

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTAMENTO DE MATEMÁTICA
VALPARAÍSO-CHILE



Estimación de los efectos recíprocos entre Felicidad y Desempeño Laboral en Chile

Tesis presentada por:

Francisco Antonio Alfaro Medina

*Como requisito para optar al grado de Magister en Ciencias, Mención
Matemática*

Profesor Guía:

Ronny Vallejos Arriagada

Valparaíso, 2019

Índice general

Índice general	I
Índice de figuras	III
Índice de cuadros	IV
1. Introducción	1
1.1. Vínculo entre la felicidad y los resultados del trabajo: una perspectiva floreciente	1
1.2. Felicidad desde la perspectiva hedónica y eudaimónica.	2
1.3. La hipótesis del trabajador productivo feliz	3
1.4. Problemas de los modelos de paneles cruzados	4
2. Preliminares y Métodos	6
2.1. Análisis Factorial	6
2.1.1. Modelos de factores aleatorios no restringidos en la población	6
2.1.2. Máxima verosimilitud de los modelos de factores no restringidos	9
2.1.2.1. El modelo de Lawley	10
2.1.2.2. Modelo de ecuaciones de mínimos cuadrados generalizados	11
2.1.3. Test de Significancia	12
2.1.3.1. Test Chi-cuadrado	13
2.1.4. Criterio de Información	14
2.1.5. Análisis Factorial Confirmatorio	15
2.2. Modelo de ecuaciones estructurales (SEM)	17
2.2.1. Notación del modelo	18
2.2.1.1. Modelo de variable latente	18
2.2.1.2. Modelo de medición	18
2.2.2. Análisis de rutas	19
2.2.2.1. Diagrama de rutas	19
2.2.2.2. Descomposición de covarianzas y correlaciones	20
2.2.2.3. Efectos directos, indirectos y totales	21
2.2.3. Relaciones de causalidad entre las variables	22
2.3. Modelos de panel cruzados (CLPM)	23
2.3.1. Problemas y suposiciones en el análisis de paneles cruzados	25

2.3.1.1.	Problemas con el muestreo en los modelos CLMP . . .	25
2.3.1.2.	Problemas con la interacción de las variables en los modelos CLMP	26
2.3.2.	Enfoques alternativos para los modelos de panel	27
2.3.2.1.	Modelo de trayectorias autorregresivas (ALT) y mo- delo RI-CLPM	28
2.3.2.2.	Modelo de rasgo autorregresivo (STARTS)	29
2.3.2.3.	Modelo del puntaje de cambio latente (LCT)	29
2.3.2.4.	Modelo de estado latente modificado (LST)	30
2.3.2.5.	Resumen de los enfoques alternativos	30
3.	Desarrollo	32
3.1.	Modelos CLPM versus RI-CLPM	32
3.1.1.	Los modelos CLPM	32
3.1.2.	Los modelos RI-CLPM	34
4.	Aplicaciones	39
4.1.	Conjunto de Datos	39
4.1.1.	Medidas del bienestar	39
4.2.	Análisis descriptivo	41
5.	Conclusiones y Trabajos Futuros	46
	Bibliografía	47
	Anexo A	49
A.1.	Lenguaje de Programación R	49
A.1.1.	Lenguaje R y su entorno	49
A.1.2.	Paquetes de modelación de ecuaciones estructurales en R	50
A.1.3.	Tutorial Lavaan	53
	Anexo B	57
B.2.	Rutinas en R	57
	Anexo C	64
C.3.	Cuestionario realizado a los distintos trabajadores	64

Índice de figuras

2.1. Diagrama de rutas para una variable latente con cuatro indicadores. . .	21
2.2. Diagrama de rutas para un modelo con varias variables latentes y varias variables indicadoras.	22
2.3. Relaciones causales entre variables. Formas de covariación	24
2.4. Modelo CLMP de dos olas para las variables X e Y	25
3.1. Box plot para las variables en estudio.	34
3.2. Box plot para las variables en estudio.	38
4.1. Box Plot sobre la puntuación promedio de las distintas medidas de bienestar	42
4.2. Matriz de correlación para el florecimiento y el agotamiento durante los distintos instantes de tiempo.	43
4.3. Matriz de correlación para el florecimiento y el compromiso de trabajo durante los distintos instantes de tiempo.	43
4.4. Matriz de correlación para el florecimiento y la satisfacción laboral durante los distintos instantes de tiempo.	44

Índice de cuadros

2.1. Simbolos principales usados en el análisis de rutas	20
4.1. Test de diferencias chi cuadrado al comparar florecimiento versus agotamiento	45
4.2. chi cuadrado al comparar florecimiento versus satisfacción laboral.	
4.3. chi cuadrado al comparar florecimiento versus compromiso de trabajo.	45
1. Preguntas relacionadas con perma life. . .	65
2. Preguntas relacionadas con agotamiento. .	66
3. Preguntas relacionadas con el compromiso.	67

Capítulo 1

Introducción

1.1. Vínculo entre la felicidad y los resultados del trabajo: una perspectiva floreciente

Durante las últimas décadas, ?? ha habido un consenso cada vez mayor en cuanto a los beneficios de la felicidad, incluida una mejor calidad de vida a nivel individual (por ejemplo, Diener & Tay, 2012; Lyubormirsky, King, & Diener, 2005), también. a nivel comunitario, social y nacional (por ejemplo, OCDE, 2013; Stiglitz et al., 2010; Naciones Unidas, 2011). Los beneficios de la felicidad se extienden a los entornos laborales, lo que sugiere que los trabajadores felices son más creativos y más cooperativos, y también tienen más satisfacción profesional, mejor participación, menos agotamiento y mayor compromiso organizacional (Cameron y Spreitzer, 2011; Erdogan, Bauer, Truxillo, & Mansfield, 2012). Un mayor bienestar también se asocia positivamente con un mayor rendimiento (Erdogan et al., 2012; Wright, Cropanzano y Bonett, 2007) y una productividad y ganancias potencialmente mayores para las empresas (Cropanzano y Wright, 2001; Harter, Schmidt, Asplund, Killham, & Agrawal 2010). Tales hallazgos son de gran interés para las organizaciones. La felicidad en el lugar de trabajo y la participación de los trabajadores se han convertido en palabras de moda, y muchas organizaciones están considerando formas de apoyar el bienestar de los empleados, con la esperanza de que los empleados sean más productivos y estén comprometidos con la organización, lo que a su vez dará como resultado un mayor beneficio para la empresa. Sin embargo, las definiciones de la felicidad en el lugar de trabajo son inconsistentes. La mayoría de las investigaciones que vinculan la felicidad y los resultados laborales se enfocan en una perspectiva hedónica, clasificando la felicidad en el lugar de trabajo

en términos de afecto positivo alto, afecto negativo bajo y alta satisfacción laboral (por ejemplo, Cropanzano y Wright, 2001; Fisher, 2010; Williams, Kern, & Waters, 2015). Los aspectos eudaimónicos, como las relaciones con los compañeros de trabajo, el trabajo significativo y la autonomía, también son importantes (por ejemplo, Marescaux, De Winne y Sels, 2013; Vansteenkiste, Neyrinck, Niemiec, Soenens, Witte y Van de Broeck, 2007; Wrzesniewski, Dutton, & Debebe, 2003). Sugerimos que, en lugar de centrarse en componentes hedónicos o eudaimónicos por separado, los empleados y las organizaciones se benefician de centrarse en el florecimiento, medido en múltiples dimensiones hedónicas y eudaimónicas. En este documento, consideramos que el lugar de trabajo florece específicamente en el contexto de los trabajadores latinoamericanos de Chile (sector financiero) y Paraguay (sector minorista).

1.2. Felicidad desde la perspectiva hedónica y eudaimónica.

La felicidad se ha estudiado y medido principalmente desde una perspectiva hedónica o eudaimónica (Delle Fave et al., 2011; Ryan & Deci, 2001; Waterman, 1993). Hasta la fecha, el enfoque hedónico más basado en la emoción, en el que la felicidad se define en términos de obtener placer y evitar el dolor (Ryan y Deci, 2001) ha recibido mayor atención (Gallagher, Lopez y Preacher, 2009), tal vez porque es más fácil de medir (Kashdan, Biswas-Diener, & King, 2008). Desde esta perspectiva, la felicidad, también llamada bienestar subjetivo (SWB, por sus siglas en inglés), “comprende las evaluaciones de las personas sobre sus vidas, incluidas la felicidad y la satisfacción con la vida” (Diener y Tay, 2012, pág. 1). Quizás la definición más común fue propuesta por Diener (1984), que define SWB en términos de dominios cognitivos (es decir, satisfacción con la vida) y emocionales (es decir, emociones positivas y negativas). De acuerdo con el enfoque eudaimónico, que se basa en el concepto de Aristóteles de la buena vida y en las teorías humanísticas del funcionamiento positivo, la felicidad incluye elementos como las relaciones interpersonales, el crecimiento personal y el compromiso (Delle Fave et al., 2011; Ryff & Keyes, 1995). El bienestar eudaimónico refleja la actualización de los potenciales humanos, el significado y la autorrealización (Ryan y Deci, 2001). Varias teorías definen y miden el bienestar desde esta perspectiva. Por ejemplo, el modelo de Ryff (1989) consta de seis componentes: autonomía, dominio ambiental, crecimiento personal, relaciones positivas con los demás, propósito en la vida y autoaceptación. La teoría de la autodeterminación (SDT; Deci y Ryan,

2000) sugiere que los seres humanos tienen tres necesidades básicas: autonomía, competencia y afinidad, que son nutrientes psicológicos necesarios para el funcionamiento óptimo, el crecimiento saludable y el bienestar (Reis, Sheldon, Gable, Roscoe y Ryan, 2000; Unanue, Vignoles, Dittmar y Vansteenkiste, 2014). Waterman (2008) define el bienestar eudaimónico en términos de significado, autorrealización y actualización de los potenciales humanos. Al juntar estas perspectivas, hay un impulso creciente en el campo de la psicología positiva para discutir el bienestar en términos de una construcción multidimensional más holística de florecimiento (Henderson y Knight, 2012). También existen múltiples definiciones de florecimiento.

1.3. La hipótesis del trabajador productivo feliz

Gran parte del trabajo sobre la felicidad en el lugar de trabajo, ya sea desde una perspectiva hedónica o eudaimónica, se basa en la feliz hipótesis del trabajador productivo (Harter et al., 2003; Jones, 2006; Wright et al., 2007). Desde una perspectiva organizativa, la idea de que los trabajadores serán más productivos y, por lo tanto, de aumentar las ganancias, es la única razón para centrarse en el bienestar de los empleados. Sin embargo, la asociación entre felicidad y productividad aún no es concluyente (Cropanzano y Wright, 2001; Peiró, Ayala, Tordera, Lorente y Rodríguez, 2014; Wright y Staw, 1999). Algunas investigaciones encuentran que un mayor bienestar está asociado positivamente con un mayor rendimiento (Erdogan et al., 2012; Wright et al., 2007), mientras que otros estudios sugieren que la felicidad puede estar asociada negativamente a la productividad laboral (Baron, Hmieleski y Henry, 2012; Wright & Staw, 1999), con trabajadores felices e improductivos (Peiró et al., 2014). En niveles muy altos, el efecto positivo puede interferir con aspectos específicos de la percepción, la motivación, la cognición y la autorregulación que podrían ser potencialmente perjudiciales para el desempeño de las personas (Baron et al., 2012). Esto sugiere que podría haber una asociación curvilínea entre afecto positivo y productividad (Wright y Staw, 1999). Sin embargo, la investigación hasta la fecha ha probado asociaciones lineales entre felicidad y productividad, sin considerar asociaciones cuadráticas o límites superiores a los beneficios de la felicidad. Hasta donde sabemos, aunque se ha sugerido esta posibilidad, debe ser probada empíricamente (Baron et al., 2012).

1.4. Problemas de los modelos de paneles cruzados

En 1980, se publicó el artículo seminal de Rogosa Una crítica de la correlación cruzada, que transmitió con éxito el mensaje de que la comparación de las correlaciones cruzadas entre los datos del panel longitudinal es una base inadecuada para hacer inferencias causales.¹ Una de las ideas clave derivadas de Rogosa El artículo es que, si dos constructos se caracterizan por diferentes grados de estabilidad, la comparación de las correlaciones cruzadas puede llevar a conclusiones falsas con respecto al mecanismo causal. Desde entonces, la mayoría de los investigadores interesados en la causalidad en los datos de panel han abandonado las correlaciones cruzadas y respaldaron el modelo de panel cruzado (CLPM), también conocido como el modelo de ruta cruzada o el modelo de regresión cruzada. En el CLPM, la estabilidad de los constructos se controla mediante la inclusión de relaciones autorregresivas y, por lo tanto, a menudo se cree que los parámetros de regresión cruzada obtenidos con este modelo son las medidas más apropiadas para estudiar la causalidad en datos correlacionales longitudinales (p. Ej. Deary, Allerhand, y Der, 2009; Soenens, Luyckx, Vansteekiste, Duriez y Goossens, 2008; Wood, Maltby, Gillett, Linley y Joseph, 2008). Específicamente, es una práctica común estandarizar los coeficientes de regresión cruzada y comparar su fuerza relativa para determinar qué variable tiene una influencia causal más fuerte en el otro (Bentler y Speckart, 1981). El presente artículo tiene como objetivo continuar con la advertencia de Rogosa (1980), en el sentido de que se argumentará que no solo debemos dar cuenta de la estabilidad, sino que también debemos dar cuenta del tipo correcto de estabilidad. Además, presentamos un enfoque alternativo de modelado de ecuación estructural (SEM) estrechamente relacionado que está inspirado en considerar datos de paneles cruzados desde una perspectiva multinivel, lo que implica que necesitamos distinguir entre el nivel intrapersonal y el nivel interpersonal. La estructura de trabajo es la siguiente: En el Capítulo 2 se presenta una serie de definiciones y resultados que permiten definir los modelos de factores y su relación con los modelos SEM ocupado en la literatura. Luego, se definen los modelos de panel cruzado con retraso (modelos CLPM), en donde mostramos las principales desventajas de estos a nivel de muestreo y de interacción de variables. Este capítulo finaliza presentando algunas alternativas a los modelos CLPM tradicionales. El Capítulo 3 consiste en una comparación más profunda de modelo CLPM tradicional y los modelos RI-CLPM. En el Capítulo 4 se presenta la metodología de investigación, en donde se describe detalladamente el conjunto de datos, a modo de comprender la naturaleza de nuestras variables. Luego, presentamos estadística descriptiva sobre las variables, mostrando

algunos gráficos de interés, Para finalizar se ajustan el modelo CLPM tradicional versus el modelo RI-CLPM, comparando los resultados del ajustes (a nivel de modelos y entre modelos). Finalmente, en el Capítulo 5 se concluye el presente trabajo con algunos comentarios de los resultados obtenidos y se presenta una propuesta para continuar la investigación a modo de trabajos futuros.

Capítulo 2

Preliminares y Métodos

2.1. Análisis Factorial

2.1.1. Modelos de factores aleatorios no restringidos en la población

Sea $\mathbf{X} = (y_1, \dots, y_p)^\top$ es un vector de variables aleatorias observadas, $\mathbf{\Phi} = (\phi_1, \dots, \phi_r)^\top$ es el vector de de $r < p$ variables aleatorias no observadas o variables latentes llamados factores, $\boldsymbol{\alpha}$ es la matriz $(r \times p)$ matriz de coeficientes fijos (cargas), y $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_p)^\top$ es el vector de error aleatorio. Se define el modelo de factores aleatorios por:

$$\mathbf{X} = \boldsymbol{\alpha}\mathbf{\Phi} + \boldsymbol{\epsilon} \quad (2.1)$$

Por conveniencia, suponemos que $\mathbb{E}(\mathbf{X}) = \mathbb{E}(\mathbf{\Phi}) = \mathbb{E}(\boldsymbol{\epsilon}) = 0$. El termino de error aleatorio consiste en errores de medición, junto con efectos individuales asociados con cada variable de la población x_i . Cuando una muestra aleatoria es tomada, el error residual también contiene variación muestral. Para el modelo (2.1) asumimos que $\boldsymbol{\alpha}$ es una matriz de parámetros constantes y $\mathbf{\Phi}$ es un vector de variables aleatorias. El modelo (2.1) se asemeja a una ecuación de regresión excepto por los factores $\mathbf{\Phi}$ que no son observados directamente y deben ser estimados desde los datos junto con los parámetros. Esto introduce dificultades de identificación que no existen en otros modelos estadísticos, por ejemplo, el modelo clásico de regresión de una sola ecuación.

