Universidad Nacional Experimental del Táchira

****Vicerrectorado Académico

Decanato de Docencia

Departamento de Matemática y Física

Estadística Aplicada II

**TRABAJO PARCIAL 3**

**MODELO DE REGRESIÓN LINEAL SIMPLE**

**Integrantes:**

Camargo López Melberit del Carmen C.I: 31.821.511

Ramírez Zambrano Valeria Sarahy C.I: 31.869.361

**Sección 1**

San Cristóbal, marzo de 2025

**1) ¿Qué es un modelo de regresión lineal simple?**

Un modelo de regresión lineal simple es una técnica estadística utilizada para analizar la relación entre dos variables: Una variable que es independiente (predictora) y una variable dependiente (respuesta). El objetivo de este modelo es encontrar la mejor línea recta que se ajuste a los datos, de una manera que se pueda predecir el valor de la variable dependiente a partir del valor de la variable independiente.

La ecuación general de un modelo de regresión lineal simple se expresa de tal manera:

Y = β₀ + β₁ X + ∊

La regresión lineal simple es útil en diversas aplicaciones, como la economía, la biología, la ingeniería y las ciencias sociales, ya que permite entender y cuantificar relaciones entre variables.

**2) ¿Qué utilidad tiene el ajuste de un modelo de regresión lineal simple en la psicología?**

El ajuste de un modelo de regresión lineal simple cuenta con varias utilidades en el campo de la psicología, en las cuales se destacan.

1. Predicción: Permite predecir el valor de una variable dependiente (por ejemplo, el nivel de ansiedad) a partir de una variable independiente (como el tiempo de estudio). Esto puede ayudar a identificar factores que influyen en el comportamiento o en el rendimiento de los individuos.

2. Comprensión de relaciones: Ayuda a entender la relación entre dos variables. Por ejemplo, un psicólogo puede usar la regresión lineal para investigar cómo la ansiedad se relaciona con el rendimiento en pruebas cognitivas.

3. Identificación de tendencias: Facilita la identificación de tendencias y patrones en los datos psicológicos, lo cual es fundamental para el desarrollo de teorías y modelos en la psicología.

4. Evaluación de intervenciones: Permite evaluar la efectividad de intervenciones psicológicas. Por ejemplo, se puede analizar si un programa de terapia tiene un impacto significativo en la reducción de síntomas depresivos.

5. Control de variables confusoras: A través del ajuste del modelo, se pueden controlar variables adicionales que podrían influir en la relación entre las variables de interés, proporcionando un análisis más preciso.

6. Investigación cuantitativa: Es una herramienta clave en la investigación cuantitativa, permitiendo a los psicólogos analizar datos numéricos y extraer conclusiones basadas en evidencia.

7. Toma de decisiones informadas: Los resultados de un modelo de regresión pueden informar decisiones clínicas y políticas en el ámbito de la salud mental, ayudando a diseñar programas y políticas más efectivas.

**3) Características del modelo de regresión lineal simple**

1. Linealidad: Se asume que existe una relación lineal entre la variable independiente X y la variable dependiente Y.

2. Independencia: Las observaciones deben ser independientes entre sí.

3. Homoscedasticidad: La varianza de los errores debe ser constante a lo largo de todos los niveles de la variable independiente.

4. Normalidad: Los residuos (errores) del modelo deben seguir una distribución normal.

5. No multicolinealidad: En el contexto de la regresión simple, este punto es menos relevante ya que solo hay una variable independiente. Sin embargo, en regresiones múltiples, es importante que las variables independientes no estén altamente correlacionadas.

**Modelo matemático:**

El modelo de regresión lineal simple se puede expresar mediante la siguiente ecuación:

Y = β₀ + β₁ X + ∊

Donde:

• Y es la variable dependiente.

• X es la variable independiente.

• β₀ es el intercepto (constante) del modelo.

• β₁ es el coeficiente de la variable independiente, que representa la pendiente de la línea de regresión.

• ∊ es el término de error, que captura la variabilidad en Y que no se explica por X.

