GUIA 1

1. ¿Qué se entiende por Estadística?

La estadística es una disciplina que se enfoca en recopilar, analizar y presentar datos numéricos para obtener información útil y tomar decisiones fundamentadas en diversos campos, como la investigación científica, los negocios y el gobierno. Ayuda a resumir datos, identificar patrones, hacer inferencias y evaluar la incertidumbre en los resultados, siendo esencial para la toma de decisiones basadas en evidencia cuantitativa.

1. ¿ Que es el dato?

Un dato es una unidad de información que representa un hecho, cifra o descripción particular. Puede ser numérico o no numérico y se utiliza para describir características o eventos. Los datos pueden ser recopilados, registrados y procesados para su análisis, lo que permite obtener información significativa y tomar decisiones informadas en diversas áreas de estudio y aplicación, como la ciencia, la tecnología, los negocios y más.

1. ¿Qué es una matriz de datos?

Una matriz de datos es una estructura organizada que se utiliza para almacenar y representar información tabular o numérica. Consiste en un conjunto de datos dispuestos en filas y columnas, donde cada fila suele representar una observación o un caso, y cada columna representa una variable o una característica específica de esos casos. Las matrices de datos son comunes en estadísticas, programación, análisis de datos y diversas disciplinas científicas.

En resumen, una matriz de datos es una forma de organizar y estructurar información en una tabla bidimensional que facilita la manipulación y el análisis de datos en diversas aplicaciones.

1. ¿ Que es una variable?

una variable es un elemento que puede variar y ser objeto de estudio en una investigación o experimento, y puede clasificarse en variables independientes o dependientes, así como en cualitativas o cuantitativas según su naturaleza.

* Variables independientes: Son aquellas que se manipulan o controlan en un experimento o estudio para observar su efecto en otras variables. Son las causas o factores que se estudian.
* Variables dependientes: Son aquellas que se observan o miden para evaluar cómo responden o cambian en función de las variables independientes. Son los resultados o efectos que se analizan.
* Variables cualitativas: Representan características no numéricas y se dividen en categorías o clases. Ejemplos son el género, el color de los ojos o el tipo de producto.
* Variables cuantitativas: Representan valores numéricos y se pueden medir con precisión. Ejemplos son la edad, la temperatura, el ingreso mensual, etc.

1. ¿A qué se denomina indicador?

los indicadores son medidas numéricas o variables utilizadas para evaluar y cuantificar aspectos específicos de una situación, sistema o proceso, facilitando la toma de decisiones y el seguimiento de objetivos.

1. ¿ Que es un instrumento?

un instrumento es una herramienta diseñada para llevar a cabo una tarea o función específica, y se utiliza en una variedad de aplicaciones en diferentes campos para medir, observar, analizar o realizar otras actividades específicas.

Mediana: valor que divide un conjunto de datos en dos partes iguales cuando se organizan en orden, y es una medida de tendencia central que es menos sensible a los valores extremos en comparación con la media.

Promedio: es la media aritmética de un conjunto de datos y se calcula sumando todos los valores y dividiéndolos por el número total de valores. Es una medida de tendencia central útil, pero puede verse influenciada por valores extremos en los datos.

Varianza: es una medida de la dispersión o variabilidad de un conjunto de datos que se calcula tomando las diferencias entre cada valor y la media, elevándolas al cuadrado y promediándolas. Ayuda a comprender cuán dispersos están los datos en relación con la media.

Coeficiente de Variación: es una medida que expresa la variabilidad relativa en un conjunto de datos como un porcentaje de la media, lo que facilita la comparación de la variabilidad entre diferentes conjuntos de datos.

Desvio Standard: es una medida de dispersión que indica cuánto se alejan los valores individuales de un conjunto de datos en relación con la media, proporcionando información sobre la variabilidad de los datos.

Moda / Modo: es el valor o valores que aparecen con mayor frecuencia en un conjunto de datos, lo que la convierte en una medida útil para describir la tendencia de repetición en los datos, especialmente cuando se trata de datos categóricos o discretos.

Punto Medio: es el valor que se encuentra en el centro de un intervalo de datos y se calcula promediando los valores extremos del intervalo. Se utiliza para resumir intervalos de datos de manera concisa.

Coeficiente de Asimetría: es una medida que evalúa la asimetría en la distribución de datos y ayuda a determinar si la distribución está sesgada hacia la izquierda, hacia la derecha o es aproximadamente simétrica.

