

CAPITULO 1

LA ESTADISTICA

OBJETIVOS

- Conocer las distintas posibilidades que brinda la Estadística para resolver problemas concretos.
- Aprender los importantes conceptos de población y muestra.
- Identificar las variables para cualquier tipo de estudio y su correspondiente aplicación.
- Repasar conceptos matemáticos como: proporción, porcentaje y razón, debido a su gran aplicación.

CONTENIDOS

1. La Estadística: definición y clasificación.....
 - 1.1. Población: concepto y clasificación.....
 - 1.1.1 Parámetros: concepto.....
 - 1.2. Muestra.....
 - 1.2.1 Estadísticos: concepto.....
 - 1.3. Variables: concepto, clasificación y escalas de medición....
 - 1.4. Proporciones, razones y porcentajes.....
 - 1.4.1 Proporciones.....
 - 1.4.2 Razones.....
 - 1.4.3 Porcentajes.....
 - 1.5. Indicadores de control.....
 - 1.6. Problemas propuestos.....

1. La Estadística: Definición y clasificación.

El término **estadística** tiene un uso significativo y variado en gran parte de la sociedad. Haciendo un poco de historia, los babilonios, en tablas de arcilla recopilaban datos. Los egipcios luego, mencionaban **censo** y su estudio de las poblaciones, hasta en la misma Biblia aparece la palabra censo y las familias y su linaje con su correspondiente clasificación.

En China se realizaban censos de personas por territorio ordenadamente, lo mismo que en Roma, donde aparece el concepto de censor, encargado de controlar el número de habitantes y su distribución por territorios.

En la Edad Media se crea un registro de propiedades de Carlomagno y de los bienes de la Iglesia, es la época feudal. En 1.662 John Graunt escribe un tratado donde figura cifras brutas de nacimientos y defunciones ocurridas en Londres durante el período 1.604-1.661, lo consideran como el primer trabajo estadístico sobre la población.

Se sucede el “cálculo de probabilidades” como teoría matemática de los juegos de azar, recopilación de datos y elaboración de gráficos que sirvieron de información para describir algunas cuestiones inherentes de un país. Con el fortalecimiento de los Estados Nacionales y el correr del tiempo, fue cada vez más necesaria la información estadística para planificar y conocer cuestiones que se hacen imprescindibles en la toma de decisiones, tales como la población de un país, el consumo, la inversión, la

desocupación, etc. Las empresas demandaban imperiosamente información estadística para sus inversiones y la colocación de sus productos. Los individuos requerían información para determinar sus consumos. El área de salud pública y privada necesita de la estadística para investigar enfermedades, epidemias, etc. La Estadística fue ganando imprescindibles espacios de brindar y resumir información para la sociedad en su conjunto, de ahí su evolución en los siglos posteriores frente a un mundo de incertidumbres y más complejo. La Estadística fue de la mano del surgimiento de los Estados Nacionales y sistemas económicos pertinentes. Aparecen trabajos de B.Pascal, Edmund Halley, Galton, Pearson y otros con los cuales se llegó a la Estadística actual,

El vocablo “estadística” proviene de la palabra latina STATUS, que significa estado.

Existen múltiples definiciones de diferentes autores sobre la Estadística. La definición que adoptaremos en este libro es la siguiente:

“La Estadística es una ciencia que provee métodos científicos para la recolección, organización, descripción, análisis e interpretación de los datos para la utilización de los mismos en la toma de decisiones racionales frente a contextos de incertidumbre.”

Al estudiar o investigar un fenómeno necesitamos los datos, es decir, ir en búsqueda de los mismos. Para ello comenzamos utilizando los métodos que nos provee la Estadística con la tarea de recopilación o recolección de los mismos, ya sean provenientes de la población si disponemos de los mismos (como los provenientes de un censo u otros casos particulares), o a través de una muestra que represente a la población en estudio. Si trabajáramos con una muestra, tendríamos que extraerla de la población a la que estamos por analizar por algún método adecuado al tema de estudio que nos suministra la Estadística, ese sería el primer paso. A partir de allí tenemos **los datos en bruto**. Más adelante, en este capítulo se explica los conceptos de población y muestra.

Existen dos formas fundamentales de extraer una muestra:

- a- **Muestreo probabilístico.** Se extrae en forma aleatoria o al azar que es lo mismo. Cada elemento seleccionado de la muestra, tiene la misma probabilidad de salir seleccionado, es un método científico.
- b- **Muestreo no probabilístico.** Se extrae con una clara influencia del juicio del investigador, los elementos que componen la muestra no tienen la misma posibilidad de salir seleccionados. Por ejemplo, en una calle del centro de una ciudad se encuesta a las personas que pasan por allí sobre un determinado tema y se considera que es la opinión de los pobladores de dicha ciudad, lo cual es incorrecto dado que muchos elementos de esa ciudad no entrarán en la muestra. Es un método de extraer que carece de considerarse científico. En este ejemplo, se obtienen conclusiones erróneas y subjetivas.

1.1.1 Muestreo Probabilístico

En el contexto de muestreo probabilístico, existen varias posibilidades de obtención de una muestra:

✓ Muestreo aleatorio simple

En un muestreo aleatorio simple todos los individuos tienen la misma probabilidad de ser seleccionados. La selección de la muestra puede realizarse a través de cualquier mecanismo probabilístico en el que todos los elementos tengan las mismas opciones de salir. Por ejemplo uno de estos mecanismos es utilizar una tabla de números aleatorios, o también con un ordenador generar números aleatorios, comprendidos entre cero y uno, y multiplicarlos por el tamaño de la población, este es el que vamos a utilizar.

✓ Muestreo aleatorio estratificado

Es frecuente que cuando se realiza un estudio interese estudiar una serie de subpoblaciones (estratos) en la población, siendo importante que en la muestra haya representación de todos y cada uno de los estratos considerados. El muestreo aleatorio simple no nos garantiza que tal cosa ocurra. Para evitar esto, se saca una muestra de cada uno de los estratos.

Hay dos conceptos básicos:

- **Estratificación:** El criterio a seguir en la formación de los estratos será formarlos de tal manera que haya la máxima homogeneidad en relación a la variable a estudio dentro de cada estrato y la máxima heterogeneidad entre los estratos.
- **Afijación:** Reparto del tamaño de la muestra en los diferentes estratos o subpoblaciones. Existen varios criterios de afijación entre los que destacamos:
 1. Afijación igual: Todos los estratos tienen el mismo número de elementos en la muestra.
 2. Afijación proporcional: Cada estrato tiene un número de elementos en la muestra proporcional a su tamaño.
 3. Afijación Neyman: Cuando el reparto del tamaño de la muestra se hace de forma proporcional al valor de la dispersión en cada uno de los estratos.

✓ Muestreo aleatorio sistemático

Es un tipo de muestreo aleatorio simple en el que los elementos se seleccionan según un patrón que se inicia con una elección aleatoria.

