Universidad Nacional Autónoma de México

Facultad de Ciencias

Corrupción y confianza institucional: análisis de los Estados Mexicanos mediante modelos multinivel

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Matemático

PRESENTA:

Cristobal Bautista Hernández

TUTORA

Lizbeth Naranjo Albarrán

Ciudad de México 2022

Agradecimientos

Es díficil decir como fue esta etapa para mi realizando este proyecto. Se podría decir que es solo mío, tal vez lo sea en cuanto a qué lo desarrollé y escribí, pero a su vez recibí muchos aprendizajes y apoyo tanto a nivel profesional como a nivel personal. Un trabajo puede que solo refleje el empeño que se pone a dicha tarea, pero el entorno de quien desarrolla el mismo puede hacer que el proyecto quede a la mitad, sea fantástico o nunca se terminé.

Es por ello que quiero darle las gracias a todos los que me acompañaron en este proceso, tanto en mi paso por la universidad como en mi tiempo haciendo la tesis, lo cual se aprecia más recordando que en esta época se realizaron las actividades a distancia, y aunque no pondré muchos de esos nombres que formaron parte de este proceso, me hace feliz que los haya conocido y más aún si leen esto.

A mis compañeros, quienes me acompañaron a vivir esta travesía en la universidad la cual, no fue fácil y más hablando de un área el cual la mayoría de los jóvenes se matienen alejados y la sociedad mantiene adoctrinada como algo desconocido de lo cual solo las personas más raras e ïnteligentes”participan y por ende, algo que pocos comprenden. En especial, agradezco a Rodrigo Perusquía, quien me ayudó mucho en mis primeros semestres, junto con su disciplina y habilidad me parecía espectacular y me esmeraba a querer ser tan bueno en alguna de las disciplinas como él.

A mis profesores, quienes varios de ellos fueron como hermanos mayores o padres pues me guiaron en mi futuro profesional y a mejorar a nivel personal con sus comentarios tanto positivos como negativos, y aún así considero que en ocasiones no se les da el mérito que los profesores tienen en nuestra vida y sociedad, desde la educación básica hasta la superior. Por nombrar algunos a Bibiana Obregón, Graciela Martínez y Nahiely Canseco quienes me motivaron a seguir hasta terminar esta etapa y no solo eso, si no a seguir conociendo más de mi área y de otras.

A mi universidad, la cual me dió la oportunidad de ingresar a este nivel sin tener que realizar un examen de la transición del bachillerato a universidad. Pero también por darme la oportunudad de participar en muchas actividades, así como desarrollar nuevas habilidades de tal manera que solo me preocupará en mi rendimiento y no en el costo de los mismos. Por supuesto, esto gracias a las contribuciones de los mexicanos mediante impuestos y a qué parte de ello se va a nuestra universidad, por lo cual espero ser de servicio a mi país y a mi alma mater.

A mis sinodales, gracias por los comentarios, poco o muchos me ayudaron a mejorar mi trabajo, sobretodo gracias a Ricardo pues considero que sus comentarios fueron parte de una gran retroalimentación, una lástima no haberlo conocido pero espero tener esa oportunidad en algún momento. También a Salvador, que con su muy particular manera de expresarse, sé que lo hace con el objetivo de explotar al máximo el potencial de sus estudiantes y así estar seguro de que ellos saben hacer bien su labor.

A mis amigos qué como varios profesores me lo dijeron, se cuentan con los dedos de una mano. En especial gracias a Noel quien desde el bachillerato decidió estar conmigo y que era de las pocas personas que me llegaba entender, pues partiamos de condiciones parecidas. A Monserrat con quien siempre pude pasar momentos agradables en mi estancia en la universidad.

Un muy apreciado agradecimiento a mi tutora Lizbeth, pues me permitió desarrollar mis ideas y plasmarlas en el trabajo, su tolerancia, empatía y atención, por lo que puedo decir que no pude haber escogido mejor persona para apoyarme, solo llegué a hablar con ella en este trabajo y se mostró dispuesta a enseñarme aún cuando nunca fui su alumno, pero que estoy seguro se esmera mucho en formar buenos estudiantes, no solo por mi experiencia, sino por la de muchos otros que he escuchado.

Finalmente y no por ello menos importante. Dedicó esto a mi madre quién lamentablemente ya no vió qué logré terminar mi trabajo, alguien que dió muchas de sus horas de sueño para preparme un desayuno, que aunque era sencillo, siempre lo tenía, quién me apoyaba en lo moral y psicológico, pues sus escasos estudios solo me enseñaron a hacer las cosas básicas. Agradezco a mi padre, quién me apoyo sobretodo de manera ecónomica, aunque sean pocos ingresos, gracias a ello mi madre se encargaba de administrarlo. A mis hermanas quienes estas últimas semanas me han ayudado en actividades de mi hogar para poder terminar este trabajo.

Índice general

1

Introducción

Cuando uno comienza a usar modelos estadísticos, de manera frecuente se prefiere utilizar los modelos de regresión lineal, los cuales aunque no se ignora su poder de predicción, en ocasiones son insuficientes para poder generar un buen modelo que ayude a la interpretación o se ajuste mejor a los datos. En este proyecto se desarrolla la consideración de tomar ciertas variables como parte de la variación del modelo, lo cual se mide de manera aleatoria a tráves de un efecto por niveles en cada categoría. Además, se extiende esta idea hacia los modelos con variable categórica como respuesta.

El motivo es estudiar la percepción de la corrupción y la confianza institucional en el caso de México, para ver cómo ésta se comporta midiendo el efecto aleatorio a través de los Estados. Lo anterior, mediría un efecto que se relaciona a factores propios de la entidad que los hace ser más o menos severos en la manera de evaluar. Este análisis permite estimar la manera en que se evalúa a los gobiernos partiendo de un efecto inicial, poniendo este efecto como parte de la varianza y no siendo una variable independiente en sí misma.

Se analizan dos años por separado, 2017 y 2019, para comprobar que las variables que afectan a la percepción de confianza y corrupción se mantienen y son importantes para estas variables dependientes. O por el contrario, algunas solo afectaron en cierto año. Además, inicialmente se comparan las instituciones por separado, las cuáles posteriormente con el uso del Análisis de Componentes Principales y el coeficiente Alpha de Cronbach, se muestra que las variables tienen una similitud en cuanto a la catindad de varianza explicada por un solo factor, esto generan índices para estudiar de manera general las variables dependientes. Por lo cual, también se da una breve explicación de su uso y estructura.

Todo esto, junto con el análisis del contexto en que las mediciones de interés se crean, aportan para hacer una evaluación final y proponer futuros trabajos que contemplen nuevas opciones de estudio.

2

Confianza en instituciones y los servicios públicos

2.1. Acerca de la confianza institucional en México

La confianza institucional es un elemento importante al momento de saber la capacidad de gobernar. Sin embargo, la facultad de entender cómo se compone esta confianza es realmente complicada, como señala del Tronco [9], hay distintos factores que pueden hacer que cambie, si a eso se añade que esta confianza tiende a cambiar de manera significativa a través del tiempo, nos hace pensar la relevancia de ciertos factores y su impacto.

Esta capacidad de gobernar se relaciona con el sentir de democracia. Del Tronco señala que la búsqueda de una democracia es un rasgo presente en la cultura mexicana, aunque ésta también depende de otros factores. Del Tronco plantea la visión de un México autoritario por su misma sociedad hasta antes de la transición del poder en el año 2000, puesto que en su análisis se muestra la sensación de un mayor acceso a la democracia a los ciudadanos. Esto podría añadir a la alternancia como un factor clave para la percepción de la democracia, pero no para la evaluación final del desempeño de estos mismos.

Moreno Jaimes [11], Córdova Guzmán y F. Ponce [10] explican más acerca de la relación que afectan a la calidad gubernamental que los ciudadanos perciben en sus gobiernos. De los cuales uno de estos factores que se citan es la efectividad, la cual está medida por la calidad de los servicios que provee el Estado. Otro de los factores que se marcan en relación con la calidad gubernamental es el manejo de las problemáticas sociales, que a la vez se ve reflejada en la transparencia. Moreno se centra en la relación que tiene la calidad gubernativa y la calidad de los servicios, sin embargo, también indaga en otros factores que podrían estar involucrados. Retomando a Del Tronco y la relación con la democracia, Moreno presenta parte de esta importancia de la democracia mediante el acceso que las mismas instituciones les dan a los ciudadanos.

Moreno hace un análisis con enfoque a los municipios mexicanos, por lo cual se centra en el caso local. Uno de los aspectos que varios autores consideran dentro de esta confianza que los ciudadanos ponen a las instituciones es su acceso a la posibilidad de participar de manera activa en la decisión de sus gobiernos. Sería de interés ver cómo eso afecta hoy en día al tener participación en ciertas decisiones del gobierno federal. De hecho, Moreno hace un pequeño análisis al respecto, enfocándose solo a la proporción de participación mediante distintos mecanismos.

En cuanto al tema de la transparencia, este factor podría ser relevante, es decir qué tanto acceso tienen los ciudadanos a la información de gastos que ejercen sus respectivos gobiernos. Aunque este factor podría ser aún más complicado de evaluar, el cual a su vez se relaciona con la corrupción. En este último Córdova y Ponce enfocan su estudio. Como se puede ver estos tres temas (corrupción, democracia y transparencia) se relacionan de manera estrecha.

Por lo cual, no es de extrañar que la percepción de la corrupción esté muy ligada a la calidad gubernamental que despliegue una institución. Moreno prefiere evaluar la calidad gubernamental a partir de la efectividad, es decir, de la calidad de los servicios públicos, pues considera que esto genera un mayor impacto en la calidad gubernamental que la percepción directa que se tiene mediante elecciones, este último considerado como otro factor dentro de la calidad del gobierno.

Por otro lado, Cordóva y Ponce hacen un análisis muy parecido al que se presenta en este trabajo, pero con un enfoque distinto. Ellos hacen un análisis pensando en cómo afecta la percepción de la corrupción en la calidad de los servicios públicos locales. El enfoque de estos autores parte de reactivos distintos, Cordóva y Ponce toman como punto de partida la evaluación en una escala 1 a 10 de los servicios públicos. Mientras tanto, en este trabajo, la calidad de dichos servicios se obtiene a partir de las carencias o no de ciertas características. En lo cual, se podría interpretar que la evaluación en una escala de 1 a 10 puede ser más subjetiva que señalando las características que los servicios tienen.

En el caso de la confianza institucional, es una variable que podría tener mayor volatilidad, y esto podría estar más presente en el caso de los municipios ya que, como Moreno señala, la cantidad de factores que están involucrados en esta evaluación son múltiples. Además de esto, Moreno describe que en ocasiones los municipios no obedecen a aspectos sumamente imparciales, sino que a aspectos clientelares, es decir, son municipios en los cuales figuras políticas tienen un interés particular, por lo cual se enfocan en satisfacer u ofrecer un gobierno con el cual los gobernados se sientan satisfechos. Esto no necesariamente implica que la gobernabilidad de dicho lugar sea legal o adecuada, pues también se tienen registros de que la inversión en infraestructura para los municipios y otras obras tienden a aumentar en años electorales.

2.2. La corrupción

Lo anterior lleva a tratar el tema de la corrupción. La corrupción se define a *grosso modo* como el uso ilegítimo del poder para obtener un beneficio. Este beneficio es más que solo un beneficio económico. Esto puede implicar, un beneficio político, como ascender a puestos, obtener respaldo de un grupo político o un puesto en el gobierno en turno. Y también los beneficios personales, como acceso a instituciones públicas, licitaciones directas, etc. Sin embargo, el formato más conocido es el beneficio financiero.

En este aspecto hay dos tipos de corrupción, la pequeña y la grande. Tal como lo dicen sus nombres, se caracterizan por el beneficio que se adquiere por dicho tipo de actos. En la gran corrupción se engloban aspectos como adquirir una licitación sobre una obra, implementar o ejecutar reformas o decisiones políticas en favor de un grupo de personas influyentes en el sector público o privado. Generalmente, este tipo de corrupción se relaciona a burócratas con una mayor jerarquía dentro de dichos grupos.

En contra parte, la corrupción pequeña se encuentra en burócratas los cuales son más difíciles de identificar, pues son de un rango menor, pero a su vez dichos actos obtienen beneficios considerablemente menores a los de la gran corrupción. Sin embargo, estos tipos de actos tienen una existencia en el día a día. Por ejemplo, los sobornos a policías para no ser multados o los pagos por negligencia administrativa que pueden presentarse en algunas instituciones para la adquisición obligada o acelerada de un trámite.

La manera en la que los ciudadanos pueden ver las consecuencias de actos de corrupción, es obteniendo servicios deficientes. Córdova y Ponce plantean esto desde el punto de vista de maximizar los recursos. Es decir, tratar de proveer una mejora en los servicios públicos. Retomando a Moreno, quien habla de una clientela electoral en los municipios y de una mayor inversión en fechas cercanas a las electorales, esto hace que la calificación de los ciudadanos a sus autoridades locales sean cambiantes de un momento a otro. Sin embargo, no hay algo claro con respecto a la corrupción.

Como se mostrará posteriormente, esto también recae en la complejidad para medir la corrupción. Particularmente, considero que la confianza institucional se ve más afectada por un aspecto sentimental en ese momento que la percepción de la corrupción, la cual podría pensarse con un poco más de detenimiento haciendo un análisis general en un intervalo de tiempo más amplio.

Por lo anterior, la calidad de los servicios y otros factores podrían ser más relevantes. Córdova y Ponce exponen que una deficiencia en los servicios, hace que el ciudadano piense que los recursos se están gastando de manera irresponsable. Esto también puede cambiar la preferencia electoral, lo que al final termina explicando el tema de la clientela. Sobretodo, en el caso de los gobiernos locales, ya que, como lo expresa Moreno, muchos gobiernos municipales no se ven capaces de crear grandes obras de infraestructura, ya sea por el costo que es elevado para las cuentas públicas o porque es un proyecto a largo plazo el cual podría verse afectado en una transición en el cargo.

Cordóva y Ponce también hacen un acercamiento al tema de la democracia, en qué manera se disponen los servicios públicos. Es decir, tomar en cuenta las necesidades de sus gobernados, considerándolos por algunos instrumentos de participación como mencionó Moreno: encuestas, consultas, reuniones con representantes vecinales o los mismos ciudadanos, etc. Esto con el objetivo de ver las mejoras más importantes e inmediatas que requieren sus ciudadanos.

Hasta este punto, solo se ha expuesto la manera en que la ciudadanía percibe la corrupción en sus instituciones. Sin embargo, la realidad puede diferir, el presente proyecto tiene un enfoque hacia este tema de acuerdo a los datos del INEGI, por lo cual, no se toma de un grupo específico de la sociedad como sí lo hacen otros índices que miden la corrupción, ya que lo toman desde el sector privado, que puede que tengan más conciencia y vivan más los actos de corrupción, sin embargo, sigue siendo un grupo específico, además, los datos del INEGI dan un panorama particular de lo que ocurre en México.

Medir la corrupción es complicado, ya que como se mencionó anteriormente hay pequeños actos y grandes actos de corrupción, asimismo, estos actos no necesariamente son monetarios, por lo cual, solo se puede estimar tomando en cuenta factores que podrían estar ligados al mismo.

La percepción, aunque no es la mejor medida, sí puede intuir el nivel real de la misma. Aunque el índice que genera Transparencia Internacional es una buena referencia, es un índice que se configura a partir de la percepción de un grupo específico de personas, además de que solo da un panorama general de la situación en México. De alguna manera el INEGI da más detalle acerca de lo que ocurre a nivel estatal.

Será de interés observar qué pasa en los estudios cuando se genera una transición en el poder y cuánto afecta ésta. Haciendo referencia a la transición a nivel federal. Tal vez hacer una comparación entre dos países, para ver qué tanto estos factores se mantienen sin importar el gobierno. Como menciona Moreno, parte de la percepción de confianza se debe a factores históricos. Recuerde que en algún momento se mencionó que la corrupción es “cultural”, y sin estar muy lejos, escuchar en ocasiones la frase “el que no tranza, no avanza”, nos da una visión de cómo la sociedad ve el tema. Más que ser un tema cultural, puede deberse a esos factores históricos que de cierta manera han ido normalizando un tema tan delicado.

Otros factores que pueden tener injerencia en la percepción de la corrupción podrían ser el mismo nivel de difusión o el número de investigaciones que se lleven a cabo a funcionarios por cargos de corrupción. El impacto y el tamaño de la misma, aunque el impacto puede ser otro aspecto complejo de medir. Puesto que los casos de la pequeña corrupción no son tan conocidos, o al menos sus autores son más difíciles de reconocer y por ende de condenar.

Este tipo de factores son los que podrían afectar la percepción, por lo cual faltaría obtener información al respecto, en el caso particular de México, información por Estado. Lo cual podría considerar un tipo diferente de análisis, es decir, uno por tiempo considerando un intervalo en el que se cuente cuántas de estas investigaciones se han difundido o procesado, así como el tiempo de exposición en los medios.

Finalmente, como se menciona esto sería un acercamiento hacia lo que pasa realmente respecto al tema. En caso de ser realmente relevantes los factores anteriormente mencionados, significaría que no importa si los actos de corrupción son pequeños o grandes. Es decir, no importaría la cantidad del beneficio que se lleven los actores de dicho acto ilícito, sino su exposición ante el público.

Estos son solo algunos factores que se podrían agregar al estudio del tema, que a su vez considero que está relacionado a la transparencia, lo cual algunos estudios ya consideran, a partir de la medición de otros factores que generan una variable de grado de transparencia. A continuación, el presente trabajo muestra solo una forma de estudiar la confianza en las instituciones y el nivel de corrupción que se percibe en ellas mediante el modelo multinivel, lo cual valdría la pena considerar en análisis a nivel internacional, ya que, como se señala, esto puede depender de distintos contextos. Incluso, usando más que solo el nivel de países o estados.

3

Preliminares

Antes de comenzar con el tema central del presente trabajo, se introducirán algunos temas que serán de importancia al momento de realizar el análisis. Los cuales serán el análisis de componentes principales, ACP o PCA (por sus siglas en inglés), y el Alpha de Cronbach. Estos temas tienen el objetivo de mostrar la relación que tienen variables explicativas y su posible redundancia a la hora de generar el modelo final.

3.1. Análisis de componentes principales

El objetivo del ACP es reducir las dimensiones de las observaciones, es decir, reducir el número de variables que representan cada una de las muestras sin afectar la información que se tiene de las mismas. La manera en que este análisis reduce las variables no es solo tomar una de manera aleatoria, tampoco tomarlas y darles la misma importancia; sino que lo que hace es estudiar el peso promedio, denotado de la siguiente manera:

$$\delta' X = \sum_{j=1}^{p} \delta_j x_j \quad \text{tal que} \quad \sum_{j=1}^{p} \delta_j^2 = 1, \tag{3.1}$$

el cual se denomina combinación lineal estandarizada con $\delta = (\delta_1, \delta_2, \ldots, \delta_p)^T$, el vector de pesos , mientras que la matriz X es el conjunto de n observaciones con p dimensiones, es decir, una matriz de tamaño $n \times p$ y en donde x_j es un vector con las entradas x_{ij}, para $i = 1, \ldots, n$, de la matriz X. Esto tiene el objetivo de maximizar la varianza de la proyección $\delta' X$, es decir, δ tal que:

$$\max_{\{\delta:||\delta||=1\}} Var(\delta' X) = \max_{\{\delta:||\delta||=1\}} \delta' Var(X) \delta. \tag{3.2}$$

Las direcciones de δ se encuentran mediante la descomposición espectral de la matriz de covarianzas, de la cual se obtiene la dirección δ que está dada por el vector γ_1 correspondiente al valor propio más grande λ_1 de la matriz de covarianzas $\Sigma = Var(X)$, de esta manera se genera la primer componente dada por $y_1 = \gamma_1' X$. Por otro lado, una dirección ortogonal a γ_1 describe la segunda componente principal, esta componente principal está asociada al segundo valor propio más grande λ_2, con su respectivo vector propio γ_2 que es ortogonal a γ_1, de tal manera que $y_2 = \gamma_2' X$. Este proceso se sigue hasta generar p componentes principales, generando vectores ortogonales a partir de la combinaciones lineales que maximizan la varianza [6] (págs: 216-217), los nuevos valores generados $Y = (y_1, y_2, \ldots, y_p)$ representan las nuevas coordenadas de las muestras generadas por los factores, de esta manera Y es una matriz de tamaño $n \times p$, mientras que el conjunto de $(\gamma_1, \gamma_2, \ldots, \gamma_p)$ conforman una matriz Γ. Finalmente, la matriz Λ es una matriz diagonal con los valores propios $(\lambda_1, \lambda_2, \ldots, \lambda_p)$ como elementos en dicha diagonal.

De esta manera, para una variable aleatoria X con media $E(X) = \mu$ y varianza en su descomposición espectral $Var(X) = \Sigma = \Gamma\Lambda\Gamma'$, dicha descomposición existe por el Teorema de la Descomposición y en este caso es Γ ya que Γ y Γ' son ortogonales. Esta variable tiene la transformación de componentes principales:

$$Y = \Gamma'(X - \mu). \tag{3.3}$$

Con lo cual, a partir de la matriz X se obtiene una nueva matriz Y cuyas componentes principales son de media cero.

Dada estas definiciones se pueden probar las siguientes propiedades:

$$\begin{aligned}
&E[Y_j] = 0 \quad j = 1, \ldots, p \\
&Var(Y_j) = \lambda_j, \quad j = 1, \ldots, p \\
&Cov(Y_i, Y_j) = 0, \quad i \neq j \\
&Var(Y_1) \geq Var(Y_2) \geq \ldots \geq Var(Y_p) \geq 0 \\
&\sum_{j=1}^{p} Var(Y_j) = tr(\Sigma) \\
&\prod_{i=1}^{p} Var(Y_j) = |\Sigma|.
\end{aligned}$$

La prueba de estas propiedades se encuentran en el libro sobre análisis multivariado de Härdle [6] (pág: 218). De manera general, lo que muestran estas propiedades es que los valores propios están ordenados de tal manera que los primeros son los más grandes y los que más porcentaje de varianza explican, por ende los más relevantes, pues explican de mayor manera el comportamiento de los datos. Además que las variables resultantes de Y no están relacionadas, esto es, son independientes.

La idea principal de este análisis es generar nuevas variables que parten de las proyecciones de las originales maximizando el porcentaje de varianza explicada, los valores propios más grandes indicarían que capturan más porcentaje de dicha variación a partir de sus vectores propios asociados. Esto se puede ver mediante la proporción relativa de la varianza explicada por las primeras q componentes principales, lo cual se hace de la siguiente manera:

$$\psi = \frac{\sum_{j=1}^{q} \lambda_j}{\sum_{j=1}^{p} \lambda_j} = \frac{\sum_{j=1}^{q} Var(Y_j)}{\sum_{j=1}^{p} Var(Y_j)}.$$

Usando las propiedades mostradas anteriormente y dado que la multiplicación de dos matrices ortogonales es la identidad, entonces $\Gamma\Gamma' = I$. Por ende, la covarianza entre el nuevo vector generado Y y el original X es:

$$\begin{aligned} Cov(X,Y) &= E[XY'] - E[X]E[Y'] = E[XY'] = E[X(X'\Gamma - \mu'\Gamma)] \\ &= E[XX'\Gamma] - E[X\mu'\Gamma] = E[XX']\Gamma - \mu\mu'\Gamma = Var(X)\Gamma \\ &= \Sigma\Gamma = \Gamma\Lambda\Gamma'\Gamma = \Gamma\Lambda, \end{aligned}$$

con un coeficiente de correlación entre ambas variables dado por:

$$\rho_{X_iY_j} = \frac{\gamma_{ij}\lambda_j}{(\sigma_{X_iX_i}\lambda_j)^{1/2}} = \gamma_{ij}\left(\frac{\lambda_j}{\sigma_{X_iX_i}}\right)^{1/2}.$$

De esta relación se puede extraer un segunda manera de evaluar la proporción de la varianza explicada por las componentes principales, pues vea la siguiente igualdad:

$$\sum_{j=1}^{p} \rho^2_{X_iY_j} = \frac{\sum_{j=1}^{p} \gamma'_{ij}\lambda_j\gamma_{ij}}{\sigma_{X_iX_i}} = \frac{\gamma'_i\lambda\gamma_i}{\sigma_{X_iX_i}} = \frac{\sigma_{X_iX_i}}{\sigma_{X_iX_i}} = 1. \tag{3.4}$$

De la ecuación (3.4), se puede deducir que $\rho^2_{X_iY_j}$ se puede apreciar como la proporción de la varianza explicada de X_i por Y_j.

