

Taller 1

Integrantes

Laura Alejandra Becerra Feo
Laura Jisley Gutiérrez Ballesteros
Karim Liseth Lopez Lesmes

Docente

Fredy Alexander Orjuela Lopez

Sistemas avanzados de producción
Universidad ECCI
Ingeniería industrial

Introducción

El presente trabajo tiene como propósito analizar la relación entre la inversión publicitaria en televisión, radio y prensa escrita, y el comportamiento de las ventas, utilizando el conjunto de datos *Advertising*. A través de herramientas de estadística descriptiva y análisis, se caracterizó la distribución de las variables, evaluando medidas de tendencia central, dispersión, forma y presencia de valores atípicos. Posteriormente, se estimó un modelo de regresión lineal múltiple con el fin de cuantificar el efecto marginal de cada medio publicitario sobre las ventas y evaluar la capacidad explicativa del modelo.

El análisis no se limita a la obtención de resultados numéricos, sino que busca interpretar los hallazgos desde una perspectiva estadística y de negocio, permitiendo comprender qué medio genera mayor impacto y qué tan robusta es la relación entre la inversión publicitaria y el desempeño comercial.

Fase 1: Estadística Descriptiva y Análisis Exploratorio

Medidas de Tendencia Central y Dispersión

Variable	Media	Mediana	Desv. Estándar	Mínimo	Máximo	(skewness)	Curtosis
TV	147.0425	149.75	85.854236	0.7	296.4	-0.069853	-1.226495
Radio	23.2640	22.90	14.846809	0.0	49.6	0.094175	-1.260401
Newspaper	30.5540	25.75	21.778621	0.3	114.0	0.894720	0.649502
Sales	14.0225	12.90	5.217457	1.6	27.0	0.407571	-0.408869

Tabla [1]: En esta tabla se presentan las medidas de tendencia central y de dispersión correspondientes a las variables independientes, así como a la variable dependiente *Sales*, con el fin de describir su comportamiento estadístico.

Análisis de distribución y atípicos

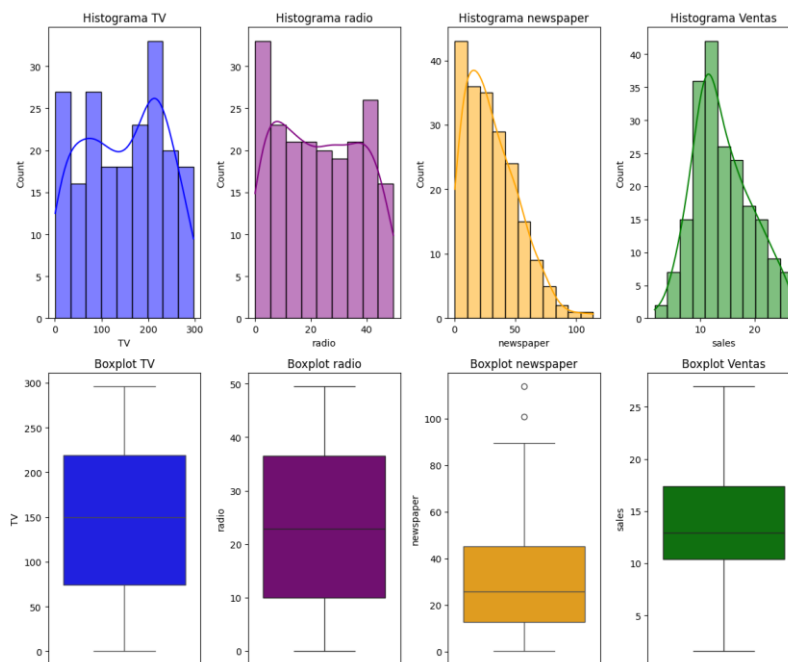


Figura [1]: La figura muestra, en la sección superior, los histogramas de las variables *TV*, *Radio*, *Newspaper* y *Ventas*, permitiendo analizar su comportamiento distributivo.

En la sección inferior se presentan los diagramas de caja, los cuales permiten evaluar la variabilidad, la simetría y la presencia de posibles valores atípicos.

