

## **Probabilidad y Estadística. Unidad 1: La Estadística.**

**Año de cursado: 2024**

### **Clase N.º 1: La Estadística, definiciones y conceptos básicos.**

**Contenido:** Definición de estadística. Etapas del método estadístico. Conceptos básicos. Organización y presentación de datos. Gráficos.

#### **1. Presentación:**

**Bienvenidos y bienvenidas a la primera clase de Probabilidad y Estadística.** Comenzaremos a descubrir qué es eso que llamamos Estadística y Probabilidad, aunque existe una aproximación conceptual que retomaremos para direccionar e intentar precisar las definiciones dentro del campo específico para constituirse, de manera paulatina y progresiva, en herramientas útiles para el trabajo que realiza el Técnico en Ciencia de Datos e Ingeniería Artificial. Esta propuesta pretende comenzar a trabajar con los conceptos básicos e introducir algunas precisiones respecto del vocabulario técnico que emplearemos dentro de este campo.

La estadística es una rama de las matemáticas, y es por ello que utilizaremos conocimientos provenientes de este campo disciplinar que desarrollaremos y retomaremos en el transcurso de la cursada. Por supuesto, este espacio está en estrecha vinculación con los espacios que cursaran en este cuatrimestre de la carrera.

Por otra parte, conoceremos un software que nos acompañará a lo largo del cursado de este espacio. La particularidad de esta herramienta, a diferencia de otras que existen y son útiles para el trabajo estadístico, fue desarrollado

por docentes investigadores con un de la Universidad Nacional de Córdoba en nuestro país. InfoStat como herramienta con múltiples usos es ampliamente aceptada por docentes y profesionales de la estadística.

## **2. Desarrollo:**

### **Una aproximación al concepto de estadística.**

Es ampliamente aceptado por el común de las personas que la estadística consiste en un montón de números ordenados en filas o en gráficos que expresan una caracterización respecto de un fenómeno natural o social, sin embargo, la estadística llega a tener un sentido más amplio. La función principal de la estadística es la elaboración de métodos, técnicas y procedimientos que nos permitan tomar decisiones frente a la incertidumbre. Por ejemplo, una empresa que desea invertir en un determinado sector realiza un estudio de mercado para evaluar los riesgos asociados a la inversión para tomar una decisión con determinado nivel de certeza, aunque no siempre se puede estar totalmente seguro de lo que vaya a ocurrir.

La estadística es ampliamente utilizada en los diferentes campos científicos para tomar decisiones basadas en los datos provenientes de la experimentación. Entonces podemos decir que la estadística comienza a constituirse como una herramienta de análisis. Para este curso diremos que *“La estadística es el estudio de los métodos y procedimientos para recoger, clasificar, resumir y analizar datos y para hacer inferencias partiendo de tales datos que nos permitirá tomar decisiones en determinados ámbitos”*.

De esta definición se distinguen dos orientaciones, una que llamaremos Estadística Descriptiva y se encarga de la recolección, organización y

presentación de datos para poner de manifiesto sus características distintivas para elaborar conclusiones sobre ese conjunto de datos; la otra llamada Estadística Inferencial pretende hacer extensiva las conclusiones provenientes de una muestra, a toda la población de estudio, es decir realizar inferencias. Bien, más adelante retomaremos estos dos conceptos que incorporamos para caracterizar a la Estadística Inferencial.

Continuando, para poder operar la estadística se vale de un método científico que se denomina **Método Estadístico**.

### **Etapas del Método Estadístico.**

El método estadístico consiste en *una serie de pasos ordenados para el manejo de datos numéricos y no numéricos en una investigación*. Las etapas que se presentan y desarrollan de manera ordenada son:

1. **Recolección:** también llamada etapa de medición, en ella se recoge la información en consonancia con el método definido por el modelo de investigación puesto los datos pueden asumir diferentes características en relación al fenómeno que se estudia. Algunos métodos más usuales pueden ser la observación directa, la entrevista o la encuesta.
2. **Organización:** en esta etapa, los datos obtenidos en la recolección se someten a revisión para realizar las correcciones pertinentes o eliminar aquellos que resulten imprecisos. En muchos casos, se suelen utilizar algunos programas para el manejo de los datos.
3. **Presentación:** en esta etapa se elabora los cuadros (también llamados tablas) y gráficos que casi siempre se presentan en pares para ofrecer una clara idea de la distribución de los datos. La elaboración de

gráficos estadísticos facilita la inspección visual del conjunto de datos que se está estudiando.

