

# Análisis de caso — Taller II

1<sup>st</sup> Mauro Alonso Gonzalez Figueroa  
Universidad Tecnologica de Bolivar  
UTB  
Cartagena, Colombia  
maugonzalez@utb.edu.co  
T00067622

2<sup>st</sup> María Valentina Serna González  
Universidad Tecnologica de Bolivar  
UTB  
Cartagena, Colombia  
maserna@utb.edu.co  
T00067756

**Abstract**—This study presents a statistical analysis of the given case, examining the relationship between sales and advertising expenditures. Data on *Sales* and *Advertising* were collected for a sample of 60 companies and analyzed using Excel and Statgraphics.

The results of the study showed a positive and moderate correlation between sales and advertising, with a determination coefficient ( $R^2$ ) of 0,6343. The covariance was 5,3493, and Pearson's test indicated a statistical significance of 0,7964.

Based on these results, we conclude that sales and advertising expenditures are positively associated. This study provides empirical evidence on the connection between advertising and sales in this case.

**Index Terms**—Determination coefficient, Covariance, Pearson Test, Statistical Analysis

## I. INTRODUCCIÓN

El presente estudio tiene como objetivo analizar la relación entre dos variables: Ventas y Publicidad, utilizando un conjunto de datos de 60 empresas. Se aplicarán métodos estadísticos para determinar si existe una correlación significativa entre las variables y comprender la naturaleza de la relación.

El análisis se basa en los siguientes métodos:

- Coeficiente de determinación: Medirá la fuerza y la dirección de la relación lineal entre las variables [1].
- Covarianza: Indicará la variación conjunta de las variables [2].
- Test de Pearson: Evaluará la significancia estadística de la correlación [3].

Los resultados de este estudio proporcionarán información valiosa sobre la relación entre las variables en el contexto específico del conjunto de datos. Se espera que la investigación contribuya a una mejor comprensión de cómo la publicidad puede influir en las ventas, con aplicaciones potenciales para la toma de decisiones estratégicas en el ámbito empresarial.

## II. FORMULAS UTILIZADAS

- Covarianza:

$$C_{xy} = \frac{\sum xy - N\bar{x}\bar{y}}{N - 1} \quad (1)$$

- Test de Pearson:

$$r_{xy} = \frac{C_{xy}}{S_x S_y} \quad (2)$$

$$S_x^2 = \frac{\sum x^2 - N(\bar{x})^2}{N - 1} \quad (3)$$

$$S_y^2 = \frac{\sum y^2 - N(\bar{y})^2}{N - 1} \quad (4)$$

$$\therefore r_{xy} = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum(x^2) - (\sum x)^2][N\sum(y^2) - (\sum y)^2]}} \quad (5)$$

- Coeficiente de determinación:

$$R^2 = r_{xy}^2 \quad (6)$$

- Recta modelo:

$$y = mx + b \quad (7)$$

$$m = \frac{S_{xy}}{S_x} = \frac{N\sum xy - (\sum x)(\sum y)}{N\sum(x^2) - (\sum x)^2} \quad (8)$$

$$b = \bar{y} - m\bar{x} = \frac{\sum y}{N} - m \frac{\sum x}{N} \quad (9)$$

## III. PREGUNTAS A RESPONDER

- 1) El diagrama de dispersión o puntos para las dos variables, y observar algún patrón entre ellas, como se están correlacionando, por ejemplo, si puede considerarse lineal (Copiar como imagen de la herramienta de cómputo Statgraphics o Excel).
- 2) El coeficiente de correlación, tratando de obtener un coeficiente numérico que nos indique como se están correlacionando las dos variables (cálculos en tabla construida con Excel para aplicar en fórmula).
- 3) El modelo de regresión lineal que mejor se ajuste a los datos (ecuación de la recta estimada y la imagen de la herramienta de cómputo con la recta obtenida), si las variables están correlacionadas linealmente (cálculos manuales de las constantes, aplicar fórmula).
- 4) El coeficiente de determinación o bondad de ajuste,  $R^2$ , de plantear un modelo de regresión lineal simple ¿Se considera un buen modelo? (calcular manual, aplicar fórmula).

