

CLASE

5

ESTADÍSTICA II

Regresión lineal simple I y II

Educación de calidad

INTRO DUCCIÓN

DE LA CLASE

GRADO / POSGRADO / TECNOLOGÍAS

Estudia a tu tiempo
sin interrupciones



En esta quinta semana se explorará la regresión lineal simple como herramienta fundamental para analizar la relación entre variables cuantitativas, iniciando con la comprensión de la correlación, la dispersión de los datos y la construcción de la recta de regresión, profundizando en la interpretación de los coeficientes, la evaluación de su significancia estadística y la aplicación de estos conceptos a situaciones prácticas, de manera que los estudiantes puedan desarrollar habilidades para modelar relaciones lineales, comprender patrones en conjuntos de datos y generar análisis precisos que respalden decisiones fundamentadas en evidencia, fortaleciendo así su capacidad para interpretar resultados de manera crítica y aplicar métodos estadísticos en contextos administrativos, investigativos y profesionales.

De manera complementaria, se abordará la evaluación y aplicación de los modelos de regresión lineal simple mediante indicadores como el coeficiente de determinación R^2 , el error estándar y las técnicas de pronóstico, incorporando la predicción y la interpretación de resultados para la toma de decisiones estratégicas, de modo que los estudiantes puedan consolidar competencias en el análisis predictivo, utilizar los modelos para anticipar comportamientos de variables dependientes y aplicar sus conocimientos en la generación de información confiable y pertinente, promoviendo un enfoque integral que articule teoría, práctica y análisis crítico en la gestión de datos dentro de organizaciones y entornos de investigación.

Educación de calidad

INTRO DUCCIÓN DE LA CLASE

GRADO / POSGRADO / TECNOLOGÍAS

Estudia a tu tiempo
son interrupciones



Resultado o resultados de aprendizaje
que será abordado con el contenido
de la clase.

- 1) Identificar los conceptos, definiciones, propiedades y teoremas de la Estadística Inferencial como base del análisis de datos en el contexto de su área de conocimiento.
- 2) Analizar las técnicas de muestreo, estimadores y los contrastes de hipótesis, en el contexto de su área del conocimiento.
- 3) Interpretar los resultados de la aplicación de técnicas de la Estadística Inferencial en el contexto de su área de conocimiento.

9. Regresión lineal simple I

Esta sección introduce la regresión lineal simple como herramienta estadística para estudiar la relación entre dos variables cuantitativas, permitiendo analizar cómo cambios en la variable independiente afectan a la variable dependiente. Se enfatiza la importancia de comprender la correlación, la dispersión de los datos y la construcción de la recta de regresión como elementos fundamentales para modelar relaciones lineales de manera precisa.

Además, se destaca que la regresión lineal simple proporciona un marco conceptual para interpretar la fuerza y dirección de la relación entre variables, facilitando la toma de decisiones basada en datos y la evaluación de patrones que pueden ser útiles en investigaciones y aplicaciones prácticas.

9.1 Relación entre variables

La regresión lineal simple es un modelo estadístico que describe la relación entre una variable independiente (X) y una variable dependiente (Y) mediante una ecuación lineal. Su objetivo principal es explicar o predecir el comportamiento de Y a partir de X.

En el ámbito empresarial, la regresión lineal simple permite cuantificar cómo un factor específico influye en un resultado clave, facilitando la predicción y la toma de decisiones. Por ejemplo, una empresa puede usar la regresión para estimar cómo el gasto en publicidad (X) afecta las ventas (Y), identificando si un incremento en inversión genera un aumento proporcional en ingresos. Este enfoque proporciona a los gestores herramientas objetivas para planificar presupuestos, proyectar resultados y evaluar estrategias, convirtiendo datos históricos en información accionable que respalde decisiones estratégicas.

9.1.1 Correlación

Antes de ajustar un modelo de regresión, es fundamental analizar la **correlación** entre las dos variables. El coeficiente de correlación de Pearson (r) mide la fuerza y la dirección de la relación lineal:

- r cercano a $+1$ → correlación positiva fuerte
- r cercano a -1 → correlación negativa fuerte
- r cercano a 0 → sin correlación lineal

Sin embargo, la correlación no implica causalidad. Una correlación alta solo indica que las variables se mueven juntas, no que una cause la otra.

