



Análisis de ventas mensuales mediante Modelos Lineales Mixtos

Integrantes: Jherson Guzman Ramirez, Santiago Florez Carranza

.....

Resumen

En este proyecto se realiza un análisis de las ventas mensuales de una cadena de supermercados utilizando modelos lineales mixtos (MLM) aplicados a un conjunto de datos longitudinales para capturar la variabilidad intra como interhogar en el comportamiento de compra. Los MLM permiten incorporar *efectos fijos*, como el mes (o temporada) de compra, el ingreso monetario, edad promedio y tamaño del hogar, entre otras variables demográficas que son comunes en todos los individuos de la población, por otro lado los *efectos aleatorios* capturan la variabilidad asociada a cada hogar, reflejando las diferencias individuales en los patrones de compra. Los resultados muestran la importancia de considerar la estructura jerárquica de los datos usando MLM mejora significativamente el ajuste del modelo y permite identificar patrones clave en las ventas, los cuales pueden ser utilizados en modelos de marketing u optimización de ventas.

Abstract

This study conducts an analysis of the monthly sales of a supermarket chain using linear mixed models (LMM) applied to a longitudinal dataset to capture both intra-household and inter-household variability in purchasing behavior. LMMs allow for the incorporation of *fixed effects*, such as the month (or season) of purchase, monetary income, and household size, among other demographic variables. Conversely, *random effects* capture the variability associated with each household, reflecting individual differences in purchasing patterns. The results demonstrate that considering the hierarchical structure of the data using LMM significantly improves model fit and enables the identification of key sales patterns, which can be utilized in marketing models and sales optimization strategies

Introducción

El análisis de ventas es fundamental para las estrategias comerciales y de marketing de las empresas. Comprender cómo varían las ventas a lo largo del tiempo y entre diferentes hogares puede proporcionar información valiosa para la toma de decisiones. Los datos longitudinales ofrecen la posibilidad de analizar estas variaciones considerando la dependencia temporal y la heterogeneidad entre unidades de observación.

Los MLM son una herramienta estadística poderosa para analizar datos longitudinales y jerárquicos. Permiten modelar efectos fijos, comunes a todos los individuos, y efectos aleatorios, que capturan la variabilidad individual o grupal. En este contexto, es posible identificar patrones generales en las ventas mensuales y, al mismo tiempo, considerar las diferencias específicas de cada hogar.

Palabras claves: datos longitudinales, modelos lineales mixtos, partial pooling, efectos aleatorios

Marco teórico

El análisis de datos longitudinales mediante medidas repetidas es crucial en diversas disciplinas, como la psicología, la medicina y las ciencias sociales, al permitir examinar la evolución de variables a lo largo del tiempo en sujetos individuales. Según Bates, [1] los MLM son especialmente útiles en este contexto, ya que capturan la variabilidad tanto intra como intersujeto.

Un hallazgo significativo de estudios en este ámbito es la notable variabilidad en las respuestas entre sujetos. En un experimento que evaluó el impacto de la privación del sueño en el rendimiento cognitivo, se observó que los interceptos y pendientes variaban significativamente entre los participantes, lo que resalta la importancia de incorporar efectos aleatorios para representar adecuadamente esta heterogeneidad. Aunque una regresión lineal simple puede ser adecuada en ciertos casos, no capta la complejidad de los datos, especialmente ante patrones no lineales o inconsistentes. El uso de MLM permite modelar interacciones y variaciones en las respuestas que podrían pasarse por alto.

Mahr,[2] basándose en el trabajo de Bates [1] también analizó este conjunto de datos y observó que permitir que las estimaciones individuales se ajusten a promedios poblacionales aumenta la robustez de los parámetros estimados. Este fenómeno, conocido como "shrinkage", es particularmente relevante cuando hay pocas observaciones por sujeto, ya que las estimaciones de efectos aleatorios se ven influenciadas por las medias de la población, mejorando así la validez de las inferencias.

De este modo, se concluye que los modelos lineales mixtos no solo aportan flexibilidad en la representación de datos longitudinales, sino que son fundamentales para abordar la variabilidad inherente en las respuestas de diferentes sujetos. Esta metodología es aplicable a una amplia gama de áreas de investigación donde las medidas repetidas son una característica clave.

