

Mariano Méndez-Suárez

# ESTADÍSTICA DESCRIPTIVA E INFERENCIAL

EJERCICIOS  
CON SOLUCIONES



# **Estadística descriptiva e inferencial**

Ejercicios con soluciones

Madrid, 2025

Mariano Méndez-Suárez

# **Estadística descriptiva e inferencial**

Ejercicios con soluciones

Abril, 2025

*Estadística descriptiva e inferencial: Ejercicios con soluciones*  
Mariano Méndez-Suárez

Todos los derechos reservados.

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo las excepciones previstas por la ley.

Diríjase a CEDRO (Centro Español de Derechos Reprográficos) si necesita fotocopiar o escanear algún fragmento de esta obra ([www.cedro.org](http://www.cedro.org)).

© 2025, ESIC Editorial  
Avda. de Valdenigrales, s/n  
28223 Pozuelo de Alarcón (Madrid)  
Tel.: 91 452 41 00  
[www.esic.edu/editorial](http://www.esic.edu/editorial)  
@EsicEditorial

ISBN: 978-84-1192-154-1  
Depósito Legal: M-5077-2025

Diseño de cubierta: Zita Moreno Puig  
Maquetación: Santiago Díez Escribano  
Lectura: Balloon Comunicación  
Impresión: Gráficas Dehon

Un libro de



Impreso en España – *Printed in Spain*

*Este libro ha sido impreso con tinta ecológica y papel sostenible.*

A Pilar

# Índice

Capítulo 1. <b>Introducción a la estadística descriptiva</b> .....	19
1.1. Conceptos clave y técnicas para las tablas de frecuencias .....	19
Descripción de la tabla de frecuencias .....	19
Cálculos clave.....	20
Ejemplo de cálculo paso a paso de una tabla de frecuencias .....	20
Solución paso a paso .....	22
1.2. Ejercicios de estadística descriptiva .....	24
Ejercicio 1. Visitas diarias a un sitio web.....	24
Ejercicio 2. Número de productos vendidos por día (marketing) .....	26
Ejercicio 3. Opiniones de clientes por día (marketing) .....	27
Ejercicio 4. Transacciones diarias en un negocio .....	28
Ejercicio 5. Unidades producidas por día (negocios) .....	30
1.3. Ejercicios propuestos de estadística descriptiva .....	31
Ejercicio 1. Visitas diarias a un sitio web de comercio electrónico ....	31
Ejercicio 2. Número de productos vendidos por semana (marketing) .....	32
Ejercicio 3. <i>Tickets</i> de soporte al cliente por día (negocios digitales) .....	32
Ejercicio 4. Suscripciones mensuales a un servicio digital.....	33
Ejercicio 5. Transacciones diarias en una tienda en línea .....	33
Capítulo 2. <b>Medidas de tendencia central</b> .....	35
2.1. Las medidas de tendencia central .....	35
2.2. Ejercicios sobre la media, mediana, moda y percentiles .....	36
Ejercicio 1. Tiempo de pantalla diario del cliente en una aplicación móvil .....	36

Ejercicio 2. Tiempo de pantalla diario del cliente en una página web de deportes.....	37
Ejercicio 3. Tiempo de pantalla diario del cliente en una página web de moda.....	38
Ejercicio 4. Número de pedidos procesados por día en una tienda en línea .....	39
Ejercicio 5. Registros de usuarios diarios en una plataforma digital..	41
2.3. Ejercicios propuestos sobre media, mediana, moda y percentiles....	42
Ejercicio 1. Visitas diarias a un blog de marketing .....	42
Ejercicio 2. Ventas semanales de productos en una tienda <i>online</i> ...	43
Ejercicio 3. Suscripciones mensuales a un servicio digital .....	43
Ejercicio 4. Ingresos diarios de un sitio de comercio electrónico .....	44
Ejercicio 5. Usuarios activos semanales en una plataforma de redes sociales .....	45
2.4. Medidas de forma: asimetría y curtosis .....	45
Asimetría.....	46
Curtosis .....	46
Ejemplos .....	47
Capítulo 3. <b>Medidas de concentración: curva de Lorenz e índice de Gini..</b>	49
3.1. Curva de Lorenz.....	49
Puntos clave.....	50
3.2. Índice de Gini .....	50
Aplicaciones .....	50
Ejemplo 1. Rangos salariales en una población .....	51
Ejemplo 2. Ventas diarias en una tienda minorista .....	53
Ejemplo 3. Gastos mensuales de una campaña de marketing .....	55
Ejemplo 4. Ventas semanales de un producto en una tienda de comercio electrónico .....	58
Ejemplo 5. Ingresos mensuales de servicios digitales.....	60
Ejemplo 6. Ganancias trimestrales de una empresa de marketing ..	63
Ejemplo 7. Ingresos anuales de un producto digital.....	65
Capítulo 4. <b>Medidas de dispersión: varianza y desviación típica .....</b>	69
4.1. Cálculo de la varianza y desviación típica o estándar .....	69
4.2. Varianza y desviación típica con variables con frecuencia 1.....	70
Ejercicio 1. Ingresos diarios de una campaña de marketing .....	70
Ejercicio 2. Ventas semanales de productos .....	72
Ejercicio 3. Suscripciones mensuales a un servicio digital .....	73
Ejercicio 4. Usuarios activos diarios en una plataforma de redes sociales .....	74
Ejercicio 5. Ingresos semanales de un sitio web de comercio electrónico .....	75