GRÁFICO DE DISPERSIÓN CON DIFERENTES NIVELES DE CORRELACIÓN

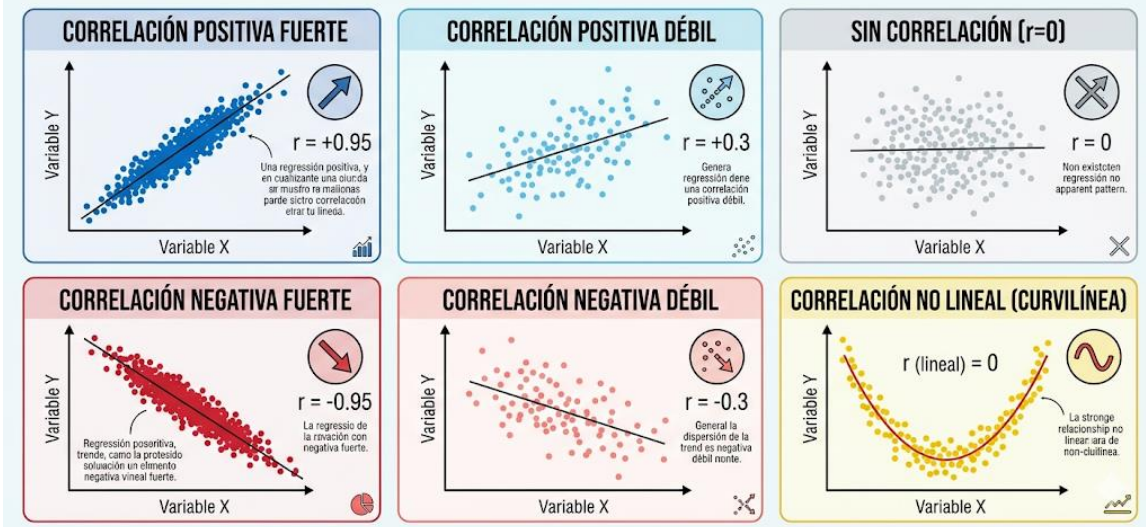


Imagen 1. Gráfico de dispersión con diferentes niveles de correlación (creación propia del autor).

En un contexto empresarial, el análisis de correlación ayuda a identificar relaciones potencialmente útiles entre variables antes de construir un modelo de regresión. Por ejemplo, si se observa una correlación positiva fuerte entre la inversión en publicidad y las ventas, esto sugiere que la publicidad podría ser un predictor relevante para planificar campañas futuras. Sin embargo, es crucial recordar que una correlación alta no garantiza causalidad; factores externos, tendencias de mercado o variables omitidas podrían estar influyendo en la relación observada. Por ello, la correlación sirve como una primera guía para explorar patrones y orientar análisis más profundos, como la regresión, de manera informada.

9.1.2 Dispersión

La dispersión de los puntos alrededor de la recta es clave para evaluar la calidad del modelo. Cuando los puntos están muy cercanos a la recta, la relación es fuerte y el modelo explica bien la variabilidad de Y. Si los puntos están muy dispersos, el modelo tiene menor poder predictivo.

Se utiliza el coeficiente de determinación (R^2) para cuantificar qué porcentaje de la variabilidad de Y es explicada por X. Un R^2 de 0.85 significa que el 85% de los cambios en Y se explican por X.

En el ámbito empresarial, un alto valor de R^2 indica que la variable independiente elegida es un buen predictor de la variable dependiente, lo que fortalece la confianza en las decisiones basadas en el modelo. Por ejemplo, si una empresa encuentra que el 85% de la variabilidad en las ventas se explica por la inversión en publicidad, puede planificar presupuestos y estrategias de marketing con mayor seguridad. Sin embargo, un R^2 bajo no necesariamente invalida el análisis; puede indicar que otras variables importantes no están incluidas en el modelo, lo que sugiere la necesidad de explorar modelos multivariantes o incorporar factores adicionales para mejorar la capacidad predictiva.

9.1.3 Recta

La recta de regresión se representa con la ecuación:

$$\hat{Y} = a + bX$$

donde:

- a = intercepto (valor de Y cuando $X = 0$)
- b = pendiente (cambio en Y por cada unidad de cambio en X)

La recta se obtiene mediante el método de mínimos cuadrados, que minimiza la suma de las distancias verticales (residuos) entre los puntos observados y la recta.

