



UNIVERSIDAD INTERNACIONAL DE ANDALUCÍA

UNIA 10

INTRODUCCIÓN AL ANÁLISIS CUANTITATIVO EN LAS CIENCIAS SOCIALES

POSGRADO EN DESARROLLO ECONÓMICO PARA AMÉRICA LATINA

APUNTES TEÓRICO-PRÁCTICOS

Equipo docente:

Eduardo F. Simonetti

María Fernanda Niño

2006

INDICE

INTRODUCCIÓN	1
I. Aspectos Básicos en el Proceso de Trabajo Estadístico	3
1. Reflexiones previas sobre el uso y la producción de datos	3
2. Algunos conceptos estadísticos básicos	3
2.1. Población	4
2.2. Unidad de análisis	4
2.3. Las variables	4
2.4. Los valores de una variable.....	5
2.5 El dato	6
2.6. Formas de medición de una variable.....	6
2.7. La obtención de los datos.....	7
3. La Matriz de Datos.....	8
II. Primeros Resúmenes de los Datos	11
1. Distintas Formas de Análisis de Datos.....	11
2. Las Distribuciones de Frecuencias Univariadas	12
2.1. Las frecuencias absolutas	12
2.2. Las frecuencias relativas	14
2.3. Los recursos gráficos	14
2.4. Las frecuencias acumuladas	17
2.5. Un caso especial: las variables numéricas con “muchos” valores diferentes	18
III. Valores que Caracterizan una Distribución	27
1. Medias de Tendencia Central	27
1.1. La media aritmética	27
1.2. La mediana	28
1.3. El modo	29
2. Medidas de Posición	30
2.1. Los cuartiles.....	30
2.2. Los quintiles	31
2.3. Los deciles y centiles	31
3. Construcción de una “buena imagen” de la distribución	31
4. Medidas de Variabilidad	34
4.1. El rango.....	34
4.2. Variancia y desviación estándar.....	35
4.3. Las medidas relativas de variabilidad	36
5. La Curva de Lorenz y el Índice de Gini.....	37
IV. La Relación entre Variables	41
1. La Relación entre Variables Categóricas	41
3. La Relación entre Variables Categóricas y Numéricas	44
4. La Relación entre Variables Numéricas.....	46
V. Los Números Índices	51
1. Los Números Índices Simples.....	51

2. Los Números Índices Compuestos.....	52
2.1. Los índices de precios.....	52
2.2. Los índices de cantidad.....	54
3. La Elección del Período Base	54
BIBLIOGRAFÍA.....	59

ACTIVIDADES

ACTIVIDAD N° 1.....	22
ACTIVIDAD N° 2.....	29
ACTIVIDAD N° 3.....	32
ACTIVIDAD N° 4.....	36
ACTIVIDAD N° 5.....	40
ACTIVIDAD N° 6.....	43
ACTIVIDAD N° 7.....	45
ACTIVIDAD N° 8.....	49
ACTIVIDAD N° 9.....	54

INTRODUCCIÓN

Con este material introductorio al análisis de datos cuantitativos en ciencias sociales, pretendemos que los participantes en el postgrado adquieran los conocimientos necesarios para interpretar críticamente información expresada en datos estadísticos.

En este sentido, el curso ha puesto especial énfasis en la comprensión de las ideas generales que guían el trabajo estadístico, las que -asociadas a los interrogantes que se pueden responder desde esta disciplina- derivan en la presentación de técnicas específicas. En cuanto a esas técnicas de organización y resumen de datos, privilegiamos explicar la lógica que las sustenta, sus alcances y limitaciones, evitando en la medida de lo posible transformar este material en un “manual” de cálculos o procedimientos estadísticos.

En consecuencia, esperamos que luego de una lectura detenida de este curso introductorio y apoyándose en las actividades que se van presentando secuencialmente, los participantes puedan estar en condiciones de comprender *por qué* se utilizan esos recursos estadísticos, *cuándo* recurrir a ellos, *qué vínculos* de complementariedad presentan con otros recursos y técnicas consideradas y *qué información* se obtiene a partir de ellos acerca del tema en que se aplica.

Algunas consideraciones para la administración del material del curso:

En este curso introductorio al "Análisis Cuantitativo en Ciencias Sociales", se incluye el siguiente material:

- el presente documento de “notas teórico-prácticas” en el que se desarrollan todos los contenidos teóricos que entendemos necesarios para esta instancia de posgrado, con sus actividades correspondientes, que tienen el propósito de ofrecerles algunas situaciones de trabajo que les ayude a comprender los contenidos y evaluar el aprendizaje individual. Las respuestas que Uds. construyan para estas actividades, no es necesario remitirlas.
- un documento con las “respuestas de las actividades”, a los efectos de que puedan "controlar" sus logros en la utilización/interpretación de los temas estadísticos presentados. Consideramos importante señalar que, en muchos casos, la respuesta *es una de las posibles* (básicamente aquellas en las que hay que redactar un texto que describa un conjunto de datos).
- un archivo con la "Matriz de datos de los países de América" (en formato Excel).
- un archivo que contiene la "Evaluación del curso" (formato word). Esta evaluación es **el único material que deberá ser remitido** a la dirección electrónica y en la fecha que establezca la coordinación del posgrado.

I. Aspectos Básicos en el Proceso de Trabajo Estadístico

1. Reflexiones previas sobre el uso y la producción de datos

Los participantes en esta maestría acordaron con nosotros en que en la cotidianeidad de nuestra vida profesional y ciudadana, tomamos decisiones, interpretamos ciertos hechos y prevemos situaciones futuras basándonos en información actualizada proveniente de distintas fuentes: el periódico, revistas especializadas, informes producidos en el marco de nuestra profesión y los datos que nosotros mismos generamos.

Tampoco debe escapar a nuestra comprensión que esa información siempre está basada en cierto punto de vista que impregna todo el proceso de producción de la información. Cuando esa información es producida a partir de datos estadísticos, debe considerarse que la obtención de estos datos es el fruto de un proceso de construcción metodológica, que supone tomar decisiones en relación a qué es lo que se observa, cuándo, cómo y dónde. Es decir, ya sea que nos ubiquemos en la posición de productores de la información o usuarios de información producida por otros, debemos explicitar en un caso o conocer en el otro, la forma de construcción de esos datos.

Si somos usuarios de información producida por otros y no tenemos esta precaución, corremos el riesgo de sacar conclusiones erróneas. Una apropiada interpretación de la información requiere tanto conocer cómo fueron construidos los datos como conocer las técnicas estadísticas utilizadas en el análisis.

Esta presentación introductoria al "análisis de datos" tiene el propósito de proveerles los saberes mínimos necesarios para abordar adecuadamente la generación y/o interpretación de información.

2. Algunos conceptos estadísticos básicos

Cualquier referencia a un concepto estadístico requiere conocer algunas cuestiones básicas que hacen al concepto de "dato estadístico". Estas cuestiones están referidas a distinguir los **elementos con los que se estructura un dato**. Esto es:

CUESTIONES A CONSIDERAR	CONCEPTO ASOCIADO
¿a quiénes se refieren los datos?	Población y unidad de análisis ("individuo")
¿qué característica interesa registrar?	Variable
¿cuál es el resultado de medir/observar esa característica?	Valor (de la variable)

Además, al enfrentarnos a un conjunto de datos será necesario conocer los aspectos básicos que hacen a la **forma de observación**:

CUESTIONES A CONSIDERAR	CONCEPTO ASOCIADO
¿cuál es el conjunto de unidades observadas?	Población / muestra
¿cómo se midieron las características de interés?	Escalas de medición (tipos de variable) Datos longitudinales o Transversales

A los efectos de ejemplificar la presentación de los conceptos asociados a los aspectos básicos de un trabajo estadístico, supongamos dos situaciones de trabajo:

Caso 1: Queremos analizar las **características económico-sociales de la población desocupada de Cochabamba-Bolivia** en el año 2006. Para ello, *de 700 personas seleccionadas al azar*, registraremos las siguientes características:

- ⇒ **Edad**
- ⇒ **Sexo**
- ⇒ **Nivel de educación**
- ⇒ **Cantidad de años de educación formal**
- ⇒ **Número de hijos**
- ⇒ **Última ocupación realizada**

Caso 2: Queremos hacer una **descripción económico-social de los países de América Latina y el Caribe**. Para ello, disponemos de indicadores básicos referidos a la situación de cada uno de los países. Algunos de estos indicadores son:

- ⇒ **Esperanza de Vida al nacer** (años)
- ⇒ **Tasa de alfabetización de adultos** (% 15 años y mayores)
- ⇒ **PBI per capita** (PPA US\$)
- ⇒ **Gasto Público en salud** (% de PBI)

En base a estos casos, definiremos los principales conceptos.

2.1. Población

Es el conjunto de todos los elementos que interesan estudiar. Su especificación exige explicitar la naturaleza de los elementos que la conforman, el espacio en el que se ubican y el momento del tiempo en el que se realiza la medición/observación.

En los casos presentados como ejemplo, podemos acordar que las poblaciones en estudio, son respectivamente:

Caso 1: "los desempleados de Cochabamba-Bolivia en el año 2006".

Caso 2: "los países de América Latina y el Caribe en el año 2004"

2.2. Unidad de análisis

Cada uno de los **elementos que constituyen la población se denominan "unidad de análisis" o "individuo"**; la **unidad de análisis** se designa con la naturaleza de esos elementos que se están estudiando. Es así que una unidad de análisis puede ser una persona, una entidad abstracta (escuela, una asociación gremial, vivienda, etc.) o cualquier objeto (árbol, animal, parque nacional, etc.).

En los casos presentados las unidades de análisis son:

Caso 1: "las personas"

Caso 2: "los países"

2.3. Las variables

Se denominan **variables en estudio** a las **características o propiedades que interesa estudiar** en cada unidad de análisis,.

En los ejemplos que hemos presentado, las variables de interés para caracterizar a las unidades de análisis, son:

Caso 1:

- ⇒ **Edad**
- ⇒ **Sexo**
- ⇒ **Nivel de educación**
- ⇒ **Cantidad de años de educación formal**
- ⇒ **Número de hijos**
- ⇒ **Última ocupación realizada**

Caso 2:

- ⇒ **Esperanza de Vida al nacer** (años)
- ⇒ **Tasa de alfabetización de adultos** (% 15 años y mayores)
- ⇒ **PBI per capita** (PPA US\$)
- ⇒ **Gasto Público en salud** (% de PBI)

2.4. Los valores de una variable

Los valores de una variable son los números o categorías que constituyen el resultado de observar una cierta característica en una unidad de análisis.

En los ejemplos anteriores, los valores posibles de algunas de las variables observadas, serían:

Caso 1:

Variable	Valores posibles
<i>Edad</i>	15, 16, 17, ..., 45, 46, 47,
<i>Sexo</i>	Masculino Femenino
<i>Nivel de educación</i>	Sin Instrucción; Primario Inc.; Primario Comp.; Secundario Inc.; Secundario Comp.; Superior Inc.; Superior Comp.
<i>Número de hijos</i>	0, 1, 2, 3,

Caso 2:

Variable	Algunos valores posibles
<i>Esperanza de Vida al nacer</i>	49,4; 67,1; ... 74,1 ...
<i>PBI per capita (PPA US\$)</i>	2540, ... 3520, 10.920
<i>Bloque Comercial</i>	Nafta, MERCOSUR,

PARA DESTACAR

Observe que algunos conceptos básicos utilizados en Estadística, dan otro sentido a palabras que se usan cotidianamente:

- 1) **Población** en estadística *no* remite solamente a un conjunto de personas, sino que puede referirse a un conjunto de cualquier tipo de objetos.
- 2) **Individuo** se refiere en estadística a cualquier tipo de objetos y *no* sólo a personas.
- 3) **Valor** lo usamos en estadística para referirnos tanto a números como a categorías y *no* sólo a números.

2.5. El dato

Un dato refiere a los tres conceptos ya definidos: *unidad de análisis*, *variable* y *valor*. El dato será todo registro que indique el valor que asume una variable para una determinada unidad de análisis.

Por ejemplo:

Caso 1:

El nivel de educación de Javier Suárez es Secundario Completo.
Variable
Unidad de análisis
Valor

Caso 2:

En 2002, Brasil tenía una tasa de alfabetización de adultos del 86%.
Unidad de análisis
Variable
Valor

2.6. Formas de medición de una variable

La medición de una característica en una unidad de análisis puede realizarse -como hemos visto- de diferentes maneras: con una escala numérica o a través de categorías, lo cual permite distinguir dos grandes grupos de variables según los valores que asuman: **numéricas o cuantitativas y categóricas o cualitativas**.

Por ejemplo: la variable edad podría medirse en *cantidad de años cumplidos*, en cuyo caso sus valores serían números; pero también podríamos registrar la edad según la etapa vital del individuo. En este caso sus valores podrían ser: *niños, adolescentes, jóvenes, adultos, adultos mayores*.

Entre las **variables numéricas** se pueden distinguir dos tipos:

- las que toman como valores, números enteros. ➔ **Numéricas discretas**
- las que pueden asumir valores no enteros. ➔ **Numéricas continuas**

Por ejemplo, la edad en años cumplidos sólo asume valores enteros, pero la edad real a la fecha de la entrevista puede ser tomada con años y la fracción correspondiente (años, meses y días).

A su vez, las **variables categóricas**, se clasifican en:

- aquellas cuyos valores son categorías no jerarquizables ➔ **Categóricas Nominales**
- aquellas cuyos valores son categorías que asumen una jerarquía ➔ **Categóricas Ordinales**

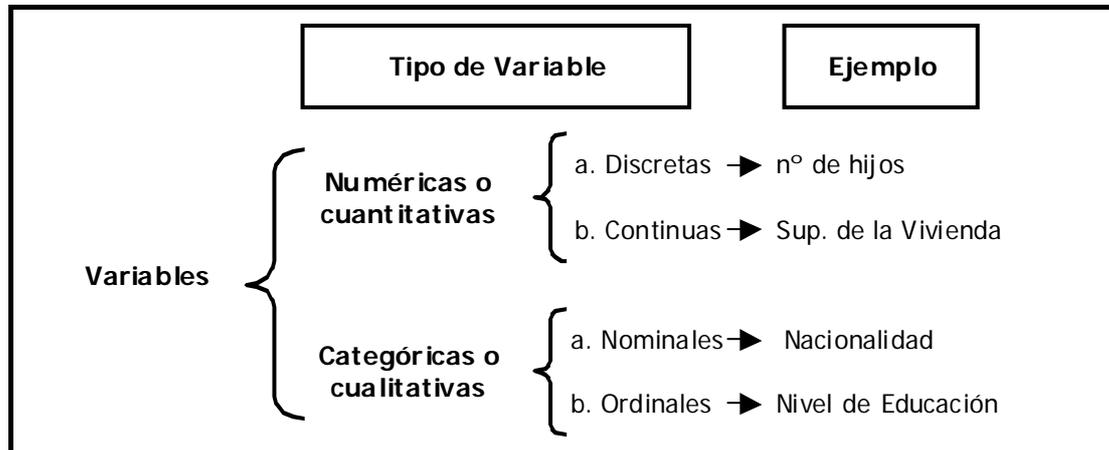
Considerando las variables registradas en los casos que venimos ejemplificando, tenemos:

Caso 1:

Variable	Valores	Tipo de Variable
<i>Edad (años cumplidos)</i>	15, 16, 17, ..., 45, 46, 47,	Numérica Discreta
<i>Sexo</i>	Masculino Femenino	Categórica Nominal
<i>Nivel de educación</i>	Sin Inst.; Prim. Inc.; Prim. Comp.; Sec. Inc.; Sec. Comp.; Sup. Inc.; Sup. Comp.	Categórica Ordinal
<i>Número de hijos</i>	0, 1, 2, 3,	Numérica Discreta

Caso 2:

Variable	Valores	Tipo de Variable
Esperanza de Vida al nacer	49,4; ... 67,1; ... 74,1 ...	Numérica Continua
PBI per capita (PPA US\$)	2540, ...3520, ... 10.920 ...	Numérica Continua
Bloque Comercial	Nafta, MERCOSUR, ...	Categórica Nominal

Esquema 1. Tipos de Variables ¹**2.7. La obtención de los datos**

Hasta aquí hemos presentado dos casos en los que tenemos el propósito de analizar la variabilidad de ciertas características (variables) en un conjunto de "individuos". Sin embargo, se podría plantear una situación diferente en la cual lo que nos interesa es analizar la variabilidad de una o más características a lo largo del tiempo para una misma unidad de análisis. Sería el caso, por ejemplo de ver la evolución que ha tenido la esperanza de vida en los últimos 20 años, en un país determinado (por ejemplo: Uruguay).

La primer situación en la que se observan una serie de variables en un conjunto ("relativamente grande") de individuos, con el propósito de describir la variabilidad de esas características entre las unidades observadas, corresponde a lo que se denomina "**datos transversales**". La observación transversal supone registrar esos datos en un período de tiempo determinado.

En cambio, la segunda situación planteada, que implica la necesidad de describir la variabilidad de esas características entre diferentes momentos de tiempo (años, semestres, meses, etc.), corresponde a lo que se llaman "**datos longitudinales**". La observación longitudinal, demanda la repetición de la observación en el mismo individuo a través del tiempo.

A su vez, podemos distinguir dos situaciones de relevamiento que deben ser consideradas al momento de construir los datos estadísticos o utilizar los producidos por otros. Observen que, en el segundo caso que hemos presentado como ejemplo, los datos disponibles corresponden a la totalidad de cada uno de los "individuos" (países) que conforman la población (países de América Latina y el Caribe). Es decir, se trata de un relevamiento exhaustivo de todos los elementos de la población, forma de observación que se conoce como "**enumeración completa**" o "**censo**".

¹ Tomado de Aguirre, C.; Niño, F.; Simonetti, E. (2004): *Estadística aplicada en las Ciencias Sociales y Humanas. Estadística I*. Colección Cuadernos de Cátedra. Editorial Universitaria de Misiones. Posadas, Misiones, Argentina. Pág. 16.

En contraste, el primer caso, se refiere a un relevamiento realizado sobre 700 personas desocupadas del total de las personas en esa condición que viven en Cochabamba-Bolivia. Así, en esta situación, si bien lo que interesa es caracterizar a los desocupados de esa ciudad, para ese propósito nos basamos en el relevamiento de los datos relativos a un subconjunto de ese total de personas que nos interesan; a esta forma de recolección de datos se la conoce como "**observación por muestreo**" y el conjunto de unidades observadas constituye la "**muestra**".

Observación:

Por Enumeración completa o Censo: se observan las variables de interés en todos los individuos de la población.

Por Muestreo: se observan las variables de interés en un subconjunto de la población.

En general, la observación por muestreo tiene como propósito final describir a la población de la cual fue obtenida. Pero esa posibilidad de *generalizar* los resultados obtenidos en la muestra, a la población dependerá de si consideramos que esa muestra "reproduce bien" las características de la población que nos interesan estudiar (es representativa). Asumido que la muestra tiene esa propiedad, debemos señalar que, aún así, siempre existe el riesgo de cometer un *error* en este proceso de generalización.

En consecuencia cuando analizamos o utilizamos datos muestrales, debemos hacerlo conscientes de la existencia de un error debido al muestreo. Existen otros tipos de errores (de relevamiento, de volcado de datos, etc.) de los cuales no está exento tampoco un relevamiento exhaustivo.

3. La Matriz de Datos

En una situación bastante frecuente, es posible recurrir a datos secundarios² para responder las preguntas que nos planteamos en una investigación o trabajo concreto. En otras, Ud. dispondrá de los datos para cada unidad de análisis³, los que deberá "procesar" en función de la información que necesita producir. En situaciones como esta última, esos datos sin resumir se organizan de forma tal que faciliten su procesamiento (manual o por ordenador).

A esta forma de organización de los datos se la denomina "**Matriz de datos**".

Se trata de una "planilla" con filas y columnas, en la cual se disponen los datos según la siguiente estructura:

- las **filas** representan a cada *unidad de análisis*
- las **columnas** representan a cada una de las *variables*
- en las **celdas** se registrará el *valor* de cada unidad de análisis para cada variable.

² Son aquellos producidos por terceros y de los cuales disponemos en forma ya resumida (tablas, medidas descriptivas varias, gráficos, etc.); es decir no contamos con los datos de cada una de las unidades de análisis. Por ejemplo los cuadros producidos por los Organismos Nacionales de Estadística.

³ Pueden ser de producción propia o contruidos por otros.