Algunas propiedades asintóticas de las componentes principales ayudan a construir un intervalo de confianza para los valores propios de las componentes. Para ello primero se introducen algunas definiciones, se denominan a las matrices $\mathcal{G}$ y $\mathcal{L}$, como la estimación de las matrices Γ y Λ, respectivamente. Dichas matrices están conformadas por elementos $(\ell_1, \ell_2, \ldots, \ell_p)$ y $(g_1, g_2, \ldots, g_p)$, respectivamente. Por otra parte, se define a $W_p(\Sigma, m)$ como la distribución Wishart con matriz de varianzas Σ, de p variables y m grados de libertad. De aquí se desprenden las siguientes propiedades:

Teorema 1 *Dada la matriz $\Sigma > 0$, es decir, positiva definida con vectores propios distintos y sea $U \sim m^{-1}W_p(\Sigma, m)$ con descomposiciones espectrales $\Sigma = \Gamma\Lambda\Gamma^T$ y $U = \mathcal{GLG}$. Entonces:*

(a) $\sqrt{m}(\ell - \lambda) \xrightarrow{\mathcal{L}} N_p(0, 2\lambda^2)$ *donde* $\ell = (\ell_1, \ell_2, \ldots, \ell_p)'$ *y* $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_p)'$ *son diagonales de* $\mathcal{L}$ *y* Λ *respectivamente, es decir,* $\sqrt{m}(\ell - \lambda)$ *converge asintóticamente a* $N_p(0, 2\lambda^2)$*,*

(b) $\sqrt{m}(g_j - \gamma_j) \xrightarrow{\mathcal{L}} N_p(0, \mathcal{V}_j)$*, con* $\mathcal{V}_j = \lambda_j \sum\limits_{k \neq j} \frac{\lambda_k}{(\lambda_k - \lambda_j)^2} \gamma_k \gamma_k^T$*,*

(c) $Cov(g_j, g_k) = \mathcal{V}_{jk}$*, donde el elemento* (r, s) *de la matriz* $\mathcal{V}_{jk}(p \times p)$ *es* $-\frac{\lambda_j \lambda_k \gamma_{rk} \gamma_{sj}}{m(\lambda_j - \lambda_k)^2}$*,*

(d) Los elementos de ℓ *son asintóticamente independientes de los elementos de* $\mathcal{G}$*.*

Si $X_1, \ldots, X_p$ están descritas de la forma $N(0, \mu)$, entonces:

$$\sqrt{n-1}(\ell - \lambda) \xrightarrow{\mathcal{L}} N_p(0, 2\lambda^2),$$

donde $N_p(0, 2\lambda^2)$ es la distribución normal multiple de p variables aleatorias. Además, si se considera $f(\ell) = \log(\ell)$, entonces:

$$\sqrt{n-1}(\log(\ell) - \log(\lambda)) \longrightarrow N_p(0, 2),$$

esto dado que se tiene un teorema correspondiente a las tranformaciones estadísticas.

Teorema 2 *Si* $\sqrt{n}(t - \mu) \xrightarrow{\mathcal{L}} N_p(0, \Sigma)$ *y si* $f = (f_1, f_2, \ldots, f_q)' : \mathbb{R}^p \longrightarrow \mathbb{R}^q$*, son funciones con valores en los reales y diferenciables en* $\mu \in \mathbb{R}^p$*, entonces* $f(t)$ *es normal de manera asintótica con media* $f(\mu)$ *y covarianzas* $\mathcal{D}'\Sigma\mathcal{D}$*, es decir:*

$$\sqrt{n}\{f(t) - f(\mu)\} \xrightarrow{\mathcal{L}} N_q(0, \mathcal{D}'\Sigma\mathcal{D}) \quad para \quad n \longrightarrow \infty$$

donde,

$$\mathcal{D} = \left(\frac{\partial f_j}{\partial t_i}\right)(t)\Big|_{t=\mu}$$

Luego,

$$\sqrt{\frac{n-1}{2}}(\log(\ell) - \log(\lambda)) \xrightarrow{\mathcal{L}} N_p(0, 1),$$

de esta manera, para un nivel de confianza $\alpha = 0.95$, un intervalo de confianza para el j-ésimo valor propio es:

$$\left(\ell_j \exp\left(-1.96\sqrt{\frac{2}{n-1}}\right), \ell_j \exp\left(1.96\sqrt{\frac{2}{n-1}}\right)\right).$$

Retomando la manera en que se calcula la varianza explicada por las primeras q componentes principales, de lo cual en la práctica se obtiene:

$$\widehat{\psi} = \frac{\sum_{i=1}^{q} \ell_i}{\sum_{i=1}^{p} \ell_i}.$$

Además, usando el mismo teorema se puede obtener que:

$$\sqrt{n-1}(\widehat{\psi} - \psi) \xrightarrow{\mathcal{L}} N(0, \mathcal{D}'\mathcal{V}\mathcal{D}),$$

donde $\mathcal{V} = 2\Lambda^2$ y $\mathcal{D} = (d_1, \ldots, d_p)'$ con:

$$d_j = \frac{\partial \psi}{\partial \lambda_j} \begin{cases} \frac{1-\psi}{tr(\Sigma)} & \text{para } 1 \leq j \leq q, \\ \frac{-\psi}{tr(\Sigma)} & \text{para } q+1 \leq j \leq p. \end{cases}$$

Con este resultado, se puede generar un tercer teorema.

Teorema 3

$$\sqrt{n-1}(\widehat{\psi} - \psi) \xrightarrow{\mathcal{L}} N(0, \omega^2),$$

donde

$$\omega^2 = \mathcal{D}'\mathcal{V}\mathcal{D} = \frac{2}{\{tr(\Sigma)\}^2}\{(1-\psi^2)\sum_{i=1}^{q}\lambda_i^2 + \psi^2 \sum_{i=q+1}^{p}\lambda_i^2\}\frac{2tr(\Sigma^2)}{\{tr(\Sigma)\}^2}(\psi^2 - 2\beta\psi - \beta)$$

y

$$\beta = \frac{\sum_{i=1}^{q} \lambda_i^2}{\sum_{i=1}^{p} \lambda_i^2}.$$

Lo cual lleva a estimar un intervalo de confianza para la varianza explicada por las primeras q componentes principales a un nivel $\alpha = 0.95$:

$$\left(\widehat{\psi} - 1.96\sqrt{\frac{\frac{2tr(S^2)}{\{tr(S)\}^2}(\widehat{\psi}^2 - 2\widehat{\beta}\widehat{\psi} + \widehat{\beta})}{n-1}}, \widehat{\psi} + 1.96\sqrt{\frac{\frac{2tr(S^2)}{\{tr(S)\}^2}(\widehat{\psi}^2 - 2\widehat{\beta}\widehat{\psi} + \widehat{\beta})}{n-1}}\right).$$

Para más detalles de cómo se obtienen estos intervalos de confianza, se recomienda consultar a Härdle [6] (pág: 227).

En ocasiones los datos son heterogéneos al tener distintas escalas de medición (kg, cm, km, volumen, etc.) por lo cual, es necesario estandarizar los datos. Para ello, se aplica una transformación a los datos determinada de la siguiente manera:

$$\mathcal{X}_S = \mathcal{H}\mathcal{X}\mathcal{D}^{1/2},$$

donde $\mathcal{D} = diag(s_{X_1X_1}, s_{X_1X_2}, \ldots, s_{X_pX_p})'$. Note que la variable x_S tiene que $\overline{x}_S = 0$ y $S_{\mathcal{X}_S} = \mathcal{R}$, la matriz de correlación. Las matrices transformadas de las componentes principales de $\mathcal{X}_S$ son comúnmente llamadas *Componentes Principales Normalizadas (NCPs, por sus siglas en inglés)*. La descomposición espectral de dicha matriz es:

$$\mathcal{R} = \mathcal{G}_{\mathcal{R}}\mathcal{L}_{\mathcal{R}}\mathcal{G}'_{\mathcal{R}},$$

donde $\mathcal{L}_{\mathcal{R}} = diag(\ell_1^{\mathcal{R}}, \ldots, \ell_p^{\mathcal{R}})$ y $\ell_1^{\mathcal{R}} \geq \ldots \geq \ell_p^{\mathcal{R}}$ son los valores propios de $\mathcal{R}$ relacionados a los vectores propios $g_1^{\mathcal{R}}, \ldots, g_p^{\mathcal{R}}$.

Los NCPs proveen una representación individual dada por:

$$Z = \mathcal{X}_{\mathcal{R}}\mathcal{G}_{\mathcal{R}} = (z_1, \ldots, z_p).$$

Las NCPs dan una perspectiva similar a los ACPs, pero con posiciones relativas de los individuos, puesto que en este caso todas las variables tienen el mismo peso. En ocasiones este tipo de análisis es mejor, sobretodo cuando se quiere que las variables se piensen con la misma importancia, más información acerca de este tipo de análisis, así como propiedades se encuentran en Härdle [6].

3.2. Alpha de Cronbach

El coeficiente alpha de Cronbach es un valor estadístico muy bien conocido en el ámbito de la Psicología, este coeficiente se caracteriza por varias razones, una de ellas es que es un valor con características relacionadas a su confiabilidad, es decir el porcentaje de variabilidad que engloba varios item o reactivos, en este caso, dichos items son variables bianrias que formarán una variable latente. El valor alpha se calcula de la siguiente forma:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum_{i=1}^{n} V_i}{V_t}\right),$$

donde V_t es la varianza de todos los valores observados, mientras V_i es la varianza del item i. En el caso donde las variables sean binarias, Cronbach muestra que esta fórmula se puede simplificar a:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum_{i=1}^{n} p_i q_i}{\sigma_t^2}\right),$$

donde p_i representa la proporción de valores 1 para el item i, q_i, la proporción de valores distintos de 1 y σ_t^2 es la varianza total. De esta manera, α representa de manera general, el promedio de todos los posibles coeficientes que se contemplen como parte de la medición de un mismo fénomeno.

En el artículo de Cronbach [7] se encuentra que mediante una serie de igualdades, se obtiene la siguiente expresión la cual tiene una interpretación más sencilla:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum_j \sum_i^n C_{ij}}{V_t}\right) \quad (para \quad i, j = 1, 2, \ldots, n; i \neq j)$$

donde C_{ij} es la covarianza entre los items i y j. De aquí, se puede observar que α evalúa la relación entre los items, por ende se usa como una medida para estimar qué tan confiable es agrupar un conjunto de items en una sola variable.

Al ser un valor que estima la relación entre las variables, lo que se espera es que este valor esté próximo a 1 si las variables están altamente correlacionadas. En un caso contrario, es posible obtener valores negativos, lo cual indicaría que habrá variables las cuales estén relacionadas de manera negativa.

Cronbach describe este coeficiente como “la proporción de la variable atribuible entre los factores comunes de los items”, es decir, un índice de concentración común.

3.3. Instrumentos de medición

La utilización al Alpha de Cronbach provee la confiabilidad de un instrumento de medición, es decir, da a saber si se puede confiar que una nueva medida creada por un conjunto de variables podrá resumir esta información de manera adecuada sin que se tenga que estudiar cada uno de estos reactivos, esto se hace suponiendo que las variables sean consistentes, es decir, que formen parte de un mismo concepto y vayan en la misma dirección en cuanto a sus valores.

A menudo lo que se usa son las escalas y los índices, para conocer de manera más profunda estos conceptos, consulte Montero [8]. En el presente trabajo solo se presenta el índice.

Inicialmente, la conformación de un índice puede ser el consenso de varios factores que no necesariamente tienen la misma naturaleza. Como se verá posteriormente, los índices en este proyecto son sobre la percepción de la calidad de los servicios públicos locales, tomando de referencia la calidad de cada uno de esos servicios, tales como el agua, el drenaje, el señalamiento, las policías, etc. Esto conlleva al índice de los servicios públicos. Pero así como dicho índice, se plantea la construcción de otros, como uno que mida la confianza institucional o la percepción de la corrupción institucional, a partir de un conjunto de items que evalúan la percepción de la corrupción en instituciones específicas; por ejemplo: la percepción en el gobierno federal, en el estatal, en el municipal, en los jueces, etc.

De aquí, es importante destacar que la conformación de índices no se da solo por obtener la relación de una prueba que indique que las variables tienen una relación significativa para poder crear un índice, sino que también se debe de mantener una cierta relación en el marco teórico. Por lo cual, a pesar de que la educación, longevidad y poder adquisitivo pueden tener una naturaleza distante de manera subjetiva, éstos pueden analizarse de manera conjunta para entender la capacidad de un individuo para obtener un buen desarrollo dentro del entorno donde se desenvuelve.

La medida de confiabilidad más usada para este tipo de casos es el coeficiente *alpha de Cronbach*. En adición a esto, para los reactivos binarios es usual aceptar que cada uno de dichos reactivos tienen el mismo peso, por lo cual, la medida compuesta (el índice) generada es solo una suma simple de los puntajes de los reactivos.

Sin embargo, el implementar otro tipo de ponderación para cada uno de los reactivos no suele ser una tarea sencilla, pues puede partir de la información en el marco teórico y los juicios de especialistas en el tema. Otro tipo de medida compuesta usual es el estandarizar las medidas, en el caso binario, éstas ya lo están. Sin embargo, en el caso de los porcentajes, tasas o años transucurridos, estos datos se les haría una transformación para que las variables tengan la misma variación y posteriormente generar una suma asumiendo el mismo peso.

4

Modelos multinivel

4.1. Modelo de regresión lineal

Antes de adentrarse de lleno a este tipo de modelos hace falta recordar lo que son los modelos de regresión. Inicialmente, se puede pensar en el modelo de regresión lineal. El cual supone un conjunto Y de k elementos y a su vez un conjunto de p arreglos con k elementos cada uno, de tal manera que se tiene una matriz X de dimensión $(k \times p)$. A estos elementos se les demonina variable dependiente y el conjunto de variables independientes, Y y X, respectivamente. Además, frecuentemente se le suele agregar a X un arreglo más que indica la constante en el modelo, el cual representa el valor estimado cuando las variables independientes valen 0, el arreglo que se agrega a dicha matriz contiene a 1 como elemento en cada una de sus entradas, es decir:

$$x_0 = (\overbrace{1, \ldots .1}^{k\ veces})$$

.

De esta manera, se puede diseñar un modelo para observar si existe una relación entre el conjunto Y y cada una de las variables X, $x_0, x_1, x_2, \ldots, x_p$. Dicho modelo es:

$$Y = X\beta + \epsilon,$$

donde β es un vector de tamaño $p+1$, donde cada β_j es una constante que indica la relación que tiene cada variable x_{ij} con la variable dependiente y_i, es decir, los coeficientes de regresión del modelo. Además, ϵ es un vector de errores, donde cada $\epsilon_i \sim N(0, \sigma^2)$ para $i = 1, \ldots, k$. De esta manera, el mismo modelo se puede escribir de la forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i \quad para \quad i = 1, \ldots, k.$$

En este caso $x_{i0} = 1$ por lo cual $\beta_0 x_{i0} = \beta_0$. Los valores de β son desconocidos inicialmente, sin embargo, pueden ser estimados, la manera más común de estimarlos es minimizando la suma de errores cuadrados, $\sum_{i=1}^{k}(y_i - X_i'\beta)^2$. De esta manera, la estimación para estos coeficientes está dada por $\widehat{\beta} = (X'X)^{-1}X'Y$, con una matriz de varianzas y covarianzas igual a $\sigma^2(X'X)^{-1}$.

El modelo genera estimaciones que no siempre representan el comportamiento de los datos, por lo cual se genera una diferencia entre el valor estimado del modelo y el valor real, dicha diferencia es llamada *residuo*, $r_i = y_i - X_i'\widehat{\beta}$, del cual se puede decir que tiene media 0 y desviación estándar igual a $\widehat{\sigma} = \sqrt{\sum_{i=1}^{k} r_i^2/(k-p)}$.

Una vez estimado este modelo, se puede verificar la bondad de ajuste, la cual mide qué tan buenos son los regresores de X para ajustarse al modelo y hacer una buena estimación de Y, esto se hace mediante la relación $R^2 = 1 - \widehat{\sigma}^2/s_y$, y s_y es la desviación estándar de los datos de Y. Si nuestro modelo es bueno, $R^2 = 1$ o R^2 sería cercano a 1, por el contrario el modelo va a ser deficiente si este tiende a 0. Esto da paso a que se tenga que implementar un modelo que mejore esta representación. He ahí donde surgen transformaciones de los datos, tales como variables polinómicas, exponenciales u otra transformación de tal manera que se le pueda asociar un modelo mejor:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_t x_{it}^t + \ldots + \beta_p x_{ip}^p + \epsilon_i \qquad \text{para algún} \quad t \in \{1, \ldots, p\}.$$

El modelo anterior es llamado polinomial, por otro lado el siguiente modelo es logarítmico:

$$\log(y_i) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \epsilon_i.$$

Las estimaciones para estos modelos se siguen obteniendo de la misma manera, y por tanto la información que nos da es la misma, más no así la interpretación, ya que en relación a los valores de las β's esta relación cambia. En el ejemplo del modelo logarítmico, se puede decir que por cada incremento que tenga la variable x_{1i}, el valor de y_i incrementa exponencialmente, en proporción e^{β_1}. A este tipo de transformaciones se les denomina transformaciones lineales, acerca de este tema puede encontrarse un poco más en el libro de Gelman y Hill [1].

Como se puede observar hay distintas maneras de medir el efecto de varios factores en una variable. Lo más usado en estos casos son los modelos de regresión. Aún más, los modelos de regresión simple y múltiple. Sin embargo, éstos se caracterizan principalmente por generar una relación lineal entre la variable que se estudia y sus posibles factores que la explican.

Otras opciones que se pueden considerar son aquellas que en vez de tener una relación lineal, tienen una relación de tipo polinomial. Sin embargo, esto también puede depender del tipo de estructura que exista en los datos que se están estudiando.

Para el caso en donde estos datos se dividen por categorías, tienen un trato distinto, en el cual, la estimación va cambiando de acuerdo al grupo que se está evaluando. Regularmente se generan nuevas variables a las cuales se les denomina variables *ficticias* o *dummy*, estas nuevas variables son binarias por cada nivel, es decir que a partir de cada valor que puede tomar la variable asociada a β_j se le crea una nueva variable que es 1, si toma el valor de dicho nivel o 0 si no lo es, esto quiere decir:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_{j_1} x_{ij_1} + \ldots + \beta_{j_{s-1}} x_{ij_{s-1}} + \ldots + \beta_p x_{ip} + \epsilon_i \text{ para algún } j \in \{1, \ldots, p\},$$

donde $x_{ij_l} = 1$ si $x_{ij} = l$, en caso contrario $x_{ij_l} = 0$, para $l \in \{1, \ldots, s-1\}$ y s es el número de categorías que tiene la variable j. Además, $\beta_{j_0} = 0$, en lo que se denomina la categoría base o de referencia.

4.2. El modelo multinivel

Una visión distinta con la que se puede tratar este tipo de información, son con los modelos *multinivel*, los cuales generan una relación con las demás variables pero estima el efecto que tiene al localizarse en cierto grupo. Este tipo de modelos reciben varios nombres tales como *modelos con efectos aleatorios*, *de efectos mixtos* o *modelos jerárquicos*. Lo cual, está justificado, son modelos con *efectos aleatorios*, ya que como se verá posteriormente, la descripción del modelo en el cual los coeficientes del modelo de regresión se consideran aleatorios. En este mismo sentido, va el nombre de modelos de *efectos mixtos*, para este caso se considera que hay coeficientes que dependen de la aleatoriedad y otros que se mantienen fijos. Finalmente, los modelos *jerárquicos* hacen referencia a que el modelo está clasificado en niveles según la categoría que puede tener una o más variables, siendo posible ordenarlas, en ese mismo sentido, podemos ver el porqué del modelo *multinivel*.

En comparación con el modelo de regresión clásico, el cual toma una categoría base, este modelo toma en cuenta cada una de las categorías, como si hiciera una regresión por cada conjunto de base de datos, que tiene a la población contenida en dicha categoría.

Considere que se quiere observar los efectos de una variable, x, en cada una de las c categorías que toma con respecto al valor que registra la variable respuesta. En este caso se dice que el modelo varía el intercepto, dicho modelo se escribe de la siguiente manera:

$$y_{ij} = \alpha_j + \beta x_{ij} + \epsilon_{ij} \quad \ldots \quad \text{para} \quad i = 1, \ldots, k \quad \text{y} \quad j = 1, \ldots, c,$$

donde $\epsilon_{ij} \sim N(0, \sigma^2_{\epsilon_{ij}})$. Este es el caso simple de un modelo multinivel en el cual la variable α_j cambia de acuerdo a la categoría j de la muestra i (subíndice ij). De este modo se puede reescribir a la variable $\alpha_j = \beta_0 + u_{0j}$, donde el error $u_{0j} \sim N(0, \sigma^2_{u_{0j}})$.

Como se puede observar, u_{0j} es una variable que tiene una distribución de probabilidad, lo cual, cobra sentido al llamarlo modelo con *efectos aleatorios*.

Ahora, suponga el caso en el cual, lo que varía no es el intercepto sino el valor de una variable de acuerdo al grupo que representa. Es decir, varía una pendiente del modelo de regresión, entonces, se tiene un modelo como:

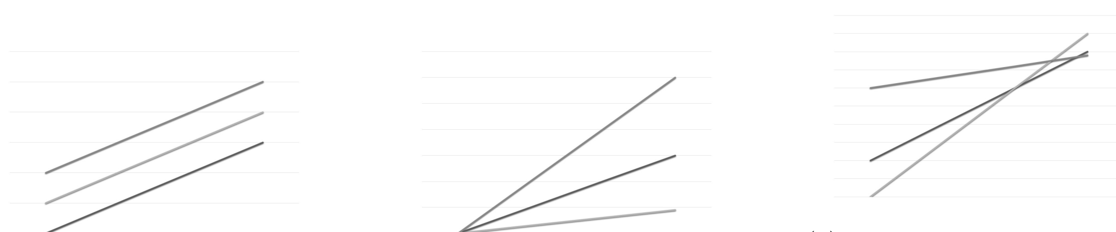
$$y_{ij} = \alpha + \beta_j x_{ij} + \epsilon_{ij} \qquad \ldots \text{ para } i = 1, \ldots, k \quad \text{y} \quad j = 1, \ldots, c.$$

En este caso se sigue suponiendo $\epsilon_{ij} \sim N(0, \sigma^2_{\epsilon_{ij}})$. Pero ahora la variable con efecto aleatorio no se localiza en el intercepto, de este modo la variable β_j genera que la pendiente del modelo se modifique según la categoría en la que se encuentra la muestra i. En consecuencia, se puede ver a la variable de la siguiente manera $\beta_j = \beta_1 + u_{1j}$, donde $u_{1j} \sim N(0, \sigma^2_{u_{1j}})$.

El modelo multinivel se puede extender aún más a un modelo en el cual tanto el intercepto como una de las pendientes pueden cambiar de acuerdo a la categoría en que esté la muestra, de esta manera y haciendo el cambio de variables que se ha propuesto, el modelo queda de la forma:

$$y_{ij} = \beta_0 + \beta_1 x_i + (u_{0j} + u_{1j} x_{ij} + \epsilon_{ij}). \tag{4.1}$$

La figura 4.1 muestra un ejemplo de lo que genera un modelo multinivel para un intercepto o una pendiente.



(a) Modelo de regresión en el cual se varía el intercepto (intercepto aleatorio), como se observa, es como si se desplazara la recta hacia arriba o abajo.

(b) Modelo de regresión en el cual se varía la pendiente (pendiente aleatoria), como se observa, es como si se desplazara la recta e hiciera una rotación.

(c) Modelo de regresión en el cual se varían tanto el intercepto como la pendiente (intercepto y pendiente aleatorios), como se observa, en este caso las rectas tienen distintas direcciones.

Figura 4.1: Tipos de modelos multinivel.

