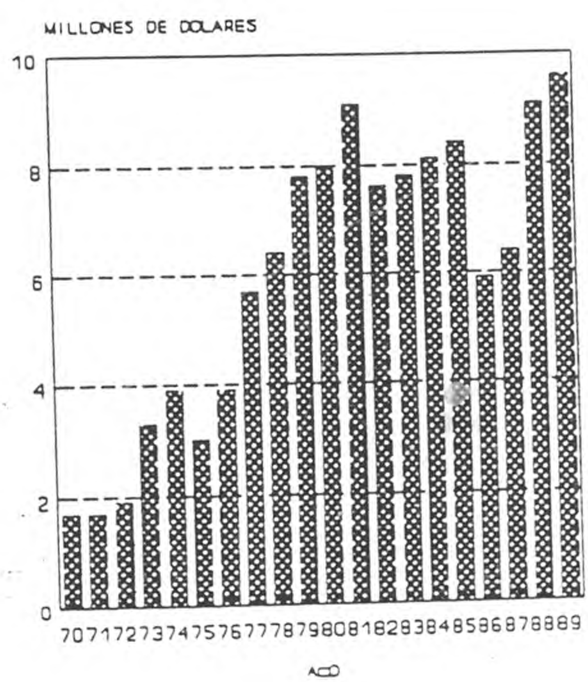
Además de cumplir con las pautas enumeradas anteriormente, se pueden distinguir distintos tipos de gráfico de barras que persiguen una mejor representación del fenómeno que se está analizando. Generalmente los gráficos de barras se utilizan para la representación de variables nominales.

Veamos a continuación estos gráficos:

A) *Barras simples o bastones:* están construídos por barras del mismo ancho, en los cuales la longitud muestra los datos representados.

Ejemplo:

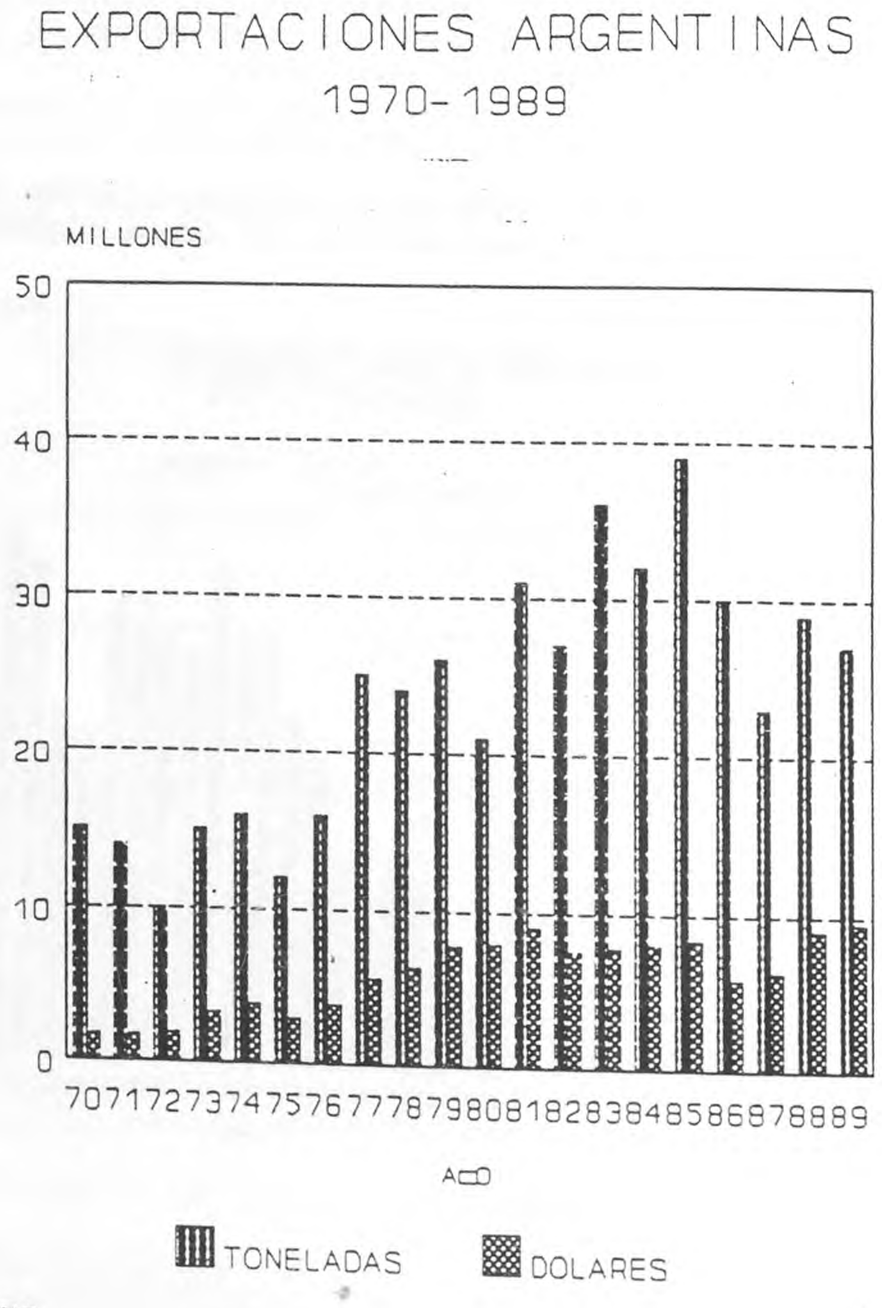
EXPORTACIONES ARGENTINAS
1970-1989



INDEC

B) ***Barras compuestas o adyacentes:*** *se utilizan para representar datos clasificados por más de una variable. Este tipo de representación permite hacer comparaciones para los niveles de clasificación de una de las variables y dentro de las categorías de ésta, permite comparar la segunda variable en estudio.*

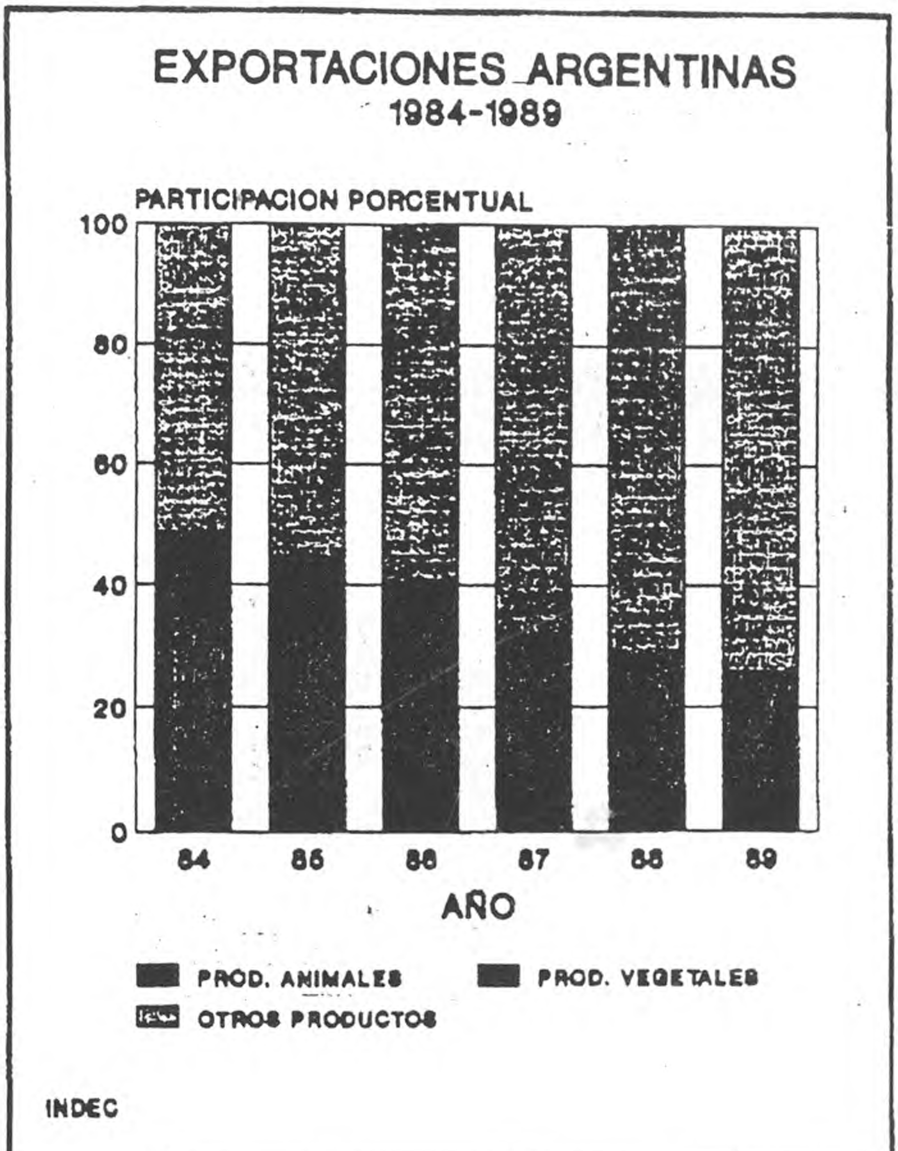
Ejemplo:



INDEC

C) *Barras subdivididas:* al igual que las gráficas compuestas, permite representar más de una variable. Usualmente se los utiliza para mostrar la importancia relativa de los valores de una de las variables dentro de una categoría de la otra variable.

Ejemplo:



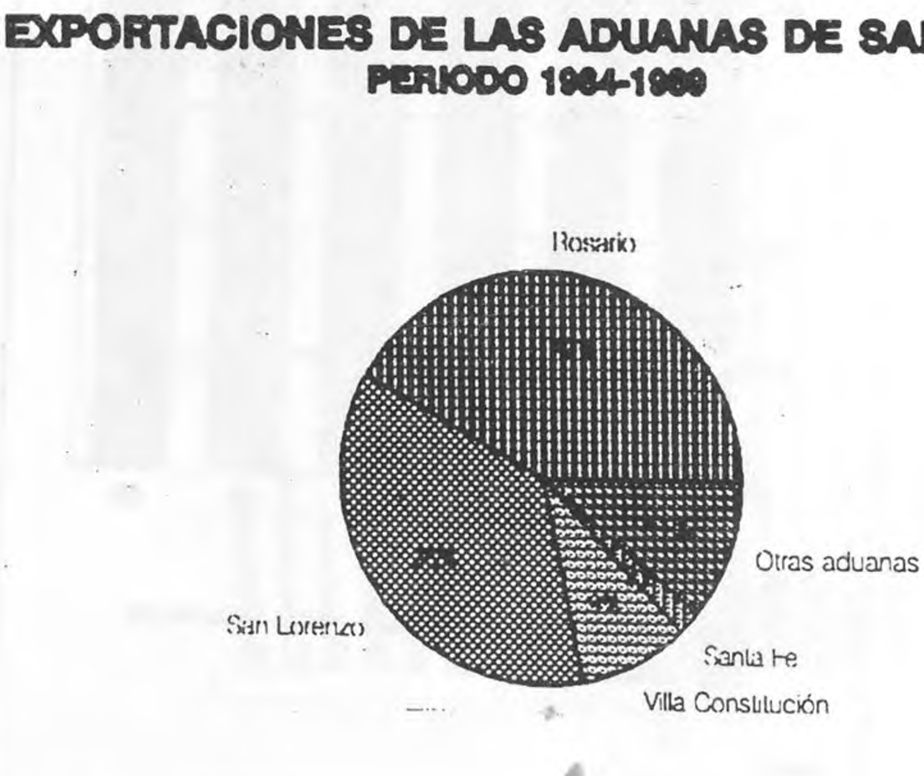
3.3. - *Gráfico de sectores*

Sobre un círculo que representa a la población total o universo, se dividen sectores que representan la participación relativa de un valor de la variable sobre el total de la población.

La amplitud de los sectores se obtiene por una regla de 3 simple, teniendo en cuenta que el ángulo central del círculo completo mide 360 grados.

Por lo general se utilizan para representar variables medidas en escala nominal u ordinal.

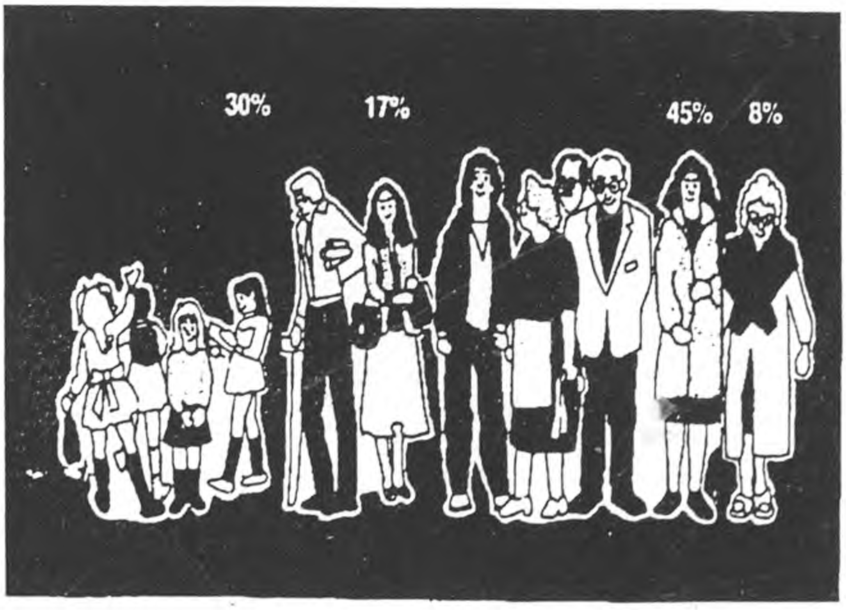
Ejemplo:



3.4 - Pictograma

En este tipo de gráficos se representan los datos mediante símbolos figurativos de la variable en análisis. Su aplicación está condicionada al hecho de que los valores de la variable deben ser susceptibles de adecuarse a formas gráfico-simbólicas.

Ejemplo:

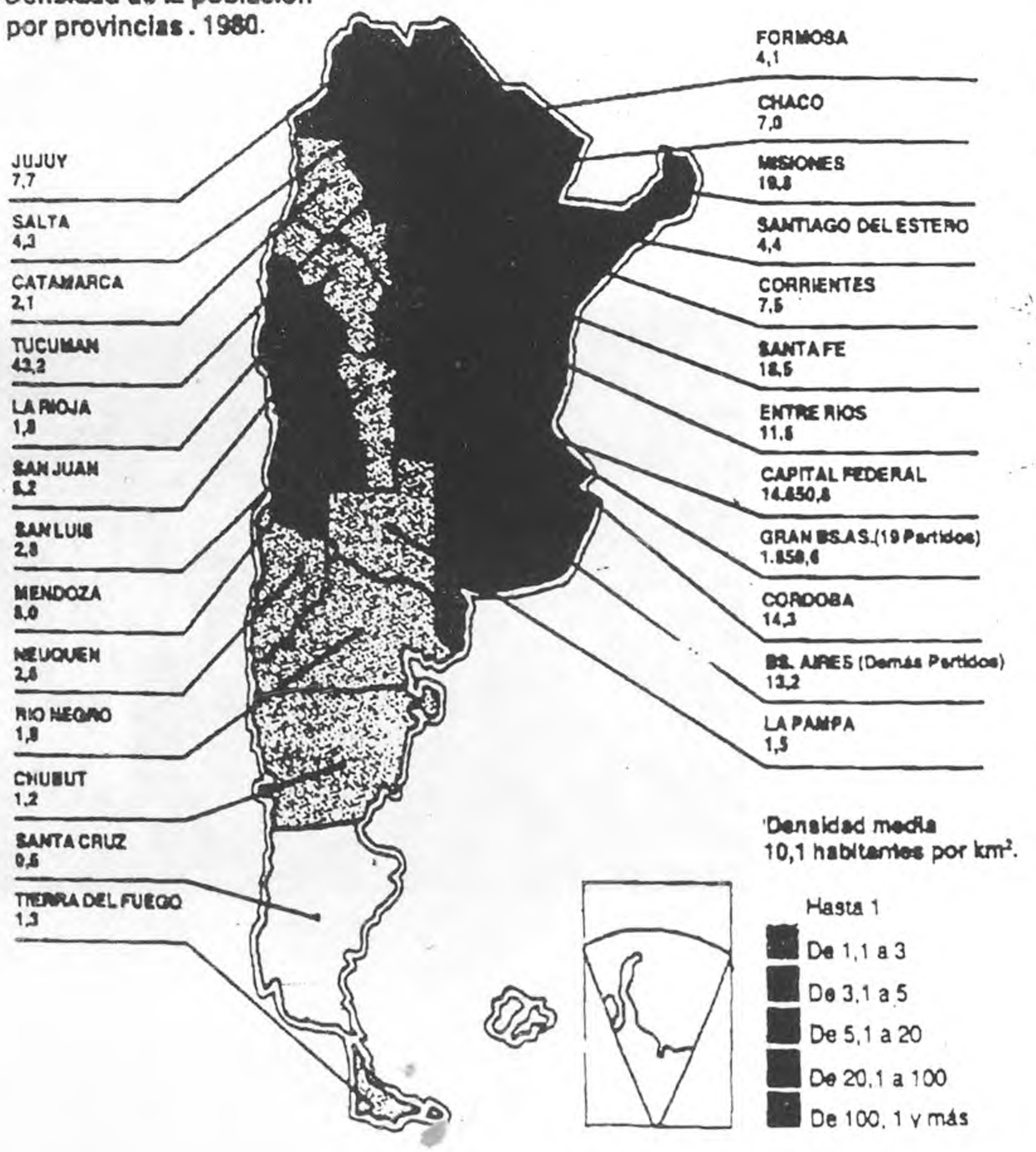


3.5 - *Mapa estadístico*

Es un artificio gráfico para mostrar datos o información cuantitativa sobre una base geográfica. Permite representar simultáneamente variables cuantitativas con su correspondiente distribución geográfica.

Ejemplo:

Densidad de la población por provincias. 1980.



Ejercicio 4.1

Entre las variables que se indagaron en el Censo de Población y Vivienda de 1991 podemos enumerar las siguientes:
- provincia
- tipo de vivienda (casa, rancho o casilla, departamento, casa de inquilinato, pensión u hotel, local no construído para habitación, vivienda móvil).
- cantidad de cuartos de uso exclusivo del hogar (ninguno, uno, dos, tres, cuatro, cinco, seis, siete, ocho o más).
- edad en años cumplidos de todas las personas.
- sexo.
- lugar de nacimiento (en esta provincia, en otra provincia, en un país limítrofe, en otro país, ignorado).
- asistencia a algún establecimiento educacional (asiste a establecimiento público, asiste a establecimiento privado, no asiste pero asistió, nunca asistió, ignorado).
- tenencia de obra social (sí, no, ignorado).
- sabe leer y escribir (sí, no, ignorado).
- durante la semana pasada, trabajó aunque sea por pocas horas? (sí, no, ignorado).
- a qué se dedica o qué produce el lugar o establecimiento donde trabaja?

Elija algunas de las variables presentadas y realice:

a) Una presentación de datos en forma de texto.
b) Una presentación de datos en forma de cuadro.
c) Un gráfico de barras y otro de sectores.

Resolución del ejercicio 4.1

a)

b)

c)

UNIDAD 5

TRATAMIENTO DE VARIABLES CUALITATIVAS

Cuando se trabaja con datos medidos en escalas nominales u ordinales *-variables cualitativas-*, el análisis de la información es mucho más sencillo que cuando los datos están medidos en escala intervalar. Como ya vimos, la medición en escala intervalar es más potente que las otras escalas, ya que permite establecer distancias entre las categorías de medición.

La primera operación a realizar con variables cualitativas es contabilizar el número de casos que pertenecen a cada una de las categorías de la variable. Un grupo de 60 personas clasificado por sexo, puede constar de 36 varones y 24 mujeres. Si se quiere comparar este grupo con otros grupos en relación a su distribución por sexo; es necesario tener en cuenta el número total de personas que conforman cada grupo.

Los indicadores descriptivos que veremos en esta unidad permiten efectuar comparaciones entre diversos grupos, basándose esencialmente en el tamaño de los mismos.

Estos indicadores son aplicables a variables medidas en cualquier escala, pero son de fundamental utilidad cuando las mismas son medidas en escalas nominales u ordinales, ya que son la única herramienta para el tratamiento estadístico de las mismas.

En esta unidad, veremos el cálculo de tres medidas descriptivas para el tratamiento de variables cualitativas. Ellas son:

 # PROPORCIONES

 # PORCENTAJES

 # RAZONES

1 - Proporciones

Para el cálculo de una proporción, la clasificación de los datos en las distintas categorías de la variable debe ser mutuamente excluyente y exhaustivo. Con esto, estamos pidiendo que cada elemento clasificado debe pertenecer a una y sólo una de las categorías, y que todos los elementos deben poder ser clasificados en alguna de las categorías previstas.

