

Guía de Problemas Regresión Lineal Multiple

Formulación. Estimación. Validación.

Problema 1

Un econometrista desea estimar la relación entre la demanda Y de un bien, su precio X_1 , el precio X_2 de un bien sustituto (o de la competencia) y el ingreso medio X_3 de los consumidores. Se propuso el siguiente modelo lineal:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

A partir de una muestra de 30 períodos de tiempo, obtuvo los resultados de las tablas siguientes:

<i>Estadísticas de la regresión</i>	
Coefficiente de correlación múltiple	
Coefficiente de determinación (R^2)	
Error típico	
Observaciones	30

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>
Regresión			
Residuos		28.810.216	
Total		3.649.538.600	

	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>
X_1	-14.870	1594	
X_2	-4.986	4141	
X_3	146,7	25,8	

- a) Complete los cuadros en blanco e interprete la no significatividad de la variable X_2 .

Problema 2

Se han tomado datos de las siguientes variables temporales de un país para los últimos 10 años: - Y : PBI agropecuario, pesca y caza (Miles de millones moneda constante) - X_1 : Área sembrada de oleaginosas (Millones de Ha) - X_2 : Área sembrada de trigo (Millones de Ha) - X_3 : Cabezas de ganado faenadas (Millones)

Con la función de excel se obtuvo el siguiente resultado:

<i>Estadísticas de la regresión</i>	
Coefficiente de correlación múltiple	
Coefficiente de determinación (R^2)	
Error típico	
Observaciones	30

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>
Regresión			
Residuos		28.810.216	
Total		3.649.538.600	

	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>
X_1	-14.870	1594	
X_2	-4.986	4141	
X_3	146,7	25,8	

- a) Complete los cuadros en blanco.
 b) Indique si el modelo puede o no ser válido.
 c) Interprete los coeficientes de regresión, indique si puede haber multicolinealidad y si puede considerarse quitar alguna variable; en caso afirmativo, indique cuál y por qué.

Problema 3

Se estudia el calor desprendido durante la fase de fraguado de un cemento Pórtland en función de su composición. Para ello se tomaron 13 muestras con distintas concentraciones de dos

compuestos (A y B). El ingeniero a cargo del estudio aplicó un modelo de regresión lineal y obtuvo la siguiente ecuación:

$$Y = 52.5773 + 1.4683 X_A + 0.6622 X_B$$

Donde el calor desprendido se mide en calorías por gramo de cemento y las concentraciones se expresan en porcentaje (%)

Información útil:

Suma de Cuadrados de los residuos = 57,90

Suma de Cuadrados Total = 2715,76

Coefficiente de correlación entre las concentraciones: 0.2286

$$(X'X)^{-1} = \begin{bmatrix} 0.9026 & -0.0084 & -0.0158 \\ -0.0084 & 0.0025 & -0.0002 \\ -0.0158 & -0.0002 & 0.0004 \end{bmatrix}$$

- Investigue la validez del modelo y evalúe la multicolinealidad.
- Para un cemento con una concentración del 8% del compuesto A y 70% del B ¿Cuál será el valor del calor desprendido promedio superado con un 95% de probabilidad?

Problema 4

En períodos de fuerte inflación, las entidades financieras pagaban elevados intereses a los ahorristas. Una consecuencia de esto era una fuerte depresión de la demanda de ciertos bienes; así teníamos inflación con recesión, fenómeno sobre el cual no había experiencia y costaba mucho manejar. En una fábrica de automóviles se habían registrados datos de las variables:

- X_1 : Tasa de interés anual para plazo fijo de 30 días al comienzo del mes, expresada en tanto por uno.
- X_2 : Gastos de publicidad del mes anterior en miles dólares.
- Y : Ventas mensuales de automóviles en miles de unidades.

Para un período de 10 meses se obtuvieron los siguientes datos:

Datos: *Datos_Guia_RLM.xlsx* Hoja *Problema4*

- Investigue la validez del modelo lineal general (14-1) para estas variables y compruebe que el modelo que presenta mejor ajuste es el que tiene la variable X_2 (Gastos de publicidad) sola.
- Calcule un intervalo de predicción del 90% para las ventas de un mes en que la empresa gastó, en el mes anterior, 18000 dólares en publicidad.

- c) Analice si el modelo mejora agregando una variable temporal, codificada correlativamente. Interprete en términos del problema el significado del coeficiente que acompaña a la variable temporal.

Problema 5

Un importante laboratorio de capitales europeos estudia las ventas (en cientos de miles de unidades) de uno de sus medicamentos (“Caprex”) en los que es líder en el mercado. Se relevan las ventas mensuales del año pasado, el precio de lista del Caprex (X_2) y el precio del mismo medicamento fabricado por un competidor local (X_3). Un analista propone predecir las ventas mensuales empleando un modelo de regresión lineal múltiple, incluyendo una variable temporal X_1 indicadora del mes (1, 2, 3, ..., 12) obteniendo:

$$\hat{Y} = -2.269 + 0.0858 X_1 - 0.0157 X_2 + 0.128 X_3$$

$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = 0.0358$$

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = 0.2999$$

El determinante de la matriz de correlaciones es 0,038.

$$(X'X)^{-1} = \begin{bmatrix} 66.51263 & -1.31074 & 0.46468 & -3.03748 \\ -1.31074 & 0.03205 & -0.02470 & 0.06962 \\ 0.46468 & -0.02470 & 0.29621 & -0.23449 \\ -3.03748 & 0.06962 & -0.23449 & 0.29460 \end{bmatrix}$$

- Indicar el porcentaje de la variación de las ventas mensuales que el modelo explica y calcular los intervalos de confianza para los coeficientes de regresión (NC = 95%).
- Validar el modelo. ¿Hay multicolinealidad? ¿Puede eliminar alguna variable?. Justifique claramente.
- ¿Cuál es la venta mínima que puede asegurarse con 90% de probabilidad para el mes 14, suponiendo un precio de lista de \$22 y un precio del competidor de \$19?

