Unidad III: Estadística descriptiva

3.1 Conceptos básicos de estadística: Definición, Teoría de decisión, Población, Muestra aleatoria, Parámetros aleatorios

TEORÍA DE DECISIÓN

Estudio formal sobre la toma de decisiones. Los estudios de casos reales, que se sirven de la inspección y los experimentos, se denominan teoría descriptiva de decisión; los estudios de la toma de decisiones racionales, que utilizan la lógica y la estadística, se llaman teoría preceptiva de decisión. Estos estudios se hacen más complicados cuando hay más de un individuo, cuando los resultados de diversas opciones no se conocen con exactitud y cuando las probabilidades de los distintos resultados son desconocidas. La teoría de decisión comparte características con la teoría de juegos, aunque en la teoría de decisión el "adversario" es la realidad en vez de otro jugador o jugadores.

Al hacer un análisis sobre esta teoría, y mirándola desde el punto de vista de un sistema, se puede decir que al tomar una decisión sobre un problema en particular, se debe tener en cuenta los puntos de dificultad que lo componen, para así empezar a estudiarlos uno a uno hasta obtener una solución que sea acorde a lo que se esta esperando obtener de este, y sino, buscar otras soluciones que se acomoden a lo deseado.

La teoría de decisión, no solamente se puede ver desde el punto de vista de un sistema, sino en general, porque esta se utiliza a menudo para tomar decisiones de la vida cotidiana, ya que muchas personas piensan que la vida es como una de las teorías; La teoría del juego, que para poder empezarlo y entenderlo hay que saber jugarlo y para eso se deben conocer las reglas de este, para que no surjan equivocaciones al empezar la partida.

Se puede decir que la Teoría de decisión es una de las ramas que sirve para que al dar un paso, no se vaya a dar en falso, porque si se conoce de esta no hay el porque de equivocarse.

POBLACIÓN

El concepto de población en estadística va más allá de lo que comúnmente se conoce como tal. Una población se precisa como un conjunto finito o infinito de personas u objetos que presentan características comunes. "Una población es un conjunto de todos los elementos que estamos estudiando, acerca de los cuales intentamos sacar conclusiones". Levan & Rubin (1996). "Una población es un conjunto de elementos que presentan una característica común". Cadenas (1974). Ejemplo: Los miembros del Colegio de Ingenieros del Estado Cojedes. El tamaño que tiene una población es un factor de suma importancia en el proceso de investigación estadística, y este tamaño vienen dado por el número de elementos que constituyen la población, según el número de elementos la población puede ser finita o infinita. Cuando el número de elementos que integra la población es muy grande, se puede considerar a esta como una población infinita, por ejemplo; el conjunto de todos los números positivos. Una población finita es aquella que está formada por un limitado número de elementos, por ejemplo; el número de estudiante del Núcleo San Carlos de la Universidad Nacional Experimental Simón Rodríguez. Cuando la población es muy grande, es obvio que la observación de todos los elementos se dificulte en cuanto al trabajo, tiempo y costos necesarios para hacerlo. Para solucionar este inconveniente se utiliza una muestra estadística. Es a menudo imposible o poco práctico observar la totalidad de los individuos, sobre todos si estos son muchos. En lugar de examinar el grupo entero llamado población o universo, se examina una pequeña parte del grupo llamada muestra.

3.2 Descripción de datos: Datos agrupados y no agrupados, Frecuencia de clase, Frecuencia relativa, Punto medio, Límites

La estadística descriptiva es una parte de la estadística que se dedica a analizar y representar los datos. Este análisis es muy básico, pero fundamental en todo estudio. Aunque hay tendencia a generalizar a toda la población las primeras conclusiones obtenidas tras un análisis descriptivo, su poder inferencia es mínimo y debería evitarse tal proceder. Otras ramas de la estadística se centran en el contraste de hipótesis y su generalización a la población.

Algunas de las técnicas empleadas en este primer análisis de los datos se enumeran más abajo en el listado de conceptos básicos. Básicamente, se lleva a cabo un estudio calculando una serie de medidas de tendencia central, para ver en qué medida los datos se agrupan o dispersan en torno a un valor central.

DATOS AGRUPADOS Y NO AGRUPADOS

Cuando la muestra que se ha tomado de la población o proceso que se desea analizar, es decir, tenemos menos de 20 elementos en la muestra, entonces estos datos son analizados sin necesidad de formar clases con ellos y a esto es a lo que se le llama tratamiento de datos no agrupados.

Cuando la muestra consta de 30 o más datos, lo aconsejable es agrupar los datos en clases y a partir de estas determinar las características de la muestra y por consiguiente las de la población de donde fue tomada. Antes de pasar a definir cuál es la manera de determinar las características de interés (media, mediana, moda, etc.) cuando se han agrupado en clases los datos de la muestra, es necesario que sepamos como se agrupan los datos.

FRECUENCIA DE CLASE

Marca de clase (punto medio): punto que divide a la clase en dos partes iguales. Es el promedio entre los límites superior e inferior de la clase.

Intervalo de clase: para una distribución de frecuencias que tiene clases del mismo tamaño, el intervalo de clase se obtiene restando el límite inferior de una clase del límite inferior de la siguiente.