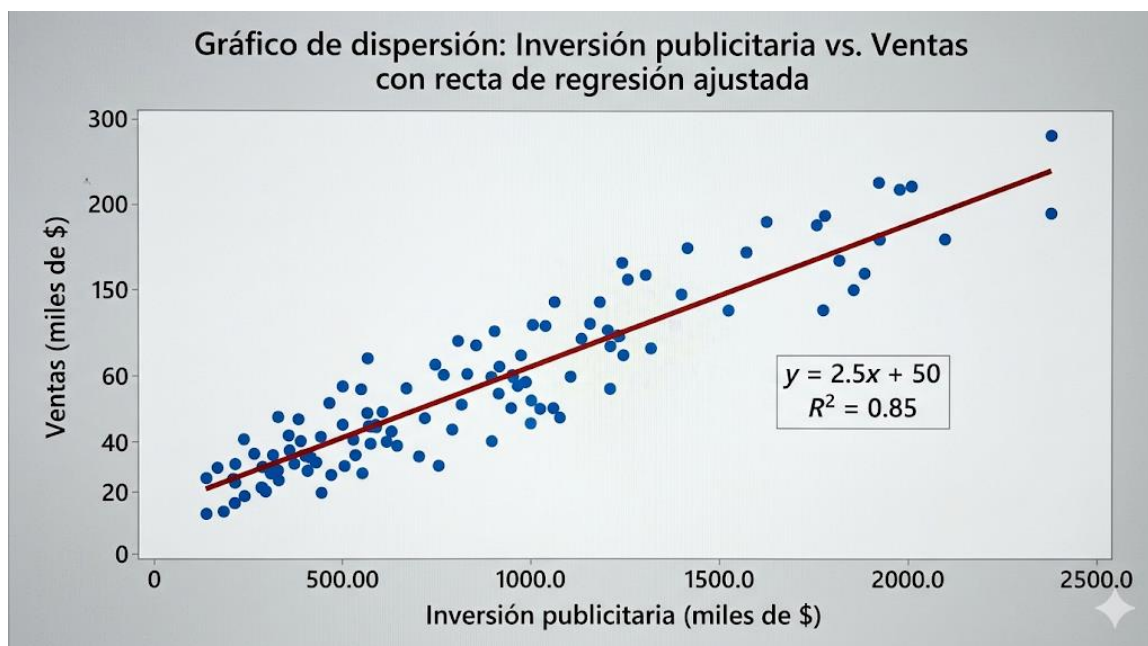


Imagen 2. Recta de regresión ajustada sobre datos de inversión publicitaria vs. ventas (creación propia del autor).

La pendiente b indica cuánto aumenta o disminuye la variable dependiente por cada unidad adicional en la variable independiente, lo que permite cuantificar el impacto real en términos monetarios o de eficiencia.

9.2 Interpretación

Se aborda el análisis de los coeficientes de la recta de regresión y la significancia estadística, lo que permite determinar si la relación observada entre variables es confiable y relevante para los objetivos del estudio. La correcta interpretación de estos resultados es fundamental para reportar conclusiones precisas y fundamentadas.

Además, se presentan aplicaciones prácticas de la regresión lineal simple, mostrando cómo los modelos construidos pueden servir para predicciones, análisis de tendencias y respaldo de decisiones estratégicas en distintos ámbitos académicos y profesionales.

9.2.1 Coeficientes

La interpretación de los coeficientes debe hacerse en el contexto del negocio:

- **Intercepto (a):** Valor esperado de Y cuando $X = 0$. A veces no tiene sentido práctico (por ejemplo, ventas cuando la publicidad es cero), pero sirve como punto de partida.
- **Pendiente (b):** Por cada unidad adicional en X, Y cambia en b unidades (manteniendo todo lo demás constante). Ejemplo: "Por cada \$1.000 adicionales en publicidad, las ventas aumentan en \$4.500".

La interpretación de los coeficientes permite a los gestores cuantificar el impacto de acciones específicas y planificar estrategias basadas en datos. Por ejemplo, conocer que un incremento de \$1.000 en publicidad genera \$4.500 adicionales en ventas ayuda a calcular el retorno sobre

la inversión (ROI) y a priorizar campañas más rentables. Además, analizar el intercepto y la pendiente dentro del contexto del negocio permite ajustar expectativas, identificar límites prácticos del modelo y tomar decisiones más informadas, evitando conclusiones erróneas basadas únicamente en la estadística sin considerar la realidad operativa de la empresa.

