



# **Casa abierta al tiempo**

## **MODELO DE REGRESIÓN LINEAL SIMPLE**

Madin Rivera, Alberto.

Nov 26, 2023

# Índice

<b>1. ¿QUÉ ES EL MODELO DE REGRESIÓN LINEAL SIMPLE?</b>	<b>4</b>
<b>2. LA NATURALEZA DEL MODELO DE REGRESIÓN LINEAL</b>	<b>6</b>
<b>3. EL OBJETIVO DE USAR EL MODELO DE REGRESIÓN LINEAL SIMPLE</b>	<b>8</b>
3.1 CIENCIAS SOCIALES: . . . . .	9
3.2 INVESTIGACIÓN Y RECOLECCIÓN DE DATOS: . . . . .	9
3.3 EMPRESAS Y CONSULTORIAS: . . . . .	9
<b>4. ECUACIÓN POBLACIONAL Y ECUACIÓN MUESTRAL DEL MODELO DE REGRESIÓN LINEAL SIMPLE</b>	<b>10</b>
4.1 ECUACIÓN POBLACIONAL DEL MODELO DE REGRESIÓN LINEAL SIMPLE: . . . . .	11
4.2 ECUACIÓN MUESTRAL DEL MODELO DE REGRESIÓN LINEAL SIMPLE: . . . . .	11
<b>5. ECUACIONES PARA OBTENER LOS PARÁMETROS POBLACIONAL Y MUESTRAL DEL MODELO DE REGRESIÓN LINEAL SIMPLE</b>	<b>13</b>
5.1 ECUACIÓN POBLACIONAL: . . . . .	14
5.2 ECUACIÓN MUESTRAL: . . . . .	15
<b>6. PREDICTOR Y ERROR DEL MODELO DE REGRESIÓN LINEAL SIMPLE POBLACIONAL Y MUESTRAL</b>	<b>16</b>
6.1 PREDICTOR POBLACIONAL: . . . . .	17
6.2 ERROR POBLACIONAL: . . . . .	17
6.3 PREDICTOR MUESTRAL . . . . .	18
6.4 ERROR MUESTRAL: . . . . .	18
<b>7. EL COEFICIENTE DE CORRELACIÓN DEL MODELO DE REGRESIÓN LINEAL SIMPLE</b>	<b>19</b>
7.1 FÓRMULA DEL COEFICIENTE DE CORRELACIÓN (PEARSON): . . . . .	20
7.2 SIGNIFICADO DEL COEFICIENTE DE CORRELACIÓN: . . . . .	20
7.3 INTERPRETACIÓN: . . . . .	20
<b>8. EL COEFICIENTE DE DETERMINACIÓN DEL MODELO DE REGRESIÓN LINEAL SIMPLE</b>	<b>21</b>
8.1 FÓRMULA DEL COEFICIENTE DE DETERMINACIÓN: . . . . .	22
8.2 INTERPRETACIÓN DEL COEFICIENTE DE DETERMINACIÓN: . . . . .	22
<b>9. EJEMPLO VISUAL DEL ANÁLISIS E INTERPRETACIÓN DE LA REGRESIÓN LINEAL SIMPLE</b>	<b>23</b>
9.1 CORRELACIÓN POSITIVA . . . . .	24
9.2 CORRELACIÓN NEGATIVA: . . . . .	24
9.3 CORRELACIÓN NEUTRA: . . . . .	25

<b>10. CASO DE UN MODELO DE REGRESIÓN LINEAL SIMPLE</b>	<b>26</b>
<b>11. LOS ERRORES DENTRO DEL MODELO DE REGRESIÓN LINEAL SIMPLE</b>	<b>29</b>
<b>12. PRUEBAS ESTADÍSTICAS DEL MODELO DE REGRESIÓN LINEAL SIMPLE</b>	<b>32</b>
<b>13. CONCLUSIONES DEL MODELO DE REGRESIÓN LINEAL SIMPLE</b>	<b>34</b>
13.1 CONCLUSIONES DEL ANÁLISIS DE REGRESIÓN LINEAL SIMPLE . .	35
<b>14. BIBLIOGRAFÍA</b>	<b>36</b>

# **1. ¿QUÉ ES EL MODELO DE REGRESIÓN LINEAL SIMPLE?**

La regresión lineal simple es un modelo estadístico que busca establecer la relación lineal entre una variable dependiente (la variable que queremos predecir) y una variable independiente (la variable que utilizamos para hacer la predicción). La forma más simple de la regresión lineal se llama “*regresión lineal simple*”, que implica una única variable independiente.

La ecuación general de la regresión lineal simple es:

$$y = mx + b \quad (1)$$

Donde:

- $y$  es la variable dependiente.
- $x$  es la variable independiente.
- $m$  es la pendiente de la línea de regresión, que representa la tasa de cambio de  $y$  con respecto a  $x$ .
- $b$  es la ordenada al origen, que indica el valor de  $y$  cuando  $x$  es igual a cero.

El objetivo de la regresión lineal simple es encontrar los valores de  $m$  y  $b$  que minimizan la suma de los cuadrados de las diferencias entre los valores observados de  $y$  y los valores predichos por la ecuación de la regresión lineal. Este método se conoce como el método de mínimos cuadrados.

El modelo de regresión lineal simple nace de la necesidad de entender y cuantificar la relación entre dos variables continuas. Es útil cuando se sospecha que hay una relación lineal entre la variable independiente y la variable dependiente.

