

Universidad Nacional de Colombia – Sede Medellín

Escuela de Estadística – Semestre 2021-01

Informe 2 – Estadística II - David Arango Londoño

Objetivo: Resolver un problema práctico usando lo aprendido del análisis de regresión lineal.

Problema. En un estudio a gran escala realizado en EE.UU sobre la eficacia en el control de infecciones hospitalarias se recogió información en 113 hospitales, los datos se encuentran en publicados junto con el este archivo (datos2.txt). La base de datos contiene las siguientes columnas (**variables**):

Abr. y Nombre de la variable	Descripción
Y: Riesgo de infección	Probabilidad promedio estimada de adquirir infección en el hospital (en porcentaje).
X1: Duración de la estadía	Duración promedio de la estadía de todos los pacientes en el hospital (en días).
X2: Rutina de cultivos	Razón del número de cultivos realizados en pacientes sin síntomas de infección hospitalaria, por cada 100.
X3: Número de camas	Número promedio de camas en el hospital durante el periodo del estudio.
X4: Censo promedio diario	Número promedio de pacientes en el hospital por día durante el periodo del estudio.
X5: Número de enfermeras	Número promedio de enfermeras, equivalentes a tiempo completo, durante el periodo del estudio.

Preguntas a resolver.

1. Estime un modelo de regresión lineal múltiple que explique el Riesgo de Infección en términos de todas las variables predictoras. Analice la significancia de la regresión y de los parámetros individuales. Interprete los parámetros estimados. Calcule e interprete el coeficiente de determinación múltiple R^2 . Comente los resultados.
2. Use la tabla de todas las regresiones posibles, para probar la significancia simultánea del subconjunto de tres variables con los valores p mayores del punto anterior. Según el resultado de la prueba es posible descartar del modelo las variables del subconjunto?.
3. Plantee una pregunta donde su solución implique el uso **exclusivo** de una prueba de hipótesis lineal general de la forma $H_0 : \mathbf{L}\beta = \mathbf{0}$ (solo se puede usar este procedimiento y no SSextra), donde especifique claramente la matriz \mathbf{L} , el modelo reducido y la expresión para el estadístico de prueba.
4. Realice una validación de los supuestos en los errores y examine si hay valores atípicos, de balanceo e influenciales. Qué puede decir acerca de la validez de éste modelo?. Argumente.
5. Verificar la presencia de multicolinealidad usando graficos y/o indicadores apropiados.

Instrucciones:

- Se deben resolver cada uno de los puntos y publicar los resultados en Rpubs y compartir el link al correo del profesor davarangolon@unal.edu.co