

MODELADO CON ECUACIONES ESTRUCTURALES PARA EVALUAR EL APRENDIZAJE EN INGENIERÍA

STRUCTURAL EQUATION MODELING TO ASSESS ENGINEERING LEARNING

Sergio F. Yapur^{a,b,c}

^a*Centro de Investigación de Métodos Computacionales, Predio CONICET Santa Fe "Dr. Alberto Cassano" Colectora Ruta Nac N° 168, Km 0, Paraje El Pozo. 3000 Santa Fe, Argentina, sergio.yapur@cimec.unl.edu.ar, <http://www.cimec.org.ar>*

^b*Facultad de Ingeniería en Ciencias Hídricas, Universidad Nacional del Litoral, Ciudad Universitaria. Ruta Nacional N° 168 - Km 0, 3000 Santa Fe, Argentina, <http://fich.unl.edu.ar>*

^c*Facultad de Ingeniería Química, Universidad Nacional del Litoral, Santiago del Estero 2829, 3000 Santa Fe, Argentina, <http://fiq.unl.edu.ar>*

Palabras clave: modelado con ecuaciones estructurales, estadística multiparamétrica, educación en ingeniería, computación, optimización, métodos numéricos, matemática.

Resumen. Este artículo emplea modelado por ecuaciones estructurales (SEM por *Structural Equation Modeling*) aplicado a un curso universitario de grado en ingeniería. El objetivo es estimar las relaciones subyacentes entre variables relevantes del proceso de aprendizaje, con el fin de cuantificar sus efectos y habilitar una estrategia más efectiva de enseñanza. Se discute la especificación del modelo, su implementación, identificación y estimación de parámetros. Los datos se obtienen mediante un relevamiento de datos con alumnos de un curso básico de grado durante el primer semestre del año 2018.

Keywords: structural equation modeling, multiparametric statistics, engineering education, computation, optimization, numerical methods, mathematics.

Abstract. Structural Equation Modeling (SEM) is applied to a university grade course in engineering. The objective is to estimate the underlying relationships among variables that are relevant to the learning process, in order to quantify their effects and enable a more effective teaching strategy. The specification of the model, its implementation, identification, and estimation of parameters is discussed. The data is obtained from a survey of the students taking a basic degree course during the first semester of 2018.

1. INTRODUCCIÓN

La formación académica en ingeniería es central en un país que fomente un desarrollo industrial y tecnológico. Es vital también para el crecimiento de las comunidades científicas, ya que una experiencia positiva en el aprendizaje de grado promueve el camino hacia la investigación.

Al día de hoy, prevalecen problemáticas que obstaculizan la efectividad de la enseñanza. Los programas de contenidos de los cursos universitarios están definidos principalmente por las necesidades de cada carrera objetivo. Sin embargo, no se realiza un análisis previo de las necesidades y carencias de los estudiantes, especialmente en la transición de la educación media a la superior. Como consecuencia, puede darse un desfase en lo que se pretende enseñar y lo que el alumno está preparado para asimilar al momento de comenzar un determinado curso. Este inconveniente solo puede agravarse si se deteriora el nivel de la educación media e inicial. Ante este desfase, el alumno debe compensar sus falencias de forma autónoma, al tiempo que incorpora todos los contenidos nuevos. Esta situación aumenta el número de alumnos recursantes y eventualmente incrementa la deserción en la educación superior. Todo esto se traduce en una menor eficiencia del sistema educativo y en un mayor gasto en educación pública por parte del estado.

Según se ha demostrado a través de las encuestas PISA ([Organisation for Economic Co-operation and Development, 2017](#)) sobre Argentina en el periodo 2001-2012¹, nuestro país persiste en el último o penúltimo lugar en las tres categorías evaluadas: Ciencias, Matemáticas y Habilidad Lectora, si se hace una comparación con los países que representan las principales economías iberoamericanas: España, Portugal, México, Brasil y Chile. Adicionalmente, otros estudios de la OECD revelan que Argentina tiene aproximadamente la mitad de egresados en ciencias e ingenierías que el promedio de los países evaluados por la OECD.

En virtud de lo anterior, en este trabajo se implementa un análisis estadístico que por su complejidad, no suele emplearse en el ámbito de la educación superior. Para esto, se relevó información en un curso de Teoría de la Computación, obligatorio en la carrera de Ingeniería en Informática. Es claro que las notas de por sí constituyen una medida global de la efectividad del curso. Pero dicha medida es producto a su vez de otras variables de importancia, muchas de las cuales no son directamente observables. A estas variables las llamaremos latentes, y la interrelación entre ellas resulta crucial para definir una estrategia que mejore el proceso de aprendizaje.