Para el modelo (2.1) realizamos los siguientes supuestos adicionales.

$$\begin{aligned}
& \text{(i)} \quad \rho(\boldsymbol{\alpha}) = r < p \\
& \text{(ii)} \quad \mathbb{E}(\mathbf{X}|\boldsymbol{\Phi}) = \boldsymbol{\alpha}\boldsymbol{\Phi} \\
& \text{(iii)} \quad \mathbb{E}(\mathbf{X}\mathbf{X}^\top) = \boldsymbol{\Sigma}, \mathbb{E}(\boldsymbol{\Phi}\boldsymbol{\Phi}^\top) = \boldsymbol{\Omega}, \text{ y } \mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top) = \boldsymbol{\Psi} = \text{diag}(\sigma_1^2, \dots, \sigma_p^2) \\
& \text{(iv)} \quad \mathbb{E}(\boldsymbol{\Phi}\boldsymbol{\epsilon}^\top) = 0.
\end{aligned} \tag{2.2}$$

Considerando la ecuación (2.1) y los supuestos (i)-(iv), tenemos que

$$\begin{aligned}
\mathbb{E}(\mathbf{X}\mathbf{X}^\top) &= \boldsymbol{\Sigma} = \mathbb{E}(\boldsymbol{\alpha}\boldsymbol{\Phi} + \boldsymbol{\epsilon})(\boldsymbol{\alpha}\boldsymbol{\Phi} + \boldsymbol{\epsilon})^\top \\
&= \mathbb{E}(\boldsymbol{\alpha}\boldsymbol{\Phi}\boldsymbol{\Phi}^\top\boldsymbol{\alpha}^\top + \boldsymbol{\alpha}\boldsymbol{\Phi}\boldsymbol{\epsilon}^\top + \boldsymbol{\epsilon}\boldsymbol{\Phi}^\top\boldsymbol{\alpha}^\top + \boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top) \\
&= \boldsymbol{\alpha}\mathbb{E}(\boldsymbol{\Phi}\boldsymbol{\Phi}^\top)\boldsymbol{\alpha}^\top + \boldsymbol{\alpha}\mathbb{E}(\boldsymbol{\Phi}\boldsymbol{\epsilon}^\top) + \mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\Phi}^\top\boldsymbol{\alpha}^\top) + \mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top) \\
&= \boldsymbol{\alpha}\boldsymbol{\Omega}\boldsymbol{\alpha}^\top + \mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top) \\
&= \boldsymbol{\Gamma} + \boldsymbol{\Psi}.
\end{aligned} \tag{2.3}$$

donde $\boldsymbol{\Gamma} = \boldsymbol{\alpha}\boldsymbol{\Omega}\boldsymbol{\alpha}^\top$ y $\boldsymbol{\Psi} = \mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top)$, son las matrices de varianza real y el error, respectivamente. También, post multiplicando la ecuación (2.1) por $\boldsymbol{\Phi}^\top$, tenemos que

$$\mathbb{E}(\mathbf{X}\boldsymbol{\Phi}^\top) = \mathbb{E}(\boldsymbol{\alpha}\boldsymbol{\Phi}\boldsymbol{\Phi}^\top + \boldsymbol{\epsilon}\boldsymbol{\Phi}^\top) = \boldsymbol{\alpha}\boldsymbol{\Omega}. \tag{2.4}$$

Un caso importante ocurre cuando \mathbf{X} es normal multivariante, en este caso los segundos momentos de la ecuación (2.3) contiene toda la información sobre el modelo factorial. También se sigue que el modelo factorial (2.1) es lineal, y que las variables son condicionalmente independientes de los factores $\boldsymbol{\Phi}$. Sea $\boldsymbol{\Phi} \sim \mathcal{N}(0, 1)$, entonces la distribución condicional de \mathbf{X} viene dada por

$$\begin{aligned}
\mathbf{X}|\boldsymbol{\Phi} &\sim \mathcal{N}(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Phi}, (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^\top)) \\
&\sim \mathcal{N}(\boldsymbol{\alpha}\boldsymbol{\Phi}, (\boldsymbol{\Sigma} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\top)) \\
&\sim \mathcal{N}(\boldsymbol{\alpha}\boldsymbol{\Phi}, \boldsymbol{\Psi}).
\end{aligned} \tag{2.5}$$

La independencia condicional se sigue de la diagonal de $\boldsymbol{\Psi}$. Por lo tanto, los factores comunes de $\boldsymbol{\Phi}$ reproducen toda la covarianza (correlaciones) entre las variables, pero representan una parte de las covarianzas. Es posible derivar los estimadores ML de $\boldsymbol{\alpha}, \boldsymbol{\Phi}$ y $\boldsymbol{\Psi}$ los cuales tienen las ventajas de ser asintóticamente eficientes, los cuales sirven para realizar test de hipótesis y construir intervalos de confianzas para las estimaciones ML de $\boldsymbol{\alpha}$. Resulta que siempre es posible seleccionar coeficientes de $\boldsymbol{\alpha}$

de manera tal que represente correlación entre \mathbf{X} y Φ (asumiendo que los factores existen) si Σ representa la matriz de varianzas o correlaciones. Finalmente, a partir de la distribución de \mathbf{X} , el modelo de factor (2.1) es asumido lineal en el coeficiente α , pero no necesariamente en los factores Φ . A diferencia del modelo de regresión, la variable independiente Φ , no se conoce, por lo tanto la estimación inicial del modelo se basa en (2.3) en lugar de (2.1) asumiendo que Σ es conocida (o estimada). Esto normalmente se logra establecer si $\Omega = \mathbf{I}$ y la estimación de α y Ψ es conjunta, a menudo por métodos iterativos. Cuando una muestra finita es válida, el valor de los factores Φ puede ser estimado por los procedimientos presentados más adelante. Dado que los factores no se observan directamente surge la pregunta si los factores son reales. Claramente, la interpretabilidad de los factores no se puede para todos los casos o aplicaciones, y la identificación de los factores debe permanecer, esencialmente, como una idea sustantiva según la naturaleza de la aplicación, tipo de variables y datos utilizados. Supongamos que existen $1 < r < p$ factores comunes tales que $\Gamma = \alpha\Omega\alpha^\top$ y Ψ es una de Gram diagonal, con elementos en la diagonal $\sigma_i^2 > 0$ ($i = 1, \dots, p$), entonces la matriz de covarianza Σ tiene $\binom{p}{2} + p = p(p+1)/2$ elementos distintos, los cuales son iguales al número de ecuaciones normales a resolver. El número de soluciones, sin embargo, es infinita, lo cual podemos observar a continuación. Sea Ω una matriz definida positiva, entonces debe existir una matriz \mathbf{B} de tamaño $r \times r$ no singular tal que $\Omega = \mathbf{B}^\top \mathbf{B}$, entonces

$$\begin{aligned}\Sigma &= \alpha\Omega\alpha^\top + \Psi \\ &= \alpha(\mathbf{B}^\top \mathbf{B})\alpha^\top + \Psi \\ &= (\alpha\mathbf{B}^\top)(\alpha\mathbf{B}^\top)^\top + \Psi \\ &= \alpha^* \alpha^{*\top} + \Psi.\end{aligned}\tag{2.6}$$

Tanto (2.4) como (2.6), representan factorizaciones de la matriz Σ que dejan el mismo error residual Ψ y por tanto deben representar soluciones igualmente válidas. Bajo esta aseveración se concluye que α y α^* son estadísticamente equivalentes y no se puede distinguir entre si o identificar de manera única. Se debe poner más restricciones sobre Ψ , la matriz de covarianza de los factores. La restricción más usual es considerar $\Psi = \mathbf{I}$, así definimos los factores como vectores unitarios. El número de m parámetros libres implicados en la ecuación (2.6) es igual a $pr + p$ parámetros desconocidos de α y Ψ , menos el número de restricciones impuestas sobre los elementos fuera de la diagonal de Ω , lo cual es igual a $(r^2 - r)/2$, donde Ω es simétrica (asumiendo r factores

comunes). Entonces, tenemos que

$$m = (pr + p) - 1/2(r^2 - r) = p(r + 1) - 1/2r(r - 1). \quad (2.7)$$

donde las columnas de α se asumen ortogonales. El número de grados de libertad df es un número estrictamente positivo, el cual está dado por el número de ecuaciones implicadas por (??), esto es, el número de elementos distintos en Σ menos el número de parámetros m , es decir,

$$\begin{aligned} df &= 1/2p(p + 1) - (p(r + 1) - 1/2r(r - 1)) \\ &= 1/2(p - r)^2 - (p + r). \end{aligned} \quad (2.8)$$

Esto impone un límite superior en el número de factores comunes r que pueden ser observados en la práctica, que es un número más pequeño que la variable p . Además, en la ecuación (2.8), suponemos que las ecuaciones son linealmente independientes, esto es, Σ debe ser no singular y $\rho(\alpha) = r$.

2.1.2. Máxima verosimilitud de los modelos de factores no restringidos

Cuando la población es normal multivariada, se puede ocupar el principio de máxima verosimilitud (ML) para derivar las ecuaciones normales. La ventaja de los estimadores ML es que son eficientes y consistentes, permitiendo pruebas estadísticas sobre los parámetros. Las primeras contribuciones significativas en esta área se deben a Lawley (1963,1966), y a pesar de las dificultades de la convergencia numérica, la estimación ML es ampliamente utilizada. Actualmente, varias especificaciones sobre la máxima verosimilitud en el análisis de factores han sido propuestas, conduciendo a procedimientos de estimación ligeramente diferentes a las estimaciones de máxima verosimilitud, sin embargo, solo se puede ocupar para modelos de factores aleatorios, ya que cuando los factores son fijos, estrictamente hablando los estimadores ML no existen (ver, Solari, 1969). Esta clase de modelos es conocida como máxima verosimilitud en análisis factorial no restringido, para diferenciarlo de los modelos de factores restringidos o confirmatorios (ver Lawley & Maxwell, 1971).

2.1.2.1. El modelo de Lawley

El modelo de factor ML originalmente fue presentado por Lawley (1940,1941), es este modelo la matriz de dispersión es invariante respecto al cambio de escala, así que la matriz de correlación, la matriz de covarianza y $\mathbf{X}^\top \mathbf{X}$ producen coeficientes de cargas idénticos. Sea $\mathbf{X} = (x_1, \dots, x_p) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces usando la distribución Wishart, la log-verosimilitud puede ser expresada por

$$L(\boldsymbol{\Sigma}) = \ln(c) - \frac{n}{2} \ln|\boldsymbol{\Sigma}| + \frac{1}{2}(n - p - 1) \ln|\mathbf{S}| + \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}). \quad (2.9)$$

Omitiendo funciones constantes, la ecuación (2.9) puede ser expresada por

$$\begin{aligned} L(\boldsymbol{\Sigma}) &= \frac{n}{2} (\ln|\boldsymbol{\Sigma}| + \text{tr}(\mathbf{S} \boldsymbol{\Sigma}^{-1})) \\ &= \frac{n}{2} [\ln|\boldsymbol{\alpha} \boldsymbol{\alpha}^\top + \boldsymbol{\Psi}| + \text{tr}(\mathbf{S}(\boldsymbol{\alpha} \boldsymbol{\alpha}^\top + \boldsymbol{\Psi})^{-1})]. \end{aligned} \quad (2.10)$$

La expresión anterior depende solo de $\boldsymbol{\alpha}$ y $\boldsymbol{\Psi}$, entonces

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\Psi}} &= \frac{\partial \ln|\boldsymbol{\alpha} \boldsymbol{\alpha}^\top + \boldsymbol{\Psi}|}{\partial \boldsymbol{\Psi}} + \frac{\partial [\text{tr}(\mathbf{S}(\boldsymbol{\alpha} \boldsymbol{\alpha}^\top + \boldsymbol{\Psi})^{-1})]}{\partial \boldsymbol{\Psi}} \\ &= \text{diag}(\boldsymbol{\alpha} \boldsymbol{\alpha}^\top + \boldsymbol{\Psi})^{-1} - (\boldsymbol{\alpha} \boldsymbol{\alpha}^\top + \boldsymbol{\Psi})^{-1} \mathbf{S} (\boldsymbol{\alpha} \boldsymbol{\alpha}^\top + \boldsymbol{\Psi})^{-1} \frac{\partial \boldsymbol{\Psi}}{\partial \boldsymbol{\Psi}} \\ &= \text{diag}(\boldsymbol{\alpha} \boldsymbol{\alpha}^\top + \boldsymbol{\Psi})^{-1} - (\boldsymbol{\alpha} \boldsymbol{\alpha}^\top + \boldsymbol{\Psi})^{-1} \mathbf{S} (\boldsymbol{\alpha} \boldsymbol{\alpha}^\top + \boldsymbol{\Psi})^{-1} \\ &= 0. \end{aligned} \quad (2.11)$$

lo anterior equivale a decir que

$$\text{diag}(\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\Sigma}} - \mathbf{S})\hat{\boldsymbol{\Sigma}}^{-1}) = 0. \quad (2.12)$$

donde $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\alpha}} \hat{\boldsymbol{\alpha}}^\top + \hat{\boldsymbol{\Psi}}$. Por lo tanto, la ecuación normal (2.12) es equivalente a

$$\text{diag}(\hat{\boldsymbol{\Sigma}}) = \text{diag}(\hat{\mathbf{S}}). \quad (2.13)$$

Lo condición anterior significa que las varianzas muestrales reproducidas por los estimadores ML $\hat{\boldsymbol{\alpha}}$ y $\hat{\boldsymbol{\Psi}}$ son iguales a la varianzas muestrales observadas en la muestra.

Ahora, diferenciado la ecuación (2.10) con respecto a α tenemos que

$$\begin{aligned}\frac{\partial L}{\partial \alpha} &= \frac{\partial \ln|\alpha\alpha^\top + \Psi|}{\partial \alpha} + \frac{\partial[\text{tr}(\mathbf{S}(\alpha\alpha^\top + \Psi)^{-1})]}{\partial \alpha} \\ &= -\Sigma^{-1}\alpha + \Sigma^{-1}\mathbf{S}\Sigma^{-1}\alpha \\ &= 0,\end{aligned}\tag{2.14}$$

luego, la ecuación normal se puede escribir como

$$(\hat{\Sigma} - \mathbf{S})\hat{\Sigma}^{-1}\hat{\alpha}.\tag{2.15}$$

Las ecuaciones (2.12) y (2.15) constituyen las ecuaciones normales para el modelo factorial ML de Lawley. Donde (2.15) no tiene solución única, se requieren condiciones adicionales para que el modelo sea identificable. Las restricciones que utiliza Lawley especifican que $\alpha^\top \Psi^{-1} \alpha = \eta$ debe ser diagonal, y esto fijaría las bases del espacio solución, arrojando una correlación única para las cargas α .

2.1.2.2. Modelo de ecuaciones de mínimos cuadrados generalizados

Sea \mathbf{S}^{-1} la matriz de pesos (Joreskog & Goldberger, 1972; Joreskog, 1977). El modelo de mínimos cuadrados se puede generalizar para minimizar

$$\begin{aligned}\mathbf{G} &= \text{tr}[\mathbf{S}^{-1}(\mathbf{S} - \Sigma)\mathbf{S}^{-1}(\mathbf{S} - \Sigma)] \\ &= \text{tr}[\mathbf{S}^{-1}(\mathbf{S} - \Sigma)]^2 \\ &= \text{tr}(\mathbf{I} - \mathbf{S}^{-1}\Sigma)^2.\end{aligned}\tag{2.16}$$

La razón de ser del criterio expuesto se asemeja al criterio por regresión de mínimos cuadrados generalizados para términos de error heteroscedástico. Tenemos que el diferencial total es

$$\begin{aligned}d\mathbf{G} &= d[\text{tr}(\mathbf{I} - \mathbf{S}^{-1}\Sigma)^2] \\ &= \text{tr}[d(\mathbf{I} - \mathbf{S}^{-1}\Sigma)^2] \\ &= 2\text{tr}[(\mathbf{I} - \mathbf{S}^{-1}\Sigma)d(\mathbf{I} - \mathbf{S}^{-1}\Sigma)] \\ &= 2\text{tr}[(\mathbf{I} - \mathbf{S}^{-1}\Sigma)\mathbf{S}^{-1}d\Sigma].\end{aligned}$$

Fijando Ψ , tenemos que

$$d\Sigma = d\alpha\alpha^\top + \Psi = 2\alpha(d\alpha^\top).$$

Por lo tanto,

$$d\mathbf{G} = 4\text{tr}[(I - \mathbf{S}^{-1}\Sigma)\mathbf{S}^{-1}d2\alpha(d\alpha^\top)].$$

Si derivamos \mathbf{G} respecto a α tenemos que

$$\frac{\partial \mathbf{G}}{\partial \alpha} = 4\mathbf{S}^{-1}(\Sigma - \mathbf{S})\mathbf{S}^{-1}\alpha = 0$$

por lo tanto, la ecuación normal nos queda

$$\hat{\Sigma}\mathbf{S}^{-1}\hat{\alpha} = \hat{\alpha}. \quad (2.17)$$

Premultiplicando por $\hat{\Sigma}^{-1}$ y reorganizando algunos términos, obtenemos una expresión equivalente a las ecuaciones normales en el modelo de Lawley.

2.1.3. Test de Significancia

Cuando $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, la hipótesis de los r factores comunes puede ser testada usando teoría asintótica, esto es, estudiar el test de hipótesis:

$$\begin{aligned} H_0 : \Sigma &= \alpha\alpha^\top + \Psi, \text{ versus} \\ H_a : \Sigma &\neq \alpha\alpha^\top + \Psi, \end{aligned} \quad (2.18)$$

para determinar si Σ contiene $r > 0$ factores comunes (más una matriz de varianza de error diagonal) versus la alternativa de que Σ es una matriz de covarianza arbitraria.