La proporción de casos de una categoría se define como el número de casos en esa categoría dividido por el número total de casos.

El valor de una proporción no puede ser mayor que 1, ya que el numerador a lo sumo puede ser igual al denominador, y si calculamos las proporciones para cada una de las categorías, las sumas de dichas proporciones será igual a 1.

En efecto, si contamos con una población de N elementos, donde los mismos están clasificados en 3 categorías mutuamente excluyentes y exhaustivas $-N_1$, N_2 *y* N_3-, *y calculamos la proporción de elementos que pertenecen a cada una de las categorías, tenemos:*

Proporción de elementos que pertenecen a N_1:

$$p_1=\frac{N_1}{N}$$

de igual forma será:

$$p_2=\frac{N_2}{N}$$

$$p_3=\frac{N_3}{N}$$

donde p_1, p_2 *y* p_3 *son menores o iguales a 1 y la sumatoria de las proporciones correspondientes a todas las categorías es igual a 1, es decir:*

$$\sum_{1}^{3} p_i=p_1+p_2+p_3=\frac{N_1}{N}+\frac{N_2}{N}+\frac{N_3}{N}=\frac{N_1+N_2+N_3}{N}=\frac{N}{N}=1$$

Ilustremos el empleo de las proporciones mediante el siguiente ejemplo:

CUADRO: Cantidad de alumnos aprobados y no aprobados en 2 cursos de informática.

	CURSO 1	CURSO 2
APROBADOS	24	34
Nota >= 7	15	10
Nota < 7	9	24
NO APROBADOS	8	11
TOTAL	32	45

Dado que los dos cursos son de distinto tamaño, resulta difícil extraer conclusiones acerca de la comparación del número de aprobados o no aprobados en cada uno de los cursos.

Si expresamos los datos en términos de proporciones, podemos establecer una comparación directa, logrando independencia sobre el tamaño total de cada uno de los cursos.

CUADRO: Proporción de alumnos aprobados y no aprobados en 2 cursos de informática.

	CURSO 1	CURSO 2
APROBADOS	0.75	0.76
Nota >= 7	0.63	0.29
Nota < 7	0.37	0.71
NO APROBADOS	0.25	0.24
TOTAL	1	1

Este cuadro nos permite apreciar que el número relativo de personas aprobadas es casi igual en cada uno de los cursos, pero que en el CURSO 1 los aprobados con nota superior a 7 puntos son, en términos relativos, muchos más que en el CURSO 2.

En el ejemplo analizado se trabajó con proporciones calculadas sobre 2 conjuntos diferentes. Como puede observarse, se calcularon proporciones sobre el total de personas en cada uno de los cursos, relativas a si habían aprobado o no aprobado el curso , y además sobre los aprobados se calculó la proporción que aprobó con nota superior o inferior a 7 puntos.

Por lo tanto, constituye un buen principio, acostumbrarse a determinar siempre las categorías que se hallan comprendidas en el número total de casos que se usará como denominador de la proporción.

Al calcular una proporción siempre deberá preguntarse:

ESTO ES LA PROPORCION DE QUE?

Recuerde siempre que una proporción es un cociente entre una "parte" y el "todo" de donde proviene esa parte.

2 - Porcentajes

Los porcentajes se obtienen de las proporciones multiplicando a éstas por 100. La palabra porcentaje significa "por ciento".

Al hablar de porcentajes, estamos normalizando la información en relación con el volumen total, calculando el número de elementos que pertenecen a una categoría determinada si el total de casos fuera 100, permaneciendo inalterada la proporción en cada categoría.

Así como la suma de las proporciones dan la unidad, es obvio que la suma de los porcentajes será 100.

Si retomamos el ejemplo anterior, calculando los porcentajes tenemos:

CUADRO: Porcentaje de alumnos aprobados y no aprobados en 2 cursos de informática.

	CURSO 1	CURSO 2
APROBADOS	75	76
Nota >= 7	63	29
Nota < 7	37	71
NO APROBADOS	25	24
TOTAL	100	100

En el ejemplo, los porcentajes se calcularon en sentido vertical, es decir, la suma 100 se obtiene sumando a través de las categorías "aprobados" y "no aprobados".

Veamos cómo se leen los porcentajes del cuadro. Del total de alumnos del CURSO 1, vemos que el 75% de ellos aprobó el curso de *informática, mientras que el 25% restante no lo hizo. A su vez, de los alumnos aprobados, un 63% lo hizo con nota superior a 7 y el 37% con nota inferior a 7.*

De igual forma se podrían calcular porcentajes con suma 100 en sentido horizontal, es decir sumando a través de los cursos.

Un estudio alternativo puede consistir en considerar categorías conjuntas para ambas variables -curso y aprobación- donde el 100% se obtiene para la suma de las personas, cualquiera sea el curso y su resultado.

Veamos cómo quedaría el cuadro en este caso:

CUADRO: Porcentaje de alumnos aprobados y no aprobados en 2 cursos de informática.

	CURSO 1	CURSO 2	TOTAL
APROBADOS	31.2	44.2	75.4
Nota >= 7	19.5	13.0	
Nota < 7	11.7	31.2	
NO APROBADOS	10.4	14.2	24.6
TOTAL	41.6	58.4	100

Observando los porcentajes del cuadro, vemos que, del total de alumnos, un 31,2% son aprobados y del CURSO 1; a su vez de éstos el 19,5% aprobó con nota superior a 7 y el 11,7% lo hizo con nota inferior a 7. También surge, por ejemplo, que el 58,4% de los alumnos pertenecen al CURSO 2, o que el 75,4% aprobó el curso, etc.

3 - Razones

El término razón se emplea cuando los casos que estamos contabilizando representan categorías no contenidas una en la otra.

La razón de un número A con respecto a un número B se define como A dividido B. La cantidad que precede se pone en el numerador y la que sigue en el denominador.

Si tomamos el ejemplo de los cursos, podemos calcular la razón de los "aprobados" a los "no aprobados", siendo ésta:

$$R=\frac{58}{19}=3.05$$

Observe que a diferencia de la proporción, la razón es un número que puede ser mayor que 1.

Es conveniente expresar la razón en términos de un denominador que sea la unidad con el fin de facilitar la interpretación. Por ejemplo, la razón de los "aprobados" a los "no aprobados" la expresamos como:

$$R=\frac{3.05}{1}$$

La lectura de esta razón es la siguiente: "por cada 3.05 personas aprobadas hay 1 no aprobado".

Las proporciones representan un caso particular de las razones, en la que el denominador es el número total de los casos y el numerador una fracción del total. En una proporción, siempre el numerador es una cantidad que está contenida en el denominador.

Refiriéndonos al ejemplo, podemos calcular la razón de alumnos en el CURSO 1 respecto al CURSO 2, siendo:

$$R=\frac{32}{45}=0.71$$

Podemos decir que por cada 0.71 alumnos en el CURSO 1 hay 1 alumno en el CURSO 2.

Generalmente, cuando la razón es un número menor que 1, se lleva a base 100, multiplicando el numerador y el denominador de la razón por 100, quedando:

$$R=\frac{71}{100}$$

Vemos que la relación no cambia, pero las cifras obtenidas guardan relación con las unidades y podemos decir, ya que hablamos de cantidad de elementos, que hay 71 alumnos en el CURSO 1 por cada 100 alumnos en el CURSO 2.

Ejercicio 5.1:

Con el fin de conocer la opinión de las mujeres respecto del divorcio, se realizó una encuesta en la Capital Federal. Los resultados de la misma fueron los siguientes:

ACUERDO CON EL DIVORCIO	ESTADO CIVIL						
	SOLTERA		CASADA		SEPARADA		TOTAL
	C/HIJ	S/HIJ	C/HIJ	S/HIJ	C/HIJ	S/HIJ	
SI	373	62	129	576	89	99	1328
NO	255	10	77	318	4	8	672
TOTAL	628	72	206	894	93	107	2000

Responda las siguientes preguntas a partir de los datos del cuadro:

a) Qué proporción de mujeres está a favor del divorcio?
b) Entre las mujeres casadas, qué proporción está de acuerdo con el divorcio?
c) Del total de mujeres que están a favor del divorcio, qué proporción son solteras?
d) Del total de mujeres con hijos, qué proporción está en contra del divorcio?
e) Cuál es el porcentaje de mujeres solteras con hijos?
f) Cuál es la razón de las mujeres solteras en relación a las casadas?
g) Cuál es la razón de las que están a favor del divorcio respecto a las que no lo están?
h) De las mujeres casadas a favor del divorcio, cuál es la razón de las que tienen hijos respecto a las que no tienen?

Resolución del ejercicio 5.1:

a)

b)

c)

d)

e)

f)

g)

h)

Ejercicio 5.2:

Un sociólogo ha estudiado la participación sindical entre el personal de una empresa obteniendo los siguientes datos:

NUMERO DE EMPLEADO	SEXO	CATEGORIA OCUPACIONAL	PARTICIPACION SINDICAL
1	F	empleado	no
2	F	empleado	no
3	M	promotor	si
4	M	jefe	si
5	F	jefe	no
6	F	empleado	si
7	F	empleado	si
8	M	promotor	si
9	M	promotor	si
10	M	promotor	si
11	F	jefe	no
12	F	promotor	si
13	M	empleado	si
14	F	empleado	no
15	M	jefe	no
16	M	promotor	no
17	M	empleado	si
18	M	promotor	si
19	F	empleado	no
20	F	empleado	no

21	F	empleado	no
22	M	jefe	no
23	F	promotor	si
24	M	jefe	no
25	M	empleado	si
26	M	promotor	no
27	M	empleado	no
28	F	promotor	no
29	F	empleado	no
30	F	empleado	si

a) Confeccione 2 cuadros, uno clasificado por sexo y participación sindical, y el otro clasificado por categoría ocupacional y participación sindical.
b) De acuerdo al resumen de la información, existe alguna relación evidente entre la participación sindical y el sexo?
c) Entre las personas con participación sindical, cuál es la razón de los empleados respecto a los jefes?
d) Se puede afirmar que el grado de participación sindical tiene alguna relación con la categoría ocupacional?

Resolución del ejercicio 5.2:
a)

b)

c)

d)

UNIDAD 6

TRATAMIENTO DE VARIABLES CUANTITATIVAS

En esta unidad nos ocuparemos de métodos para el resumen de datos medidos en escala de intervalo -*variables cuantitativas*-. Las técnicas que desarrollaremos tienden a reducir la información relativa a un número muy grande de casos a una forma muy simple, que permite representar de qué manera están distribuídos los casos.

En la unidad anterior nos hemos encontrado con algunas medidas para resumir los datos. Se trataba de variables cualitativas, donde las categorías estaban determinadas y lo único que había que hacer era contabilizar el número de casos pertenecientes a cada categoría y normalizar en relación al número total de casos, calculando una proporción, un porcentaje o una razón.

En cambio, si los datos medidos en escala de intervalo -*variables cuantitativas*- han de resumirse de igual modo, hay que tener en cuenta si la variable es *discreta* o *continua*.

Dicho resumen, consiste en organizar tablas que sintetizan los datos originales o valores observados. Estas tablas se denominan *distribuciones de frecuencia*.

Una distribución de frecuencias es una tabla que presenta en forma ordenada los distintos valores de una variable y sus correspondientes frecuencias.

Definimos como *frecuencia* al número de veces que se presenta cada valor de la variable.

1 - Tablas de frecuencias simples

Mediante un ejemplo veremos la presentación de una tabla de frecuencias simples para una variable discreta.

Consideremos la variable "número de cuartos por hogar" según datos de la E.P.H. para la ciudad de Formosa.

NUMERO DE CUARTOS POR HOGAR (1)	FRECUENCIA ABSOLUTA (2)	FRECUENCIA RELATIVA (3)	FRECUENCIA ACUMULADA (4)	FRECUENCIA RELATIVA ACUMULADA (5)
1	144	0,214	144	0,214
2	225	0,334	369	0,548
3	174	0,259	543	0,807
4	88	0,131	631	0,938
5	42	0,062	673	1,000
	673	1,000	-	-

En la columna (1) se observan los valores que toma la variable "número de cuartos por hogar", cuyo campo de variabilidad o recorrido es de 1 a 5.

En la columna (2) se ha colocado la cantidad de hogares u observaciones correspondientes a cada valor de la variable, es decir la *frecuencia absoluta* que presenta cada valor de la misma. Si sumamos esta columna obtenemos la cantidad total de hogares bajo estudio.

Luego, en la columna (3) calculamos el cociente de cada uno de los valores de la columna (2) respecto al total de hogares. Llamamos a estos valores *frecuencias relativas*. Las frecuencias relativas no son más que proporciones, ya que representan la importancia relativa de cada valor de la variable en el total de casos.

En la columna (4) sumamos los hogares acumulados hasta cada uno de los valores de la variable. Por ejemplo: si queremos saber cuántos hogares hay que tienen como máximo 2 cuartos, vemos que se acumulan 369, o sea, 144 hogares con un cuarto y 225 con 2 cuartos. Llamamos a estos valores *frecuencias acumuladas*.

Finalmente, en la columna (5) efectuamos el cociente entre los valores de la columna (4) dividido por el total de hogares, lo que nos indica el peso relativo de los casos acumulados hasta cada uno de los valores de la variable, y llamamos a esta columna *frecuencia relativa acumulada*.

1.1 - Simbología

Para generalizar el ejemplo a cualquier estudio, llamaremos:

Xi: a los valores que asume la variable, donde "i" es un subíndice que indica el número de orden de los distintos valores de la variable ordenados en forma ascendente. Si i = 1 indica que es el primer valor que toma la variable; si i = 10 indica que es el décimo valor de la variable (tengamos siempre en cuenta que no es el orden en que se efectuó la observación, sino el número que le corresponde después de ordenar los distintos valores de la variable en forma ascendente).

El subíndice "i" varía de "*1 a n*", siendo "n" el último valor distinto que toma la variable.

fai: frecuencia absoluta para el i-ésimo valor de la variable. Por ejemplo fa3 = 174 indica que el tercer valor de la variable tiene una frecuencia absoluta igual a 174. En otras palabras, hay 174 elementos de la población que toman el valor x3 de la variable. La suma de todas las "fai" reproduce el número total de elementos del conjunto estudiado.

$$\sum_{i=1}^{n} fa_i = n$$

fri: **frecuencia relativa correspondiente al i-ésimo valor de la variable. La suma de todas las "fri" dá como resultado 1.**

$$\sum_{i=1}^{n} fr_i = 1$$

Fai: **frecuencia absoluta acumulada hasta el i-ésimo valor de la variable. Por ejemplo Fa3 = 543 indica que hasta el tercer valor de la variable se acumularon 543 casos. El valor de Fa3 resulta de sumar las "fai" anteriores o iguales a i=3, es decir fa1 + fa2 + fa3.**

$$Fa_3 = \sum_{i=1}^{3} fa_i = 543$$

La frecuencia absoluta acumulada correspondiente al último valor de la variable es igual al número total de casos.

Fri: frecuencia relativa acumulada hasta el i-ésimo valor de la variable. La Fri correspondiente al último valor de la variable es siempre igual a 1.

NOTA: las frecuencias relativas "fri" y relativas acumuladas "Fri" suelen expresarse en porcentajes. Dichos valores son los correspondientes a estas frecuencias multiplicados por 100.

1.2 - Representación gráfica

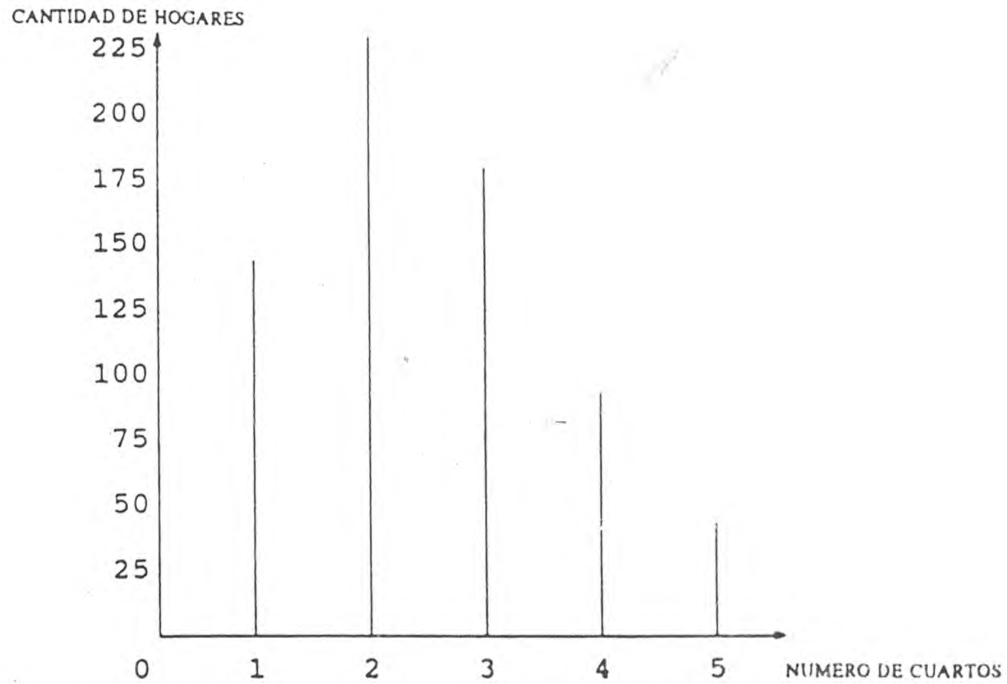
En general, la representación gráfica de una tabla de frecuencias permite percibir con mayor claridad algunas características de la masa de datos que se investiga. Por ello, resulta bastante más fácil transmitir conclusiones a personas no habituadas a la interpretación de distribuciones de frecuencias cuando se utilizan gráficos.

Para representar gráficamente una distribución de frecuencias se utiliza un par de ejes de coordenadas. En el eje de las abscisas se representará la variable estudiada y en el eje de ordenadas, las correspondientes frecuencias - absolutas o relativas-.

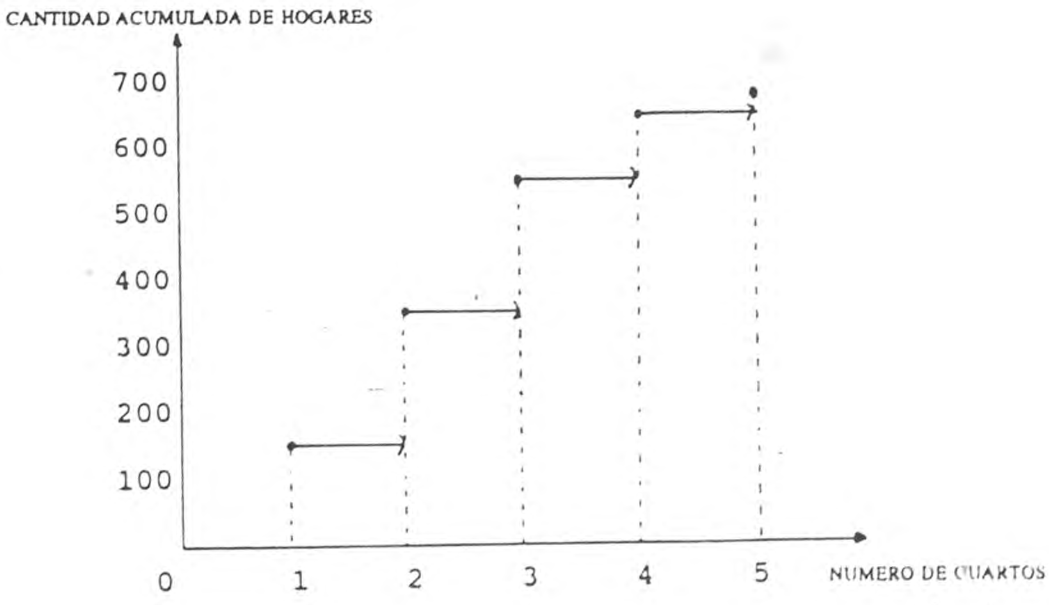
Cuando se grafican datos provenientes de una variable discreta, como la frecuencia corresponde a cada valor de la variable, las mismas se representan por un bastón vertical, construyéndose lo que habitualmente llamamos un *Gráfico de Bastones*.

El gráfico de bastones es la representación gráfica de las frecuencias de una variable discreta, cuyas abscisas son los valores de la variable y cuyas ordenadas son las frecuencias absolutas o relativas.

El siguiente es un gráfico de frecuencias absolutas confeccionado con las datos del ejemplo anterior.



Para graficar las frecuencias acumuladas de una variable discreta lo hacemos mediante segmentos paralelos al eje de abscisas. Cada segmento se extiende entre dos valores consecutivos de la variable, siendo las respectivas ordenadas las frecuencias acumuladas correspondientes al valor de la variable que es abscisa del punto inicial del segmento. A este gráfico se lo denomina ***Gráfico Escalonado***.



