



Nombre del estudiante		Edgardo de los Ángeles Gutiérrez Vega
Título de la tesis		Comparación de modelos de ecuaciones estructurales bayesianos con muestras pequeñas de comunidades microbianas composicionales.
Introducción	Justificación/importancia del tema	<p>Los análisis estadísticos aplicados a datos composicionales requieren una cuidadosa consideración de sus características con el fin de evitar resultados y conclusiones engañosos. Uno de los peligros más notables al analizar datos composicionales son las correlaciones espurias, hecho que fue detectado por Karl Pearson en 1897 (Pawlowsky-Glhan y Buccianti, 2011). La existencia de correlaciones espurias entre las partes de una composición, lleva a que dentro de las dificultades más relevantes se encuentre la interpretación de las covarianzas y coeficientes de correlación, que son elementos esenciales en la estimación de modelos de ecuaciones estructurales (SEM) (Mateu <i>et al.</i>, 2003).</p> <p>La definición de datos composicionales ha evolucionado progresivamente de vectores con componentes positivos, cuya suma es una constante, a definiciones generales más recientes basadas en equivalencia de clases. En este último caso, una composición es definida por un vector de <math>D</math> componentes estrictamente positivos, que al sumarse no necesariamente suman una constante. De esta manera, el concepto clave es el carácter relativo de la información de interés, de manera que la multiplicación de la composición por cualquier constante no cambia la información contenida en las razones entre componentes (Egozcue <i>et al.</i>, 2018).</p> <p>Las apariciones frecuentes de datos composicionales en los estudios realizados en diversas disciplinas, como las ciencias de la tierra, biología, química, ciencias ambientales, economía, medicina, sociología e ingeniería, entre otras, ponen de manifiesto el interés por disponer de herramientas adecuadas para su análisis, ya que si no se dispone de técnicas adecuadas para el análisis de datos composicionales es posible que los investigadores lleguen a conclusiones incorrectas (Mateu <i>et al.</i>, 2003).</p> <p>Específicamente, en los campos de la microbiología y epidemiología, durante la última década, se ha generado un incremento en los estudios sobre los microbiomas presentes en distintas zonas del cuerpo humano (boca, estómago e intestinos, entre otros) y su relación con diversas condiciones de salud (JVCI, 2018).</p> <p>De acuerdo con Weiss <i>et al.</i> (2017), en los estudios de microbiomas, para poder tomar en cuenta los distintos tamaños de las librerías de secuenciación 16S de ARN ribosomal (ARNr), que varían en amplios rangos de magnitud e incluso suelen tener muchos ceros, las lecturas de secuenciación (entendidos como datos de conteo) son usualmente normalizadas a proporciones.</p> <p>Estudios como el de Sohn y Li (2017), que discuten sobre la naturaleza composicional y multi-dimensional de los datos de comunidades microbianas, dejan en evidencia que los estudios sobre microbiomas y su influencia en condiciones de salud, pueden abordarse mediante el establecimiento de modelos de ecuaciones estructurales con el fin de generar evidencia de relaciones causales entre constructos complejos que representan el microbioma del individuo, el ambiente y la condición de salud, cuyos indicadores son de tipo composicional, dicotómico o continuo.</p> <p>Sin embargo, los modelos de ecuaciones estructurales tienen sus propias restricciones al considerar tamaños de muestra cuando se utilizan variables cuantitativas continuas</p>



multinormales como indicadores del modelo de medición. Wolf *et al.* (2013) realizaron estudios de simulación y determinaron que para este tipo de modelos no se puede establecer un requisito del tamaño de muestra que satisfaga los ajustes para cualquier modelo, de manera que en sus resultados los tamaños de muestra variaron desde 30 a 460 observaciones, tomando en cuenta la complejidad del modelo y las cargas factoriales que se deseaban obtener. De esta forma, la determinación del tamaño de muestra recomendado, cuando se trabaja con datos composicionales, implica todo un reto al estimar modelos de ecuaciones estructurales, debido a que, aunado a las limitaciones encontradas por Wolf *et al.* (2013), la normalización de los datos de comunidades microbianas introduce la limitación de que el resultado de sumar las proporciones es la unidad, lo cual transforma el espacio Euclidiano  $k$ -dimensional ( $\mathbb{R}^k$ ) en un espacio simplex  $k-1$  ( $S^{k-1}$ ) y, por lo tanto, los modelos para datos no-limitados son inapropiados para los datos composicionales.

