

Universidad Nacional de Colombia – Sede Medellín
Escuela de Estadística – Semestre 2023-1S

Trabajo Corto 01 – Estadística II

Fecha de entrega: hasta las 11:59 p.m. del 30 de Marzo de 2023 (al e-mail del profesor, de máximo 12 páginas y marcado en la portada con el número de equipo asignado)

Objetivo: Usar de manera eficiente las herramientas del análisis de regresión para resolver un problema práctico.

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. A su equipo de trabajo le corresponde analizar una muestra aleatoria de n hospitales, que están dentro de un archivo de texto etiquetado **EquipoXX.txt**, donde n es el número de registros en el archivo asignado y **XX** es el número de equipo asignado por su profesor. Las bases de datos, la asignación de los equipos y los archivos necesarios para resolver el trabajo, serán enviados por su profesor. Cada base de datos contiene las siguientes columnas (**variables**):

Variable	Descripción
Y: Riesgo de infección	Probabilidad promedio estimada de adquirir infección en el hospital (en porcentaje).
X_1 : Duración de la estadía	Duración promedio de la estadía de todos los pacientes en el hospital (en días).
X_2 : Rutina de cultivos	Razón del número de cultivos realizados en pacientes sin síntomas de infección hospitalaria, por cada 100.
X_3 : Número de camas	Número promedio de camas en el hospital durante el periodo del estudio.
X_4 : Censo promedio diario	Número promedio de pacientes en el hospital por día durante el periodo del estudio.
X_5 : 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 las variables restantes (actuando como 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 .
2. Use la tabla de todas las regresiones posibles, para probar la significancia simultánea del subconjunto de tres variables con los valores p más grandes del punto anterior. Según el resultado de la prueba es posible descartar del modelo las variables del subconjunto? Explique su respuesta.
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). Especifique claramente la matriz \mathbf{L} , el modelo reducido y la expresión para el estadístico de prueba (no hay que calcularlo).
4. Realice una validación de los supuestos en los errores y examine si hay valores atípicos, de balanceo e influencias. Qué puede decir acerca de la validez de éste modelo?. Argumente su respuesta.

Instrucciones:

- Junto a los ya mencionados archivos del trabajo se entrega el archivo `Trab1_Rcode.R` que incluye el código *R* necesario para el trabajo y que puede abrirse en **RStudio** o en el editor de scripts de *R*.
 - Descargue los archivos del trabajo en su PC: `Trabajo_Corto_01.pdf`, `Trab1_Rcode.R`, `Functions.R` y la base de datos correspondiente `EquipoXX.txt`, donde `XX` es el número de equipo asignado por su profesor.
 - Abra el archivo `Trab1_Rcode.R` en **RStudio** o en el editor de scripts de *R*. Dentro de este archivo, busque y reemplace a `XX` por el número de equipo asignado por su profesor. Finalmente, ejecute en su totalidad el código que contiene sin hacer más cambios al mismo (debe contar con conexión a internet para la instalación de las librerías requeridas).
 - En la lectura de la base de datos, se le pedirá seleccionar el archivo `TXT` de la base de datos, asegúrese de seleccionar la base de datos asignada por el profesor para su equipo de trabajo.
 - Con los resultados producidos resuelva las preguntas del trabajo.
-