

UNIVERSIDAD MAYOR DE SAN ANDRÉS
FACULTAD DE CIENCIAS PURAS Y NATURALES
CARRERA DE MATEMÁTICA
MAESTRIA EN MODELAJE MATEMÁTICO PRIMERA VERSIÓN



**MODELO DE ECUACIONES ESTRUCTURALES – SEM TIPO
LISREL, PARA MODELAR LA CAUSALIDAD DE LA
DEMOCRACIA AL DESARROLLO**

Tesis de Postgrado para Obtener el Título de Magíster Scientiarum

POR: ROLLY ROGER VASQUEZ MACEDO

TUTOR: PHD. RIMER ZURITA ORELLANA

La Paz - Bolivia

Noviembre, 2021

Dedicatoria

Filipenses 4:13 todo lo puedo en cristo que me fortalece, gracias a Dios por permitirme haber llegado hasta este momento tan importante en mi formación profesional. A mi querido hijo Jesus Daniel, y toda mi familia.

Agradecimientos

Este trabajo no habría sido posible sin el apoyo de mi Tutor, Doctor Rimer Zurita Orellana, bajo cuya supervisión realice esta tesis, así también agradezco a mis profesores de la maestría en Modelaje Matemático que su enseñanza fue crucial en la culminación de esta tesis.

Índice general

Dedicatoria	2
Agradecimientos	3
1. Aspectos Generales	11
1.1. Antecedentes	11
1.2. Justificación	13
1.3. Planteamiento del problema	13
1.3.1. Preguntas específicas	13
1.4. Objetivo General	14
1.4.1. Objetivos específicos	14
1.5. Alcances de la investigación	14
2. Marco Teórico	16
2.1. Índice de Democracia - IDEM	16
2.2. Índice de Desarrollo Humano - IDH	17
2.3. Desarrollo humano y Democracia	18
2.4. Curvas de Andrews	19
2.5. Ecuaciones Estructurales	20
2.5.1. Introducción	20
2.5.2. Causalidad	20
2.5.3. Variables del modelo de ecuaciones estructurales	21
2.5.4. Diagrama de trayectorias	23
2.5.5. Especificación del modelo	24
2.5.6. Supuestos del modelo LISREL	25

2.5.7. Parámetros de las ecuaciones	28
2.5.8. Estimación	29
3. Marco Metodológico	31
3.1. Diseño de investigación	31
3.2. Definición de las variables	31
3.2.1. Diagrama de trayectorias del modelo	33
3.3. Estructura del modelo	34
3.3.1. Identificación	35
3.3.2. Estimación del modelo	37
4. Resultados	38
4.1. Programa para la modelación	38
4.2. Normalidad multivariante	39
4.3. Análisis de los datos	40
4.3.1. Datos atípicos	41
4.3.2. Gráfico de dispersión	41
4.4. Resultados del modelo	42
4.4.1. Diagrama de trayectorias del modelo	43
4.4.2. Ajuste del modelo	44
4.4.3. Ecuaciones de medida	45
4.4.4. Modelo Estructural	45
5. Conclusiones y recomendaciones	47
5.1. Conclusiones	47
5.2. Recomendaciones	49
Anexos	50
Bibliografía	55

Índice de figuras

2.1. Curvas de Andrews	19
2.2. Diagrama de trayectorias	24
3.1. Diagrama de trayectorias del modelo	33
4.1. Gráfico de Andrews	41
4.2. Dispersograma	42
4.3. Diagrama de trayectorias del modelo	43
A.1. Modelo estandarizado	50
A.2. Modelo estandarizado y eliminando ingreso	51
A.3. Modelo estandarizado aplicado ln a los componentes del IDH	52

Índice de cuadros

4.1. Resultado de las ecuaciones de medida	44
4.2. Resultado de las ecuaciones de medida	45
4.3. Resultado del modelo estructural	46

Resumen

Los modelos de ecuaciones estructurales son parte de la estadística multivariante, tienen aplicaciones a diferentes áreas de las ciencias se pueden ver en aquellas donde el interés sea ver la causalidad y/o impacto de variables latentes. En este trabajo se propone la construcción de un modelo de ecuaciones estructurales para determinar la causalidad de la democracia hacia el desarrollo humano (variables no observables directamente llamadas también variables latentes), manejamos como desarrollo humano al Índice de Desarrollo Humano - IDH y como democracia al Índice de Democracia - IDEM que permitirán mostrar el impacto, causalidad del IDEM hacia el IDH, como se indicó este modelo quiere ver el grado de la relación, así como el sentido de la causalidad, pero también permitirá verificar que las variables que hacen a los constructos o variables latentes IDH y el IDEM pertenecen al modelo estructural propuesto. Es así que se realiza el Modelo de Ecuaciones Estructurales - SEM, del tipo LISREL, el modelo se realizó, con información de los diferentes países del mundo, y el uso del paquete estadístico R versión 4.1. Los resultados son confiables del tipo descriptivo tanto para para las ecuaciones de medida como para el modelo estructural, se evidencia una alta relación de la democracia hacia el desarrollo humano. El aporte es metodológico aplicando modelación de ecuaciones estructurales para el estudio del fenómeno en cuestión, así como ampliar material de investigación sobre este tema, para que otros estudios puedan utilizarlo, replicarlo, confirmarlo o refutarlo.

Clasificación JEL: C360, 0150

Palabras clave: Ecuaciones estructurales, LISREL, R, índice desarrollo humano, índice de democracia.

Introducción

Si planteamos las siguientes interrogantes ¿como medir el bienestar de las poblaciones del mundo en un contexto múltiple?, ¿cómo la democracia en los países del mundo pueden sustentar al bienestar de las poblaciones?.

Al respecto se puede expresar que durante las últimas tres décadas, en el comportamiento mundial una oleada democratizadora relativamente universal ha venido cubriendo los distintos sistemas políticos a lo largo y ancho de las diversas regiones geográficas y culturales a nivel mundial [1]. Este fenómeno ha sido investigado por diferentes autores para determinar la institucionalidad sobre el bienestar de las poblaciones.

Amartya Sen, precursor en la comprensión del desarrollo humano es quien tuvo la influencia principal sobre el cálculo del IDH que actualmente se mantiene como el principal indicador de desarrollo de las poblaciones del mundo, la misma elaborada por el Programa de las Naciones Unidas para el Desarrollo - PNUD, así mismo en sus cuadernos de economía, Amartya Sen, defiende la postura de la democracia como influyente hacia el desarrollo humano [1].

El IDH puede entenderse como el indicador para medir el desarrollo humano en términos de salud, educación e ingreso, estos indicadores de bienestar los desarrolla el PNUD; y el IDEM puede entenderse como el indicador para medir la democracia en términos de libertades civiles, procesos electorales y pluralismo, funcionamiento del gobierno, participación política y cultura política, estos indicadores son generados y publicados por The Economist. En adelante usaremos The Economist para referirnos a la Unidad de Inteligencia de The Economist - EIU, por sus siglas en inglés.

Tomando en cuenta distintos indicadores del IDH que son publicadas por el PNUD, y distintos indicadores del IDEM publicadas por The Economist, se busca

ver cuál es el impacto del IDEM hacia el IDH en 165 países de las diversas regiones geográficas del mundo. Este trabajo tiene por objetivo, determinar la relación y fuerza entre la valoración de la democracia sobre el bienestar, esto se desarrollará mediante un análisis de ecuaciones estructurales.

La organización de esta tesis comprende un capítulo I en el que se presenta los antecedentes y justificación de la investigación, el planteamiento del problema, los objetivos general y específico del estudio y los alcances de la investigación. En el capítulo II se exponen los fundamentos que involucra el marco conceptual, referencial, que corresponde al modelo de ecuaciones estructurales y la notación empleada, así como los métodos de estimación de parámetros y las medidas de ajuste del modelo. En el capítulo III se presenta la metodología de la investigación (método y tipos de investigación, sujetos de la investigación, la definición de las fuentes). Los resultados, el análisis de los datos investigados, y el resultado modelo se exponen en el capítulo IV. Y finalmente, la tesis se completa con el capítulo V que corresponde a conclusiones y recomendaciones donde además se plantean propuesta de mejoramiento.

Capítulo 1

Aspectos Generales

En este capítulo se delimitará el tema de investigación, así como la caracterización del problema y planteamiento de los objetivos. Se desarrollará el alcance y justificación del tema, y finalmente se definirá la metodología de la investigación a seguir.

1.1. Antecedentes

En esta sección se delimita el contexto externo del tema de la tesis, lo cual implica desarrollar el marco referencial en el cual se desenvuelve el tema de investigación, que corresponde en la revisión de trabajos que han tenido que ver con intentos de solución, trabajos teóricos o algunas ideas respecto al tema de investigación que se propone en este documento.

Se puede mencionar el trabajo Temkin y Del Tronco (2006) [1] donde los autores de manera preliminar indican, sobre un análisis a los países del mundo, que existen evidencias para afirmar que el nivel de desarrollo humano esta estrecha y significativamente relacionado con algunas variables político-institucionales, como las libertades civiles o los derechos políticos. En este sentido califican, que un desarrollo humano alto está positiva y significativamente correlacionado con un nivel elevado de libertades ciudadanas y con una institucionalidad política más sólida.

También refieren, que los indicadores como el de libertades civiles y políticas, y de buen gobierno están altamente relacionados entre sí, de igual forma para el caso de

los indicadores de bienestar. Así mismo toman en cuenta como bienestar objetivo al IDH y toman en cuenta como bienestar subjetivo a variables como aspectos afectivos emocionales, aspectos cognitivos valorativos [1].

En el trabajo de Temkin y Del Tronco (2006) [1] los autores indican que en cuanto no se alcanzan logros socioeconómicos, la valoración del régimen disminuye y su estabilidad corre peligro, se extrae el siguiente el texto que expresa, "Los enfoques *racionalistas* tanto en su versión *rational culturalist* como en la de *political economy*, sostienen que la estabilidad del régimen democrático está muy relacionada con su efectividad".

Los enfoques *racionalistas* tanto en su versión *rational culturalist* como en la de *political economy*, sostienen que la estabilidad del régimen democrático está muy relacionada con su efectividad.

En el Informe de Desarrollo Humano 2010 [15], sobre los Objetivos del Desarrollo del Milenio, se indica que para encontrar al desarrollo humano de manera más completa están los objetivos del milenio que son el énfasis en los derechos humanos, la democracia y la participación, esto permite avances en los fundamentos del desarrollo humano.

Así mismo en Altman et al. (2009) [2] se realizó la construcción de un indicador de democracia con la metodología de análisis factorial que es la base en el modelo de ecuaciones estructurales, dichos indicadores fueron desarrollados para estudiar la calidad democrática, trabajaron un análisis factorial con variables gasto social, rol del estado, gasto público, identidad nacional, nacionalismo, posición frente a U.S.A., autoridad tradicional, ubicación izquierda derecha, posición comercio exterior, crecimiento redistribución, pobreza redistribución, pobreza seguridad, carga tributaria, proteccionismo interacción, eje valórico, democracia, empresas públicas, eutanasia, posición económica, aborto.

Berggren [12] mediante datos de panel analiza los efectos de la democracia en niveles de desarrollo humano y crecimiento económico para el periodo 1980-2010; la democracia lo mide con los derechos políticos y las libertades civiles de fuente Freedom House, y el desarrollo humano con el IDH. Las mismas regresiones son también realizadas con el PIB per cápita como la variable dependiente para 146 países, los resultados encontrados respaldan su hipótesis que la democracia es más importante para el desarrollo humano con un impacto positivo, que para el desarrollo económico este podría tener impactos negativos.

En Gerring, Tracker y Alfaro [7] una de sus hipótesis es que la democracia podría mejorar el bienestar social mediante series de datos de panel de 1960-2000 con muestras de alrededor de 159 países modelaron con efectos fijos, efectos dinámicos, probaron di-

cha hipótesis, empero toman en cuenta la tasa de mortalidad infantil como indicador de desarrollo humano y la democracia es extraída de Polity IV, la variable democracia que toman mide el grado de patrones de autoridad democráticos que están institucionalizados en un país determinado; también tomaron cuatro variables de control PIB per cápita, analfabetismo femenino, inestabilidad y urbanización.

1.2. Justificación

Esta investigación es pertinente dentro del contexto de las problemáticas que los países del mundo viven actualmente en un entorno de democracia y desarrollo, y conocer cual es la causalidad y su impacto, puede contribuir a mejores políticas del desarrollo como realizar políticas macroeconómicas socialmente responsables con perspectiva más amplia, más allá de la estabilización y las metas de inflación y crecimiento, gasto social adecuado, crecimiento del empleo y el objetivo inter-temporal del desarrollo humano, así como generar empleo más productivo y reducir la desigualdad [11].

Determinar la referida relación causal que busca la presente investigación tiene impacto de suma importancia para la visualización y puesta en práctica de políticas para mejorar indicadores de desarrollo y democracia de los diferentes países del mundo.

Acotando al problema, la perspectiva del desarrollo del trabajo lleva un enfoque aplicado de la matemática, con el modelo de Ecuaciones Estructurales ella parte de la minería de datos, reflejará un aporte a la investigación presentando este enfoque con el uso del software R.

1.3. Planteamiento del problema

¿La democracia en los países causa a su nivel del desarrollo humano?.

1.3.1. Preguntas específicas

Las preguntas específicas de investigación que se pretenden responder son las siguientes:

- ¿Es posible que la cultura y participación política, libertades civiles, procesos electorales y pluralismo, y el funcionamiento del gobierno sean determinantes del IDEM?.
- ¿Es posible que la educación, la vida saludable y prolongada, y niveles de vida digno sean determinantes del IDH?.

1.4. Objetivo General

El objetivo general es, determinar la causalidad de las variables latentes democracia al desarrollo humano de los países del mundo a través de un modelo de ecuaciones estructurales tipo LISREL. Además de estudiar y explicar los resultados de las variables con relación al IDEM e IDH, y la causalidad de las variables latentes.

1.4.1. Objetivos específicos

Los objetivos específicos son:

- Determinar un indicador sintético que hace a la democracia, a partir de un conjunto de variables que pertenecen a los diferentes países del mundo mediante un análisis factorial.
- Determinar un indicador sintético que hace al desarrollo humano, a partir de un conjunto de variables que pertenecen a los diferentes países del mundo mediante un análisis factorial.
- Realizar un modelo de ecuaciones estructurales para ver la causalidad de la democracia al desarrollo humano.

