

UNIDAD 1. LOS CONCEPTOS BASICOS DE LA ESTADÍSTICA.

1.1. La Estadística y sus aplicaciones

¿Qué es la Estadística?

La Estadística es una ciencia que estudia las características de un conjunto de casos para hallar en ellos regularidades en el comportamiento, que sirven para describir el conjunto y para efectuar predicciones.

La Estadística tiene por objeto recolectar, organizar, resumir, presentar y analizar datos relativos a un conjunto de objetos, personas, procesos, etc. A través de la cuantificación y el ordenamiento de los datos intenta explicar los fenómenos observados, por lo que resulta una herramienta de suma utilidad para la toma de decisiones.

Bajo este contexto, la Estadística se divide en dos áreas: Estadística Descriptiva y Estadística Inferencial.

ESTADÍSTICA DESCRIPTIVA: permite organizar y presentar un conjunto de datos de manera que describan en forma precisa las variables analizadas haciendo rápida su lectura e interpretación. Obviamente, la materia prima de la Estadística Descriptiva la constituyen los datos, que son el resultado de las observaciones y/o experimentos.

Ejemplos:

1. Durante los últimos dos días se ha informado de un total de trece homicidios diarios.
2. La encuesta Gallup informa una ventaja de 25% para el candidato de izquierda.

ESTADÍSTICA INFERENCIAL: Generaliza los resultados de una muestra a los de una población total, es cuando de los datos estadísticos obtenidos de una muestra se infiere o se deduce una observación la cual se generaliza sobre la población en total. Para determinar la confiabilidad de la inferencia de los datos estadísticos de una muestra, se hace necesario comprobar la misma para poder asegurar que lo que se observa en una muestra se observará también en la población. Generalmente el análisis estadístico inferencial se lleva a cabo para mostrar relaciones de causa y efecto, así como para probar hipótesis y teorías científicas.

IMPORTANCIA DE LA ESTADISTICA.

En las últimas décadas la estadística ha alcanzado un alto grado de desarrollo, hasta el punto de incursionar en la totalidad de las ciencias; inclusive, en la lingüística se aplican técnicas estadísticas para esclarecer la paternidad de un escrito o los caracteres más relevantes de un idioma.

La estadística es una ciencia auxiliar para todas las ramas del saber; su utilidad se entiende mejor si tenemos en cuenta que los quehaceres y decisiones diarias embargan cierto grado

de incertidumbre... y la Estadística ayuda en la incertidumbre, trabaja con ella y nos orienta para tomar las decisiones con un determinado grado de confianza.

Los críticos de la estadística afirman que a través de ella es posible probar cualquier cosa, lo cual es un concepto profano que se deriva de la ignorancia en este campo y de lo polifacético de los métodos estadísticos. Sin embargo muchos "investigadores" tendenciosos han cometido abusos con la estadística, elaborando "investigaciones" de intención, teniendo previamente los resultados que les interesan mostrar a personas ingenuas y desconocedoras de los hechos. Otros, por ignorancia o negligencia, abusan de la estadística utilizando modelos inapropiados o razonamientos ilógicos y erróneos que conducen al rotundo fracaso de sus investigaciones

Aplicaciones de la Estadística.

Mucha gente piensa que la Estadística no tiene nada que ver con otras disciplinas que no sean las ingenierías y economía. Otros nunca le encuentran aplicaciones útiles, y por eso tampoco les gusta. Pero, en realidad tiene infinitas aplicaciones en todo el conocimiento adquirido por la humanidad, partiendo ante todo, lo relacionado con las ingenierías, economía, las ciencias biológicas, ciencias sociales e incluso en algunas ramas del área Jurídica. Y, en definitiva, casi todos los campos de las ciencias emplean instrumentos estadísticos de importancia fundamental para el desarrollo de sus modelos de trabajo.

En nuestros días, la estadística se ha convertido en un método efectivo para describir con exactitud los valores de los datos económicos, políticos, sociales, psicológicos, biológicos y físicos, entre otros, y sirven como herramienta para relacionar y analizar dichos datos. El trabajo del experto estadístico no consiste ya sólo en reunir y tabular los datos, sino el proceso de interpretación de esa información a través de modelos estadísticos-matemáticos, aumentando el alcance de las aplicaciones de la estadística, esto se ha logrado, gracias al desarrollo de la teoría de probabilidad.

Dentro de las Aplicaciones de la Estadística se destacan las siguientes:

1. La Estadística en el Periodismo

En general, los periodistas además de dedicarse al ámbito de la noticia, realizan crónicas y estudios de investigación, que nos entregan preguntas y respuestas frente a determinados sucesos o situaciones de interés público. Algunos de los estudios más frecuentes realizados por los periodistas son sobre alcoholismo, enfermedades, sexualidad, delincuencia, política, etc. Para ello, hacen uso de las encuestas u otros instrumentos técnicos de medición propios de la estadística, a través de dichos estudios es posible conocer la opinión de la gente y con ello informar a la opinión pública, a través de los medios de comunicación, desde donde las autoridades pertinentes e interesadas en estos estudios pueden adoptar las medidas correctivas, si es el caso. Tal es así, que la estadística forma parte importante del periodismo investigativo.

