

Capítulo 1

Metodología

1.1. Antecedentes

El incremento de la participación cultural forma parte tanto de las políticas públicas de diversos países como de los objetivos de las instituciones. Diversas razones, algunas veces muy interrelacionadas, son esgrimidas para diseñar estrategias que fomenten la participación del público, sea en calidad de asistentes, como actores o en forma mixta, entre las cuales se encuentran:

- La *sostenibilidad económica* de las instituciones culturales: una mayor asistencia asegura los ingresos necesarios para mantener o ampliar la oferta de servicios y productos culturales.
- La *inclusión social*. En naciones occidentales la asistencia a determinados eventos culturales, como conciertos y salones de arte, proviene principalmente de personas provenientes de hogares con holgados niveles de ingresos. De modo que se trazan estrategias para atraer a personas de menores ingresos y las instituciones se involucran de manera más cercana con las comunidades ofreciéndoles actividades más atractivas.
- *Integración cultural*: También se observa que la asistencia a exposiciones en museos, proviene principalmente de personas descendientes de europeos [2]. El progresivo cambio en la composición étnica de la población reduce a mediano y largo plazo, la audiencia, los ingresos y hasta la razón de ser de dicha instituciones, de modo que se trazan estrategias para atraer a personas de otros grupos étnicos, provenientes de emigrantes y refugiados. Para ello varían su personal y oferta, incluso en ciertas naciones, como Australia, se interesan por la integración cultural plena de inmigrantes y refugiados, conduciendo estudios sistemáticos para analizar los posibles obstáculos y tratar de reducirlos o eliminarlos.
- *Expectativa de vida y Bienestar social*. Ciertas naciones han encontrado una correlación positiva entre la expectativa de vida y la sensación de bienestar de las personas y su participación en actividades culturales. De allí el interés por promover y facilitar a sus ciudadanos el acceso a los bienes y servicios culturales.

A los fines de evaluar la efectividad de las políticas públicas dirigidas al incremento de la participación cultural, los estados periódicamente generan estadísticas y construyen modelos lineales.

1.2. Modelos Generales Lineales Multinivel

Los MGLM consideran un conjunto de datos jerárquicos consistentes de sujetos anidados dentro de grupos, con una variable de respuesta medida al nivel más bajo y variables explicativas en todos los niveles existentes. Es útil visualizar un MGLM como un sistema jerárquico de ecuaciones de regresión. Una forma general de presentar los modelos corresponde a Lee et al [5] quien presenta los doble MGLM, siglas en inglés de “double hierarchical generalized linear models”, con la siguiente estructura (los superíndices indican a qué parte del modelo pertenece la matriz de diseño o parámetro):

$$\mathbf{y} = \mathbf{X}^{(\mu)}\beta^{(\mu)} + \mathbf{Z}^{(\mu)}\mathbf{v}^{(\mu)} + \epsilon \quad (1.1)$$

Donde $\mathbf{X}^{(\mu)}$ y $\mathbf{Z}^{(\mu)}$ son matrices de diseño, mientras que $\beta^{(\mu)}$ y $\mathbf{v}^{(\mu)}$ son efectos fijos y aleatorios respectivamente, para el modelo de componentes de μ . Los modelos de este tipo admiten efectos aleatorios para todos los componentes de los modelos, es decir permiten:

1. Efectos aleatorios para la media μ proporcionando modelado de covarianza y variación extra-Poisson¹ ; o extra-binomial² .
2. Los efectos aleatorios para ϕ , por ejemplo, ofreciendo una distribución de cola pesada para la varianza residual para dar un análisis robusto contra valores atípicos o contaminaciones de datos.
3. Efectos aleatorios para λ , por ejemplo $v_i^{(\mu)} \sim N(0, \lambda)$, suministrando distribuciones de cola pesada para los efectos aleatorios a fin de proporcionar un análisis robusto contra las especificaciones erróneas de los supuestos de distribución sobre los efectos aleatorios. Además, algunos supuestos de distribución proporcionan una selección de variables.