Matriz de datos con variables económico-sociales para países de América

iden	País	Variables				
		Esperanza de Vida al nacer (años)	Tasa de alfabetización de adultos (% 15 años y mayores)	PBI per capita (PPA US\$)	...	Gasto Público en salud (% de PBI)
		2002	2002	2002	...	2001
1	Antigua y Barbuda	73,9	85,8	10.920	...	3,4
2	Argentina	74,1	97	10.880	...	5,1
3	Bahamas	67,1	95,5	17.280	...	3,2
4	Barbados	77,1	99,7	15.290	...	4,3
5	Belice	71,5	76,9	6.080	...	2,3
6	Bolivia	63,7	86,7	2.460	...	3,5
7	Brasil	68	86,4	7.770	...	3,2
8	Canadá	79,3	..	29.480	...	6,8
9	Chile	76	95,7	9.820	...	3,1
10	Colombia	72,1	92,1	6.370	...	3,6
11	Costa Rica	78	95,8	8.840	...	4,9
12	Cuba	76,7	96,9	5.259	...	6,2
.
.
.
32	Surinam	71	94	6.590	...	5,7
33	Trinidad y Tobago	71,4	98,5	9.430	...	1,7
34	Uruguay	75,2	97,7	7.830	...	5,1
35	Venezuela	73,6	93,1	5.380	...	3,7

En la matriz anterior⁴, el "individuo" o "unidad de análisis" es un país americano y las variables son -entre otras- la *Esperanza de vida al nacer*, la *Tasa de alfabetización de adultos*, el *PBI per cápita*, *Gasto público en Salud*. Por ejemplo, el "individuo" *Belice* (línea grisada) registra una Esperanza de vida de *71,5 años*; una Tasa de alfabetización de adultos de *76,9%*; un ingreso *per cápita* de *6080 US\$PPA* (Paridad de poder adquisitivo).

Si se tratara de la matriz de datos para la muestra de desocupados en Cochabamba, las características de cada desocupado (unidad de análisis) se podrán leer en cada una de las filas de la matriz. En cada columna, podríamos leer los valores que toma una variable en el conjunto de los 700 "individuos" de la muestra.

⁴ En un archivo Excel se incluye la matriz completa (archivo: Matriz países de América.xls) la cual se reproduce aquí parcialmente.

II. Primeros Resúmenes de los Datos

1. Distintas formas de Análisis de Datos

Una vez resuelta la necesaria etapa de organizar los datos en forma práctica (la matriz), estamos en condiciones de retomar el propósito que generó la necesidad de un tratamiento estadístico. Tomemos el caso de querer describir la situación económico-social de los países de América Latina y el Caribe; es fácil ver que ese propósito de descripción, supone responder interrogantes de diferente complejidad.

En el nivel más general, procuramos saber: *¿la situación de desarrollo económico-social de los países de América refleja cierto nivel de heterogeneidad/diversidad?* Esta pregunta conduce a una comparación de los países en términos de *todos* aquellos indicadores que hemos seleccionado como pertinentes para caracterizar su desarrollo. Es decir, que requeriríamos de herramientas estadísticas que nos permitan el análisis de la matriz en su conjunto. Esta forma de abordar la matriz de datos que permite una descripción de los individuos observados considerando **un gran número de variables simultáneamente**, se conoce como ***análisis estadístico multivariado***.

Al mismo tiempo que nos planteamos esta pregunta de nivel general, aparecen otras cuestiones más específicas: *¿los niveles de educación en estos países están relacionados con logros en materia de ciencia y tecnología?* Para responder a este interrogante, será necesario conocer si efectivamente *hay diferencias "considerables" en materia de logros educativos*.

Este tipo de interrogantes nos conducen a tratamientos parciales de la matriz: algunos requerirán el análisis de una única variable mientras que, para otros, será necesario considerar a los países según dos características simultáneamente. En el primer caso, estaríamos realizando un ***análisis univariado***, mientras que en el segundo, se trataría de un ***análisis bivariado***.

En función de lo dicho, procure Ud. identificar qué tipo de tratamiento estadístico será necesario para responder las siguientes preguntas.

- ✓ ¿es heterogénea la expectativa de vida?
- ✓ ¿por debajo de qué ingreso *per cápita* están los países más pobres?
- ✓ ¿los niveles de alfabetización se relacionan con los niveles de gasto en educación?
- ✓ ¿los niveles de gasto en investigación y desarrollo se relacionan con los niveles de exportación de manufacturas y los niveles de desempleo?

Con independencia de lo que hemos señalado arriba, cabe mencionar que en la práctica se procede partiendo de un análisis univariado. Esta práctica que se explica porque así, de una manera rápida y sencilla, tenemos una primera aproximación al problema que estamos abordando y podemos detectar aquellas variables que aparecen como más "interesantes" desde el punto de vista del problema.

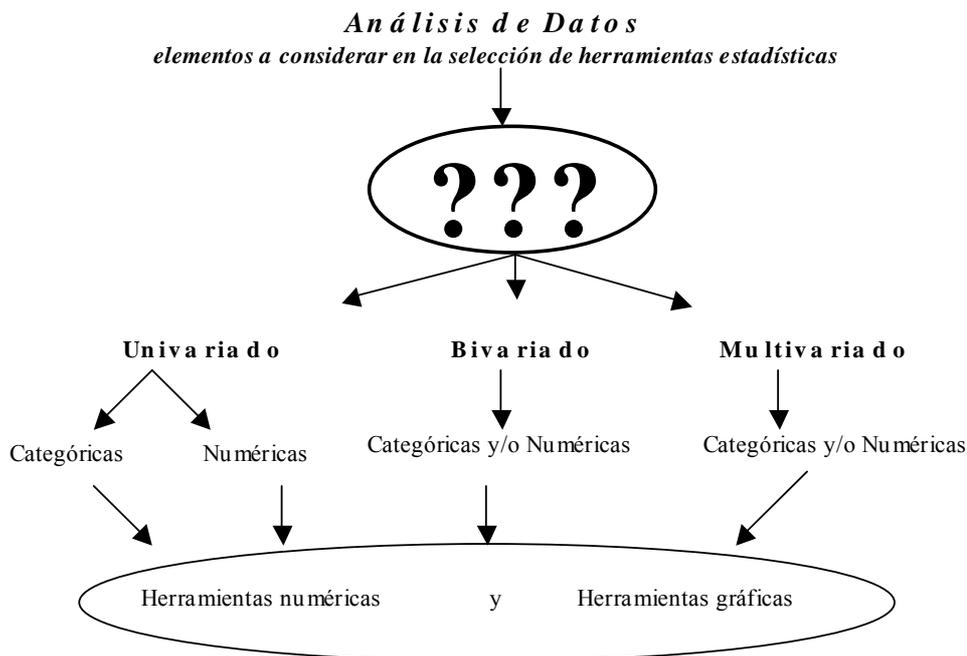
Por las razones mencionadas precedentemente, comenzaremos esta presentación con el desarrollo de las herramientas de análisis univariado que son de uso más frecuente en las Ciencias Sociales; posteriormente, presentaremos las herramientas de análisis bivariado⁵.

Hasta aquí hemos presentado una clasificación de las herramientas estadísticas que toman como criterio los interrogantes acerca del fenómeno que estamos analizando. Sin embargo, la elección del tipo de herramientas estadísticas a utilizar dependerá también de considerar el tipo de variable que se quiere analizar (numéricas o categóricas) y la forma en que se quiere (o conviene) resumir la información: recursos numéricos y/o gráficos.

Esta clasificación está sintetizada en el siguiente esquema del análisis de datos⁶.

⁵ La complejidad de los métodos multivariados demanda una experiencia en el análisis estadístico y un tiempo de presentación, que impide su inclusión en un curso de estas características.

⁶ Extraído de Aguirre, Niño, Simonetti (2004, 29).



2. Las Distribuciones de Frecuencias Univariadas

2.1. Las frecuencias absolutas

Las *distribuciones de frecuencias univariadas* constituyen la forma más sencilla e intuitiva de *resumir los datos*. Son el resultado de contar el número de veces que se repite cada valor de la variable en estudio en el conjunto de todos los individuos. Por ejemplo, si consideramos la variable *Bloques comerciales* de la matriz de datos *económico-sociales para países de América*, contamos el número de veces que se presenta el valor “MERCOSUR”, el valor “CARICOM”, etc. en el conjunto de los 35 países. Así, resulta que 4 es el número de veces que se repite la categoría MERCOSUR y 14 la categoría CARICOM. Este número de repeticiones que corresponde a cada *valor* de la variable recibe el nombre de *frecuencia absoluta*.

Frecuencia absoluta

Es la cantidad de individuos que presentan un mismo valor de la variable (una misma categoría si se trata de una variable categórica, un mismo número si la variable es numérica).

Calculadas las frecuencias absolutas para los diferentes valores de una variable de la matriz de datos, se puede *construir una tabla* que exprese esa distribución de los individuos según la característica analizada.

... en el caso de variables categóricas

Consideremos la variable *bloques comerciales*:

Distribución de Países Americanos según Bloque Comercial al que pertenece-2005.

Nombre de la variable	Bloque Comercial ⁽¹⁾	n° de países	Cantidad de integrantes del
	CAN	5	CAN
Valores de la variable	CARICOM	14	Frecuencias absolutas
	MCCA	5	
	MERCOSUR	4	
	NAFTA	3	
	Otros	4	
	Total	35	Total de "individuos" observados

(1) CAN: Comunidad Andina de Naciones; CARICOM: Comunidad del Caribe; MCCA: Mercado Común Centroamericano; MERCOSUR: Mercado Común del Sur; NAFTA: Tratado de Libre Comercio de América del Norte.

Fuente: Datos ec.-soc. para países de América

PARA DESTACAR

Observe que es indispensable consignar en la tabla:

- a) Un **título** que sintetice el contenido de la tabla: qué se distribuye, cómo, cuándo y dónde.
- b) El **nombre de la variable** y los **valores** registrados (en la primera columna de la izquierda).
- c) La **Fuente** de los datos. Aquí se registrara el responsable del relevamiento y/o de los datos.
- d) El **Total** de datos observados.

En la tabla anterior hemos resumido la columna de "Bloques comerciales" correspondiente a la matriz de datos. Con este resumen, y sin perder información, hemos ganado en la posibilidad de comprender mejor la distribución de los países en los diferentes bloques.

... en el caso de variables numéricas

Cuando se trata de una **variable numérica con pocos valores diferentes**, la construcción de la tabla será análoga a la anterior. Este es el caso de la variable **Niños con bajo peso al nacer (%)** en la matriz de datos económico sociales de países americanos. La construcción de la distribución de frecuencias para esta variable será:

Países americanos según el porcentaje de niños con bajo peso al nacer. 1998-2002.

Porcentaje de Niños con bajo peso al nacer	n° países	
5	1	
6	3	
7	4	
8	4	
9	7	Hay 7 países con un 9% de niños con bajo peso al nacer
10	5	
11	1	
12	1	
13	4	
14	2	
16	1	
21	1	
23	1	
Total	35	

Fuente: Datos ec.-soc. para países de América

2.2. Las frecuencias relativas

Generalmente interesa presentar y analizar las distribuciones de frecuencias en términos relativos al total de datos (sean las variables, numéricas o categóricas), para señalar la importancia relativa de cada valor en el conjunto de los individuos observados. Esta necesidad es aún mayor cuando se quiere hacer una comparación entre conjuntos de "individuos" que tienen diferentes tamaños. Por ejemplo:

- “El 40% de los países pertenecen al bloque comercial Comunidad del Caribe”, para señalar el peso relativo de esta asociación en términos del número de países miembros.
- “El 37% de los países de América presentan un IDH Alto, mientras que entre los Estados Árabes se incluyen en esta categoría un 22% de estos estados”, para comparar el logro de una buena calidad de vida entre dos regiones del mundo.

Además, cuando trabajamos con datos muestrales, las cantidades absolutas de "individuos" en la muestra carecen de interés. En tanto que, la proporción sugiere igual presencia de esa característica en la población.

Frecuencia relativa

Generalmente expresada en porcentajes, indica la proporción de individuos que presentan un mismo valor de la variable.

Se la obtiene como el cociente entre la frecuencia absoluta y el total de datos, multiplicado por 100.

Distribución de Países Americanos según Bloque Comercial al que pertenece-2005.

Bloque Comercial ⁽¹⁾	nº de países	% de países
CAN	5	14,3
CARICOM	14	40,0
MCCA	5	14,3
MERCOSUR	4	11,4
NAFTA	3	8,6
Otros	4	11,4
Total	35	100,0

Porcentaje de Países de América que integran el CAN

Frecuencias relativas

⁽¹⁾ CAN: Comunidad Andina de Naciones; CARICOM: Comunidad del Caribe; MCCA: Mercado Común Centroamericano; MERCOSUR: Mercado Común del Sur; NAFTA: Tratado de Libre Comercio de América del Norte.

Fuente: Datos ec.-soc. para países de América

Las frecuencias relativas también se utilizan en el caso de variables numéricas. Complete Ud. la tabla de la distribución de la variable **Niños con bajo peso al nacer (%)**, agregando las correspondientes frecuencias relativas.

2.3. Los recursos gráficos

... para variables categóricas

Las dos formas gráficas más utilizadas para presentar distribuciones de frecuencias de variables categóricas son: el **gráfico de barras** y el **gráfico de sectores**.

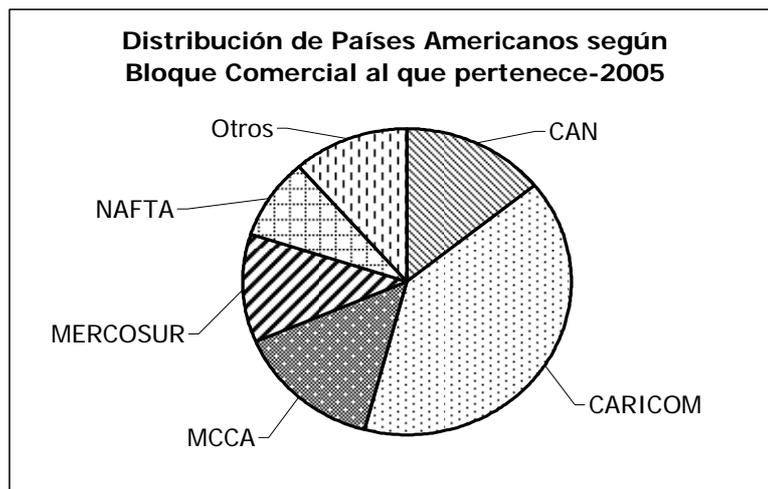


Gráfico de Sectores

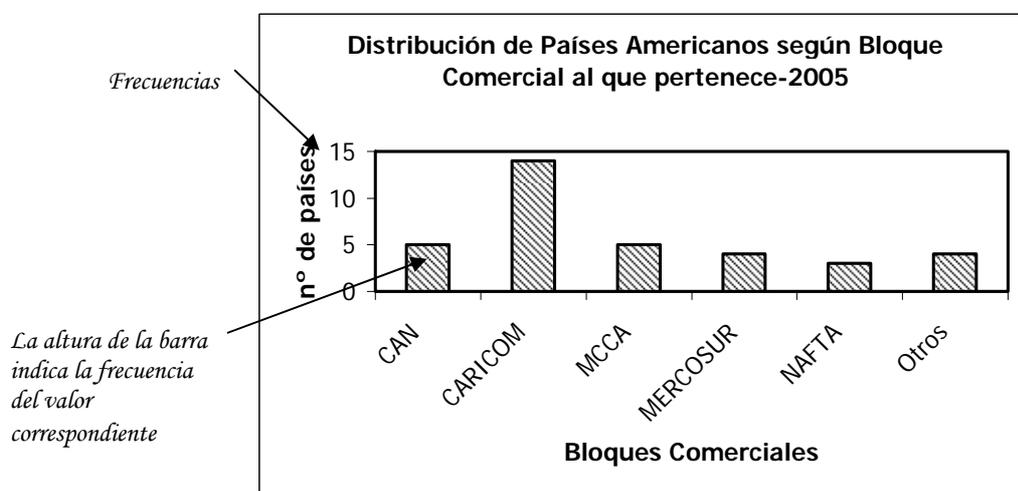


Gráfico de Barras

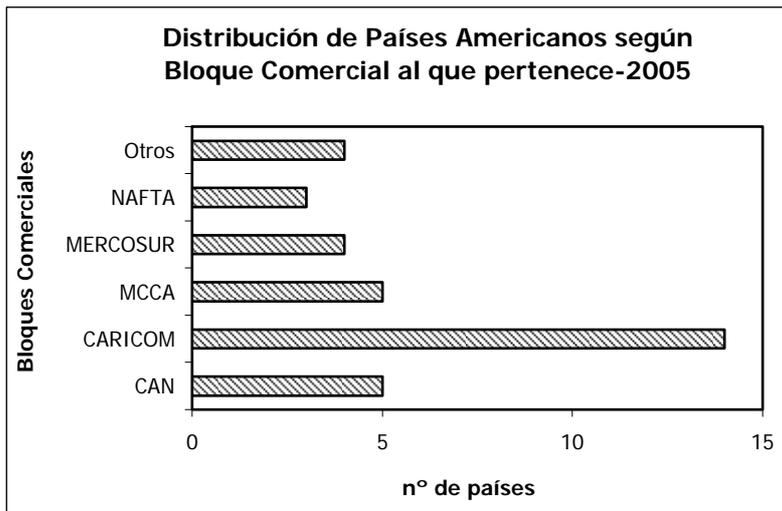
Fuente: Datos ec.-soc. para países de América

A simple vista, los gráficos construidos nos permiten captar rápidamente la desigual distribución por bloques comerciales de los países americanos. Esta característica de las herramientas gráficas hacen que las mismas sean apropiadas como:

- un recurso de análisis de los datos, y
- una forma efectiva de presentar y comunicar los resultados.

En ambos casos y al igual que en las tablas, es imprescindible que se consigne un título indicando lo que se está representando.

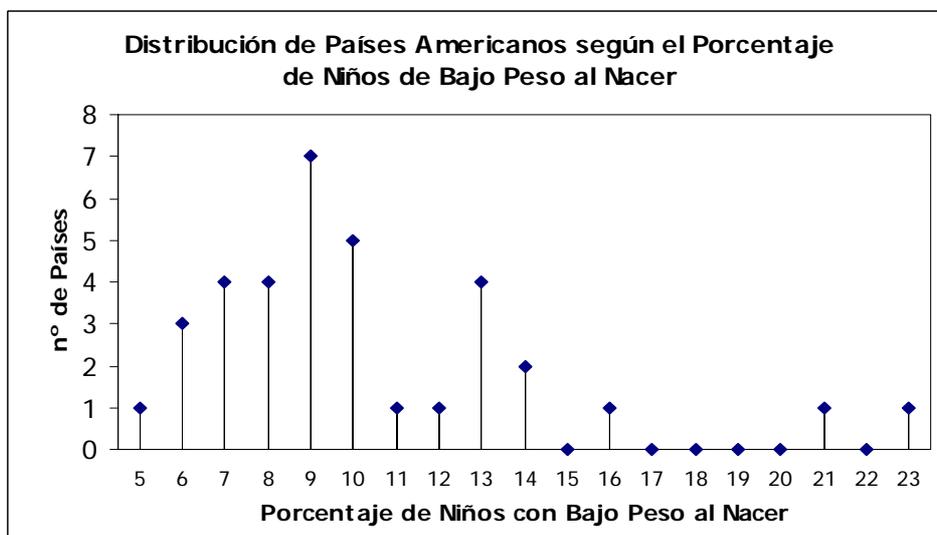
En particular, en el **gráfico de barra** se pueden representar las frecuencias absolutas o relativas. Además, en este tipo de gráfico (basado en dos ejes), es igualmente necesario que se adjunte un texto en el cual se consigne lo que se representa en cada eje (la variable y el tipo de frecuencias respectivamente). Por último, este gráfico puede ser construido representando en el eje horizontal las frecuencias y en el vertical las categorías, recurso que resulta particularmente útil cuando la variable tiene muchas categorías.



Con respecto al **gráfico de sectores**, debemos señalar que su expresividad se reduce notablemente (se hace muy difícil de leer) cuando estamos en presencia de una variable con muchas categorías. En estos casos es recomendable un gráfico de barras con las frecuencias en el eje horizontal.

... para Variables Numéricas

En el caso de variable numéricas con pocos valores diferentes, se utiliza un gráfico similar al de barras utilizado para las variables categóricas. Sin embargo, existe aquí una diferencia destacable, dada por el hecho de que los valores son números y en consecuencia, a cada uno le corresponde un punto en un eje numérico (eje X). Consecuentemente, a cada valor se le asocia un bastón cuya altura expresa la frecuencia (absoluta o relativa) correspondiente al mismo.



Fuente: Datos ec.-soc. para países de América

El gráfico anterior (de bastones) nos permite apreciar más rápidamente (aunque sin mucho detalle) la diversidad de situaciones que presentan los países de América en relación a los niños con bajo al nacer. Así, podríamos decir que: el porcentaje de niños con bajo peso al nacer en los países de América, varía entre el 5 y el 23%. Lo más frecuente es entre un 9 y 10% de niños con bajo peso al nacer, pero hay en general una mayor concentración de países entre 7 y 13 %. Son unos pocos países los que superan el 14%.

A pesar de las ventajas de describir la información a partir de un gráfico como éste, es evidente que para comunicar de manera más sintética y precisa la información que podemos "ver" en el gráfico, será

necesario recurrir a otros instrumentos estadísticos; las *frecuencia acumuladas* constituyen una respuesta técnica adecuada a esta necesidad.

