República Bolivariana de Venezuela.

Universidad Nacional Experimental del Táchira.

Colegio de Psicólogos de Venezuela

San Cristóbal - Estado Táchira.

**Regresión Lineal Simple**

**Estudiantes:**

Loriangel Vargas Pérez.

31.356.899

Marzo, 2025

**Estadística II – sección 02**

**Regresión Lineal**

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

Un enfoque de regresión lineal simple es un método estadístico que se aplica para analizar la conexión entre dos variables: una variable que no cambia (o que se usa para predecir) y una variable que se ve afectada (o que responde). La meta es identificar la línea recta que se adapta mejor a los datos, lo que facilita la predicción del valor de la variable que responde, a partir del valor de la variable que no cambia.

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

En el campo de la psicología, la aplicación de un modelo de regresión lineal simple tiene múltiples propósitos, ayudando a investigadores y expertos a entender y prever el comportamiento humano. A continuación, te presento algunas de las aplicaciones más importantes:

**1. Anticipación de comportamientos y estados psicológicos:**

* Conexión entre variables:

La regresión lineal facilita la identificación de si hay una conexión lineal entre dos variables psicológicas. Por ejemplo, es posible examinar cómo se relacionan el nivel de estrés (variable independiente) y la calidad del sueño (variable dependiente). Al definir esta conexión, se pueden hacer predicciones sobre cómo alteraciones en una variable influirán en la otra.

* Predicción:

Permite estimar el valor de una variable psicológica tomando como base otra variable. Por ejemplo, se podría proyectar el nivel de ansiedad de un estudiante según su rendimiento académico anterior.

**2. Detección y entendimiento de las relaciones:**

* Análisis de factores determinantes:

Ayuda a descubrir qué factores (variables independientes) tienen un efecto importante en un comportamiento o estado psicológico específico (variable dependiente). Por ejemplo, se podría investigar de qué manera el uso de redes sociales afecta la autoestima de los jóvenes.

* Examen de correlaciones:

Facilita la medición de la intensidad y la dirección de la relación entre dos variables. Esto resulta valioso para entender cómo distintos elementos psicológicos están conectados entre sí.

**3. Evaluación de intervenciones y tratamientos:**

* Medición de la eficacia:

La regresión lineal se puede aplicar para determinar si una intervención psicológica o un tratamiento tiene un impacto notable en una variable importante. Por ejemplo, se podría examinar si un programa de terapia disminuye los síntomas de depresión en los pacientes.

* Seguimiento de progresos:

Facilita el monitoreo del avance de los pacientes con el tiempo y analiza la efectividad de las intervenciones.

Ejemplo, un psicólogo educativo podría emplear la regresión lineal para anticipar el desempeño académico de los alumnos según sus métodos de estudio. Un psicólogo clínico podría investigar la conexión entre el apoyo social y la recuperación de personas con problemas de salud mental. Un psicólogo organizacional podría usar esta herramienta para evaluar la relación que existe entre la motivación de sus trabajadores y su rendimiento en el trabajo.

En resumen, la regresión lineal simple es una herramienta útil en psicología para comprender, anticipar y evaluar las interrelaciones entre variables, lo que ayuda a tener una mejor comprensión del comportamiento humano y a tomar decisiones bien fundamentadas en la práctica profesional.

**Características del modelo de regresión lineal simple. El modelo matemático, los coeficientes del modelo y la forma correcta de interpretarlos.**

**1. Modelo matemático:**

La expresión general del modelo de regresión lineal simple se presenta en la siguiente fórmula:

Y = β₀ + β₁X + ε

Donde:

Y es la variable de respuesta (la que se pretende predecir).

X es la variable explicativa (la que se usa para hacer la predicción).

β₀ representa el valor de Y cuando X es igual a cero, conocido como el intercepto.

β₁ indica la inclinación, que muestra el cambio en Y por cada unidad que cambia X.

ε es el error o residuo, que mide la variabilidad de Y que no puede ser explicada por X.

**2. Coeficientes del modelo:**

Los coeficientes β₀ y β₁ son los valores que se calculan a partir de los datos disponibles.

β₀ (Intercepto):

Este indica el valor esperado de la variable dependiente (Y) cuando la variable independiente (X) está en cero.

A veces, el intercepto puede no tener un significado práctico relevante, especialmente si X=0 no se encuentra dentro del rango de datos que se han observado.

β₁ (Pendiente):

Esto muestra el cambio promedio en la variable dependiente (Y) por cada unidad que cambia la variable independiente (X).

Si β₁ es positivo, hay una relación lineal positiva entre X e Y (cuando X aumenta, también lo hace Y).

Por otro lado, si β₁ es negativo, existe una relación lineal negativa (si X crece, Y disminuye).

