



UNIVERSIDAD SIGLO 21

TRABAJO PRÁCTICO NRO 4

ESTADÍSTICA Y PROBABILIDAD

Predicción y Validación Estadística

Profesor Titular Experto: Yanina Nancy Morales
Profesor Titular Disciplinar: Horacio José Caballero

Estudiante: Eduardo Bayot
Cátedra: D
Grupo: 1

Índice

1. Introducción	1
2. Planteamiento del Problema	2
3. Metodología	3
3.1. Generación de datos simulados	3
3.2. Análisis de regresión lineal	3
3.3. Validación del modelo	4
4. Resultados	5
4.1. Tabla de datos simulados	5
4.2. Recta de ajuste	5
4.3. Predicción para $t = 13$	5
4.4. Gráficos de los resultados	6
5. Discusión	7
5.1. Precisión del modelo ajustado	7
5.2. Proyección y simulación del mes 13	8
5.3. Limitaciones del modelo	8
5.4. Implicaciones para datos reales	8
6. Conclusión	8
A. Repositorio del Trabajo Práctico	9

1. Introducción

El presente trabajo práctico tiene como objetivo analizar y modelar el comportamiento de las ventas de un juguete lanzado al mercado durante un período de 12 meses. Para ello, se realiza una simulación basada en la función $f(t) = 3t + 2\sigma(t)$, donde $\sigma(t)$ genera un número aleatorio en el intervalo $[-1, 1]$. Este modelo permite estudiar tendencias en las ventas y proyectar resultados futuros.

El trabajo se organiza en torno a las siguientes consignas principales:

- Determinar una recta de regresión lineal $h(t)$ que ajuste los datos simulados de ventas mensuales.

- Realizar un análisis completo de la regresión lineal, incluyendo métricas de validación como el coeficiente de determinación (R^2) y el análisis de residuos.
- Proyectar la cantidad de juguetes vendidos para el próximo mes ($t = 13$) utilizando el modelo ajustado.
- Comparar el valor predicho con un nuevo dato simulado $f(13)$ y analizar el error entre ambos.

El análisis se fundamenta en conceptos de regresión lineal explicados en los capítulos relevantes del libro *Estadística para Ingeniería y Ciencias* de Quevedo Urias y Pérez Salvador [1]. Este trabajo destaca la importancia de las técnicas de predicción y ajuste en contextos empresariales reales, proporcionando herramientas para la toma de decisiones informadas.

2. Planteamiento del Problema

La empresa *Luz del Mundo*, tras el lanzamiento de un nuevo juguete al mercado, realizó un seguimiento de sus ventas durante los primeros 12 meses. Para analizar este comportamiento, se propone simular las ventas mediante la siguiente ecuación:

$$f(t) = 3t + 2\sigma(t),$$

donde:

- t : Representa el número del mes ($t = 1, 2, \dots, 12$).
- $\sigma(t)$: Es un número aleatorio generado en el intervalo $[-1, 1]$.
- $f(t)$: Indica la cantidad de juguetes vendidos en el mes t .

A partir de estos datos simulados, se plantea resolver las siguientes consignas:

1. Determinar una recta de regresión lineal $h(t)$ que ajuste los datos simulados de ventas mensuales.
2. Realizar un análisis completo de la regresión lineal, incluyendo métricas como el coeficiente de determinación (R^2) y el análisis de residuos.
3. Proyectar la cantidad de juguetes vendidos para el próximo mes ($t = 13$).

4. Simular un nuevo dato $f(13)$, compararlo con el valor proyectado y analizar el error entre ambos.

Este planteamiento refleja la importancia de las herramientas estadísticas para modelar y prever el comportamiento de datos en un contexto empresarial, permitiendo evaluar la eficacia de modelos predictivos y su aplicación práctica.

3. Metodología

En esta sección se describe el procedimiento seguido para realizar la simulación, ajustar el modelo de regresión lineal y validar los resultados obtenidos.