El proceso de construir un modelo de regresión lineal implica recopilar datos, ajustar el modelo a esos datos y luego evaluar la calidad del ajuste. En la práctica, se utilizan herramientas estadísticas y algoritmos para encontrar los mejores valores para los coeficientes  $m$  y  $b$ .

Es importante destacar que la regresión lineal simple asume que la relación entre las variables es realmente lineal, y puede no ser el modelo más adecuado si la relación es no lineal. En tales casos, se pueden considerar modelos de regresión más complejos, como la regresión polinómica o modelos no lineales.

## **2. LA NATURALEZA DEL MODELO DE REGRESIÓN LINEAL**

La regresión lineal es un modelo estadístico que busca establecer la relación lineal entre una variable dependiente (la variable que se quiere predecir) y una o más variables independientes (las variables utilizadas para hacer la predicción). La naturaleza del modelo de regresión lineal es matemática y estadística. Aquí hay algunos aspectos clave de la naturaleza del modelo de regresión lineal:

1. **Linealidad:** La regresión lineal asume que la relación entre las variables es lineal. Esto significa que se representa mediante una línea recta en un espacio bidimensional o un hiperplano en espacios de dimensiones superiores. La ecuación de la regresión lineal es una combinación lineal de los coeficientes y las variables independientes.
2. **Objetivo de Ajuste:** El objetivo principal del modelo de regresión lineal es ajustar la línea (o hiperplano) de manera que minimice la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos. Este método se conoce como el método de mínimos cuadrados.
3. **Coefficientes:** La regresión lineal busca encontrar los coeficientes (pendientes y términos constantes) que mejor ajusten los datos observados. Estos coeficientes indican la fuerza y dirección de la relación entre las variables independientes y la variable dependiente.
4. **Supuestos:** El modelo de regresión lineal se basa en ciertos supuestos, como la linealidad de la relación, la independencia de los errores, la homocedasticidad de los errores y la normalidad de los errores. Estos supuestos deben ser evaluados y cumplidos para que los resultados del modelo sean válidos.
5. **Interpretación:** Los coeficientes de la regresión lineal tienen interpretaciones específicas. La pendiente indica el cambio en la variable dependiente por unidad de cambio en la variable independiente, y la ordenada al origen indica el valor de la variable dependiente cuando todas las variables independientes son cero.
6. **Inferencia Estadística:** Además de predecir valores, la regresión lineal también permite realizar inferencias estadísticas sobre la significancia de los coeficientes, la bondad del ajuste del modelo y otras propiedades.

Es importante destacar que la regresión lineal es un modelo simplificado y asume una relación lineal entre las variables, lo cual puede no ser siempre el caso en la realidad. En situaciones donde la relación es no lineal, pueden ser más apropiados modelos más complejos, como la regresión polinómica o modelos no lineales.

### **3. EL OBJETIVO DE USAR EL MODELO DE REGRESIÓN LINEAL SIMPLE**



El modelo de regresión lineal simple se utiliza en diversas áreas, incluyendo las ciencias sociales, la investigación y recolección de datos para empresas y consultoras, debido a sus ventajas y su capacidad para modelar relaciones lineales entre variables. Aquí se presentan algunos de los objetivos específicos de utilizar el modelo de regresión lineal simple en estos contextos:

### 3.1 CIENCIAS SOCIALES:

1. **Análisis de Relaciones:** En las ciencias sociales, la regresión lineal simple puede utilizarse para analizar la relación entre dos variables continuas. Por ejemplo, se podría estudiar la relación entre la cantidad de horas de estudio y el rendimiento académico.
2. **Predicción:** El modelo puede ser empleado para predecir valores de una variable en función de otra. Por ejemplo, prever el ingreso de un individuo basado en su nivel educativo.
3. **Identificación de Factores:** Permite identificar y cuantificar la influencia de variables independientes en una variable dependiente. Esto es útil para comprender mejor los factores que afectan ciertos resultados sociales.
4. **Validación de Teorías:** La regresión lineal simple puede ayudar a validar teorías existentes o a generar nuevas hipótesis sobre relaciones causales en ciencias sociales.

### 3.2 INVESTIGACIÓN Y RECOLECCIÓN DE DATOS:

1. **Modelado de Datos:** La regresión lineal simple proporciona un modelo matemático simple que puede ayudar a modelar y entender la estructura de los datos recopilados.
2. **Identificación de Tendencias:** Permite identificar tendencias y patrones en los datos, lo que puede ser crucial para la toma de decisiones informada.
3. **Estimación de Impacto:** Se puede utilizar para estimar el impacto de una variable en otra, lo cual es valioso en la evaluación de intervenciones o cambios.
4. **Optimización:** En empresas y consultoras, la regresión lineal simple se puede utilizar para optimizar procesos y tomar decisiones basadas en datos.

### 3.3 EMPRESAS Y CONSULTORIAS:

1. **Planificación Estratégica:** Ayuda en la planificación estratégica al proporcionar información sobre cómo ciertas variables impactan en los resultados deseados.
2. **Pronósticos y Predicciones:** Facilita la realización de pronósticos y predicciones basadas en datos históricos, lo que es valioso para la planificación y gestión de recursos.
3. **Análisis de Desempeño:** Permite analizar el desempeño de productos, servicios o empleados en función de variables relevantes.
4. **Optimización de Recursos:** Ayuda en la toma de decisiones para la asignación óptima de recursos, como presupuestos de marketing, personal, etc.

## **4. ECUACIÓN POBLACIONAL Y ECUACIÓN MUESTRAL DEL MODELO DE REGRESIÓN LINEAL SIMPLE**

La regresión lineal simple tiene dos ecuaciones principales: la ecuación poblacional y la ecuación muestral.