#### IV. DATOS

X	Y	XY	X <sup>2</sup>	Y <sup>2</sup>
2,80	18,00	50,4	7,84	324,00
1,70	9,00	15,3	2,89	81,00
4,50	24,00	108	20,25	576,00
3,50	19,00	66,5	12,25	361,00
4,30	29,00	142,1	24,01	841,00
3,80	28,00	106,4	14,44	784,00
2,50	7,00	17,5	6,25	49,00
5,00	30,00	150	25,00	900,00
4,00	23,00	92	16,00	529,00
3,40	19,00	64,6	11,56	361,00
2,20	13,00	28,6	4,84	169,00
4,30	24,00	103,2	18,49	576,00
2,60	10,00	26	6,76	100,00
4,50	25,00	112,5	20,25	625,00
2,80	12,00	33,6	7,84	144,00
4,20	23,00	96,6	17,64	529,00
3,80	19,00	72,2	14,44	361,00
2,10	17,00	35,7	4,41	289,00
3,10	17,00	52,7	9,61	289,00
4,70	30,00	141	22,09	900,00
2,80	18,00	50,4	7,84	324,00
1,70	8,00	13,6	2,89	64,00
4,50	20,00	90	20,25	400,00
3,00	19,00	57	9,00	361,00
4,30	28,00	137,2	24,01	784,00
3,80	28,00	106,4	14,44	784,00
2,50	6,00	15	6,25	36,00
5,00	31,00	155	25,00	961,00
4,00	24,00	96	16,00	576,00
3,40	20,00	68	11,56	400,00
2,20	15,00	33	4,84	225,00
4,30	14,00	60,2	18,49	196,00
2,60	15,00	39	6,76	225,00
4,50	22,00	99	20,25	484,00
2,80	13,00	36,4	7,84	169,00
4,20	23,00	96,6	17,64	529,00
3,80	19,00	72,2	14,44	361,00
2,10	18,00	37,8	4,41	324,00
3,10	16,00	49,6	9,61	256,00
4,70	30,00	141	22,09	900,00
1,80	10,00	18	3,24	100,00
1,50	8,00	12	2,25	64,00
4,00	20,00	80	16,00	400,00
3,10	19,00	58,9	9,61	361,00
4,50	28,00	126	20,25	784,00
3,20	28,00	89,6	10,24	784,00
2,50	6,00	15	6,25	36,00
5,00	31,00	155	25,00	961,00
4,10	24,00	98,4	16,81	576,00
3,40	20,00	68	11,56	400,00
2,20	15,00	33	4,84	225,00
4,30	14,00	60,2	18,49	196,00
2,60	15,00	39	6,76	225,00
4,50	22,00	99	20,25	484,00
2,80	13,00	36,4	7,84	169,00
4,20	22,00	92,4	17,64	484,00
3,80	18,00	68,4	14,44	324,00
2,10	18,00	37,8	4,41	324,00
3,10	16,00	49,6	9,61	256,00
4,70	18,00	84,6	22,09	324,00

Fig. 1. Tabla de datos

#### V. DESARROLLO

1)

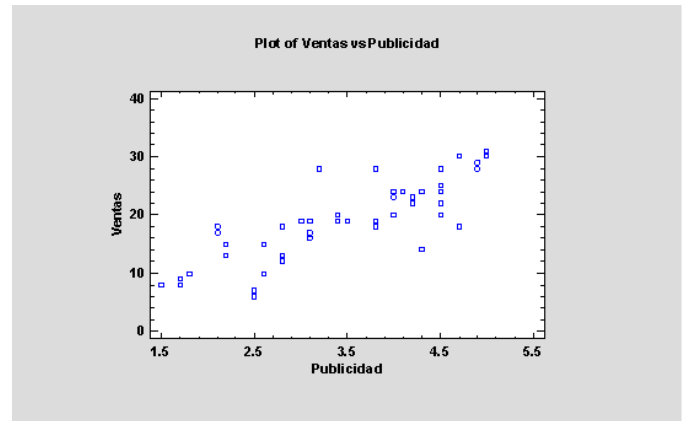


Fig. 2. Diagrama de dispersion de los datos

- El diagrama de dispersión proporciona evidencia de una relación positiva y lineal entre las variables “Ventas” y “Publicidad”.
- La mayoría de los puntos se ajustan a la tendencia general, lo que indica una fuerte correlación entre las variables.
- Hay algunos puntos atípicos que pueden ser causados por errores en la medición o por factores externos.

2) Para términos del análisis, se tomo como variable independiente ( $x$ ) a los valores de “Publicidad”, mientras que como variable dependiente ( $y$ ), las “Ventas”.