En donde, $u_{0j} \sim N(0, \sigma^2_{u_{0j}})$, $u_{1j} \sim N(0, \sigma^2_{u_{1j}})$ y $\epsilon_{ij} \sim N(0, \sigma^2_{\epsilon})$, son errores aleatorios del modelo y son independientes entre ellos. Considerando un modelo general y que contenga además variables fijas, este modelo se expresa:

$$y_{ij} = \beta_0 + \beta_1 x_{1i} + \sum_{h=2}^{p} \beta_h x_{hi} + (u_{0j} + u_{1j} x_{1i} + \epsilon_{ij}), \tag{4.2}$$

o escrito de otra manera:

$$y_{ij} = X_i\beta + \sum_{h=0}^{1} u_{hj} z_{hi} + \epsilon_{ij}, \tag{4.3}$$

donde $z = \{z_0, z_1\}$ y z_0 es un vector con unos, mientras z_1 es un vector de los datos de la primer variable, de la cual se estimarán los efectos aleatorios.

Al primer modelo, donde solo hay un intercepto, se le suele llamar modelo nulo, pues solo se toma en cuenta el intercepto, el efecto aleatorio de este intercepto y el error. Si bien, este modelo no aporta mucho, sí ofrece lo que se conoce como la *correlación intra-clase*, que se define como:

$$\rho = \frac{Var(u_{0j})}{Var(u_{0j}) + Var(\epsilon_{ij})},$$

dicho coeficiente indicaría la proporción de la varianza que es explicada por cada grupo j con respecto a la varianza total de las observaciones.

Haciendo el modelo de manera más general, podemos tener un modelo como el siguiente:

$$y_{ij} = X_i\beta + Z_i u_i + \epsilon_{ij}, \tag{4.4}$$

donde se definen:

$$Z_i = \left(z_i^{(1)} z_i^{(2)} \dots z_i^{(q)}\right) = \begin{bmatrix} z_{i1}^{(1)} & z_{i1}^{(2)} & \dots & z_{i1}^{(q)} \\ z_{i2}^{(1)} & z_{i2}^{(2)} & \dots & z_{i2}^{(q)} \\ \vdots & \vdots & \ddots & \vdots \\ z_{in_i}^{(1)} & z_{in_i}^{(2)} & \dots & z_{in_i}^{(q)} \end{bmatrix} \quad y \quad u_i = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iq} \end{bmatrix}.$$

Y en la cual los vectores $u_i \sim N_q(0, \mathcal{D})$ y $\epsilon_i \sim N_{n_i}(0, \mathcal{R}_i)$, en donde estos errores son independientes, es decir, $\epsilon_i \perp u_i$.

Y a su vez se especifica que:

$$\mathcal{D} = \sigma^2 D \quad y \quad \mathcal{R}_i = \sigma^2 R_i.$$

Donde $\mathcal{D}$ y $\mathcal{R}_i$ son matrices definidas positivas. Además, la matriz Z_i es un subconjunto de columnas seleccionadas apropiadamente de X_i, dicho subconjunto se selecciona de las variables que se consideran en los efectos aleatorios.

Tomando esta información se pueden generar las siguientes definiciones: $y = (y_1', y_2', \dots, y_N')'$ que contiene a todos los valores observados de las variables independientes, $n = \sum_{i=1}^{N} n_i$. De manera similar, $u = (u_1', u_2', \dots, u_N')'$ y $\epsilon = (\epsilon_1', \epsilon_2', \dots, \epsilon_N')'$. Finalmente se tiene:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \quad y \quad Z = \begin{bmatrix} Z_1 & 0 & \dots & 0 \\ 0 & Z_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Z_N \end{bmatrix},$$

donde 0 es una matriz que solo contiene valores 0, en sus entradas y las matrices X y Z de tamaños $n \times p$ y $n \times Nq$ respectivamente.

De esta manera (4.4) puede escribirse como:

$$y = X\beta + Zu + \epsilon, \tag{4.5}$$

con

$$u \sim N_{Nq}(0, \sigma^2 \mathbf{D}) \quad y \quad \epsilon \sim N_n(0, \sigma^2 \mathbf{R}),$$

donde

$$\mathbf{D} = \begin{bmatrix} D & 0 & \dots & 0 \\ 0 & D & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & D \end{bmatrix} \quad y \quad \mathbf{R} = \begin{bmatrix} R_1 & 0 & \dots & 0 \\ 0 & R_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & R_N \end{bmatrix}.$$

4.3. Estimación de parámetros

Se sabe que el vector de efectos aleatorios tiene la matriz de varianzas y covarianzas $\mathcal{D}$, mientras que su media es 0. Tomando esto en cuenta, se tiene:

$$\mathcal{D}(\sigma^2, \theta_D) = \sigma^2 D(\theta_D), \tag{4.6}$$

donde θ_D es un vector de parámetros que representan los elementos de la varianza y covarianza de u. Algo que no es necesario, es que cualesquiera dos elementos de u estén correlacionados. De esta forma, $\mathcal{D}$ tiene $q(q+1)/2$ elementos diferentes, correspondientes a las q varianzas y $q(q-1)/2$ covarianzas de u, en consecuencia, θ_D tiene $q(q+1)/2$ elementos distintos.

Finalmente se puede definir a la función de *varianzas independiente* como sigue:

$$Var(\epsilon_i \mid u_i) = Var(\epsilon_i) = \sigma^2 R_i(\theta_R; v_i), \tag{4.7}$$

esto último dado que se supone a ϵ_i independiente de u_i, de esta manera, $\mathcal{R}_i = \sigma^2 R_i(\theta_R; v_i)$ donde θ_R es un vector de parámetros de la varianza y a la estrucutra de la matriz de correlatión y $v_i = (v'_{i1}, v'_{i2}, \dots, v'_{in_i})$ es un vector de varianzas de covariables para la muestra del i-ésimo grupo.

Con esto como punto de partida, es de interés obtener una estimación para los parámetros β, σ^2, θ_D y θ_R. Con lo que se obtiene la siguiente información, a partir del modelo de (4.4):

$$E[Y_i] = X_i\beta$$

$$Var(Y_i \mid u_i) = \mathcal{V}_i(\sigma^2, \theta; v_i) = \sigma^2 V_i(\theta; v_i) = Var(u_i) + Var(\epsilon_i) = \sigma^2[Z_i D(\theta_D) Z'_i + R_i(\theta_R; v_i)].$$

Donde $\theta' = (\theta'_D, \theta'_R)'$. Lo anterior lleva a que y_i se distribuye normal, con distribución parecida a la siguiente suma de variables con distribución normal $X_i\beta + Z_i u_i + \epsilon_i$.

$$y_i \sim N_{n_i}(X_i\beta, \sigma^2 Z_i D Z_i' + \sigma^2 R_i). \tag{4.8}$$

Como se puede observar, la matriz de varianzas y covarianzas de y_i depende de dos componentes. El primero $\sigma^2 Z_i D Z_i'$ que es la contribución de los efectos aleatorios u_i y el segundo $\sigma^2 R_i$ que se relaciona al residuo ϵ_i. Note que lo anterior no está asociado a u_i. Por tanto, la matriz de $\mathcal{D}$ no tiene que ser tratada como una matriz de varianzas y covarianzas. No tiene que ser definida positiva si $\mathcal{V}_i$ es definida positiva, sin embargo, $\mathcal{D}$ necesita ser simétrica para asegurar que $\mathcal{V}_i$ sea simétrica.

La estimación por máxima verosimilitud, se hace basada en la distribución marginal de y_i, la cual coincide con (4.8), de esta manera, se puede usar el método de máxima verosimilitud o su modificación REML (estimación de máxima verosimilitud restringido) la cual se diferencia ya que la máxima verosimilitud usa la función marginal de y_i lo que genera un sesgo en la matriz de varianzas y covarianzas, lo cual es corregido por el método restringido, del cual se puede leer aún más en Andrzej [2].

Usando esta información, la estimación basada en el método de máxima verosimilitud genera la siguiente función de log-verosimilitud:

$$l_{Full}(\beta, \sigma^2, \theta) = -\frac{N}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\log[det(V_i)] - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - X_i\beta)'V_i^{-1}(y_i - X_i\beta), \tag{4.9}$$

los resultados para β y σ^2 obtenidos a partir de dicha estimación son:

$$\widehat{\beta}(\theta) = \left(\sum_{i=1}^{N} X_i'V_i^{-1}X_i\right)^{-1}\sum_{i=1}^{N} X_i'V_i^{-1}y_i, \tag{4.10}$$

$$\widehat{\sigma}^2_{ML}(\theta) = \sum_{i=1}^{N} r_i'V_i^{-1}r_i/n, \tag{4.11}$$

donde $r_i = r_i(\theta) = y_i - X_i\widehat{\beta}(\theta)$. Sin embargo, la estimación mediante este método genera parámetros de la varianza y covarianza sesgados. Por lo cual, se recurre al método REML, de esta manera la nueva σ^2 queda estimada de la siguiente manera:

$$\widehat{\sigma}^2_{REML} = \sum_{i=1}^{N} r_i'V_i^{-1}r_i/(n-p). \tag{4.12}$$

Con la misma r_i que se menciona en (4.11). Lo cual, lleva a la función de *log-verosimilitud perfil restringida*:

$$l^*_{REML} = -\frac{n-p}{2}\log\left(\sum_{i=1}^{N} r_i'r_i\right) - \frac{1}{2}\sum_{i=1}^{N}\log[det(V_i)] - \frac{1}{2}\log\left[\det\left(\sum_{i=1}^{N} X_i'V_i^{-1}X_i\right)\right].$$

Maximizando esta función nos llevará a la misma estimación de β y σ^2 que se describen en (4.10) y (4.12).

Finalmente para estimar el valor de u_i se tiene que dicha variable no depende de y_i, por lo cual, se puede suponer que los valores de y_i son iguales a los de $y_i^{(obs)}$. La distribución de u_i dado y_i puede ser calculada de la siguiente forma:

$$f_{u|y}(u_i \mid y_i) = f_{u|y}(u_i \mid y_i = y_i^{(obs)}) = \frac{f_{y|u}(y_i \mid u_i) f_u(u_i)}{\int f_{y|u}(y_i \mid u_i) f_u(u_i) du}. \tag{4.13}$$

Suponiendo que β y θ son conocidos, usualmente la distribución se estima de la siguiente manera:

$$\widehat{u}_i(\beta, \theta) = \widehat{u}_i = DZ_i' V_i^{-1}(y_i^{(obs)} - X_i\beta). \tag{4.14}$$

Para más detalles consulte Andrzej [2] en las páginas 252-254. Mientras que la matriz de varianzas y covarianzas es igual a

$$Var(\widehat{u}_i) = \sigma^2 DZ_i' \left\{ V_i^{-1} - V_i^{-1} X_i \left(\sum_{i=1}^{N} X_i' V_i^{-1} X_i \right)^{-1} X_i' V_i^{-1} \right\} Z_i D.$$

La solución a estos problemas de maximizar son numéricamente difíciles, ya que se debe de encontrar una solución a las matrices definidas positivas $\mathcal{D}$ y $\mathcal{R}_i$. Para eso es necesario parametrizar los valores de D y R_i. Primero, se toma el caso para la matriz R_i, donde se usará una descomposición de la misma:

$$R_i(\mu_{ij}, \theta_R; v_{ij}) = \Lambda_i(\mu_{ij}, \delta, v_{ij}) C_i(\varrho) \Lambda_i(\mu_{ij}, \delta, v_{ij}), \tag{4.15}$$

donde δ es un vector de parámetros de varianzas, v_{ij} es un vector conocido de covariables de varianzas, mientras que ϱ es una correlación uniforme de simetría compuesta y $\theta_R = (\delta', \varrho')'$, Λ_i es una matriz diagonal con la función varianza como elementos en la diagonal, la cual depende de δ. Una solución rápida a este componente es $\delta_s = e^{\log(\delta_s)}$ y se usa $\delta_s^* = \log(\delta_s)$, como parámetro para la función varianza. Para el caso de la matriz C_i se le hace una transformación a los elementos de C_i la cual asegurará que sea definida positiva.

Para esto se aplicará la transformación de *z de Fisher* a ϱ:

$$\varrho^* = \log\left(\frac{1+\varrho}{1-\varrho}\right),$$

lo cual lleva a que los elementos de C_i tengan un parámetro no restringido ϱ. Al mismo tiempo, se usa la transformación *black*:

$$\varrho = \frac{e^{\varrho^*} - 1}{e^{\varrho^*} + 1},$$

lo que garantiza que $\varrho \in (-1, 1)$ y que la matriz C_i sea definida positiva. Además, para garantizar que es definida positiva y una matriz de correlación de $n_i \times n_i$, correspondiendo a la estructura de correlación de simetría compuesta, sus eigenvalores, iguales a $1 + (n_i - 1)\varrho$ y $1 - \varrho$, necesitan ser positivos, siguiendo con la transformada z de Fisher para ϱ:

$$\varrho^* = \log\left(\frac{\frac{1}{n^*-1} + \varrho}{1 - \varrho}\right),$$

donde $n_i^* = \max(n_i)$. Así, se expresa la matriz C_i con elementos no restringidos ϱ^*. Y utilizando de nuevo la transformación *black* de ϱ^* a ϱ se asegura a C_i como matriz positiva definida.

Por otro lado, para la matriz $\mathcal{D}$, se podría considerar una parametrización en términos de varianzas y correlaciones al usar la log-transformación y la transformación z de Fisher. Esta parametrización debería tener restricciones individuales, es decir, necesita tener varianzas positivas y correlaciones que estén en el intervalo $[0, 1]$. Aunque en general, ésta condición no reflejaría la restricción conjunta referente a la matriz definida positiva. Mientras esta parametrización de la matriz pueda usarse como definida positiva en algunas estructuras, es decir, con elementos diagonales iguales y elementos en el exterior de la misma, no sería adecuado para los propósitos de optimizar.

Una opción, considera la representación de D en términos de la descomposición de Cholesky en la que se considera una matriz triangular superior U tal que $D = U'U$, la cual es más simple y estable, sin embargo, la matriz U necesita elementos positivos en la diagonal. Aunque solo se tiene una frágil relación entre la matriz D y U, a excepción de que $\mid U_{1,1} \mid = \sqrt{D_{1,1}}$, esto permite obtener intervalos de confianza para las varianzas de D, pero no para los demás elementos.

Otra forma de parametrizar a D, es usando la matriz logaritmo. Se usa la descomposición del valor singular para D:

$$D = QTQ',$$

donde T es una matriz diagonal de elementos positivos y Q una matriz ortogonal. Considerando a $\log(T)$ con elementos de la diagonal igual al logaritmo de los elementos de T, entonces:

$$D^* = Q\log(T)Q'.$$

En consecuencia la matriz $D = \exp(D^*)$, donde:

$$\exp(D^*) = \sum_{k=0}^{\infty} \frac{(D^*)^k}{k!}.$$

Aunque en general hay varias opciones para dar solución a $\mathcal{D}$.

De manera similar al modelo de regresión lineal la matriz de varianzas y covarianzas para $\widehat{\beta}$ está dada por:

$$\widehat{Var}(\widehat{\beta}) = \widehat{\sigma}^2 \left(\sum_{i=1}^{N} X_i' \widehat{V}_i^{-1} X_i \right)^{-1},$$

donde $\widehat{V}$ y $\widehat{\sigma}^2$ fueron previamente estimados.

Las matrices de varianzas y covarianzas para los parámetros $\widehat{\sigma}^2$ y $\widehat{\theta}$ se estima utilizando la inversa negativa de la Hessiana de la log-verosimilitud restringida, evaluados en el valor estimado de σ^2 y θ. Donde la log-verosimilitud restringida es:

$$l^*_{REML}(\sigma^2, \theta_r) = l_{Full}(\widehat{\beta}(\theta_R), \sigma^2, \theta_R) + \frac{p}{2}\log(\sigma^2) - \frac{1}{2}\log\left(det \left[\sum_{i=1}^{N} X_i' R_i^{-1} X_i \right] \right),$$

y en donde:

$$l_{Full}(\beta, \sigma^2, \theta_R) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\log(det[R_i]) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - X_i\beta)' R_i^{-1}(y_i - X_i\beta)$$

4.4. Diagnóstico del modelo

Como en los modelos lineales, en los multinivel también se presenta un análisis posterior al ajustar la regresión. Recuerde que uno de los supuestos del modelo hace referencia a la normalidad de u_i, normalidad de media cero y con matriz de varianzas y covarianzas $\sigma^2 D$, lo cual señala que el modelo es de varianza constante. De esta manera, las esperanzas condicionales de los efectos aleatorios u_i dado la respuesta observada y_i:

$$u_i = \widehat{D} Z_i' \widehat{V}_i^{-1}(y_i - X_i\widehat{\beta}), \tag{4.16}$$

es comúnmente llamada estimaciones empíricas de Bayes, pues éstos son obtenidos a partir del valor de los efectos fijos $\widehat{\beta}$ y los parámetros de $\widehat{\theta}$. Estrictamente hablando, los efectos u_i no son parámetros, pues se están estimando, prediciendo. Por lo cual, se refieren a ellos como *mejores predictores lineales insesgados* (BLUPs o EBLUPs, por sus siglas en inglés). La contracción de estos estimadores denota una desigualdad:

$$Var(\lambda' \widehat{u}_i) \leq \lambda' \widehat{D} \lambda,$$

para cualquier transformación lineal λ, lo cual muestra la contracción de los efectos aleatorios hacia la media a priori, es decir, hacia el valor 0.

Para el análisis de los residuos de ϵ_i, la estructura de un modelo multinivel es distinta, por lo que se pueden calcular distintos tipos de residuos. El primero de ellos, es el *residuo condicional*, cuya representación se define como:

$$\widehat{\epsilon}_{(c)i} = y_i - X_i\widehat{\beta} - Z_i\widehat{u}_i, \tag{4.17}$$

con la u_i previamente definida en (4.16).

Mientras que el segundo tipo de residuo es llamado el *residuo marginal*, que se define de la siguiente manera:

$$\widehat{\epsilon}_{(m)i} = y_i - X_i\widehat{\beta}. \tag{4.18}$$

Estos tipos de residuos sirven para comprobar heterogeneidad, sin embargo, no es lo mejor para el caso de normalidad. Esto porque los residuos estarían correlacionados y las varianzas diferirían. Sin embargo, los residuales *studentizados* y de Pearson suelen ser más usados, pero aún así, podrían no ser apropiados para validar los supuestos de normalidad. Una aproximación a la solución es considerar la descomposición de Cholesky de las matrices de varianzas y covarianzas estimadas de los residuos condicional o marginal, $\sigma^2 R_i$ y $\sigma^2 V_i$ respectivamente. Esto define a:

$$\widehat{\epsilon}^*_{(c)i} = \left(\widehat{\sigma}\widehat{U}'_{(c)i}\right)^{-1}\widehat{\epsilon}_{(c)i}$$

o

$$\widehat{\epsilon}^*_{(m)i} = \left(\widehat{\sigma}\widehat{U}'_{(m)i}\right)^{-1}\widehat{\epsilon}_{(m)i}$$

donde la matrices diagonales superiores son definidas por $\widehat{U}'_{(c)i}\widehat{U}_{(c)i} = R_i$ y $\widehat{U}'_{(m)i}\widehat{U}_{(m)i} = V_i$, respectivamente. Entonces, $\widehat{\epsilon}^*_{(c)i}$ y $\widehat{\epsilon}_{(m)i}$ deberían ser aproximadamente de distribución normal con media cero y matriz de varianzas y covarianzas igual a la matriz identidad, es decir, una matriz de varianzas y covarianzas constante.

Por otra parte, para hacer la selección del modelo se puede comprobar su importancia de los valores de β con las mismas pruebas estadísticas que se realizan en el modelo de regresión lineal. En particular, usar la *prueba F* que está determinada por:

$$H_0: \quad L\beta = c_0 \qquad vs \qquad H_1: \quad L\beta \neq c_0, \tag{4.19}$$

donde L es una matriz de tamaño q ($q \leq p$) y c_0 es un vector conocido. Usualmente $L = I_q$, mientras $c_0 = 0$, aunque la prueba F en general se realiza:

$$F = \frac{(L\widehat{\beta} - c_0)'(L\widehat{Var}(\widehat{\beta})L')^{-1}(L\widehat{\beta} - c_0)}{rango(L)}. \tag{4.20}$$

Si la hipótesis sobre β no permite generar la expresión de un modelo bajo las hipótesis nula o alternativa, se pueden aplicar otros criterios, como el AIC o el BIC, para seleccionar el modelo que mejor se ajusta a los datos. Estos toman la función de verosimilitud y una función $f(\cdot)$, la cual tiene un par de propuestas más adelante. La función de verosimilitud usa los parámetros estimados de acuerdo a la hipótesis que se tiene. De esta forma se tiene que l_0 representa al modelo nulo y l_1 representa un modelo alternativo. La hipótesis se rechaza si:

$$l_1 - l_0 > f(p_1) - f(p_0),$$

donde p_0 y p_1 son el número de parámetros que usa el modelo nulo y alternativo, respectivamente. Mientras que para $f(\cdot)$ se puede definir:

$$f(p) = p$$

o

$$f(p) = 0.5p\log(N^*),$$

donde $N^* = N$ para ML y $N^* = N - p$ para REML. En el caso de la primera función es lo que se denomina el *criterio de información de Akaike (AIC)*, mientras en la segunda función se llama *criterio de información Bayesiano (BIC)*.

Finalmente, los intervalos de confianza de β se pueden construir a partir de la *prueba t*,

$$\left[\widehat{\beta} - t_{1-\alpha/2,n-p}\sqrt{\widehat{Var}(\widehat{\beta})}, \widehat{\beta} + t_{1-\alpha/2,n-p}\sqrt{\widehat{Var}(\widehat{\beta})}\right].$$

4.5. Modelos multinivel con respuesta categórica

El modelo multinivel hasta ahora supone que la variable respuesta es continua, sin embargo, eso no siempre es así. En general, una extensión de los modelos contempla a los modelos categóricos. En el presente trabajo solo se expone un resumen de lo que son los modelos categóricos y sus variantes en el caso de modelos multinivel, para saber más acerca del tema, se puede consultar los libros de Hill [1], Goldstein [3] y Agresti [4].

Los modelos categóricos son parte de un grupo de modelos denominados, *modelos lineales generalizados*, en este grupo hay modelos para los casos en el que la regresión predice una respuesta en la que el valor es binario, en este tipo de modelos se encuentran el logístico (*logit*) y *probit*. Tomando de ejemplo el modelo logit:

$$\pi_i = P(y_i = 1) = logit^{-1}(X_i\beta),$$

donde la función logit se define de la siguiente manera:

$$logit(x) = \log\left(\frac{x}{1-x}\right),$$

dicha función tiene dominio en el (0,1) e imagen en el intervalo $(-\infty, \infty)$.

El modelo probit es parecido al anterior en cuanto a su manera de predecir, aunque este modelo tiene una desviación estándar del residuo distinta, pues el modelo logit tiene 1.6, mientras que en probit es de 1. Pues vea que el modelo probit usa Φ, la función de distribución acumulativa de una normal estándar, con media 0 y desviación estándar 1:

$$P(y_i = 1) = \Phi(X_i\beta).$$

Este tipo de modelos como se dijo, solo predicen una respuesta binaria, para el caso de más categorías se puede seguir considerando un modelo parecido, en este caso es llamado modelo de regresión logística multinomial.