4. **Síntesis:** en etapa se pretende resumir la información en medidas que permitan expresar de manera sintética las principales propiedades numéricas de grandes series o agrupamientos de datos.
5. **Análisis:** en esta etapa, mediante el uso de fórmulas estadísticas y tablas específicamente diseñadas se efectúa la comparación de datos en relación con las medidas de resumen previamente calculadas, por ejemplo, se mide imagen de una empresa antes y después de un proyecto publicitario y se efectúa el análisis para establecer el impacto del proyecto, por supuesto este análisis dentro de un proceso de investigación.
6. **Interpretación:** esta etapa está presente en todo el trabajo estadístico y está ligada a una apreciación de los resultados del trabajo.

Esta breve descripción del método estadístico pretende presentar a grandes rasgos las características de un proceso de investigación. A lo largo del curso trabajaremos sobre las herramientas propias de cada una de las etapas.

### **Conceptos básicos.**

A continuación, presentaremos algunos conceptos básicos que son usuales cuando trabajamos en estadística y nos acompañaran lo largo de todo el espacio. Cabe aclarar que con la explicitación de cada concepto presentaremos una breve explicación para facilitar la comprensión de los mismos. Comencemos:

1. **Individuo, unidad de observación o experimental:** es cualquier elemento que porte información acerca del fenómeno que se estudia.

Por ejemplo, si se realiza un estudio sobre las estaturas de los niños y niñas de una escuela cada niño o niña es una unidad de observación, o si se estudia el nivel de conformidad de los usuarios de un servicio brindado por una empresa telefónica, cada uno de los usuarios sería una unidad de observación; si se estudia la duración de la batería de un teléfono de una determinada marca y modelo la unidad de observación sería cada uno de los teléfonos de dicha marca y modelo.

2. **Población:** es el conjunto total de las unidades de observación (pueden ser personas u objetos de cualquier naturaleza) que porten información sobre el fenómeno de estudio. Por ejemplo, si estudiamos el acceso al estudio de nivel superior en una ciudad, la población serán todos los habitantes de dicha ciudad mayor de 18 años. La muestra se denotará con la letra **N**.
3. **Muestra:** la muestra es una parte representativa de la población seleccionada para el estudio. Por lo general, se recurre al estudio de una muestra cuando estudiar a la población no es viable por cuestiones de accesibilidad, presupuesto, o tamaño, entre otras. Por ejemplo, en el caso del acceso a los estudios de nivel superior resultaría impráctico trabajar con la población de provincia mayor a 18 años, entonces necesitamos recurrir a una muestra que sea representativa (en las unidades siguientes veremos algunos métodos para realizar esta selección). Simbolizamos a la muestra con la letra **n**.
4. **Variable:** es una característica (numérica o no) que varía de un individuo a otro en la muestra o población. En el caso de estudiar el acceso al estudio de nivel superior, la característica que cambia o varía de un individuo a otro sería si estudia o no en el nivel superior; otro



ejemplo podría ser estudiar las edades de los ingresantes a una determinada carrera en un centro de formación superior y en ese caso la variable sería la edad. Como podemos notar las variables pueden ser numéricas o no, detallaremos sus características distintivas:

**4.1. Variable cuantitativa:** son aquellas variables que se representan por medio de un valor numérico, y puede provenir de un proceso de conteo o medición; en el caso que provengan de un proceso de conteo las llamaremos **variables discretas** (por ejemplo, cantidad de mascotas, cantidad de hermanos, cantidad de materia aprobadas, entre otras), mientras que en el caso de provenir de un proceso de medición se llamarán **variables continuas** (por ejemplo, el peso, la talla, el índice de masa corporal, el tiempo de descanso diario, entre otras).