#### 4.1 ECUACIÓN POBLACIONAL DEL MODELO DE REGRESIÓN LINEAL SIMPLE:

La ecuación poblacional describe la relación verdadera entre las variables en la población completa. En el contexto de la regresión lineal simple, la ecuación poblacional es:

$$Y = \beta_0 + \beta_1 + \varepsilon \quad (2)$$

Donde:

- $Y$  es la variable dependiente.
- $X$  es la variable independiente.
- $\beta_0$  es la ordenada al origen (Intercepto).
- $\beta_1$  es la pendiente de la línea de regresión.
- $\varepsilon$  es el término de error, que representa la variabilidad no explicada por la relación lineal.

La ecuación poblacional establece la relación promedio entre las variables de toda la población y, debido al término de error ( $\varepsilon$ ), reconoce que hay variabilidad en  $Y$  que no puede explicarse completamente mediante  $X$ .

#### 4.2 ECUACIÓN MUESTRAL DEL MODELO DE REGRESIÓN LINEAL SIMPLE:

La ecuación muestral es una estimación basada en una muestra de la población y se utiliza para realizar inferencias sobre la ecuación poblacional. La ecuación muestral es similar a la ecuación poblacional, pero utiliza estimaciones de los parámetros  $\beta_0$  y  $\beta_1$  basadas en los datos de la muestra. La ecuación muestral es:

$$y = b_0 + b_1x + e \quad (3)$$

Donde:

- $y$  es la variable dependiente observada en la muestra.
- $x$  es la variable independiente observada en la muestra.
- $b_0$  es el estimador del intercepto ( $\beta_0$ ).
- $b_1$  es el estimador de la pendiente ( $\beta_1$ ).
- $e$  es el residuo, que representa la diferencia entre los valores observados y los valores predichos ( $y - \hat{y}$ ).

En la práctica, los coeficientes  $b_0$  y  $b_1$  se obtiene a través del método de **mínimos cuadrados ordinarios**, que minimiza la suma de los cuadrados de los residuos<sup>1</sup>.

Estas ecuaciones son fundamentales para entender la relación lineal entre las variables en el modelo de regresión lineal simple y para realizar inferencias sobre la población a partir de datos muestrales.

---

<sup>1</sup>El método de Mínimos Cuadrados Ordinarios (MCO) fue desarrollado de manera independiente por varios matemáticos en el siglo XIX. Uno de los primeros en trabajar en este método fue el matemático francés Adrien-Marie Legendre. Adrien-Marie Legendre presentó el método de Mínimos Cuadrados Ordinarios en su obra “*Nouvelles méthodes pour la détermination des orbites des comètes*” (Nuevos métodos para la determinación de las órbitas de los cometas), publicada en 1805. El objetivo original de Legendre era encontrar una solución para el problema astronómico de determinar las órbitas de los cometas a partir de observaciones astronómicas. Sin embargo, el método resultó ser de aplicación mucho más amplia y se convirtió en una herramienta fundamental en estadísticas y regresión. Aunque Legendre contribuyó significativamente al desarrollo del método de Mínimos Cuadrados, es importante mencionar que otros matemáticos, como Carl Friedrich Gauss, también trabajaron en problemas similares al mismo tiempo. Gauss desarrolló de manera independiente un enfoque similar y publicó sus resultados en 1809 en “*Theoria Motus Corporum Coelestium*” (Teoría del Movimiento de los Cuerpos Celestiales).

## **5. ECUACIONES PARA OBTENER LOS PARÁMETROS POBLACIONAL Y MUESTRAL DEL MODELO DE REGRESIÓN LINEAL SIMPLE**

El método de **Mínimos Cuadrados Ordinarios** (MCO) se utiliza para encontrar los valores de los parámetros ( $\beta_0$  y  $\beta_1$  en la ecuación poblacional) que minimizan la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por el modelo de regresión lineal simple. Aquí la representación de las ecuaciones para obtener los parámetros de la regresión lineal simple poblacional y muestral por el método del MCO:

### 5.1 ECUACIÓN POBLACIONAL:

Dada la ecuación de regresión lineal simple poblacional:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (4)$$

Para obtener los valores de  $\beta_0$  y  $\beta_1$ , se utilizan las siguientes fórmulas derivadas del método de mínimos cuadrados ordinarios:

$$\beta_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (5)$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X} \quad (6)$$

Donde:

- $n$  es el número de observaciones.
- $X_i$  e  $Y_i$  son las observaciones individuales de las variables  $X$  e  $Y$ .
- $\bar{X}$  y  $\bar{Y}$  son las medias de las variables  $X$  e  $Y$ .

## 5.2 ECUACIÓN MUESTRAL:

La ecuación muestral se refiere a los estimadores de los parámetros poblacionales basados en una muestra. Los estimadores son  $b_0$  y  $b_1$ , y se obtienen mediante las siguientes fórmulas:

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (7)$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \quad (8)$$

Donde:

- $n$  es el tamaño de la muestra.
- $x_i$  e  $y_i$  son las observaciones individuales de las variables  $x$  e  $y$ .
- $\bar{x}$  y  $\bar{y}$  son las medias de las variables  $x$  e  $y$ .

Estas fórmulas son utilizadas para calcular los valores específicos de  $\beta_0$  y  $\beta_1$  en el contexto poblacional, o  $b_0$  y  $b_1$  en el contexto muestral, y así ajustar la línea de regresión a los datos. La regresión lineal busca minimizar la suma de los cuadrados de las diferencias entre los valores observados y los predichos, y estas fórmulas son fundamentales para lograrlo mediante el método de mínimos cuadrados ordinarios.