El modelo multinomial considera una categoría base, usualmente la última, en este caso suponga c, donde se cumple la siguiente relación:

$$\log\left(\frac{\pi_s}{\pi_c}\right) = \sum_{r=1}^{p} \beta_{sr} x_{ir} = x_i'\beta_s, \text{ para } s = 1, \ldots, c-1, \tag{4.21}$$

del cual se tiene que:

$$\pi_s = \pi_c \exp\left(\sum_{r=1}^{p} \beta_{sr} x_{ir}\right). \tag{4.22}$$

Ahora, recuerde las propiedad de la función de probabilidad, pues usando una de las propiedades se obtiene lo siguiente:

$$\begin{aligned}
\sum_{k=1}^{c-1} \pi_k = \pi_c \sum_{k=1}^{c-1} \exp\left(\sum_{r=1}^{p} \beta_{kr} x_{ir}\right) &\implies \sum_{k=1}^{c-1} \pi_k + \pi_c - \pi_c = \pi_c \sum_{k=1}^{c-1} \exp\left(\sum_{r=1}^{p} \beta_{kr} x_{ir}\right) \\
&\implies \sum_{k=1}^{c} \pi_k = \pi_c + \pi_c \sum_{k=1}^{c-1} \exp\left(\sum_{r=1}^{p} \beta_{kr} x_{ir}\right) \\
&\implies 1 = \pi_c \left(1 + \sum_{k=1}^{c-1} \exp\left(\sum_{r=1}^{p} \beta_{kr} x_{ir}\right)\right),
\end{aligned}$$

con lo cual se obtiene:

$$\pi_c = \frac{1}{1 + \sum_{k=1}^{c-1} \exp\left(\sum_{r=1}^{p} \beta_{kr} x_{ir}\right)}, \tag{4.23}$$

retomando (4.22) y (4.23):

$$P(Y_{ij} = y_{ij}) = \pi_j = \frac{\exp\left(\sum_{r=1}^{p} \beta_{jr} x_{ir}\right)}{1 + \sum_{k=1}^{c-1} \exp\left(\sum_{r=1}^{p} \beta_{kr} x_{ir}\right)}. \tag{4.24}$$

Con lo anterior como base, se introduce el modelo multinivel con múltiples respuestas categóricas, la cual se parece a (4.21):

$$\log\left(\frac{\pi_{ij}^{(s)}}{\pi_{ij}^{(c)}}\right) = X_i\beta_j^{(s)} + Z_i u_j^{(s)} + \epsilon_i^{(s)} \text{ para } s = 1, \ldots, c-1, \tag{4.25}$$

donde c es el número de categorías respuesta. Donde esto sería una extensión del modelo logit, sin embargo también se puede hacer con otro tipo de modelo, como el log-complementario.

Como se puede ver, en este caso las categorías no necesitan llevar un orden para ser modelado, sin embargo, sí existe la posibilidad de generar un modelo que suponga tal hecho. Dicho modelo cumple la siguiente propiedad:

$$P(Y \leq s) = \pi_1 + \ldots + \pi_s,\ s = 1, \ldots, c, \tag{4.26}$$

y recordando que esto es acumulativo:

$$P(Y \leq 1) \leq P(Y \leq 2) \leq \ldots \leq P(Y \leq c) = 1. \tag{4.27}$$

Entonces la probabilidad *logit acumulativa* queda de la siguiente manera:

$$logit[P(Y \leq s)] = \log\left(\frac{P(Y \leq s)}{1 - P(Y \leq s)}\right) = \log\left(\frac{\pi_1 + \ldots + \pi_s}{\pi_{s+1} + \ldots + \pi_c}\right), \tag{4.28}$$

para $s = 1, \ldots, c$. En consecuencia, el modelo logit acumulativo se describe de la siguiente forma:

$$logit[P(Y \leq s)] = x_i'\beta_s,\ s = 1, \ldots, c-1. \tag{4.29}$$

Para entender lo que se está haciendo en este modelo, primero, observe que en estos casos las variables respuesta manejan una probabilidad, la probabilidad de que un individuo pertenezca a la categoría s. Segundo, el valor de los coeficientes no tiene la misma interpretación que en un modelo lineal, si se tiene un coeficiente negativo, la curva de probabilidad es decreciente y en consecuencia la probabilidad de pertenecer a la categoría, suponga s, disminuye entre más grande sea el valor del individuo en dicho coeficiente.

Finalmente, suponga que se quiere comparar un individuo m y un individuo n en donde lo único en que se diferencian es en la variable asociada al coeficiente β_{st}. Si $x_{mt} > x_{nt}$, habrá un valor de $\exp(\beta_{st})$ veces más de posibilidades que el individuo m pertenezca a la categoría s sobre la del individuo n.

De esta manera, la expresión del modelo de probabilidades acumulativas queda de la siguiente forma:

$$P(Y \leq s) = \frac{\exp\left(\sum_{r=1}^{p} \beta_{sr} x_{ir}\right)}{1 + \exp\left(\sum_{r=1}^{p} \beta_{sr} x_{ir}\right)}. \tag{4.30}$$

Además se cumple que:

$$P(Y = a) = P(Y \leq a) - P(Y \leq a - 1) \text{ para cualquier } a \in \{1, \ldots, c\}. \tag{4.31}$$

Finalmente, para el caso de un modelo multinivel se escribe de la siguiente manera:

$$logit(P(Y \leq s)) = \log\left(\frac{\pi_1 + \ldots + \pi_s}{\pi_{s+1} + \ldots + \pi_c}\right) = \alpha^{(s)} + X_i\beta_j + Z_i u_j + \epsilon_i, \tag{4.32}$$

donde $\alpha^{(s)}$ es lo que se conoce como umbral o punto de corte entre el nivel $s-1$ y s.

5

Presentación de los datos

5.1. Descripción de los datos

Para el presente trabajo se usaron las bases de datos disponibles del INEGI principalmente se usó la Encuesta Nacional de Calidad e Impacto Gubernamental (`https://www.inegi.org.mx/programas/encig/2017/#Microdatos`) de los años 2017 y 2019, en específico las secciones 1 y 3 que tratan sobre la evaluación de los servidores públicos (presidente, gobernadores, congreso, etc.) de los encuestados, el acceso a servicios otorgados por el Estado (agua, drenaje, seguridad, etc.) y sus datos demográficos, aquí cabe señalar que los datos sobre la percepción de la corrupción se encuentra en los datos donde también se evalúa la calidad de los servicios. También se consultaron los datos referentes a la tasa de homicidios por entidad federativa y sus respectivas tasas de delitos, este último en un intervalo de los últimos tres años (2014-2017) y (2015-2018).

Haciendo el uso de datos externos, se consultaron los datos del Instituto Mexicano para la Competitividad (IMCO[1]), donde se obtuvieron los datos del crecimiento del PIB por entidad federativa.

Se hizo una extracción de los datos, los cuales se trataron de distintas maneras. Por un lado, los datos que califican a los servidores se trataron de la misma manera que vienen dados, así como los que evalúan la percepción de la corrupción en las instituciones, mientras los datos demográficos en general se usaron de la misma manera, a excepción de los datos de escolaridad, población e incidencia delictiva. En el caso de la primera variable, se crearon tres más: educación básica, media superior, y universidad o más; donde se toma el valor 1 si se tenía dicho grado de escolaridad o 0 si no era así. Mientras que para las otras dos variables, solo se les aplicó la transformación logaritmo. Esto se aplica para ambas bases de datos del INEGI, tanto para la del 2017 como la del 2019.

[1] Índice de Competitividad Estatal 2018.

5.1.1. Variable dependiente

En el caso de la variable dependiente, se usará la calificación que los ciudadanos dan a sus servidores públicos, puesto que se creará un modelo para cada evaluación a dichas instituciones, además se creará un índice que evaluará la calificación general que los ciudadanos dan a sus servidores, en este caso se usará un promedio de la calificación de cada ciudadano dada a cada una de las instituciones. Las razones por lo cual se puede hacer esto se ve un poco más en el capítulo de Análisis de Componentes Principales, en donde también se explica acerca de los índices.

Además de la variable dependiente referente a la evaluación de los servidores dada por los ciudadanos, se usará una segunda variable dependiente para producir un tipo diferente de respuesta, solo que en este caso la variable a utilizar va a ser del tipo categórica, la cual muestra el nivel de corrupción que los ciudadanos perciben en sus gobernantes, tienen valores entre 1 y 4, donde 1 significa que el ciudadano percibe que es muy frecuente la corrupción de dicha institución y 4 que es nada frecuente. Además, basado en estos mismo reactivos, también se creó un segundo índice que contempla de manera general, la percepción de la corrupción. Dichos índices se describen a continuación.

Índice de confianza institucional

En la creación de estos índices se tiene una tabla que muestra a partir de qué medidas se deicidió formarlos, esto se hace mediante un análisis de componentes principales en paralelo a la utilización del coeficiente *alpha* de Cronbach.

Este índice engloba la evaluación que la población da a sus instituciones gubernamentales (ICI). Mediante el análisis mencionado se concluirá que se puede crear una métrica para realizar una evaluación general de los servidores públicos en la población, estas instituciones son: Presidencia, Gubernatura, Presidencia Municipal, Jueces y magistrados, Congreso, Ministerios públicos y Partidos políticos. Este mismo índice se usa como una variable dependiente más, además de cada una de las instituciones que ya se nombraron anteriormente y se obtuvo asumiendo que cada una de las variables tienen el mismo peso.

Tanto en este caso con en los posteriores (ICI e IPCSL), en el análisis de componentes solo se toma el porcentaje de la varianza explicada por la primer componente, ya que como se menciona en la sección de ACP en Preliminares, este análisis genera que las primeras componentes, las cuales son una combinación lineal de las variables originales sean capaces de reducir las dimensiones de un modelo. Para efecto de ello, la primer componente se usa para ver si una componente puede resumir de manera significativa las variables transformadas en un solo factor.

Este análisis se complementa analizando el alpha de Cronbach generado por las mismas variables las cuales se estudian en el ACP. Teniendo como resultado una manera de ver si nuestras variables están midiendo el mismo fénomeno y por ende, puede generarse una variable consenso.

Índice de percepción de la corrupción institucional

Por otro, el índice de percepción de la corrupción institucional (IPCI) como se hizo con el índice anterior, se prueba mediante un análisis de componentes y su valor alpha que señala que la calificación que los ciudadanos asignan a las instituciones está estrechamente relacionada, por lo que se puede crear un índice continuo que evalúa la frecuencia de la corrupción donde el valor 1 representaría que se percibe que la corrupción es muy frecuente en las instituciones gubernamentales, mientras 4 respondería que la corrupción que se percibe es nula, igualmente se usa la suma de las variables para crear dicho índice. Las instituciones que componen este índice son las mismas que conformaron el índice de confianza.

5.1.2. Variables independientes

Es de interés saber qué es lo que provoca que una persona tenga más o menos confianza en sus servidores públicos, en este caso se considera un conjunto de reactivos que representan la información demográfica de esta persona y su información de acceso a servicios públicos y si estos son de calidad, dichos reactivos se encuentran en la tabla 5.1, cada uno de estos reactivos se evalúan de manera binaria 1 dice que el servicio es así y 2 dice que el servicio no tiene esa cualidad.

En el caso de los datos demográficos, la variable es categórica donde solo se tiene el nombre del estado donde habita el individuo; la edad está medida por un número entero; la actividad que desarrolla la persona está medida por nueve categorías, donde solo se tomó las categorías 1 y 2, pues en éstas la persona señala trabajar o tener trabajo aunque no lo ejerza, ya que en las demás categorías son actividades que no señalan una condición de laborar; el sexo, es binario donde 1 pertenece a una mujer y 2 a un hombre; finalmente, para la categoría de escolaridad se tienen diez categorías.

A algunas de estas variables se les hace una recodificación. Inicialmente los datos faltantes, en blanco o en el cual el encuestado no responde, son *imputados*, y se acoplan a alguna de las otras categorías, y a partir de ello algunas de las variables se recodifican. Para imputar los datos, se utilizó la función *amelia* del software *R*, la cual se basa en la utilización del algoritmo EM, tomando muestras multiples *bootstrap* que completan los datos faltantes, dichas muestras son una aproximación de la distribución de los datos, para más información a cerca del algoritmo se recomienda consultar Honaker [12]. Además, solo se hizo la imputación a los datos después de realizar el análisis por institución y antes de crear los índices. Esto no afecta gravemente a la información ya que del total de muestras que se tiene, solo alrededor de un 7 % de ellas tienen datos faltantes, mientras que solo el 2 % se imputaron, esto con el objetivo de intentar tomar el mismo número de muestras que se tiene en el artículo de Moreno.

Es el caso de la variable de actividad, las categorías 1 y 2 se transformaron a 1 para mostrar que la persona hace una actividad económica y 0 para las demás; la variable sexo conservó sus valores originales 1 para los hombres y 2 para la mujeres; mientras que para la variable de escolaridad se extrajeron tres nuevas variables relacionadas a si la persona tiene el grado básico, medio superior o universitario o más.

Para la variable *Primaria* las categorías 0, 1 y 2 se les asignó el valor 1 que denota a la persona encuestada con un nivel de instrucción de primaria o menos, mientras que toma el valor 0 en otro caso; la variable que evalúa la educación básica (*Edu_bas*) toma en cuenta las categorías 3, 4 y 5 y las transforma a 1, lo cual quiere decir que el individuo encuestado tiene un nivel de educación básico concluido, en otro caso se le asigna el valor 0; para las categorías 6 y 7 toman el valor 1 en la variable de educación media (*Edu_med*) si el individuo contestó que terminó los estudios en el nivel medio-superior; finalmente, para la variable de educación superior (*Edu_sup*) las categorías 8 y 9 se les da el valor uno pues señalan que tienen un nivel universitario o más, y 0 en otro caso.

Para los servicios, se les aplicó una recodificación, donde 1 representa que el servicio cumple la condición y 0 si no es así. Una vez se han aplicado estas recodificaciones se crea un índice que evalúa la calidad general del servicio, en donde cada servicio está compuesto por el grupo de variables como se muestra en la tabla 5.1. Lo anterior se puede aplicar en ambas bases de datos del INEGI al estar ordenados de la misma manera, dicho índice está formado solo por las sumas de los reactivos, tomando en cuenta que cada pregunta tiene el mismo peso.

Finalmente, para los datos externos, el impacto delictivo de los últimos tres años se promediaron por estado y se le aplicó la transformación logaritmo al igual que los datos de la población por estado del año 2015. La tasa de homicidios son originales a los datos del INEGI en los años 2017 y 2018, respectivamente. En cuanto a los datos de Alternancia, se consultaron los cargos anteriores a la última elección gubernamental y la posterior (hasta el 2019) y se asignó 1 si hubo alternancia de partido en el poder o 0 si no lo hubo, como estas variables se toman por estado, los estados de dichas variables corresponden a los estados que forman parte del efecto aleatorio del modelo.

Es importante señalar que estas variables independientes se usan en ambos modelos, tanto para el modelo que evalúa la relación con la evaluación de los servidores públicos como con la percepción de la corrupción dentro de dichas instituciones.

Índice de acceso a servicios públicos (IPCSL)

En las variables independientes también se crea un último índice, la creación de estos índices estará mejor explicada posteriormente, pero cada uno de los reactivos está relacionado con los demás de acuerdo al servicio que se especifique, lo cual, respalda el resumir dicha información en un índice. Para este caso particular se usará el promedio, como cada reactivo tiene un valor 0 o 1, al final dicho promedio también va a estar en el intervalo $[0, 1]$.

Tabla 5.1: Variables independientes de los servicios públicos.

Agua: "De acuerdo a su experiencia, el agua potable de su ciudad..."
1. Llega de forma constante sin interrupciones y con presión.
2. Es pura y cristalina.
3. Es bebible sin temor a enfermarse.
4. Cuando hay fugas son arregladas.
Drenaje y alcantarillado.
1. Están conectados a su vivienda de tal forma que los desechos se descargan adecuadamente.
2. Reciben mantenimiento frecuente que evita olores desagradables y plagas.
3. Se limpian constantemente de tal forma que evitan inundaciones y encharcamientos.
Alumbrado público.
1. Ilumina adecuadamente las calles y áreas públicas.
2. Cuenta con mantenimiento.
3. Cuenta con atención inmediata de las fallas para conservar su buen estado.
Los parques y jardines.
1. Son accesibles en horario.
2. Se encuentran cerca (máximo 15 minutos caminando).
3. Están limpios y tienen buena imagen.
4. Son seguros en términos de delincuencia para usted y su familia.
La policía.
1. Contribuye a que usted y su familia se sientan seguros dentro y fuera de su casa.
2. Está dispuesta a ayudarte.
Las calles y avenidas.
1. Se encuentran en buen estado, libres de baches, coladeras hundidas o abiertas.
2. Ante la existencia de coladeras abiertas o baches se reparan de manera inmediata.
Señalamiento urbano.
1. Tienen semáforos funcionales.
2. Tienen señalamientos claros.
Carreteras y caminos sin cuota.
1. Se encuentran en buen estado, libres de baches, deslaves, etc.
2. En términos de delincuencia, son seguras.
3. Comunican a todo el estado de manera rápida.
4. Cuentan con señalamientos claros.

Entonces, de los 24 reactivos iniciales que se tienen en la base de datos, se transforman en solo 8 variables que evalúan la calidad de cada servicio que tiene el individuo encuestado. Además de esto, se crea un índice que engloba toda esta información para mostrar la calidad de los servicios.

Este índice es reforzado con el objetivo de evaluar de manera general la calidad de los servicios, cada una de las ocho variables formadas se suponen con un mismo peso, por lo cual, el índice termina estando compuesto por la suma de las variables.

5.2. Análisis exploratorio de los datos

El siguiente análisis toma de referencia al artículo de Monsiváis Carrillo [5], varias de las modificaciones a la base de datos para generar nuevas variables se hicieron a partir de este artículo. Una vez configurados los servicios de dicha manera, se tienen los resultados de los análisis de ACP y su respectiva Alpha de Cronbach.

	2017		2019	
Servicio	**ACP**	**Alpha**	**ACP**	**Alpha**
Agua	0.47	0.61	0.47	0.61
Drenaje y alcantarillado	0.61	0.57	0.62	0.59
Alumbrado público	0.71	0.79	0.72	0.81
Jardines y parques	0.47	0.61	0.47	0.61
Policía	0.78	0.72	0.78	0.71
Calles y avenidas	0.76	0.68	0.77	0.70
Señalamiento urbano	0.77	0.70	0.76	0.69
Carreteras y caminos	0.53	0.70	0.52	0.69
IPCSL	0.35	0.71	0.34	0.71
ICI	0.67	0.91	0.59	0.88
IPCI	0.54	0.86	0.49	0.82

Tabla 5.2: Análisis de los grupos de servicios públicos.

Como se puede ver en la tabla 5.2, en el caso de los servicios, los porcentajes de variación que tienen las primeras componentes son mayores al 47 %, mientras los valores de alpha son mayores a 0.57, pero este valor incrementa a 0.71 cuando se toma en cuenta el índice generado por cadas una de las componentes del servicio, desde este punto de vista, resulta mejor g generar una variable índice que para cada uno de estos servicios a partir de sus reactivos. Para los índices, la primer componente del IPCSL explica solo el 35 % de la variación de los datos, mientras que en el ICI e IPCI son mayores al 50 % y los valores alpha mayores a 0.70. Por un lado, se observa consistencia en los reactivos relacionados a los servicios públicos, sobre todo al evaluar el IPCSL, pero la primer componente principal solo reune el 35 % de la variación. Esta consistencia es mejor en los índices ICI e IPCI, además la primer componente explica más del 50 %.

Finalmente, esto aunque sea una manera de generar menos variables en los análisis se observará si estas mismas variables son significativas o no dentro del análisis. En cuanto a ICI e IPCI, permite que se pueda ver la confianza y la percepción de la corrupción en instituciones sin que se tenga que evaluar una por una.

La cantidad de datos que se usaron así como un análisis general de dicha base se presenta en las tablas 5.3 y 5.4.

Variables	N	Media	Desviación Estándar	Mínimo	Máximo
IPCI	38850	1.69	0.58	1	4
Presidencia y secretarías	37468	1.53	0.73	1	4
Gubernatura	37709	1.67	0.79	1	4
Presidencia Municipal	37550	1.77	0.82	1	4
Congreso	36908	1.55	0.73	1	4
Jueces y magistrados	35662	1.93	0.86	1	4
Partidos Políticos	38276	1.47	0.69	1	4
Ministerios Públicos	37112	1.77	0.83	1	4
ICI	38148	3.9	2.1	1	10
Presidencia y secretarías	37815	3.8	2.6	1	10
Gubernatura	37861	4.1	2.6	1	10
Presidencia Municipal	37788	4.3	2.6	1	10
Congreso	36905	3.5	2.4	1	10
Jueces y magistrados	35973	4.4	2.5	1	10
Partidos Políticos	38148	3.2	2.3	1	10
Ministerios Públicos	37189	4.1	2.6	1	10
IPCSL	39164	0.43	0.20	0	1
Agua	39164	0.44	0.31	0	1
Drenaje y alcantarillado	39164	0.47	0.31	0	1
Alumbrado Público	39164	0.37	0.40	0	1
Parques y jardines	39164	0.58	0.31	0	1
Policías	39164	0.37	0.42	0	1
Calles y avenidas	39164	0.14	0.30	0	1
Señalamiento urbano	39164	0.58	0.43	0	1
Carreteras públicas	39164	0.47	0.35	0	1
IPCSL Estatal	32	0.43	0.07	0.30	0.59
Escolaridad: Primaria o menos	39164	0.21	0.40	0	1
Educación básica	39164	0.27	0.44	0	1
Educación media	39164	0.22	0.42	0	1
Universidad o más	39164	0.28	0.44	0	1
Sexo	39164	1.53	0.49	1	2
Edad	39164	43.63	17.0	18	98
Actividad económica	39164	0.64	0.47	0	1
Población 2015 (ln)	39164	14.96	0.78	13.4	16.6
Porcentaje de crecimiento del PIB estatal	39164	3.25	2.22	-5.9	7.9
Alternancia	39164	0.47	0.49	0	1
Homicidios 2016	39164	21.52	18.39	2.9	82.3
Incidencia Delictiva (ln)	39164	10.43	0.32	9.8	11.1

Tabla 5.3: Estadísticas descriptivas de los datos 2017.

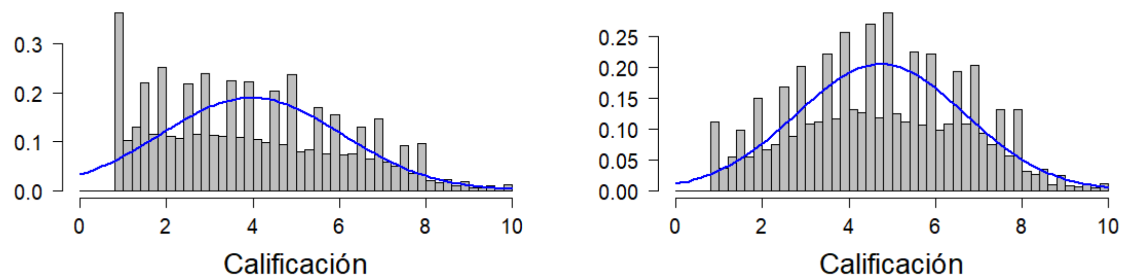
Vea que la información que se tiene en la tabla 5.3 que representa la información en el año 2017, no cambia mucho con respecto a la información que se tiene del año 2019 que se encuentra en la tabla 5.4, principalmente en el caso de las variables explicativas. Un aspecto interesante, es que la calificación media de las instituciones políticas es distinta de acuerdo a la institución que se evalúa. Por un lado, se tiende a tener mejor calificados a los jueces, por el contrario los partidos políticos y el congreso tienden a tener calificaciones más bajas. Algo muy parecido pasa con la percepción de la corrupción, en general se considera a los jueces como una institución que practica la corrupción de manera frecuente, sin embargo a los partidos políticos se percibe que eso tiende a ser aún más frecuente.