1.5. Alcances de la investigación

El presente trabajo tiene un alcance específico, que abarca a 165 países del mundo, este espacio geográfico se debe a que se cuenta con información para estos países en dimensiones de democracia y de desarrollo humano en fuentes de organismos internacionales PNUD [18] y The Economist [3].

Como alcance económico, realizar un estudio objetivo, para los países del mundo, el cual posibilite una solución al problema de identificar la relación causal de la democracia al desarrollo humano.

Sobre el alcance del tema, pertenece al área de la Matemática y Economía, por lo tanto el alcance temático de este trabajo implica llevar a cabo todo un proceso de modelaje matemático con el fin de llegar al objetivo general, a través del cumplimiento de los objetivos específicos.

El alcance metodológico, del desarrollo de la tesis, tiene un enfoque cuantitativo, precisamente la propuesta de aproximación es causal, descriptivo, se busca determinar la

causalidad entre los fenómenos de la democracia y el desarrollo humano para los países de los que se cuenta con información.

Capítulo 2

Marco Teórico

En este capítulo se desarrollará conceptos que se abordarán en la investigación, el mismo servirá para el proceso del marco metodológico; así también el abordaje de los aspectos teóricos que permitirá dar respuesta al problema de investigación que se planteó.

2.1. Índice de Democracia - IDEM

El IDEM, tiene la finalidad de evaluar la calidad de las democracias en los diferentes países del mundo, se basa en las calificaciones de 60 preguntas, recogidas de entrevistas a expertos, y el resto de las preguntas son respondidas a través de encuestas realizadas en los diferentes países, donde cada pregunta puede ser constestada con 0 ó 1 y 0.5 como respuesta intermedia. Al final se suma la puntuación total de las 60 preguntas que corresponde a 60 puntos como máximo, finalmente la puntuación se establece de 0 - 10 puntos [3]. Así entonces podemos expresar al IDEM como:

$$IDEM = \frac{\sum_{i=1}^{n=60} x_i}{n} \times 10 \quad (2.1)$$

Las cinco dimensiones que hacen al IDEM, se basan en las calificaciones de n_i preguntas de un total de 60, las dimensiones se expresan como:

$$IPRL = \frac{\sum_{i=1}^{n_1=12} x_{1i}}{12} \times 10$$

$$IFNC = \frac{\sum_{i=1}^{n_2=14} x_{2i}}{14} \times 10$$

$$IPRT = \frac{\sum_{i=1}^{n_3=9} x_{3i}}{9} \times 10$$

$$ICLT = \frac{\sum_{i=1}^{n_4=8} x_{4i}}{8} \times 10$$

$$ILBC = \frac{\sum_{i=1}^{n_5=17} x_{5i}}{17} \times 10$$

Donde IPRL=Índice de proceso electoral y pluralismo, IFNC=Índice del funcionamiento del gobierno, IPRT=Índice de participación política, ICLT=Índice de cultura política, ILBC=Índice de libertades civiles.

2.2. Índice de Desarrollo Humano - IDH

El IDH se entiende como un indicador que determina el nivel del desarrollo de los países del mundo, que toma en cuenta las dimensiones de: i) la esperanza de vida al nacer, que refleja la capacidad de llevar una vida larga y saludable, ii) los años promedio de escolaridad y los años esperados de escolaridad, que reflejan la capacidad de adquirir conocimientos, iii) el ingreso nacional bruto per cápita, que refleja el acceso a los recursos económicos necesarios para que las personas puedan tener un nivel de vida decente [17].

El IDH es la media geométrica de los índices normalizados que miden los logros en cada dimensión. En cada una de estas dimensiones se determinan mínimos (valores de subsistencia) para transformar los indicadores entre 0 y 1, fijando así un mínimo de 20 años para la esperanza de vida, cero años para las variables de educación y 100 USD para el ingreso nacional bruto per cápita. Así entonces podemos expresar al IDH como:

$$IDH = \sqrt[3]{IEV \times IE \times II} \quad (2.2)$$

donde las dimensiones del índice se expresan de la siguiente forma,

$$IEV = \frac{EV - 20}{\text{máx } EV - 20}$$

$$IE = \frac{\sqrt{IAPE \cdot IAEE}}{\text{máx } \sqrt{IAPE \cdot IAEE}}$$

$$II = \frac{\ln(Ipc) - \ln(100)}{\ln(40000) - \ln(100)}$$

tal que,

$$IAPE = \frac{APE}{\text{máx } APE - 0}$$

$$IAEE = \frac{AEE}{\text{máx } AEE - 0}$$

Donde: IEV = Índice de esperanza de vida, IE = Índice de educación, II = Índice de ingreso, $IAPE$ = Índice de años promedio de escolaridad, $IAEE$ = Índice de años esperados de escolaridad. De la misma manera se tiene AEE = Número de años esperados de escolaridad, Ipc = Ingreso PIB per cápita, EV = Esperanza de vida de un país expresada en años, APE = Número de años promedio de escolaridad [16].

2.3. Desarrollo humano y Democracia

Hyden et al. (2008) mediante metodologías comparativas de estudios de casos plantea que los avances en materia de desarrollo han sido explicados por la calidad de la gobernanza en términos de sociedad civil, sociedad política, eficacia del gobierno, calidad de la burocracia, la sociedad económica y el poder judicial [4].

Perspectivas que se plasman sobre la relación democracia y desarrollo a lo largo en los estudios del Programa de Naciones Unidas para el Desarrollo, como los informes de desarrollo humano, el PNUD (2002) en su informe “Profundizando la democracia en un mundo fragmentado”, ofrece un análisis de acerca del papel importante de la política en el desarrollo humano, examinando diferentes modos en que las instituciones democráticas pueden ayudar al fomento del progreso social y crecimiento económico [14].

Lipset (1994), indica que hay fuerte relación entre desarrollo y democracia, pero que la dirección de desarrollo hacia la democracia no podría ser porque pues las democracias se pueden dar en otros escenarios por ejemplo las revueltas sociales violentas o pacíficas, que no necesariamente repercuten en incremento en el nivel de vida de la población [20].

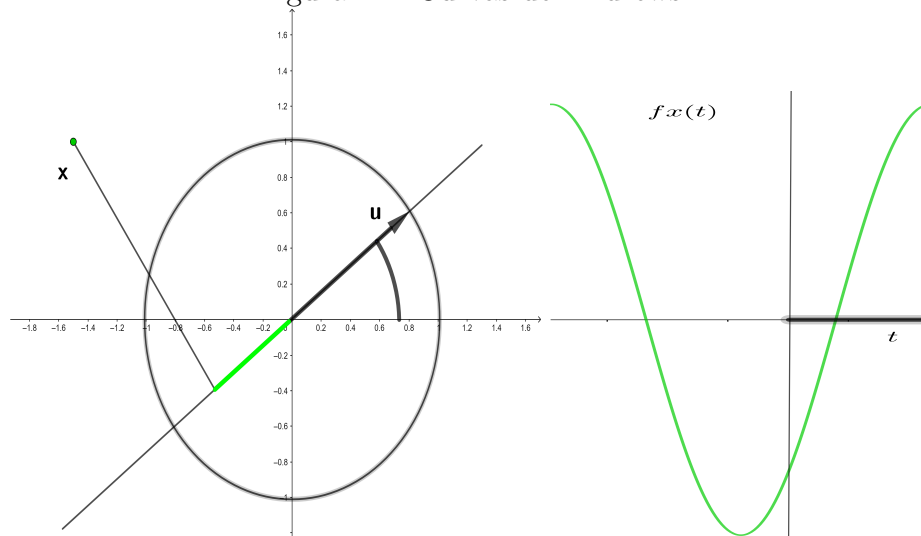
Tomando información empírica Bergger(2012) mediante modelos de regresión lineal múltiple de paneles transversales y agrupados, con período 1980-2010. Donde la democracia se mide por los derechos políticos y las libertades civiles de Freedom House, y el desarrollo humano se mide por el IDH. Los resultados indican que la democracia y la democratización tienen efectos positivos en los cambios en el desarrollo humano [12].

Buscando el enfoque de PNUD, Hyden y tomando en cuenta las definiciones antes desarrolladas como la de Lipset y acorde a lo que sugieren los diferentes autores se busca la relación de la democracia al desarrollo humano, si se genera democracia para la población se estaría generando desarrollo económico como consecuencia. Esto desarrollada mediante modelo de ecuaciones estructurales, donde previamente se contruye las variables tanto de democracia como de desarrollo humano, y luego buscar la relación de la democracia al desarrollo humano.

2.4. Curvas de Andrews

Dado un conjunto $x'_i = [x_{i1} \ x_{i2} \ \dots \ x_{ip}]$ Andrews permite visualizarlo en dos dimensiones, el gráfico de curvas de Andrews tiene como frecuencia el ángulo, y su amplitud corresponde a las longitudes de las proyecciones de los puntos k dimensional en el vector u de norma 1. La combinación de ángulo t y la proyección (es una posición en la curva de Andrews), variando t se tiene diferentes posiciones en el gráfico uniendo cada posición se encuentra la gráfica de Andrews ver figura (2.1).

Figura 2.1: Curvas de Andrews



Ese método se puede escribir basado en la transformación de Fourier para representar datos multivariantes en las dos dimensiones. Cada punto k dimensional es representado por la función

$$f_i(t) = \frac{x_{i1}}{\sqrt{2}} + x_{i2} \sin(t) + x_{i3} \cos(t) + x_{i4} \sin(2t) + x_{i5} \cos(2t) + \dots + \cos(4t)$$

en un rango $-\pi \leq t \leq \pi$ para permitir a $x'_i = [x_{i1} \ x_{i2} \dots \ x_{ip}]$ donde los coeficientes constituyen una serie de Fourier, y tomando el producto punto de funciones, cada una de las funciones en el dominio son ortogonales. Cada variable de cada observación es representado por una componente individual en la suma de la transformada de Fourier. La magnitud de cada variable de un sujeto o elemento particular afecta la frecuencia, la amplitud y la periodicidad de la función, dando una representación única para cada sujeto [5]. Por lo que el gráfico de curvas de Andrews sirve para detectar datos atípicos que aparecen como curvas únicas de Andrews que se ven diferentes del resto.

2.5. Ecuaciones Estructurales

2.5.1. Introducción

Los Modelos de Ecuaciones Estructurales en variables latentes tipo LISREL (Linear Structure Relationships), que expresaremos como SEM por sus siglas en inglés, permite estudiar las relaciones causales entre variables latentes, asumiendo linealidad en las relaciones existentes, se puede entender de forma esquemática como:

$$\text{Datos observados} \Rightarrow \text{Factores no observables} \Rightarrow \text{Datos observados}$$

Donde se asume que todos los datos observados deben ser estandarizados, con media cero, eso define un SEM LISREL sin intercepto. En este trabajo se desarrollará el SEM recursivo en los que la causalidad es unidireccional no hay causalidad de caminos bidireccionales.

2.5.2. Causalidad

Los conceptos sobre causalidad son varios, sin embargo nos basaremos en Bollen (1989) [10], así entonces se puede explicar lo siguiente.

Aislamiento

Sea x, y dos variables donde x es la variable independiente, y la variable dependiente, si la variable y esta aislada de toda influencia, salvo la influencia que ejerce x entonces se dice que x causa a y (esto en la realidad es imposible). Una relación causal se debe generalmente a factores no observables.

Asociación

Se dice que cuando, una supuesta causa y su efecto están aislados de otras influencias, entonces podrían estar asociados. Una asociación bivariada no es una condición necesaria ni suficiente para determinar una relación causal. En cambio la asociación junto con otros factores si podría considerarse como una relación causal. Las relaciones causales son estudiadas por las ecuaciones estructurales, es complicado tener certeza de una causa y su efecto empero los modelos de ecuaciones estructurales buscan plantear una aproximación lineal a tal situación.

Dirección de la influencia

En la relación causal el componente importante es la dirección de la causa $x_{i-h} \rightarrow y_i$, donde la variable que produce la causa requiere una prioridad temporal como una condición de causalidad $y_i = f(x_{i-h})$ tal que $i > h$.

2.5.3. Variables del modelo de ecuaciones estructurales

Las variables observadas

Son variables de manifiesto que son directamente observables y de tamaño $(p \times 1)$ se conoce como vector de variables observables que tienen relación con una o algunas variables latentes (no observables) endógenas, \mathbf{x} de tamaño $(q \times 1)$ se conoce como vector de variables observables que tienen relación con una o algunas variables latentes (no observables) exógenas.

Las variables latentes

Las variables latentes denominadas constructos, factores o variables no observables, son conceptos abstractos, que pueden ser observados indirectamente a través de sus efectos

en las variables observables, para encontrar dichas variables latentes se explica el análisis factorial exploratorio con las ecuaciones (2.3) y (2.4).

Las variables latentes exógenas

Son variables latentes independientes, que afectan a otras variables latentes endógenas y no reciben ningún efecto de ellas, se denotan por el vector $\xi_{(n \times 1)}$.

Las variables latentes endógenas

Son variables latentes dependientes que reciben el efecto de otras variables latentes exógenas y en particular son variables afectadas por perturbaciones o términos de error, se denota por el vector $\eta_{(m \times 1)}$.

Las variables de error

Son aquellas variables, que toman en cuenta todas las fuentes de variación y muchas veces se las conoce como variables latentes al no ser observables, se denotan por los vectores $\varepsilon_{(p \times 1)}$, $\delta_{(q \times 1)}$, $\zeta_{(m \times 1)}$ que corresponden a los vectores \mathbf{y} , \mathbf{x} , y η respectivamente.