Con respecto a los tamaños de muestra, Hackshaw (2008) señala fortalezas y limitaciones de los estudios a pequeña escala que en general cuentan con un tamaño de muestra pequeño. Según este mismo autor, dentro de los estudios a pequeña escala se pueden tener estudios de investigación clínica como estudios observacionales, ensayos clínicos, ensayos basados en análisis de laboratorios, entre otros. Sin embargo, Hackshaw señala que para definir qué tan pequeño es “pequeño”, se debe tomar en cuenta el objetivo principal del estudio, por ejemplo, al comparar características entre 2 o más grupos de sujetos, el tamaño del estudio dependerá del tamaño de efecto esperado, teniendo en cuenta que mientras menor sea el tamaño del efecto, más individuos debe incluir el estudio.

Hackshaw (2008) también señala que los estudios a pequeña escala permiten que la pregunta de investigación pueda ser respondida en un espacio de tiempo relativamente corto, por lo que son excelentes para probar nuevas hipótesis de investigación antes de realizar un estudio a gran escala. Los estudios a pequeña escala también evitan el gasto de demasiados recursos, como investigadores, tiempo y costos financieros; al determinar asociaciones entre un factor y un desorden de salud cuando no hay ningún efecto. Sin embargo, si la asociación se encuentra, es importante dejar en claro dentro de las conclusiones que estas se generaron a partir de un estudio a pequeña escala y que un estudio confirmatorio a mayor escala sería necesario.

El Centro de Investigaciones en Biología Celular y Molecular (CIBCM) de la Universidad de Costa Rica (UCR), realiza estudios a pequeña escala, justamente para dar respuesta a preguntas de investigación a corto plazo y optimizando los recursos económicos disponibles. Algunos de estos estudios están relacionados con la descripción de comunidades microbianas y su implicación en condiciones de salud como la enfermedad periodontal o el cambio de peso en infantes. Los datos de comunidades microbianas son multidimensionales, por lo que presentan gran cantidad de indicadores, lo que aumentaría los requerimientos de muestra al necesitar la formulación de un modelo más complejo, ya que se necesitaría estimar parámetros para cada uno de ellos. Sin embargo, en concordancia con la realización de estudios a pequeña escala, la muestra suele ser relativamente pequeña. Por lo tanto, para mitigar el impacto de esta complicación se recurre a la parcelación, que consiste en el cálculo de sumas o promedios de un conjunto de indicadores para formar un nuevo indicador que representaría al conjunto en el modelo (Orcan, 2013).