Variables	**N**	**Media**	**Desviación Estándar**	**Mínimo**	**Máximo**
IPCI	39118	1.90	0.59	1	4
Presidencia y secretarías	36122	2.04	0.93	1	4
Gubernatura	36854	1.88	0.82	1	4
Presidencia Municipal	36642	1.96	0.83	1	4
Congreso	35508	1.8	0.79	1	4
Jueces y magistrados	34943	1.95	0.84	1	4
Partidos Políticos	37815	1.63	0.75	1	4
Ministerios Públicos	36662	1.82	0.83	1	4
ICI	35784	4.7	1.9	1	10
Presidencia y secretarías	37719	5.6	2.6	1	10
Gubernatura	37549	4.9	2.5	1	10
Presidencia Municipal	37515	5.1	2.5	1	10
Congreso	35784	4.3	2.4	1	10
Jueces y magistrados	35385	4.7	2.4	1	10
Partidos Políticos	37764	3.7	2.4	1	10
Ministerios Públicos	36991	4.5	2.5	1	10
IPCSL	39614	0.45	0.21	0	1
Agua	39614	0.46	0.32	0	1
Drenaje y alcantarillado	39614	0.49	0.31	0	1
Alumbrado Público	39614	0.37	0.41	0	1
Parques y jardines	39614	0.58	0.31	0	1
Policías	39614	0.37	0.42	0	1
Calles y avenidas	39614	0.15	0.31	0	1
Señalamiento urbano	39614	0.60	0.42	0	1
Carreteras públicas	39614	0.49	0.34	0	1
IPCSL Estatal	32	0.45	0.07	0.30	0.62
Escolaridad: Primaria o menos	39614	0.20	0.40	0	1
Educación básica	39614	0.27	0.44	0	1
Educación media	39614	0.23	0.42	0	1
Universidad o más	39614	0.29	0.45	0	1
Sexo	39614	1.54	0.49	1	2
Edad	39614	44.87	17.5	18	98
Actividad económica	39614	0.62	0.48	0	1
Población 2015 (ln)	39614	14.96	0.78	13.4	16.6
Porcentaje de crecimiento del PIB estatal	39614	3.25	2.22	-5.9	7.9
Alternancia	39614	0.47	0.49	0	1
Homicidios 2018	39614	30.13	18.39	2.7	98.2
Incidencia Delictiva (ln)	39614	10.43	0.32	9.8	11.1

Tabla 5.4: Estadísticas descriptivas de los datos 2019.

Además, véase que la calidad de los servicios tendió a subir ligeramente, lo cual puede ser un factor por el que la calificación a las instituciones también lo hayan hecho, y no solo eso, sino también su percepción de la corrupción en dichas instituciones. Sobre todo, se muestra un mayor incremento en la confianza hacia la Presidencia. Esto a pesar que otros parámetros que afecten a la sociedad hayan ido a peor.

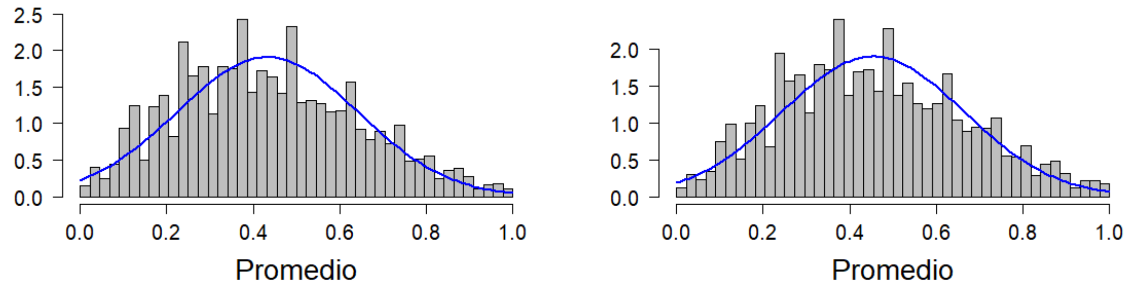
Véase una comparación gráfica de dicha calificación a las instituciones en la figura 5.1, donde la población pasó de calificar con las puntuaciones más bajas en el año 2017 a tener una puntuación más cercana a el punto medio de la evaluación que podían dar los encuestados. Como se mencionó, y como se verá en las figuras 5.3, la corrupción tiene un cambio similar.



(a) Confianza Institucional 2017 (b) Confianza Institucional 2019

Figura 5.1: Confianza institucional en los años 2017 y 2019.

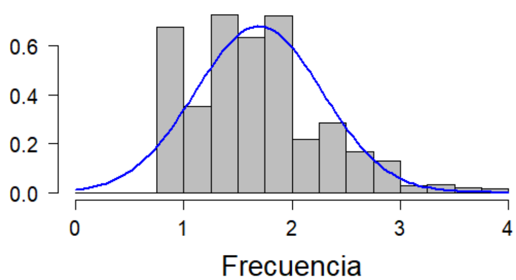
Véase en las siguientes gráficas 5.2, donde se muestra la evaluación a los servicios en el IPCSL, donde no es claro que la calidad de los servicios hayan incrementado de manera drástica en ese lapso de dos años, aunque como se observó en las estadísticas previas, sí se dio un ligero cambio. Entonces, es posible que haya más razones por la cual la evaluación de los ciudadanos a sus gobernantes haya cambiado notoriamente en dicho lapso.



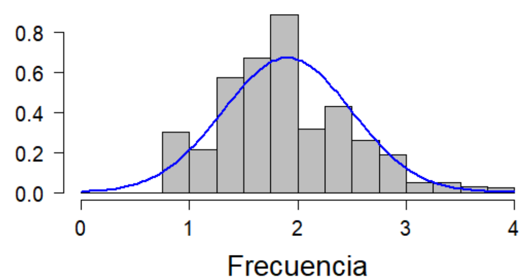
(a) IPCSL 2017 (b) IPCSL 2019

Figura 5.2: Percepción de la calidad de los servicios en los años 2017 y 2019.

Y por último observe en las gráficas 5.3 que la percepción de la corrupción de las instituciones cambia aunque no de manera tan significativa como lo fue en la evaluación de las mismas. Recuerde que entre más alta sea la evaluación en el IPCI, significa que la percepción de la corrupción es mínima, mientras que el valor 1 indica que se perciben como actos muy frecuentes. Por lo anterior, se pasa de evaluar entre frecuente y muy frecuente esa percepción a solo pensar que es frecuente dichos actos en las instituciones. Observe que el cambio es ligeramente parecido al cambio que se tiene al evaluar las instituciones por desempeño, aunque en este caso se tiene un intervalo más pequeño para medir la percepción de la corrupción.



(a) Índice de corrupción 2017



(b) Índice de corrupción 2019

Figura 5.3: Percepción de la corrupción en instituciones gubernamentales de los años 2017 y 2019.

Por otra parte, si se realiza la prueba U de Mann-Whitney, se comprueba para un nivel 0.001 de confianza que los datos del año 2017 al 2019 hay una diferencia significativa con respecto a su evaluación en los índices. En la siguiente tabla se puede ver un resumen de estas pruebas.

	Valor W	p-value
IPCSL	731388089	$2.2e^{-16}$
ICI	559473393	$2.2e^{-16}$
IPCI	594773499	$2.2e^{-16}$

Antes de pasar al análisis multinivel, en las últimas gráficas 5.4 se muestra un análisis de la calidad de los servicios por estado el cual se hizo mediante el promedio del IPCSL de los encuestados por estado. Lo cual, nos hace ver un poco acerca de la justificación de usar un análisis multinivel en este caso. Puesto que la calidad de los servicios está asociada a la evaluación que los ciudadanos dan a las instituciones y al cambiar con respecto a cada estado, la evaluación que una persona puede dar a sus instituciones estaría relacionada al estado en donde habita.

Finalmente, observe que la calidad de los servicios en los estados también tiende a incrementar en la mayoría de los estados del año 2017 al 2019, principalmente en aquellos que eran de los que mejor calidad tenían o en los que eran deficientes en el año 2017 con respecto al 2019. Mientras que en la parte intermedia de los estados con mejor servicio la diferencia tiende a ser mínima, pero otros como el caso de Chiapas, al parecer esa calidad de los servicios no ha mejorado, pero al menos no han ido empeorando.

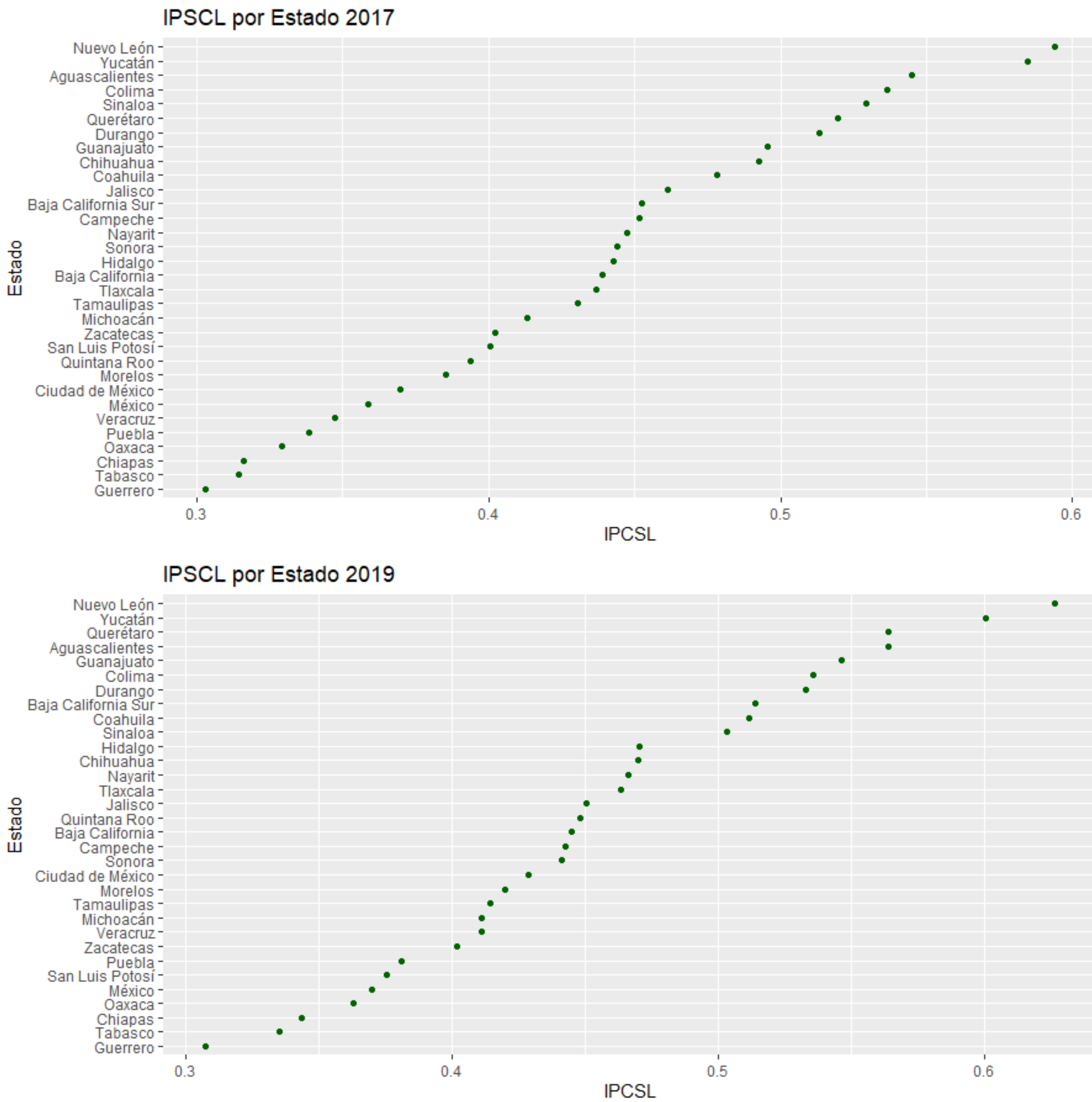


Figura 5.4: IPCSL por entidad federativa en los años 2017 y 2019.

En este caso, se podría pensar que una persona que habita en Chiapas podría evaluar de manera pésima a sus instituciones gubernamentales a comparación de una persona que habita en el estado de Nuevo León que tendería a evaluar de manera menos negativa a sus gobernantes.

6

Análisis de resultados

6.1. Modelo multinivel, confianza institucional

Ahora que ya se sabe que la manera en la que los ciudadanos evalúan a sus instituciones puede estar afectado por la entidad federativa en la que viven, se puede mostrar cómo esto afecta a la evaluación final de las mismas.

El análisis se hace para cada una de las instituciones, sin embargo se considera un caso en el que la variable independiente es aquella a la que se decidió denominar índice de confianza institucional, la cual resume de manera general la evaluación promedio que los ciudadanos dan a todas sus instituciones.

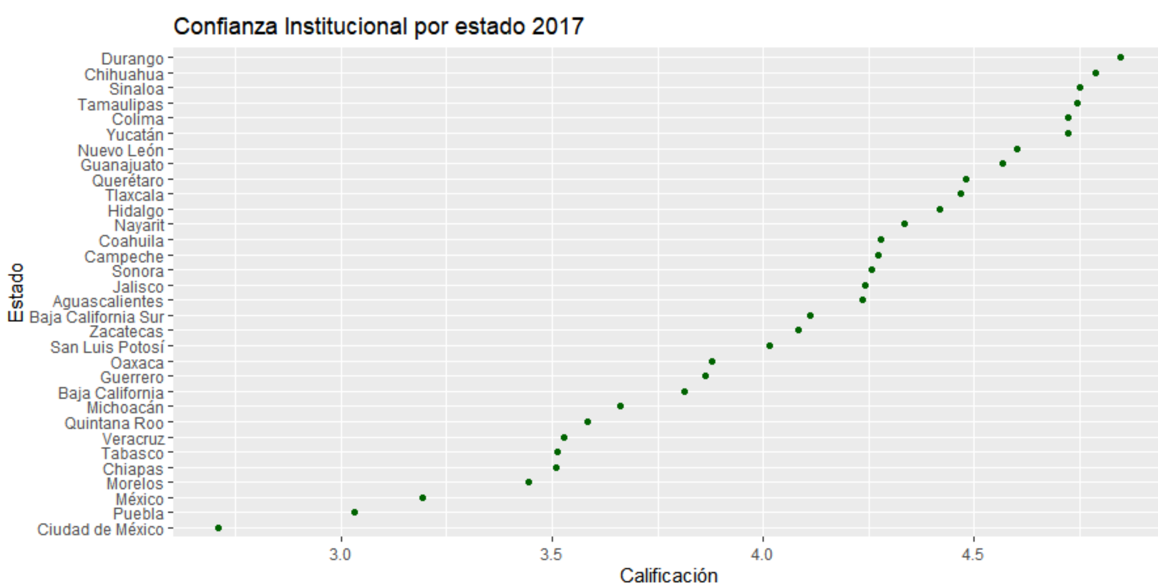


Figura 6.1: Confianza institucional por estado en el año 2017.

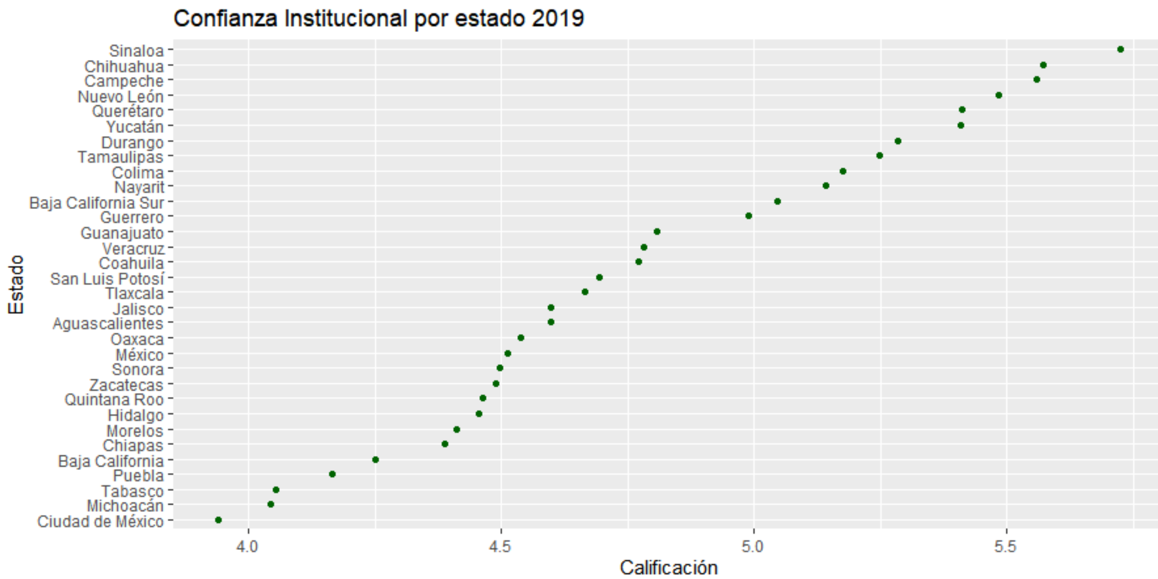


Figura 6.2: Confianza institucional por estado en el año 2019.

Primero, vea que la evaluación promedio que dan a las instituciones se diferencia por estado, tanto en el año 2017 6.1 como en el año 2019 6.2, aunque gráficamente se puede notar cierta diferencia al momento de evaluar a sus instituciones, es necesario probar estas hipótesis.

Lo cual se comprueba generando un modelo multinivel nulo, el cual consiste de solo tener el efecto aleatorio como variable explicativa, observe en la tabla 6.1 que el valor de la correlación intraclase está entre un 2.2 % y 7 %, lo cual indica que el porcentaje de la varianza del modelo se explica por los estados en los porcentajes anteriormente mencionados. Estos procentajes a pesar de ser relativamente pequeños al contrastar los valores BIC de dicho modelo nulo con un modelo de regresión que solo tiene la variable del Estado, el modelo nulo parece tener un mejor ajuste que el segundo modelo.

Solo para representar, el modelo nulo sería:

$$Confianza_{Presidencia} = \beta_{0_j} + \epsilon = 3.9 + u_j,$$

donde u_j es el parámetro relacional al efecto aleatorio que tiene una varianza estimada de 0.32 (ver parámetros estimados en la Tabla 6.1). Una formación análoga se utiliza para la Confianza de las demás entidades de Gobernatura, Presidencia municipal, etc. para 2017 y 2019.

Siguiendo, aquellos modelos que tienen como variable respuesta al índice de confianza institucional señalan que este tipo de modelo mejora en contraste a los modelos que solo tienen a la institución como variable respuesta. Sin embargo, la razón principal para implementar este tipo de modelo es por la característica de los datos, pues están caracterizados por la entidad en que se localizan los encuestados.

2017	**Presidencia**	**Gobernatura**	**Presidencia municipal**	**Congreso**
Constante fija	3.9 (0.10)***	4.2 (0.12)***	4.5 (0.12)***	3.6(0.09)***
Var. intercepto aleatorio	0.32 (0.07)	0.47 (0.09)	0.47 (0.08)	0.26 (0.07)
Residuo	6.37 (0.01)	6.40 (0.01)	6.32 (0.01)	5.68 (2.38)
BIC	177497.8	177930.5	177091.5	169014.9
Correlación intraclase	0.047	0.068	0.069	0.043
N	37815	37861	37788	36905
Grupos	32	32	32	32
2017	**Jueces y magistrados**	**Partidos políticos**	**Ministerios públicos**	**ICI**
Constante fija	4.6 (0.09)***	3.3 (0.08)***	4.4 (0.11)***	4.1 (0.09)***
Var. intercepto aleatorio	0.29 (0.07)	0.24 (0.07)	0.39 (0.08)	0.30 (0.08)
Residuo	6.18 (0.01)	5.34 (0.01)	6.15 (0.01)	3.97 (0.01)
BIC	167789.2	172341	173292.6	155444.5
Correlación intraclase	0.044	0.043	0.059	0.070
N	35973	38148	37189	38148
Grupos	32	32	32	32
2019	**Presidencia**	**Gobernatura**	**Presidencia municipal**	**Congreso**
Constante fija	5.7 (0.07)***	4.9 (0.09)***	5.1 (0.09)***	4.4 (0.07)***
Var. intercepto aleatorio	0.16 (0.07)	0.28 (0.09)	0.25 (0.08)	0.19 (0.7)
Residual	7.12 (0.01)	6.43 (0.01)	6.07 (0.01)	5.81 (0.01)
BIC	181229.5	176640.6	174314.9	164681.7
Correlación intraclase	0.022	0.042	0.039	0.031
N	37719	37549	37515	35784
Grupos	32	32	32	32
2019	**Jueces y magistrados**	**Partidos políticos**	**Ministerios públicos**	**ICI**
Constante fija	4.8 (0.08)***	3.9 (0.08)***	4.7 (0.10)***	4.8 (0.07)***
Var. intercepto aleatorio	0.21 (0.07)	0.23 (0.07)	0.34 (0.08)	0.18 (0.07)
Residuo	5.84 (0.01)	5.71 (0.01)	6.09 (0.01)	3.57 (0.01)
BIC	163057.1	173134.3	172000.3	155375.2
Correlación intraclase	0.034	0.038	0.052	0.048
N	35385	37764	36991	37764
Grupos	32	32	32	32

Errores estándar entre paréntesis. Significancia:* $p < 0.1$, ** $p < 0.01$, *** $p < 0.001$

Tabla 6.1: Modelos multinivel nulos: el efecto estatal sobre la confianza institucional en los años 2017 y 2019.

2017	**Presidencia**	**Gobernatura**	**Presidencia municipal**	**Congreso**
Servicios públicos				
Agua	0.39 (0.04)***	0.39 (0.04)***	0.38 (0.04)***	0.39 (0.04)***
Drenaje y alcantarillado	0.21 (0.04)***	0.21 (0.04)***	0.29 (0.04)***	0.21 (0.04)***
Alumbrado público	0.32 (0.03)***	0.36 (0.03)***	0.34 (0.03)***	0.32 (0.03)***
Parques y jardines	0.19 (0.04)***	0.27 (0.04)***	0.38 (0.04)***	0.16 (0.04)***
Policías	1.35 (0.03)***	1.41 (0.03)***	1.33 (0.03)***	1.15 (0.03)***
Calles y avenidas	0.49 (0.04)***	0.41 (0.04)***	0.44 (0.04)***	0.41 (0.04)***
Señalamiento urbano	0.07 (0.03)*	0.13 (0.03)***	0.22 (0.03)***	0.09 (0.03)**
Carreteras públicas	0.62 (0.04)***	0.70 (0.04)***	0.65 (0.04)***	0.58 (0.04)***
Controles individuales				
Educación: primaria o menos				
Básica	−0.10 (0.04)**	−0.02 (0.04)	−0.03 (0.03)	−0.15 (0.04)***
Media superior	−0.40 (0.04)***	−0.16 (0.04)***	−0.20 (0.04)***	−0.45 (0.04)***
Universidad o más	−0.65 (0.04)***	−0.20 (0.04)***	−0.31 (0.04)***	−0.67 (0.04)***
Sexo: mujer	−0.05 (0.02)*	−0.10 (0.02)***	−0.02 (0.02)	−0.02 (0.02)
Edad	−0.01 (0.00)***	0.00 (0.00)	0.00 (0.00)	−0.02 (0.00)***
Edad^2	0.00 (0.00)***	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)***
Ocupación: en actividad económica	−0.13 (0.03)***	−0.05 (0.02)*	0.00 (0.02)	−0.07 (0.02)*
Controles estatales				
Población 2015 (log)	−0.04 (0.09)	−0.09 (0.13)	−0.06 (0.11)	−0.07 (0.08)
PIB %	0.01 (0.02)	0.02 (0.03)	0.03 (0.03)	0.01 (0.02)
Alternancia	−0.21 (0.13)	0.00 (0.19)	0.00 (0.16)	−0.02 (0.12)
Homicidios 2016	0.01 (0.00)*	0.00 (0.00)	0.00 (0.00)	0.01 (0.00)*
Incidencia delictiva (log)	−0.12 (0.26)	−0.24 (0.35)	−0.35 (0.32)	−0.02 (0.25)
Intercepto	5.61 (2.65)*	6.82 (3.57)*	7.71 (3.26)*	4.55 (2.48)*
Componentes de la varianza				
Varianza del intercepto aleatorio	0.11 (0.05)	0.23 (0.07)	0.17 (0.06)	0.09 (0.05)
Residuo	5.58 (0.01)	5.58 (0.01)	5.50 (0.01)	5.05 (0.01)
N	37815	37861	37788	36905
Grupos	32	32	32	32
BIC	172764.7	173001.8	172134.1	164946.7

Modelos lineales multinivel. Errores estándar entre paréntesis. Significancia: * $p < 0.1$, ** $p < 0.01$, *** $p < 0.001$

Tabla 6.2: Análisis de la calidad percibida en los servicios públicos locales y la confianza en las instituciones políticas (2017).