**4.2. Variable cualitativa:** son aquellas características variables que no se pueden representar por medio de un valor numérico, es decir, enuncia una cualidad de la unidad de observación (por ejemplo, el color de un objeto, la opinión respecto de un tema específico, el sexo, el estado civil, entre otras). Estas variables se pueden reclasificar en otras dos que responden a un ordenamiento jerárquico o no, es decir, existen variables cualitativas que no se pueden ordenar de acuerdo a un orden de importancia (por ejemplo: los nombres, la nacionalidad, la preferencia de un deporte, etc.) a las cuales llamaremos **variables cualitativas no ordenables**; mientras aquellas que se pueden ordenar de acuerdo a un rango de importancia (por

ejemplo, los cargos en una empresa, el nivel de instrucción, etc.) se denominan **variables cualitativas ordenables**.

**5. Escala de medición:** para poder llevar adelante el proceso de recolección de información o medición necesitamos establecer un parámetro de medición de acuerdo a la naturaleza de los datos que vamos a estudiar. Para este curso solo nombraremos algunas de las escalas más usuales: **Escala nominal** (sus valores se pueden organizar en categorías no ordenables), **Escala ordinal** (sus valores pueden ordenarse siguiendo una relación de jerarquía), **Escala intervalar** (en esta escala los valores tienen un orden natural y además es posible establecer una diferencia entre dos valores del intervalo, pueden ser continuos o discretos) y **Escala de razón** (tiene las mismas características de una escala de intervalos con la diferencia de que el cero indica la ausencia de la variable, por ejemplo la velocidad cero indicaría la ausencia de movimiento).

**6. Datos:** son los registros de los valores de la variable de estudio y para convertirse en información deben pasar por un proceso, en este caso, estadístico.

Veamos ahora un ejemplo.

Se realiza un estudio para determinar las imperfecciones en los rollos de telas producidas en Tierra del Fuego, para ello, se seleccionan dos fabricas al azar y se lleva adelante el estudio.

Entonces, cuando nos referimos a la **población**, hablamos de todas las fabricas textiles de la provincia pues el enunciado nos indica que se lleva adelante un estudio provincial que involucra a las fabricas textiles. Ahora

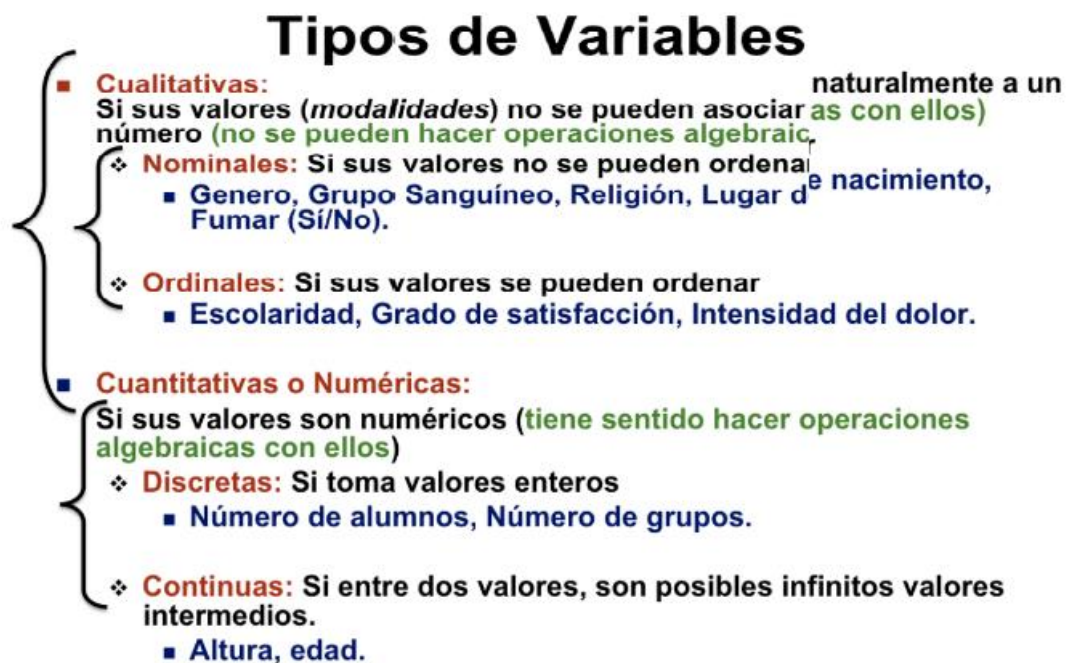
bien, por cuestiones de costos, tiempo, manejo de la cantidad de información, etc., se decide estudiar a solo dos fábricas, lo cual constituye a la muestra para la situación que estamos analizando, es decir, la **muestra** están compuestas por esas dos fabricas seleccionadas. Por otra parte, para poder identificar cual es la variable, debemos hacernos la siguiente pregunta: ¿qué es lo que cambia entre una fábrica y otra? La respuesta es evidente, la cantidad de rollos con imperfecciones y es lo que constituye nuestra **variable** de estudio. La **unidad de análisis** será aquel objeto o persona quien nos brinde información acerca de la variable que para nuestro caso es cada una de las fábricas.

