

Guía de Problemas Regresión Lineal Multiple

Parte 2

Problemas de Análisis confirmatorio

Problema 1

- 1) Un estadístico efectúa un análisis exploratorio del modelo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \epsilon$$

Plantea todos los posibles modelos, calcula los indicadores de la tabla y luego de una primera selección se queda con los siguientes:

Modelo	R ²	S ²	DET	PRESS	p	C _p
X1 X2 X3 X4 X6	0,981258	0,671667	0,013287	33,72427	6	10,9309
X1 X2 X3 X4 X5 X6	0,984113	0,586595	0,00643	34,09261	7	7
X1 X2 X3 X4	0,975207	0,863134	0,014217	41,15499	5	21,50009
X1 X2 X3 X5 X6	0,980009	0,716428	0,329847	41,64223	6	13,52534
X1 X2 X3 X4 X5	0,976354	0,847442	0,007274	43,90814	6	21,11909
X1 X2 X3 X6	0,975119	0,866224	0,645986	44,957	5	21,68445
X1 X2 X3	0,967983	1,083689	0,686584	52,62116	4	34,50722

- a. ¿Cuántos posibles modelos se plantearon? ¿Qué modelo elegiría? Justifique detalladamente su decisión.
- b. Explique que son el DET, el PRESS y el C_p.

Problema 2

En una auditoría solicitada por la Secretaría de Pesca fue necesario relacionar las variables: - Y: Volumen de pesca de merluza registrada en partes de pesca de buques. - X1: Capacidad de

bodega del buque - X2: Potencia en HP - X3: Eslora - X4: Manga - X5: Días de navegación - X6: Otras especies capturadas - X7: Total de merluza en el período

El modelo se utilizará para analizar partes futuros y, en caso de que el volumen denunciado en un parte presente una diferencia significativa con la estimación del modelo, auditar el buque. Para el conjunto de 19541 partes de los años 1989 a 1996, se corrieron los 127 modelos que se ordenaron por varianzas crecientes, obteniendo la siguiente tabla, limitada a los primeros 20.

TABLA DE MODELOS		PROBLEMA 153			
Modelo	R ²	S ²	DET	PRESS	CP/P
X1, X2, X3, X4, X5, X6, X7	0,7387	1,4083E+09	0,0080	2,6581E+13	1,00
X1, X2, X3, X4, X6, X7	0,7387	1,4083E+09	0,0095	2,6576E+13	1,01
X1, X2, X3, X4, X5, X6	0,7356	1,4248E+09	0,0100	2,6891E+13	32,49
X1, X2, X3, X4, X6	0,7355	1,4250E+09	0,0100	2,6889E+13	38,14
X1, X3, X4, X5, X6, X7	0,7356	1,4251E+09	0,0430	2,6881E+13	32,90
X1, X3, X4, X6, X7	0,7355	1,4252E+09	0,0510	2,6878E+13	38,49
X2, X3, X4, X6, X7	0,7334	1,4369E+09	0,0359	2,7041E+13	64,45
X2, X3, X4, X5, X6, X7	0,7334	1,4369E+09	0,0303	2,7048E+13	55,51
X1, X3, X4, X5, X6	0,7326	1,4411E+09	0,0442	2,7181E+13	73,78
X1, X3, X4, X6	0,7325	1,4414E+09	0,0520	2,7182E+13	89,29
X1, X2, X3, X6, X7	0,7314	1,4471E+09	0,0855	2,7298E+13	87,23
X1, X2, X3, X5, X6, X7	0,7315	1,4472E+09	0,0720	2,7305E+13	75,01
X2, X3, X4, X6	0,7303	1,4534E+09	0,0369	2,7350E+13	121,23
X2, X3, X4, X5, X6	0,7303	1,4535E+09	0,0313	2,7357E+13	101,33
X3, X4, X6, X7	0,7293	1,4584E+09	0,1897	2,7431E+13	134,66
X3, X4, X5, X6, X7	0,7293	1,4585E+09	0,1606	2,7438E+13	112,55
X1, X2, X3, X5, X6	0,7271	1,4706E+09	0,0730	2,7745E+13	139,43
X1, X2, X3, X6	0,7271	1,4707E+09	0,0867	2,7741E+13	167,37
X3, X4, X6	0,7264	1,4744E+09	0,1950	2,7731E+13	221,39
X3, X4, X5, X6	0,7264	1,4744E+09	0,1656	2,7737E+13	177,35

a. Indique el subconjunto de variables que Ud. considere más apropiado. Justifique.

