

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Encicla versión gringa

Juliana Zuluaga Serna
Laura Tatiana Riveros Quintero

Universidad Nacional de Colombia
juzuluagase@unal.edu.co, ltriverosq@udea.edu.co

1. Capital Bikeshare

Este póster presenta los resultados de la encuesta del uso Capital Bikeshare de noviembre de 2011 y 2012 realizada para el programa Capital Bikeshare, un programa propiedad y patrocinado conjuntamente por el Distrito de Columbia, el Condado de Arlington, y la Ciudad de Alexandria, operado por Alta Bicycle Share, Inc., ofrece el uso a corto plazo de más de 1,650 bicicletas registradas. miembros y usuarios de pase diario en más de 175 estaciones. Capital Bikeshare permite desbloquear una bicicleta en cualquier estación. Los usuarios pueden devolver la bicicleta a la misma estación o a cualquier otra estación en la red. Hoy en día, existe un gran interés en estos sistemas debido a su importante papel en el tráfico, el medio ambiente y la salud. Además de las interesantes aplicaciones del mundo real de los sistemas para compartir bicicletas, las características de los datos que generan estos sistemas los hacen atractivos para la investigación.

2. Introduction

En este trabajo se busca analizar los datos de Capital Bikeshare con el objetivo de estudiar cómo influyen algunas características sobre el recuento total de alquiler de bicicletas. Variables a estudiar:

1. Season: Primavera, Verano, Otoño, Invierno
2. mnth: 1 a 12 meses
3. weekday: 0 a 6 días
4. workingday: 1: jornada laboral, 0: de lo contrario
5. weathersit: 1: Agradable, 2: Nublado, 3: Mojado
6. temp: temperatura normalizada en Celsius
7. atemp: temperatura de sensación normalizada en Celsius
8. hum: humedad normalizada
9. windspeed: velocidad del viento normalizada
10. cnt: recuento de bicicletas de alquiler totales

La base de datos original contiene 731 observaciones con 16 variables, con 8 cualitativas y 7 cuantitativas para los años 2011 y 2012. De esta base se extraerá otra base donde solo se trabajará con el año 2012, con esto nos quedan 366 observaciones.