Ahora bien, como mencionamos anteriormente, hay distintos tipos de variables, para poder identificarlas debemos preguntar, nuestra variable se responde con un número, y si la respuesta es sí, se trata de una variable cuantitativa, caso contrario es cualitativa, en la situación que estamos trabajando la variable corresponde a un valor numérico, por lo tanto, la variable es cuantitativa. Luego, sabiendo que la variable es cuantitativa, debemos preguntar si este valor proviene de un proceso de medición o de conteo, o sea, para tomar ese valor, ¿se debe contar o medir?. Si es producto del conteo entonces es discreta y si proviene de un proceso de medición entonces, corresponde a una variable continua, en nuestro caso particular, corresponde a una variable discreta pues para saber la cantidad de rollos con imperfecciones que produce cada fábrica debemos simplemente contarlos.

Cabe aclarar que para el caso de las variables cualitativas, las preguntas que hacemos refieren a si es ordenable o no.



Te invito mirar el siguiente cuadro que te servirá como guía para clasificar las variables, sería conveniente realizar un cuadro con las preguntas que nos permiten realizar la clasificación.



### Actividad 1.

En consonancia con el trabajo que venimos desarrollando, se propone como actividad para esta clase clasificar las variables que se presentan a continuación, identificando también las escalas de medición que correspondan:

- a) Edad gestacional de la madre.
- b) Salud dental (dientes careados, perdidos, obturados)
- c) Agudeza visual (normal, disminuida, perdida).
- d) Estado de nutrición.

- e) Duración de la lactancia en días.
- f) Nivel de escolaridad materna.
- g) Cobertura social (si – no)
- h) Coeficiente intelectual.
- I) Temperatura corporal.
- j) Tensión arterial
- k) Concentración de Azufre en el crudo de petróleo.
- l) Número de operaciones

### **Organización y presentación de datos.**

Cuando en estadística hablamos de organización y representación de datos hacemos referencia a los cuadros, tablas y gráficos estadísticos. Entonces, los datos pueden ser organizados según algún criterio de referencia y presentarlos en forma bruta, arreglo ordenado, exhibición de tallo y hoja, y distribuir de frecuencias.

Una vez organizados los datos, estos deben ser presentados. La información puede ser presentada en forma textual, en cuadros o tablas, y mediante gráficos.

**Textual.** Los datos se presentan en forma literal. Entre las ventajas de esta presentación podemos encontrar que los datos de importancia se pueden resaltar mediante subrayados, palabras o cifras en negrita, palabras en

cursiva o entre comillas. Entre las desventajas podemos nombrar que solo se puede utilizar esta forma cuando los datos a presentar son pocos o reducidos.

**Tabular.** Cuando los datos a presentar son abundantes, estos se presentan mediante cuadros estadísticos. Como ventaja encontramos que los cuadros son más breves, claros y fáciles de leer. Además, facilita las comparaciones.