## **6. PREDICTOR Y ERROR DEL MODELO DE REGRESIÓN LINEAL SIMPLE POBLACIONAL Y MUESTRAL**



El predictor en la regresión lineal simple se denota como  $\hat{Y}$  para la población, y  $\hat{y}$  para la muestra. Estos predictores representan las estimaciones de la variable dependiente ( $Y$ ) basadas en los valores de la variable independiente ( $X$ ), utilizando los coeficientes estimados mediante el método de Mínimos Cuadrados Ordinarios. A continuación, se presentan las ecuaciones y el significado asociado:

### 6.1 PREDICTOR POBLACIONAL:

Dada la ecuación del predictor poblacional es:

$$\hat{y} = \beta_0 + \beta_1 X \quad (9)$$

Donde:

- $\hat{Y}$  es el valor predicho de la variable dependiente para la población.
- $X$  es el valor de la variable independiente para la población.
- $\beta_0$  es el estimador del intercepto para la población.
- $\beta_1$  es el estimador de la pendiente para la población.

#### Significado e Interpretación:

- $\hat{Y}$  representa la estimación del valor medio de la variable dependiente para un valor específico de la variable independiente  $X$ .
- $\beta_0$  es el valor esperado de la variable dependiente cuando  $X$  es igual a cero.
- $\beta_1$  indica cómo cambia, en promedio, la variable dependiente por unidad de cambio en la variable independiente.

### 6.2 ERROR POBLACIONAL:

El error en la regresión lineal simple es el término  $\varepsilon$ , que representa la variabilidad no explicada en la relación entre  $X$  e  $Y$ . La ecuación completa incluyendo el error es:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (10)$$

#### Significado e Interpretación:

- $\varepsilon$  es la diferencia entre el valor observado de la variable dependiente y el valor predicho por la regresión.
- $\varepsilon$  captura la variabilidad en  $Y$  que no es explicada por la relación lineal con  $X$ .

### 6.3 PREDICTOR MUESTRAL

Dada la ecuación del predictor muestral es:

$$\hat{y} = b_0 + b_1x \quad (11)$$

- $\hat{y}$  es el valor predicho de la variable dependiente para la muestra.
- $x$  es el valor de la variable independiente en la muestra.
- $b_0$  es el estimador del intercepto en la muestra.
- $b_1$  es el estimador de la pendiente en la muestra.

### 6.4 ERROR MUESTRAL:

El error en la regresión lineal simple para la muestra es el residuo ( $e$ ), que es la diferencia entre el valor observado y el valor predicho por el modelo. La ecuación completa incluyendo el residuo es:

$$y = b_0 + b_1x + e \quad (12)$$

#### Significado e Interpretación:

- $e$  es la diferencia entre el valor observado de la variable dependiente en la muestra y el valor predicho por la regresión.
- $e$  captura la variabilidad en  $\bar{y}$  que no es explicada por la relación lineal con  $\bar{x}$

Estas ecuaciones indican cómo se predice el valor de la variable dependiente para un valor dado de la variable independiente. El objetivo del ajuste de la regresión lineal es minimizar la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos ( $Y_i - \hat{Y}_i$  o  $y_i - \hat{y}_i$ )<sup>2</sup>.

---

<sup>2</sup>El predictor representa la “*mejor estimación*” de la variable dependiente basada en la relación lineal establecida por el modelo de regresión. Es importante señalar que, debido a la presencia del término de error en la regresión, los predictores no son perfectos y siempre habrá algún grado de variabilidad no explicada en los datos.

## **7. EL COEFICIENTE DE CORRELACIÓN DEL MODELO DE REGRESIÓN LINEAL SIMPLE**

El coeficiente de correlación es una medida estadística que describe la fuerza y la dirección de una relación lineal entre dos variables. Este coeficiente, comúnmente denotado como  $r$ , oscila entre -1 y 1. Un valor de 1 indica una correlación positiva perfecta, -1 indica una correlación negativa perfecta, y 0 indica que no hay correlación lineal<sup>3</sup>.

## 7.1 FÓRMULA DEL COEFICIENTE DE CORRELACIÓN (PEARSON):

La fórmula para el coeficiente de correlación de Pearson ( $r$ ) entre dos variables,  $X$  e  $Y$ , es:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (13)$$

Donde:

- $n$  es el número de observaciones.
- $X_i$  e  $Y_i$  son las observaciones individuales de las variables  $X$  e  $Y$ .
- $\bar{X}$  y  $\bar{Y}$  son las medias de las variables de  $X$  e  $Y$ .

## 7.2 SIGNIFICADO DEL COEFICIENTE DE CORRELACIÓN:

- $r = 1$ : Correlación positiva perfecta. A medida que una variable aumenta, la otra también aumenta en una relación perfecta.
- $r = -1$ : Correlación negativa perfecta. A medida que una variable aumenta, la otra disminuye en una relación lineal perfecta.
- $r = 0$ : No hay correlación lineal. No hay una relación lineal sistemática entre las dos variables.
- $0 < |r| < 1$ : Correlación positiva o negativa. Cuanto más cercano a 1 o -1, más fuerte es la correlación lineal.

## 7.3 INTERPRETACIÓN:

- **Signo de  $r$** : El signo indica la dirección de la relación. Positivo significa que las variables se mueven en la misma dirección, y negativo significa que se mueven en direcciones opuestas.
- **Magnitud de  $r$** : Cuanto más cercano a 1 o -1, más fuerte es la relación lineal. Cuanto más cercano a 0, más débil es la relación.

