

UNIVERSIDAD TECNOLÓGICA DE TABASCO



**DIVISION DE TECNOLOGIAS DE LA INFORMACIÓN Y
COMUNICACIÓN.**

Ingeniería en Desarrollo y Gestión de Software

3-A Vespertino

EXTRACCIÓN DEL CONOCIMIENTO EN BASES DE DATOS

Actividad No. 04

Estadística

ALUMNO:

Jesus Alejandro Ortiz Trejo

422110834_i

PROFESOR

Octavio Elías Sánchez Aquino

CONTENIDO

ESTADÍSTICA	3
¿Qué Es?.....	3
Aplicaciones.....	3
CONCEPTOS BÁSICOS.....	4
Desviación	4
Dispersión.....	4
Representación Grafica	4
Tendencias.....	4
ESTADÍSTICA INFERENCIAL.....	5
¿Qué Es?.....	5
Objetivo.....	5
GRÁFICOS.....	6
¿Qué Es?.....	6
Tipos	6
Gráfico De Barras	6
Gráfico Circular O Por Sectores.....	6
Histograma.....	7
Gráfico De Líneas	7
Gráfico De Dispersión	8
Gráfico De Áreas.....	8
MEDIDAS DE TENDENCIA CENTRAL	9
¿Qué Son?	9
Tipos	9
Media	9
Mediana	10
Moda	10

ESTADÍSTICA

¿Qué Es?

La estadística es una ciencia que se ocupa del estudio de los datos. Esto implica recopilar, analizar y describir datos para sacar conclusiones sobre un fenómeno particular. Utilizar herramientas matemáticas y de probabilidad para desarrollar métodos y modelos de análisis de datos. Estos métodos estadísticos se utilizan en muchos campos de las ciencias naturales y sociales, la publicidad y el marketing, la industria y el entretenimiento. Gracias a esto, luego del análisis estadístico, se logra: comprender una realidad, tomar decisiones, investigar problemas sociales, proponer soluciones en casos específicos, determinar relaciones en la población (Equipo de Enciclopedia Significados, 2023).

Aplicaciones

La estadística ofrece muchas aplicaciones útiles en una variedad de situaciones:

- Identifica los productos más vendidos en una tienda.
- Crea previsiones meteorológicas basadas en datos climáticos.
- Analizar resultados de equipos deportivos. Esto se aplica a victorias, derrotas y empates. Esta información es muy útil para predecir resultados en diferentes juegos.
- Ayuda a los gobiernos a comprender las características de sus poblaciones y economías, permitiéndoles tomar decisiones informadas basadas en datos reales.
- Permita que las empresas conozcan a sus competidores y mejoren su desempeño en el mercado.
- Sirven como base indispensable para que los investigadores científicos analicen datos y saquen conclusiones de ellos.

CONCEPTOS BÁSICOS

Desviación

La desviación es una medida de cuán diferentes son los valores individuales en un conjunto de datos de su valor medio o mediano. Puede ser positivo o negativo, dependiendo de si el valor está por encima o por debajo de la media.

Dispersión

La varianza se refiere a la variación o cantidad de datos en un conjunto. Mide el grado de dispersión o agrupación de valores alrededor de la media. Algunas medidas comunes de dispersión son la desviación estándar, el rango y el coeficiente de variación.

Representación Gráfica

La representación gráfica es la visualización de datos mediante gráficos o tablas. Estos gráficos pueden incluir, entre otros, histogramas, diagramas de dispersión, líneas de tendencia, etc. Le ayudan a comprender patrones y relaciones en sus datos.

Tendencias

Las tendencias son patrones o direcciones observadas en los datos a lo largo del tiempo o en función de otra variable. Pueden aumentar (positivos), disminuir (negativos) o permanecer estables. Identificar tendencias es importante para la toma de decisiones y la planificación.

ESTADÍSTICA INFERENCIAL

¿Qué Es?

La estadística inferencial es una rama de la estadística que se centra en sacar conclusiones y generalizaciones sobre una población a partir de la información obtenida de una muestra de esa población.

Imaginemos que queremos saber la altura promedio de todos los estudiantes de una escuela, pero sería difícil medir la altura de cada estudiante. En cambio, podemos medir la altura de una muestra de estudiantes y usar esta información para inferir la altura promedio de todos los estudiantes de la escuela.