* Coeficiente de asimetría positivo (mayor que cero): Indica que la distribución está sesgada hacia la derecha, lo que significa que hay valores atípicos o extremadamente altos en el conjunto de datos que hacen que la cola derecha sea más larga que la izquierda.
* Coeficiente de asimetría negativo (menor que cero): Indica que la distribución está sesgada hacia la izquierda, lo que significa que hay valores atípicos o extremadamente bajos en el conjunto de datos que hacen que la cola izquierda sea más larga que la derecha.
* Coeficiente de asimetría cercano a cero: Indica que la distribución es aproximadamente simétrica, es decir, que no muestra un sesgo claro hacia la izquierda o hacia la derecha.

Percentil: son medidas que dividen un conjunto de datos en partes iguales, ayudando a entender la posición relativa de un valor en relación con el conjunto de datos y a evaluar la distribución de los datos.

* El percentil 80 (valor que separa el 20% superior de los datos del 80% inferior en un conjunto de datos ordenados)
* el percentil 25 (primer cuartil)
* el percentil 75 (tercer cuartil)
* el percentil 90

Cuartil: son valores que dividen un conjunto de datos en cuatro partes iguales, proporcionando información sobre la distribución y la posición relativa de los valores en los datos.

Los cuartiles más comunes son:

* Primer Cuartil (Q1): Es el valor que divide el conjunto de datos en el 25% inferior y el 75% superior. Es equivalente al percentil 25.
* Segundo Cuartil (Q2): Es el valor que divide el conjunto de datos en el 50% inferior y el 50% superior. Es igual a la mediana y al percentil 50.
* Tercer Cuartil (Q3): Es el valor que divide el conjunto de datos en el 75% inferior y el 25% superior. Es equivalente al percentil 75.

GUIA 4

Probabilidad Condicional: es la probabilidad de que ocurra un evento dado que otro evento ya ha ocurrido, y se calcula utilizando la probabilidad conjunta de ambos eventos y la probabilidad del evento condicional. La fórmula para calcular la probabilidad condicional es:

P(A|B) = P (A y B) / P(B)

Donde:

* P(A|B) es la probabilidad condicional de A dado B.
* P (A y B) es la probabilidad de que ocurran ambos eventos A y B.
* P(B) es la probabilidad de que ocurra el evento B.

Sucesos Independientes: son aquellos en los que la ocurrencia o no ocurrencia de uno no afecta la probabilidad del otro, y se relacionan mediante la fórmula P (A y B) = P(A) \* P(B).

Matemáticamente, si A y B son dos sucesos independientes, entonces se cumple la siguiente relación:

P (A y B) = P(A) \* P(B)

Donde:

* P (A y B) es la probabilidad de que ambos sucesos A y B ocurran al mismo tiempo.
* P(A) es la probabilidad de que ocurra el suceso A.
* P(B) es la probabilidad de que ocurra el suceso B.

Teorema de la probabilidad total: es un principio fundamental en la teoría de la probabilidad que permite calcular la probabilidad de un evento mediante la suma de probabilidades condicionales ponderadas por las probabilidades de eventos relacionados.

Si {B₁, B₂, ..., Bₙ} es una partición de un espacio muestral S (lo que significa que los eventos Bᵢ son mutuamente excluyentes y su unión es igual a S), entonces para cualquier evento A, la probabilidad de A se puede calcular como: P(A) = ∑ [P(A|Bᵢ) \* P(Bᵢ)]

Supongamos que tenemos dos máquinas (M₁ y M₂) que producen tornillos. El 80% de los tornillos provienen de la Máquina 1 (M₁) y el 20% provienen de la Máquina 2 (M₂). Sabemos que el 2% de los tornillos producidos por M₁ son defectuosos, mientras que el 5% de los tornillos producidos por M₂ son defectuosos.

Queremos calcular la probabilidad de que un tornillo elegido al azar sea defectuoso.

* Definimos los eventos:
* A: Un tornillo es defectuoso.
* B₁: El tornillo proviene de M₁.
* B₂: El tornillo proviene de M₂.
* Sabemos:
* P(B₁) = 0.80 (probabilidad de que el tornillo provenga de M₁)
* P(B₂) = 0.20 (probabilidad de que el tornillo provenga de M₂)
* P(A|B₁) = 0.02 (probabilidad de que un tornillo sea defectuoso dado que proviene de M₁)
* P(A|B₂) = 0.05 (probabilidad de que un tornillo sea defectuoso dado que proviene de M₂)

Usando el Teorema de la Probabilidad Total:

P(A) = P(A|B₁) \* P(B₁) + P(A|B₂) \* P(B₂)

P(A) = (0.02 \* 0.80) + (0.05 \* 0.20)

P(A) = 0.016 + 0.010

P(A) = 0.026

La probabilidad de que un tornillo elegido al azar sea defectuoso es del 2.6%.