2.4. Las frecuencias acumuladas

Para lograr una descripción más sintética (como planteáramos arriba) podríamos indicar *el número de países que no superan el 10% de niños con bajo peso al nacer*; asimismo, podríamos señalar *cuántos son los que están por encima del 14%*.

El número (o porcentaje) de unidades de análisis que registran un determinado valor de la variable o cualquier valor menor que él, se denomina **frecuencia acumulada "menos que"**. Análogamente se puede definir la **frecuencia acumulada "más que"**, como el número (o porcentaje) de observaciones que tienen un determinado valor de la variable o cualquier valor mayor que él.

Países americanos según el porcentaje de niños con bajo peso al nacer. 1998-2002.

Porcentaje de Niños con bajo peso al nacer	países (n°)	Fa-	Fa+
5	1	1	35
6	3	4	34
7	4	8	31
8	4	12	27
9	7	19	23
10	5	24	16
11	1	25	11
12	1	26	10
13	4	30	9
14	2	32	5
16	1	33	3
21	1	34	2
23	1	35	1
Total	35		

Hay 24 países con un 10% o menos de niños con bajo peso al nacer

Fuente: Datos ec.-soc. para países de América

En la tabla anterior, podemos precisar que son 24 los países que no superan el 10% de niños con bajo peso al nacer, estos países representan el 69% de todos los países de América. A su vez, podemos ver que sólo 5 de los países tienen tasas de bajo peso al nacer de 14% o más.

Las frecuencias acumuladas son un valioso recurso para el análisis de los datos que permiten resumir aún más lo que figura en la tabla de frecuencias. Pero, cabe señalar que para la presentación en un informe, no es necesaria la inclusión de las columnas que indican cada una de estas frecuencias acumuladas, aún cuando figuren algunas de ellas en el texto de interpretación de la distribución.

Este recurso de lectura, que transforma las frecuencias absolutas o relativas, es **válido también para las variables de tipo ordinal**. Por ejemplo, en la distribución de los países según su *nivel de IDH*, podemos leer que *un 63% de los países presentan a lo sumo niveles medio de desarrollo humano*.

Distribución de Países de América según Nivel de IDH. 2002.

Nivel de IDH	Países n°	Países (%)
Bajo	1	2,9
Medio	21	60,0
Alto	13	37,1
Total	35	100,0

Fuente: Datos ec.-soc. para países de América

2.5. Un caso especial: las variables numéricas con "muchos" valores diferentes.**... el recurso numérico**

Todo lo señalado precedentemente en cuanto los conceptos relacionados a las distribuciones de frecuencias, resultan válidos también para este caso especial. Lo que ocurre cuando la variable numérica toma un gran número de valores diferentes, es que nos encontramos con la necesidad de realizar un "tratamiento" particular de los valores de la variable. Para comprender esta idea, veamos el siguiente ejemplo.

Consideremos ahora la variable *Porcentaje de Población con Acceso a Fuentes de Agua Mejorada*⁷.

Distribución de los Países según Porcentaje de Población con Acceso a Fuentes de Agua Mejorada. 2000.

% de población con acceso a agua mejorada	n° de países
46	1
77	2
78	1
80	1
82	1
83	2
85	1
86	1
87	1
88	2
90	2
91	3
92	3
93	2
94	1
95	2
97	2
98	3
99	1
100	3
Total	35

Fuente: Datos ec.-soc. para países de América

⁷ Porcentaje de la población con un acceso razonable a cualquiera de los siguientes tipos de fuentes de agua potable: conexiones en la vivienda, fuentes públicas, pozos perforados, pozos excavados protegidos, manantiales protegidos y depósitos de agua de lluvia. Se define como acceso razonable la disponibilidad de al menos 20 litros por persona diarios de una fuente situada en un radio de un kilómetro de distancia desde la vivienda del usuario.

La distribución resultante, al listar los diferentes valores que registran el conjunto de los países de América y asignarle a cada uno la correspondiente frecuencia, **constituye un resumen de la información** que no mejora demasiado la posibilidad de comprender el comportamiento de los datos. Entre los países se registran 20 valores diferentes, lo que resulta en una tabla extensa que dificulta aprehender la tendencia general de esta característica en el conjunto de países. En consecuencia, la tabla anterior no resulta un buen recurso para el análisis.

Imagine Ud. la interpretación de los datos con una distribución de frecuencias en la que se listan los ingresos familiares de una muestra de 600 familias de la ciudad de Encarnación-Paraguay. En esa muestra, seguramente habrá tantos valores diferentes de la variable que resultará entonces una tabla muy extensa en la que se resume muy poco la información de la matriz.

En este tipo de situaciones, para lograr un buena síntesis que exprese más claramente el comportamiento de los datos, lo que se hace es agrupar los valores de la variable y computar las frecuencias para cada uno de esos grupos definidos.

Distribución de los Países según Porcentaje de Población con Acceso a Fuentes de Agua Mejorada. 2000.

	% de población con acceso a agua mejorada	n° de países
<i>Clase abierta</i> →	menos de 76 (*)	1
Intervalos de igual amplitud {	76-80	3
	80-84	4
	84-88	3
	88-92	7
	92-96	8
	96-100	9
	Total	35

Son 7 los países en los que, la población que accede a agua mejorada es de 88 a menos de 92%.

(*) Haití presenta un 46% de población con acceso a agua mejorada.

Fuente: Datos ec.-soc. para países de América

A cada uno de esos **grupos de valores** para los cuales computamos las frecuencias, los denominamos **intervalos de clase** y a la distribución resultante se la llama **distribución en intervalos de clase**.

El logro de un mayor resumen de los datos, ha redundado (necesariamente como en cualquier resumen) en cierta **pérdida de información**. Por ejemplo: la tabla indica que son 4 los países que tienen *entre* 80 y 84 % de población con acceso a agua mejorada, sin que sea posible precisar el valor exacto que asume cada uno de ellos. Además, cuando las *clases son abiertas*⁸ la pérdida de información es aún mayor en la medida en que estas clases no están totalmente acotadas. Por ello, en la tabla anterior resulta conveniente agregar la información referida al país que presenta un valor muy atípico.

Acerca de la construcción y lectura de las distribuciones en intervalos

Los intervalos pueden ser contruidos de diversas maneras: **cerrados o abiertos y continuos o discontinuos**. La decisión del tipo de intervalos a utilizar está vinculada a las características del conjunto de datos con el que estamos trabajando.

Cerrados o abiertos:

En el ejemplo anterior todos los intervalos menos el primero, son cerrados (tienen un límite inferior y uno superior). Los intervalos abiertos, como el primero (menos de 76) son necesarios

⁸ La necesidad de utilizar este tipo de clases es frecuente cuando existen valores atípicos (muy diferentes del conjunto).

cuando existen algunos datos muy atípicos. En este caso, el valor registrado para Haití (46%), llevaría a incluir 7 intervalos vacíos (si conservamos la amplitud) o bien, un gran intervalo de una amplitud mayor que los otros (46 – 76); y ninguna de estas dos situaciones es deseable.

Continuos o discontinuos

Estas distintas formas de construir los intervalos están determinadas por la coincidencia entre el límite superior de una clase y el inferior de la siguiente.

En el ejemplo, los intervalos son continuos. Cuando esto ocurre, debe considerarse la siguiente convención para “leerlo”: el intervalo incluye todos los valores que van desde el límite inferior inclusive hasta el **límite superior sin incluirlo**. Por ejemplo: el intervalo 80-84 se leería: son 4 los países en los que acceden a agua mejorada entre un 80 y un 83,99% de su población.

Los intervalos discontinuos, son aquellos en los que el límite superior de una clase **no es igual al límite inferior de la siguiente** (ejemplo: 76-79, 80-83, etc.). Este tipo de intervalos podrá utilizarse siempre que se trate de una variable discreta.

Consideremos ahora el análisis el *Producto Bruto Interno anual per cápita (US\$ PPA)*⁹ de los países de América. La cantidad de valores diferentes que presenta esta variable (ver matriz de datos) nos exige construir una distribución de frecuencias en intervalos. Utilizando las diferentes herramientas vistas hasta aquí (frecuencias absolutas, relativas y acumuladas) podemos hacer una buena descripción de la diversidad de producción de riqueza por habitantes en los países de América.

Distribución de Países de América según el PBI per cápita. 2002

PBI p.c (US\$ PPA)	Países nº	Países (%)	Países Acum. nº	Países Acum. (%)
1500-7500	22	62,8	22	62,8
7500-13500	9	25,7	31	88,5
13500-19500	2	5,7	33	94,2
19500-25500	0	0,0	33	94,2
25500-31500	1	2,9	34	97,1
31500-37500	1	2,9	35	100,0
Total	35	100,0		

$22 + 9$
 $\frac{22 + 9}{35} \times 100$

De la información resumida en el cuadro anterior, se puede concluir que:

la mayoría (63%) de los países americanos tiene un ingreso per cápita de entre 1500 y 7500 US\$ PPA. Además, se aprecia una concentración de países en los primeros tramos de ingreso per cápita, que se traducen en que aproximadamente el 89% no superan los 13500 US\$ PPA, en tanto que sólo el 6% (2 países) generan una riqueza por habitante que supera los 25500 US\$ PPA.

... el recurso gráfico

A las distribuciones de frecuencias organizadas en intervalos de clase se les asocia un gráfico que se conoce como *histograma* y al cual se le asocia otro gráfico conocido como *polígono de frecuencias*.

⁹ **PIB per cápita (PPA en USD)**. Es la producción total de bienes y servicios de una economía dividido por la población a mitad de año expresada en un valor dólar (US\$ PPA) que refleja las diferencias de precios entre países y permite la comparación internacional. (1 USD de PPA tiene el mismo poder adquisitivo en la economía nacional que 1 USD en los Estados Unidos).

Histograma

Análogamente a los gráficos de bastones, el histograma utiliza un sistema de ejes cartesianos en el cual en el eje de abscisas se representa a la variable y en el de ordenadas las frecuencias (absolutas o relativas).

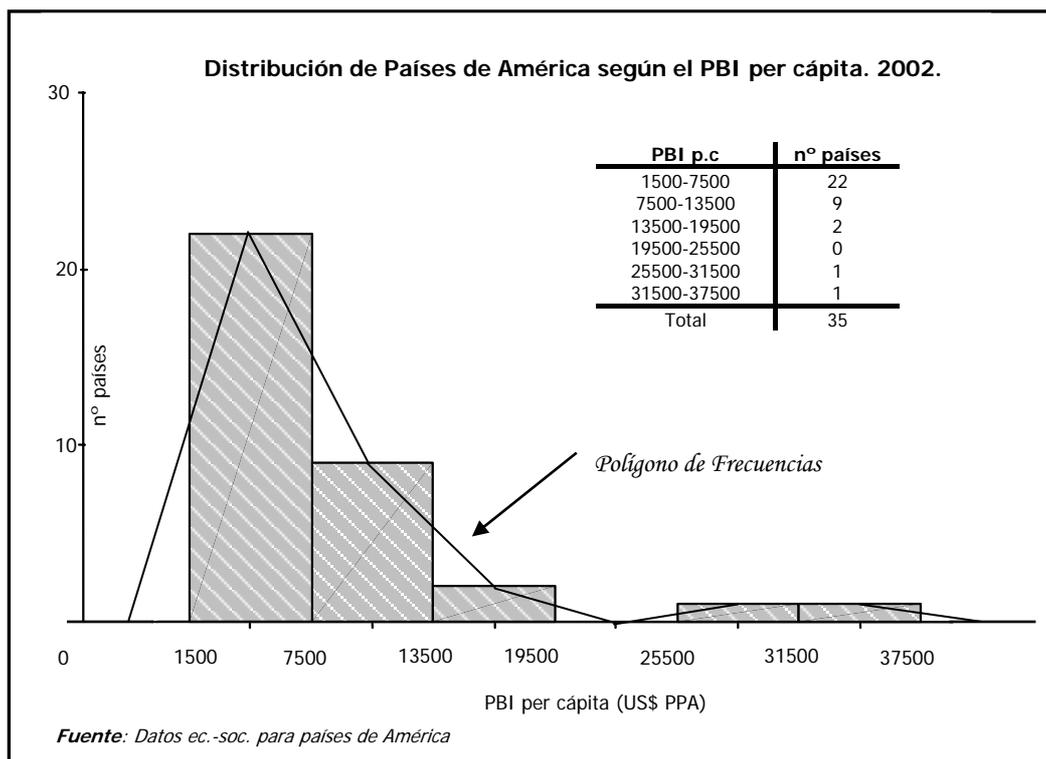
La diferencia con el gráfico de bastones es que aquí, no conocemos las frecuencias de los valores individuales, sino la de intervalos. De modo que, para representar la frecuencia de cada intervalo se utiliza una **barra cuya base está dada por la amplitud del intervalo y la altura coincide con la frecuencia de la clase.**

Polígono de Frecuencias.

Se construye mediante segmentos que unen la parte superior de las barras del histograma en su punto medio. El polígono se completa uniendo los extremos de la primer barra con el punto medio de un primer intervalo hipotético y análogamente con otro segmento se une el punto medio de la última barra con un último intervalo hipotético.

Tanto el histograma como el polígono son recursos con una misma finalidad: explorar y presentar el comportamiento de variables numéricas con muchos valores diferentes.

Siguiendo con el ejemplo del *Producto Bruto Interno anual per cápita (US\$ PPA)* de los países de América el histograma y polígono correspondientes serían:



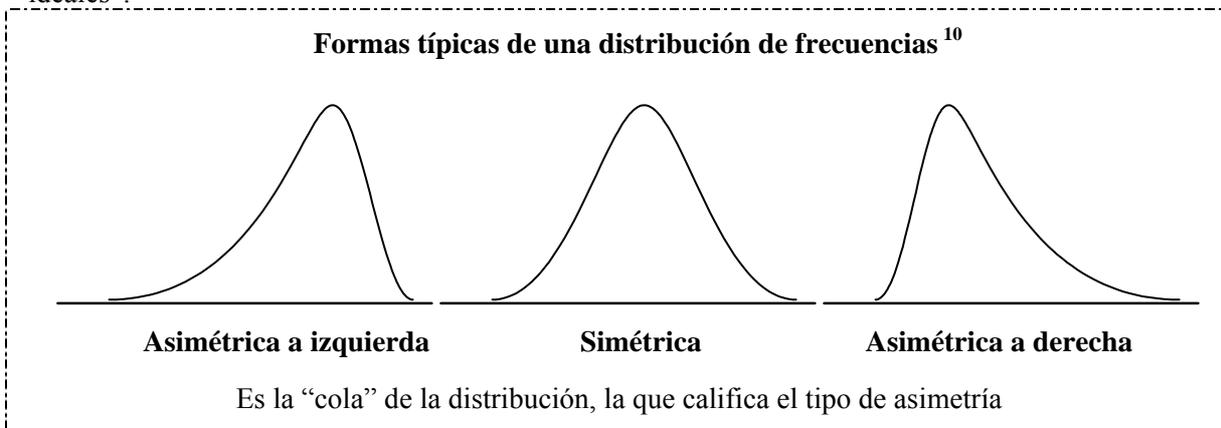
Se puede ver que la concentración de países se produce en los primeros intervalos (1500-7500 y 7500-13500). Además, se aprecia que hay una gran heterogeneidad en el ingreso *per cápita* debida fundamentalmente a la presencia de dos países con PBI *per cápita* superior a los 25500 US\$PPA anuales.

El recurso gráfico nos permite apreciar más claramente la *diversidad* de situaciones en la “riqueza” generada por los países de América. Describir esa variabilidad implica considerar tanto las concentraciones como las singularidades (valores poco frecuentes y atípicos).

La idea general que pudimos construir a partir del gráfico nos orienta en la lectura de la tabla para hacer una descripción sintética de la variable en estudio. De modo que, el uso complementario del recurso numérico y el gráfico, nos permiten concluir:

Los datos revelan que los países de América presentan una situación diversa en términos de la producción de bienes y servicios que generan en relación a su población. Es posible distinguir dos grupos de países marcadamente diferentes: el primero donde se concentra la mayoría de ellos (33 países) que presentan ingresos per cápita que no superan los 19500 US\$ PPA y un segundo (2 países) con ingresos superior a los 25500 US\$ PPA que pueden considerar atípicos. Es de destacar que, en el primer grupo, 22 países no superan los 7500 US\$.

Los siguientes gráficos son polígonos "estilizados" que representan las formas típicas que puede asumir una distribución: *asimétricas a derecha e izquierda* y *simétrica*. En la práctica, difícilmente podremos identificar formas tan claras, sino diferentes niveles de aproximación a estas distribuciones "ideales".



La forma de la distribución nos indica en qué valore se produce la mayor concentración de los datos:

Asimetría a izquierda: los datos se concentran en los mayores valores de la variable y existen datos poco frecuentes y/o atípicos en los valores más bajos.

Asimetría a derecha: los datos se concentran en los valores menores de la variable y a la derecha se observa la presencia de valores poco frecuentes y/o atípicos.

Simetría: la concentración se produce en los valores centrales con valores igualmente poco frecuentes en ambos extremos. Se trata de una distribución "equilibrada".

En el ejemplo del PBI *per cápita* para los países de América, la distribución es muy asimétrica a la derecha.



ACTIVIDAD N° 1

PRIMERA PARTE

Encuesta Nacional de Juventud (2000)- Secretaría de Educación Pública, Instituto Mexicano de la Juventud, Centro de Investigación y Estudios sobre Juventud, México 2002.

El Instituto Mexicano de la Juventud (IMJ) se propuso como tarea central actualizar y profundizar en el conocimiento sobre las y los jóvenes mexicanos, como un elemento fundamental para la construcción de políticas públicas en beneficio del sector juvenil. Para ello se realizó un relevamiento con representatividad a nivel nacional (n= 23129).

¹⁰ Extraído de: Aguirre, Niño, Simonetti (2004: 42)

Una de las dimensiones consideradas fue la relativa a la inserción y trayectoria laboral de los jóvenes. De las tablas incluidas en el informe hemos seleccionado aquellas que se refieren al primer empleo que han tenido.

Edad a la que los jóvenes obtuvieron su primer trabajo. México, 2000.

Grupos de edad	Jóvenes (%)	Far (%)
Menos de 11 años	13,2	13,2
De 12 a 14 años	28,8	42,0
De 15 a 19 años	49,0	91,0
De 20 a 24 años	8,3	99,3
De 25 a 29 años	0,7	100,0
Total	100,0 (23129)	

Número de horas al día que laboraban los jóvenes en su primer trabajo. México, 2000.

Horas al día	Jóvenes (%)	Far (%)
De 1 a 3 horas	7,7	7,7
De 4 a 6 horas	30,8	38,5
De 7 a 9 horas	47,4	85,9
De 10 a 12 horas	12,4	98,3
12 y más horas	1,7	100,0
Total	100,0 (23129)	

Jóvenes que al comenzar a trabajar continuaban estudiando. México, 2000.

Continuaban estudiando	Jóvenes (%)
Sí	54,7
No	44,9
No contesta	0,4
Total	100,0 (23129)

Mecanismo por el cual los jóvenes obtuvieron su primer trabajo. México, 2000.

Mecanismo de obtención	Jóvenes (%)
Por los periódicos	7,1
En una bolsa de trabajo	3,2
Por un amigo	35,6
Me contrató un familiar	32,2
Por la escuela	2,0
Por recomendación	11,7
Otra	7,5
No especificado	0,7
Total	100,0 (23129)

A partir de la información resumida en los cuadros anteriores, acerca del primer trabajo de los jóvenes mexicanos. Responda:

- ¿qué proporción de jóvenes comenzó a trabajar entre los 12 y los 14 años?
- ¿cuáles han sido los mecanismos más frecuentes de obtención del primer trabajo?
- ¿qué proporción de los jóvenes continuaron estudiando cuando empezaron a trabajar?
- ¿qué parte de los jóvenes mexicanos, en su primer trabajo, desarrollaron una actividad de hasta 6 horas por día?
- ¿qué parte lo hicieron entre 7 y 12 horas diarias?

- f) ¿qué porcentaje de los jóvenes mexicanos comenzó a trabajar antes de los 15 años?
- g) ¿qué proporción lo hizo a partir de los 20 años?
- h) ¿cómo se podrían describir las condiciones del primer empleo entre los jóvenes mexicanos?

SEGUNDA PARTE

Primera Encuesta Nacional de Percepción Pública de la Ciencia y la Tecnología. Principales evidencias. Observatorio de Ciencia, Tecnología e Innovación Productiva, Secretaría de Ciencia y Tecnología. Argentina, 2004.

