

Trabajo 4

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. A su equipo de trabajo le corresponde analizar una muestra aleatoria de n hospitales, que están dentro de un archivo de texto adjunto, donde n es el número de registros en el archivo asignado y X es el número de equipo asignad. La base de datos contiene las siguientes columnas (variables):

- 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.
1. Emplee el análisis de regresión lineal múltiple que explique el riesgo de infección en términos de las variables restantes (actuando como predictoras X_i).
 2. Identifique observaciones que puedan considerarse problemáticas (datos atípicos, puntos de balanceo e influyentes) y analice si debe eliminarlas de su conjunto de datos o no, justifique. Repita la construcción del modelo de regresión si eliminó observaciones.
 3. Realice la prueba de significancia del modelo, interprete.
 4. Obtener el coeficiente de determinación y el coeficiente de determinación ajustado. Interprete.
 5. Analice si hay problemas de multicolinealidad.
 6. Realice una selección e variables por el método que prefiera, tome decisiones, explique.
 7. Realice una predicción utilizando el modelo seleccionado, interprete.