Considerando una población de N elementos, si queremos extraer una muestra de

tamaño n , partimos de un número $h=N/n$, llamado coeficiente de elevación y tomamos un número al azar a comprendido entre 1 y h que se denomina arranque u origen.

La muestra estará formada por los elementos: $a, a+h, a+2h, \dots, a+(n-1)h$.

De aquí se deduce que un elemento poblacional no podrá aparecer más de una vez en la muestra. La muestra será representativa de la población pero introduce algunos sesgos cuando la población está ordenada en función de determinados criterios.

✓ Muestreo aleatorio por conglomerados o áreas

Mientras que en el muestreo aleatorio estratificado cada estrato presenta cierta homogeneidad, un conglomerado se considera una agrupación de elementos que presentan características similares a toda la población.

Por ejemplo, para analizar los gastos familiares o para controlar el nivel de audiencia de los programas y cadenas de televisión, se utiliza un muestreo por conglomerados-familias que han sido elegidas aleatoriamente. Las familias incluyen personas de todas las edades, muy representativas de las mismas edades y preferencias que la totalidad de la población.

Una vez seleccionados aleatoriamente los conglomerados, se toman todos los elementos de cada uno para formar la muestra. En este tipo de muestreo lo que se elige al azar no son unos cuantos elementos de la población, sino unos grupos de elementos de la población previamente formados. Elegidos estos grupos o "conglomerados" en un número suficiente, se pasa posteriormente a la elección, también al azar, de los elementos que han de ser observados dentro de cada grupo, o bien, según se desee, a la observación de todos los elementos que componen los grupos elegidos.

✓ Muestreo no Probabilístico

Existen otros procedimientos para seleccionar las muestras, que son menos precisos que los citados y que resultan menos costosos. El procedimiento más utilizado es el muestreo no probabilístico, denominado opinático consistente en que el investigador selecciona la muestra que supone sea la más representativa, utilizando un criterio subjetivo y en función de la investigación que se vaya a realizar.

Con el muestreo opinático la realización del trabajo de campo puede simplificarse enormemente pues se puede concentrar mucho la muestra. Sin embargo, al querer concentrar la muestra, se pueden cometer errores y sesgos debidos al investigador y, al tratarse de un muestreo subjetivo (según las preferencias del investigador), los resultados de la encuesta no tienen una fiabilidad estadística exacta.

Un muestreo no probabilístico muy utilizado hoy en día por los institutos de opinión es el de itinerarios, consistente en facilitar al entrevistador el perfil de las personas que tiene que entrevistar en cada uno de los itinerarios en que se realizan las entrevistas.

El muestreo denominado de cuotas, utiliza en sucesivos sondeos al mismo conjunto muestral (inicialmente seleccionado de forma aleatoria) y es el empleado para medir índices de audiencia de programas televisivos

Para el relevamiento de datos se extrajo una muestra de la población a través de la teoría del muestreo probabilístico, cuando se puede determinar de antemano la probabilidad de selección de cada uno de los elementos de la población, la selección debe ser realizada al azar.

Dentro del muestreo probabilístico es un muestreo aleatorio simple, porque cada uno de los elementos de una población tiene la misma posibilidad de ser elegido, la selección es al azar, debe emplearse algún método aleatorio. (Bolillas, tablas)

Ya contamos con todos los datos a nuestra disposición, y nos damos cuenta que están desordenados. ¿Qué hacemos con ellos? Así como están no podemos trabajarlos ni analizarlos, el paso siguiente es ordenarlos. Supongamos que deseamos investigar en un colegio de nivel medio (población) el consumo de cigarrillos en los alumnos, seleccionamos una muestra representativa y obtenemos, por ejemplo, la respuesta de 100 alumnos a través de un cuestionario (hemos extraído los datos).

La tarea siguiente es trabajar los datos, organizándolos y clasificándolos, los que obtuvimos de los cuestionarios. ¿Cómo hacemos? Para ello la Estadística nos proporciona los métodos y técnicas que nos posibilita realizar esa tarea.

Una vez que ya pudimos organizarlos de forma correcta (tablas de frecuencias y gráficos), nuestra próxima labor es describirlos para conocer lo que nos están expresando. ¿Cómo los describimos? La Estadística nos suministra los métodos y técnicas para poder describirlos, a través de diferentes medidas, las cuales nos brindan la información necesaria. Siguiendo el ejemplo, si quisiéramos saber el consumo promedio de cigarrillos de los alumnos de ese colegio, utilizaríamos una medida adecuada, en este caso sería la media aritmética

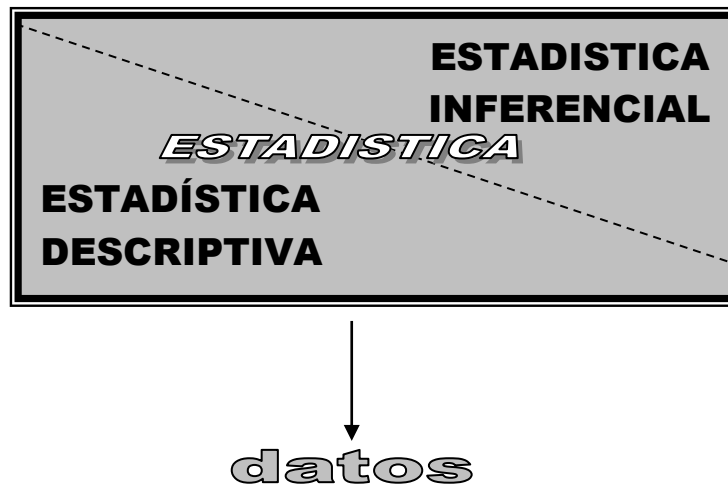
Continuando el ejemplo, supongamos que nos proponemos determinar la relación que hay entre el consumo de alcohol y el nivel socioeconómico de los alumnos seleccionados. Queremos alcanzar de dicho análisis una confiabilidad de los resultados con un 95% de confianza si trabajáramos con una muestra al tratar de estudiar a la población. Para llegar a ese resultado la estadística nos brinda los métodos determinarlos, que nos permitirán adoptar decisiones al respecto. Gran parte de estos métodos y técnicas mencionadas las aprenderemos a través de este libro.

La Estadística se clasifica en dos grandes áreas: **Estadística Descriptiva** y **Estadística Inferencial**.

La **Estadística Descriptiva** nos provee los métodos de recolección, organización, descripción, análisis e interpretación de todo un conjunto de datos.

La **Estadística Inferencial** nos brinda los métodos que permiten estimar o estudiar una característica de una población en base a resultados muestrales. La muestra o las muestras deben ser representativas de la población para que las conclusiones logradas sean fehacientes.

Gráficamente, podemos observar la división en dos áreas:



Los datos a través de los cuales se alimenta la Estadística, son la materia prima de ella. Pasamos a explicar dos conceptos importantes: población y muestra.