3. Analisis de la base de datos

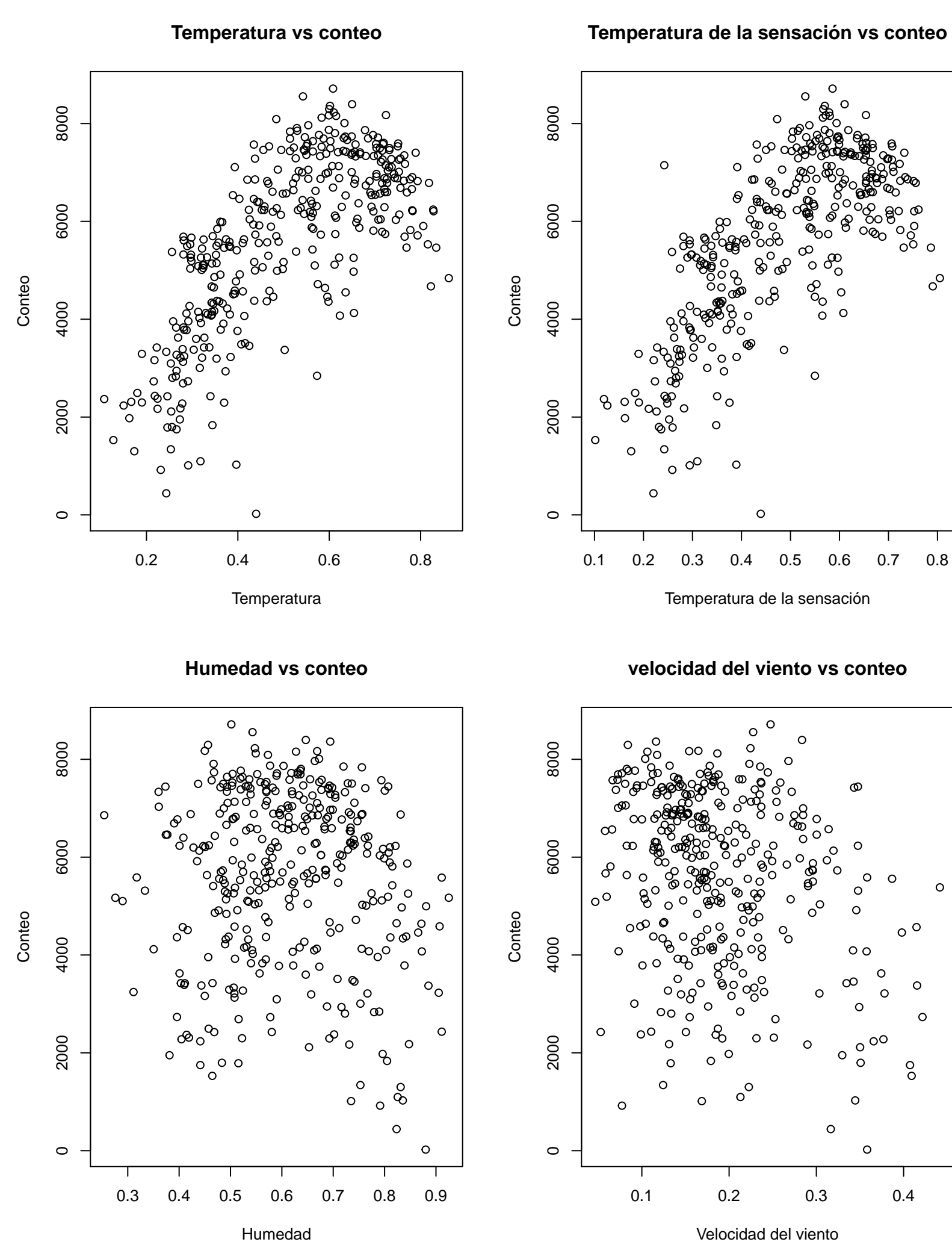


Figura 1: Gráficos de dispersión de variables cuantitativas

Weathersit	Temp mean	Cnt
1	0.523	1422885
2	0.472	613934
3	0.415	12757

Cuadro 1: Clima, media temperatura y conteo

Season	Temp mean	Cnt
1	0.322	321248
2	0.555	571273
3	0.711	641479
4	0.419	515476

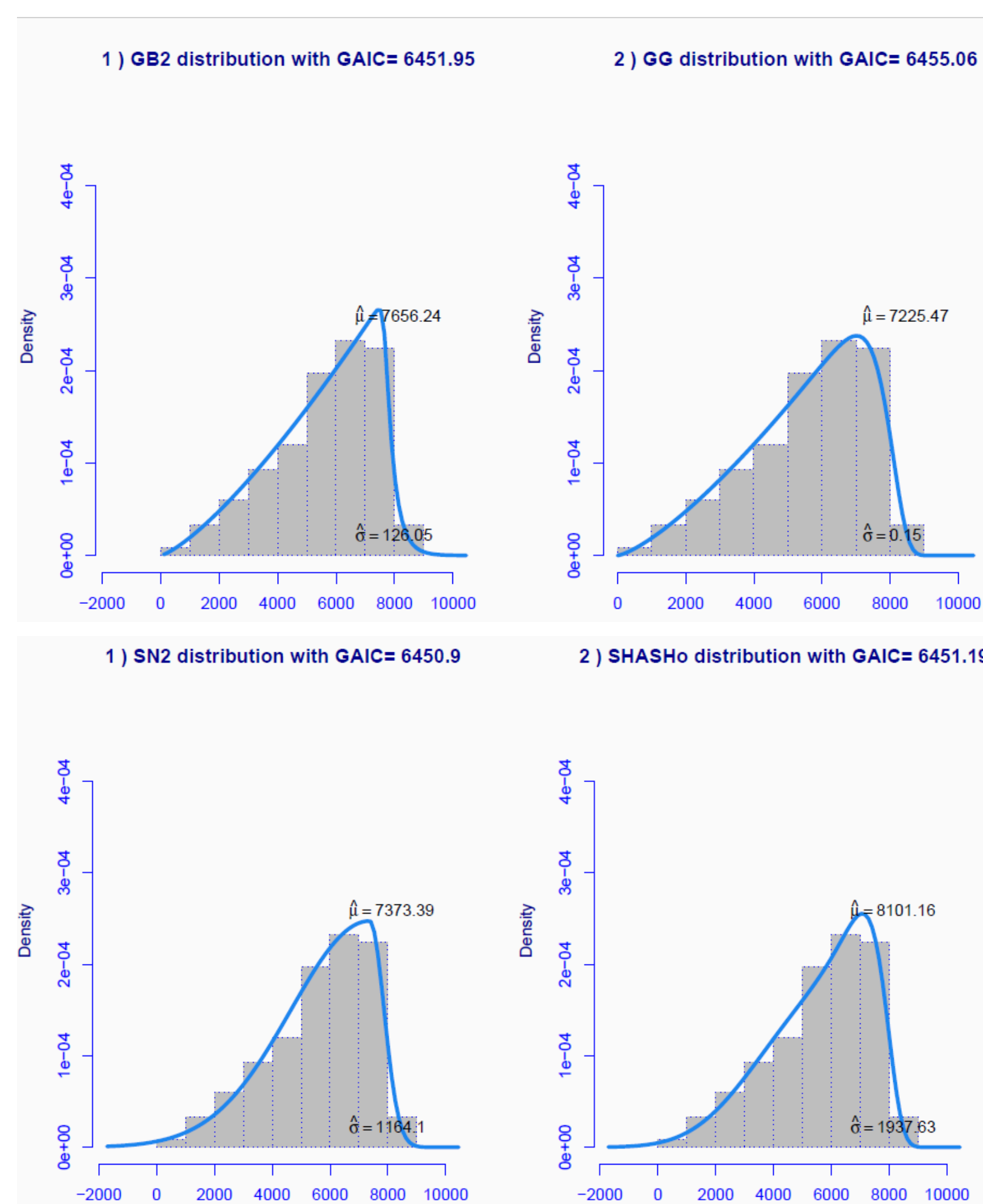
Cuadro 2: Estación, media temperatura y conteo