2. La Estadística en la Política

Conocidas son las famosas encuestas de tipo político, que entregan una orientación de la intención de voto, de la aceptación de un candidato, del impacto de un programa o proyecto de estado, etc. de una muestra estadística representativa, sobre la opinión de las personas en un tiempo determinado, teniendo esta herramienta una gran confiabilidad. Así es que el uso de la estadística es imprescindible para determinar caminos a seguir para los candidatos de elección popular.

3. La Estadística en la Publicidad

Cuando las grandes marcas trasnacionales y/o nacionales como Coca-cola, Pepsi, Nice, Adidas, Laboratorios López, etc. nos llenan de slogans, música y colores en sus comerciales, lo único que buscan es que la gente adquiera los productos y/o servicios que ofrecen.

Se dedican, entonces, a realizar las llamadas “campañas publicitarias”, y, antes de lanzar una campaña, hacen un estudio de mercado para encontrar las mejores alternativas posibles a fin de lograr el éxito de ventas deseado. Estos estudios son de carácter estadístico, es decir, hacen un diseño muestral y seleccionan una muestra para inferir las características de la población.

4. La Estadística en la economía y las finanzas.

En la administración es una herramienta del control, como parte del proceso administrativo (o lo que es lo mismo: planeación, organización, dirección y control) ya que la estadística ayuda a recolectar, estudiar y al final interpretar los datos que obtienen al terminar el proceso administrativo, retroalimenta con esta información y al final se observa en que pueden mejorar y que se está haciendo bien.

En la mercadotecnia es una herramienta muy importante cuando tienes la necesidad de trabajar con datos de muestreo para conocer los gustos y preferencias de las personas. Igualmente en la auditoría administrativa cuando recabas datos para conocer en que puede mejorar una organización. En pocas palabras te puede servir en cualquier área de una organización debido a que muestra los resultados de las actividades que bien se están realizando.

En la economía se utiliza como una herramienta de predicción para pronosticar el comportamiento futuro, por ejemplo, de los precios de los metales (oro, plata, cobre) tomando como base el comportamiento pasado de los precios de los mismos. También puede servir para estudiar el comportamiento de la bolsa de valores, de ciertos productos básicos, los economistas por lo regular se sienten magos que creen predecir cosas. En general, la Estadística suministra los valores que ayudan a descubrir interrelaciones entre múltiples parámetros macro y microeconómicos, a través de la evaluación de modelos econométricos para el establecimiento de políticas económicas; análisis del costo de la canasta básica, el poder adquisitivo de la población, etc.

5. La estadística en la Banca y Seguros

El profesional del seguro de vida ha de ser capaz de asignar primas suficientes para cubrir las cantidades que habrá de pagar la compañía en el caso de muerte del asegurado. En consecuencia, la predicción adecuada de las probabilidades de muerte constituye uno de los ejes centrales de la reducción del riesgo que se asume. Por ello, el objetivo de la Estadística de Seguros es una presentación exhaustiva de los métodos disponibles para ajustar tablas de mortalidad y tablas de seguros no vida, ejemplo, aseguramiento de vehículos, viviendas, etc.

Por otra parte, algunas de las aplicaciones concretas de la Estadística en el sistema bancario son las siguientes:

- Sistemas de concesión de tarjetas de crédito y fijación de su límite.
- Sistemas de estimación del potencial económico de los clientes.
- Definición de tipologías comerciales de clientes.
- Determinación del público objetivo en campañas comerciales.
- Modelización del riesgo según las características de los clientes.
- Aplicación de la teoría de colas para brindar un servicio de calidad.
- Finalmente, es de mucha utilidad la técnica de minería de datos para el análisis de bases de datos del sistema bancario.

6. La estadística en ciencias humanas y sociales

La investigación en Psicología, Sociología y Educación, al igual que ocurre en otras ciencias, en buena medida se basa en el manejo de recursos estadísticos como elementos indispensables para llegar a conclusiones aceptables por el resto de la comunidad científica. Dada la peculiaridad de su objeto de estudio, inabordable en la mayoría de los casos sino es a través de perspectivas complejas de relación entre variables, la atención de los investigadores en las ciencias humanas y sociales se concentra cada vez más en la llamada Estadística Multivariante (Análisis Cluster, Factorial, Discriminante, etc.).

Las ciencias sociales se han visto apabulladas en los últimos años por avances vertiginosos en informática y aplicaciones estadísticas, por ejemplo, en El Salvador, se ha elaborado el Mapa de Pobreza, donde se ha interrelacionado una gran variedad de variables de diferentes áreas: Educativas, Económicas, Salud, entre otras. También, se ha hecho investigación sobre los factores que están asociados al rendimiento académico de los estudiantes, finalmente, se han utilizado diferentes técnicas para el análisis de los resultados académicos de los estudiantes.