1.1. Población

Una población puede definirse como el conjunto de todas las observaciones posibles de medidas o resultados. Otra definición: **Población** es la colección de todos los individuos, objetos u observaciones que poseen al menos una característica común. Ejemplos de población son:

- a. Los ingresos de los habitantes de un país en un período determinado
- b. Las edades de los estudiantes del nivel medio en una localidad.
- c. Las remuneraciones que perciben los empleados de una fábrica
- d. Las patentes de las motos que circulan en un país.
- e. Los pacientes con la enfermedad del dengue sometido a una determinada técnica de curación.
- f. Los enfermos de alcoholismo de la provincia de Jujuy.
- g. Habitantes de un determinado país.
- h. El peso de animales en todos los zoológicos de un país.

- Como vemos en Estadística difiere el concepto de población no queda acotado en relación a Demografía, donde los elementos de la población son las personas únicamente. En Estadística el concepto de población es más abarcativo, ya que los elementos de la población pueden ser no solamente las personas, sino también los animales, las plantas, automóviles etc...

La población puede clasificarse en: **finita** e **infinita**.

Población Finita: Una población finita es aquella que tiene un número limitado de elementos. Por ejemplo, al hablar de “los estudiantes de una escuela”, definimos una población acotada, por lo tanto finita. Podemos contar los elementos sin ninguna inconveniente ni problema.

Población Infinita: Una población infinita es aquella que no tiene límite o cotas, es decir, tiene un número ilimitado de elementos. Por ejemplo, al hablar de los “pacientes con poliomielitis, sometidos a una determinada terapia de rehabilitación” definimos una población no limitada, cuyos elementos son, tanto las personas que ya han sido sometidos a la referida terapia, como los que sean abordados en el futuro.

1.1.1 Parámetros

Un parámetro es una medida resumen que describe una característica de toda la población. Los parámetros necesitan para determinar su valor, utilizar toda la información de la población. Por ejemplo, si queremos determinar la edad promedio (parámetro) de los alumnos de un colegio, necesitamos considerar todas las edades de los alumnos del colegio. Si deseamos determinar la proporción de enfermos de gripe en una localidad, debemos considerar toda la población que posee esa característica en dicha localidad.

Los parámetros que más se utilizan son:

- La media poblacional: μ
- La proporción poblacional : π
- El desvío estándar poblacional: σ
- Los parámetros de regresión y correlación

Es obvio que si deseamos calcular un parámetro de una población infinita no es posible. Si se trata de una población finita y quisiéramos determinar un parámetro de la misma nos resultaría muy oneroso y nos ocuparía mucho tiempo. Es por ello que en la práctica, se utiliza sólo una parte de la población que sea representativa de la misma, a lo que llamaremos **muestra** y a partir de la misma estimaremos el parámetro de la población con un determinado nivel de confiabilidad. Un clásico ejemplo donde se considera a la población, es el “Censo Nacional de Población”, se realiza cada diez años en nuestro país, y se consideran todas las personas con sus respectivas características tales como edades, vivienda, sexo, etc. Dicho censo lleva un considerable empleo de tiempo para planificarlo, y al implementarlo, insume una suma considerable de dinero, dado que se debe emplear recursos humanos para recabar los datos y luego procesarlos y finalmente publicarlos. Es por ello que se realiza cada diez años. El mismo está a cargo del Instituto Nacional de Estadística y Censo (I.N.D.E.C.) en nuestro país. Se realiza en un día determinado únicamente y resulta ser una foto de la población en el día previsto.

1.2. Muestra

Es un conjunto de medidas o resultados seleccionados a partir de la población. Es un subconjunto representativo de la población. Desde luego, el número de observaciones en una muestra es menor que el número de observaciones posibles de la población. Al seleccionar una muestra, ella debe ser representativa de la población, es decir debe reunir todas las características de la población para que tengan veracidad los datos que obtenemos de ella, con un margen de error. Si la muestra reúne esas condiciones, es la población en miniatura y es solamente a través de ella, que podemos realizar inferencias en forma válida, los resultados obtenidos de la población en su totalidad.

Por ejemplo, si se desea estimar el gasto promedio anual de los estudiantes universitarios de Jujuy, se extraería una muestra formada por cierto número de estudiantes, en seguida se determinaría el gasto anual correspondiente a cada uno de

ellos y después se obtendría el promedio de estos gastos. Se utiliza una muestra debido a que simplemente no se tiene el tiempo y los recursos para establecer contacto con todos los estudiantes universitarios de Jujuy. Aun cuando es posible hacerlo, tomaríamos una muestra dado que los resultados que obtenemos mediante un método apropiado, son confiables con un pequeño error. En base al promedio surgido de la muestra (promedio muestral), se realiza una inferencia acerca del gasto promedio de todos los estudiantes universitarios de Jujuy. La muestra no sería representativa si obviáramos alguna facultad.

Otro ejemplo, para determinar la proporción de piezas defectuosas producidas en determinado proceso de fabricación, los técnicos de control de calidad examinan un lote de unidades producidas para determinar el número de defectuosas contenidas en él (el lote constituye una muestra). La proporción de la población, la cual es un parámetro que se desconoce, es la proporción de todas las unidades defectuosas producidas en el proceso; se la estima mediante la proporción de la muestra, la cual es la proporción de las unidades defectuosas contenidas en la misma. Así, si en un lote de 300 unidades producidas en el proceso, el ingeniero de control de calidad encuentra 45 defectuosas, entonces la proporción de defectuosas en la muestra será

$$45/300 = 0.15 \text{ (15\%)}$$

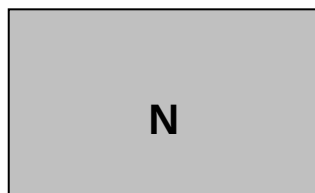
Obsérvese que tanto el promedio de la muestra como la proporción de la muestra en los ejemplos anteriores, son características medibles surgidas de las muestras.

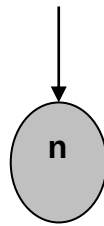
La selección de la muestra o muestras se realiza por diferentes métodos que suministra la estadística, y cada método va a depender del fenómeno en estudio. Se debe ser cuidadoso al seleccionar una muestra, dado que los elementos seleccionados deben contener las características de la población para que la misma sea representativa con un pequeño error. Existen variados ejemplos de selección de muestra en forma incorrecta. A mediados del siglo pasado en Estados Unidos, para la elección de presidente, se realizó una encuesta telefónica. Los resultados de dicha encuesta resultaron inversos a los resultados del escrutinio. La explicación de ello radica en que no se consideraron todos los elementos de la población, más precisamente no todos tienen teléfono y más en esa época, dejando de lado a sectores de la población fundamentalmente aquellos de bajos ingresos. Últimamente en nuestro país, a partir del restablecimiento de la democracia en 1983, es notoria la cantidad de encuestas de intención de voto que se realizan frente a un escrutinio. También es permanente las encuestas de opinión que encarga el gobierno de turno, para algunos temas que figuran en la agenda oficial que son de su interés.