2.1.3.1. Test Chi-cuadrado

El test clásico para verificar la existencia de r factores comunes es el test chi-cuadrado. Primero, considere el tes de independencia completa de las variables (Lawley & Maxwell, 1971). Bajo H_0 (ecuación (2.18)) la función de verosimilitud es:

$$\begin{aligned} L(\omega) &= c|\hat{\Sigma}|^{-n/2} \exp\left[-\frac{n}{2}(\hat{\Sigma}^{-1}\mathbf{S})\right] \\ &= c|\hat{\alpha}\hat{\alpha}^\top + \hat{\Psi}|^{-n/2} \exp[(tr(\hat{\alpha}\hat{\alpha}^\top + \hat{\Psi})\mathbf{S})], \end{aligned} \quad (2.19)$$

y bajo H_a (ecuación (2.18)) tenemos que,

$$\begin{aligned} L(\Omega) &= c|\mathbf{S}|^{-n/2} \exp\left[-\frac{n}{2}(\mathbf{S}^{-1}\mathbf{S})\right] \\ &= c|\mathbf{S}|^{-n/2} \exp\left(-\frac{np}{2}\right), \end{aligned} \quad (2.20)$$

donde bajo H_a se tiene que : $\hat{\Sigma} = \mathbf{S}$. El cuociente de verosimilitud para testear H_0 es $\lambda = L(\omega)/L(\Omega)$, donde $-2\ln(\lambda)$ es asintóticamente chi-cuadrado. Para n suficientemente grande, tenemos que

$$\begin{aligned} \chi^2 &= -2\ln(\lambda) \\ &= -2\ln L(\omega) + 2\ln L(\Omega) \\ &= n[\ln|\hat{\Sigma}| + tr(\mathbf{S}\hat{\Sigma}^{-1}) - \ln|\mathbf{S}| - p]. \end{aligned}$$

De la ecuación (2.13), $\text{diag}\hat{\Sigma} = \text{diag}\mathbf{S}$, por lo cual el criterio se reduce a

$$\begin{aligned} \chi^2 &= n[\ln|\hat{\Sigma}| + p - \ln|\mathbf{S}| - p] \\ &= n \ln\left(\frac{|\hat{\Sigma}|}{|\mathbf{S}|}\right). \end{aligned} \quad (2.21)$$

Donde el test de completa independencia es equivalente a testear si $r = 0$, de la ecuación (2.21) para grandes muestras se tiene que

$$\chi^2 = n \ln\left(\frac{|\hat{\Sigma}|}{|\mathbf{S}|}\right) = n \ln|\mathbf{R}|, \quad (2.22)$$

donde al ser $r = 0$ tenemos que $\Sigma = \text{diag}(\mathbf{S})$, y el cuociente de determinantes es igual a $|\mathbf{R}|$, el determinante de la matriz de correlación. Para un nivel de significancia $\alpha = 0.05$, la regla de desición rechaza H_0 si $p_{\text{value}} < 0.05$. El estadístico de la ecuación (2.22) es válido para grandes tamaños de n . Para tamaños muestrales pequeños el

estadístico chi-cuadrado puede ser mejorado si n es reemplazado por $(n-1) - 1/6(2p+5)$ (Box, 1949) para $df = (1/2)[(p-r)^2 - (p+r)]$ grados de libertad, la diferencia entre el número de parámetros en Σ y el número de restricciones lineales impuestas por la hipótesis nula. En la práctica el test de razón de verosimilitud a menudo indicará un número mayor de factores que pueden ser interpretados en un entorno significativo, y a veces es mejor llevar una rotación de las cargas antes de decidir sobre el valor de r . En cualquier caso, el número de factores no puede exceder tal número para el cual df no es positivo.

2.1.4. Criterio de Información

Los criterios de información son ocupados para determinar el número correcto u óptimo de factores. De hecho el estadístico AIC fue inicialmente desarrollado para series de tiempo y modelos de factores ML (Akaike, 1974; 1987). La idea principal es usar funciones que sean similares al estadístico C_p de Mallow, el cual penaliza una cantidad excesiva de parámetros ajustado, algo que el test chi-cuadrado no hace. A continuación, basamos nuestra discusión en Bozdogan (1987). Considere un vector aleatorio absolutamente continuo \mathbf{X} , el cual tiene una densidad $f(\mathbf{X}|\boldsymbol{\theta})$ y donde $\boldsymbol{\theta}$ es algún vector de parámetros. Si existe un vector de parámetros verdadero $\boldsymbol{\theta}^*$, entonces la medida de bondad de ajuste utilizada es la cantidad de información general de Boltzmann

$$\mathbf{B} = \mathbb{E}[\ln f(\mathbf{X}|\boldsymbol{\theta}) - \ln f(\mathbf{X}|\boldsymbol{\theta}^*)] = -\mathbf{I}, \quad (2.23)$$

también conocido en la literatura como el criterio de información Kullback-Leibler. Bajo el principio de maximización de la entropía deseamos estimar $f(\mathbf{X}|\boldsymbol{\theta}^*)$ por medio de $f(\mathbf{X}|\boldsymbol{\theta})$ tal que la entropía esperada

$$\begin{aligned} \mathbb{E}_x(\mathbf{B}) &= \mathbb{E}_x\{\mathbb{E}[\ln f(\mathbf{X}|\boldsymbol{\theta})] - \mathbb{E}[\ln f(\mathbf{X}|\boldsymbol{\theta}^*)]\} \\ &= \mathbb{E}_x\{\mathbb{E}[\ln f(\mathbf{X}|\boldsymbol{\theta})]\}, \end{aligned} \quad (2.24)$$

sea maximizada. El valor esperado de la ecuación (2.23) puede ser ocupado como una función de riesgo para medir el error de estimación promedio del modelo ajustado. Por lo tanto el criterio AIC es entonces un estimador muestral de $\mathbb{E}[\ln f(\mathbf{X}|\boldsymbol{\theta})]$, el valor esperado de la log verosimilitud, y esta dado por la expresión general

$$\text{AIC}(r) = -2\ln L(r) + 2m, \quad (2.25)$$

donde m es el número de parámetros libres después de ser ajustado el modelo. y $L(r)$ es la verosimilitud. La ecuación (2.25) fue originalmente desarrollada por Akaike(1974a) en el contexto de series de tiempo, pero pudo ser expresada de igual manera para el máxima verosimilitud de análisis de factores (Akaike,1987). Donde para r factores comunes tenemos $L(r) = (n/2) \sum_{i=r+1}^p \ln \hat{\theta}_i$ y el número de parámetros libres $m = p(r+1) - (1/2)r(r-1)$, por lo tanto la ecuación (2.25) puede ser expresada por

$$AIC(r) = -2\left(\frac{n}{2} \sum_{i=r+1}^p \ln \hat{\theta}_i\right) + [2p(r+1) - r(r-1)], \quad (2.26)$$

donde $\hat{\theta}_{r+1}, \hat{\theta}_{r+2}, \dots, \hat{\theta}_p$ representan las raíces latentes más pequeñas. La idea de usar la ecuación (2.26) es variar el número de factores comunes, comenzando con $r = 1$, y escoger el valor de r el cual minimice el $AIC(r)$.

2.1.5. Análisis Factorial Confirmatorio

Los modelos de factores descritos anteriormente se conocen como modelos de factores exploratorios, ya que aparte de una especificación de r (o Ψ) no se requiere otra información a priori. El objetivo del análisis de factores no difiere mucho con el de componentes principales. Al usar una estructura de error más general, el análisis exploratorio busca obtener una mayor comprensión de la estructura multivariante de los datos. A veces, sin embargo, el investigador pudo haber llevado a cabo un análisis factorial de un conjunto de variables de una muestra diferente, en cuyo caso la información disponible se puede ocupar para más muestras. Por lo tanto, es posible que ya se conozcan la cantidad de factores comunes o que ya hayan probado los coeficientes de carga y se haya determinado que algunos son insignificantes diferentes de cero. Si se toman muestras sucesivas, sería un desperdicio ignorar dicha información previa, y entonces podemos desear imponer cero restricciones en las cargas. También se pueden usar restricciones distintas de cero, pero los ceros son más comunes en la práctica. Sea Ω la matriz de correlación de los factores. El número total de parámetros estimados en α , Ω y Ψ es

$$(pr + \frac{1}{2}r)(r+1) + p = \frac{1}{2}(2p+r)(r+1). \quad (2.27)$$

Sea n_α, n_Ω y n_Ψ el número de parámetros en α , Ω y Ψ respectivamente, donde $m = n_\alpha + n_\Omega + n_\Psi$. Entonces el número de parámetros no restringidos es

$$\frac{1}{2}(2p+r)(r+1) - m, \quad (2.28)$$

el cual debe ser un número más pequeño que $p(p+1)/2$, el número de elementos distintos de Σ , para reproducir una solución no trivial. Así

$$r^2 + m > \frac{1}{2}(p+r)(p+r+1), \quad (2.29)$$

y una condición necesaria para determinar de manera única los parámetros de factores es que

$$n_\alpha + n_\Omega \geq r^2. \quad (2.30)$$

Sin embargo las condiciones de suficiencia son difíciles de definir, ya que no se debe tener en cuenta solo el número sino también la posición de los parámetros con restricción cero (Joreskog & Lawley, 1968). También se pueden tener restricciones igual a cero en los elementos fuera de la diagonal de Ω , lo cual resultaría un modelo de factores oblicuos/ortogonales mixto. Claramente, si se usa un número suficiente de restricciones, se vuelve innecesario rotar los factores. Los elementos diagonales de Ψ pueden ser pre-especificados, pero esto es menos habitual en la práctica (ver más detalles en Lawley & Maxwell (1971)). Al igual que para el modelo exploratorio, la máxima verosimilitud en los modelos confirmatorio se calculan de manera iterativa, pero Bentler (1982) propuso un modelo computacional no iterativo para tal modelo. Sin embargo, una dificultad con el análisis factorial confirmatorio, como con cualquier otro modelo que incorpore restricciones a priori, es la introducción de sesgo en las estimaciones cuando se usan restricciones incorrectas. A veces, las restricciones cero surgen naturalmente en circunstancias específicas. Así Anderson (1976), por ejemplo, ha desarrollado un modelo de factor útil para confirmar ensayos clínicos, que utiliza una matriz triangular de bloques de cargas factoriales para lograr una rotación única de los ejes.

2.2. Modelo de ecuaciones estructurales (SEM)

El concepto de modelos de ecuaciones estructurales (SEM) fue introducido por Sewall Wrigth durante los años 1921-1934 en un trabajo sobre las características de las co-bayas. En este trabajo desarrolló herramientas para establecer las relaciones causales de las variables representadas en un diagrama de rutas (del inglés path diagram) por lo que su método se llegó a conocer como análisis de rutas (del inglés path analysis) [Wrigth 1921, 1934]. Años más tarde, este método fue re-descubierto y desarrollado por economistas y sociólogos, entre los que se destacan los trabajos de Jöreskog (1973, 1977 y 1981) y Jöreskog & Sörbom (1982). En ellos, transformaron el análisis de rutas de Wright en un nuevo método denominado modelos de ecuaciones estructurales (del inglés structural equation modelling), en el que se combinan técnicas del análisis factorial con el análisis de rutas. El objetivo de los modelos de ecuaciones estructurales es estudiar las relaciones causales entre datos directamente observables, asumiendo la existencia de relaciones lineales, aunque también se pueden establecer modelos con relaciones no lineales. La ventaja de este tipo de modelos es que permiten proponer el tipo y dirección de las relaciones que se espera encontrar entre las diversas variables contenidas en él, para pasar posteriormente a estimar los parámetros que vienen especificados por las relaciones propuestas a nivel teórico. También, permiten la comparación de modelos alternativos (competidores). Los puntos fuertes de estos modelos son: el desarrollo de convenciones que permiten su representación gráfica, la posibilidad de analizar efectos causales entre las variables, permitir la concatenación de efectos entre variables y permitir relaciones recíprocas entre ellas. La aplicación de este tipo de modelos requiere de un diseño a priori, para el cual el investigador se apoya en la teoría de aquello que busca explicar. Este diseño a priori se conoce como “modelo teórico”, este consistente en un conjunto sistemático de relaciones (entre variables) que proporcionan una explicación consistente y comprensiva del fenómeno que se pretende estudiar. Dichas relaciones están definidas por una serie de ecuaciones que describen las estructuras de las relaciones establecidas (relaciones causales, de regresión con respuesta múltiple, correlaciones, etc.). Este modelo teórico se puede representar bien a través del conjunto de ecuaciones que lo estructuran o bien a través de los diagramas de rutas, que sirven además, para resumir gráficamente, el conjunto de hipótesis sobre las que se asienta el modelo.

2.2.1. Notación del modelo

El modelo consiste en un sistema de ecuaciones estructurales, conformadas por variables aleatorias, parámetros estructurales y, en algunos casos, variables no aleatorias. Los tres tipos de variables aleatorias son variables latentes, observadas y de error. Las variables no aleatorias son variables explicativas cuyos valores permanecen iguales en el muestreo aleatorio repetido (variables fijas o no estocásticas). Estos son menos comunes que las variables explicativas aleatorias. Las relaciones entre las variables se resumen en los parámetros estructurales. Los parámetros estructurales son constantes invariantes que proporcionan la relación causal entre las variables en estudio. Los parámetros estructurales pueden describir el vínculo causal entre variables no observadas, entre variables observadas, o entre variables no observadas y observadas. El sistema de ecuaciones estructurales tiene dos subsistemas principales: el modelo de la variable latente y el modelo de medición.

2.2.1.1. Modelo de variable latente

Sea $\boldsymbol{\eta} = (\eta_1, \dots, \eta_m)^\top$ es un vector de variables endógenas latentes, $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)^\top$ es el vector de variables exógenas latentes, \mathbf{B} es la matriz $(m \times m)$ de coeficientes que relacionan las variables latentes endógenas entre sí, $\mathbf{\Gamma}$ es la matriz $(m \times n)$ de coeficientes que relacionan las variables latentes exógenas con las endógenas y $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_m)^\top$ es el vector de error aleatorio. Se define el modelo de variables latentes por

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \mathbf{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}. \quad (2.31)$$

Por conveniencia, suponemos que

- (i) $\mathbb{E}(\boldsymbol{\eta}) = \mathbb{E}(\boldsymbol{\xi}) = \mathbb{E}(\boldsymbol{\zeta}) = 0$
 - (ii) $\boldsymbol{\eta}$ no está correlacionado con $\boldsymbol{\xi}$
 - (iii) $(\mathbf{I} - \mathbf{B})$ es una matriz no singular.
- (2.32)

2.2.1.2. Modelo de medición

Sea $\mathbf{Y} = (y_1, \dots, y_p)^\top$ es un vector de variables observadas que actúan como indicador de la variable $\boldsymbol{\eta}$, $\mathbf{X} = (x_1, \dots, x_q)^\top$ es un vector de variables observadas que actúan como indicador de la variable $\boldsymbol{\xi}$, $\boldsymbol{\Lambda}_y$ es la matriz $(p \times m)$ de coeficientes que relaciona

\mathbf{Y} con $\boldsymbol{\eta}$, Λ_y es la matriz $(q \times n)$ de coeficientes que relaciona \mathbf{Y} con $\boldsymbol{\eta}$, $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_p)^\top$ es la medida de error para la variable \mathbf{Y} y $\boldsymbol{\delta} = (\delta_1, \dots, \delta_q)^\top$ es la medida de error para la variable \mathbf{X} . Se define el modelo de medición por

$$\begin{aligned}\mathbf{Y} &= \Lambda_y \boldsymbol{\eta} + \boldsymbol{\epsilon} \\ \mathbf{X} &= \Lambda_x \boldsymbol{\xi} + \boldsymbol{\delta}.\end{aligned}\tag{2.33}$$

Por conveniencia, suponemos que

$$\begin{aligned}\text{(i)} \quad & \mathbb{E}(\boldsymbol{\eta}) = \mathbb{E}(\boldsymbol{\xi}) = \mathbb{E}(\boldsymbol{\epsilon}) = \mathbb{E}(\boldsymbol{\delta}) = 0 \\ \text{(ii)} \quad & \boldsymbol{\epsilon} \text{ no está correlacionado con } \boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\delta} \\ \text{(iii)} \quad & \boldsymbol{\delta} \text{ no está correlacionado con } \boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\epsilon}.\end{aligned}\tag{2.34}$$


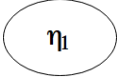
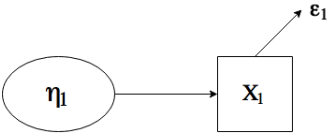
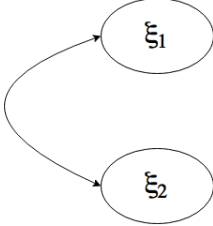
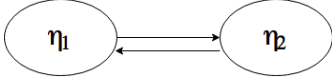
2.2.2. Análisis de rutas

El análisis de rutas es una metodología para analizar los sistemas de ecuaciones estructurales. Este tipo de análisis enfatiza tres componentes: (1) diagrama de rutas, (2) descomposición de covarianzas y correlaciones en términos de parámetros de modelo, y (3) las distinciones entre los efectos directos, indirectos y totales de una variable sobre otra.

2.2.2.1. Diagrama de rutas

Un diagrama de rutas es una representación gráfica (o imagen) de un sistema simultáneo ecuaciones. Para muchos investigadores esta imagen puede representar las relaciones más claramente que el ecuaciones, aunque en ocasiones, esto puede ser engorroso si se tiene un sistema de ecuaciones estructurales con muchas variables. En la tabla 2.1 se definen la nomenclatura para describir las distintas componentes de los modelos SEM. Las variables observadas están encerradas en cuadros. Las variables no observadas o latentes están rodeadas por un círculo, con la excepción de los términos de error que no están encerrados. Las flechas rectas de una sola cabeza representan relaciones causales entre las variables conectadas por las flechas. Una flecha curva de dos puntas indica una asociación entre dos variables. Las variables pueden estar asociadas por varias razones. La asociación puede deberse a que ambas variables dependen de alguna tercera (s) variable (s), o las variables pueden tener una relación causal, pero esto permanece sin especificar.