9.2.2 Significancia

Para determinar si la relación es estadísticamente significativa se realiza una prueba de hipótesis sobre la pendiente:

- H_0 : $b = 0$ (no existe relación lineal)
- H_1 : $b \neq 0$ (existe relación lineal)

Se utiliza el valor p asociado al coeficiente b o el estadístico t. Si $p < 0.05$, se concluye que la pendiente es significativa y que X influye en Y.

También se evalúa la significancia global del modelo mediante el ANOVA de regresión (F calculado).

En un contexto empresarial, la significancia estadística de la pendiente indica si el factor analizado realmente tiene un efecto relevante sobre la variable de interés, respaldando decisiones basadas en evidencia. Por ejemplo, si la pendiente del gasto en publicidad sobre ventas es significativa ($p < 0.05$), la empresa puede justificar incrementos presupuestales o ajustes en campañas, confiando en que estos cambios impactarán positivamente en los resultados. Asimismo, evaluar la significancia global mediante el ANOVA de regresión permite verificar si el modelo en su conjunto es útil para predecir la variable dependiente, asegurando que las estrategias derivadas del análisis se apoyen en un modelo sólido y confiable.

Tabla de Coeficientes de Regresión							
Modelo	Coeficientes No Estandarizados (B)	Error Estándar	Coeficientes Estandarizados (Beta)	t	p-valor	Intervalo de Confianza para B (95%)	
						Límite Inferior	Límite Superior
1 Constante	21,633	0,069		19,77	0,000**	2,8333	12,4769
Variable X1	0,1369	0,040	0,336	-7,97	0,000**	0,8900	0,9363
Variable X2	0,1466	0,000	0,301	6,42	0,000**	0,7540	0,8070
Variable X3	0,667	0,032	0,726	2,72	0,000**	0,5456	0,7483
Variable X4	-0,224	0,042		-3,38	0,000**	-0,3995	-0,3499
Variable X5	-1,033	0,020		6,63	0,000**	-0,3766	0,1559
Variable X7	0,464	0,013	0,233	-1,72	0,003**	-1,0054	-1,0732
Variable X8	-0,638	0,014	0,089	-1,96	0,000**	-0,2998	-0,3409
Variable X10	-0,658	0,020		-3,58	0,001**	-0,1373	-0,1838
Variable X11	-0,873	0,020		-2,72	0,000**	-0,1272	-0,1366

Imagen 3. Tabla de coeficientes con valores t, p-valor y confianza (creación propia del autor).

9.2.3 Aplicaciones

La regresión lineal simple se aplica ampliamente en:

- Marketing: Predecir ventas a partir de inversión publicitaria o precio.

- Finanzas: Relacionar gastos operativos con ingresos.
- Operaciones: Estimar tiempo de entrega según distancia o volumen.
- Recursos Humanos: Predecir productividad según horas de capacitación o salario.

En la práctica empresarial, la regresión no solo describe relaciones, sino que permite pronosticar escenarios futuros y optimizar presupuestos.

Enlace 1: Este video explica de forma clara y visual la correlación, el diagrama de dispersión y los conceptos básicos de regresión lineal simple. [Ver en YouTube: Regresión Lineal Simple - Explicación Completa](#) **Descripción del enlace:** Este recurso visual te ayudará a entender la diferencia entre correlación y regresión y cómo interpretar la recta de mejor ajuste.

10. Regresión lineal simple II

Esta sección profundiza en la evaluación de la calidad del modelo de regresión lineal simple mediante indicadores como el coeficiente de determinación R^2 , el error estándar y su capacidad de pronóstico. Se destaca la importancia de estos criterios para garantizar que el modelo represente adecuadamente la relación entre variables y sea útil para la predicción y análisis de resultados.

Además, se enfatiza la aplicación del modelo en situaciones prácticas, mostrando cómo la regresión lineal simple permite realizar predicciones confiables, interpretar tendencias de manera fundamentada y tomar decisiones estratégicas basadas en información cuantitativa.

