John Sebastián Marín Benítez

# Examen final

La variable más importante que explica el número de bicicletas alquiladas en un día determinado (cnt) es “casual” (Cantidad de usuarios casuales).

# Análisis

## Limpieza

Se elimina la variable “instant” puesto que simplemente es un consecutivo que representa a la observación.

Se convierte la variable “dteday” de tipo Factor a tipo Date.

## Multicolinealidad

Se ejecuta el análisis VIF y se obtiene la siguiente tabla:

|  |  |
| --- | --- |
| Variable | VIF |
| Dteday | 584,61 |
| Season | 4,11 |
| Yr | 447,15 |
| Mnth | 148,61 |
| Holiday | 1,1 |
| Weekday | 1,05 |
| Workingday | 3,11 |
| Weathersit | 1,93 |
| Temp | 63,82 |
| Atemp | 64,94 |
| Hum | 1,96 |
| Windspeed | 1,25 |
| Casual | 3,53 |
| Registered | 6,03 |

Tabla 1. Análisis VIF

Las variables dteday, season, yr, mnth, temp, atemp y registered son variables que tienen un VIF mayor a 4, por lo tanto presentan multicolinealidad.

Se eliminan las variables yr y mnth porque la misma información se puede obtener de dteday.

Se elimina la variable atemp porque la misma información se puede obtener de temp.

Se elimina la variable registered porque la misma información se puede obtener de casual.

Ecuación 1. Ecuación de registered

Se ejecuta el análisis VIF con el nuevo modelo y se obtiene la siguiente tabla:

|  |  |
| --- | --- |
| Variable | VIF |
| Dteday | 1,34 |
| Season | 1,41 |
| Holiday | 1,1 |
| Weekday | 1,04 |
| Workingday | 2,05 |
| Weathersit | 1,76 |
| Temp | 2,1 |
| Hum | 1,9 |
| Windspeed | 1,2 |
| Casual | 3,09 |

Tabla 2. Análisis VIF después de la corrección

Ya no hay más variables con un VIF mayor a 4, por lo tanto ya no hay un problema grave de multicolinealidad.

## Heteroscedasticidad

Primero se verifica el supuesto de normalidad de los errores para saber qué método es el apropiado. Aplicando varios test se obtiene la siguiente tabla:

|  |  |
| --- | --- |
| Test | p-valor |
| Shapiro-Wilk | ~ 0 |
| Kolmogorov-Smirnov | 0,0034 |
| Cramer-von Mises | 0,0012 |
| Anderson-Darling | ~ 0 |

Tabla 3. Análisis de normalidad de los errores

En todos los casos podemos rechazar que los errores sigan una distribución normal, por lo tanto se usará la versión studentizada del método Breusch-Pagan propuesta por Koenker para verificar heteroscedasticidad. Esta prueba da un p-valor de casi 0, por lo que rechazamos la homoscedasticidad.

## Autocorrelación

## Estandarización de coeficientes