2. MODELADO POR ECUACIONES ESTRUCTURALES

2.1. Introducción

La teoría de Modelado por Ecuaciones Estructurales (SEM, por su sigla en inglés) agrupa varias técnicas de análisis estadístico multivariable, que se utilizan en conjunto para contrastar modelos, los cuales se construyen a partir de relaciones causales o de covarianza entre las variables que los componen ([Ruiz et al., 2010](#)). Esta característica lo distingue de otros enfoques estadísticos clásicos, como ser el de análisis de varianza o de regresión, ya que no se estiman respuestas individuales, sino las covarianzas entre variables.

2.2. Definiciones Básicas

A continuación se lista una serie de definiciones básicas con el fin de facilitar el resto de la lectura ([Ruiz et al., 2010](#)).

¹La Argentina no participó en las últimas encuestas, realizadas durante el 2015.

- *Variable latente*: También llamados constructos o factores, son variables que no pueden observarse o medirse directamente.
- *Indicadores*: Son las variables medibles u observables.
- *Variable error*: Representa tanto los errores asociados a la medición de una variable como a los errores asociados al conjunto de variables que se contemplan en el modelo, y que pueden afectar a la medición de una variable observada.
- *Variable exógena*: Variable que afecta a otras variables y que no recibe efecto de ninguna variable. Las variables independientes son exógenas.
- *Variable endógena*: Variable que recibe efecto de otra variable. Las variables dependientes son endógenas.

Cabe mencionar que la teoría SEM admite dos tipos de relaciones, las de causalidad y las bidireccionales. En las primeras, el cambio de una variable afecta a otra, pero no se observa el efecto inverso. En las relaciones bidireccionales, también llamadas de correlación o covarianzas, las variables crecen o decrecen simultáneamente. Ocasionalmente, una relación bidireccional entre dos variables puede indicar que las mismas son consecuencia de una variable común, normalmente de tipo latente.

2.3. Marco Teórico

En este trabajo, se usan relaciones lineales para determinar la funcionalidad entre las variables, de forma que el modelo queda determinado como un conjunto de matrices. En particular, se emplea la notación asociada con el código LISREL (Lineal Structural Relationships) descrito por Jöreskog y Sörbom ([Schumacker y Lomax, 2004](#)):

$$\begin{aligned}\eta &= \mathbf{B}\eta + \mathbf{\Gamma}\xi + \zeta, \\ \mathbf{Y} &= \mathbf{\Lambda}_y\eta + \epsilon, \\ \mathbf{X} &= \mathbf{\Lambda}_x\xi + \delta.\end{aligned}\tag{1}$$

Donde la primer ecuación expresa las relaciones entre variables latentes, mientras que la segunda y la tercera son llamados modelos de medida; estos representan las relaciones entre indicadores y variables latentes. Así, η es el vector de variables latentes endógenas, ξ es el vector de variables latentes exógenas, ζ es el error estructural del modelo, Y es el vector de indicadores de las variables latentes endógenas, X es el vector de indicadores de las variables latentes exógenas, ϵ es el vector de errores de medida de Y y δ es el vector de errores de medidas de X . El resto de los símbolos representan matrices. Así, B es la matriz que representa las interacciones entre los elementos de η , y su diagonal es nula. Además Λ_y y Λ_x son las matrices que relacionan indicadores con variables latentes, y Γ es la matriz que vincula variables latentes exógenas con endógenas.

Un modelo estructural implica una estructura para las covarianzas entre las variables observadas. Es decir, a partir de las relaciones que plantea el modelo es posible derivar una matriz de covarianzas, que depende de un conjunto de parámetros. La hipótesis básica de la teoría SEM es que una buena estimación de los parámetros se asocia a la idoneidad del modelo para representar las relaciones entre variables ([Hox y Bechger, 2009](#)), ya que la matriz de covarianzas del modelo se asemeja a la matriz de covarianzas calculada a partir de las variables observadas.

Nótese que el modelo contempla tanto los errores ϵ , δ en los modelos de medida como un error asociado a las relaciones planteadas, que definen la estructura del modelo. En la práctica, se suele asumir que los errores son independientes entre sí, y por lo tanto tienen covarianzas nulas.

3. DATOS DE ENTRADA

3.1. Sondeo de Datos

Como en todo estudio estadístico, se comenzó con un relevamiento de datos. Los mismos fueron obtenidos mediante una encuesta realizada al comienzo del primer cuatrimestre del año 2018, en el curso de Teoría de la Computación de la Facultad de Ingeniería y Ciencias Hídricas de la Universidad Nacional del Litoral. Dicho curso se dicta para las carreras de Ingeniería Informática e Ingeniería Industrial. Los datos relevados con las encuestas fueron complementados con los registros de asistencia a clases de teoría y de práctica, así como también con las notas alcanzadas en dos exámenes parciales durante el cuatrimestre.

El programa del curso de Teoría de la Computación involucra mayormente tópicos de Matemática Discreta con diversas aplicaciones a la computación. Así, se enseña lógica proposicional, métodos de demostración, combinatoria, grafos y árboles, entre otros tópicos.