---

<sup>3</sup>Este coeficiente se puede calcular tanto para el modelo poblacional como para el modelo muestral.

## **8. EL COEFICIENTE DE DETERMINACIÓN DEL MODELO DE REGRESIÓN LINEAL SIMPLE**

El coeficiente de determinación, comúnmente representado como  $R^2$ , es una medida estadística que proporciona una indicación de la proporción de la variabilidad en la variable dependiente que es explicada por el modelo de regresión. En otras palabras,  $R^2$  indica la bondad de ajuste del modelo y qué tan bien las variaciones en la variable independiente explican las variaciones en la variable dependiente.

## 8.1 FÓRMULA DEL COEFICIENTE DE DETERMINACIÓN:

El  $R^2$  se calcula utilizando la siguiente fórmula:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (14)$$

Donde:

- $n$  es el número de observaciones.
- $Y_i$  son los valores observados de la variable dependiente.
- $\hat{Y}_i$  son los valores predichos por el modelo de regresión.
- $\bar{Y}$  es la media de los valores observados de la variable dependiente.

## 8.2 INTERPRETACIÓN DEL COEFICIENTE DE DETERMINACIÓN:

- $0 \leq R^2 \leq 1$ :
  - $R^2 = 0$ : El modelo no explica ninguna variabilidad en la variable dependiente.
  - $0 < R^2 < 1$ : El modelo explica una proporción de la variabilidad en la variable dependiente. Cuanto más cercano a 1, mejor es el ajuste del modelo.

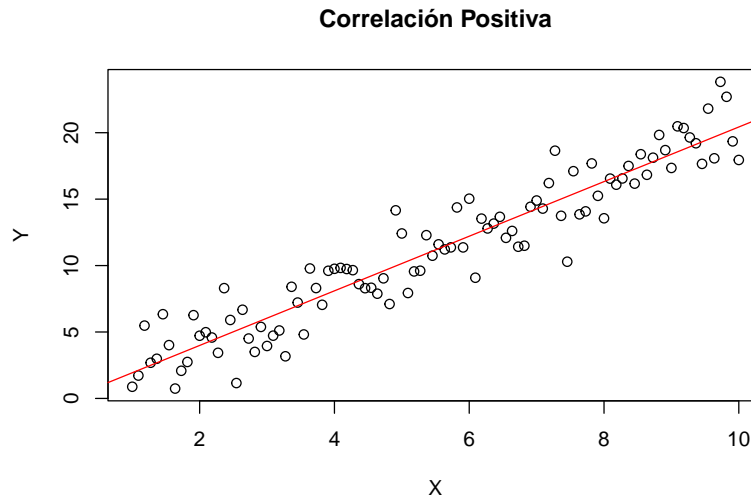
**Significado:**

- Un  $R^2$  cercano a 1 indica que una gran proporción de la variabilidad en la variable dependiente es explicada por el modelo.
- Un  $R^2$  cercano a 0 indica que el modelo no explica bien la variabilidad observada.

## **9. EJEMPLO VISUAL DEL ANÁLISIS E INTERPRETACIÓN DE LA REGRESIÓN LINEAL SIMPLE**

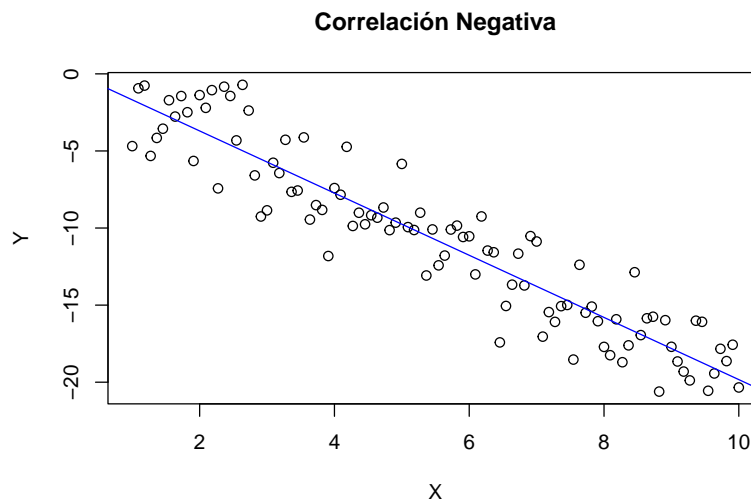
Utilizando ejemplos sencillos de regresión lineal para demostrar los tres tipos de correlaciones, positivo, negativo y neutro, se crearán unos modelos de regresión lineal simple para tener una mejor visualización de datos para el modelo:

## 9.1 CORRELACIÓN POSITIVA



Este es un ejemplo sencillo de 100 observaciones, en el cual se muestra cómo existe una relación **positiva** entre la variable dependiente ( $Y$ ) y la variable independiente ( $X$ ) en cada una de las observaciones de datos.