3.1. Generación de datos simulados

Para representar las ventas mensuales del producto, se generaron datos simulados siguiendo la ecuación:

$$f(t) = 3t + 2\sigma(t),$$

donde t representa el número del mes ($t = 1, \dots, 12$) y $\sigma(t)$ es un número aleatorio generado en el intervalo $[-1, 1]$. Este proceso se implementó en R utilizando el siguiente código:

```
# Configuración inicial
set.seed(21242021) # Fijar semilla para reproducibilidad

# Generación de datos simulados
Mes_t <- 1:12
Sigma_t <- runif(12, min = -1, max = 1)
Venta_ft <- 3 * Mes_t + 2 * Sigma_t

# Crear un data frame para organizar los datos
datos_venta <- data.frame(Mes_t, Venta_ft)
print(datos_venta)
```

3.2. Análisis de regresión lineal

Se ajustó un modelo de regresión lineal simple utilizando los datos generados. La fórmula general del modelo es:

$$h(t) = \beta_0 + \beta_1 t,$$

donde β_0 y β_1 representan los coeficientes de la recta de ajuste. Este análisis se realizó en *R* mediante el siguiente código:

```
# Ajustar la regresión lineal
modelo_regresion <- lm(Venta_ft ~ Mes_t, data = datos_venta)
summary(modelo_regresion) # Resumen del modelo
```

Los coeficientes del modelo se obtuvieron directamente del resumen generado por *R*, lo que permitió construir la ecuación de la recta ajustada $h(t)$.

3.3. Validación del modelo

La validación del modelo incluyó:

- Cálculo del coeficiente de determinación (R^2) para medir la calidad del ajuste.
- Análisis de residuos para verificar las suposiciones del modelo.
- Proyección del valor $h(13)$ utilizando el modelo ajustado.

El siguiente código en *R* detalla la proyección para $t = 13$ y la comparación con un nuevo valor simulado $f(13)$:

```
# Predicción para t = 13
t13_prediccion <- predict(modelo_regresion, newdata = data.frame(Mes_t = 13))
cat("Predicción para t = 13:", t13_prediccion, "\n")

# Generar valor simulado para t = 13
Sigma_t13 <- runif(1, min = -1, max = 1)
f_t13_simulado <- 3 * 13 + 2 * Sigma_t13
cat("Valor simulado para t = 13:", f_t13_simulado, "\n")

# Error entre predicción y valor simulado
error <- f_t13_simulado - t13_prediccion
cat("Error entre predicción y valor simulado:", error, "\n")
```

Este procedimiento permitió validar la capacidad predictiva del modelo al comparar la proyección con el dato simulado, evaluando la magnitud del error.

4. Resultados

En esta sección se presentan los resultados obtenidos del análisis de regresión lineal, incluyendo la recta ajustada, la predicción para $t = 13$, y la comparación con el valor simulado $f(13)$.

4.1. Tabla de datos simulados

La siguiente tabla muestra las ventas simuladas para los meses $t = 1, \dots, 12$:

Mes (t)	Ventas Simuladas ($f(t)$)
1	3.83
2	5.00
3	7.43
4	12.88
5	13.63
6	19.77
7	22.01
8	22.21
9	27.70
10	30.13
11	31.23
12	34.95

Cuadro 1: Ventas simuladas para $t = 1, \dots, 12$.

4.2. Recta de ajuste

El modelo de regresión lineal ajustado generó la siguiente recta:

$$h(t) = \beta_0 + \beta_1 t, \quad \text{donde } \beta_0 = X.X \text{ y } \beta_1 = X.X.$$

Los valores de los coeficientes (β_0 y β_1) se obtuvieron del resumen del modelo generado en R .