Originalmente, la nómina de inscriptos totalizaba 130 alumnos. No obstante, solo una fracción del total optó por completar la encuesta. Además, se descartaron encuestas que no estaban completamente contestadas. Finalmente, el tamaño de la muestra resultó ser de 41 individuos. Este tamaño muestral resulta ser relativamente pequeño. No obstante, este autor considera que es suficiente a los fines de elaborar los algoritmos necesarios y obtener algunos resultados preliminares. Futuros relevamientos de datos permitirán mejorar la calidad de los resultados.

4. MODELADO

4.1. Métodos y Aspecto Computacional

Existen varios programas en la actualidad que fueron específicamente desarrollados para aplicar SEM, como LISREL, EQS o AMOS. Sin embargo, los mismos son comerciales y resulta imposible tener acceso a la organización del algoritmo, métodos y parámetros del cálculo numérico involucrado. Como consecuencia, se optó por desarrollar un código propio en el entorno Matlab[®].

El primer paso es adoptar los constructos o variables latentes que se desean estudiar. Luego se seleccionan los indicadores adecuados y finalmente se trabaja sobre la estructura del modelo, según las ecuaciones (1). De este proceso se dimensionan las matrices B , Λ_x , Λ_y y Γ . Los parámetros desconocidos de estas matrices se agrupan en el vector θ . Luego es posible construir la matriz de covarianzas inferida $\Sigma(\theta)$, a partir de (1), de forma que quede expresada en términos de θ . La expresión resultante es

$$\Sigma(\theta) = \begin{bmatrix} \Lambda_y(I - B)^{-1}(\Gamma\Phi\Gamma' + \Psi)((I - B)^{-1})'\Lambda_y' + \theta_\epsilon & \Lambda_y(I - B)^{-1}\Gamma\Phi\Lambda_x' \\ \Lambda_x\Phi\Gamma'((I - B)^{-1})'\Lambda_y' & \Lambda_x\Phi\Lambda_x' + \theta_\delta \end{bmatrix} \quad (2)$$

donde I representa la identidad, Ψ es la matriz de covarianzas de los errores estructurales ζ , Φ es la matriz de covarianzas de las variables latentes ξ , θ_ϵ es la matriz de covarianzas de los errores de medida ϵ , y θ_δ es la matriz de covarianzas de los errores de medida δ .

Un modelo bien especificado aproximaría la matriz de covarianzas muestral S independientemente de la muestra tomada, ya que se modelan relaciones sobre la población y no sobre una

muestra en particular. Para lograr esto se maximiza la verosimilitud, es decir, la plausibilidad de que el modelo represente la población bajo el supuesto de que sus parámetros $\hat{\theta}$ siguen una distribución normal multivariable², dada la observación muestral S . Este tipo de estimación es el método por defecto en SEM (Kline, 2011). Maximizar la verosimilitud equivale a minimizar la Función de Máxima Verosimilitud (MLF). En el contexto SEM, este estimador tiene la siguiente forma

$$MLF = \log(|\Sigma(\theta)|) + \text{tr}(S\Sigma^{-1}(\theta)) - \log(|S|) - (p + q) \quad (3)$$

donde S es la matriz de covarianzas muestral, p es el número de indicadores de variables latentes exógenas, es decir $p = \dim(X)$ y q es el número de indicadores de variables latentes endógenas, $q = \dim(Y)$.

De esta forma, se aplica un proceso de optimización para estimar los parámetros. Así, una estimación óptima $\hat{\theta}$ maximiza la verosimilitud del modelo, que está asociado a minimizar la distancia $d(\Sigma(\hat{\theta}), S)$ entre la matriz de covarianzas muestral S y las covarianzas poblacionales inferidas $\Sigma(\hat{\theta})$, aunque no debe confundirse con un ajuste de datos (Kline, 2011).

En este trabajo la optimización sobre la función MLF fue de tipo no lineal, con un algoritmo de punto fijo con restricciones sobre las varianzas, que por su naturaleza son no negativas³.

4.2. Hipótesis y Reglas de Modelado

Con el fin de asegurar una correcta especificación del modelo, se adoptaron las siguientes hipótesis de modelado:

- La covarianza entre errores estructurales ζ es nula. Esto es, $COV(\zeta_i, \zeta_j) = 0 \quad \forall i \neq j$.
- La covarianza entre errores ϵ en el modelo de medida de Y es nula
- La covarianza entre errores δ en el modelo de medida de X es nula
- Todos los coeficientes asociados a errores se fijan en 1 (es decir, no se normaliza la varianza)

A las hipótesis anteriores se agregan las siguientes reglas heurísticas, que facilitan la identificabilidad (Hildreth, 2013), que asegura la unicidad de la solución:

- Se evitan ciclos o bucles entre las relaciones propuestas
- Cada indicador es causado por exactamente una variable latente

5. MODELO

Debido a la naturaleza ya mencionada de la teoría SEM, se ensayó con más de veinte modelos distintos, de complejidad creciente. Aquí se presenta un solo modelo, que puede representarse gráficamente como un grafo dirigido, como se muestra en la Figura 1. Las variables seleccionadas para estudiar mediante el modelo se presentan en la Tabla 1, donde se establece además el tipo de variable junto con una breve descripción de su significado.