2017	**Jueces y magistrados**	**Partidos políticos**	**Ministerios públicos**	**ICI**
Servicios públicos				
Agua	0.38 (0.05)***	0.34 (0.04)***	0.37 (0.04)***	0.37 (0.03)***
Drenaje y alcantillado	0.23 (0.04)**	0.20 (0.04)***	0.24 (0.04)***	0.22 (0.03)***
Alumbrado público	0.23 (0.03)***	0.28 (0.03)***	0.28 (0.03)***	0.31 (0.03)***
Parques y jardines	0.31 (0.04)***	0.16 (0.04)***	0.25 (0.04)***	0.24 (0.03)***
Policías	1.23 (0.03)***	1.09 (0.03)***	1.43 (0.03)***	1.28 (0.02)***
Calles y avenidas	0.27 (0.04)***	0.47 (0.04)***	0.36 (0.04)***	0.40 (0.03)***
Señalamiento urbano	0.15 (0.03)***	0.05 (0.03)	0.14 (0.03)***	0.12 (0.02)***
Carreteras públicas	0.69 (0.04)***	0.51 (0.04)***	0.67 (0.04)***	0.63 (0.03)***
Controles individuales				
Educación: primaria o menos				
Básica	0.00 (0.03)	−0.07 (0.03)*	−0.04 (0.04)	−0.05 (0.03)*
Media superior	−0.10 (0.04)*	−0.34 (0.04)***	−0.28 (0.04)***	−0.27 (0.03)***
Universidad o más	−0.12 (0.04)**	−0.59 (0.03)***	−0.50 (0.04)***	−0.43 (0.03)***
Sexo: mujer	−0.02 (0.02)	0.07 (0.02)**	0.05 (0.02)*	−0.01 (0.02)
Edad	−0.01 (0.00)***	−0.01 (0.00)***	−0.02 (0.00)***	−0.01 (0.00)
Edad^2	0.00 (0.00)*	0.00 (0.00)***	0.00 (0.00)***	0.00 (0.00)***
Ocupación: en actividad económica	−0.02 (0.03)	−0.04 (0.03)	−0.07 (0.03)**	−0.05 (0.02)*
Controles estatales				
Población 2015 (log)	−0.11 (0.08)	−0.07 (0.08)	−0.12 (0.09)	−0.08 (0.08)
PIB %	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.02 (0.02)
Alternancia	−0.10 (0.12)	0.01 (0.12)	−0.05 (0.13)	−0.06 (0.12)
Homicidios 2016	0.00 (0.00)	0.01 (0.00)*	0.00 (0.00)	0.01 (0.00)*
Incidencia delictiva (log)	−0.00 (0.20)	−0.06 (0.24)	−0.06 (0.29)	−0.13 (0.25)
Intercepto	5.40 (2.45)*	4.58 (2.38)*	6.26 (2.91)*	5.86 (2.50)*
Componentes de la varianza				
Varianza del intercepto aleatorio	0.08 (0.04)	0.09 (0.04)	0.10 (0.05)	0.09 (0.04)
Residuo	5.55 (0.01)	4.79 (0.01)	5.35 (0.01)	3.28 (0.01)
N	35973	38148	37189	38148
Grupos	32	32	32	32
BIC	164182.6	168455.7	168321.9	153976.1

Tabla 6.3: Análisis de la calidad percibida en los servicios públicos locales y la confianza en las instituciones políticas (2017).

2019	**Presidencia**	**Gobernatura**	**Presidencia municipal**	**Congreso**
Servicios públicos				
Agua	0.25 (0.05)***	0.45 (0.04)***	0.41 (0.04)***	0.44 (0.04)***
Drenaje y alcantarillado	0.10 (0.05)*	0.16 (0.04)***	0.26 (0.04)***	0.20 (0.04)***
Alumbrado público	0.18 (0.04)***	0.32 (0.03)***	0.37 (0.03)***	0.29 (0.03)***
Parques y jardines	0.32 (0.05)***	0.32 (0.04)***	0.32 (0.04)***	0.25 (0.04)***
Policías	0.92 (0.03)***	1.17 (0.03)***	1.17 (0.03)***	0.97 (0.03)***
Calles y avenidas	0.21 (0.05)***	0.44 (0.04)***	0.44 (0.04)***	0.33 (0.04)***
Señalamiento urbano	0.15 (0.03)***	0.19 (0.03)***	0.22 (0.03)***	0.16 (0.03)***
Carreteras públicas	0.65 (0.04)***	0.71 (0.04)***	0.64 (0.04)***	0.68 (0.04)***
Controles individuales				
Educación: primaria o menos				
Básica	−0.09 (0.04)*	−0.19 (0.04)***	−0.11 (0.03)**	−0.15 (0.04)***
Media superior	−0.21 (0.04)***	−0.29 (0.04)***	−0.15 (0.04)***	−0.28 (0.04)***
Universidad o más	−0.51 (0.04)***	−0.34 (0.04)***	−0.26 (0.04)***	−0.45 (0.04)***
Sexo: mujer	−0.28 (0.03)***	−0.05 (0.02)*	0.02 (0.03)	−0.02 (0.03)
Edad	0.02 (0.00)***	0.00 (0.00)	0.00 (0.00)	−0.02 (0.00)***
Edad^2	0.00 (0.00)*	0.00 (0.00)*	0.00 (0.00)	0.00 (0.00)***
Ocupación: en actividad económica	−0.12 (0.03)***	−0.09 (0.03)**	−0.01 (0.03)	−0.08 (0.03)**
Controles estatales				
Población 2015 (log)	0.01 (0.09)	0.08 (0.12)	0.08 (0.09)	0.11 (0.09)
PIB %	−0.11 (0.03)***	−0.01 (0.03)	−0.02 (0.03)	−0.05 (0.03)*
Alternancia	0.16 (0.14)	0.06 (0.17)	0.01 (0.13)	0.04 (0.13)
Homicidios 2018	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.005 (0.003)*
Incidencia delictiva (log)	−0.12 (0.26)	−0.40 (0.30)	−0.15 (0.24)	−0.10 (0.27)*
Intercepto	3.18 (2.55)	7.10 (3.01)*	3.89 (2.45)	4.24 (2.74)
Componentes de la varianza				
Varianza del intercepto aleatorio	0.11 (0.05)	0.19 (0.06)	0.10 (0.05)	0.12 (0.05)
Residuo	6.60 (0.01)	5.70 (0.01)	5.35 (0.01)	5.26 (0.01)
N	37719	37549	37515	35784
Grupos	32	32	32	32
BIC	178656.1	172388.3	169860.3	161391.4

Tabla 6.4: Análisis de la calidad percibida en los servicios públicos locales y la confianza en las instituciones políticas (2019).

2019	**Jueces y magistrados**	**Partidos políticos**	**Ministerios públicos**	**ICI**
Servicios públicos				
Agua	0.39 (0.05)***	0.37 (0.04)***	0.38 (0.04)***	0.39 (0.03)***
Drenaje y alcantarillado	0.22 (0.04)***	0.19 (0.04)***	0.20 (0.04)***	0.18 (0.03)***
Alumbrado público	0.24 (0.04)***	0.27 (0.03)***	0.29 (0.03)***	0.27 (0.02)***
Parques y jardines	0.31 (0.04)***	0.17 (0.04)***	0.27 (0.04)***	0.28 (0.03)***
Policías	1.06 (0.03)***	1.02 (0.03)***	1.33 (0.03)***	1.09 (0.02)***
Calles y avenidas	0.25 (0.04)***	0.43 (0.04)***	0.32 (0.04)***	0.32 (0.03)***
Señalamiento urbano	0.19 (0.03)***	0.10 (0.03)**	0.10 (0.03)***	0.15 (0.02)***
Carreteras públicas	0.65 (0.04)***	0.53 (0.04)***	0.65 (0.04)***	0.64 (0.03)***
Controles individuales				
Educación primaria o menos				
Básica	−0.10 (0.04)*	−0.11 (0.04)**	−0.23 (0.04)***	−0.12 (0.03)***
Media superior	−0.19 (0.04)***	−0.30 (0.04)***	−0.34 (0.04)***	−0.23 (0.03)***
Universidad o más	−0.24 (0.04)***	−0.55 (0.04)***	−0.55 (0.04)***	−0.39 (0.03***
Sexo: mujer	0.08 (0.02)**	0.08 (0.02)**	0.11 (0.03)***	−0.01 (0.02)
Edad	−0.03 (0.00)***	−0.02 (0.00)***	−0.04 (0.00)***	−0.01 (0.00)***
Edad^2	0.00 (0.00)***	0.00 (0.00)***	0.00 (0.00)	0.00 (0.00)
Ocupación: en actividad económica	−0.05 (0.03)*	−0.08 (0.03)**	−0.09 (0.03)**	−0.07 (0.02)**
Controles estatales				
Población 2015 (log)	0.02 (0.08)	0.14 (0.10)	−0.06 (0.10)	0.05 (0.08)
PIB %	−0.03 (0.02)	−0.03 (0.03)	−0.01 (0.03)	−0.04 (0.02)
Alternancia	0.01 (0.13)	0.02 (0.15)	0.05 (0.15)	0.06 (0.12)
Homicidios 2018	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Incidencia delictiva (log)	−0.05 (0.25)	−0.16 (0.29)	−0.06 (0.31)	−0.11 (0.24)
Intercepto	5.46 (2.50)*	3.74 (2.89)	7.12 (3.06)*	4.92 (2.36)*
Componentes de la varianza				
Varianza del intercepto aleatorio	0.10 (0.05)	0.14 (0.05)	0.14 (0.05)	0.07 (0.04)
Residuo	5.29 (0.01)	5.19 (0.01)	5.34 (0.01)	3.00 (0.01)
N	35385	37764	36991	37764
Grupos	32	32	32	32
BIC	159778.7	169793.5	167411.7	149076.7

Tabla 6.5: Análisis de la calidad percibida en los servicios públicos locales y la confianza en las instituciones políticas (2019).

En consecuencia, se sugiere usar el modelo multinivel, en donde las variables dependientes son las mismas, pero en este caso sí se añadirán todas las variables independientes que se han mencionado en la sección que trata dichas variables, vea las tablas 6.2 a 6.5. Este análisis permite monstrarnos que los servicios públicos tienen un papel importante en la evaluación hacia las instituciones, asimismo se ve que el ajuste que tienen los modelos con las variables explicativas es mejor que cuando se genera solo el modelo multinivel nulo.

Vea por ejemplo el modelo generado para la confianza en la presidencia en el año 2019 (ver coeficientes estimados en la segunda columna de la Tabla 6.4):

$$
\begin{aligned}
Confianza_{Presidencia} = 3.18 &+ 0.25Agua + 0.10Drenaje + 0.18Alumbrado \\
&+ 0.32Parques + 0.92Policias + 0.21Calles \\
&+ 0.15Senalamiento + 0.65Carreteras - 0.09Edu\_Bas \\
&- 0.21Edu\_Med - 0.51Edu\_Uni - 0.28Mujer + 0.02Edad \\
&- 0.12Act\_eco + 0.01Poblacion - 0.11PIB \\
&+ 0.16Alternancia - 0.12Inc\_delictiva + Estado.
\end{aligned}
$$

En este caso, vemos que la variable que mayor peso tiene es la evaluación hacia los policías incrementando su confianza en casi el valor de 1 si esta institución ofrece un buen servicio. Recordemos que dicha variable está entre 0 y 1. Otro aspecto a considerar es la edad, que incrementa un número en la evaluación si el individuo tiene 50 años. Sin embargo, no en todas las instituciones la edad resulta ser un factor importante a la hora de evaluar una institución. Vea que al final se agrega una variable llamada Estado, la cual no cuenta con un coeficiente, pues es el efecto aleatorio, es decir, que el valor va cambiando de acuerdo al estado. Este modelo es idéntico en cada una de las instituciones así como el ICI, tanto para el año 2017 como el año 2019, formulaciones análogas se obtienen para la Confianza para las demás entidades de Gobernatura, Presidencia Municipal, etc. para el 2017 y 2019, usando los coeficientes estimados correspondientes de las Tablas 6.2, 6.3, 6.4 y 6.5.

Todos los servicios públicos llevan a evaluar de manera más positiva a las instituciones, esto dará pie a generar un tercer modelo que considere la calidad promedio de los servicios por estado (un modelo multinivel con una pendiente aleatoria).

Además de los servicios públicos locales que se podrían considerar que están relacionados al estado en que habita el encuestado, se observa una relación entre el grado de estudios de una persona y su manera de evaluar a las instituciones. Una persona va a evaluar de peor manera a sus instituciones si ésta tiene más grado de estudios. También se puede ver que casi en todos los casos una persona que realiza una actividad económica también tiende a evaluar con un puntaje menor a sus instituciones que los que no realizan una actividad económica.

Asimismo, observe que tener un alto índice de homicidios en general no tiene ningún impacto en la evaluación, a excepción del año 2017, en donde se observa una mayor calificación en algunos casos.

Variable	**Primer modelo**	**Segundo modelo**	**Tercer modelo**
IPCSL	3.82 (0.05)***	3.77 (0.05)***	3.78 (0.12)***
IPCSL (Promedio estatal)			2.05 (0.86)*
Controles individuales			
Educación: primaria o menos			
Básica		−0.07 (0.03)*	−0.06 (0.03)*
Media superior		−0.29 (0.03)***	−0.29 (0.03)***
Universidad o más		−0.46 (0.03)***	−0.46 (0.03)***
Sexo: mujer		−0.03 (0.02)	−0.03 (0.02)
Edad		−0.01 (0.00)*	−0.01 (0.00)***
Edad^2		0.00 (0.00)***	0.00 (0.00)**
Ocupación: en actividad económica		−0.07 (0.02)**	−0.07 (0.02)**
Controles estatales			
Población 2015 (log)		−0.10 (0.08)	−0.06 (0.08)
PIB %		0.01 (0.02)	−0.00 (0.03)
Alternancia		−0.05 (0.12)	−0.02 (0.12)
Homicidios 2016		0.00 (0.00)	0.01 (0.00)*
Incidencia delictiva (log)		−0.10 (0.25)	−0.10 (0.22)
Intercepto	2.42 (0.07)***	5.89 (2.53)*	4.36 (2.32)*
Componentes de la varianza			
Pendiente aleatoria(IPCSL)			0.39 (0.11)
Varianza del intercepto aleatorio	0.14 (0.05)	0.09 (0.04)	0.14 (0.07)
Covarianza (pendiente, intercepto)			−0.14
Residuo	3.42 (0.01)	3.39 (0.01)	3.38 (0.14)
Correlación intraclase	0.039	0.026	0.036
N	38148	38148	38148
Grupos	32	32	32
BIC	155382.5	155185.8	155213.8

Tabla 6.6: Modelos multinivel: el efecto estatal sobre la confianza institucional en el año 2017.

También, es interesante que en instituciones como la presidencia, gobierno estatal y el congreso, la mujeres tienden a dar una menor calificación, mientras que a los ministerios públicos y partidos políticos es lo contrario, pues los hombres suelen evaluar de peor manera. Mientras que la edad muestra que una persona con mayor edad evalúa con menor calificación que una más joven, aunque con un factor de una décima por año.

Variable	**Primer modelo**	**Segundo modelo**	**Tercer modelo**
IPCSL	3.53 (0.05)***	3.48 (0.05)***	3.52 (0.09)***
IPCSL (Promedio estatal)			0.18 (1.00)
Controles individuales			
Educación: primaria o menos			
Básica		−0.14 (0.03)***	−0.13 (0.03)***
Media superior		−0.24 (0.03)***	−0.24 (0.03)***
Universidad o más		−0.41 (0.03)***	−0.40 (0.03)***
Sexo: mujer		−0.01 (0.02)	−0.01 (0.02)
Edad		−0.01 (0.00)***	−0.01 (0.00)***
Edad^2		0.00 (0.00)***	0.00 (0.00)***
Ocupación: en actividad económica		−0.08 (0.02)**	−0.07 (0.02)**
Controles estatales			
Población 2015 (log)		0.02 (0.09)	0.03 (0.10)
PIB %		−0.04 (0.03)	−0.04 (0.03)
Alternancia		0.07 (0.13)	0.08 (0.14)
Homicidios 2016		0.00 (0.00)	0.00 (0.00)
Incidencia delictiva (log)		−0.11 (0.25)	−0.10 (0.26)
Intercepto	3.22 (0.07)***	5.48 (2.54)*	5.20 (2.76)*
Componentes de la varianza			
Pendiente aleatoria(IPCSL)			0.18 (0.14)
Varianza del intercepto aleatorio	0.12 (0.04)	0.11 (0.05)	0.15 (0.13)
Covarianza (pendiente, intercepto)			−0.07
Residuo	3.10 (0.01)	3.08 (0.01)	3.07 (0.57)
Correlación intraclase	0.037	0.034	0.044
N	37764	37764	37764
Grupos	32	32	32
BIC	150035.4	149989.6	149953.5

Tabla 6.7: Modelos multinivel: el efecto estatal sobre la confianza institucional en el año 2019.

Aunque estas variables son interesantes la mayoría de ellas, sobre todo las que están relacionadas a las condiciones estatales, se puede atribuir estos ligeros cambios, y en la mayoría de ellos su insignificancia, a que son cambios que están condicionados al estado en que residen.

De manera general, se puede ver que en ocasiones el tener una mejor calidad de servicios ayuda más a una institución que a otra. Sin embargo, este servicio puede no estar directamente relacionado con la institución, es decir, que la calidad de un servicio incremente la confianza a una institución que ni siquiera la provee. Por ejemplo, en el año 2019, los partidos políticos se ven principalmente beneficiados por la calidad de las calles y avenidas, o el caso de los parques y jardines, donde un mejor servicio de éstos generan el mismo incremento en la confianza tanto a los gobiernos estatal, federal y municipal como a los jueces y magistrados.

Estas pequeñas variaciones se pueden explicar de distintas maneras, pero para ver de mejor manera la calidad de los servicios se genera una última tabla que compara tres tipos de modelos. El primer modelo evalúa solo el efecto aleatorio de residir en un estado con respecto a la confianza institucional. El segundo modelo evalúa la misma confianza en las instituciones pero con el IPCSL. Finalmente, el tercer modelo evalúa lo mismo solo que en este caso se añade una variable que mide el IPCSL por estado, así como una pendiente aleatoria que está relacionada al IPCSL que se registra en cada estado.

Tomando de ejemplo los modelos para el año 2017. El primer modelo sería de la forma (ver parámetros estimados en la primera columna de la Tabla 6.6):

$$ICI = 2.42 + 3.82 IPCSL + Estado.$$

El modelo 2 (ver parámetros estimados en la segunda columna de la Tabla 6.6):

$$\begin{aligned} ICI = 5.89 &+ 3.77 IPCSL - 0.07 Edu\_Bas - 0.29 Edu\_med - 0.46 Edu\_Uni \\ &- 0.03 Mujer - 0.01 Edad - 0.07 Act\_eco - 0.10 Poblacion + 0.01 PIB \\ &- 0.05 Alternancia - 0.66 Inc\_delictiva + Estado. \end{aligned}$$

Y el modelo 3 (ver parámetros estimados en la tercera columna de la Tabla 6.6):

$$\begin{aligned} ICI = 4.36 &+ 3.78 IPCSL - 0.06 Edu\_Bas - 0.24 Edu\_med - 0.41 Edu\_Uni \\ &- 0.01 Mujer - 0.01 Edad - 0.08 Act\_eco + 0.02 Poblacion - 0.04 PIB \\ &+ 0.07 Alternancia - 0.11 Inc\_delictiva + Estado + 2.05 IPCSL_{Estado}. \end{aligned}$$

Así como para el caso del 2017, los tres modelos correspondientes al 2019 son análogos, usando las parámetros estimados correspondientes.

En estos tres modelos se valida una hipótesis que se tenía acerca de que los servicios públicos se pueden resumir en un índice. Pues solo con el índice se muestra que este factor es significativo para la confianza en las instituciones. En cuanto a los factores asociados al género, solo el sexo no es significativo, mientras que las demás variables tienen un impacto negativo en dicha confianza, por ejemplo, entre mayor sea la edad y más grados de estudios se tenga, la confianza decae, así como el estar realizando una actividad económica. Sin embargo, en los controles estatales, solo la variable de delitos tienen un valor significativo en la calificación de confianza institucional.

Las variables relacionadas a los controles estatales no son significativas esto ya que esta información puede estar contenida en el estado o al IPCSL que está relacionado al estado.

Tanto el primer modelo como el segundo modelo se pueden considerar como buenos, pues en la prueba BIC ambos mejoran, aunque el primer modelo resulta ser mejor sin las variables relacionadas a las características de los encuestados en el año 2017, lo cual no sucede en el año 2019. Por otra parte, el IPCSL estatal no es significativa en ninguno de los dos años, esto se puede ver en el tercer modelo de cada tabla, además que solo parece ajustar mejor el modelo en el año 2019. Por lo cual, es posible omitir esta variable, pues no es significativa en ninguno de los modelos de estos dos años.

Para finalizar, los residuos de cada uno de los modelos fueron evaluados bajo la prueba de Brown-Forsyth para determinar si había diferencias entre grupos (Estados), lo cual fue así para un nivel de significancia menor a 0.001, además bajo la prueba de normalidad de Kolmogorov-Smirnov se prueba la normalidad de dichos residuos, los cuáles también resultaron tener una distribución normal a un nivel de confianza menor a 0.001 la cual comprueba el supuesto de normalidad de los residuos en ded modelo. Por lo cual, se puede concluir que los modelos son adecuados para representar los datos.

6.2. Modelo multinivel con datos categóricos corrupción institucional

Para este segundo análisis se presentan dos tipos de análisis, uno multinivel lineal, como en la sección anterior y modelos multinivel con respuesta ordinal. La variable que se predice en los modelos ordinales es la frecuencia de actos de corrupción en cada una de las instituciones clasificadas en los siguientes valores: *1 (muy frecuente), 2 (frecuente), 3 (poco frecuente) y 4 (nunca)*. Siendo de esta manera, se obtiene un orden donde el valor 1 implica una alta frecuencia alta, y si el valor incrementa esta frecuencia disminuye.

Por otro lado, para el modelo que tendrá una variable continua se basa en el IPCI, vea una comparación de dicho IPCI de acuerdo a cada estado en los años 2017 y 2019, figuras 6.3 y 6.4 en donde se puede observar que hay una variación por estado y esto se comprueba al implementar los modelos nulos como se realizó en la sección pasada.

Además, es interesante que la percepción de la corrupción se sienta más frecuente en los estados del centro del país: México, Morelos y la Ciudad de México. Mientras que en los estados fronterizos de Tamaulipas y Durango se perciba una menor corrupción. Esto se conserva en los dos años donde se evaluó esta percepción. Aunque como se mencionó, hay un incremento, pero este incremento pasa en prácticamente todos los estados.

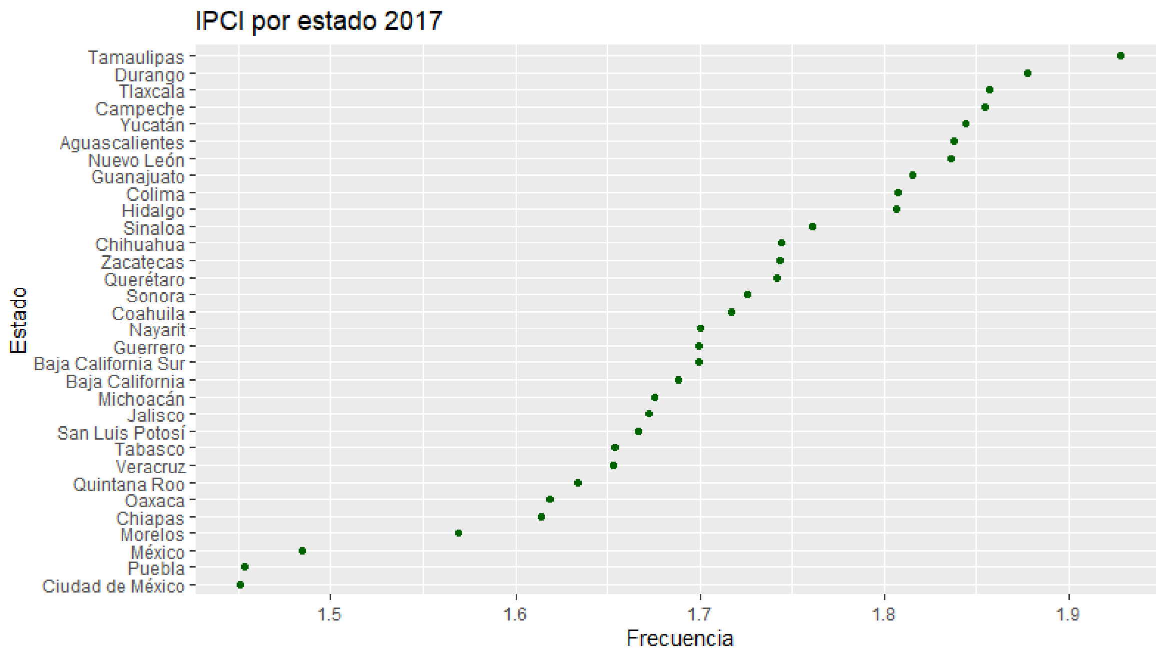


Figura 6.3: Índice de la percepción de la corrupción institucional por estado en el año 2017.