4.3. Ejercicios de medidas de dispersión cuando cada observación tiene una frecuencia mayor de 1.....	77
Ejercicio 1. Análisis de ventas de un producto digital .....	77
Ejercicio 2. Tasa de clics (CTR) en campañas publicitarias .....	78
Ejercicio 3. Tasas de <i>engagement</i> en redes sociales .....	79
Ejercicio 4. Ingresos mensuales por suscripciones .....	80
Ejercicio 5. Tasas de retención de clientes .....	81
4.4. Ejercicios propuestos sobre medidas de dispersión con frecuencia absoluta igual a 1.....	82
Ejercicio 1. Ventas mensuales de un producto .....	82
Ejercicio 2. Visitantes diarios a un sitio web .....	82
Ejercicio 3. Ingresos semanales de un servicio de suscripción.....	83
Ejercicio 4. Usuarios activos mensuales en una aplicación .....	83
Ejercicio 5. Ingresos trimestrales de un negocio .....	84
4.5. Ejercicios propuestos con frecuencia absoluta mayor de 1.....	84
Ejercicio 1. Tasa de conversión de una campaña .....	84
Ejercicio 2. Valoraciones de clientes para un nuevo producto .....	85
Ejercicio 3. Impresiones de anuncios .....	85
Ejercicio 4. Vistas de página.....	85
Ejercicio 5. Frecuencia de compra de clientes .....	86
Capítulo 5. <b>Números índice</b> .....	87
5.1. Visión general de los números índice .....	87
5.2. Índices simples .....	87
5.3. Índices agregados y ponderados.....	88
Índices agregados.....	88
Índices ponderados .....	88
5.4. Índice de Laspeyres $L_0^t$ .....	88
Definición .....	88
Uso práctico .....	89
Ejemplo de cálculo .....	89
5.5. Índice de Paasche $P_0^t$ .....	89
Definición .....	90
Uso práctico .....	90
Ejemplo de cálculo .....	90
5.6. Comparación de los índices de Laspeyres y Paasche .....	90
5.7. Cómo se calcula el índice de precios al consumo (IPC) en España ...	91
Pasos para calcular el IPC en España .....	91
Ejemplo de cálculo del IPC.....	92
5.8. Cómo se calcula el índice NASDAQ .....	92



Tipo de índice utilizado .....	92
Pasos para calcular el índice NASDAQ .....	92
¿Qué es el divisor del índice? .....	93
Ejemplo de cálculo .....	93
5.9. Ejercicios de cálculo de los índices de Laspeyres y Paasche .....	93
Ejercicio 1. Cesta de bienes de consumo .....	93
Ejercicio 2. Cesta de acciones .....	94
Ejercicio 3. Cesta de activos financieros energéticos .....	96
Ejercicio 4. Cesta de acciones tecnológicas .....	97
Ejercicio 5. Cesta de bienes agrícolas .....	98
Resumen .....	99
5.10. Ejercicios propuestos de cálculo de los índices de Laspeyres y Paasche .....	99
Ejercicio propuesto 1. Cesta de <i>gadgets</i> tecnológicos .....	99
Ejercicio propuesto 2. Cesta de alimentos .....	100
Ejercicio propuesto 3. Cesta de activos financieros energéticos .....	100
Ejercicio propuesto 4. Cesta de acciones .....	101
Ejercicio propuesto 5. Cesta de automóviles .....	101
Capítulo 6. <b>Análisis estadístico bidimensional</b> .....	103
6.1. Covarianza .....	103
Definición .....	103
Fórmula .....	103
Pasos de cálculo .....	104
Ejemplo .....	104
6.2. Correlación .....	105
Definición .....	105
Fórmula .....	105
Pasos de cálculo .....	105
Ejemplo .....	105
Interpretación de los valores de correlación .....	106
Ejemplos de valores de correlación .....	106
Ejemplo .....	107
6.3. Tabla de contingencia .....	108
Ejemplo .....	108
Conceptos clave .....	108
Aplicación en negocios .....	108
6.4. Ejercicios .....	108
Ejercicio 1. Satisfacción del cliente vs. frecuencia de compra .....	108
Ejercicio 2. Clics en anuncios vs. tipo de dispositivo .....	109
Ejercicio 3. Canal de marketing vs. generación de <i>leads</i> .....	110

Ejercicio 4. Edad del cliente vs. categoría de producto preferida ....	110
Ejercicio 5. Tipo de campaña de <i>email</i> vs. tasa de clics .....	111
Capítulo 7. <b>Análisis de regresión</b> .....	113
7.1. Conceptos clave .....	113
7.2. Regresión lineal simple .....	114
Ecuación .....	114
Interpretación .....	114
7.3. Ejercicios de regresión .....	114
Ejercicio 1. Predicción de ventas mensuales en función del gasto en publicidad.....	114
Ejercicio 2. Optimización del presupuesto de marketing .....	117
Ejercicio 3. Predicción de la rotación de empleados basada en la satisfacción laboral .....	119
Ejercicio 4. Predicción de ventas de productos basada en valoraciones de clientes.....	122
Ejercicio 5. Predicción de ingresos basada en el número de productos vendidos .....	125
Ejercicio 6. Predicción de las tasas de conversión basada en la calidad de las campañas de <i>email</i> .....	128
Ejercicio 7. Predicción del tráfico en el sitio web basada en el gasto en marketing .....	130
Ejercicio 8. Predicción de retención de clientes basada en la calidad del servicio.....	133
Ejercicio 9. Predicción de la tasa de abandono de clientes basada en la duración media de las llamadas.....	136
Ejercicio 10. Predicción del rendimiento de los empleados basada en horas de formación .....	139
7.4. Ejercicios propuestos sobre regresión .....	142
Ejercicio 1. Predicción de ventas mensuales basada en el gasto en publicidad.....	142
Ejercicio 2. Predicción de la satisfacción del cliente basada en el tiempo de servicio.....	142
Ejercicio 3. Estimación de la demanda del producto basada en el precio .....	143
Ejercicio 4. Análisis del desempeño de los empleados basado en horas de formación .....	143
Ejercicio 5. Pronóstico de ingresos basado en menciones en redes sociales .....	143
Capítulo 8. <b>Probabilidad</b> .....	145
8.1. Conceptos básicos .....	145
8.2. Probabilidad condicional.....	145

Definición .....	146
Ejemplo en marketing .....	146
Interpretación .....	146
8.3. Teorema de Bayes .....	147
Definición .....	147
Ejemplo en empresa .....	147
Interpretación .....	147
8.4. Aplicaciones avanzadas de probabilidad y el teorema de Bayes .....	148
Segmentación de clientes usando probabilidad .....	148
Aplicación .....	148
Usando el teorema de Bayes .....	148
Interpretación .....	148
8.5. Predicción de la pérdida de clientes con probabilidad condicional ...	148
Contexto .....	148
Aplicación .....	149
Usando el teorema de Bayes .....	149
Interpretación .....	149
8.6. Optimización de campañas de marketing con atribución multicanal	149
Contexto .....	149
Aplicación .....	149
Solución .....	150
Interpretación .....	150
8.7. Ejercicios resueltos .....	150
Ejercicio 1. Efectividad de la campaña de marketing .....	150
Ejercicio 2. Programa de recomendación de clientes .....	151
Ejercicio 3. Retirada de un producto del mercado .....	151
Ejercicio 4. Tasa de éxito de una campaña de <i>email</i> .....	152
Ejercicio 5. Encuesta de satisfacción del cliente .....	152
8.8. Ejercicios propuestos .....	153
Ejercicio 1. Programa de recomendación de clientes .....	153
Ejercicio 2. Producto defectuoso en su producción .....	153
Ejercicio 3. Tasa de conversión de clics en anuncios .....	154
Ejercicio 4. Éxito de un modelo de aprendizaje automático .....	155
Ejercicio 5. Detección de errores de <i>software</i> .....	156
Capítulo 9. <b>Variables aleatorias</b> .....	157
9.1. Variable discreta .....	157
Distribución binomial .....	157
Aplicaciones en negocios y marketing .....	157
Ejemplo. <i>Overbooking</i> en reservas de aerolíneas .....	158