²No debe confundirse la distribución de los parámetros del modelo θ con las distribuciones asociadas a los indicadores X, Y .

³Este criterio de restringir las varianzas a valores no negativos es aplicado también por el software EQS.

Nº	Variable	Tipo	Descripción	Sigla
1	η_1	Latente Endógena	Aprendizaje	Ap
2	η_2	Latente Endógena	Compromiso	C
3	η_3	Latente Endógena	Carga Externa	CE
4	ξ	Latente Exógena	Antecedentes	At
5	y_1	Indicador	Nota parcial 1	N ₁
6	y_2	Indicador	Nota parcial 2	N ₂
7	y_3	Indicador	Asistencia a Práctica	AP
8	y_4	Indicador	Asistencia a Teoría	AT
9	y_5	Indicador	Personas a Cargo	PC
10	y_6	Indicador	Trabaja	T
11	x_1	Indicador	Relación histórica con las matemáticas	RH
12	x_2	Indicador	Recursante	R

Tabla 1: Descripción de variables de modelo SEM

La representación del modelo en la Figura 1 sigue la convención establecida en la literatura. Las variables indicadoras se representan dentro de rectángulos, mientras que las variables latentes se representan dentro de círculos. Además, las flechas simples indican relaciones de causalidad. De esta forma, las flechas del diagrama permiten reconstruir las relaciones del modelo. Por ejemplo, $x_1 = \lambda_1 \xi + \delta_1$. Como puede verse en la Figura 1, se plantean tres variables latentes y una variable exógena. La notación de las variables responde a la dada en la Ecuación (1), con los parámetros disgregados por elemento en lugar de usar representación vectorial, para mayor claridad. Nótese que se los coeficientes λ_i , de un solo subíndice, son elementos de la matriz Λ_x , mientras que los coeficientes λ_{ij} de doble subíndice, son elementos de la matriz Λ_y .

Desde el punto de vista del problema de identificación el modelo hace uso de $p = 8$ variables observadas, con lo cual se tienen $p(p + 1)/2 = 36$ varianzas y covarianzas distintas. Por otro lado, el modelo tiene 20 incógnitas. Esto da como resultado un número de grados de libertad $d_f = 16$. Como consecuencia se tiene un modelo sobre-especificado, lo cual es deseable y resulta una condición necesaria para satisfacer identificabilidad (Loehlin, 2004).

Para asegurar la identificabilidad se impusieron las pautas mencionadas en la Sección 4.2. En primer lugar, se fija uno de los coeficientes que asocia cada variable latente con sus indicadores en uno, para resolver la ambigüedad por escala. Esto puede verse en la Figura 1. Además, se adoptó la regla recursiva, que establece que los errores estructurales no están correlacionados (Ψ es diagonal). Finalmente, todas las relaciones causales son unidireccionales, es decir, no hay ciclos de realimentación o caminos recíprocos.

En la práctica, la identificabilidad se comprobó corriendo varias veces el código con distintos valores iniciales de los parámetros. Estos valores fueron generados aleatoriamente en cada ocasión.

6. RESULTADOS

La minimización sobre la función MLF dada por la Ecuación 3 alcanzó un valor de 0.61 luego de 157 iteraciones. En la Tabla 2 se muestran los parámetros $\hat{\theta}_i$ estimados, truncados al segundo decimal. Las primeras dos columnas dan los valores de los coeficientes del modelo, que pueden verse en la Figura 1, mientras que las últimas dos columnas dan las varianzas de los errores asociados al modelo.

Las varianzas de los errores en general son relativamente pequeñas, lo cual indica poca dis-

