

Regresión Lineal Simple

Ejercicio 1: Fundamentos Conceptuales

Basándote en el texto, explica con tus propias palabras por qué un coeficiente de correlación de Pearson (r) alto no es suficiente para modelar una relación y por qué la regresión lineal es un paso más allá. Menciona al menos dos cosas que el modelo de regresión proporciona y que la correlación por sí sola no ofrece.

Ejercicio 2: Interpretación de Coeficientes

Un analista ajusta un modelo para predecir el gasto anual en compras online (**gasto**, en euros) basándose en la edad del cliente (**edad**). El modelo ajustado es:

$$\text{gasto} = 1500 + 12 * \text{edad}$$

- ¿Cuál es el gasto predicho para un cliente de 30 años?
- Interpreta el significado de la pendiente (12) en el contexto específico de este problema.
- Interpreta el significado del intercepto (1500). ¿Crees que esta interpretación tiene sentido práctico en el mundo real? ¿Por qué?

Ejercicio 3: Aplicación Práctica con R (Ajuste e Inferencia)

Utiliza el conjunto de datos **pressure** de R, que contiene mediciones de temperatura y presión de vapor de mercurio.

- Ajusta un modelo de regresión lineal simple para predecir la presión (**pressure**) en función de la temperatura (**temperature**). Guarda el modelo en un objeto.
- Utiliza la función **summary()** sobre el objeto del modelo.
- Interpreta el valor del **coeficiente de determinación R^2** . ¿Qué porcentaje de la variabilidad de la presión es explicado por la temperatura?
- Interpreta el **p-valor del estadístico F**. ¿Es el modelo útil en su conjunto?
- ¿Es el coeficiente de la temperatura estadísticamente significativo a un nivel de $\alpha = 0.05$? Justifica tu respuesta basándote en el p-valor del test t.

Ejercicio 4: Intervalos de Confianza y Predicción

Usando el modelo del ejercicio anterior (`lm(pressure ~ temperature, data = pressure)`):

- Calcula el **intervalo de confianza al 95%** para la *presión media* esperada cuando la temperatura es de 250 grados.
- Calcula el **intervalo de predicción al 95%** para la presión de una *única y nueva* medición realizada a 250 grados.
- ¿Cuál de los dos intervalos es más ancho? Explica la razón teórica de esta diferencia.

Ejercicio 5: Supuestos del Modelo

Enumera los cuatro supuestos del modelo de regresión lineal clásico (también conocidos como supuestos de Gauss-Markov) y explica brevemente la importancia de cada uno.

Ejercicio 6: Diagnóstico de Linealidad y Homocedasticidad

Para el modelo del ejercicio 3:

- Genera y muestra el gráfico de **Residuos vs. Valores Ajustados**. Basándote en este gráfico, ¿se cumple el supuesto de **linealidad**? Explica en qué te basas.
- Genera y muestra el gráfico **Scale-Location**. Basándote en este gráfico, ¿se cumple el supuesto de **homocedasticidad**? Describe el patrón que indicaría un problema de heterocedasticidad.

Ejercicio 7: Diagnóstico de Normalidad

Para el modelo del ejercicio 3:

- Genera un gráfico **Normal Q-Q** de los residuos. ¿Parecen seguir los residuos una distribución normal?
- Realiza un **test de Shapiro-Wilk** sobre los residuos del modelo. ¿Qué concluyes a partir del p-valor?

Ejercicio 8: Descomposición de la Varianza (ANOVA)

Explica qué representan la **Suma de Cuadrados Total (SST)**, la **Suma de Cuadrados de la Regresión (SSR)** y la **Suma de Cuadrados del Error (SSE)**. ¿Cuál es la ecuación fundamental que las relaciona?

Ejercicio 9: Observaciones Influyentes

Basado en la teoría de los apuntes:

- a) Explica la diferencia entre un residuo simple (e_i), un residuo estandarizado y un residuo estudentizado. ¿Por qué se prefieren los estudentizados para el diagnóstico?
- b) ¿Qué mide el **leverage** (h_{ii})? ¿Y la **distancia de Cook** (D_i)? ¿Puede una observación tener un leverage alto y no ser influyente?

Ejercicio 10: Relación entre Pruebas de Hipótesis

En el contexto **exclusivo** de la regresión lineal simple, ¿qué relación matemática existe entre el estadístico **F** del test ANOVA y el estadístico **t** del test para la pendiente β_1 ? ¿Qué implica esto para sus respectivos p-valores?