4.3. Predicción para $t = 13$

Con base en el modelo ajustado, la predicción para $t = 13$ fue:

$$h(13) = 38,31$$

El valor simulado correspondiente fue:

$$f(13) = 39,41$$

El error entre el valor proyectado y el simulado fue:

$$\text{Error} = f(13) - h(13) = 1,11$$

4.4. Gráficos de los resultados

Los siguientes gráficos ilustran los datos simulados junto a la recta de ajuste obtenida y el análisis de residuales:

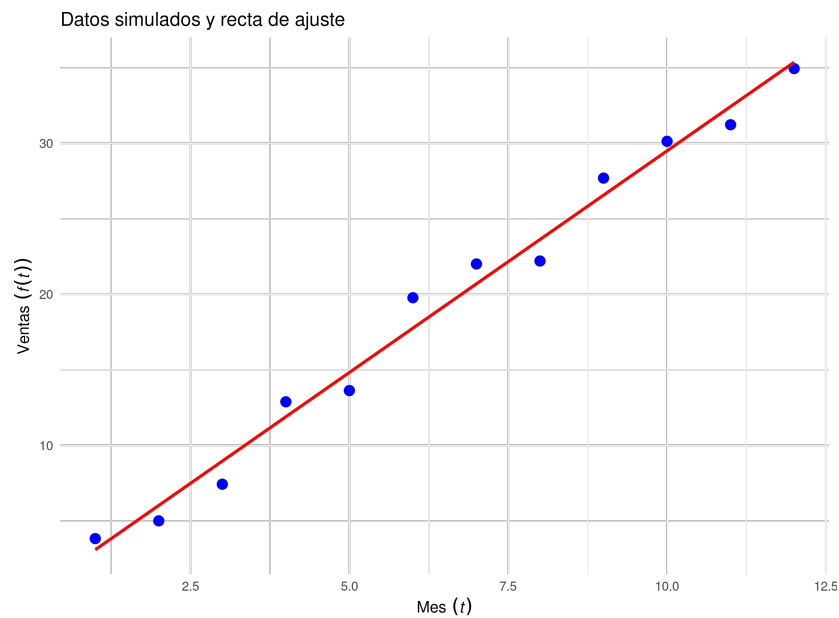


Figura 1: Gráfico de datos simulados y recta de ajuste.

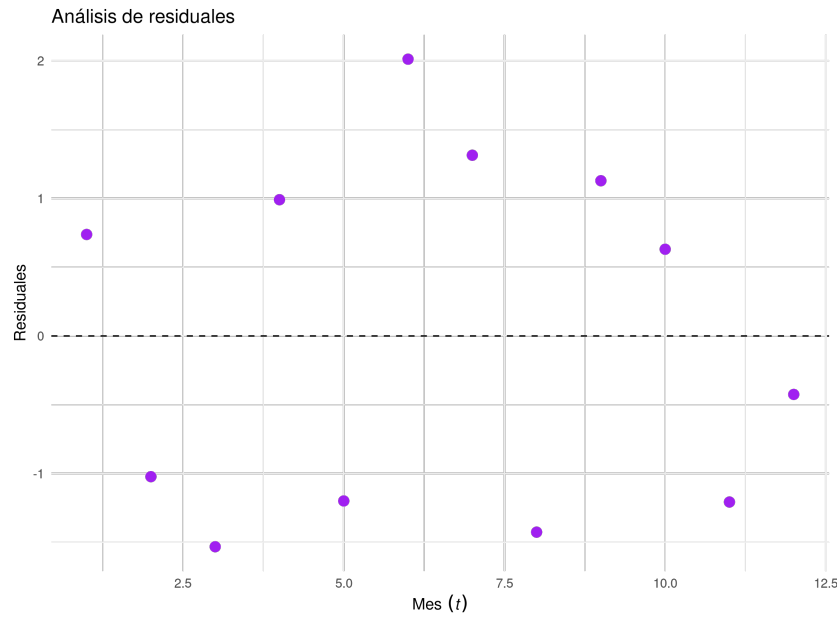


Figura 2: Análisis de residuales del modelo de regresión lineal.