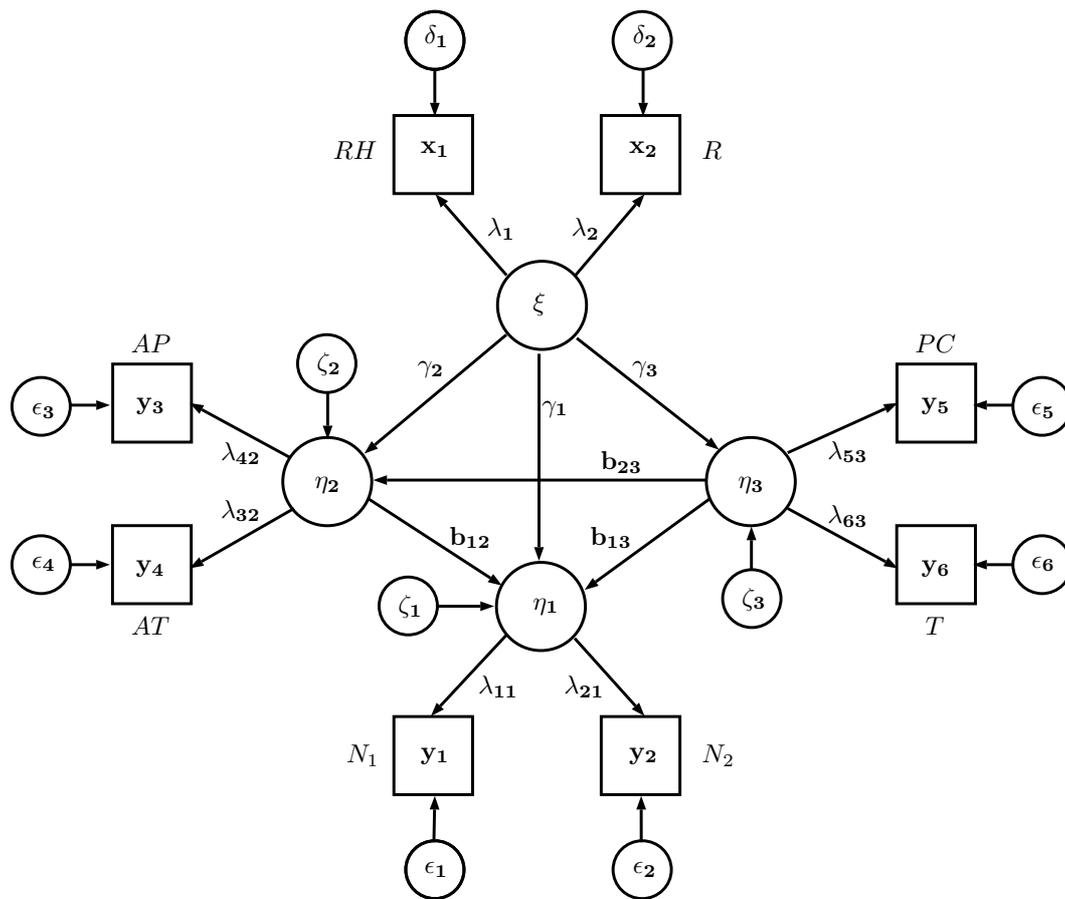


Figura 1: Modelo Propuesto

persión. Esto es especialmente cierto para los errores estructurales ζ_i y para ξ . Estas características apoyan las interrelaciones planteadas entre las variables latentes, como suficientes para representar los datos. Por otro lado, las varianzas asociadas a los indicadores son algo mayores. Esto puede estar asociado a que se usaron solo dos indicadores por variable latente, que es el menor número posible. En general, una menor varianzas están asociadas a una mayor confiabilidad del coeficiente estimado (Schumacker y Lomax, 2004).

Por otro lado, en la Figura 2 se muestra la matriz S de covarianzas muestral junto a la matriz $\Sigma(\hat{\theta})$ de covarianzas estimadas para la población. Además, en la Figura 3 se presentan los residuos de la matriz de covarianzas, es decir los elementos $|S_{ij} - \hat{\Sigma}_{ij}|$, para $i, j = 1, 2, 3$.

Coefficiente	Valor	Varianza	Valor
λ_1	1.00	$\text{Var}(\xi)$	0.12
λ_2	-1.35	$\text{Var}(\zeta_1)$	0.16
λ_{11}	1.00	$\text{Var}(\zeta_2)$	0.00
λ_{21}	0.82	$\text{Var}(\zeta_3)$	0.01
λ_{32}	1.00	$\text{Var}(\delta_1)$	0.82
λ_{42}	0.71	$\text{Var}(\delta_2)$	0.77
λ_{53}	1.00	$\text{Var}(\epsilon_1)$	0.00
λ_{63}	7.40	$\text{Var}(\epsilon_2)$	0.32
b_{12}	1.00	$\text{Var}(\epsilon_3)$	0.69
b_{13}	-1.95	$\text{Var}(\epsilon_4)$	0.87
b_{23}	-2.30	$\text{Var}(\epsilon_5)$	0.98
γ_1	1.00	$\text{Var}(\epsilon_6)$	0.00
γ_2	1.15		
γ_3	0.01		