Conclusión .....	159
Distribución de Poisson .....	159
Aplicaciones en negocios digitales y marketing .....	160
9.2. Ejercicios de distribución binomial .....	160
Ejercicio 1. Campaña de marketing por correo electrónico .....	160
Ejercicio 2. Tasa de clics en anuncios en línea .....	161
Ejercicio 3. Servicio de suscripción .....	161
Ejercicio 4. Devoluciones de productos .....	161
Ejercicio 5. Retroalimentación de clientes .....	162
9.3. Ejercicios de distribución de Poisson.....	162
Ejercicio 1. Tráfico en sitios web .....	162
Ejercicio 2. Registros de correos electrónicos.....	162
Ejercicio 3. Quejas de clientes .....	163
Ejercicio 4. Descargas de aplicaciones .....	163
Ejercicio 5. Menciones en redes sociales.....	163
9.4. Ejercicios propuestos de distribución binomial .....	164
Ejercicio 1. Tasa de conversión en un sitio web .....	164
Ejercicio 2. Devoluciones de productos .....	164
Ejercicio 3. Registros de clientes .....	164
Ejercicio 4. Tasa de clics en una campaña de correo electrónico ....	165
Ejercicio 5. Control de calidad.....	165
9.5. Ejercicios propuestos de distribución de Poisson.....	165
Ejercicio 1. Visitas a un sitio web .....	165
Ejercicio 2. Llamadas a servicio al cliente .....	166
Ejercicio 3. Instalaciones de aplicaciones.....	166
Ejercicio 4. Pedidos en línea .....	166
Ejercicio 5. Informes de errores en <i>software</i> .....	167
Capítulo 10. <b>Distribución normal</b> .....	169
10.1. Características clave .....	169
Fórmula de la distribución normal.....	169
10.2. Aplicaciones prácticas.....	169
10.3. Distribución normal estándar .....	170
10.4. Características de la distribución normal estándar.....	170
10.5. Estandarización de cualquier distribución normal .....	171
Ejemplo 1. Control de calidad en fabricación .....	171
Ejemplo 2. Evaluación del rendimiento de los empleados .....	171
10.6. Interpretación de los Z-scores .....	172
10.7. Uso de la tabla de distribución normal estándar (bilateral) .....	172
Ejemplo 1. Probabilidad de un valor menor que un valor dado .....	172

Ejemplo 2. Probabilidad entre dos valores .....	173
Ejemplo 3. Probabilidad mayor que un valor dado.....	173
10.8. Ejercicios sobre distribución normal .....	174
Ejercicio 1. Puntuaciones de satisfacción del cliente.....	174
Ejercicio 2. Tiempos de carga del sitio web .....	174
Ejercicio 3. Tasas de apertura de correos electrónicos .....	175
Ejercicio 4. Tasas de defectos de productos .....	176
Ejercicio 5. Clics en anuncios digitales.....	176
Ejercicio 6. Rendimiento de ventas .....	177
Ejercicio 7. Tiempo de uso de la aplicación .....	178
Ejercicio 8. Retorno de la inversión (ROI) .....	179
10.9. Ejercicios propuestos sobre la distribución normal con solución .....	179
Ejercicio 1. Tasa de clics en campaña de correo electrónico .....	179
Ejercicio 2. Retención de clientes .....	180
Ejercicio 3. Interacción en redes sociales .....	180
Ejercicio 4. Encuesta de reconocimiento de marca .....	180
Ejercicio 5. Recuerdo publicitario .....	181
Ejercicio 6. Tasa de devolución de productos.....	181
Ejercicio 7. Rotación de inventario .....	181
Ejercicio 8. Ventas trimestrales.....	182
Ejercicio 9. Productividad de los empleados .....	182
Ejercicio 10. Tasa de defectos en producción .....	182
Ejercicio 11. Tasa de rebote en el sitio web .....	183
Ejercicio 12. Tasa de descargas de la aplicación .....	183
Ejercicio 13. Tiempo de inactividad del servidor.....	183
Ejercicio 14. Volumen de transacciones en línea.....	184
Ejercicio 15. Tasa de conversión .....	184
Capítulo 11. <b>Muestreo</b> .....	185
11.1. Teorema central del límite (TCL).....	185
Puntos clave del teorema central del límite .....	185
11.2. Muestreo de la media .....	186
Propiedades de la media muestral.....	186
Ejemplo práctico.....	186
11.3. Muestreo de proporciones .....	186
Propiedades de la proporción muestral.....	187
Ejemplo práctico.....	187
11.4. Intervalos de confianza .....	187
Entender los intervalos de confianza.....	188
Conceptos clave.....	188
¿Cuándo se utilizan los intervalos de confianza en la vida real? .....	188