Las ciencias sociales: es un pilar básico del desarrollo de la demografía y la sociología aplicada, lo que conlleva a:

- Definición de indicadores de fenómenos sociales.
- Medición de constructos o variables no directamente observables (la satisfacción, la inteligencia, ...)

- Medición de los efectos entre constructos no observables para establecer políticas sociales.
- Estudio de la evolución de la demografía.
- Estudios sociales sobre la integración de la población inmigrada.
- Fenómenos sociales como las pandillas, criminalidad, delincuencia, contaminación, entre otros.

7. La estadística en las ciencias químicas.

En Química, la estadística se aplica en varias áreas: En el diseño de experimentos se usan métodos estadísticos, en el control de procesos y control de calidad (o gerenciamiento de calidad) de procesos y productos. En EEUU está muy de moda el sistema seis sigma, creado por general electric, que utiliza algunos conceptos estadísticos para lograr el aseguramiento de la calidad.

Por otra parte la producción química tiene su costo económico y financiero que también requiere mucho uso de estadística, por ejemplo, si no se elabora un buen diseño experimental, se necesitarán muchas réplicas para validar una formulación química.

Algunas de las aplicaciones concretas que podemos mencionar:

- Utilización de diseños experimentales para optimizar la composición de productos alimenticios.
- Evaluación de la superficie de respuesta de una reacción química según determinados factores.
- Predicción del comportamiento de un componente no sintetizado a partir de las propiedades moleculares de sus descriptores.
- Control de procesos de producción para detectar problemas evitando a su vez falsas alarmas.

8. La estadística en Ciencias biológicas.

En el área de las ciencias biológicas, interesa estudiar el comportamiento de ciertas plantas y sus cruces a fin de determinar cómo se relacionan genéticamente los padres con los hijos, hablando de Genotipo y Fenotipo. En esta categoría es también donde se realizan los mayores avances de la humanidad, en descubrimientos. Cada año se descubren miles de fórmulas científicas que relacionan fenómenos de la naturaleza con modelos matemáticos.

Los científicos se dedican a realizar estudios estadísticos, recogiendo datos y muestras, investigando el tiempo de reproducción de un virus, el comportamiento migratorio de algunas aves o insectos, además de factores de tamaño y volumen del crecimiento de ciertas especies de animales o vegetales. Todo esto funciona con la idea de recopilar información, muestrear ciertas áreas para ver cómo se han comportado algunas aves, por ejemplo, se pueden dibujar o simular curvas que se supone que son relativamente parecidas al comportamiento migratorio de aves. Con esta herramienta se podrían determinar también

las épocas de mayor probabilidad de contagio, diseminación de algún virus o bien enfermedades transmitidas por insectos.

Algunas aplicaciones concretas en esta área son: Determinación del tamaño de poblaciones naturales en una región; efectividad de la utilización de barreras naturales (filas de árboles plantados en los límites del terreno) como medio de prevenir las plagas de insectos o aves sobre las plantaciones y así disminuir la utilización de pesticidas; y, determinación de los niveles óptimos de utilización de los fertilizantes. Obviamente, en esta área es donde más se hace usos de la teoría de Diseños de Experimentos.

9. La estadística en las ciencias médicas.

Permite establecer pautas sobre la evolución de las enfermedades y los enfermos, los índices de mortalidad asociados a procesos morbosos, el grado de eficacia de un medicamento, etcétera.

Bajo este conjunto de ideas, se ha presentado un panorama de la utilidad de la estadística, haciendo un recorrido por diversas áreas del conocimiento humano, con el fin, de conocer cómo se relacionan con las diversas ciencias, formando una sola verdad. Evidentemente, existen, muchas disciplinas donde se aplica la estadística, que no han sido consideradas en este apartado, por ejemplo, las ingenierías, la pintura, la música, etc. Finalmente, es de resaltar que en todos los temas donde se analice información está presente la estadística.

1.2. Conceptos Básicos

VARIABLES

Las variables son las características observables de un objeto, problema o evento que se puede describir según un esquema de medición bien definido. Cada rasgo o aspecto de una población constituye una variable. La edad de unas personas, su sexo, color de su piel, nacionalidad, su nivel de motivación, niveles de ansiedad, el número de nacimientos, número de matrimonios, frecuencia de suicidios, estatura, peso, niveles de inteligencia, actitudes, entre muchas otras.

Escalas de medición de las variables.

1. Escala Nominal

Consiste en clasificar en categorías a los sujetos u objetos, de modo que todos los sujetos u objetos clasificados dentro de una misma categoría sean equivalentes respecto a la variable o propiedad que se está midiendo. Tras esto, se asignan símbolos o números a las categorías establecidas. Las categorías utilizadas (que serán tantas como niveles o categorías tenga la variable que se desea medir) deben reunir tres propiedades: exhaustividad (cada sujeto u objetos pueden ser clasificado en alguna de las categorías establecidas), complementarias (todos los sujetos u objetos pueden ser clasificados y cada uno de ellos será el complemento del conjunto universo) y exclusividad (cada sujeto u objeto pueden ser clasificados en sólo una de las categorías establecidas; las categorías no se repiten).