1.2.1 Estadístico

El estadístico es una medida resumen que describe una característica de la muestra. Por ejemplo, la estatura promedio de los alumnos de una escuela. Se selecciona una muestra que los represente a los alumnos de dicha escuela, se mide la estatura de los alumnos pertenecientes a la muestra, y se calcula el promedio de las estaturas. Ese promedio es un estadístico.

Los métodos de selección de muestras los veremos más adelante en este texto.



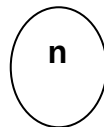


Simbolizamos a la población y a la muestra de esta manera:

N: Población
n: Muestra

En esta figura, una forma gráfica de visualizar cuando una muestra es representativa, es que la parte extraída de la población guarda el mismo “color” de la misma. Si la muestra no fuera representativa tendría un “color” distinto al de la población. Ahora, cuando seleccionamos una muestra, ésta no es exacta, puede haber un pequeño error al extraerla, de allí que trabajamos con un cierto nivel de confianza, que precisamente no es del 100 %. El resto que no entra en el nivel de confianza lo llamaremos error de muestreo.

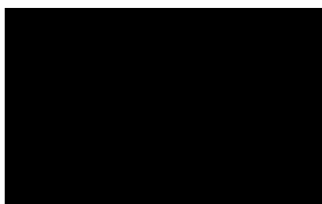
Muestra no representativa



Ejemplo 1.0 Se realiza una encuesta de intención de voto previa a una elección, sobre las preferencias de los electores en un departamento de la provincia, para votar concejales. Se presentan en la elección dos listas pertenecientes a distintos partidos políticos. Se selecciona una muestra y a cada persona que salió seleccionada se la entrevista. La muestra seleccionada fue de 1800 electores, de los cuales 1100 manifestaron estar a favor de la lista 1.

- ¿Cuál es la población en estudio?
- ¿Cómo está constituida la muestra?
- ¿La población es finita o infinita?
- ¿Cuál es el parámetro de la población?
- ¿Se conoce el parámetro poblacional?
- ¿Cuál es el estadístico de la muestra?
- ¿Para qué calculo el estadístico muestral?

Respuesta:



LISTA 2

?

- a) La población en estudio está constituida por todos los electores registrados en el municipio.
- b) La muestra seleccionada está constituida por 1800 electores. Otra forma de decir lo mismo es que el tamaño de la muestra es de 1800 electores.
- c) La población es finita, es mensurable y la puedo obtener de los padrones electorales.
- d) El parámetro de la población es la proporción de electores de la población, que están a favor de la lista 1 y la proporción de electores que tienen preferencias por la lista 2.
- e) El parámetro poblacional es desconocido, ya que no se entrevistó a todos los electores de la población y por ende no se calculó la proporción de electores que tienen determinadas preferencias de la misma.
- f) El estadístico es **la proporción** de electores a favor de alguna lista. Por ejemplo, la proporción de electores que votarían la lista 1 es:

$$p = \frac{1100}{1800} = 0.61$$
$$p = 0.61$$

- g) Calculamos el estadístico dado que trabajamos con la muestra e inferimos o estimamos a través del estadístico de la muestra, el parámetro poblacional.

Ejemplo 1.1 En un colegio se está organizando para fin de año un concurso de baile, para ello, se desea saber las preferencias de los alumnos por los distintos géneros musicales. De ésta forma se organizarán los números musicales por género musical y se tomarán los tres más elegidos de cada uno. Se selecciona una muestra de 450 alumnos y los tres géneros musicales más votados siguen en el siguiente orden:

- 1. Rock & roll (190 alumnos)
- 2. Folklore (160 alumnos)
- 3. Tango (100 alumnos)

- 1) ¿Cuál es la población en estudio?
- 2) ¿Cómo está constituida la muestra?
- 3) ¿La población es finita o infinita?
- 4) ¿Cuál es el parámetro de la población?
- 5) ¿Se conoce el parámetro poblacional?
- 6) ¿Cuál es el estadístico de la muestra?

Respuesta:

- 1) La población en estudio son los estudiantes del colegio.

- 2) La muestra está constituida por 450 estudiantes seleccionados del colegio.
- 3) La población es finita.
- 4) El parámetro de la población es la **proporción** de alumnos que tienen preferencia por un determinado género musical.
- 5) No se conoce, porque no tengo la información de la cantidad de alumnos del colegio, puedo inferirlo a través de la muestra.
- 6) El estadístico muestral es la **proporción** de alumnos de la muestra que tienen preferencia por un género musical. Por ejemplo, la proporción de alumnos que tienen preferencia por el tango es:

$$100/450 = 0.22$$

Ejemplo 1.2 Un equipo de pediatras pertenecientes al Ministerio de Bienestar Social de la provincia realizará un trabajo de investigación, sobre los recién nacidos en los hospitales públicos de la provincia. Uno de los objetivos de dicho trabajo es determinar el peso promedio de los recién nacidos. Para ello se realiza un registro del peso de todos los recién nacidos en cada hospital público de la provincia durante un mes. Finalmente se calcula el peso promedio.

1. ¿Cuál es la población en estudio?
2. ¿La población es finita o infinita?
3. ¿Cuál es el parámetro de la población?
4. ¿Se podría haber tomado una muestra para inferir el peso promedio de los recién nacidos?

Respuesta:

1. La población en estudio son los recién nacidos de los hospitales públicos de la provincia.
2. La población es finita.
3. El parámetro poblacional es el peso **promedio** de los recién nacidos.
4. Se podría haber tomado una muestra en hospitales públicos de cada localidad de la provincia.

Actividad 1:



En un museo se realiza un estudio para determinar el promedio de edad de los visitantes y las preferencias por las pinturas de arte moderno. Se tomó una muestra de 500 personas dando el promedio de edad de 30 años y 300 personas manifestaron su preferencia por las pinturas de arte moderno.

- a. ¿Cuál es la población en estudio?
- b. ¿La población es finita o infinita?
- c. ¿Cuál o cuáles son los parámetros poblacionales?
- d. ¿Cómo está constituida la muestra?
- e. ¿Cuál es el estadístico o estadísticos de la muestra?

1.3 Variables: concepto, clasificación y escalas de medición.

El término variable proviene del latín **variabilis**.

Es una característica que se puede medir o contar y que varía de un elemento a otro de la población o de la muestra. En todo trabajo de investigación trabajamos con variables, es fundamental comprender su significado y sus aplicaciones. Para entender el concepto recurramos a un ejemplo de variable:

“Número de hijos por familia, con necesidades básicas insatisfechas que viven en un barrio capitalino”

Siguiendo los distintos componentes de la definición propuesta, observaremos como el ejemplo queda encuadrado en la misma:

✓ ¿Cuál es la característica que podemos medir o contar?

