

CONCEPTOS BÁSICOS DE PROBABILIDAD Y ESTADÍSTICA

Teoría Probabilística: Es la rama de las Matemáticas que proporciona los fundamentos, modelos matemáticos y el lenguaje que se usa en la Estadística. Se encarga de modelar matemáticamente la incertidumbre o aleatoriedad de ciertas características de un fenómeno de interés.

Probabilidad: medida de la certidumbre que se le asocia a la ocurrencia u observación de un fenómeno o al hecho de que una característica de interés tome cierto valor.

Estadística: rama de la ciencia que estudia las reglas para diseñar, planear, recolectar, capturar, organizar, presentar, procesar y analizar los datos obtenidos al realizar varios ensayos repetidos de un experimento y para inferir conclusiones acerca de este último. Proporciona además, los métodos para el diseño estadístico de experimentos y para tomar decisiones cuando aparecen situaciones de incertidumbre. Algunos autores establecen que la estadística no es ciencia ya que algunas de las reglas que emplea son empíricas.

Ramas relacionadas directamente con la Estadística:

1. Teoría Probabilística
2. Inferencia Estadística
 - a. Estimación
 - b. Pruebas de Hipótesis
3. Muestreo
4. Regresión y Correlación
5. Estadística Descriptiva
6. Diseño de Experimentos
7. Estadística Bayesiana
8. Estadística Demográfica
9. Estadística Paramétrica
10. Estadística No Paramétrica

Teoría Probabilística: rama de la matemática que proporciona los fundamentos, modelos matemáticos y el lenguaje que se usa en la estadística.

Estadística Descriptiva: rama de la Estadística que se encarga de analizar las reglas para recolectar, presentar y procesar los datos obtenidos al hacer una medición u observación de una característica de

interés de un objeto, con la finalidad de conocer su comportamiento sujeto a incertidumbre.

Las escalas de medición de una característica de interés de un objeto pueden clasificarse en cuatro grandes rubros:

1. Nominal
2. Ordinal
3. De intervalo
4. De razón

Medición Nominal. En este nivel de medición se establecen categorías distintivas que no implican un orden específico. Por ejemplo, si la unidad de análisis es un grupo de personas, para clasificarlas se puede establecer la categoría sexo con dos niveles, masculino (M) y femenino (F), los respondientes solo tienen que señalar su género, no se requiere de un orden real. Así, si se asignan números a estos niveles solo sirven para identificación y puede ser indistinto: 1=M, 2=F o bien, se pueden invertir los números sin que afecte la medición: 1=F y 2=M. En resumen en la escala nominal se asignan números a eventos con el propósito de identificarlos. No existe ningún referente cuantitativo. Sirve para nombrar las unidades de análisis en una investigación. Ejemplos de características de interés que utilizan escalas nominales: las placas de los carros, el número de cuenta de un alumno, el tipo de sangre, etcétera.

Medición Ordinal. Se establecen categorías con dos o más niveles que implican un orden inherente entre sí. La escala de medición ordinal es cuantitativa porque permite ordenar a los eventos en función de la mayor o menor posesión de un atributo o característica. Por ejemplo, en las instituciones escolares de nivel básico suelen formar por estatura a los estudiantes, se desarrolla un orden cuantitativo pero no suministra medidas de los sujetos. Estas escalas admiten la asignación de números en función de un orden prescrito. Las formas más comunes de variables ordinales son ítems (reactivos) actitudinales estableciendo una serie de niveles que expresan una actitud de acuerdo o desacuerdo con respecto a algún referente. Por ejemplo, ante el ítem: La economía mexicana debe dolarizarse, el respondiente puede marcar su respuesta de acuerdo a las siguientes alternativas: Totalmente de acuerdo; De acuerdo; Indiferente; En desacuerdo; Totalmente en desacuerdo. Las anteriores alternativas de respuesta pueden codificarse con números que van del uno al cinco que sugieren un orden preestablecido pero no implican una distancia entre un número y otro. Las escalas de actitudes son ordinales pero son tratadas como variables continuas (Therese L. Baker, 1997).

Medición de Intervalo. La medición de intervalo posee las características de la medición nominal y ordinal. Establece la distancia entre una

medida y otra. La escala de intervalo se aplica a variables continuas pero carece de un punto cero absoluto. El ejemplo más representativo de este tipo de medición es un termómetro, cuando registra cero grados centígrados de temperatura indica el nivel de congelación del agua y cuando registra 100 grados centígrados indica el nivel de ebullición, el punto cero es arbitrario, no real, lo que significa que en este punto no hay ausencia de temperatura. Una persona que en un examen de matemáticas obtiene una puntuación de cero no significa que carezca de conocimientos, el punto cero es arbitrario por que sigue existiendo la característica medida.