Si β₁ es cero, indica que no hay una relación lineal entre las variables.

**3. Interpretación de los coeficientes:**

Es fundamental entender los coeficientes en relación con un problema concreto.

Ejemplo:

Imaginemos que estamos estudiando la conexión entre el tiempo dedicado al estudio (X) y la calificación en un examen (Y).

Si nuestro modelo muestra β₀ = 50 y β₁ = 5, esto implica:

El intercepto (β₀ = 50) sugiere que un estudiante que no estudia (0 horas) tendría una calificación de 50 (lo cual podría no ser realista en la vida real).

La pendiente (β₁ = 5) indica que, en promedio, la calificación del examen sube 5 puntos por cada hora extra de estudio.

* Consideraciones importantes:

La interpretación de los coeficientes es válida solamente dentro del rango de datos que se han analizado.

La regresión lineal simple supone que la relación entre las variables es lineal.

Es esencial verificar las suposiciones del modelo de regresión lineal para asegurar que los resultados sean válidos.

**El principio del método de mínimos cuadrados.**

El enfoque de mínimos cuadrados es una técnica clave en los campos de la estadística y las matemáticas, especialmente relevante cuando se trata de la regresión lineal. Su fundamento radica en identificar la mejor representación de un conjunto de datos, buscando minimizar la suma de los cuadrados de los errores o residuos. A continuación, detallo los puntos fundamentales:

1. Concepto principal:

Cuando se dispone de un conjunto de datos que ilustra la relación entre dos variables (como la altura y el peso), no es común que exista una línea recta que trace perfectamente todos los puntos. El método de mínimos cuadrados tiene como objetivo encontrar la línea recta (o una curva en regresiones más elaboradas) que ofrezca el mejor ajuste a esos datos. El término "mejor ajuste" se refiere a que la línea se aproxima lo más posible a todos los puntos en conjunto.

2. Definición de residuos:

Un residuo se entiende como la discrepancia entre el valor real de un dato y el valor estimado por la línea de ajuste. En términos simples, representa la distancia vertical entre un punto y la línea correspondiente.

3. Proceso de minimización de cuadrados:

Este método se enfoca en reducir no solo los residuos, sino específicamente los cuadrados de esos residuos. Esta selección se justifica por diversas razones:

Elimina los resultados negativos, puesto que tanto los residuos positivos como negativos reflejan un error.

Proporciona mayor importancia a los errores más significativos, implicando que la línea de ajuste se ve más influenciada por los puntos más distantes.

El objetivo es determinar los valores de los coeficientes de la línea (la pendiente y el intercepto) de tal manera que la suma de todos los residuos al cuadrado sea la menor posible.

4. Uso en regresión lineal simple:

En el contexto de la regresión lineal simple, el método de mínimos cuadrados se aplica para calcular los coeficientes de la pendiente (β₁) y el intercepto (β₀) de la línea recta que se adapta mejor a los datos. Existen fórmulas matemáticas específicas que posibilitan la obtención directa de estos coeficientes a partir de los datos, minimizando así la función que agrupa la suma de los residuos al cuadrado.

En conclusión, el método de mínimos cuadrados representa una herramienta eficaz para determinar la mejor aproximación lineal a un conjunto de datos. Al concentrarse en reducir la suma de los cuadrados de los residuos, se logra una línea que refleja la relación entre las variables de manera más precisa.

**El diagrama de dispersión ¿Qué es y cuál es su utilidad?**

Un diagrama de dispersión es una herramienta gráfica crucial en el análisis de datos, particularmente valiosa para investigar la conexión entre dos variables cuantitativas. A continuación, se explican sus características y funciones:

1. Definición de un diagrama de dispersión:

* Representación gráfica:

Un diagrama de dispersión ilustra la relación entre dos variables mediante la colocación de puntos en un sistema de coordenadas cartesianas.

Cada punto en el gráfico simboliza un par de valores asociativos de las dos variables.

La variable independiente, también conocida como la predictora, suele representarse en el eje horizontal (eje X), mientras que la variable dependiente, o respuesta, se ubica en el eje vertical (eje Y).

1. Funciones del diagrama de dispersión:

* Visualización de relaciones:

Facilita una representación clara y sencilla de la posible relación entre las dos variables.

Contribuye a la identificación de patrones y tendencias presentes en los datos.

1. Clasificación de tipos de relación:

* Relación lineal:

Cuando los puntos tienden a alinearse en una línea recta, esto sugiere una relación lineal.

Esta relación puede ser positiva, donde los puntos ascienden de izquierda a derecha, o negativa, en la cual los puntos descienden en la misma dirección.

