

Actividad. Regresión Lineal Múltiple

En un fichero Rmarkdown (‘.Rmd’) realice los ejercicios que se plantean a continuación:

1. Cargue el fichero “spear.xls”
2. Realice un análisis exploratorio de las variables. Comente brevemente los resultados.
3. Verifique los supuestos que han de cumplir las variables implicadas en un análisis de la regresión lineal múltiple. Comente los resultados.
4. Visualice la distribución de frecuencia de las variables.
5. ¿Qué valores ha de registrar la prueba de Shapiro-Wilk para indicar que se trata de una variable con distribución normal?
6. ¿Cómo verificaría el supuesto de linealidad?
7. Visualice la matriz de dispersión entre variables para averiguar la intensidad de la relación entre dos variables.
8. Seleccione y visualice las variables que mayor correlación poseen en una nueva matriz de dispersión.
9. Realice la prueba de correlación lineal de Pearson. Comente los valores arrojados por los distintos coeficiente que arroja esta prueba.
10. ¿Qué nos indica el p-value?
11. Defina brevemente “nivel de significancia”.
12. ¿Cuáles son la hipótesis planteadas en esta prueba (Pearson)?
13. ¿Qué quiere decir “Subset selection”? ¿Qué importancia tiene en la regresión lineal múltiple?
14. Dígame una estrategia eficaz en el contexto de “Subset Selection”.
15. Defina “modelo anidado”.
16. ¿Qué es una iteración?
17. ¿Qué mide el error de entrenamiento?
18. ¿En qué consiste la estrategia “stepwise”? Implemente dicha estrategia en sus 3 vertientes.
19. Evalúe los modelos creados en base al Criterio de Información de Akaike.
20. ¿Cuándo ingresa una variable en el modelo? Justifique su respuesta.
21. Ejecute un modelo con las variables que posean una mayor significancia.
22. Calculad la varianza del modelo.
23. Calculad los intervalos de confianza de los coeficientes del modelo lineal previamente calculado. Interprete los resultados.
24. Visualice en un gráfico de dispersión 3D el modelo lineal.
25. Añada el plano de regresión lineal en un gráfico 3D. Interprete los resultados.
26. Valide la robustez del modelo lineal.
27. Verifique que se cumplen las condiciones del modelo de regresión lineal múltiple. Justifique su respuesta.
28. ¿Emplearía el Factor de Inflación de la Varianza para verificar alguno de las condiciones? Si es así diga cual.
29. Defina parsimonia.
30. Verifique la normalidad de los residuos. Dígame la diferencia entre residuos, residuos estandarizados y residuos estudentizados.
31. ¿Cómo interpretaría el test de Breusch-Pagan?
32. Verifique si existen o no observaciones influyentes.
33. Diferencia entre observación atípica y observación que produzca “apalancamiento” del modelo.
34. Calcule la distancia de Cook. ¿Cómo viene expresada? ¿Cómo la interpretaría?
35. ¿Qué valores *hat* superan la media?
36. ¿Cómo calcularía el límite *hat* que determina que observaciones son influyentes?
37. Elimine las observaciones influyentes del modelo y repetir todo el proceso comprobando el peso de esta observación en la regresión lineal. Comente brevemente los resultados.

Además, confeccione un documento (HTML, Word y/o PDF) en el cual se detallen los procedimientos computacionales llevados a cabo. Han de adjuntarse pantallazos de los procesos implementados comentados y justificados.