Para llegar a esta conclusión, utilizamos métodos estadísticos para muestrear los datos y estimar el tamaño de la población desconocida (en este caso, la altura promedio de todos los estudiantes). Estos métodos pueden incluir la estimación de parámetros de distribución de probabilidad, el cálculo de intervalos de confianza o la realización de pruebas de hipótesis (Ortega, s.f.).

Objetivo

El objetivo principal de la estadística inferencial es hacer generalizaciones precisas sobre una población basándose en una muestra de datos obtenidos de esa población.

Las estadísticas inferenciales son útiles porque no siempre es posible medir todos los elementos de una población. Por lo tanto, la inferencia estadística nos permite tomar decisiones y predicciones basadas en una muestra representativa de la población, en lugar de medir todo en la población.

GRÁFICOS

¿Qué Es?

A nivel estadístico y matemático, los gráficos son representaciones visuales comúnmente utilizadas para representar e interpretar valores numéricos. Entre la diversa información que se obtiene al observar el gráfico, podemos detectar la existencia de una relación entre las variables y el grado, frecuencia o tasa de aparición de valores particulares (Mimenza, 2017).

Tipos

Existen tipos de gráficos muy diferentes, a menudo se utilizan uno u otro dependiendo del contenido que se representa o simplemente de la preferencia del autor. A continuación, señalaremos algunos de los más famosos y populares.

Gráfico De Barras

El tipo de gráfico más famoso y utilizado es el histograma o gráfico. En este caso, los datos se presentan como columnas contenidas en dos ejes cartesianos (coordenadas y abscisas), que representan valores diferentes. El aspecto visual representado por los datos es la longitud de las franjas indicadas, su grosor no es importante.

A menudo se utiliza para representar la frecuencia de diferentes estados o variables distintas (por ejemplo, la frecuencia de diferentes colores del iris en una muestra determinada, que solo puede tener valores específicos). Sólo se observa una variable en el eje horizontal y la frecuencia en la ordenada.

Gráfico Circular O Por Sectores

También es muy popular el gráfico del "queso", en el que los datos se presentan dividiendo el círculo en partes iguales al valor de la variable en estudio, teniendo cada parte un tamaño proporcional a la frecuencia de la variable en estudio en total. la cantidad de datos. Cada área representará el valor de la variable con la que estás trabajando.

Este tipo de cuadro o gráfico se utiliza a menudo para mostrar la proporción de casos con respecto a un número total, utilizando porcentajes (porcentaje de cada valor) para representarlo.

Histograma

Aunque a primera vista un gráfico se parece mucho a un histograma, es uno de los tipos de gráficos más importantes y fiables a nivel estadístico. En este caso, las barras también sirven para representar la frecuencia de valores particulares usando un eje cartesiano, pero en cambio se limitan a determinar la frecuencia de un valor particular de la variable que se está evaluando, reflejan todo el rango. De esta forma se observa una amplia gama de valores, que también pueden reflejar intervalos de diferentes longitudes.

Esto nos permite observar no solo la frecuencia sino también la varianza de una serie continua de valores, lo que puede ayudarnos a inferir la probabilidad. A menudo se aplica a variables continuas como el tiempo.

Gráfico De Líneas

En este tipo de gráfico, se utilizan líneas para delimitar el valor de una variable dependiente en relación con otra variable independiente. También se puede utilizar para comparar valores de una misma variable o de diferentes estudios usando el mismo gráfico (usando diferentes líneas). A menudo se utiliza para observar cambios en una variable a lo largo del tiempo.

Un ejemplo destacado de este tipo de gráfico es el polígono de frecuencias. Funciona de forma casi idéntica a un histograma, aunque utiliza puntos en lugar de barras, excepto que permite determinar la pendiente entre dos puntos específicos y comparar diferentes variables asociadas a los puntos de forma independiente o entre los resultados de diferentes ensayos con la misma variable; como estudiar el índice de efectividad del tratamiento, observar los datos de cambio antes y después del tratamiento.

Gráfico De Dispersión

Un diagrama de dispersión o diagrama xy es un tipo de gráfico en el que, utilizando un eje cartesiano, todos los datos obtenidos de las observaciones se representan como puntos. Los ejes x e y muestran el valor de la variable dependiente y otra variable independiente, o dos variables observadas si representan algún tipo de relación. Los puntos representan el valor reflejado en cada observación, lo que muestra visualmente la nube de puntos para que podamos observar el grado de dispersión de los datos.