Tabla 2: Estimación de parámetros del modelo SEM

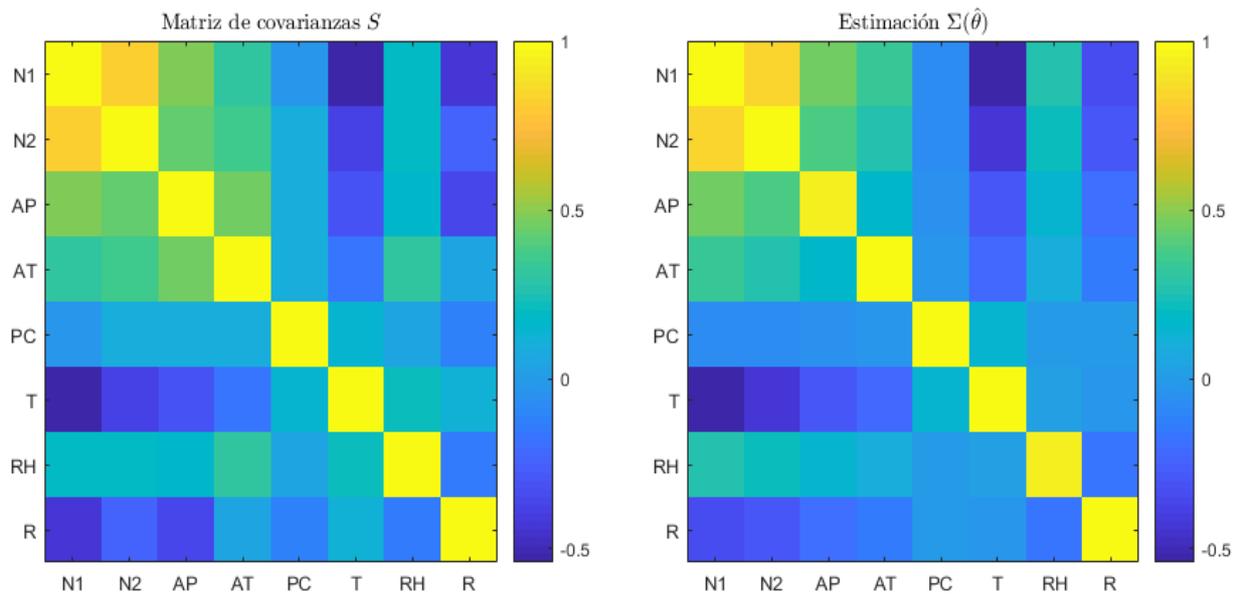


Figura 2: Comparación entre matriz de covarianzas muestral y estimación poblacional

Se observa que hay una correspondencia aproximada entre ambas matrices, con una desviación absoluta máxima de 0.28, correspondiente a la covarianza entre AT y AP. Esto es, el modelo subestima la relación entre estas variables con respecto al valor derivado de la muestra, que es de $S_{34} = 0,46$. En segundo lugar, con una desviación absoluta de 0.21, el modelo subestima la relación entre AT y RH con respecto al valor $S_{47} = 0,21$ obtenido de la muestra. El resto de las desviaciones es menor a 0.18. Como una medida más global del grado de similitud entre S y $\hat{\Sigma}$, puede usarse la norma-2, que da una distancia $\|S - \hat{\Sigma}\|_2 = 0,53$.

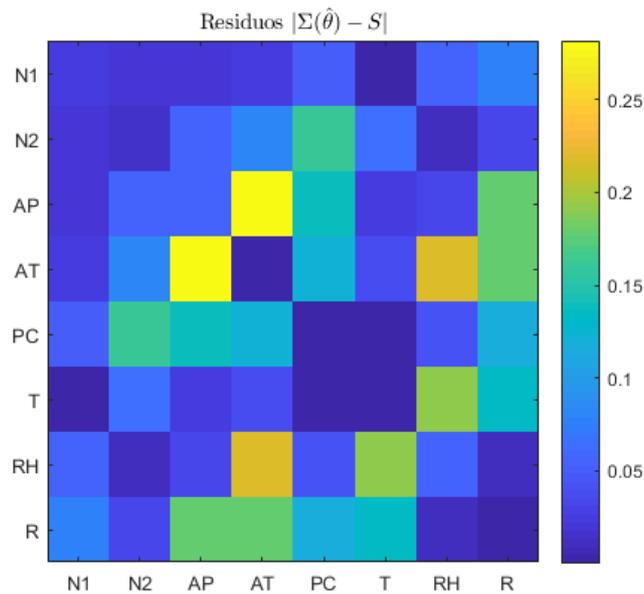


Figura 3: Residuos numéricos de la estimación

Si bien se considera conveniente presentar estos resultados de forma gráfica, se debe recordar que no se intenta aquí hacer un ajuste de datos, sino de inferir relaciones sobre la población. De hecho, cualquier modelo, incluso uno mal especificado, puede ajustarse a los datos simplemente agregando parámetros libres (Kline, 2011). El objetivo del análisis SEM es probar un modelo o teoría desde un punto de vista estadístico.