Simetría de las distribuciones

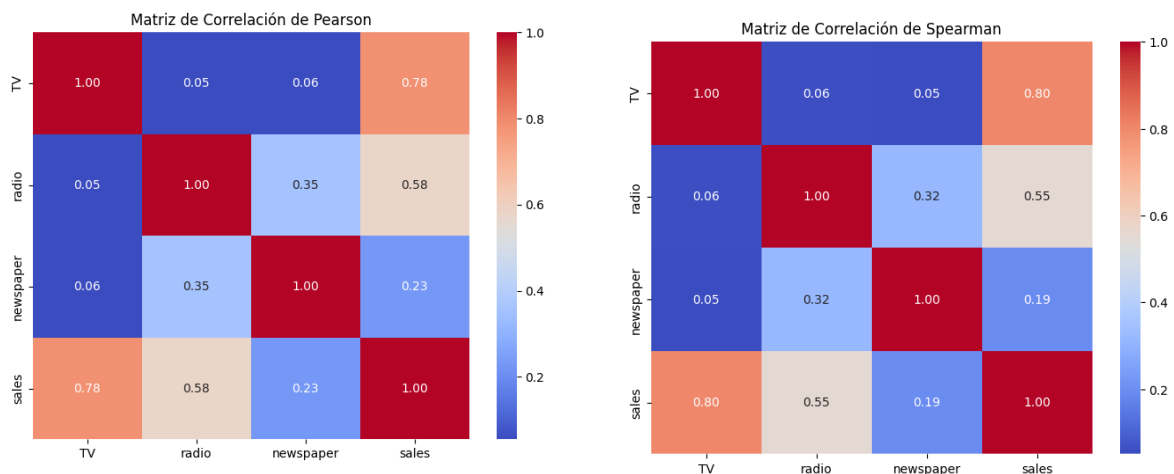
- Para la variable de TV (Skewness = -0.0699) indica una distribución aproximadamente simétrica con una ligera simetría negativa, donde este valor al estar muy cercano a cero sugiere que los datos se distribuyen de manera equilibrada alrededor de la media.
- Para la variable Radio (Skewness = 0.0942) se refleja una asimetría positiva leve, el valor es cercano a cero y la distribución es simétrica, pero con una ligera concentración de valores en el extremo inferior y una pequeña cola hacia la derecha.
- Para la variable Newspaper (skewness = 0.8947) el coeficiente tiene una asimetría positiva moderada donde indica que la distribución tiene una cola extendida hacia la derecha y sugiere la presencia de valores altos que se alejan del centro de la distribución.

- Para la variable Sales (skewness = 0.4076) se presenta una asimetría positiva leve donde indica que la mayoría de los datos se concentran en valores inferiores o cercanos a la media con unos valores mayores que generan inclinación a la derecha.

¿Se identifican datos atípicos u observaciones que se alejan significativamente de la masa de los datos? Explique cómo podrían influir estas observaciones en la trayectoria de una línea de regresión

Al realizar el análisis de los histogramas y los diagramas de caja (boxplots), se identifican datos atípicos en la variable newspaper, se observan valores que se alejan considerablemente del conjunto principal de los datos, especialmente en el extremo superior de la distribución. En las variables TV, radio y sales no se evidencian valores atípicos significativos, ya que los datos se encuentran distribuidos de manera homogénea dentro de los límites establecidos por el rango. La presencia de estas observaciones atípicas puede influir de manera importante en la estimación de un modelo de regresión lineal. Dado que este modelo se fundamenta en el método de mínimos cuadrados, los valores extremos tienden a ejercer una influencia desproporcionada sobre la pendiente y el intercepto de la recta de regresión. Como consecuencia, la línea ajustada puede desplazarse hacia dichos puntos atípicos, generando una representación distorsionada de la tendencia real de los datos.

Evaluación de Asociación



Interpretación de Resultados

- **¿Qué significa un coeficiente cercano a 1 o -1 en el contexto de la inversión publicitaria y el retorno en ventas?**

Un coeficiente cercano a 1 indica la existencia de una relación fuerte y positiva entre la inversión publicitaria y las ventas. Esto significa que, a medida que aumenta la inversión en publicidad, las ventas también tienden a incrementarse de manera consistente.

Por otro lado, un coeficiente cercano a -1 señala una relación fuerte y negativa, lo que implica que, al aumentar la inversión en publicidad, las ventas tienden a disminuir de forma significativa.

- **¿Como se interpreta un valor de correlación cercano a 0? Compare ambos coeficientes (Pearson y Spearman).**

Un valor cercano a 0 indica que no existe una relación clara entre las variables. La diferencia entre ambos coeficientes radica en que Pearson mide la intensidad de una relación lineal, mientras que Spearman evalúa una relación monotónica, es decir, si las variables tienden a aumentar o disminuir conjuntamente, aunque no sigan necesariamente una línea recta.