10.1 Evaluación

Se estudian los métodos para evaluar el desempeño del modelo, incluyendo la interpretación del R^2 , el error estándar de estimación y la capacidad de pronóstico. Estos indicadores permiten medir la precisión del modelo y determinar su utilidad para análisis predictivos y toma de decisiones.

Además, se analiza cómo aplicar estos criterios en distintos contextos, asegurando que las conclusiones obtenidas sean válidas y que el modelo pueda ser utilizado de manera efectiva para anticipar resultados futuros.

10.1.1 Supuestos del modelo (evaluación de residuos)

La evaluación comienza con la verificación de los supuestos clásicos del modelo de regresión lineal simple mediante el análisis de residuos (diferencia entre el valor observado y el valor predicho: $e_i = y_i - \hat{y}_i$):

- Linealidad: La relación entre X e Y debe ser lineal (gráfico de residuos vs. valores ajustados debe mostrar patrón aleatorio).
- Normalidad de residuos: Los errores deben distribuirse aproximadamente normal (histograma o gráfico Q-Q).
- Homocedasticidad: La varianza de los residuos debe ser constante (no formar “embudo”).
- Independencia: Los residuos no deben mostrar patrones de autocorrelación.

Si estos supuestos se violan gravemente, el modelo puede requerir transformaciones de variables o considerar alternativas.

Gráficos de Diagnóstico de Residuos: Residuos vs. Valores Ajustados y Q-Q Plot

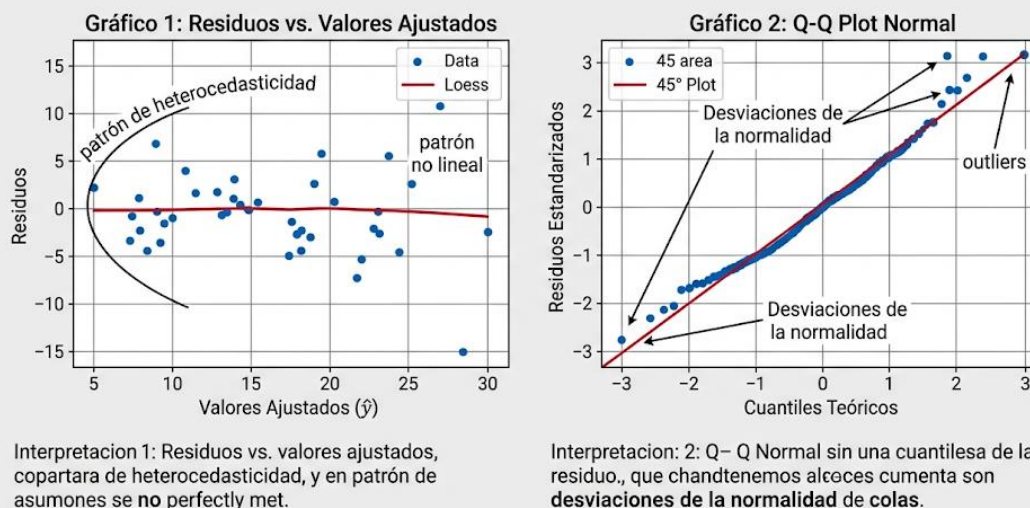


Imagen 4. Gráficos de diagnóstico de residuos: residuos vs. valores ajustados y Q-Q plot (creación propia del autor).

La evaluación de residuos no solo es un requisito estadístico, sino que tiene implicaciones directas en la confiabilidad de las decisiones empresariales basadas en el modelo. Por ejemplo, si se detecta heterocedasticidad o autocorrelación en los residuos, las predicciones sobre ventas, costos o demanda podrían ser inexactas, llevando a presupuestos mal planificados o estrategias ineficaces. Identificar patrones en los residuos permite ajustar el modelo: mediante transformaciones, inclusión de variables adicionales o métodos robustos, asegurando que las estimaciones sean precisas y que las decisiones derivadas, como asignación de recursos o planificación de campañas, estén respaldadas por un análisis sólido y confiable.

10.1.2 Error estándar

El **error estándar de la estimación** (también llamado error típico o residual standard error) mide la precisión del modelo. Se calcula como la desviación estándar de los residuos:

$$s_e = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - 2}}$$

Un valor pequeño de s_e indica que las predicciones están, en promedio, muy cercanas a los valores reales. Este error se utiliza para construir intervalos de confianza y de predicción.