**Cuadro general**

| Provincias del país que conforman distintas regiones   |   |           |   |
|--|---|-----------|---|
| Región   | Provincias  | Región    | Provincias  |
| Metropolitana  | Ciudad de Buenos Aires<br>24 Partidos del Gran Buenos Aires               | Pampeana  | Córdoba<br>Entre Ríos<br>La Pampa<br>Santa Fe<br>Resto de la Pcia. de Bs. Aires                       |
| Cuyo   | Mendoza<br>San Juan<br>San Luis   | Nordeste  | Chaco<br>Corrientes<br>Formosa<br>Misiones  |
| Noroeste   | Catamarca<br>Jujuy<br>La Rioja<br>Salta<br>Santiago del Estero<br>Tucumán | Patagonia | Chubut<br>Neuquén<br>Rio Negro<br>Santa Cruz<br>Tierra del Fuego, Antártida e Islas del Atlántico Sur |
| Fuente: Proyecto Apoyo al último año del nivel medio / polimodal para la articulación con el nivel superior. Resolución de problemas. Cuaderno de trabajo para los alumnos de matemática distribuido por el Ministerio de Educación, Ciencia y Tecnología. |   |           |   |

Un cuadro como el que muestra en la imagen consta de las siguientes partes:

**Título:** debe dar una descripción del contenido del cuadro. Además, debe responder a las preguntas: ¿Qué datos son los incluidos en el cuerpo de la tabla? ¿Dónde está el área representada por los datos? ¿Cómo están clasificados los datos? ¿Cuándo ocurrieron los datos?

**Encabezado:** es el título de la parte superior de las columnas. Una tabla puede tener sub-encabezado.

**Cuerpo o columna matriz:** son las descripciones en hileras colocadas en el lado izquierdo de la tabla. Representan las clasificaciones de las cifras incluidas en el cuerpo de la tabla. Pueden ser divididos en subconceptos.

**Cuerpo:** es el contenido de los datos estadísticos.

**Nota de encabezado:** es usado para expresar ciertos puntos relacionados con el cuadro total y que no han sido incluidos en el título ni en el encabezado ni en los conceptos.

**Nota de pie:** se utiliza para clasificar algunas partes incluidas en el cuadro que no son explicados en otras partes.

**Fuente:** debe ser declarada al pie del cuadro. Es el dato en el cual se informa de donde o a quien corresponde la información.

**Gráficos.** Los datos se presentan en gráficos circulares, de barras, de líneas, etc., esto dependerá de lo que se quiera informar. Las partes principales de una gráfica son:

**Título:** es la descripción del contenido de la gráfica. Puede encontrarse en la parte superior o en la parte inferior.

**Diagrama:** representa, al igual que en un cuadro, los datos mostrados en la gráfica.

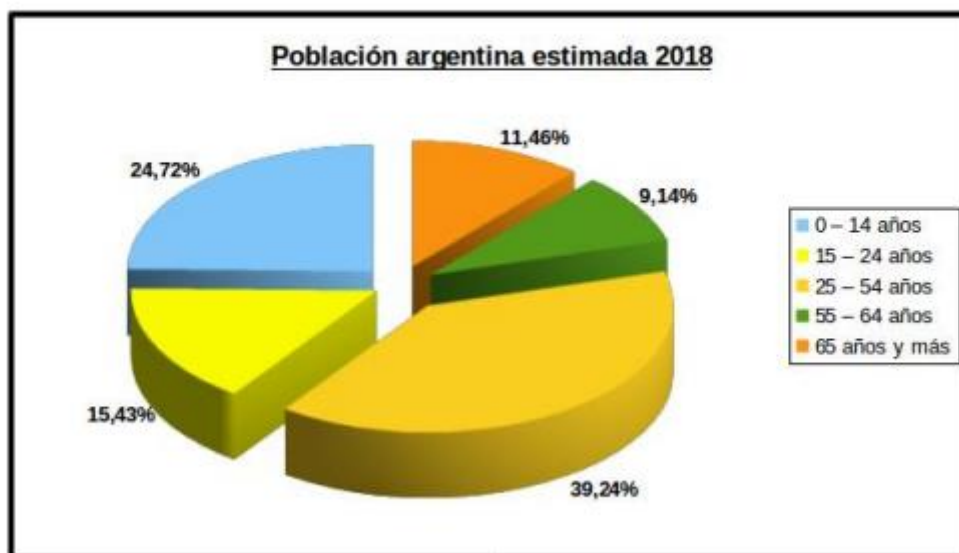
**Escala:** frecuentemente se coloca en el eje de las abscisas la clasificación y en el eje de las ordenadas las magnitudes.