La característica en estudio que se puede medir o contar, es el número de hijos de familias con necesidades básicas insatisfechas. Por ejemplo: 0, 1, 2, 3, 4 hijos

✓ ¿Cómo varía de un elemento a otro?

El número de hijos va a variar de una familia a otra con la condición mencionada. Una familia tendrá 2 hijos, otra familia ningún hijo, otra familia 6 hijos, es decir, el número de hijos va a ir variando según cada familia.

✓ ¿Esas familias provienen de una población o de una muestra?

Si consideramos todo el barrio, éste será nuestra población. En cambio, si seleccionamos una muestra del barrio, las familias escogidas provendrán de una muestra.

Queda probado que el ejemplo encierra satisfactoriamente lo que se denomina variable.

Si no hay variación de un elemento a otro, estamos en presencia de lo que denominamos **constantes**. Ello lo podemos observar a través de un ejemplo:

“Exportaciones de carne de Argentina en el año 2005”

Es un ejemplo de constante, supongamos que las exportaciones de carne en el año 2005 fueron de 100000 toneladas. La característica es la exportación de carne, pero no hay variación, por lo tanto no es una variable sino una constante. Para que sea una variable el ejemplo dado diría así: exportaciones de carne de Argentina en el año 2005, **por mes**. Es decir varía de un mes a otro del año mencionado. Queda clara la diferencia entre variable y constante.

Las variables pueden clasificarse de dos formas: **cualitativas** y **cuantitativas**.

Las variables **cualitativas** son aquellas que están definidas por las categorías que la componen. Es decir adoptan categorías y no valores numéricos. Algunos autores también las llaman variables categóricas. Ejemplos:

<u>variable</u>	<u>categorías</u>	<u>variable</u>	<u>categorías</u>
Sexo →	masculino	Altura de las plantas	→ alta
  →	femenino		→ media
			→ baja

Dentro de este tipo de variable se encuentra como caso particular, aquella que asume sólo dos categorías: se denomina variable **dicotómica** o **binaria**. Un ejemplo es la variable sexo. También las variables cualitativas adoptan más de dos categorías.

Las categorías que componen una variable cualitativa deben cumplir las siguientes condiciones:

- **Mutuamente excluyentes:** debe ser incluida en una sola categoría, ya sea la persona o el objeto. En el ejemplo de la variable sexo, la persona debe aparecer en una sola categoría (varón o mujer).
- **Exhaustivas:** el individuo u objeto debe aparecer en una categoría, es decir, si por ejemplo una persona pertenece a una determinada categoría, debe estar presente dicha categoría. Por ejemplo si la variable fuera tipo de religión y hay una persona que es musulmán, debe aparecer esta categoría para que dicha persona esté considerada.

Las variables **cuantitativas**, son como su nombre lo indica, se expresan en forma numérica, sus valores provienen de medir o contar las observaciones. Dentro de este tipo de variable, se encuentran dos clasificaciones:

_____ variable discreta
_____ variable continua

Una variable **continua** es una variable que puede tomar cualquier valor sobre el eje numérico o una parte de él y sus valores provienen de medir los elementos. Ejemplos de variables continuas son: la altura, el peso, la temperatura, etc.

1 1.2 1.65 - 0,8 ...

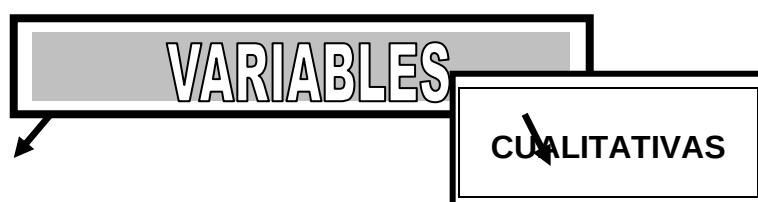
Contrariamente a las variables continuas, las variables **discretas** sólo pueden tomar algunos valores determinados sobre la escala numérica es decir, valores enteros y se puede contar.

10 12 20 ...

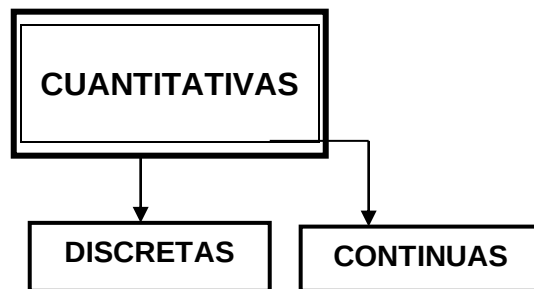
Ejemplos de variables discretas son: número de hijos, cantidad de alumnos, etc. No podemos decir un hijo y medio, un alumno y tres cuartos, carece de sentido, es por ello que adoptan valores enteros.

La diferencia entre ambas variables cuantitativas, es que una puede tomar valores enteros (discreta) y la otra no solo valores enteros, sino cualquier tipo de valor numérico (continua).

La clasificación de variables se presentan en el siguiente diagrama:



Escalas de medición



Una escala se define como un instrumento de medición.

Los datos que se obtienen para una variable cualitativa, se miden en una escala nominal u ordinal.

Escala nominal: cuando los datos observados simplemente se clasifican en distintas categorías que **no implican un orden preestablecido**, se tiene un nivel de medición nominal. Ejemplo:

Variable	Categorías
Afiliación política	Partido Radical, Socialista, Peronista, etc
Lugar de nacimiento	Jujuy, Mendoza, Capital Federal, Salta, etc

Escala ordinal: si los datos observados se clasifican en distintas categorías en las que **admiten algún orden**, se tiene una medición ordinal. Ejemplos:

Variable	Categorías
Calificación de Estudiantes	(ordenadas de mayor a menor) A B C D

El orden de las categorías puede ser de mayor a menor, o de menor a mayor indistintamente.

Escala de razón y escala de intervalo

De razón: corresponde o se utiliza cuando el cero es natural, es decir, significa ausencia de algo. Permite establecer proporciones entre los valores. Ej : número de cuentas a cobrar por mes, número de hijos.

De intervalo: se utilizan cuando el cero es convencional (no representa ausencia) y no admite proporciones entre los valores. Ej temperaturas registradas por día

Ejemplo 1.3 De las siguientes variables cualitativas indique su escala de medición:

- Países del mundo.
- Animales en el Amazonas.
- Empresas TOP de la Argentina por sus ventas registradas.
- Los 40 temas musicales o canciones más votados en FM 100.

- v. Tipo de fobias que sufren determinadas personas.

Respuestas:

- i. Nominal.
- ii. Nominal.
- iii. Ordinal.
- iv. Ordinal.
- v. Nominal.