NOTA: los mismos tipos de gráficos se utilizan para representar las frecuencias relativas. La diferencia es un cambio de escala en las ordenadas.

2 - *Tablas de frecuencias para valores agrupados*

En el caso de tener muchos datos de una variable discreta, o de tratarse de una variable continua, será necesario fijar *intervalos de clase* para llegar a un resumen efectivo de la información original. Esta información se presenta en una tabla de frecuencias para datos agrupados.

Definimos como intervalos de clase a las subdivisiones o intervalos en que se ha dividido al dominio o campo de variabilidad de la variable, de modo tal que cada intervalo estará compuesto por tramos del recorrido de la variable.

Por ejemplo, si estudiamos la distribución por edad de la población de un país estamos en presencia de una variable que toma muchos valores distintos. Estos valores se pueden agrupar en intervalos de clase tomando tramos de edades que cubran todo el recorrido de la variable. Se podrían definir los siguientes intervalos de clase: 0 a 10 años, 11 a 20 años, 21 a 30 años, etc.

Llamamos *límites de clase* a los valores que definen los extremos de un intervalo. Por ejemplo, el intervalo 0 a 10 años, tiene como límites a los valores 0 y 10.

Por lo tanto, tendremos para cada intervalo un límite inferior y un límite superior que los simbolizaremos Li y Ls respectivamente.

La *amplitud* del intervalo vendrá dada por la diferencia entre el límite superior y el límite inferior. A la amplitud la llamamos "H", siendo:

$$H=L_s-L_i$$

Cada intervalo tendrá también, lo que llamamos *marca de clase*, que es el punto medio o valor central del intervalo. Lo simbolizamos con Xi'.

$$Xi'=\frac{L_i+L_s}{2}$$

Cuando los datos se agrupan en intervalos, el problema fundamental es pensar en una amplitud adecuada para los mismos. Generalmente se *aconseja una cantidad de entre 10 y 15 de intervalos de clase, de modo que no haya tantos como para que no sea manejable la tabla, ni tan pocos como para que la amplitud sea tan grande que haga perder información.*

Para calcular la amplitud del intervalo se busca primero la amplitud o rango de la variable, es decir la diferencia entre el mayor y el menor de los valores que toma la variable, y luego, el resultado se divide por la cantidad de intervalos que se quieren formar.

$$R = máx(Xi) - mín(Xi)$$

$$H = \frac{R}{cantidad.de.intervalos}$$

Donde **H** es la amplitud del intervalo de clase.

Puede ocurrir que se necesite la información agrupada en intervalos con una amplitud determinada; en ese caso, conociendo la amplitud, se divide el rango y se obtiene la cantidad de intervalos. Es decir:

$$Cantidad.de.intervalos = \frac{R}{H}$$

Vamos a ver cómo queda determinada una tabla de frecuencias cuando los datos originales se agruparon en intervalos de clase.

Consideremos el siguiente ejemplo: en una empresa se obtuvo la edad de los empleados del sector productivo. Las observaciones se ordenaron y organizaron en la siguiente tabla:

Xi	Xi'	fai	fri	Fai	Fri
[20-24)	22	1	0,005	1	0,005
[24-28)	26	3	0,015	4	0,020
[28-32)	30	9	0,045	13	0,065
[32-36)	34	30	0,150	43	0,215
[36-40)	38	60	0,300	103	0,515
[40-44)	42	52	0,260	155	0,775
[44-48)	46	35	0,175	190	0,950
[48-52)	50	10	0,050	200	1,000
		200	1,000	-	-

NOTA: cuando escribimos un intervalo [Li-Ls), el símbolo "[" indica que el valor que le sucede está contenido en dicho intervalo y el símbolo ")" indica que el valor que le precede no está contenido en el intervalo. Por ejemplo, en la tabla el valor X = 24 pertenece al segundo intervalo.

En esta serie tenemos:

$$R = máx(Xi) - mín(Xi) = 52 - 20 = 32 \text{años}$$

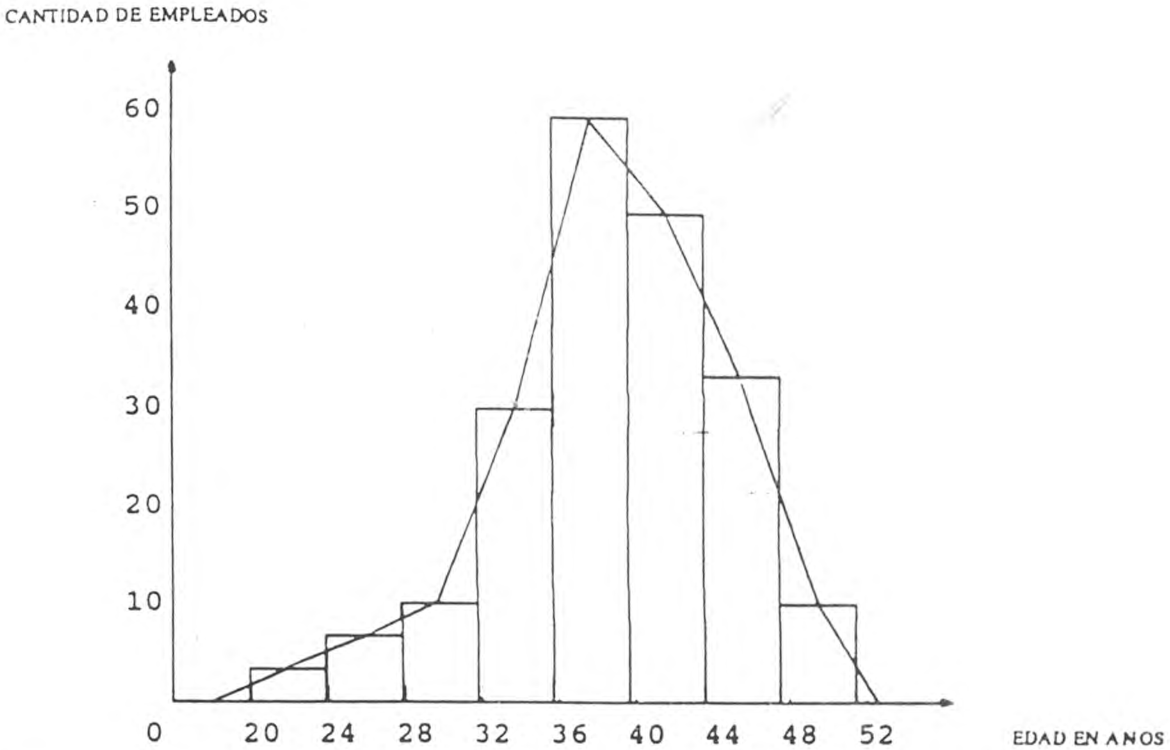
$$H = 24 - 20 = 4 \text{años}$$

2.1 - Representación gráfica

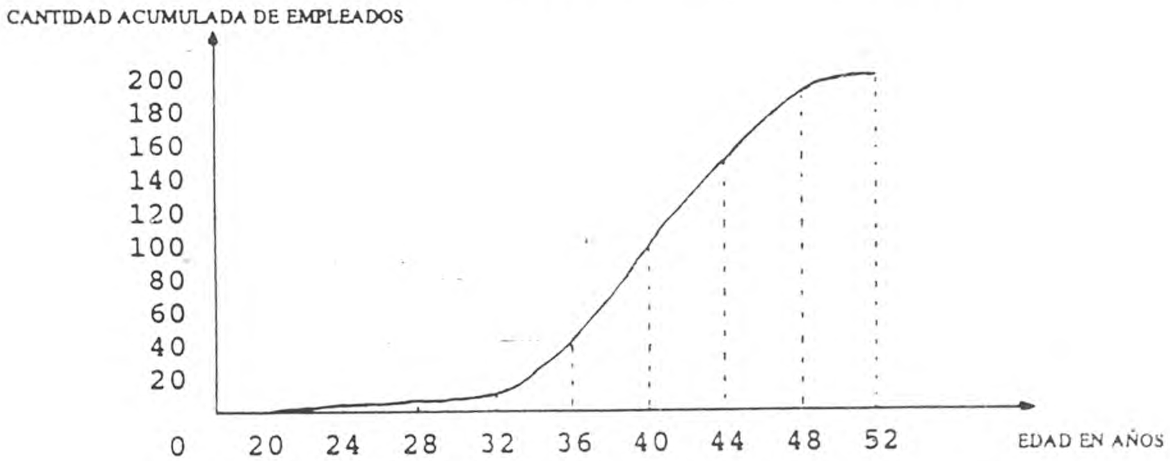
Para representar gráficamente una distribución de frecuencias para datos agrupados usamos el *Histograma* y el *Polígono de frecuencias*.

Histograma: es la representación en un sistema de coordenadas cartesianas ortogonales de la distribución de frecuencias -absolutas o relativas-, de una variable agrupada en intervalos, mediante un gráfico de superficies. Sobre el eje de las abscisas se presentan los intervalos y se levanta sobre cada una de ellos un rectángulo cuyo área es igual a la respectiva frecuencia.

Polígono de frecuencias: es una línea poligonal obtenida en un histograma, uniendo los puntos medios de los lados superiores de los rectángulos. Los dos extremos del polígono pueden o no ser conectados con el eje de la variable. Cuando ello ocurre, se crean dos intervalos hipotéticos con frecuencia cero, colocado cada uno de ellos en ambos extremos del histograma y con amplitud igual a la del primer y último intervalo respectivamente.



Ojiva: *es la representación gráfica de las frecuencias acumuladas -relativas o absolutas- de una variable agrupada en intervalos, mediante una línea poligonal obtenida uniendo los puntos que tienen por abscisa los límites superiores del intervalo y por ordenadas las respectivas frecuencias acumuladas. A este gráfico también se lo conoce como polígono de frecuencias acumuladas.*



NOTA: los mismos tipos de gráficos se realizan para representar las frecuencias relativas, cambiando la escala en el eje de ordenadas.

Ejercicio 6.1:

Las ventas de automóviles efectuadas por una concesionaria, en los 24 días hábiles del mes de junio de 1992 fueron:

3 - 6 - 1 - 2 - 3 - 3 - 2 - 6 - 1 - 4 - 2 - 3 - 2 - 5 - 2 - 1 - 4 - 2 - 3 - 5 - 4 - 2 - 3 - 2

a) Cúal es la variable en estudio?
b) La variable en estudio es continua o discreta?
c) Construya una distribución de frecuencias para resumir la información.
d) Qué porcentaje de días tuvieron una venta menor o igual a 2 unidades?
e) Qué proporción de días tuvieron ventas superiores o iguales a 4 unidades?
f) Represente gráficamente los datos resumidos en la tabla de frecuencias.

Resolución del ejercicio 6.1:

a)

b)

c)

d)

e)

f)

Ejercicio 6.2:

Los siguientes datos representan las importaciones de 42 países, realizadas durante 1989, en miles de dólares:

72333	98854	31415	27600	38847	105084
17312	49718	24469	27430	11525	22240
83933	19399	57615	27500	15420	18750
62712	29920	65777	43232	34451	10090
98003	88932	88255	56611	45098	43886
13343	11564	24874	32955	44809	90405
21442	56832	49022	47531	39920	37772

a) Agrupe los datos en intervalos de clase y ordénelos en una distribución de frecuencias.
b) Qué graficaciones aplicaría? Represéntelas.
c) Qué proporción de países tuvieron importaciones por menos de 20.000 dólares?

Resolución del ejercicio 6.2:

a)

b)

c)

Ejercicio 6.3:

Los siguientes datos están agrupados en intervalos de clase y corresponden a los precios de un conjunto de 40 productos de verdulería.

Precio en $ por kg.	fai	fri	Fai	Fri
menos de 1.0	5			
[1.0 - 1.5)	12			
[1.5 - 2.0)	13			
[2.0 - 2.5)	7			
más de 2.5	3			

a) Complete la información de la tabla precedente.
b) Qué significa el valor 12 en la columna fai?
c) Qué significa el valor 17 en la columna Fai?
d) Qué significa el valor 0.125 en la columna fri?

Resolución del ejercicio 6.3:

b)

c)

d)

Ejercicio 6.4:

En una encuesta se midió la altura de un grupo de 133 alumnos de una Universidad. Los datos se presentan en la siguiente tabla:

X: alt. en cm.	fai	X: alt. en cm.	fai
(158-159]	1	(167-168]	10
(159-160]	2	(168-169]	14
(160-161]	3	(169-170]	20
(161-162]	5	(170-171]	17
(162-163]	6	(171-172]	9
(163-164]	7	(172-173]	7
(164-165]	14	(173-174]	5
(165-166]	6	(174-175]	2
(166-167]	4	(175-176]	1

a) Construya una tabla de distribución de frecuencias con intervalos de clase de amplitud 2 y otra con intervalos de clase de amplitud 3.
b) Grafique los histogramas para cada una de las dos distribuciones.
c) A partir de los histogramas dibujados, qué distribución elegiría como más representativa del fenómeno en estudio? Tenga en cuenta para esto el tipo de variable en estudio y las características de la población.

Resolución del ejercicio 6.4:

a)

b)

c)

UNIDAD 7

MEDIDAS DE TENDENCIA CENTRAL

Vimos que las variables medidas en escalas nominales u ordinales pueden resumirse fácilmente en términos de porcentajes, proporciones o razones. En el caso de variables medidas en escalas de intervalo, a su vez, vimos que los datos pueden describirse por medio de una distribución de frecuencias.

Al obtener de una población la distribución de frecuencias de una variable, lo que se persigue es reducir o condensar en pocas cifras el conjunto de observaciones relativas a dicha variable.

Este proceso de reducción debe continuarse hasta su grado máximo, es decir, hasta sustituir todos los valores observados por uno solo, que se llama **promedio**. Cuando esto se hace se logra, por una parte, una visión más clara del nivel que alcanza la variable y, por otra, una mayor facilidad al hacer comparaciones.

La noción de promedio lleva implícita la idea de variación, ya que no tiene sentido promediar cantidades invariables. Este número promedio que sustituye al conjunto de observaciones de la población debe cumplir la condición de ser **representativo** del conjunto, para lo cual ha de reflejar la tendencia de las observaciones. Es por esto que cuando hablamos de promedios nos referimos a ellos como ***medidas de tendencia central***.

El carácter representativo del promedio exige, evidentemente, que su valor esté comprendido, al menos, entre los valores extremos observados de la variable. En consecuencia, existen numerosas maneras de calcular promedios que cumplan esta condición. El más conocido de ellos es la ***media aritmética***, que es el de uso más frecuente. Además veremos en este curso, la ***mediana***, el ***modo***, la ***media geométrica*** y la ***media armónica***.

La existencia de varios promedios no implica que todos se deban calcular sobre la misma variable. En determinadas circunstancias, dichos promedios dan resultados muy distintos. Cada uno de éstos tienen ventajas y desventajas que los hacen más adecuados para ser aplicados a cada variable en particular.

1 - Media aritmética

La media aritmética es el número que se obtiene al dividir la suma de todas las observaciones por la cantidad de observaciones sumadas. Por lo tanto, la media aritmética es un valor de la variable, posiblemente no observable, que viene dado en la misma unidad de medida de la variable.

A la media aritmética la simbolizamos con $\bar{X}$.

Veamos a continuación su forma de cálculo, que varía de acuerdo al modo en que estén organizados los datos que queremos promediar.

1.1 - Cálculo de la $\bar{x}$ para datos simples o sin agrupar

Sea una variable "x" cualquiera, que toma los siguientes valores:

9 - 8 - 5 - 10 - 13 - 11 - 22 - 15 - 16 - 12 - 14 - 18

La media aritmética para estos valores es:

$$\bar{X}=\frac{\sum_{i=1}^{12} x_i}{12}=\frac{9+8+5+10+13+11+22+15+16+12+14+18}{12}=12,75$$

Si generalizamos este cálculo para cualquier variable "x" y los valores que ésta puede tomar, tenemos:

$$\bar{X}=\frac{x_1+x_2+\ldots+x_n}{n}=\frac{\sum_{i=1}^{n} x_i}{n}$$

La expresión precedente es la fórmula de cálculo para la media aritmética cuando los datos a promediar no están agrupados o resumidos en una distribución de frecuencias.

1.2 - Cálculo de la $\bar{x}$ para datos agrupados

A) Sin intervalos de clase

Cuando la cantidad de observaciones es muy grande y las mismas están organizadas en una tabla de frecuencias, calculamos la media aritmética haciendo uso de las frecuencias absolutas observadas -*fai*- para cada valor de la variable, logrando de esta manera una reducción de la operación de cálculo.

La expresión de la media aritmética en este caso es la siguiente:

$$\overline{X}=\frac{\sum_{i=1}^{n} x_i \times fa_i}{\sum_{i=1}^{n} fa_i}$$

Donde "i" es la cantidad de valores distintos que toma la variable.

Además, como ya vimos,

$$\sum_{i=1}^{n} fa_i = n$$

entonces:

$$\overline{X}=\frac{\sum_{i=1}^{n} x_i \times fa_i}{n}$$

*Cuando calculamos la media aritmética, multiplicando a cada valor de la variable por su correspondiente frecuencia, decimos que dicha media está "**ponderada**".*

*La **ponderación** es un coeficiente numérico que expresa la importancia absoluta de cada valor de la variable.*

*En este caso particular usamos como ponderación a las frecuencias absolutas -**fai**-.*

Tomando el ejemplo presentado en la UNIDAD 6 para distribuciones de frecuencias, calcularemos la media ponderada de acuerdo a la fórmula presentada:

NUMERO DE CUARTOS POR HOGAR Xi	FRECUENCIA ABSOLUTA fai	FRECUENCIA RELATIVA fri	FRECUENCIA ABSOLUTA ACUMULADA Fai	FRECUENCIA RELATIVA ACUMULADA Fri
1	144	0,214	144	0,214
2	225	0,334	369	0,548
3	174	0,259	543	0,807
4	88	0,131	631	0,938
5	42	0,062	673	1,000
	673	1,000	-	-

Procedamos a calcular la media aritmética para la variable "X: número de cuartos por hogar", usando como ponderador las frecuencias absolutas. La expresión de cálculo que utilizamos en este caso es:

$$\bar{X}=\frac{\sum_{i=1}^{n} x_i \times fa_i}{\sum_{i=1}^{n} fa_i}$$

Si reemplazamos los valores observados en la expresión anterior, tenemos:

$$\bar{X}=\frac{1\times144+2\times225+3\times174+4\times88+5\times42}{673}=2,49$$

Podemos decir entonces que en promedio hay 2,49 cuartos por hogar.

Del mismo modo, podríamos haber usado como ponderación a las frecuencias relativas -fri-. En este caso, el ponderador nos está indicando la importancia "relativa" de cada valor de la variable sobre el total de las observaciones.

La expresión de cálculo en este caso es:

$$\bar{X}=\sum_{i=1}^{n} x_i \times fr_i$$

El cálculo de la media aritmética para el ejemplo presentado, puede realizarse también utilizando como ponderador a las frecuencias relativas.

Reemplazando los datos en la expresión anterior tenemos:

$$\bar{X}=1\times 0,214+2\times 0,334+3\times 0,259+4\times 0,131+5\times 0,062=2,49$$

B) Con intervalos de clase

Si los datos a promediar se encuentran agrupados en intervalos de clase, no podemos contar con el valor exacto que toma la variable para cada observación ya que tenemos frecuencias para un rango de la variable. En consecuencia, el cálculo de la media aritmética en este caso no es exacto.

Una buena aproximación se logra tomando para cada intervalo el punto medio -marca de clase-, como valor observado de la variable. Con esta decisión, estamos suponiendo que todas las observaciones que pertenecen a un intervalo toman como valor la marca de clase -Xi'-.

La expresión de cálculo en este caso es la misma que para datos sin agrupar, donde las "Xi" son reemplazadas por "Xi'". Si ponderamos por las fai tenemos entonces:

$$\bar{X}=\frac{\sum_{i=1}^{n} x_i' \times fa_i}{\sum_{i=1}^{n} fa_i}$$

En el caso de ponderar por las fri tenemos:

$$\bar{X}=\sum_{i=1}^{n} x_i' \times fr_i$$

Observemos la tabla de frecuencias obtenida para el ejemplo desarrollado en la UNIDAD 6, donde se presentaba la distribución por edad de los empleados de una empresa.

Xi	Xi'	fai	fri	Fai	Fri
[20-24)	22	1	0,005	1	0,005
[24-28)	26	3	0,015	4	0,020
[28-32)	30	9	0,045	13	0,065
[32-36)	34	30	0,150	43	0,215
[36-40)	38	60	0,300	103	0,515
[40-44)	42	52	0,260	155	0,775
[44-48)	46	35	0,175	190	0,950
[48-52)	50	10	0,050	200	1,000
		200	1,000	-	-

Si nos manejamos con las frecuencias absolutas -fai- y reemplazamos en la expresión de cálculo los valores observados en la tabla, resolveremos el cálculo de la media aritmética del siguiente modo:

$$\bar{X}=\frac{22\times1+26\times3+30\times9+34\times30+38\times60+42\times52+46\times35+50\times10}{200}$$

$$\bar{X}=\frac{22+78+270+1020+2280+2184+1610+500}{200}=39,82$$

Podemos concluir entonces, que la edad promedio de los empleados de la empresa es de 39,82 años.