Metodología

Los datos utilizados en este estudio provienen del paquete *completejourney* [3] de R, el cual proporciona un conjunto amplio de transacciones hechas por hogares que compran frecuentemente en supermercados a lo largo de un año. Este conjunto de datos incluye todas las compras de cada hogar, sin limitarse a ciertas categorías de productos. Además, para algunos hogares, se cuenta con información demográfica y el historial de contacto de campañas de marketing directo. A continuación, podemos observar el diagrama relacional del dataset.

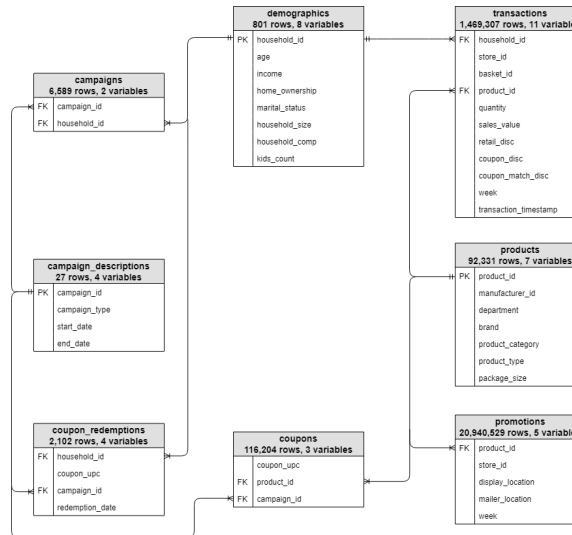


Figura 1: Diagrama relacional dataset *completejourney*

Para el análisis, se utilizaron dos de las tablas del modelo relacional: **transacciones** y variables **demográficas**, las cuales fueron integradas mediante las funciones `left_join()` e `inner_join()` del paquete `dplyr`, permitiendo consolidar la información necesaria para el análisis posterior. Este proceso permite trabajar con una cantidad total de hogares de **801 unidades**, en los cuales se puede observar el **valor de las compras** realizadas mes a mes.

Se realiza un breve **análisis descriptivo y exploratorio** del conjunto de datos.

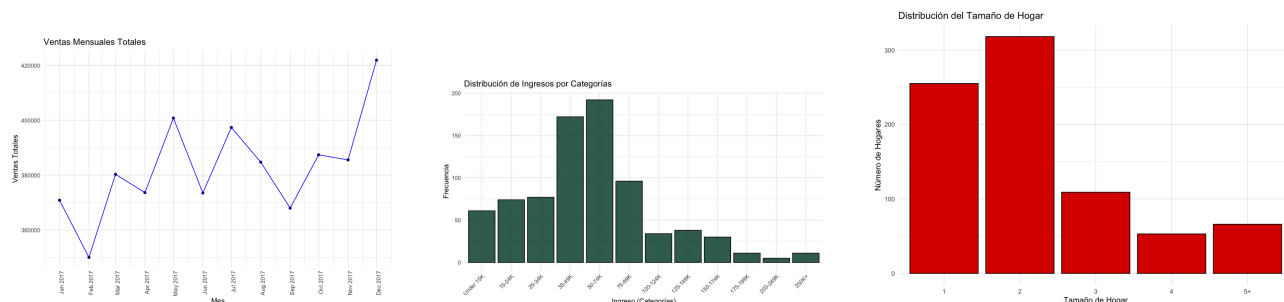


Figura 2: Ventas totales por mes

Figura 3: Distribución de ingresos

Figura 4: Distribución tamaño hogar

De las imágenes anteriores, se observa un patrón o tendencia de ventas significativamente mayor durante el mes de diciembre, lo que probablemente se debe a las festividades de fin de año, un periodo caracterizado por un aumento en el consumo. En contraste, el mes de febrero muestra las ventas más bajas, posiblemente debido a la reducción del gasto tras las festividades. Estos factores deben considerarse al planificar estrategias de marketing, permitiendo ajustar campañas promocionales y recursos según los picos y caídas estacionales del comportamiento de compra.

La distribución de ingresos revela que la mayoría de los hogares que realizan compras pertenecen a las clases baja y media, dado que la distribución está sesgada a la derecha, indicando una mayor concentración de hogares con ingresos menores. En cuanto al tamaño de los hogares, predominan aquellos conformados por 1 o 2 personas. Este análisis es crucial para la toma de decisiones comerciales, ya que permite ajustar tanto el tipo de productos ofrecidos como sus presentaciones, optimizando el tamaño y la cantidad de acuerdo con las características y necesidades de estos hogares.