		<p>Por otro lado, Muthén y Asparowhov (2010), señalan que uno de los puntos que motivan el uso de inferencia Bayesiana al estimar modelos de ecuaciones estructurales, es que es posible obtener un mejor rendimiento para muestras pequeñas. Además, proveen mayor flexibilidad en los supuestos y permiten incluir información previa disponible sobre los parámetros para obtener estimaciones más precisas. Por este motivo, el enfoque Bayesiano, parece ser prometedor al estimar modelos de ecuaciones estructurales con las limitaciones muestrales señaladas, como en los estudios a pequeña escala, con datos composicionales de comunidades microbianas realizados en el CIBCM.</p> <p>En resumen, algunos estudios a pequeña escala con datos composicionales de comunidades microbianas realizados en el CIBCM buscan proveer evidencia de asociaciones de comunidades microbianas con condiciones de salud, optimizando recursos como tiempo y costo. Dados los aspectos, tanto de los estudios a pequeña escala descritos por Hackshaw (2008), como de los datos composicionales detallados en Egozcue <i>et al.</i> (2018), la estimación precisa de parámetros e identificación de modelos representan retos estadísticos-matemáticos, los cuales podrían ser abordados mediante los modelos de ecuaciones estructurales Bayesianas (BSEM).</p> <p>Finalmente, los estudios a pequeña escala con datos composicionales de comunidades microbianas y su implicación en condiciones de salud, son esenciales para promover la producción de conocimiento sobre el impacto de los microbiomas sobre condiciones de salud en Costa Rica y, por lo tanto, corresponde a un trabajo interdisciplinario entre los profesionales en estadística y los científicos expertos en comunidades microbianas, contribuir a la conciliación entre los recursos existentes y los datos obtenidos, de manera que se pueda tener en consideración, el impacto del tamaño de muestra al diseñar este tipo de estudios.</p>
	Contribución de la tesis a la Estadística como disciplina	La contribución principal de la tesis a la disciplina es aportar evidencia, mediante un estudio de simulación, acerca del efecto del tamaño de muestra en la precisión de las estimaciones de los parámetros en modelos de ecuaciones estructurales Bayesianos. Específicamente, en BSEMs que incluyen indicadores de tipo composicional y a los cuales se les aplica parcelación.
Objetivos	Objetivo general	Evaluar el efecto del tamaño de la muestra en la recuperación de los parámetros de los modelos de ecuaciones estructurales Bayesianos con datos composicionales.
	Objetivos específicos	<ol style="list-style-type: none"> <li>1. Comparar el efecto del tamaño de muestra en la precisión de la estimación de los parámetros de modelos de ecuaciones estructurales Bayesianos con indicadores parcelados de datos composicionales.</li> <li>2. Estimar un modelo de ecuaciones estructurales Bayesiano para datos recolectados en el estudio “Descripción de las comunidades microbianas del tracto gastrointestinal (GI) y su implicación en el peso de una muestra de niños costarricenses menores de un año de edad (801-B7-612)” del CIBCM.</li> </ol>
Metodología	Referentes o elementos teóricos que va a utilizar	<p>Dunson <i>et al.</i> (2005) presentan con gran detalle los elementos teóricos y prácticos que deben tomarse en cuenta para la ejecución de un análisis de ecuaciones estructurales utilizando el enfoque bayesiano (BSEM). En general, respecto a los modelos de ecuaciones estructurales clásicos, estos autores plantean las siguientes ecuaciones para el modelo de medición:</p> $y_i = v_y + \Lambda_y \eta_i + \delta_i^y$ $x_i = v_x + \Lambda_x \xi_i + \delta_i^x$



Donde:

$y$  = vector de  $p$  indicadores de constructos endógenos.

$x$  = vector de  $q$  indicadores de constructos exógenos.

$v$  = parámetros de intercepto modelo medición.

$\Lambda$  = matriz de cargas factoriales.

$\eta$  = vector de  $m$  variables latentes endógenas.

$\xi_i$  = vector de  $n$  variables latentes exógenas, con distribución:

$$\xi_i \sim N_n(\mu_\xi, \Omega_\xi)$$

$\delta$  = vectores de residuales del modelo de medición, con distribuciones:

$$\delta_i^y \sim N_p(0, \Sigma_y)$$

$$\delta_i^x \sim N_q(0, \Sigma_x)$$

Por otro lado, se plantea la ecuación del modelo estructural de la siguiente forma:

$$\eta_i = \alpha + B\eta_i + \Gamma\xi_i + \zeta_i$$

En el cual:

$\alpha$  = vector de interceptos estructurales.

$B$  = matriz que describe relaciones entre constructos.

$\Gamma$  = matriz que cuantifica la influencia de  $\xi$  en  $\eta$ .