Ejemplo 1.4 Clasifique las siguientes variables e indique su escala de medición:

- 1. Estrellas más famosas de la televisión.
- 2. Libros más vendidos en el 2011.
- 3. Tipo de conductas de algunos enfermos.
- 4. Evolución de las exportaciones en el año 2010, por mes.
- 5. Montos de facturas impagas de una empresa.
- 6. Agrupaciones de desocupados en la Argentina.
- 7. Número de muertos por día en África.
- 8. Número de hijos en barrios con N.B.I. de una localidad.

Respuestas:

- 1. Variable cualitativa. Escala Ordinal.
- 2. Variable cualitativa. Escala Ordinal.
- 3. Variable cualitativa. Escala Nominal.
- 4. Variable cuantitativa continua. Escala de Razón.
- 5. Variable cuantitativa continua. Escala de Razón.
- 6. Variable cualitativa. Escala Nominal.
- 7. Variable cuantitativa discreta. Escala de Razón.
- 8. Variable cuantitativa discreta. Escala de Razón.

Actividad 2:



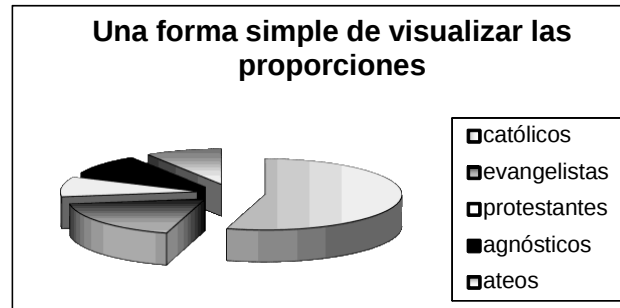
Clasifique las siguientes variables e indique su escala de medición.

- a) Promedios de calificación de estudiantes de los cursos de un colegio.
- b) Distancias que los estudiantes recorren para llegar a clase.
- c) Calificaciones de los estudiantes en el examen de Estadística
- d) Lugar de nacimiento de los alumnos de un colegio.
- e) Número de horas de estudio de los estudiantes a la semana.
- f) Cantidad de alumnos según el año escolar donde se encuentran.
- g) Número de inasistencias de los alumnos por curso de un colegio.
- h) Grado de aprendizaje de los alumnos en una materia.
- i) Motivos de pérdida del año de los alumnos de un colegio.
- j) Prioridades que consideran los profesores en la enseñanza.
- k) Tipo de conductas de los alumnos.
- l) Tipo de patologías que sufren pacientes que iniciaron un proceso terapéutico.

1.4 Proporciones, razones y porcentajes

Dada la utilización de estos conceptos matemáticos, se incluyen en este texto para refrescarlos.

1.4.1 Proporciones



Un grupo determinado de alumnos puede constar de 30 varones y de 25 mujeres, o de 30 católicos, 10 evangelistas, 5 protestantes, 5 agnósticos y 5 ateos. Sin embargo, para poder establecer comparaciones con otros grupos, hay que tener en cuenta el número de casos en cada uno de los grupos considerados. Las medidas que sirven para comparar dichos grupos son: las proporciones, las razones y los porcentajes.

Para poder aplicar las proporciones, suponemos que el método de clasificación fue tal que las categorías son mutuamente exclusivas y exhaustivas: cada individuo ha sido puesto en una categoría y en una sola. Si N es el número total de casos y tomamos una variable, por ejemplo, que consta de tres categorías: N_1 , N_2 y N_3 , **la proporción de casos en cualquier categoría dada está definida como el número en la categoría dividido entre el número de total de casos**. Por lo tanto, la proporción de individuos de la primera categoría esta dada por N_1 / N y así con los otros casos.

El valor de una proporción no puede ser mayor que la unidad. Esto es:

$$N_1 + N_2 + N_3 = N$$

Tenemos que

$$\frac{N_1}{N} + \frac{N_2}{N} + \frac{N_3}{N} = \frac{N}{N} = 1$$

Si sumamos las proporciones de los casos de todas las categorías (mutuamente excluyentes), el resultado es la unidad.

Ejemplo 1.5:



Se obtuvieron los siguientes datos, a partir de una encuesta en dos localidades de una región del país. La finalidad de la encuesta era conocer el hábito a la lectura de los habitantes de esas dos localidades, para tomar medidas concretas sobre el tema y poder comparar entre las dos localidades.

HABITANTES	LOCALIDAD 1	LOCALIDAD 2
LEE POCO	25	60
LEE RAZONABLEMENTE	30	80
LEE MUCHO	45	150
TOTAL	100	290

Los datos que se obtienen para cada categoría de la variable, se denominan frecuencias absolutas. En este caso, al tratarse de variable cualitativa, las frecuencias absolutas nos indican la cantidad de veces que se repite cada categoría. Por ejemplo, 25 habitantes de la localidad 1 leen poco, 30 habitantes de la localidad 1 leen razonablemente, y así sucesivamente.

Dado que los datos obtenidos de las muestras en las dos localidades, no me permite comparar, es necesario transformar la tabla en proporciones.

La siguiente tabla está expresada en proporciones y de ella puedo inferir cuestiones importantes sobre el tema.

LEE POCO	0.25	0.21
LEE RAZONABLEMENTE	0.30	0.28
LEE MUCHO	0.45	0.51
TOTAL	1.00	1.00

Al haber transformado los valores en proporciones, puedo concluir:

- ❖ Tanto en la localidad 1 y 2, es mayor la proporción de habitantes que leen mucho.
- ❖ Es mayor la proporción de habitantes que leen mucho en la localidad 2, en relación a la localidad 1.
- ❖ La proporción de los habitantes que leen razonablemente en la localidad 1, es de 0.45.

Y así, sucesivamente puedo inferir más cosas. De esta forma se ve la utilidad de las proporciones. A veces por problemas de decimales, en el total no da exactamente 1. Una forma de resolver esta cuestión, es agregar o disminuir en la última cifra decimal de la proporción mayor, dado que no afecta como si hiciéramos lo mismo en una proporción pequeña, donde la incidencia es mayor.

Los valores entre los que varía la proporción son:

$$0 \leq p \leq 1$$

1.4.2 Razones

El cociente entre dos frecuencias da por resultado una razón. Las proporciones representan un tipo especial, donde el denominador es el número total de casos y el numerador una fracción de él. A diferencia de las proporciones la razón puede adoptar un número mayor a 1.