Ahora, con base en la literatura sobre MLM y análisis de datos longitudinales, se desarrolla un primer modelo en el que se establece que el valor medio de las compras realizadas depende de la variable "mes" como efecto fijo, mientras que se incorporan los identificadores de los hogares como efectos aleatorios.

A partir de este momento, se lleva a cabo una evaluación exhaustiva de la importancia y la multicolinealidad de diversas variables demográficas, utilizando los valores de inflación de varianza (VIF) para identificar posibles colinealidades, en el contexto de los efectos fijos significativos dentro del modelo. Además, se analiza la influencia de los efectos aleatorios, no solo en los interceptos, sino también en las pendientes asociadas a cada hogar.

De este modo, se desarrollan múltiples modelos en los cuales se realiza una validación de supuestos sobre los residuos marginales, definidos como:

$$\epsilon^* = Y - X\hat{\beta}$$

Estos residuos se utilizan para verificar los supuestos relacionados con los efectos fijos del modelo, tales como la normalidad y la homocedasticidad.

Residuos Condicionales: Los residuos condicionales permiten evaluar tanto los efectos fijos como los aleatorios del modelo. Su análisis es fundamental para identificar observaciones atípicas dentro de los clústeres y verificar la constancia de la varianza y la normalidad de los errores. Se definen como:

$$\hat{\varepsilon} = Y - X\hat{\beta} - Z\hat{\gamma}$$

Residuos BLUP (Best Linear Unbiased Prediction): Los residuos BLUP son utilizados para comprobar la normalidad y detectar posibles grupos atípicos en los efectos aleatorios. Se obtienen restando la predicción de los efectos aleatorios de la media ajustada, de la siguiente manera:

$$Z\hat{\gamma} = \mathbb{E}[Y|\gamma] - \mathbb{E}[Y]$$

Para evaluar el rendimiento de los modelos, se aplica validación cruzada, utilizando un 80 % de los datos para entrenamiento y un 20 % para prueba. A través de este proceso, se calculan métricas de ajuste, tales como el logaritmo de la verosimilitud (logLik), el criterio de información de Akaike (AIC), el criterio de información bayesiano (BIC) y el criterio REML (Restricted Maximum Likelihood), que permitieron seleccionar el modelo más adecuado.

Resultados

A continuación se presenta la tabla de resultados para los modelos desarrollados, con el objetivo de comparar las métricas de ajuste obtenidas a partir de distintas especificaciones. Cada modelo considera diferentes combinaciones de efectos fijos y aleatorios, permitiendo así evaluar la influencia de variables como el mes, el ingreso y el tamaño del hogar en las ventas, así como la capacidad predictiva de los hogares como efectos aleatorios.

Modelo	#. obs	σ	logLik	AIC	BIC	REMLcrit	df.res	
Modelo 1 - Ventas por mes ~ Mes Fijo / Hogares Aleatorios	19504	105.76	-122112.74	244253.48	244363.78	244225.48	19490	
Modelo 2 - log(Ventas por mes) ~ Mes Fijo / Hogares Aleatorios	19504	0.78	-25941.81	51911.63	52021.92	51883.63	19490	
Modelo 3 - log(Ventas por mes) ~ Mes, Ingreso, Tamaño Hogar Fijos / Hogares Aleatorios	7489	0.62	-8017.03	16092.06	16292.78	16034.06	7460	*
Modelo 4 - Ventas por mes ~ Mes, Ingreso, Tamaño Hogar Fijos / Hogares Aleatorios	7489	124.08	-47793.37	95644.74	95845.45	95586.74	7460	

Cuadro 1: - Comparación de modelos

En los modelos que consideran la variable de ventas transformada mediante logaritmo (Modelos 2 y 3), la transformación logarítmica se aplicó con el fin de estabilizar la varianza y mejorar la normalidad de los residuos, lo cual es fundamental para cumplir con los supuestos inherentes a MLM. En particular, esta transformación ha mostrado ser efectiva en la mejora de la calidad de ajuste en términos de las métricas de selección de modelos como AIC, BIC y logLik.