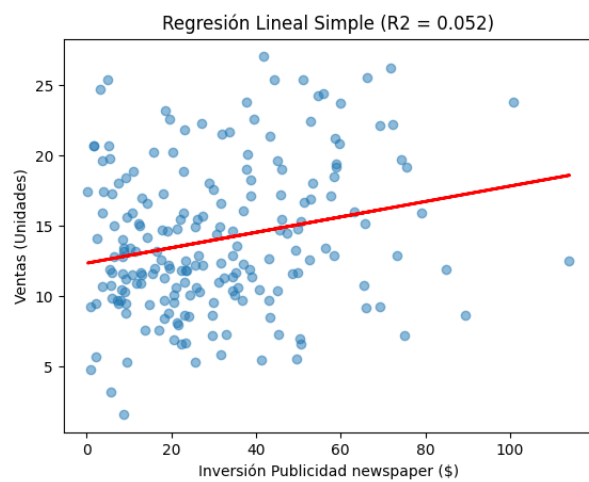
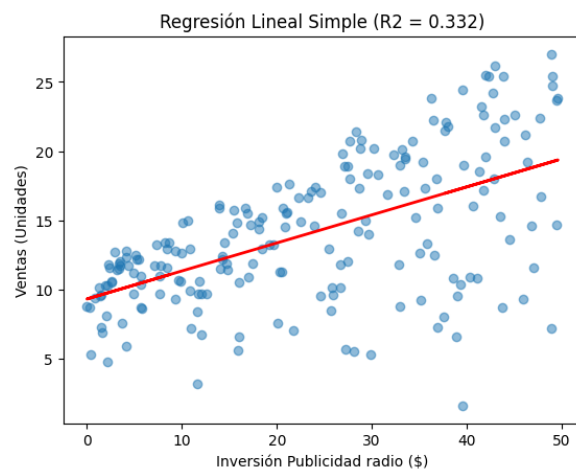
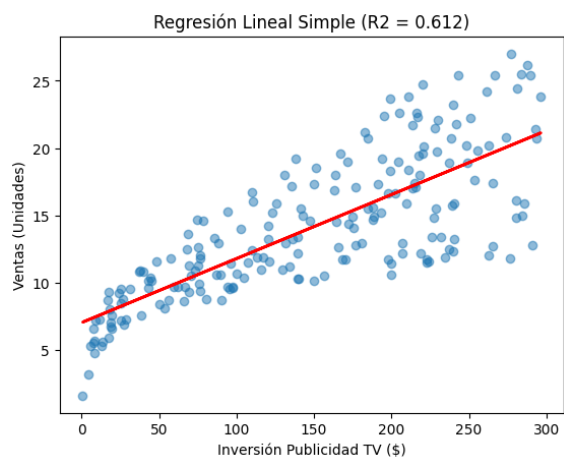
Por lo tanto, Spearman no requiere que la relación sea estrictamente lineal, sino que exista una tendencia consistente (creciente o decreciente).

- **¿Sugieren estos resultados que las relaciones son estrictamente lineales? Justifique su respuesta basándose en la forma de la nube de puntos observada.**

Los resultados indican que las relaciones son principalmente lineales, aunque no lo sean estrictamente. Esto se debe a que los coeficientes de Pearson y Spearman presentan valores muy similares en todos los casos. Cuando ambos coeficientes son cercanos, significa que la relación no solo es monotónica (es decir, siempre creciente), sino que además se ajusta adecuadamente a un patrón lineal.

Al observar la nube de puntos, se aprecia una tendencia ascendente con una distribución de los datos cercana a una línea recta y sin presencia de curvaturas marcadas. Esto permite interpretar que la relación entre las variables es aproximadamente lineal.