Un caso particular, sería la variable color de ojos con las siguientes categorías establecidas **café, azul, verde, negro**. A cada uno se le puede asignar símbolos que servirá como etiquetas para sustituir los nombres que permiten facilitar la estructura de una base de datos y los análisis respectivos.




Y podemos usar las siguientes etiquetas: A = café, B = azul, C = verde, D = negro; también podría ser 1 = café, 2 = azul, 3 = verde, 4 = negro.

2. Escala Ordinal

Sea X una característica observable en una población. Diremos que X está medida en la escala ordinal si, además de la relación de igualdad, la única relación que podemos establecer entre las modalidades es la relación de orden

Consiste en asignar a los sujetos u objetos medidos un número que permite ordenarlos según la cantidad en la variable que poseen. En la escala ordinal, además de estar presente la relación de igualdad- desigualdad propia de la escala nominal, los números asignados permiten afirmar si la cantidad de variable que posee un sujeto u objeto es mayor que o menor que la cantidad de variable que posee otro sujeto u objeto cualquiera.

En las ciencias sociales y de la salud es frecuente encontrarse con variables en las que resulta apropiado utilizar una escala de medida ordinal por ejemplo: Cuando se realizó el Mapa de Pobreza en El Salvador¹ en el año 2004 se clasificó la pobreza del país de la siguiente manera

-  No Pobres
-  Pobreza Relativa
-  Pobreza Extrema

Se puede ordenar en forma ascendente o descendente, pero no se puede saber si la diferencia entre dos valores es la misma o diferente entre otros dos valores.

Aquí no hay sentido operaciones algebraicas tales como $x_i + x_j$ ó $x_i - x_j$. **Sólo tienen sentido las relaciones de igualdad y orden.**

3. Escala de intervalos

Este tipo de escala posee las características de ser nominal y ordinal; además la principal característica que la define es que se puede determinar la magnitud de la diferencia existente entre dos objetos medidos, los objetos de la variable se pueden distribuir en intervalos igualmente espaciados, es decir, la unidad de medida está claramente determinada y se asigna a cada objeto medido un número indicativo que lo ubica en uno y solo un intervalo.

¹

En la escala de intervalo no existe el cero absoluto, es decir, no existe un valor numérico que indique ausencia total de cantidad de variable

La temperatura, por ejemplo, es una variable que se mide utilizando una escala de intervalo. Cuando se dice, en escala Celsius, que ayer hubo 20 grados de temperatura máxima y hoy 25, se está diciendo no sólo que hoy hubo más temperatura que ayer (afirmación propia de la escala ordinal), sino que hoy hubo 5 grados más de temperatura que ayer. Del mismo modo, 20 grados son 5 más que 15. La diferencia entre 15 y 20 grados es la misma que entre 20 y 25, y esto va más allá de lo que puede afirmarse con una escala ordinal. Sin embargo, no es posible afirmar que 20 representen el doble de temperatura que 10. Esto es debido a que, en la escala Celsius, el punto cero es un punto arbitrario de la escala y, por tanto, no indica ausencia de cantidad de variable.

4. Escala de Razón

Diremos que una variable esta medida en una escala de razón si a través de un proceso de medición es posible asignar valores numéricos (valores reales) a la variable, en los que el cero representa la ausencia de dicha variable. Este tipo de escala posee todas las características de las escalas anteriores, es decir, es la escala más completa de medición.

Ejemplos de escala de razón son las siguientes: el tiempo, la extensión, el peso, la longitud, la masa, la intensidad de la corriente eléctrica y otras variables del mundo físico

Tipos de variables:

Las variables pueden adquirir diferentes valores o clasificarse en diferentes categorías según la naturaleza o tipo de estudio que se lleve a cabo. Entre éstas tenemos las siguientes clasificaciones:

VARIABLES CUALITATIVAS - son aquellas que se expresan en forma verbal como categorías o atributos y no pueden ordenarse. Por ejemplo, el sexo, color, afiliación política, nacionalidad, motivación, área académica o profesión de una persona. Estas a su vez se clasifican en ***Dicotómicas*** cuando tienen solo dos categorías y ***politómicas*** cuando hay más de dos categorías

VARIABLES CUANTITATIVAS - son las que varían en términos de cantidad y se registran o expresan en forma numérica. Por ejemplo, edad, promedio académico, puntuaciones de exámenes, frecuencia de delitos, temperatura, ingresos anuales o salarios por hora. Hay algunas características que pueden clasificarse o expresarse como variable cuantitativa y transformarla a cualitativa o viceversa. Por ejemplo, nivel de aprovechamiento académico estudiantes de 4:00 puntos, o estudiantes de 3:00 puntos y así

sucesivamente. El investigador puede expresar mediante una escala numérica el aprovechamiento académico al clasificar a los estudiantes, como también puede clasificarlos como variable ordinal en las categorías de excelentes, buenos, regulares y deficientes.

Las VARIABLES CUANTITATIVAS se dividen en:

VARIABLE DISCRETA – cuando el conjunto de valores que toma pueden ser finito o infinito numerable.