- Modelo seleccionado

$$\text{log_valor_ventas}_{ij} = \beta_0 + \beta_1 \cdot \text{ingreso}_i + \beta_2 \cdot \text{mes}_j + \beta_3 \cdot \text{tamaño_hogar}_i + b_{0i} + \varepsilon_{ij}$$

siendo:

- **log_valor_ventas_{ij}**: Variable de respuesta logarítmica de las ventas totales para el hogar i en el mes j
- **β_0 (intercepto)**: Valor esperado de log_valor_ventas_{ij} para un hogar con ingreso promedio y tamaño de hogar promedio en el mes de referencia (enero).

- $\beta_1 \cdot \text{ingreso}_i$: Efecto Fijo del ingreso monetario del hogar i sobre las ventas logarítmicas.
- $\beta_2 \cdot \text{mes}_j$: Efecto Fijo del Mes, captura las variaciones mensuales en las ventas relativo al mes de referencia (enero).
- $\beta_3 \cdot \text{tamaño_hogar}_j$: Efecto Fijo del Tamaño del Hogar i sobre las ventas logarítmicas.
- b_{0i} : Efecto Aleatorio del Hogar i , captura las desviaciones específicas de cada hogar i respecto al intercepto fijo β_0
- ε_{ij} : Término de Error Aleatorio: Representa la variabilidad residual no explicada por el modelo.

Este enfoque proporciona una adecuada representación de la estructura jerárquica de los datos, permitiendo modelar la heterogeneidad existente entre los hogares y capturar tanto las variaciones específicas de los sujetos (hogares) como las generales del conjunto de datos.

- Validaciones

Grafico de efectos aleatorios BLUP

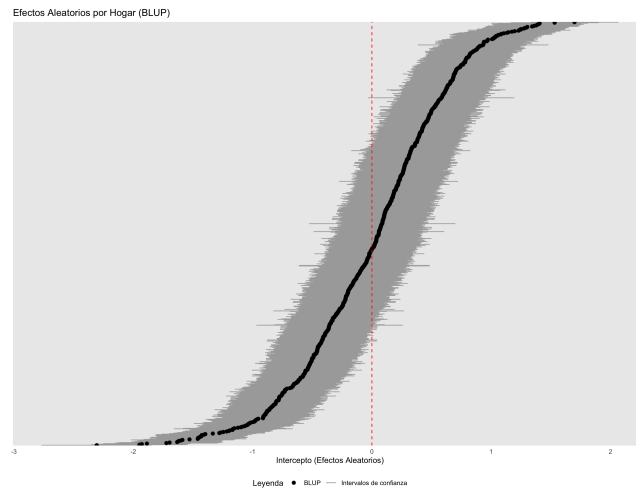


Figura 5: Efectos aleatorios (BLUPs) por hogar con intervalos de confianza

Podemos observar que la mayoría de los efectos aleatorios están distribuidos alrededor de 0, lo cual es un comportamiento esperado si no existe un sesgo considerable en la estructura aleatoria de los hogares. Esto indica que, en promedio, las desviaciones aleatorias de las ventas entre los hogares no son significativamente grandes. Además, al analizar los intervalos de confianza, se observa que prácticamente todos los hogares están dentro de estos intervalos, lo que refuerza la idea de que las estimaciones son confiables y bien ajustadas, lo que proporciona solidez al modelo.

Residuos

A continuación, se presentan dos gráficos que ilustran los residuales condicionales y marginales, respectivamente. En el primer gráfico se muestran los *residuales condicionales* frente a los valores ajustados que incluyen tanto los efectos fijos, como los aleatorios en el modelo mixto. Se puede ver que la mayor parte de los puntos se concentran alrededor de la línea de cero, en general, el modelo está capturando las tendencias generales de los datos. Aun así, el ensanchamiento de los puntos hacia la derecha (para valores ajustados más altos) indica que el modelo tiene una peor capacidad predictiva para ciertos valores, además de ser un indicador de heterocedasticidad.

En cuanto al segundo gráfico se ven los *residuales marginales* frente a los valores ajustados de los efectos fijos, se pueden observar una cantidad de residuales lejos de la línea de cero mostrando que hay observaciones mal ajustadas por el modelo, además de la posibilidad de encontrar alta dispersión en los residuales, lo cual lleva a un aumento en la variabilidad de los residuales a medida que los valores ajustados aumentan. Por lo tanto se puede sospechar que estamos bajo heterocedasticidad.