Fase 2: Regresión Lineal y Diagnostico



Modelamiento múltiple

- **Interpretación de parámetros:** El modelo de regresión lineal múltiple estimado es:

$$\widehat{Sales} = 2.9389 + 0.0458(TV) + 0.1885(Radio) - 0.0010(Newspaper)$$

Desde una perspectiva estadística, la interpretación de los coeficientes es la siguiente:

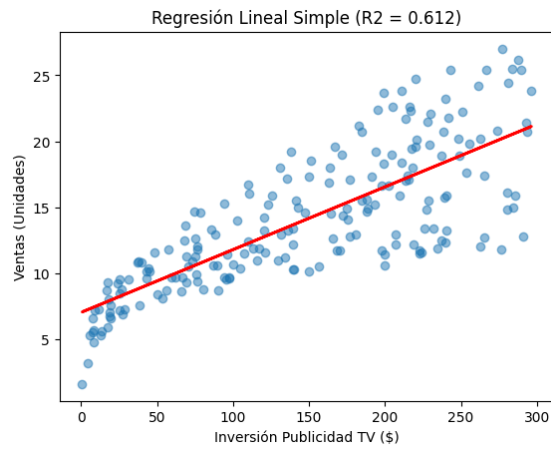
El intercepto ($\beta_0 = 2.9389$) representa el valor esperado de la variable dependiente (*Sales*) cuando todas las variables explicativas (TV, Radio y Newspaper) toman el valor cero. En este contexto, el modelo estima que, en ausencia de inversión publicitaria, las ventas promedio serían aproximadamente 2.94 unidades.

El coeficiente asociado a TV ($\beta_1 = 0.0458$) indica el efecto marginal parcial de la inversión en televisión sobre las ventas. Es decir, manteniendo constantes las demás variables, un incremento de una unidad monetaria en la inversión en TV genera un aumento promedio de 0.0458 unidades en las ventas. Esto evidencia una relación lineal positiva entre ambas variables.

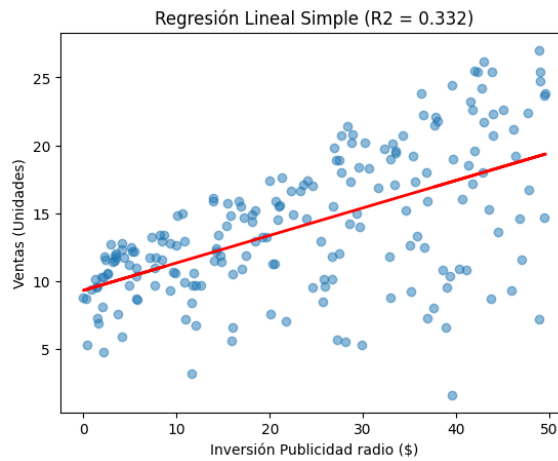
El coeficiente asociado a Radio ($\beta_2 = 0.1885$) representa el cambio esperado en las ventas ante un incremento de una unidad en la inversión en radio, manteniendo constantes las demás variables. Estadísticamente, por cada unidad adicional invertida en radio, las ventas aumentan en promedio 0.1885 unidades. Este es el coeficiente de mayor magnitud, lo que sugiere que, dentro del modelo, radio tiene el mayor impacto marginal sobre las ventas.

Finalmente, el coeficiente de Newspaper ($\beta_3 = -0.0010$) indica que, manteniendo constantes las demás variables, un incremento de una unidad en la inversión en periódico se asocia con una disminución promedio de 0.0010 unidades en las ventas. Sin embargo, debido a su magnitud cercana a cero, este efecto puede considerarse prácticamente nulo en términos prácticos, lo que sugiere una influencia muy débil dentro del modelo.

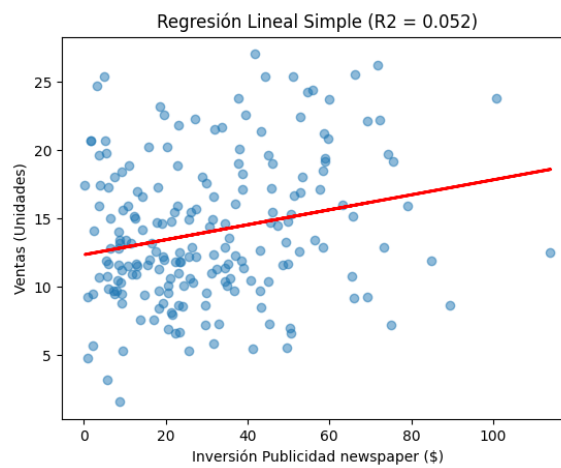
- **Análisis de Bondad de ajuste**



TV



Radio



Newspaper

El Error Cuadrático Medio ($MSE = 2.7841$) indica que, en promedio, el modelo comete un error cuadrático de aproximadamente 2.78 unidades al estimar las ventas. Este valor refleja la variabilidad de los errores y penaliza con mayor peso los errores grandes debido a que se encuentran elevados al cuadrado. Por su parte, la Raíz del Error Cuadrático Medio ($RMSE = 1.6686$) representa el error promedio de predicción en las mismas unidades de la variable *Sales*. En términos prácticos, esto significa que las predicciones del modelo se desvían, en promedio, alrededor de 1.67 unidades respecto a los valores reales de ventas, lo que permite dimensionar de manera más clara la precisión del modelo.

