

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.

$$registered = cnt - casual$$

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.

Usando la corrección de estimadores HC3 obtenemos la siguiente tabla:

Variable	p-valor
Intercept	~ 0 ***
Dteday	~ 0 ***
Season	0,03 *
Holiday	0,61
Weekday	0,15
Workingday	~ 0 ***
Weathersit	~ 0 ***
Temp	~ 0 ***
Hum	0,01 *
Windspeed	0,003 **
Casual	~ 0 ***

Tabla 4. Análisis de coeficientes con la corrección HC3

Con esta corrección se puede ver que las variables holiday y weekday no son significativas.

Se compara el modelo sin las dos variables con el modelo sin restringir mediante la prueba de Wald y se obtiene un p-valor de 0,28. Por lo tanto el modelo restringido es mejor.

Autocorrelación

Se ejecuta el análisis de Durbin – Watson para verificar si existe autocorrelación. Primero se verifica la hipótesis de no autocorrelación y se obtiene un p-valor de casi 0, por lo que se concluye que sí hay algún tipo de autocorrelación.

Después se verifica la hipótesis alternativa de autocorrelación negativa y se obtiene un p-valor de 1, por lo que no hay este tipo de autocorrelación.

Por último, se verifica la hipótesis alternativa de autocorrelación positiva y se obtiene un p-valor de casi 0. Por lo tanto, se concluye que sí se presenta autocorrelación positiva.

Usando las correcciones de Newey - West, Andrews y Lumley – Heagerty se puede concluir que las variables season y hum no son significativas.

Se compara el modelo sin las dos variables con el modelo sin restringir mediante la prueba de Wald con las tres correcciones y se concluye que el modelo restringido es mejor.

Estandarización de coeficientes

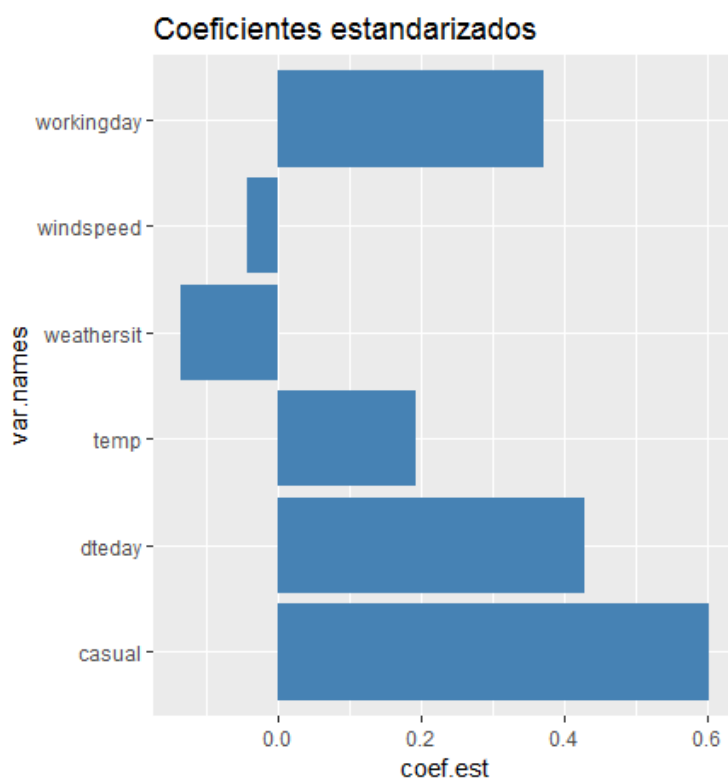


Ilustración 1. Coeficientes estandarizados.

Con esta gráfica se puede observar que la variable que más afecta es “casual”, es decir la cantidad de usuarios casuales en un día.