

Examen Final Regresión Avanzada

Universidad Austral

Diciembre de 2022

Apellido y Nombre:.....

Ejercicio 1 En el archivo **preciocasas.xlsx** se han registrado respecto de 100 viviendas las siguientes variables:

- **impuestos:** valor de impuesto anual de la vivienda.
- **dormitorios** cantidad de ambientes de la vivienda.
- **banios:** cantidad de baños del inmueble.
- **estrena:** si es a estrenar.
- **precio:** valor del alquiler de la vivienda
- **tamano:** superficie total de la vivienda.

1. Construir un modelo lineal simple para explicar el precio en función de la superficie y evaluar la bondad del ajuste.
2. Realizar un análisis diagnóstico y de puntos influyentes e indicar si el modelo es adecuado.
3. Realizar una transformación de la variable respuesta para lograr normalidad en la distribución de los residuos. Indicar si el modelo con esta transformación resulta adecuado.
4. Eliminar la observación 64 y ajustar nuevamente el segundo modelo evaluando su validez.
5. Ajustar un modelo robusto y evaluar el promedio de los errores absolutos cometidos. Comparar con el mejor modelo lineal disponible.
6. Utilizar un método de selección de variables para proponer un modelo multivariado. Analizar el cumplimiento de los supuestos.
7. Le parece adecuado un modelo GAMLSS en este caso? Justifique.
8. Resuma sus conclusiones.

Ejercicio 2 Se desea saber si la dosis de ácido ascórbico y el tipo de bebida en la cual se lo administró a ciertos animales de laboratorio logró mayor desarrollo de los dientes en los mismo. Se utilizaron 60 replicaciones del experimento y se tienen grupos balanceados. La variable respuesta de interés es la longitud de los dientes frontales(len). Los resultados están en el archivo **odonto.csv** Se pide analizar, analítica y gráficamente, si:

1. Existen diferencias estadísticamente significativas respecto de las dosis administradas?
2. Existen diferencias estadísticamente singificativas respeccto del tipo de vehículo de administración?
3. La interacción entre estas variables es significativa?
4. Se satisfacen los supuestos del modelo?
5. Puede realizar una recomendación?

Ejercicio 3 En el archivo **morosos.xlsx** se encuentran los registros de 10 mil clientes de un banco para los cuales se relevaron las siguientes variables:

- mora: si está en mora con el saldo de su tarjeta de crédito.
- estudiantes: si es estudiante o no.
- balance: el saldo al 31/12 próximo pasado.
- ingreso: ingreso mensual medio del cliente.

Se desea modelar la probabilidad de mora de un cliente en función de las variables disponibles en la base.

Se pide:

1. Ajustar un modelo logístico para predecir la probabilidad de incurrir en mora.
2. Evaluar la calidad de ajuste del modelo con al menos dos criterios distintos.
3. Interpretar los coeficientes del modelo elegido.
4. Evaluar la calidad de clasificación y compararlo con otro método de clasificación.