Además, vea que en estos modelos multinivel nulos presentados en la tabla 6.8, a partir de los interceptos se puede inferir la probabilidad de que una persona pertenezca a cierta categoría, dichos interceptos se decriben como umbrales en la tabla, estos valores hacen que la probabilidad de clasificar como no corrupto vaya incrementando. Lo cual está asociado a la cantidad de personas que percibieron la corrupción con cierta frecuencia, de esta manera, se tendría que lo que piensan la mayoría de las personas tendría más probabilidad de ocurrir que las otras opciones. Recuerde que de (4.31), al aplicarse al modelo para la percepción de la corrupción en la presidencia se tiene lo siguiente (ver parámetros estimados en la Tabla 6.8):

$$P[Y = 4] = P[Y \leq 4] - P[Y \leq 3] = 1 - \frac{e^{\beta_1^{(3)}}}{1 + e^{\beta_1^{(3)}}} = 1 - \frac{e^{3.96}}{1 + e^{3.96}} = 1 - 0.98 = 0.02.$$

En general el modelo multinivel nulo estaría escrito de las siguiente manera:

$$P[Y = s] = P[Y \leq s] - P[Y \leq s-1] = \frac{e^{Umbral(s|s+1)+Estado}}{1 + e^{Umbral(s|s+1)+Estado}} - \frac{e^{Umbral(s-1|s)+Estado}}{1 + e^{Umbral(s-1|s)+Estado}}.$$

Del primer caso se puede decir que la probabilidad de que una persona perciba que la corrupción en la presidencia es de nivel 4 (nunca se cometen actos de corrupción) es de tan solo 2 %. Si esto se reproduce con los siguientes modelos, se observará que la probabilidad de que una persona perciba con 1 (muy frecuente) la corrupción en el año 2017 es mayor en casi todas las instituciones, mientras que en el año 2019 es más probable que la persona perciba como frecuente la corrupción en las instituciones. Lo cual, se asemeja al comportamiento de la figura 5.3.

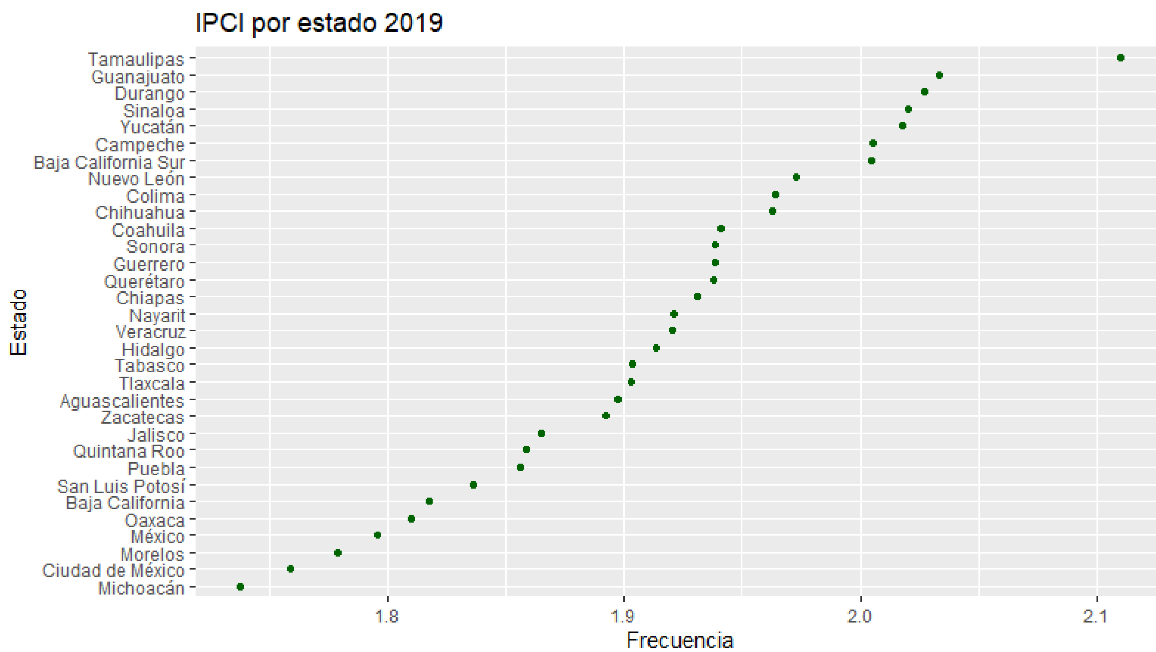


Figura 6.4: Índice de percepción de la corrupción institucional por estado en el año 2019.

Esta información se trata de resumir con el modelo lineal multinivel que usa el IPCI, vea la tabla 6.9.

Como se puede ver el intercepto varía de manera considerable, mientras que en el año 2017 se obtenía un IPCI de 1.71, lo que indica una percepción entre frecuente y muy frecuente en los actos de corrupción; en el año 2019, el IPCI alcanza hasta 3.09, lo que indica más bien una percepción de poco frecuente, con una ligera variación de los estados que representa el 2.9 % y 2.8 %, con respecto a los años mencionados. Recuerde que los primeros valores representan un valor promedio de la percepción de la corrupción, sin importar el estado.

De esta manera tenemos el siguiente modelo para el IPCI, usando el año 2017 (ver estimaciones de los parámetros de la tabla 6.9):

$$IPCI = 1.71 + Estado$$

2017	**Presidencia**	**Gobernatura**	**Presidencia municipal**	**Congreso**
Umbral (1|2)	−0.35 (0.05)	−0.03 (0.06)	−0.31 (0.06)	0.26 (0.05)
Umbral (2|3)	2.06 (0.05)	1.72 (0.07)	1.48 (0.07)	2.11 (0.06)
Umbral (3|4)	3.96 (0.06)	3.66 (0.07)	3.37 (0.07)	4.06 (0.07)
Varianza del intercepto aleatorio	0.08	0.13	0.13	0.09
BIC	71214.7	78950	83264.3	70497.1
N	37468	37709	37550	36908
2017	**Jueces y magistrados**	**Partidos políticos**	**Ministerios públicos**	
Umbral (1|2)	−0.63 (0.06)	0.48 (0.05)	−0.31 (0.07)	
Umbral (2|3)	1.09 (0.06)	2.35 (0.05)	1.44 (0.07)	
Umbral (3|4)	3.02 (0.06)	4.20 (0.07)	3.28 (0.07)	
Varianza del intercepto aleatorio	0.09	0.08	0.13	
BIC	84403.7	68164.9	81775.9	
N	35662	38276	37468	
2019	**Presidencia**	**Gobernatura**	**Presidencia municipal**	**Congreso**
Umbral (1|2)	−0.69 (0.04)	−0.54 (0.05)	−0.77 (0.04)	−0.46 (0.04)
Umbral (2|3)	0.85 (0.04)	1.32 (0.05)	1.14 (0.04)	1.54 (0.04)
Umbral (3|4)	2.43 (0.04)	3.33 (0.06)	3.05 (0.05)	3.39 (0.05)
Varianza del intercepto aleatorio	0.04	0.08	0.06	0.06
BIC	91624.9	85014.6	86600.1	78620.8
N	36128	36854	36642	35508
2019	**Jueces y magistrados**	**Partidos políticos**	**Ministerios públicos**	
Umbral (1|2)	−0.81 (0.04)	0.02 (0.04)	−0.46 (0.06)	
Umbral (2|3)	1.10 (0.05)	1.97 (0.04)	1.34 (0.06)	
Umbral (3|4)	3.03 (0.05)	3.80 (0.05)	3.24 (0.06)	
Varianza del intercepto aleatorio	0.06	0.06	0.10	
BIC	82013.7	76806.7	82623.6	
N	34943	37815	36662	

Tabla 6.8: Modelos multinivel nulos: el efecto estatal sobre la percepción de la corrupción institucional en los años 2017 y 2019.

	IPCI 2017	IPCI 2019
Intercepto	1.71 (0.02)***	3.09 (0.02)***
Varianza del intercepto aleatorio	0.01 (0.01)	0.01 (0.01)
Residual	0.33 (0.00)	0.34 (0.00)
Correlación intraclase	0.029	0.028
BIC	66857.7	69387.8
N	38850	39118
Grupos	32	32

Tabla 6.9: Modelos multinivel nulos: el efecto estatal sobre el IPCI en los años 2017 y 2019.

Esta información inicial nos ayuda a ver que hubo un cambio en la percepción sobre qué tan frecuente son los actos de corrupción de sus instituciones. Sin embargo, no se tiene información acerca de qué es lo que provoca este cambio. Lo que sí se puede apreciar, es la relación que guarda esta percepción con los servicios o las condiciones de vida del individuo. De la cual se puede ver cuáles son los factores que hacen que una persona perciba más o menos corrupta una institución, y esto ayudaría a entender un poco a cerca de porqué se da el cambio anteriormente mencionado. Es por ello que se realizan análisis introduciendo más variables al modelo.

Pasando al segundo análisis, vea esta información contenida en las tablas de 6.10 a 6.13, en el cual se observa que hay un comportamiento parecido a la información sobre la confianza institucional.

Tomemos de ejemplo el modelo para la percepción de la corrupción en la Presidencia del año 2017 (vea la primer fila de la Tabla 6.10):

$$P[Perc\_corrup_{Presidencia} = s] = P[Y_s \leq s] - P[Y_{s-1} \leq s-1] = \frac{e^{Y_s}}{1+e^{Y_s}} - \frac{e^{Y_{s-1}}}{1+e^{Y_{s-1}}}.$$

donde s es la percepción de con qué tanta frecuencia se realizan actos de corrupción en dicha institución, mientras que Y_s tiene la siguiente expresión:

$$\begin{aligned} Y_s = Umbral(s|s+1) &+ 0.18Agua + 0.14Drenaje + 0.16Alumbrado + 0.64Policias \\ &+ 0.24Calles + 0.33Carreteras - 0.27Edu\_Bas - 0.56Edu\_Med \\ &- 0.79Edu\_Uni - 0.09Mujer - 0.012Edad - 0.05Act\_eco \\ &- 0.12Poblacion - 0.02PIB + 0.05Alternancia \\ &- 0.31Inc\_delictiva. \end{aligned}$$

2017	Presidencia	Gobernatura	Presidencia municipal	Congreso
Servicios públicos				
Agua	0.18 (0.04)***	0.19 (0.04)***	0.23 (0.04)***	0.19 (0.04)***
Drenaje y alcantarillado	0.14 (0.04)***	0.14 (0.04)***	0.10 (0.04)**	0.10 (0.04)**
Alumbrado público	0.16 (0.03)***	0.17 (0.03)***	0.20 (0.03)***	0.14 (0.03)***
Parques y jardines	0.00 (0.04)	0.06 (0.04)	0.18 (0.04)***	0.08 (0.04)*
Policías	0.64 (0.03)***	0.67 (0.03)***	0.71 (0.03)***	0.62 (0.03)***
Calles y avenidas	0.24 (0.04)***	0.26 (0.04)***	0.31 (0.04)***	0.21 (0.04)***
Señalamiento urbano	0.03 (0.03)	0.05 (0.03)*	0.09 (0.03)***	0.05 (0.03)*
Carreteras públicas	0.33 (0.04)***	0.45 (0.03)***	0.36 (0.03)***	0.35 (0.04)***
Controles individuales				
Educación: primaria o menos				
Básica	−0.27 (0.03)***	−0.17 (0.03)***	−0.15 (0.03)**	−0.31 (0.03)***
Media superior	−0.56 (0.03)***	−0.35 (0.03)***	−0.35 (0.03)***	−0.57 (0.04)***
Universidad o más	−0.79 (0.03)***	−0.50 (0.03)***	−0.53 (0.03)***	−0.83 (0.03)***
Sexo: mujer	−0.09 (0.02)***	−0.14 (0.02)***	−0.07 (0.02)***	−0.04 (0.02)*
Edad	−0.012 (0.003)***	0.00 (0.00)	−0.005 (0.003)*	−0.02 (0.00)***
Edad^2	0.00 (0.00)***	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)***
Ocupación: en actividad económica	−0.05 (0.02)*	−0.06 (0.02)*	−0.05 (0.02)*	−0.05 (0.02)*
Controles estatales				
Población 2015 (log)	−0.08 (0.25)	−0.10 (0.25)	−0.11 (0.25)	−0.10 (0.25)
PIB %	−0.02 (0.08)	−0.02 (0.08)	−0.01 (0.08)	−0.03 (0.08)
Alternancia	0.06 (0.37)	0.11 (0.37)	0.16 (0.37)	0.08 (0.37)
Homicidios 2018	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)
Incidencia delictiva (log)	−0.34 (0.44)	−0.33 (0.40)	−0.39 (0.45)	−0.29 (0.52)
Umbral (1|2)	−4.60 (3.08)	−4.63 (1.97)	−5.73 (2.96)	−4.58 (4.35)
Umbral (2|3)	−2.82 (3.08)	−2.81 (1.97)	−3.86 (2.96)***	−2.66 (4.35)
Umbral (3|4)	−0.88 (3.08)	−0.83 (1.97)	−1.91 (2.96)***	−0.67 (4.35)
Componentes de la varianza				
Varianza del intercepto aleatorio	1	1	1	1
N	37468	37709	37550	36908
Grupos	32	32	32	32
BIC	69176.4	76666.5	80835.1	68429.6

Tabla 6.10: Análisis de la calidad percibida en los servicios públicos locales y la percepción en las instituciones políticas (2017).

2017	**Jueces y magistrados**	**Partidos políticos**	**Ministerios públicos**
Servicios públicos			
Agua	0.17 (0.04)***	0.20 (0.04)***	0.25 (0.04)***
Drenaje y alcantarillado	0.11 (0.04)***	0.15 (0.04)***	0.13 (0.04)***
Alumbrado público	0.10 (0.03)***	0.18 (0.03)***	0.15 (0.03)***
Parques y jardines	0.14 (0.04)	−0.06 (0.04)	0.08 (0.04)*
Policías	0.65 (0.03)***	0.62 (0.03)***	0.74 (0.03)***
Calles y avenidas	0.19 (0.04)***	0.29 (0.04)***	0.19 (0.04)***
Señalamiento urbano	0.13 (0.03)***	0.00 (0.03)	0.09 (0.03)***
Carreteras públicas	0.40 (0.03)***	0.39 (0.04)***	0.42 (0.03)***
Controles individuales			
Educación: primaria o menos			
Básica	−0.15 (0.03)***	−0.25 (0.03)***	−0.25 (0.03)***
Media superior	−0.31 (0.03)***	−0.52 (0.04)***	−0.49 (0.03)***
Universidad o más	−0.45 (0.03)***	−0.82 (0.03)***	−0.71 (0.03)***
Sexo: mujer	−0.08 (0.02)***	0.03 (0.03)	−0.01 (0.02)
Edad	−0.012 (0.003)***	−0.015 (0.003)***	−0.031 (0.003)***
Edad^2	0.00 (0.00)*	0.00 (0.00)***	0.00 (0.00)***
Ocupación: en actividad económica	−0.05 (0.03)*	−0.06 (0.02)*	−0.07 (0.02)**
Controles estatales			
Población 2015 (log)	−0.12 (0.25)	−0.08 (0.25)	−0.16 (0.25)
PIB %	−0.02 (0.08)	−0.01 (0.08)	−0.01 (0.08)
Alternancia	0.05 (0.37)	0.05 (0.37)	0.17 (0.37)
Homicidios 2018	0.00 (0.01)	0.00 (0.01)	−0.64 (0.41)
Incidencia delictiva (log)	−0.31 (0.42)	−0.31 (0.56)	−0.31 (0.45)*
Umbral (1|2)	−5.60 (2.60)	−4.06 (4.92)	−9.58 (2.08)
Umbral (2|3)	−3.82 (2.60)	−2.12 (4.92)	−7.73 (2.08)
Umbral (3|4)	−1.84 (2.60)	−0.24 (4.92)	−5.83 (2.08)
Componentes de la varianza			
Varianza del intercepto aleatorio	1	1	1
N	37468	37709	37550
Grupos	32	32	32
BIC	82579.4	66078.3	79184.9

Tabla 6.11: Análisis de la calidad percibida en los servicios públicos locales y la percepción en las instituciones políticas (2017).

2019	**Presidencia**	**Gobernatura**	**Presidencia municipal**	**Congreso**
Servicios públicos				
Agua	0.04 (0.04)	0.08 (0.04)*	0.21 (0.04)***	0.15 (0.04)***
Drenaje y alcantarillado	0.06 (0.03)*	0.12 (0.03)***	0.15 (0.03)***	0.17 (0.04)***
Alumbrado público	0.10 (0.03)***	0.17 (0.03)***	0.20 (0.03)***	0.17 (0.03)***
Parques y jardines	0.01 (0.04)	0.08 (0.04)*	0.14 (0.04)***	0.03 (0.04)
Policías	0.43 (0.03)***	0.68 (0.03)***	0.67 (0.03)***	0.61 (0.03)***
Calles y avenidas	0.24 (0.03)***	0.35 (0.03)***	0.37 (0.03)***	0.29 (0.04)***
Señalamiento urbano	0.02 (0.03)	0.02 (0.03)	0.06 (0.03)*	0.04 (0.03)
Carreteras públicas	0.29 (0.03)***	0.37 (0.03)***	0.37 (0.03)***	0.31 (0.03)***
Controles individuales				
Educación: primaria o menos				
Básica	−0.15 (0.03)***	−0.24 (0.03)***	−0.22 (0.03)**	−0.25 (0.03)***
Media superior	−0.36 (0.03)***	−0.42 (0.03)***	−0.38 (0.03)***	−0.52 (0.03)***
Universidad o más	−0.49 (0.03)***	−0.55 (0.03)***	−0.51 (0.03)***	−0.70 (0.03)***
Sexo: mujer	−0.34 (0.02)***	−0.15 (0.02)***	−0.06 (0.02)**	−0.07 (0.02)**
Edad	0.021 (0.003)***	0.005 (0.003)*	−0.00 (0.00)	−0.009 (0.003)***
Edad^2	0.00 (0.00)***	0.00 (0.00)	0.00 (0.00)*	0.00 (0.00)**
Ocupación: en actividad económica	−0.13 (0.02)***	−0.13 (0.02)***	−0.09 (0.02)***	−0.15 (0.02)***
Controles estatales				
Población 2015 (log)	0.00 (0.14)	0.03 (0.25)	0.02 (0.25)	0.01 (0.25)
PIB %	−0.07 (0.07)	−0.01 (0.08)	−0.03 (0.08)	−0.05 (0.08)
Alternancia	0.06 (0.26)	−0.09 (0.36)	0.03 (0.36)	0.10 (0.37)
Homicidios 2018	0.00 (0.00)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)
Incidencia delictiva (log)	0.04 (0.20)	0.00 (0.40)	−0.24 (0.45)	−0.38 (0.39)
Umbral (1|2)	−0.22 (0.17)	0.02 (2.13)	−2.60 (2.92)	−4.34 (1.79)
Umbral (2|3)	1.39 (0.17)	1.97 (2.13)	−0.59 (2.92)	−2.25 (1.79)
Umbral (3|4)	3.09 (0.17)	4.05 (2.13)	1.39 (2.92)	−0.35 (1.79)
Componentes de la varianza				
Varianza del intercepto aleatorio	1	1	1	1
N	36128	36854	36642	35508
Grupos	32	32	32	32
BIC	89518.3	82288.6	83929.9	76489.7

Tabla 6.12: Análisis de la calidad percibida en los servicios públicos locales y la percepción en las instituciones políticas (2019).

2019	Jueces y magistrados	Partidos políticos	Ministerios públicos
Servicios públicos			
Agua	0.13 (0.04)***	0.18 (0.04)***	0.21 (0.04)***
Drenaje y alcantarillado	0.12 (0.04)***	0.14 (0.04)***	0.15 (0.04)***
Alumbrado público	0.17 (0.03)***	0.18 (0.03)***	0.17 (0.03)***
Parques y jardines	0.02 (0.04)	−0.04 (0.04)	0.03 (0.04)
Policías	0.61 (0.03)***	0.57 (0.03)***	0.72 (0.03)***
Calles y avenidas	0.03 (0.06)	0.35 (0.03)***	0.33 (0.04)***
Señalamiento urbano	0.13 (0.03)***	0.07 (0.03)**	0.00 (0.03)
Carreteras públicas	0.37 (0.03)***	0.29 (0.03)***	0.37 (0.03)***
Controles individuales			
Educación: primaria o menos			
Básica	−0.20 (0.03)***	−0.27 (0.03)***	−0.28 (0.03)***
Media superior	−0.40 (0.03)***	−0.52 (0.03)***	−0.46 (0.03)***
Universidad o más	−0.54 (0.03)***	−0.78 (0.03)***	−0.69 (0.03)***
Sexo: mujer	0.01 (0.02)***	0.01 (0.02)	0.00 (0.02)
Edad	−0.021 (0.003)***	−0.009 (0.003)**	−0.026 (0.003)***
Edad^2	0.00 (0.00)***	0.00 (0.00)**	0.00 (0.00)***
Ocupación: en actividad económica	−0.07 (0.02)**	−0.10 (0.02)***	−0.14 (0.02)***
Controles estatales			
Población 2015 (log)	−0.06 (0.25)	0.07 (0.25)	−0.09 (0.25)
PIB %	−0.03 (0.08)	−0.04 (0.08)	−0.01 (0.08)
Alternancia	0.09 (0.37)	0.05 (0.37)	0.15 (0.37)
Homicidios 2018	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)
Incidencia delictiva (log)	−0.44 (0.43)	−0.34 (0.41)	−0.67 (0.44)
Umbral (1|2)	−6.47 (2.56)	−2.48 (2.21)	−8.95 (2.86)
Umbral (2|3)	−4.49 (2.56)	−0.44 (2.21)	−7.05 (2.86)
Umbral (3|4)	−2.52 (2.56)	1.43 (2.21)	−5.09 (2.86)
Componentes de la varianza			
Varianza del intercepto aleatorio	1	1	1
N	34943	37815	36662
Grupos	32	32	32
BIC	80217.3	74654.5	79938.5

Tabla 6.13: Análisis de la calidad percibida en los servicios públicos locales y la percepción en las instituciones políticas (2019).

En el cual los umbrales tienen los siguientes valores:

$$Umbral(1|2) = -4.60$$

$$Umbral(2|3) = -2.82$$

$$Umbral(3|4) = -0.88$$

Este modelo se realiza de manera análoga con las otras instituciones en las tablas 6.10, 6.11, 6.12 y 6.13.

De lo cual podemos concluir varias cosas, primero, los factores estatales resultan ser no tan relevantes en la percepción de la corrupción, esta información podría estar contenida en el efecto aleatorio de los estados. Por otro lado, en los factores individuales, el nivel de instrucción es relevante y entre más grados de estudio tiene la persona, valora de manera más negativa a las instituciones. Finalmente, los servicios públicos son los más relevantes para valorar de manera positiva, excepción de la calidad de los parques y los señalamientos públicos que en varios casos parecen ser un factor no tan relevante. A comparación de los modelos implementados en el modelo que evaluaba la confianza en instituciones, en este segundo caso la percepción de la corrupción parece que solo se centra en las condiciones iniciales de cada individuo, sus características demográficas y su acceso a los servicios públicos. Mientras que los factores estatales resultan irrelevantes.

Para apreciar esto de manera general, en las tablas 6.14 y 6.15 se analiza la información de la percepción de la corrupción con el IPCI, en este caso se crea una variable continua, por lo cual el modelo ya no es multinomial logístico como en las tablas anteriormente mencionadas, sino que es uno lineal multinivel.