Análisis factorial exploratorio

El análisis factorial exploratorio es una técnica que busca encontrar relaciones entre muchas variables que estén explicadas a través de un número reducido de factores comunes, no se conoce a priori la cantidad de factores que pueden existir. El modelo utilizado para relacionar los factores observados es el modelo de análisis de factores lineales y puede describirse de la siguiente forma.

$$\mathbf{x} = \mathbf{\Lambda}_x \xi + \delta \quad (2.3)$$

Donde \mathbf{x} es un vector $(q \times 1)$ de variables aleatorias, ξ es un vector $(n \times 1)$ de factores comunes, $n < q$, $\mathbf{\Lambda}_x$ es una matriz $(q \times n)$ de cargas factoriales y δ es un vector $(q \times 1)$ de factores específicos o errores de cada una de las variables. Se supone que,

$$\mathbf{E}(\delta) = \mathbf{0} \quad \mathbf{E}(\xi) = \mathbf{0}$$

$$\mathbf{E}(\delta\delta') = \mathbf{\Theta}_\delta \quad \mathbf{E}(\xi\xi') = \mathbf{0} \quad \mathbf{E}(\xi\xi') = \mathbf{\Phi}$$

donde Θ_δ es una matriz diagonal. Aplicando los supuestos correspondientes se tiene, la matriz de covarianza de los datos observados que se puede escribir en forma de la ecuación analítica factorial de la siguiente manera

$$\begin{aligned}\Sigma_{\mathbf{xx}} &= \mathbf{cov}(\mathbf{xx}') \\ \Sigma_{\mathbf{xx}} &= \mathbf{cov}[(\Lambda_{\mathbf{x}}\xi + \delta)(\xi'\Lambda'_{\mathbf{x}} + \delta')] \\ \Sigma_{\mathbf{xx}} &= \Lambda_{\mathbf{x}}\mathbf{E}(\xi\xi')\Lambda'_{\mathbf{x}} + \mathbf{E}(\delta\delta')\end{aligned}$$

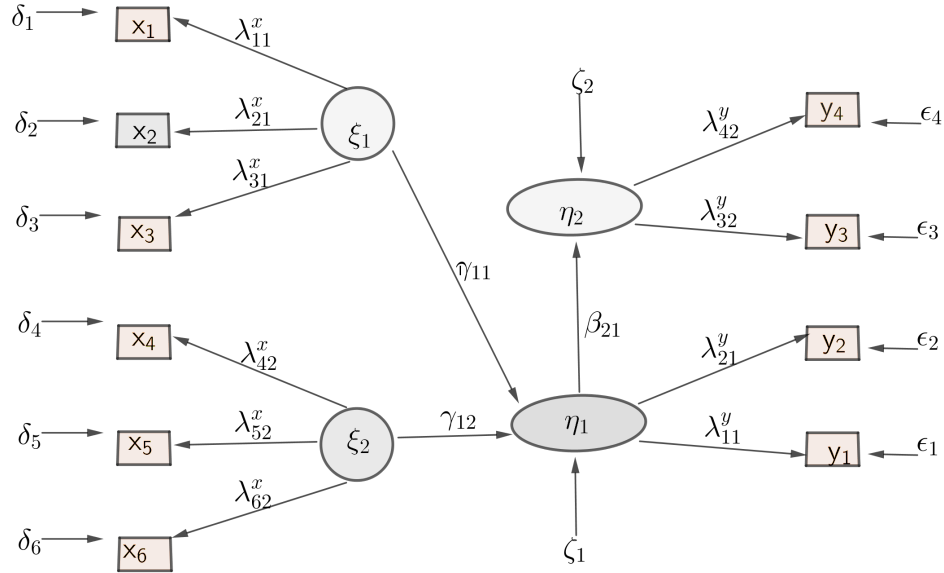
$$\Sigma_{\mathbf{xx}} = \Lambda_{\mathbf{x}}\Phi\Lambda'_{\mathbf{x}} + \Theta_\delta \quad (2.4)$$

Donde $\Sigma_{\mathbf{xx}}$ es una matriz $(q \times q)$ de covarianza poblacional, Φ matriz $(n \times n)$ de varianzas y covarianzas de los factores, y Θ_δ matriz $(q \times q)$ diagonal de varianzas únicas.

2.5.4. Diagrama de trayectorias

El diagrama de trayectorias conocido también como diagrama de senderos, permite diferenciar entre variables que no son influenciadas por otras (variables exógenas), y variables que son afectadas por otras (variables endógenas). Para tal efecto puede ver en la figura (2.2) los diagramas de ecuaciones estructurales, estos permiten la utilización de símbolos tales como el cuadrado que representa a las variables observadas, las variables no observables se representan con: un círculo para las exógenas y un ovalo para las endógenas, los errores no se representan ni por círculos, óvalos o cuadrados. Las relaciones unidireccionales se representan por flechas \mapsto . Así entonces se expresa el ejemplo de un diagrama de trayectorias que se observa en la figura (2.2).

Figura 2.2: Diagrama de trayectorias



2.5.5. Especificación del modelo

Luego y dado algunas características al respecto, el modelo de ecuaciones estructurales se define como:

$$\mathbf{y}_{\mathbf{p} \times 1} = \mathbf{\Lambda}_{\mathbf{y}_{\mathbf{p} \times \mathbf{m}}} \eta_{\mathbf{m} \times 1} + \epsilon_{\mathbf{p} \times 1} \quad (2.5)$$

$$\mathbf{x}_{\mathbf{q} \times 1} = \mathbf{\Lambda}_{\mathbf{x}_{\mathbf{q} \times \mathbf{n}}} \xi_{\mathbf{n} \times 1} + \delta_{\mathbf{q} \times 1} \quad (2.6)$$

$$\eta_{\mathbf{m} \times 1} = \mathbf{B}_{\mathbf{m} \times \mathbf{m}} \eta_{\mathbf{m} \times 1} + \mathbf{\Gamma}_{\mathbf{m} \times \mathbf{n}} \xi_{\mathbf{n} \times 1} + \zeta_{\mathbf{m} \times 1} \quad (2.7)$$

Donde (2.5) y (2.6) representan las ecuaciones de medida, también llamado primer modelo factorial para \mathbf{y} , y primer modelo factorial para \mathbf{x} respectivamente. Así \mathbf{y} , \mathbf{x} son vectores de variables observables de los factores η y ξ ; y $\mathbf{\Lambda}_{\mathbf{y}_{\mathbf{p} \times \mathbf{m}}}$, $\mathbf{\Lambda}_{\mathbf{x}_{\mathbf{q} \times \mathbf{n}}}$ son matrices de coeficientes que relaciona \mathbf{y} con η y \mathbf{x} con ξ respectivamente. Finalmente ϵ y δ son vectores que miden los errores, donde $\epsilon \sim \mathbf{N}_{\mathbf{p}}(\mathbf{0}, \mathbf{\Theta}_{\epsilon})$, $\delta \sim \mathbf{N}_{\mathbf{q}}(\mathbf{0}, \mathbf{\Theta}_{\delta})$ y donde ϵ, δ, ζ son mutuamente independientes.

De esta manera (2.7) representa el modelo estructural, donde $\eta_{(\mathbf{m} \times \mathbf{n})}$, $\xi_{(\mathbf{n} \times \mathbf{1})}$, y $\zeta_{(\mathbf{m} \times \mathbf{1})}$ representan vectores aleatorios de variables latentes endógena, exógena, y de error respectivamente; las matrices \mathbf{B} , $\mathbf{\Gamma}$, $\mathbf{\Lambda}$ son matrices de efectos directos; donde $\xi_{(\mathbf{n} \times \mathbf{1})} \sim \mathbf{N}_{\mathbf{n}}(\mathbf{0}, \mathbf{\Phi})$, $\zeta_{(\mathbf{m} \times \mathbf{1})} \sim \mathbf{N}_{\mathbf{m}}(\mathbf{0}, \mathbf{\Psi})$, además $\xi_{(\mathbf{n} \times \mathbf{1})}$ y $\zeta_{(\mathbf{m} \times \mathbf{1})}$ están incorrelacionadas.

Si trabajamos con la ecuación (2.7), restamos $\mathbf{B}\eta$ y realizamos operaciones se tiene la ecuación en terminos reducidos de la siguiente manera.

$$\begin{aligned} \eta - \mathbf{B}\eta &= \mathbf{\Gamma}\xi + \zeta \\ (\mathbf{I} - \mathbf{B})\eta &= \mathbf{\Gamma}\xi + \zeta \\ \eta &= (\mathbf{I} - \mathbf{B})^{-1}(\mathbf{\Gamma}\xi + \zeta) \end{aligned} \tag{2.8}$$

Donde suponemos que $(\mathbf{I} - \mathbf{B})$ es no singular, ahora la ecuación (2.8) permitirá operar y reducir una serie de operaciones más adelante.

2.5.6. Supuestos del modelo LISREL

Dos conjuntos de variables observadas \mathbf{x}, \mathbf{y} siguen un modelo factorial ξ, η en la que existe una relación de dependencia estructural entre los factores de ambos conjuntos, los datos provienen por un lado de \mathbf{x} asociada al factor independiente ξ expresada en la ecuación $\mathbf{x}_{\mathbf{q} \times \mathbf{1}} = \mathbf{\Lambda}_{\mathbf{x}_{\mathbf{q} \times \mathbf{n}}} \xi_{\mathbf{n} \times \mathbf{1}} + \delta_{\mathbf{q} \times \mathbf{1}}$, y por otro lado de \mathbf{y} asociada al factor dependiente η expresada en la ecuación $\mathbf{y}_{\mathbf{p} \times \mathbf{1}} = \mathbf{\Lambda}_{\mathbf{y}_{\mathbf{p} \times \mathbf{m}}} \eta_{\mathbf{m} \times \mathbf{1}} + \varepsilon_{\mathbf{p} \times \mathbf{1}}$, donde la relación estructural entre ξ, η se observa en la ecuación (2.7).

La hipótesis del modelo de ecuaciones estructurales, es probar que la matriz de varianzas y covarianzas poblacional $\mathbf{\Sigma}$, sea igual a la matriz de varianzas y covarianzas asociada al modelo planteado $\mathbf{\Sigma}(\theta)$, donde θ representa el vector parámetros del modelo $\theta = (\mathbf{\Lambda}_{\mathbf{x}}, \mathbf{\Lambda}_{\mathbf{y}}, \mathbf{B}, \mathbf{\Gamma}, \mathbf{\Phi}, \mathbf{\Theta}_{\varepsilon}, \mathbf{\Theta}_{\delta}, \mathbf{\Psi})$.

$$\mathbf{H}_0 : \mathbf{\Sigma} = \mathbf{\Sigma}(\theta) \tag{2.9}$$

En la práctica no se puede probar dicha igualdad como tal, entonces el objetivo queda en encontrar $\hat{\theta}$, de tal forma que $\mathbf{\Sigma}$ sea lo más parecido a $\mathbf{\Sigma}(\hat{\theta})$

$$\mathbf{H}_0 : \mathbf{\Sigma} = \mathbf{\Sigma}(\hat{\theta}) \tag{2.10}$$

Usaremos a la matriz de varianzas y covarianzas muestral \mathbf{S} , en lugar de $\mathbf{\Sigma}$, puesto que

Σ no es conocida explícitamente, así entonces \mathbf{S} permitirá comparar,

$$\mathbf{H}_0 : \mathbf{S} = \Sigma(\hat{\theta}) \quad (2.11)$$

Teorema 2.1 *Matriz de varianzas y covarianzas asociadas al modelo SEM*

$$\Sigma(\theta) = \left[\begin{array}{c|c} \Lambda_y(I-B)^{-1} [\Gamma\Phi\Gamma' + \Psi] (I-B)^{-1}\Lambda_y' + \Theta_\varepsilon & \Lambda_y(I-B)^{-1}\Gamma\Phi\Lambda_x' \\ \hline \Lambda_x\Phi\Gamma'(I-B)^{-1}\Lambda_y' & \Lambda_x\Phi\Lambda_x' + \Theta_\delta \end{array} \right]$$

Donde θ representa el vector parámetros del modelo.

$$\theta = (\Lambda_x, \Lambda_y, \mathbf{B}, \Gamma, \Phi, \Theta_\varepsilon, \Theta_\delta, \Psi)$$

Demostración. Partimos de la matriz de varianzas y covarianzas particionada de los vectores de variables observables \mathbf{y} y \mathbf{x} ,

$$\Sigma = \left[\begin{array}{c|c} \Sigma_{yy} & \Sigma_{yx} \\ \hline \Sigma_{xy} & \Sigma_{xx} \end{array} \right]$$

esta última expresión puede presentarse como,

$$\Sigma = \left[\begin{array}{c|c} \mathbf{cov}(\mathbf{y}) & \mathbf{cov}(\mathbf{y}, \mathbf{x}) \\ \hline \mathbf{cov}(\mathbf{x}, \mathbf{y}) & \mathbf{cov}(\mathbf{x}) \end{array} \right]$$

entonces los componentes de la matriz a desarrollar son Σ_{yy} , Σ_{yx} , Σ_{xy} , y Σ_{xx} .

Primero desarrollamos para,

$$\Sigma_{yy} = \mathbf{cov}(\mathbf{y}) = \mathbf{E} [\mathbf{y}\mathbf{y}'] = \mathbf{cov} [\Lambda_y\eta + \varepsilon]$$

usando definiciones y conocido que η , ξ son independientes tenemos,

$$\Sigma_{yy} = \Lambda_y \mathbf{cov}(\eta) \Lambda_y' + \mathbf{cov}(\varepsilon)$$

$$\Sigma_{yy} = \Lambda_y \mathbf{cov}[(\mathbf{I} - \mathbf{B})^{-1}(\Gamma\xi + \zeta)] \Lambda_y' + \Theta_\varepsilon$$

así entonces obtenemos,

$$\Sigma_{yy} = \Lambda_y(\mathbf{I} - \mathbf{B})^{-1} [\Gamma\Phi\Gamma' + \Psi] (\mathbf{I} - \mathbf{B})^{-1}\Lambda_y' + \Theta_\varepsilon \quad (2.12)$$