$\zeta$  = vector de residuales del modelo estructural, con distribución:

$$\zeta_i \sim N_m(0, \Omega_\zeta)$$

Tomando estos planteamientos en cuenta, la distribución posterior conjunta del SEM Bayesiano tiene la siguiente estructura:

$$\pi(\theta, \xi, \eta | y, x) = \frac{L(y, x, \eta, \xi; \theta) \pi(\theta)}{\int L(y, x, \eta, \xi; \theta) \pi(\theta) d\eta d\xi d\theta}$$

Donde:

$$L(y, x, \eta, \xi; \theta) = \prod_{i=1}^N \left\{ N_p(y_i; v_y + \Lambda_y \eta_i, \Sigma_y) N_p(x_i; v_x + \Lambda_x \xi_i, \Sigma_x) \times \right. \\ \left. \times N_m(\eta_i; \alpha + B\eta_i + \Gamma\xi_i, \Omega_\zeta) N_n(\xi_i; \mu_\xi, \Omega_\xi) \right\}$$

Y:

$$\theta = \alpha, b, \gamma, v_y, v_x, \lambda_y, \lambda_x, \sigma_y^2, \sigma_x^2, \omega_\zeta^2, \mu_\xi, \omega_\xi^2$$

Que es el vector de parámetros libres del modelo. Los demás parámetros se fijan en el proceso de especificación del modelo. De esta forma,  $\pi(\theta)$  representa la previa de dichos parámetros libres. Generalmente, por conveniencia se eligen previas normales o normales truncadas para los elementos de los términos de intercepto ( $v_y, v_x, \alpha$ ), las cargas factoriales ( $\lambda_y, \lambda_x$ ) y los parámetros estructurales ( $b, \gamma$ ). Por otro lado, para los parámetros que componen la matriz de varianzas y covarianzas se eligen previas gammas-inversas independientes, específicamente para los elementos de las diagonales de  $\Sigma_y, \Sigma_x, \Omega_\zeta$  y  $\Omega_\xi$ .

Kaplan y Depaoli (2012) también brindan su perspectiva sobre la estimación de BSEM, otorgando también una perspectiva clara sobre los indicadores de ajuste que deben ser tomados en cuenta al evaluar los modelos obtenidos.

Asparowhov y Muthén (2010) ofrecen un detalle técnico de la implementación utilizada en el programa Mplus para ejecutar análisis utilizando BSEM, por lo que pueden tomarse como referencia para establecer mayor flexibilidad en el modelo, ya



	<p>que incluyen diferentes enfoques para el manejo de las matrices de varianzas y covarianzas.</p> <p>Muthén y Asparowhov (2010) brindan algunas ideas sobre estudios de simulación aplicados a BSEM, y los indicadores que deben ser tomados en cuenta para su valoración.</p> <p>En cuanto a la estimación mediante R, Merkle y Rosseel (2016), brindan una guía bastante clara sobre el procedimiento que se debe seguir para la estimación de BSEMs utilizando la biblioteca <i>blavaan</i> como base.</p>
Fuente(s) de datos (si ya están recolectados) o diseño del estudio (si no están recolectados)	<p>Los datos pertenecen al Centro de Investigaciones en Biología Celular y Molecular (CIBCM) y provienen del estudio denominado “Descripción de las comunidades microbianas del tracto gastrointestinal (GI) y su implicación en el peso de una muestra de niños costarricenses menores de un año de edad (801-B7-612)”, en el que se desea generar evidencia para relaciones causales entre microbiomas, ambiente y aumento de peso en infantes.</p> <p>El estudio estuvo compuesto por una muestra de 60 individuos menores de 1 año, a los que se les realizó la medición de peso y toma de heces y oral cada 3 meses, durante 9 meses. El objetivo de este estudio fue evidenciar relaciones causales entre el microbioma oral e intestinal y la tendencia a ganar peso de los infantes, considerando factores ambientales como el contacto con mascotas, la alimentación con leche materna, entre otros.</p>
Definición de variable(s) de estudio	<p>Se consideran las siguientes variables latentes: composición y diversidad del microbioma, ambiente y cambio de peso. A continuación se presenta una breve descripción sobre estas:</p> <p><b>Microbioma:</b> en general, la composición de los diferentes microbiomas está dada por la abundancia relativa de los diversos géneros y especies bacterianas, determinada mediante secuenciación de la región hipervariable v3-v4 del gen ARN ribosomal 16S, de manera que los indicadores que conforman este constructo son datos de tipo composicional. Es muy posible que este constructo esté conformado por subconstructos denominados Microbioma Oral y Microbioma Fecal, ya que se tomaron muestras orales y fecales, lo que ocasiona que la diversidad bacteriana sea diferente para ambas muestras pero posiblemente estén correlacionadas.</p> <p><b>Ambiente:</b> Están dados principalmente por variables dicotómicas, ya que generalmente se considera la presencia o ausencia de ciertos hábitos como el consumo de leche materna, la tenencia de mascotas, la interacción con otros niños fuera del hogar, el consumo de probióticos después de la primera revisión, el consumo de medicamentos debido a resfríos o alergias después de la primera revisión, vacunación y el nivel de estrés de la mamá.</p> <p><b>Cambio de Peso:</b> Se determinó el peso en gramos (g) del individuo cada 3 meses y constituye una variable cuantitativa.</p>
Evidencias de calidad de la medición para la(s) variable(s) del estudio	<p>Las variables incluidas en el estudio del CIBCM fueron recolectadas por expertos de campo y por profesionales en medicina.</p> <p>En general, los aspectos relacionados a la entrevista a las madres de los niños, así como las mediciones de los diferentes componentes antropomórficos y fisiológicos de los infantes fueron realizados por los profesionales en medicina.</p>