**Coeficientes del modelo:**

1. Intercepto (β₀):

• Representa el valor esperado de Y cuando X = 0.

• Es el punto donde la línea de regresión cruza el eje Y.

2. Coeficiente (β₁):

• Representa el cambio esperado en Y por cada unidad de cambio en X.

• Si β₁ es positivo, indica que a medida que X aumenta, Y también tiende a aumentar. Si es negativo, indica que a medida que X aumenta, Y tiende a disminuir.

**Interpretación de los coeficientes:**

• Interpretación del intercepto (β₀):

• Si β₀ = 5, esto significa que cuando X = 0, se espera que Y = 5. Sin embargo, es importante considerar si tiene sentido interpretar este valor en el contexto del problema, ya que puede no ser un valor realista.

• Interpretación del coeficiente (β₁):

• Si β₁ = 2, esto significa que por cada incremento de 1 unidad en X, se espera que Y aumente en 2 unidades. Por ejemplo, si X representa horas de estudio y Y representa la calificación en un examen, esto indicaría que cada hora adicional de estudio se asocia con un aumento esperado de 2 puntos en la calificación.

**4) El principio del método de mínimos cuadrados**

El principio de métodos de mínimos cuadrados es una técnica estadística utilizada para ajustar un modelo a un conjunto de datos. Su objetivo es encontrar la mejor aproximación de una función a un conjunto de observaciones, minimizando la suma de los cuadrados de las diferencias (residuos) entre los valores observados y los valores predichos por el modelo.

**Conceptos Clave:**

1. Modelo: Generalmente se utiliza un modelo lineal, que puede representarse como y = mx + b , donde y es la variable dependiente, x es la variable independiente, m es la pendiente y b es la intersección con el eje y.

2. Residuos: Son las diferencias entre los valores observados ( yᵢ ) y los valores predichos ( ^yᵢ ) por el modelo. Se definen como: rᵢ = yᵢ - ^yᵢ

3. Suma de los Cuadrados de los Residuos: La función objetivo que se busca minimizar es la suma de los cuadrados de estos residuos: S = ∑ᵢ₌₁ⁿ (yᵢ - ^yᵢ)²

4. Minimización: Para encontrar los parámetros del modelo (por ejemplo, m y b en el caso de un ajuste lineal), se derivan las expresiones de S respecto a esos parámetros y se igualan a cero para encontrar los valores óptimos.

**Aplicaciones:**

• Regresión Lineal: Es una de las aplicaciones más comunes del método de mínimos cuadrados, donde se busca ajustar una línea a un conjunto de datos.

• Análisis de Tendencias: Se utiliza para identificar tendencias en datos históricos.

• Estimación de Parámetros: En modelos estadísticos más complejos, el método se puede extender para estimar parámetros en modelos no lineales.

**Ventajas:**

• Es un método sencillo y ampliamente utilizado en estadística y econometría.

• Proporciona estimaciones que tienen propiedades deseables bajo ciertas condiciones, como ser insesgadas y consistentes.

**Desventajas:**

• Sensible a valores atípicos, ya que los residuos se elevan al cuadrado.

• Asume que los errores son homocedásticos (tienen varianza constante) y normalmente distribuidos.

**5) ¿Qué es el diagrama de dispersión y cuál es su utilidad?:**

Un diagrama de dispersión, también conocido como gráfico de dispersión o scatter plot en inglés, es una representación gráfica que muestra la relación entre dos variables numéricas. En este tipo de gráfico, cada punto representa un par de valores, donde uno se asigna al eje X (horizontal) y el otro al eje Y (vertical).

**Utilidad del diagrama de dispersión:**

1. Visualización de relaciones: Permite observar si existe una relación entre las dos variables, ya sea positiva, negativa o nula. Por ejemplo, se puede ver si a medida que una variable aumenta, la otra también lo hace.

2. Identificación de patrones: Ayuda a identificar patrones o tendencias en los datos, como la linealidad, la curvatura o la dispersión.