Ejemplo de intervalos de confianza en la práctica. Estimación de ventas mensuales medias .....	189
11.5. Intervalos de confianza para proporciones .....	189
Ejemplo: Estimación de la satisfacción del cliente .....	190
11.6. Ejercicios propuestos sobre la estimación de intervalos de confianza con muestreo de medias .....	190
Ejercicio 1. Estimación del tiempo medio de entrega .....	190
Ejercicio 2. Estimación del gasto medio en marketing por campaña .....	191
Ejercicio 3. Estimación del tiempo medio de carga de páginas .....	192
Ejercicio 4. Estimación del valor de vida del cliente (CLV) medio.....	193
Ejercicio 5. Estimación de las ventas mensuales medias .....	194
Ejercicio 6. Estimación de la tasa de clics media .....	194
Ejercicio 7. Estimación del peso medio de productos .....	195
Ejercicio 8. Estimación del tiempo medio en un sitio web .....	196
Ejercicio 9. Estimación de la renta mensual media.....	197
Ejercicio 10. Estimación de las ventas diarias medias .....	197
Ejercicio 11. Estimación del tiempo de viaje medio.....	198
Ejercicio 12. Estimación de la vida útil media de las baterías .....	199
11.7. Ejercicios propuestos sobre la estimación de intervalos de confianza con muestreo de proporciones.....	200
Ejercicio 1. Estimación de la proporción de clientes satisfechos.....	200
Ejercicio 2. Estimación de la proporción de anuncios con clics.....	201
Ejercicio 3. Estimación de la proporción de compradores recurrentes .....	201
Ejercicio 4. Estimación de la proporción de productos defectuosos..	202
Ejercicio 5. Estimación de la proporción de clientes que regresan (negocios).....	203
Ejercicio 6. Estimación de la proporción de campañas exitosas (marketing) .....	204
Ejercicio 7. Estimación de la proporción de usuarios con alta participación (negocios digitales) .....	205
Ejercicio 8. Estimación de la proporción de productos devueltos (negocios).....	206
Ejercicio 9. Estimación de la proporción de reseñas positivas (marketing) .....	207
Ejercicio 10. Estimación de la proporción de tasas de apertura de correos electrónicos (negocios digitales) .....	207
Ejercicio 11. Estimación de la proporción de conversiones en una aplicación móvil (negocios digitales) .....	208
Ejercicio 12. Estimación de la proporción de suscriptores a un boletín (marketing) .....	209
Capítulo 12. <b>Contraste de hipótesis</b> .....	211
12.1. Valor p para el contraste de hipótesis .....	212
¿Cómo interpretar el valor p? .....	213

12.2. Contraste de hipótesis .....	214
Test Bilateral ejemplo. Evaluación de una nueva estrategia de marketing .....	214
Test unilateral (cola superior) ejemplo. Prueba de una nueva funcionalidad en un producto digital.....	215
Test unilateral (cola inferior) ejemplo. Evaluación de la reducción de costes en un proceso empresarial.....	216
12.3. Ejercicios contraste de hipótesis para la media .....	217
Ejercicio 1. Ventas mensuales medias .....	217
Ejercicio 2. Mejora de la tasa de conversión promedio .....	218
Ejercicio 3. Tiempo medio en la aplicación .....	218
Ejercicio 4. Reducción de costes de fabricación .....	219
Ejercicio 5. Prueba del valor medio de satisfacción del cliente.....	220
Ejercicio 6. Precio medio de producto.....	221
Ejercicio 7. Tiempo de carga medio del sitio web .....	222
Ejercicio 8. Ventas diarias medias .....	223
Ejercicio 9. Valoración de satisfacción del cliente .....	223
Ejercicio 10. Gasto medio por campaña de marketing.....	224
12.4. Ejercicios contraste de hipótesis para proporciones .....	225
Ejercicio 1. Proporción de clientes satisfechos .....	225
Ejercicio 2. Proporción de clientes recurrentes .....	226
Ejercicio 3. Proporción de anuncios clicados .....	227
Ejercicio 4. Proporción de correos electrónicos abiertos.....	227
Ejercicio 5. Proporción de reseñas positivas .....	228
Ejercicio 6. Proporción de campañas exitosas .....	229
Ejercicio 7. Proporción de quejas de clientes .....	230
Ejercicio 8. Proporción de productos devueltos .....	231
Ejercicio 9. Proporción de usuarios comprometidos.....	231
Ejercicio 10. Proporción de productos defectuosos .....	232
12.5. Ejercicios propuestos contraste de hipótesis.....	233
Ejercicio 1. Tiempo medio de entrega .....	233
Ejercicio 2. Tiempo medio de uso de la plataforma .....	233
Ejercicio 3. Número medio de artículos vendidos por día .....	234
Ejercicio 4. Tiempo medio de producción.....	234
Ejercicio 5. Tasa de apertura de correos electrónicos.....	234
Ejercicio 6. Satisfacción del cliente .....	234
Ejercicio 7. Interacción con una nueva característica .....	235
Ejercicio 8. Productos devueltos .....	235
Ejercicio 9. Tiempo de corrección de errores.....	235
Ejercicio 10. Preferencia de los consumidores .....	235
Apéndice. <b>Tabla de distribución normal estándar</b> .....	236

7.1. Conceptos clave. | 7.2. Regresión lineal simple. | 7.3. Ejercicios de regresión. | 7.4. Ejercicios propuestos sobre regresión.

El análisis de regresión, también conocido como *supervised machine learning*, es una de las principales técnicas estadísticas empleadas para comprender las relaciones entre variables, predecir resultados futuros y tomar decisiones basadas en datos. Ayuda a identificar tendencias y patrones, a comprender cómo los cambios en una variable pueden afectar a otra. Además, la regresión permite a los analistas predecir tendencias futuras basándose en datos históricos, determinar el impacto de diversos factores en un resultado y evaluar la fuerza de las relaciones.

En la práctica, el método de regresión tiene una gran variedad de aplicaciones en distintos contextos empresariales. Por ejemplo, puede utilizarse para predecir las ventas en función de la inversión publicitaria, lo que ayuda a las empresas a asignarla de forma más eficaz. También desempeña un papel importante en la comprensión de cómo la satisfacción del cliente afecta a las tasas de recompra, proporcionando información sobre la lealtad del cliente y las posibles estrategias de retención. Además, la regresión es fundamental para optimizar las estrategias de precios mediante el análisis de la elasticidad de la demanda, que ayuda a determinar el precio óptimo de los productos para maximizar los ingresos. Otra aplicación importante de la regresión es el análisis de las métricas de *engagement* de los usuarios para comprender su efecto en las tasas de conversión de los sitios web. Este análisis es esencial para mejorar la experiencia del usuario y aumentar las ventas. Además, la regresión puede mejorar la orientación de las campañas promocionales mediante la predicción del comportamiento de los clientes a través de datos demográficos, lo que hace que los esfuerzos de marketing sean más eficientes y eficaces.

## 7.1. Conceptos clave

- **Variable dependiente (Y):** la variable que se intenta predecir o explicar. También se conoce como variable de respuesta o de resultado.
- **Variable independiente (X):** la variable que se utiliza para predecir la variable dependiente. También se conoce como variable predictora o explicativa.
- **Regresión lineal:** un tipo de análisis de regresión donde la relación entre las variables dependiente e independiente se modela como una línea recta.



