Unidad 3 - VARIABLES ALEATORIAS

1. **Definición de Variables Aleatorias**

Una **variable aleatoria** es una función matemática que asigna un valor numérico a cada posible resultado de un experimento aleatorio. Los resultados son inciertos, pero la variable aleatoria los modela numéricamente.

Ejemplo práctico 1

Lanzamiento de una moneda

- Imagina que lanzas una moneda. Los posibles resultados son **cara (C)** y **cruz (X)**.
- Si asignamos valores numéricos a los resultados, tenemos que:
  - **C = 1** (cara),
  - **X = 0** (cruz).

Entonces, podemos definir una variable aleatoria **X** que depende del resultado del lanzamiento de la moneda.

2. Tipos de Variables Aleatorias

A. Variables Aleatorias Discretas

Las **variables aleatorias discretas** toman un número finito o contable de valores.

Ejemplo práctico 2

Número de clientes en una tienda

Supongamos que deseas modelar el número de clientes que ingresan a una tienda durante una hora. Los valores posibles de la variable aleatoria **X** podrían ser: **0, 1, 2, 3, 4, ...**, es decir, valores enteros no negativos.

Una distribución discreta común para este tipo de problemas es la **distribución de Poisson**. La fórmula de la distribución de Poisson es:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Donde:

- $P(X = k)$ es la probabilidad de que haya **k** clientes en una hora,
- λ es el promedio de clientes por hora (por ejemplo, 5 clientes),
- k es el número de clientes (0, 1, 2, 3, ...).

Si $\lambda = 5$, la probabilidad de que haya exactamente 3 clientes en una hora sería:

$$P(X = 3) = \frac{5^3 e^{-5}}{3!} = 0.1404$$

B. Variables Aleatorias Continuas

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Donde:

- μ es la media de la distribución (tiempo promedio de espera),
- σ es la desviación estándar (variabilidad del tiempo de espera),
- x es el tiempo de espera en minutos.

Si la media es 10 minutos y la desviación estándar es 2 minutos, la probabilidad de que un cliente espere exactamente 12 minutos sería:

$$f(12) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(12-10)^2}{2\cdot 2^2}} = 0.176$$

3.

Determinación del Tipo de Distribución de un Conjunto de Datos

A. Prueba Chi-Cuadrada

La **prueba de chi-cuadrada** se utiliza para comparar la distribución observada de un conjunto de datos con una distribución esperada.

Ejemplo práctico 4: Distribución de los colores de bolas en una urna

Supón que tienes una urna con 3 colores de bolas: **rojo, verde y azul**. Lanzas 100 bolas y obtienes las siguientes frecuencias observadas:

- 40 bolas rojas,
- 30 bolas verdes,
- 30 bolas azules.

El supuesto es que la distribución de bolas debería ser uniforme, es decir, $\frac{100}{3} \approx 33.33$ bolas para cada color.

La fórmula de la estadística de chi-cuadrada es:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Donde:

- O_i son los valores observados,
- E_i son los valores esperados.

Entonces, calculamos la chi-cuadrada para cada color:

$$\chi^2 = \frac{(40-33.33)^2}{33.33} + \frac{(30-33.33)^2}{33.33} + \frac{(30-33.33)^2}{33.33}$$

El valor de χ^2 se compara con el valor crítico de la tabla para decidir si se rechaza la hipótesis nula (que la distribución es uniforme).

B. Prueba de Kolmogorov-Smirnov

La **prueba de Kolmogorov-Smirnov** se utiliza para comparar una muestra de datos con una distribución teórica, o para comparar dos muestras entre sí.

Ejemplo práctico 5

Verificar si una muestra sigue una distribución normal

Supongamos que tienes un conjunto de datos de tiempos de espera en una fila (valores continuos). Quieres verificar si estos datos siguen una distribución normal. La prueba de Kolmogorov-Smirnov compara la **función de distribución empírica** de los datos con la **función de distribución acumulada** (CDF) de la distribución normal.

La prueba devuelve un valor DDD que se compara con un valor crítico para determinar si se rechaza o no la hipótesis nula.

4. Ajuste de Datos con Stad::Fit

Stad::Fit es una herramienta que ajusta los datos a diferentes distribuciones de probabilidad, para que puedas elegir cuál distribuye mejor tus datos.

Ejemplo práctico 6:

Ajuste de datos con una distribución exponencial

Imagina que tienes un conjunto de datos que representa los **tiempos entre llegadas de clientes** en una tienda. Para ajustar estos datos a una **distribución exponencial**, podrías usar Stad::Fit, que calcularía el parámetro **lambda** (λ) de la distribución exponencial.

Si el valor de $\lambda = 0.1$, la función de densidad de probabilidad sería:

$$f(x) = \lambda e^{-\lambda x} = 0.1e^{-0.1x}$$

Esta fórmula te permitiría calcular la probabilidad de que un cliente llegue después de un tiempo x.

5. Generación de Variables Aleatorias

Las **variables aleatorias** se generan utilizando algoritmos de números pseudoaleatorios, que generan secuencias de números que parecen aleatorios, pero en realidad están determinados por un algoritmo.

Ejemplo práctico 7:

Generación de números aleatorios con un generador congruencial lineal

Supón que estás utilizando un **Generador Congruencial Lineal (LCG)** para generar números

$$X_{n+1} = (aX_n + c) \mod m$$

Donde:

- X_n es el número anterior,
- a es un multiplicador,
- c es un incremento,
- m es el módulo.

Si usas valores $a = 5$, $c = 1$, $m = 16$ y $X_0 = 7$, el siguiente número aleatorio sería:

$$X_1 = (5 \times 7 + 1) \mod 16 = 36 \mod 16 = 4$$

6. Variables Continuas y Discretas

Ya se discutieron los ejemplos anteriores, donde **variables continuas** como el tiempo de espera o la temperatura se modelan usando distribuciones como la normal, mientras que **variables discretas** como el número de clientes o el número de bolas de un color en una urna se modelan con distribuciones como la Poisson.

Bibliografía:

1. **Ross, S. M. (2014).** *Simulation (5th ed.)*. Academic Press.
2. **Montgomery, D. C. (2009).** *Design and Analysis of Experiments (7th ed.)*. Wiley.
3. **Law, A. M., & Kelton, W. D. (2000).** *Simulation Modeling and Analysis (3rd ed.)*. McGraw-Hill.
4. **Kachitvichyanukul, V., & Schmeiser, B. W. (2001).** *Simulation: A Modeler's Approach*. Prentice Hall.