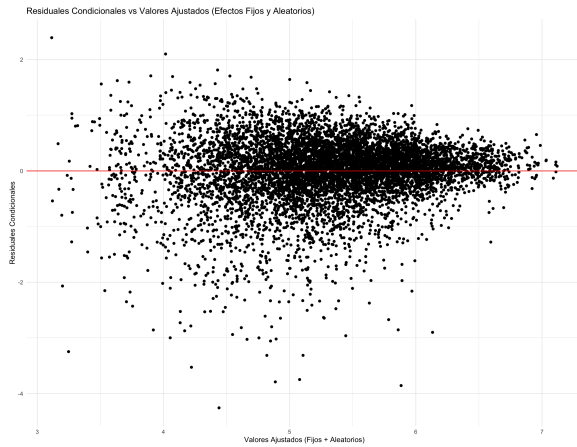


Figura 6: Residuos condicionales

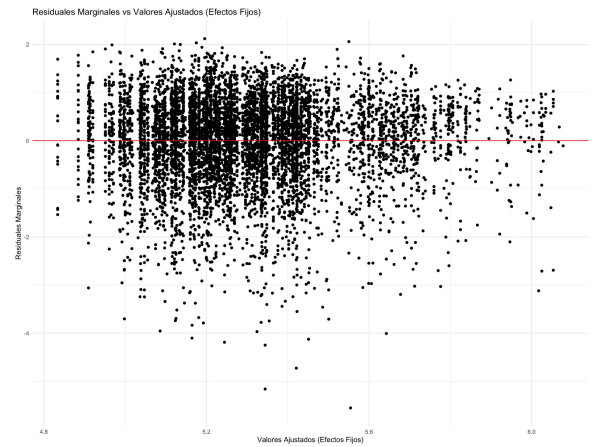


Figura 7: Residuos marginales

A continuación realizamos un test de Breusch-Pagan, adaptándolo a nuestro caso de modelo mixto lineal del cual obtenemos:

```
data: residuos_cond ~ valores_ajustados + I(valores_ajustados^2)
BP = 374.56, df = 2, p-value < 2.2e-16
```

lo cual es un resultado altamente significativo que nos permite rechazar la hipótesis nula H_0 de homocedasticidad, así concluimos heterocedasticidad en el modelo.

Por lo tanto podemos realizar una revisión de nuestro modelo, tratando de ajustar un modelo robusto mediante el paquete **robustlmm**, de esta manera nuestra especificación de modelo sería la misma, pero con la función **robustlmm::rlmer()**.

Observamos de nuevo las gráficas, viendo que no hay una mejoría sustancial en nuestros residuos aplicando este modelo lineal mixto robusto

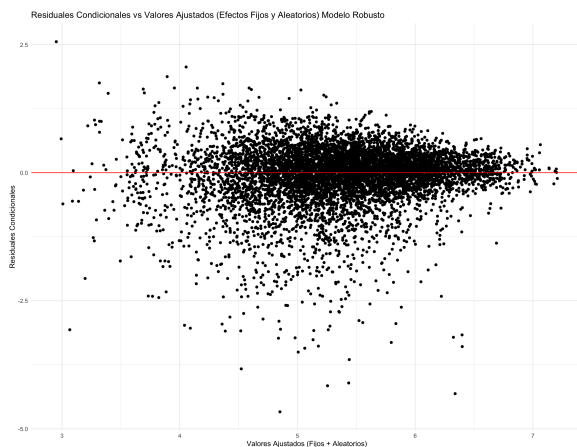


Figura 8: Residuos condicionales modelo robusto

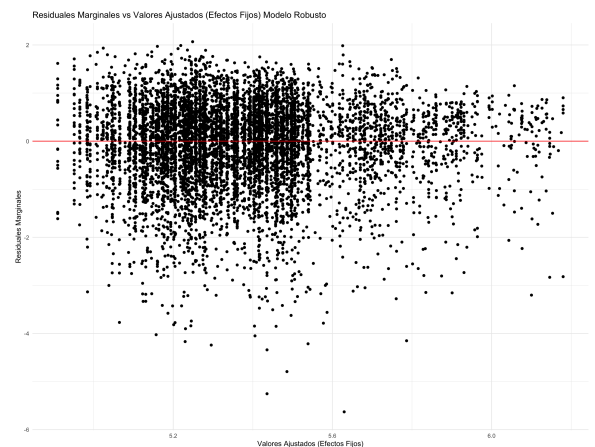


Figura 9: Residuos marginales del modelo robusto

Al realizar test de Breusch-Pagan, observamos que sigue la heterocedasticidad

```
data: residuos_cond_rob ~ valores_ajustados_rob + I(valores_ajustados_rob^2)
BP = 188, df = 2, p-value < 2.2e-16
```