CUADRO 2.1: Símbolos principales usados en el análisis de rutas

	una caja rectangular o cuadrada corresponde a una variable observada
	un círculo o elipse corresponde a una variable latente
	una variable no cerrada significa un término de perturbación (error en la ecuación o medida) una flecha recta significa la suposición de que la variable en la base de la flecha causa una variable en la punta de la flecha
	una flecha curva de dos cabezas significa una asociación no analizada entre dos variables
	dos flechas rectas de una sola cabeza que conectan dos variables significa una relación de retroalimentación o causalidad recíproca

2.2.2.2. Descomposición de covarianzas y correlaciones

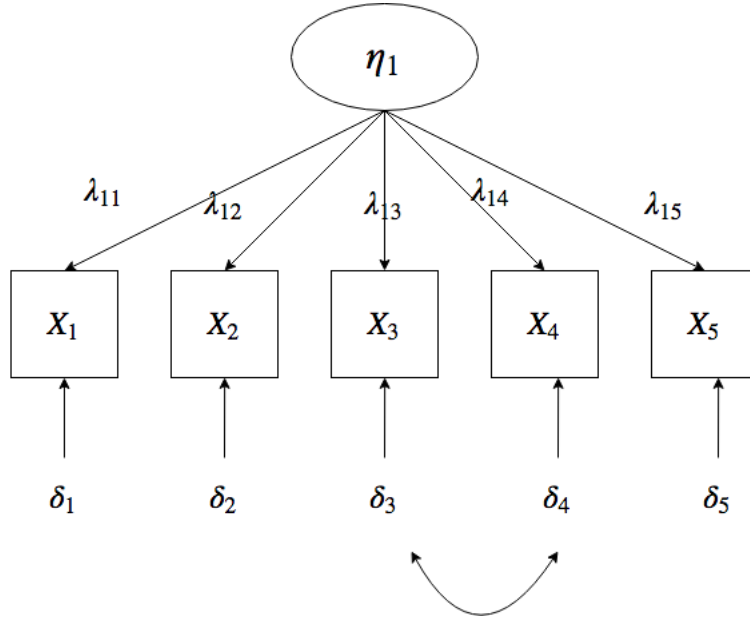
El análisis de ruta permite escribir la covarianza o correlación entre dos variables como funciones de los parámetros del modelo mediante álgebra de covarianza. Para ilustrar esto, consideremos el modelo de la Figura 3.1. Este se encuentra representado por una variable latente (η_1) que tiene cinco indicadores (x_1 a x_5). Todos los errores de medición (δ_i 's) son no correlacionados, excepto para δ_3 y δ_4 . Los errores de medición (δ_i 's) son asumidos no correlacionados con ξ_1 y $\mathbb{E}(\delta_i) = 0$, para todo $i = 1, 2, 3, 4, 5$.

La descomposición de la $COV(x_1, x_4)$ es

$$\begin{aligned} COV(x_1, x_5) &= COV(\lambda_{11}\eta_1 + \delta_1, \lambda_{51}\eta_1 + \delta_5) \\ &= \lambda_{11}\lambda_{51}\phi_{11}. \end{aligned}$$

El lado derecho de la ecuación anterior, indica que la $COV(x_1, x_5)$ se encuentra en función de los efectos de η_1 sobre x_1 y x_5 (i.e. λ_{11} y λ_{51}) y la varianza de la variable latente η_1 .

FIGURA 2.1: Diagrama de rutas para una variable latente con cuatro indicadores.



2.2.2.3. Efectos directos, indirectos y totales

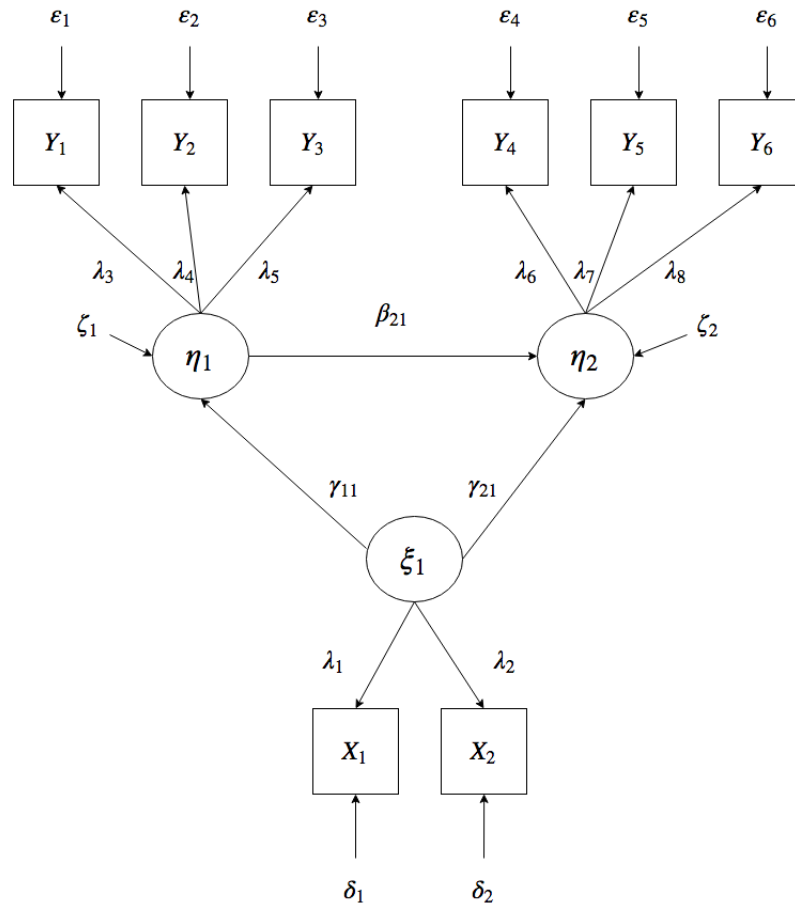
El análisis de ruta distingue tres tipos de efectos: efectos directos, indirectos y totales. El efecto directo es la influencia de una variable sobre otra que no está mediada por ninguna otra variable en un modelo de ruta. Los efectos indirectos de una variable están mediados por al menos una variable intermedia. La suma de los efectos directos e indirectos son los efectos totales:

$$\text{efectos totales} = \text{efecto directo} + \text{efectos indirectos}$$

La descomposición de los efectos siempre es con respecto a un modelo específico. Si el sistema de ecuaciones se altera incluyendo o excluyendo variables, las estimaciones de los efectos totales, directos e indirectos pueden cambiar. Para ilustrar, considere el modelo de la Figura 3.2, este modelo relaciona la industrialización de 1960 con la democracia política de 1960 y 1965 en los países en desarrollo. Estudiemos el efecto total de la variable ξ_1 sobre η_2 :

$$\begin{aligned} \text{efecto total} &= \text{efecto directo} + \text{efecto indirecto} \\ &= \gamma_{21} + \gamma_{11}\beta_{21}. \end{aligned}$$

FIGURA 2.2: Diagrama de rutas para un modelo con varias variables latentes y varias variables indicadoras.



El efecto total de ξ_1 sobre y_6 es:

$$\begin{aligned} \text{efecto total} &= \text{efecto directo} + \text{efecto indirecto} \\ &= 0 + (\gamma_{21}\lambda_8 + \gamma_{11}\beta_{21}\lambda_8). \end{aligned}$$

En este caso, el efecto total de ξ_1 sobre y_6 se compone de efectos indirectos.

2.2.3. Relaciones de causalidad entre las variables

Un aspecto importante de los sistemas de ecuaciones estructurales es el tipo de relación causal que pueden tener las variables involucradas en el modelo. Para explicar este tipo de relaciones, consideremos dos variables, digamos v_1 y v_2 . Entonces los tipos de relaciones entre v_1 y v_2 puede ser

- i) **relación directa:** la variable v_1 causa v_2 , es decir, asumir un modelo de regresión en el que v_2 fuese la variable respuesta y v_1 la variables predictora (ver Figura 2.3a). De igual forma, pueden covariar si v_2 causa v_1 (ver Figura 2.3b).
- ii) **relación recíproca:** la variable v_1 causa v_2 y la variable v_2 causa v_1 al mismo tiempo (ver Figura 2.3c).
- iii) **relación espúrea:** las variables v_1 y v_2 tienen una causa común, digamos una tercera variable v_3 (ver Figura 2.3d).
- iv) **relación indirecta:** las variables v_1 y v_2 pueden covariar a través de una tercera variable interviniente v_3 ; estableciéndose entre la primer y la segunda de ellas (Figura 2.3e).
- v) **efecto conjunto:** las variables v_1 y v_2 pueden tener una relación indirecta y espúrea al mismo tiempo (ver Figura 2.3f).

Cabe destacar que la diferencia entre la Figura 2.3d, la Figura 2.3e y la Figura 2.3f, es que en esta última, tanto v_1 como v_3 son variables exógenas y carecen de mecanismo causal explícito que las relacionen entre sí, tampoco se especifica qué tipo de relación existe entre v_1 y v_3 , dejando la covariación entre ambas variables como no explicada. El coste que se deriva de esta indeterminación se refleja en la imposibilidad de discernir si v_3 contribuye a la covariación entre v_1 y v_2 por vía indirecta o espúrea.

2.3. Modelos de panel cruzados (CLPM)

Los modelos de panel cruzados (sus siglas en inglés son **Cross-Lagged Panel Analysis**) fueron estudiados por David Kenny en el año 1975, con el fin de examinar las relaciones causales entre dos o más variables a lo largo del tiempo. El número de veces que se miden las interacción entre las variables lo definiremos como número de olas. A modo de ejemplo, mostraremos un modelo CLPM de dos olas medidos sobre dos variables. Sea X e Y dos variables de interés medidas en el tiempo i , entonces el modelo se puede escribir por

$$\begin{aligned} X_2 &= \beta_1 X_1 + \beta_2 Y_1 + \epsilon_x, \\ Y_2 &= \beta_3 Y_1 + \beta_4 X_1 + \epsilon_y, \end{aligned} \tag{2.35}$$

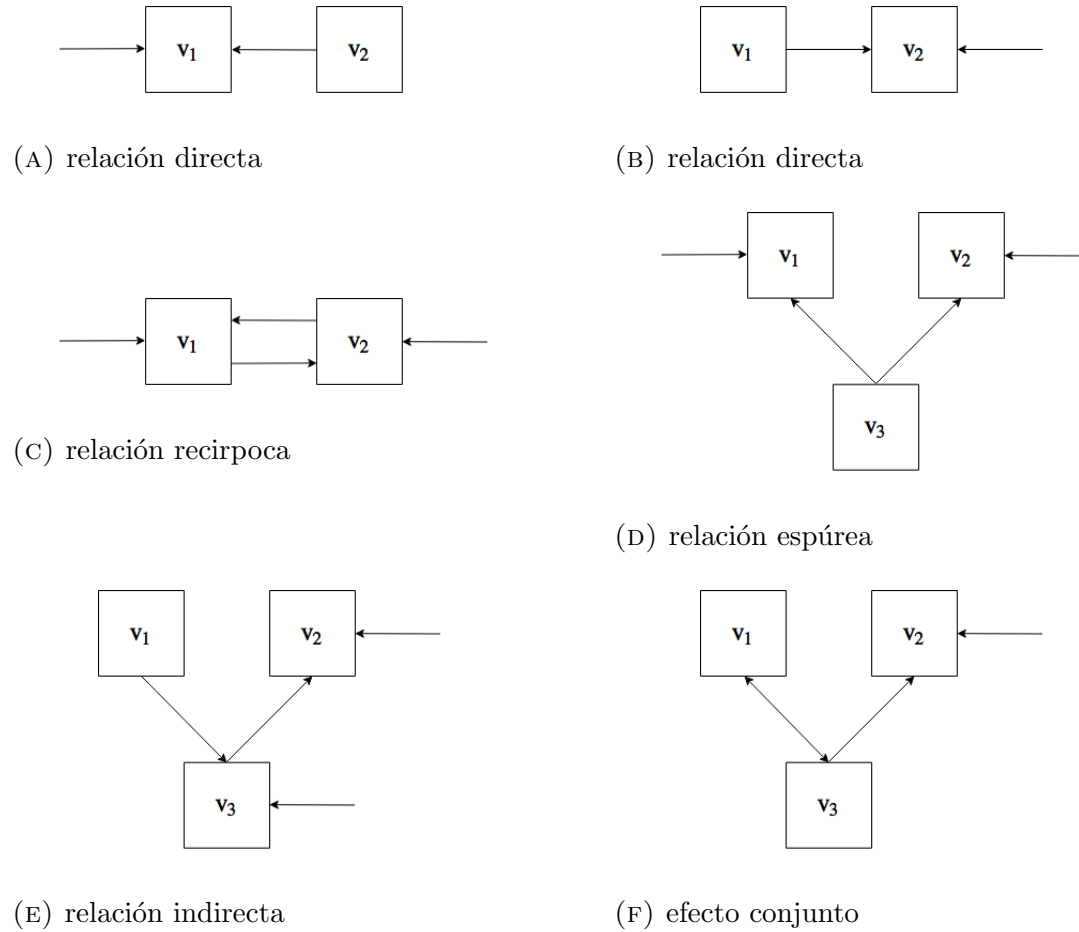
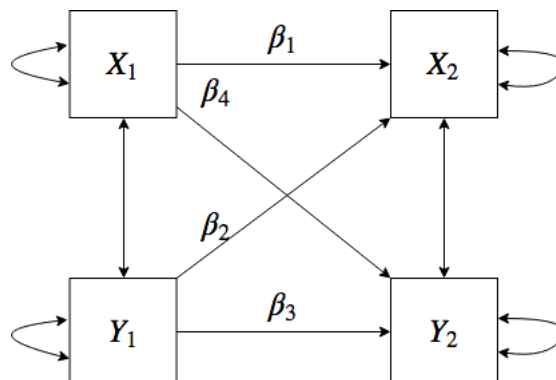


FIGURA 2.3: Relaciones causales entre variables. Formas de covariación

donde β_1 y β_3 son los efectos autorregresivos y β_2 y β_4 son los efectos cruzados. Los coeficientes de regresión β_1 y β_3 describen la estabilidad de las diferencias individuales de un instante a otro. Entre más cercano este a uno, más estable es el orden jerárquico de los individuos de un instante a otro. Los coeficientes de regresión β_2 y β_4 describen el predominio causal que puede tener una variable sobre otra. Por ejemplo, si el efecto cruzado de X en el tiempo 1 sobre Y en el tiempo 2 es grande, mientras que el efecto cruzado de Y en el tiempo 1 sobre X en el tiempo 2 es cero, la variable X sería la variable fuente y la variable Y sería la variable de efecto. La Figura 2.4 muestra un diagrama de ruta para un modelo de panel de dos olas sobre las variables X e Y .

Por otro lado, los modelos CLPM se pueden extender a más de dos variables y a un número de olas mayor a dos, sin embargo, el costo de aumentar el número de olas ocasiona las estimaciones de los parámetros resulten engorrosa y difícil de interpretar. Debido a esta dificultad, es que en la mayoría de los estudios en que se involucran modelos del tipo CLPM, se suelen ocupara un número de olas igual a dos o tres.

FIGURA 2.4: Modelo CLMP de dos olas para las variables X e Y .

2.3.1. Problemas y suposiciones en el análisis de paneles cruzados

En el año 1980, David Rogosa publica su artículo denominado "Una crítica de la correlación cruzada". Tal publicación dejó en claro que las correlaciones cruzadas entre los datos del panel longitudinal es una base inadecuada para hacer inferencias causales. Desde entonces, los investigadores se han respaldado en los modelos de panel cruzado (CLPM) para estudiar causalidad, sin embargo, este tipo de modelos presenta varias dificultades, tanto en el muestreo realizado como en los supuestos que se deben cumplir para que los modelos CLPM sean interpretables. A continuación presentamos un breve reseña de las dificultades y problemas que suelen tener estos modelos.

2.3.1.1. Problemas con el muestreo en los modelos CLMP

(i) **Sincronicidad:** el supuesto de sincronicidad asume que las mediciones observadas en cada punto de tiempo se realizaron exactamente en el mismo momento. Aunque la mayoría de los estudios están diseñados para medir las variables de manera simultánea, las complicaciones durante la recopilación de datos con frecuencia viola esta suposición.

(ii) **Confiabilidad de la medición:** en reiteradas ocasiones, los investigadores asumen que las variables observadas en los modelos CLMP fueron medidas sin error, lo

que conlleva a resultados sesgados. Sin embargo, el error de medición puede ser abordado utilizando variables latentes con múltiples indicadores (Brown,2006).

(iii) **Selectividad de la muestra:** la muestra obtenida no es representativa para la población en estudio. Esto se debe a una deficiencia realizada por el experimentador o al hecho de cómo fue tomada la muestra.

(iv) **Desgaste selectivo:** esto ocurre cuando los participantes abandonan un estudio de manera predecible por distintos problemas (por ejemplo, bajo nivel socio-económico, problemas psiquiátricos, muertes inoportuno en el estudio de una enfermedad, etc.), lo que nos lleva a un problema de datos perdidos. Sin embargo, existen varias metodologías para solucionar esta problemática, pero esto se complica si el conjunto de datos obtenidos de la muestra es significativamente pequeño.

(v) **Efectos pruebas repetitivas:** al trabajar con datos de panel, necesitamos medir la misma información para distintos instantes de tiempo, llevándonos a inconvenientes tales como: los encuestados recuerdan el material presentado, reaccionan a preguntas repetidas o simplemente se pierde el interés en el estudio, por nombrar algunos ejemplos. Muchas veces estos efectos son ignorados y pocas veces son estimados, lo bueno que existen alternativas para el efecto de reprobación (ver Enders,2010).