	<p>Por otro lado, las lecturas de los ARN ribosomales 16S fueron generadas en un secuenciador de nueva generación MiSeq de la casa Illumina adquirido por el CIBCM. Todos los protocolos para la generación de los datos se realizaron en el CIBCM. La normalización y estructuración de estos datos fue llevada a cabo por los investigadores del CIBCM</p> <p>Los riesgos principales provienen de la aplicación de las entrevistas por parte de los médicos, ya que por restricciones en tiempo podrían no llenarlas con suficiente detalle.</p>
Modelo(s) estadísticos o técnica(s) estadística(s) de análisis a emplear	<p>En general, el enfoque de este estudio de simulación requiere la estimación de modelos de ecuaciones estructurales utilizando inferencia Bayesiana y, por lo tanto, considerando métodos de Cadenas de Markov Monte Carlo (MCMC) como la técnica de muestreo de Gibbs.</p> <p>De acuerdo con Dunson <i>et al.</i> (2005), la técnica de muestreo de Gibbs muestrea alternativamente de la distribución posterior conjunta de cada elemento desconocido, o conjunto de elementos desconocidos, incluyendo parámetros y variables latentes. Cada elemento o conjunto de estos se actualiza antes del siguiente paso. Usualmente, estas muestras convergen a una distribución estacionaria que es la distribución posterior conjunta. De esta manera, se puede correr el algoritmo de Gibbs y descartar una porción de estimaciones iniciales para permitir que los diagnósticos de convergencia puedan funcionar y así calcular las medidas de resumen basados en las muestras recolectadas.</p> <p>Para el caso de las distribuciones posteriores conjuntas en SEM, se utiliza la siguiente forma del teorema de Bayes:</p> $\pi(\Theta, \xi, \eta   y, x) \propto L(y, x, \eta, \xi; \Theta) \pi(\Theta)$ <p>Ya que el denominador es una constante de normalización denominada verosimilitud marginal.</p> <p>Tomando este planteamiento como referencia, el algoritmo de Gibbs puede seguir un orden similar al siguiente:</p> <ol style="list-style-type: none"> <li>1. Se determinan los valores iniciales <math>\Theta^0, \xi^0, \eta^0</math></li> <li>2. Se inicia el proceso de k iteraciones</li> <li>3. Para cada una de las <math>i</math> de <math>N</math> observaciones, se sigue el siguiente proceso de muestreo: <ol style="list-style-type: none"> <li>a. <math>\eta_i^k \sim \pi(\eta_i   \tilde{\mu}_{\eta_i}^{k-1}, \tilde{\Omega}_{\eta}^{k-1}, \xi_i^{k-1}, \Theta^{k-1}, y_i)</math></li> <li>b. <math>\xi_i^k \sim \pi(\xi_i   \eta_i^k, \Theta^{k-1}, x_i)</math></li> <li>c. <math>(v_y^k, \lambda_y^k) \sim \pi(v_y, \lambda_y   \Theta^{k-1}, \eta^k, y)</math></li> <li>d. <math>(v_x^k, \lambda_x^k) \sim \pi(v_x, \lambda_x   \Theta^{k-1}, \xi^k, x)</math></li> <li>e. <math>(\alpha^k, b^k, \gamma^k) \sim \pi(\alpha, b, \gamma   \eta^k, \xi^k, \Theta^{k-1}, x, y)</math></li> <li>f. <math>(\sigma_y^2)^k \sim \pi(\sigma_y^2   \eta^k, v_y^k, \lambda_y^k, \Theta^{k-1}, y)</math></li> <li>g. <math>(\sigma_x^2)^k \sim \pi(\sigma_x^2   \eta^k, v_x^k, \lambda_x^k, \Theta^{k-1}, x)</math></li> <li>h. <math>\mu_{\xi}^k \sim \pi(\mu_{\xi}   \xi^k, m, M)</math></li> <li>i. <math>\Omega_{\xi}^k \sim \pi(\Omega_{\xi}   \xi^k, \alpha_{\xi}, \beta_{\xi})</math></li> </ol> </li> </ol>