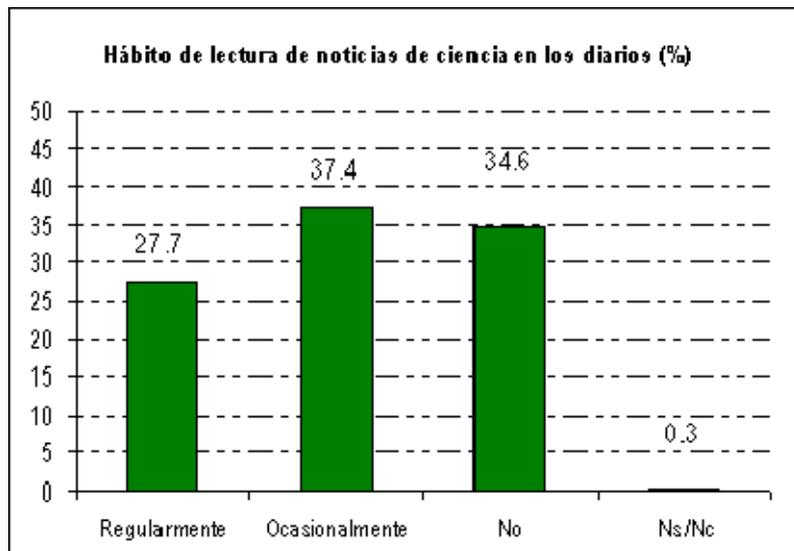
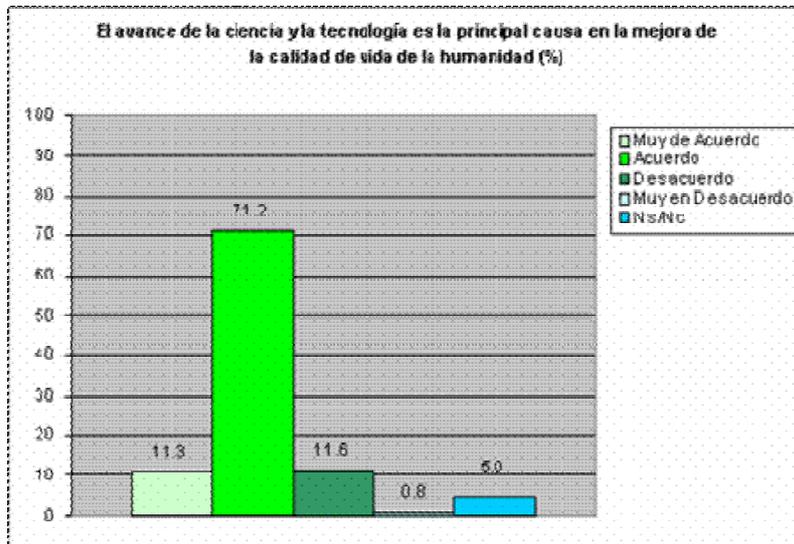
La ficha técnica de la Encuesta es la siguiente:

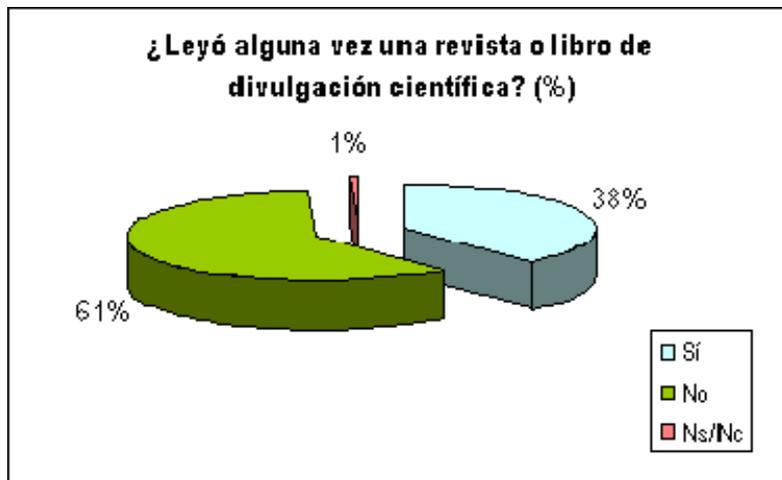
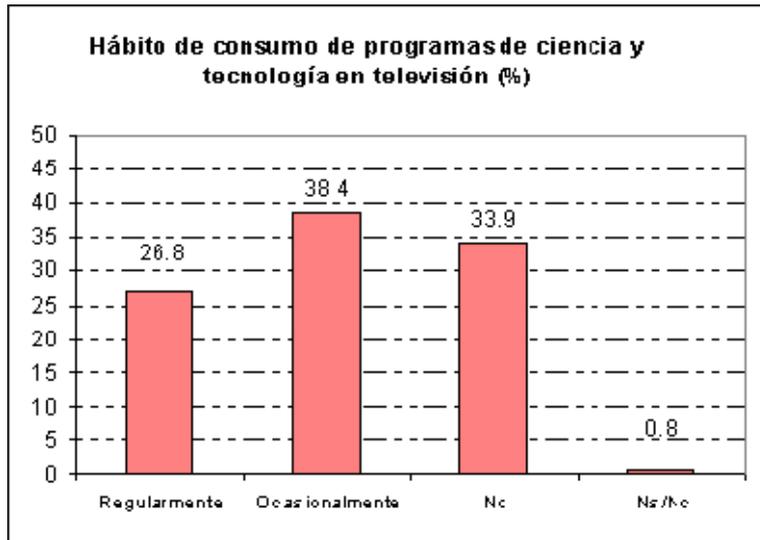
Características generales de la Muestra: Muestra nacional aleatoria y domiciliaria de población adulta urbana, estratificada según región, sexo y edad. Se considera como población urbana a toda aquella residente en localidades de 50.000 habitantes o más.

Tamaño: 1.744 casos, distribuidos en 17 localidades.

De los resultados hemos seleccionado algunos resúmenes relativos a la *valoración y la información sobre ciencia y tecnología*.

A continuación incluimos algunos gráficos correspondientes al informe de este trabajo.





A partir de los gráficos anteriores que han sido extraídos del informe preliminar de este trabajo, describa la valoración que hacen y el consumo de información científica-tecnológica que realizan los habitantes urbanos de la Argentina.



III. Valores que caracterizan una distribución

Una vez organizados los datos en tabla de frecuencias, la que nos ha permitido hacernos una primera idea del comportamiento de los mismos, podemos ahora recurrir a una serie de valores característicos que se utilizan corrientemente para describir la distribución y hacer comparaciones.

Por ejemplo:

- ✓ “El nivel de desarrollo medio es el más frecuente entre los países de América”.
- ✓ “Los países de América, tienen un PBI per cápita promedio de 8297 US\$ PPA”.
- ✓ “La mitad de los países americanos tiene una esperanza de vida de 71,5 años o menos”
- ✓ El 10% más pobre de los ocupados gana en promedio \$210, mientras el 10% más rico tiene un ingreso promedio de \$2380
- ✓ La máxima diferencia que se registra entre los países de América, en términos de su PBI per cápita es de 34.140 US\$ PPA.

En todos los casos se trata de afirmaciones que, o bien describen a toda la distribución con un único número o comparan partes de una distribución.

Distintos criterios que definen diferentes características de la distribución, dan lugar a medidas de distinto tipo.

1. **Medidas de Tendencia Central:** caracterizan la distribución usando un valor que tiende a ubicarse en el centro del conjunto de datos.
2. **Medidas de Posición:** caracterizan la distribución, utilizando valores que la dividen en un determinado número de partes iguales (10%, 20%, etc.).
3. **Medida de variabilidad:** caracterizan la distribución según el nivel de heterogeneidad /dispersión de los datos.

Estos diferentes tipos de medidas se presentan a lo largo de esta unidad.

1. Medidas de Tendencia Central

En los primeros tres ejemplos, cada uno de los conjuntos de datos analizados, para las variables: *Nivel de Desarrollo Humano*, *PBI per cápita*, *Esperanza de vida*, quedan **resumidos/expresados** por un único valor de la variable en estudio: “Medio”, “8297 US\$ PPA” y “71,5 años”. Estas son medidas estadísticas denominadas “**de tendencia central**”.

En este curso analizaremos solo las tres de uso más común:

- el promedio aritmético o “*media aritmética*”,
- la *mediana*, y
- la *moda o modo*.

1.1. La media aritmética

La media aritmética (\bar{x}) es el resultado de **sumar** los valores de la variable observados en cada una de las unidades de análisis y **dividir esa suma** por el total de observaciones.

La consecuencia inmediata de esta definición (que supone la realización de operaciones aritméticas) es que su uso queda **restringido a las variables numéricas**.

La existencia de valores atípicos (distribuciones muy asimétricas) hará que el valor de \bar{x} se desplace del centro de la distribución, de modo tal que –en estos casos– la media aritmética no resultará una medida que represente bien el conjunto de datos.

Por ejemplo, en el caso del PBI *per cápita* el promedio que se obtiene (8297 US\$ PPA)¹¹ es superior a los valores registrados por casi las dos terceras partes del conjunto de países observados que tienen un ingreso muy inferior al promedio obtenido (no superan los 7500 US\$ PPA). De manera que, la media aritmética está desplazada hacia los valores mayores del PBI per cápita. Situación que se corresponde con su forma marcadamente asimétrica a la derecha.

PARA DESTACAR

En el caso de distribuciones marcadamente asimétricas, la media aritmética deja de expresar la tendencia central y consecuentemente deja de ser representativa del conjunto de datos. En estos casos, habrá que recurrir a otras medidas que la reemplacen o la complementen. Generalmente se utiliza con ese fin a la mediana.

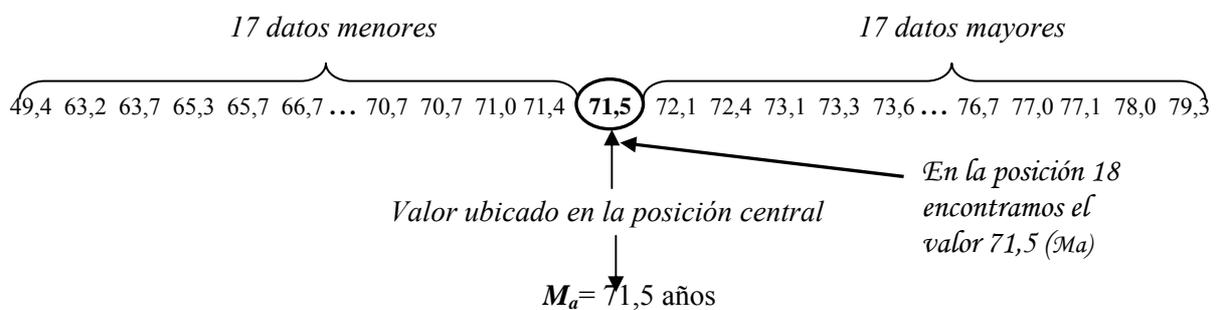
En consecuencia, para decidir si la media aritmética es una buena medida para representar al conjunto de datos, convendrá realizar previamente un análisis gráfico de la distribución.

1.2. La mediana

La mediana (M_a) de una distribución es **el valor de la variable** que ocupa la **posición central** del conjunto de observaciones previamente **ordenadas**. En consecuencia, la mitad de las observaciones serán iguales o inferiores al valor mediana y el otro cincuenta por ciento iguales o superiores a él.

A partir de la definición, se hace evidente que se podrá determinar la mediana en aquellas variables cuyos valores admitan un orden o jerarquía: numéricas y categóricas ordinales.

Considerando la variable *Esperanza de vida* para los países de América (ver matriz de datos) y *ordenando* esos valores en forma ascendente¹², resulta:



Este valor de la mediana nos está diciendo que;
la mitad de los países de América tienen una Esperanza de vida de 71,5 años o menos.

A diferencia de la media aritmética, esta medida **resiste la presencia de valores atípicos**. Retomando el ejemplo anterior, supongamos que Haití hubiese registrado una esperanza de 20 años (en lugar de los 49,4 años), en este caso la mediana del conjunto seguiría siendo la misma. Lo mismo vale para valores atípicos a la derecha (muy altos).

Esta propiedad de la mediana la convierte en una medida de tendencia central que **representa mejor** al conjunto de datos, en el caso de distribuciones muy asimétricas.

¹¹ Este promedio lo obtuvimos a partir de los datos originales de la matriz, sumando todos los valores del PBI *per cápita* y dividiendo esa suma por 35 (número de países en América).
¹² Idéntico resultado se obtendría si el orden en los datos fuera descendente.

1.3. El modo

El modo (M_0) de un conjunto de observaciones es el **valor de la variable que más veces aparece en el conjunto de datos.**

Como consecuencia de la definición, esta medida puede ser obtenida para cualquier tipo de variable. Bastará con identificar en la distribución aquel valor de la variable al que le corresponde la mayor frecuencia. Probablemente, esta sea la medida de resumen a la que generalmente recurrimos casi por intuición, cuando señalamos –mirando una tabla de frecuencia– el valor que más se repite.

En relación a los países de América y utilizando esta medida podemos señalar:

El nivel medio de IDH es el más frecuente entre los países de América. Asimismo, analizando la riqueza generada por habitante, lo más común es encontrar un PBI per cápita de entre 1500 y 7500 US\$ PPA¹³.

PARA DESTACAR

Si bien el modo es una medida *resistente* a valores atípicos, puede ocurrir que **dos o más** valores del conjunto de datos presenten **frecuencias máximas** muy próximas, situación en la cual el modo pierde importancia para resumir el conjunto.



ACTIVIDAD N° 2

En el cuadro siguiente se presentan las medidas de tendencia central más características, correspondientes a las variables "Número de habitantes" y "Gasto en Salud" registrados en la matriz de datos económico-sociales para los países de América.

Medidas de tendencia central	Población (millones)	Gasto Salud (US\$ PPA)
n	34 *	35
Media	25,074	631,7
Mediana	6,050	376,0
Modo	0,100	**

* Saint Kitts y Nevis no registra datos de población.

** no existe un valor más frecuente

Fuente: Datos ec.-soc. para países de América

- a) Utilice estas medidas para hacer una primera descripción de los países de América en términos de sus poblaciones y gasto en salud.
- b) Atendiendo a las propiedades de estas medidas y a los valores registrados para cada una, exprese:
 - b.1. ¿cómo supone Ud. que serán las formas que presentarían estas distribuciones (simétrica, asimétrica)? y,
 - b.2. ¿cuáles de las medidas le resulta más confiable como expresión de la tendencia central?



¹³ Ud. puede notar que en este último caso estamos utilizando un "intervalo modal" en lugar de un único valor. El valor modal podría obtenerse tomando el *punto medio* de este "intervalo modal" (la suma de los límites dividido 2). En el ejemplo podría afirmarse "los más frecuente en América es un PBI per cápita de 4500 US\$ PPA".

2. Medidas de Posición

Las medidas de tendencia central tienen el propósito de representar a la distribución con un único número. Sin embargo, esa descripción que puede resultar útil por constituir la máxima síntesis de una distribución, no da cuenta de la variabilidad del conjunto de datos. Por ejemplo, podemos señalar que:

“el promedio del PBI per cápita es de 8297 US\$ PPA aunque la mitad de los países generan una riqueza por habitante de 6170 US\$ PPA o menos y lo más frecuente es un PBI per cápita de 4500 US\$ PPA.”

Esta información, que nos permite describir sintéticamente al conjunto de los países y eventualmente realizar comparaciones con otras regiones del mundo, no permite analizar diferencias al interior del conjunto de datos. Por ejemplo:

- ✓ ¿cuál es el PBI per cápita de los países más ricos?
- ✓ ¿por debajo de que valor de la riqueza generada por habitante se ubican los países más pobres?
- ✓ etc.

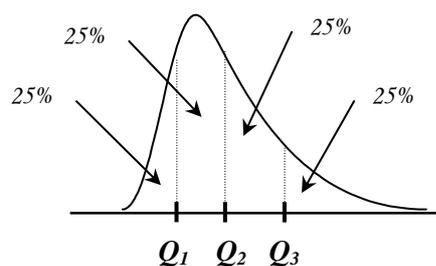
En efecto, muchas veces estamos interesados en distinguir “grupos” al interior del conjunto de datos, que nos faciliten un análisis más detallado de la variable que estamos estudiando e incluso realizar comparaciones entre esos grupos. Para este propósito, disponemos de otras medidas estadísticas, conocidas como: **medidas de posición** (*deciles, quintiles, cuartiles, centiles*).

2.1. Los cuartiles

Los cuartiles son **tres valores** de la variable que dividen en cuatro partes iguales al conjunto de datos previamente ordenado de menor a mayor.

El **primer cuartil** (Q_1) es el valor que deja a su izquierda el 25% de los datos (menores o iguales a él). El **segundo cuartil** (Q_2) coincide con la mediana. El **tercer cuartil** (Q_3) es el valor que deja a su izquierda el 75% de las observaciones (menores o iguales a él).

Gráficamente



Calculados los cuartiles para la distribución de los países de América, según el PBI *per cápita*, estos resultan:

$$Q_1 = 4750 \text{ US\$ PPA} \quad Q_2 = \text{Ma} = 6170 \text{ US\$ PPA} \quad Q_3 = 9200 \text{ US\$ PPA}$$

Los datos correspondientes al año 2002, revelan que una cuarta parte de los países (los 9 de menor riqueza generada por habitante) registran un PBI per cápita de 4750 \$US PPA o menos, además el 25% más rico produce 9200 \$US PPA o más por habitante. La mitad de los países en situación intermedia registran un PBI per cápita de entre 4750 y 9200 US\$ PPA.

2.2. Los quintiles

Son los *cuatro valores* de la distribución, ordenada en forma ascendente, que la dividen en *cinco partes iguales*, cada una de ellas con el 20% de los datos.

El primer Quintil (K_1) es el valor de la variable que deja a su izquierda un 20% de observaciones menores o iguales a él.

El segundo Quintil (K_2) deja a su izquierda un 40% de datos menores o iguales a él.

El tercer Quintil (K_3) deja a su izquierda un 60% de datos menores o iguales a él.

El cuarto Quintil (K_4) deja a su izquierda un 80% de datos menores o iguales a él.

En el caso de la variable PBI *per cápita* los quintiles son:

$$K_1 = 4224 \text{ US\$ PPA} \quad K_2 = 5428 \text{ US\$ PPA} \quad K_3 = 6896 \text{ US\$ PPA} \quad K_4 = 10032 \text{ US\$ PPA}$$

El 20% más pobre entre los países de América generan una riqueza por habitante de 4224 US\\$ PPA o menos;

El mismo quintil se también podría interpretar como:

un 80% de los países de América generan una riqueza de 4224 US\\$ PPA o más.

2.3 Los deciles y centiles

Son los valores de la distribución ordenada en forma ascendente que la dividen en diez o cien partes iguales respectivamente, cada una de ellas con la misma cantidad de datos (10% y 1% respectivamente).

3. Construcción de una “buena imagen” de la distribución

Las medidas de resumen presentadas hasta aquí (frecuencias, medidas de tendencia central y de posición) nos han permitido resumir el conjunto de datos apoyándonos en diferentes criterios. Todas esas formas de resumen pueden contribuir de distintas manera a la descripción de una característica (variable) de los individuos analizados.

Ahora bien, ante una situación concreta no se utilizan todas estas medidas simultáneamente para realizar la descripción, sino que habrá que seleccionar aquellas que resultan pertinentes en función de: el tipo de variable, la forma de la distribución y las preguntas de interés.

Independientemente de las aplicaciones particulares de cada medida (comentadas oportunamente), cuando queremos describir una distribución buscaremos expresar los rasgos más salientes del conjunto de datos para “generar una buena imagen” del mismo. Esta intención exige “*hacernos cargo*” de *todo el conjunto de datos*, dando cuenta de la variabilidad observada.

Una forma aceptada para describir adecuadamente un conjunto de datos numéricos, es la que se conoce como el “resumen de los cinco números”.

El **resumen de los cinco números** consiste en la utilización de las siguientes medidas:

X_{\min} : el mínimo Q_1 : el cuartil 1 M_a : la mediana Q_3 : el cuartil 3 X_{\max} : el máximo

Observe que estos números nos permiten indicar entre qué valores varía la variable (X_{\min} y X_{\max}), nos da una medida del centro de la distribución, y nos permite describir usando los cuarteles 1 y 3, tanto el 50% de valores centrales como la cuarta parte de valores que se ubican en los extremos de la distribución.

El PBI *per cápita* de los países de América, puede ser descripto como sigue, por el resumen de los cinco números:

$$X_{\min} = 1610 \text{ US\$PPA} \quad Q_1 = 4750 \text{ US\$PPA} \quad M_a = 6170 \text{ US\$PPA} \quad Q_3 = 9200 \text{ US\$PPA} \quad X_{\max} = 35750 \text{ US\$PPA}$$

La mitad de los países americanos no superan los 6170 US\$ PPA, aunque la riqueza generada por habitante varía entre 1610 y 35750 US\$ PPA. Por otro lado, el 50% de los países que tienen un PBI per cápita intermedio se ubica entre 4750 y 9200 US\$ PPA. Además, el cuarto de países “más pobres” genera una riqueza por habitante entre 1610 y 4750 US\$PPA y en el 25% “más rico” el PBI per cápita está entre 9200 y 35750 US\$PPA.