El mismo valor para la media aritmética puede obtenerse trabajando con las frecuencias relativas -fri-, usando la expresión correspondiente.

1.3 - Propiedades de la $\overline{X}$

- *La media aritmética es un valor de la variable comprendido entre el máximo y el mínimo valor de la variable.*

- *La unidad de medida de la media aritmética, es igual a la unidad de medida de la variable. Esto significa que si la unidad de medida de la variable viene dada, por ejemplo en kg., la media aritmética también será un valor expresado en kg.*

- *Si la variable toma siempre el mismo valor, la media aritmética es igual a dicho valor.*

- *La suma de las distancias de cada valor de la variable a la media aritmética es igual a "0". Esta propiedad demuestra el efecto compensador que tiene este promedio respecto a la distribución de los datos. Es decir, la media aritmética es un valor que compensa las distancias de los valores que quedan a izquierda y a derecha del valor medio.*

2 - Mediana

Si todos los valores observados de la variable se ordenan en sentido creciente (o decreciente), la mediana es el valor de la variable que ocupa el lugar central, es decir, el que deja a un lado y a otro el mismo número de observaciones.

*La mediana no se expresa, entonces, por una fórmula. Se representa mediante el símbolo **Mna**. Para su obtención se considerará la forma en que estén disponibles los datos.*

2.1 - Cálculo de la Mna para datos simples o sin agrupar

Cuando trabajamos con datos sin agrupar, si el número de observaciones es impar, y tales observaciones se ordenan en forma creciente, la mediana es el valor central de dichas observaciones.

Si el número de observaciones es par, se toma como mediana a la media aritmética de los dos valores centrales.

Por ejemplo, si tomamos las edades de un grupo de 9 personas y las ordenamos de menor a mayor, la mediana será:

X: 17 - 17 - 19 - 19 - $\boxed{22}$ - 22 - 23 - 28 - 29

$$Mna = 22 años$$

Si sólo hubiésemos tenido ocho observaciones, la mediana sería:

X: 17 - 17 - 19 - $\boxed{19 - 22}$ - 22 - 23 - 28

$$Mna = \frac{19+22}{2} = 20{,}5 años$$

2.2 - Cálculo de la Mna para datos agrupados

A) Sin intervalos de clase

A.1) Determinación analítica

Hay que tener siempre presente, que para el cálculo de la Mna es condición previa que los datos estén ordenados. Cuando los datos están agrupados en una distribución de frecuencias, este orden siempre existe. El problema, consiste en hallar el valor de la variable que corresponde a la observación central.

El número total de observaciones es "n", por lo tanto la observación central será "n/2" .

En consecuencia, el valor de la Mna será el correspondiente a la posición "n/2" en la distribución de la variable.

Para encontrar tal valor, el procedimiento consiste en buscar sobre las frecuencias absolutas acumuladas "Fai", el valor de la variable que acumula por lo menos "n/2" observaciones.

Veamos el cálculo de la *Mna* para la distribución de la variable "X: número de cuartos por hogar" que se muestra en el siguiente cuadro:

NUMERO DE CUARTOS POR HOGAR	FRECUENCIA ABSOLUTA	FRECUENCIA RELATIVA	FRECUENCIA ACUMULADA	FRECUENCIA RELATIVA ACUMULADA
1	144	0,214	144	0,214
2	225	0,334	369	0,548
3	174	0,259	543	0,807
4	88	0,131	631	0,938
5	42	0,062	673	1,000
	673	1,000	-	-

La primera operación que hay que realizar es obtener las Fai, que en nuestro ejemplo ya están calculadas.

La segunda operación es calcular "n/2", que en este caso es:

$$\frac{n}{2}=\frac{673}{2}=336,5$$

El tercer paso es localizar la primera Fai que acumule por lo menos 336,5 observaciones -en este caso 336,5 hogares-. En el ejemplo, la Fai correspondiente es 369.

Por lo tanto, la Mna será el valor de la variable correspondiente a dicha Fai, es decir:

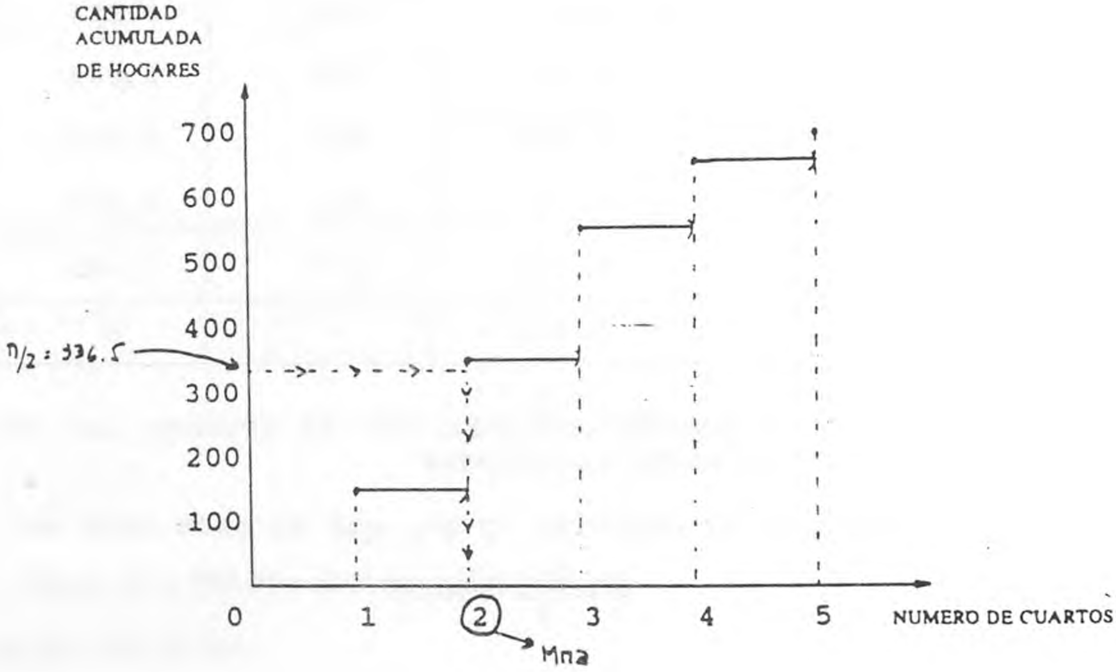
Mna = 2 cuartos por hogar

A.2) Determinación gráfica

El valor de la Mna también puede obtenerse a partir del gráfico de frecuencias absolutas acumuladas, teniendo en cuenta los siguientes pasos:

- *Ubicamos el resultado de hacer "n/2" sobre el eje de ordenadas.*
- *Trazamos una línea horizontal a la altura de dicho valor, hasta tocar el gráfico.*

- Luego, bajamos hasta el eje de abscisas. El punto que encontramos, es el valor correspondiente a la Mna.



B) Con intervalos de clase

B.1) Determinación analítica

Cuando la distribución de frecuencias viene dada para valores de la variable agrupados en intervalos de clase, no puede obtenerse exactamente el valor de la Mna, porque se desconocen las observaciones individuales de la variable.

No obstante, puede obtenerse una aproximación del valor de la Mna mediante la siguiente expresión:

$$Mna = L_i + \frac{\frac{n}{2} - Fa_{i-1}}{fa_i} \times h_i$$

Siendo:

Li: el límite inferior del intervalo correspondiente a la frecuencia absoluta acumulada que contiene a la cantidad "n/2".

Fa(i-1): la frecuencia absoluta acumulada hasta el intervalo anterior al que contiene la Mna.

fai: la frecuencia absoluta del intervalo en el que ubicamos a la Mna.

hi: la amplitud del intervalo en el que se encuentra la Mna.

Esta fórmula se obtiene bajo el supuesto de que las observaciones pertenecientes a cada intervalo se reparten uniformemente dentro de él. Esto significa suponer que todos los valores de la variable contenidos en un intervalo tienen la misma frecuencia absoluta.

Este supuesto rara vez se cumple, por lo tanto, el valor obtenido para la Mna es un valor aproximado.

Retomemos el ejemplo de la distribución de las edades de los empleados de una empresa:

Xi	Xi'	fai	fri	Fai	Fri
[20-24)	22	1	0,005	1	0,005
[24-28)	26	3	0,015	4	0,020
[28-32)	30	9	0,045	13	0,065
[32-36)	34	30	0,150	43	0,215
[36-40)	38	60	0,300	103	0,515
[40-44)	42	52	0,260	155	0,775
[44-48)	46	35	0,175	190	0,950
[48-52)	50	10	0,050	200	1,000
		200	1,000	-	-

El primer paso que debemos cumplir para el cálculo de la Mna es encontrar el valor que divide a las observaciones en 2 partes iguales, es decir:

$$\frac{n}{2} = \frac{200}{2} = 100$$

Luego ubicamos el intervalo de clase que contiene por lo menos a las 100 primeras observaciones. Observamos que dicho intervalo es el correspondiente al tramo de edades [36 - 40).

Entonces:

$$Mna = 36 + \frac{100-43}{60} \times 4$$

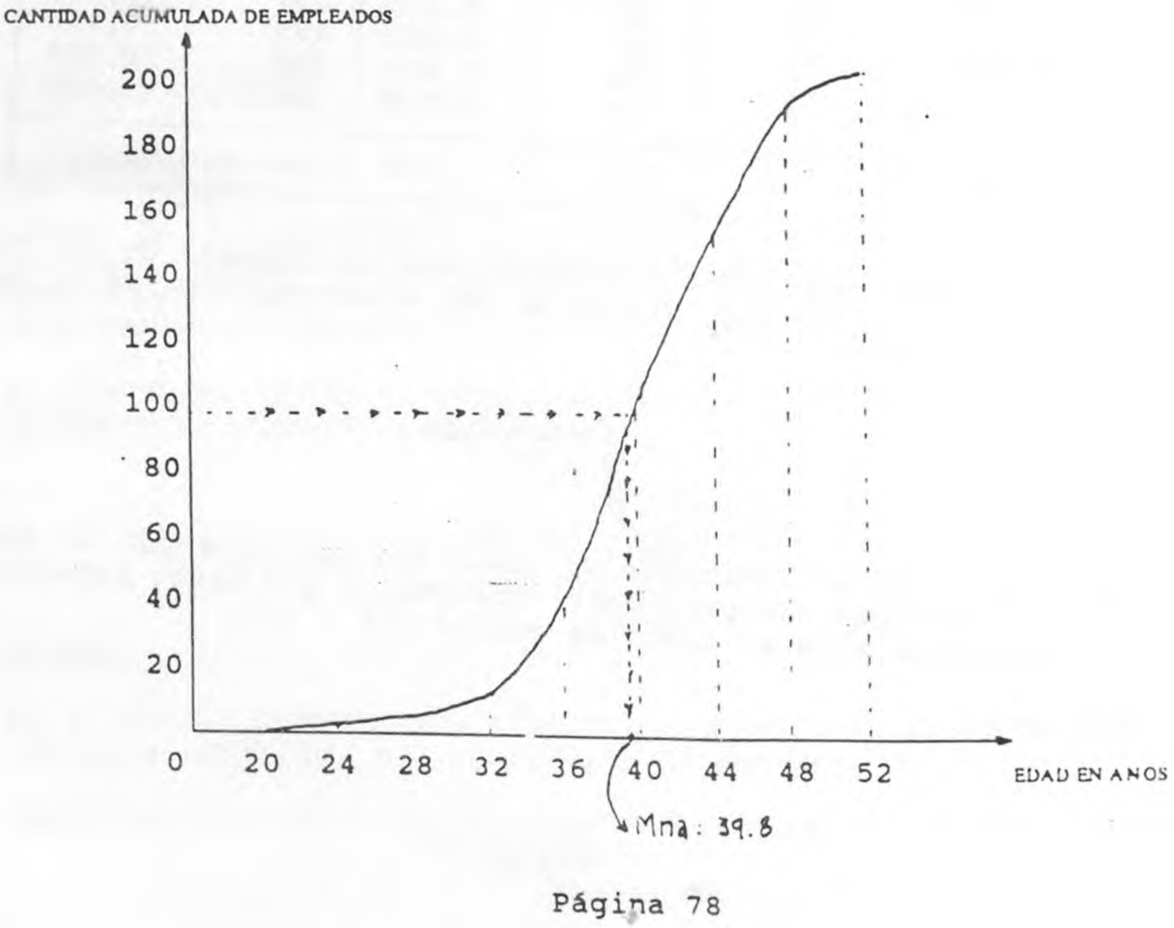
$$Mna = 36 + \frac{57}{60} \times 4 = 36 + 0,95 \times 4 = 36 + 3,8 = 39,8$$

Por lo tanto, la Mna para las edades de los 200 empleados de la empresa es 39,8 años.

B.2) Determinación gráfica

Al igual que para el caso A) utilizamos el gráfico de frecuencias acumuladas, en este caso el correspondiente para datos agrupados en intervalos de clase.

Siguiendo el mismo procedimiento que para datos sin agrupar, y utilizando el ejemplo de las edades de los empleados de una empresa, tenemos:



3 - *Modo*

El modo es el valor de la variable que más veces se repite, o sea, el valor que presenta mayor frecuencia.

Es útil como medida de tendencia central, sólo en aquellos casos en que un valor de la variable es mucho más frecuente que el resto. Se basa en la idea de "lo que es moda" o en el "comportamiento de la mayoría" para tomar a cierto valor como representativo del comportamiento de los datos.

En el caso del modo, no existe una fórmula general para expresarlo. Lo simbolizamos con *Mo*.

Veamos cómo se encuentra el *Mo* para los distintos tipos de disposición de los datos.

3.1 - Cálculo del Mo para datos simples o sin agrupar

Cuando el número de observaciones es pequeño, como es el caso de datos sin agrupar, no debe calcularse el Mo porque no puede apreciarse si existe una decidida tendencia de los datos a concentrarse en un sólo valor de la variable.

De todos modos, su cálculo es inmediato. Basta con localizar el valor que más veces se ha presentado.

3.2 - Cálculo del Mo para datos agrupados

A) Sin intervalos de clase

En este caso, el Mo se obtiene rápidamente. En la distribución de frecuencias se observa cuál es la mayor frecuencia absoluta. Entonces, el Mo será el valor de la variable correspondiente a dicha frecuencia.

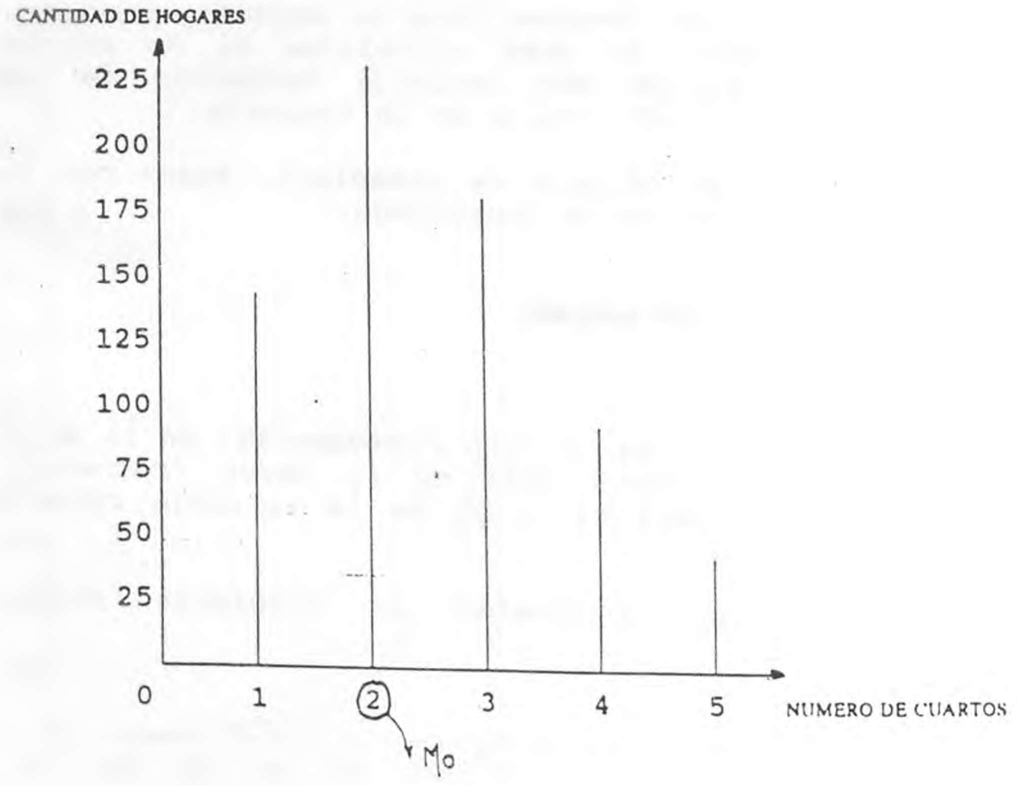
Por ejemplo, si observamos la siguiente distribución de frecuencias:

NUMERO DE CUARTOS POR HOGAR	FRECUENCIA ABSOLUTA	FRECUENCIA RELATIVA	FRECUENCIA ACUMULADA	FRECUENCIA RELATIVA ACUMULADA
1	144	0,214	144	0,214
2	225	0,334	369	0,548
3	174	0,259	543	0,807
4	88	0,131	631	0,938
5	42	0,062	673	1,000
	673	1,000	-	-

Observamos que el Mo corresponde al valor de la variable x=2, ya que éste es el valor de la variable con mayor frecuencia absoluta.

El Mo también puede obtenerse gráficamente observando el gráfico de frecuencias absolutas para datos sin agrupar. En este caso, el Mo corresponde al valor de la variable que tiene mayor ordenada.

Veamos el gráfico correspondiente a la distribución de la variable "X: número de cuartos por hogar".



B) Con intervalos de clase

Al igual que para las otras medidas de tendencia central, cuando trabajamos con datos agrupados en intervalos de clase, no puede calcularse exactamente el Mo, por desconocerse los valores observados de la variable.

Una aproximación del Mo se obtiene mediante la siguiente expresión:

$$Mo = L_i + \frac{d_1}{d_1 + d_2} \times h_i$$

Siendo:

Li: límite inferior del intervalo de clase al que corresponde la mayor frecuencia absoluta.

d1: diferencia absoluta entre la frecuencia absoluta del intervalo de mayor frecuencia o intervalo modal respecto a la frecuencia absoluta del intervalo anterior.

d2: diferencia absoluta entre la frecuencia absoluta del intervalo de mayor frecuencia o intervalo modal respecto a la frecuencia absoluta del intervalo posterior.

hi: amplitud del intervalo modal.

NOTA: esta expresión es aplicable solamente en el caso de que todos los intervalos tengan la misma amplitud.

Ejemplo:

Xi	Xi'	fai	fri	Fai	Fri
[20-24)	22	1	0,005	1	0,005
[24-28)	26	3	0,015	4	0,020
[28-32)	30	9	0,045	13	0,065
[32-36)	34	30	0,150	43	0,215
[36-40)	38	60	0,300	103	0,515
[40-44)	42	52	0,260	155	0,775
[44-48)	46	35	0,175	190	0,950
[48-52)	50	10	0,050	200	1,000
		200	1,000	-	-

Vemos en la tabla que el valor fai = 60 es la mayor frecuencia absoluta observada.