$$j. \quad \Omega_{\zeta}^k \sim \pi(\Omega_{\zeta} | \eta^k, \alpha_{\eta}, \beta_{\eta})$$

4. Se reinicia el ciclo con los nuevos parámetros  $\eta^k, \xi^k, \theta^k$

#### Estudio de Simulación

##### *Diseño*

En cuanto al estudio de simulación, se desea evaluar el efecto de los siguientes factores:

1. **Tamaño de muestra:** 15, 30 y 60 individuos,
2. **Cargas factoriales:** 0.1, 0.3 y 0.6.
3. **Efectos de las relaciones estructurales:** 0.1 y 0.3

##### *Generación de datos*

La generación de datos se realizará mediante programación de una rutina en R, produciendo 100 conjuntos de datos para cada una de las 18 combinaciones de los factores descritos en el apartado de diseño. De esta forma se deberán estimar modelos bayesianos de ecuaciones estructurales para un total de 1800 conjuntos de datos, con el fin de obtener los conjuntos de distribuciones empíricas para los parámetros estimados. Hallgren (2013) expone por qué R es un ambiente de programación adecuado para conducir estudios de simulación y provee algunas consideraciones a tener en cuenta.

##### *Análisis de datos*

La comparación del impacto en la precisión de los parámetros estimados mediante BSEM, para los factores citados anteriormente, se realizará mediante las medidas de rendimiento propuestas por Burton *et al.* (2006) en el diseño de estudios de simulación para estadísticas médicas, las cuales se detallan en el Cuadro 1.

**Cuadro 1.** Medidas de rendimiento para la evaluación de diferentes métodos estadísticos.

Criterio de Evaluación	Fórmula
Sesgo	$\delta = \bar{\beta} - \beta$
Porcentaje de Sesgo	$\left( \frac{\bar{\beta} - \beta}{\beta} \right) * 100$
Sesgo Estandarizado	$\left( \frac{\bar{\beta} - \beta}{SE(\bar{\beta})} \right) * 100$
Error Cuadrático Medio (MSE)	$(\bar{\beta} - \beta)^2 + (SE(\bar{\beta}))^2$
Cobertura	Proporción de veces en que el intervalo de credibilidad 95% determinado por $(P_{0.025, \hat{\beta}_i}, P_{0.975, \hat{\beta}_i})$ incluye a $\beta$ , para $i = 1, \dots, B$
Longitud Promedio del Intervalo de Credibilidad 95%	$\frac{\sum_{i=1}^B (P_{0.975, \hat{\beta}_i} - P_{0.025, \hat{\beta}_i})}{B}$