* Relación no lineal:

Si los puntos adoptan una forma curva, esto indica una relación no lineal.

* Falta de relación:

Cuando los puntos están dispuestos de manera aleatoria, sugiere que no hay una relación evidente entre las variables.

* Identificación de valores atípicos:

Permite el descubrimiento de valores atípicos, puntos que se apartan significativamente del patrón general.

Esos valores atípicos pueden afectar el análisis y deben ser evaluados cuidadosamente.

* Evaluación de la fuerza de la relación:

La distribución y el formato de la nube de puntos pueden ofrecer información sobre la fortaleza de la relación.

Una nube de puntos compacta y bien definida indica una relación robusta, mientras que una nube más dispersa sugiere una relación débil.

* Paso previo al análisis de regresión:

Antes de llevar a cabo un análisis de regresión, es aconsejable generar un diagrama de dispersión para confirmar la existencia de una relación lineal y verificar que se cumplan los supuestos del modelo de regresión.

En conclusión, el diagrama de dispersión es una herramienta esencial para investigar y entender la relación entre dos variables, facilitando la identificación de patrones, la detección de valores atípicos y la evaluación de la intensidad de la relación.

**Ajustar un modelo de regresión lineal simple utilizando datos reales. En este sentido deberán buscar una base de datos real y ajustar un modelo de regresión lineal simple usando los datos de dos variables, la cual deben garantizar sea accesible para la consulta y verificación de los datos.**

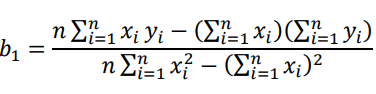
**El INEGI hizo un estudio sobre la situación conyugal de la población en México, registrando el numero de divorcios en 6 años y obteniendo los siguientes resultados.**

https://www.inegi.org.mx/temas/nupcialidad/

|  |  |
| --- | --- |
| **Periodo** | **Divorcios** |
| 2018 | 156,556 |
| 2019 | 160,107 |
| 2020 | 92,739 |
| 2021 | 149,675 |
| 2022 | 166,766 |
| 2023 | 163,587 |

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Periodo (X)** | **Divorcios (Y)** | **XY** | **X^2** | **Y^2** |
|  | 2018 | 156,556 | 315930,008 | 4072324 | 24509,7811 |
|  | 2019 | 160,107 | 323256,033 | 4076361 | 25634,2514 |
|  | 2020 | 92,739 | 187332,78 | 4080400 | 8600,52212 |
|  | 2021 | 149,675 | 302493,175 | 4084441 | 22402,6056 |
|  | 2022 | 166,766 | 337200,852 | 4088484 | 27810,8988 |
|  | 2023 | 163,587 | 330936,501 | 4092529 | 26760,7066 |
| **TOTAL** | **12123** | **889,43** | **1797149,35** | **24494539** | **135718,766** |



3,2019

336,204

105

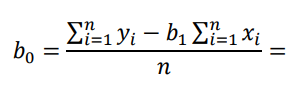
6\*1797149,35-(12123) (889,43)

6\*24494539-(12123)2

=

=

=



-6321,287

=

-37927,723

6

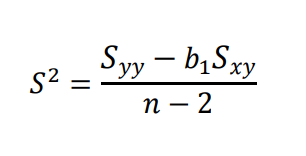
889,43-(3,2019\*12123)

6

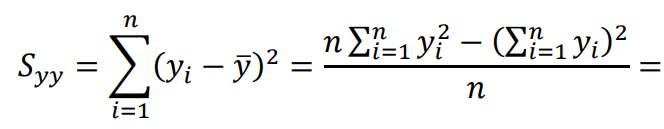
=

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

Estimación de σ es.

****

**Donde:**

****

=

23226,869

6

3871,1448

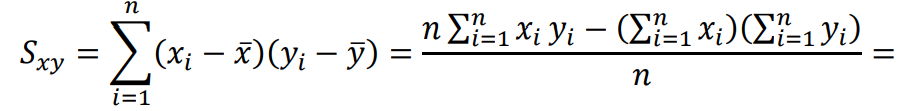
6\*135718,766-(889,43)2

6

=

6\*1797149,35-(12123) (889,43)

6

****

336,204

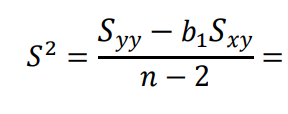
6

=

56,034

=

Así,



922,932

=

3691,727

6-2

=

3871,1448-(3,2019) \*(56,034)

6-2

Entonces, el error estándar de E esta dado por:

****

S= √ 922,932 = 30,380

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 los divorcios promedio tendrá el año 2024**

Para X = Periodo = 2024 los divorcios se estiman a

****

****

= -6318,0853

3,2019(2024) + (-6321,287)