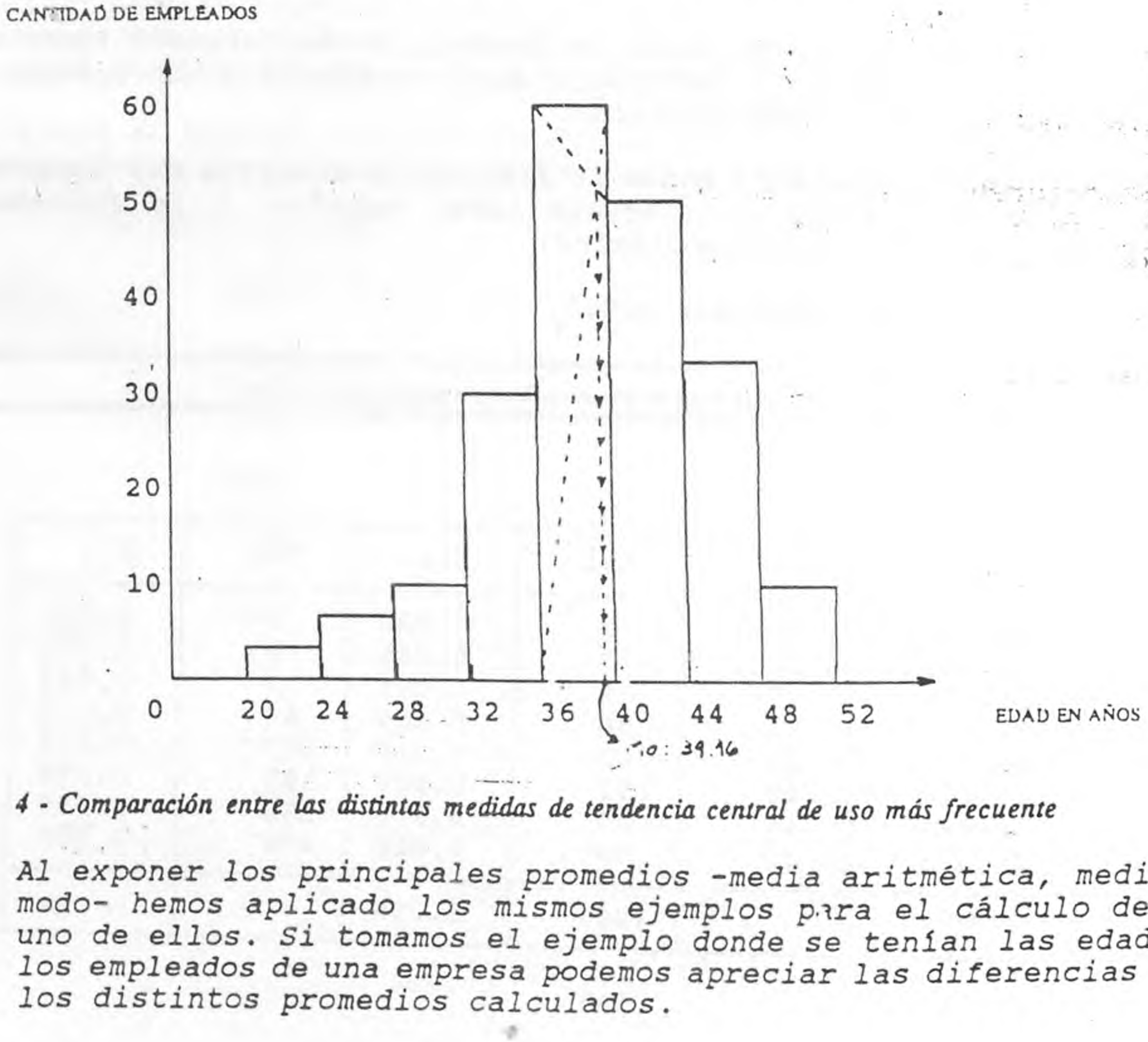
Por lo tanto, el Mo corresponde al intervalo [36 - 40) de la variable.

Entonces:

$$Mo = 36 + \frac{30}{30+8} \times 4 = 36 + 0,79 \times 4 = 36 + 3,16 = 39,16$$

Es decir, el Mo para la variable edad de los empleados de la empresa es Mo = 39,16 años.

También podemos obtener el Mo gráficamente utilizando el HISTOGRAMA. Lo vemos en el siguiente ejemplo:



4 - Comparación entre las distintas medidas de tendencia central de uso más frecuente

Al exponer los principales promedios -media aritmética, mediana y modo- hemos aplicado los mismos ejemplos para el cálculo de cada uno de ellos. Si tomamos el ejemplo donde se tenían las edades de los empleados de una empresa podemos apreciar las diferencias entre los distintos promedios calculados.

Recordemos cuales fueron dichos valores:

$\overline{X}$ = 39,82 años

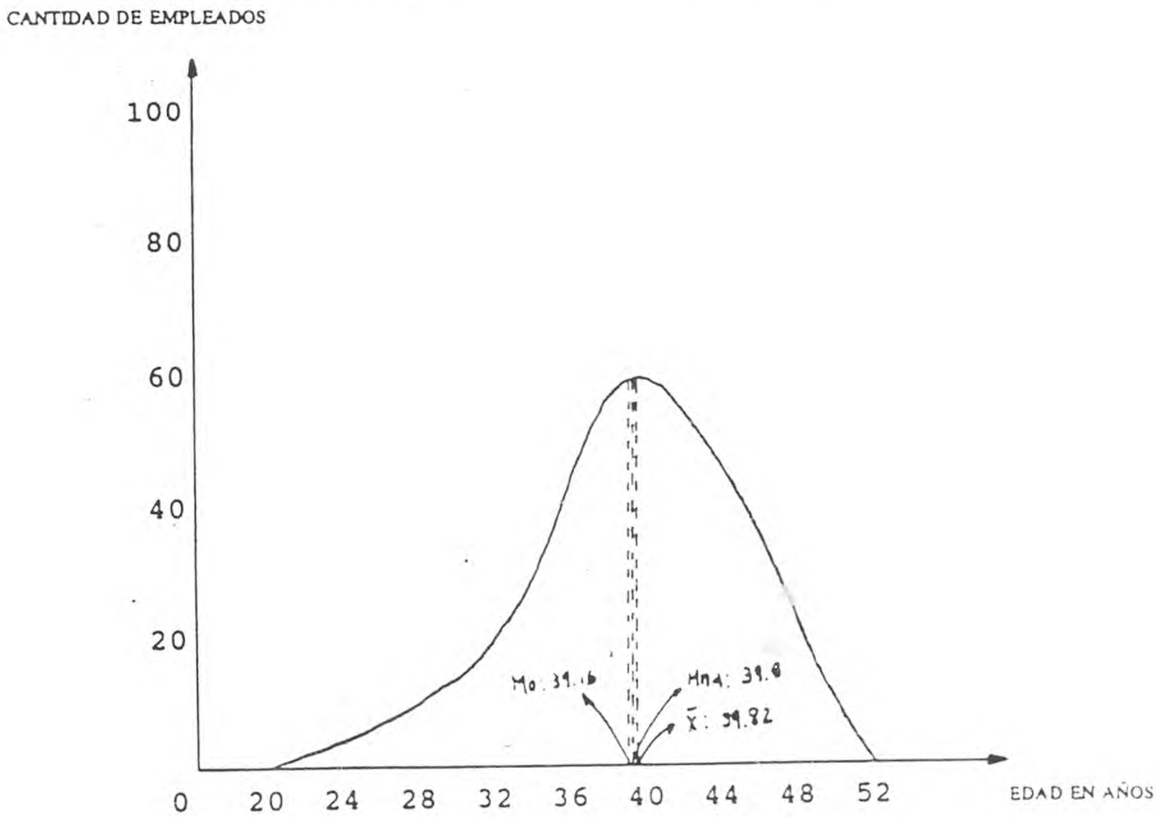
Mna = 39,8 años

Mo = 39,16 años

Puede observarse que para una misma distribución rara vez coinciden los valores obtenidos mediante los tres promedios. Esto nos plantea la siguiente pregunta:

QUE PROMEDIO DEBE USARSE EN CADA CASO?

A continuación vamos a resumir las características de cada uno de los tres promedios considerados, así como sus ventajas e inconvenientes. Para ello vamos a representar gráficamente la línea suavizada del polígono que define la distribución de la variable "edad", que consideramos levemente asimétrica.



► *Media aritmética: la X es el centro de gravedad de la distribución. El punto X es el punto de equilibrio de la figura.*

La media aritmética es un valor de la variable que depende de todas las observaciones, porque en su cálculo intervienen todas ellas. Por lo tanto, la presencia de un valor observado anormalmente grande o anormalmente chico influye sensiblemente en el valor del promedio, lo cual evidentemente, es un inconveniente de la media aritmética.

Frente a este inconveniente tiene la ventaja de utilizar "toda" la información recogida.

En estadística se trabaja frecuentemente con muestras. Con una muestra no puede obtenerse el valor exacto de un promedio de la población, sólo se obtiene una estimación de él. Una condición esencial de cualquier promedio es que su valor en la muestra no varíe mucho al pasar de una muestra a otra, es decir, que el promedio calculado sea lo más estable posible.

Esta condición de la máxima estabilidad la posee la media aritmética.

Finalmente, la $\bar{X}$, por venir definida mediante una expresión algebraica, puede someterse a cálculos matemáticos necesarios para deducir cuestiones importantes. Por ejemplo, si tenemos 2 valores medios calculados con la media aritmética, dichos valores podrán ser comparados haciendo uso de operaciones algebraicas.

► ***Mediana:*** por definición, sabemos que la mediana es el valor de la variable que deja a un lado y a otro de dicho valor el mismo número de observaciones, bajo el supuesto de que los datos están ordenados en sentido creciente o decreciente. En la gráfica, la ordenada correspondiente a la Mna, divide el área total en dos partes iguales.

Para determinar el valor de la mediana, no es necesario conocer el valor de todas las observaciones, sólo es preciso saber el valor de la observación central y que las restantes son mayores o menores que ésta. No se utiliza, pues, toda la información recogida para su cálculo, lo cual es un inconveniente.

En cambio, tiene la ventaja de que los valores observados anormalmente grandes o anormalmente chicos no influyen en el promedio.

Otra ventaja es que puede obtenerse con datos incompletos, por ejemplo, en las distribuciones de frecuencias con intervalos de clase que comienzan con un intervalo " menos de ..." o finalizan con intervalos "más de ...".

Un serio inconveniente es que la Mna no viene definida mediante una expresión matemática. La fórmula de aproximación es, simplemente un artificio que se utiliza en el caso de las distribuciones para datos agrupados en intervalos de clase. En consecuencia, no puede someterse al cálculo algebraico para deducir cuestiones importantes de comportamiento. Por ejemplo, no podrán compararse dos valores de Mna utilizando cálculos algebraicos.

▸ *Modo:* como ya vimos, es el valor más frecuente, es decir, el punto donde se concentra el mayor número de observaciones. En la gráfica, el modo es el punto de la variable al cual le corresponde la altura máxima de la curva.

Este promedio, tampoco utiliza toda la información pues, basta con saber tan sólo qué valor de la variable es el más frecuente. Esto hace, al igual que en el caso de la Mna, que éste promedio no se vea afectado por los valores anormalmente grandes o anormalmente chicos.

Tampoco el Mo se define algebraicamente y por ello no puede utilizarse para la obtención de deducciones matemáticas.

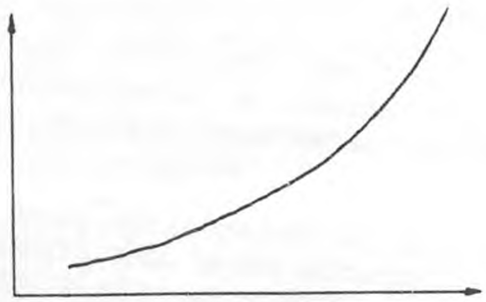
El modo es un promedio muy interesante cuando existe en la distribución una clara y decidida tendencia a que los valores se concentren alrededor de un solo valor.

Una vez vistas las propiedades de cada promedio separadamente, conviene repasar algunas cuestiones que afectan a todos ellos.

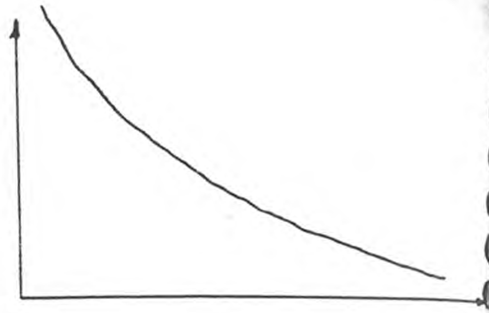
Recordemos primeramente que un promedio tiene por objeto obtener un valor de la variable alrededor del cual se distribuyen las observaciones. Esta condición se cumple muy bien en las distribuciones simétricas o moderadamente asimétricas. Si la distribución de la variable es de este tipo, los tres promedios -media aritmética, mediana y modo- son perfectamente representativos del conjunto de observaciones. En este caso es difícil señalar una preferencia de uno sobre otro desde el punto de vista de su representatividad. Si tomamos en cuenta, las restantes propiedades, el mejor promedio, es la media aritmética por sus propiedades matemáticas y de estabilidad en el muestreo.

Si la distribución es fuertemente asimétrica, es decir tiene forma de "J" o de "L", entonces, la mediana es el promedio más apto.

Distribución con forma de "J"

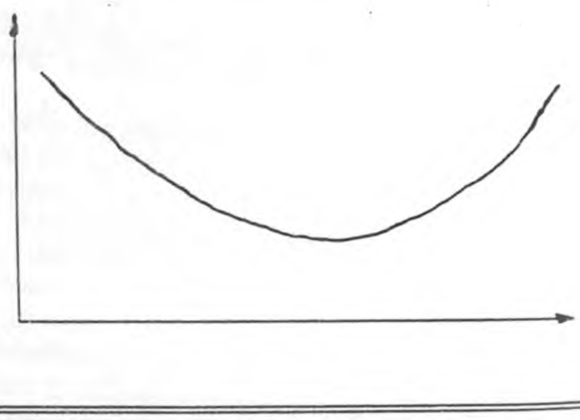


Distribución con forma de "L"



Si la distribución tiene forma de "U", los tres promedios tiene poca fuerza representativa. Generalmente, las distribuciones d esta forma suelen ser difíciles de tratar desde el punto de vist de los promedios.

Distribución con forma de "U"



NOTA: recuerde siempre que el tipo de distribución que representan los datos es importante para la selección del promedio más adecuado. En caso de duda, debe seguir siempre la misma regla: EMPLEAR LA MEDIA ARITMETICA.

5 - La media geométrica y la media armónica

La media aritmética ($\bar{X}$), es sin duda, el promedio más utiliza siguen a éste la mediana (Mna) y el modo (Mo).

Pero como se dijo al principio, existen al menos teóricam infinitos promedios, entre los cuales, algunos se emplean algunas situaciones particulares.

5.1 - *Media geométrica*

Cuando trabajamos con una variable cuyos valores crecen en forma geométrica o casi geométrica, por ejemplo, X: 2,4,8,15,29..... no podemos utilizar para calcular el promedio ninguna de las medidas de tendencia central vistas hasta el momento, pues ninguna de ellas es representativa del comportamiento de los datos.

La media geométrica es un tipo de promedio que se utiliza siempre y cuando las observaciones de una variable tengan un comportamiento geométrico.

La media geométrica de n observaciones de una variable, que simbolizamos con $\bar{X}g$, se calcula como:

$$\bar{X}g=\sqrt[n]{X_1 \times X_2 \times X_3 \ldots\ldots \times X_n}$$

Si los datos están agrupados en intervalos de clase, con sus correspondientes frecuencias, la expresión de cálculo es la siguiente:

$$\bar{X}g=\sqrt[n]{X_1^{fa1} \times X_2^{fa2} \times X_3^{fa3} \ldots\ldots \times X_n^{fan}}$$

La media geométrica no puede utilizarse cuando la variable, en algún caso, toma el valor "0" o valores negativos.

Veamos un ejemplo:

Una ciudad turística se puso de moda en los últimos años a partir de una fuerte campaña publicitaria. La afluencia de turistas en los últimos 9 años fue la siguiente:

AÑO	TURISTAS EN MILES
1984	3
1985	4,6
1986	7
1987	10
1988	15
1989	22
1990	34
1991	40
1992	60

Si queremos calcular el promedio anual de turistas que arribaron a dicha ciudad y observamos el comportamiento de la variable a través del tiempo, podemos observar que hubo un crecimiento casi geométrico en los valores observados.

Por lo tanto, dicho promedio lo calculamos utilizando la media geométrica, es decir:

$$\bar{X}g=\sqrt[9]{3\times4,6\times7\times10\times15\times22\times34\times40\times60}$$

$$\bar{X}g=\sqrt[9]{26012448000}=14,363$$

Por lo tanto, el promedio anual de turistas que arribaron a dicha ciudad es de 14363 turistas.

5.2 - Media armónica

La media armónica es un promedio que se utiliza cuando los valores de la variable que queremos promediar no se obtienen por observación directa de dicha variable, sino que resultan del cociente de los valores observados de 2 variables.

La media armónica, que simbolizamos con $\bar{X}a$, de n observaciones de una variable se calcula como:

$$\bar{X}a=\frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

Por ejemplo, si quisiéramos obtener la velocidad promedio de un viaje de ida y vuelta desde Buenos Aires a la ciudad de Mar del Plata, donde el viaje de ida demandó 4:30 hs. y el de vuelta 5:40 hs., utilizaríamos la media armónica.

Esto es así, debido a que la velocidad se calcula como un cociente entre el espacio recorrido y el tiempo utilizado en recorrerlo.

Si llamamos:

V = velocidad
E = espacio recorrido
T = tiempo utilizado

Entonces:

$$V=\frac{E}{T}$$

La velocidad de ida, a la que llamamos "X1", la calculamos como:

$$X_1=V_{(ida)}=\frac{400km}{4,5hs}=88,9km/h$$

La velocidad de vuelta, a la que llamamos "X2", la calculamos como:

$$X_2=V_{(vuelta)}=\frac{400km}{5,67hs}=70,5km/h$$

La velocidad promedio, utilizando la media armónica resulta:

$$\overline{X}g=\frac{2}{\frac{1}{88,9}+\frac{1}{70,5}}=\frac{2}{0,0112+0,0142}=\frac{2}{0,0254}=78,74km/h$$

Ejercicio 7.1:

Los sueldos que cierta companía paga a su personal se muestran en la siguiente tabla:

X: Sueldo en $	fai: cantidad de personas
5600	4
4800	10
3300	15
2450	21
1900	30
1200	40
900	50
500	8

a) Calcule el sueldo promedio por persona utilizando la media aritmética.
b) Puede afirmar que este promedio es representativo de los sueldos de este grupo de personas?

Resolución del ejercicio 7.1:
a)   b)

Ejercicio 7.2:

La Dirección de Recursos Humanos del INDEC evalúa a los solicitantes de empleo mediante un test de capacidad. El tiempo en minutos que empleó para su realización un grupo de 100 personas que dió el test con resultado "muy satisfactorio", se muestra en la siguiente tabla:

X: minutos	fai: cantidad de personas
(24-28]	2
(28-32]	3
(32-36]	8
(36-40]	12
(40-44]	45
(44-48]	14
(48-52]	10
(52-56]	6

a) Cuál es el tiempo promedio empleado por estas personas para la realización del test?

b) Cuánto tardaron como máximo el 50% de las personas que realizaron el test en menos tiempo? Realice el cálculo analítica y gráficamente.

c) Cuál es el modo de la distribución? Calcúlelo analítica y gráficamente.

d) Si se decide que ingresarán aquellas personas que realizaron el test en menos de 40 minutos, qué porcentaje de los mismos ingresará al INDEC?
e) Si usted tuviera que determinar un valor promedio del tiempo que tardan las personas que dan el test con resultado "muy satisfactorio", qué medida de tendencia central utilizaría?

Resolución del ejercicio 7.2:
a)
b)

c)

d)

e)

Ejercicio 7.3:

Utilizando los datos del Ejercicio 6.4, saque conclusiones acerca de la altura promedio de los alumnos de dicha Universidad.

Resolución del ejercicio 7.3:

Ejercicio 7.4:

Se ha lanzado al mercado una nueva marca de vino, cuyo consumo fue aumentando a medida que la gente lo fue conociendo. Desde la fecha de su lanzamiento transcurrieron 7 meses y las ventas efectuadas en *dicho lapso son las siguientes:*

Meses	Cantidad consumida en hl.
mayo	50
junio	65
julio	84
agosto	109
setiembre	162
octubre	211
noviembre	274

*a) Obtenga el promedio mensual de consumo de este vino utilizando la medida de tendencia central **más** adecuada. Justifique su elección.*
b) Si no contara con el dato de consumo del mes de setiembre, como estimaría dicho valor para incluírlo en la serie?

Resolución del ejercicio 7.4:
a)
b)

Ejercicio 7.5:

La variación porcentual de los índices de precios al consumidor durante los últimos 9 meses de 1991 fueron las siguientes:

Meses	Variación porcentual
abril	5,5
mayo	2,8
junio	3,1
julio	2,6
agosto	1,3
setiembre	1,8
octubre	1,4
noviembre	0,4
diciembre	0,6

Calcule la variación porcentual promedio del índice de precios al consumidor en los últimos 9 meses de 1991 utilizando la medida de tendencia central más adecuada.

Resolución del ejercicio 7.5:

UNIDAD 8

OTRAS MEDIDAS DE POSICION Y SU UTILIDAD

Así como vimos las medidas de tendencia central, existen otras medidas que indican una posición determinada de la variable y que son de suma utilidad en el análisis de los datos.

Estas medidas son conceptualmente análogas a la mediana. Así como ésta busca el valor de la variable que reparte las observaciones en un 50% a la izquierda y un 50% a la derecha de la distribución, se puede determinar algún valor particular de la variable que deje a su izquierda y derecha porcentajes distintos al 50%.

Por ejemplo, nos puede interesar saber qué valor de la variable deja a su izquierda el 25% de las observaciones, o cuál es el valor que a su derecha contiene al 40% de las observaciones.