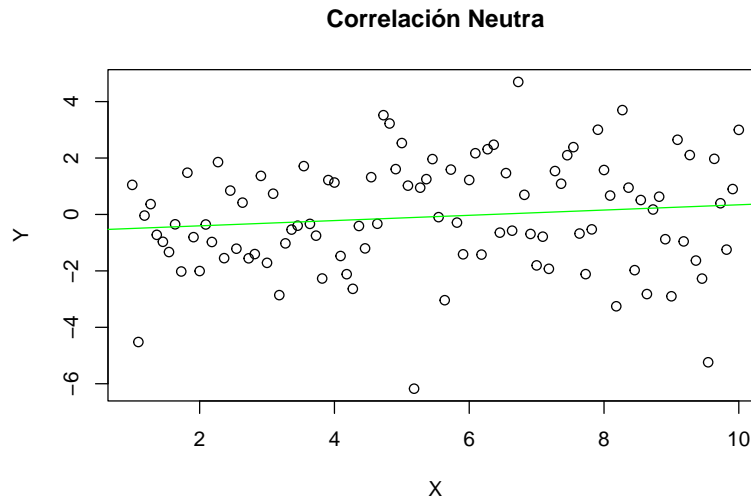
## 9.2 CORRELACIÓN NEGATIVA:



Este es un ejemplo sencillo de 100 observaciones, en el cual se muestra cómo existe una relación **negativa** entre la variable dependiente ( $Y$ ) y la variable independiente ( $X$ ) en cada una de las observaciones de datos.



### 9.3 CORRELACIÓN NEUTRA:

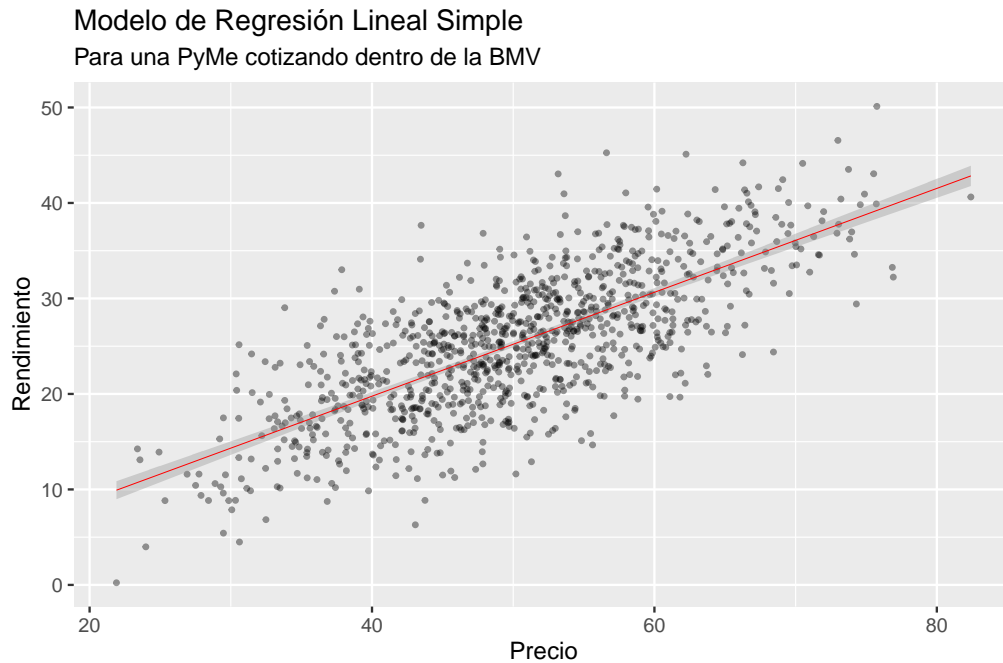


Este es un ejemplo sencillo de 100 observaciones, en el cual se muestra cómo existe una relación **neutra** entre la variable dependiente ( $Y$ ) y la variable independiente ( $X$ ) en cada una de las observaciones de datos.

Estos ejemplos utilizan datos generados artificialmente para ilustrar correlaciones positivas, negativas y neutras. Los gráficos muestran los puntos de datos y las líneas de regresión ajustadas por el modelo lineal. En la práctica, los datos provendrían de observaciones reales y se aplicarían métodos estadísticos para evaluar la fuerza y la significancia de la relación.

## 10. CASO DE UN MODELO DE REGRESIÓN LINEAL SIMPLE

Una vez realizando y revisado toda la metodología para el modelo de regresión lineal simple, se tomará un caso práctico con datos reales. En este caso, se tiene un conjunto de datos de dos variables de una PyMe que cotiza dentro de la Bolsa Mexicana de Valores: “*Rendimiento*” y “*Precio*”. La idea es simular una relación lineal entre el rendimiento y el precio



Fuente: Elaboración propia a partir de datos de BMV (2023)

Este es un caso donde se tiene el “*Precio*” y “*Rendimiento*”. La relación entre ellas es lineal, y se utiliza el método de Mínimos Cuadrados Ordinarios para obtener la ecuación de regresión lineal. En este caso, el modelo fue calculado en lenguaje de programación R y estos son los datos:

#### Output console

Call:

```
lm(formula = Rendimiento ~ Precio, data = finanzas_df)
```

Residuals:

Min	1Q	Median	3Q	Max
-15.1394	-3.4572	0.0213	3.5436	16.4557

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1.99596	0.82085	-2.432	0.0152 *
Precio	0.54402	0.01605	33.888	<2e-16 ***