2.3.1.2. Problemas con la interacción de las variables en los modelos CLMP

(i) **Teoría del cambio:** en el contexto de datos longitudinales, la teoría de cambio se enfoca en el cambio que tiene un variable en el tiempo y en el cambio inducido de una variable sobre otra. Los modelos CLPM generalmente carecen de teorías explícitas de cambio, asumiendo que no hay diferencias inter-individuales, o diferencias entre-personas, a lo largo del tiempo en el sentido de estabilidad. Las diferencias inter-individuales que existen, como las influencias o dependencias tipo rasgo no observadas, pueden sesgar los resultados. Este problema lo mostraremos con más detalles en el capítulo 2.

(ii) **Estacionariedad:** los modelos CLPM suponen que las variables y las relaciones permanecen iguales a través del tiempo. Es te supuesto se relaciona con la estabilidad de una construcción, así como con la naturaleza de las relaciones entre construcciones a lo largo del tiempo. Existen distinto grados de estacionariedad para los modelos de tres o más olas, sin embargo, existe pocas teorías que nos orienten en este aspecto.

(iii) **Comparación de coeficientes cruzados:** Para hacer afirmaciones sobre el predominio causal, el análisis de trayectorias cruzada incluye la comparación de los tamaños relativos de los coeficientes cruzados. Esto se logra mediante la estandarización de variables. Aunque cada una de las variables estandarizadas se puede describir en términos similares, la estandarización no necesariamente aborda las distribuciones de diferencias fundamentales. En algunos casos, puede no ser apropiado suponer que las variables se midieron en la misma escala.

(iv) **Variables omitidas:** en teoría, los modelos CLPM suponen que todas las variables importantes se midieron e incluyeron en el modelo. Sin embargo, la omisión de variables puede conducir a resultados sesgado. Por ejemplo, para un modelo de dos olas en el que se estudian las interacciones entre las variables X e Y , los efectos cruzados entre ambas variables puede explicarse por la dependencia mutua de una tercera variable Z .

2.3.2. Enfoques alternativos para los modelos de panel

Debido a las serias limitaciones de los modelos CLPM, muchos expertos han propuesto enfoques alternativos sobre los modelos de panel. Estos enfoques tratan de ser consistente con la teoría de cambio de las variables, abordan el tema de las diferencias interindividuales e intraindividual, realizar estudios sobre la descomposición de las matrices de covarianza, entre otros aspectos. Nosotros consideramos los siguientes modelos a discutir: (1) modelo de trayectorias autorregresivas latent (modelos ALT) y modelos RI-CLPM, (2) modelo de rasgo autorregresivo (modelos STARTS) ,(3) modelo del puntaje de cambio latente (modelos LCT),(4) modelo de estado latente modificado (modelos LST). Esta sección expone los aspectos más importantes de cada modelo, mostrando las ventajas que tienen sobre los modelos CLPM. Para más detalles en cada modelos se dejan las citas correspondientes.

2.3.2.1. Modelo de trayectorias autorregresivas (ALT) y modelo RI-CLPM

Los modelos ALT, desarrollado por Curran y Bollen (2001; ver también Bollen & Curran, 2006), permiten que las personas se caractericen por su propia trayectoria en el tiempo, mientras que sus observaciones también puedan exhibir algún efecto de arrastre de un instante al siguiente. En la extensión bivariada, los efectos aleatorios que describen las trayectorias individuales pueden correlacionarse entre sí a través de las variables, y también puede haber ser influencias cruzadas entre las observaciones (como en el CLPM). Por otro lado, dado que las relaciones autorregresivas se incluyen entre las observaciones, existe recursividad en el modelo, que tiene algunos efectos adversos. En primer lugar, implica que el proceso debe ser **iniciado**, por lo que se proponen dos soluciones: la primera observación se trata como exógena, o se imponen restricciones no lineales sobre los coeficientes de cargas para el primer instante. Los autores Jongerling y Hamaker (2011) muestran que el considerar la primera observación como exógena puede conducir a curvas de crecimiento bastante inesperadas. Por ejemplo, consideremos un modelo ALT con solo una constante aleatoria (es decir, sin parámetro de tendencia lineal), esto puede estar modelando una tendencia creciente o decreciente a lo largo del tiempo. Por otro lado, estas problemáticas quedan exentas si se utilizan las restricciones no lineales para iniciar el proceso, sin embargo, estas suponen que los efectos de retraso son constantes en el tiempo, algo que es más difícil de imponer. La recursividad en el modelo ALT implica que la constante aleatoria y el parámetro de cambio aleatorio ya no tienen el papel original de las intersecciones y pendientes de los individuos (Hamaker, 2005). Por ejemplo, la constante aleatoria no solo afecta una observación directamente, sino también indirectamente a través de (todas) las ocasiones anteriores. Considerando la parametrización anterior, se definen los modelos RI-CLPM (desarrollados por Hamaker, Kuipers & Grasman, 2015) como un caso particular de los modelos ALT bivariados. Estos modelos solo consideran los interceptos aleatorios y no las pendientes aleatorias. Los modelos RI-CLPM tienen ventajas sobre los modelos ALT, se implementa de mejor manera sobre datos bivariados (o multivariados), permite la existencia de cambios no lineales de los individuos a lo largo del tiempo y la matriz de covarianza es más fácil de estimar. Por otro lado, estos modelos requieren al menos tres olas de datos y suponen que a lo largo del tiempo los individuos tienden a su propia media, supuesto que no siempre se cumple.

2.3.2.2. Modelo de rasgo autorregresivo (STARTS)

Los modelos STARTS de Kenny y Zautra (2001), también conocidos en la literatura por el nombre de modelos Trait State Error (TSE) (Kenny y Zautra, 1995), permiten descomponer la varianza observada en tres componentes: (a) rasgo estable, que no cambia; (b) rasgo autorregresivo, que cambia según un proceso autorregresivo; y (c) el estado o error, que es exclusivo de la ocasión. En un comienzo, estos modelos incluyeron restricciones en el tiempo, de modo que las contribuciones relativas de estas tres componentes permanezcan estables en el tiempo, sin embargo, tales restricciones pueden relajarse si existe un número de suficientes de olas en las mediciones disponibles (ver Lucas & Donnellan, 2007). La mayoría de las aplicaciones de estos modelos se basan en mediciones repetidas univariadas, sin embargo, muchos autores también han considerados extensiones bivariadas de estos modelos. Por otro lado este modelo presenta serias desventajas, asume la no existe error en las mediciones y requiere al menos 10 olas de mediciones, complejizando la estimación de los parámetros. Los autores Cole, Martin y Steiger (2005) realizaron diferentes simulaciones, concluyendo que estos modelos arrojaban soluciones incorrectas difíciles de interpretar, por ejemplo, estimación de varianza negativa, problemas de singularidad, entre otros.

2.3.2.3. Modelo del puntaje de cambio latente (LCT)

Los modelos LCS, también conocido como modelos de puntuación de diferencia latente (LDS), fueron propuestos por McArdle y Hamagami en el año 2001. Los modelos LDS consideran que los cambios latentes (es decir, los puntajes de diferencias corregidos por el error de medición), de un instante de tiempo al siguiente son modelados como una función de un parámetro de cambio constante y un parámetro de cambio proporcional los cuales dependen del puntaje anterior. Por este motivo, estos modelos también se suelen denominar modelos Dual Change Score (McArdle, 2009). En la extensión bivariada de este modelo, el cambio no es solo una función de un parámetro de cambio constante y el parámetro de cambio proporcional, sino también del puntaje anterior en la otra variable. En los modelos LCS las correlaciones cruzadas se denominan parámetros de acoplamiento. Si bien la interpretación es la misma, los parámetros de acoplamiento importantes implican que una variable tiene la tendencia a afectar los cambios en la otra variable. En lugar ocupar los coeficientes para determinar la causalidad entre las variables, los parámetros de acoplamiento son usados para configurar un campo vectorial que representa los cambios esperados de un instante

a la siguiente en ambas variables como una función del estado actual (ver McArdle y Grimm, 2010). El modelo LCS se ha ampliado con lo que se ha denominado “error dinámico” para distinguirlo del error de medición (ver McArdle, 2001): mientras que el error de medición solo afecta a la observación del instante actual, el error dinámico afecta la trayectoria del sistema y lo convierten en un proceso estocástico en lugar de determinista.

2.3.2.4. Modelo de estado latente modificado (LST)

Los modelos LST fueron desarrollado originalmente para hacer distinción entre el error de medición y el puntaje o score verdadero (es decir, una variable latente), y para descomponer aún más el score verdadero en una parte similar a un rasgo y un estado (Steyer, Mayer, Geiser, y Cole, 2015). En la práctica, se supone que existe un constructo subyacente, que se mide mediante indicadores múltiples. Este constructo subyacente en un instante de tiempo particular se conoce como estado, que luego se descompone en una parte de rasgos y una parte ocasional. La parte de rasgos es a menudo incluido como un factor de segundo orden que relaciona los estados -que están representados por los factores de primer orden- entre sí. La parte específica de la ocasión es la parte residual del factor de estado, que no fue contabilizada por el rasgo. El modelo LST se ha ampliado con relaciones autorregresivas entre los factores de estado (introduciendo una recursividad similar a la existente en el modelo ALT y el modelo LCS), o entre los componentes específicos de la ocasión (para evitar la recursividad perjudicial en el modelo), este último se ha denominado modelo Trait State Occasion (TSO) (Cole et al., 2005). En este modelo modificado, se omite el término de error de medición, el factor de rasgo se modela como un factor separado con cargas de factores libres a lo largo del tiempo (en lugar de un factor secundario), y se incluyen relaciones autorregresivas de segundo orden. Tenga en cuenta que si se mantuviera el término de error de medición (y se omitieran las relaciones autorregresivas de segundo orden), el modelo sería idéntico al modelo STARTS.

2.3.2.5. Resumen de los enfoques alternativos

Los enfoques alternativos descrito anteriormente muestran mejoras sustanciales respecto a los modelos CLPM, en aspecto tales como la descomposición de la varianza y el estudio de los efectos intraindividual-interindividual. Si los investigadores están específicamente interesados en descomponer la varianza en componentes de rasgos y

estado, los modelos STARTS y los modelos basados en el modelo LST son los más relevantes. Por el contrario, si el interés está en trayectorias individuales de desarrollo, los modelos RI-CLPM, los modelos ALT y los modelos LCS son más apropiados, ya que se basan en modelar tanto la estructura media como la estructura de covarianza y permiten a los individuos tener sus propias curvas de crecimiento. Por otro lado, el modelo STARTS, el modelo ALT y el modelo LST generalmente se aplican a datos univariados (aunque el modelo LST original usa indicadores múltiples); si bien las extensiones bivariadas (o multivariadas) son posibles, no forman el foco central y los parámetros de regresión cruzada no son de interés clave. Por el contrario, el modelo LCS se utiliza principalmente para investigar cómo se influyen dos variables entre sí (en función del cambio esperado descrito con el campo vectorial), aunque también se puede aplicar a datos univariados. Sin embargo, estos enfoques requieren de otros supuestos para que puedan presentar mejoras respecto a los modelos CLPM. Por ejemplo, los modelos STARTS requieren al menos 10 olas de datos, lo cual muchas veces es imposible de tratar. La mayoría de las investigaciones en el ámbito de las ciencias sociales solo utilizan dos o tres olas de datos, rara vez se ocupan un número elevado de olas. Otro ejemplo, serían los modelos RI-CLPM, estos suponen que las variaciones intraindividuales caerán a su valor medio en el tiempo, algo que no siempre es así. Se advierte a los investigadores que antes de aplicar estos modelos se tenga pleno conocimientos de la información disponible, desde el número de olas hasta el tipo de comportamiento a priori de las variables en estudio.

Capítulo 3

Desarrollo

3.1. Modelos CLPM versus RI-CLPM

En esta sección se presenta de manera formal los modelos de panel cruzados (CLPM) y una alternativa a este tipo de modelos, los modelos de panel cruzados con efectos aleatorios (RI-CLPM), en donde se muestran su formulación matemática, sus respectivos diagramas de rutas, sus principales características y sus principales diferencias. Para esto, nos basamos las investigaciones realizados por Hamaker en el año 2015 para mostrar los puntos descrito con anterioridad.

3.1.1. Los modelos CLPM

Como explicamos en la sección pasada, los modelos de panel cruzados (CLPM) se aplican para datos del tipo longitudinales, en donde estudiamos la interacción de dos o más variables en el tiempo, con el fin de establecer relaciones causales. Para definir este tipo de modelo, a diferencia con el ejemplo de la Figura 2.4 es que acá consideramos como variable la media de las distintos constructos, como una forma de considerar el efecto inter-individual. Una representación gráfica de este modelo se da en la Figura 3.1. Las ecuaciones de medición para el CLPM con los medios se pueden expresar por

$$\begin{aligned}x_{it} &= \mu_t + p_{it}, \\y_{it} &= \pi_t + q_{it},\end{aligned}\tag{3.1}$$

donde μ_t y π_t representan la medias grupales en el tiempo t para las dos variables, respectivamente. Si se trabajan con datos centrados, tenemos que $\mu_t = \pi_t = 0$, tal que $p_{it} = x_{it}$ y $q_{it} = y_{it}$. Sin embargo, cuando se incluyen los medias, p_{it} y q_{it} representan las desviaciones temporales de los individuos de las medias grupales. Por otro lado, debemos especificar los errores de mediciones para las ecuaciones (2) mediante las siguientes ecuaciones estructurales:

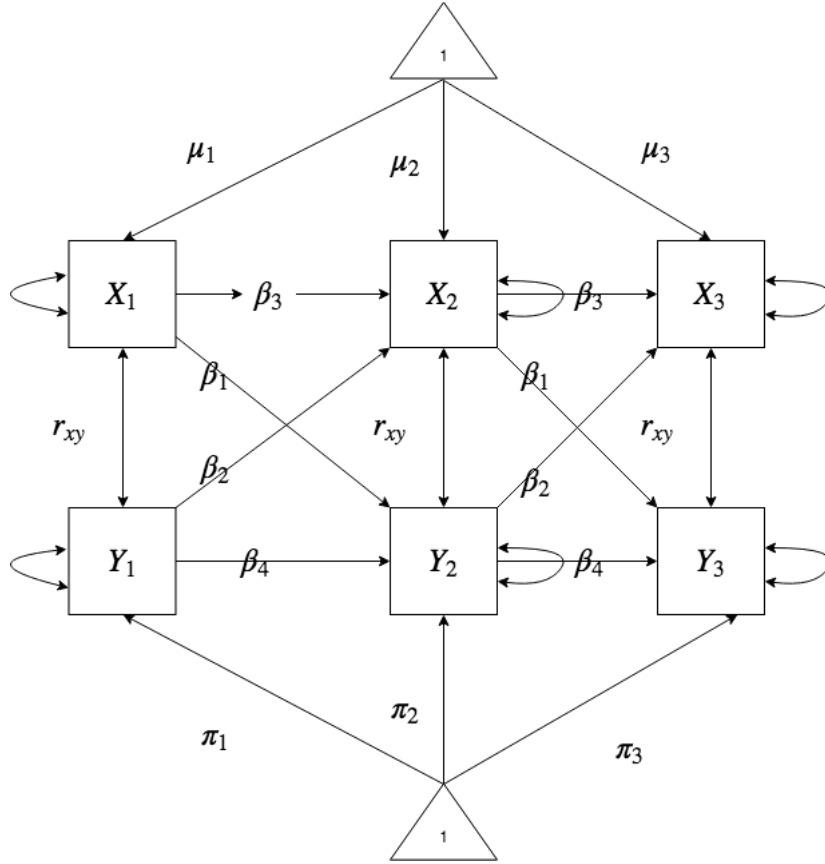
$$\begin{aligned} p_{it} &= \alpha_t p_{i,t-1} + \beta_t q_{i,t-1} + u_{it}, \\ q_{it} &= \delta_t q_{i,t-1} + \gamma_t p_{i,t-1} + v_{it}. \end{aligned} \quad (3.2)$$

Los parámetros autorregresivos α_t y δ_t se incluyen para tener en cuenta la estabilidad de los constructos: cuanto más cercanos son estos parámetros autorregresivos a uno, más estable es el orden jerárquico de los individuos de una ocasión a la siguiente. Sin embargo, incluso cuando los coeficientes de estabilidad son muy altos, para un tiempo suficiente, se perderá la orden de rango original. Los parámetros cruzados β_t y γ_t forman la clave para investigar los efectos causales recíprocos en este modelo (Rogosa, 1980). Mediante la estandarización de estos parámetros, se puede hacer una comparación de los efectos relativos de X e Y entre sí, que luego se puede utilizar para determinar el predominio causal (Bentler y Speckart, 1981). Estos parámetros a menudo se interpretan en términos de predicción de cambios (por ejemplo, Finkel, 1995; Ribeiro et al., 2011; Rindermann, 2008). Para mostrar el razonamiento detrás de esta interpretación, escribimos

$$\begin{aligned} y_{it} - y_{i,t-1} &= (\pi_t + q_{it}) - (\pi_{t-1} + q_{i,t-1}) \\ &= (\pi_t - \pi_{t-1}) + (\delta_t - 1)q_{i,t-1} + \gamma_t p_{i,t-1} + v_{it}, \end{aligned} \quad (3.3)$$

el cual muestra que el parámetro cruzado con retraso γ_t indica el grado en que el cambio en Y puede predecirse a partir de la desviación previa del individuo de la media del grupo en X (es decir, $p_{i,t-1} = x_{i,t-1} - \mu_{t-1}$). En esta expresión, también se controla el cambio estructural en y (es decir, $\pi_t - \pi_{t-1}$), y la desviación de la media del grupo en y (es decir, $q_{i,t-1} = y_{i,t-1} - \pi_{t-1}$). Incluir la desviación previa de las personas mediante media del grup, se considera una forma de controlar el sesgo debido a la regresión hacia la media (Liker, Augustyniak y Duncan, 1985).