VARIABLE CONTINUA - Son las que pueden asumir cualquier valor en un intervalo real. Por ejemplo, la edad, altura, peso, etc.

En el campo de la investigación, que se suele examinar las relaciones entre dos o más variables al investigar un asunto o problema, se clasifican las variables como:

VARIABLES INDEPENDIENTES - son las características controladas por el investigador y que se supone tendrán efectos sobre otras variables.

VARIABLES DEPENDIENTES - son las características o aspectos que se alteran por consecuencia del control que ejerce el investigador sobre otras variables.

POBLACIÓN - es el conjunto total de individuos, objetos o medidas que poseen algunas características comunes observables en un lugar y en un momento determinado. Cuando se vaya a llevar a cabo alguna investigación debe de tenerse en cuenta algunas características esenciales al seleccionarse la población bajo estudio. Entre éstas tenemos:

Homogeneidad - que todos los miembros de la población tengan las mismas características según las variables que se vayan a considerar en el estudio o investigación. Por ejemplo, si se fuera a investigar la incidencia de la drogadicción entre jóvenes mujeres adolescentes, entonces hay que definir claramente las edades que comprenden la adolescencia y cuando se seleccione la población asegurarse de que todas las personas entrevistadas sean de la edad determinada y del sexo femenino. (La adolescencia se define operacionalmente como el periodo comprendido de edad que fluctúa entre 12 y 21 años.)

Tiempo - se refiere al período de tiempo donde se ubicaría la población de interés. Determinar si el estudio es del momento presente o si se va a estudiar a una población de cinco años atrás o si se van a entrevistar personas de diferentes generaciones.

Espacio - se refiere al lugar donde se ubica la población de interés. Un estudio no puede ser muy abarcador y por falta de tiempo y recursos hay que limitarlo a un área o comunidad en específico.

Cantidad - se refiere al tamaño de la población. El tamaño de la población es sumamente importante porque ello determina o afecta al tamaño de la muestra que se vaya a seleccionar, además que la falta de recursos y tiempo también nos limita la extensión de la población que se vaya a investigar.

Censo: En ocasiones resulta posible estudiar cada uno de los elementos que componen la población, realizándose lo que se denomina un censo, es decir, el estudio de todos los elementos que componen la población.

Parámetro: Cualquier medida matemática o función cuyo valor depende de la información de toda la población.

MUESTRA - La muestra es un subconjunto fielmente representativo de la población.

Hay diferentes tipos de muestreo. El tipo de muestra que se seleccione dependerá de la calidad y cuán representativo se quiera que sea el estudio de la población, en este sentido, la muestra a seleccionar, es recomendable que sea:

MUESTRA ALEATORIA SIMPLE (MAS): Cuando se selecciona al azar cada miembro y cada miembro en la población tiene igual oportunidad de ser incluido en la muestra. Esta selección puede ser CON REEMPLAZO o SIN REEMPLAZO

Muestreo con reemplazo: Es aquel en que un elemento puede ser seleccionado más de una vez en la muestra para ello se extrae un elemento de la población se observa y se devuelve a la población, por lo que de esta forma se pueden hacer infinitas extracciones de la población aun siendo esta finita.

Muestreo sin reemplazo: No se devuelve los elementos extraídos a la población hasta que no se hallan extraídos todos los elementos de la población que conforman la muestra.

Un **procedimiento** de extraer una muestra aleatoria simple de una población finita es: enumerar todos los elementos que conforman la población, escribir esos números en papelitos y echarlos en una urna o bolsa mezclarlos bien removiéndolos y sacar uno a uno tantos como lo indique el tamaño de la muestra. En este caso los elementos de la muestra lo constituirán los elementos de la población cuyos números coincidan con los extraídos de la bolsa o urna.

Estadístico o estadígrafo: Cualquier medida matemática o función que depende únicamente de al menos un valor de la muestra.

MARCO MUESTRAL

Una vez definido el universo, se debe recabar información, lo más exacta posible, de sus dimensiones y distribución espacial y temporal, para con ello poder construir el marco muestral, que es la base para hacer el diseño de muestreo. El marco muestral es la información que ubica y dimensiona al universo y puede consistir de censos de vivienda y mapas agrupados por localidades, barrios, repartos, etc.; mapas de cobertura forestal con agrupamientos por tipos de vegetación o usos del suelo; listados de viviendas en localidades pequeñas, etc.

UNIDAD DE MUESTREO

Un concepto básico en la teoría del muestreo, que debe definirse claramente para construir el marco muestral, es la *unidad de muestreo*, que es la unidad mínima de observación de la que se obtendrá información de las variables útiles. Por convención estadística, se usará "N" para referirnos al número de unidades de muestreo que integran el Universo y "n" para el número de unidades de muestreo en la muestra. En la siguiente tabla, para cada sector la unidad de muestreo más conveniente se indica:

Grupo	Sector o ramo	Unidad de muestreo
Demanda	Residencial	Vivienda
	- urbana	
	- rural	
	Industrial	Establecimiento
	Comercial	
	Institucional	
Oferta	Directa	Parcela
	Indirecta	Establecimiento
Abastecimiento	Productores	Productores individuales, empresas
	Transportistas	
	Comercializadores	

UNIDAD DE ANÁLISIS

Cada una de las entidades a la cual se refiere un dato determinado en un instante dado respecto de una característica en estudio se denomina Unidad de Análisis. Expresado más coloquialmente, la unidad de análisis es cada una de las entidades que presenta un valor de alguno de los atributos observados en el estudio

Tamaños de muestra en poblaciones homogéneas.