Análogamente, se pueden incorporar otros números que permitan describir partes más específicas de la distribución. Por ejemplo, si incorporamos los deciles 1 y 9, en lo que constituiría un resumen de los 7 números, tendríamos.

$$X_{min} = 1610 \text{ US\$ PPA} \quad M_a = 6170 \text{ US\$ PPA} \quad X_{max} = 35750 \text{ US\$ PPA}$$
$$D_1 = 2992 \text{ US\$ PPA} \quad Q_1 = 4750 \text{ US\$ PPA} \quad Q_3 = 9200 \text{ US\$ PPA} \quad D_9 = 14142 \text{ US\$ PPA}$$

Con esta nueva información podemos señalar además que:

El 10% más pobre de los países americanos, producen anualmente por habitante, 2992 US\$ PPA o menos, mientras que un 10% más rico tienen un PBI per cápita de 14142 US\$ PPA o más.

La presentación de estas formas establecidas de describir una distribución, tiene el propósito de mostrar combinaciones posibles de algunas medidas de resumen para construir una buena imagen de la distribución.

Sin embargo, cada investigador podrá recurrir a otras combinaciones de medidas que expresen bien los rasgos que se consideren salientes en función de las particularidades de cada distribución. Por ejemplo: si la distribución es asimétrica, podrá utilizarse la media con la precaución de complementarla por ejemplo con la mediana e incluso algunas frecuencias que resulten significativas.



ACTIVIDAD N° 3

PRIMERA PARTE

Sobre informes, técnicas y subjetividades

Dos asesores políticos, analizan la situación social de la ciudad de Formosa tomando como referencia los datos disponibles sobre Ingresos Familiares y Distribución del Ingreso según Nivel Educativo a los efectos de informar a sus respectivos bloques legislativos. Los informes resultantes, en sus aspectos más salientes, señalaban lo siguiente:

Informe 1:
... es importante destacar que el ingreso medio de las familias de la ciudad de Formosa es de \$684 (Cuadro N° 2), valor que supera los \$550 necesarios para cubrir la canasta básica de una familia tipo en esta ciudad. Además el 30% de las familias con mayores ingresos superan los \$720, e incluso el 10% más favorecido tiene un ingreso medio de \$2219.
Por otro lado, la distribución de los ingresos es bastante equitativa, lo que se evidencia al analizar la poca variación existente entre los ingresos medios del 50% central de las familias (decil 3 a 7). Las familias del decil 7, tienen un ingreso medio que supera en sólo 2,2 veces el promedio para las familias del decil 3.
Por otro lado, es interesante rescatar que entre aquellos individuos de mejores ingresos, más del 22% ha cursado a lo sumo el nivel primario (Cuadro N° 1), lo que estaría expresando condiciones de mercado laboral que ofrecen buenas oportunidades aún a los que no poseen un alto nivel de educación.
Así, sería recomendable establecer políticas que den algún tipo de contención (planes trabajar, bolsas del programa SOMA, refuerzo alimentario, etc.) para aquel 20% de familias que por situaciones diversas se encuentran en una situación de desprotección.

Informe 2:
... es importante destacar que el ingreso promedio de las familias de la ciudad de Formosa es de \$684, pero al mismo tiempo el 70% de las familias tienen un ingreso inferior a ese valor y el 60% no alcanzaría

a cubrir los \$550 necesarios para solventar la canasta básica de una familia tipo, e incluso el 20% de las familias (deciles 1 y 2) no alcanzan a cubrir el 50% de esa canasta.

La desigual distribución del ingreso se hace evidente cuando se observa que el ingreso medio de las familias más favorecidas (último decil) es 17,5 veces mayor que el ingreso del 10% de menor ingreso (decil 1).

Por otra parte, es interesante destacar la fuerte relación que se observa entre el nivel de educación y los ingresos individuales. Se puede ver que entre aquellos de menores ingresos (decil 1 a 4), el 51% a lo sumo completó el nivel primario, y sólo el 5,8% han alcanzado un nivel superior de educación. En tanto que estos valores para los estratos de mayores ingresos son: el 22,7% han cursado solamente el nivel primario y el 38,4% tienen un nivel de educación superior.

De todo esto surge que una política orientada a mejorar las condiciones de ingreso de las familias está íntimamente vinculada, con una mayor oportunidad de acceso a la educación.

Cuadro 1- Población Económicamente Activa Clasificada por Ingreso Total Individual según Nivel de Educación

Ingreso Total Individ.	TOTAL GENERAL	Nivel de Educación							
		Primario		Secundario		Sup. y Univ.		Sin Instruc.	
		Incomp.	Comp.	Incomp.	Comp.	Incomp.	Comp.	Incomp.	Comp.
TOTAL	100,0	12,0	27,6	20,5	22,1	5,6	11,0	1,3	
Entre el decil 1 y 4	100,0	19,9	31,1	26,3	14,2	3,4	2,4	2,6	
Entre el decil 5 y el 8	100,0	10,1	30,9	18,3	23,8	5,4	10,6	0,8	
Entre decil 9 y el Máx.	100,0	2,4	20,3	13,5	25,4	7,4	31,0	--	
Sin Ingresos	100,0	9,8	11,1	23,4	40,8	11,3	2,5	1,2	

-- Valor cero.

Las estimaciones de este tabulado han sido ajustadas de acuerdo al valor de la proyección demográfica elaborada por INDEC para este aglomerado.

Fuente: INDEC - EPH FORMOSA - OCTUBRE DE 1997

Cuadro 2- Hogares según Escala de Ingreso Total Familiar

Número de DECIL	Ingreso Desde	Ingreso Hasta	Población por DECIL	Porcentaje de personas	Ingreso Tot por DECIL (miles)	Porcentaje del ingreso	Ingreso medio por DECIL
1	20	200	13.235	7,4	549	1,9	127
2	200	250	16.945	9,4	976	3,3	225
3	250	330	14.353	8,0	1.281	4,3	296
4	330	400	18.019	10,0	1.603	5,4	371
5	400	500	17.267	9,6	1.901	6,4	439
6	500	600	19.125	10,6	2.316	7,8	533
7	600	710	21.374	11,9	2.796	9,4	652
8	720	980	21.493	12,0	3.584	12,1	830
9	980	1.330	18.312	10,2	4.935	16,7	1.134
10	1.330	10.449	19.626	10,9	9.668	32,7	2.219
Hog.			179.749	100,0	29.609	100,0	684

Total de Hogares: 45.527 - Población: 187.989 - Declaran ingresos: 43.288 (95,1%)

Las estimaciones de este tabulado han sido ajustadas de acuerdo al valor de la proyección demográfica elaborada por INDEC para este aglomerado.

Fuente: INDEC - EPH FORMOSA - OCTUBRE DE 1997

A partir de la lectura de los informes 1 y 2, responda:

- o ¿Existen afirmaciones que aporten los analistas y no se revelen en los cuadros?
- o ¿Cuál de estos informes, se ajusta más a la realidad que se desprende de los datos? ¿Por qué?

SEGUNDA PARTE

Complete la descripción de los países de América realizada en la actividad n° 3, con la información presentada en la siguiente tabla.

Medidas de Posición	Población (millones)	Gasto Salud (US\$PPA)
Cuartil 1	0,500	242,0
Cuartil 3	14,900	574,5
Quintil 1	0,300	226,2
Quintil 2	3,540	344,6
Quintil 3	8,520	416,8
Quintil 4	25,840	649,6
Mínimo	0,100	56,0
Máximo	291,000	4887,0



4. Medidas de Variabilidad

Hemos desarrollado hasta aquí una serie de recursos estadísticos que constituyen resúmenes de una distribución que utilizamos para describir lo observado; la integración de estas medidas nos permiten dar cuenta de la variabilidad de los datos. Aún así, en diferentes situaciones resulta conveniente contar con un único número que exprese la variabilidad de los datos que estamos describiendo.

Estas medidas de variabilidad completan la descripción que podríamos realizar del conjunto de datos utilizando las medidas de resumen ya presentadas.

Ahora bien, en muchas situaciones usaremos para resumir una distribución, solamente una medida de tendencia central (por ejemplo la media aritmética). En esos casos, las medidas de variabilidad adquieren particular importancia para indicar sintéticamente la dispersión de los datos en torno a ese valor central y, consecuentemente, evaluar la calidad de representación del mismo.

Medir la variabilidad supone encontrar un número que exprese cuán diferentes son entre sí los datos del conjunto. Hay diferentes formas de hacerlo y presentaremos aquí las medidas más utilizadas en el caso de variables numéricas.

4.1. El rango: indica la *amplitud* o *recorrido* de variación de los datos y expresa la mayor diferencia que puede observarse en el conjunto.

El Rango es la diferencia entre el máximo y el mínimo valor de la variable ¹⁴.

$$R = x_{\text{máx}} - x_{\text{mín}}$$

La riqueza generada por los países de América se registra en una amplitud de variación de 34140 US\$PPA (R = 35750 – 1610). A su vez, entre el país de América con mayor esperanza de vida y aquél que registra la menor expectativa de sobrevivencia, hay una diferencia de 29,9 años (R=79,3-49,4).

¹⁴ Cuando los datos están agrupados en intervalos de clase el rango se *estima* haciendo la diferencia entre el límite superior de la última clase y el límite inferior de la primera (aunque éstos no sean estrictamente el máximo y el mínimo valor observado).

Aceptando que estos valores de rango obtenidos expresan diferencias reales entre la mejor y la peor situación de los países americanos, debemos observar que *pueden producirnos una imagen distorsionada* en relación al conjunto de los países analizados. En el caso del PBI *per cápita*, el campo de variación que expresa esta medida está muy afectado por los altos niveles de riqueza que presentan Estados Unidos y Canadá. A su vez en el caso de la esperanza de vida, la impactante diferencia de casi 30 años que se observa, se debe a la baja esperanza de vida de Haití que no alcanza a los 50 años¹⁵.

Otra **desventaja** de esta medida es que, al tomar en cuenta solamente dos valores (los extremos) de la distribución, no está dando cuenta de la dispersión que puede ocurrir de diferentes maneras si consideramos la **totalidad de los datos**. Dos conjuntos de datos pueden tener la misma amplitud de variación (rango) pero en uno de ellos la mayoría de los datos están muy concentrados en torno a los valores centrales y en el otro no.

4.2. Variancia y desviación estándar:

Estas medidas -a diferencia del rango- toman en cuenta todo el conjunto de datos y se construyen a partir de la diferencia entre el valor observado en cada individuo y la media del conjunto (**desvíos individuales**: $d_i = x_i - \bar{x}$).

Para el ejemplo del PBI *per cápita* de los países americanos, hemos calculado los 35 desvíos individuales a la media general, de los cuales presentamos sólo algunos en la siguiente tabla:

País	Haití	Bolivia		Panamá		Canadá	EEUU	Media
PBI per cápita	1610	2460	...	6170	...	29480	35750	8297
Desvíos individuales a la \bar{x}	-6687	-5837		-2127		21183	27453	

Se puede ver que, mientras Haití está 6.687 US\$PPA por debajo de la media, Estados Unidos tiene un PBI per cápita que está 27.453 US\$PPA por encima de ese valor. Así, habrá un conjunto de países cuyos valores son inferiores a la media y en consecuencia registran desvíos individuales negativos, mientras que otro conjunto de países (los que están por encima de la media) presentarán desvíos individuales positivos. Tenemos así que los desvíos están indicando la variabilidad existente en los datos en relación a una medida característica del conjunto. Entonces el problema de medir la variabilidad se reduce a obtener un único número que exprese esa variabilidad observada a través de los desvíos.

Promediar los 35 desvíos con respecto a la media presenta el problema de *que la suma de los desvíos a la media siempre es cero*¹⁶; una forma de resolver esto es elevar al cuadrado los desvíos y, recién entonces promediarlos, obteniendo de esta manera una medida de variabilidad que se conoce como **Variancia**.

Variancia (σ^2): es el promedio de los cuadrados de los desvíos a la media aritmética.

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

El cálculo de la variancia para el ejemplo anterior

$$\text{es: } \sigma^2 = \frac{\sum \text{desvíos}^2}{n} = \frac{(-6687)^2 + (-5837)^2 + \dots + (-2127)^2 + (21183)^2 + (27453)^2}{35}$$

$$\sigma^2 = 48.149.291$$

Al elevarse al cuadrado los desvíos, la unidad de medida de la variancia hace que el resultado obtenido no sea interpretable en términos de la variable que se está analizando. Para obtener una

¹⁵ Muchas veces, por esta razón se recurre a otras medidas como por ejemplo el rango intercuartil ($RQ = Q_3 - Q_1$), que mide la extensión de variación del 50% central de los datos.

¹⁶ Una propiedad de la media es que la suma de los desvíos individuales con respecto a ella, siempre es cero. $\sum_{i=1}^n (x_i - \bar{x}) = 0$

medida de la variabilidad expresada en la unidad de medida original, se calcula la *raíz cuadrada de la variancia*, que resulta en una nueva medida llamada **desvío estándar** (σ) : $\sigma = \sqrt{\sigma^2}$

En el ejemplo, el desvío estándar será: $\sigma = \sqrt{48.149.291} \Rightarrow \sigma = 6939\text{US\$PPA}$

“El PBI per cápita de los países de América, se dispersan en promedio en torno a la media en 6939 US\$PPA”.

4.3. Las medidas relativas de variabilidad

En aquellas situaciones en que necesitamos decidir si un conjunto de datos es más homogéneo (o heterogéneo) que otro, podría resultar engañoso recurrir a medidas absolutas como las presentadas precedentemente.

Tenga en cuenta que –por ejemplo– un desvío estándar de \$99 en la variable ingreso de un conjunto cuyo promedio es de \$800 representa una variabilidad mayor a un desvío de \$150 si el ingreso promedio del conjunto es de \$2600. Ante este tipo de situaciones, **la comparación requiere construir medidas relativas de variabilidad.**

La más utilizada es el **coeficiente de variación**, la cual pone en relación el desvío estándar y la media aritmética del conjunto de datos.

Coeficiente de Variación (CV): $CV = \frac{\sigma}{\bar{x}} \cdot 100$

Generalmente expresado en porcentaje, es una **medida relativa de variabilidad que se obtiene como el cociente entre el desvío estándar y la media aritmética.**

	América	CEE
N	35	15
\bar{x}	8297 US\$PPA	28779 US\$ PPA
σ	6939 US\$PPA	9703 US\$ PPA
CV	83,6%	33,7%

Comparando el PBI *per cápita* de los países de América con los que conformaban la Comunidad Económica Europea a 1995, podemos ver que la riqueza generada en la CEE es –en promedio– muy superior a la de los países americanos. Además, si bien en términos absolutos la dispersión es mayor entre

los países europeos, el coeficiente de variación pone en evidencia la homogeneidad relativa que existe entre los mismos (33,7%) frente a la heterogeneidad de los países americanos (83,6%).



ACTIVIDAD N° 4

Con la información presentada en la siguiente tabla:

- Expresar en una frase el significado de 873,1 US\$ PPA registrado como valor del desvío estándar de la variable Gasto en Salud.
- ¿Qué objeción podría hacer al rango como medida de variabilidad de estas variables? (*sugerencia*: explore los valores originales de esta variables en la matriz de datos).
- ¿en cuál de las dos variables son más homogéneos los países americanos?

Medidas de variabilidad	Población (millones)	Gasto Salud (US\$PPA)
Desvío Estándar	57,279	873,1
Rango	291,000	4831,0
CV	228,444	138,2



5. La Curva de Lorenz y el Índice de Gini

La *Curva de Lorenz* y el *coeficiente de Gini* son recursos ampliamente utilizados en el análisis de las desigualdades de la distribución de ciertas variables como por ejemplo el ingreso, el consumo o la tierra.

Para presentar estas herramientas, analizaremos la *distribución del ingreso per cápita familiar* en la Argentina, a partir de datos de la Encuesta Permanente de Hogares correspondiente al primer trimestre del año 2005. Los datos se resumen en una tabla que muestra a los hogares distribuidos en diez grupos, de modo tal que cada uno de ellos incluye un 10% de los hogares. Esos grupos quedan determinados por los deciles de la variable ingreso.

Distribución de los Hogares según ingreso *per cápita* familiar – Argentina. 1er. Trimestre 2005

Decil	Escala Ingresos (*)	Hogares (en miles)	Ingreso total por Decil (miles \$)	Porcentaje de Ingreso	Ingreso medio por decil (**)
1	2 - 98	681	215.664	2,4	60
2	98 - 150	681	366.609	4,0	123
3	150 - 200	682	492.846	5,4	174
4	200 - 266	680	589.546	6,5	230
5	266 - 325	680	637.955	7,0	296
6	325 - 410	681	763.352	8,4	369
7	410 - 525	681	905.956	9,9	462
8	525 - 703	682	1.073.808	11,8	611
9	703 - 1075	680	1.370.674	15,1	851
10	1075 - 13013	680	2.689.753	29,5	1.845
Total		6808	9.106.163	100,0	396

(*) Cada intervalo en la escala de ingresos tiene por límite superior al valor del decil.

(**) El ingreso medio de cada decil se obtiene como el cociente entre el ingreso total del decil y el número total de hogares.

Fuente: INDEC – EPH. 2005

Si no existiera desigualdad entre los hogares (*situación de equidistribución*), a cada hogar le correspondería el mismo ingreso. En ese caso, al primer 10% de los hogares le correspondería un 10% del ingreso total; a un 20% de los hogares le correspondería el 20% del total de los ingresos *per cápita*, al 30% el 30% y así sucesivamente. En la tabla puede apreciarse rápidamente la desigualdad en los ingresos *per cápita* de los hogares. Se puede apreciar que el 10% de los hogares “más pobres” tienen un ingreso *p.c.* promedio de \$60 mientras, en el otro extremo de la escala, el 10% más rico, alcanza un ingreso medio *p.c.* de \$1845. Esta situación se suele resumir en un concepto que se denomina “brecha” y que mide la diferencia entre los ingresos de los dos extremos de la escala¹⁷. Así, la brecha de ingresos se puede medir como:

$$Brecha = \frac{\text{Ingreso medio del 10\% más rico}}{\text{Ingreso medio del 10\% más pobre}}$$

En nuestro ejemplo:

$$Brecha = \frac{1845}{60} = 30,8$$

¹⁷ La brecha de ingreso se puede medir calculando “cuánto veces más gana el 10% más rico con respecto al 10% más pobre, o bien la diferencia entre el 20% más rico y el 20% más pobre.

Así, en el primer trimestre de 2005 en Argentina, la brecha entre ricos y pobres es de 30,8. Es decir, en el 10% de los hogares más ricos el ingreso promedio p.c. es casi 31 veces superior al ingreso medio p.c. del 10% más pobre. Lo cual también puede verse en la tabla observando que el 10% de hogares (681) con menores ingresos acumula \$215.664.000, es decir, el 2,4% del total de los mismos, en tanto que al 10% de los hogares más ricos le corresponden el 29,5%.

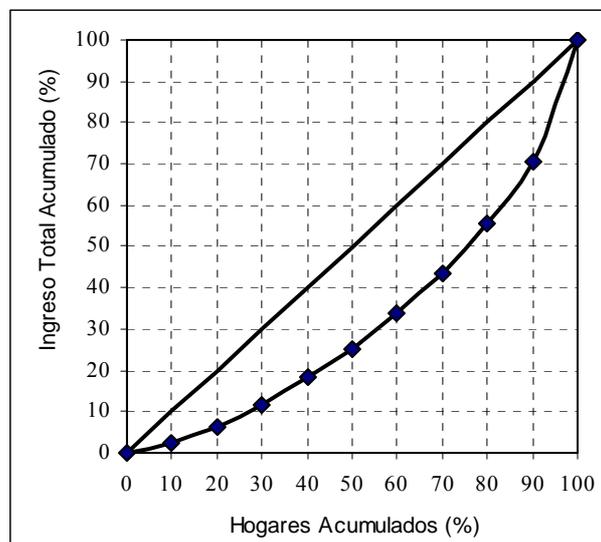
Esta forma de dar cuenta de las desigualdades a través de la brecha, toma en cuenta sólo las situaciones extremas. Una manera de evaluar la equidad/inequidad del ingreso pero teniendo en cuenta la totalidad de los hogares, se logra con el uso de la **gráfica de Lorenz** y el **coeficiente de Gini**.

La gráfica que se conoce como **curva de Lorenz**, pretende mostrar en que medida la situación que se está analizando se aleja de la condición planteada como de equidistribución/equidad. Para lograr esto, se utiliza un sistema de ejes cartesianos en el cual cada punto representado queda definido por el *porcentaje acumulado de hogares* y sus correspondientes *porcentajes acumulados de ingresos*.