A continuación se desarrollará, sin pérdida de generalidad, un sistema simplificado de dos niveles extensible a otros, similar al de Hox *et al* [[4], pp. 8-13). Se modela la *asistencia* a los eventos culturales en función de varias variables. En el nivel 2 hay J entidades federales (J= 1,2,..24 estados) con n_j municipios en cada uno de ellos. La variable dependiente es la *asistencia a eventos culturales en el nivel 1*, es decir en el municipio (Y), la cual es una variable que toma valores numéricos iguales o mayores a cero. Para simplificar, se la hace depender de dos variables explicativas en el nivel 1: si el municipio tiene *Presencia sedes culturales* X_1 (X_1 : Sin sedes=0, Con sedes=1) y la *cantidad de actividades realizadas en el año* X_2 , la cual toma valores numéricos mayores o iguales a cero; y una variable explicativa en el nivel 2 de los estados: el *nivel de desarrollo socioeconómico del estado*

¹Es decir con sobredispersión debido a una varianza superior a la media por lo cual se usa una distribución gamma para modelar el carácter aleatorio de esta última.

²También con sobredispersión, en la cual algunas observaciones no son Bernoulli y se modela el carácter aleatorio del parámetro de la binomial con una distribución beta.

Z , el cual de acuerdo al trabajo de Siso Lucena y Mac-Quhae [7] toma cinco valores de 1-5. Hay datos correspondientes a $i=1,2,\dots,335$ municipios en $j=1,2,3,\dots,24$ estados. Se escriben ecuaciones de regresión separadas para cada estado:

$$y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \epsilon_{ij} \quad (1.2)$$

Usando etiquetas de variables en lugar de símbolos algebraicos, la ecuación es la siguiente:

$$Asistencia_{ij} = \beta_{0j} + \beta_{1j} Sedes_{1ij} + \beta_{2j} Actividades_{2ij} + \epsilon_{ij} \quad (1.3)$$

Donde:

β_{0j} : intercepto

β_{1j} : Coeficiente o pendiente de regresión para la variable explicativa numérica *Presencia de sedes culturales municipales*,

β_{2j} : Coeficiente de regresión (pendiente) para la variable explicativa numérica *cantidad de actividades realizadas en el municipio*, y

ϵ_{ij} : Error residual. Los errores residuales tienen una media de cero $\mu_\epsilon = 0$ y una varianza σ_ϵ^2 que debe estimarse.

A diferencia de la regresión habitual cada clase tiene un coeficiente de intercepción diferente, β_{0j} , y diferentes coeficientes de pendiente, β_{1j} y β_{2j} , de allí el índice subscrito j .

El siguiente esquema muestra la interacción entre el nivel 2 y el nivel 1:

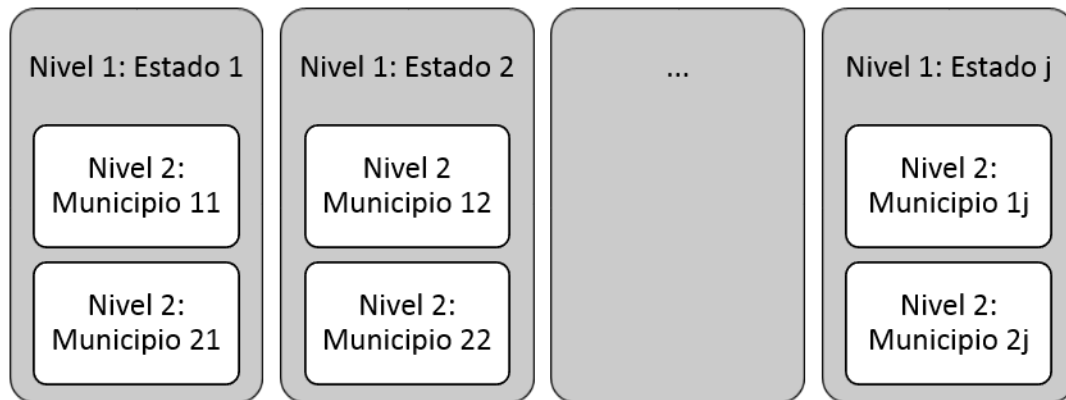


Ilustración 1. Esquema de dos niveles: Estados en el nivel 2 y Municipios en el Nivel 1. Autor)

Los coeficientes de intercepto y pendiente son los *coeficientes aleatorios*. Su variación no debe ser totalmente aleatoria, por lo que se puede explicar parte de la variación vía la introducción de variables de un nivel superior. Generalmente se explica solo parte de las variaciones, por lo que se espera que permanezca alguna variación residual inexplicable.

