

Proyecto - Análisis Numérico

Daniel Fernández¹, David Herrera², and Pablo Pulido³

^{1,2,3}Departamento de Ingeniería de Sistemas, Pontificia Universidad
Javeriana

Abril 2019

1. Introducción

La televisión en Colombia hace parte del diario vivir de muchas familias. Es por esto, que son varias las programaciones que se realizan en un solo día, para poder ofrecer una variedad de opciones con respecto a lo que entretenimiento se refiere. Existen cinco canales principales nacionales en Colombia, dos privados (RCN y CARACOL) y tres operados por el estado (Canal Uno, Señal Colombia y Señal Institucional), los cuales nos ofrecen 5 opciones para canales nacionales. [1] A estos se le suman los canales regionales y los comunitarios, lo que nos deja con una enorme variedad de canales que son vistos diariamente por los colombianos. Colombiamania (2017)

En el año 2016, se hizo un estudio el cual confirmó que el 93 % de los hogares en Colombia cuenta con al menos un televisor. Este tipo de dato demuestra la alta cobertura que posee la televisión nacional en el país, así como también demuestra la importancia de este mismo medio para otros usos además del entretenimiento. El poder promocionar un producto o un servicio a más de 12,090,000(cifra 2018) familias es algo muy deseable para cualquier empresa, por lo cual se crean métodos de medición para conocer cuántos de ese "mercado meta" hace parte de la sintonía de un canal. ConexiónCapital (2017)

En Colombia se sigue un modelo de muestreo de datos similar a lo que se hace en el resto de Latinoamérica, en donde se selecciona una muestra representativa de la población colombiana, y se les hace medición de los canales que sintonizan día a día. Obtener estos datos sirve para que los canales sepan el agrado de cierto programa, lo cual permite que se realicen ciertas decisiones como: seguir con la serie o no, promover más propagandas en esas franjas, entre otros.

Debido a que el actual método puede resultar algo costoso por la frecuencia con la cual se toman los datos, Se propone un modelo de ajuste de curvas,

desarrollado en R, usando datos de rating de un día para poder predecir el comportamiento de la audiencia con respecto a que franjas son las más vistas, y en qué cantidad son vistas por los Colombianos.

2. Metodología

2.1. Metodología Numérica

Para el desarrollo de el aplicativo, se implementó el ajuste de curvas, por método de mínimos cuadrados. Además, ya que la problemática habla sobre si un programa se ve o no, se usará un modelo probit que asemeje una curva sigmoidea.

EL ajuste de curvas será desarrollado con la ayuda de la libreria GGplot2, la cual contiene una función llamada `stat_smooth()`, esta funcion recibe, además de los datos proporcionados por GGplot2, un modelo y un objeto de tipo formula. El Modelo, por cuestiones de conveniencia con la herramienta, debió ser lineal. Ya que el resto de modelos no permitía un ajuste de curvas como tal, mas bien realizaba una interpolación. Para contrarestar este inconveniente(el modelo de rating no sigue un comportamiento lineal), se pasó el parámetro formula, la cual seguía un planteamiento no lineal. El resultado fue exitoso ya que, como se verá en la aplicación web, se ve una curva ajustada con una incertidumbre de los datos relativamente pequeña

El modelo probit es usado para datos cualitativos de una variable dependiente que solo pueden tomar valores binarios. En este caso, los datos serian sobre si el televidente está observando cierto canal o no. A la hora de poner en practica este modelo, se tienen las siguientes equivalencias:

$$Pr(Y = 1|X) = F(X^tb)$$

$$F(X^tb) = Pr(Y^* > 0)$$

$$Y^* = bX^t + e$$

Donde Y es la variable binaria, X es el conjunto de variables independientes, F es la función de distribución acumulativa de la distribución normal, b es un parámetro estimado por máxima verosimilitud y e es el error que sigue una distribución normal $N(0, 1)$. Garza (2008)

Es importante mencionar este modelo porque, generalmente, sigue un comportamiento de datos semejante a una curva sigmoidea, el cual puede ser de gran ayuda a la hora de entender el comportamiento de la sintonía de un programa de televisión.