Interpretación del Error Estándar en un Modelo de Ventas vs. Publicidad

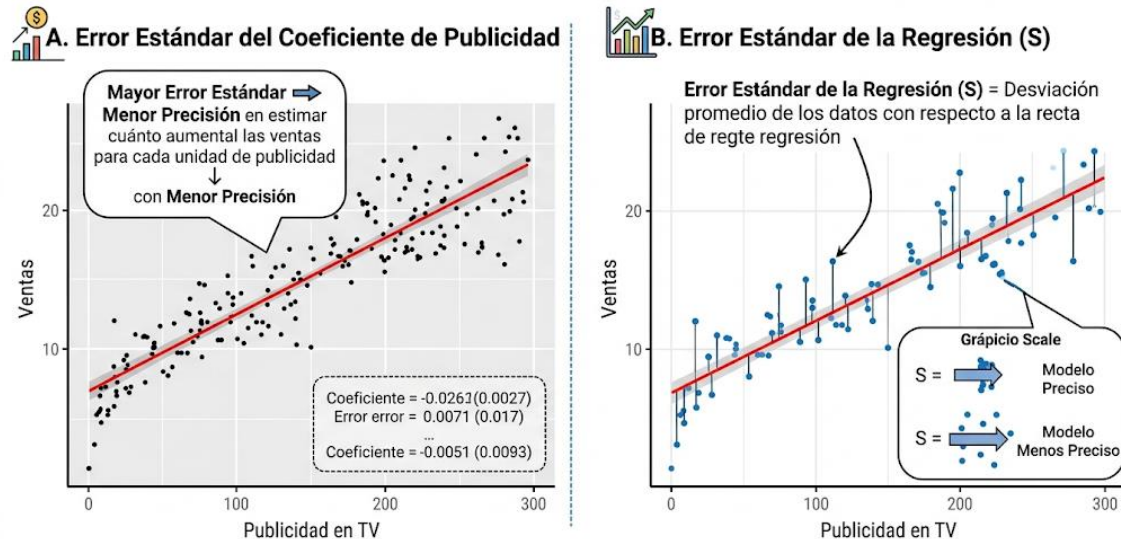


Imagen 5. Interpretación del error estándar en un modelo de ventas vs. publicidad (creación propia del autor).

El error estándar nos indica cuánto error promedio podemos esperar en nuestras predicciones; mientras menor sea, más preciso y confiable será el modelo para la toma de decisiones empresariales.

10.1.3 Pronóstico

El pronóstico (o predicción) consiste en usar la ecuación de regresión para estimar el valor de Y para un valor dado de X. Existen dos tipos:

- **Intervalo de confianza** para la media de Y (más estrecho).
- **Intervalo de predicción** para un valor individual de Y (más ancho, porque incluye la variabilidad individual).

Cuanto más lejos esté el valor de X del promedio de la muestra, mayor será la incertidumbre del pronóstico.

En la práctica empresarial, los pronósticos permiten planificar y tomar decisiones estratégicas basadas en datos históricos. Por ejemplo, una empresa puede estimar las ventas futuras según distintos niveles de inversión en publicidad o prever la demanda de un producto según la tendencia de consumo. La distinción entre intervalo de confianza y de predicción es clave: el primero ayuda a estimar el comportamiento promedio esperado, útil para decisiones estratégicas de planificación, mientras que el segundo considera la variabilidad individual, siendo fundamental para evaluar riesgos o establecer márgenes de seguridad en inventarios y producción. Este enfoque cuantitativo fortalece la capacidad de anticiparse a escenarios futuros y optimizar recursos.

10.2 Aplicación

Se enfatiza la utilización del modelo de regresión lineal simple para realizar predicciones y apoyar decisiones basadas en datos, destacando la relevancia de interpretar correctamente los resultados. La aplicación práctica permite proyectar tendencias, evaluar escenarios futuros y respaldar estrategias en ámbitos académicos y profesionales.

Asimismo, se presentan ejemplos de cómo las predicciones obtenidas a través de la regresión pueden guiar acciones estratégicas y optimizar procesos, demostrando la utilidad de esta herramienta estadística en la resolución de problemas reales.

