

1. Resumen

Se empleo metodología lme4 para realizar el problema propuesto por la empresa Tienda Registrada S.A.S la cual es una empresa que logra generar información de manera oportuna y con la base estadística adecuada según se va consolidando en las principales ciudades de Colombia; se planteo una opción para explicar el numero de registros mediante variables especificas de una tienda de barrio las cuales eran en mayor parte cualitativas. Encontrando con esta metodología una mejor distribución y adecuación para la variable respuesta, los resultados fueron obtenidos empleando el programa estadístico R

2. Modelos GAMLSS

El paquete lme4 (Bates, Maechler, Bolker y Walker 2014a) para R (R Core Team 2015) proporciona funciones para ajustar y analizar modelos mixtos lineales, modelos mixtos lineales generalizados y modelos mixtos no lineales. En cada uno de estos nombres, el término "mixto" o, más completamente, "efectos mixtos" denota un modelo que incorpora términos de efectos fijos y aleatorios en una expresión predictiva lineal a partir de la cual se puede evaluar la media condicional de la respuesta.

3. Introducción

Tienda registrada capta la información del comportamiento de los productos de consumo masivo dentro de la tienda de barrio a través del código de barras. Se creo un modelo que explica en numero de registros(R) de una tienda del valle de aburra en función de las covariables: (I)id, (M)mes, (U)ubicación, (S)sector, (D)domicilio, (N)numero de personas, (C)categoría, (E)estrato, (T)tamaño medio, (MS)municipio y (TS)tiempo de servicio de las tiendas . Construir un modelo de regresión para tratar de estimar cual es la cantidad de registros da cada tienda. Este modelo se puede utilizar para obtener un indicador de subregistro, para validar la información que llega de las tiendas y así poder analizar que tiendas dejan de registrar en comparación de las otras.

4. Análisis descriptivo

R	I	M	U	S	D	N	C	E	T	MU	TS
25245	18	1	CABECERA DE BARRIO	Residencial	NO	2	TIENDA MOSTRADOR	3	45	Medellin	13
21883	18	2	CABECERA DE BARRIO	Residencial	NO	2	TIENDA MOSTRADOR	3	45	Medellin	13
47698	18	3	CABECERA DE BARRIO	Residencial	NO	2	TIENDA MOSTRADOR	3	45	Medellin	13
22462	18	4	CABECERA DE BARRIO	Residencial	NO	2	TIENDA MOSTRADOR	3	45	Medellin	13
24433	18	5	CABECERA DE BARRIO	Residencial	NO	2	TIENDA MOSTRADOR	3	45	Medellin	13
23610	18	6	CABECERA DE BARRIO	Residencial	NO	2	TIENDA MOSTRADOR	3	45	Medellin	13

Tabla 1: Datos del artículo

Se presenta la base de datos del experimento, en la cual se aprecian las primeras 6 observaciones.