Una curva sigmoidea, es la cual gráficamente tiene una forma de S y es comúnmente vista en modelos para el área de biología Caglar y cols. (s.f.), así como es conocida como "la curva del aprendizaje". El principal motivo por

el cual se tiene en cuenta este tipo de comportamientos es porque, un modelo de rating de un canal comparte similitudes con este tipo de gráficas. Como

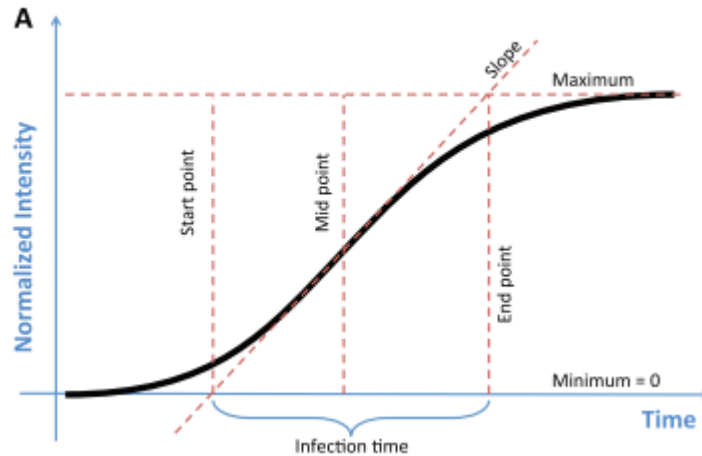


Figura 1: Ejemplo gráfica Sigmoide

podemos observar en la Figura 1, existen 4 principales momentos de la curva, y si hacemos e símil con el rating, serian los siguientes: primeras semanas no siempre tan positivas, momento después cuando el programa se vuelve popular, semanas después van saliendo de la audiencia personas no tan interesadas, final del programa donde ya poco a poco se deja de ver en esa franja.

Si bien no es un comportamiento exactamente igual en todos los casos, si se puede tener en cuenta este modelo para el ajuste de curvas. Cabe resaltar de que existe un paquete en R llamado Sicegar el cual facilita estos ajustes. Caglar y cols. (s.f.)

2.2. Obtención y trato de datos

Para la obtención de datos, se usó la página rating colombia” para obtener datos de sintonía en distintos días. La forma como se interpretan recogidos de esta página es a manera de porcentaje y son publicados en una lista de los 12 programas más vistos en la franja nocturna (después de las 7 pm y antes de las 11 pm) ya que este es el horario en donde hay más audiencia. Estos datos los extraemos por medio del paquete *rvest* de el lenguaje R. RatingColombia (2019) Wickham (2019) Wickham y cols. (2019)

Es deseable crear categorías propias de los horarios de televisión, en el cual también se denoten días donde hay más audiencia en la televisión colombiana. Por ejemplo, es más probable que un viernes se vea más televisión que un lunes,

o que un domingo no exista tanta audiencia por diferencias de contenido. Teniendo esto en cuenta, se crearon las siguientes categorías: Días entre semana, Sábados y Domingos. Como se mencionó anteriormente, el hecho de que exista una programación distinta en algunos días de la semana, hace que una segmentación como esta sea de ayuda a la hora de analizar la sintonía de cada programa al rededor de la semana. Finalmente, se decide tomar una muestra de datos quincenales (cada 15 días) para poder predecir la sintonía a un plazo de 6 meses.

3. Objetivos

Se pretende ofrecer una alternativa a un proceso que probablemente puede ser optimizado para reducir costos de operación, así como obtener información futura, lo cual servirá para que los canales que hagan parte de este estudio puedan conocer las estadísticas de audiencia. Portafolio (2013)

Además, se espera poder hacer un buen uso de los métodos numéricos de ajuste de curvas, el cual logre facilitar la predicción con el menor error posible en comparación con los datos reales. Así como poder aprender conceptos de otras disciplinas como lo son el modelo probit y la curva sigmoidea. De esta forma lo que se quiere es determinar la aproximada aplicando ajuste de curvas por mínimo cuadrado que nos ayude a determinar el porcentaje de audiencia en una franja determinada.

4. Antecedentes

La empresa que ofrece el servicio de recopilación de la estadística de rating a nivel latinoamericana es IBOPE, una empresa de origen brasileña que se especializa en medición, monitoreo y planificación de los medios. IBOPE (s.f.)