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```
Residual standard error: 5.032 on 998 degrees of freedom
Multiple R-squared: 0.535, Adjusted R-squared: 0.5346
F-statistic: 1148 on 1 and 998 DF, p-value: < 2.2e-16
```

El modelo de ecuación de regresión lineal simple para la anterior gráfica sería el siguiente:

1. **Coeficientes:**

- Para la variable “*Precio*”, el coeficiente es la pendiente de la línea de regresión ( $b_1$ ). En este caso, el coeficiente de “*Precio*” en el modelo es aproximadamente 0.51.
- El intercepto ( $b_0$ ), que representa el valor esperado de “*Rendimiento*” cuando “*Precio*” es cero, es aproximadamente 1.15.
- La ecuación de regresión lineal sería algo como la siguiente:

$$\text{Rendimiento} \approx 1.15 + 0.51(\text{Precio}) \quad (15)$$

2. **Estadísticas de Prueba**<sup>4</sup>:

- La estadística  $t$  para “*Precio*” y el intercepto proporciona información sobre si estos coeficientes son significativamente diferentes de cero. Un valor  $p$  bajo (generalmente  $< 0.05$ ) sugiere significancia estadística.

3. **Coeficiente de Determinación ( $R^2$ ):**

- El  $R^2$  indica la proporción de variabilidad en “*Rendimiento*” explicada por “*Precio*”. En este caso, el  $R^2$  es un valor que va de 0 a 1, y un  $R^2$  es de 0.535.

La ecuación de regresión lineal sugiere que, en este modelo específico, el “*Rendimiento*” está positivamente relacionado con el “*Precio*”. Cada unidad adicional en el “*Precio*” se asocia, en promedio, con un aumento de aproximadamente 0.51 unidades en el “*Rendimiento*”. El intercepto de 1.15 indica que, cuando el “*Precio*” es cero (lo cual puede no tener un significado práctico en este contexto), el “*Rendimiento*” esperado es de aproximadamente 1.15 unidades.

El  $R^2$  indica que al rededor del porcentaje  $R^2 \times 100$  de la variabilidad en el “*Rendimiento*” se explica por la variabilidad en el “*Precio*” en este modelo. El resto se explica por factores fuera del modelo.

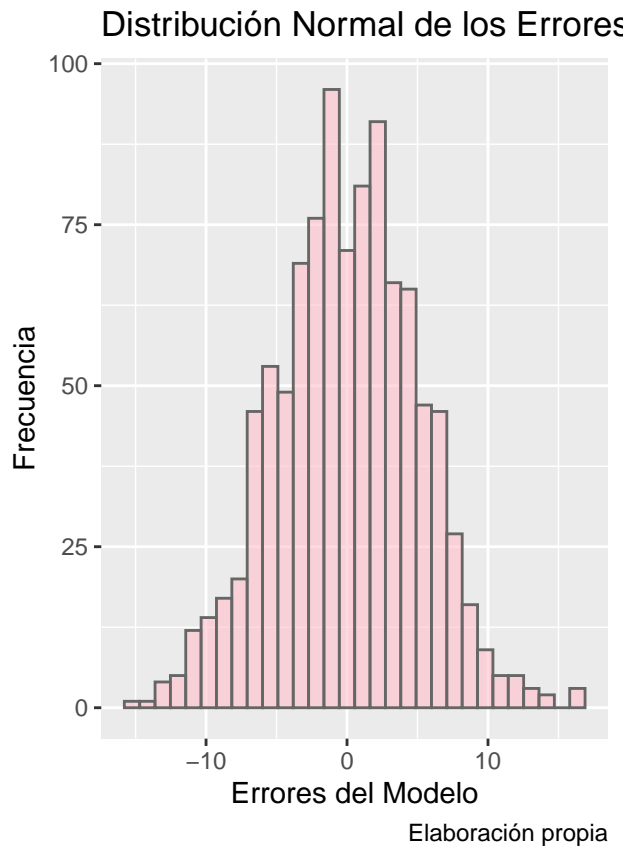
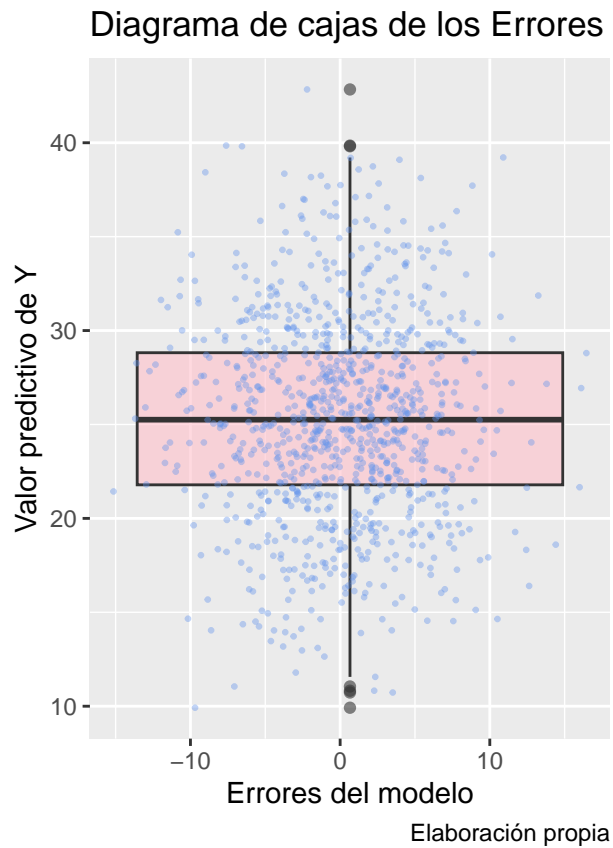
Es importante considerar que estos datos de basan generados para la práctica de una PyMe, las interpretaciones deben de ser considerando el contexto específico de los datos y el problema financiero en cuestion.

---

<sup>4</sup>Las pruebas estadísticas son herramientas fundamentales en la interpretación de los coeficientes de regresión lineal simple. Estas pruebas ayudan a determinar si los coeficientes son significativamente diferentes de cero, lo que implica si hay una relación estadísticamente significativa entre la variable independiente y la variable dependiente. Aquí hay dos pruebas estadísticas comunes utilizadas en la interpretación de los coeficientes de regresión lineal simple.

## **11. LOS ERRORES DENTRO DEL MODELO DE REGRESIÓN LINEAL SIMPLE**

Recordemos que los errores del modelo de regresión lineal simple, son todos aquellos que se obtienen al realizar la diferencia entre el valor observado y el valor predicho,  $y_i - \hat{y}_i$ .



Interpretación:

**1. BOXPLOT (GRÁFICO DE CAJAS):**

- El boxplot te proporciona una visualización de la distribución de los residuos. Puedes observar la mediana, los cuartiles y los valores atípicos potenciales.
- Si el boxplot muestra simetría alrededor de cero y no hay patrones notables, indica que los residuos están distribuidos de manera relativamente uniforme.

**2. HISTOGRAMA:**

- El histograma te permite visualizar la forma de la distribución de los residuos. Puedes observar si siguen una distribución normal.
- La línea roja en el histograma indica la media de los residuos. Si la distribución es aproximadamente normal, la media debería estar cerca de cero.

Estas visualizaciones son útiles para evaluar la validez de las suposiciones del modelo de regresión lineal, como la normalidad y homocedasticidad de los residuos. Si observas patrones no deseados en estas gráficas, puede indicar que el modelo no está capturando completamente la estructura de los datos y que podrían ser necesarios ajustes. Además, los valores atípicos en los residuos pueden indicar observaciones inusuales o errores en el modelo.