10.2.1 Predicción

En el contexto empresarial, la regresión se usa para predecir escenarios futuros:

- Predecir ventas del próximo trimestre según presupuesto de marketing.
- Estimar costos operativos según volumen de producción.
- Pronosticar tiempo de entrega según distancia o cantidad de pedidos.

Es fundamental acompañar toda predicción con sus intervalos para comunicar el nivel de incertidumbre.

10.2.2 Interpretación

Una buena interpretación combina:

- El coeficiente de determinación R^2 (porcentaje de variabilidad explicada).
- La significancia de la pendiente.
- El error estándar.
- El análisis de residuos (¿el modelo es válido?).
- Limitaciones: correlación no es causalidad, extrapolación fuera del rango de datos puede ser riesgosa.

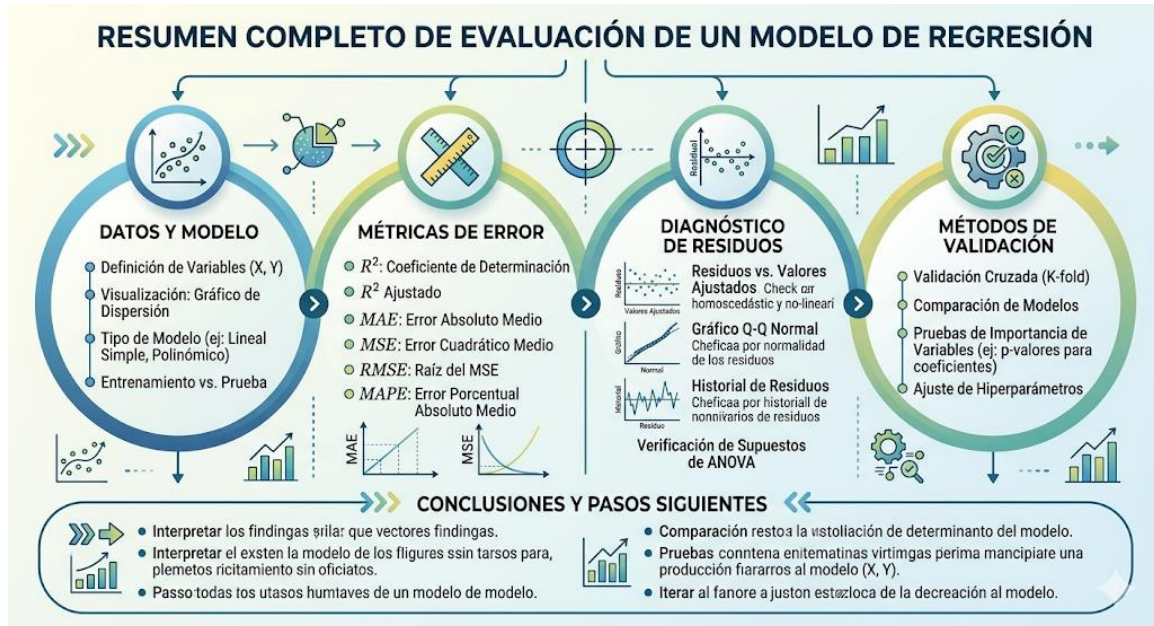


Imagen 6. Resumen completo de evaluación de un modelo de regresión (creación propia del autor).

Siempre recuerda: un modelo con alto R^2 pero residuos con patrones claros no es confiable. La validación mediante residuos es tan importante como los coeficientes.

10.2.3 Decisiones

La regresión lineal simple apoya decisiones estratégicas tales como:

- Asignación óptima de presupuesto publicitario para maximizar ventas.
- Determinación de precios que equilibren volumen y margen.
- Evaluación del retorno de inversión en programas de capacitación.
- Planificación de capacidad productiva basada en pronósticos de demanda.

La clave es combinar el modelo estadístico con el conocimiento del negocio y considerar escenarios sensibles (¿qué pasa si la relación cambia?).