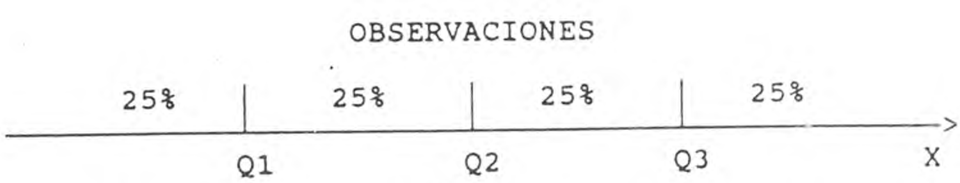
Entre las medidas de posición que veremos en esta oportunidad, se encuentran los *cuartiles, deciles* y *percentiles*.

1 - Cuartiles

Los cuartiles son aquellos valores de la variable que dividen a las observaciones en "4" partes iguales. Los mismos se simbolizan con Q_i, donde el subíndice "i" toma los valores 1, 2 y 3.

Es decir, existen tres cuartiles: Q_1, Q_2 y Q_3.

Gráficamente, siendo "X" una variable cualquiera, tenemos:



Veamos cómo se procede para el cálculo de los cuartiles.

1.1 - Cálculo de los Q_i para datos simples o sin agrupar

En este caso, se divide al total de observaciones "n", por cuatro.

Para el cálculo del primer cuartil $-Q_1-$, simplemente tomamos el valor $n/4$, es decir:

$$Q_1=\frac{n}{4}$$

El segundo cuartil *-Q2-*, se obtiene a partir del *Q1*, multiplicando a éste por 2:

$$Q_2=Q_1\times 2=\frac{n}{4}\times 2=\frac{n}{2}$$

Como puede observarse:

$$Q_2=Mna$$

El tercer cuartil *-Q3-*, se obtiene haciendo:

$$Q_3=Q_1\times 3=\frac{n}{4}\times 3$$

1.2 - Cálculo de los Qi para datos agrupados

Cuando los datos se encuentran agrupados en distribuciones de frecuencia, ya sea con o sin intervalos de clase, el procedimiento es análogo al del cálculo de la Mna, teniendo en cuenta, que la partición de las "n" observaciones debe hacerse por "4" y no por "2", y que según sea el número de cuartil que busquemos, deberá multiplicarse el resultado del cociente por dicho número. En forma general tenemos:

$$Q_i=\frac{n}{4}\times i$$

Al resultado de esta operación le aplicamos el procedimiento que corresponde según sea el agrupamiento de los datos.

2 - Deciles

Los deciles son aquellos valores de la variable que dividen a las observaciones en 10 partes iguales. Los mismos se simbolizan con *Di*, donde el subíndice "i" toma los valores 1 a 9.

Es decir, existen 9 deciles: D1, D2, D3, D4, D5, D6, D7, D8 y D9.

Veamos gráficamente la posición de los deciles sobre la distribución de una variable "X" cualquiera:

OBSERVACIONES

10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	
D1	D2	D3	D4	D5	D6	D7	D8	D9		X

Para el cálculo de los deciles se divide a la cantidad de observaciones por "10", siendo:

$$D_i = \frac{n}{10} \times i$$

Los procedimientos a aplicar, según sea la disposición de los datos, son los mismos que para los cuartiles.

3 - Percentiles

Como la palabra lo indica, los percentiles dividen al conjunto de observaciones en "100" partes iguales. A los percentiles los simbolizamos con Pi, donde el subíndice "i" toma los valores 1 a 99.

Existen entonces 99 percentiles: P1, P2, P3,.......,P98 y P99.

Para el cálculo, se procede de la misma manera que para los cuartiles o deciles.

4 - Consideraciones generales

Cualquiera de estas medidas nos permite encontrar posiciones de la variable que nos brindan información acerca de la distribución de los datos, pudiendo obtener valores particulares de la variable que acumulan hasta un porcentaje determinado de las observaciones.

Además, podemos deducir que:

$$Mna = Q_2 = D_5 = P_{50}$$

$$Q_1 = P_{25}$$

$$Q_3 = P_{75}$$

Ejercicio 8.1:

A partir de los datos del ejercicio 6.4 que muestra las alturas de un grupo de estudiantes de una Universidad, responda:
a) Cuál es la altura máxima del 10% de los estudiantes más bajos?
b) Cuánto miden como mínimo el 25% de los estudiantes más altos?
c) Entre qué rango de altura se encuentra el 50% de los estudiantes con altura intermedia?

Resolución del ejercicio 8.1:
a)

b)

c)

UNIDAD 9

MEDIDAS DE DISPERSION

Las medidas de tendencia central, como las restantes medidas de posición que desarrollamos hasta ahora, caracterizan la distribución de una variable.

Estas medidas nos dan idea respecto a cómo se posiciona algún valor de la variable, y en el caso de las medidas utilizadas como promedios, nos indican cuál es el centro de la distribución de una variable.

Esta información por sí sola no alcanza, ya que nos remite a un número que resume el comportamiento de los datos, pero ese número no refleja qué es lo que pasa con los valores de la variable que se distribuyen alrededor de él. Dos distribuciones pueden tener iguales medidas de tendencia central, por ejemplo, pueden coincidir sus medias aritméticas y sin embargo los datos pueden distribuirse de forma diferente en cada una de ellas.

Para dar una idea más clara del problema vamos a recurrir al siguiente ejemplo. Supongamos dos familias de dos personas cada una. El consumo diario de calorías de las personas de la primera familia es de 1500 y 4500, y el de las personas de la segunda familia es de 2900 y 3100. En ambas familias el consumo medio diario de calorías es 3000, por lo tanto cabe concluir que las dos tienen el mismo nivel calórico.

Esto es cierto, pero también es cierto que el promedio de calorías consumidas es mucho más representativo en la segunda familia que en la primera. Esta conclusión puede sacarse porque se dispone de las observaciones primarias y porque éstas son pocas y puede fácilmente deducirse la característica de variabilidad de los valores originales.

Cuando se concluye sobre el comportamiento de una variable, a partir del promedio de la misma, no debe prescindirse de la variabilidad o dispersión de los datos que dieron origen a ese promedio.

Hemos hablado de la dispersión como algo ligado al promedio, como una medida para conocer su representatividad. Sin embargo, la dispersión existe con independencia de los promedios.

Las variables no permanecen constantes al pasar de un elemento a otro de la población -de ahí que así se llamen-. Todas las personas no tienen la misma estatura, ni la misma edad, ni los mismos ingresos, etc.

Si no existiera la variabilidad, la estadística no tendría sentido. Puede decirse que la variabilidad es la esencia de la estadística, que persigue conocer las regularidades existentes en dicha variabilidad.

Cuando se quiere conocer la dispersión de una variable, lo que se intenta es obtener una medida, un número, que indique el mayor o menor grado de dispersión.

Así como vimos la existencia de varias medidas que apuntaban a describir el valor central de una distribución, veremos ahora que para describir la variabilidad de los datos, también contamos con muchas medidas, de las cuales, en este curso veremos las siguientes:

- *RANGO O AMPLITUD*
- *VARIANCIA Y DESVIO STANDARD*
- *COEFICIENTE DE VARIACION*

1 - Rango o Amplitud

Es la diferencia entre el máximo y el mínimo de los valores observados de la variable. Si bien brinda una primera idea acerca de la variabilidad, tiene el inconveniente de que sólo toma los valores extremos, descuidando el conjunto de valores intermedios. De esta forma, un valor extremo muy distanciado del grueso de los valores observados aumentaría considerablemente el rango.

Al Rango lo simbolizamos con **R** y se calcula como:

$$R=Máx(x_i)-Mín(x_i)$$

Supongamos dos variables, "X" e "Y", donde los valores observados para cada una de ellas son los siguientes:

X: 3 - 4 - 5 - 1 - 2 - 12

Y: 6 - 8 - 7 - 3 - 4 - 2 - 1 - 12 - 9 - 10

Como podemos observar, ambas variables tienen igual rango, ya que:

$$R(x)=Máx(x_i)-Mín(x_i)=12-1=11$$

$$R(y)=Máx(y_i)-Mín(y_i)=12-1=11$$

Esto no significa que las variables "X" e "Y" tengan similar grado de heterogeneidad. Vemos, que en la variable "Y", los datos se distribuyen en toda la amplitud, mientras que la variable "X", para un mismo rango, presenta una visible concentración en los valores menores, y el rango se vé afectado por un sólo valor, que es el extremo superior que toma la variable.

2 - *Variancia y desvío standard*

Una medida numérica de la dispersión la podemos obtener a partir de las diferencias de cada valor de la variable con respecto a su media aritmética. Simbólicamente estas diferencias, que se llaman **desviaciones**, se expresan así:

$$d_i=x_i-\bar{X}$$

Como la cantidad total de observaciones es "n", habrá por lo tanto "n" desviaciones.

Si los valores de la variable están muy concentrados alrededor de $\bar{X}$, entonces las desviaciones "di" serán pequeñas, y si están muy dispersos, las desviaciones serán grandes. Por lo tanto, un promedio de tales desviaciones da un solo número, que es expresivo del grado de concentración o dispersión de las observaciones. Si hay concentración, el promedio de las "di" será un número pequeño, y si hay dispersión, el promedio de las "di" será un número grande.

Hay, pues, que elegir un promedio para aplicarlo a las "n" desviaciones. Este podría ser la media aritmética, pero en este caso no es aplicable. En efecto, para obtener la media aritmética de "n" observaciones hay que sumarlas y dividirlas por "n". Pero la suma de las "di" es siempre igual a cero, porque se compensan las desviaciones positivas -valores Xi mayores a la media- con las desviaciones negativas -valores Xi menores a la media-.

Por lo tanto, la desviación media de todas las observaciones no puede ser una medida de la dispersión porque siempre es igual a "0".

Para evitar que la suma de las "di" sea igual a cero, se recurre a un artificio matemático que convierte a cada uno de los valores "di" en números positivos. Consiste en elevar al cuadrado a todos los valores "di".

Al hacer esto, perdemos la unidad de medida de la variable, ya que la misma ahora queda elevada al cuadrado.

Luego de hacer esta operación calculamos entonces el promedio de las "di" elevadas al cuadrado. A este resultado se lo denomina ***variancia*** *de la variable y se simboliza con la letra "V".*

Su expresión de cálculo es la siguiente:

$$V=\frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n}$$

Ahora bien, esta medida no tiene interpretación respecto a las desviaciones de la variable, ya que la unidad de medida de la variable no es la original. Nosotros necesitamos un número que venga expresado en la misma unidad de medida para poder interpretar la desviación de los valores con respecto a la media aritmética.

Para ello, realizamos la operación inversa a la de elevar al cuadrado, que es obtener la raíz cuadrada.

Cuando le aplicamos la raíz cuadrada a la variancia "V", obtenemos el ***desvío standard*** *que lo representamos con la letra S y lo calculamos de la siguiente forma:*

$$S=\sqrt{V}=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n}}$$

Como vemos, el desvío standard es el resultado de aplicar una raíz cuadrada. Con él obtenemos un número que puede tomarse con signo positivo y negativo. En el caso de la desviación standard se toma siempre el resultado positivo.

De esta forma, "S" es un número que varía desde cero en adelante. Cuando S = 0, todos los valores Xi son iguales, entonces no hay dispersión. A medida que S es mayor, la desviación también lo es. En consecuencia, la desviación standard es una medida de la dispersión de los valores de la variable con respecto a su media aritmética.

Para el cálculo del desvío standard, como lo hicimos para los promedios, consideraremos separadamente los distintos tipos de distribución en que pueden estar presentados los datos.

2.1 - Cálculo del S para datos simples o sin agrupar

La fórmula vista puede aplicarse directamente para calcular la desviación standard, pero es mucho más sencillo utilizar la que se deduce de ella, que es la siguiente:

$$S=\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}-\bar{X}^2}$$

Veamos con un ejemplo la forma de cálculo del desvío standard. Supongamos una variable "x" cualquiera que toma los siguientes valores:

X: 9 - 8 - 5 - 10 - 13 - 11 - 22 - 15 - 16 - 12 - 14 - 18

De la fórmula se deduce que es necesario obtener la suma de las Xi elevadas al cuadrado, la media aritmética elevada al cuadrado y la cantidad de observaciones "n". Entonces:

$$\sum_{i=1}^{n} x_i^2=2189$$

$$\bar{X}^2=162,56$$

$$n=12$$

Entonces, la desviación standard es:

$$S=\sqrt{\frac{2189}{12}-162,56}=4,46$$

Esto significa que los datos se encuentran en promedio alejados a izquierda o a derecha en 4,46 unidades de la media aritmética.

Recuerde siempre que la desviación standard viene dada en la misma unidad de medida que la variable.

2.2 - Cálculo del S Para datos agrupados

A) Sin intervalos de clase

La fórmula, en este caso, debe contemplar las frecuencias absolutas que presentan cada uno de los valores de la variable. Por lo tanto, cada valor de la variable ha de multiplicarse por la frecuencia correspondiente, que es lo que indica el número de veces que aparece repetido ese valor distinto. La fórmula será ahora:

$$S=\sqrt{\frac{\sum_{i=1}^{n} x_i^2 \times fa_i}{n}-\bar{X}^2}$$

Si retomamos el ejemplo que utilizamos en la UNIDAD 6 para distribuciones de frecuencia tenemos:

NUMERO DE CUARTOS POR HOGAR	FRECUENCIA ABSOLUTA	Xi^2	Xi^2 x fai
1	144	1	144
2	225	4	900
3	174	9	1566
4	88	16	1408
5	42	25	1050
	673		5068

Calculemos el desvío standard para la variable "X:número de cuartos por hogar".

$$\sum_{i=1}^{n} x_i^2 \times fa_i = 5068$$

$$\overline{X}^2=6,2$$

$$n=673$$

Entonces, la desviación standard es:

$$S=\sqrt{\frac{5068}{673}-6,2}=1,15$$

Esto significa, que en promedio, el desvío de las observaciones es de 1,15 cuartos respecto del valor de la media aritmética.

B) Con intervalos de clase

Si los datos se encuentran agrupados en intervalos de clase, al no contar con el valor exacto de la variable, ya que tenemos frecuencias para un rango de la variable, el cálculo del desvío standard no es exacto.

Una buena aproximación, al igual que cuando calculábamos la media aritmética, se logra tomando para cada intervalo, como valor observado de la variable, el punto medio de cada intervalo -marca de clase-.

La expresión de cálculo en este caso es la misma que para datos sin agrupar, dónde las "Xi" son reemplazadas por "Xi'".

Entonces:

$$S=\sqrt{\frac{\sum_{i=1}^{n} x'^{2}_{i}\times fa_i}{n}-\overline{X}^2}$$

Observemos la tabla de frecuencias obtenida para el ejemplo desarrollado en la UNIDAD 6, donde se presentaba la distribución por edad de los empleados de una empresa.

Xi	Xi'	fai	Xi'^2	$Xi'^2 \times fai$
[20-24)	22	1	484	484
[24-28)	26	3	676	2028
[28-32)	30	9	900	8100
[32-36)	34	30	1156	34680
[36-40)	38	60	1444	86640
[40-44)	42	52	1764	91728
[44-48)	46	35	2116	74060
[48-52)	50	10	2500	25000
				322720

Se tiene pues, que la desviación standard es:

$$S=\sqrt{\frac{322720}{200}-1585,6}=5,29\,años$$

Esto significa que las edades están alejadas a izquierda y a derecha, en promedio en 5,29 años respecto al valor de la media aritmética que es, según vimos, de 39,82 años.

3 - Coeficiente de variación

Con gran frecuencia se presenta el problema de comparar la dispersión de dos o más distribuciones. Por ejemplo, se quiere saber si la variabilidad de la temperatura de un lugar es mayor o menor que la de otro, si los salarios varían más en un grupo de obreros que en otro, si los ingresos están más equitativamente distribuídos en un país que en otro, etc.

En todos estos casos de comparación de dispersiones, ésta puede hacerse utilizando la desviación standard, si las variables comparadas tienen sus medias aritméticas iguales o aproximadamente iguales y si dichas variables vienen expresadas en la misma unidad de medida.

Si estas condiciones no ocurren, que suele ser lo más habitual, la comparación debe hacerse mediante el denominado coeficiente de variación, que se simboliza con "CV" y se define así:

$$CV=\frac{S}{\bar{X}}\times 100$$

El "CV" no es otra cosa que la desviación standard expresada como porcentaje de la media aritmética. Por lo tanto, el "CV" no viene expresado en la unidad de medida de la variable, ya que al hacer su cálculo ésta se pierde al simplificarse en el numerador y denominador.

Veamos la utilización del "CV" con un ejemplo. Supongamos que queremos comparar la dispersión entre las edades de los alumnos de una carrera universitaria que cursan en el turno mañana y en el turno noche. Las observaciones arrojan como resultado:

TURNO MANANA: $\bar{X}$ = 23 años S = 2,4 años

TURNO NOCHE: $\bar{X}$ = 25 años S = 3,9 años

Si tomamos las desviaciones standard para comparar la dispersión de las edades no llegamos a ninguna conclusión correcta, debido a que los dos grupos no tienen la misma edad promedio.

Calculando los "CV" para ambas distribuciones tenemos:

$$CV_{(TM)} = \frac{2,4 años}{23 años} \times 100 = 10,4\%$$

$$CV_{(TN)} = \frac{3,9 años}{25 años} \times 100 = 15,6\%$$

Ahora, puede afirmarse, que las edades en el turno mañana tienen menor dispersión que en el turno noche, ya que su coeficiente de variación es menor.

Siempre, a mayor valor del coeficiente de variación, mayor es la heterogeneidad en los valores de la variable; y a menor valor del coeficiente de dispersión, mayor es la homogeneidad de los valores de la variable.

Como conclusión podemos decir que el coeficiente de variación depende del valor de "S" y en mayor medida del valor de "$\bar{X}$". Cuando el valor de la "$\bar{X}$" es cero o muy próximo a cero, el coeficiente de variación pierde significado, ya que puede dar valores muy grandes que no necesariamente implican una gran dispersión de los datos.

Ejercicio 9.1:

Calcule el rango y el desvío standard para la variable considerada *en el ejercicio 7.2. Interprete los resultados.*

Resolución del ejercicio 9.1:

Ejercicio 9.2:

El ejercicio 7.1 muestra los sueldos de un grupo de empleados de una empresa. Calcule el desvío standard de la variable e interprete su resultado.

Resolución del ejercicio 9.2:

Ejercicio 9.3:

Para comparar la variabilidad en las edades y los ingresos de un grupo de profesores universitarios de dos universidades privadas se relevaron los siguientes datos:

Universidad A			Universidad B		
Profesor	Edad	Ingreso	Profesor	Edad	Ingreso
1	28	170	1	33	190
2	40	220	2	36	260
3	31	350	3	55	540
4	48	350	4	36	290
5	44	330	5	50	420
6	41	370	6	42	370

Calcule el coeficiente de variación para las variables edad e ingreso en ambas poblaciones y comente los resultados.

Resolución del ejercicio 9.3:

Ejercicio 9.4:

Los automóviles de la marca Fiat tuvieron durante el mes de julio de 1992 un precio promedio de 14814 u$s con un desvío standard de 1043 u$s. Para los de marca Peugeot el precio promedio para el mismo período fue de 21662 u$s con un desvío standard de 1895 u$s. Cuál de las dos marcas tiene automóviles con precios más homogéneos?

Resolución del ejercicio 9.4:

UNIDAD 10

ESTANDARIZACION DE VARIABLES

Muchas veces, a una misma unidad de análisis le medimos más de una variable, o también puede ocurrir que midamos la misma variable a unidades de análisis provenientes de distintas poblaciones.

La medición de una variable implica que la observación venga expresada en alguna unidad de medida. Estas unidades de medida suelen no coincidir de observación a observación, cuando se trata de una misma variable, o directamente, puede tratarse de variables distintas observadas en diferentes unidades de medida.

Por ejemplo, si la variable en estudio es el tiempo, puede ocurrir que las observaciones se realicen en años, días, minutos, etc., difiriendo las unidades de una observación a otra. Este caso no nos presentaría ningún problema ya que podemos convertir los valores y llevarlos a todos a una misma unidad de medida.

Si a una misma unidad elemental, le medimos más de una variable, por ejemplo, si a una persona le medimos peso y altura, estas variables seguramente vendrán dadas en Kg y cm. Como podemos observar, estas medidas están expresadas en unidades de medida distintas.

Si quisiéramos hacer alguna comparación entre la altura y el peso de las personas, tendríamos la dificultad de querer comparar características no comparables, ya que se refieren a dimensiones diferentes.

Lo que nos proponemos aquí, es encontrar algún método que nos permita prescindir de la unidad de medida original de la variable, y encontrar una nueva unidad de medida que sea "standard", cualquiera sea la variable en estudio.

La estandarización de la unidad de medida se obtiene a partir de la tendencia central y de la variabilidad que presenta la variable.