## 7.2. Regresión lineal simple

### Ecuación

La ecuación para una línea de regresión lineal simple es:

$$Y = \alpha + \beta X + \varepsilon$$

- $Y$  es la variable dependiente.
- $X$  es la variable independiente.
- $\alpha$  es la intersección en el eje  $Y$ .
- $\beta$  es la pendiente de la línea de regresión.
- $\varepsilon$  es el término de error.

### Interpretación

- **Intersección en  $Y$  ( $\alpha$ ):** el valor de  $Y$  cuando  $X = 0$ . Representa el punto de inicio de la línea de regresión. En términos matemáticos, esto se puede interpretar como la derivada de  $Y$  con respecto a  $X$ , denotada como  $\frac{dY}{dX}$ . La pendiente representa la tasa de cambio de la variable dependiente  $Y$  por cada incremento de una unidad en la variable independiente  $X$ .
- **Pendiente ( $\beta$ ):** el cambio en  $Y$  por un cambio de una unidad en  $X$ . Indica la fuerza y dirección de la relación entre  $X$  e  $Y$ .

**Nota.** Debido a la gran cantidad de pasos de cálculo y posibles errores de redondeo en los decimales, los resultados calculados pueden diferir levemente.

## 7.3. Ejercicios de regresión

### Ejercicio 1. Predicción de ventas mensuales en función del gasto en publicidad

Un gerente de marketing quiere predecir las ventas mensuales en función del gasto en publicidad. Se han recopilado los siguientes datos.

*Datos*

Gasto en publicidad (X)	Ventas mensuales (Y)
8	20
12	25
18	30
22	32
30	35

Crea un modelo y evalúalo en función de  $R^2$  y RMSE.

### Solución

- Paso 1: Calcular las medias de  $X$  e  $Y$ .

$$\bar{X} = \frac{8 + 12 + 18 + 22 + 30}{5} = 18$$

$$\bar{Y} = \frac{20 + 25 + 30 + 32 + 35}{5} = 28,4$$

- Paso 2: Calcular la covarianza de  $X$  e  $Y$ .

$$\text{Cov}(X, Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
8	20	-10	-8,4	84,0
12	25	-6	-3,4	20,4
18	30	0	1,6	0,0
22	32	4	3,6	14,4
30	35	12	6,6	79,2

$$\text{Cov}(X, Y) = \frac{84 + 20,4 + 0 + 14,4 + 79,2}{5} = 39,6$$

- Paso 3: Calcular la varianza de  $X$ .

$$\sigma_X^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

X	$X - \bar{X}$	$(X - \bar{X})^2$
8	-10	100
12	-6	36
18	0	0
22	4	16
30	12	144

$$\sigma_X^2 = \frac{100 + 36 + 0 + 16 + 144}{5} = 59,2$$

- Paso 4: Calcular la pendiente e intersección.

$$\beta = \frac{\text{Cov}(X, Y)}{\sigma_X^2} = \frac{39,6}{59,2} = 0,6689$$

$$\alpha = \bar{Y} - \beta \cdot \bar{X} = 28,4 - (0,6689 \times 18) = 16,37$$

- Paso 5: Ecuación de la línea de regresión.

$$Y = 16,37 + 0,6689X$$

- *Paso 6: Calcular el RMSE.*

$$e_i = Y_i - (\alpha + \beta X_i)$$

X	Y	$\hat{Y}$	Residual ( $e_i$ )	Residual cuadrado ( $e_i^2$ )
8	20	21,71	-1,71	2,92
12	25	24,39	0,61	0,37
18	30	28,37	1,63	2,65
22	32	31,12	0,88	0,77
30	35	36,44	-1,44	2,07

$$RMSE = \sqrt{\frac{2,92 + 0,37 + 2,65 + 0,77 + 2,07}{5}} = 1,58$$

- *Paso 7: Calcular el coeficiente de determinación ( $R^2$ ).*

$$R^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2}$$

$$R^2 = 1 - \frac{2,92 + 0,37 + 2,65 + 0,77 + 2,07}{(20 - 28,4)^2 + (25 - 28,4)^2 + (30 - 28,4)^2 + (32 - 28,4)^2 + (35 - 28,4)^2}$$

$$R^2 = 1 - \frac{8,78}{(70,56 + 11,56 + 2,56 + 12,96 + 43,56)} = 1 - \frac{8,78}{141,2} = 0,9378$$

- *Paso 8: Evaluación del modelo.*

El modelo tiene un  $R^2$  de 0,9378, lo que significa que el 93,78% de la variabilidad en las ventas mensuales puede explicarse por el gasto en publicidad. El RMSE de 1,58 indica la magnitud promedio del error de predicción.

### Interpretación de los resultados

1. **Pendiente ( $\beta$ ).** La pendiente de 0,6689 sugiere que, por cada unidad de aumento en el gasto en publicidad, las ventas mensuales aumentan en aproximadamente 0,6689 unidades en promedio. Esto indica una relación positiva entre el gasto en publicidad y las ventas.
2. **Intersección ( $\alpha$ ).** La intersección de 16,37 representa el valor estimado de las ventas mensuales cuando el gasto en publicidad es cero.
3. **Coeficiente de determinación ( $R^2$ ).** Un  $R^2$  de 0,9378 indica que aproximadamente el 93,78% de la variabilidad en las ventas mensuales puede explicarse por el gasto en publicidad.
4. **Raíz del error cuadrático medio (RMSE).** Un RMSE de 1,58 sugiere que, en promedio, las predicciones de ventas están a 1,58 unidades de las ventas reales.

### Conclusiones

El análisis de regresión indica una relación positiva y significativa entre el gasto en publicidad y las ventas mensuales. Aumentar el gasto en publicidad probablemente resultará

en un aumento de las ventas. Sin embargo, para un modelo más robusto, podrían incluirse otros factores que influyan en las ventas.

## Ejercicio 2. Optimización del presupuesto de marketing

Una empresa quiere optimizar su presupuesto de marketing para aumentar la conciencia de producto. Se han recopilado datos sobre el gasto en marketing y el número resultante de clientes potenciales generados. Utilizando análisis de regresión, el objetivo es determinar cómo los cambios en el gasto en marketing afectan el número de clientes potenciales generados y establecer la asignación óptima del presupuesto.