**Fuente:** Burton *et al.* (2006)



		<p>Donde:</p> <ul style="list-style-type: none"> <li>• <math>\beta</math> es el verdadero valor para el parámetro de interés</li> <li>• <math>B</math> es el número de iteraciones realizadas</li> <li>• <math>\hat{\beta}_i</math> es el estimador de interés en cada una de las <math>i = 1, \dots, B</math> iteraciones.</li> <li>• <math>SE(\hat{\beta})</math> es el error estándar empírico para el parámetro de interés calculado sobre todas las iteraciones.</li> <li>• <math>P_{0.025, \hat{\beta}_i}</math> y <math>P_{0.975, \hat{\beta}_i}</math> son los percentiles 2.5 y 97.5 de la distribución del estimador Bayesiano</li> </ul>
	Software estadístico a utilizar	El software a utilizar será R versión 3.5. Entre las bibliotecas de R a utilizar se encuentran lavaan, blavaan, sem, phyloseq, psych, entre otras que serán detalladas conforme avance el proyecto.
Referencias bibliográficas		<p>Asparowhov, T. y Muthén, B. 2010. <i>Bayesian Analysis using Mplus: Technical Implementation</i>. Disponible en <a href="http://www.statmodel.com/download/Bayes3.pdf">http://www.statmodel.com/download/Bayes3.pdf</a></p> <p>Burton, A., Altman, D., Royston, P. y Holder, R. 2006. The design of simulation studies in medical statistics. <i>Statistics in Medicina</i>, 25:4279-4292. Disponible en <a href="http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.457.5976&amp;rep=rep1&amp;type=pdf">http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.457.5976&amp;rep=rep1&amp;type=pdf</a></p> <p>Dunson, D., Palomo, J. y Bollen, K. 2005. <i>Bayesian Structural Equation Modeling</i>. Statistical and Applied Mathematical Sciences Institute. Disponible en <a href="https://www.samsi.info/wp-content/uploads/2010/08/tr2005-05.pdf">https://www.samsi.info/wp-content/uploads/2010/08/tr2005-05.pdf</a></p> <p>Egozcue, J.J., Parlowsky-Glahn, V. y Gloor, G.B. 2018. Linear Association in Compositional Data Analysis. <i>Austrian Journal of Statistics</i>, 47. Disponible en <a href="https://www.researchgate.net/profile/Juan_Jose_Egozcue/publication/322809225_Linear_Association_in_Compositional_Data_Analysis/links/5a720b440f7e9b20d48e0102/Linear-Association-in-Compositional-Data-Analysis.pdf">https://www.researchgate.net/profile/Juan_Jose_Egozcue/publication/322809225_Linear_Association_in_Compositional_Data_Analysis/links/5a720b440f7e9b20d48e0102/Linear-Association-in-Compositional-Data-Analysis.pdf</a></p> <p>Hackshaw, A. 2008. <i>Small studies: strengths and limitations</i>. <i>Eur Respir J</i>, 32; 1141 – 1143. Disponible en <a href="https://erj.ersjournals.com/content/erj/32/5/1141.full.pdf">https://erj.ersjournals.com/content/erj/32/5/1141.full.pdf</a></p> <p>J. Craig Venter Institute (JCVI). 2018. <i>Human Microbiome Studies: Research Projects</i>. Disponible en <a href="https://www.jcvi.org/human-microbiome-studies">https://www.jcvi.org/human-microbiome-studies</a></p> <p>Kaplan, D. y Depaoli, S. 2012. <i>Chapter 38: Bayesian Structural Equation Modeling</i>. Handbook of Structural Equation Modeling. The Guildford Press. Disponible en <a href="https://www.statmodel.com/download/Kaplan_Depaoli.pdf">https://www.statmodel.com/download/Kaplan_Depaoli.pdf</a></p> <p>Hallgren, A. 2013. <i>Conducting Simulation Studies in the R Programming Environment</i>. <i>Tutor Quant Methods Psychol.</i> 9 (2); 43-60. Disponible en <a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4110976/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4110976/</a></p> <p>Mateu-Figueras, G., Martín-Fernández, J.A., Pawlosky-Glahn, V. y Barceló-Vidal, C. 2003. <i>El problema del análisis estadístico de datos composicionales</i>. 27 Congreso Nacional de Estadística e Investigación Operativa. Disponible en <a href="http://ima.udg.edu/~jamf/paper_INTRO_MMPB.pdf">http://ima.udg.edu/~jamf/paper_INTRO_MMPB.pdf</a></p>