Para el muestreo aleatorio simple(MAS):

Tamaño muestral n para estimar una media en una población de tamaño N y asumiendo normalidad para \bar{x} (Teorema central de limite)

$$n = \frac{\left(\frac{z_{1-\alpha/2} S_x}{d} \right)^2}{1 + \left(\frac{z_{1-\alpha/2} S_x}{d} \right)^2 \frac{1}{N}}, \quad S_x \text{ se obtiene de una muestra piloto de tamaño } n', \text{ si } n > n', \text{ se}$$

complementa la muestra piloto, si no, la muestra piloto es suficiente, el valor d es el error

máximo de estimación ($d \geq |\bar{x} - \mu|$), los valores de z con confianza del 95%, 90%, 99% son $z = 1.96, 1.645, 2.576$ respectivamente.

Si N es grande ($N \rightarrow \infty$): $n = \left(\frac{z_{1-\alpha/2} S_x}{d} \right)^2$

Si no se admite normalidad, entonces se utiliza Tchebycheff:

$$n = \frac{\left(\frac{S_x}{d\sqrt{\alpha}} \right)^2}{1 + \left(\frac{S_x}{d\sqrt{\alpha}} \right)^2 \frac{1}{N}}$$

Donde $\alpha = 1\%, 5\%, 10\%$ son errores permitidos de probabilidad.

Tamaño muestral n para estimar una proporción P en una población de tamaño N con un error máximo de d porcentual ($|p - P| \leq d$) y asumiendo normalidad en p .

$$n = \frac{(d^2 + pqz_{1-\alpha/2}^2)N}{d^2N + pqz_{1-\alpha/2}^2}$$

Donde d es el error porcentual de estimación y p es la proporción con éxito en la muestra y $q = 1 - p$ y N es el tamaño de la población.

Tamaños de muestra en poblaciones no homogéneas

Muestreo estratificado con asignación proporcional. la muestra se reparte entre los estratos proporcionalmente a los tamaños de éstos. Este tipo de asignación se utiliza cuando los costos y las varianzas de los estratos no son muy diferentes.

Tamaño de muestra para estimar el promedio poblacional con asignación proporcional.

$$n = \frac{\sum_{k=1}^L N_k \hat{S}_k^2}{N \frac{B^2}{k^2} + \frac{1}{N} \sum_{k=1}^L N_k \hat{S}_k^2} \quad (6.26)$$

Donde: B = error de estimación

k = percentil que se halla en la tabla de la distribución normal y depende del nivel de confianza

Una vez determinado el tamaño de la muestra, para repartirla proporcionalmente al tamaño de los estratos se utiliza la siguiente expresión:

$$n_h = n \frac{N_h}{N} \quad (6.27)$$

EJEMPLO

Se desea estimar la nota promedio de los estudiantes de administración de empresas diurna y nocturna en una universidad. En la carrera diurna (estrato 1) hay 280 estudiantes y en la nocturna (estrato 2) hay 200 estudiantes. Determine el tamaño de muestra necesario para cumplir el objetivo con un error máximo de 0,15 y una confiabilidad del 95 por ciento.

Por un estudio realizado tiempo atrás se conocen las varianzas de las notas de administración diurna y nocturna, las que respectivamente son: 0,31 y 0,28.

Solución

Considerando que las varianzas son similares, se trabaja con muestreo estratificado con asignación proporcional. El error (B) es 0,15 y para una confiabilidad del 95 por ciento el valor correspondiente en la distribución normal es 1,96, entonces, k = 1,96:

$$N_1 = 280 \quad N_2 = 200 \quad N = 480 \quad s_1^2 = 0,31 \quad s_2^2 = 0,28$$

Para hallar el tamaño de muestra se utiliza la ecuación 6.26

$$n = \frac{280(0,31) + 200(0,28)}{480 \frac{0,15^2}{1,96^2} + \frac{1}{480} [280(0,31) + 200(0,28)]} = 45,93$$

El tamaño de la muestra es de 46 estudiantes. Esta muestra se reparte proporcionalmente al tamaño de los estratos, con la ecuación 6.27

$$n_1 = 46 \frac{280}{480} = 26,83 \quad n_2 = 46 \frac{200}{480} = 19,17$$

Se deben seleccionar 27 estudiantes de administración de empresas diurna y 19 de la nocturna.