Problema 3

Los gasoductos de acero transportan, aparte de gas, corriente eléctrica a efectos de impedir la corrosión. El ente regulatorio de nuestro país contrató los servicios de una empresa para que efectuase mediciones del potencial de los caños, que debe tener un nivel mínimo con respecto a tierra, a efectos de garantizar una adecuada protección. El procedimiento empleado fue medir una serie de puntos de la red de gasoductos de cada ciudad y luego armar un modelo para estimar el potencial de cualquier otro punto de la red. Para dicho modelo, del tipo (14-1), se consideraron las siguientes variables:

- Y: Potencial del punto a estimar
- X1: Potencial del rectificador más próximo (mv)
- X2: Corriente del rectificador más próximo (Amp)
- X3: Tensión del rectificador más próximo (Volts)
- X4: Potencial medido en el punto más próximo

- X_5 : Relación X_3/X_2

Datos: *Datos_Guia_RLM_p2.xlsx* Hoja *Problema3*

Efectúe una corrida con excel para las 5 variables explicativas y luego analice un modelo con 4 variables y seleccione uno de los dos.

Problema 4

Ventas de Lotrial. En un estudio de mercado, un laboratorio de productos medicinales desea formular un modelo para pronosticar las ventas de Lotrial, que es un producto antihipertensivo, cuya droga base es el Maleato de Enalapril. A tal efecto, registró datos de las siguientes variables:

- Y : Ventas mensuales de Lotrial en cientos de miles de unidades
- X_1 : Tiempo en meses (1, 2, ..., 12), donde 1 es enero 1987
- X_2 : Precio unitario en marcos
- X_3 : Precio de la competencia en marcos
- X_4 : Precio relativo Lotrial/Competencia
- X_5 : Muestras médicas en miles de unidades
- X_6 : Gastos en publicidad en miles de marcos

Datos: *Datos_Guia_RLM_p2.xlsx* Hoja *Problema4*

- a. No es necesario decir que el modelo que se ajusta mejor a los datos es el 1, 3, que tiene la menor varianza y también el menor press. Córralo con la planilla excel y obtenga la ecuación predictora.

Problema 5

Para poder determinar el impacto de distintos atributos de teléfonos móviles sobre la valoración que de los mismos, se realizó una encuesta a 150 usuarios de teléfonos móviles registrándose las características de sus equipos celulares y la evaluación que hacen de los mismos (medidas en una escala de 1-10).

- y : Evaluacion [1-10]
- x_1 : Definicion [Mpx]
- x_2 : Zoom total: optico x digital [x]
- x_3 : Memoria [Mb]
- x_4 : Peso [g]
- x_5 : Tamano pantalla: diagonal [mm]

- a. Encontrar el mejor modelo, a su criterio, para explicar la evaluación de los usuarios.

- b. Valide el modelo elegido e interprete sus coeficientes en términos del problema
- c. Posee este modelo puntos que puedan ser considerados outliers o influyentes? Justifique

Problema 6

Una empresa constructora está preocupada por la cantidad de penalizaciones que paga por demoras en la entrega de los proyectos. Como en el corto plazo debe hacer una propuesta para la construcción de un puente, desea construir un modelo predictivo para tener una estimación del tiempo de entrega del proyecto. Para eso relevó las especificaciones los últimos proyectos similares junto con el tiempo real de entrega. Las variables son las siguientes:

- *Case*: Número de proyecto
- *Time*: Tiempo de entrega en días
- *DArea*: Área de la superficie transitable del puente
- *CCost*: Costo con el que se ganó la licitación
- *Dwgs*: Cantidad de planos en el pliego de licitación
- *Length*: Longitud del puente
- *Spans*: Cantidad de Pilares

Datos: *Datos_Guia_RLM_p2.xlsx* Hoja: *Problema6*

- a. Corra el siguiente modelo: `modelo1 <- lm(CCosto ~ Case, data = datos)`. Analice la significancia del modelo. ¿Le parece un modelo razonable? Si fuera así, ¿cómo lo utilizaría? ¿Qué interpretación tendría la pendiente? Grafique el diagrama de dispersión.
- b. Genere un modelo completo con todas las variables relevantes para el problema. ¿Es un buen modelo? Justifique. Piensa que puede haber algún problema de multicolinealidad? Mas allá de la significancia o no del modelo evalúe la presencia de multicolinealidad.
- c. Aplique un método stepwise hacia atrás, hacia adelante y hacia ambos lados. ¿Los modelos son iguales? ¿Los modelos son buenos? ¿Significativos? Evalúe nuevamente la multicolinealidad.
- d. Utilice la librería `leaps` de R para estimar todos los modelos posibles. Seleccione el modelo mas razonable.
- e. De acuerdo a lo analizado, seleccione el mejor modelo. Realice un análisis de la normalidad de los residuos. Se cumple el supuesto de normalidad?
- f. Realice un histograma de la variable *Time*, observe la asimetría, es positiva pero no tan fuerte como una lognormal. Para normalizar un poco la variable podría aplicar alguna transformación no tan fuerte como el logaritmo (por ejemplo la raíz cuadrada o cúbica). Pruebe aplicar alguna de estas transformaciones, hacer el histograma de la variable *Time* para confirmar la corrección, volver a estimar el modelo y analizar nuevamente los residuos. ¿Qué conclusiones puede sacar?

- g. Utilice el software Regre.exe para estimar todos los modelos posibles y seleccionar el mas apropiado, llega a la misma conclusión?