FIGURA 3.1: Box plot para las variables en estudio.



3.1.2. Los modelos RI-CLPM

Como se describimos anteriormente, el CLPM solo tiene en cuenta la estabilidad temporal mediante la inclusión de parámetros autorregresivos. Esto implica que en este modelo se supone implícitamente que cada persona varía con el tiempo en los mismos términos μ_t y π_t , y que no existen diferencias individuales de tipo rasgo que perduren, es decir, no existe un fundamento explícito respecto a la teoría del cambio entre las variables. Esta suposición es bastante problemática, puesto que es difícil imaginar una construcción psicológica, ya sea conductual, cognitiva, emocional o psicofisiológica, que no se caracterice hasta cierto punto por las diferencias individuales. Los datos longitudinales en realidad se pueden considerar como datos multinivel, en los cuales las mediciones se anidan dentro del individuos. Al considerar esta perspectiva, se pone en evidencia la necesidad de separar el nivel dentro de la persona del nivel entre personas. Esta idea fue el motor para estudiar alternativas a los modelos CLPM, en particular, los modelos RI-CLPM aquí estudiados, que se presentan como una extensión de los

modelos CLPM que no solo se da cuenta de la estabilidad temporal, sino también de la estabilidad de rasgo invariante en el tiempo mediante la inclusión de un intercepto aleatorio. Las ecuaciones de medición para el RI-CLPM con las medias grupales se pueden expresar por

$$\begin{aligned}x_{it} &= \mu_t + \kappa_i + p_{it}^*, \\y_{it} &= \pi_t + \omega_i + q_{it}^*,\end{aligned}\tag{3.4}$$

donde μ_t y π_t son nuevamente la media del grupo temporal. Los términos adicionales κ_i y ω_i son las desviaciones de los rasgos individuales respecto de los rasgos de estas medias grupales. También se pueden considerar como variables latentes o factores cuyas cargas de factores están todas limitadas a 1, como en el caso de intersecciones aleatorias en la curva de crecimiento latente (LGC) modelado (con la diferencia de que aquí el grupo significa que pueden variar libremente con el tiempo). Se ha agregado un asterisco a los términos de desviación temporal p_{it}^* y q_{it}^* , para enfatizar que estos términos son diferentes de los términos de desviación individual en los modelos CLPM tradicionales. En el modelo actual, estas variables representan las desviaciones temporales individuales de sus puntajes o score esperados (es decir, $\mu_t + \kappa_i$ y $\pi_t + \omega_i$), en lugar de las medias grupales (es decir, μ_t y π_t). Posteriormente, estas desviaciones son

$$\begin{aligned}p_{it}^* &= \alpha_t^* p_{i,t-1}^* + \beta_t^* q_{i,t-1}^* + u_{it}^*, \\q_{it}^* &= \delta_t^* q_{i,t-1}^* + \gamma_t^* p_{i,t-1}^* + v_{it}^*,\end{aligned}\tag{3.5}$$

donde los parámetros de regresión autorregresivos y cruzados difieren de los del CLPM, como lo indican los asteriscos. Es decir, los parámetros autorregresivos α_t^* y δ_t^* no representan la estabilidad del orden jerárquico de las personas de una ocasión a la siguiente, sino la cantidad de efecto de arrastre dentro de la persona (véase Hamaker, 2012; Kuppens, Allen y Sheeber, 2010; Suls, Green, & Hillis, 1998). Si estos efectos son positivos, implica que las ocasiones en las que una persona obtuvo un puntaje o score por encima de su puntaje esperado, es probable que sean seguidas las ocasiones en las que todavía obtenga puntajes por encima del puntaje esperado nuevamente, y viceversa. Sin embargo, el principal interés aquí está en los parámetros cruzados β_t^* y γ_t^* , que indican el grado en que las dos variables se influyen mutuamente. Específicamente, γ_t^* indica el grado en que las desviaciones del puntaje esperado de un individuo en Y (es decir, $q_{it}^* = y_{it} - \{\pi_t + \omega_i\}$) pueden predecirse a partir de las desviaciones precedentes de la puntuación esperada en X (es decir, $p_{i,t-1}^* = x_{i,t-1} - \{\mu_{t-1} + \kappa_i\}$),

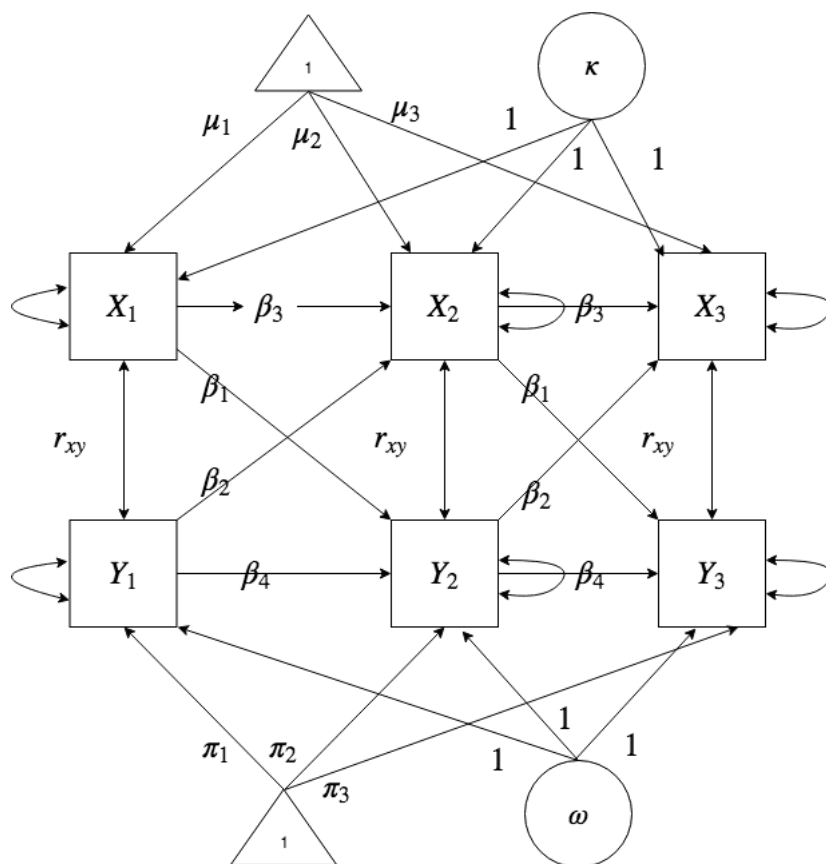
Mientras controla la desviación individual de la puntuación esperada anterior en Y (es decir, $q_{i,t-1}^* = y_{i,t-1} - \{\pi_{t-1} + \omega_i\}$). Las relaciones entrecruzadas pertenecen a un proceso que tiene lugar en el nivel intra-personal. Una representación gráfica de este modelo se da en el panel derecho de la Figura 3.2. Al igual que en el modelo CLPM, analizaremos el cambio que se produce al analizar la diferencia $y_{it} - y_{i,t-1}$ en los modelos RI-CLPM.

$$\begin{aligned} y_{it} - y_{i,t-1} &= (\pi_t + \omega_i + q_{it}^*) - (\pi_{t-1} + \omega_i + q_{i,t-1}^*) \\ &= (\pi_t - \pi_{t-1}) + (\delta_t^* - 1)q_{i,t-1}^* + \gamma_t^* p_{i,t-1}^* + v_{it}^*. \end{aligned} \quad (3.6)$$

La ecuación anterior muestra que el parámetro cruzado con retrasa indica la medida en que el cambio en Y puede predecirse a partir de la desviación previa del individuo de su puntaje o score esperado en la otra variable (es decir, $p_{i,t-1}^* = x_{i,t-1} - \{\mu_{t-1} + \kappa_i\}$), mientras controla el cambio estructural en Y (es decir, $q_{i,t-1}^* = y_{i,t-1} - \{\pi_{t-1} + \omega_i\}$), y la desviación previa de la puntuación esperada en Y (es decir, $q_{it}^* = y_{it} - \{\pi_t + \omega_i\}$). Al tomar la diferencia, la Ecuación (7) ya no incluye el componente individual estable, tipo rasgo ω_t . Esto ilustra el hecho de que los puntajes de diferencia son una forma de eliminar el efecto de variables estables no observadas, lo que a veces se considera una ventaja importante del modelo de puntuación de diferencia sobre otros enfoques (véase Allison, 2009; Liker et al., 1985). Debemos tener en cuenta que tanto ω_i como κ_i todavía están implícitamente presentes en la expresión de la ecuación (7), mediante la inclusión de $q_{i,t-1}^*$ y $p_{i,t-1}^*$. Por lo tanto, a pesar de la similitud entre la Ecuación (4) (basada en el CLPM) y en la Ecuación (7) (basada en el RI-CLPM), estos dos modelos predicen el cambio desde otros aspectos, a menos que ω_i y κ_i sean cero. De hecho, es fácil ver que el CLPM tradicional se anida bajo el modelo actual, ya que se puede obtener a partir de este último fijando las varianzas y la covarianza de ω_i y κ_i al igualarlas a 0. Sin embargo, comparar estos modelos usando una prueba de diferencia chi-cuadrado es complicado por el hecho de que requiere que fijemos dos parámetros en los límites del espacio de parámetros (es decir, las dos varianzas se fijan a cero), como resultado, la probabilidad de registro la diferencia de estos dos modelos anidados no tiene una distribución regular de chi-cuadrado, sino que sigue una distribución de chi-bar-cuadrado, que es una suma ponderada de diferentes distribuciones de chi-cuadrado (Stoel, Galindo Garre, Dolan, & van den Wittenboer, 2006). El cálculo de los pesos requeridos y, posteriormente, la determinación de los valores de p reales puede ser problemático (Silvapulle & Sen, 2004; Stoel et al., 2006). Sin

embargo, podemos hacer uso del hecho de que la prueba de diferencia chi-cuadrado regular es conservadora en este contexto, lo que significa que, si es significativa, estamos seguros de que la prueba correcta (es decir, diferencia de chi-cuadrado) será significativo también. Mientras que el CLPM requiere solo dos ondas de datos, el RI-CLPM requiere al menos tres ondas de datos, en cuyo caso hay 1 grado de libertad (df). Si los intervalos (es decir, retrasos) entre las ocasiones 1 y 2 y entre las ocasiones 2 y 3 son las mismas (es decir, las observaciones están espaciadas por igual en el tiempo), que podemos probar si los efectos que las variables tienen entre sí permanecen estables a lo largo del tiempo al restringir los parámetros rezagados con el tiempo y al hacer un chi prueba de diferencia cuadrada Este último modelo nos daría 4 df adicionales (es decir, 5 df en total). Además, es posible que deseemos investigar si los medios se pueden restringir en el tiempo, de modo que obtenemos otros 4 df (lo que da como resultado 9 df en total). Si estas restricciones no son sostenibles (por ejemplo, porque los intervalos entre las observaciones varían con el tiempo, o porque el proceso subyacente cambia con el tiempo), y no estamos seguros de si el efecto de los componentes de estabilidad invariantes en el tiempo ω_i y κ_i son iguales a largo del tiempo, podemos desear eliminar la restricción en las cargas de factores. Esta relajación puede ser especialmente interesante cuando las observaciones se separan aún más en el tiempo, y esperamos que también estemos midiendo algunos cambios estructurales. Sin embargo, esto implicaría que ω_i y κ_i ya no representan intersecciones aleatorias (como en el modelado multinivel), sino que representan variables latentes o rasgos (como es común en SEM). Aún más, implicaría que necesitamos más olas de datos para estimar este modelo.

FIGURA 3.2: Box plot para las variables en estudio.



Capítulo 4

Aplicaciones

4.1. Conjunto de Datos

En 2013, se reclutaron empleados chilenos para completar una encuesta sobre la vida y el trabajo. Se envió a los participantes un correo electrónico introductorio con una breve descripción del estudio junto con un enlace web a la encuesta. Los encuestados fueron invitados a participar voluntariamente en un proyecto de investigación completando una encuesta en línea e informados sobre el propósito de la investigación. El estudio se realizó de acuerdo con las pautas de la Asociación Británica de Psicología y la Asociación Americana de Psicología. La muestra chilena consistió en 742 adultos que trabajaban en una gran empresa chilena de servicios financieros, con edades comprendidas entre 18 y 62 años (media = 31.80; SD = 7.57). Los encuestados (72.5 % mujeres) fueron reclutados en línea a través de la unidad de recursos humanos (HR). El estudio se llevó a cabo en las 15 regiones geográficas del país (desde Arica hasta Punta Arenas) donde la compañía brinda servicios a la población chilena. En total, encuestamos a las 36 unidades de la compañía y estudiamos tareas muy diferentes (por ejemplo, gerentes generales, gerentes, supervisores, asistentes de clientes, guardias). Finalmente, el cuestionario realizado a cada uno de los individuos se realizó en tres instantes de tiempo, con un periodo de ?? meses entre un instante y otro.

4.1.1. Medidas del bienestar

Los participantes de Chile completaron un cuestionario que incluía medidas de florecimiento, compromiso de trabajo, agotamiento y satisfacción laboral. La sección

final de la encuesta evaluó los detalles demográficos, incluidos la edad y el género. La encuesta también incluyó otras medidas que son irrelevantes para la investigación actual y se excluyen aquí. El cuestionario se tradujo al español para los participantes chilenos, y se verificó la equivalencia de significado con la versión en inglés mediante procedimientos de traducción inversa en cada país (Brislin, 1970).

FloreCIMIENTO

El florecimiento es una medida relacionada con el bienestar individual. Para capturar esta medida, se utiliza el PERMA-Profiler (Butler & Kern, 2015). La medida captura específicamente los cinco componentes propuestos por Seligman (2011; emoción positiva, compromiso, relaciones, significado, logros), con tres preguntas por construcción. Los participantes respondieron en una escala de 11 puntos que oscila entre 0 y 10, y los elementos dentro de cada construcción se promediaron juntos para crear un indicador de ese dominio. Las preguntas realizadas para esta medida se encuentran en el Anexo C, tabla C.3.1..

Agotamiento

El Agotamiento es una medida relacionada con el bienestar laboral. El agotamiento se conceptualizó como agotamiento emocional (5 preguntas) y despersonalización (5 preguntas), y se evaluó con el Estudio general de inventario de agotamiento de Maslach (MBI-GS; Schaufeli, Leiter, Maslach y Jackson, 1996). Los participantes respondieron una escala de 7 puntos que va desde nunca (0) a siempre (6). Las preguntas realizadas para esta medida se encuentran en el Anexo C, tabla C.3.2..

Compromiso de trabajo

El compromiso de trabajo es una medida relacionada con el bienestar laboral. El compromiso laboral se conceptualizó en términos de vigor y dedicación. El vigor (6 preguntas) y la dedicación (5 preguntas) se midieron a través de la Escala de compromiso laboral de Utrecht (Schaufeli, Bakker y Salanova, 2006). Los participantes respondieron en una escala de 7 puntos desde nunca (0) hasta siempre (6). Las preguntas realizadas para esta medida se encuentran en el Anexo C, tabla C.3.3..

Satisfacción laboral

La satisfacción laboral es una medida relacionada con el bienestar laboral. La satisfacción en el trabajo se evaluó con una sola pregunta: “en general, ¿qué tan satisfecho está con su trabajo?”, Que los participantes calificaron en una escala de 11 puntos (0 = “extremadamente insatisfecho, 10 = “extremadamente satisfecho ”). Investigaciones anteriores encontraron que el artículo se relaciona fuertemente con una evaluación de múltiples elementos de la satisfacción laboral (Van den Broeck, Vansteenkiste, Witte, Soenens, & Lens, 2010; Vansteenkiste et al., 2007; Wanous, Reichers y Hudy, 1997), sugiriendo que un solo elemento para esta construcción es suficiente (Dolbier, Webster, McCalister, Mallon, & Steinhardt, 2005).