Medición de Razón. Una escala de medición de razón incluye las características de los tres anteriores niveles de medición anteriores (nominal, ordinal e intervalo). Determina la distancia exacta entre los intervalos de una categoría. Adicionalmente tiene un punto cero absoluto, es decir, en el punto cero no existe la característica o atributo que se mide. Las variables de ingreso, edad, número de hijos, etcétera, son ejemplos de este tipo de escala. El nivel de medición de razón se aplica tanto a variables continuas como discretas.

Las características de interés de un objeto que sólo pueden ser medidas utilizando las escalas nominal u ordinal se clasifican como atributos; por ejemplo, en un tornillo, la apariencia, el tipo de cabeza. En cambio, las características de interés de un objeto que pueden ser medidas utilizando las escalas por intervalos y de razón, se clasifican como variables; por ejemplo, en un tornillo, el diámetro, la longitud, la resistencia al torque.

Si se conocen con certeza los valores que tomará una variable de interés, previamente a medirla o realizar un experimento, a dicha variable se le denomina determinística. En este caso se puede conocer su comportamiento sin necesidad de hacer el experimento; si es el caso, el experimento se realizará con la finalidad de comprobar los resultados esperados. Si los valores que tomará la variable no pueden predecirse con certeza, antes del experimento, a dicha variable se le denomina aleatoria. Por otra parte, dentro del estudio de variables aleatorias, se puede ver que existen dos tipos: a) aquella cuyos resultados pueden ser medidos en forma discreta; por ejemplo: el número de llegadas a una cola, el número de defectos en un lote, el número de ases que se obtienen en un juego de pocker, etcétera, a éstas se les denomina variables aleatorias discretas; b) variables aleatorias continuas, son aquellas que tienen unidades de medida continua; por ejemplo: la cantidad de leche que produce una vaca diariamente, el tiempo de vida de un producto, el tiempo de espera en una cola, etcétera.

Una Población se define como la totalidad de valores posibles (mediciones o conteos) de una característica particular de un grupo específico de objetos. Al conjunto total de objetos se le denomina Universo. Por ejemplo: el conjunto de estudiantes de ingeniería puede ser un universo; dentro de este universo puede haber muchas poblaciones, la población de tipos de sangre, la población de estaturas, la población de mujeres, etcétera. Una Muestra es una parte de la población, seleccionada de acuerdo a una regla o plan. La estadística descriptiva se puede dividir en dos grandes rubros: la descripción tabular y gráfica de los datos de una muestra, y, la definición de algunos parámetros descriptivos de dicha muestra; posteriormente se hará una breve descripción del significado de cada uno de estos rubros.

Inferencia Estadística: rama de la Estadística que proporciona las reglas para estimar ciertos valores de una población, con base en los resultados de una muestra, formular hipótesis sobre la verdad de estas estimaciones y tomar decisiones con base en estos resultados. La inferencia estadística, siguiendo el texto anterior, puede dividirse por una parte en estimación puntual y por intervalos, y por otra parte en pruebas de hipótesis.

El Muestreo es la rama de la Estadística que se encarga de definir las reglas para tomar muestras de una población específica, el tamaño de dichas muestras y los parámetros que indicarán la representatividad de éstas. La primera finalidad del muestreo es obtener muestras representativas de la población en estudio. Una muestra es representativa si es obtenida aleatoriamente. Se dice que el Muestreo es Aleatorio si cumple las siguientes características:

1. Todos los posibles resultados del experimento deben tener la misma posibilidad de ocurrir.
2. Los resultados deben ser independientes entre sí.

El experimentador controla la cantidad de información contenida en la muestra por medio del número de unidades muestrales que incluye en la muestra y por el método usado para seleccionar los datos muestrales.

¿Cómo se puede determinar cuál procedimiento usar y el número de elementos a elegir de la muestra?. La respuesta depende de dos factores: ¿qué tanta representatividad se desea? y ¿qué tan seguro se requiere estar de esta representatividad?, es decir,

1. Si u es la variable de interés y \hat{u} es un estimador de u entonces se debe especificar un límite para el error de estimación, esto es $|u - \hat{u}| < B$.
2. Se debe fijar la probabilidad de que efectivamente el error de estimación sea menor de B , esto es, la fracción de las veces en que el muestreo tiene como error de estimación un valor menor a B , $P[E.E. < B] = 1 - \alpha$.

Usualmente se selecciona B como dos veces la desviación estándar del estimador de u , es decir, $B = 2\sigma_{\hat{u}}$. Si la distribución de los datos tiene forma de campana esto equivale a tomar $\alpha = 0.05$.