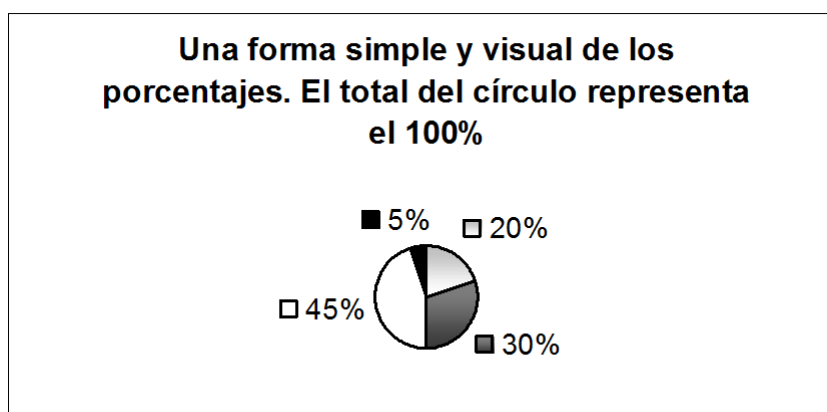
- De la observación en la localidad 2, podemos afirmar: La razón entre los habitantes que leen mucho y los que leen poco es $150:60 = 2.5$. En algunos casos es conveniente expresar la razón, en términos de un denominador formado por la unidad. En el ejemplo la razón de los que leen mucho y los que leen poco puede escribirse como 2.5 a 1. Es decir que de cada 1 habitante que lee poco 2.5 habitantes leen mucho.
- Si la variable es dicotómica, es posible calcular la proporción directamente a través de la razón o viceversa. Ejemplo: si la razón de los varones a las mujeres es de 2:4, entonces de cada seis personas se va a dar un promedio de dos varones y cuatro mujeres. Si lo enfocamos como razón podemos afirmar que de cada cuatro mujeres hay dos varones.
- Cuando la variable contiene muchas categorías es preferible usar las proporciones o los porcentajes, por razones económicas, de comprensión y de simplicidad.
- Las razones pueden expresarse en términos de cualquier base que resulte más conveniente. La base está indicada por el denominador. Por ejemplo podemos elegir la base 100 para denotar la razón de la variable sexo. Una razón de 84 en materia de sexo nos estaría indicando que el número de varones es notoriamente inferior al de las mujeres.
- Las tasas de crecimiento constituyen otro tipo de razón. Al calcular las tasas, consideramos el crecimiento real durante el periodo que analizamos y lo dividimos entre el volumen al principio del periodo. Por ejemplo, si la población de una ciudad aumenta de 50000 a 65000 habitantes entre 1990 y el año 2000, la tasa de crecimiento será de:

$$\frac{65000 - 50000}{50000} = 0.30$$

La tasa de crecimiento fue del 30%.

1.4.3 Porcentajes

Los porcentajes pueden obtenerse de las proporciones multiplicando simplemente por 100.



La palabra porcentaje significa por ciento. Por lo tanto, calculamos el número de individuos o cosas que habría en una categoría determinada si el total de casos fuera 100, permaneciendo inalterada la proporción en cada categoría. Los porcentajes sumados dan 100, salvo que las categorías no sean mutuamente excluyentes. El uso

de los porcentajes es más frecuente que el de las proporciones, dado que para los usuarios de los resultados de cualquier estudio, le resulta más comprensible interpretar. Con el ejemplo anterior de proporciones, expresamos los resultados en porcentajes:

HABITANTES	LOCALIDAD 1	LOCALIDAD 2
LEE POCO	25	21
LEE RAZONABLEMENTE	30	28
LEE MUCHO	45	51
TOTAL	100	100

Las mismas interpretaciones la referimos en porcentajes.

- ❖ Tanto en la localidad 1 y 2, es mayor el porcentaje de habitantes que leen mucho. Estaríamos ante una población proclive a la lectura.
- ❖ Es mayor el porcentaje de habitantes que leen mucho en la localidad 2, en relación a la localidad 1.
- ❖ El porcentaje de habitantes que leen razonablemente en la localidad 1, es del 45%

Por último, los porcentajes pueden calcularse en forma vertical u horizontal.

Ejemplo 1.6

Supongamos tener como población de estudio un grupo de 35 adolescentes varones que tienen entre 14 y 18 años y los entrevistamos para averiguar el deporte que más les gusta.

Una vez elegida la población (en este caso los 35 adolescentes), seleccionamos la variable (en esta caso “deporte”) que queremos estudiar y el instrumento que nos permitirá recolectar el conjunto de datos para analizar (en este caso la “entrevista”).

Las categorías que admite esta variable pueden ser: fútbol, básquet, tenis etc. Es decir que **deporte** es una variable **categorica que no admite orden**.

Volviendo al ejemplo, luego de las entrevistas supongamos tener los siguientes resultados:

Ninguno – Fútbol – Fútbol – Tenis – Básquet – Fútbol – Fútbol – Fútbol – Ninguno – Tenis - Básquet – Fútbol - Básquet - Básquet – Ninguno – Tenis – Fútbol – Fútbol - Básquet – Fútbol – Fútbol – Fútbol – Fútbol – Ninguno – Ninguno – Tenis - Básquet - Básquet – Fútbol - Básquet - Básquet – Fútbol – Ninguno – Fútbol – Fútbol

Organizamos los datos obtenidos en una tabla denominada **tabla de distribución de frecuencias**, donde en la primera columna colocamos las categorías que toma la variable y en la segunda la cantidad de veces que aparece cada una de esas categorías.

Deporte X	Cantidad f	Proporciones	Porcentajes
Ninguno	6	0,1714 (6/35)	17,14
Fútbol	16	0,4572	45,72

Tenis	4	0,1143	11,43
Básquet	9	0,2571	25,71
Total (n)	35	1	100

f se denomina **frecuencia absoluta** e indica la cantidad de veces que aparece cada valor o categoría de la variable. La suma de todos los valores de **f** da como resultado el total de las observaciones. A este valor se lo denomina **n**.

Algunos resultados

De la observación de la columna de las f_i podemos decir que

- la proporción de chicos que prefieren el fútbol es de 0,4572
- la proporción de jóvenes que no les gusta algún deporte es 0,1714
- la proporción de chicos que elige algún deporte es 0,8286

De la observación de la columna de las $f\%$ podemos decir que

- el 25,71% de los jóvenes encuestados prefiere el básquet.
- sólo el 11,43% de los jóvenes elige el tenis.

El cociente entre dos frecuencias da por resultado una razón.

la razón entre los que prefieren el fútbol y los que prefieren básquet es $16:9 = 1,8$. Es decir que aproximadamente de cada 1 chico que prefiere el básquet 1,8 prefiere el fútbol (la razón es de casi 2 a 1)

1.5 Indicadores de control

Actividad 1:

- a. La población en estudio son las personas que visitan el museo
- b. La población es finita.
- c. Los parámetros poblacionales son: el promedio de edad y la proporción de las personas que visitan el museo.
- d. La muestra está constituida por 500 personas.
- e. Los estadísticos son: el promedio de edad de las personas es de 30 años, y la proporción de personas que prefieren el arte moderno es $300/500 = 0.6$

Actividad 2:

- a) Variable cuantitativa continua. Escala de razón.
- b) Variable cuantitativa continua. Escala de razón.
- c) Variable cuantitativa continua. Escala de razón.
- d) Variable cualitativa. Escala nominal.
- e) Variable cuantitativa continua. Escala de razón.
- f) Variable cuantitativa discreta. Escala de razón.
- g) Variable cuantitativa discreta. Escala de razón.
- h) Variable cualitativa. Escala nominal.
- i) Variable cualitativa. Escala nominal.
- j) Variable cualitativa. Escala ordinal.
- k) Variable cualitativa. Escala nominal.
- l) Variable cualitativa. Escala nominal.