6.1. Interpretación

Los coeficientes del modelo revelan cuantitativamente cuales son los factores que más afectan el aprendizaje η_1 , así como las interrelaciones entre los factores. En la discusión que sigue, debe tenerse en cuenta que todos los coeficientes son valores relativos a la estructura del modelo. Así, por ejemplo, un modelo que involucre más variables podría presentar distintos coeficientes, con respecto al presente modelo, asociados a las mismas variables. No obstante, las observaciones cuantitativas tienen validez en el contexto del modelado que se discute en este trabajo.

En primer lugar, se observa que los antecedentes ξ tienen una influencia un 15 % mayor sobre el compromiso η_2 que sobre el aprendizaje η_1 . Sin embargo, la diferencia de signo entre λ_1 y λ_2 hace que los indicadores afectados se interpreten de distinta forma. Así, la condición de recurrente R afecta negativamente el compromiso con el curso de acuerdo al factor combinado $\lambda_2 \times \gamma_2 = -1,55$ (ver Figura 1). Esto podría explicar por qué existen alumnos que recursan varias veces. La condición de recurrente implica un menor compromiso con seguir el curso, que a su vez empeora el aprendizaje, resultando en menores notas y una mayor probabilidad de volver a recursar. Por otro lado, una buena relación previa con el área de matemáticas RH ayuda a consolidar un mayor compromiso con el curso. Sin embargo, este efecto positivo es cuantitativamente menor que el efecto negativo de R sobre el compromiso η_2 . Por otro lado, no hay un efecto apreciable entre los antecedentes ξ y la carga externa η_3 , como era de esperarse.

Por su lado, la carga externa η_3 afecta negativamente tanto al compromiso como al aprendizaje, con un mayor efecto sobre el primero. Así, el factor $b_{23} = -2,30$ constituye una medida del efecto de la carga externa η_3 sobre el compromiso η_2 , mientras que su efecto sobre el aprendi-

zaje η_1 se cuantifica en el factor $b_{13} = -1,95$. Estos valores son relativos al valor fijado $b_{12} = 1$, que mide el efecto del compromiso sobre el aprendizaje.

De los coeficientes obtenidos, se debe destacar el indicador T de η_3 , que presenta un valor $\lambda_{63} = 7,40$. Así, el mayor efecto sobre η_1 se encuentra sobre aquellos alumnos que trabajan. Esto se da mediante dos caminos. El camino directo que tiene un coeficiente combinado de $\lambda_{63} \times b_{13} = -14,43$ entre T y η_1 . El camino directo podría asociarse al menor tiempo que el individuo tiene para estudiar debido a su trabajo. Pero aun mayor es el efecto indirecto mediante el compromiso, con un coeficiente combinado de $\lambda_{63} \times b_{23} \times b_{12} = -17,02$. El efecto total del T sobre η_1 se da por la suma de ambos coeficientes combinados, que da un valor de $-31,45$. Claramente, el modelo sugiere que hay un efecto negativo dominante cuando el individuo trabaja.

Se pueden resumir los mayores efectos positivos sobre el aprendizaje como sigue: en primer lugar, RH afecta a η_1 con un coeficiente combinado total de 2.13 (camino directo e indirectos). En segundo y tercer lugares, hay un efecto positivo de 1.00 entre AP y η_1 , y de 0.7 entre AT y η_1 . Esto indicaría que la predisposición previa del estudiante con el área de estudio es de la mayor importancia, ya que condiciona directa e indirectamente la asimilación de contenidos. Nuevamente, aquí surge naturalmente el concepto de realimentación positiva, debido a que la variable RH se define en buena medida por la experiencia previa que el alumno tuvo con las matemáticas. Así, experiencias tempranas positivas son determinantes en futuros desempeños asociados con matemáticas.

También se pueden ordenar por importancia los efectos negativos sobre η_1 . En primer lugar se encuentra T, como ya se detalló anteriormente. En segundo lugar se encuentra la condición de recurrente R, con un coeficiente combinado total de aproximadamente -2.9. Se observa que el efecto de R es de un orden de magnitud menor que el de T sobre el aprendizaje. En tercer lugar, se encuentra el efecto de tener personas a cargo PC, con un factor de -1.9.

Finalmente, notese que N_2 es un indicador de menor importancia de η_1 con respecto a N_1 . Esto puede explicarse por el hecho de que algunos de los alumnos que no tuvieron el resultado que esperaban en el primer parcial, no se presentan al segundo. De esta forma, una mayor ausencia en el segundo parcial disminuye la correlación de esta N_2 con el resto de las variables.