### Datos

Gasto en marketing (X)	Cientes potenciales (Y)
12	145
15	160
8	120
18	180
14	170
20	190

### Solución

- Paso 1: Calcular las medias de X e Y.

$$\bar{X} = \frac{12 + 15 + 8 + 18 + 14 + 20}{6} = 14,5$$
$$\bar{Y} = \frac{145 + 160 + 120 + 180 + 170 + 190}{6} = 160,83$$

- Paso 2: Calcular la covarianza de X e Y.

$$\text{Cov}(X,Y) = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
12	145	-2,5	-15,83	39,58
15	160	0,5	-0,83	-0,42
8	120	-6,5	-40,83	265,42
18	180	3,5	19,17	67,10
14	170	-0,5	9,17	-4,58
20	190	5,5	29,17	160,42

$$\text{Cov}(X,Y) = \frac{39,58 - 0,42 + 265,42 + 67,10 - 4,58 + 160,42}{6} = 87,25$$

- Paso 3: Calcular la varianza de X.

$$\sigma_X^2 = \frac{\sum(X_i - \bar{X})^2}{n}$$

X	$X - \bar{X}$	$(X - \bar{X})^2$
12	-2,5	6,25
15	0,5	0,25
8	-6,5	42,25
18	3,5	12,25
14	-0,5	0,25
20	5,5	30,25

$$\sigma_X^2 = \frac{6,25 + 0,25 + 42,25 + 12,25 + 0,25 + 30,25}{6} = 15,25$$

- Paso 4: Calcular la pendiente e intersección.

$$\beta = \frac{\text{Cov}(X, Y)}{\sigma_X^2} = \frac{87,25}{15,25} = 5,77$$

$$\alpha = \bar{Y} - \beta \cdot \bar{X} = 160,83 - (5,77 \times 14,5) = 77,24$$

- Paso 5: Ecuación de la línea de regresión.

$$Y = 77,24 + 5,77X$$

- Paso 6: Calcular el RMSE.

$$e_i = Y_i - (\alpha + \beta X_i)$$

X	Y	$\hat{Y}$	Residual ( $e_i$ )	Residual Cuadrado ( $e_i^2$ )
12	145	146,51	-1,51	2,28
15	160	163,84	-3,84	14,74
8	120	123,37	-3,37	11,37
18	180	181,12	-1,12	1,25
14	170	158,61	11,39	129,74
20	190	192,76	-2,76	7,61

$$\text{RMSE} = \sqrt{\frac{2,28 + 14,74 + 11,37 + 1,25 + 129,74 + 7,61}{6}} = 5,47$$

- Paso 7: Calcular el coeficiente de determinación ( $R^2$ ).

$$R^2 = \frac{\text{Varianza explicada por el modelo}}{\text{Varianza total de Y}}$$

$$R^2 = 0,94$$

- Paso 8: Evaluación del modelo. El modelo tiene un  $R^2$  de 0,94, lo que indica que el 94% de la variabilidad en los clientes potenciales puede explicarse por el gasto en marketing. El RMSE de 5,47 muestra que, en promedio, las predicciones están a 5,47 unidades de los valores reales, lo que sugiere un alto nivel de precisión en la predicción.

Conclusiones

El análisis sugiere que un aumento en el gasto en marketing impacta positivamente el número de clientes potenciales generados. El valor de  $R^2$  indica que el modelo explica una parte significativa de la variabilidad en los clientes potenciales basándose en el gasto en marketing. Sin embargo, otros factores como las condiciones del mercado, la efectividad de la campaña y las preferencias de los clientes también pueden influir y deben ser considerados para una optimización completa.

Ejercicio 3. Predicción de la rotación de empleados basada en la satisfacción laboral

Un gerente de recursos humanos desea predecir la rotación de empleados en función de la satisfacción laboral. Se recopilan los siguientes datos.

Datos

Satisfacción laboral (X)	Rotación de empleados (%) (Y)
2	25
5	22
7	20
9	15
11	10

Crear un modelo y evaluarlo en función de  $R^2$  y RMSE.

Solución

- Paso 1: Calcular las medias de X y Y.

$$\bar{X} = \frac{2 + 5 + 7 + 9 + 11}{5} = 6,8$$
$$\bar{Y} = \frac{25 + 22 + 20 + 15 + 10}{5} = 18,4$$

- Paso 2: Calcular la covarianza de X y Y.

$$\text{Cov}(X, Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
2	25	-4,8	6,6	-31,68
5	22	-1,8	3,6	-6,48
7	20	0,2	1,6	0,32
9	15	2,2	-3,4	-7,48
11	10	4,2	-8,4	-35,28

Suma de  $(X - \bar{X})(Y - \bar{Y})$ :

$$\sum(X_i - \bar{X})(Y_i - \bar{Y}) = -31,68 - 6,48 + 0,32 - 7,48 - 35,28 = -80,6$$

Covarianza:

$$\text{Cov}(X, Y) = \frac{-80,6}{5} = -16,12$$

- *Paso 3: Calcular la varianza de X.*

$$\text{Var}(X) = \frac{\sum(X_i - \bar{X})^2}{n}$$

Suma de  $(X - \bar{X})^2$ :

$$\sum(X_i - \bar{X})^2 = 23,04 + 3,24 + 0,04 + 4,84 + 17,64 = 48,8$$

Varianza:

$$\text{Var}(X) = \frac{48,8}{5} = 9,76$$

- *Paso 4: Calcular la pendiente ( $\beta$ ).*

$$\beta = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{-16,12}{9,76} \approx -1,65$$

- *Paso 5: Calcular la intersección ( $\alpha$ ).*

$$\alpha = \bar{Y} - \beta \bar{X}$$

$$\alpha = 18,4 - (-1,65) \times 6,8 \approx 29,62$$

- *Paso 6: Formular la ecuación de regresión.*

$$Y = 29,62 - 1,65X$$

- *Paso 7: Calcular  $R^2$ .*

Primero, calcular el coeficiente de correlación ( $\rho$ ) usando la fórmula:

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Donde  $\sigma_X$  y  $\sigma_Y$  son las desviaciones estándar de X y Y.