## **12. PRUEBAS ESTADÍSTICAS DEL MODELO DE REGRESIÓN LINEAL SIMPLE**



Las pruebas estadísticas son herramientas fundamentales en la interpretación de los coeficientes de regresión lineal simple. Estas pruebas ayudan a determinar si los coeficientes son significativamente diferentes de cero, lo que implica si hay una relación estadísticamente significativa entre la variable independiente y la variable dependiente. Aquí hay dos pruebas estadísticas comunes utilizadas en la interpretación de los coeficientes de regresión lineal simple:

**1. Prueba de Hipótesis para Coeficientes Individuales:**

- La hipótesis nula ( $H_0$ ) es que el coeficiente de la variable independiente es igual a cero ( $b_1 = 0$ ). La hipótesis alternativa ( $H_1$ ) es que el coeficiente es diferente de cero ( $b_1 \neq 0$ ).
- La estadística de prueba  $t$  se calcula dividiendo el estimador del coeficiente ( $b$ ) entre su error estándar. La fórmula es:  $t = \frac{b_1}{sb_1}$
- Si el valor  $p$  asociado con la estadística  $t$  es menor que un nivel de significancia (comúnmente 0.05), rechazamos la hipótesis nula y concluimos que el coeficiente es significativamente diferente de cero.

**2. Prueba de Hipótesis para el Modelo en Conjunto:**

- En lugar de probar un coeficiente específico, esta prueba evalúa la significancia global del modelo de regresión lineal simple.
- La hipótesis nula ( $H_0$ ) es que todos los coeficientes del modelo son igual a cero ( $b_0 = b_1 = 0$ ). La hipótesis alternativa ( $H_1$ ) es que al menos un coeficiente es diferente de cero.
- La estadística de prueba  $F$  se utiliza para esta prueba. Un valor grande de la estadística  $F$  con un valor  $p$  bajo sugiere que al menos una de las variables independientes es significativa en la explicación de la variabilidad en la variable dependiente.

## **13. CONCLUSIONES DEL MODELO DE REGRESIÓN LINEAL SIMPLE**

Las conclusiones de un modelo de regresión lineal simple pueden ser extensas y dependen en gran medida del contexto específico de los datos y el problema que estás abordando. A continuación, proporciono un ejemplo general de cómo podrías estructurar las conclusiones para un modelo de regresión lineal simple. Ten en cuenta que este es un ejemplo genérico y es importante adaptar las conclusiones al propio análisis y contexto:

### **13.1 CONCLUSIONES DEL ANÁLISIS DE REGRESIÓN LINEAL SIMPLE**

En el presente estudio, se llevó a cabo un análisis de regresión lineal simple para investigar la relación entre la variable independiente  $X$  y la variable dependiente  $Y$ . A continuación, se presentan las principales conclusiones derivadas de este análisis:

1. Significancia Estadística.
2. Interpretación de Coeficientes.
3. Bondad de Ajuste.
4. Diagnóstico de Residuos.
5. Limitaciones y Consideraciones Adicionales.
6. Recomendaciones para Futuras Investigaciones.

En resumen, el análisis de regresión lineal simple proporciona evidencia de una relación significativa entre las variables  $X$  e  $Y$ , y el modelo propuesto puede ser útil para predecir  $Y$  en función de  $X$ . Sin embargo, es crucial interpretar estos resultados dentro del contexto específico del problema y considerar las limitaciones del modelo.

## 14. BIBLIOGRAFÍA

- Stock, J. H., & Watson, M. W. (2007). Introduction to econometrics (Vol. 3). Addison-Wesley.
- Gujarati, D. N., & Porter, D. C. (2009). Basic econometrics. McGraw-Hill Education.
- Greene, W. H. (2011). Econometric analysis (7th ed.). Pearson.
- Hayashi, F. (2000). Econometrics. Princeton University Press.
- Crawley, Michael (2013), The R Book, 2a. Ed., Wiley, United Kingdom.
- Quintana, L. y M. A. Mendoza (2008), Econometría Básica. Modelos y aplicaciones a la economía mexicana, Plaza y Valdés Editores, México.

### Recursos en línea

- [Khan Academy](#). 2023.
- [Coursera - Machine Learning y Regresión Lineal](#). 2023