La siguiente tabla se obtiene acumulando los porcentajes de hogares y porcentajes de ingreso presentados precedentemente. En la curva de Lorenz que acompaña a la tabla, **la recta** representa la situación de equidistribución en la cual al primer 10% de los hogares le corresponde el 10% de los ingresos, al primer 20% de los hogares el 20% de los ingresos y así sucesivamente. A su vez la curva se construye a partir de los ingresos observados en los hogares; así tenemos que el primer punto de abajo hacia arriba queda definido por el primer 10% de los hogares los cuales acumulan el 2,4% de los ingresos, el segundo punto esta definido por el 20% más pobre, que acumula el 6,4% de los ingresos, y así sucesivamente hasta llegar al 100% de los hogares que acumulan el 100% de los ingresos.

Distribución de los Hogares según deciles de ingreso p.c. familiar - Argentina. 1er. Trim. 2005.

Decil	Escala Ingresos	Hogares Acum. (%)	Ingresos Acum. (%)
1	2 - 98	10	2,4
2	98 - 150	20	6,4
3	150 - 200	30	11,8
4	200 - 266	40	18,3
5	266 - 325	50	25,3
6	325 - 410	60	33,7
7	410 - 525	70	43,6
8	525 - 703	80	55,4
9	703 - 1075	90	70,5
10	1075 - 13013	100	100,0



En el gráfico se observa que la curva (lo observado) se aleja de la recta (situación teórica de perfecta igualdad en la distribución de los ingresos). Intuitivamente podemos apreciar que, cuanto más alejada esté la curva de la recta (mayor sea el *área de concentración*¹⁸), mayor será la desigualdad y en consecuencia mayor la concentración de los ingresos. Es decir, la curva de Lorenz nos permite hacer una apreciación global del grado de concentración de la variable en estudio, en este caso de los ingresos.

¹⁸ Se conoce como área de concentración al área que queda encerrada entre la recta de equidistribución y la curva.

El **índice de Gini** (I_G) mide el grado de concentración / distribución que apreciamos visualmente en la curva de Lorenz. Este índice toma valores entre 0 y 1. Un **valor de 0** expresa la situación en que la curva coincide exactamente con la recta de equidistribución (área de concentración nula) y por lo tanto debe interpretarse como una situación en la que se da la perfecta igualdad en la distribución. En el otro extremo, un **valor de 1**, expresa la situación de concentración absoluta o máxima desigualdad que se corresponde con una curva con alejamiento máximo con respecto a la recta (la curva coincide con el triángulo inferior definido por la recta de equidistribución y el eje horizontal).

$$0 \leq I_G \leq 1$$

Estas situaciones, donde el coeficiente asume los valores 0 o 1, son teóricas. En la práctica, nos encontraremos con valores intermedios, los cuales indicarán situaciones de mayor desigualdad o concentración cuanto más próximos estén del valor 1.

La ventaja de contar con este índice, es la de constituir un indicador objetivo para apreciar el nivel de concentración que, a su vez, facilita las comparaciones (a través del tiempo, de distintas poblaciones e incluso de distintos recursos en una misma población), principalmente cuando las diferencias entre las curvas de Lorenz no son muy claras.

Indicadores de la distribución del ingreso en cinco países de América Latina

Países	Años	Coefficiente de Gini	Participación del primer decil (%)	Participación del último decil (%)
Argentina (Bs. As.)	1980	0,365	3,0	30,6
	1986	0,393	2,6	34,0
	1990	0,422	2,3	34,2
	1992	0,408	2,3	31,6
Brasil (San Pablo-Río de Janeiro)	1979	0,448	2,1	36,1
	1987	0,526	1,4	44,0
	1990	0,503	1,5	37,6
Colombia (Bogotá)	1980	0,520	1,3	40,8
	1985	0,457	1,5	36,2
	1990	0,478	1,4	36,0
	1992	0,448	1,8	34,0
Chile (Gran Santiago)	1987	0,507	1,4	40,0
	1990	0,487	1,7	39,6
	1992	0,512	1,7	44,4
México (Zona de alta densidad)	1984	0,324	3,1	26,7
	1989	0,432	2,4	37,6
	1992	0,426	2,4	36,2

Fuente: CEPAL Boletín n° 612-Feb. 1998, sobre la base de tabulados especiales de las encuestas de hogares.

El análisis de los datos de la tabla anterior -entre otras cosas- revela que:

"Durante el período analizado, la concentración del ingreso se redujo en Colombia, aumentó en tres países (Argentina, Brasil y México) y se mantuvo con ciertas oscilaciones en Chile.

En el caso de Colombia, se aprecia una redistribución de alrededor del 7 % del ingreso desde los grupos superiores hacia los grupos medios altos. La participación de los grupos de menores ingresos se eleva en una proporción muy reducida. Se concluye que allí se dio un proceso redistributivo pero que este no alcanzó significativamente a los estratos pobres.

En los países en que aumentó la concentración (Argentina, Brasil y México) se observan reducciones en la participación del estrato más bajo y, en ciertos casos, significativas redistribuciones a favor del estrato superior, ..." (CEPAL, Boletín n° 612-Feb. 1998).



ACTIVIDAD Nº 5

PRIMERA PARTE

Lea el artículo siguiente publicado en el diario Clarín de Buenos Aires y:

- a. Identifique las medidas estadísticas que se utilizaron en el análisis.
- b. Reflexione sobre los distintos pasos que fueron necesarios realizar a partir de los datos originales para obtener la información presentada en el artículo.
- c. De acuerdo a la información presentada para el período analizado, ¿qué piensa Ud. que ha ido ocurriendo con la curva de Lorenz?

LA CRISIS EN LOS BOLSILLOS: LA PEOR DISTRIBUCION DE LA RIQUEZA DESDE QUE HAY DATOS

Los más ricos ganan 28 veces más que los más pobres

El 10% de la población con más recursos recibe el 37,3% de los ingresos. El 10% más pobre, sólo el 1,3%. La brecha creció por la caída bajo la línea de pobreza de los sectores medios.

La Argentina tiene hoy la peor distribución del ingreso desde que el INDEC mide, a partir de la década del 70, cuánto ganan y cómo se reparte la riqueza de los argentinos.

De acuerdo a las cifras oficiales de Capital y Gran Buenos Aires, en octubre pasado (última medición) el 10% más rico de la población recibió el 37,3% de los ingresos totales. En tanto, al 10% más pobre le correspondió apenas el 1,3%. Así, los más ricos recibieron 28,7 veces más que los más pobres. En 1974 esa brecha era de 12,3 veces, según los datos elaborados por la consultora Equis.

Como desde octubre en adelante creció el desempleo y la inflación deterioró el poder adquisitivo de los salarios y de la gente con ingresos fijos, se estima que ahora esa brecha debe superar las 30 veces.

Aún así, el INDEC y los analistas privados reconocen que esta brecha es superior porque los sectores más ricos, aunque la encuesta es confidencial, declaran menos ingresos por razones impositivas, mientras los más pobres, por el llamado efecto "vergüenza", declaran más de lo que reciben. Por eso se estima que la brecha entre ricos y pobres sería de más de 45 veces: el 10% más pobre recibiría menos del 1% y los más ricos, 40% del ingreso total.

El aumento de la brecha entre ricos y pobres significa que se incrementó la transferencia de riqueza de los sectores de menores recursos hacia los estratos más altos de la sociedad. Y que "se consolidó un perfil de distribución del ingreso regresivo que se explica por la caída bajo la línea de pobreza de los sectores medios que se encontraban en riesgo de empobrecimiento y que no han logrado sostener sus ingresos familiares sobre la línea de 120 pesos mensuales por integrante", señala Artemio López, de Equis.

El primer vuelco se produjo a partir de 1986, con la crisis del Plan Austral, y alcanzó su pico con la hiperinflación de 1989. Ese año, la brecha saltó a 23,1 veces por el fuerte deterioro que tuvieron los ingresos de los sectores con ingresos fijos ante la velocidad hiperinflacionaria. Así la brecha de 1989 fue de 23,1 veces cuando el año anterior era de 18 veces.

A partir de los años 90 hubo una mejora en la distribución de los ingresos que duró poco. Ya en 1993 los más pobres vieron reducida su porción de la torta, que se acentuó con la recesión de 1994, la crisis del Tequila de 1995 y el aumento del desempleo y el trabajo no registrado. La distancia entre ricos y pobres volvió a crecer y ya en 1998 y 1999, con deflación de precios, superó la marca de 1989. Así, Carlos Menem concluyó su mandato con una distribución del ingreso peor que la que había heredado y luego Fernando de la Rúa superó la marca de su antecesor.

Así, por la alta inflación de los 80 y por el desempleo, el trabajo en negro y la recesión en los 90, siguió aumentando la desigualdad social. Y ahora con la vuelta de la inflación, en un contexto que sigue siendo recesivo, se estima que la brecha volverá a pegar un salto: es sabido que los salarios y los ingresos fijos son los grandes perjudicados por los aumentos de precios.

Diario Clarín, Buenos Aires. Domingo 31 de marzo de 2002



IV. La Relación Entre Variables

Superado el primer análisis exploratorio de los datos, que nos permitiera hacer una primera descripción del conjunto de individuos en términos de cada una de las variables observadas y seguramente responder algunas preguntas iniciales, estamos ahora en condiciones de responder interrogantes que ponen en juego dos variables simultáneamente. Por ejemplo: ¿cuánto mayor es la riqueza generada por un país, es menor el analfabetismo?; ¿hay relación entre el nivel de analfabetismo y la esperanza de vida de los países?, ¿los niveles de alfabetización se relacionan con los niveles de gasto en educación?, ¿los niveles de gasto en investigación y desarrollo se relacionan con los niveles de exportación de manufacturas y los niveles de desempleo?

Con respecto al caso de los desempleados en Cochabamba, sería válido preguntarnos ¿el tiempo de permanencia en situación de desempleado tiene relación con el tipo de actividad que busca?, ¿y con el nivel de educación?

Estas preguntas conducen a realizar un **análisis de la relación entre dos variables**. Decimos que, para un mismo conjunto de individuos:

Dos variables están relacionadas o asociadas, cuando ciertos valores de una de ellas tienden a presentarse en forma conjunta con ciertos valores de la otra.

Por ejemplo: “Valores altos de PBI se corresponden con valores bajos de analfabetismo”; “entre los niveles altos de educación se observan los menores tiempos de permanencia como desocupado”; “entre las mujeres es más frecuente el trabajo informal”, etc.

La relación entre dos variables puede ser el resultado de que los valores de una de ellas causan o provocan los valores de la otra, en estos casos hablamos de una relación causal. Ahora bien, el encontrar relaciones estadísticas entre dos variables no implica necesariamente la existencia de causalidad. Ej. podemos suponer que el nivel de educación “explica” la condición de ocupación de un individuo, en cambio, no es tan evidente que exista una relación de causalidad entre la disponibilidad de agua dentro de la vivienda y el estado de nutrición de los niños que habitan en ella, seguramente aquí, se podrían identificar “terceras variables” (la pobreza) que explican esa asociación.

En este capítulo presentaremos un conjunto de recursos estadísticos básicos que nos permitirán analizar la relación entre dos variables. Esas herramientas son diferentes según el tipo de variables que se ponen en juego: *Tablas de contingencia* (para dos variables categóricas), *Comparación de medias* (una categórica y una numérica), *Análisis de correlación* (dos variables numéricas).

1. La Relación entre Variables Categóricas

Cualquier análisis de la relación entre dos variables implica la reducción de la información contenida en la matriz para ese par de variables, de manera tal que permita analizar el comportamiento simultáneo de las dos variables en ese conjunto de individuos.

En el caso en que ambas variables son categóricas, la herramienta consiste –inicialmente– en construir una tabla en la cual se pueden observar las frecuencias de ambas variables simultáneamente. Para presentar este recurso consideremos los 177 países que aparecen en el informe del IDH-2004 clasificados según el Nivel de Esperanza de vida (Bajo, Medio y Alto) y el Tipo de asentamiento predominante en la población (Rural, Compensado y Urbano).

Esta información la podemos organizar contando en la matriz de datos el número de países que responde a la doble condición: *Esperanza de vida baja* y *Asentamiento predominantemente rural*; *Esperanza de vida baja* y *Asentamiento predominante compensado*, ... ; *Esperanza de vida alta* y *Asentamiento predominante urbano*. Tenemos así, los 177 países distribuidos en las 9 combinaciones posibles de las categorías de *Nivel de Esperanza de Vida* y *Tipo de Asentamiento Predominante*.

Si realizado el conteo en la matriz de datos, observamos que fueron 29 los países que tiene un predominio de población en asentamiento rurales y una esperanza de vida baja, 15 los países que tiene un predominio de población en asentamiento rurales y una esperanza de vida media, 2 los países que tienen un predominio de población en asentamiento rurales y una esperanza de vida alta, y así sucesivamente, podríamos organizar estos datos en una tabla que se conoce como “**tabla de contingencia**”:

Países según el Tipo de asentamiento predominante y el Nivel de esperanza de vida. 2002.

Asentamiento	Nivel de Esperanza de Vida			Total	<i>Marginal: Dist. según Tipo de Asentamiento</i>
	Bajo	Medio	Alto		
Rural	29	15	2	46	<i>Cuerpo de la tabla</i>
Compensado	18	35	30	83	
Urbano	2	8	38	48	
Total	49	58	70	177	<i>Marginal: Dist. según Nivel de Esperanza de Vida</i>

30 países tienen alta esperanza de vida y un tipo de asentamiento compensado

En el cuerpo de la tabla aparecen los 177 países distribuidos según el *Tipo de asentamiento* y el *Nivel de esperanza de vida*. En cada uno de los marginales, se presentan las distribuciones de frecuencias (univariada) de las variables *Esperanza de vida* (última fila) y *Tipo de asentamiento* (última columna).

Podemos leer en la tabla que,

hay 30 países con Asentamiento compensado y Nivel de Esperanza de vida Alto. A su vez también podemos ver que, son 83 el total de países con Asentamiento compensado y 70 los que presentan un nivel de Esperanza de vida Alto.

Una vez que organizamos los datos en una tabla como la anterior, tenemos que poder establecer si existe algún tipo de relación entre las condiciones de asentamiento predominante (de países más rurales a los más urbanos) y la expectativa de vida de sus poblaciones. Como interrogante de trabajo se podría plantear: *¿varía la expectativa de vida según sea el tipo de asentamiento predominante?*

Si al comparar la **distribución** de la variable esperanza de vida entre los distintos tipos de asentamiento no observáramos diferencias, **diríamos que no existe** relación entre las variables. La esperanza de vida es "indiferente" al (independiente del) tipo de asentamiento predominante. En caso contrario, estaríamos en presencia de una relación entre ambas variables que será tanto más definida (más fuerte) cuánto mayor sean las diferencias observadas.

Pero la comparación de esas distribuciones condicionadas (de la esperanza de vida según asentamiento), no se puede realizar en valores absolutos cuando el número total de efectivos de cada una de esas distribuciones es diferente. Por ejemplo, para valorar si la esperanza de vida media es más frecuente entre los países con asentamiento compensado que entre los países con asentamiento rural, no podemos hacerlo a simple vista comparando los 35 países que se registran entre los de asentamiento compensado y los 15 países observados entre los de asentamiento rural. Esto es así porque el número de países con asentamiento compensado es muy superior al total de la otra categoría.

Este tipo de situación, que se presenta muy frecuentemente, nos lleva al cálculo de porcentajes, como los presentados en la siguiente tabla¹⁹.

¹⁹ Según el tipo de interrogante que nos planteamos puede ser necesario calcular los porcentajes de otra manera. Por ejemplo: si nos preguntáramos *¿cómo se distribuye el asentamiento según sea el nivel de esperanza de vida?* Evidentemente en este caso, deberán calcularse los porcentajes en el sentido de las columnas, es decir calcular la proporción que representa cada tipo de asentamiento en el total de cada una de las categorías de Nivel de esperanza de vida.

Países distribuidos por Nivel de Esperanza de Vida según el Tipo de asentamiento predominante. 2002. (%)

Asentamiento	Nivel de Esperanza de Vida			Total
	Bajo	Medio	Alto	
Rural	63	33	4	100 (46)
Compensado	22	42	36	100 (83)
Urbano	4	17	79	100 (48)
Total	28	33	40	100 (177)

$\frac{15}{46} \cdot 100 = 33\%$ de los países rurales tienen un nivel de esperanza de vida medio

Comparando en la tabla, la distribución de la Esperanza de Vida según el Tipo de Asentamiento Predominante, se hace evidente que el comportamiento varía en cada caso: **a mayor nivel de urbanización se observa una mayor esperanza de vida.** Es así que, entre los países predominantemente rurales, es marcadamente más frecuente encontrar países con esperanza de vida baja (63%), mientras que entre los países con asentamiento más equilibrado (compensados) esta condición se da en un 22% de los casos y entre los urbanos involucra sólo el 4%. Una situación inversa surge de comparar la presencia de expectativa de vida alta: un 79% de los países con asentamiento urbano tienen una expectativa de vida alta, proporción que decrece progresivamente con un aumento de los niveles de ruralidad. Puede decirse entonces que **existe una relación entre ambas variables** o que la Esperanza de Vida y el Tipo de Asentamiento Predominante **no son independientes**.



ACTIVIDAD N° 6

Encuesta nacional sobre las tecnologías de la información y comunicación. Ministerio de Educación- Programa Huascarán. Lima, Perú, 2002.

La información presentada se obtuvo del informe final de la Encuesta Nacional sobre Tecnologías de Información y Comunicación (TIC) 2002, que se realizara con el propósito de conocer la situación actual del acceso, conocimiento y uso de las TIC por los docentes en el servicio educativo. También, conocer la forma como los docentes han incorporado el uso de las TIC en el proceso de aprendizaje e interacción con sus educandos.

Conocimiento de algún programa de informática según ámbito

RESPUESTAS	TOTAL	URBANO	RURAL
SI	59.2 %	69.3 %	44.2 %
NO	40.8 %	30.7 %	55.8 %
TOTAL	100.0 %	100.0 %	100.0 %

TOTAL ENTREVISTADOS: 4.014

Conocimiento de algún programa de informática según gestión

RESPUESTAS	TOTAL	ESTATAL	NO ESTATAL
SI	59.2 %	56.9 %	75.8 %
NO	40.8 %	43.1 %	24.2 %
TOTAL	100.0 %	100.0 %	100.0 %

TOTAL ENTREVISTADOS: 4.014

Con la información presentada en las tablas anteriores:

- a) ¿qué porcentaje del total de los docentes conocen algún programa de informática?
- b) Esta proporción ¿difiere según se trate de docentes del ámbito urbano o rural?
- c) ¿qué parte de los docentes que trabajan para el estado declaran no conocer algún programa de informática?
- d) ¿Existe relación entre el conocimiento de algún programa informático y el ámbito y/o el tipo de gestión en que se desempeñan?



3. La Relación entre Variables Categóricas y Numéricas

En este caso se trata de establecer si los valores de una variable numérica varían según sean los valores de una variable categórica. Conducen a este tipo de análisis, preguntas como: A igualdad de tareas, ¿los salarios que perciben las mujeres son similares a los que perciben los hombres?; ¿las familias tienen diferente número de hijos según habiten en áreas rurales y urbanas?, ¿qué tan grande es la diferencia en la riqueza producida por habitante entre los países centrales y periféricos?, etc.

Si, como esperamos, encontramos que el salario de las mujeres –en promedio– resulta inferior al de los hombres, diríamos que sexo e ingreso están relacionados. Del mismo modo, si el asentamiento y el número de hijos por mujer estuviéramos relacionados, esperaríamos que entre las mujeres de áreas rurales se registre un mayor número hijos (en general) que entre aquellas que viven en áreas urbanas.

De modo que:

Para analizar la existencia de asociación entre una variable numérica y una categórica, **tenemos que comparar los grupos de individuos definidos por la variable categórica, en términos de los valores promedio²⁰ que registran en la variable numérica.**

En los ejemplos anteriores, compararíamos el ingreso medio de las mujeres con el ingreso medio de los hombres; el promedio de hijos de las mujeres rurales con el ese mismo promedio entre las mujeres urbanas; compararíamos el ingreso medio *per cápita* de los países centrales, con este mismo ingreso de los países periféricos.

A los efectos de ejemplificar el razonamiento propio de este análisis, nos proponemos estudiar la relación entre el PBI *per cápita* y el Tipo de Asentamiento predominante de la población de los países. A continuación presentamos las *medidas características* de la distribución del PBI *per cápita* para cada una de las subpoblaciones que quedan determinadas por las categorías de la variable *Tipo de asentamiento predominante*.

PBI per cápita (US\$ PPA) según Tipo de Asentamiento de los países. 2002.

Medidas	Tipo de Asentamiento		
	Rural	Compensado	Urbano
Media	2404,1 US\$ PPA	8248,3 US\$ PPA	17348,8 US\$ PPA
Mediana	1690,0 US\$ PPA	5440,0 US\$ PPA	16950,0 US\$ PPA
Número de casos	45 países	83 países	47 países

²⁰ Esos valores promedios que se utilizan para la comparación deberán elegirse tomando en cuenta las consideraciones realizadas al presentar las medidas de tendencia central. Ej. si la distribución es marcadamente asimétrica, optaríamos por la mediana y no por la media aritmética.