Calcular  $\sigma_X$ :

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{9,76} \approx 3,12$$

Calcular la varianza y desviación estándar de Y:

$$\text{Var}(Y) = \frac{\sum(Y_i - \bar{Y})^2}{n}$$

Suma de  $(Y - \bar{Y})^2$ :

$$\sum(Y_i - \bar{Y})^2 = 42,56 + 12,96 + 2,56 + 11,56 + 70,56 = 140,2$$

Varianza:

$$\text{Var}(Y) = \frac{140,2}{5} = 28,04$$

Desviación estándar:

$$\sigma_Y = \sqrt{28,04} \approx 5,30$$

Calcular el coeficiente de correlación:

$$\rho = \frac{-16,12}{3,12 \times 5,30} \approx -0,98$$

Calcular  $R^2$ :

$$R^2 = \rho^2 = (-0,98)^2 = 0,96$$

• *Paso 8: Calcular RMSE*

1. Calcular los residuos (errores):

$$e_i = Y_i - \hat{Y}_i$$

X	Y	$\hat{Y}_i$	Residuo ( $e_i$ )	$e_i^2$
2	25	29,62 - 1,65(2)	25 - 26,32	1,74
5	22	29,62 - 1,65(5)	22 - 21,37	0,40
7	20	29,62 - 1,65(7)	20 - 18,07	3,72
9	15	29,62 - 1,65(9)	15 - 14,77	0,05
11	10	29,62 - 1,65(11)	10 - 11,47	2,16

2. Sumar los residuos al cuadrado:

$$\sum e_i^2 = 1,74 + 0,40 + 3,72 + 0,05 + 2,16 = 8,07$$

3. Calcular el error cuadrático medio (MSE):

$$MSE = \frac{\sum e_i^2}{n} = \frac{8,07}{5} = 1,61$$

4. Calcular el error cuadrático medio de la raíz (RMSE):

$$RMSE = \sqrt{MSE} = \sqrt{1,61} \approx 1,27$$

*Resumen de resultados*

- Intersección ( $\alpha$ ): 29,62
- Pendiente ( $\beta$ ): -1,65
- $R^2$ : 0,96
- RMSE: 1,27



### Interpretación

La ecuación de regresión  $Y = 29,62 - 1,65X$  sugiere que, por cada unidad adicional de satisfacción laboral, la rotación de empleados disminuye aproximadamente en 1,65%. El modelo explica el 96% de la variabilidad en la rotación de empleados, con un error de predicción promedio de 1,27%.

## Ejercicio 4. Predicción de ventas de productos basada en valoraciones de clientes

Un gerente de producto desea predecir las ventas de un producto en función de las valoraciones de los clientes. Se recopilan los siguientes datos.

### Datos

Calificación de clientes (X)	Ventas de productos (Y)
1	10
3	15
4	20
6	25
8	28

Crear un modelo y evaluarlo en función de  $R^2$  y RMSE.

### Solución

- Paso 1: Calcular las medias de  $X$  y  $Y$ .

$$\bar{X} = \frac{1 + 3 + 4 + 6 + 8}{5} = 4,4$$

$$\bar{Y} = \frac{10 + 15 + 20 + 25 + 28}{5} = 19,6$$

- Paso 2: Calcular la covarianza de  $X$  y  $Y$ .

$$\text{Cov}(X, Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
1	10	-3,4	-9,6	32,64
3	15	-1,4	-4,6	6,44
4	20	-0,4	0,4	-0,16
6	25	1,6	5,4	8,64
8	28	3,6	8,4	30,24

Suma de  $(X - \bar{X})(Y - \bar{Y})$ :

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = 32,64 + 6,44 - 0,16 + 8,64 + 30,24 = 77,8$$

Covarianza:

$$\text{Cov}(X, Y) = \frac{77,8}{5} = 15,56$$

- *Paso 3: Calcular la varianza de X.*

$$\text{Var}(X) = \frac{\sum (X_i - \bar{X})^2}{n}$$

Suma de  $(X - \bar{X})^2$ :

$$\sum (X_i - \bar{X})^2 = 11,56 + 1,96 + 0,16 + 2,56 + 12,96 = 29,2$$

Varianza:

$$\text{Var}(X) = \frac{29,2}{5} = 5,84$$

- *Paso 4: Calcular la pendiente ( $\beta$ ).*

$$\beta = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{15,56}{5,84} \approx 2,66$$

- *Paso 5: Calcular la intersección ( $\alpha$ ).*

$$\alpha = \bar{Y} - \beta \bar{X}$$

$$\alpha = 19,6 - 2,66 \times 4,4 \approx 7,86$$

- *Paso 6: Formular la ecuación de regresión.*

$$Y = 7,86 + 2,66X$$

- *Paso 7: Calcular  $R^2$ . Primero, calcular el coeficiente de correlación ( $\rho$ ) usando la fórmula:*

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Donde  $\sigma_X$  y  $\sigma_Y$  son las desviaciones estándar de X y Y. Calcular  $\sigma_X$ :

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{5,84} \approx 2,42$$

Calcular la varianza y desviación estándar de Y:

$$\text{Var}(Y) = \frac{\sum (Y_i - \bar{Y})^2}{n}$$

Suma de  $(Y - \bar{Y})^2$ :

$$\sum (Y_i - \bar{Y})^2 = 92,16 + 21,16 + 0,16 + 29,16 + 69,16 = 211,8$$

Varianza:

$$\text{Var}(Y) = \frac{211,8}{5} = 42,36$$

Desviación estándar:

$$\sigma_Y = \sqrt{42,36} \approx 6,51$$

Calcular el coeficiente de correlación:

$$\rho = \frac{15,56}{2,42 \times 6,51} \approx 0,98$$

Calcular  $R^2$ :

$$R^2 = \rho^2 = 0,98^2 = 0,96$$

- *Paso 8: Calcular RMSE.*

1. Calcular los residuos (errores):

$$e_i = Y_i - \hat{Y}_i$$

X	Y	$\hat{Y}_i$	Residuo ( $e_i$ )	$e_i^2$
1	10	$7,86 + 2,66(1)$	$10 - 10,5$	0,27
3	15	$7,86 + 2,66(3)$	$15 - 15,8$	0,71
4	20	$7,86 + 2,66(4)$	$20 - 18,5$	2,25
6	25	$7,86 + 2,66(6)$	$25 - 23,8$	1,39
8	28	$7,86 + 2,66(8)$	$28 - 29,1$	1,30

2. Sumar los residuos al cuadrado:

$$\sum e_i^2 = 0,27 + 0,71 + 2,25 + 1,39 + 1,30 = 5,92$$

3. Calcular el error cuadrático medio (MSE):

$$MSE = \frac{\sum e_i^2}{n} = \frac{5,92}{5} = 1,18$$

4. Calcular el error cuadrático medio de la raíz (RMSE):

$$RMSE = \sqrt{MSE} = \sqrt{1,18} \approx 1,09$$

### Resumen de resultados

- Intersección ( $\alpha$ ): 7,86.
- Pendiente ( $\beta$ ): 2,66.
- $R^2$ : 0,96.
- RMSE: 1,09.