Tamaño de muestra para estimar la proporción con asignación proporcional

$$n = \frac{\sum_{k=1}^L N_k p_k q_k}{N \frac{B^2}{k^2} + \frac{1}{N} \sum_{k=1}^L N_k p_k q_k} \quad (6.29)$$

Una vez determinado el tamaño de la muestra, se reparte entre los estratos utilizando la expresión 6.27

EJEMPLO

Tomado de¹. En vista de la recesión económica existente, una empresa textil pretende reducir el número de días laborables por semana a cuatro. Otra alternativa consiste en clausurar una de sus tres plantas y despedir a los trabajadores. Para tener una idea de la opinión de los trabajadores, el gerente de personal de la empresa desea seleccionar una muestra de empleados de las tres plantas para estimar la proporción de trabajadores que prefieren la reducción de la semana de trabajo, con un error de estimación máximo de 0,1.

La empresa emplea 150 personas en la planta 1, 65 en la planta 2 y 40 en la 3. Se estima que cerca del 75 por ciento de los de la planta tres están a favor de la reducción de la semana de trabajo, mientras que en las otras plantas este porcentaje parece corresponder al 50 por ciento. Encuentre el tamaño de muestra y la asignación necesaria en cada estrato.

Solución

$$\begin{aligned} N_1 &= 150 & N_2 &= 65 & N_3 &= 40 & N &= 255 \\ p_1 &= 0,5 & p_2 &= 0,5 & p_3 &= 0,75 & B &= 0,1 \end{aligned}$$

Por la diferencia en el tamaño de las plantas, se utiliza el muestreo estratificado con asignación proporcional.

Asumiendo un nivel de confianza del 95 por ciento, el valor correspondiente en la distribución normal es 1,96 ($k=1,96$).

Para determinar el tamaño de la muestra se utiliza la ecuación 6.29 y se reparte en los estratos con la ecuación 6.27

$$n = \frac{150(0,5)(0,5) + 65(0,5)(0,5) + 40(0,75)(0,25)}{255 \frac{0,1^2}{1,96^2} + \frac{1}{255} [150(0,25) + 65(0,25) + 40(0,1875)]} = 67,76$$

$$n_1 = 68 \frac{150}{255} = 40 \quad n_2 = 68 \frac{65}{255} = 17,33 \quad n_3 = 68 \frac{40}{255} = 10,67$$

ETAPAS DEL MÉTODO ESTADÍSTICO

Las etapas son las siguientes:

1. PLANTEAMIENTO DEL PROBLEMA

Al abordar una investigación se debe tener bien definido qué se va a investigar y por qué se pretende estudiar algo. Es decir, se debe establecer una delimitación clara, concreta e inteligible sobre el o los fenómenos que se pretenden estudiar, para lo cual se deben tener en cuenta, entre otras cosas, la revisión bibliográfica del tema, para ver su accesibilidad y consultar los resultados obtenidos por investigaciones similares, someter nuestras proposiciones básicas a un análisis lógico; es decir, se debe hacer una ubicación histórica y teórica del problema.

2. FIJACIÓN DE LOS OBJETIVOS

Luego de tener claro lo que se pretende investigar, Debemos presupuestar hasta dónde queremos llegar; en otras palabras, debemos fijar cuales son nuestras metas y objetivos. Estos deben plantearse de tal forma que no haya lugar a confusiones o ambigüedades y debe, además, establecerse diferenciación entre lo de corto, mediano y largo plazo, así como entre los objetivos generales y los específicos.

3. FORMULACIÓN DE LAS HIPÓTESIS.

Una hipótesis es ante todo, una explicación provisional de los hechos objeto de estudio, y su formulación depende del conocimiento que el investigador posea sobre la población investigada. Una hipótesis estadística debe ser susceptible de docimar, esto es, debe poderse probar para su aceptación o rechazo.

Una hipótesis que se formula acerca de un parámetro (media, proporción, varianza, etc.), con el propósito de rechazarla, se llama Hipótesis de Nulidad y se representa por H_0 ; a su hipótesis contraria se le llama Hipótesis Alternativa (H_1)

4. DEFINICIÓN DE LA UNIDAD DE OBSERVACIÓN Y DE LA UNIDAD DE MEDIDA

La Unidad de Observación, entendida como cada uno de los elementos constituyentes de la población estudiada, debe definirse previamente, resaltando todas sus características; pues, al fin de cuentas, es a ellas a las que se les hará la medición.

La unidad de observación puede estar constituida por uno o varios individuos u objetos y denominarse respectivamente simple o compleja.

El criterio sobre la unidad de medición debe ser previamente definido y unificado por todo el equipo de investigación. Si se trata de medidas de longitud, volumen, peso, etc., debe establecerse bajo qué unidad se tomarán las observaciones ya sea en metros, pulgadas, libras, kilogramos, etc.

Asociado a la unidad de medida, deben establecerse los criterios sobre las condiciones en las cuales se ha de efectuar la toma de la información.

5. DETERMINACIÓN DE LA POBLACIÓN Y DE LA MUESTRA

Estadísticamente, la población se define como un conjunto de individuos o de objetos que poseen una o varias características comunes. No se refiere esta definición únicamente a los seres vivos; una población puede estar constituida por los habitantes de un país o por los peces de un estanque, así como por los establecimientos comerciales de un barrio o las unidades de vivienda de una ciudad.