De igual forma para el segundo caso,

$$\Sigma_{\mathbf{y}\mathbf{x}} = \mathbf{cov}(\mathbf{y}\mathbf{x}) = \mathbf{E} [\mathbf{y}\mathbf{x}']$$

reemplazando las ecuaciones de medida (2.5) y (2.6) se tiene,

$$\Sigma_{\mathbf{y}\mathbf{x}} = \mathbf{E} [(\Lambda_{\mathbf{y}}\eta + \varepsilon)(\Lambda_{\mathbf{x}}\xi + \delta)']$$

aplicamos distribución y transpuesta,

$$\Sigma_{\mathbf{y}\mathbf{x}} = \mathbf{E} [\Lambda_{\mathbf{y}}\eta\xi' \Lambda_{\mathbf{x}}' + \Lambda_{\mathbf{y}}\eta\delta' + \varepsilon\xi' \Lambda_{\mathbf{x}}' + \varepsilon\delta']$$

las cargas factoriales Λ son constantes, tenemos,

$$\Sigma_{\mathbf{y}\mathbf{x}} = \Lambda_{\mathbf{y}}\mathbf{E}[\eta\xi']\Lambda_{\mathbf{x}}' + \Lambda_{\mathbf{y}}\mathbf{E}(\eta\delta') + \mathbf{E}[\varepsilon\xi']\Lambda_{\mathbf{x}}' + \mathbf{E}(\varepsilon\delta')$$

$$\Sigma_{\mathbf{y}\mathbf{x}} = \Lambda_{\mathbf{y}}\mathbf{E}[\eta\xi']\Lambda_{\mathbf{x}}'$$

reemplazando la ecuación estructural en su forma (2.8),

$$\Sigma_{\mathbf{y}\mathbf{x}} = \Lambda_{\mathbf{y}}\mathbf{E}[(\mathbf{I} - \mathbf{B})^{-1}(\Gamma\xi + \zeta)\xi']\Lambda_{\mathbf{x}}'$$

aplicando distribución,

$$\Sigma_{\mathbf{y}\mathbf{x}} = \Lambda_{\mathbf{y}}\mathbf{E}[(\mathbf{I} - \mathbf{B})^{-1}\Gamma\xi\xi' + (\mathbf{I} - \mathbf{B})^{-1}\zeta\xi']\Lambda_{\mathbf{x}}'$$

$$\Sigma_{\mathbf{y}\mathbf{x}} = \Lambda_{\mathbf{y}}(\mathbf{I} - \mathbf{B})^{-1}\Gamma\mathbf{E}(\xi\xi')\Lambda_{\mathbf{x}}' + (\mathbf{I} - \mathbf{B})^{-1}\mathbf{E}(\zeta\xi')\Lambda_{\mathbf{x}}'$$

se tiene,

$$\Sigma_{\mathbf{y}\mathbf{x}} = \Lambda_{\mathbf{y}}(\mathbf{I} - \mathbf{B})^{-1}\Gamma\Phi\Lambda_{\mathbf{x}}' \quad (2.13)$$

Para el caso del tercer componente, este corresponde a la transpuesta de (2.13), entonces tenemos que

$$\Sigma_{\mathbf{x}\mathbf{y}} = \Lambda_{\mathbf{x}}\Phi\Gamma'(\mathbf{I} - \mathbf{B})^{-1}\Lambda_{\mathbf{y}}' \quad (2.14)$$

Finalmente para el cuarto caso,

$$\Sigma_{\mathbf{xx}} = \text{cov}(\mathbf{x}) = \mathbf{E}[\mathbf{xx}'] = \text{cov}[\Lambda_{\mathbf{x}} \xi + \delta]$$

$$\Sigma_{\mathbf{xx}} = \Lambda_{\mathbf{x}} \text{cov}(\xi) \Lambda_{\mathbf{x}}' + \text{cov}(\delta)$$

obtenemos por (2.4),

$$\Sigma_{\mathbf{xx}} = \Lambda_{\mathbf{x}} \Phi \Lambda_{\mathbf{x}}' + \Theta_{\delta} \quad (2.15)$$

así entonces,

$$\Sigma = \left[\begin{array}{c|c} \Lambda_{\mathbf{y}}(I - B)^{-1} [\Gamma \Phi \Gamma' + \Psi] (I - B)^{-1} \Lambda_{\mathbf{y}}' + \Theta_{\varepsilon} & \Lambda_{\mathbf{y}}(I - B)^{-1} \Gamma \Phi \Lambda_{\mathbf{x}}' \\ \hline \Lambda_{\mathbf{x}} \Phi \Gamma' (I - B)^{-1} \Lambda_{\mathbf{y}}' & \Lambda_{\mathbf{x}} \Phi \Lambda_{\mathbf{x}}' + \Theta_{\delta} \end{array} \right]$$

Donde θ representa el vector parámetros del modelo.

$$\theta = (\Lambda_{\mathbf{x}}, \Lambda_{\mathbf{y}}, \mathbf{B}, \Gamma, \Phi, \Theta_{\varepsilon}, \Theta_{\delta}, \Psi)$$

Definición 2.1 Matriz de varianzas y covarianzas muestral

Si se cuentan con n vectores de observaciones $[\mathbf{y}'_j \ \mathbf{x}'_j]$ con $j = 1, 2, \dots, n$ la matriz de covarianzas muestral es,

$$\mathbf{S} = \left[\begin{array}{c|c} \mathbf{S}_{\mathbf{yy}} & \mathbf{S}_{\mathbf{yx}} \\ \hline \mathbf{S}_{\mathbf{xy}} & \mathbf{S}_{\mathbf{xx}} \end{array} \right]$$

Tomando en cuenta la definición (2.1) la información contenida en \mathbf{S} puede ser utilizada para estimar los parámetros desconocidos del modelo, la estimación de dichos parámetros son representados por

$$\hat{\theta} = (\hat{\Lambda}_{\mathbf{x}}, \hat{\Lambda}_{\mathbf{y}}, \hat{\mathbf{B}}, \hat{\Gamma}, \hat{\Phi}, \hat{\Theta}_{\varepsilon}, \hat{\Theta}_{\delta}, \hat{\Psi})$$

2.5.7. Parámetros de las ecuaciones

Para estimar los parámetros de los modelos se necesita más ecuaciones que incógnitas por lo tanto, si \mathbf{t} es el número de parámetros desconocidos, y \mathbf{p} , \mathbf{q} son las dimensiones conocidas de \mathbf{y} , \mathbf{x} , entonces toda esta información debe satisfacer:

$$\mathbf{t} \leq (\mathbf{p} + \mathbf{q})(\mathbf{p} + \mathbf{q} + 1)/2$$

Esta condición es necesaria mas no suficiente para garantizar la identificabilidad del modelo.

2.5.8. Estimación

Dado un modelo identificado, lo siguiente que se debe hacer es estimar los parámetros del modelo en $\Sigma(\theta)$ para el modelo de ecuaciones estructurales. Para estimar los parámetros en $\Sigma(\theta)$, se obtiene una estimación muestral de $\Sigma(\theta)$ y se elige una función $\mathbf{F}(\mathbf{S}, \Sigma(\theta)) \geq \mathbf{0}$ que minimice el error, la estimación es un proceso de aproximación de $\hat{\theta}$, aproximamos F utilizando el método de máxima verosimilitud que se expresa como.

$$\mathbf{F}_{\text{ML}} = \log |\Sigma(\theta)| + \text{tr}(\mathbf{S}\Sigma^{-1}(\theta)) - \log |\mathbf{S}| - (\mathbf{p} + \mathbf{q}) \quad (2.16)$$

Como la estimación es un procedimiento iterativo se generaran $\theta_1, \theta_2, \dots$, tal que F es una función de discrepancia.

$$\mathbf{F}(\mathbf{S}, \Sigma(\theta_{r+1})) < \mathbf{F}(\mathbf{S}, \Sigma(\theta_r))$$

Para elegir el punto inicial θ_1 , la regla de detención es que las magnitudes absolutas de las derivadas de primer orden de \mathbf{F} con respecto a los parámetros sean todas menores que algún pequeño valor positivo, y así el procedimiento debe continuar hasta que no sea posible una disminución adicional de \mathbf{F} , para esto utilizamos el algoritmo de Newton-Raphson, entonces son necesarias la primera y segunda derivada de \mathbf{F} , para ello tomaremos en cuenta Lee (2007) [21] entonces, para la primera derivada tenemos,

$$\frac{\partial \mathbf{F}}{\partial \theta} = \dot{\mathbf{F}}(\theta) = \frac{\partial}{\partial \theta} [\log |\Sigma| + \text{tr}(\mathbf{S}\Sigma^{-1})]$$

$$\dot{\mathbf{F}}(\theta) = \frac{\partial \log |\Sigma|}{\partial \theta} + \dot{\Sigma}(\theta) \frac{\partial \Sigma^{-1}}{\partial \Sigma} \frac{\partial \text{tr}(\mathbf{S}\Sigma^{-1})}{\partial \Sigma^{-1}}$$

$$\dot{\mathbf{F}}(\theta) = \dot{\Sigma}(\theta) \text{vec}(\Sigma^{-1}) + \dot{\Sigma}(\theta) (-\Sigma^{-1} \otimes \Sigma^{-1}) (\mathbf{S} \otimes \mathbf{I}_p) \text{vec} \mathbf{I}_p$$

donde \otimes representa el producto kronecker, y vec representa la vectorización de matrices.

$$\dot{\mathbf{F}}(\theta) = \dot{\Sigma}(\theta) \text{vec}(\Sigma^{-1}) - \dot{\Sigma}(\theta) \text{vec}(\Sigma^{-1} \mathbf{S} \Sigma^{-1})$$

$$\dot{\mathbf{F}}(\theta) = -\dot{\Sigma}(\theta)\text{vec}[\Sigma^{-1}(\mathbf{S} - \Sigma)\Sigma^{-1}]$$

así entonces la primera derivada queda expresada como

$$\dot{\mathbf{F}}(\theta) = -\dot{\Sigma}(\theta)(\Sigma^{-1} \otimes \Sigma^{-1})\text{vec}(\mathbf{S} - \Sigma) \quad (2.17)$$

Para la segunda derivada, partimos de la primera derivada,

$$\frac{\partial^2 \mathbf{F}}{\partial \theta \partial \theta} = \ddot{\mathbf{F}}(\theta) = \frac{\partial}{\partial \theta} \{-\dot{\Sigma}(\theta)\text{vec}[\Sigma^{-1}(\mathbf{S} - \Sigma)\Sigma^{-1}]\}$$

$$\ddot{\mathbf{F}}(\theta) = -\ddot{\Sigma}(\theta)\{\mathbf{I}_q \otimes \text{vec}[\Sigma^{-1}(\mathbf{S} - \Sigma)\Sigma^{-1}]\} - \frac{\partial \Sigma^{-1}(\mathbf{S} - \Sigma)\Sigma^{-1}}{\partial \theta} [\dot{\Sigma}(\theta)' \otimes \mathbf{1}]$$

$$\ddot{\mathbf{F}}(\theta) = \dot{\Sigma}(\theta)(\Sigma^{-1} \otimes \Sigma^{-1})\dot{\Sigma}(\theta)' - \ddot{\Sigma}(\theta)[\mathbf{I}_q \otimes (\Sigma^{-1} \otimes \Sigma^{-1})\text{vec}(\mathbf{S} - \Sigma)]$$

$$-\dot{\Sigma}(\theta)\{[\Sigma^{-1} \otimes \Sigma^{-1}(\mathbf{S} - \Sigma)\Sigma^{-1}] + [\Sigma^{-1}(\mathbf{S} - \Sigma)\Sigma^{-1} \otimes \Sigma^{-1}]\}\dot{\Sigma}(\theta)'$$

así entonces tenemos la segunda derivada, queda expresada como

$$\begin{aligned} \ddot{\mathbf{F}}(\theta) &= \dot{\Sigma}(\theta)(\Sigma^{-1} \otimes \Sigma^{-1})\dot{\Sigma}(\theta)' - 2\dot{\Sigma}(\theta)[\Sigma^{-1} \otimes \Sigma^{-1}(\mathbf{S} - \Sigma)\Sigma^{-1}]\dot{\Sigma}(\theta)' \\ &\quad - \dot{\Sigma}(\theta)[\mathbf{I}_q \otimes (\Sigma^{-1} \otimes \Sigma^{-1})\text{vec}(\mathbf{S} - \Sigma)] \end{aligned} \quad (2.18)$$

y el método de Newton-Raphson consiste en iterar

$$\theta_{(\mathbf{r}+1)} = \theta_{(\mathbf{r})} + \lambda_{(\mathbf{r})}[\mathbf{H}_{(\mathbf{r})}(\theta_{(\mathbf{r})})]^{-1}\mathbf{g}(\theta_{(\mathbf{r})})$$

donde $\theta_{(\mathbf{r})}$ contiene el parámetro para la r -ava iteración y como el procedimiento es iterativo $\lambda_{\mathbf{r}}$ es un parámetro de paso inicial (denominado punto de arranque o valor supuesto) para poder encontrar estimaciones, toma un valor supuesto de tamaño $\lambda_{\mathbf{r}} = \mathbf{1}$; $\mathbf{H}_{(\mathbf{r})}(\theta_{(\mathbf{r})})$ es una aproximación a la matriz Hessiana que contiene las segundas derivaciones de F con respecto a los parámetros evaluados en $\theta_{(\mathbf{r})}$ y $\mathbf{g}(\theta_{(\mathbf{r})})$ es el gradiente negativo de F evaluado en $\theta_{(\mathbf{r})}$.

Capítulo 3

Marco Metodológico

3.1. Diseño de investigación

El modelo metodológico escogido, trata de un estudio explicativo causal y descriptivo; como proceso formal se trata del método deductivo i.e. que parte de una premisa general para obtener las conclusiones de un caso en particular; la dimensión cronológica es de corte transversal. El enfoque es la matemática aplicada, que trata de identificar las relaciones causales entre factores que no son medibles de manera directa como el desarrollo humano y la democracia, que es de interés de los policy makers y de los diferentes organismos internacionales.

3.2. Definición de las variables

Las variables a utilizar para poder modelar la ecuación estructural son 5 variables que están relacionadas con la democracia: x_1 procesos electoral y pluralismo, x_2 libertades civiles, x_3 funcionamiento del gobierno, x_4 participación política, y x_5 cultura política, estas variables están medidas en una escala de 0-10 son continuas.

Por otro lado, otro grupo de 4 variables relacionadas con el desarrollo humano: y_1 esperanza de vida, y_2 años promedio de escolaridad, y_3 años esperados de escolarización, y_4 ingresos per cápita, estos índices están medidos en años excepto el ingreso per cápita que esta en dólares a precios constantes del 2017 y el IDH tiene una escala de 0-1 son continuas.

Estas variables pueden conformarse en vectores de variables \mathbf{y} de tamaño (4×1) y \mathbf{x} de tamaño (5×1) , al respecto se cuenta con información disponible de 165 países del mundo.

La información de \mathbf{x} es publicada por The Economist, la información de \mathbf{y} es publicada por el Programa de Naciones Unidas PNUD. La información disponible de manera pública que el PNUD presenta es para el año 2019, esta es la característica que hace que, el corte transversal de estudio sea para ese año 2019. La información recolectada entonces corresponde a una matriz $\mathbf{W}_{\mathbf{n} \times \mathbf{p}}$ con $n = 165$ países y $p^* = 9$ variables observables donde $p^* = p + q$.

La información recogida en la matriz $\mathbf{W}_{\mathbf{n} \times \mathbf{p}^*}$ es estandarizada en una matriz $\mathbf{Z}_{\mathbf{n} \times \mathbf{p}^*}$, como indica el modelo de ecuaciones estructurales.

