



INTRODUCCIÓN A LA ESTADÍSTICA

Luis Nava P., Douglas Rivas O., Victor Márquez P.

INTRODUCCIÓN A LA ESTADÍSTICA.

Luis Nava P., Douglas Rivas O. y Victor Márquez P.
Escuela de Estadística
Universidad de Los Andes

MÉRIDA - VENEZUELA
2010

Agradecimiento

Nuestros profundos agradecimiento a Dios primeramente, a nuestras esposas, Naelys, Marienella y Ana por su apoyo incondicional, a los profesores Franklin Leal y Hanzel Larez, por su valioso aporte en algunos consejos sobre latex, a nuestros compañeros de trabajo y a nuestros estudiantes que influyeron de una manera determinante en la realización de este trabajo.

1. Aspectos Básicos	1
1.1. Introducción.	1
1.2. Importancia.	1
1.3. Conceptos Básicos	2
1.4. Clasificación de la Estadística	10
1.5. Ejercicios	10
2. Estadística Descriptiva	13
2.1. Introducción.	13
2.2. Organización de los Datos	14
2.2.1. Organización de Datos Cualitativos	14
2.2.2. Organización de Datos Cuantitativos	16
2.3. Presentación de los Datos	26
2.3.1. Representación Tabular	26
2.3.2. Representación Gráfica	27
2.4. Medidas Descriptivas Numéricas	36
2.4.1. Medidas de Localización	37
2.4.2. Medidas de Dispersión.	52
2.4.3. Medidas de Forma	56
2.5. Ejercicios	60
3. Probabilidades	65
3.1. Introducción.	65
3.2. Conceptos Básicos	66
3.2.1. Teoría de Conjuntos y Técnicas de Conteo	66
3.2.2. Experimentos Aleatorios	71

3.3.	Probabilidades	74
3.4.	Probabilidad Condicional	78
3.4.1.	Introducción.	78
3.4.2.	Probabilidad Condicional	79
3.4.3.	Teoremas de Probabilidad Condicional	81
3.5.	Teorema de Probabilidad Total	84
3.6.	Regla de Bayes	85
3.7.	Eventos Independientes	85
3.8.	Ejercicios	87
4.	Variable Aleatoria	91
4.1.	Introducción	91
4.2.	Clasificación de las Variables Aleatorias	93
4.3.	Distribución de Probabilidad	93
4.3.1.	Distribución de Probabilidad de una Variable Aleatoria Discreta	93
4.4.	Valor Esperado y Varianza de una Variable Aleatoria Discreta	97
4.4.1.	Valor Esperado de una Variable Aleatoria Discreta	97
4.4.2.	Valor Esperado de una Función de una Variable Aleatoria Discreta	98
4.4.3.	Varianza de una Variable Aleatoria Discreta	100
4.5.	Distribución de Probabilidad Multivariable	101
4.5.1.	Introducción	101
4.5.2.	Distribución de Probabilidad Conjunta	101
4.5.3.	Distribuciones de Probabilidad Marginal y Condicional	104
4.5.4.	Variables Aleatorias Independientes	106
4.5.5.	Valor esperado de una función de 2 variables aleatorias discretas	107
4.5.6.	Covarianza de dos variables aleatorias	110
4.5.7.	Ejercicios	113
5.	Distribuciones de Probabilidad Discretas	117
5.1.	Introducción	117
5.2.	Distribución Uniforme Discreta	117
5.3.	Distribución Bernoulli	119
5.4.	Distribución Binomial	121
5.5.	Distribución Hipergeométrica	127
5.6.	Distribución Poisson	129
5.7.	Aproximación de la distribución Binomial usando la distribución Poisson	133
5.8.	Distribución Binomial Negativa	134
5.9.	Ejercicios	136

6. Distribuciones de Probabilidad Continuas	141
6.1. Introducción	141
6.2. Distribución Normal	141
6.3. Aproximación de la distribución binomial a la normal	146
6.4. Aproximación de la distribución poisson a la normal	148
6.5. Distribuciones derivadas de la normal	149
6.5.1. Distribución Chi-Cuadrado	149
6.6. Ejercicios	155
7. Distribuciones en el muestreo	159
7.1. Introducción	159
7.2. Muestreo y tipos muestreo	160
7.2.1. Métodos para seleccionar una muestra aleatoria.	162
7.3. Distribución muestral de la media muestral	166
7.4. Distribución muestral de la diferencia de medias muestrales	169
7.5. Distribución muestral de varianzas muestrales	171
7.6. Distribución muestral de la proporción muestral	173
7.7. Distribución muestral de la diferencia de proporciones muestrales	174
7.8. Ejercicios	176
Bibliografía	181

Llegará el día en el que la Estadística será una condición tan necesaria para la convivencia como la capacidad de leer y escribir. (Anónimo.)

Esta obra representa la primera de una serie de textos que tienen como objetivo ofrecer un material de apoyo a los estudiantes de métodos estadísticos y que además, sea útil a toda aquella persona que tenga algún interés por la estadística. El mismo es el resultado de una extensa revisión bibliográfica y de la preparación de los cursos de estadística básica durante varios semestres.

En el capítulo 1 se presenta los conceptos básicos de estadística, destacando la importancia de la misma en las Ciencias. de la estadística descriptiva.

En el capítulo 2 se estudian los procedimientos relacionados con estadística descriptiva, ofreciendo al estudiante mediante una amplia variedad de ejemplos, el como y cuando aplicar un procedimiento en particular.

En los capítulos 3, 4, 5, 6 y 7 se hace una exposición de temas de gran importancia en inferencia estadística: probabilidad, variable aleatoria, distribuciones de probabilidad (normal, t de student, Chi-cuadrado y F de snedecor) y distribuciones muestrales.

Luis Nava P.

Douglas Rivas O.

Víctor Márquez P.

CAPÍTULO 1

ASPECTOS BÁSICOS

1.1. Introducción.

Cada vez resulta más difícil tomar decisiones apropiadas sobre situaciones que inciden en forma significativa por ejemplo en el éxito de un producto o quizás, en nuestro bienestar. La estadística es esencial en la investigación de tales situaciones pues permite caracterizar y/o explicar su comportamiento. Es una disciplina de aplicabilidad en cualquier campo de estudio. Este capítulo tiene como propósito establecer el marco de referencia para el estudio de la estadística. En el mismo se destacará la importancia y campo de acción de esta y se introducen algunas definiciones básicas.

1.2. Importancia.

En el mundo real dos tipos de fenómenos pueden identificarse; fenómenos determinísticos y fenómenos aleatorios. Los fenómenos determinísticos son aquellos que bajo las mismas condiciones producen siempre los mismos resultados, mientras que los aleatorios se refiere a aquellas situaciones donde está presente la incertidumbre en cuanto al resultado del mismo. Durante mucho tiempo se considero que el campo propio de la investigación científica era exclusivo de fenómenos determinísticos. Sin embargo, los fenómenos aleatorios, es decir, aquellos fenómenos en los cuales no se puede predecir el resultado de cada experiencia particular, son de gran importancia y la información derivada de los mismos, es objeto de estudio por parte de la Estadística.

Ejemplo 1.1 : *Son ejemplos de fenómenos aleatorios:*

1. Lanzamiento de un dado.
2. Cantidades vendidas de artículos en un Supermercado en días sucesivos.

3. La duración de los equipos eléctricos en un lote producido por determinada empresa.
4. Precio del barril de petróleo.
5. Nivel de aceptación de un determinado candidato a Gobernador.
6. Proporción de artículos defectuosos en una línea de producción.

En el estudio de fenómenos como los mencionados arriba, la Estadística constituye una herramienta valiosa, resultando de fundamental importancia en los procesos de toma de decisión. La estadística es de gran utilidad para evaluar por ejemplo, la posibilidad de que un producto nuevo tenga éxito. Igualmente, si se desea tomar la decisión acerca de que tipo de maquinaria adquirir, los métodos estadísticos resultan indispensables para tal fin.

1.3. Conceptos Básicos

Cuando se habla de Estadística, tradicionalmente se hace referencia a números presentados ordenada y sistemáticamente. Esta idea es consecuencia del concepto popular que existe sobre esta ciencia y que cada vez se extiende más debido a la influencia de nuestro entorno. Por ejemplo, es muy común escuchar información denominada estadística de accidentes, y sólo se hace referencia al número de accidentes. Sin embargo cuando profundizamos en el campo de la investigación podemos entender que la estadística no solo son números, sino que representa la principal y a veces, la única herramienta para la obtención y análisis de la información en cualquier tipo de investigación, y cuyo comportamiento no puede ser abordado desde el punto de vista determinísticos. Se puede decir entonces, que la estadística es la ciencia que establece como obtener y usar la información referente a una investigación y como actuar en situaciones practicas donde esta presente la incertidumbre.

Definición 1.1 (Estadística) *Es la ciencia de recolectar, ordenar, presentar y describir la información asociada a un fenómeno aleatorio, con el objeto de caracterizarlo y deducir la ley que rige dicho fenómeno y así poder obtener conclusiones o, tomar decisiones.*

Ejemplo 1.2 *Las siguientes son situaciones en las que la Estadística juega un papel importante:*

1. Para el calculo del Índice de Precios al Consumidor (IPC), el organismo encargado aplica una encuesta que permite **recolectar** información sobre el consumo de las familias venezolanas. Luego esta información es **resumida** en un índice que representa el consumo familiar promedio ponderado.
2. Para **estimar** la inflación los organismos encargados usan el IPC.
3. La **predicción** o **pronóstico** de ventas por parte de una empresa.
4. Para determinar si las políticas económicas han sido efectivas, el gobierno quiere **contrastar** la **hipótesis** de que la tasa de inflación anual no supera el 15%.

5. Una empresa estudia el histórico de sus ganancias para **tomar decisiones** respecto a la aplicación de nuevas estrategias de mercado.

Para el estudio de fenómenos como los enumerados arriba, necesitamos contar con información relacionada con el mismo. Esta información, obtenida bien sea experimentalmente o, mediante la observación, esta dada por datos. Estos datos son el resultado de medir en un conjunto de elementos o individuos, una o varias características de interés y que son objeto de análisis en una investigación. De esta forma, en la estadística interesa las características y no los elementos o individuos en si.

Definición 1.2 (Elemento) *Es un ser vivo, objeto o cosa que posee características que se desean investigar.*

Ejemplo 1.3 *Son ejemplos de elementos y sus características:*

1. Elemento: Persona. Característica: Sueldos.
2. Elemento: Persona. Característica: Edad y sexo.
3. Elemento: Hogar. Característica: Consumo e ingresos.
4. Elemento: Hogar. Característica: Número de miembros, oficio y sexo del jefe.
5. Elemento: Supermercado. Característica: ventas.
6. Elemento: Fabrica. Característica: Proporción de artículos defectuosos por lote de producción.

Definición 1.3 (Universo Estadístico) *Se denomina universo estadístico a cualquier conjunto finito o infinito de individuos o elementos, sobre los cuales están definidas características comunes y que pueden ser objeto de análisis.*

Ejemplo 1.4 *Son ejemplos de universo estadístico:*

1. Los Habitantes de la ciudad de Mérida.
2. Los estudiantes de la Facultad de Ciencias Económicas y Sociales.
3. Los trabajadores de una empresa.
4. Los animales en un bosque.
5. Los carros que entran en un estacionamiento al día.
6. La producción de envases plásticos durante un periodo de tiempo (día, semana).

Como se dijo antes, los elementos del universo estadístico tienen una serie de características que pueden ser objeto del estudio estadístico. Así por ejemplo, si el elemento considerado es una persona, podemos distinguir en ella las siguientes características: Sexo, Edad, Nivel de estudios, Profesión, Peso, Estatura, Color del cabello, etc. Obsérvese que todos y cada uno de los elementos de la población, poseen las características de interés. Por lo tanto, de cada elemento del universo se puede estudiar uno o más aspectos, cualidades o caracteres.

El universo estadístico puede clasificarse de acuerdo al número de elementos que contiene en finito e infinito. Es finito cuando el número de elementos que lo conforman es finito (el número de alumnos de un centro de enseñanza, los niños de una comunidad, el número de plantas en un vivero). Es infinito si el número de elementos que lo conforman es infinito, o tan grande que pudiese considerarse infinito (todas los individuos con estaturas entre 1 y 2 metros).

Hay situaciones en las que aún cuando el universo es finito, contiene tantos elementos que el mismo puede considerarse infinito. Por ejemplo, si se realizase un estudio sobre los productos que hay en el mercado, hay un número finito de los mismos, pero son tantos y de tantas calidades que este universo podría considerarse infinito.

En un estudio estadístico, el investigador está interesado en la colección o conjunto de observaciones asociadas con una o varias características de los elementos del Universo estadístico. A este conjunto de observaciones se le denomina población estadística.

Definición 1.4 (Población) *Es el conjunto de todas las posibles mediciones que pueden hacerse de una o varias características comunes a todos los elementos del universo y que son de interés en una investigación dada.*

Ejemplo 1.5 *Son ejemplos de población estadística:*

1. El sexo de los habitantes de la ciudad de Mérida
2. La edad y el sexo de los estudiantes de la Facultad de Ciencias Económicas y Sociales.
3. El sueldo, el nivel instrucción y el cargo que ocupan los trabajadores de una empresa.
4. El color de ojos de los animales en un bosque.
5. La marca de los carros que entran en un estacionamiento al día.

De esta forma, puede observarse que la población estadística está constituida por valores o datos bien sea numéricos o no.

Se puede notar que sobre un mismo Universo Estadístico se pueden definir una o varias poblaciones. Además una población puede estar conformada por las mediciones asociadas con una o varias características. Así, una población que considera sólo valores asociados con una característica se denomina población univariante, en caso contrario, se denomina población multivariante.

Al igual que el universo, la población puede ser finita o infinita, dependiendo del número de valores que la constituyen. En el caso de que la población sea finita, se dice que esta tiene tamaño N .

Ahora bien, cuando se desea investigar un fenómeno, una interrogante a responder es la relacionada con la cantidad de información a usar en dicha investigación. Si se usa toda la información disponible, es decir la población, se dice que se ha realizado una investigación exhaustiva o total (Censo). Por lo general el desarrollo de este tipo de investigaciones no siempre es posible, por diferentes motivos: a) economía; b) pruebas destructivas; c) población infinita o tan grande que exceda las posibilidades del investigador. En su lugar, se plantea una investigación parcial o por muestreo. Esta investigación se basa en la información correspondiente a una porción de la población, denominada muestra. Esta muestra debe ser generalmente suficiente para evaluar o estudiar toda la población y el método que se use para su selección es determinante en el uso que pueda hacerse de la misma.

Definición 1.5 (Muestra) *Es una parte o porción extraída de una población, mediante métodos de selección científicos que permiten considerarla preferiblemente representativa de dicha población.*

Ejemplo 1.6 : *Son ejemplos de muestras:*

1. Un grupo de 200 personas seleccionadas de todos los habitantes de la ciudad de Mérida.
2. 5 estudiantes elegidos de un grupo de 20 alumnos de estadística II.
3. Un grupo de 20 Profesores seleccionados de la planta profesoral de la Universidad de Los Andes.
4. Una selección de 25 trabajadores de la nómina de una gran empresa.
5. Los vehículos que entran a un estacionamiento en una hora.

Las poblaciones tienen características que las describen numéricamente. El ingreso promedio de los habitantes del Estado Mérida, la proporción de hombres blancos en una comunidad, la variabilidad en la estatura de los alumnos de una institución educativa, son algunos ejemplos. A estos valores o medidas se les denomina parámetros.

Definición 1.6 (Parámetro) *Representa cualquier característica medible de la población. Es una función de los valores de la población que sirve para sintetizar alguna característica relevante de la misma. Ejemplos de parámetros son: La media poblacional, La proporción poblacional, la varianza poblacional, entre otros.*

Una población puede estar caracterizada por uno o más parámetros. Para conocer el valor de los parámetros de una población, es necesario usar toda la información disponible, es decir, es necesario el desarrollo de una investigación exhaustiva, lo que resulta por lo general inconveniente o imposible. Sin embargo, se pueden establecer aproximaciones a dichos valores a través de funciones de los elementos en la muestra.

Definición 1.7 (Estadístico) *Dada una determinada población y una muestra de tamaño n , se define como estadístico a cualquier función que se defina a partir de los n componentes que integran dicha muestra.*

De la definición 1.7 se puede resaltar el hecho de que cualquier función de la muestra, sin condición alguna, es un estadístico. De todas estas funciones, un subconjunto de ellas es de especial interés en estadística; los **estimadores**.

Definición 1.8 (Estimador) *Estadístico utilizado para estimar un parámetro de una población. Es una función de los valores de la muestra que sirve para sintetizar alguna característica relevante de la misma.*

Nótese que un estimador es un estadístico que permite la estimación del parámetro. Esto significa que no todo estadístico es un estimador pero, todo estimador es un estadístico. La media muestral, La proporción muestral, la varianza muestral, entre otros, son ejemplos de estimadores.

Como se ha dicho anteriormente, la estadística se encarga del estudio de un fenómeno a través del manejo de la información que se tiene sobre una o más características del mismo. En el lenguaje estadístico al igual que en el matemático a las características se les conocen como variables y a las distintas formas en que pueden presentarse, modalidades o valores de las variables.

Definición 1.9 (Variable) *Se refiere a una característica que poseen todos los elementos del universo y que se mide en el estudio. Representa las mediciones de una característica que poseen los elementos del universo.*

Ejemplo 1.7 *Son ejemplos de variables:*

- El color de ojos de las personas.
- La edad de las personas.
- El sueldo de un empleado.
- La raza de los perros.
- La nota de los alumnos de Métodos Estadísticos I.

Una variable puede tomar distintas modalidades o valores. Al conjunto de las modalidades o valores de una variable se le denomina **Escala de Medida**. Las Escalas de Medida pueden clasificarse de acuerdo a las relaciones que existen entre los valores y las operaciones aritméticas que pueden realizarse entre las mismas en: nominal, ordinal, de intervalos, de razón y absolutas.

1. Escala Nominal: Son aquellas en que la única relación que se define entre sus valores es la igualdad o diferencia, es decir solo podemos decir que dos valores de una variable son iguales o diferentes. No hay operaciones aritméticas definidas, por lo tanto, los números no tienen sentido como magnitudes.

Ejemplo 1.8 *Son ejemplos de medidas de escala nominal:*

- El grupo sanguíneo.
 - El sexo.
 - El color de ojos.
 - El estado civil.
 - Los números que llevan los atletas en la espalda
2. Escala Ordinal: Son aquellas en las que entre sus valores están definidas las relaciones de igualdad, diferencia, mayor que o menor que, es decir solo podemos decir que dos valores de una variable son iguales, diferentes y en el caso de que sean diferentes se puede establecer un orden entre ellos. No hay operaciones aritméticas definidas.

Ejemplo 1.9 *Las siguientes medidas corresponden a una escala ordinal:*

- Dureza de los minerales.
 - Grado de satisfacción.
 - Intensidad de un dolor.
 - Rango militar.
 - Nivel de educación.
3. Escala de Intervalo: Los valores de las variables son números y entre ellos tienen sentido las relaciones de igualdad, de orden y de las distancias. La resta es la única operación aritmética definida.

Esta escala posee una propiedad de gran importancia; Posee un cero u origen relativo (arbitrario). El cero no significa ausencia de la característica.

Ejemplo 1.10 *Los siguientes son ejemplos de medidas pertenecientes a una escala de intervalo:*

- Puntuación obtenida en una evaluación.
- La temperatura.
- La distancia sobre el nivel del mar.

4. Escala de Razón o Escala proporcional: Los valores de la variable son números y entre ellos tienen sentido las relaciones de igualdad, orden y están definidas las operaciones aritméticas de suma, diferencia y proporciones (múltiplos). Estas escalas tienen un cero absoluto, el cual representa la ausencia de la característica. Es una escala de intervalo con un cero absoluto.

Ejemplo 1.11 *Las siguientes medidas pertenecen a una escala de razón:*

- El Sueldo de los habitantes del Estado Mérida.
- La edad de los alumnos de Métodos Estadísticos I.
- El nivel de hemoglobina.
- Presión arterial.
- Peso en gramos o en kilos.

5. Escala Absoluta: Los valores que puede tomar la variable son el resultado de un conteo, por lo tanto, esta escala está constituida por todos los número enteros positivos y el cero.

Ejemplo 1.12 *Son ejemplos de mediciones pertenecientes a una escala absoluta:*

- Número de accidentes automovilísticos el fin de semana.
- Número de integrantes de una familia.
- Número de alumnos en un salón de clase.

Tipos de Variables

En una investigación, se puede clasificar las variables según la escala de medición o la influencia que se asigne a unas variables sobre otras. De acuerdo a su escala de medida las variables se pueden clasificar en cualitativas y cuantitativas.

Definición 1.10 (Variable Cualitativa) *Son aquellas cuya escala de medida es nominal u ordinal, es decir, una variable es cualitativa si sus valores representan una cualidad o atributo del elemento en estudio.*

Según el número de categorías o modalidades, pueden ser dicotómicas cuando sólo pueden tomar dos valores posibles como hombre y mujer o, politómicas cuando pueden tomar tres o más valores como leve, moderado y grave.

Ejemplo 1.13 *Las siguientes variables son de tipo cualitativo:*

- El sexo de las personas.
- El Tipo de sangre.

- La nacionalidad.
- Lugar de residencia.
- El color de los ojos.
- Escalafón del profesor universitario.
- Grados militares.

Definición 1.11 (Variable Cuantitativa) *Hablamos de variables cuantitativas cuando los valores posibles son cantidades numéricas con las que podemos hacer operaciones aritméticas. Es decir, son aquellos cuya escala de medidas es de intervalos, de proporción o absoluta.*

Las variables cuantitativas pueden ser discretas o continuas. Si la variable presenta separaciones o interrupciones en la escala de valores que puede tomar se dice que es discreta. Si la variable puede tomar cualquier valor dentro de un intervalo especificado de valores se dice que es continua.

Definición 1.12 (Variables Cuantitativas Discretas) *Generalmente son el resultado de un conteo o enumeración. Son aquellas que pueden tomar un número finito o numerable de valores, es decir, se puede establecer correspondencia entre los valores que puede tomar la variable y el conjunto de los números enteros. Son variables que no admiten valor alguno entre dos valores consecutivos de las mismas.*

Ejemplo 1.14 *Son ejemplos de variables discretas:*

- Número de integrantes de una familia.
- Número de artículos defectuosos.
- Número de pares de zapatos que compran las mujeres al mes.
- Número de personas a favor de un candidato.

Definición 1.13 (Variables Cuantitativas Continuas) *Son aquellas formadas por un conjunto no numerable de puntos, es decir, se puede establecer correspondencia entre los valores que puede tomar la variable y el conjunto de los números reales. Para dos valores cualesquiera, siempre hay un tercer valor entre ellos.*

Ejemplo 1.15 *Las siguientes son variables continuas:*

- La edad.
- El peso.
- La temperatura.
- La distancia.

1.4. Clasificación de la Estadística

La Estadística puede clasificarse de acuerdo a su función en el tratamiento de los datos en estadística descriptiva y estadística inferencial.

Definición 1.14 (Estadística Descriptiva) *Denominada también Estadística Deductiva. Es la encargada de describir, analizar y representar un conjunto de datos, utilizando métodos numéricos, tablas y gráficos que resumen y presentan la información contenida en ellos. Puede llevarse a cabo sobre una muestra o sobre toda una población.*

La estadística descriptiva se emplea simplemente para resumir de forma numérica o gráfica un conjunto de datos. Si se aplican las herramientas ofrecidas por la estadística descriptiva se logra describir los datos que se analizan.

Definición 1.15 (Estadística Inferencial) *Denominada también Inferencia Estadística o Estadística Inductiva. Es la que apoyándose en la Teoría de Probabilidades y la Teoría del Muestreo, se encarga de efectuar estimaciones, permitir la toma de decisiones, predicciones u otras generalizaciones sobre una población partiendo del estudio de una muestra.*

La estadística inferencial permite obtener conclusiones o hacer inferencias, basándose en los datos simplificados y analizados de una muestra hacia la población.

La estadística descriptiva e inductiva pueden ser usadas separadas o conjuntamente, dependiendo de los objetivos y tipo de investigación. Lo usual es que en una investigación parcial participen las dos.

1.5. Ejercicios

1. Se realiza un estudio en el municipio Libertador del Estado Mérida sobre el tipo de transporte utilizado por sus residentes, para lo cual se encuesta a un grupo de ellos, obteniéndose

Tipo de Transporte	N° de Residentes
Particular	45
Taxi	25
Trolebús	50
Bus	60
Otros	10

Identifique:

- a) Universo

-
- b)* Población
c) Muestra
d) Variable y tipo de variable.
e) Tipo de escala.
2. Un fabricante produce tornillos para los cuales existen estrechos márgenes de tolerancia en sus diámetros. El departamento de Control de Calidad selecciona la producción de un día y la somete a proceso de control. Identifique:
- a)* Universo
b) Población
c) Muestra
d) Variable y tipo de variable.
e) Tipo de escala.
3. De un lote de 1000 piezas defectuosas se toman al azar 150 de ellas encontrándose con 1,2,3 ó 4 y más defectos, 15, 52, 46 y 37 piezas respectivamente.
Identifique:
- a)* Universo
b) Población
c) Muestra
d) Variable y tipo de variable.
e) Tipo de escala.
4. Identifique el tipo de variable en cada uno de los siguientes casos:
- a)* La resistencia a la ruptura de un determinado tipo de cuerda.
b) El color del cabello de los niños que estén viendo por televisión una película.
c) El número de señales de tránsito en poblados con menos de 500 habitantes.
d) Si una llave de lavamanos esta defectuosa o no.
e) El número de preguntas contestadas correctamente en un examen.
f) El tiempo que se necesita para contestar una llamada telefónica en un a oficina de bienes raíces.
g) El resultado de la encuesta hecha a un grupo de votantes posibles acerca del candidato de su preferencia.
h) El gasto en que incurre una empresa al mes en el pago de la nomina.

- i)* El número de empleados del sexo femenino que hay en una empresa.
 - j)* El precio de un producto en el mercado.
5. Para cada uno de los ítem del ejercicio 5, identifique el tipo de escala más adecuada para realizar la medición.

2.1. Introducción.

Aun cuando en la actualidad la mayor parte del uso de la estadística esta dirigido a la Inferencia, la Estadística descriptiva tiene una utilidad importante fundamentalmente en la primera fase de una investigación.

La estadística descriptiva se refiere al proceso en el que los datos son ordenados, resumidos y clasificados con objeto de tener una visión más precisa y conjunta de las observaciones, intentando descubrir posibles relaciones entre los datos, observando similitudes y diferencias entre los mismos, destacando hechos de posible interés, entre otras cosas. Esto es, tiene como objetivo caracterizar, describir y extraer conclusiones sobre los datos de forma tal que permitan sugerir cuestiones a analizar con mayor profundidad, llegando en ocasiones a ayudar en el establecimiento de las primeras hipótesis acerca de la naturaleza del fenómeno que se estudia o investiga.

La Estadística Descriptiva además permite estudiar si pueden mantenerse algunos supuestos necesarios para procesos de inferencia, tales como la de simetría, normalidad, homocedasticidad, etc.

Este tema tiene como propósito principal, introducir técnicas que permitan tanto matemática como gráficamente describir apropiadamente un conjunto de datos y al finalizar el mismo, el estudiante debe estar en capacidad, una vez coleccionados los datos, de:

- Ordenarlos y clasificarlos,
- presentarlos a través de cuadros estadísticos y gráficos,
- calcular medidas descriptivas numéricas y
- analizar la información obtenida en los pasos anteriores.

2.2. Organización de los Datos

La organización de los datos consiste en una agrupación apropiada de los mismos. Es importante dicha agrupación, ya que por lo general la información obtenida de un estudio implica gran cantidad de datos que no es fácil interpretar directamente. Esta organización depende del tipo de variable que se maneje. Por lo tanto, se debe estudiar como realizar dicha agrupación cuando la variable es cualitativa y cuando es cuantitativa.

Los datos se organizan en una distribución de frecuencias, la cual es una tabla resumen en la que los datos se disponen en agrupamientos o categorías convenientemente establecidas de clases ordenadas numéricamente. Su estructura dependerá del tipo de variable a analizar.

2.2.1. Organización de Datos Cualitativos

Cuando los datos son cualitativos de escala nominal, la organización consiste en la construcción de una tabla de frecuencias con los siguientes columnas: la enumeración de las distintas modalidades que presenta la variable, el número de datos que corresponde a cada modalidad (frecuencia absoluta, f_i) y la proporción que cada uno de ellos representa con respecto al total (frecuencia relativa, fr_i). La tabla 2.1 muestra la estructura de una tabla de frecuencias para este caso.

Tabla 2.1: Tabla de Frecuencias para datos cualitativos de escala nominal

Modalidades	f_i	fr_i
1	f_1	fr_1
2	f_2	fr_2
\vdots	\vdots	\vdots
k	f_k	fr_k

donde

$$\sum_{i=1}^k f_i = n: \text{ representa el número total de datos.}$$

$$fr_i = \frac{f_i}{n} \text{ y debe cumplirse que } \sum_{i=1}^k fr_i = 1$$

Ejemplo 2.1 *A continuación se muestran los resultados obtenidos al aplicar una encuesta a 50 estudiantes de FACES donde se les preguntó sobre la carrera que estudiaban (A: Administración, C: Contaduría, E: Economía, ES: Estadística):*

La variable en este ejemplo es la carrera que estudian las personas, la cual es cualitativa de escala nominal. Al organizar los datos en una distribución de frecuencia se obtiene la tabla 2.2

C	A	A	C	C	A	A	E	A	C
E	E	C	ES	E	A	C	C	A	C
C	A	ES	C	E	A	A	C	A	C
C	C	A	E	E	A	C	C	C	A
C	C	A	C	C	C	C	ES	A	E

Tabla 2.2: Tabla de frecuencia para los datos del ejemplo 2.1

Carrera	f_i	fr_i
Administración	16	0.32
Contaduría	23	0.46
Economía	8	0.16
Estadística	3	0.06

Si los datos son cualitativos de escala ordinal, su organización implican dos cosas: en primer lugar, las clases llevan un orden preestablecido por las modalidades de la variable; en segundo lugar se incorporan a la tabla 2.1, columnas que muestren la frecuencia absoluta acumulada, F_i , y la proporción que cada uno de ellos representa con respecto al total, frecuencia relativa acumulada, Fr_i . La tabla 2.3 muestra la estructura de una tabla de frecuencias para este caso.

Tabla 2.3: Tabla de Frecuencias para datos cualitativos de escala ordinal

Modalidades	f_i	fr_i	F_i	Fr_i
1	f_1	fr_1	F_1	Fr_1
2	f_2	fr_2	F_2	Fr_2
\vdots	\vdots	\vdots	\vdots	\vdots
k	f_k	fr_k	F_k	Fr_k

donde

$$F_l = \sum_{i=1}^l f_i$$

$$Fr_l = \sum_{i=1}^l fr_i = \frac{F_l}{n} \text{ y debe cumplirse que } Fr_k = 1$$

Ejemplo 2.2 *Los siguientes datos corresponden a una consulta realizada a 45 pacientes acerca de su percepción sobre la atención prestada en el Instituto Autónomo Hospital Universitario de Los Andes, IAHULA (MB: Muy Bueno, B: Bueno, A: Aceptable, M: Malo, MM: Muy Malo):*

MB	B	B	A	A	M	A	MM	B	A
B	B	MM	MB	A	A	M	M	B	B
M	A	MM	MB	B	A	B	MB	A	B
B	M	M	B	B	A	B	B	M	A
MB	B	M	MM	A					

Aquí la variable es la percepción de los pacientes, la cual es cualitativa de escala ordinal. En la tabla 2.4 se muestra la organización de estos datos en una tabla de frecuencias.

Tabla 2.4: Percepción de la calidad de atención en pacientes del IAHULA

Percepción	f_i	fr_i	F_i	Fr_i
Muy buena	5	0.1111	5	0.1111
Buena	16	0.3555	21	0.4667
Aceptable	12	0.2667	33	0.7333
Mala	8	0.1778	41	0.9111
Muy mala	4	0.0889	45	1

2.2.2. Organización de Datos Cuantitativos

Si los datos son cuantitativos, los mismos pueden ser discretos o continuos. Para su organización se usa un procedimiento similar al utilizado con los datos cualitativos, considerando otros aspectos que la hacen más laboriosa.

Ejemplo 2.3 En un curso de estadística I se observa el número de asignaturas aprobadas por cada uno de los alumnos:

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

Ejemplo 2.4 Para los alumnos del ejemplo 2.3, se obtiene su estatura:

1.55	1.55	1.58	1.57	1.59	1.60	1.65	1.70	1.73	1.57
1.56	1.60	1.61	1.62	1.69	1.68	1.71	1.71	1.74	1.79
1.77	1.67	1.65	1.65	1.59	1.58	1.55	1.63	1.62	1.61
1.64	1.68	1.70	1.72	1.72	1.76	1.74	1.71	1.75	1.75
1.58	1.71								

La variable número de asignaturas en el ejemplo 2.3 es discreta, mientras que en el ejemplo 2.4, la variable estatura es continua. En estos casos la tabla de frecuencias contiene los siguientes elementos:

- **Intervalos de Clase:** El intervalo total en que están repartidas las observaciones es dividido en k intervalos parciales. A estos intervalos se les denomina intervalos de clase o, simplemente clases. Deben ser excluyentes
- **Límites de Clase:** Extremos de los intervalos de clase. Al menor de estos valores se le llama límite inferior y al mayor, límite superior.
- **Marcas de Clase (m_i):** Punto medio o centro de intervalo. Es una forma abreviada de representar el intervalo. De esta forma, todos los cálculos que se realizan como si en lugar de tener n_i valores en la clase i , se tiene n_i veces el mismo valor, m_i
- **Frecuencia Absoluta (f_i):** Número de observaciones contenidas o incluidas en una clase. Se debe satisfacer la siguiente igualdad

$$n = \sum_{i=1}^k f_i$$

donde n es el número total de datos.

- **Frecuencia Relativa (fr_i):** Proporción de los datos contenidos en la clase. Se obtiene al dividir la frecuencia absoluta entre el número total de observaciones. Debe cumplirse que

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

- **Frecuencia Absoluta Acumulada (F_i):** Suma de frecuencias absolutas hasta la clase correspondiente. De esta forma, la frecuencia acumulada para la clase k es el número total de datos, n .
- **Frecuencia Relativa Acumulada (Fr_i):** Suma de las Frecuencias Relativas hasta la clase correspondiente. Se pueden obtener dividiendo la frecuencia absoluta acumulada entre el número total de observaciones. Para la clase k se cumple que $1 = Fr_k$.

Nota: En el caso discreto, cuando el número de valores diferentes que puede tomar la variable es pequeño, entonces cada uno de ellos representa una clase. De esta forma las marcas de clase coinciden con las clases. Lo mismo es válido en el caso continuo, cuando el número de datos es pequeño.

Para construir una tabla o distribución de frecuencias, en el caso de variables cuantitativas se debe seguir el siguiente procedimiento:

1. Obtener los extremos del intervalo total (V_{max} y V_{min}).
2. Obtener el rango o recorrido de la variable, $R = V_{max} - V_{min}$.
3. Determinar el número de clases y la amplitud de las mismas. Para determinar el número de clases no existe una regla fija. Una primera aproximación es tomar

$$K = N^{\circ} \text{ de clases} = \sqrt{n}$$

Esta aproximación no siempre es conveniente, sobre todo cuando n es grande.

Existe una fórmula para calcular el número óptimo de clases, denominada fórmula de Stugers

$$K = N^{\circ} \text{ de clases} = 1 + 3,3 \log n$$

Cuando se particionan los datos en clases, es generalmente recomendado usar entre 5 y 15 clases. Fuera de estos extremos, la organización resulta poco eficiente. Si hay pocas clases la pérdida de información es por lo general significativa. Si hay muchas clases y adicionalmente el número de datos es pequeño, las frecuencias de clases tienden a subir y bajar de un manera desordenada evitando que se produzca una distribución ideal de los datos.

Una vez que se toma una decisión en cuanto al número de clases, la amplitud de las clases, es simplemente

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

Esto permite en resumen, particionar los datos en K clases, cada una con amplitud A . Es importante hacer notar que, no siempre es posible contar con clases de igual amplitud.

Si la amplitud de los intervalos no es constante, se debe corregir entonces las frecuencias, dividiendo las mismas por la amplitud del intervalo.

4. Construir los Intervalos de Clase: Para construir la primera clase, seleccionamos como un límite inferior el valor mínimo (V_{min}). El límite superior se obtiene al sumarle al límite inferior la amplitud, A . Para la segunda clase se tiene que el límite inferior es el límite superior de la primera clase y el límite superior, resulta de sumarle a este, A . Siguiendo este procedimiento construimos las k clases. Como el límite superior de una clase representa el límite inferior de la clase siguiente, conviene considerar las clases como intervalos del tipo $[L_{inf} - L_{sup})$; esto es, intervalos cerrados por la izquierda y abiertas por la derecha.

5. Calcular las marcas de clase (m_i): Las marcas de clase están representadas por los puntos medios de los intervalos de clase, es decir, $m_i = ls_i - li_i$
6. Obtener las frecuencias absolutas, relativas, absolutas acumuladas y relativa acumulada. La tabla 2.5 muestra la estructura de una tabla de frecuencias para datos cuantitativos

Tabla 2.5: Tabla de Frecuencias para datos cuantitativos

Clases	m_i	f_i	fr_i	F_i	Fr_i
$[li_1 - ls_1)$	m_1	f_1	fr_1	F_1	Fr_1
$[li_2 - ls_2)$	m_2	f_2	fr_2	F_2	Fr_2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$[li_k - ls_k)$	m_k	f_k	fr_k	F_k	Fr_k

Ejemplo 2.5 *A continuación se muestra la información sobre el número de hijos que tienen 40 Mujeres extraídas al azar de la ciudad de Mérida.*

1	1	3	3	2	4	4	1
1	2	1	3	3	2	1	3
2	1	2	2	4	3	4	4
4	0	3	0	4	1	5	2
2	3	3	4	4	4	1	2

Antes de organizar los datos en una distribución de frecuencia, observemos que la variable es discreta y además posee pocos valores diferentes, pues su rango está dado por $\{0, 1, 2, 3, 4, 5\}$. Entonces las clases de la distribución de frecuencia están dadas por los valores individuales de la variable. En la tabla 2.6 se presenta la organización de estos datos.

Tabla 2.6: Distribución del N° de Hijos que tienen 40 Mujeres

N° de Hijos	f_i	fr_i	F_i	Fr_i
0	2	0,050	2	0,050
1	9	0,225	11	0,275
2	9	0,225	20	0,500
3	9	0,225	29	0,725
4	10	0,250	39	0,975
5	1	0,025	40	1