De la comparación de medias surge que existe una relación bien definida entre el tipo de asentamiento y el PBI per cápita. Es así que, **a mayor nivel de urbanización, crece significativamente la riqueza promedio generada** por los países con ese asentamiento predominante: mientras que entre los países rurales el PBI per cápita promedio es de 2400 US\$ PPA, entre los de asentamiento compensado supera los 8200 US\$ PPA y entre los urbanos los 17300 US\$ PPA. Si alguna de las distribuciones consideradas presentaran una fuerte asimetría que restaran representatividad a la media, podemos utilizar la mediana como medida representativa y -en este caso- observamos que se corrobora la relación descripta anteriormente.

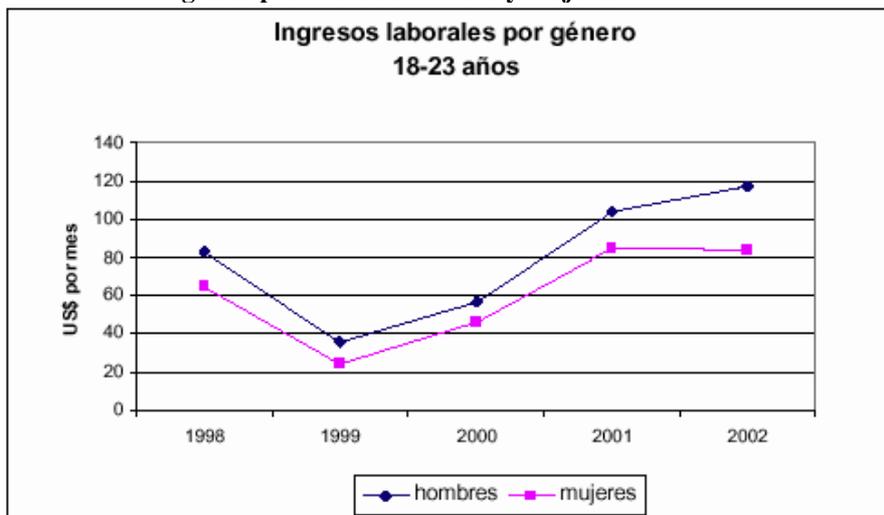


ACTIVIDAD N° 7

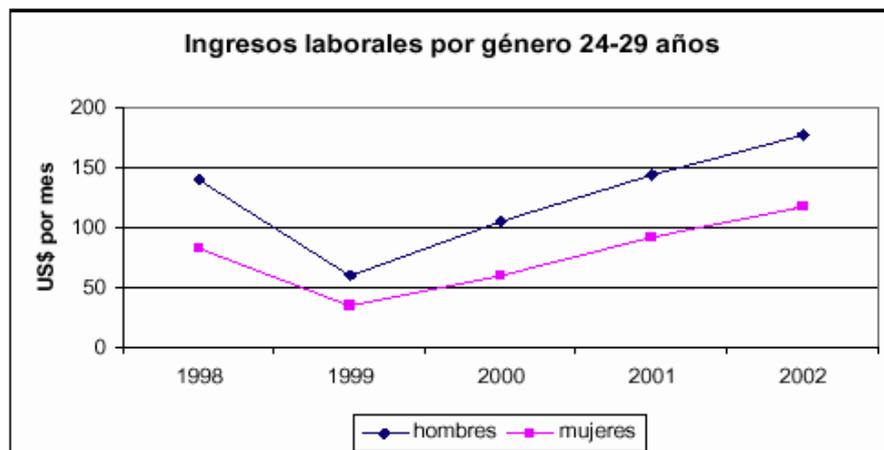
PRIMERA PARTE

Los siguientes gráficos fueron extraídos del informe final: VASCONEZ, A.; TRUJILLO, A.: *Incorporación de los jóvenes al mercado laboral. Facultad Latinoamericana de Ciencias Sociales FLACSO-Sede Ecuador. Contrato de consultoría cepal/gtz. Quito, Ecuador, 2004.*

Brecha de género en ingresos laborales por grupos de edad. Ecuador, 2004.
Ingresos promedio de hombre y mujeres. 1996-2002.



Fuente: Encuesta de empleo FLACSO-Banco Central
 Elaboración: autoras



Fuente: Encuesta de empleo FLACSO-Banco Central
 Elaboración: autoras

Analice los gráficos anteriores y describa comparativamente la evolución de los ingresos y la brecha existente entre hombres y mujeres.

SEGUNDA PARTE

Encuesta Nacional de Hogares sobre Medición de Niveles de Vida. (Enniv) - 2000

A continuación incluimos una tabla extraída del informe de la ENNIV, presentado por el Instituto Cuánto en el año 2000. El objetivo de esta encuesta fue medir y analizar las condiciones de vida de la población residentes en Perú y producir información a nivel de los hogares y personas sobre impactos de las políticas gubernamentales en los niveles de bienestar de la población.

AÑOS PROMEDIO DE EDUCACION DE LA POBLACIÓN DE 6 A MAS AÑOS

	1994	1997	2000
TOTAL	7.2	7.3	7.7
Lima Metropolitana	8.7	9.0	9.2
Resto urbano	7.9	8.1	8.5
Rural	5.0	5.1	5.6

Fuente: Instituto Cuánto. Encuesta Nacional de Niveles de Vida, 1994-1997-2000.

Analizando la tabla anterior, compare los niveles medios de educación entre las tres regiones del Perú durante el período considerado.



4. La Relación entre Variables Numéricas

La situación en este caso consiste en analizar la relación entre dos variables numéricas.

Dos variables numéricas **están relacionadas cuando al cambiar los valores de una de ellas, se puede identificar un patrón en el comportamiento de variación de la otra.**

Por ejemplo: si observamos que cuanto mayor es el PBI *per cápita* de los países, mayores son los porcentajes de alfabetismo; si al aumentar el número de años de educación formal, también aumenta –en promedio– el ingreso de las personas; en las zonas rurales, si cuánto mayor es la distancia al establecimiento escolar, menor es –en promedio– el porcentaje de asistencia de los niños; etc.

En el caso de variables numéricas, para resumir los datos de la matriz sin perder el valor de las observaciones individuales, **el recurso gráfico es un poderoso y expresivo instrumento para analizar el patrón de comportamiento que observan las variables.**

Ese recurso gráfico –conocido como **diagrama de dispersión** o **nube de puntos**– consiste en la representación, en un sistema de ejes cartesianos, de los individuos observados según las variables consideradas.

Consideremos para un análisis de este tipo, las variables *PBI per cápita* y *Porcentaje de población desnutrida* para un grupo de países de América.

Países de América. PBI per cápita en US\$PPA (2002) y Población desnutrida en % (2001).

País	PBI per cápita	Pob. desnutrida
Bolivia	2460	22
Brasil	7770	9
Chile	9820	4
Colombia	6370	13
Costa Rica	8840	6
Cuba	5259	11
Ecuador	3580	4
El Salvador	4890	14
Guatemala	4080	25
Guyana	4260	14
Haití	1610	49
Honduras	2600	20
Jamaica	3980	9
México	8970	5
Nicaragua	2470	29
Panamá	6170	26
Paraguay	4610	13
Perú	5010	11
Rep. Dominicana	6640	25
Surinam	6590	11
Trinidad y Tobago	9430	12
Uruguay	7830	3
Venezuela	5380	18

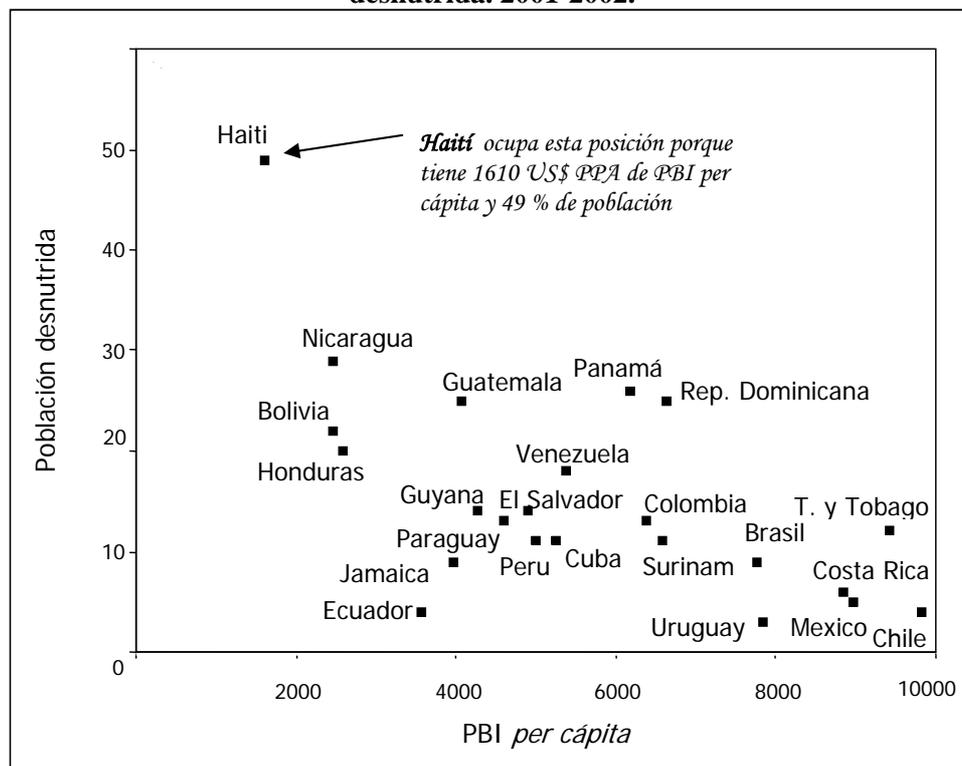
Una mirada atenta del comportamiento de estas dos variables, revela cierto patrón según el cual a los países con mayor PBI per cápita le corresponden menores niveles de población desnutrida. Este comportamiento se corresponde con lo que podemos esperar, dado que –en alguna medida– la generación de mayor riqueza por habitante se “debería” traducir en mejores condiciones de vida de la población. Sin embargo, sería aventurado simplificar la explicación de los niveles de desnutrición únicamente en el nivel de riqueza producido por habitante. En la desnutrición seguramente estarán incidiendo además otras variables, entre las que debemos destacar el nivel de redistribución de la riqueza generada.

Indudablemente esta lectura ha sido posible por dos razones fundamentales: el número de unidades de observación no es demasiado grande y tenemos una hipótesis en relación al comportamiento conjunto de las variables. Convengamos a su vez que no siempre se dan estas condiciones de trabajo; es allí donde el *diagrama de dispersión* adquiere una función

exploratoria de los patrones que estamos buscando.

A partir de los datos correspondientes a 23 países de América, podemos construir el siguiente diagrama de dispersión.

Distribución de los países americanos según PBI per cápita y Porcentaje de población desnutrida. 2001-2002.



En este gráfico cada variable se representa sobre uno de los ejes. Preferentemente, en el eje X se representa aquella variable que consideramos actúa como “independiente” (PBI *per cápita*)²¹.

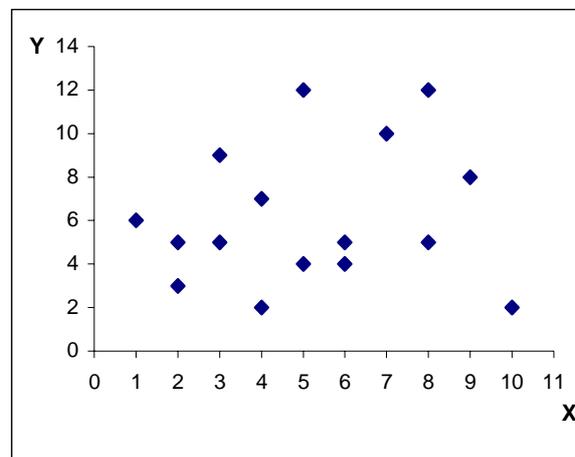
Cada individuo se representa por un punto cuya ubicación en el plano está determinada por los valores que asume en ambas variables. Así, en nuestro ejemplo, República Dominicana, aparece ubicado en el plano con una coordenada en el eje X igual a 6640 y una coordenada en el eje Y de 25. Representados los 23 países en este gráfico, es posible apreciar más claramente (que en la matriz) el comportamiento conjunto de ambas variables y encontrar –si existe– el patrón de asociación.

Así, en el gráfico se puede ver que entre los países considerados, Chile, Costa Rica, México y Trinidad Tobago presentan los mayores niveles de PBI per cápita y simultáneamente tasa de desnutrición bajas en relación al resto. Además, existe otro grupo de países (Bolivia, Honduras y Nicaragua) que producen una riqueza relativamente baja (inferior a los 2500 US\$ PPA) pero presentan tasas de desnutrición de 20% o más. En particular, se destaca la situación de Haití que presenta los mayores niveles de desnutrición y el menor PBI per cápita.

*Globalmente, apreciamos que al aumentar el PBI per cápita se produce -en promedio- una disminución progresiva de los niveles de desnutrición. Esto estaría indicando un comportamiento relativamente coordinado/sincronizado entre las variables y en consecuencia la **existencia de una relación entre PBI per cápita y Desnutrición de la población.***

En el ejemplo, si la desnutrición estuviera “explicada” exclusivamente por el PBI *per cápita*, observaríamos un diagrama de dispersión en el que los puntos se ubican sobre una recta. Pero esta situación no corresponde al diagrama de dispersión obtenido; en éste, si bien se aprecia un patrón de comportamiento como el descrito precedentemente, los puntos se desvían bastante de la situación perfectamente lineal.

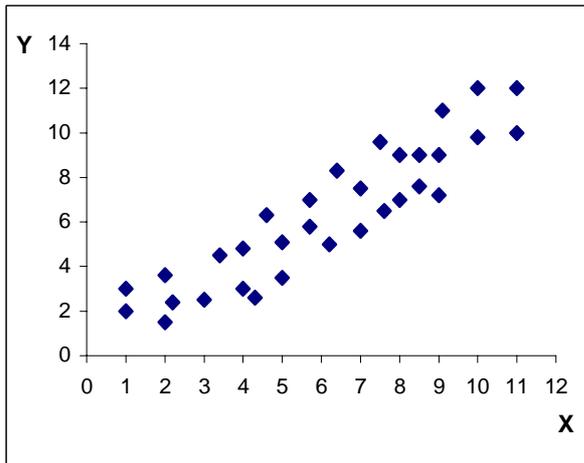
Consideremos en términos generales, algunas situaciones “típicas” con las que nos podemos encontrar en la práctica.



En el diagrama anterior no se observa ningún patrón de comportamiento: valores altos de X se corresponden tanto con valores altos, como con valores bajos de Y; esto indica que las variables **no están relacionadas**.

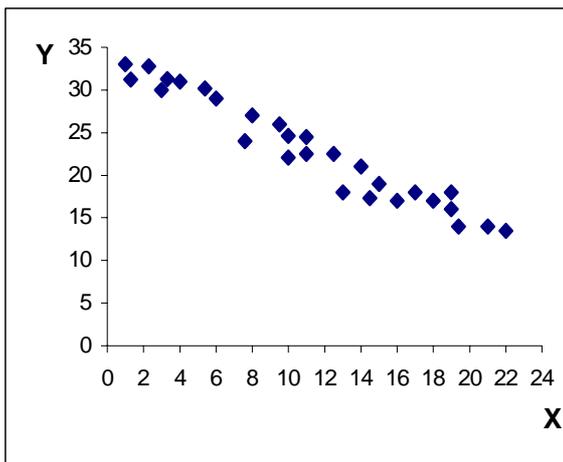
En cambio, en los diagramas que presentamos a continuación es posible identificar claramente una asociación lineal entre las variables analizadas.

²¹ En general, cuando se trata de una relación causal, se representa en el eje X la variable *explicativa* y en el Y, la variable *explicada*. Cuál es –en cada caso– la variable que actúa como explicativa es una cuestión que debe decidir quién analiza los datos y no depende de ningún criterio estadístico.



En este diagrama se observa que a medida que aumentan los valores de X también aumentan –en promedio– los valores de Y.

En estas situaciones se dice que existe una **relación lineal directa o positiva** entre las variables.



En este otro caso tenemos que, a valores crecientes de X le corresponden valores decrecientes de Y. Decimos aquí, que **estamos ante una relación lineal inversa o negativa** entre las variables.

La “proximidad” de los puntos a una recta imaginaria nos habla también de una relación más “fuerte” entre las variables que el que se observa en el diagrama anterior.

En el ejemplo del PBI per cápita y la Desnutrición, vemos que a medida que aumenta el PBI, disminuye –en promedio– los porcentajes de desnutrición y, en correspondencia con esto, la figura que describen los puntos se asimila a la presentada en este último gráfico, es decir que estamos en presencia de una relación lineal decreciente.



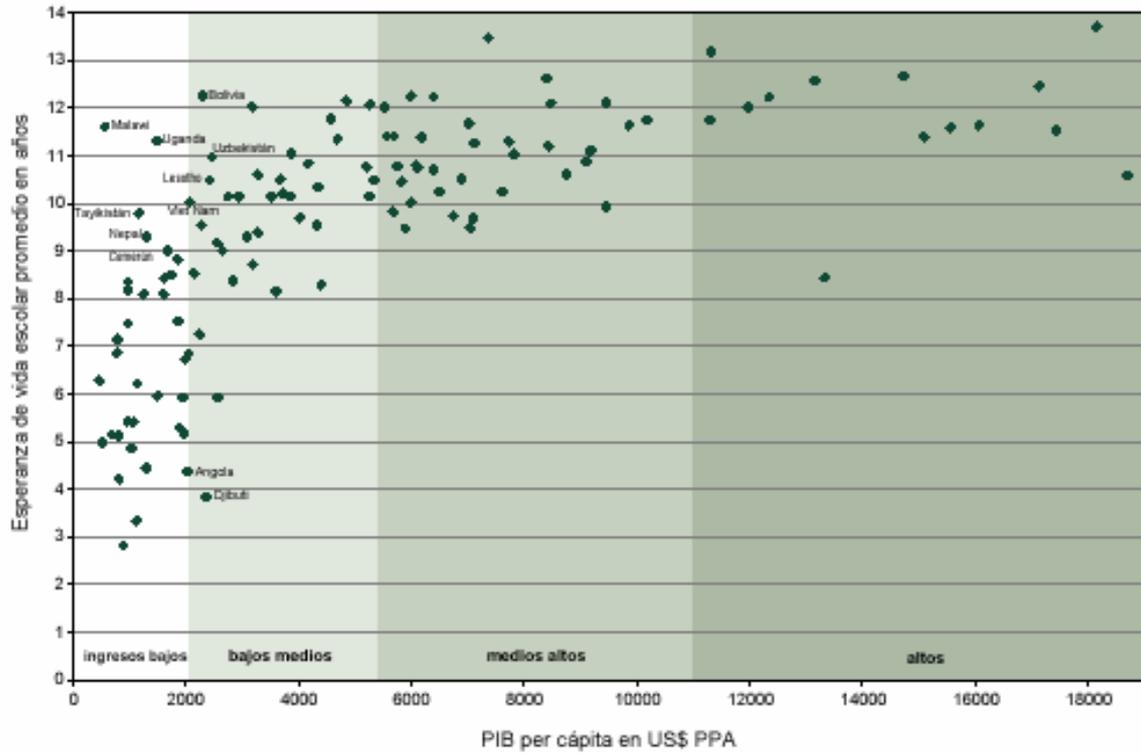
ACTIVIDAD N° 8

En el gráfico que se presenta a continuación, extraído del informe de la UNESCO: **COMPENDIO MUNDIAL DE LA EDUCACIÓN 2004. Comparación de las estadísticas de educación en el mundo**, se representan los países según la “*Esperanza promedio de vida escolar primaria y secundaria*” (en años)²² y el “*PBI per cápita*” (US\$ PPA).

A partir del análisis del gráfico: describa la relación entre la duración esperada de la escolarización y la riqueza nacional.

²² La **esperanza de vida escolar** (EVE) se define como el total de años de escolarización que un niño de 4 años puede esperar recibir en el futuro, suponiendo que la probabilidad de matrícula escolar a una edad determinada sea igual a la tasa de matriculación actual para esa edad. Indica la duración promedio de la escolarización, no el número de grados (años escolares) alcanzado. También puede definirse como el promedio de años que es probable que un niño pase en el sistema educativo.

Promedio de esperanza de vida escolar en años para la educación primaria y secundaria y PIB per cápita (PPA convertido en dólares EE.UU.), 2001



Fuentes: Instituto de Estadística de la UNESCO, Cuadro 4. Banco Mundial, 2003.

Cobertura: 147 países, el 94% de la población mundial (Incluidos los países de altos ingresos no mostrados).