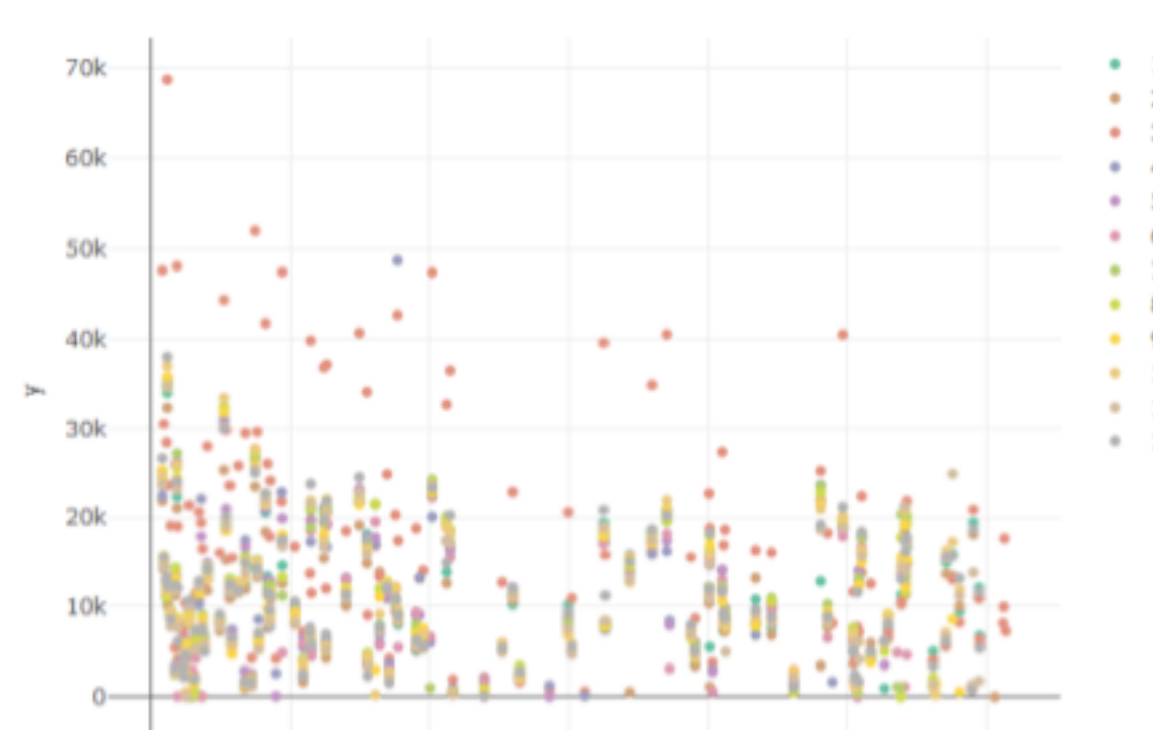


Figura 1: gráfico de dispersión Id vs numero de registros, agrupado por meses

Figure 2: gráfico de dispersión Id vs numero de registros, agrupado por meses

En la imagen anterior podemos ver como la variable Id que se encuentran en el eje horizontal

vs el numero de registros en el eje vertical están agrupadas de acuerdo a los meses que es el color que identifica a cada punto, lo que nos indica que se debe de plantear un modelo con efectos mixtos.

5. Metodología

Se buscó usar metodología lme4, en donde se plantearon los mejores modelos respecto a este tipo de método, que pudieran modelar la variable respuesta en nuestro caso el número de registro de una tienda (R). Dentro del desarrollo del trabajo se propusieron diferentes tipo de modelo entre ellos lm, transformaciones, splines, GAMLSS y lmer. Para la selección de los mejores modelos se utilizaron los criterios: Normalidad de los errores, AIC, Correlación con la variables respuesta. En busca del mejor modelo posible, se realizo un proceso de selección de variables con la función Step.AICALL.A(), pero el modelo final se eligió mediante un proceso en donde solo se tuvo en cuenta los anteriores criterios.

Modelos	Metodologia	Correlacion	AIC	R^2
mod	GAMLSS	0.46	25797	0.2576
mod1	GAMLSS	0.47	25241	0.5485
mod2	GAMLSS	0.69	24886	0.6717
mod3	lmer	0.62		

Tabla 2: Modelos propuestos

6. Resultados

Después de analizar los modelos con los criterios de AIC, correlación y normalidad de los errores, aunque en el ultimo modelo al usar metodología lmer no se obtuviera el AIC y el R^2 , descartamos los modelos que tenían la correlación mas baja o el modelo que no se podía interpretar y por ultimo elegimos el modelo con la mejor capacidad de explicar la variable respuesta en términos de correlación.

Scaled residuals:

