Modelos de Regresión Aplicados

Lab de práctica.

Esta práctica la van a realizar en clase. Se hace en grupos. Para obtener una calificación de 100, se tiene que entregar antes de las 11:50 del jueves 10 de noviembre. Si no terminan antes, pueden enviar la respuesta a [Gilbert.brenes.camacho@gmail.com](mailto:Gilbert.brenes.camacho@gmail.com), pero la nota base es de 90.

Toman el archivo antropometria.Rdata, con la submuestra del mes que le tocó a cada persona o grupo (Tienen que hacer un subset, tomando la variable id0a, que es numérica y en formato añomesdia).

Para ello, haga un subconjunto de datos. Por ejemplo,

datos <-antropometria[(anyomes==201012 | anyomes==201101),]

Si son un grupo, juntan los dos o tres meses que les corresponde.

Recuerden que el objetivo principal del análisis es estudiar qué variables antropométricas logran predecir el VEF, controlando además por edad, sexo y condición de fumador. Con el archivo de datos, tiene que estimar modelos con todas las variables y modelos reducidos (a partir de la selección de variables usando stepwise), con las medidas remediales solicitadas más abajo. Para el modelo reducido, primero lo escoge sin ninguna medida remedial, y después le aplica las medidas remediales. Para todos los modelos tiene que evaluar normalidad, homoscedasticidad y calcular el VIF más alto. Puede apoyar sus decisiones en gráficos, que le puede enseñar al profesor para justificar su decisión. Los resultados de los modelos se tienen que usar para llenar las tablas 1 y 2 que están más abajo. Los modelos con medidas remediales son:

1. Modelo con la variable dependiente transformada con una transformación de BoxCox.
2. Modelo con la transformación de la variable dependiente sugerida por el gráfico spread.level.plot().
3. Modelo heteroscedástico
4. Modelo con estimaciones robustas a presencia de valores extremos.
5. Modelo corrigiendo los errores estándar con mínimos cuadrados ponderados (use los predichos para las ponderaciones).

Al final, tiene que contestar cuál modelo escoge y por qué?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estudiante | Mes |  | Estudiante | Mes |
| AMANDA | 200411 |  | JOSTIN | 200605 |
| ANA PAULA | 200412 |  | JOSUE DAVID | 200606 |
| ANDREA | 200501 |  | JUAN JOSE | 200607 |
| BRANDON JOSUE | 200502 |  | KEYLA | 200608 |
| CESAR JOSUE | 200503 |  | LAURA VANESSA | 200609 |
| DANIEL ALFARO | 200504 |  | LUIS FERNANDO | 200411 |
| DANIEL SIBAJA | 200505 |  | MANRIQUE | 200412 |
| FELIPE | 200506 |  | MIRIAM | 200501 |
| IVAN DANIEL | 200507 |  | NATASHA | 200502 |
| JAMES | 200508 |  | OLGER LEONARDO | 200503 |
| JOSE PABLO | 200509 |  | PAMELA | 200504 |
| JOSEPH | 200510 |  | PATSY | 200505 |
|  | 200511 |  | YARLIN | 200506 |
|  | 200512 |  |  | 200507 |
|  | 200601 |  |  | 200508 |
|  | 200602 |  |  | 200509 |
|  | 200603 |  |  | 200510 |
|  | 200604 |  |  | 200511 |
|  |  |  |  |  |

Tabla 1. Modelo para predecir el VEF con todo el conjunto de variables predictoras en la base de datos, usando distintas medidas remediales. (En la columna prob solo pongan un asterisco si el coeficiente es significativamente distinto de cero).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Modelo sin corregir | | Mod. con transf. Box-Cox | | Mod. con transf spread.level.plot | | Mod. heteroscedástico | | Mod. estimac. robustas | | Mod. Mínimos Cuadrados Pond | |
| Variable | Coef | prob | Coef | prob | Coef | prob | Coef | prob | Coef | prob | Coef | prob |
| age |  |  |  |  |  |  |  |  |  |  |  |  |
| cinturar |  |  |  |  |  |  |  |  |  |  |  |  |
| caderar |  |  |  |  |  |  |  |  |  |  |  |  |
| grip |  |  |  |  |  |  |  |  |  |  |  |  |
| pantorrilla |  |  |  |  |  |  |  |  |  |  |  |  |
| brazo |  |  |  |  |  |  |  |  |  |  |  |  |
| tricipital |  |  |  |  |  |  |  |  |  |  |  |  |
| subscap |  |  |  |  |  |  |  |  |  |  |  |  |
| pesokg |  |  |  |  |  |  |  |  |  |  |  |  |
| tallacm |  |  |  |  |  |  |  |  |  |  |  |  |
| smoker |  |  |  |  |  |  |  |  |  |  |  |  |
| hombre |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| Prueba de normalidad |  |  |  |  |  |  |  |  |  |  |  |  |
| Prueba de Variancia constante (ncvTest) |  |  |  |  |  |  |  |  |  |  |  |  |
| R2 |  |  |  |  |  |  |  |  |  |  |  |  |
| AIC |  |  |  |  |  |  |  |  |  |  |  |  |
| P2 |  |  |  |  |  |  |  |  |  |  |  |  |

Nota: \*: p<0.05

Tabla 2. Modelo reducido para predecir el VEF con todo el conjunto de variables independientes sugeridas en la Tarea 2, usando distintas medidas remediales. (En la columna prob solo pongan un asterisco si el coeficiente es significativamente distinto de cero).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Modelo sin corregir | | Mod. con transf. Box-Cox | | Mod. con transf spread.level.plot | | Mod. heteroscedástico | | Mod. estimac. robustas | | Mod. Mínimos Cuadrados Pond | |
| Variable | Coef | prob | Coef | prob | Coef | prob | Coef | prob | Coef | prob | Coef | prob |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| Var1 |  |  |  |  |  |  |  |  |  |  |  |  |
| Var2 |  |  |  |  |  |  |  |  |  |  |  |  |
| Var3 |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| Prueba de normalidad |  |  |  |  |  |  |  |  |  |  |  |  |
| Prueba de Variancia constante (ncvTest) |  |  |  |  |  |  |  |  |  |  |  |  |
| R2 |  |  |  |  |  |  |  |  |  |  |  |  |
| AIC |  |  |  |  |  |  |  |  |  |  |  |  |
| P2 |  |  |  |  |  |  |  |  |  |  |  |  |

Nota: \*: p<0.05

Conteste: Cuál modelo escoge y por qué?