Los coeficientes de intercepto y de pendiente son una característica de estado, uno con un intercepto alto tendrá más asistencia que otro con un valor bajo. Hay una variable *dummy* para la *Presencia de sedes*, el valor de la intercepción reflejará el valor predicho para aquellas sin sedes. Las distintas intercepciones desplazan el valor medio de todo el estado, tanto de los *Sin sedes* como *Con Sedes*. Las diferencias en el coeficiente de pendiente para la *presencia de sedes* o la *cantidad de actividades realizadas*, indican que la relación entre estas predictoras y la asistencia predicha no será la misma en todos los estados. La figura presenta un ejemplo con dos grupos:

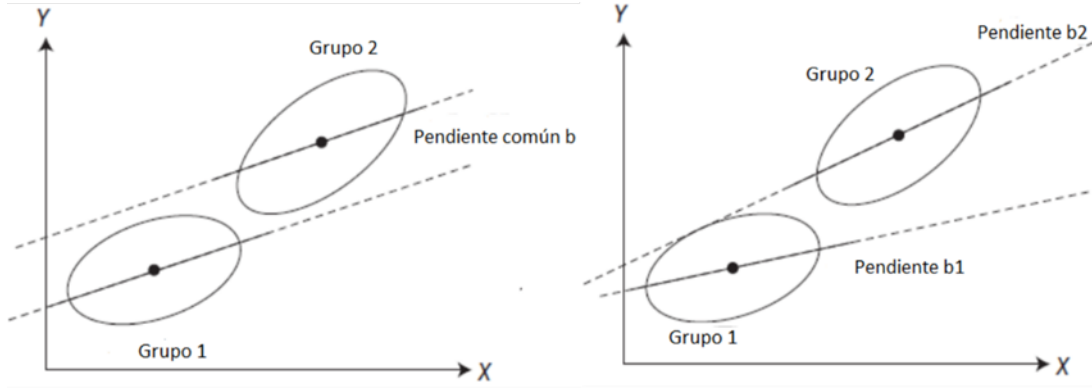


Ilustración 1. Esquema de dos niveles: Estados en el nivel 2 y Municipios en el Nivel 1.

A la izquierda aparecen 2 grupos sin variación de pendiente, y como resultado las dos pendientes son paralelas. Las intercepciones para ambos grupos son diferentes, el de la derecha muestra dos grupos con diferentes pendientes. La variación en las pendientes también tiene un efecto en la diferencia entre las intercepciones. En todas los estados, los coeficientes de regresión $\beta_{0j} \dots \beta_{2j}$, tienen una distribución de la familia exponencial. El siguiente paso es explicar la variación de los coeficientes de regresión $\beta_{0j} \dots \beta_{2j}$, introduciendo variables explicativas en la variable al nivel de estado, para el intercepto.

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j} \quad (1.4)$$

Y para las pendientes:

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j} \quad (1.5)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}Z_j + u_{2j} \quad (1.6)$$

La ecuación 1.4 predice la *asistencia media* en un estado (el intercepto β_{0j} según el *nivel de desarrollo socioeconómico del estado* (Z)). Por lo tanto, si γ_{01} es positiva, la *asistencia media* es mayor en los estados con mayor nivel de desarrollo socioeconómico. Por el contrario, si γ_{01} es negativa, la *asistencia media* es menor en los estados con mayor nivel de desarrollo socioeconómico. La ecuación 1.5 establece que la relación, expresada por el coeficiente de la pendiente β_{1j} , entre la *asistencia* (Y) y la *Presencia sedes culturales* (X) en el municipio, depende del *nivel de desarrollo socioeconómico del estado* (Z). Si γ_{11} es positivo, el efecto de *Presencia sedes culturales* en la *asistencia* es más grande con estados con mayor *nivel de desarrollo socioeconómico*. Por el contrario, si γ_{11} es negativo, el efecto de *Presencia sedes culturales* en la *asistencia* es más pequeño en estados con mayor *nivel de desarrollo socioeconómico*. La ecuación 1.6 establece que si γ_{21} es positivo, entonces el efecto de la *cantidad de actividades* realizadas en el municipio es mayor en estados más desarrollados socioeconómicamente. El *nivel de desarrollo socioeconómico del estado* actúa como variable moderadora de la relación entre *asistencia* y *Presencia sedes culturales* o *cantidad de actividades realizadas en el municipio*.