1.6



PROBLEMAS PROPUESTOS

GUIA DE PREGUNTAS:

- ✓ ¿Cuáles son las áreas de la Estadística?
- ✓ ¿Cuál es la diferencia entre población y muestra?
- ✓ ¿Cómo se clasifican las variables?
- ✓ ¿Cuáles son las escalas de medición para las variables cualitativas y cuantitativas?. Explique la diferencia.
- ✓ ¿Cuál es la diferencia entre razón y proporción?
- ✓ Dé dos ejemplos de población y muestra.

Problema 1: Suponga que los ingresantes al primer año de su carrera consta de 3000 estudiantes, todos los cuales han dado un examen único de selección. Se seleccionan 120 estudiantes para conocer la calificación promedio. Indique cual es la población y cual es la muestra.

Problema 2: Se va a implementar un programa de salud mental en la provincia. Se realiza una prueba piloto en hospitales públicos y se seleccionan 300 adultas. Indicar cual seria la población y cual la muestra en este caso.

Problema 3: Se realiza una votación preliminar para determinar las preferencias de los electores en una elección municipal. Con este fin se entrevistan 1500 electores registrados y entre ellos 860 están a favor del candidato A. Responda lo siguiente:

- a) ¿Qué constituye la muestra?
- b) ¿Qué constituye la población?
- c) ¿La población es finita o infinita?

Problema 4: Supóngase que el 60% de todos los electores registrados en un país son integrantes de un partido A y el 40% no. A partir de una muestra de 500 electores, se encuentra que 250 pertenecen al partido A. Responda lo siguiente:

- a) ¿Cuál es la proporción de electores en la muestra que pertenecen al partido A?
- b) ¿Cuál es la proporción de electores en la población que pertenecen al partido A?
- c) ¿Cuál es la población? ¿es finita o infinita?

Problema 5: Establezca la veracidad o falsedad de cada una de las proposiciones siguientes. Reemplace cada enunciado falso por la proposición verdadera correspondiente.

1. La población es una colección de todos los elementos que estamos estudiando.
2. Un estadístico es una característica de la población.
3. La Inferencia Estadística, formula inferencias con respecto a una muestra.

Problema 6: Para cada una de las siguientes proposiciones, defínase la población que se está muestreando.

- a- Se entrevista a 800 estudiantes de la carrera de Profesorado en Psicología y Trabajo Social para determinar sus opiniones (respuestas si o no) acerca de la conveniencia de implementar un programa de prevención contra el alcoholismo.
- b- Se selecciona una muestra de recién nacidos en un hospital en el año 2000, y se lleva un registro de sus pesos al nacer, para luego determinar el peso promedio de los bebés.

Problema 7:

Clasificar las siguientes variables en cuantitativas o cualitativas e indicar la escala de medición:

- a) Días de la semana.
- b) Ingresos de los empleados públicos.
- c) Kilos de verduras vendidos por día en un supermercado.
- d) Estaciones del año.
- e) Temperatura en grados centígrados de un grupo de enfermos.
- f) Signos del zodiaco.
- g) Estatura en centímetros de los alumnos de un colegio.
- h) Meses del año.
- i) Tipos de tabaco producidos en la provincia de Jujuy.
- j) Producción de neumáticos en Argentina durante el año 2010.
- k) Destino de las exportaciones argentinas de trigo durante 2010.
- l) Tipos de programas sociales implementados en Argentina en el año 2005.
- m) Tipos de patologías sufridas por un grupo de personas que llevan a cabo un proceso terapéutico.
- n) Tipo de discriminaciones denunciadas en la Argentina en el año 2010.
- ñ) Tipo de desnutrición en niños de hasta 4 años en la provincia de Jujuy en el año 2009.
- o) Número de recitales de rock en Buenos Aires por mes en el año 2010.

Problema 8:

Clasificar las siguientes variables cuantitativas en continuas o discretas e indicar la escala de medición:

- a) Cantidad de azúcar producido en la provincia de Jujuy en los últimos 20 años
- b) Número de alumnos matriculados en 1995 en distintas escuelas secundarias
- c) Temperatura en grados centígrados necesarias para la cocción de diferentes tipos de pan
- d) Número de jefes de familia desocupados en las ciudades capitales de las provincias argentinas
- e) Número de personas afectadas por el cólera en la provincia de Jujuy en el año 1996 por meses
- f) Toneladas de trigo exportadas al Brasil durante los últimos 10 años

Problema 9: Enumere cinco variables que utiliza con frecuencia en su carrera. Clasifíquelas.

Problema 10: Supóngase que se da el siguiente cuadro que muestra la relación entre la asistencia a la biblioteca y el año de clase en una determinada institución educativa:

Asistencia a la biblioteca	1er. Año	Año de clase 2º Año	3er. Año	4to. Año	TOTAL
Asistencia regular	83	71	82	59	295
Asistencia irregular	31	44	61	78	214
TOTAL	114	115	143	137	509

- Construya dos tablas: una que indique las proporciones y otra los porcentajes.
- ¿Cuál es el porcentaje de asistencia regular en el conjunto?
- ¿Cuál es la proporción de los asistentes irregulares entre los estudiantes de cuarto año? ¿La proporción de estudiantes de cuarto año?
- ¿Hay relativamente más asistentes irregulares entre los estudiantes de 1º y 2º años que entre los de 3º y 4º años?. Exprésense los resultados en porcentajes.
- ¿Cuál es la razón de los estudiantes de primer año a los del año superior?
- Resúmanse los datos en varias proposiciones.

Problema 11:

Responder:

- Si se tiene una variable categórica que no admite orden ¿cuál será la escala de medición que le corresponde?
- Si se tiene una variable categórica que sí admite orden ¿cuál será la escala de medición que le corresponde?
- Si se tiene una variable cuantitativa ¿cuál será la escala de medición que le corresponde?
- Decidir las escalas de medición para cada una de las siguientes variables. (Sugerencia: es conveniente identificar primero el tipo de variable). Entre paréntesis se indican los valores o categorías de las mismas.
 - Sexo (F, M)
 - Estatura (en centímetros)
 - Grado de conformismo con la tarea realizada (muy disconforme, disconforme, conforme., muy conforme)
 - Nivel de escolaridad (Primario incompleto, primario completo, sec. incomp., sec. comp., univ. incomp., univ. comp.)
 - Profesión de algunas personas (abogado, economista, contador, médico)
 - Cantidad de hermanos en diferentes familias (0, 1, 2, 3, 4, 5,6)