Se realizó un análisis exhaustivo de las variables meses y días del año con gráficos vs el conteo, la conclusión de la gráfica meses comparada con la gráfica de SvsC nos arrojó curvaturas similares para ambos, por esto se decidió trabajar solo con la variable estación. Para la variable días por semana el gráfico nos arrojó una constancia en los días y con un análisis resultó no significativa con una confianza del 90 %. Se realizó un análisis de multicolinealidad y basándonos en el Variance Inflation Factor (VIF), Debido a esto se decide tomar solo la variable temperatura. Se muestra el coeficiente de correlación de Pearson para el conteo de alquiler de bicicletas

Var	VIF
Temp	37.4603
Atemp	38.0145

Var	Corr
Temp	0.714
Hum	-0.089
Windspeed	-0.279

4. Grafico de las mejores Distribuciones para la variable respuesta



Se eligieron estas cuatro distribuciones para el análisis de los modelos, con estas distribuciones se analizaron 37 modelos variando los parámetros de los cuales se eligieron los mejores con el criterio AIC.

5. Resultados

MODELO	DIS	AIC	COR	R ²	N.VAR	N.VAR SIG
Modelo 1	SHASHo	5869.8	0.884	0.80	8	8
Modelo 2	GB2	5951.5	0.900	0.76	12	10
Modelo 3	GG	6000.6	0.881	0.72	11	10
Modelo 4	SN2	5878.9	0.907	0.80	12	9
Modelo 5	SHASHo	5836.8	0.904	0.82	12	9

Los modelos mencionados en la tabla no se muestran en el póster debido al exceso de variables y betas estimados, lo cual generaba saturación de texto. Se obtuvo como modelo final el modelo 1

$$\hat{C} = 4624,12 - 50383,27T^3 + 53417,85T^2 - 6998,23T + 730,22S^2 + 1240,10S^3 + 1555,22S^4 - 3053,54H - 3231,38W.$$

Donde T= Temperatura, S= Estación, H= Humedad, W= Velocidad del viento

Concluimos al hacer la selección de variables que la iteración de temp*season no mejoraba el modelo y resultó no significativa eliminando nuestra hipótesis del comienzo de que podía influenciar. Así mismo la iteración de la temp*weathersit fue significativa para el clima 3 (Mod5) pero al hacer el análisis de los modelos con y sin esa variable nos mostró que no variaba mucho el AIC (Mod1) ni las otras pruebas que se hicieron, así por menor cantidad de parámetros se eligió el modelo sin esta variable.