Para la estandarización de $\mathbf{W}_{\mathbf{n} \times \mathbf{p}^*}$,

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1p} \\ w_{21} & w_{22} & \cdots & w_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ w_{n1} & w_{n2} & \cdots & w_{np} \end{bmatrix}$$

denotamos al vector de medias como,

$$\bar{\mathbf{w}} = \mathbf{W}' \mathbf{1}_n / \mathbf{n}$$

y a \mathbf{X} como la matriz centrada, que se obtiene de la siguiente forma,

$$\mathbf{X} = \mathbf{W} - \mathbf{1}_n \bar{\mathbf{w}}'$$

$$\mathbf{X} = \mathbf{W} - \mathbf{1}_n (\mathbf{W}' \mathbf{1}_n / \mathbf{n})'$$

$$\mathbf{X} = \mathbf{W} - \mathbf{1}_n \mathbf{1}_n' \mathbf{W} / \mathbf{n}$$

$$\mathbf{X} = (\mathbf{I} - \mathbf{1}_n \mathbf{1}_n' / \mathbf{n}) \mathbf{W}$$

$$\mathbf{X} = \mathbf{H} \mathbf{W}$$

donde \mathbf{H} es una matriz idempotente. Así entonces se define a \mathbf{Z} , como la matriz estandarizada que se expresa como,

$$\mathbf{Z}_{\mathbf{n} \times \mathbf{p}^*} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{bmatrix}$$

$$\mathbf{Z} = \mathbf{HXD}^{-\frac{1}{2}} \tag{3.1}$$

donde $\mathbf{D} = \text{diag}(\mathbf{S})$ y \mathbf{D} se expresa como,

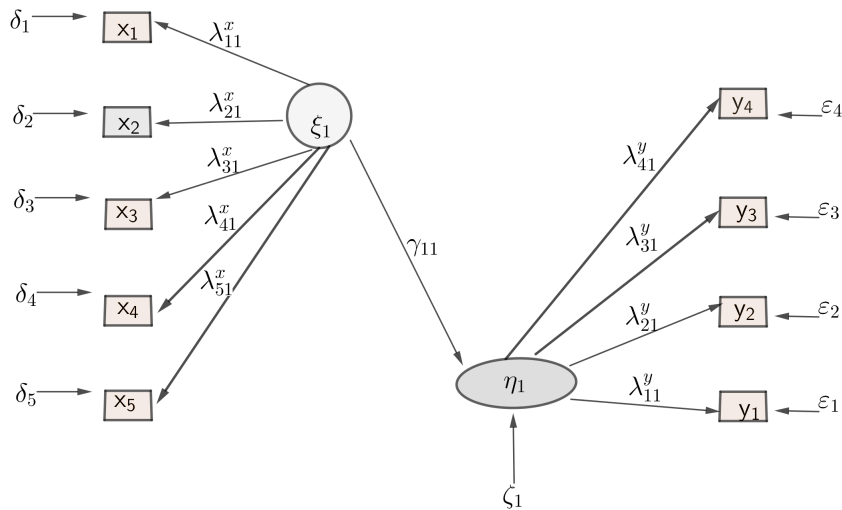
$$\mathbf{D} = \begin{bmatrix} s_{11} & 0 & \cdots & 0 \\ 0 & s_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & s_{np} \end{bmatrix}$$

$$\text{tal que } s_{ij} = \sum_{i=1}^n (\mathbf{w}_i - \bar{\mathbf{w}})(\mathbf{w}_i - \bar{\mathbf{w}})' \mathbf{1}/n_j$$

3.2.1. Diagrama de trayectorias del modelo

El modelo presenta la estructura de la figura (4.3) que describe las trayectorias de las variables latentes ξ , η hacia las variables de manifiesto del modelo \mathbf{x} , \mathbf{y} , así también la dirección de la causalidad de la variable latente ξ hacia η , y como es debido los errores ε y δ con las direcciones hacia las variables observables \mathbf{x} , \mathbf{y} , finalmente el error ζ hacia la variable η .

Figura 3.1: Diagrama de trayectorias del modelo



3.3. Estructura del modelo

Apriori, se conoce que \mathbf{y} es utilizada por PNUD para calcular el índice de desarrollo humano -IDH, y \mathbf{x} es utilizada por The economist para calcular el índice de democracia - IDEM. Así también se persigue el siguiente supuesto, la democracia causa al desarrollo i.e. IDEM \mapsto IDH entonces podemos expresar las ecuaciones de medida de la siguiente manera:

Ecuaciones de medida, de variable endógena

$$\mathbf{y}_1 = \lambda_{11}^y \eta_1 + \varepsilon_1$$

$$\mathbf{y}_2 = \lambda_{21}^y \eta_1 + \varepsilon_2$$

$$\mathbf{y}_3 = \lambda_{31}^y \eta_1 + \varepsilon_3$$

$$\mathbf{y}_4 = \lambda_{41}^y \eta_1 + \varepsilon_4$$

$$\mathbf{y} = \Lambda_y \eta + \varepsilon$$

Ecuación de medida de variable exógena

$$\mathbf{x}_1 = \lambda_{11}^x \xi_1 + \delta_1$$

$$\mathbf{x}_2 = \lambda_{21}^x \xi_1 + \delta_2$$

$$\mathbf{x}_3 = \lambda_{31}^x \xi_1 + \delta_3$$

$$\mathbf{x}_4 = \lambda_{41}^x \xi_1 + \delta_4$$

$$\mathbf{x}_5 = \lambda_{51}^x \xi_1 + \delta_5$$

$$\mathbf{x} = \Lambda_x \xi + \delta$$

A partir de la información apriori, se construirán los factores **IDH** e **IDEM** que se denominan variables latentes η , ξ respectivamente, en adelante se expresa el modelo de ecuaciones estructurales:

$$\mathbf{y}_{4 \times 1} = \Lambda_{y_{4 \times 1}} \eta_{1 \times 1} + \varepsilon_{4 \times 1} \quad (3.2)$$

$$\mathbf{x}_{5 \times 1} = \Lambda_{x_{5 \times 1}} \xi_{1 \times 1} + \delta_{5 \times 1} \quad (3.3)$$

$$\eta_{1x1} = \mathbf{\Gamma}_{1x1}\xi_{1x1} + \varsigma_{1x1} \quad (3.4)$$

Para poder desarrollar el supuesto del modelo de ecuaciones estructurales, expresaremos las ecuaciones (3.2) y (3.3) respectivamente de la siguiente forma,

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \lambda_{11}^y \\ \lambda_{21}^y \\ \lambda_{31}^y \\ \lambda_{41}^y \end{bmatrix} \eta_{11} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} \lambda_{11}^x \\ \lambda_{21}^x \\ \lambda_{31}^x \\ \lambda_{41}^x \\ \lambda_{51}^x \end{bmatrix} \xi_{11} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{bmatrix}$$

y el modelo de ecuación estructural queda de la forma,

$$\eta_{11} = \gamma_{11}\xi_{11} + \varsigma_{11} \quad (3.5)$$

3.3.1. Identificación

Para estimar los parámetros se necesita más ecuaciones que incógnitas, tomando en cuenta la ecuación (3.4) donde se identifica que no participa el coeficiente \mathbf{B} , desarrollamos la matriz de varianzas y covarianzas de la siguiente manera

$$\Sigma(\hat{\theta}) = \left[\begin{array}{c|c} \mathbf{\Lambda}_y [\mathbf{\Gamma}\mathbf{\Phi}\mathbf{\Gamma}' + \mathbf{\Psi}] \mathbf{\Lambda}'_y + \mathbf{\Theta}_\varepsilon & \mathbf{\Lambda}_y \mathbf{\Gamma}\mathbf{\Phi}\mathbf{\Lambda}'_x \\ \hline \mathbf{\Lambda}_x \mathbf{\Phi} \mathbf{\Gamma}' \mathbf{\Lambda}'_y & \mathbf{\Lambda}_x \mathbf{\Phi}\mathbf{\Lambda}'_x + \mathbf{\Theta}_\delta \end{array} \right]$$

De tal forma que,

$$\Sigma_{yy} = \mathbf{\Lambda}_y [\mathbf{\Gamma}\mathbf{\Phi}\mathbf{\Gamma}' + \mathbf{\Psi}] \mathbf{\Lambda}'_y + \mathbf{\Theta}_\varepsilon$$

$$\Sigma_{yy} = \begin{bmatrix} \lambda_{11}^y \\ \lambda_{21}^y \\ \lambda_{31}^y \\ \lambda_{41}^y \end{bmatrix} \left(\gamma_{11}^2 \phi_{11} + \psi_{11} \right) \begin{bmatrix} \lambda_{11}^y & \lambda_{21}^y & \lambda_{31}^y & \lambda_{41}^y \end{bmatrix} + \begin{bmatrix} \theta_{11}^\varepsilon & 0 & 0 & 0 \\ 0 & \theta_{22}^\varepsilon & 0 & 0 \\ 0 & 0 & \theta_{33}^\varepsilon & 0 \\ 0 & 0 & 0 & \theta_{44}^\varepsilon \end{bmatrix}$$

$$\Sigma_{yy} = \gamma_{11}^2 \phi_{11} + \psi_{11} \begin{bmatrix} \lambda_{11}^{2y} & \lambda_{11}^y \lambda_{21}^y & \lambda_{11}^y \lambda_{31}^y & \lambda_{11}^y \lambda_{41}^y \\ \lambda_{11}^y \lambda_{21}^y & \lambda_{21}^{2y} & \lambda_{21}^y \lambda_{31}^y & \lambda_{21}^y \lambda_{41}^y \\ \lambda_{11}^y \lambda_{31}^y & \lambda_{21}^y \lambda_{31}^y & \lambda_{31}^{2y} & \lambda_{31}^y \lambda_{41}^y \\ \lambda_{11}^y \lambda_{41}^y & \lambda_{21}^y \lambda_{41}^y & \lambda_{31}^y \lambda_{41}^y & \lambda_{41}^{2y} \end{bmatrix} + \begin{bmatrix} \theta_{11}^\varepsilon & 0 & 0 & 0 \\ 0 & \theta_{22}^\varepsilon & 0 & 0 \\ 0 & 0 & \theta_{33}^\varepsilon & 0 \\ 0 & 0 & 0 & \theta_{44}^\varepsilon \end{bmatrix}$$

$$\Sigma_{yy} = \gamma_{11}^2 \phi_{11} + \psi_{11} \begin{bmatrix} \lambda_{11}^{2y} + \theta_{11}^\varepsilon & \lambda_{11}^y \lambda_{21}^y & \lambda_{11}^y \lambda_{31}^y & \lambda_{11}^y \lambda_{41}^y \\ \lambda_{11}^y \lambda_{21}^y & \lambda_{21}^{2y} + \theta_{22}^\varepsilon & \lambda_{21}^y \lambda_{31}^y & \lambda_{21}^y \lambda_{41}^y \\ \lambda_{11}^y \lambda_{31}^y & \lambda_{21}^y \lambda_{31}^y & \lambda_{31}^{2y} + \theta_{33}^\varepsilon & \lambda_{31}^y \lambda_{41}^y \\ \lambda_{11}^y \lambda_{41}^y & \lambda_{21}^y \lambda_{41}^y & \lambda_{31}^y \lambda_{41}^y & \lambda_{41}^{2y} + \theta_{44}^\varepsilon \end{bmatrix}$$

de la misma forma,

$$\Sigma_{yx} = \Lambda_y \Gamma \Phi \Lambda_x'$$

$$\Sigma_{yx} = \begin{bmatrix} \lambda_{11}^y \\ \lambda_{21}^y \\ \lambda_{31}^y \\ \lambda_{41}^y \end{bmatrix} (\gamma_{11}^2 \phi_{11}) \begin{bmatrix} \lambda_{11}^x & \lambda_{21}^x & \lambda_{31}^x & \lambda_{41}^x & \lambda_{51}^x \end{bmatrix}$$

$$\Sigma_{yx} = \gamma_{11}^2 \phi_{11} \begin{bmatrix} \lambda_{11}^x \lambda_{11}^y & \lambda_{11}^y \lambda_{21}^x & \lambda_{11}^y \lambda_{31}^x & \lambda_{11}^y \lambda_{41}^x & \lambda_{11}^y \lambda_{51}^x \\ \lambda_{11}^x \lambda_{21}^y & \lambda_{21}^x \lambda_{21}^y & \lambda_{21}^x \lambda_{31}^y & \lambda_{21}^x \lambda_{41}^y & \lambda_{21}^x \lambda_{51}^y \\ \lambda_{11}^x \lambda_{31}^y & \lambda_{21}^x \lambda_{31}^y & \lambda_{31}^x \lambda_{31}^y & \lambda_{31}^x \lambda_{41}^y & \lambda_{31}^x \lambda_{51}^y \\ \lambda_{11}^x \lambda_{41}^y & \lambda_{21}^x \lambda_{41}^y & \lambda_{31}^x \lambda_{41}^y & \lambda_{41}^x \lambda_{41}^y & \lambda_{41}^x \lambda_{51}^y \end{bmatrix}$$

así mismo,

$$\Sigma_{xy} = \Lambda_x \Phi \Gamma' \Lambda_y'$$

$$\Sigma_{xy} = \begin{bmatrix} \lambda_{11}^x \\ \lambda_{21}^x \\ \lambda_{31}^x \\ \lambda_{41}^x \\ \lambda_{51}^x \end{bmatrix} (\phi_{11} \gamma_{11}^2) \begin{bmatrix} \lambda_{11}^y & \lambda_{21}^y & \lambda_{31}^y & \lambda_{41}^y \end{bmatrix}$$

$$\Sigma_{\mathbf{xy}} = \gamma_{11}^2 \phi_{11} \begin{bmatrix} \lambda_{11}^x \lambda_{11}^y & \lambda_{11}^x \lambda_{21}^y & \lambda_{11}^x \lambda_{31}^y & \lambda_{11}^x \lambda_{41}^y \\ \lambda_{11}^y \lambda_{21}^x & \lambda_{21}^y \lambda_{21}^x & \lambda_{21}^y \lambda_{31}^x & \lambda_{21}^y \lambda_{41}^x \\ \lambda_{11}^y \lambda_{31}^x & \lambda_{21}^y \lambda_{31}^x & \lambda_{31}^y \lambda_{31}^x & \lambda_{31}^y \lambda_{41}^x \\ \lambda_{11}^y \lambda_{41}^x & \lambda_{21}^y \lambda_{41}^x & \lambda_{31}^y \lambda_{41}^x & \lambda_{41}^y \lambda_{41}^x \\ \lambda_{11}^y \lambda_{51}^x & \lambda_{21}^y \lambda_{51}^x & \lambda_{31}^y \lambda_{51}^x & \lambda_{41}^y \lambda_{51}^x \end{bmatrix}$$

