

TAREA 1

Modelo de regresión lineal múltiple

1. Base de datos

Se desea construir un modelo de regresión lineal múltiple para predecir la variable “median house value” a partir de seis variables económicas y demográficas que se describen en la archivo adjunto “cadata.txt”.

2. Asignación

Seleccione 500 observaciones consecutivas. Para elegir la primera de sus 500 observaciones, escoja un número aleatorio entre 1 y el total de observaciones. Observe que aunque es posible, es poco probable que dos grupos elijan el mismo conjunto de datos. Reporte los números de las observaciones elegidas, por ejemplo, “elegimos las observaciones 1 a 500” (lo que ocurriría únicamente si su generador de números aleatorios escoge el número 1). Si el número que escoge es muy alto, de tal manera que al llegar a la última observación aún no tiene 500, regrese a la primera observación y agregue consecutivamente las que le hagan falta.

Una vez tenga su subconjunto de 500 observaciones, conduzca una análisis exploratorio de los datos, incluyendo todas las 9 variables. Una vez crea que ya está familiarizado con la información, ajuste un modelo de regresión múltiple usando las variables “median income”, “housing median age”, “total rooms”, “total bedrooms”, “population” y “households” como variables explicativas y la variable “median house value” como variable de respuesta.

Reporte el modelo de regresión ajustado, sin hacer transformaciones de variables y sin hacer validación de supuestos.

3. El informe

- Se entrega en grupos de 2 integrantes.
- Se escribe en L^AT_EX.
- El análisis exploratorio se hace con R.
- El modelo de regresión también se ajusta con R.
- El código de R no hace parte del informe.
- No debe tener más de 6 páginas, incluida la bibliografía.
- Se entrega en el Campus virtual, a través de un enlace dispuesto para ello.
- Cada grupo debe subir al Campus una presentación elaborada con Beamer, conforme con lo previsto en el programa del curso.

4. Algunas recomendaciones

- Estudie con cuidado qué busca el estudio y cuál es el significado de las variables **antes** de iniciar el análisis exploratorio y antes de intentar ajustar el modelo de regresión.
- Renombre sus variables con nombres en español que reflejen lo que cada variable mide.
- Evite referirse a las variable con nombres genéricos tales como Y o X_4 .
- Recuerde: reporte el modelo ajustado con las variables tal como aparecen en el archivo de datos.
- En todo caso, opine sobre sus resultados usando los conceptos estadísticos ya aprendidos.
- En la medida de lo posible, limite el número de figuras en su informe escrito al mínimo razonable para justificar sus interpretaciones. Recuerde que solamente dispone de seis páginas.
- Al elaborar su presentación oral, tenga presente que el tiempo máximo de exposición es de 15 minutos.
- Elabore buenas diapositivas, con gráficos de alta resolución. Si desea mostrar más de una figura para ilustrar uno de sus resultados, use una diapositiva para cada una, a menos que usted crea que se deben mostrar juntas por alguna razón. Incluso en este caso, podría mostrarlas juntas en una diapositiva y luego individualmente.
- Sea imaginativo y creativo, sin exageraciones. Todos aprenderemos de su iniciativa.