6.2. Validación de Modelo por Prueba de Hipótesis

La validación de modelos estadísticos mediante pruebas de hipótesis es ampliamente aceptado en la literatura. En particular, la prueba del cociente de máxima verosimilitud es uno de los métodos más populares (Pinto y Mingoti, 2015). Esta prueba asume que las variables están normalmente distribuidas. La hipótesis nula y la alternativa son, respectivamente

$$H_0 : \Sigma = \Sigma_0 \quad (4)$$

$$H_1 : \Sigma \neq \Sigma_0 \quad (5)$$

donde Σ_0 es la matriz de covarianzas poblacional estimada y el estadístico de cociente de verosimilitud está dado por

$$W = -kn + kn \ln(n) - n \ln(|A\Sigma_0^{-1}|) + tr(\Sigma_0^{-1}A) \quad (6)$$

con $A = (n - 1)S$, mientras que $k = 8$ es el número de variables involucradas en la matriz de covarianzas y $n = 41$ es el tamaño de la muestra. La distribución de W bajo la hipótesis nula es asintóticamente χ^2 . Esta prueba se realiza con un nivel de significación $\alpha = 0,05$ y $k(k + 1)/2$

grados de libertad. Con estos valores, se obtiene $W = 25,3$ y $\chi_{k(k+1)/2}^2 = 50,9$, con lo cual $W < \chi_{k(k+1)/2}^2$ y no se rechaza la hipótesis nula.

De la prueba de hipótesis se puede concluir que hay 95 % de probabilidad de que el modelo represente la población bajo estudio.

7. CONCLUSIONES

Se presentó un modelo de ecuaciones estructurales que representa la relación de distintas variables sobre el proceso de aprendizaje para un curso universitario de Teoría de Computación. Se obtuvieron resultados preliminares que indican la importancia relativa de cada factor considerado sobre dicho proceso, así como la importancia relativa de los indicadores individuales.

Concretamente, el modelo establece que el mayor impacto relativo viene dado por el condicionamiento previo que el estudiante tiene con el área, si bien el grado de compromiso con el cursado también resulta ser un factor importante. Esto revela la importancia de proyectar la educación matemática como una concatenación de experiencias positivas desde un principio. Pero lo anterior resulta prácticamente imposible si no se nivela adecuadamente a los ingresantes. Una estrategia para esto puede ser extender el ciclo de ingreso. Asimismo, enfatizar la conexión de los temas impartidos en las distintas cátedras podría tener un efecto favorable sobre el aprendizaje.

Por otro lado, el mayor efecto negativo está dado por la condición de trabajar y estudiar en simultáneo, seguido por la situación de recursante y finalmente por la circunstancia de tener personas a cargo. En este caso, es posible que una atención más personalizada sobre estos estudiantes los ayude a aumentar su nivel de compromiso con el curso.

En el futuro se espera poder ampliar el modelo actual con un mayor número de variables empleando, de ser posible, un tamaño muestral mayor. Sin embargo, estos resultados preliminares confirman el potencial de la teoría SEM como herramienta de análisis del sistema educativo universitario. Otras líneas de trabajo futuro pueden incluir la aplicación de este tipo de códigos en otros problemas en multifísica o ingeniería, e.g. análisis de señales en mediciones de laboratorio para la detección de estructuras turbulentas en flujos de fluidos en régimen turbulento (Leo et al., 2017).

AGRADECIMIENTOS

Este trabajo ha sido financiado por la Universidad Nacional del Litoral (UNL, proyectos CAI+D 2016, códigos: 502-201501-00012-LI y 504-201501-00112-LI). Se agradece al Centro de Investigación de Métodos Computacionales CIMEC (UNL-CONICET), a la cátedra de Teoría de la Computación (FICH-UNL), y al Grupo de Control de Procesos (FIQ-UNL) por el apoyo recibido. Finalmente, este trabajo ha sido parcialmente realizado con recursos de código abierto, tales como \LaTeX y LaTeXDraw.

REFERENCIAS

- Hildreth L. *Residual Analysis for Structural Equation Modeling*. Tesis de Doctorado, Iowa State University, 2013.
- Hox J.J. y Bechger T.M. An introduction to structural equation modeling. *Family Science Review*, 11:357–373, 2009.
- Kline R.B. *Principles and Practice of Structural Equation Modeling*. The Guilford Press, 2011.
- Leo J.M.D., Calandra M., y Delnero J. Algoritmos de punto de cambio aplicados a la detección

- de estructuras vorticosas en flujos turbulentos. *Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería*, 33(3):225–234, 2017.
- Loehlin J.C. *Latent Variable Models*. Lawrence Erlbaum Associates, 2004.
- Organisation for Economic Co-operation and Development. Argentina: Multi-dimensional economic survey. Informe Técnico, 2017.
- Pinto L.P. y Mingoti S.A. On hypothesis tests for covariance matrices under multivariate normality. *Sociedade Brasileira de Pesquisa Operacional*, 35:123–142, 2015.
- Ruiz M.A., Pardo A., y Martín R.S. Modelos de ecuaciones estructurales. *Papeles del Psicólogo*, 21:34–45, 2010.
- Schumacker R.E. y Lomax R.G. *A Beginner's Guide to Structural Equation Modeling*. Lawrence Erlbaum Associates, 2004.