Existen desde el punto de vista de su manejabilidad poblaciones finitas e infinitas. Aquí el término infinito no está siendo tomado con el rigor semántico de la palabra; por ejemplo, los peces dentro de un estanque son un conjunto finito; sin embargo, en términos estadísticos, puede ser considerado como infinito.

Muestra es un subconjunto de la población a la cual se le efectúa la medición con el fin de estudiar las propiedades del conjunto del cual es obtenida. En la práctica, estudiar todos y cada uno de los elementos que conforman la población no es aconsejable, ya sea por la poca disponibilidad de recursos, por la homogeneidad de sus elementos, porque a veces es necesario destruir lo que se está midiendo, por ser demasiado grande el número de sus componentes o no se pueden controlar; por eso se recurre al análisis de los elementos de una muestra con el fin de hacer inferencias respecto al total de la población. Existen diversos métodos para calcular el tamaño de la muestra y también para tomar los elementos que la conforman, pero no es el objetivo de este curso estudiarlos. Diremos solamente que la muestra debe ser representativa de la población y sus elementos escogidos al azar para asegurar la objetividad de la investigación.

6. LA RECOLECCIÓN

Una de las etapas más importantes de la investigación es la recolección de la información, la cual ha de partir, a menos que se tenga experiencia con muestras análogas, de una o varias muestras piloto en las cuales se pondrán a prueba los cuestionarios y se obtendrá una aproximación de la variabilidad de la población, con el fin de calcular el tamaño exacto de la muestra que conduzca a una estimación de los parámetros con la precisión establecida.

El establecimiento de las fuentes y cauces de información, así como la cantidad y complejidad de las preguntas, de acuerdo con los objetivos de la investigación son decisiones que se han de tomar teniendo en cuenta la disponibilidad de los recursos

financieros, humanos y de tiempo y las limitaciones que se tengan en la zona geográfica, el grado de desarrollo, la ausencia de técnica, etc.

Es, entonces, descubrir dónde está la información y cómo y a qué "costo" se puede conseguir; es determinar si la encuesta se debe aplicar por teléfono, por correo, o si se necesitan agentes directos que recojan la información; establecer su número óptimo y preparar su entrenamiento adecuado.

7. CRITICA, CLASIFICACIÓN Y ORDENACIÓN

Después de haber reunido toda la información pertinente, se necesita la depuración de los datos recogidos. Para hacer la crítica de una información, es fundamental el conocimiento de la población por parte de quien depura para poder detectar falsedades en las respuestas, incomprensión a las preguntas, respuestas al margen, amén de todas las posibles causas de nulidad de una pregunta o nulidad de todo un cuestionario.

Separado el material de "desecho" con la información depurada se procede a establecer las clasificaciones respectivas y con la ayuda de hojas de trabajo, en las que se establecen los cruces necesarios entre las preguntas, se ordenan las respuestas y se preparan los modelos de tabulación de las diferentes variables que intervienen en la investigación.

El avance tecnológico y la popularización de los computadores hacen que estas tareas, manualmente dispendiosas, puedan ser realizadas en corto tiempo.

8. LA TABULACIÓN.

Una tabla es un resumen de información respecto a una o más variables, que ofrece claridad al lector sobre lo que se pretende describir; para su fácil interpretación una tabla debe tener por lo menos: Un título adecuado el cual debe ser claro y conciso. La Tabla propiamente dicha con los correspondientes subtítulos internos y la cuantificación de los diferentes ítems de las variables, y las notas de pie de cuadro que hagan claridad sobre situaciones especiales de la tabla, u otorguen los créditos a la fuente de la información.

9. LA PRESENTACIÓN

Una información estadística adquiere más claridad cuando se presenta en la forma adecuada. Los cuadros, tablas y gráficos facilitan el análisis, pero se debe tener cuidado con las variables que se van a presentar y la forma de hacerlo. No es aconsejable saturar un informe con tablas y gráficos redundantes que, antes que claridad, crean confusión.

Además la elección de determinada tabla o gráfico para mostrar los resultados, debe hacerse no sólo en función de las variables que relaciona, sino del lector a quien va dirigido el informe.

10. EL ANÁLISIS.

La técnica estadística ofrece métodos y procedimientos objetivos que convierten las especulaciones de primera mano en aseveraciones cuya confiabilidad puede ser evaluada y ofrecer una premisa medible en la toma de una decisión.

Es el análisis donde se cristaliza la investigación. Esta es la fase de la determinación de los parámetros y estadísticos muestrales para las estimaciones e inferencias respecto a la población, el ajuste de modelos y las pruebas de las hipótesis planteadas, con el fin de establecer y redactar las conclusiones definitivas.

11. PUBLICACIÓN

Toda conclusión es digna de ser comunicada a un auditorio. Es más, hay otros estudiosos del mismo problema a quienes se les puede aportar información, conocimientos y otros puntos de vista acerca de él.