finalmente,

$$\Sigma_{\mathbf{xx}} = \Lambda_{\mathbf{x}} \Phi \Lambda_{\mathbf{x}}' + \Theta_{\delta}$$

$$\Sigma_{\mathbf{xx}} = \begin{bmatrix} \lambda_{11}^x \\ \lambda_{21}^x \\ \lambda_{31}^x \\ \lambda_{41}^x \\ \lambda_{51}^x \end{bmatrix} (\phi_{11}) \begin{bmatrix} \lambda_{11}^x & \lambda_{21}^x & \lambda_{31}^x & \lambda_{41}^x & \lambda_{51}^x \end{bmatrix} + \begin{bmatrix} \theta_{11}^{\delta} & 0 & 0 & 0 & 0 \\ 0 & \theta_{22}^{\delta} & 0 & 0 & 0 \\ 0 & 0 & \theta_{33}^{\delta} & 0 & 0 \\ 0 & 0 & 0 & \theta_{44}^{\delta} & 0 \\ 0 & 0 & 0 & 0 & \theta_{55}^{\delta} \end{bmatrix}$$

$$\Sigma_{\mathbf{xx}} = \phi_{11} \begin{bmatrix} \lambda_{11}^{2x} + \theta_{11} & \lambda_{11}^x \lambda_{21}^x & \lambda_{11}^x \lambda_{31}^x & \lambda_{11}^x \lambda_{41}^x & \lambda_{11}^x \lambda_{51}^x \\ \lambda_{11}^x \lambda_{21}^x & \lambda_{21}^{2x} + \theta_{22} & \lambda_{21}^x \lambda_{31}^x & \lambda_{21}^x \lambda_{41}^x & \lambda_{21}^x \lambda_{51}^x \\ \lambda_{11}^x \lambda_{31}^x & \lambda_{21}^x \lambda_{31}^x & \lambda_{31}^{2x} + \theta_{33} & \lambda_{31}^x \lambda_{41}^x & \lambda_{31}^x \lambda_{51}^x \\ \lambda_{11}^x \lambda_{41}^x & \lambda_{21}^x \lambda_{41}^x & \lambda_{31}^x \lambda_{41}^x & \lambda_{41}^{2x} + \theta_{44} & \lambda_{41}^x \lambda_{51}^x \\ \lambda_{11}^x \lambda_{51}^x & \lambda_{21}^x \lambda_{51}^x & \lambda_{31}^x \lambda_{51}^x & \lambda_{41}^x \lambda_{51}^x & \lambda_{51}^{2x} + \theta_{55} \end{bmatrix}$$

Así entonces dado que \mathbf{t} es el número de parámetros desconocidos y \mathbf{p} , \mathbf{q} son las dimensiones de \mathbf{y} , \mathbf{x} que satisfacen $\mathbf{t} \leq (\mathbf{p} + \mathbf{q})(\mathbf{p} + \mathbf{q} + 1)/2$ tenemos $\Lambda_{\mathbf{y}} : 4$, $\Lambda_{\mathbf{x}} : 5$, $\Gamma : 1$, $\Phi : 1$, $\Psi : 1$, $\Theta_{\varepsilon} : 4$, $\Theta_{\delta} : 5$, por lo que $2\mathbf{1} \leq (4 + 5)(4 + 5 + 1)/2$, entonces $2\mathbf{1} \leq 45$.

3.3.2. Estimación del modelo

Para la estimación \mathbf{F} utilizamos el método de máxima verosimilitud

$$\mathbf{F}_{\text{ML}} = \log |\Sigma(\theta)| + \text{tr}(\mathbf{S}\Sigma^{-1}(\theta)) - \log |\mathbf{S}| - (\mathbf{p} + \mathbf{q}) \quad (3.6)$$

fijamos valores iniciales para $\lambda_{11}^y = 1$ y $\lambda_{11}^x = 1$.

Capítulo 4

Resultados

Este capítulo contiene los resultados de estimar las ecuaciones de medida y el modelo estructural entre las variables latentes IDEM y el IDH realizado con datos de los diferentes países del mundo, donde las variables relacionadas con el desarrollo humano son: y_1 es esperanza de vida, y_2 años promedio de escolaridad, y_3 años esperados de escolarización, y_4 ingresos per cápita. Y cinco variables relacionadas con la democracia: x_1 procesos electoral y pluralismo, x_2 libertades civiles, x_3 funcionamiento del gobierno, x_4 participación política, y x_5 cultura política. Así entonces la matriz de datos se puede expresar de la siguiente forma:

$$X_{n(p+q)}^* = [Y:X] = \begin{bmatrix} y_{11} & \cdots & y_{1p} & x_{11} & \cdots & x_{1q} \\ y_{21} & \cdots & y_{2p} & x_{21} & \cdots & x_{2q} \\ \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{np} & x_{n1} & \cdots & x_{nq} \end{bmatrix} \text{ con } p = 4, q = 5$$

4.1. Programa para la modelación

El tratamiento de los datos y la modelación de las ecuaciones estructurales se realizó con el Software i) R lenguaje de código abierto que dispone de amplias comunidades de desarrolladores que lo sostienen en constante crecimiento; podemos indicar algunos otros que permiten trabajar con este tipo de modelos, con los que se ha tenido alguna experiencia básica al momento de elegir i) R como son los siguientes ii) AMOS (Analysis of Moment Structures) de SPSS – (Statistical Package for the Social Sciences) AMOS con SPSS son

productos con licencia es bastante utilizado en investigaciones de este tipo de modelaciones por su facilidad de uso al disminuir el nivel de la sintaxis, iii) LISREL – (Linear Structural Relations) es un software para modelos de ecuaciones estructurales es un programa con licencia la forma de trabajar con este programa es a través de sintaxis propia, iv) Stata paquete de software estadístico puede realizar ecuaciones estructurales trabaja con su propia sintaxis es un software con licencia, v) PLS – (Partial Least Squares) permite trabajar con modelo de ecuaciones estructurales trabaja similar a SPSS disminuye la sintaxis bastante utilizado en el área de investigación de mercados, psicología.

En general los programas mencionados manejan la gráfica de senderos que es una forma de simplificar las expresiones matemáticas unas con mejor aceptabilidad que otras por ejemplo el software SPSS es más amigable con los usuarios por lo que tiene bastante demanda está en su versión 26 lo mismo con su extensión AMOS, no sucede ese hecho con el programa LISREL, el resto de los programas en el caso de Stata es cuestión de gusto o también de que tan familiarizado este el usuario con uno u otro software, o su nivel de experiencia recae también en su formación, PLS al igual que SPSS muy similares en cuanto a su uso por la facilidad de manejo que tiene es un software con licencia con mayores restricciones que SPSS como tamaño de muestra número de variables, facilidades en el uso.

El lenguaje i) R tiene un enfoque matemático estadístico, en ese sentido es más fácil de utilizar por profesionales con conocimientos matemáticos o estadístico. Tomando en cuenta que ha sido necesario realizar análisis estadístico, R facilitó en el análisis pues cuenta con varias librerías como es el caso de Lavaan librería para modelo de ecuaciones estructurales, la escritura de código tiene la sencillez de implementación en pocas líneas de código. Eso hace que R sea la elección así también como tener una amplia comunidad para poder soportar diferentes consultas en el proceso de la modelación.

4.2. Normalidad multivariante

La matriz de datos X^* debe seguir una distribución normal multivariante para poder hacer la inferencia del modelo, empero se advierte que los datos no puedan seguir dicha distribución como es común en la mayoría de los modelos en los que se manejan datos empíricos.

Si se cuentan con $[y'_j : x'_j]$ $j = 1, 2, \dots, n$ vectores de observaciones que son vectores aleatorios e idénticamente distribuidos en R^{p^*} , con $p^* = p + q$ se plantea la siguiente

hipótesis:

Ho: $[y'_j : x'_j]$ $j = 1, 2, \dots, n$ sigue una distribución normal multivariante $N^{p^*}(\mu, \Sigma)$

Donde μ y Σ son desconocidos, [6] la prueba de Shapiro - Wilk corresponde a evaluar $W^* = \frac{1}{p^*} \sum_{i=1}^{p^*} W z_i$, donde $W z_i$ es el estadístico Shapiro - Wilk evaluado en la i -ésima coordenada de la observación transformada z_{i1}, \dots, z_{in} con $i = 1, \dots, p^*$. Los resultados son los siguientes:

Shapiro-Wilk normality test: W = 0.80718, p-value = 1.748e-13

Se observó que el valor de probabilidad no es estadísticamente significativo a un nivel de significancia del 0.05, no se cumple la normalidad multivariante de la matriz de datos X^* por lo que los resultados en adelante serán del tipo descriptivo, cabe recalcar que los datos disponibles (165 países del mundo) son tomados de las fuentes PNUD y The Economist, para comparar la magnitud de la información podemos mencionar que el Programa de Naciones Unidas para el Desarrollo esta presente en unos 170 países [13].

4.3. Análisis de los datos

Las variables que hacen al IDEM tienen una escala entre 0 y 10, las variables que hacen al IDH están en diferentes medidas además el ingreso per cápita esta representada en unidades monetarias¹ que por su naturaleza presentan alta variabilidad, es así que a las variables se aplica normalización. En ese sentido y para trabajar con la matriz de varianzas y covarianzas se realizó la estandarización de la matriz de datos.

$$\mathbf{Z}_{n \times p^*} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{bmatrix}$$

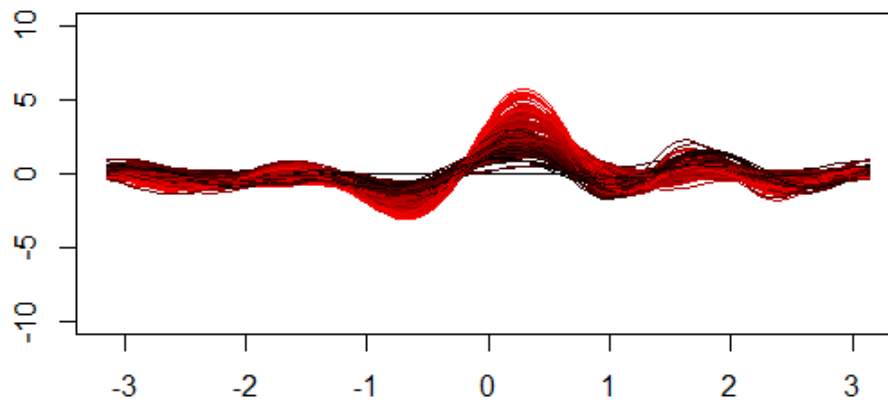
Con $p^*=9$ y $n=165$, en adelante se trabaja con esta matriz empero la notación que le asignaremos sera X_{np^*} .

¹El PIB per cápita que maneja el PNUD es el ingreso per cápita a precios constantes o reales del año 2017, éste ingreso per cápita quita el efecto de la inflación.

4.3.1. Datos atípicos

Para ver los datos atípicos se utilizó la gráfica de Andrews que representa al vector de variables $x'_i = [x_{i1} \ x_{i2} \ \dots \ x_{ip}]$ mediante la siguiente función $f_i(t) = \frac{x_{i1}}{\sqrt{2}} + x_{i2} \sin(t) + x_{i3} \cos(t) + x_{i4} \sin(2t) + x_{i5} \cos(2t) + \dots + \cos(4t)$, con $-\pi \leq t \leq \pi$.

Figura 4.1: Gráfico de Andrews

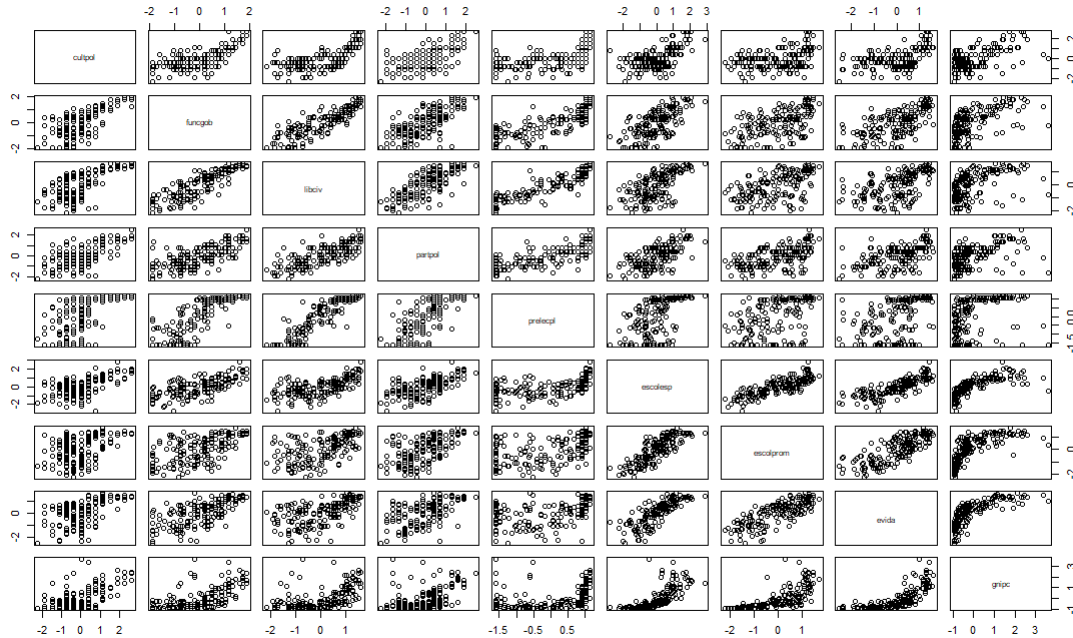


En la gráfica se observó que se produce una sola curva para cada variable de la matriz, donde la influencia de cada variable determina la forma de curva, no se apreciaron datos atípicos presentes puesto que no hay curvas con formas muy diferentes a la generalidad del conjunto de las curvas.

4.3.2. Gráfico de dispersión

Podemos ver en el dispersograma como las relaciones son altas entre los componentes del IDEM: cultpol, funcgob, libciv, partpol, prelecpl, empero estas no se ven asociadas con los componentes del IDH. Lo mismo sucede si vemos del lado del IDH: escolesp, escolprom, evida, gnip. Esto ayuda a tratar el modelo de ecuaciones estructurales creando las variables latentes y luego relacionarlas para ver la causalidad.