El procedimiento de estandarización de variables es particularmente importante cuando se comparan características que pertenecen a dimensiones cualitativamente diferentes.

Hay muchos y diferentes métodos de estandarización de unidades de medida. En este curso veremos el más común y usado de ellos, que es la escala "Z".

La escala "Z" es una nueva escala de medida, obtenida a partir de *la distribución de la variable "X", que logra despojar a las observaciones de la variable "X" de su unidad de medida original. La escala "Z" se construye a partir de una transformación en la que interviene la media aritmética ($\bar{X}$) y el desvío standard (S) de la variable "X".*

Así, cualquier valor observado de la variable "X" puede transformarse en un valor de "Z" de la siguiente forma:

Sea "Xi" una observación de la variable "X" en una unidad de análisis cualquiera, entonces:

$$Z=\frac{x_i-\bar{X}}{S}$$

De esta forma, la variable "X" que estaba expresada en una unidad de medida cualquiera, se transforma en "Z" perdiendo la unidad de medida original.

Siempre, la unidad de "Z" es igual a un desvío de la variable "X", cualquiera sea ésta.

Vamos a desarrollar un ejemplo, que nos permita aclarar este concepto.

*Retomando las alturas y el peso de las personas, supongamos que le pregunte a cualquier persona de un grupo si se considera **más alto que flaco** o si es **más petiso que gordo**.*

Evidentemente, esta es una pregunta difícil de responder, ya que ellos pueden conocer su altura y su peso, pero desconocen cómo comparar estas medidas, ya que una de ellas está expresada en cm. y la otra en kg.

Además, para tener una respuesta a esta cuestión, debemos referirnos a una altura promedio y a un peso promedio de los individuos, que nos den un marco de referencia del comportamiento de las variables que queremos comparar en el conjunto de las personas.

Supongamos que la altura media de las personas adultas es de 170 cm., con un desvío de 3 cm. y que el peso medio es de 78 kg. con un desvío de 9 kg.

Uno de ellos se interesa en saber si realmente es más petiso que gordo y dice que su estatura es de 173 cm. y su peso es de 96 kg.

Para responderle, necesito cambiar las unidades de medida para que sean las mismas, y poder hacer la comparación entre las alturas y el peso. Lo que hago, entonces, es transformar la variable "X:altura" y "X:peso" en dos variables "Z" que tendrán la misma unidad de medida.

Calculamos primero a Z para la altura haciendo:

$$Z_{(alt)} = \frac{173\,cm - 170\,cm}{3\,cm} = \frac{3\,cm}{3\,cm} = 1$$

Luego, Z para el peso será:

$$Z_{(peso)} = \frac{96\,kg - 78\,kg}{9\,kg} = \frac{18\,kg}{9\,kg} = 2$$

Ahora sí, puedo comparar los valores Z(alt) y Z(peso), y responder:

$$Z_{(alt)} = 1 < Z_{(peso)} = 2$$

Por lo tanto, la respuesta es, el interesado es más pesado que alto.

Z es una variable que puede tomar valores positivos o negativos. Cualquier valor de Z negativo implica que dicho valor es menor que un valor de Z positivo.

Ejercicio 10.1:

La Encuesta Industrial del INDEC releva información sobre el volumen físico de la producción en los establecimientos industriales. Las empresas productoras de pinturas declaran su producción indistintamente en kilogramos o en litros, según sea el tipo de pintura. Se quiere determinar si la empresa "A" es más importante en la producción de pinturas por litro que por Kilogramo, sabiendo que su producción mensual es de 10345 lt. y de 21338 kg.

Por otra parte, se sabe que las empresas del sector producen en promedio 9550 l. por mes con un desvío standard de 925 l. y 24340 kg. con un desvío de 4820 kg.

Qué puede decir acerca de la producción de la empresa "A" ?.

Resolución del ejercicio 10.1:

Ejercicio 10.2:

Una persona se presenta en dos universidades extranjeras para solicitar su ingreso a las mismas. En ambas le toman un test de inteligencia cuyos resultados se miden en escalas distintas. Los puntajes en cada uno de los test fueron diseñados de manera arbitraria, pero se sabe que en la universidad "A" el promedio de puntaje de los test realizados es de 90 puntos con un desvío de 12 puntos, mientras que en la universidad "B" el puntaje promedio es de 45 puntos con un desvío de 4 puntos. La persona en cuestión sacó 92 puntos en el test de la universidad "A" y 48 puntos en el test de la universidad "B". En cuál de los dos test le fue mejor? Justifique su respuesta.

Resolución del ejercicio 10.2:

UNIDAD 11

SERIES CRONOLOGICAS

Casi todas las ciencias prestan considerable atención a los fenómenos que permiten observar variaciones cuantitativas a lo largo del tiempo. Dichos datos constituyen *series cronológicas*. Por lo tanto podemos definir como serie cronológica o serie de tiempo a los datos de una variable cuantitativa a través del tiempo, es decir, los valores sucesivos que toma una variable en intervalos mensuales, anuales o de cualquier otro período de tiempo. Son ejemplos de series de tiempo los datos de ventas semanales de una empresa, datos anuales de producción, de migración, etc.

Al graficar en un eje de coordenadas los datos de una serie cronológica:

► el tiempo se simboliza con "t" y sus valores se representan en la abscisa.

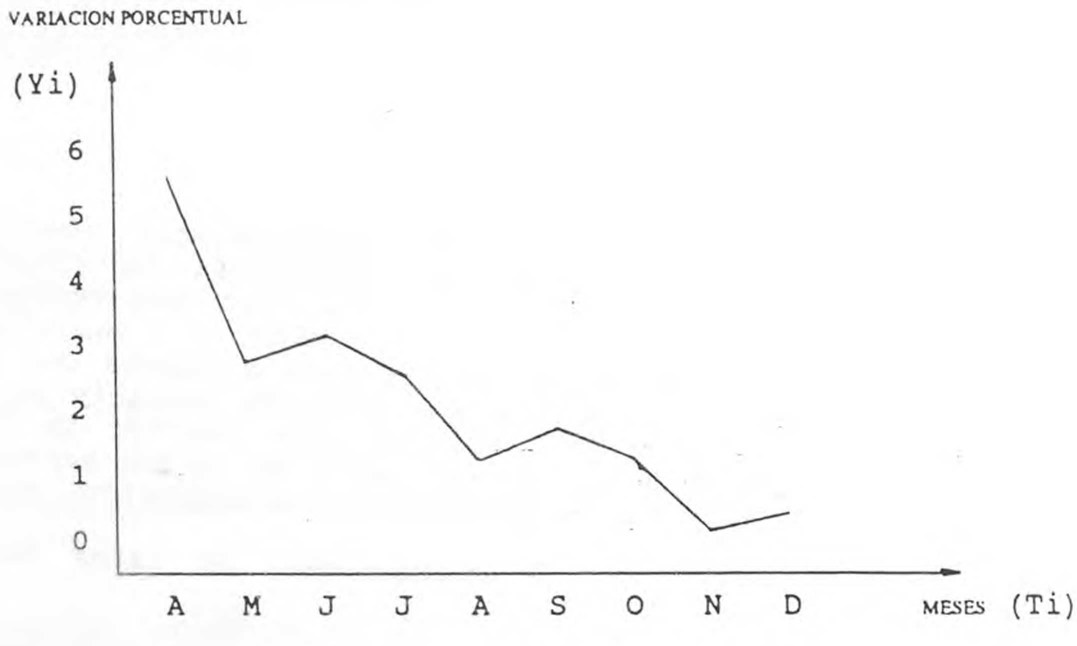
► la variable cuantitativa referida al tiempo se simboliza con "y" y sus valores se representan en la ordenada.

Representaremos la siguiente serie cronológica:

INDICE DE PRECIOS MINORISTAS

VARIACION PORCENTUAL MENSUAL

PERIODO (Ti)	VARIACION % (Yi)
abril 1991	5,5
mayo 1991	2,8
junio 1991	3,1
julio 1991	2,6
agosto 1991	1,3
setiembre 1991	1,8
octubre 1991	1,4
noviembre 1991	0,4
diciembre 1991	0,6



Cada punto del gráfico representa una dupla de valores (Yi,Ti). Un análisis estadístico de los datos de una serie cronológica permite:

- *determinar la evolución o el comportamiento de la característica en estudio durante el período de referencia o período de tiempo al que están referidos los datos o valores observados.*

- *realizar pronósticos sobre la evolución de la característica en estudio para períodos de tiempo futuros, usando la tendencia.*

En una serie cronológica estamos estudiando dos variables: Ti (tiempo) e Yi (ventas, producción, etc.).

Yi: es una variable cuantitativa.

Ti: es una variable cualitativa o atributo (meses, bimestres, años, etc.) transformable a variable cuantitativa. En la práctica, para analizar una serie cronológica se cuantifica a la variable Ti, asignándole números correlativos a cada valor de Ti.

En el ejemplo, podríamos cuantificar los meses dando valores correlativos a los meses, siendo abril 1991=1, mayo 1991=2, etc. Se podría haber comenzado con 0 para abril 1991, pero esto es arbitrario. Generalmente se comienza con el valor 1 porque de esa forma, rápidamente se puede saber con cuántos datos se está trabajando.

1 - Componentes de una serie cronológica

Una serie cronológica que representa una actividad particular de una organización, es el resultado de interacciones de numerosos tipos de fuerzas cambiantes.

Estas fuerzas pueden ser económicas, políticas e influencias sociales, lo mismo que las fuerzas de la naturaleza. Las fuerzas son usualmente investigadas para tomar decisiones después de que los movimientos de una serie de tiempo se descomponen en las siguientes componentes:

- *TENDENCIA SECULAR O TENDENCIA*
- *ESTACIONALIDAD*
- *VARIACION CICLICA*
- *VARIACION ALEATORIA O IRREGULAR*

1.1 - Tendencia secular

La tendencia secular o tendencia señala la dirección del movimiento de una serie de tiempo sobre un largo período de tiempo. Puede ser un movimiento ascendente, descendente o constante. Cuando se muestra gráficamente, es usualmente representada por una línea recta o por una curva suave.

Por ejemplo, para la economía la tendencia de una serie significa el desarrollo, el estancamiento o depresión de su marcha.

1.2 - Variación estacional

Las variaciones estacionales se refieren a las idénticas, o casi idénticas variaciones que una serie de tiempo parece seguir durante los meses correspondientes de años sucesivos. Estos movimientos se deben a sucesos recurrentes que se repiten anualmente, como por ejemplo, el incremento en el consumo de gas domiciliario en el invierno, o el incremento del consumo de helados en el verano.

1.3 - Variaciones cíclicas

Las variaciones o fluctuaciones cíclicas, también llamadas ciclos, son movimientos periódicos que se observan en las series cronológicas, pero no dentro del año, sino en períodos mayores.

Se refieren a las oscilaciones de larga duración alrededor de la tendencia, o sea que indican las expansiones y las contracciones de las actividades de los negocios en torno a la línea de tendencia. La duración de cada ciclo no es fija.

Las fuerzas que son responsables de las fluctuaciones cíclicas son numerosas y complejas, pero son principalmente factores económicos.

1.4 - Variaciones aleatorias o irregulares

Las variaciones aleatorias o los movimientos irregulares representan todos los tipos de movimientos de una serie de tiempo que no son debidos a la tendencia, variaciones estacionales ni variaciones cíclicas. Las fuerzas que causan las irregularidades en las actividades de los negocios son numerosas y son de naturaleza aleatoria, por ejemplo, huelgas, guerras, inundaciones, etc.

2 - Valor de la serie cronológica

Un valor de la serie de tiempo (Yi) puede calcularse como el producto o la suma de las cuatro componentes mencionadas. Es decir, existen modelos matemáticos -multiplicativos o aditivos-, que permiten calcular el valor de una serie cronológica.

Siendo,

Yi: valor real u observado en el período Ti.
T: valor observado de la tendencia de Yi.
E: variación estacional de Yi.
C: variación cíclica de Yi.
A: variación aleatoria de Yi.

Podemos expresar en símbolos el valor de una serie de la siguiente manera:

- Modelo multiplicativo:

$$Y_i = T \times E \times C \times A$$

- Modelo aditivo:

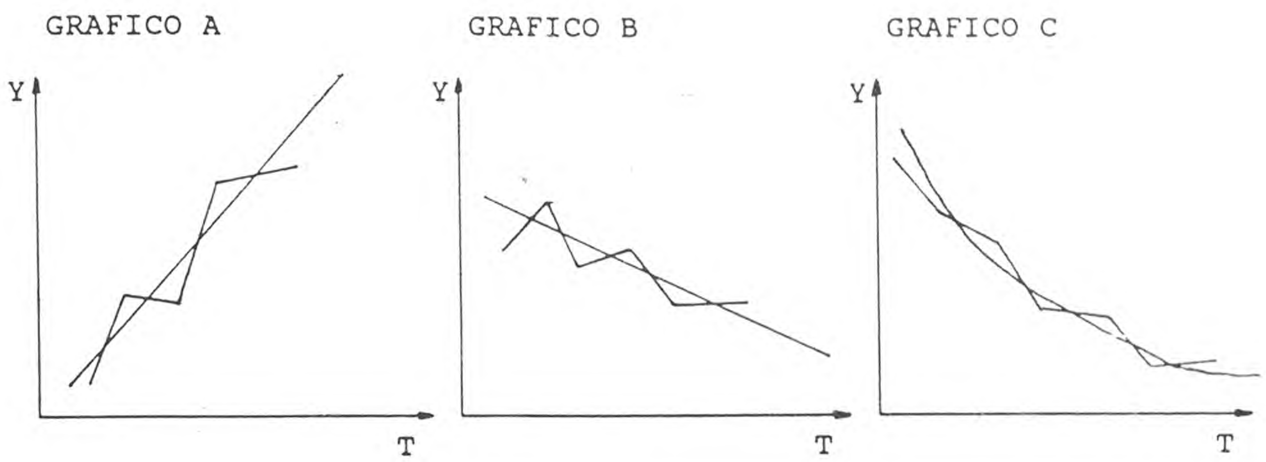
$$Y_i = T + E + C + A$$

3 - Tendencia

Al representar datos de una serie cronológica, se obtienen líneas quebradas. Sin embargo, es posible observar una tendencia en la evolución de los datos a través del tiempo.

La representación gráfica de esa tendencia puede ser una línea recta o curva.

Veamos los siguientes gráficos:



El GRAFICO A tiene una tendencia lineal positiva: a un incremento del tiempo (T) se produce un incremento de la variable cuantitativa (Y).

El GRAFICO B tiene una tendencia lineal negativa: a un incremento del tiempo (T) se produce un decremento de la variable cuantitativa (Y).

El GRAFICO C tiene una tendencia curvilínea.

En este curso sólo nos referiremos a la tendencia lineal.

3.1 - Tendencia lineal

La tendencia lineal se representa en la gráfica como una línea recta, por lo tanto, puede ser expresada mediante la siguiente ecuación:

$$Y=a+b\times T$$

donde:

Y: variable cuantitativa relacionada con el tiempo (variable dependiente del tiempo).

T: es el tiempo (variable independiente).

a: es la ordenada al origen (valor que toma Y cuando T vale 0).

b: es la pendiente (indica cuanto crece o decrece Y cuando T crece en una unidad).

3.2 - Método para determinar la tendencia

Determinar la tendencia significa encontrar la línea recta y la ecuación asociada a la misma.

En este curso presentaremos sólo el método gráfico, pero existen también métodos matemáticos complejos que permiten establecer con mayor precisión la tendencia de una serie cronológica.

El método consiste en:

- Marcar la serie cronológica sobre una gráfica.

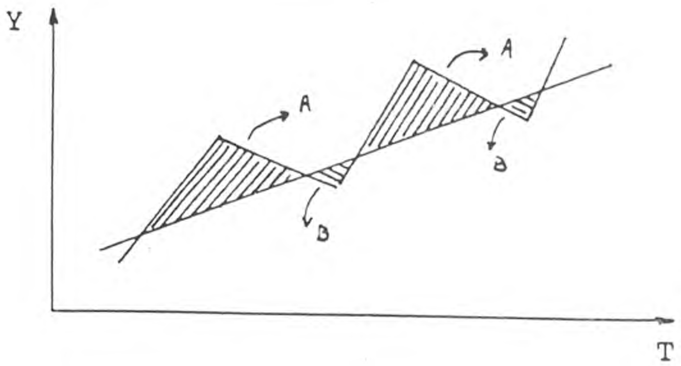
- Examinar cuidadosamente la dirección de la tendencia, basados en la información marcada.

- Dibujar una línea recta que sea el mejor ajuste a los datos de acuerdo con el juicio personal del dibujante.

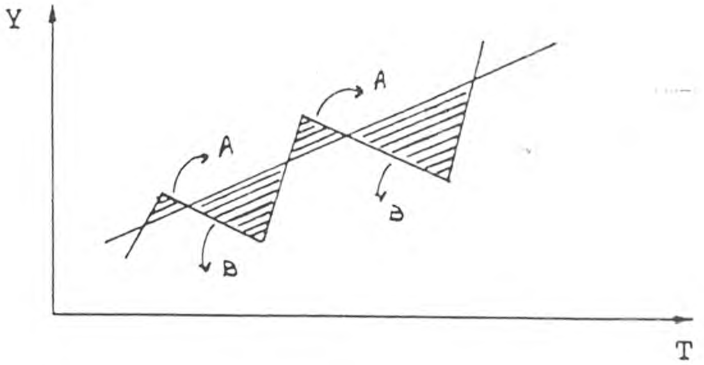
Este método es muy simple y fácil de aplicar pero es demasiado subjetivo. Líneas dibujadas por diferentes personas para la misma información pueden tener distintas localizaciones sobre la gráfica.

Una forma de reducir la subjetividad y lograr un mejor ajuste es trazar la recta tratando de compensar las áreas que quedan determinadas entre la línea de tendencia y la línea quebrada (gráfico de la serie), por encima y por debajo de la recta.

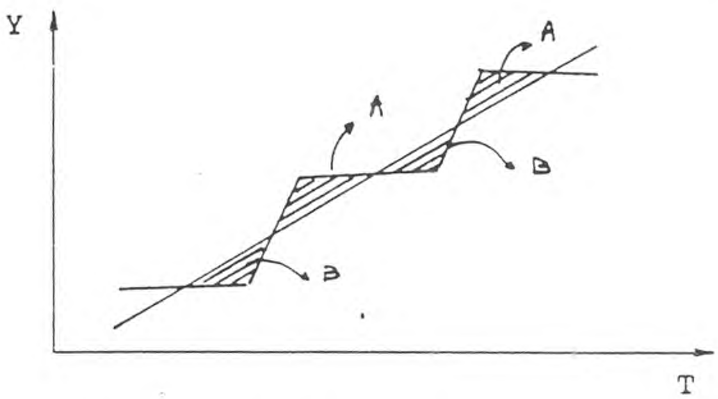
Veamos algunos ejemplos de trazado gráfico de líneas de tendencia:



área A > área B



área A < área B



área A = área B

De las 3 rectas de tendencia marcadas, la tercera es la que da un mejor ajuste, puesto que es la que más compensa las distancias entre los datos y la recta, por encima y por debajo de ella.

Veamos ahora como se procede para determinar el valor de la tendencia en un período de tiempo Ti cualquiera:

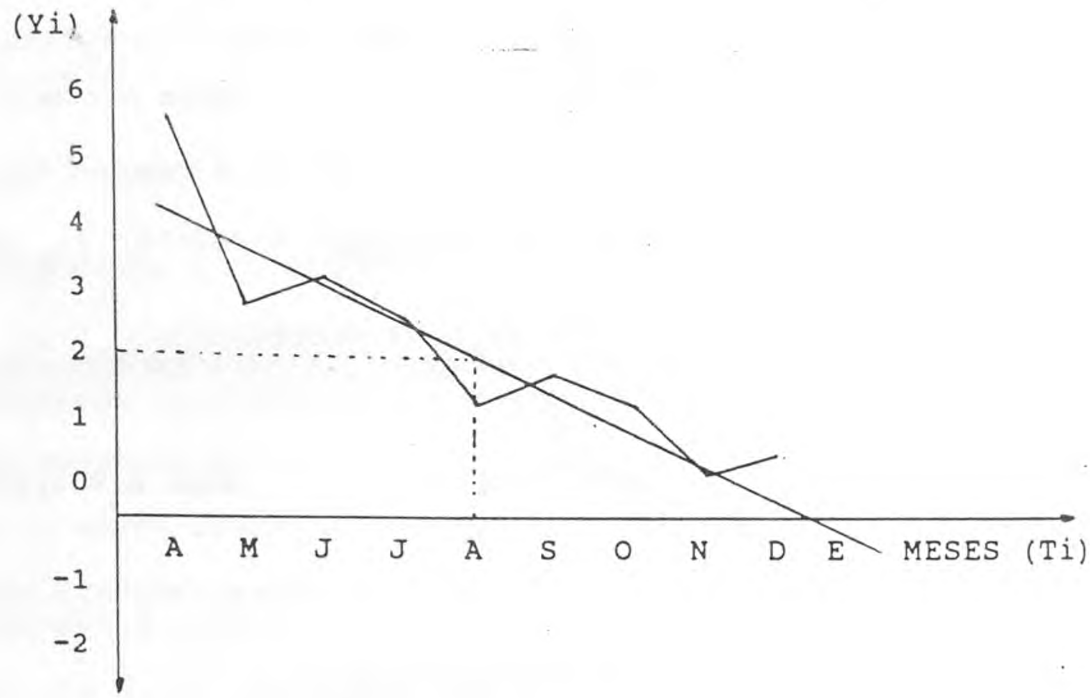
- *para el período Ti elegido, trazamos una perpendicular al eje de abscisas hasta tocar la línea de tendencia dibujada.*

- *cuando tocamos la línea de tendencia nos desplazamos hasta el eje de ordenadas dónde encontramos un valor Yi que es el valor estimado correspondiente a la tendencia.*

► *para determinar o pronosticar los valores de tendencia para otros períodos (pasados o futuros), prolongamos la línea de tendencia hasta el período deseado y leemos sobre el eje de ordenadas el valor de tendencia correspondiente.*

Analicemos el ejemplo presentado en la introducción, graficando la línea de tendencia y calculando su valor para el período "agosto 1991". Además hagamos un pronóstico del valor de la tendencia para el período "enero 1992".