FRECUENCIA RELATIVA

Es la relación o cociente entre la frecuencia absoluta y el número total de observaciones. Es la proporción entre la frecuencia de un intervalo y el número total de datos.

PUNTO MEDIO

Punto medio es el punto que divide a un segmento en dos partes iguales.

El punto medio de un segmento, es único y equidista de los extremos del segmento. Cumpliendo esta última condición, pertenece a la mediatriz del segmento.

La fórmula para determinar el punto medio de un segmento en el plano, con coordenadas: (x1,y1) y (x2,y2) es: [(x1 + x2) / 2] + [(y1 + y2) / 2]

LIMITES

Son los valores extremos que tiene el intervalo de clase, inferior y superior, entre los cuales van a estar los valores de los datos agrupados en ese intervalo de clase.

3.3 Medidas de tendencia central: Media aritmética, geométrica y ponderada, Mediana, Moda, Medidas de dispersión, Varianza, Desviación estándar, Desviación media, Desviación mediana, Rango

También se les conoce como medidas de posición o promedios son los valores que se utilizan para representar el conjunto de observaciones. Tienden a situarse en el centro del conjunto de los datos, previamente ordenados.

Las principales medidas centrales son: La media, la mediana y la moda.

MEDIA ARITMETICA GEOMETRICA PONDERADA

Media aritmética:

Es uno de los promedios de mayor utilización. Su generalización se debe a las propiedades que posee, que la convierten en un indicador muy representativo.

Se obtiene sumando los productos de cada valor de la variable por su respectiva frecuencia, y dividiendo esta suma por el total de observaciones.

MEDIANA EN ESTADISTICA

Mediana:

Es el valor de la variable que divide a la distribución de frecuencias en dos partes iguales.

Para hallar la mediana ordenamos las frecuencias de mayor a menor o viceversa, y si hay un número impar de valores la mediana es el central y si es par será la semisuma de los dos valores.

3.5 Parámetros para datos agrupados

Pasos para agrupar datos.

Determinar el rango o recorrido de los datos.

Rango = Valor mayor – Valor menor

Establecer el número de clases (k) en que se van a agrupar los datos tomando como base para esto la siguiente tabla.

Tamaño de muestra o No. De datos	Número de clases
Menos de 50	5 a 7
50 a 99	6 a 10
100 a 250	7 a 12
250 en adelante	10 a 20

El uso de esta tabla es uno de los criterios que se puede tomar en cuenta para establecer el número de clases en las que se van a agrupar los datos, existen otros para hacerlo.

Determinar la amplitud de clase para agrupar (C).

d. Formar clases y agrupar datos.

Para formar la primera clase, se pone como límite inferior de la primera clase un valor un poco menor que el dato menor encontrado en la muestra y posteriormente se suma a este valor C, obteniendo de esta manera el límite superior de la primera clase, luego se procede a obtener los límites de la clase siguiente y así sucesivamente.

3.6 Distribución de frecuencias

Distribución de frecuencia para datos no Agrupados:

Es aquella distribución que indica las frecuencias con que aparecen los datos estadísticos, desde el menor de ellos hasta el mayor de ese conjunto sin que se haya hecho ninguna modificación al tamaño de las unidades originales. En estas distribuciones cada dato mantiene su propia identidad después que la distribución de frecuencia se ha elaborado. En estas distribuciones los valores de cada variable han sido solamente reagrupados, siguiendo un orden lógico con sus respectivas frecuencias.

Distribución de frecuencia de clase o de datos Agrupados:

Es aquella distribución en la que la disposición tabular de los datos estadísticos se encuentran ordenados en clases y con la frecuencia de cada clase; es decir, los datos originales de varios valores adyacentes del conjunto se combinan para formar un intervalo de clase. No existen normas establecidas para determinar cuándo es apropiado utilizar datos agrupados o datos no agrupados; sin embargo, se sugiere que cuando el número total de datos (N) es igual o superior 50 y además el rango o recorrido de la serie de datos es mayor de 20, entonces, se utilizará la distribución de frecuencia para datos agrupados, también se utilizará este tipo de distribución cuando se requiera elaborar gráficos lineales como el histograma, el polígono de frecuencia o la ojiva.

La razón fundamental para utilizar la distribución de frecuencia de clases es proporcionar mejor comunicación acerca del patrón establecido en los datos y facilitar la manipulación de los mismos. Los datos se agrupan en clases con el fin de sintetizar, resumir, condensar o hacer que la información obtenida de una investigación sea manejable con mayor facilidad.

Componentes de una distribución de frecuencia de clase

1.- Rango o Amplitud total (recorrido).- Es el límite dentro del cual están comprendidos todos los valores de la serie de datos, en otras palabras, es el número de diferentes valores que toma la variable en un estudio o investigación dada. Es la diferencia entre el valor máximo de una variable y el valor mínimo que ésta toma en una investigación cualquiera. El rango es el tamaño del intervalo en el cual se ubican todos los valores que pueden tomar los diferentes datos de la serie de valores, desde el menor de ellos hasta el valor mayor estando incluidos ambos extremos. El rango de una distribución de frecuencia se designa con la letra R.