4.2. Análisis descriptivo

Antes de aplicar el análisis descriptivo, es necesario definir las variables de interés. Como se dijo en el comienzo de esta sección, se busca relacionar el desempeño personal (felicidad) con el desempeño laboral (trabajo) en el tiempo, no obstante, tenemos distintas maneras de medir el desempeño laboral (en este caso, tres maneras diferentes). Por lo cual, tendremos tres casos de interés: a) florecimiento versus agotamiento, b) florecimiento versus compromiso de trabajo y c) florecimiento versus satisfacción laboral. Por otro lado, se considerará cada una de las preguntas igual de importantes (es decir, cada uno tiene un peso ponderado igual a 1) para las distintas medidas de bienestar (personal o laboral). En base a esto, nuestras nuevas variables de interés quedan definidas por el promedio de las puntuaciones obtenidas en cada pregunta según el tipo de medida de bienestar (personal o laboral), es decir,

a) florecimiento:

$$PERMA_i = \frac{1}{15} \left(\sum_{j=1}^3 P_{ji} + \sum_{j=1}^3 E_{ji} + \sum_{j=1}^3 R_{ji} + \sum_{j=1}^3 M_{ji} + \sum_{j=1}^3 A_{ji} \right),$$

para todo $i = 1, 2, 3$.

b) agotamiento:

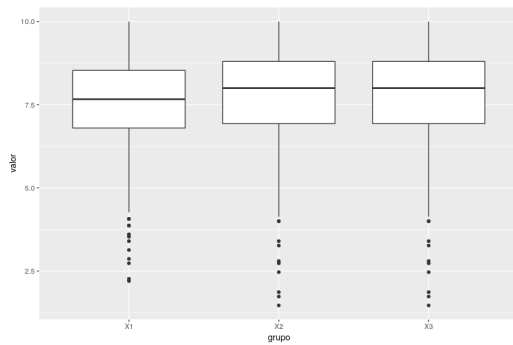
$$BOUW_i = \frac{1}{10} \left(\sum_{j=1}^5 EE_{ji} + \sum_{j=1}^5 DP_{ji} \right), \text{ para todo } i = 1, 2, 3.$$

c) compromiso de trabajo:

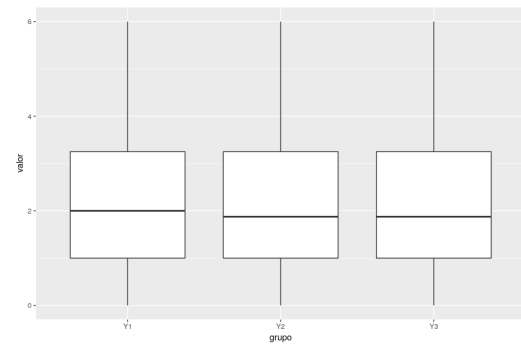
$$ENGW_i = \frac{1}{10} \left(\sum_{j=1}^6 VIG_{ji} + \sum_{j=1}^5 DED_{ji} \right), \text{ para todo } i = 1, 2, 3.$$

d) satisfacción laboral: En este caso, al tener solo una pregunta, se considera la puntuación obtenida en la pregunta realizada. La variable queda definida por WSG_i , para todo $i = 1, 2, 3$.

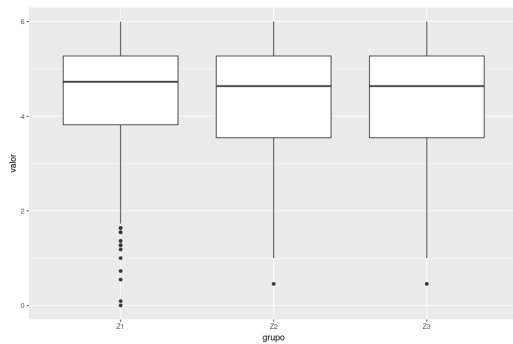
En la Figura 4.1 se muestra un diagrama de box-plot para las distintas medidas de bienestar en los tiempos $i = 1, 2, 3$. En promedio, la medida de bienestar se mantienen constantes para los distintos instantes de tiempo. Por otro lado, para el florecimiento (ver Figura 4.1a) y el compromiso de trabajo (ver Figura 4.1c) se observa la presencia de outliers. Además, el agotamiento (ver Figura 4.1b) es la única de las cuatro medidas que entre menor sea su puntuación, mejor será su desempeño laboral.



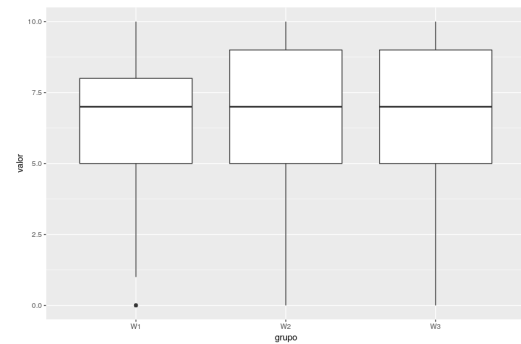
(A) florecimiento



(B) agotamiento



(C) compromiso de trabajo



(D) satisfacción laboral

FIGURA 4.1: Box Plot sobre la puntuación promedio de las distintas medidas de bienestar

Por otro lado, analizamos las correlaciones de las medidas de bienestar personal y laboral para los distintos escenarios (Ver Figuras ??, ?? y ??). Al considerar las variables de bienestar por si sola (es decir, considerar las medidas de bienestar florecimiento, agotamiento, compromiso de trabajo y satisfacción laboral independientemente), se aprecia que las variables se encuentran fuertemente correlacionadas ($\rho > 0.7$) con ellas mismas para los distintos instantes de tiempos. Las Figuras ?? y ?? indican que el florecimiento está fuertemente correlacionado con compromiso de trabajo y satisfacción laboral, caso contrario sucede con el agotamiento, donde las variables están correlacionadas de manera inversa (ver Figura ??).

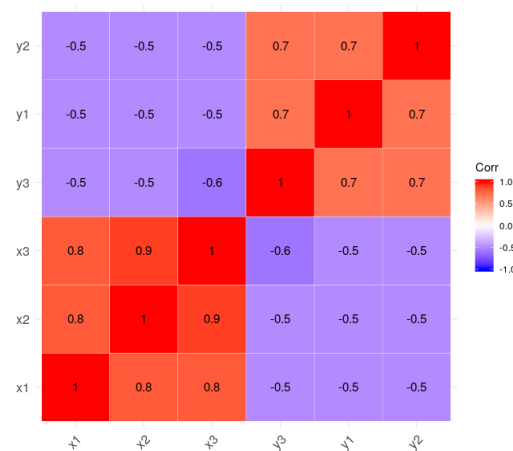


FIGURA 4.2: Matriz de correlación para el florecimiento y el agotamiento durante los distintos instantes de tiempo.

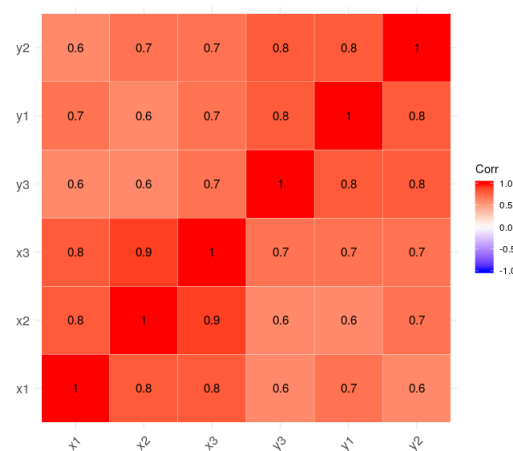


FIGURA 4.3: Matriz de correlación para el florecimiento y el compromiso de trabajo durante los distintos instantes de tiempo.

Con esta información, procedemos a aplicar los modelos estudiados en el capítulo 3: los modelos CLPM y los modelos RI-CLPM. La implementación de los modelos se

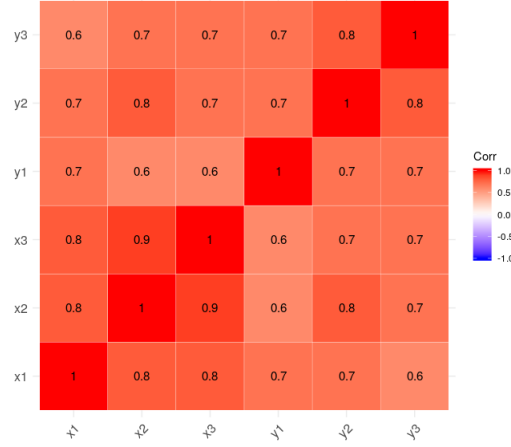


FIGURA 4.4: Matriz de correlación para el florecimiento y la satisfacción laboral durante los distintos instantes de tiempo.

realiza en el lenguaje de programación R. Dentro de este lenguaje, nos centramos en las librerías **Laavan** (librería especializada en problemas de ecuaciones estructurales y en modelos de regresión multinivel) y **ggplot2** (librería especializada en la visualización de datos). Las rutinas se encuentran disponibles en Rutina 1 (Anexo B). Para ajustar los distintos modelos es necesario definir el par ordenado $(\mathbf{X}_i, \mathbf{Y}_i)$, donde \mathbf{X}_i es el bienestar personal medido en los instantes $i = 1, 2, 3$ e \mathbf{Y}_i es el bienestar personal medido en los instante $i = 1, 2, 3$. En este caso, tendremos los siguientes casos: a) florecimiento versus agotamiento, b) florecimiento versus compromiso de trabajo y c) florecimiento versus satisfacción laboral. Basado en los resultados obtenidos sobre el ajuste de los modelos para los distintos escenarios, tanto los efectos autorregresivos como los efectos cruzados (ver Capítulo 3), son significativos (es decir, los valores de $|\beta| > 0.7$, donde $-1 \leq \beta \leq 1$). Esto es un resultado bastante interesante, ya que ambos modelos (para los distintos escenarios) nos muestran la existencia de una relación recíproca entre el bienestar personal y el desempeño laboral, sin embargo, debemos corroborar que la inclusión del efecto intraclase (es decir, el efecto del mismo individuo en el tiempo) del modelo RI-CLPM nos proporciona una mejora respecto al modelo CLPM, el cual deja aislado este efecto. Cabe recordar, que los modelos CLPM y RI-CLPM son modelos anidados, puesto que los modelos CLPM son un caso particular del modelo RI-CLPM cuando $\kappa_i = \omega_i = 0$. Para comparar que modelo es mejor uno respecto del otro, realizamos un test de diferencia chi-cuadrado, donde la hipótesis nula del test es $H_0: \Sigma_{clpm} = \Sigma_{riclpm}$. Los resultados de las Tablas 4.1, 4.2 y 4.3, muestran que ambos modelos (para los distintos escenarios) son significativamente distintos entre ellos (basándonos en el p-valor), por otro lado, los criterios de AKAIKE

y BIC nos muestra que el ajuste de los datos mediante los modelos RI-CLPM son mejores respecto de CLPM, lo que significa que los efectos intra-individuo mejoran los resultados, en vez de dejar estos efectos aislados. Por otro lado, los modelos RI-CLPM considera el efecto intra-individua como la media del individuo para los distintos instantes de tiempo. Este comportamiento a nivel individual se podría complejizar, por ejemplo, agregando una pendiente que dependa del tiempo más un intercepto de la media individual (Modelos ALT). Queda como trabajo futuro implementar y comparar las nuevas alternativas a los modelos CLPM (ver Capítulo 2).

CUADRO 4.1: Test de diferencias chi cuadrado al comparar florecimiento versus agotamiento .

Ajuste	df	AIC	BIC	Chisq	ChiDiff	DfDiff	Pr(> Chisq)
RI-CLPM	1	7994.7	8113.1	0.000			
CLPM	4	8063.2	8167.9	74.494	74.494	3	4.651e-16

CUADRO 4.2: chi cuadrado al comparar florecimiento versus satisfacción laboral.

Ajuste	df	AIC	BIC	Chisq	ChiDiff	DfDiff	Pr(> Chisq)
RI-CLPM	1	6910.3	7028.7	4.4677			
CLPM	4	6975.9	7080.7	76.0704	71.603	3	1.936e-15

CUADRO 4.3: chi cuadrado al comparar florecimiento versus compromiso de trabajo.

Ajuste	df	AIC	BIC	Chisq	ChiDiff	DfDiff	Pr(> Chisq)
RI-CLPM	1	8995.9	9114.4	0.9159			
CLPM	4	9073.4	9178.2	84.3925	83.477	3	2.2e-16

Capítulo 5

Conclusiones y Trabajos Futuros

En el presente artículo se estudio a profundidad los problemas encontrado por Rogosa (1980) sobre los modelos CLPM, en donde se pone en manifiesto los problemas tanto a nivel de muestreo como de interacción de variables. Por otro lado, se propone una alternativa denominada modelos RI-CLPM. Los modelos RI-CLPM mejoran las interacción de las variables, siendo consistente con la teoría del cambio (Capítulo 2) y captando las interacciones a nivel intraindividual e inter clase. Por otro lado se muestra que los modelos CLPM son un caso particular de los modelos RI-CLPM cuando $\kappa_i = \omega_i = 0$ (Capítulo 3). Finalmente, para analizar ambos modelos, realizamos una aplicación sobre un conjunto de trabajadores chilenos, midiendo su escala de bienestar personal y laboral basado en distintas medidas (florecimiento, agotamiento, compromiso de trabajo y satisfacción laboral). Se realiza un análisis descriptivo del conjunto de datos, para luego dar paso al ajuste de los distintos modelos. Los resultados muestran que a nivel de modelo, los ajustes similares, vale decir, que los parámetros autorregresivos tienen sentido con lo se esperaba. Por otro lado, al comparar ambos modelos, el modelo RI-CLPM se ajusta de mejor manera respecto a los modelos CLPM tradicional (bajo un test de diferencia chi cuadrado y analizando los estadísticos AKAIKE Y BIC). A modo de trabajos futuros, se espera analizar el bienestar personal y laboral bajo distintas medidas. Se busca estudiar a fondo algunos modelos propuestos en el Capítulo 2, a modo de mejorar los resultados de los modelos CLPM. Además, se estaría interesado en comprender la complejidad computacional para este tipo de problemas pero para un conjunto de datos de mayor volumen, puesto que el conjunto de datos estudiado tiene una orden de magnitud de $\sim 10^2$.

Bibliografía

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trarrs. Autom. Control*, (AC-19):716–723.
- Akaike, H., 1987. Factor analysis and aic. *Psychometrika*, (52):317–332.
- Anderson, J. A., 1977. Multivariate methods in clinical assessment,. *Journal of the Royal Statistical Society*., (139):161–182.
- Bentler, P. M., 1982. Confirmatory factor analysis via noniterative estimation: A fast, inexpensive method. *International Journal of Research in Marketing*, (19):417–424.
- Bozdogan, H., 1987. Model selection and akaike’s Information criterion (aic): The general theory and its analytical extensions. *Psychometrika*., (52):345–370.
- Genton, M. G. y Kleiber, W., 2015. Cross-covariance functions for multivariate geostatistics. *Statist. Sci.*, 30(2):147–163.
- Jiirekog, K. G. . y Goldherger, A. S., 1972. Factor analysis by generalized least squares. *Psychometrika*., (37):243–260.
- Joreskog, G., K y Lawley, D. N., 1968. New methods in maximum likelihood, factor analysis. *British Journal of Mathematical and Statistical Psychology*, (21):85–96.
- Joreskog, K. G., 1977. “factor analysis by least squares and maximum-likelihood methods. *Statistical Methods for Digital Computer*, 3:125–153.
- Joreskog, K. G., 1962. On the statistical treatment of residuals in factor analysis. *Psychometrika*, (27):335–354.
- Joreskog, K. G., 1963. *Statistical Estimation in Factor Analysis*. Almqvist & Wiksell, Stockholm.
- Joreskog, K. G., 1973. A general method for estimating a linear structural equation system. *Structural Equation Models in the Social Sciences*, 85–112.

- Joreskog, K. G., 1981. Analysis of covariance structures. *Scandinavian Journal of Statistics*, (8):65–92.
- Lawley, D. N., 1940. The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh*, (60):64–82.
- Lawley, D. N., 1941. Further investigation in factor estimation. *Proceedings of the Royal Society of Edinburgh*, (61):176–185.
- Lawley, D. N., 1942. Further investigations in factor estimation. *Proceedings of the Royal Society of Edinburgh*, (62):176–185.
- Lawley, D. N., 1953. A modified method of estimation in factor analysis and some large sample results. *Uppsala Symposium on Psychological Factor Analysis: Nordisk Psykologi Monograph*, (3):35–42.
- Lawley, D. N. y Maxwell, A. E., 1971. *Factor Analysis as a Statistical Method*. Butterworths, London.
- Lawley, D. N. y Maxwell, A. E., 1973. Regression and factor analysis. *Biometrika*, (60):331–338.
- Solari, M. E., 1969. The ‘maximum likelihood solution of the problem of estimating a linear functional relationship,. *Journal of the Royal Statistical Society*, (31):372–375.
- Wright, S., 1921. Correlation and causation. *Journal of Agricultural Research*, (20):557–585.
- Wright, S., 1934. The method of path coefficients. *Annals of Mathematical Statistics*, (5):557–585.

Anexo A

A.1. Lenguaje de Programación R

A.1.1. Lenguaje R y su entorno

R (R Development Core Team, 2010; ver también Ihaka y Gentleman, 1996) es un lenguaje y entorno para computación y gráficos estadísticos basados en el lenguaje y el entorno S (Becker & Chambers, 1984; ahora disponible en el programa TIBCO Spotfire S +; TIBCO Software Inc., 2010). El código fuente de R está libremente disponible bajo la Licencia Pública General de GNU (GPL) en <http://www.r-project.org>. Existen binarios precompilados para distintas plataformas, tales como: Unix, Microsoft Windows y Mac. R es altamente extensible gracias a más de 4.000 librerías disponibles, llamadas paquetes, que representan principalmente funciones especializadas del lenguaje R.

Las principales ventajas de R son que está disponible de forma gratuita, es extremadamente dinámico en su desarrollo (varios análisis estadísticos solo se pueden calcular en R), tiene excelentes capacidades gráficas, tiene una buena sistema de ayuda incorporado (basta con escribir ? tema cuando se busca ayuda para un tema), permite funciones escritas por el usuario (que luego pueden compartirse con la comunidad R general) y está orientado a objetos (elementos de operaciones) se crean para existir de forma independiente y los comandos generales se adaptan a la naturaleza de todos los elementos). Como consecuencia, todos los paquetes se benefician de estas características al atribuir porciones complicadas de sintaxis a los objetos. La tarea del usuario se simplifica enormemente y consiste en aprender los objetos de los paquetes, sin tener que preocuparse por la sintaxis subyacente. Aunque R usa una interfaz de línea de comandos, lo

que podría desalentar a algunos novatos para que la adopten, hay varias buenas interfaces gráficas de usuario disponibles. Estos facilitan enormemente el aprendizaje de los conceptos básicos del lenguaje R (véase http://www.sciviews.org_rgui/index.html).

