

Universidad Nacional de Colombia – Sede Medellín

Escuela de Estadística – Semestre 2022-01

Trabajo 1 – Estadística II

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

Problema. Con el fin de modelar el rendimiento de un supervisor y la satisfacción de sus empleados, una empresa realizó un estudio a 50 empleados. Para la realización de dicho estudio la empresa tuvo en cuenta las siguientes variables: la calificación global del trabajo hecho por el Supervisor (Y); la tasa de manejo de quejas de los empleados (X_1); la tasa de no permitir privilegios especiales (X_2); la tasa de oportunidad de aprender cosas nuevas (X_3); la tasa de avance del supervisor a mejores puestos (X_4). Los datos correspondientes a su equipo de trabajo están dentro de un archivo de texto etiquetado con el nombre: `EquipoXX.txt`, donde XX es el número de equipo asignado por el profesor (la asignación de los equipos está disponible en el archivo `Asignación_Equipos_Estadística_II_Grupos_01_y_02.pdf`). Todas las bases de datos, la asignación de los equipos y los archivos necesarios para resolver el trabajo, están disponibles en la carpeta del DRIVE compartida con los estudiantes.

Preguntas a resolver.

1. Estime un modelo de regresión lineal múltiple que explique la calificación global del trabajo hecho por el Supervisor 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 las dos 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`, `Functions.R`, `Asignación_Equipos_Estadística_II_Grupos_01_y_02.pdf` y la base de datos correspondiente `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.