Formula de Bayes: es un concepto fundamental en la teoría de la probabilidad que se utiliza para actualizar probabilidades cuando se dispone de nueva información o evidencia. Permite calcular la probabilidad condicional de un evento A dado un evento B, cuando se conoce la probabilidad de B dado A y la probabilidad a priori de A y B.

Si A y B son eventos, la probabilidad condicional de que ocurra A dado que ha ocurrido B se calcula de la siguiente manera:

P (A|B) = [P (B|A) \* P (A)] / P (B)

Donde:

* P(A|B) es la probabilidad condicional de A dado B.
* P(B|A) es la probabilidad condicional de B dado A.
* P(A) es la probabilidad a priori de A.
* P(B) es la probabilidad a priori de B.

Supongamos que una clínica médica ofrece pruebas para detectar una enfermedad rara. La prueba tiene una tasa de falsos positivos del 5% y una tasa de falsos negativos del 1%. La enfermedad es realmente rara, y solo el 0.1% de la población la tiene. Queremos calcular la probabilidad de que una persona que ha dado positivo en la prueba realmente tenga la enfermedad.

* Definimos los eventos:
* A: La persona tiene la enfermedad.
* B: La persona da positivo en la prueba.
* Sabemos:
* P(A) = 0.001 (probabilidad a priori de tener la enfermedad).
* P(B|A) = 0.99 (probabilidad de dar positivo en la prueba dado que tiene la enfermedad).
* P(B|¬A) = 0.05 (probabilidad de dar positivo en la prueba dado que no tiene la enfermedad, ¬A es la negación de A).
* Queremos calcular P(A|B) (probabilidad de tener la enfermedad dado que dio positivo en la prueba).

Usando la Fórmula de Bayes:

P (A|B) = [P (B|A) \* P (A)] / [P (B|A) \* P (A) + P (B|¬A) \* P (¬A)]

P(A|B) = (0.99 \* 0.001) / [(0.99 \* 0.001) + (0.05 \* 0.999)]

P(A|B) ≈ 0.0196

La probabilidad de que una persona que ha dado positivo en la prueba realmente tenga la enfermedad es aproximadamente 1.96%.

Diagrama de Árbol: es una representación gráfica utilizada en teoría de la probabilidad y toma de decisiones para visualizar diferentes posibles resultados o secuencias de eventos en un proceso o experimento aleatorio. Elementos clave de un diagrama de árbol:

* Nodo Raíz: Representa el evento inicial o el punto de partida en un proceso.
* Ramas: Conectan los nodos y muestran las posibles trayectorias o resultados.
* Nodos Hoja: Son los extremos de las ramas y representan resultados finales o terminales.
* Probabilidades: Se asignan a cada rama para indicar la probabilidad de que ocurra ese resultado particular.
* Estructura Jerárquica: Los nodos se organizan de manera jerárquica, con el nodo raíz en la parte superior y los nodos hoja en la parte inferior.

Sistema Exhaustivo: es un conjunto de eventos mutuamente excluyentes y colectivamente exhaustivos que cubren todas las posibilidades en un experimento aleatorio, lo que garantiza que al menos uno de los eventos debe ocurrir en cada realización del experimento.

Elementos clave de un sistema exhaustivo:

* Mutuamente excluyente: Los eventos en el sistema exhaustivo no pueden ocurrir simultáneamente; son disjuntos entre sí.
* Colectivamente exhaustivo: La unión de todos los eventos en el sistema cubre todas las posibles ocurrencias en el experimento.

Prueba de Hipótesis Estadística: Una prueba de hipótesis estadística es un proceso utilizado en estadística inferencial para evaluar y tomar decisiones basadas en datos y evidencia empírica. Sus elementos clave son:

* Hipótesis nula (H0): Una afirmación inicial que se asume como verdadera antes de la prueba y que se busca refutar.
* Hipótesis alternativa (H1 o Ha): La afirmación que se está tratando de respaldar con evidencia y que contradice la hipótesis nula.
* Estadístico de prueba: Un valor calculado a partir de los datos de la muestra utilizado para evaluar la hipótesis nula.
* Nivel de significancia (α): La probabilidad de cometer un error de Tipo I al rechazar incorrectamente la hipótesis nula.
* Región crítica: El rango de valores del estadístico de prueba que conduciría al rechazo de la hipótesis nula.
* Valor p: La probabilidad de obtener un estadístico de prueba al menos tan extremo como el observado bajo la hipótesis nula.
* Toma de decisiones: Basada en el valor p y el nivel de significancia, se decide si se rechaza o no se rechaza la hipótesis nula.
* Conclusión: Se comunica si se rechaza o no se rechaza la hipótesis nula y se interpreta el resultado en el contexto de la investigación.