Como factor final de comparación entre el modelo simple y el modelo robusto, utilizamos el RMSE, dado que

ambos modelos fueron evaluados mediante validación cruzada. Este enfoque permite comparar el rendimiento predictivo de los modelos en términos de la magnitud de los errores en las predicciones.

Es importante destacar que, a diferencia del modelo simple, el modelo robusto no proporciona medidas como el AIC, BIC o el log-likelihood debido a su naturaleza específica. Por ello, el RMSE es una métrica adecuada para evaluar el ajuste de ambos modelos y comparar su capacidad de predicción, en especial cuando el enfoque robusto busca minimizar el impacto de los outliers o datos ruidosos en las predicciones.

De esta manera, optamos nuevamente por nuestro modelo simple (**Modelo 3** en la tabla), ya que el modelo robusto no mostró una mejora significativa en los residuos ni en la heterocedasticidad. Además, los valores de RMSE de ambos modelos son bastante cercanos, lo que indica que ambos poseen una capacidad predictiva similar. Esta elección se fundamenta en la simplicidad y eficiencia del modelo simple, que proporciona resultados comparables sin la complejidad adicional del enfoque robusto.

- Factor de la inflación de la varianza (VIF):

Una de las verificaciones que se realizaron al momento de incluir las variables demográficas, fue la del VIF para evitar una multicolinealidad, a continuación mostramos su resultado, que indica que las variables pueden ser incluidas en nuestro modelo, de tal manera que se realizó

- Q-Q plot

Comparamos los cuantiles teóricos de una distribución normal estándar con los cuantiles de las muestras de los residuos del modelo.

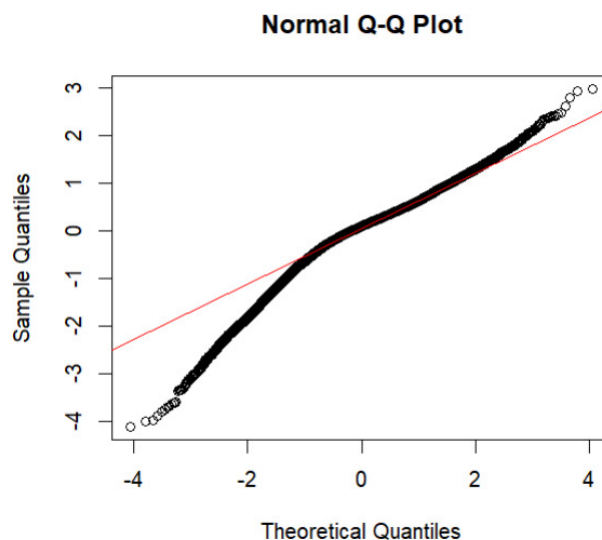


Figura 10: QQ Plot

RMSE	Modelo simple	Modelo robusto
Entrenamiento	0.5849	0.5947
Validación	0.6533	0.6558

Cuadro 2: RMSE modelos

	GVIF	Df	$GVIF^{1/(2 \cdot Df)}$
mes	1.00	11.00	1.00
income	1.08	11.00	1.00
household_size	1.08	4.00	1.01

Cuadro 3: Factor de la inflación de varianza (VIF)

Vemos que la mayoría de los residuos parecen ser aproximadamente normales, aunque las desviaciones en las colas sugieren que hay algunos valores atípicos y que la distribución de los residuos no es perfectamente normal en sus extremos. Debemos considerar esto, ya que la no normalidad de los residuos afectar la validez de las inferencias estadísticas del modelo.