En los tres párrafos anteriores cabe señalar que esto se verá con detenimiento a lo largo del curso, son conceptos que todavía no se manejan.

En la Teoría del Muestreo se definen algunos tipos de muestreo para que la muestra obtenida sea representativa de la población bajo estudio:

1. Muestreo Aleatorio Simple. Cuando se selecciona un grupo de n unidades muestrales de tal manera que cada muestra de tamaño n tenga la misma posibilidad de ser seleccionada. Este tipo de muestreos se aplica cuando todos los elementos de la población bajo estudio se encuentran agrupados de la misma forma, sin distinguos de ninguna especie. Por ejemplo, un lote de artículos agrupados a granel en un solo contenedor.
2. Muestreo Aleatorio Estratificado. Consiste en separar los elementos de la población en grupos que no presenten traslapes, llamados estratos, y la selección posterior de una muestra aleatoria simple de cada estrato. La estratificación es la separación de los datos en grupos de la misma especie o que tienen el mismo origen. Se usa el término estratificación por una similitud con la manera en que las capas terrestres o estratos, forman la corteza de la tierra; por lo que cuando un grupo de datos con características importantes comunes se separa del total de datos disponibles, se dice que se estratifica. Algunos ejemplos de estratificación pueden ser: por turno, por proveedor de materia prima, por operario, por máquina, por semana, por método de trabajo, por molde, etcétera. Por ejemplo, los seres humanos pueden ser estratificados por sexo, por edad, por lugar de origen, etcétera. El muestreo aleatorio estratificado es aplicable cuando la población bajo estudio se

encuentra agrupada en bloques perfectamente distinguibles y sin traslapes; por ejemplo, un lote de artículos que llega en m cajas. Un grupo de alumnos en un salón no es homogéneo, está dividido por el sexo de cada alumno, la carrera que estén estudiando, el semestre de avance, la edad, etcétera; en este caso, si se va a realizar un estudio con ellos, conviene estratificarlos.

3. Muestreo Sistemático. Se obtiene al seleccionar aleatoriamente un elemento de los primeros k elementos en el marco y después seleccionar cada k-ésimo elemento a partir del primero. El muestreo sistemático es aplicable cuando la población bajo estudio se encuentra ordenada de alguna forma sistemática; por ejemplo, si al hacer un experimento, la característica que se está analizando, depende de la estatura de los miembros, es conveniente entonces, antes de aplicar el muestreo, ordenar a los miembros de la población por estatura (puede ser de mayor a menor o de menor a mayor) y aplicar el muestreo sistemático. También es aplicable el muestreo sistemático, cuando la población bajo estudio está surgiendo de una línea de producción; por ejemplo, si están saliendo de la línea de producción 10000 contactos de plata por hora y deseamos obtener una muestra de tamaño $n=100$, que sea representativa de la población producida en todo un turno de trabajo de 8 horas; entonces, tendríamos que sacar un contacto cada 800 producidos, lo cual significa sacar un contacto cada 5 minutos aproximadamente, a lo largo de todo el turno.
4. Muestreo por Conglomerados. Cuando cada unidad de muestreo aleatorio es a su vez una colección o conglomerado de elementos. En Geología se dice que un conglomerado es una colección de arenas de diferente tamaño. Las colonias de la Zona Metropolitana de la Ciudad de México, no pueden considerarse como estratos, porque las características de los habitantes en cada una de ellas, no pueden considerarse como semejantes; por ejemplo, el nivel socioeconómico de las familias no puede analizarse tomando a las colonias como estratos, porque en cada colonia existen niveles socioeconómicos diversos.

Hojas de Verificación. Para analizar a una población se debe tener presente lo dicho anteriormente; para muestrear se deben recoger datos, los cuales proporcionan evidencia de los hechos y por ello deben ser recolectados en forma cuidadosa. También el propósito para el cual se recolectan debe estar muy claro, si no es así, lo que sucederá es que

se obtendrá mucha información, pero no se tomarán acciones debido a que no se sabe para que se quieren los datos. Lo esencial al recolectar datos es que el propósito de hacerlo esté bien claro y que los datos reflejen la verdad. El siguiente paso es hacer que los datos sean fáciles de recolectar y de ser usados. Las hojas de verificación son formatos usados para capturar fácil y rápidamente los datos que se obtienen al realizar un experimento o medir cierta característica de interés de un proceso. Algunos de los usos de las hojas de verificación son:

1. Verificar la frecuencia de repetición de actividades, defectos o sucesos.
2. Verificar la localización de actividades, defectos o sucesos.
3. Verificar posibles causas de problemas a resolver para investigar si son las causas raíz.
4. Verificar las acciones a tomar previamente enlistadas.