**Fuente:** debe ser colocada al pie de la gráfica e indicar cual fue el origen de los datos a partir de los cuales la gráfica fue construida.

Actividad 2:

Indicá en cuales son las partes de los siguientes gráficos:

### Gráfico de sectores circulares



Para no extendernos mucho en la escritura de esta clase, les dejo un link para revisar cuestiones sobre los cuadros y gráficos. Además, les dejo un pequeño video tutorial sobre InfoStat para que puedan descargarlos en sus computadoras y comenzar a familiarizarse con él.

Tipos de gráficos estadísticos: <https://youtu.be/rJPYV7V7ssc>

Instalación de InfoStat. <https://youtu.be/8CStN1tijTQ>

Introducción a InfoStat: <https://youtu.be/OwNXPgA4r0s>

### 3. Actividad integradora:

Como actividad integradora te proponemos que respondas el cuestionario de autoevaluación en el que se incluirán los conceptos trabajados en esta clase.



#### 4. Cierre

Entonces, en esta clase hicimos un recorrido desde una definición formal de estadística que involucra aquellos aspectos del conocimiento común acerca de la estadística para precisar, avanzado sobre una descripción de las etapas del método estadístico, como así también el trabajo sobre los conceptos básicos que nos acompañaran a lo largo del cursado de este espacio y que además conforman el vocabulario específico dentro de la estadística. También, avanzamos sobre el estudio de los aspectos y características principales acerca de la organización y presentación de datos.

Por último, la propuesta de actividades en relación al desarrollo de esta clase pretende brindar un espacio para la elaboración de material estudio personal.

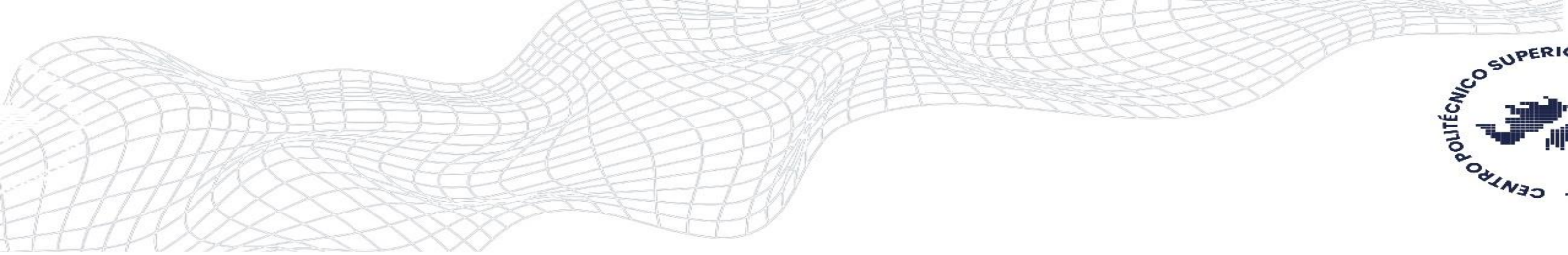
Aprovecho este espacio para saludarlos y alentarlos para el desarrollo del trabajo que se propone para esta clase. Nos seguimos encontrando en el aula. Saludos cordiales.

#### 5. Bibliografía:

García, J; López, N; Calvo, J. (2011). Estadísticas Básicas para Estudiantes de Ciencias. Facultad de Ciencias Físicas Universidad Complutense de Madrid. España.

Wackerly, D; Mendenhall, W; Schaeffer, R. (2010). Estadística Matemática con Aplicaciones. 7<sup>ma</sup> Ed. Cengage Learning. Santa Fe, México.

|  |                |
|--|----------------|
| - Probabilidad y Estadística                         | Versión: 0.0.1 |
| - Probabilidad y Estadística – Clase 2               | Fecha:         |
| Elaborado por: Equipo de Probabilidad y Estadística. | Estado:        |



|                               |                               |
|-------------------------------|-------------------------------|
| <b>Revisado por:</b><br>..... | <b>Aprobado por:</b><br>..... |
|-------------------------------|-------------------------------|