Los términos u , es decir u_{0j} , u_{1j} y u_{2j} son términos de error residual al nivel de estado. Los errores residuales tienen media cero y son independientes de los errores residuales ϵ_{ij} al nivel de los municipios. La varianza de los errores residuales u_{0j} es especificada como $\sigma_{u_0}^2$ y la varianza de los errores residuales u_{1j} y u_{2j} son especificadas como $\sigma_{u_1}^2$ y $\sigma_{u_2}^2$. Las

covarianzas entre los términos de los errores residuales son denotadas por $\sigma_{u_{01}}$, $\sigma_{u_{02}}$ y $\sigma_{u_{12}}$, las cuales generalmente se suponen distintas de cero.

Los coeficientes de regresión γ no varían entre los estados, por lo cual carecen del subscrito j , y por aplicarse a los estados, se denominan *coeficientes fijos* del modelo. Toda la variación entre estados que queda en los coeficientes de β , después de predecirlos con la variable de estado Z_j , se supone que es una variación del error residual. Esta es capturada por los términos de error residual u_j los cuales sí llevan el subscrito j para indicar el estado al que pertenecen.

El modelo con dos variables explicativas a nivel de municipio y una a nivel de estado puede ser escrito como una ecuación compleja de regresión sustituyendo las ecuaciones 1.5 y 1.6 en la ecuación 1.4 obteniéndose:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{1ij} + \gamma_{20}X_{2ij} + \gamma_{01}Z_j + \gamma_{11}X_{1ij}Z_j + \gamma_{21}X_{2ij}Z_j + u_{ij}X_{1ij} + u_{2j}X_{2ij} + u_{0j} + \epsilon_{ij} \quad (1.7)$$

Usando etiquetas de variables en lugar de símbolos algebraicos, se obtiene:

$$Y_{ij} = \gamma_{00} + \gamma_{10}Sede_{ij} + \gamma_{20}Actividad_{ij} + \gamma_{01}Desarrollo_j + \gamma_{11}Sede_{ij}Desarrollo_j + \gamma_{21}Actividad_{ij}Desarrollo_j + u_{ij}Sede_{ij} + u_{2j}Actividad_{ij} + u_{0j} + \epsilon_{ij} \quad (1.8)$$

El segmento $[\gamma_{00} + \gamma_{10}X_{1ij} + \gamma_{20}X_{2ij} + \gamma_{01}Z_j + \gamma_{11}X_{1ij}Z_j + \gamma_{21}X_{2ij}Z_j]$ contiene los coeficientes fijos siendo la parte *fija o determinista* del modelo. La restante, con los errores aleatorios, es la parte *aleatoria o estocástica* del modelo. Los términos $X_{1ij}Z_j$ y $X_{2ij}Z_j$ son de interacción y aparecen como consecuencia de modelar β_{1j} de una variable del municipio X_{ij} con la variable de estado Z . Así, el efecto moderador de Z sobre la relación entre la variable dependiente Y y el predictor X , se expresa en la versión de ecuación simple del modelo como una interacción entre niveles. La interpretación de los términos de interacción en el análisis de regresión múltiple es compleja, pero lo relevante a tener en cuenta es que la interpretación sustantiva de los coeficientes en modelos con interacciones es mucho más simple si las variables que configuran la interacción se expresan como desviaciones de sus respectivas medias.

En el análisis de regresión múltiple los coeficientes de regresión se estandarizan a menudo porque se facilita la interpretación cuando se desea comparar los efectos de diferentes variables dentro de una muestra. Pero si el objetivo es comparar las estimaciones de los parámetros de diferentes muestras entre sí, se deben utilizar coeficientes no estandarizados. Para estandarizar lo mejor es derivar los coeficientes de la regresión estandarizada a partir de los coeficientes no estandarizados. Si existe una parte aleatoria complicada, que incluye componentes aleatorios para pendientes de regresión, se debería pensar cuidadosamente acerca de la escala de las variables explicativas.