ACF residuales

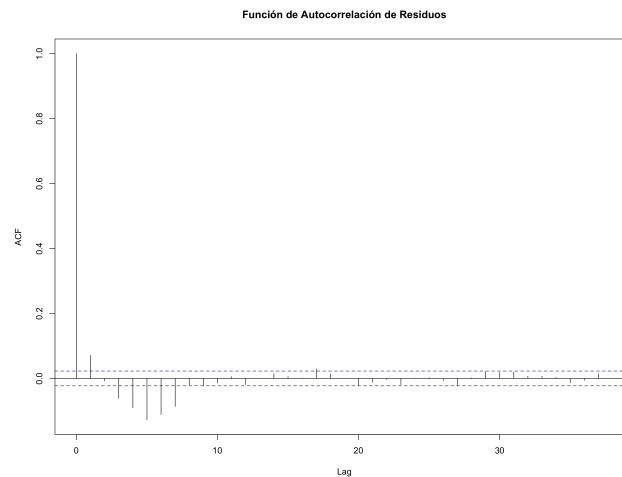


Figura 11: ACF residuales

Estamos en una estructura de autocorreación ya que se analizan los meses

En vista de las anteriores consideraciones, podemos proyectar el trabajo realizado anteriormente a una exploración **posterior** de un modelo mixto en el cual se puedan incluir algunos factores aleatorios adicionales para modelar de mejor manera, revisar igualmente los factores fijos y como estos afectan el modelo. También debemos tener en cuenta que al tratar con una estructura autocorrelacionada al tratarse del tiempo se pueden realizar aproximaciones desde un enfoque de series de tiempo, por ejemplo.

Finalmente revisamos del summary de nuestro modelo

```
Formula: log_sales_value ~ income + mes + household_size + (1 | household_id)
Data: train_data1
```

REML criterion at convergence: 16034.1

Scaled residuals:

Min	1Q	Median	3Q	Max
-6.9095	-0.3895	0.1113	0.5533	3.8854

Random effects:

Groups	Name	Variance	Std.Dev.
--------	------	----------	----------


```

household_id (Intercept) 0.4297    0.6555
Residual                  0.3798    0.6163
Number of obs: 7489, groups:  household_id, 801

```

Podemos ver que el 65,5 % de la desviación estándar en efecto aleatorio del hogar muestra que en nuestro modelo este porcentaje de variabilidad puede ser explicado por las diferencias entre los hogares (nuestras unidades), por su lado los residuales indican que hay un 61,6 % de variabilidad no puede ser explicada por el modelo

```

Fixed effects:
              Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   5.403e+00  5.199e-02  1.293e+03  103.929 < 2e-16 ***
income.L      6.579e-01  1.743e-01  7.732e+02   3.774 0.000173 ***
...
income^4      3.456e-01  1.585e-01  7.748e+02   2.181 0.029506 *
income^5      3.027e-01  1.686e-01  7.709e+02   1.795 0.073025 .
income^6      4.285e-01  1.684e-01  7.703e+02   2.545 0.011115 *
mes2         -7.571e-02  3.601e-02  6.712e+03  -2.103 0.035539 *
mes7          7.853e-02  3.589e-02  6.714e+03   2.188 0.028713 *
mes5          8.807e-02  3.572e-02  6.711e+03   2.465 0.013710 *
mes9          7.665e-02  3.567e-02  6.713e+03   2.149 0.031661 *
mes10         7.834e-02  3.583e-02  6.713e+03   2.186 0.028823 *
mes12         1.068e-01  3.604e-02  6.714e+03   2.963 0.003057 **
household_size.L 1.377e-01  6.961e-02  7.894e+02   1.979 0.048200 *

```

Podemos ver los β significativos dentro de nuestro modelo lo igualmente en los efectos fijos.

Conclusiones

El ajuste de modelos lineales mixtos para datos longitudinales de ventas mensuales permitió capturar hasta cierto punto tanto las variaciones individuales entre los hogares como los patrones generales observados en la población. Este enfoque jerárquico fue importante para reflejar la estructura inherente de los datos, donde la variabilidad entre los hogares desempeñó un papel clave en la predicción de las ventas.

El **Modelo 3** fue seleccionado como el más adecuado tras la evaluación de múltiples modelos, presentando los valores más bajos en los criterios de información de Akaike (AIC) y bayesiano (BIC). Estos indicadores sugirieron que el modelo alcanzó un equilibrio óptimo entre la complejidad y la calidad del ajuste, minimizando el riesgo de sobreajuste sin comprometer la precisión predictiva.

La transformación logarítmica aplicada a las ventas estabilizó la varianza y mejoró la normalidad de los residuos, lo cual fue esencial para cumplir con los supuestos de los modelos lineales mixtos. Esta transformación permitió además identificar relaciones no lineales y facilitó una interpretación más clara de los efectos de las variables explicativas.