Figura 4.2: Dispersograma

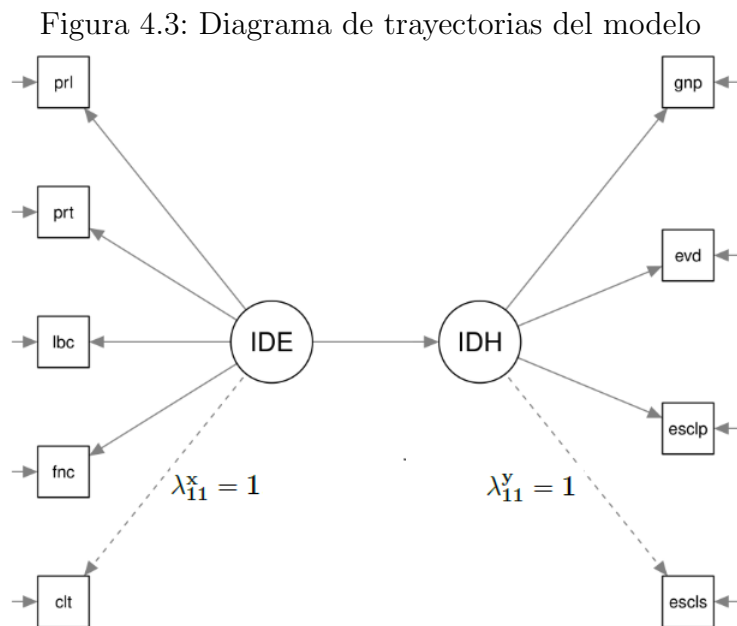


4.4. Resultados del modelo

Se desarrolló cuatro modelos (ver anexo) en función a las variables propuestas, la primera toma en cuenta la matriz estandarizada, la segunda toma en cuenta matriz estandarizada eliminando del modelo a la variable ingreso percápita, la tercera toma en cuenta la matriz estandarizada en la que previamente se aplicó logaritmos a los componentes del IDH, la cuarta toma en cuenta la matriz estandarizada en la que previamente se aplicó logaritmos a los componentes del IDH y se elimino al ingreso percápita. Si bien los resultados para los modelos dos y cuatro indicados son mejores estadísticamente al resto de los modelos no se tomaron en cuenta porque no hay un sustento teórico para poder tomar dicho modelo, es decir no hay una teoría que respalde que el IDH debe ajustarse sin tomar el ingreso percápita, la mejor opción es entonces el modelo que toma en cuenta la matriz estandarizada en la que previamente se aplicó logaritmos.

4.4.1. Diagrama de trayectorias del modelo

El modelo desarrollado presentó la estructura de la figura (4.3) que describe las trayectorias de las variables latentes IDEM ó IDE y IDH hacia las variables de manifiesto². Así mismo también describe la trayectoria de IDEM hacia IDH que representa la relación estructural. Las líneas discontinuas representan los valores iniciales fijados para resolver el modelo $\lambda_{11}^y = 1$ y $\lambda_{11}^x = 1$.



²Donde prl: procesos electoral y pluralismo, lbc: libertades civiles, fnc: funcionamiento del gobierno, prt: participación política, clt: cultura política. Y evd: esperanza de vida, esclp: años promedio de escolaridad, escls: años esperados de escolarización, gnp: ingresos percápita.

4.4.2. Ajuste del modelo

User model versus baseline model:	
Comparative Fit Index (CFI)	0.948
Tuker-Lewis Index (TLI)	0.928
Information Criteria:	
Akaike (AIC)	2774.506
Bayesian (BIC)	2833.519
Root Mean Square Error of Approximation:	
RMSEA	0.137
Standardized Root Mean Square Residual:	
SRMR	0.054

Cuadro 4.1: Resultado de las ecuaciones de medida

Se evaluó el ajuste del modelo con una combinación de índices siguiendo el criterio de Nuñez et al. (2005) [9]. La condición de la Raíz cuadrada del error medio cuadrático $RMSEA < 0.08$ no se cumple por lo que indica que el modelo basado en la muestra no representa a la población. Sin embargo para la condición de la normalización de la raíz cuadrada media residual $SRMR \leq 0.06$, se obtuvo un valor menor a 0.06 que minimiza el problema derivado del tamaño de la muestra indicando así un buen ajuste. Sobre los índices comparativo de ajuste por grados de libertad del modelo propuesto y el modelo nulo³, se tiene el índice de ajuste comparativo que satisface $CFI > 0.9$ entonces se considera mínimamente como aceptable el modelo; según el índice de Tucker-Lewis se evidenció que este índice cumple la condición $TLI > 0.9$ por lo que el modelo se considera aceptable.

El criterio de información de Akaike permite elegir en comparación con otros modelos con diferente número de constructos indicar cual es el mejor [8], la condición es elegir el modelo que tenga el menor valor, el modelo propuesto presenta un valor de 2774.506 menor de las diferentes alternativas antes indicadas (ver anexos), de la misma manera el criterio de información bayesiano presenta un valor 2833.519 menor que alternativas realizadas.

³El modelo nulo es aquel que tiene de ausencia de relación entre las variables.

4.4.3. Ecuaciones de medida

Observado el cuadro (4.2) se evidenció primero que para el caso de la ecuación de medida IDH las variables escolprom, evida, pnipc, son estadísticamente significativas a un 0.05 de nivel de significancia las variables propuestas aportan a la construcción del IDH. De forma similar se observó que para la segunda ecuación de medida se evidenció que las variables funcgob, livciv, partpol, prelecpl son estadísticamente significativo al 0.05, así entonces se demuestra la construcción de las ecuaciones de medida IDH y IDEM también llamado variables latentes.

	Estimate	Stdr.Err	z-value	$P(> z)$
IDH= $\tilde{}$				
escolsp	1			
escolprom	1.023	0.060	17.195	0.000
evida	1.000	0.061	16.353	0.000
gnipc	1.063	0.057	18.716	0.000
IDEM= $\tilde{}$				
cultpol	1.000			
funcgob	1.299	0.123	10.563	0.000
libciv	1.419	0.125	11.377	0.000
partpol	1.187	0.122	9.744	0.000
prelecpl	1.351	0.124	10.938	0.000

Cuadro 4.2: Resultado de las ecuaciones de medida

Las variables con mayor influencia sobre el IDH son escolprom y gnipc por último evida, para el caso del IDEM con mayor fuerza sobre la variable latente son libciv, prelecpl y con menor fuerza la influencia funcgob y partpol.

4.4.4. Modelo Estructural

Se observó el cuadro (4.3) donde la relación de IDEM a IDH es positiva y acorde a lo que se esperaba es un coeficiente cercano a 1, lo que quiere decir que una mejora en la democracia mejora en un punto el índice de desarrollo humano la dirección es adecuada acorde con lo planteado en los objetivos, con el coeficiente significativo al 0.05.

	Estimate	Stdr.Err	z-value	$P(> z)$
IDH \sim				
IDEM	0.854	0.113	7.530	0.000
Variances:				
escolsp	0.221	0.030	7.363	0.000
escolprom	0.185	0.027	6.903	0.000
evida	0.221	0.030	7.362	0.000
gnipc	0.121	0.022	5.523	0.000
cultpol	0.531	0.060	8.793	0.000
funcgob	0.214	0.028	7.766	0.000
libciv	0.062	0.016	3.901	0.000
partpol	0.342	0.041	8.429	0.000
prelecpl	0.149	0.022	6.903	0.000
IDH	0.436	0.064	6.831	0.000
IDEM	0.463	0.094	4.938	0.000

Cuadro 4.3: Resultados del modelo estructural

Capítulo 5

Conclusiones y recomendaciones

En adelante se detallan las conclusiones y recomendaciones luego del desarrollo de la investigación realizada.

5.1. Conclusiones

- Como uno de los objetivos específicos fué determinar un indicador sintético para el IDH, se ha demostrado que las variables esperanza de vida al nacer, años promedio de escolaridad, años esperados de escolarización, ingreso per cápita hacen al índice de desarrollo humano, eso se ha corroborado con los resultados adecuados de las ecuaciones de medida.
- Como uno de los objetivos específicos fué determinar un indicador sintético para el IDEM, se ha demostrado que las variables proceso electoral y pluralismo, libertades civiles, funcionamiento del gobierno, participación política, cultura política hacen al índice de democracia, éllo corroborado con los resultados adecuados de las ecuaciones de medida.
- Las variables tomadas en cuenta que hacen al IDEM e IDH no cumplen el supuesto de normalidad multivariante, este fué un inconveniente.
- La hipótesis planteada es ver la relación causal del IDEM al IDH, empero se ha evidenciado esta relación de manera descriptiva, concluyendo que existe una relación causal positiva alta, del índice de democracia al índice de desarrollo humano para

los países de los cuales existía la información (165 países del mundo), sustentado por los resultados de los ajustes de los modelos que fueron aceptables más no óptimos.

- Así mismo lo que más contribuyó a decidir el cumplimiento de la hipótesis fue el demostrar los dos objetivos principales, construir las ecuaciones de medida IDH y IDEM, también el resultado de un coeficiente alto en la relación estructural, así como las significancias estadísticas de los coeficientes del modelo estructural y de las ecuaciones de medida.
- Organizar tres alternativas adicionales del modelo expuestas en el anexo, ayudo a tener un mejor panorama para decidir que el modelo propuesto es el más adecuado.
- Considerar la comparación de un modelo sin el ingreso per cápita con mejores resultados ayudo a decidir sobre el modelo, desechando la propuesta de eliminación del ingreso per cápita porque no existe la teoría que respalde esta situación, así se acabo de enfocar al modelo inicialmente propuesto, siendo el mejor y con resultados aceptables.
- Si bien trabajar con información actualizada o tenerla con un rezago es óptimo, al respecto se sostiene que la información al 2019 utilizada, es adecuada es la última disponible en PNUD, y tomar información al 2020 sería necesario tomar en cuenta el contexto de crisis de salud global.
- Lo gratificante de la investigación, es el conocimiento de la teoría que hace a estos modelos de ecuaciones estructurales, en el cual existen muchos índices de ajuste y que puedes tomar algunos para una forma de trabajar y otros para otra forma; donde la naturaleza de los datos es importante, para clasificar si son categóricos o continuos, y verificar que los datos estén en las mismas escalas.
- El aporte de la tesis, es en el campo de aplicación por el hecho de analizarlo de una manera distinta, de abordar el problema de forma estructural con la relación de variables latentes Democracia y Desarrollo Humano, esta relación estructural es la modelación donde la construcción de variables latentes es a través de variables observables, y la relación de variables latentes desde este enfoque estructural permite entender el fenómeno con esta modelación matemática.
- El aporte en el campo de aplicación amplia el material científico sobre este tema, para que otros estudios puedan utilizarla, replicarla, confirmarla o refutarla. Es básicamente la proposición de la modelación mediante ecuaciones estructurales. Esto

en respuesta a que no hay estudios que analicen este problema con esta modelación propuesta en esta tesis, no hay estudios que tomen en cuenta la modelación de ecuaciones estructurales para abordar este tema en particular.

5.2. Recomendaciones

- Aún sin el cumplimiento de la normalidad multivariante estos modelos pueden ser representados como resultados descriptivos.
- Como aplicaciones posibles a la vida cotidiana sugiero, hacer encuestas con variables que tengan relación entre si para crear constructos y luego ver las relaciones, para ello se debe tomar en cuenta la experiencia de un experto del área sobre el que se quiera investigar ó revisar sobre tal especialidad para tener conocimiento y poder fundamentar las relaciones. Un ejemplo para el área financiera sería crear encuestas sobre satisfacción del cliente ó crear un indicador de satisfacción luego crear encuestas de variables relacionadas con la antigüedad y/o lealtad del cliente y luego estas dos variables latentes relacionarlas para ver la causalidad.
- De todo lo que se trabajó considero que el modelo de ecuaciones estructurales se puede expandir, esto a más factores latentes depende de la teoría que respalde dichas relaciones o propuestas bien sustentadas.
- Los méritos de trabajar en este tipo de investigaciones, es la aplicación matemática a la investigación económica.
- Para un análisis más minucioso se recomienda utilizar otras fuentes sobre el índice de democracia por ejemplo Freedom house, Polity IV, The Database of Political Institutions.
- Se asume a la democracia como exógena, al tomar el ingreso per cápita y educación como componentes del IDH; existen trabajos que podrían entender a estos factores como determinantes de la democracia, Barro (1999) determinantes of democracy [19]. Es decir democracia puede ser considerada como endógena, por lo que bajo ese criterio sería adecuado utilizar los SEM no recursivos.
- Sería bueno controlar en la regresión estructural algunas características, por ejemplo las diferencias regionales, desigualdad y/o urbanización.

Anexos

Anexos A

Resultados del modelo realizado y las alternativas al modelo.