### Interpretación

La ecuación de regresión  $Y = 7,86 + 2,66X$  sugiere que, por cada incremento adicional en la valoración del cliente, las ventas del producto aumentan aproximadamente en 2,66 unidades. El modelo explica el 96% de la variabilidad en las ventas de productos, con un error de predicción promedio de 1,09 unidades.

## Ejercicio 5. Predicción de ingresos basada en el número de productos vendidos

Un analista de negocios desea predecir los ingresos en función del número de productos vendidos. Se recopilan los siguientes datos.

### Datos

Número de productos vendidos (X)	Ingresos (Y)
5	50
12	70
19	85
25	90
30	95

Crear un modelo y evaluarlo en función de  $R^2$  y RMSE.

### Solución

- Paso 1: Calcular las medias de X y Y.

$$\bar{X} = \frac{5 + 12 + 19 + 25 + 30}{5} = 18,2$$

$$\bar{Y} = \frac{50 + 70 + 85 + 90 + 95}{5} = 78$$

- Paso 2: Calcular la covarianza de X y Y.

$$\text{Cov}(X, Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
5	50	-13,2	-28	369,6
12	70	-6,2	-8	49,6
19	85	0,8	7	5,6
25	90	6,8	12	81,6
30	95	11,8	17	200,6

Suma de  $(X - \bar{X})(Y - \bar{Y})$ :

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = 369,6 + 49,6 + 5,6 + 81,6 + 200,6 = 707$$

Covarianza:

$$\text{Cov}(X, Y) = \frac{707}{5} = 141,4$$

- Paso 3: Calcular la varianza de X.

$$\text{Var}(X) = \frac{\sum (X_i - \bar{X})^2}{n}$$

Suma de  $(X - \bar{X})^2$ :

$$\sum (X_i - \bar{X})^2 = 174,24 + 38,44 + 0,64 + 46,24 + 139,24 = 398,8$$

Varianza:

$$\text{Var}(X) = \frac{398,8}{5} = 79,76$$

- Paso 4: Calcular la pendiente ( $\beta$ ).

$$\beta = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{141,4}{79,76} \approx 1,77$$

- Paso 5: Calcular la intersección ( $\alpha$ ).

$$\alpha = \bar{Y} - \beta \bar{X}$$

$$\alpha = 78 - 1,77 \times 18,2 \approx 46,76$$

- Paso 6: Formular la ecuación de regresión.

$$Y = 46,76 + 1,77X$$

- Paso 7: Calcular  $R^2$ .

Primero, calcular el coeficiente de correlación ( $\rho$ ) usando la fórmula:

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Donde  $\sigma_X$  y  $\sigma_Y$  son las desviaciones estándar de  $X$  y  $Y$ .

Calcular  $\sigma_X$ :

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{79,76} \approx 8,93$$

Calcular la varianza y desviación estándar de  $Y$ :

$$\text{Var}(Y) = \frac{\sum (Y_i - \bar{Y})^2}{n}$$

Suma de  $(Y - \bar{Y})^2$ :

$$\sum (Y_i - \bar{Y})^2 = 784 + 64 + 49 + 144 + 289 = 1.330$$

Varianza:

$$\text{Var}(Y) = \frac{1.330}{5} = 266$$

Desviación estándar:

$$\sigma_Y = \sqrt{266} \approx 16,31$$

Calcular el coeficiente de correlación:

$$\rho = \frac{141,4}{8,93 \times 16,31} \approx 0,98$$

Calcular  $R^2$ :

$$R^2 = \rho^2 = 0,98^2 = 0,96$$

- *Paso 8: Calcular RMSE.*

1. Calcular los residuos (errores):

$$e_i = Y_i - \hat{Y}_i$$

X	Y	$\hat{Y}_i$	Residuo ( $e_i$ )	$e_i^2$
5	50	46,76 + 1,77(5)	50 - 55,61	31,47
12	70	46,76 + 1,77(12)	70 - 68,96	1,08
19	85	46,76 + 1,77(19)	85 - 80,39	21,24
25	90	46,76 + 1,77(25)	90 - 91,01	1,02
30	95	46,76 + 1,77(30)	95 - 99,86	23,62

2. Sumar los residuos al cuadrado:

$$\sum e_i^2 = 31,47 + 1,08 + 21,24 + 1,02 + 23,62 = 78,43$$

3. Calcular el error cuadrático medio (MSE):

$$MSE = \frac{\sum e_i^2}{n} = \frac{78,43}{5} = 15,69$$

4. Calcular el error cuadrático medio de la raíz (RMSE):

$$RMSE = \sqrt{MSE} = \sqrt{15,69} \approx 3,96$$

### Resumen de resultados

- Intersección ( $\alpha$ ): 46,76.
- Pendiente ( $\beta$ ): 1,77.
- $R^2$ : 0,96.
- RMSE: 3,96.

### Interpretación

La ecuación de regresión  $Y = 46,76 + 1,77X$  sugiere que, por cada producto adicional vendido, los ingresos aumentan aproximadamente en 1,77 unidades. El modelo explica el 96% de la variabilidad en los ingresos, con un error de predicción promedio de 3,96 unidades.

## Ejercicio 6. Predicción de las tasas de conversión basada en la calidad de las campañas de *email*

Un gerente de marketing digital quiere predecir las tasas de conversión en función de las puntuaciones de calidad de las campañas de *email*. Se han recopilado los siguientes datos.

### Datos

Calidad de la campaña de <i>email</i> (X)	Tasa de conversión (Y)
3	5
6	10
9	12
12	18
15	20

Crear un modelo y evaluarlo en función de  $R^2$  y RMSE.

### Solución

- *Paso 1: Calcular las medias de X y Y.*

$$\bar{X} = \frac{3 + 6 + 9 + 12 + 15}{5} = 9$$

$$\bar{Y} = \frac{5 + 10 + 12 + 18 + 20}{5} = 13$$

- *Paso 2: Calcular la covarianza de X y Y.*

$$\text{Cov}(X, Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
3	5	-6	-8	48
6	10	-3	-3	9
9	12	0	-1	0
12	18	3	5	15
15	20	6	7	42

Suma de  $(X - \bar{X})(Y - \bar{Y})$ :

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = 48 + 9 + 0 + 15 + 42 = 114$$

Covarianza:

$$\text{Cov}(X, Y) = \frac{114}{5} = 22,8$$

- *Paso 3: Calcular la varianza de X.*

$$\text{Var}(X) = \frac{\sum (X_i - \bar{X})^2}{n}$$