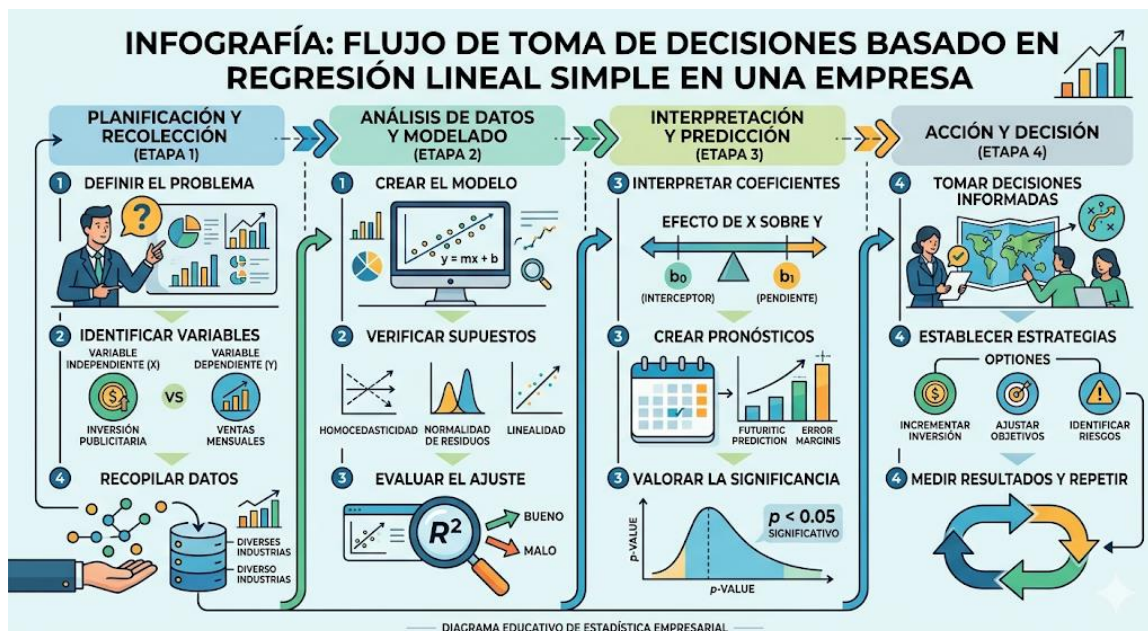


Imagen 7. Infografía: Flujo de toma de decisiones basado en regresión lineal simple en una empresa (creación propia del autor).

Enlace 2: Tutorial práctico sobre regresión lineal simple aplicada a ventas versus publicidad, con énfasis en predicciones e interpretación de resultados. [Ver en YouTube: Regresión Lineal Simple - Ventas vs Publicidad](#) Descripción del enlace: Ideal para ver un caso completo de aplicación empresarial, incluyendo evaluación del modelo y uso de pronósticos para decisiones.

RE FE REN CIAS

- Field, A. (2018). Discovering statistics using IBM SPSS statistics (5th ed.). SAGE Publications.
- Montgomery, D. C., & Runger, G. C. (2018). Applied statistics and probability for engineers (7th ed.). Wiley.
- Levine, D. M., Stephan, D. F., Krehbiel, T. C., & Berenson, M. L. (2021). Statistics for managers using Microsoft Excel (9th ed.). Pearson.
- Draper, N. R., & Smith, H. (1998). Applied regression analysis (3rd ed.). Wiley.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). Applied linear statistical models (5th ed.). McGraw-Hill/Irwin.

Definición de los términos citados en la semana 5

- **Correlación:** Medida estadística que indica la fuerza y dirección de la relación lineal entre dos variables cuantitativas; no implica causalidad.
- **Error estándar de la estimación:** Medida de la precisión del modelo de regresión que representa la desviación estándar de los residuos; indica cuánto se alejan, en promedio, las observaciones reales de los valores predichos por el modelo.

Recursos de Profundización

1. Simulador de Regresión Lineal

Descripción del enlace: Permite visualizar la relación entre variables, la recta de regresión y la correlación.

Autor: Carlos Julio Mayorga Arias

Enlace: [RETO 1](#)

2. Análisis de Correlación y Dispersión

Descripción del enlace: Herramienta para explorar relaciones lineales mediante diagramas de dispersión.



Autor: Carlos Julio Mayorga Arias

Enlace: [RETO 2](#)

3. Evaluación del Modelo (R^2 y Error)

Descripción del enlace: Simulador que permite analizar el ajuste del modelo mediante R^2 y error estándar.

Autor: Carlos Julio Mayorga Arias

Enlace: [RETO 3](#)



síguenos en:



www.pucevirtual.puce.edu.ec