En el caso de Colombia, como se mencionó en la introducción, se dispone de una muestra significativa (en este caso de 1100 hogares al rededor de toda Colombia) a la cual se le hará la medición de rating. Para esto, se usan dispositivos llamados people Meter, el cual es instalado en el televisor para poder tener una medición verídica de si un programa es visto o no. Portafolio (2013)

En el ámbito de los modelos matemáticos del comportamiento del rating se realizó un estudio por parte de un grupo del politécnico de Milán en el cual se propuso un modelo dinámico el cual brinda una solución a esta misma problemática pero considerando muchas mas variables como los hábitos familiares, el contenido del programa sintonizado entre otros Pagano y cols. (2015) Si bien la solución propuesta en este documento no tiene en cuenta tantas variables, requiere de muchos menos conocimientos avanzados de métodos numéricos. Principalmente el tema de un modelo multinivel es algo que puede hacer que el

desarrollo de este aplicativo se extienda a mas de lo permitido, pero podría ser tenido en cuenta para futuras mejoras.

5. Diagrama de flujo

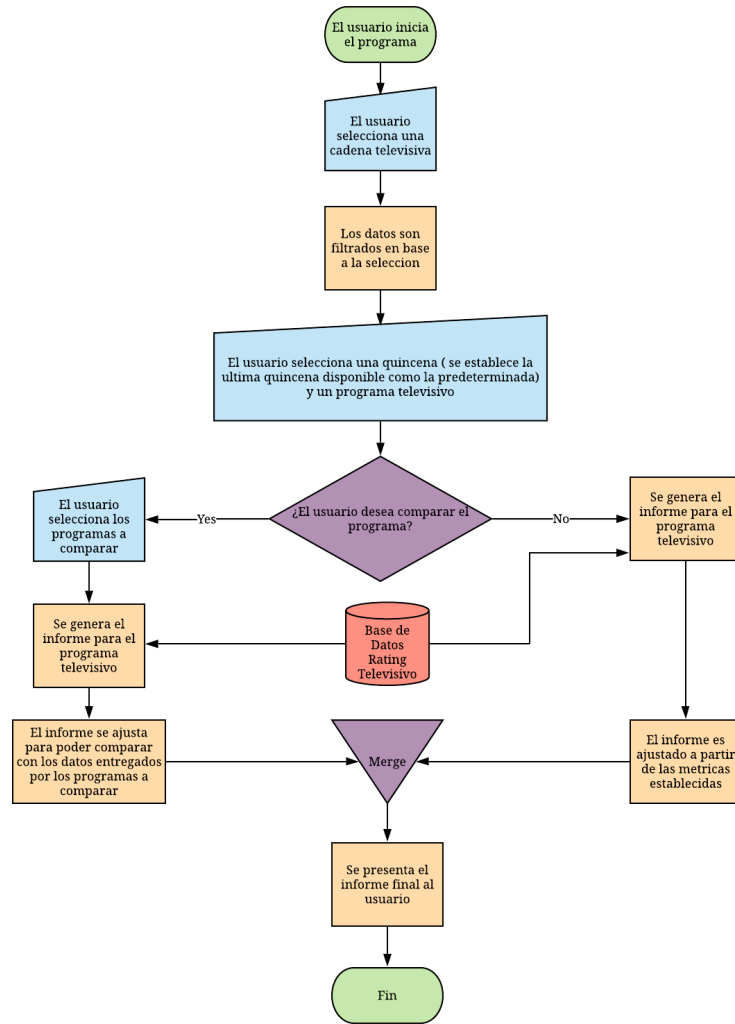


Figura 2: Diagrama de flujo

6. Cronograma

CRONOGRAMA DE ACTIVIDADES										
ACTIVIDAD	Tareas	Responsables	R.	ABRIL				MAYO		Observaciones
				SEMANA				SEMANA		
				1	2	3	4	1	2	
Definición del alcance del Proyecto.	Establecer los objetivos, propósito y alcance del proyecto	Todo el equipo								
	Definir los supuestos y restricciones	Todo el equipo								
Levantamiento de información respecto a variables adyacentes	Establecer que información se desea obtener a partir de los diferentes modelos de ecuaciones diferenciales y ajustes de curvas	Todo el equipo								
	Definir los segmetos por el cual se van a tipar los horarios de rating	Todo el equipo								
Corrección primera entrega	Realizar las correcciones pertinentes al documento, definir en su totalidad los objetivos y el modelo a utilizar	Todo el equipo								
Diseño de la aplicación	Definir los posibles casos de uso	Todo el equipo								
	Establecer el catálogo de estadísticas que se desea obtener	Todo el equipo								
	Definir las diferentes funciones de las que hara uso el aplicativo	Todo el equipo								
Realización del aplicativo	Implementación de la codificación para solucionar el problema	Todo el equipo								
	Realización de la interfaz	Todo el equipo								