Nota: Se muestran sólo los países con un PIB per cápita inferior a 19.000 dólares PPA. Los países están clasificados por grupos de ingresos sobre la base del PIB per cápita en dólares EE.UU. convertidos usando las Paridades del Poder Adquisitivo (PPA). Los países están agrupados por cuartiles de ingresos. El PIB per cápita de los países de bajos ingresos es inferior a 2.055 dólares PPA, el de los países de ingresos bajos-medios oscila entre 2.055 y 5.415 dólares PPA, el de países de ingresos medios-altos oscila entre 5.415 y 11.010 dólares PPA, y el de los países de altos ingresos supera este valor. Esta agrupación difiere del método empleado por el Banco Mundial basado en el Ingreso Nacional Bruto (INB) per cápita convertido a dólares EE.UU. utilizando los tipos de cambio del mercado.



V. Los Números Índices

Al analizar la evolución de las condiciones socioeconómicas de un país o región, es de uso común recurrir a comparaciones relativas de los cambios que experimenta una variable o un conjunto de variables que resultan de interés.

De esta manera es frecuente encontrar publicaciones de las oficinas de estadísticas, u otros organismos oficiales, que incluyen las variaciones relativas de precios, de volúmenes producidos, de montos recaudados por alguna actividad comercial, de cantidades exportadas, etc.

A veces, esas variaciones *resumen en un único valor* los cambios de un conjunto importante de variables. Un ejemplo muy conocido de elaboración de este tipo de información es el *índice de precios al consumidor*, con el cual se valora la variación promedio que experimenta una canasta de artículos de consumo de cierta población.

Esta forma de expresar los cambios de una variable o grupo de variables (a lo largo del tiempo o en el espacio) se conoce como números índices.

Los **números índices** constituyen una forma de establecer la magnitud del cambio de una o más variables en el tiempo o en el espacio.

Los **números índices se clasifican en Simples y Compuestos**, según se trate de medir los cambios de una única variable (simples) o de un grupo de variables (compuestos)

1. Los Números Índices Simples

Los **índices relativos simples** se construyen para *mostrar los cambios porcentuales que experimenta una variable* con respecto a un valor de la serie seleccionado como base.

A manera de ejemplo, consideremos la serie de *precios de exportación de la yerba mate* durante el período comprendido entre los años 1990 y 2000. En este caso deseamos medir la variación relativa de los precios de cada período de la serie, con respecto al valor del año 1990 (año base elegido arbitrariamente en este ejemplo)²³. Para ello construimos un *Índice de Precios Relativo Simple (IPRS)*

Precios de Exportación⁽¹⁾ la Yerba Mate y Variaciones de los Precios Período: 1990-2000.

Años	Precios (U\$S/kg)	IPRS (1990=100)
1990	0,835	100,0
1991	1,073	128,5
1992	1,025	122,8
1993	1,065	127,5
1994	0,943	112,9
1995	0,802	96,0
1996	0,714	85,5
1997	0,677	81,1
1998	0,663	79,4
1999	0,640	76,6
2000	0,561	67,2

Año base →

$$\text{IPRS}_{97/92} = \frac{0,943}{0,835} \cdot 100 = 112,9\%$$

El precio decreció un 32,8% (100-67,2)

⁽¹⁾ Precios FOB.

Fuente: Dirección General de Economía Agraria. Ministerio de Asuntos Agrarios. Provincia de Misiones. 2002.

²³ A los fines de este índice cualquier período de la serie puede ser adoptado como "base"; en cada problema particular de trabajo el investigador deberá decidir el período base más conveniente.

El índice relativo simple de base fija del año '95 con base en el año '90, resulta de:

$$IPRS_{95/90} = \frac{0,802}{0,835} \cdot 100 = 96,0\%$$

Es decir que en 1995 el precio de exportación de la yerba mate canchada **decreció un 4%** (96,0-100) **con respecto al valor registrado en el año base.**

Analizando los índices relativos simples para la serie completa, notamos que el precio de exportación de la yerba mate registra valores superiores al de la base durante los primeros años (1991-1994), para luego decrecer hasta el año 2000. A partir de 1995 decrece sostenidamente hasta alcanzar el menor valor en el año 2000, en el cual registra una caída del 32,8% con relación al precio de 1990.

2. Los Números Índices Compuestos

Los **índices compuestos** se construyen para expresar en un único valor, las variaciones porcentuales de un conjunto de variables. Por lo general, estos índices compuestos serán más "eficientes" cuando cada una de las variables consideradas en su construcción esté ponderada por un factor que expresa su importancia en el conjunto de variables incluidas.

2.1. Los índices de precios

En cuanto a la "importancia" que le damos a cada variable en la construcción de estos índices, es posible acordar intuitivamente que serán más importantes –por ejemplo– en un índice de costo de vida, aquellos artículos que se consumen más; es decir que en este caso la "importancia" estaría dada (como para cualquier índice de precios), por el volumen de lo consumido, producido, exportado, etc.

Al construir un índice de precios, una alternativa de ponderación muy difundida es utilizar las cantidades (compradas, vendidas, exportadas, etc.) del período base²⁴, conformando lo que se conoce como el **Índice de Laspeyres**.

A los efectos de explicar la construcción de un **índice de precios** compuesto ponderado de Laspeyres, nos planteamos la situación de tener que construir un **índice de precios al consumidor** en el cual se utiliza normalmente este método. Con este tipo de índices, se pretende medir la evolución del costo de vida a través de los cambios de precios producidos en un grupo determinado de bienes y servicios que se consideran representativos del consumo "medio" de una población.

Buscamos así responde a la pregunta: *¿cuánto se gastará en un período dado con relación al período base, si compramos los mismos artículos en las mismas cantidades del período base,?*

Para ello se debe –en primera instancia– determinar la "canasta familiar", que es ese conjunto de bienes y servicios representativo del consumo de los hogares, determinado a través de un relevamiento efectuado a tal efecto. Consideremos una situación hipotética y simplificada en la que se determinó el siguiente conjunto de bienes y servicios que constituyen la "canasta familiar" con sus correspondientes cantidades consumidas mensualmente.

Bienes y servicios	Cantidad consumida al mes
Verduras	20 kg.
Carne	10 kg.
Alquiler	1
Vestimenta	2 prendas
Transporte	40 boletos
Luz	80KW
Agua	10 m ³
Gas	45 Kg.
Libros	1

²⁴ Existen otras alternativas, como la de ponderar por las cantidades del período que se está considerando, conformando en este caso un índice conocido como de Paasche (o de precios implícitos).

¿Cuál es el costo de esta canasta? Intuitivamente se puede ver que ese costo se determinará mediante la suma de esas cantidades consumidas por los precios unitarios de esos bienes y servicios. Las *cantidades consumidas*, pasan a constituirse en el *factor de ponderación* del índice de precios al consumidor, dado que cada una de ellas expresa la "importancia" del artículo en términos del consumo de esas familias.

Cualquier variación en los precios unitarios redundará en un cambio en el costo total de la canasta, determinando cuantos pesos más (o cuantos menos) nos cuesta comprar esa misma canasta de bienes de un período a otro. El índice de precios al consumidor buscará registrar la magnitud –en promedio– de esos cambios, en términos porcentuales. Para ello se procede de la siguiente manera.

Se calcula el costo de la canasta familiar en el mes base (suma de los productos de los precios unitarios en el período base por las cantidades consumidas de cada bien) y el costo de la canasta familiar a los precios del mes que se está analizando. El índice resulta de hacer el cociente entre el costo de la canasta a los precios del período considerado y el costo de la canasta en el mes base, expresado en porcentaje.

Cálculo del costo de la Canasta familiar en los meses 1 y 2

Bienes y servicios	Cantidad consumida al mes	Precio unitario en el mes 1 (en \$)	Precio unitario en el mes 2 (en \$)	Gasto mes 1	Gasto mes 2
Verduras	20 kg.	1,25	1,40	25	28
Carne	10 kg.	3,0	3,2	30	32
Alquiler	1	300	300	300	300
Vestimenta	2 prendas	25	28	50	56
Transporte	40 boletos	0,70	0,70	28	28
Luz	120KW	0,10	0,12	12	14,4
Agua	10 m ³	3,0	3,4	30	34
Gas	45 Kg.	2,11	2,44	95	110
Libros	1	15	16	15	16
Total	----	-----	-----	585	618,4

El índice de precios para el mes 2 tomando como base el mes 1, es: $\frac{618,4}{585} \cdot 100 = 105,7\%$

En promedio, el conjunto de bienes y servicios que componen la canasta familiar, se ha incrementado durante el período analizado en 5,7%. En otros términos, comprar la misma canasta de bienes nos cuesta un 5,7% más de lo que nos costaba en el mes base.

En el marco de un análisis de las exportaciones agrarias de la provincia de Misiones-Argentina, queremos observar la evolución de los precios de dos de sus principales cultivos (yerba mate y té). Para ello utilizaremos un índice de Laspeyres.

Cantidades Exportadas y Precios de Exportación de Yerba Mate y Té. Provincia de Misiones, 1990/2000.

Años	Yerba Mate		Té		Valor de las exportaciones	IP (*90 = 100)
	Tn.	U\$\$/Tn	Tn.	U\$\$/Tn		
1990	4.266	835	42.584	770	36.351.790,0	100,0
1991	9.022	1.073	34.658	770	37.367.098,0	102,8
1992	13.491	1.025	34.809	780	37.588.170,0	103,4
1993	15.689	1.065	41.872	880	42.017.210,0	115,6
1994	15.667	943	41.188	870	41.070.918,0	113,0
1995	37.488	802	40.466	770	36.211.012,0	99,6
1996	39.499	714	39.069	750	34.983.924,0	96,2
1997	33.277	677	41.465	770	35.677.762,0	98,1
1998	34.916	663	57.738	930	42.431.478,0	116,7
1999	30.269	640	51.090	750	34.668.240,0	95,4
2000	36.528	561	49.240	760	34.757.066,0	95,6

Valor de la producción del año '90 a los precios corrientes de cada año

Fuente: Dirección de Economía Agraria. Ministerio de Asuntos Agrarios. Posadas, Misiones. 2002.

El índice de precios del año '96 tomando como referencia el año '90, se obtiene haciendo:

$$IP^L_{96/90} = \frac{714 \cdot 4.266 + 750 \cdot 42.584}{835 \cdot 4.266 + 770 \cdot 42.584} \cdot 100 = 96,2\%$$

En 1996 los precios de exportación la yerba mate y el té **decrecieron -en promedio- el 3,8% con relación a los precios registrados en el año 1990.**

El precio de estos cultivos expone –en promedio– un comportamiento inestable. Es creciente hasta 1993, posteriormente decrece hasta el año 1996 y luego muestra una recuperación hasta alcanzar en 1998 el mayor valor de la serie, que resulta superior en 16,7% al que registrara en el año base . En los dos últimos dos años, vuelve a ubicarse por debajo del precio que se registrara en el año base.

2.2. Los índices de cantidad

Los índices de cantidad se construyen de manera análoga a los índices de precio vistos precedentemente, pero ahora tomando los precios como ponderación. En el caso del índice de cantidad de Laspeyres, se tomarán como ponderación los precios del período base.

El índice de Laspeyres de cantidad para el año '96 tomando como referencia el año '90, se obtiene haciendo:

$$IQ^L_{96/90} = \frac{39.499 \cdot 835 + 39.069 \cdot 770}{4.266 \cdot 835 + 42.584 \cdot 770} \cdot 100 = 173,5\%$$

En 1996 las exportaciones de yerba mate y té crecieron -en promedio- el 73,5% con relación a las exportaciones realizadas en 1990.

3. La Elección del Período Base

La selección de la base es una decisión crucial a la hora de construir un número índice. Esto es así porque si el valor que tomamos como referencia corresponde a un período atípico (Ej. el año 2001 que coincide con la última crisis económico-social en la Argentina; o el año de la “crisis del tequila” para series de cualquier país de América Latina), puede llevar a que la evolución de la serie esté crónicamente sobreestimada dado que se ha tomado como referencia un período de gran depreciación. Una situación análoga, pero en sentido contrario, ocurriría si tomáramos como referencia un período que es atípicamente positivo. En consecuencia, es recomendable seleccionar como base períodos que resulten “**típicos o normales**”.

También, es inapropiado tomar como base períodos muy distantes en el tiempo, dado que estaríamos comparando la situación actual con momentos de la economía o la sociedad, en los cuales las condiciones contextuales (tecnológicas, ambientales, organizacionales, etc.) son muy diferentes.



ACTIVIDAD N° 9

PRIMERA PARTE

En la tabla siguiente se presentan los índices que expresan la evolución de las exportaciones de la Argentina en los últimos diecinueve años. Estos valores son provistos por el Instituto Nacional de Estadística y Censos de la Argentina.

Índices de valor, precio y cantidad de las exportaciones, importaciones, y términos del intercambio, base 1993=100

Período	Índices de las Exportaciones		
	Valor ⁽¹⁾	Precio	Cantidad
1986	52,2	79,4	65,8
1987	48,5	81,8	59,3
1988	69,6	95,3	73,1
1989	73,0	97,0	75,3
1990	94,2	97,6	96,5
1991	91,3	96,1	95,0
1992	93,3	99,8	93,5
1993	100,0	100,0	100,0
1994	120,7	102,9	117,4
1995	159,8	108,8	146,9
1996	181,5	115,9	156,6
1997	201,5	111,9	180,1
1998	201,6	100,3	201,0
1999	177,7	89,1	199,5
2000	200,8	98,0	204,9
2001	202,3	94,7	213,7
2002	196,0	91,0	215,3
2003 ⁽²⁾	224,0	99,7	224,7
2004 ⁽²⁾	226,6	109,0	214,0

⁽¹⁾ Se calcula sobre el *valor* de las exportaciones, el que resulta de multiplicar el precio y la cantidad

⁽²⁾ Datos provisorios

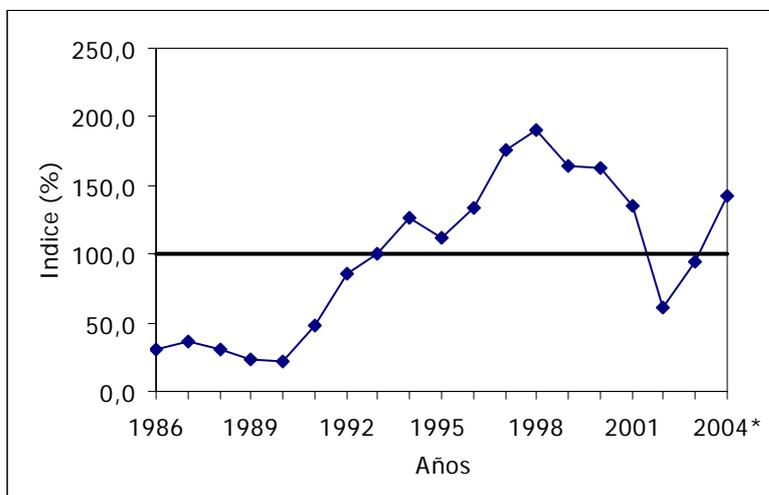
Fuente: INDEC

A partir de la información presentada en el cuadro anterior:

- Identifique el año en que se registró el mayor volumen de exportación en relación al año base.
- Interprete los índices de *valor*, *precio* y *cantidad* correspondientes al año 1989 y 1997.
- Observe que el año de mayor volumen de exportación no se corresponde con el año de mayor *valor* de las exportaciones. Explique esta situación.

SEGUNDA PARTE

Evolución de las cantidades importadas por la Argentina. 1986-2004.



A partir de la información presentada en el gráfico, describa en forma resumida la evolución de las importaciones argentinas a lo largo del período analizado.

TERCERA PARTE

El artículo que se presenta a continuación fue publicado en el diario de mayor tirada del país en un momento en el cual se estaba viviendo una crisis política-económica y social (que eclosionó en diciembre de 2001, con la caída del entonces presidente de la nación), y que se tradujo en importantes variaciones en los precios de los bienes de consumo y repercutió fuertemente en el poder adquisitivo de los ciudadanos.

A partir de la lectura del mismo, le proponemos:

- a) Identifique en el texto aquellas afirmaciones que involucran la utilización de números índices.
- b) ¿Cuál es el valor del índice en el mes de septiembre con respecto al mes anterior y con respecto al mes de diciembre del 2001?
- c) Identifique las diferencias entre la canasta básica y la canasta básica alimentaria.
- d) ¿Qué información le permite a la autora afirmar que el aumento en los bienes alimentarios "se ensañó con los sectores más bajos"?
- e) A qué atribuye la diferencia observada entre el aumento de la canasta básica durante el año y la canasta básica alimentaria.
- f) De los bienes alimentarios que se analizan en el artículo ¿cuál o cuáles tuvieron un incremento inferior al crecimiento del índice correspondiente a la canasta alimentaria?

Diario Clarín, Bs. As., 8/10/02. Sección Economía. A. Quiroga.

LA SITUACION LABORAL: DATOS OFICIALES DEL INDEC: CRECE EL NUMERO DE POBRES E INDIGENTES

La canasta familiar subió otro 2%: ya cuesta \$ 716. El aumento acumulado en lo que va del año llega a 55,2%. En diciembre, el mismo paquete de productos costaba \$ 461. Actualmente, el 70% de los asalariados gana menos de \$ 500.

Un matrimonio con dos hijos necesita 716 pesos por mes para cubrir sus necesidades básicas. Según las estimaciones del Instituto Nacional de Estadísticas y Censos (INDEC) los hogares que no llegan a tener cubrir esa suma, son considerados pobres. Y si no llegan a juntar 324 pesos al mes, la plata ni siquiera les alcanzará para comer. Entonces, serán indigentes. Así, en la Argentina de hoy hay más de 20 millones de pobres, de los cuales 9 millones son indigentes.

Según los datos que difundió ayer el INDEC, el valor de la canasta básica para una familia tipo subió en setiembre 2,05% en relación a agosto y acumula un incremento del 55,2% desde diciembre. La nueva suba de este índice, coincidió con el registro de inflación más bajo del año: en setiembre fue del 1,3%. Y la canasta alimentaria básica, que mide la indigencia, subió 3,9% en el mes y 73,5% en lo que va del año. Esto convirtió a 6 millones de argentinos en nuevos pobres en los primeros 9 meses del 2002. Y entre octubre y mayo, la pobreza en el país pasó del 40 al 54% del total de sus habitantes. Así, 3 de cada 4 chicos menores de 12 años son pobres.

En diciembre del año pasado, con ingresos familiares por 461 pesos era posible escapar de la pobreza, y con 187 pesos, se sorteaba la indigencia. Pero tras la devaluación, la inflación licuó el poder adquisitivo de todos los salarios y se ensañó con los sectores más bajos: mientras el costo de vida aumentó 39,7% en lo que va del año, los pobres soportan un incremento que casi duplica este índice.

La canasta básica total mide las necesidades de una familia tipo -compuesta por dos adultos y dos niños de 5 y 8 años- e incluye bienes y servicios no alimentarios, como transporte, vestimenta, salud y educación. A partir del valor de esta canasta, que desde marzo pasado se monitorea mensualmente, el INDEC fija la Línea de Pobreza. En el Índice de Costo de Vida, los bienes representan el 53%, y el resto lo ocupan los servicios. En cambio, en la canasta básica total los bienes tienen un peso del 65% contra el 35% de los servicios. Esto es así porque los pobres gastan la mayor parte de sus ingresos en comer y reducen al mínimo sus gastos en otros rubros también imprescindibles. Hoy el 70% de los asalariados gana menos de 500 pesos. Por eso, para no caer en la pobreza, la mayor parte de las familias tienen que tener más de un ingreso.

En cambio, la línea de indigencia se establece a partir del valor de la canasta de alimentos básica, sin contar los servicios. Y como los alimentos fueron los productos que más subieron desde enero hasta hoy -un 67,2%, contra una inflación general del 39,7%-, el dinero alcanza cada día menos y los indigentes -que a los efectos de la medición son un subgrupo dentro de los pobres- son cada día más.

La canasta básica de alimentos incluye 27 productos imprescindibles para que una persona pueda realizar una "actividad moderada". De esa canasta, se destacan los productos que precisamente más aumentaron de precio durante el año: la harina (192%), el arroz (148%), aceite (207%) , fideos (62%), carne (75%), leche (65,4%), frutas (42%) y verduras (104%). (...)



BIBLIOGRAFÍA

Aguirre, C.; Niño, F.; Simonetti, E. (2004): *Estadística aplicada en las Ciencias Sociales y Humanas. Estadística I*. Colección Cuadernos de Cátedra. Editorial Universitaria de Misiones. Posadas, Misiones, Argentina.

Baranger, D. (2000): *Construcción y Análisis de Datos. Introducción al uso de técnicas cuantitativas en la investigación social*. Editorial Universitaria de Misiones. Posadas, Misiones, Argentina.

Blalock, H.M. (1986): *Estadística Social*. FCE. México.

Moore, D. (1998): *Estadística aplicada básica*. A. Bosch Ed. Barcelona (1ra. Edición 1995).

Shao, S. (1967): *Estadística para economistas y administradores de empresas*. Herrero Hnos. S.A. México.