También observe que en este caso no todos los servicios públicos tienen una importancia en la evaluación final de la percepción de la corrupción. Como se muestra en las tablas anteriores, el servicio de parques y el de los señalamientos públicos tienen un impacto menor o no significativo en la evaluación final. Aunque sí hay un aspecto significativo que es interesante, y ese es el sexo.

Como se señala, las mujeres tienden a percibir una corrupción mayor que los hombres en sus instituciones gubernamentales, aunque es una pequeña diferencia, ésta es significativa. Además, en estos casos también comienzan a aparecer otros factores estatales, sin embargo, no se mantienen en ambos años de comparación a excepción de la tasa delictiva, la cual también señala tener un impacto negativo en su percepción final sobre la corrupción.

Como se puede observar, los factores que determinan la percepción de la corrupción en las instituciones gubernamentales son muy parecidos a aquellos que determinan su confianza, los cuales dependen más de la calidad de los servicios y su nivel de instrucción. Esto es lógico, puesto que el impacto más visible del trabajo de los servidores públicos se ve en los servicios de los cuales la mayoría de ellos están a cargo, y segundo porque se es más crítico y se logra discernir de mejor manera el trabajo del que está a cargo cada institución gubernamental.

2017	**IPCI**	**Primer modelo**	**Segundo modelo**	**Tercer modelo**
Servicios públicos				
Agua	0.09 (0.01)***			
Drenaje y alcantarillado	0.04 (0.01)***			
Alumbrado público	0.07 (0.01)***			
Parques y jardines	0.03 (0.01)***			
Policías	0.26 (0.01)***			
Calles y avenidas	0.11 (0.01)***			
Señalamiento urbano	0.01 (0.01)*			
Carreteras públicas	0.14 (0.01)***			
IPCSL		0.81 (0.01)***	0.79 (0.01)***	0.78 (0.03)***
IPCSL (promedio estatal)				0.21 (0.21)
Controles individuales				
Educación: primaria o menos				
Básica	−0.08 (0.01)***		−0.09 (0.01)***	−0.09 (0.01)***
Media superior	−0.18 (0.01)***		−0.18 (0.01)***	−0.18 (0.01)***
Universidad o más	−0.26 (0.01)***		−0.27 (0.01)***	−0.26 (0.01)***
Sexo: mujer	−0.02 (0.01)***		−0.03 (0.01)***	−0.03 (0.01)***
Edad	−0.004 (0.001)***		−0.004 (0.001)***	−0.004 (0.001)***
Edad^2	0.00 (0.00)***		0.00 (0.00)***	0.00 (0.00)***
Ocupación: en actividad económica	−0.02 (0.01)**		−0.02 (0.01)***	−0.02 (0.01)***
Controles estatales				
Población 2015 (log)	−0.04 (0.02)*		−0.04 (0.02)*	−0.04 (0.02)*
PIB %	−0.01 (0.01)		−0.01 (0.01)	−0.01 (0.01)
Alternancia	0.02 (0.03)		0.02 (0.03)	0.01 (0.03)
Homicidios 2018	0.00 (0.01)		0.00 (0.00)	0.00 0.00
Incidencia delictiva (log)	−0.12 (0.05)*		−0.12 (0.05)*	−0.11 (0.05)*
Constante	3.52 (0.51)***	1.36 (0.02)***	3.51 (0.50)***	2.60 (0.60)***
Componentes de la varianza				
Pendiente aleatoria				0.02 (0.03)
Constante aleatoria	0.005	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Covarianza (pendiente, residuo)				-0.01
Residual	0.286	0.30 (0.002)	0.29 (0.002)	0.33 (0.13)
Correlación intraclase	0.017	0.033	0.032	0.029
N	38850	38850	38850	38850
Grupos	32	32	32	32
BIC	62107.3	63701.1	62586.6	62557.7

Tabla 6.14: Análisis de la calidad percibida en los servicios públicos locales y el IPCI.

2019	**IPCI**	**Primer modelo**	**Segundo modelo**	**Tercer modelo**
Servicios públicos				
Agua	0.08 (0.01)***			
Drenaje y alcantarillado	0.05 (0.01)***			
Alumbrado público	0.07 (0.01)***			
Parques y jardines	0.02 (0.01)*			
Policías	0.26 (0.01)***			
Calles y avenidas	0.13 (0.01)***			
Señalamiento urbano	0.00 (0.01)			
Carreteras públicas	0.14 (0.01)***			
IPCSL		0.83 (0.01)***	0.79 (0.01)***	0.78 (0.03)***
IPCSL (promedio estatal)				0.21 (0.21)
Controles individuales				
Educación: primaria o menos				
Básica	−0.10 (0.01)***		−0.10 (0.01)***	−0.09 (0.01)***
Media superior	−0.20 (0.01)***		−0.21 (0.01)***	−0.18 (0.01)***
Universidad o más	−0.26 (0.01)***		−0.28 (0.01)***	−0.26 (0.01)***
Sexo: mujer	−0.03 (0.01)***		−0.03 (0.01)***	−0.03 (0.01)***
Edad	−0.001 (0.001)		0.00 (0.00)	−0.004 (0.001)***
Edad^2	0.00 (0.00)**		0.00 (0.00)**	0.00 (0.00)***
Ocupación: en actividad económica	−0.04 (0.01)***		−0.05(0.01)***	−0.02 (0.01)***
Controles estatales				
Población 2015 (log)	0.00 (0.02)		0.00 (0.01)	−0.04 (0.02)*
PIB %	−0.01 (0.01)*		0.01 (0.01)*	−0.01 (0.01)
Alternancia	0.02 (0.03)		0.02 (0.03)	0.01 (0.03)
Homicidios 2018	0.01 (0.01)		0.00 (0.00)	0.00 (0.00)
Incidencia delictiva (log)	−0.01 (0.00)*		−0.09 (0.05)*	−0.11 (0.05)*
Constante	2.85 (0.47)***	1.54 (0.02)***	2.33 (0.55)***	2.58 (0.58)***
Componentes de la varianza				
Pendiente aleatoria				0.02 (0.03)
Constante aleatoria	0.005 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Covarianza (pendiente, residuo)				-0.01
Residual	0.29 (0.00)	0.32 (0.00)	0.30 (0.00)	0.29 (0.13)
Correlación intraclase	0.016	0.030	0.033	0.032
N	39118	39118	39118	39118
Grupos	32	32	32	32
BIC	64262.5	66223.9	64693.2	62557.7

Tabla 6.15: Análisis de la calidad percibida en los servicios públicos locales y el IPCI.

Por último, el efecto global de la calidad de los servicios públicos se observa en el

IPCSL y su relación con el IPCI en las últimas dos tablas 6.14 y 6.15 se muestra cómo cuando se agrega el índice y más variables, el modelo va mejorando. Con lo cual se puede concluir que las variables que ya se habían mencionado (servicios e instrucción) son importantes, así como las variables de edad registran un efecto de ser más crítico conforme la edad va incrementando, tal como si la persona registra una actividad económica. Finalmente, en el ámbito estatal, las variables que tienen relación con el tamaño de la población y la incidencia delictiva son significativas, ambas registran un impacto negativo, mientras esa actividad vaya creciendo.

La edad podría tener esa relación ya que se podría relacionar con el nivel de instrucción, pues además la $Edad^2$ muestra que es significativa, aunque con un valor muy pequeño, los adultos más jóvenes y adultos de edad avanzada muestran ser menos críticos que aquellos que no se encuentran en estos puntos extremos. Por otra parte, tener una actividad económica estaría relacionada con la exposición a actos de corrupción o que le sea más visible la deficiencia de algunos servicios. Esto se puede reflejar un poco en el impacto negativo que muestra la incidencia delictiva, pues el tener una actividad económica podría hacerlos más vulnerables al estar expuestos a estos actos en su trayecto al trabajo. Finalmente, el sexo es un factor que realmente no se puede intuir una razón para que el factor de ser mujer tienda a percibir más actos de corrupción que en los hombres. Lo cual, sería interesante seguir de cerca este factor y entender porqué este factor parece marcar una pequeña diferencia.

Visualizando las estadísticas de cada uno de los modelos, vea que el mejor modelo es aquel que usa el índice de la percepción y que a su vez tiene a IPCSL estatal como un intercepto aleatorio, aunque éste no sea significativo.

Se hicieron pruebas de homogeneidad sobre los modelos con variable dependiente IPCI y los último tres modelos en ambos años de estudio, para comprobar si los modelos cumplían con tener diferencias entre los Estados y ser de varianza constante. Así mismo, probar la normalidad de los residuos. Con la prueba de Brown-Forsyth se pudo determinar que la diferencia de grupos es significativa a un nivel del 0.001, mientras que con la prueba de Kolmogorov-Smirnov a un nivel de significancia 0.001, se concluye que los residuos de los modelos anteriormente mencionados cumplen con dicho supuesto y por ende se válida el uso del modelo multinivel para estos casos.

7

Conclusiones

A lo largo de este trabajo el tema medular es el modelo multinivel, se mostraron dos variantes de dicho modelo, uno en el cual se toma en cuenta como una extensión del modelo lineal, mientras que en el segundo caso se implementa un modelo de respuestas categóricas añadiendo los efectos aleatorios, generando un segundo nivel.

La estructura de los datos permite apreciar la relación de varios factores que conforman el día a día de los ciudadanos con respecto a su confianza y percepción de la corrupción en las instituciones gubernamentales, el modelo multinivel considera un efecto en los estados. Esto ha llevado a apreciar que gran parte de estas evaluaciones sean explicadas por los factores locales, hablando de los servicios públicos o de la formación individual, tal vez en parte porque los factores estatales se explican por el efecto aleatorio.

Igualmente es una herramienta que en este caso midió la variabilidad de encontrarse en los estados. De aquí que los resultados puedan mostrar que particularmente los estados del norte se encuentran más satisfechos con sus gobernantes, en los cuales podrían estar involucrados factores culturales o históricos, los cuales son más complicados de medir. Sobre todo en el caso de la confianza institucional, ya que en la percepción de la corrupción, este agrupamiento no es tan claro.

Además, los resultados arrojaron un factor importante el cual podría extender su investigación a saber por qué las mujeres perciben mayores niveles de corrupción, puesto que esto pasa en ambas evaluaciones, algo que no pasa en la confianza institucional. Por una parte, el estudio continuo de estas evaluaciones con el fin de observar si esto es un factor que ya está determinando a las mujeres más críticas en este aspecto (y no menos importante, saber el por qué), o si es un factor que podría decir que las mujeres están más expuestas a actos de corrupción o a sufrir las consecuencias de los mismos.

Recuerde que este proyecto tuvo la finalidad de analizar los datos relacionados a la percepción. Así, aunque en la percepción de la corrupción se haya registrado una menor frecuencia en el año 2019 con respecto al 2017, esto se relaciona a la confianza, es necesario recordar que en este lapso paso de tener uno de los gobiernos con menor aprobación en la historia de México a uno con una muy buena evaluación por parte de la población. En el INEGI hay datos que nos puede acercar un poco más a la situación real de la corrupción en México, por lo cual sería de interés ver lo que nos ofrecen estos datos.

Aquí se mostró un acercamiento a dos variables que muestran una estrecha relación, sobre todo al presentar las variables explicativas de las mismas. Lo cual podría ser coherente, ya que al final la percepción de la corrupción es una forma de percibir qué tanto nuestros gobernantes mienten al hacer uso de los recursos. Un ciudadano no pondría toda su confianza a alguien el cual miente de manera constante y más aún, si esto perjudica sus condiciones de vida.

Por lo cual, considero que un análisis de este tipo sería de interés para observar el efecto a nivel mundial, es decir, modelo con niveles relacionado a las regiones o subcontinentes además de los países en si, al menos en el caso de la corrupción.

Con el tiempo, solo los medios de transparencia y rendición de cuentas ayudarán a saber qué pasa con la corrupción, al menos en el aspecto financiero. Otros mecanismos para medir serán mejor aceptados o sencillamente darán una mejor visión del tema.

Apéndice A

Algunas pruebas estadísticas

A.1. Prueba t

La prueba t, es una prueba estadística que contrasta dos hipótesis. Sea H_0 la hipótesis nula que indica que dos valores son iguales. Mientras que H_1 es la hipótesis alternativa la cual es cierta cuando los valores son distintos. Se supone que se tiene una muestra de tamaño N. Para realizar esta prueba se calcula el estadístico t de la siguiente manera:

$$t = \frac{b_1 - b_0}{\sigma_b},$$

donde b_1 es el valor estimado de la media de la muestra, mientras que b_0 es el valor teórico de la media, finalmente, σ_b es la desviación estándar de la muestra.

Para determinar que hipótesis es verdadera, es decir, qué hipótesis se acepta para la muestra. Se hace un contraste entre el estadístico t y el valor de la distribución t con $N-1$ grados de libertad en el cuantil de orden α, el cual se denota, $t_{(N-1,\alpha)}$.

Por lo cual, la hipótesis nula se va a aceptar cuando se cumple la siguiente condición:

$$\mid t \mid \leq t_{(N-1,\alpha)},$$

en caso contrario, rechazamos la hipótesis nula y aceptamos la hipótesis alternativa, lo que se traduce a que el valor teórico y la media son distintos.

Esta prueba aplicada a los coeficientes de regresión prueba que ellos sean distintos o iguales a un valor teórico, dicho valor teórico es usualmente *cero*. Por lo cual la prueba se reduce a hacer la misma comparación entre el estadístico y valor t en la distribución, pero con el el siguiente valor de t:

$$t = \frac{\widehat{\beta}}{\sigma_{\widehat{\beta}}}.$$

Por otra parte, cuando se quiere generar un intervalo de confianza, se quiere observar en que valores puede variar este valor a partir de su valor estimado. En este caso, diremos b a partir de b_0. Entonces:

$$| t |=| \frac{b-b_0}{\sigma_b} |\leq t_{(N-1,\alpha)}.$$

Entonces:

$$-\sigma_b t_{(N-1,1-\alpha)} \leq b - b_0 \leq \sigma_b t_{(N-1,1-\alpha)}$$

A.2. Prueba U de Mann-Withney

La prueba U de Mann-Withney, Mann-Whitney-Wilcoxon, Wilcoxon rank-sum test o Wilcoxon-Mann-Whitney es una prueba que comprueba si existe o no diferencias entre dos muestras. Suponga que se tiene las muestras X e Y. La hipótesis que se pone a prueba es comparar las medias de X e Y, μ_X y μ_y respectivamente. En este caso, la igualdad, $\mu_X = \mu_Y$, este caso se conoce como prueba de dos colas. El caso en el que solo se tiene una desigualdad, es conocido como de una cola.

Las muestras X e Y son de tamaño n_X y n_Y respectivamente. El estadístico S es la suma de los rangos de $x_1, x_2, \ldots, x_{n_X}$, en la muestra combinada de X e Y, es decir, $\{x_1, x_2, \ldots, x_{n_X}, y_1, y_2, \ldots, y_{n_Y}\}$. La j-ésima observación x_j tiene rango $R_j = k$ si es la k-ésima observación más pequeña en la muestra combinada. Eso quiere decir que $S = \sum_{j=1}^{n_X} R_j$.

Entonces, la hipótesis $H_0 : \mu_X = \mu_Y$ es rechazada a un nivel $(1-\alpha)/2 \times 100\,\%$ si $S < \kappa_{(1-\alpha)/2}$ o $S > S_{\text{máx}} - \kappa_{(1+\alpha)/2}$.

El valor máximo de S, $S_{\text{máx}} = n_X(n_X + 2n_y + 1)/2$, mientras que el mínimo es $n_x(n_x + 1)/2$.

Luego, para muestras de gran tamaño, el valor de κ toma la siguiente aproximación.

$$\kappa_p \approx n_X(n_X + n_Y + 1)/2 - Z_p\sqrt{n_X n_Y(n_X + n_Y + 1)/12}$$

.

Donde p es el valor del cuantil de la distribución normal Z.

Apéndice B

Código de R

En esta sección se añade parte del código que se ejecutó en el software R que se utilizó para adquirir los modelos.

En este caso se hicieron dos tipos de modelos, el lineal multinivel y multinivel multinomial. Entonces, para ejemplificvar ambas situaciones se mostrarán las funciones que se usaron con los datos relacionados a las Pressidencia.

B.1. Modelo para la confianza institucional

Para iniciar, se muestra un resumen de como se ve la variable dependiente de la confianza en la Presidencia.

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   1.000   3.000   3.802   6.000  10.000
```

Para este modelo se utilizó la librería lme4, entonces para el ajuste del modelo nulo, se importa la librería y el modelo se ajusta de la siguiente manera:

```
library(lme4)
regre_Pre<-lme(Presidencia ~ 1, random = ~1|Estado,
               data = Datos_Pre)
```

Como se observa, dicho modelo consiste del intercepto fijo, y el efecto aleatorio. Esto pasa con cada uno de las instituciones, al igual qué con el índice.

Para la formación del índice ICI es algo parecido al IPCI, pues solo reúne la información de las instituciones políticas. Las evaluaciones de cada institución están en las columnas, por lo cual se tiene el siguiente código

```
ICI = rowMeans(Instituciones)
regre_Pre<-lme(ICI ~ 1, random = ~1|Estado,
               data = Global)
```

Para la creación de los modelos con las variables explicativas se ejemplifica de nuevo con el caso de la Presidencia.

```
regre_Pre<-lme(Presidencia ~ Agua + Dre_alc + Alumbrado +
                 Parques + Policia + Calles + Senalamiento +
                 Carretera + Edu_bas + Edu_med + Edu_sup +
                 SEXO + EDAD + EDAD2 + Act_econo + Poblacion +
                 Crecimiento.del.PIB + Alternancia +
                 Tasa.de.homicidios + Delito,
               random = ~1|Estado,
               data = Datos_Pre)
```

Finalmente, los modelos que consideran los índices tanto de IPCSL como ICI se generan de la siguiente manera.

```
M1<-lme(Confianza ~ IPCSL, random = ~1|Estado, data = Global)
M2<-lme(Confianza ~ IPCSL + Edu_bas + Edu_med + Edu_sup +
          SEXO + EDAD + EDAD2 + Act_econo + Poblacion +
          Crecimiento.del.PIB + Alternancia +
          Tasa.de.homicidios + Delito,
        random = ~1|Estado,
        data = Global)
M3<-lme(Confianza ~ IPCSL + Medias + Edu_bas + Edu_med +
          Edu_sup + SEXO + EDAD + EDAD2 + Act_econo +
          Poblacion + Crecimiento.del.PIB + Alternancia +
          Tasa.de.homicidios + Delito,
        random = ~IPCSL|Estado,
        data = Global1)
```

M1, M2 y M3 son el primer, segundo y tercer modelo respectivamente, en este caso ICI se nombró como Confianza en los datos, por eso en la función *Confianza* está como variable dependiente. Además IPCSL se obtuvo de usando la media como en ICI, solo que de las variables que comprenden este índice.

B.2. Modelo para la percepción de la corrupción

Este modelo se genera con el índice ICI de manera sumamente parecida a como se hizo con el caso anterior. En este caso, los modelos solo son modelos multinivel multinomial. Para la implementación de los modelos multinivel multinomial se utilizó la paquetería ordinal la cual también genera modelos lineales multinivel. Sin embargo, dicha librería también genera mmodelos lineales generalizados multinivel, y para nuestro caso particular, tiene la opción de modelos multinivel multinomiales.

Tomando de nuevo el ejemplo de los datos acerca de la percepción de la corrupción de la Presidencia se muestra un nuevo resumen acerca de como está esta variable dependiente:

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   1.000   1.000   1.533   2.000   4.000
```

En el siguiente código se muestra como se generó el modelo nulo así como la importación de la librería.

```
library(ordinal)
regre_Pre<-clmm(as.factor(Presidencia) ~ 1 + (1|Estado),
                data = Datos_Pre)
```

Para los modelos completos correspondientes a cada una de las instituciones gubernamentales. Se usa esta misma función pero añadiendo a las variables explicativas.

```
regre_Pre1<-clmm(as.factor(Presidencia) ~ Agua + Dre_alc +
                   Alumbrado + Parques + Policia + Calles +
                   Senalamiento + Carretera + Edu_bas +
                   Edu_med + Edu_sup + SEXO + EDAD +
                   EDAD2 + Act_econo + Poblacion +
                   Crecimiento.del.PIB + Alternancia +
                   Tasa.de.homicidios + Delito + (1|Estado),
                  data = Datos_Pre)
```

Aunque en las instituciones gubernamentales este modelo es multinivel multinomial, en la implementación del modelo con el IPCI se hace uno modelo lineal multinivel, ya que como se mostró en la creación de los índices, estos son continuos. Por lo cual, el modelo nulo y completo son de la siguiente manera.

```
Corrupcion<-rowMeans(Instituciones)
regre<-lme(Corrupcion ~ 1, random = ~1|Estado, data = Global)
regre1<-lme(Corrupcion ~ Agua + Dre_alc + Alumbrado +
              Parques + Policia + Calles + Senalamiento +
              Carretera + Edu_bas + Edu_med + Edu_sup +
              SEXO + EDAD + EDAD2 +Act_econo + Poblacion +
              Crecimiento.del.PIB + Alternancia +
              Tasa.de.homicidios + Delito,
            random = ~1|Estado,
```

```
                  data = Global)
```

Finalmente, cómo se hizo con los últimos tres modelos en el caso de la confianza institucional. Se repitieron esos modelos, recuerde que solo se usan los índices, por lo cual las funciones en el modelo son muy parecidas al ser modelos multinivel lineales.

```
M1<-lme(Corrupcion ~ IPCSL, random = ~1|Estado, data = Global)
M2<-lme(Corrupcion ~ IPCSL + Edu_bas + Edu_med + Edu_sup +
          SEXO + EDAD + EDAD2 + Act_econo + Poblacion +
          Crecimiento.del.PIB + Alternancia +
          Tasa.de.homicidios + Delito,
        random = ~1|Estado,
        data = Global)
M3<-lme(Corrupcion ~ IPCSL + Medias + Edu_bas + Edu_med +
          Edu_sup + SEXO + EDAD + EDAD2 + Act_econo +
          Poblacion + Crecimiento.del.PIB + Alternancia +
          Tasa.de.homicidios + Delito,
        random = ~IPCSL|Estado,
        data = Global1)
```

Este código es el mismo que se usa para ambos años, solo se cambiaron los datos iniciales para adquirir los modelos de los años correspondientes.

Bibliografía

[1] Gelman Andrew y Hill Jennifer (2006). *Data Analysis using regression and multilevel hierarchical models*. Cambridge University Press.

[2] Andrzej Galecki y Tomasz Burzykowsky (2013). *Linear Mixed-Effects Models Using R*. Springer.

[3] Goldstein Harvey (2011). *Multilevel Statistical Models*. Cuarta edición. Wiley.

[4] Agresti Alan (2019). *An introduction to categorial data analysis*. Tercera edición. Wiley.

[5] Monsiváis Carrillo Alejandro (2019). La calidad percibida de los servicios públicos locales y la confianza institucional en México. *Región y sociedad, 31*, e1206. DOI: https://doi.org/10.22198/rys2019/31/1206

[6] W. Härdle y L. Simar (2017). *Applied Multivariate Statistical Analysis*. Segunda edición. Springer.

[7] Lee J. Cronbach (1951). Coefficient Alpha and the internal structure of tests. *Psychometrika, Vol. 16*, No.3, págs: 297-334

[8] Montero Rojas Eiliana (2008). Escala o índices para la medición de constructos: el dilema del analista de datos. *Avances en medición*, No.6, págs: 15-24

[9] Del Tronco José (2012). Las causas de la desconfianza política en México. *Perfiles Latinoamericanos*, No.40, págs: 227-251

[10] Córdova Guzmán y F. Ponce (2017). Los tipos de corrupción y la satisfacción con los servicios públicos. Evidencia del caso mexicano. *Región y sociedad, Vol. 29*, No.70, págs: 231-262

[11] Moreno Jaimes Carlos (2012). El nexo entre la calidad gubernativa y elecciones: discusión conceptual y aplicación al gobierno local mexicano. *Perfiles latinoamericanos, Vol. 20*, No.39, págs: 59-90

[12] Honaker James, king gary y Blackwell Mathieu (2011). Amelia II: A program for missing data. *Journal of Statistical Software, Vol. 45*,

[13] Von Storch, H., y Zwiers, F. W. (2002) Critical Values for the Mann-Whitney Test. *Statistical Analysis in Climate Research*, 437-442.