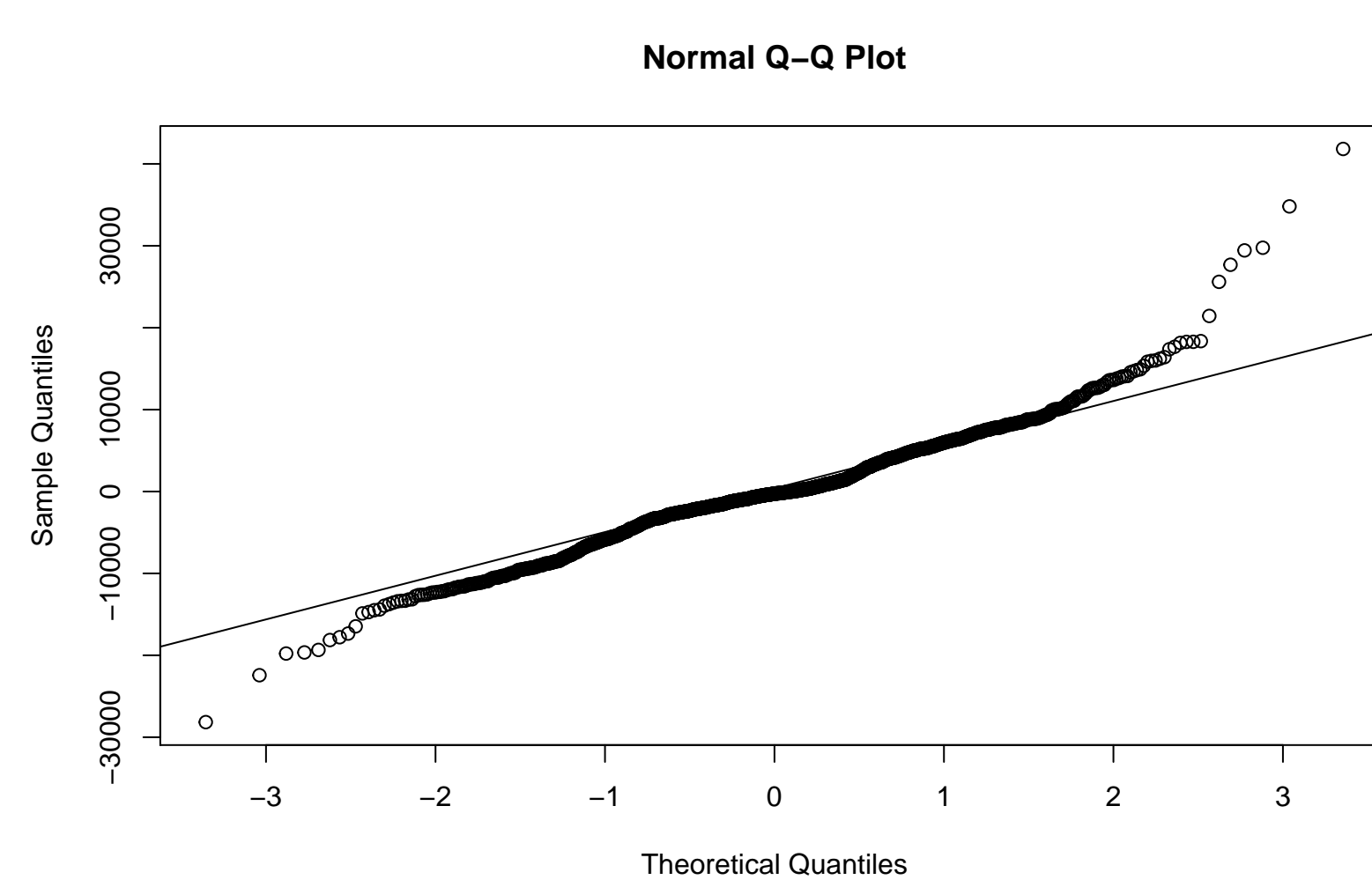
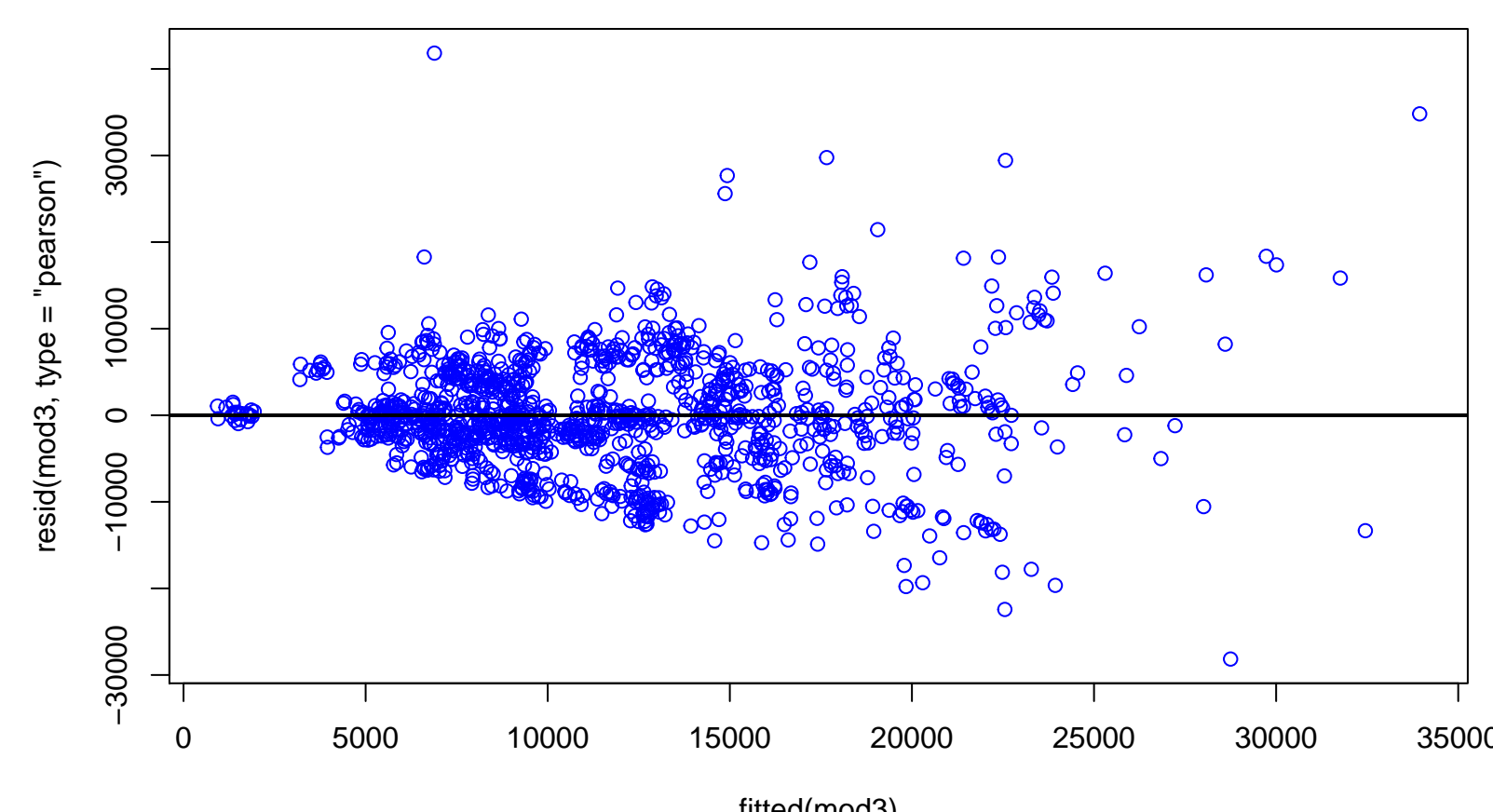
Min	1Q	Median	3Q	Max
-4.3084	-0.4837	-0.0361	0.6053	6.3320