VARIACION PORCENTUAL



Para el período "agosto 1991", según observamos en el gráfico el valor de la tendencia es T = 2,0.

El pronóstico del valor de la tendencia para el período "enero 1992" es T = -0,3.

Hay que tener presente que el pronóstico que se efectua sobre el comportamiento de la tendencia para períodos externos al período de referencia se va a cumplir si se mantienen las mismas condiciones que existían para el período de referencia.

Ejercicio 11.1:

Los siguientes factores afectan los movimientos de una serie cronológica. Responda para cada uno de ellos con qué movimientos de una serie los asociaría:
a) Un retroceso en las ventas de acero a causa de una huelga de trabajadores del sector.
b) Una declinación en las ventas de helados en el mes de junio.
c) Un aumento de la cantidad de alumnos inscriptos en el nivel primario.
d) Una disminución de la mano de obra ocupada a partir del mes de abril en la provincia de Mendoza.

Resolución del ejercicio 11.1:
a)
b)
c)
d)

Ejercicio 11.2:

Se dispone de datos sobre las ventas mensuales de libros de texto primarios y secundarios de una editorial durante los últimos 10 años.
a) Cree usted, que estos datos muestran movimientos estacionales? Cómo son?
b) Cree usted, que los datos acusan una tendencia definida? Por qué?
c) Qué factores pueden causar variaciones irregulares en los datos?
d) Suponga que usted es el dueño de la editorial y debe planificar el volumen de textos a editar para el año próximo. Le sirve esta información? Qué serie cronológica cree que es la adecuada?

Resolución del ejercicio 11.2:
a)
b)
c)
d)

Ejercicio 11.3:

Dada la siguiente tabla que muestra los datos de las ventas de una empresa que comercializa computadoras:

Período	Venta en unidades
1985	300
1986	320
1987	400
1988	550
1989	500
1990	560
1991	590

a) Grafique los datos
b) Determine gráficamente la tendencia
c) Cuál es el valor estimado de la tendencia para el año 1988?
d) Cuál es el volumen de ventas que usted pronosticaría para 1992?
e) Con los datos presentados, se arriesgaría a pronosticar las ventas para el año 2000?

Resolución del ejercicio 11.3:

a) y b)

c)

d)

e)

Ejercicio 11.4:

La siguiente tabla muestra la cantidad de pasajeros transportados por Subterráneos de Buenos Aires:

Período	Usuarios en millones
1980	201
1981	192
1982	182
1983	191
1984	191
1985	183
1986	195
1987	194

a) Grafique los datos en un eje de coordenadas.
b) Determine gráficamente la tendencia.
c) Cuál es el valor de la tendencia para 1984?
d) Qué cantidad de usuarios pronosticaría para el año 1988?
e) Compare la cifra pronosticada con la cifra real para el año 1988 que fue de 178 millones de pasajeros.

Resolución del ejercicio 11.4:

a) y b)

c)

d)

e)

UNIDAD 12

NUMEROS INDICES

Uno de los objetivos de la Estadística Descriptiva, es lograr concentrar las múltiples medidas que obtiene cuando analiza el comportamiento de un grupo de datos.

Los números índices agrupan información en una sola expresión numérica.

Un número índice es un indicador de la tendencia central de un conjunto de datos. Plantea una comparación, ya sea en el *tiempo* o en el *espacio*, respecto de un punto de referencia denominado *base del índice*.

Generalmente a un número índice se lo expresa como porcentaje. Los números índices más frecuentes son los de precios, de cantidades y de valores.

- ***Indice de precios:*** es aquel que refleja la variación de los precios de un conjunto de bienes o servicios, entre dos momentos en el tiempo o dos lugares en el espacio.

- ***Indice de cantidades:*** es aquel que refleja la variación de las cantidades de un conjunto de bienes o servicios, entre dos momentos en el tiempo o dos puntos en el espacio.

- ***Indice de valores:*** refleja la variación en el valor total (precio multiplicado por cantidad) de un conjunto de bienes o servicios, entre dos momentos en el tiempo o dos puntos en el espacio.

1 - Base de un número índice

Es el momento, período o punto con respecto al cual se establece la comparación.

Generalmente al índice en la base se le asigna el valor 100.

Por ejemplo, la actual base del Indice de Precios al Consumidor que calcula el INDEC es el año 1988, que se expresa base 1988 = 100.

Es decir, el valor del Indice de Precios al consumidor correspondiente a cualquier período se compara con el valor 100 correspondiente a la base, y según sea la relación entre esos dos valores, se establece cuál fue la variación de los precios entre esos dos períodos.

2 - Indices simples

Cuando se trata de analizar la variación en el precio de un artículo a través del tiempo, no es necesario un índice especial, basta con expresar la variación en términos porcentuales. En este caso hablamos de índices simples.

Supongamos la siguiente serie de precios de un bien determinado:

PERIODO años	PRECIO DEL BIEN en $
1987	20
1988	28
1989	40
1990	50

Si tomamos como período de referencia o año base al año 1987, el Indice de precios del bien en el período "j" con base en el período 1987 = 0 , siendo "j" cualquiera de los años de la serie, se calcula como sigue:

$$I_0^j = I_{1987}^j = \frac{precio.en.el.año.j}{precio.en.el.año.1987}$$

donde el subíndice "0" indica el año base, en este caso el año 1987, y el supraíndice "j" indica el año para el cual se efectúa la comparación.

Por ejemplo, el índice de precio del bien en 1989, con base en 1987, es:

$$I_0^j = I_{1987}^{1989} = \frac{40\$}{20\$} = 2$$

Como el numerador y el denominador están expresados en $, se simplifican. Por lo tanto, el número índice es un indicador sin unidad de medida.

El número índice puede también expresarse en porcentaje. Para ello, multiplicamos el resultado por 100:

$$I_{1987}^{1989} \times 100 = \frac{40\$}{20\$} \times 100 = 200$$

El cociente entre el precio del bien en el año "j" y el precio del bien en el año base se llama *índice simple de precio* o también relativo de precio o precio relativo y se expresa:

$$I_0^j = \frac{P_j}{P_0}$$

En el ejemplo, el resultado 2 significa que el precio del bien en el año 1989 es *2 veces* el precio del bien en el año 1987.

Tomando el índice en porcentaje, decimos que el precio del bien en el año 1989 fue un ***100% mayor*** que en el año 1987.

Si calculamos los índices para todos los años de la serie, tenemos:

PERIODO años	PRECIO DEL BIEN en $	INDICE SIMPLE base 1987=100	VARIACION PORCENTUAL
1987	20	100	-
1988	28	140	40
1989	40	200	100
1990	50	250	150

En la tabla, los índices simples están calculados en porcentaje. Por ejemplo, si tomamos el índice correspondiente al año 1990 con base 1987 = 100 podemos decir que el precio del bien tuvo un aumento del 50%.

NOTA: los índices simples que hemos visto para precios se pueden generalizar para la relación entre cualquier magnitud tomada en dos momentos o lugares diferentes.

2.1 - Propiedades de los índices simples

A) Identidad: se dice que un índice tiene la propiedad de identidad cuando toma el valor 100 en el año base. Esta propiedad puede traducirse en la siguiente igualdad:

$$I_0^0=\frac{P_0}{P_0}=1$$

o expresándolo en porcentaje:

$$I_0^0=\frac{P_0}{P_0}\times 100=100$$

B) Reversibilidad temporal: *analizaremos esta propiedad trabajando con el ejemplo de precios. Tomando como base el período 1987, el precio relativo del período 1989 es:*

$$I_{1987}^{1989}=\frac{P_{1989}}{P_{1987}}=\frac{40\$}{20\$}=2$$

Luego, el precio en 1989 es 2 veces el precio en 1987. Expresado en porcentaje, el precio en 1989 fue un 100% mayor que en 1987.

Si invertimos el índice, tomando como base el período 1989, el precio relativo en el período 1987 es:

$$I_{1989}^{1987}=\frac{P_{1987}}{P_{1989}}=\frac{20\$}{40\$}=0,5$$

Luego, el precio en 1987 es el 0,5 (la mitad) del precio en 1989. Expresado en porcentaje, el precio en 1987 es el 50% del precio en 1989.

Se puede deducir entonces que:

- *el producto de estos dos índices es igual a 1:*

$$I_{1987}^{1989}\times I_{1989}^{1987}=2\times 0,5=1$$

Generalizando esta propiedad para cualquiera sea el período "j" de referencia y siendo la base el período "0", tenemos:

$$I_0^j \times I_j^0 = 1$$

▸ los dos índices son recíprocos entre sí. Esto significa que un índice es igual a la inversa del otro:

$$I_{1989}^{1987} = \frac{1}{I_{1987}^{1989}} \rightarrow 0,5 = \frac{1}{2}$$

Si generalizamos para cualquier índice, tenemos:

$$I_j^0 = \frac{1}{I_0^j}$$

La propiedad de reversibilidad temporal es válida para todos los índices simples, y expresa que los cambios relativos de año a año que aparecen en una serie de números índices simples, no dependen del año base seleccionado.

Un número índice que cumpla con la propiedad de reversibilidad temporal estará menos afectado por la elección del año base que otro que no la cumpla.

C) Circularidad: si el precio de un producto es de 10 \$ en el primer período, 12 \$ en el segundo período y 18 \$ en el tercero; y llamamos a los tres períodos con 0, t y j respectivamente:

▸ tomando como base el período 0, el precio relativo en el período t es :

$$I_0^t = \frac{P_t}{P_0} = \frac{12\$}{10\$} = 1,2$$

▸ tomando como base el período t, el precio relativo en el período j es :

$$I_t^j = \frac{P_j}{P_t} = \frac{18\$}{12\$} = 1,5$$

▸ Tomando como base el período 0, el precio relativo en el período j es:

$$I_0^j=\frac{P_j}{P_0}=\frac{18\$}{10\$}=1,8$$

De los índices calculados se deduce que:

$$I_0^t \times I_t^j = I_0^j$$

$$\frac{P_t}{P_0}\times\frac{P_j}{P_t}=\frac{P_j}{P_0}$$

$$1,2\times 1,5=1,8$$

2.2 - Aplicación de las propiedades

A) Cambio de base: *si se conoce el índice simple de una magnitud con base 0, es posible calcular ese mismo índice sobre otra base t como un cociente entre el primer índice y el índice t con base 0. Por la propiedad de circularidad tenemos:*

$$I_t^j=\frac{I_0^j}{I_0^t}$$

Por ejemplo, si tomamos los precios del ejemplo presentado anteriormente tenemos:

$$I_{1987}^{1989}=2$$

$$I_{1987}^{1990}=2,5$$

Si queremos calcular el índice de precios de 1990 con base 1989, podemos hacer:

$$I_{1989}^{1990}=\frac{I_{1987}^{1990}}{I_{1987}^{1989}}=\frac{2,5}{2}=1,25$$

Entonces, el precio del bien aumentó un 25% en 1990 respecto al precio del mismo bien en el año 1989.

B) Encadenamiento de índices sucesivos: la propiedad de circularidad puede generalizarse para encadenar varios índices simples:

$$I_0^3 = I_0^1 \times I_1^2 \times I_2^3$$

Al encadenar índices, cada uno de los elementos de la cadena debe ser calculado con mucha precisión, ya que los errores de redondeo se van acumulando.

3 - Indices sintéticos

El problema de los números índice surge cuando se desea averiguar las variaciones de precios, cantidades, valores, etc., de un conjunto de artículos.

Este es el caso del índice de precios al consumidor, de costos de la construcción, etc.

En estos casos interesa conocer la variación de precios del *conjunto de bienes* y no de cada bien en particular.

Consideremos el siguiente conjunto de bienes:

BIENES	PRECIO (en unidades monetarias)	
	1990	1991
A (metro)	50	70
B (kilo)	30	40
C (litro)	60	90
D (tonelada)	180	250
TOTAL	320	450

Podríamos sumar los precios de todos los bienes para ambos períodos y establecer la variación entre ambos agregados. Así calcularíamos un índice agregativo simple:

$$I_{1990}^{1991}=\frac{450u.m.}{320u.m.}=1,406$$

Diríamos que el conjunto de precios ha variado un 40,6 % durante el período 1990-1991.

Generalizando, un índice agregativo simple de precios es:

$$I_0^j=\frac{\sum_{i=1}^{n} P_{ij}}{\sum_{i=1}^{n} P_{io}}$$

o expresado en porcentaje:

$$I_0^j=\frac{\sum_{i=1}^{n} P_{ij}}{\sum_{i=1}^{n} P_{io}}\times 100$$

donde "i" es un subíndice que simboliza el recorrido del conjunto de bienes considerado, "j" es el período para el que interesa estudiar la variación y "0" es el período base.

El subíndice "i" varía de "1 a n", siendo "1" el primer bien y "n" el último de todos los bienes considerados en el conjunto.

El problema de este tipo de índice es que no tiene en cuenta la importancia relativa de cada artículo dentro del conjunto. Tomar en cuenta la importancia relativa de cada bien es lo que se denomina ***ponderación*.**

Vamos a definir qué se entiende por ***canasta,*** *antes de seguir avanzando en el tema.*

Entendemos por ***Canasta*** *al conjunto de bienes y servicios seleccionados para calcular un determinado número índice.*

La selección se realiza a partir de la lista exhaustiva de bienes y servicios vinculados con el fenómeno que se desea reflejar, y debe cumplir la condición de respetar la representatividad o participación de las distintas categorías en el conjunto total.

4 - Indices ponderados

Vamos a ver la noción de índices ponderados a través de un ejemplo simplificado sobre ponderación de los rubros componentes de la canasta familiar, para el cálculo de un índice de precios al consumidor.

Supongamos el consumo mensual de dos familias, compuesto por muy pocos artículos.

FAMILIA A		
ARTICULO	CANTIDAD	GASTO
carne	30 kg	120 $
camisas	2 unid.	32 $
		152 $

FAMILIA B		
ARTICULO	CANTIDAD	GASTO
carne	20 kg.	80 $
zapatos	2 pares	40 $
		120 $

Como vemos, las dos familias tienen consumos mensuales distintos. Para que ambas estén representadas en la *Canasta* es necesario sumar sus consumos.

ARTICULO	GASTO
carne	120 + 80 = 200 $
camisas	32 + 0 = 32 $
zapatos	0 + 40 = 40 $
	272 $

A continuación dividimos el total consumido por artículo entre las dos familias:

ARTICULO	GASTO
carne camisas zapatos	200 / 2 = 100 \$ 32 / 2 = 16 \$ 40 / 2 = 20 \$
	136 \$

Es decir que la composición de la canasta es:

- ► *carne = 100 / 136 * 100 = 73 %*
- ► *camisas = 16 / 136 * 100 = 12 %*
- ► *zapatos = 20 / 136 * 100 = 15 %*

Este es un ejemplo de cómo se conforma una canasta que represente los diferentes consumos de las familias. Como puede verse, no existe una familia que consuma 100 \$ de carne, ya que una consume 120 \$ y la otra 80 \$, pero ésta cifra es representativa de los consumos de las dos familias.

Supongamos ahora que consumiendo lo mismo, los gastos de ambas familias en el mes siguiente son:

ARTICULO	GASTOS EN EL MES 1	GASTOS EN EL MES 2	% DE AUMENTO
carne camisas zapatos	100 16 20	110 19 21	10 % 19 % 5 %
TOTAL	136	150	10,29 %

Como se puede observar, el aumento de la carne fue del 10%, el de las camisas del 19% y el de los zapatos del 5%.

Si calculamos el índice sintético de la variación del conjunto de bienes entre el mes 1 y el mes 2, tenemos:

$$I_{mes1}^{mes2}=\frac{150}{136}=1{,}1029$$

Este resultado nos dice que el aumento del conjunto de bienes en el período fue del 10,29%.

En este cálculo no se tuvo en cuenta la importancia relativa de cada uno de los bienes en el conjunto. Como vimos, la carne representa el 73% del total de bienes, las camisas el 12% y los zapatos el 15%.

*Si calculamos el índice teniendo en cuenta la importancia relativa de cada bien, estamos calculando un **índice ponderado**.*

La operación que debemos hacer es la siguiente:

$$1,1 \times 0,73 + 1,19 \times 0,12 + 1,05 \times 0,15 = 1,1033$$

Es decir, multiplicar el aumento de cada uno de los bienes por la importancia relativa de cada uno de ellos.

Como se puede observar, el aumento de la carne fue menor que el de las camisas. Sin embargo, al ser la carne un artículo de mayor consumo, tiene mayor peso en la determinación de la variación de precios de la canasta, y el índice de precios del conjunto de bienes está más influído por la variación del precio de la carne.

Llevemos este ejemplo a un caso general, donde se consideran "n" artículos, cada uno de ellos con su importancia relativa en la canasta que conforman.

El primer paso consiste en calcular la variación del precio para cada uno de los artículos, que se simbolizan con un subíndice "i". Entonces:

$$I_{0i}^{1} = \frac{P_{i1}}{P_{i0}}$$

Donde se calculan tantos índices, como cantidad de bienes contenga la canasta. Esto significa que "i" varía de "1 a n".

El "0" es el período tomado como base y el "1" el período de referencia para el cual se quiere conocer la variación.

Luego, teniendo los ponderadores de cada bien a los que llamamos "Vi", el cálculo del índice se realiza de la siguiente forma:

$$I_0^1=\sum_{i=1}^{n} I_{0i}^1 \times V_i$$

Para el cálculo de un índice ponderado, se establece primero la canasta de bienes y servicios que constituyen el consumo o gasto promedio de la población.

Luego, se toma la variación en el precio de cada uno de esos bienes y servicios, y finalmente se calcula el valor del índice ponderado realizando la suma de las variaciones de todos los bienes multiplicados por la importancia relativa de cada uno de ellos.

Ejercicio 12.1:

Se tienen los siguientes precios promedios anuales de la Tn. de trigo en la República Argentina para el período 1985-1990.

Período	Precio por Tn. en miles de A.
1985	1000
1986	1500
1987	1400
1988	1600
1989	1900
1990	2100

a) Calcule el índice simple de los precios de la Tn. de trigo para todos los años de la serie con base 1987 = 100.
b) Cuál fue la variación porcentual del precio de la Tn. de trigo, para todo el período considerado?

Resolución del ejercicio 12.1:

a)

b)

Ejercicio 12.2:

En la siguiente tabla se presentan datos de población de distintas regiones de un país:

Región	Población
1	8500
2	2500
3	2500
4	4000
5	3500

a) Calcule los índices elementales de la población de esas regiones tomando como base = 100 la región 1.
b) Calcule los índices elementales de la población de esas regiones tomando como base = 100 la región 5.
c) Interprete los resultados y haga comparaciones entre los índices con ambas bases.

Resolución del ejercicio 12.2:

a)

b)

c)

Ejercicio 12.3:

Si el índice del precio de la leche para el mes de marzo de 1992, con base en julio de 1991 fue de 112. Cuál es el índice del precio de la leche para julio de 1991 con base en marzo de 1992? Explique qué propiedad aplicó.

Resolución del ejercicio 12.3:

Ejercicio 12.4:

Si en el rubro "esparcimiento" del índice de precios al consumidor, intervienen los siguientes items:
- *entrada al cine*
- *cuota societaria del club*
- *alquiler de canchas de paddle*
- *entrada a discotecas*

Simule datos para el consumo de tres familias en tres meses consecutivos y calcule los índices de precios ponderados del rubro "esparcimiento" para esos meses, tomando como base al primero de ellos.

Resolución del ejercicio 12.4: