



Identificación del Objeto de aprendizaje

Fecha	Julio 2013
Asignatura	Estadística y Probabilidad I
Unidad	Unidad 0. Introducción
Tiempo disponible	4 horas
Aprendizajes	<p>El alumno:</p> <ul style="list-style-type: none"> • Adquiere una primera noción de la Estadística y su utilidad. • Explica el significado que tienen los términos variable, población y muestra. • Conoce la importancia de trabajar con muestras seleccionadas de alguna población. • Explica la noción de variabilidad en Estadística. • Conoce la noción de azar y la necesidad de medirlo. • Conoce que es posible hacer mal uso de la información estadística
Tema	<p>1. Noción y utilidad de la Estadística. 2. Nociones básicas.</p> <ul style="list-style-type: none"> • Variable, población y muestra. • Variabilidad. • Azar y probabilidad. <p>3. Usos indebidos de la Estadística.</p>
Palabras claves	Estadística, Azar, Probabilidad, posibilidad, Población, muestra, muestra aleatoria, muestra sesgada, variable cualitativa o cuantitativa, tabla de datos, Gráficos de datos.
Autor	Tomás Espinosa Martínez.

Objetivo (para el profesor)

<p>El profesor deberá buscar que el alumno se apropie de una visión inicial de la Estadística y de la Probabilidad, a partir del planteamiento y discusión de ejemplos y problemas de su entorno que le permitan apreciar los alcances de la materia. Se sugiere iniciar una discusión con las ideas previas que tienen los estudiantes, considerando ejemplos tomados de la vida diaria tales como gráficas en artículos periódicos, resultados deportivos, etc. y con lecturas seleccionadas por el profesor.</p> <ul style="list-style-type: none"> • A partir de la discusión de esos ejemplos, hacer hincapié en la importancia de la Estadística y su aplicación en otras áreas del conocimiento. • Por medio de lluvia de ideas, construir los conceptos de variable, población y muestra, dentro del contexto estadístico. • Solicitar a los alumnos ejemplos de poblaciones y muestras, y discutirlos con el grupo. • Discutir con el grupo el hecho de que la Estadística tiene como principal aplicación inferir 	Objetivo (para el profesor)
---	------------------------------------



características de poblaciones, señalando en términos generales la secuencia del proceso estadístico.

- Plantear problemas y ejemplos en donde el estudiante tenga oportunidad de observar la homogeneidad o heterogeneidad de los valores de la variable.
- Mostrar ejemplos, de ser posible con material lúdico, en donde el estudiante tenga contacto con situaciones aleatorias.
- Se sugiere presentar al grupo modelos de información sesgada o dirigida, con el fin de que se conozca el mal uso que se hace de la Estadística. Igualmente se sugiere tomar ejemplos cotidianos, como los sondeos telefónicos de los noticieros, entre otros.

Índice de navegación del Objeto de aprendizaje

1. Estadística

1.1 Noción y utilidad de la Estadística.

- Actividad 1
- Actividad 2
- Actividad 3

1.2. Nociones básicas.

- Actividad 4
- Actividad 5

1.2.1 Variable, población y muestra.

- Actividad 6
- Actividad 7
- Actividad 8

1.2.2 Variabilidad.

- Actividad 9

1.2.3. Usos indebidos de la Estadística.

- Actividad 10
- Actividad 11

2. Actividad Final

- 3. Glosario
- 4. Referencias
- 5. Créditos



1. Introducción

1. INTRODUCCIÓN.

La Estadística es una rama de la matemática que realiza la recolección, análisis e interpretación de datos de una muestra representativa, para obtener conclusiones o para explicar condiciones regulares o irregulares de algún fenómeno o estudio aplicado, de ocurrencia en forma aleatoria o condicional.

El conocimiento de la estadística, es el vehículo que permite llevar a cabo el proceso relacionado con la investigación científica y es transversal en áreas de conocimiento tales como: las Ciencias sociales, la administración y economía, las ciencias de la salud, el control de calidad y la teoría de la confiabilidad en la ingeniería, sin descontar su gran utilidad dentro de las áreas de investigación de la física y la química entre otras. Se usa para la toma de decisiones en áreas de negocios o instituciones gubernamentales.

La estadística se clasifica en dos ramas o áreas:

1. La estadística descriptiva, se dedica a la recolección, presentación (a través de la descripción y visualización gráfica) y resumen de datos originados a partir de los fenómenos o situación bajo estudio. Los datos pueden ser resumidos de manera numérica a través de tablas o gráficamente a través de diagramas. Ejemplos básicos de los parámetros estadísticos que se obtienen como resultado de muchos datos son: la media y la desviación estándar, la moda, entre otras. Algunos ejemplos de los gráficos que pueden obtenerse para representar la información son: histograma, diagramas de barras, ojivas, diagramas de caja y bigotes, etc.
2. La estadística inferencial, permite la creación de modelos explicativos y predictivos, de diferentes situaciones económicas, sociales y en general científicas, a través de las inferencias y predicciones asociadas a los fenómenos en cuestión teniendo en cuenta la aleatoriedad de las observaciones. Esta herramienta científica se usa para modelar patrones en los datos, generando hipótesis acerca de su comportamiento y posibilitando la realización de inferencias acerca de la población de datos bajo estudio. Las herramientas estimativas de las características numéricas desembocan en la que se denomina estimación puntual y por intervalo, mientras que las posibles conclusiones son llevadas a cabo a través de un proceso llamado prueba de hipótesis donde se contrasta la teoría (hipótesis) contra la observación obtenida en la muestra. También atañe a esta área, resolver el problema de la determinación del tamaño de la muestra, que es necesaria para realizar inferencias confiables. Otras técnicas o herramientas lo constituyen los pronósticos de futuras observaciones, las descripciones de asociación correlación o la modelación de relaciones entre dos o más variables llamado análisis de regresión, aunque existen otras técnicas de modelamiento que incluyen otras herramientas, tales como: el análisis de Varianza (anova), las series de tiempo y la llamada minería de datos.

Se considera que la llamada Estadística aplicada comprende ambas ramas de la estadística (descriptiva e inferencial, cuando se investigan las bases teóricas de la materia, se dice que se está creando una estadística matemática, a menudo al considerar un conjunto de datos para su análisis se usa frecuentemente la palabra «estadística» para referirse al resultado de aplicar un algoritmo o proceso estadístico al conjunto de datos, por ejemplo el cálculo de una media o de la desviación estándar.



Actividad 1 Conteste el siguiente cuestionario:

<i>Cuestionario</i>	<i>Retroalimentación</i>
<p>1. Es la una rama de la matemática que realiza la recolección, análisis e interpretación de datos de una muestra representativa, para obtener conclusiones o para explicar condiciones regulares o irregulares de algún fenómeno o estudio aplicado, de ocurrencia en forma aleatoria o condicional.</p> <p>A) Probabilidad B) Estadística C) Lógica D) Informática</p>	<p><i>Respuesta correcta:</i> En efecto Es la rama de la matemática que realiza la recolección, análisis e interpretación de datos de una muestra representativa, para obtener conclusiones o para explicar condiciones regulares o irregulares de algún fenómeno o estudio aplicado, de ocurrencia en forma aleatoria o condicional.</p> <p><i>Respuesta incorrecta:</i> Recuerda la estadística a partir de los datos obtiene conclusiones que explican condiciones regulares o irregulares de algún fenómeno o estudio aplicado.</p>
<p>2. es el vehículo que permite llevar a cabo el proceso relacionado con la investigación científica y es transversal en áreas de conocimiento. Se usa para realizar la toma de decisiones en áreas de negocios o instituciones gubernamentales.</p> <p>A) Probabilidad B) Estadística C) Lógica D) Informática</p>	<p><i>Respuesta correcta:</i> En efecto permite llevar a cabo el proceso relacionado con la investigación científica y es transversal en áreas de conocimiento tales como: las Ciencias sociales, la administración y economía, las ciencias de la salud, el control de calidad y la teoría de la confiabilidad en la ingeniería, sin descontar su gran utilidad dentro de las áreas de investigación de la física y la química entre otras.</p> <p><i>Respuesta incorrecta:</i> Recuerda que la estadística permite llevar a cabo el proceso relacionado con la investigación científica y es transversal en áreas de conocimiento. Se usa para realizar la toma de decisiones en áreas de negocios o instituciones gubernamentales.</p>



3. se dedicada a la recolección, presentación (a través de la descripción y visualización grafica) y resumen de datos originados a partir de los fenómenos o situación bajo estudio. Los datos pueden ser resumidos de manera numérica a través de tablas o gráficamente a través de diagramas. Ejemplos básicos de los parámetros estadísticos que se obtienen como resultado de muchos datos son: la media y la desviación estándar, la moda, entre otras.

- A) *Estadística robusta*
- B) *Estadística descriptiva*
- C) *Estadística paramétrica*
- D) *Estadística inferencial*

Respuesta correcta: En efecto la estadística descriptiva está dedicada a la recolección, presentación (a través de la descripción y visualización grafica) y resumen de datos originados a partir de los fenómenos o situación bajo estudio. Los datos pueden ser resumidos de manera numérica a través de tablas o gráficamente a través de diagramas.

Respuesta incorrecta: Recuerda que la estadística descriptiva busca obtener datos y procesarlos para obtener por ejemplo: la media y la desviación estándar, la moda, entre otras. Algunos ejemplos de los gráficos que pueden obtenerse para representar la información son: histograma, diagramas de barras, ojivas, diagramas de caja y bigotes, etc.

4. permite la creación de modelos explicativos y predictivos, de diferentes situaciones económicas, sociales y en general científicas, a través de las inferencias y predicciones asociadas a los fenómenos en cuestión teniendo en cuenta la aleatoriedad de las observaciones. Esta herramienta científica se usa para modelar patrones en los datos, generando hipótesis a cerca de su comportamiento y posibilitando la realización de inferencias acerca de la población de datos bajo estudio. Las herramientas estimativas de las características numéricas desembocan en la que se denomina estimación puntual y por intervalo, mientras que las posibles conclusiones son llevadas a cabo a través de un proceso llamado prueba de hipótesis donde se contrasta la teoría (hipótesis) contra la observación obtenida en la muestra. También atañe a esta área, resolver el problema de la determinación del tamaño de la muestra, que es necesaria para realizar inferencias confiables.

- A) *Estadística robusta*
- B) *Estadística descriptiva*
- C) *Estadística paramétrica*
- D) *Estadística inferencial*

Respuesta correcta: En efecto la estadística inferencial está dedicada a la creación de modelos explicativos y predictivos, de diferentes situaciones económicas, sociales y en general científicas, a través de las inferencias y predicciones asociadas a los fenómenos en cuestión tomando en cuenta la aleatoriedad de las observaciones..

Respuesta incorrecta: Recuerda que la estadística inferencial es la herramienta científica se usa para modelar patrones en los datos, generando hipótesis a cerca de su comportamiento y posibilitando la realización de inferencias acerca de la población de datos bajo estudio.



5. comprende ambas ramas de la estadística (descriptiva e inferencial, cuando se investigan las bases teóricas de la materia, se dice que se está creando una estadística matemática, a menudo al considerar un conjunto de datos para su análisis se usa frecuentemente La palabra «estadística» para referirse al resultado de aplicar un algoritmo o proceso estadístico al conjunto de datos, por ejemplo el cálculo de una media o de la desviación estándar.

- A) Estadística robusta
- B) Estadística paramétrica
- C) Estadística aplicada
- D) Estadística muestral

Respuesta correcta: En efecto la estadística aplicada comprende ambas ramas de la estadística (descriptiva e inferencial,

Respuesta incorrecta: Recuerda que la estadística aplicada comprende ambas ramas de la estadística (descriptiva e inferencial,

1.1 Noción y utilidad de la Estadística.

La Estadística permite dar respuestas a interrogantes y necesidades planteadas por la sociedad actual. La tarea fundamental es trabajar con datos, con el objetivo de representar la realidad; predecir futuros comportamientos de situaciones o fenómenos y ayudar en la toma de decisiones.

La palabra Estadística deriva de “Estado”, etimológicamente es la ciencia del gobierno de un Estado (país). Históricamente el proceso estadístico consistía en la práctica de un recuento de características de la población humana, cada determinado tiempo (por ejemplo un censo) sirviendo estos conocimientos para la administración y *ordenamiento económico y político de todos los gobiernos del mundo. En muchos casos la estadística produce información útil para medir el llamado Estado del Bienestar de la sociedad. La estadística responde a demandas de indole social, economico, científico y tecnologico.*

Algunas de sus aplicaciones se pueden observar en procesos de tipo industrial tales como el control de calidad. En el ámbito económico, se utiliza en la investigación de mercados. En investigación médica, sirve para la creación o diseño de nuevos medicamentos que combaten algunas enfermedades. En el ámbito social sirve para investigar sobre el desarrollo de los tipos de adicciones y su relación con la edad, grados de escolaridad, o con factores económicos o bien relacionados con el entorno y la familia.

Por otra parte en los procesos de obtención de información, que generan datos, la cual es la materia prima de esta ciencia. Si se desea determinar características de una población de personas, se debe tomar en cuenta el tiempo requerido para entrevistar a todos los integrantes de esa población, su costo económico, político social y ético. Por lo que muchas veces frecuentemente se utilizara una poción de esta denominada muestra y se



deberá determinar mecanismos para determinar el tamaño de esta, de forma que la información resultante sea confiable para la investigación de la totalidad (población), tomando en cuenta los diferentes tipos de muestreo.

ACTIVIDAD 2

Lea con cuidado los enunciados de las columnas y relacione de manera correcta ambas columnas.

<p>a. Tiene por objetivo representar la realidad, predecir futuros comportamientos de situaciones o fenómenos y ayudar en la toma de decisiones.</p>		<p>(.... b) Control de calidad</p>
<p>b. proceso de tipo industrial donde interviene la estadística</p>		<p>(.... a) La estadística “ predice futuros comportamientos de situaciones o fenómenos y, tomando en cuenta la información (Datos)”.</p>
<p>c. aplicación de la estadística en el ámbito económico</p>		<p>(.... c) investigación de mercado</p>
<p>d. etimológicamente es la ciencia del gobierno. Históricamente el proceso estadístico consistía en la práctica de un recuento de características de la población humana, cada determinado tiempo (por ejemplo un censo) sirviendo estos conocimientos para la administración y <i>ordenamiento económico y político de todos los gobiernos del mundo.</i></p>		<p>(.... e) Media aritmética “al cancelarse los valores extremos , los datos intermedios, no se ven afectados por valores extremos”</p>
<p>e. “números muy Negativos o positivos grandes modifican fuertemente su valor”.</p>		<p>(... d) significado de acuerdo a Las raíces del significado de la palabra“estadística”</p>



Problema 1.

1.2. Nociones básicas.

Ya se ha comentado que la materia prima de la estadística son los datos, estos se pueden obtener en algunos casos a partir de encuestas o bien como resultado de las mediciones en un determinado experimento.

En la determinación de las características a determinar de un conjunto de datos se debe definir el marco muestral, en donde se determina que es lo que se va a medir (el para qué), es decir se determina la variable en un proceso de medición. En donde se realizara las mediciones (el cómo), con qué recursos se dispone, el alcance de la investigación (el donde), que se pretende determinar con la información recabada (datos).

Por ejemplo considere usted que se desea realizar una investigación relacionada con la una ruta de transporte público de ida y de regreso a la base donde partió, se podrían considerar varias variables; entre estas la medición de los tiempos que tarda un microbús, otra podría ser el número de pasajeros por vuelta, otra el consumo de combustible en ese trayecto. En otras palabras en la investigación se debe determinar la variable a medir. Por otra parte se debe identificar cual es la totalidad de la información disponible (a la que llamaremos población), es decir se debe preguntar o considerar si el proceso atañe una ruta en el DF, a varias rutas en el DF, o a todas las rutas de microbuses en todo el país.

Actividad 3. Identifique dos posibles marcos muestrales en cada uno de los siguientes casos:

- Se desea conocer la presión sanguínea de un grupo de personas.
- Se desea determinar la proporción de votantes están a favor de un candidato político.
- Se debe determinar el contenido promedio de botellas de refresco
- Se desea conocer la proporción de personas e determinado tipo sanguíneo
- Se desea medir la presión soportada por un empaque en una olla exprés

Una vez que se ha determinado el marco muestral, se está en posibilidades de determinar la población y la muestra.

Po ejemplo se podría determinar la estatura o peso o longitudes de los pies de los estudiantes de un grupo académico, esta podría ser la población o bien se podría indagar sobre estas características en grupo mayor de personas, por ejemplo los estudiantes de un plantel del CCH, o a un más aún, podría considerarse todos los estudiantes de nivel bachillerato en el país, es decir se debe determinar en donde se realizaran las mediciones (obtención de los datos) y los alcances de la investigación. Sin embargo, si estos datos están contenidos como subconjunto en otro de



mayor, el cual podría estar representado, por ejemplo por la población mundial (algo más o menos como 7000 millones de personas), en este caso las mediciones realizadas en los conjuntos de datos anteriores podrían considerarse como muestras o porciones de esta último conjunto de datos.

Es importante aclarar que la información que podría interesar no solamente atañe a características de una población humana, por ejemplo podríamos considerar una colección de datos referentes al voltaje que tienen cien pilas nuevas AAA, la longitud de lápices de un paquete de quinientos lápices de la marca MIRADO, el grosor de las hojas de cuaderno de la marca ESCRIBE de cien hojas, el peso de cada una de las diez piernas de pollo que venden en el POLLO FELIZ, la cantidad de grasa que tienen cada una de las rebanadas de jamón seleccionadas de empaque con contenido de tres kilos (doscientas rebanadas aproximadamente).

Es decir cuando se nombra la palabra “Población” se pueden considerar mediciones de características que no atañen a la población humana, en otras palabras, se le llamara Población Estadística para generalizar a un conjunto de datos. A continuación se presentan las definiciones:

Población estadística. En la totalidad de los datos, equivalente a decir que es el conjunto universo donde se considerará la investigación. Las conclusiones y lo que se diga referente a esta, se afirma con absoluta certeza.

Muestra es una porción de la población. Lo que se puede afirmar acerca del análisis de estos datos contiene siempre una determinada incertidumbre.

Actividad 4. Identifique dos posibles marcos muestrales en cada uno de los Enunciados

- a. Se desea conocer la presión sanguínea de un grupo de 2000 personas que acuden el día de hoy a consulta en el hospital regional ISSSTE Zaragoza, considerando realizar mediciones de la presión a los veinte pacientes atendidos en el consultorio número tres.
- b. Se desea determinar la proporción de votantes están a favor de un candidato político, para este efecto se selecciona 450 personas y se entrevistan para conocer su filiación política. El Instituto Electoral del Distrito Federal (IEDF) indicó que de acuerdo con los datos estadísticos del Padrón Electoral y la Lista Nominal con los que cuenta el IEDF 2012, la demarcación con más ciudadanos en condiciones de votar es Iztapalapa, con 1 millón 362,070 personas inscritas.
- c. Se debe determinar el contenido promedio de 10,000 botellas de refresco Jarrito, que supuestamente debieran tener exactamente 600 ml, para este efecto se revisaran 120 botellas de este producto.
- d. Se desea conocer la proporción de personas e determinado tipo sanguíneo



- e. Se desea determinar el consume promedio por hogar de agua diario, se puede elegir entre 10,000, pero para el estudio se consideran 290 hogares.
- f. Se dice que una persona parpadea aproximadamente 25 mil veces por semana, para cerciorarse de este dato se seleccionan 34 personas de un conjunto 300.

Actividad 5. Comente brevemente en los siguientes enunciados, como podría llevarse a cabo el experimento que corroboren o desechen las hipótesis:

- a. Un cocodrilo no puede sacar la lengua.
- b. Los diestros viven en promedio 9 años más que los zurdos.
- c. El corazón humano genera suficiente presión cuando bombea la sangre que podría esparcirla fuera del cuerpo hasta 10 metros de distancia.
- d. Una cucaracha puede vivir varias semanas sin cabeza.
- e. El corazón humano típico late más de 100,000 veces en un día.
- f. Más gente usa cepillos de dientes azules que de color rojo.
- g. Parpadeamos más de 10, 000,000 de veces en un año.
- h. Se producen más de 50,000 temblores terrestres en todo el mundo cada año.
- i. Las manzanas son más eficientes que la cafeína del café para mantener a la gente despierta en las mañanas.
- j. Oler plátanos y/o manzanas verdes pueden ayudar a bajar de peso.
- k. Las hormigas estiran la patas cuando despiertan en la mañana.
- l. Tomar agua después de comer reduce el ácido en la boca a un 61 por ciento, del valor inicial.
- m. El tejido de las encías no se regenera.
- n. Los gatos domésticos pueden alcanzar velocidades de hasta treinta millas por hora.
- o. La trompa de un elefante tiene hasta 40,000 músculos.

1.2.1 Variable, población y muestra.

A continuación se definen algunos conceptos que se consideran fundamentales en la estadística.

Aunque el termino población fue acuñado históricamente para medir características humanas, dentro de la estadística este término tiene un significado más amplio, de forma que un conjunto de datos de cualquier índole representa una población.



POBLACIÓN ESTADÍSTICA: Es el conjunto de elementos o datos de los que interesa estudiar alguna característica. Lo que se afirma o niegue acerca de ellos se realiza con absoluta certeza.

VARIABLE: Es la característica a investigar su comportamiento, es decir es lo que se desea estudiar. Puede ser cualitativa, cuando sus valores indican cualidad y no son numéricos (sin embargo pueden transformarse a números a través de las proporciones), o cuantitativas, cuando sus valores son numéricos. Ambos tipos de variables; cualitativas y cuantitativas también se llaman Variables estadísticas.

VARIABLES CUALITATIVAS: genero, estado civil, sabor de refrescos, color de pelo, grado de satisfacción ante un servicio, preferencia a un color en un determinado artículo, estado ocupacional de una habitación, el color de la luz de una bombilla eléctrica.

VARIABLES CUANTITATIVAS: Edad, número de llamadas realizadas desde un teléfono móvil en un año, la calificación obtenida al final del semestre en una asignatura, tiempo empleado en transportarse de casa al trabajo, la cantidad de combustible consumido por un transporte de carga para llevar productos de una ciudad a otra, el consumo de energía de determinada bombilla eléctrica. Estas se pueden clasificar discretas o continuas.

Una **variable discreta** es una variable que solo puede tomar valores dentro de un conjunto numerable, la situación inherente en estas variables se son las separaciones entre valores observables sucesivos. Por ejemplo el número de hamburguesas consumidas al año por una persona(0,1,2,3..., n) o bien cuando se compra una botella de agua potable, las capacidades de esta pueden ser 400 ml, 500 ml, 600 ml, 1000 ml, 1500 ml, 2000 ml, 2500 ml, 5000 ml , etc.

Una **variable continua**, esta puede tomar un valor cualquiera dentro de un rango predeterminado, en un intervalo acotado entre dos valores observables, y siempre es posible encontrar un tercer valor intermedio que también podría tomar la variable continua. Es decir una variable continua toma valores en un intervalo del conjunto de los números reales. Por ejemplo la cantidad de líquido que diariamente consume una persona, el tiempo que tarda una llamada telefónica, la cantidad de aire que se introduce a los pulmones al realizar una carrera, el tiempo que tarda un estudiante en resolver un ejercicio de estadística, Un atributo que se considera esencial de una variable continua es que, a diferencia de una variable discreta, nunca puede ser medida con exactitud , puesto que la primera genera de manera inevitable un error al realizar la medición. El valor obtenido a través de la medición depende de la precisión de los instrumentos de medición. Con una variable continua hay inevitablemente un error de medida. Por ejemplo el tiempo de retraso de un estudiante al llegar a la clase, considerando que el tiempo está limitado a diez minutos: 10, 7, 5:03, 5:04, 5.22, 5.223, 5.22367 etc.



MUESTRA: es una porción o subconjunto de una población.

Muestra aleatoria: es una porción o subconjunto de una población, en donde cada uno de los elementos de la población tienen una determinada probabilidad de ser seleccionados.

Algunas razones para estudiar una muestra en contraposición del estudio de una población son las siguientes:

COSTO ECONÓMICO, toda información cuesta dinero, algunos procesos son destructivos del artículo, por ejemplo medir el grado de dureza del diamante (prueba rockwell), el tiempo de duración de un televisor (se somete a pruebas que simulan su duración en el tiempo).

TIEMPO, una mayor cantidad de datos puede implicar mayor dedicación de tiempo. Carencia de personal calificado para obtener la información (a través de encuesta o de experimento).

CUESTIONES MORALES Y ÉTICAS, por ejemplo realizar experimentos que produzcan daño, dolor o muerte a personas o a animales. Por ejemplo, a los animales de laboratorio se les aplican químicos en los ojos para determinar el grado de dolor que producen el rímel y otros artículos cosméticos, o bien se les contagian enfermedades y después se les aplica un medicamento como posible cura y se investiga los resultados. Se sabe que, se regalan o aplican medicamentos y tratamientos a personas de bajos recursos económicos para determinar la bondad del medicamento en curación de enfermedades, descubriendo en el medicamento los subsecuentes efectos secundarios. En la década de los 50s en EUA e Inglaterra realizaron experimentos con personas de grupos económicamente vulnerables, engañándolos; diciéndoles que les inyectaban vitaminas, cuando en realidad se les inyectaba Yodo radioactivo para estudiar sus efectos de la radiación en seres humanos. Esta información fue muy importante desde el punto de vista de esos gobiernos, porque permitió a mandos militares saber cuántas bombas nucleares podían armar un operario altamente calificado antes de morir por los efectos de la radiación.

En los anteriores casos, es muy importante decidir a cuántos especímenes se deben sacrificar para obtener información confiable y verídica acerca del comportamiento de una población, con la finalidad de obtener conocimientos científicos resumidos en modelos predictivos (en particular el conocimiento de los parámetros de la población).

Por todas las razones anteriores, la Estadística siempre trabaja con muestras y no con una población.

EL MUESTREO Y LA GENERACIÓN DE MUESTRAS A PARTIR DE UNA POBLACIÓN ESTADÍSTICA. Se ha dicho que una muestra se obtiene de una población, esta actividad que permite seleccionar muestras de distintos tamaños con la condición de que no excedan el tamaño de la población recibe el nombre de muestreo.



Tipos de muestreo. Probabilísticos y no Probabilísticos, ambos conllevan a la obtención de la información, pero con distintos alcances.

A. muestreos probabilísticos

Son los métodos de muestreo que se basan en el principio de equiprobabilidad. Es decir, aquellos en los que todos los datos contenidos en la población tienen la misma probabilidad de ser elegidos para formar parte de una muestra (muestra aleatoria). Dentro de los métodos de muestreo probabilísticos encontramos los siguientes tipos:

A.1 MUESTREO ALEATORIO SIMPLE: El proceso para realizarlo es el siguiente: se asigna un número a cada dato de la población y a través de algún medio mecánico (papeletas dentro de una bolsa o urna, tablas de números aleatorios, números aleatorios generados con una calculadora u programa de computadora, etc.) se eligen determinado número de datos como sea necesario para completar el tamaño de muestra requerido. Este procedimiento puede presentar dificultades si la población es muy grande.

A.2 MUESTREO ALEATORIO SISTEMÁTICO: En este proceso se le utiliza en procesos industriales el área de control de calidad, en donde preferentemente se tiene una cadena de ensamblaje, los elementos de la población están numerados, pero en lugar de extraer n números aleatorios sólo se extrae uno, se dejan pasar k (Siendo k el resultado de dividir el tamaño de la población " N " entre el tamaño de la muestra " n ": $k = N/n$) elementos y nuevamente se extrae solo uno, el proceso se realiza de manera semejante hasta completar el número deseado de datos que conforman la muestra. El número i que emplea como punto de partida será un número al azar entre 1 y k . también este tipo de muestreo es muy popular en encuestas o ventas por teléfono, se toma el directorio y se deja pasar el dedo hasta encontrar el siguiente posible cliente.

A.3 MUESTREO ALEATORIO ESTRATIFICADO:

Es muy popular para fines de estudios económicos, sociales, y políticos.

Consiste en considerar categorías típicas diferentes entre sí (estratos) que poseen gran homogeneidad respecto a alguna característica (se puede estratificar, por ejemplo, según el ingreso mensual (con fines de búsqueda de compradores potenciales de autos o residencias), la profesión, el consumo de agua, consumo de energía eléctrica, el municipio de residencia, el valor catastral de algunas zonas de la ciudad, el género el estado civil, etc.). Con este tipo de muestreo todos los estratos de interés estarán representados adecuadamente en la muestra. En cada estrato por ser independiente se aplicarse el muestreo aleatorio simple para elegir los datos que formarán parte de la muestra. La distribución de la muestra en función de los diferentes estratos se denomina afijación, y puede ser de tres tipos:

Afijación Simple: A cada estrato le corresponde igual número de elementos muestrales.

Afijación Proporcional: La distribución se hace de acuerdo con el peso (tamaño) de la población en cada estrato.

Afijación Óptima: Se tiene en cuenta la previsible dispersión de los resultados, de modo que se considera la proporción y la desviación típica.



A.4 MUESTREO ALEATORIO POR CONGLOMERADOS: En este tipo de muestreo se ahorran recursos económicos por lo general, ya que mucha de la información geográfica y poblacional se puede encontrar en oficinas de gobierno tales como Instituto Nacional de Estadística y Geografía INEGI.

En el muestreo por conglomerados la unidad muestral es un grupo de elementos de la población que forman una unidad, a la que llamamos conglomerado. Las unidades hospitalarias, escuelas, manzanas o colonias de una ciudad, una caja de manzana o aguacate, una urna electoral, son ejemplos de un conglomerado.

El muestreo por conglomerados consiste en seleccionar aleatoriamente un cierto número de conglomerados (el necesario para alcanzar el tamaño muestral establecido) y en investigar después todos los elementos pertenecientes a los conglomerados elegidos.

B. métodos de muestreo no probabilísticos

A veces, para estudios exploratorios, el muestreo probabilístico resulta excesivamente costoso o bien son situaciones de carácter irrepetible (por ejemplo la explosión de una estrella (supernova), la caída de un meteorito o la erupción de un volcán. En estas situaciones se acude a métodos no probabilísticos, bajo la consideración de que sirven para generar información (denominada sesgada), y bajo la consideración de que no se deben realizar generalizaciones (estimaciones o inferencias sobre la población), esto principalmente por la carencia de certeza de que la muestra extraída sea representativa de un grupo mayor de posibles datos (población). En general en estos procesos se seleccionan a los sujetos siguiendo determinados criterios, procurando en la medida de lo posible, que la muestra sea representativa. Estos procesos permiten extraer información, por ejemplo los estudios médicos, donde los casos no son seleccionados aleatoriamente de la población, o bien lo que hace el profesor de estadística en el aula para obtener información acerca de la estatura o pesos de los alumnos. Entre estos tipos de muestreo tenemos:

B.1.- Muestreo por cuotas: También denominado en ocasiones: “accidental” o “manipulado” o “adecuado”. Por ejemplo se realiza una encuesta a compradores de un centro comercial en “Perisur” para determinar su preferencia a cierta fragancia o marca de ropa, o lo que investiga las compañías televisoras al medir el rating de ciertos programas de televisión. Este tipo de construcción de la muestra se asienta sobre la base de un buen conocimiento de los estratos de la población y de los individuos más “representativos” o “adecuados” para los fines de la investigación. Mantiene, por tanto, semejanzas con el muestreo aleatorio estratificado, pero no tiene el carácter de aleatoriedad de aquél. En este tipo de muestreo se fijan cuotas tales como género, edad, nivel económico etc. Por ejemplo si la muestra es de 60 individuos se deben seleccionar 30 hombres y treinta mujeres de nivel económico medio, si una encuesta se deberá realizarse a 100 personas se pueden seleccionar 30 de estas como compradores de Liverpool y el resto como compradores en el palacio de hierro entre las edades de 20 a 30 años. La cuota son las exigencias de que se cumplan esas características. Este método se utiliza mucho en las encuestas de opinión en los noticieros de la TV mexicana.



B.2.- Muestreo intencional o de conveniencia: Este tipo de muestreo se caracteriza por un esfuerzo deliberado de obtener muestras "representativas" mediante la inclusión en la muestra de grupos supuestamente típicos. Por ejemplo la encuestas realizadas por la empresa de CONSULTA MITOFSKI, utilización en sondeos preelectorales de zonas que en anteriores votaciones han marcado tendencias de voto por cierto partido político. Aunque algunas personas comentan acerca del posible sesgamiento de la información, y de que los resultados obtenidos, más que proporcionar conocimiento de las preferencias sirvió más bien como propaganda a ciertos partidos políticos. Este proceso se realiza al seleccionar directa e intencionadamente los individuos de la población, como lo es el caso frecuente de utilizar como muestra los individuos a los que se tiene fácil acceso, por ejemplo los profesores de estadística emplean con mucha frecuencia los datos de sus propios alumnos, tales como peso, estatura, número de materias aprobadas o reprobadas, color de ojos de los estudiantes, etc.



B.3.- Bola de nieve: Se localiza a algunos individuos, los cuales conducen a otros, y estos a otros, y así hasta conseguir una muestra suficiente. Este tipo se emplea muy frecuentemente cuando se hacen estudios con poblaciones "marginales", delincuentes, personas que pertenecen sectas, personas que consumen drogas, determinados tipos de enfermos que son segregados por la sociedad, etc.

B.4.- Muestreo Discrecional · A criterio del investigador los elementos son elegidos sobre lo que él cree que pueden aportar al estudio.

CONSTRUCCION DE UNA MUESTRA ALEATORIA SIMPLE.

A continuación deberás construir una muestra aleatoria simple. Como simulación del proceso de elección de los elementos que conformaran la muestra aleatoria

Actividad 6

Considera los siguientes datos proporcionados en la tabla como una población, dichos datos son reales y se han obtenidos como resultado de una encuesta sobre la longitud del calzado de cien estudiantes.

dato		dato		dato		dato		dato	
1	3	21	7.5	41	6	61		81	5.5
2	3.5	22	8	42	5.5	62	5.5	82	6
3	4	23	3	43	8	63	6.5	83	5
4	6	24	4	44	3	64	5.5	84	5.5
5	6	25	5.5	45	7.5	65	7.5	85	7
6	3.5	26	6	46	4	66	3	86	7
7	5.5	27	4.5	47	9	67	8	87	8.5
8	3.5	28	3.5	48	4	68	8.5	88	3.5
9	5.5	29	6.5	49	9	69	8.5	89	5
10	6	30	5.5	50	9	70	3.5	90	5
11	3.5	31	6	51	4	71	3	91	3
12	5	32	4.5	52	5.5	72	4.5	92	7
13	6	33	5	53	4.5	73	4.5	93	3
14	3.5	34	3.5	54	6	74	4	94	7
15	5	35	5	55	6.5	75	7	95	3
16	5	36	4	56	5	76	4.5	96	5
17	3.5	37	5	57	9	77	4	97	7
18	6	38	4	58	5	78	7	98	3
19	5	39	4	59	5.5	79	7	99	5.5
20	5	40	4	60	4.5	80	4.5	100	5



Ciudad de México			Mundo entero	
Tipo de sangre	Porcentaje de la Población %		Factor RH	Porcentaje de la Población %
O	65		positivo	85
A	25		negativo	15
B	8.5			
AB	1.5			

En uno de los números de la revista “muy interesante” apareció un artículo, acerca de un estudio de la Escuela de Medicina de la Universidad de Colorado (EE UU), publicado en la revista *Nature Genetics*, en donde se revela que las personas que tienen ojos azules o verdes son menos propensas a sufrir vitíligo (mal del pinto). El tener los ojos claros implica riesgo de padecer cáncer de piel. Los investigadores llegaron a estas conclusiones tras estudiar a más de 3.000 europeos.

Cuestionario 2

Se puede considerar esta muestra representativa. Explique y/o justifique.

1. ¿De la población europea?
2. ¿De la población mundial?
3. ¿De la población de México?

Actividad 8

El color de ojos es un rasgo genético que está determinado por la cantidad y la distribución de melanina en el iris, a continuación se proporciona información de los porcentajes de distribución mundial.

50% de toda la población mundial tiene los ojos castaños. 18 % nació con los ojos azules, color verde es de alrededor de 4%. El 1% de las personas en todo el mundo color negro, morado claro 0.02, rojo y naranja (personas albinas) 0.03.

Cuestionario 3

1. ¿Qué tipo de muestra es la que se ha obtenido?
2. ¿Cuántos estudiantes tienen ojos azules, cuantos se esperaban?



1.2.2 Variabilidad.

Uno de los principales objetivos de los modelos estadísticos, es conocer, predecir y controlar la variabilidad de un proceso estocástico que puede tener diferentes orígenes. Los resultados en cualquier proceso de experimentación presentan diversos tipos de variabilidad cuyas características se señalarán más adelante. El conocerlos permitirá realizar una planeación para poder controlar estos factores, de manera que los procesos sean estables, consistentes y predecibles.

Actividad 9

En el grupo académico realice las siguientes encuestas, no es necesario que sean muchas preguntas, en algunos casos basta una, no es necesario desperdiciar hojas de papel, se puede utilizar el pizarrón para consultar a los estudiantes del grupo.

- a. Se debe determinar la estatura de un estudiante.
- b. Se debe determinar la edad de una persona.
- c. Se desea conocer cómo se distribuyen los porcentajes de los grupos sanguíneos.
- d. Se desea determinar la longitud de los pies para comerciar chanclas de baño unisex.
- e. Se desea conocer la distribución en porcentaje del color de ojos verdes
- f. Se desea determinar el número de refrescos de 600 ml que consume un estudiante por semana.
- g. En qué tiempo te atenderán, si te otorgan la ficha 199, al observar la pantalla lleva el número 152, la atención en cada número es en promedio 3 minutos.
- h. El consumo promedio de agua según organismo oficial de Comisión Nacional del Agua CONAGUA es de 300 litros diarios (¿Es exagerado?, tal vez se refería a un consumo promedio por familia), sin embargo así lo dijo el titular de dicho organismo frente a cámaras de un noticiero de TV.

2.2 Tipos de variabilidad.

La variabilidad es inherente al proceso de medición (y no es posible evitarlo), de esta manera si queremos medir la longitud de una mesa, podemos realizar la medición con una regla de plástico de 30 cm, con una regla de madera de un metro de longitud o con un flexómetro con una mejor graduación, por ejemplo en milímetros, al realizar este proceso con estos instrumentos obtendremos posiblemente medidas parecidas, pero no iguales, y a veces encontraremos resultados distintos al usar el mismo instrumento.

Las causas posibles de la variabilidad de los procesos de medición se clasifican en dos grandes grupos:

- i. Causas debidas a la Variabilidad sistemática y planificada
- ii. Causas aleatorias.



Actividad 10

PROCESO PARA MEDIR EL TIEMPO DE REACCIÓN DE UNA PERSONA.

Consiga una regla de 30 cm de longitud graduada en milímetros, consiga que una persona la detenga del borde, usted deberá detenerla, con sus dos dedos separados un centímetro deberá detenerla, mida que distancia recorrió la regla hasta que la detuvo, realice este proceso unas veinte veces.

1. ¿Obtuvo el mismo resultado? Explique porque si o por que no.
2. ¿Considera que el proceso se realizó bajo las mismas condiciones en cada ensayo?

Actividad 11

Consiga dos dados de plastico, láncelos a la mesa consiga que una persona los resultados realice este proceso unas veinte veces.

1. ¿Obtuvo el mismo resultado? Explique porque si o por que no.
2. ¿Considera que el proceso se realizó bajo las mismas condiciones en cada ensayo?

I. LA VARIABILIDAD SISTEMÁTICA Y PLANIFICADA.

En la variabilidad sistemática y planificada, las causas y factores que son identificables, esta variabilidad no presenta un comportamiento estadístico y, por tanto, no son previsibles. Se debe identificar las causas y eliminarlas como paso previo a establecer el proceso bajo control; ejemplos de estas causas son;

1. roturas de herramientas utilizadas en el proceso de producción.
2. averías de máquinas
3. errores humanos
4. cambios erróneos de material
5. fallos en los sistemas de control



La imagen de esta variabilidad, no será una imagen “estadística”. Cuando en un proceso se presentan una o más de estas causas de variabilidad identificable dentro del sistema de planeación, la situación se denomina fuera de control. En un experimento en particular en el proceso de medición, se debe identificar y eliminar esas causas de variabilidad, como paso previo al control del proceso. Este tipo de variabilidad se presenta por las distintas condiciones experimentales impuestas en el diseño por expreso deseo del experimentador. Es el tipo de variabilidad que se intenta identificar con el diseño estadístico. Cuando este tipo de variabilidad está presente y tiene un tamaño importante, se espera que las respuestas tiendan a agruparse formando grupos (clústeres). En los diseños de experimentos que se estudiarán posteriormente en cursos avanzados de estadística, existen dos estrategias básicas para tratar de evitar la presencia de este tipo de variabilidad: la aleatorización y la técnica de bloques.

II. VARIABILIDAD POR CAUSAS ALEATORIAS

La segunda causa de la variabilidad está originada por factores aleatorios (desgaste de piezas, mantenimiento, personas, equipos de medida, etc.); en este caso la variabilidad tiene un comportamiento estadístico y es predecible, y se puede ejercer un control estadístico sobre el mismo. Esta variabilidad es la causante de que si en un laboratorio se toman medidas repetidas de un mismo objeto ocurra que, en muchos casos, las siguientes medidas no sean iguales a la primera. Un proceso sobre el que solo actúan causas aleatorias, se dice que está bajo control.

Un proceso productivo que presenta una variabilidad exclusivamente de carácter aleatorio se caracterizará por los valores de las mediciones se suelen distribuir alrededor de un valor central que permanece aproximadamente constante a lo largo del tiempo y porque la dispersión de estos valores también permanece aproximadamente constante en el tiempo. En procesos de medición bajo control, con una cantidad muy grande de datos, por lo general estos se distribuyen siguiendo una curva en forma de campana, donde regularmente estas mediciones tienden a agruparse o distribuirse alrededor de su valor promedio, la curva se conoce en los procesos estadísticos como la distribución normal en donde la frecuencia de las mediciones va disminuyendo hacia los extremos conforme se va alejando del promedio.

1.2.3. Usos indebidos de la Estadística.

No existen usos indebidos de la estadística sino procesos de mercadotecnia y publicidad y excelente retórica que se utilizan para engañar a las personas. Un político con supuestos conocimientos en economía puede afirmar que nuestro país mejora, porque el poder adquisitivo cuando el PIB producto interno bruto aumenta cierta cantidad de dinero. (Subió el ingreso per cápita) cuando lo que sucedió es que alguna persona es el millonario número uno del mundo, es decir su fortuna aumentó, pero el resto se mantuvo igual o empobreció. Por esta razón, más adelante consideraremos que existe la necesidad de establecer una



medida que acompañe a la medida central, es decir hace falta la medida de dispersión, por otra parte. Un aumento veinte centavos en un salario mínimo, matemáticamente existe un incremento, pero esta cantidad no es significativa.

Otro ejemplo es, el anuncio en televisión afirma que nueve de cada diez gatos prefieren Whiskas, lo más probable es que uno de los gatos prefirió morir por inanición para evitar padecer de las vías urinarias por comer Whiskas.

También a través de los gráficos se engaña a las personas, Los expertos en diseño gráfico saben que es suficiente un cambio de escalas, colores o la utilización de figuras geométricas adecuadas, dan la ilusión de mostrar que el número de accidentes o de robos disminuye.

En las encuestas que se realizan en medios de comunicación (televisión), por teléfono (encuesta Mitofski) se puede determinar que no son aleatorias sino sesgadas (manipuladas), las cuales sirven de publicidad.

Recordando lo que dijo Goebbels, el ministro de propaganda Nazi, el cual a través de su puesto político controlaba, las noticias de los periódicos la radio, la televisión, cine, y la literatura, es decir los medios de comunicación más importantes de esa época; “una mentira mil veces repetida se convierte en una verdad”, que resume el poder que tienen los medios de comunicación en cuanto al poder propagandístico de manipular el consumo, el miedo, en general el pensamiento de las personas, incidiendo fuertemente en sus decisiones.

8. ACTIVIDAD FINAL

9. GLOSARIO

10. REFERENCIAS

11. CRÉDITOS