Figura A.1: Modelo estandarizado

lavaan 0.6-9 ended normally after 32 iterations		Parameter Estimates:				
Estimator	ML	Standard errors		Standard		
Optimization method	NLMINB	Information		Expected		
Number of model parameters	19	Information saturated (h1) model		Structured		
Number of observations	165					
Model Test User Model:		Latent Variables:				
Test statistic	123.448	IDH =~	Estimate	Std.Err	z-value	P(> z)
Degrees of freedom	26	escolesp	1.000			
P-value (Chi-square)	0.000	escolprom	0.963	0.055	17.523	0.000
Model Test Baseline Model:		evida	0.964	0.055	17.547	0.000
Test statistic	1500.251	gnipc	0.878	0.061	14.294	0.000
Degrees of freedom	36	IDEM =~				
P-value	0.000	cultpol	1.000			
User Model versus Baseline Model:		funcgob	1.294	0.122	10.619	0.000
Comparative Fit Index (CFI)	0.933	libciv	1.414	0.124	11.446	0.000
Tucker-Lewis Index (TLI)	0.908	partpol	1.184	0.121	9.792	0.000
		prelecpl	1.345	0.122	10.988	0.000
Loglikelihood and Information Criteria:		Regressions:				
Loglikelihood user model (H0)	-1414.209	IDH ~	Estimate	Std.Err	z-value	P(> z)
Loglikelihood unrestricted model (H1)	-1352.485	IDEM	0.925	0.116	7.956	0.000
Akaike (AIC)	2866.417	Variances:				
Bayesian (BIC)	2925.430	.escolesp	Estimate	Std.Err	z-value	P(> z)
Sample-size adjusted Bayesian (BIC)	2865.276	.escolprom	0.156	0.027	5.727	0.000
Root Mean Square Error of Approximation:		.evida	0.216	0.032	6.830	0.000
RMSEA	0.151	.gnipc	0.215	0.032	6.817	0.000
90 Percent confidence interval - lower	0.125	.cultpol	0.348	0.044	7.950	0.000
90 Percent confidence interval - upper	0.178	.funcgob	0.528	0.060	8.789	0.000
P-value RMSEA <= 0.05	0.000	.libciv	0.213	0.027	7.767	0.000
Standardized Root Mean Square Residual:		.partpol	0.062	0.016	3.912	0.000
SRMR	0.061	.prelecpl	0.341	0.040	8.428	0.000
Parameter Estimates:		.IDH	0.151	0.022	6.947	0.000
Standard errors	Standard	IDEM	0.439	0.062	7.063	0.000
			0.466	0.094	4.960	0.000

Figura A.2: Modelo estandarizado y eliminando ingreso

lavaan 0.6-9 ended normally after 31 iterations		Parameter Estimates:			
Estimator	ML	Standard errors		Standard	
Optimization method	NLMINB	Information		Expected	
Number of model parameters	17	Information saturated (h1) model		Structured	
Number of observations	165				
Model Test User Model:		Latent Variables:			
Test statistic	85.574	IDH =~	Estimate	Std.Err	z-value P(> z)
Degrees of freedom	19	escolesp	1.000		
P-value (Chi-square)	0.000	escolprom	0.945	0.055	17.318 0.000
		evida	0.934	0.055	16.909 0.000
Model Test Baseline Model:		IDEM =~			
Test statistic	1310.176	cultpol	1.000		
Degrees of freedom	28	funcgob	1.296	0.122	10.585 0.000
P-value	0.000	libciv	1.417	0.124	11.411 0.000
		partpol	1.186	0.121	9.776 0.000
User Model versus Baseline Model:		prelecpl	1.348	0.123	10.964 0.000
Comparative Fit Index (CFI)	0.948				
Tucker-Lewis Index (TLI)	0.923	Regressions:			
Loglikelihood and Information Criteria:		IDH ~	Estimate	Std.Err	z-value P(> z)
Loglikelihood user model (H0)	-1256.686	IDEM	0.949	0.118	8.042 0.000
Loglikelihood unrestricted model (H1)	-1213.899				
Akaike (AIC)	2547.372	Variances:			
Bayesian (BIC)	2600.173	.escolesp	Estimate	Std.Err	z-value P(> z)
Sample-size adjusted Bayesian (BIC)	2546.351	.escolprom	0.131	0.029	4.471 0.000
		.evida	0.224	0.034	6.567 0.000
Root Mean Square Error of Approximation:		.cultpol	0.241	0.035	6.826 0.000
RMSEA	0.146	.funcgob	0.530	0.060	8.792 0.000
90 Percent confidence interval - lower	0.115	.libciv	0.214	0.028	7.778 0.000
90 Percent confidence interval - upper	0.178	.partpol	0.062	0.016	3.924 0.000
P-value RMSEA <= 0.05	0.000	.prelecpl	0.340	0.040	8.427 0.000
Standardized Root Mean Square Residual:		.IDH	0.150	0.022	6.928 0.000
SRMR	0.051	IDEM	0.445	0.063	7.080 0.000
			0.464	0.094	4.948 0.000

Figura A.3: Modelo estandarizado aplicado ln a los componentes del IDH

lavaan 0.6-9 ended normally after 34 iterations		Parameter Estimates:			
Estimator	ML	Standard errors		Standard	
Optimization method	NLMINB	Information		Expected	
Number of model parameters	19	Information saturated (h1) model		Structured	
Number of observations	165				
Model Test User Model:		Latent Variables:			
Test statistic	106.377	IDH =~	Estimate	Std.Err	z-value P(> z)
Degrees of freedom	26	escolesp	1.000		
P-value (Chi-square)	0.000	escolprom	1.023	0.060	17.195 0.000
		evida	1.000	0.061	16.353 0.000
		gnipc	1.063	0.057	18.716 0.000
Model Test Baseline Model:		IDEM =~			
Test statistic	1575.091	cultpol	1.000		
Degrees of freedom	36	funcgob	1.299	0.123	10.563 0.000
P-value	0.000	libciv	1.419	0.125	11.377 0.000
		partpol	1.187	0.122	9.744 0.000
		prelecpl	1.351	0.124	10.938 0.000
User Model versus Baseline Model:		Regressions:			
Comparative Fit Index (CFI)	0.948	IDH ~	Estimate	Std.Err	z-value P(> z)
Tucker-Lewis Index (TLI)	0.928	IDEM	0.854	0.113	7.530 0.000
Loglikelihood and Information Criteria:		Variances:			
Loglikelihood user model (H0)	-1368.253	.escolesp	Estimate	Std.Err	z-value P(> z)
Loglikelihood unrestricted model (H1)	-1315.064	.escolprom	0.221	0.030	7.363 0.000
		.evida	0.185	0.027	6.903 0.000
Akaike (AIC)	2774.506	.gnipc	0.221	0.030	7.362 0.000
Bayesian (BIC)	2833.519	.cultpol	0.121	0.022	5.523 0.000
Sample-size adjusted Bayesian (BIC)	2773.365	.funcgob	0.531	0.060	8.793 0.000
Root Mean Square Error of Approximation:		.libciv	0.214	0.028	7.766 0.000
RMSEA	0.137	.partpol	0.062	0.016	3.901 0.000
90 Percent confidence interval - lower	0.110	.prelecpl	0.342	0.041	8.429 0.000
90 Percent confidence interval - upper	0.164	.IDH	0.149	0.022	6.903 0.000
P-value RMSEA <= 0.05	0.000	IDEM	0.436	0.064	6.831 0.000
Standardized Root Mean Square Residual:			0.463	0.094	4.938 0.000
SRMR	0.054				

Figura A.4: Modelo estandarizado aplicado ln a los componentes del IDH, y eliminado el ingreso

lavaan 0.6-9 ended normally after 30 iterations		Parameter Estimates:			
Estimator	ML	Standard errors		Standard	
Optimization method	NLMINB	Information		Expected	
Number of model parameters	17	Information saturated (h1) model		Structured	
Number of observations	165				
Model Test User Model:		Latent Variables:			
Test statistic	81.573	IDH =~	Estimate	Std.Err	z-value P(> z)
Degrees of freedom	19	escolesp	1.000		
P-value (Chi-square)	0.000	escolprom	0.975	0.058	16.787 0.000
Model Test Baseline Model:		evida	0.936	0.060	15.540 0.000
Test statistic	1279.007	IDEM =~			
Degrees of freedom	28	cultpol	1.000		
P-value	0.000	funcgob	1.298	0.123	10.540 0.000
User Model versus Baseline Model:		libciv	1.421	0.125	11.364 0.000
Comparative Fit Index (CFI)	0.950	partpol	1.189	0.122	9.742 0.000
Tucker-Lewis Index (TLI)	0.926	prelecpl	1.353	0.124	10.928 0.000
Loglikelihood and Information Criteria:		Regressions:			
Loglikelihood user model (H0)	-1270.269	IDH ~	Estimate	Std.Err	z-value P(> z)
Loglikelihood unrestricted model (H1)	-1229.483	IDEM	0.912	0.117	7.766 0.000
Akaike (AIC)	2574.539	Variances:			
Bayesian (BIC)	2627.340	.escolesp	Estimate	Std.Err	z-value P(> z)
Sample-size adjusted Bayesian (BIC)	2573.518	.escolprom	0.164	0.032	5.052 0.000
Root Mean Square Error of Approximation:		.evida	0.205	0.034	5.951 0.000
RMSEA	0.141	.cultpol	0.267	0.038	6.947 0.000
90 Percent confidence interval - lower	0.111	.funcgob	0.532	0.060	8.795 0.000
90 Percent confidence interval - upper	0.173	.libciv	0.216	0.028	7.789 0.000
P-value RMSEA <= 0.05	0.000	.partpol	0.061	0.016	3.885 0.000
Standardized Root Mean Square Residual:		.prelecpl	0.341	0.040	8.429 0.000
SRMR	0.049	.IDH	0.149	0.022	6.901 0.000
		IDEM	0.446	0.065	6.877 0.000
			0.462	0.094	4.933 0.000

Figura A.5: Código utilizado para estimar el modelo

```

##### Cargar la data #####
library(readxl)
HBAT<- read_excel("ddh2019.xlsx")
attach(HBAT)
X<-cbind(cultpol, funcgob, libciv, partpol, prelecl, escolsp, escolprom,
        evida, gnipc)
##### Modelo con logaritmo #####
gnipc=log(gnipc)
escolsp=log(escolsp)
escolprom=escolprom(escolprom)
evida=log(evida)
Y<-cbind(cultpol, funcgob, libciv, partpol, prelecl, escolsp, escolprom,
        evida, gnipc)
Y<-scale(Y,center=T,scale=T)
#####
X<-scale(X,center=T,scale=T)
#detach(HBAT)
#head(HBAT)
head(X)
n <- dim(X)[1];n
#####
##### plot1 #####
matplot(X,type="l")
#####
##### plot2 #####
matplot(X)
#####
##### Correlaciones #####
pairs(X)
#####
##### Normalidad2 #####
library(mvnormtest)
mshapiro.test(t(X))
tabla<-X[,-c(9)]#ELIMINAMOS INGRESO PERCAPITA
mshapiro.test(t(tabla))#Prueba de normalidad mult. de Shapiro
#####
##### Andrews #####
library(andrews)
andrews(X, clr=3, type=2)
#####
##### SEM1 #####
#-----
#---- UTILIZACION DE SEM DE LAVAAN ----
#-----
library(lavaan)
library(semPlot)
library(Rcpp)
#-----
#####
#-----
# MODELO
#-----
model <- ' # latent variable definitions
IDH =~ escolsp+ escolprom+ evida+ gnipc
IDEM =~ cultpol+ funcgob+ libciv+ partpol+ prelecl
# regressions
IDH ~ IDEM'
fit <- sem(model, data=X)
summary(fit, fit.measures=TRUE)
#####
##### Phat grafico #####
semPaths(fit, style="lisrel", rotation=2)
#####
##### sem2 #####
library(lavaan)
library(semPlot)
#-----
# MODELO
#-----
model <- ' # latent variable definitions
IDH =~ escolsp+ escolprom+ evida
IDEM =~ cultpol+ funcgob+ libciv+ partpol+ prelecl
# regressions
IDH ~ IDEM'
fit <- sem(model, data=X)
summary(fit, fit.measures=TRUE)
#####
##### phat2 #####
semPaths(fit, style="lisrel", rotation=2)
#####
##### Modelo en logaritmo 1 #####
model <- ' # latent variable definitions
IDH =~ escolsp+ escolprom+ evida+ gnipc
IDEM =~ cultpol+ funcgob+ libciv+ partpol+ prelecl
# regressions
IDH ~ IDEM'
fit <- sem(model, data=Y)
summary(fit, fit.measures=TRUE)
#####
##### Modelo en logaritmo 2 #####
model <- ' # latent variable definitions
IDH =~ escolsp+ escolprom+ evida
IDEM =~ cultpol+ funcgob+ libciv+ partpol+ prelecl
# regressions
IDH ~ IDEM'
fit <- sem(model, data=Y)
summary(fit, fit.measures=TRUE)
#####

```

Bibliografía

- [1] Temkin Benjamin y Del Tronco José. Desarrollo humano, bienestar subjetivo y democracia: confirmaciones, sorpresas e interrogantes. *Instituto de Investigaciones Sociales*, 68, 2006.
- [2] Altman David, Luna Juan Pablo, Piñeiro Rafael, y Toro Sergio. Partidos y sistemas de partidos en américa latina: Aproximaciones desde la encuesta a expertos 2009. *Revista de ciencia política (Santiago)*, 29(3):775–798, 2009.
- [3] The Economist. *Indice de democracia 2019*. The Economist Intelligence Unit, 2019.
- [4] Hyden Goran, Mease Kenneth, Foresti Marta, y Fritz Verena. Governance assessments for local stakeholders: What the world governance assessment offers. *Overseas Development Institute*, 2008.
- [5] Schiattino Irene y Silva Claudio. Representación gráfica de información multivariante. aplicación al sistema de salud de chile 2010. *Rev Chil Salud Pública 2013*, 17 (2):117–123, 2013.
- [6] Porras Cerron Jaime Carlos. Comparison test multivariate normal. *Anales Científicos*, 77(2):141–146, 2016.
- [7] Gerring John, Thacker Strom C., y Alfaro Rodrigo. Democracy and human development: confirmaciones, sorpresas e interrogantes. *The Journal of Politics*, 74(1), 2012.
- [8] Hair Joseph F., Anderson Rolph E., Tatham Ronald L., y Black William C. *Análisis multivariante*, tomo 5. Prentice Hall Madrid, 1999.
- [9] Núñez Alonso Juan Luis, José Martín-Albo Lucas, y José Gregorio Navarro Izquierdo. Software, instrumentación y metodología. *Psicothema*, 17(2), 2005.

-
- [10] Bollen Kenneth. *Estructural equations with latent variables*. Wiley, 1989.
- [11] Sánchez Marco. *Políticas públicas para el desarrollo humano - Naciones Unidas*, 2011.
- [12] Berggren Olle. *Democracy and Human Development: A Cross-National Analysis*. Lund University, 2012.
- [13] PNUD. Objetivos del desarrollo sostenible. *Programa de Naciones Unidas para el Desarrollo*.
- [14] PNUD. *Informe sobre el Desarrollo Humano 2002 - Profundizando la democracia en un mundo fragmentado*”,. Programa de Naciones Unidas para el Desarrollo, 2002.
- [15] PNUD. *Informe sobre el Desarrollo Humano 2010*. Programa de Naciones Unidas para el Desarrollo, 2010.
- [16] PNUD. *Informe sobre el Desarrollo Humano 2011*. Programa de Naciones Unidas para el Desarrollo, 2011.
- [17] PNUD. *Informe sobre el Desarrollo Humano 2016*. Programa de Naciones Unidas para el Desarrollo, 2016.
- [18] PNUD. *Informe sobre el Desarrollo Humano 2020*. Programa de Naciones Unidas para el Desarrollo, 2020.
- [19] Barro Robert J. Determinants of democracy. *Journal of Political Economy*, 107(s6):s158– s183, 1999.
- [20] Lipset Seymour Martin. The social requisites of democracy. *American Sociological Association*, 59:1–22, 1994.
- [21] Lee Sik-Yum. *Structural equation modeling: A Bayesian approach*, tomo 711. John Wiley & Sons, 2007.