1.3. Selección de variables

La selección de variables busca un modelo que se ajuste bien a los datos y que a la vez sea posible buscar un equilibrio entre bondad de ajuste y sencillez. Puede verse como un problema de regularización que asegure la *ortogonalidad* de las variables, para ello se siguen los pasos siguientes:

1. Visualización y cálculo de correlación y asociación entre las posibles variables predictoras, a los fines de identificar *colinealidad* o asociación entre alguna de ellas, y eliminar las redundantes.
2. Uso de *métodos de aprendizaje* no supervisado que permitan identificar las predictoras con mayor fuerza explicativa o una combinación lineal de las mismas, por ejemplo vía un *Análisis de Componentes Principales* o ACP o un Análisis de Correspondencia para variables categóricas.
3. Uso de métodos *por paso (stepwise)* como *Eliminación hacia Atrás (Backward elimination)*, *Eliminación hacia Adelante (Forward selection)*, *Selección Paso a Paso (Stepwise selection)* y *Selección del mejor subconjunto*.

La estimación penalizada proporcionará estimaciones con cierta reducción, y puede utilizarse con poco esfuerzo adicional con procedimientos comunes. Al aplicarlo al modelo lineal estándar, donde se estima el valor de β que minimiza la pérdida del error cuadrático, se tiene:

$$\hat{\beta} = \arg \min_{\beta} \sum (y - X\beta)^2 \quad (1.9)$$

En pocas palabras, buscamos coeficientes que minimizan la suma de los residuos cuadráticos. Sólo se añade una componente de penalización al procedimiento:

$$\hat{\beta} = \arg \min_{\beta} \sum (y - X\beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (1.10)$$

λ es el término de penalización para el cual valores mayores resultarán en una mayor reducción. Se aplica a la norma $L1$ o *Manhattan* de los coeficientes, $\beta_1, \beta_2 \dots \beta_p$, es decir, sin incluir la intercepción β_0 , y es la suma de sus valores absolutos, comúnmente denominado *lasso* o *least absolute shrinkage and selection operator*, por sus siglas en inglés. Para los modelos lineales y aditivos generalizados, se puede expresar conceptualmente una verosimilitud penalizada así:

$$l_p(\beta) = l(\beta) - \lambda \sum_{j=1}^p |\beta_j| \quad (1.11)$$

Como se está maximizando la verosimilitud, la penalización es una resta, pero no se muestra nada inherentemente diferente. Esta idea básica de añadir un término de penalización se aplicará a todos los enfoques de aprendizaje automático, pero como se muestra, es posible aplicar esta herramienta a los métodos clásicos para mejorar el rendimiento de las

predicciones. Se puede proceder a la regularización de diferentes maneras. Por ejemplo, usando la norma cuadrada L2 resulta en lo que se llama regresión de *ridge* o cresta (también conocida como regularización Tikhonov), y usando una combinación ponderada de las penalizaciones de lasso y cresta nos da una regularización de *red elástica*.

1.4. Paquetes de R para Modelos Multinivel

Existen diversas opciones para el manejo de modelos multinivel pero R los aventaja por su carácter de software libre y gratuito, la permanente actualización de su comunidad de usuarios y en ser de hecho el estándar en el mundo académico. Field *et al* [3] señala varios paquetes que se pueden utilizar: nlme [8] y lme4 [1]. El paquete nlme permite modelar la estructura de covarianza. Lee, Rönnegård, & Noh [6] presentan los paquetes HGML (hierarchical generalized linear models) y Doble HGML o DHGML [5]. El paquete *dhglm* incluye varias clases de modelos incluyendo: modelos lineales generalizados (MLGe), MLGe conjuntos, MLGe con efectos aleatorios (conocidos como HGLMs) y HGLMs incluyendo modelos para los parámetros de dispersión, incluyendo HGLMs dobles (DHGLMs) introducidos posteriormente.

Para el desarrollo de la investigación se ensayará con los paquetes disponibles, pero al final se elegirá uno solo para la presentación de los resultados, de acuerdo a la facilidad de uso para los eventuales operadores de la institucionalidad cultural.