Las variables demográficas como **Ingreso** y **Tamaño del Hogar**, incluidas como efectos fijos, resultaron ser altamente significativas. Esto resaltó su influencia en la variabilidad de las ventas mensuales y subrayó la importancia de considerar factores socioeconómicos en el análisis, permitiendo una diferenciación más precisa entre los hogares.

El tratamiento de los hogares como efectos aleatorios permitió modelar la heterogeneidad entre ellos, capturando las desviaciones específicas respecto al comportamiento promedio de la población. Además, la evaluación de los efectos aleatorios no solo en los interceptos, sino también en las pendientes, identificó variaciones en la respuesta de los hogares a lo largo del tiempo, lo que subrayó la naturaleza dinámica del fenómeno.

En conjunto, los modelos lineales mixtos demostraron ser una herramienta eficaz para analizar datos longitudinales, proporcionando una visión detallada de las diferencias entre los hogares y los patrones generales de las ventas mensuales.

Limitaciones y trabajo futuro

En este estudio, se asumió una relación lineal entre las variables explicativas y la variable respuesta. Sin embargo, la presencia de patrones complejos sugiere que un modelo estrictamente lineal podría no ser adecuado para capturar completamente las dinámicas subyacentes en los datos. Aunque los MLM permitieron capturar la heterogeneidad entre los hogares, la flexibilidad en la modelización de los efectos temporales podría haberse mejorado mediante la incorporación de técnicas avanzadas de suavizado, como splines o métodos no paramétricos, que facilitan una representación más precisa de las variaciones temporales no lineales.

En este sentido, se podría explorar la representación de los P-splines utilizando B-splines como bases dentro de los modelos mixtos, lo que proporcionaría mayor estabilidad y flexibilidad en la captura de los efectos temporales, abordando de manera más eficiente las heterogeneidades entre los sujetos mediante técnicas de suavizado avanzadas. [5]

De manera complementaria, otro enfoque potencial sería la aplicación de modelos lineales mixtos a datos jerárquicos segmentados, como aquellos relacionados con categorías específicas de productos o cupones, lo que podría ofrecer una comprensión más detallada de las dinámicas de ventas a nivel de sección o tipo de producto.

Asimismo, sería pertinente contemplar la incorporación de modelos más sofisticados que permitan capturar la autocorrelación temporal en los residuales (a nivel individuo) .[4] Entre estos modelos se encuentran el autorregresivo de primer orden (AR(1)), el de media móvil de primer orden (MA(1)) y su combinación en el modelo autorregresivo de media móvil (ARMA(1,1)). Además, se sugiere el uso de matrices de covarianza más flexibles, como la matriz de Toeplitz, que captura correlaciones decrecientes entre mediciones en distintos puntos temporales. Estas alternativas permitirían una mejor representación de la estructura de los residuos, ajustándose a la naturaleza longitudinal de los datos y mejorando la precisión de los modelos.

Referencias

- [1] Bates, D. (2011). *Mixed models in R using the lme4 package: Part 2: Longitudinal data, modeling interactions*. Marzo. Recuperado de <https://lme4.r-forge.r-project.org/slides/2011-03-16-Amsterdam/2Longitudinal.pdf>
- [2] Mahr, Thilo (2017). *Plotting partial pooling in mixed effects models*. Recuperado de <https://www.tjmahr.com/plotting-partial-pooling-in-mixed-effects-models>
- [3] Complete Journey Data (2023). *The Complete Journey Dataset*. Recuperado de <https://cran.r-project.org/web/packages/completejourney/vignettes/completejourney.html>
- [4] Anderson, C. J. *Longitudinal Data Analysis via Linear Mixed Models: Edps/Psych/Soc 587*. Department of Educational Psychology, Board of Trustees, University of Illinois. Recuperado de https://education.illinois.edu/docs/default-source/carolyn-anderson/edpsy587/lectures/9_longitudinal/longitudinal_beamer_post.pdf
- [5] Durbán, M. (s.f.). *Splines con Penalizaciones: Teoría y aplicaciones*. Universidad Carlos III de Madrid. Recuperado de <https://halweb.uc3m.es/esp/Personal/personas/durban/esp/web/cursos/Psplines/material/Psplines.pdf>