

# Universidad Nacional de Colombia – Sede Medellín

## Escuela de Estadística – Semestre 2021-02

### Trabajo 1 – Estadística II

**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. 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 `XX` es el número de equipo asignado por el profesor (la asignación de los equipos está disponible en los archivos: `Asignación_Equipos_Estadística_II_Grupo_01.pdf` y `Asignación_Equipos_Estadística_II_Grupo_02.pdf`). Todas las bases de datos, la asignación de los equipos y los archivos necesarios para resolver el trabajo, están disponibles en [Archivos\\_Trabajo\\_01](#). Cada base de datos contiene las siguientes columnas (**variables**):

Abr. y Nombre de la variable	Descripción
<b>Y:</b> Riesgo de infección	Probabilidad promedio estimada de adquirir infección en el hospital (en porcentaje).
<b>X1:</b> Duración de la estadía	Duración promedio de la estadía de todos los pacientes en el hospital (en días).
<b>X2:</b> Rutina de cultivos	Razón del número de cultivos realizados en pacientes sin síntomas de infección hospitalaria, por cada 100.
<b>X3:</b> Número de camas	Número promedio de camas en el hospital durante el periodo del estudio.
<b>X4:</b> Censo promedio diario	Número promedio de pacientes en el hospital por día durante el periodo del estudio.
<b>X5:</b> 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  $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$  (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 influencias. Qué puede decir acerca de la validez de éste modelo?. Argumente.

### 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_01.pdf`, `Trab1_Rcode.R`, `funciones.R`, `Asignación_Equipos_Estadística_II_Grupo_01.pdf`, `Asignación_Equipos_Estadística_II_Grupo_02.pdf` y la base de datos correspondiente en la carpeta de nombre Bases: `EquipoXX.txt`, donde `XX` es el número de equipo asignado por el profesor.
  - Abra el archivo `Trab1_Rcode.R` en **RStudio** o en el editor de scripts de *R*, y ejecute en su totalidad el código que contiene sin hacer 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.
-