2.- Clase o Intervalo de clase.- Son divisiones o categorías en las cuales se agrupan un conjunto de datos ordenados con características comunes. En otras palabras, son fraccionamientos del rango o recorrido de la serie de valores para reunir los datos que presentan valores comprendidos entre dos limites.

Para organizar los valores de la serie de datos hay que determinar un número de clases que sea conveniente. En otras palabras, que ese número de intervalos no origine un número pequeño de clases ni muy grande. Un número de clases pequeño puede ocultar la naturaleza natural de los valores y un número muy alto

puede provocar demasiados detalles como para observar alguna información de gran utilidad en la investigación.

3.7 Técnicas de agrupación de datos

Es aquella distribución en la distribución tabular de los datos estadísticos se encuentran ordenados en clases y con la frecuencia de cada clase; es decir, los datos originales de varios valores adyacentes del conjunto se combinan para formar un intervalo de clases. No existen normas establecidas para determinar cuándo es apropiado utilizar datos agrupados o no agrupados; sin embargo, se sugiere que cuando el número total de datos (N) es igual o superior 50 y además el rango o recorrido de la serie es mayor de 20, entonces, se utilizara la distribución de frecuencia para datos agrupados, también se utilizara este tipo de distribución cuando se requiera elaborar gráficos lineales como el histograma, el polígono de frecuencia o la ojiva.

Cuando la muestra consta de 30 o más datos, lo aconsejable es agrupar los datos en clases y a partir de estas determinar las características de la muestra y por consiguiente las de la población de donde fue tomada.

Antes de pasar a definir cuál es la manera de determinar las características de interés (media, mediana, moda, etc.) cuando se han agrupado en clases los datos de la muestra, es necesario que sepamos como se agrupan los datos.

3.8 Técnicas de muestreo

Existen dos métodos para seleccionar muestras de poblaciones: el muestreo no aleatorio o de juicio y el muestreo aleatorio (que incorpora el azar como recurso en

el proceso de selección). Cuando este último cumple con la condición de que todos los elementos de la población tienen alguna oportunidad de ser escogidos en la muestra, si la probabilidad correspondiente a cada sujeto de la población es conocida de antemano, recibe el nombre de muestreo probabilístico. Una muestra seleccionada por muestreo de juicio puede basarse en la experiencia de alguien con la población. Algunas veces una muestra de juicio se usa como guía o muestra tentativa para decidir cómo tomar una muestra aleatoria más adelante.

Muestreo probabilístico

Forman parte de este tipo de muestreo todos aquellos métodos para los que puede calcular la probabilidad de extracción de cualquiera de las muestras posibles. Este conjunto de técnicas de muestreo es el más aconsejable, aunque en ocasiones no es posible optar por él. En este caso se habla de muestras probabilísticas, pues no es en rigor correcto hablar de muestras representativas dado que, al no conocer las características de la población, no es posible tener certeza de que tal característica se haya conseguido.

Sin reposición de los elementos: Cada elemento extraído se descarta para la subsiguiente extracción. Por ejemplo, si se extrae una muestra de una "población" de bombillas para estimar la vida media de las bombillas que la integran, no será posible medir más que una vez la bombilla seleccionada.

3.9 Histogramas

En estadística, un histograma es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados. En el eje vertical se representan las frecuencias, y en el eje horizontal los valores de las variables, normalmente

señalando las marcas de clase, es decir, la mitad del intervalo en el que están agrupados los datos.

Se utilizan para variables continuas o para variables discretas, con un gran número de datos, y que se han agrupado en clases.

Se utiliza cuando se estudia una variable continua, como franjas de edades o altura de la muestra, y, por comodidad, sus valores se agrupan en clases, es decir, valores continuos. En los casos en los que los datos son cualitativos (no-numéricos), como sexto grado de acuerdo o nivel de estudios, es preferible un diagrama de sectores.

Los histogramas son más frecuentes en ciencias sociales, humanas y económicas que en ciencias naturales y exactas. Y permite la comparación de los resultados de un proceso.

Tipos de histograma

- Diagramas de barras simples

Representa la frecuencia simple (absoluta o relativa) mediante la altura de la barra la cual es proporcional a la frecuencia simple de la categoría que representa.

- Diagramas de barras compuesta

Se usa para representar la información de una tabla de doble entrada o sea a partir de dos variables, las cuales se representan así; la altura de la barra representa la frecuencia simple de las modalidades o categorías de la variable y esta altura es proporcional a la frecuencia simple de cada modalidad.

- Diagramas de barras agrupadas

Se usa para representar la información de una tabla de doble entrada o sea a partir de dos variables, el cual es representado mediante un conjunto de barras como se clasifican respecto a las diferentes modalidades.

- Polígono de frecuencias

Es un gráfico de líneas que de las frecuencias absolutas de los valores de una distribución en el cual la altura del punto asociado a un valor de las variables es proporcional a la frecuencia de dicho valor.

- Ojiva porcentual

Es un gráfico acumulativo, el cual es muy útil cuando se quiere representar el rango porcentual de cada valor en una distribución de frecuencias.