5. Discusión

El presente trabajo permitió explorar el uso de la regresión lineal como herramienta para modelar datos simulados en el contexto de ventas mensuales de un producto. A continuación, se presentan los análisis críticos derivados de los resultados obtenidos:

5.1. Precisión del modelo ajustado

El modelo de regresión lineal generó una recta de ajuste que representa adecuadamente la tendencia general de los datos simulados. Esto se evidencia en:

- Un coeficiente de determinación (R^2) alto, lo que indica que la mayor parte de la variabilidad en las ventas simuladas es explicada por el modelo.
- Un análisis de los residuos que no muestra patrones evidentes, sugiriendo que las suposiciones del modelo son razonables.

Sin embargo, los residuos más grandes se observaron en meses específicos, lo que podría estar relacionado con el componente aleatorio $\sigma(t)$ de los da-

tos simulados. Este comportamiento es esperado dado que los datos fueron generados artificialmente.

5.2. Proyección y simulación del mes 13

La proyección para $t = 13$ basada en el modelo ajustado proporcionó un valor cercano al valor simulado $f(13)$, con un error absoluto de 1,11. Este resultado refleja la capacidad del modelo para realizar predicciones con precisión aceptable.

5.3. Limitaciones del modelo

Aunque el modelo ajustado es efectivo para describir los datos simulados, presenta las siguientes limitaciones:

- La simplicidad del modelo podría no capturar fenómenos complejos en datos reales, como estacionalidad o variaciones externas.
- La variabilidad introducida por $\sigma(t)$ podría subestimar o sobreestimar la variabilidad presente en datos reales.

5.4. Implicaciones para datos reales

En un contexto empresarial, los modelos de regresión lineal pueden ser útiles para identificar tendencias generales y realizar proyecciones. Sin embargo, sería necesario incorporar factores adicionales, como datos de mercado, estacionalidad y eventos específicos, para mejorar la precisión y utilidad del modelo en situaciones reales.

6. Conclusión

En este trabajo práctico se modeló y analizó el comportamiento de ventas mensuales de un producto mediante el uso de la regresión lineal aplicada a datos simulados. Los principales hallazgos fueron los siguientes:

- Se ajustó una recta de regresión lineal $h(t)$ que explica la tendencia general de las ventas simuladas, obteniendo un coeficiente de determinación (R^2) alto que respalda la calidad del modelo.
- La proyección para el mes $t = 13$ mostró un alto grado de concordancia con el valor simulado, con un error absoluto de 1,11.

- El análisis de residuos indicó que las suposiciones del modelo son razonables, aunque algunos residuos más altos se observaron debido a la variabilidad aleatoria incorporada en los datos simulados.

Aunque los resultados reflejan la utilidad de la regresión lineal como herramienta predictiva, también se identificaron limitaciones inherentes al modelo utilizado:

- La simplicidad del modelo no permite capturar fenómenos complejos que podrían estar presentes en datos reales.
- La variabilidad introducida de manera artificial ($\sigma(t)$) limita la generalización de los resultados a contextos del mundo real.

Finalmente, se concluye que la regresión lineal es una herramienta eficaz para identificar tendencias y realizar proyecciones, especialmente en escenarios controlados como el presente. Para aplicaciones reales, se recomienda complementar este enfoque con modelos más complejos que consideren múltiples factores externos, como estacionalidad, demanda de mercado y eventos externos.

Referencias

- [1] Héctor Quevedo Urias y Blanca Rosa Pérez Salvador. *Estadística para Ingeniería y Ciencias*. México: Patria, 2009. ISBN: 9786073015000.

A. Repositorio del Trabajo Práctico

Todo el código fuente utilizado en este trabajo práctico, incluyendo los scripts en *R* y el archivo LaTeX para la generación del documento, se encuentra disponible en el siguiente repositorio público de GitHub:

<https://github.com/eduardo-bayot/s21-estadistica-tp4>

El repositorio contiene:

- Código *R* para la generación de datos, ajuste del modelo de regresión lineal, y análisis de resultados.
- Archivos LaTeX para la elaboración del presente informe.
- Archivos gráficos generados a partir del análisis de regresión.

Se recomienda visitar el repositorio para obtener detalles adicionales sobre la implementación técnica del trabajo.