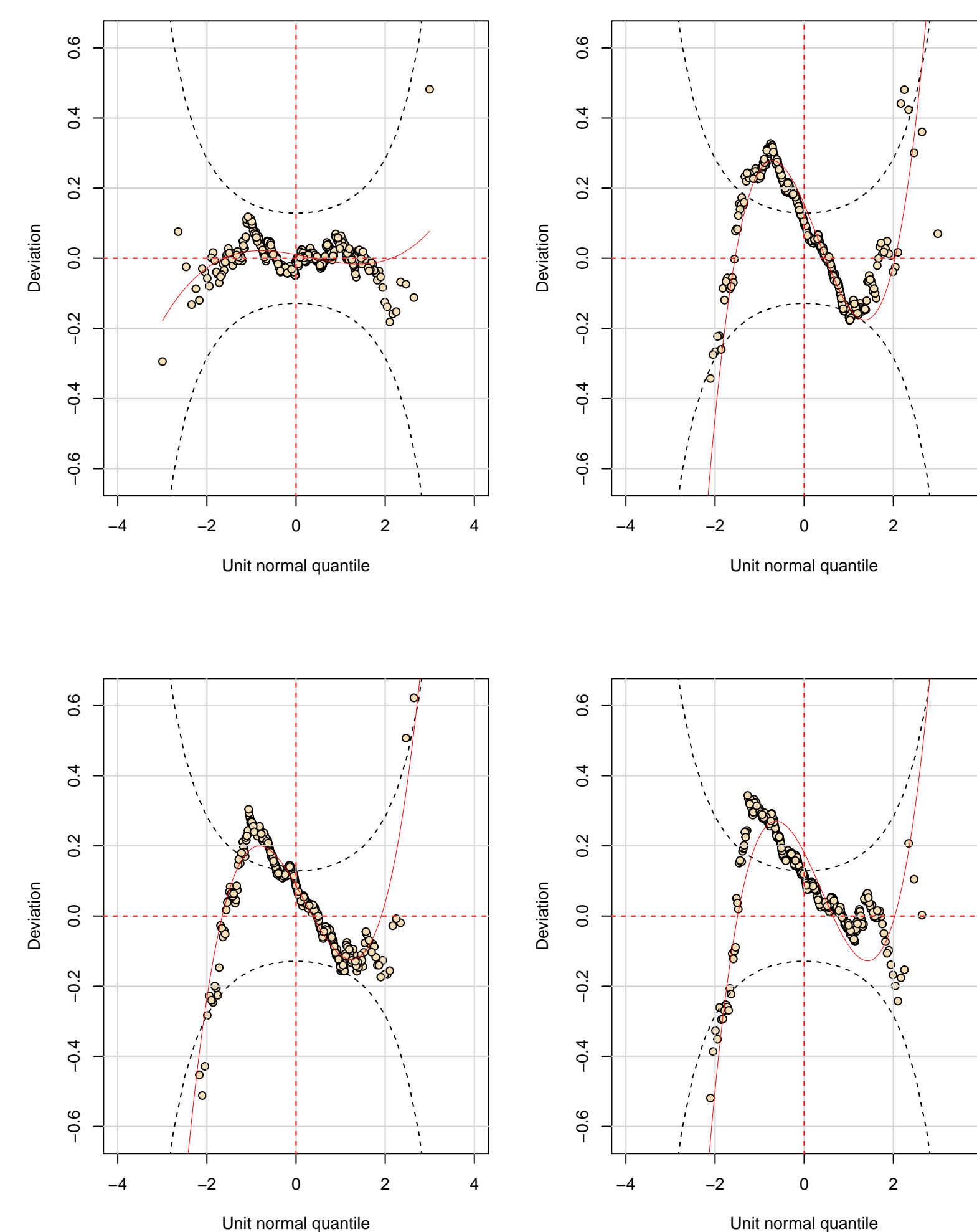


Figura 3: Grafico de gusano para los mejores modelos

Gráfico 1- dist SHASHo, 2- dist GG, 3- dist SN2, 4- dist GB2. Con este gráfico se puede ver claramente que el modelo SHASHo es el que mejor ajusta debido los puntos no sobresalen las bandas de confianza y se acerca mas a la recta horizontal.

6. Conclusiones

Para el conteo de alquiler de bicicletas en Capital Bikeshare fue influenciada en gran parte por la temperatura, la estación del año, la humedad y la velocidad del viento. Indicándonos que a menor temperatura menor número de alquiler de bicicletas, que en las estaciones del año Otoño fue de las más significativas para el conteo de alquiler de bicicletas confirmando las tablas analizadas donde esta estación es la de mayor número de bicicletas que se alquilaron.

Referencias

Base de datos: 'https://data.world/uci/bike-sharing-dataset'
Programa para trabajar los codigos RStudio

Mikis Stasinopoulos, Bob Rigby and Calliope Akantziliotou (January 11, 2008) Instructions on how to use the gamlss package in R Second Edition A.3.5 Sinh-Arcsinh (SHASH) página 186

Para conocer más sobre los coeficientes de los modelos y/o el código utilizado para el análisis de datos, ingresar a: <https://github.com/julianazuluaga487/codigos/blob/master/posterbici.R>