Problema 6

Considere los siguientes datos (reales) para distintas ciudades de nuestro país:

- Y : Miles de Mw facturados en 1998
- X_1 : Miles de habitantes año 2000
- X_2 : Porcentaje de individuos pobres

Datos: *Datos_Guia_RLM.xlsx* Hoja *Problema6*

- Corra el modelo tipo Cobb-Douglas y verifique su validez.

Problema 7

En una fábrica de conservas de pescado se desea estimar el peso de los pescados en función de sus dimensiones. Las dimensiones las genera un sistema automatizado de cámaras que se encuentra en el ingreso de los pescados en la cinta transportadora, por lo que es más sencillo obtener estas variables que medir el peso de cada pescado. El mismo registra la siguiente información:

- Y : Peso del pescado en gramos
- X_1 : Longitud vertical en pulgadas
- X_2 : Ancho en pulgadas

Datos: *Datos_Guia_RLM.xlsx* Hoja *Problema7*

- Represente el Peso de los pescados vs su longitud y ancho en un gráfico de 3 dimensiones. Entrene y valide un modelo para estimar la variable correspondiente según las otras, y luego agregue el plano de regresión estimado al gráfico.

```
library(rgl)
library(readxl)

datos <- read_excel("Datos_Guia_RLM.xlsx", sheet = "Problema7")
names(datos) <- c("peso", "ancho", "longitud")

plot3d(x=datos$ancho, y=datos$longitud, z=datos$peso,
       col = "blue",
       xlab = "Ancho", ylab = "Longitud", zlab = "Peso",
       type = "s")

modelo_rlm <- lm(formula = peso ~ ancho + longitud, data = datos)
resumen_rlm <- summary(modelo_rlm)
```

```
# Ahora agregamos el plano de regresion
b_0 <- resumen_rlm$coefficients[1,1]
b_ancho <- resumen_rlm$coefficients[2,1]
b_longitud <- resumen_rlm$coefficients[3,1]

# plano de regresion estimado
planes3d(a = -b_ancho, b = -b_longitud, c = 1, d = -b_0, col = "lightgrey") #el software gra
```

- Realice un diagramas de dispersión entre las dos variables explicativas. Calcule el coeficiente de correlación entre ambas variables. ¿Como se relacionan estos resultados con el modelo de regresión?
- Represente el Peso de los pescados vs su longitud en un gráfico. Y en otro gráfico, represente el Peso de los pescados vs su ancho. Entrene y valide dos modelos de regresión lineal simple, uno en que la variable explicativa sea la longitud y otro que la variable explicativa sea el ancho.
- ¿Cambian los coeficientes de regresión estimados de cada variable para los modelos de regresión lineal simple frente a los coeficientes estimados del modelo de regresión lineal múltiple? ¿Por qué cree que suceda esto?
- Analice la multicolinealidad de las variables explicativas, mediante indicadores adecuados. ¿Cuál sería el principal inconveniente? ¿Qué otros inconvenientes conlleva este? Tome una decisión respecto a la multicolinealidad y sobre qué hacer con el modelo.

Problema 8

En una empresa de productos de retail, se desea poder predecir las ventas trimestrales en base a variables como los gastos realizados en el trimestre anterior y el precio estimado de un producto sustituto. Del sector de Ventas se obtuvieron datos de unidades vendidas durante los últimos 5 años; mientras que del área de Costos se obtuvieron datos de gastos de publicidad y gastos de desarrollo de producto. Además, el área Comercial le brinda las estimaciones realizadas del precio del principal producto sustituto. Obteniéndose datos de las siguientes variables:

- Y : ventas de este trimestre [miles]
- X_1 : gastos publicidad del trimestre anterior [miles de \$]
- X_2 : gastos desarrollo del trimestre anterior [miles de \$]
- X_3 : precio estimado de producto sustituto de este trimestre

Datos: *Datos_Guia_RLM.xlsx* Hoja *Problema8*

- a. Estime un modelo para estimar la variable correspondiente según las otras. Luego, estime un modelo de RLS por cada variable explicativa. Observe que sucede con los coeficientes de regresión estimados.
- b. Represente la variable de respuesta vs cada una de las explicativas en distintos gráficos.
- c. Represente la relación entre las variables explicativas en tres gráficos de dispersión (X_1 vs X_2 , X_1 vs X_3 , X_2 vs X_3).
- d. Analice la multicolinealidad de las variables explicativas, mediante indicadores adecuados. ¿Cuál sería el principal inconveniente? ¿Qué otros inconvenientes conlleva este?