3. Detección de outliers: Facilita la identificación de valores atípicos (outliers) que pueden influir en el análisis y la interpretación de los datos.

4. Análisis de correlación: Se puede utilizar para calcular y visualizar la correlación entre las variables, lo que es útil en estudios estadísticos y científicos.

5. Modelado predictivo: Sirve como una herramienta preliminar para el desarrollo de modelos estadísticos o de machine learning, donde se busca predecir una variable en función de otra.

En conclusión, el diagrama de dispersión es una herramienta fundamental en el análisis de datos que ayuda a entender la relación entre variables y a tomar decisiones informadas basadas en esa información.

**6. Ajustar un modelo de regresión lineal simple utilizando datos reales.**

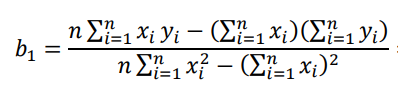
- Los investigadores del INEGI se han encargado de registrar la cantidad de suicidios ocurridos en México en los últimos años, obteniendo de esta manera los siguientes resultados.

(Fuente: https://www.inegi.org.mx/temas/salud/)

|  |  |
| --- | --- |
| **AÑO** | **SUICIDIOS REGISTRADOS** |
| 2017 | 6559 |
| 2018 | 6808 |
| 2019 | 7223 |
| 2020 | 7896 |
| 2021 | 8432 |
| 2022 | 8239 |
| 2023 | 9072 |

**a.- Obtener la recta de regresión lineal mediante el método de lo mínimos cuadrado para predecir la temperatura.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **AÑO (X)** | **SUICIDIOS REGISTRADOS (Y)** | **XY** | **X^2** | **Y^2** |
|  | 2017 | 6559 | 13229503 | 4068289 | 43020481 |
|  | 2018 | 6808 | 13738544 | 4072324 | 46348864 |
|  | 2019 | 7223 | 14583237 | 4076361 | 52171729 |
|  | 2020 | 7896 | 15949920 | 4080400 | 62346816 |
|  | 2021 | 8432 | 17041072 | 4084441 | 71098624 |
|  | 2022 | 8239 | 16659258 | 4088484 | 67881121 |
|  | 2023 | 9072 | 18352656 | 4092529 | 82301184 |
| **SUMA** | **14140** | **54229** | **109554190** | **28562828** | **425168819** |



=

=

81270

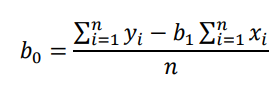
196

=

7 \* 109554190 – (14140) (54229)

7 \* 28562828 – (14140)2

414,64



-829831,57

-5808821

7

=

=

54229 – (414,64) \* 14140

7

=

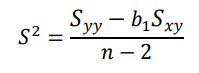
La línea de regresión está dada por:

=

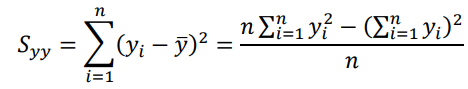


414,64*x* + (-829831,57)

**b.- Calcular el error estándar del modelo e interpretarlo**

Estimación de σ2

Donde



35397292

7

7 \* 425168819 – (54229)2

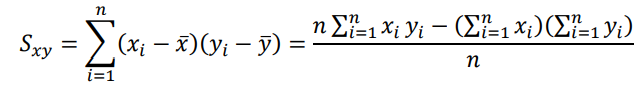
7

=

5056756

=

=



7 \* 109554190 – (14140) (54229)

7

=



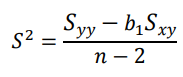
11610

=

81270

7

=



48550,4857

=

242752,429

5

=

5056756 – (414,64) \* (11610)

7 – 2

=

Entonces, el error estándar de Ԑ está dado por:



220,34

=

=

Y representa la variabilidad de las perturbaciones no explicadas por el modelo de regresión.

**c.- Uso del modelo encontrado para a partir de él estimar que cantidad de suicidios habrá en el año 2024.**

Para x = Año = 2024 los suicidios que se estiman son:



9405,57 suicidios

=

414,64 \* (2024) + (-829831,57)

=