Figura 3: Cronograma

7. Diseño de la aplicación

Para la interfaz grafica de usuario, GUI por sus siglas en inglés, se hará uso de Shiny, una librería de R la cual implementa HTML y CSS para realizar aplicativos web sin necesidad de conocer de esos lenguajes, es decir, directamente con R. Se tiene planeado que en la aplicación web se tenga acceso a una importación de datos al sistema, una graficación de los mismos, un calculo por medio de ajuste de curvas, y, dependiendo de la fecha de los datos ingresados, se hará una comparativa entre los datos ajustados y los reales.

GGPlot2 es una librería de R la cual sirve para realizar gráficas de datos que son mas vistosas y fueron escogidas por el grupo por su llamativo diseño. Generalmente, se graficaría los datos obtenidos, junto con los datos ajustados para un futuro. Ya si se quiere hacer una comparativa con los datos reales, en caso de que los datos originales fueran antiguos, pues se recurrirá a graficar los valores reales juntos con los calculados. Para comprender de mejor manera la diferencia entre estos datos, se recurre al calculo del error, donde se comparan valores teóricos y los reales de la siguiente forma:

$$Error = \frac{|aprox - real|}{real}$$

Para la captura de datos, se implementó un método en R, el cual, dada unas fechas especificas, revisará en la pagina [ratingcolombia](#), los datos de rating de un programa en especifico.

Referencias

- Caglar, M. U., Teufel, A. I., y Wilke, C. O. (s.f.). Sicegar: R package for sigmoidal and double-sigmoidal curve fitting.(Ingles)[sicegar: Paquete en r para ajuste de curvas sigmoidales y doble sigmoidales], url=.
- Colombiamania. (2017). Canales de television nacionales y regionales en colombia. *Colombiamania*. Descargado de <http://www.colombiamania.com/medios/television/index.htm>
- ConexiónCapital. (2017). En colombia hay más televisores que personas con servicios básicos. *Conexión Capital*. Descargado de <https://conexioncapital.co/colombia-mas-televisores-personas-servicios-basicos/>
- Garza, O. (2008). Bienestar subjetivo, ingreso y autonomía. apendice c: Modelo probit. Descargado de http://catarina.udlap.mx/u_dla/tales/documentos/lec/garza_vor/apendiceC.pdf
- IBOPE. (s.f.). QuiÉnes somos. *KANTAR IBOPE MEDIA*. Descargado de <http://www.kantaribopemedia.com.co/ibope.html>
- Pagano, R., Quadrana, M., Cremonesi, P., Bittanti, S., Formwentin, S., y Mosconi, A. (2015). Prediction of TV ratings with dynamic models.(Ingles)[predicción de ratings de tv usando un modelo dinámico]. *RecSysTV 2015 2nd Workshop on Recommendation Systems for TV and Online Video*. Descargado de <https://home.deib.polimi.it/pagano/portfolio/papers/PredictionofTVratingswithdynamicmodels.pdf>
- Portafolio. (2013). Así se mide el 'rating' en colombia. *Portafolio*. Descargado de <https://www.portafolio.co/tendencias/mide-rating-colombia-87718>
- RatingColombia. (2019). Rating colombia lunes 8 de abril. *Rating Colombia*. Descargado de <https://www.ratingcolombia.com/search/label/Rating%20Colombia>
- Wickham, H. (2019). rvest: Easily harvest (scrape) web pages [Manual de software informático]. Descargado de <https://CRAN.R-project.org/package=rvest> (R package version 0.3.4)
- Wickham, H., François, R., Henry, L., y Müller, K. (2019). dplyr: A grammar of data manipulation [Manual de software informático]. Descargado de <https://CRAN.R-project.org/package=dplyr> (R package version 0.8.1)