1.5. Operacionalización del indicador

A continuación se detalla la operacionalización del indicador o variable dependiente del modelo:

Variable dependiente

Asistencia: Mide los niveles de participación de la población en eventos municipales culturales, como exposiciones, ferias, talleres, exposiciones en museos, visitas a bibliotecas, sitios históricos-culturales, entre otros, como asistentes en calidad de público o consumidores los bienes y servicios culturales ofrecidos en el año 2018. La información es recopilada a través de los entes institucionales regionales de la cultura. La unidad de análisis es el municipio como primer nivel y el estado como segundo nivel.

Variables independientes

Por su parte las variables independientes se extraen de los diversos registros de datos disponibles en los entes oficiales a cargo de la cultura y entes relacionados con el presupuesto y finanzas públicas, estadísticas nacionales, turismo y división político administrativa del país. Algunas de estas variables ya han sido analizadas en investigaciones preliminares. Siguiendo las orientaciones de la UNESCO y su adaptación al caso venezolano, se agrupan por las siguientes dimensiones:

Sociales

- *Educación*: Población escolarizada/ Niveles de alfabetización por estado
- *Pobreza*: Indicador de niveles de pobreza por municipios.

Económico-Financiero

- Actividad Económica por Entidad Federal: establecimientos económicos por estados según último censo industrial de 2012.
- Presupuesto de Municipios: Total y Sector 09 Cultura y Comunicación, por cada año y total acumulado en el período 2008-2018.
- Presupuesto de Entidades Federales proyectado a los municipios de acuerdo a su población: Total y Sector 09 Cultura y Comunicación, por cada año y total acumulado en el período 2008-2018.
- PIB total y pér cápita correlacionado con la creación de Sedes de Sedes Culturales.

Infraestructural cultural

- Sedes culturales en 2018 discriminadas por Estado, Municipio, Parroquia, Poblado o Ciudad, ente, tipo de sede, cantidad de trabajadores, beneficiarios, naturaleza de la propiedad, año de fundación, Área Cultural y Plataforma Cultural de adscripción.
- Espacios de la Misión Cultura en 2018 por Estado, Municipio, Parroquia, Poblado o Ciudad, Tipo de Sede, cantidad de animadores, tutores y beneficiarios, año de fundación, Área Cultural y Plataforma Cultural de adscripción.

Turismo

- Afluencia turística por Entidad Federal en el período 2015-2018.

Actividad Cultural

- Actividades de Gabinetes Culturales por fecha, ubicación, beneficiarios, área y Plataforma Cultural.
- Cultores: Todo aquel dedicado a determinada actividad cultural de manera permanente y que se considere a sí mismo como: Cultor o Cultora, Artista, Intelectual, Artesano o Artesana. Incluye a las personas dedicadas a la promoción y difusión cultural.

Geografía

- Extensión territorial de Estados y Municipios.
- Población de Estados, Municipios y Parroquias proyectada quinquenalmente de 2000 a 2050.

Organización

- Plataformas Culturales, Entes adscritos y Gabinetes Culturales Territoriales.
- Personal por entes y jerarquía.

Política Cultural

- Descripción de las Actividades de Gabinetes en 2018.
- Parroquias priorizadas.
- La cultura en Plan de la Patria 2013-2019.

- [1] Bates, D., Maechler, M., Bolker, B. and Walker, S. 2019. *Linear Mixed-Effects Models using Eigen and S4*. R package version 1.
- [2] Coates, C. 2019. *Culture for all – museum attendance in an age of diversity*. Disponible en <https://blooloop.com/features/museum-attendance-diverse-audience/>.
- [3] Field, A., Miles, J. and Field, Z. 2012. *Discovering statistics using r*. Sage.
- [4] Hox, J., Moerbeek, M. and Schoot, R. 2018. *Multilevel analysis, techniques and applications*. Routledge.
- [5] Lee, Y. and Noh, M. 2018. Double hierarchical generalized linear models. *R package version 20*. (2018).
- [6] Lee, Y., Ronnegard, L. and Noh, M. 2017. *Data analysis using hierarchical generalized linear models with r*. Taylor & Francis Group.
- [7] Lucena, C. and Mac-Quhae, R. 2018. Desigualdad Territorial en Venezuela: Una aproximación a través de indicadores socioeconómicos. *Terra Nueva Etapa*. XXXIV, 55 (2018), 1–24.
- [8] Pinheiro, J., Bates, D., DebRoy, S. and Sarkar, R. package version 3.-1., D 2019. *Linear and nonlinear mixed effects models*.