R² interpretación

El coeficiente de determinación ($R^2 = 0.8972$) indica que el 89.72% de la variación total observada en la variable *Sales* es explicada por el conjunto de variables independientes incluidas en el modelo (TV, Radio y Newspaper). En términos prácticos, esto significa que el modelo logra capturar casi el 90% del comportamiento de las ventas a partir de la inversión en publicidad en estos medios.

Este valor se considera alto, lo que sugiere que existe una fuerte relación lineal entre las variables explicativas y las ventas, y que el ajuste del modelo es adecuado. Sin embargo, el porcentaje restante (10.28%) corresponde a factores que no están siendo considerados en el modelo, como otras estrategias de marketing, condiciones del mercado, competencia o variaciones aleatorias.

Es importante tener en cuenta que un R^2 alto no garantiza que el modelo sea perfecto ni que exista causalidad, sino que indica qué tan bien las variables incluidas explican la variabilidad de la variable dependiente dentro de la muestra analizada.

- **Estimación Matricial**

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix}, \quad Y = \begin{pmatrix} 2 \\ 4 \\ 6 \\ 7 \\ 9 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 1+1+1+1+1 & 1+2+3+4+5 \\ 1+2+3+4+5 & 1^2+2^2+3^2+4^2+5^2 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 5 & 15 \\ 15 & 55 \end{pmatrix}$$

Inversa

$$|X'X| = (5)(55) - (15)(15)$$

$$|X'X| = 275 - 225 = 50$$

Matriz 2X2

$$(X'X)^{-1} = \frac{1}{50} \begin{pmatrix} 55 & -15 \\ -15 & 5 \end{pmatrix}$$

$$(X'X)^{-1} = \begin{pmatrix} 1.1 & -0.3 \\ -0.3 & 0.1 \end{pmatrix}$$

Se calcula el producto X'Y

$$X'Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \\ 6 \\ 7 \\ 9 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 2+4+6+7+9 \\ (1)(2) + (2)(4) + (3)(6) + (4)(7) + (5)(9) \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 28 \\ 101 \end{pmatrix}$$

Vector de parámetros

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$\hat{\beta} = \begin{pmatrix} 1.1 & -0.3 \\ -0.3 & 0.1 \end{pmatrix} \begin{pmatrix} 28 \\ 101 \end{pmatrix}$$

$$\hat{\beta}_0 = (1.1)(28) + (-0.3)(101)$$

$$\hat{\beta}_0 = 30.8 - 30.3 = 0.5$$

$$\hat{\beta}_1 = (-0.3)(28) + (0.1)(101)$$

$$\hat{\beta}_1 = -8.4 + 10.1 = 1.7$$

$$\hat{\beta} = \begin{pmatrix} 0.5 \\ 1.7 \end{pmatrix}$$

$$\hat{y} = 0.5 + 1.7x$$

Conclusiones

- El análisis exploratorio permitió identificar que las variables TV y Radio presentan distribuciones estables y sin valores atípicos relevantes, mientras que Newspaper muestra cierta asimetría positiva y presencia de valores extremos que podrían influir en la estimación del modelo.
- Las correlaciones evidenciaron una relación positiva entre la inversión publicitaria y las ventas, especialmente en los medios electrónicos. La similitud entre los coeficientes de Pearson y Spearman confirmó que la relación observada es predominantemente lineal.
- El modelo de regresión lineal múltiple mostró que la inversión en radio tiene el mayor impacto marginal sobre las ventas, seguida por televisión, mientras que la inversión en prensa presenta un efecto prácticamente nulo. El coeficiente de determinación ($R^2 = 0.8972$) indica que el modelo explica el 89.72% de la variabilidad de las ventas, lo que refleja un ajuste fuerte.

En conclusión, la inversión en radio y televisión resulta más efectiva para incrementar las ventas dentro del contexto analizado, y la regresión lineal múltiple se consolida como una herramienta adecuada para apoyar la toma de decisiones estratégicas en la asignación del presupuesto publicitario.