Random effects:

Groups	Name	Varianza	Std.Dev	Corr
M	(Intercept)	1.795e+07	4236.735	
	I	7.669e+07	2.769	-1.00
Residual		4.3663e+07	6605.548	

Tabla 3: Resumen de los datos

En la tabla anterior podemos observar como los residuales están agrupados alrededor del cero , el Min con un resultado de -4.3084 y en el Max 6.3320 lo cual nos indica que hay una sobre estimación , también podemos observar la varianza de (M) y de (I) que son números muy bajos.



En la gráfica anterior podemos observar como los residuales se encuentran alrededor de la linea cero y un breve patrón.

En la gráfica de normalidad podemos ver que en los extremos los puntos comienzan a dispersarse un poco pero no lo suficiente como para concluir que no existe normalidad.



Figure 6: Trazos de oruga de efectos aleatorios

En esta gráfica vemos los efectos aleatorios del intercepto y de la pendiente, luego este tambien nos dice como fueron ajustados los grupos, por medio de intervalos de confianza.

7. Conclusiones

Se obtiene como resultado un modelo con cual se puede predecir el numero de registros de una tienda de acuerdo a una correlación de 0.62 lo cual nos indica que es un modelo que puede ser confiable ya que la variables respuesta es cuantitativa y la mayoría de las covariables con cualitativas lo que genera entre ellas poca correlación.

También notamos como el paquete lme es mas fuerte que GAMLSS , ya que comparando el mismo modelo con estos dos paquetes obtuvimos una mejor correlación con el lme.

Referencias

Bates,D Walker,S;Fitting Linear Mixed-Effects Models Using lme4 (2015). <https://cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf>
A Visual Introduction to Hierarchical Models Mfviz.com <http://mfviz.com/hierarchical-models/>
Cran.r-project.org <https://cran.r-project.org/web/packages/lme4/vignettes/>



Escanea el codigo y para ver los codigos en R