- $\Sigma x = 207,70$
- $\Sigma y = 1148,00$
- $\Sigma xy = 4289,60$
- $\Sigma(x^2) = 778,05$
- $\Sigma(y^2) = 24624,00$
- $\bar{x} = 3,46$
- $\bar{y} = 19,13$

Por lo tanto, utilizando la formula (5), se obtiene un valor de

$$r_{xy} = \frac{(60 \cdot 4289,60) - (207,70 \cdot 1148,00)}{\sqrt{(60 \cdot 778,05 - (207,70)^2)(60 \cdot 24624,00 - (1148,00)^2)}} = 0,796414177$$

Resultando en una correlación lineal moderada positiva.

3) Aplicando las formulas (8) y (9),

$$m = \frac{60 \cdot 4289,60 - (207,70 \cdot 1148,00)}{60 \cdot 778,05 - (207,70)^2} = 5,343665255$$

$$b = 19,13 - (5,343665255 \cdot 3,46) = 0,635345443$$

Resultando en la recta,

$$y = 5,343665255x + 0,635345443$$

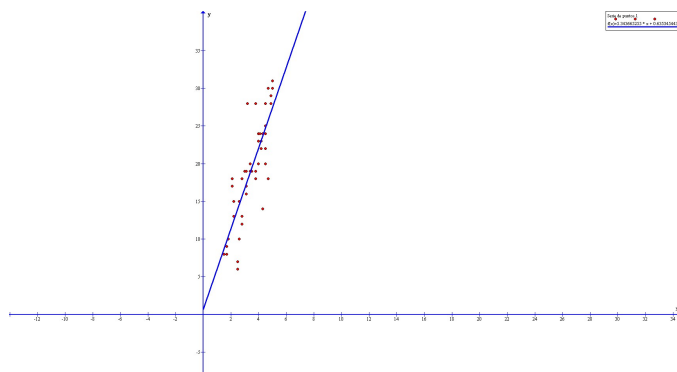


Fig. 3. Recta modelo en relación a los datos

- 4) Con el valor obtenido del test de Pearson, aplicando la formula (6), se obtiene

$$\begin{aligned} R^2 &= (0,796414177)^2 \\ &= (0,634275541) \\ &\approx 63\% \end{aligned}$$

Lo cual nos dice que los datos tienen un ajuste muy bueno.

## VI. SOPORTE DE LA HERRAMIENTA STATGRAPHICS

Simple Regression - Ventas vs. Publicidad

Dependent variable: Ventas

Independent variable: Publicidad

Linear model:  $Y = a + b \cdot X$

Number of observations: 60

Coefficients

Parameter	Least Squares Estimate	Standard Error	T Statistic	P-Value
Intercept	0.635345	1.91863	0.331145	0.7417
Slope	5.34367	0.532799	10.0294	0.0000

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	1686.5	1	1686.5	100.59	0.0000
Residual	972.437	58	16.7662		
Total (Corr.)	2658.93	59			

Correlation Coefficient = 0.796414

R-squared = 63.4276 percent

R-squared (adjusted for d.f.) = 62.797 percent

Standard Error of Est. = 4.09465

Mean absolute error = 3.01212

Durbin-Watson statistic = 2.49092 (P=0.9757)

Lag 1 residual autocorrelation = -0.279317

Fig. 4. Valores calculados

### The StatAdvisor

The output shows the results of fitting a linear model to describe the relationship between Ventas and Publicidad. The equation of the fitted model is

$$\text{Ventas} = 0.635345 + 5.34367 \cdot \text{Publicidad}$$

Since the P-value in the ANOVA table is less than 0.05, there is a statistically significant relationship between Ventas and Publicidad at the 95.0% confidence level.

Fig. 5. Recta modelo calculada

## REFERENCES

- [1] J. F. López, "Coeficiente de determinación (r cuadrado)," 2 2021. [Online]. Available: <https://economipedia.com/definiciones/r-cuadrado-coeficiente-determinacion.html>
- [2] —, "Covarianza," 11 2022. [Online]. Available: <https://economipedia.com/definiciones/covarianza.html>
- [3] C. Ortega, "¿qué es el coeficiente de correlación de pearson?" 2 2023. [Online]. Available: <https://www.questionpro.com/blog/es/coeficiente-de-correlacion-de-pearson/>