Al utilizar cálculos, puede observar si existe una relación entre las variables. Este es un procedimiento comúnmente utilizado, por ejemplo, para establecer la existencia de líneas de regresión lineal, que nos permiten determinar si existe relación entre variables o incluso qué tipo de relación existe o no.

Gráfico De Áreas

Este tipo de gráfico observa la relación entre variables dependientes e independientes, similar a un gráfico de líneas. Inicialmente, se dibuja una línea que conecta los puntos que representan diferentes valores de la variable medida, pero también se incluye todo lo que está debajo: este tipo de gráfico permite ver la acumulación (en este punto, también se incluyen los puntos a continuación).

Se puede utilizar para medir y comparar los valores de diferentes muestras (por ejemplo, comparar los resultados obtenidos por dos personas, empresas, países, dos registros con el mismo valor...). Los diferentes resultados se pueden resumir observando fácilmente las diferencias entre diferentes muestras.

MEDIDAS DE TENDENCIA CENTRAL

¿Qué Son?

Las medidas de tendencia central o medidas de centralización son medidas estadísticas que muestran la centralidad de una distribución. Esto significa que la medida de tendencia central se utiliza para encontrar el valor que representa el centro de un conjunto de datos (Fernanado, 2024).

Tipos

Las medidas de tendencia central son valores estadísticos que resumen un conjunto de datos en un valor representativo.

Media

Para calcular la media se deben sumar todos los valores y luego dividir entre el número total de datos. Por lo tanto, la fórmula de la media es la siguiente:

El símbolo de la media es una ralla horizontal encima de la letra x (\overline{x}). Aunque también se puede diferenciar entre la media muestral y la media poblacional con el símbolo de la media: la media de una muestra se expresa con el símbolo (\overline{x}), mientras que la media de una población se utiliza la letra griega μ .

La media también se conoce como media aritmética o promedio. Además, la media de una distribución estadística es equivalente a su esperanza matemática.

Mediana

La mediana es el valor del medio de todos los datos ordenados de menor a mayor. Es decir, la mediana divide todo el conjunto de datos ordenados en dos partes iguales.

El cálculo de la mediana depende de si el número total de datos es par o impar:

- Si el número total de datos es impar, la mediana será el valor que está justo en el medio de los datos. Es decir, el valor que está en la posición $(n+1)/2$ de los datos ordenados.
- Si el número total de datos es par, la mediana será la media de los dos datos que están en el centro. Esto es, la media aritmética de los valores que están en las posiciones $n/2$ y $n/2+1$ de los datos ordenados.

Moda

En estadística, la moda es el valor del conjunto de datos que tiene una mayor frecuencia absoluta, es decir, la moda es el valor que más se repite de un conjunto de datos.

Por lo tanto, para calcular la moda de un conjunto de datos estadísticos basta con contar el número de veces que aparece cada dato en la muestra, y el dato más repetido será la moda.

La moda también se puede decir moda estadística o valor modal.

Se pueden distinguir tres tipos de modas según el número de valores que están más repetidos:

- Moda unimodal: solo hay un valor con el máximo número de repeticiones. Por ejemplo, [1, 4, 2, 4, 5, 3].
- Moda bimodal: el máximo número de repeticiones se produce en dos valores diferentes y ambos valores se repiten el mismo número de veces. Por ejemplo, [2, 6, 7, 2, 3, 6, 9].
- Moda multimodal: tres o más valores tienen el mismo número máximo de repeticiones. Por ejemplo, [3, 3, 4, 1, 3, 4, 2, 1, 4, 5, 2, 1].

Referencias

Equipo de Enciclopedia Significados. (8 de Mayo de 2023). *Qué es la Estadística: conceptos, tipos y ejemplos*. Obtenido de Significados.com: <https://www.significados.com/estadistica/>

Fernanado. (11 de Mayo de 2024). *Medidas de tendencia central*. Obtenido de probabilidadyestadistica.net: <https://www.probabilidadyestadistica.net/medidas-de-tendencia-central/#%c2%bfque-son-las-medidas-de-tendencia-central>

Mimenza, O. C. (24 de Agosto de 2017). *Tipos de gráficas: las diversas maneras de representar datos visualmente*. Obtenido de Psicologiaymente: <https://psicologiaymente.com/miscelanea/tipos-de-graficas>

Ortega, C. (s.f.). *Estadística inferencial: Qué es, importancia y ejemplos*. Obtenido de Questionpro: https://www.questionpro.com/blog/es/estadistica-inferencial/#Que_es_la_estadistica_inferencial