Merkle, E. y Rosseel, Y. 2015. *blavaan: Bayesian Structural Equation Models via Parameter Expansion*. Journal of Statistical Software, 85(4). Disponible en <https://arxiv.org/pdf/1511.05604.pdf>

Muthén, B. y Asparowhov, T. 2010. *Bayesian SEM: A more flexible representation of substantive theory*. Disponible en <https://www.statmodel.com/download/BSEMV4.pdf>

Orcan, F. 2013. *Use of ítem parceling in structural equation modeling with missing data*. Florida State University. Disponible en <https://diginole.lib.fsu.edu/islandora/object/fsu:185150/datastream/PDF/download/citation.pdf>

Pawlowsky-Glahn, V. y Buccianti, A. 2011. *Compositional Data Analysis: Theory and Applications. First Edition*. West Sussex: JohnWiley & Sons,Ltd.

Sohn, M. y Li, H. 2017. Compositional Mediation Analusis for Microbiome Studies. bioRxiv The preprint server for biology. Disponible en <https://pdfs.semanticscholar.org/0e37/7bed365a2d49ebc2154956bfd0759b3760e2.pdf>

Weiss, S., Zech, Z., Peddada, S., Amir, A., Bittinger, K. Gonzalez, A. Lozupone, C., et al. 2017. *Normalization and microbial differential abundance strategies depend upon data characteristics*. Microbiome, 5:27. Disponible en [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5335496/pdf/40168\\_2017\\_Article\\_237.pdf](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5335496/pdf/40168_2017_Article_237.pdf)

Westland, J.C. 2010. *Lower Bounds on Sample Size in Structural Equation Modeling*. Electronic Commerce Research and Applications, 9:6.

Wolf, E., Harrington, K., Clark, S. y Miller, M. 2013. *Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety*. Educ. Psychol. Meas., 76:6. Disponible en <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4334479/pdf/nihms661496.pdf>



### Cronograma

Actividad	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Set	Oct	Nov
Presentación de la propuesta											
Introducción y Justificación											
Marco Teórico											
Estudio de Simulación											
Análisis de Estudio de Simulación											
Elaboración de Resultados de Estudio de Simulación											
Análisis de Datos Empíricos											
Elaboración de Resultados de Datos Empíricos											
Discusión de Resultados y Conclusiones											
Primera Presentación y Revisión de Comité											
Aplicación de Correcciones											
Segunda presentación y Revisión de Comité											
Aplicación de Correcciones Finales											
Presentación de Tesis											

**Artículo 15.** Para el Programa de Maestría Académica el periodo máximo entre el ingreso del estudiante a la segunda etapa (fecha del primer curso matriculado) y la presentación del examen de candidatura es de cuatro años. Si no lo aprobara en ese periodo quedará automáticamente fuera del Programa, pudiendo solicitar a la Comisión traslado al Programa de la Maestría Profesional en Estadística. El estudiante tendrá un plazo de tres ciclos lectivos para completar la tercera etapa, a partir de la fecha de aprobación del examen de candidatura. Si al cabo de este periodo el estudiante no ha presentado la tesis, la Comisión podrá conceder una única prórroga de un ciclo lectivo, al cabo del cual, el estudiante que no ha defendido su tesis será separado del programa.

**FIRMA DEL ESTUDIANTE**

Edgardo Gutiérrez Vega

B38558