

## Taller RLM (parte 1)- Grupo 4

Fecha de entrega: Hasta las 5:00 pm del 13 de enero de 2022

El trabajo de RLM tiene dos partes, la primera parte tiene un valor del 60 % y la segunda del 40 %. Cada una debe ser presentada por todos los miembros del equipo. Las presentaciones de esta primera parte se realizarán los días 14 y 19 de enero de 2022.

En la primera parte se hará el análisis de la base de datos `AdmissionPredict.csv`. Se tendrá con variable de respuesta `Chance.of.Admit` y cinco covariables `GRE.Score`, `TOEFL.Score`, `SOP`, `LOR`, `CGPA`. **Únicamente deben realizar el ajuste con las primeras 100 observaciones.** Realice las siguientes actividades:

1. Realice una descripción de la base de datos. Contextualice el problema y explique cada una de las variables involucradas en el modelo. (<https://tanwirkhan.medium.com/predicting-graduate-admissions-using-linear-regression-bab05b6988b5>).
2. Realice un análisis descriptivo de las variables que se van a tener en cuenta en el modelo. Concluya.
3. Ajuste un modelo de regresión lineal múltiple, muestre la tabla de parámetros ajustados y escriba la ecuación ajustada. Calcule la Anova del modelo ¿Es significativo el modelo? ¿Qué proporción de la variabilidad total de la respuesta es explicada por el modelo? Opine sobre esto último.
4. Calcule los coeficientes de regresión estandarizados y concluya acerca de cuál de las variables aporta más a la respuesta según la magnitud en valor absoluto de tales coeficientes (cuidado, no confunda esto con la significancia de los coeficientes de regresión).
5. Pruebe la significancia individual de cada uno de los parámetros del modelo (excepto intercepto), usando la prueba t. Establezca claramente la prueba de hipótesis y el criterio de decisión.
6. Teniendo en cuenta los resultados anteriores, realice una prueba con sumas de cuadrados extras con test lineal general; especifique claramente el modelo reducido y completo, estadístico de la prueba, su distribución, cálculo de valor P, decisión y conclusión a la luz de los datos. Justifique la hipótesis que desea probar en este numeral.
7. Calcule las sumas de cuadrados tipo I (secuenciales) y tipo II (parciales) ¿Cuál de las variables tienen menor valor en tales sumas? ¿Qué puede significar ello?
8. Construya y analice gráficos de los residuales estudentizados vs. Valores ajustados y contra las variables de regresión utilizadas. ¿Qué información proporcionan estas gráficas?
9. Construya una gráfica de probabilidad normal para los residuales estudentizados. ¿Existen razones para dudar de la hipótesis de normalidad sobre los errores en este modelo?
10. Diagnostique la presencia de observaciones atípicas, de balanceo y/o influencias y concluya.
11. Ajuste el modelo de regresión sin las observaciones 10, 38 y 92, suponga que se establece que hay un error de digitación con estas dos observaciones, presente sólo la tabla de parámetros ajustados resultante ¿Cambian notoriamente las estimaciones de los parámetros, sus errores estándar y/o la significancia? ¿Qué concluye al respecto? Evalúe el gráfico de normalidad para los residuales estudentizados para este ajuste ¿mejoró la normalidad? Concluya sobre los efectos de este par de observaciones.
12. Para el modelo con todas las variables y sin las observaciones 10, 38 y 92, realice diagnósticos de multicolinealidad mediante
  - a) Matriz de correlación de las variables predictoras

b) VIF's

c) Proporciones de varianza

13. En el modelo ajustado sin las observaciones 10, 38 y 92, construya modelos de regresión utilizando los métodos de selección (muestre de cada método sólo la tabla de resumen de este y la tabla ANOVA y la de parámetros estimados del modelo finalmente resultante):

a) Selección según el  $R^2_{adj}$

b) Selección según el estadístico  $C_p$

c) Stepwise

d) Selección hacia adelante o forward

e) Selección hacia atrás o backward

14. Con base en los anteriores numerales, ¿Cuál modelo sugiere para la variable respuesta? ¿por qué?

A seguir una ayuda R.

---

```
#LEER DATOS EN archivo asignado a su grupo, así
datos=read.table(file.choose(),header=T,sep=" ",dec=".")
datos=datos[1:100,-c(1,4,8)]
names(datos)
dim(datos)
```

---