A.1.2. Paquetes de modelación de ecuaciones estructurales en R

Tres paquetes son particularmente útiles para estimar modelos de ecuaciones estructurales en R: *sem*, disponibles a través del sitio web R; *lavaan* (para el análisis de variables latentes), disponible a través del sitio web R, pero también en <http://lavaan.ugent.be>, y *OpenMx*, (la versión R de Mx; Neale, Boker, Xie y Maes, 2006), disponible en <http://openmx.psyc.virginia.edu/>.

Los tres son de código abierto y adaptaron el lenguaje R para estimar una gran variedad de modelos de ecuaciones estructurales. Por lo tanto, los tres paquetes no solo son atractivos porque son de código abierto, sino también porque permiten una fácil integración de sus resultados en el entorno R para futuras operaciones. También aceptan objetos R para integrarse en su sintaxis. *sem* fue desarrollado por John Fox, quien lo mantiene, con contribuciones de Adam Kramer y Michael Friendly; *lavaan* es desarrollado y mantenido por Yves Rosseel; *OpenMx* está financiado por el Instituto Nacional de Salud e involucra a un equipo de varias personas: Steve Boker (investigador principal); Michael Neale, Hermine Maes y Michael Wilde (investigadores co-principales); y Tim Brick, Jeff Spies, Michael Spiegel, Ryne Eastabrook, Michael Hunter, Sarah Kenny, Paras Mehta, Timothy Bates, John Fox y Zhiyong Zhang. Proporcionamos una discusión sucinta de cada paquete solamente. Se anima al lector interesado a explorar las referencias de cada paquete para obtener más información.

El paquete *sem* en R

En el momento de escribir esto usamos la versión 0.9-21 (Fox, 2010). Fox describió su paquete en un número anterior de esta revista (Fox, 2006, donde también ofrece una introducción a R). Este fue el primer paquete R para SEM y permite estimar parámetros de ecuaciones estructurales de variables manifiestas por mínimos cuadrados de dos etapas y de modelos de ecuaciones estructurales generales por máxima verosimilitud (ML). En el momento de escribir este documento, el paquete no permite cálculos en el caso ML; es decir, una función de verosimilitud en el nivel individual

(datos brutos), en lugar de en el grupo (matriz de momentos). Por lo tanto, los datos incompletos solo pueden manejarse con un enfoque de grupos múltiples al aplicar ML a la matriz de covarianzas de cada grupo (si el número de patrones es bajo, ver más adelante) o por imputación (varios paquetes a este efecto están disponibles en R).

El paquete `sem` utiliza la formulación del modelo de acción reticular (RAM, McArdle, 1980; véase también Boker et al., 2002), que distingue tres elementos de cualquier SEM: una matriz con flechas de una sola cabeza (pesos de regresión y carga de factores), una matriz con flechas de dos cabezas (varianzas y covarianzas) y una matriz de filtro, en la que se distinguen las variables manifiesta y latente. Si los medios también se incluyen en un modelo, se debe analizar una matriz de suma de cuadrados y productos cruzados en lugar de una matriz de covarianza. El paquete ofrece una función para este efecto. Las variables endógenas y exógenas no se distinguen, ni el orden cero de las varianzas residuales.

El análisis de varios grupos no está integrado en la versión actual de `sem`. Sin embargo, si los grupos son idénticos en tamaño, se pueden analizar dentro de `sem` utilizando la noción de modelos apilados, tal como lo describe Evermann (2010). Por lo tanto, los datos incompletos en una serie de patrones limitados solo pueden tratarse mediante la estimación de ML si los grupos son iguales en tamaño.

`sem` produce archivos de puntos para generar diagramas en un programa externo (por ejemplo, GraphViz, un software de visualización de gráficos de código abierto disponible en <http://www.graphviz.org/>). La sintaxis de punto se puede editar para implementar cambios en la apariencia. Producimos las Figuras 1 y 2 editando su sintaxis de puntos.

El paquete de lavaan en R

En el momento de escribir esto usamos la versión 0.4-7 (Rosseel, 2011). Este paquete es más reciente y permite más opciones que `sem`. Hay tres estimadores disponibles para datos continuos: ML, mínimos cuadrados generalizados y mínimos cuadrados ponderados (también llamada distribución asintótica libre). Además, para ML están disponibles cuatro métodos de estimación de errores estándar (SE): el SE convencional basado en la inversión de la matriz de información (ya sea observada o esperada), SE robusta (con una estadística chi-cuadrado escalada de Satorra-Bentler), basada en SE en derivados de primer orden, y robusto Huber-White SE (con una estadística

escalada de Yuan-Bentler). lavaan también contiene soporte completo para el análisis de medios / interceptos y grupos múltiples. La sintaxis de lavaan es bastante simple y contiene una serie de funciones populares. Por ejemplo, las funciones dedicadas se definen para análisis factoriales confirmatorios, modelos de curvas latentes, varios grados de invarianza, etc. Cada función dedicada implementa un conjunto específico de valores predeterminados. Sin embargo, también hay una función que evita todos los valores predeterminados (esta es una característica deseable en modelos avanzados de ecuaciones estructurales). Por último, el usuario puede solicitar que la salida aparezca en formato Mplus (Muthén & Muthén, 1998-2010) o EQS (Bentler, 1995).

Al momento de escribir este artículo, lavaan no permite el análisis de variables categóricas o censuradas, modelos de mezclas y datos multinivel. Según el autor, estas características deberían incluirse en el futuro cercano. El sitio web de Lavaan ofrece una breve descripción del paquete, la sintaxis R mínima necesaria para ejecutar un modelo de ecuaciones estructurales, un ejemplo de modelo de ecuación estructural simple, pero también un tutorial muy útil. El paquete no produce sintaxis para dibujar diagramas. Sin embargo, el paquete psych (Revelle, 2010) incluye una función para este fin.

El paquete OpenMx en R

Este paquete se describe en Boker et al. (2011). En el momento de escribir esto usamos la versión 1.0.6-1581 (OpenMx Development Team, 2010). Este paquete aún no se encuentra en el sitio web de R debido a una restricción de licencia en una parte del código (el optimizador de NPSOL, que no es de código abierto). Según los autores, esto debería remediarse en el futuro cercano.

Este paquete es sin duda el más completo en R para SEM y esto podría deberse principalmente a dos razones. En primer lugar, OpenMx se basa en el software Mx ya disponible, que fue pionero en varias funciones avanzadas de SEM (Neale et al., 2006). Estos incluyen la implementación del algoritmo ML a nivel de datos individuales (ML mayúsculas), análisis de variables no continuas, análisis multinivel y la posibilidad de incluir cualquier cálculo de álgebra matricial dentro del modelo (una característica muy poderosa para el tipo no lineal o de cualquier tipo) de la restricción algebraica de parámetros) y para especificar la propia función de ajuste (además del ML ya existente y ML de información completa). De hecho, Mx y OpenMx no son solo paquetes para SEM, sino más generalmente para la optimización del álgebra matricial.

La segunda razón que diferencia este paquete de los dos anteriores es que, gracias a su financiación, OpenMx es mantenido por un equipo grande y muy activo que también creó un sitio web muy rico con mucha documentación, varios tutoriales, presentaciones didácticas, wikis, y foros.

Juntas, estas características permiten una retroalimentación inmediata a los usuarios y hacen de este uno de los paquetes de software SEM más completos disponibles, incluso cuando se compara con contrapartes comerciales.

A.1.3. Tutorial Lavaan

Si eres nuevo en lavaan, este es el lugar para comenzar. En este tutorial, presentamos los componentes básicos de lavaan: la sintaxis del modelo, las funciones de ajuste (cfa, sem y crecimiento) y las funciones principales del extractor (resumen, coef, tted, inspeccionar). Después de haber proporcionado dos ejemplos simples, discutiremos brevemente algunos temas importantes: estructuras de medios, grupos múltiples, modelos de curvas de crecimiento, análisis de mediación y datos categóricos. En el camino, esperamos darte la información suficiente para que comiences (pero nada más).

Antes de comenzar

Antes de comenzar, lea estos puntos atentamente:

- a) En primer lugar, debe tener una versión reciente (3.0.0 o superior) de R instalada. Puede descargar la última versión de R desde esta página: <http://cran.r-project.org/>.
- b) El paquete de lavaan aún no está terminado. Pero ya es muy útil para la mayoría de los usuarios, o eso esperamos. Sin embargo, algunas características importantes que actualmente NO están disponibles en lavaan son: - soporte completo para conjuntos de datos jerárquicos / multinivel (cfa multinivel, sem multinivel); sin embargo, la versión 0.6-1 admite cfa / sem de dos niveles solo con intercepciones aleatorias, para datos completos continuos -soporte completo para el muestreo (y otros) ponderaciones (0.6-1 tiene ponderaciones de muestreo, pero solo para datos completos continuos y no agrupados) -support para variables discretas latentes (modelos de mezcla, clases latentes) -La estimación bayesiana (pero vea el paquete blavaan)

Esperamos agregar estas características a lavaan en el futuro cercano (pero por favor no pregunte cuándo).

- c) Consideramos que la versión actual es un software beta. Esto NO significa que no puede confiar en los resultados. Creemos que los resultados son precisos. Significa que las cosas pueden cambiar cuando salgan nuevas versiones. Por ejemplo, podemos cambiar el nombre de los argumentos en llamadas a funciones. Y cambiamos las partes internas del código fuente constantemente. Sin embargo, la sintaxis del modelo es bastante madura y se ha mantenido estable durante un tiempo.
- d) No esperamos que sea un experto en R. De hecho, el paquete lavaan está diseñado para ser utilizado por usuarios que normalmente nunca usarían R. Sin embargo, puede ser útil familiarizarse un poco con R, solo para sentirse cómodo con eso. Quizás la habilidad más importante que deba aprender es cómo importar sus propios conjuntos de datos (quizás en formato SPSS) en R. Hay muchos tutoriales en la web que le enseñan exactamente eso. Una vez que tenga sus datos en R, puede comenzar a especificar su modelo. Hemos intentado hacer todo lo posible para que los usuarios puedan utilizar sus modelos. Por supuesto, si tiene sugerencias sobre cómo podemos mejorar las cosas, háganoslo saber.
- e) Este documento está escrito para los usuarios del paquete lavaan por primera vez. No es un manual de referencia, ni contiene material técnico sobre cómo se hacen las cosas en el paquete de Lavaan. Estos documentos están actualmente en preparación.
- f) El paquete lavaan es un software libre de código abierto. Esto significa (entre otras cosas) que no hay garantía alguna.

Instalación del paquete

Desde mayo de 2010, el paquete lavaan está disponible en CRAN. Por lo tanto, para instalar lavaan, simplemente inicie R y escriba:

```
1 install.packages("lavaan", dependencies=TRUE)
```

Puede verificar si la instalación fue exitosa escribiendo

```
1 library(lavaan)
```

Cuando se carga el paquete, se mostrará un mensaje de inicio que muestra el número de versión y un recordatorio de que se trata de un software beta:

This is lavaan 0.6-1 lavaan is BETA software! Please report any bugs.

Si ves este mensaje, estás listo para comenzar.

Sintax del modelo

En el corazón del paquete de lavaan está la ‘sintaxis del modelo’. La sintaxis del modelo es una descripción del modelo a estimar. En esta sección, explicamos brevemente los elementos de la sintaxis del modelo lavaan. Se dan más detalles en los ejemplos que siguen. En el entorno R, una fórmula de regresión tiene la siguiente forma:

```
1 y ~ x1 + x2 + x3 + x4
```

En esta fórmula, la tilde (\sim) es el operador de regresión. En el lado izquierdo del operador, tenemos la variable dependiente (y), y en el lado derecho, tenemos las variables independientes, separadas por el operador $+$. En lavaan, un modelo típico es simplemente un conjunto (o sistema) de fórmulas de regresión, donde algunas variables (comenzando con una ‘f’ a continuación) pueden estar latentes. Por ejemplo:

```
1 y ~ f1 + f2 + x1 + x2
2 f1 ~ f2 + f3
3 f2 ~ f3 + x1 + x2
```

Si tenemos variables latentes en cualquiera de las fórmulas de regresión, debemos ‘de ne ’ellos enumerando sus indicadores (manifiestos o latentes). Hacemos esto usando el operador especial $=$, que se puede leer como se mide por. Por ejemplo, para definir las tres variables latentes f1, f2 y f3, podemos usar algo como:

```
1 f1 =~ y1 + y2 + y3
2 f2 =~ y4 + y5 + y6
3 f3 =~ y7 + y8 + y9 + y10
```

Además, las varianzas y las covarianzas se especifican usando un operador ‘double tilde’, por ejemplo:

```
1 y1 ~~ y1 # variance
2 y1 ~~ y2 # covariance
3 f1 ~~ f2 # covariance
```

Y, por último, las interceptaciones para variables observadas y latentes son fórmulas de regresión simples con solo un intercepto (indicado explícitamente por el número ‘1’) como el único predictor:

1	$y^1 \sim 1$
2	$f^1 \sim 1$

Usando estos cuatro tipos de fórmula, se puede describir una gran variedad de modelos de variables latentes. El conjunto actual de tipos de fórmula se resume en la tabla a continuación.

Formula tipo	Operador	Significado
Definición de variable latente	$=\sim$	se mide por
Regresión	\sim	tiene una relación con
(residuo) covarianza	$\sim\sim$	se correlaciona con
intercepto	~ 1	intercepto

Anexo B

B.2. Rutinas en R

```
1 # -----
2 # Generate dataset for models
3 # -----
4
5 #####
6 # 1.- load libraries ###
7 #####
8
9 library(foreign)
10 library(lavaan)
11 library(psych)
12
13 #####
14 # 2.- aux. functions ###
15 #####
16
17 # load db
18
19 show_data_as_df <- function(string_name){
20   df <- read.spss(string_name,
21                   header = TRUE,
22                   to.data.frame=TRUE)
23 }
24
25 chilean_woker <- function(df){
26   FILTRO <- c("FILT_1C")
27   mask <- df[,FILTRO] == 1
28   df_work <- df[mask,]
29
30   return(df_work)
31 }
32 }
33
34 # selecct specifics columns
35
36 select_columns <- function( df,columns){
37   df_work <- df[,columns]
38
39   return(df_work)
40 }
41
42 # cronbach alpha per wave
43
44 cronbach_alpha_wave <- function(df,columns){
45
46   df1 <- df[,columns[grepl("_1", columns)]]
47   df2 <- df[,columns[grepl("_2", columns)]]
48   df3 <- df[,columns[grepl("_3", columns)]]
49
50   v1 <- alpha(df1)$total$raw_alpha
51   v2 <- alpha(df2)$total$raw_alpha
52   v3 <- alpha(df3)$total$raw_alpha
```

```

53 |
54 |   return(c(v1,v2,v3))
55 | }
56 |
57 | # mean df with nan values
58 |
59 | mean_nan <- function(df){
60 |
61 |   l_aux <- c()
62 |   n <- dim(df)[1]
63 |   m <- dim(df)[2]
64 |
65 |   for(i in 1:n){
66 |     num <- mean(as.numeric(df[i,]))
67 |     l_aux <- c(l_aux,num)
68 |
69 |   }
70 |
71 |
72 |   return(l_aux)
73 | }
74 |
75 | # create df with mean
76 |
77 | df_mean_nan <- function(df,columns,name_columns){
78 |   df_work <- df[,columns]
79 |   df1 <- df_work[,columns[grepl("_1", columns)]]
80 |   df2 <- df_work[,columns[grepl("_2", columns)]]
81 |   df3 <- df_work[,columns[grepl("_3", columns)]]
82 |
83 |   c1_mean <- mean_nan(df1)
84 |   c2_mean <- mean_nan(df2)
85 |   c3_mean <- mean_nan(df3)
86 |
87 |   df_mean <- data.frame(c1_mean,c2_mean,c3_mean)
88 |   colnames(df_mean) <- c(paste(name_columns,'_1',sep=''),
89 |                          paste(name_columns,'_2',sep=''),
90 |                          paste(name_columns,'_3',sep=''))
91 |
92 |   return(df_mean)
93 |
94 |
95 | }
96 |
97 | # crete df to working
98 |
99 | df_model <- function(df1,df2){
100 |   df <- data.frame(df1,df2)
101 |   colnames(df) <- c('x1','x2','x3','y1','y2','y3')
102 |   return(df)
103 | }
104 |
105 | # clpm model
106 |
107 | clpm_model <- function(df){
108 |
109 |   # fit CLPM model
110 |
111 |   clpmModel <- #yes, "Model" is redundant
112 |   ,
113 |   #Note, the data contain x1-3 and y1-3
114 |   #Latent mean Structure with intercepts
115 |
116 |   kappa =~ 1*x1 + 1*x2 + 1*x3
117 |   omega =~ 1*y1 + 1*y2 + 1*y3
118 |
119 |   x1 ~ mu1*1 #intercepts
120 |   x2 ~ mu2*1
121 |   x3 ~ mu3*1
122 |   y1 ~ pi1*1
123 |   y2 ~ pi2*1
124 |   y3 ~ pi3*1
125 |
126 |   kappa ~~ 0*kappa #variance nope

```

```

127 omega ~~ 0*omega #variance nope
128 kappa ~~ 0*omega #covariance not even
129
130 #latent vars for AR and cross-lagged effects
131 p1 =~ 1*x1 #each factor loading set to 1
132 p2 =~ 1*x2
133 p3 =~ 1*x3
134 q1 =~ 1*y1
135 q2 =~ 1*y2
136 q3 =~ 1*y3
137
138 p3 ~ alpha3*p2 + beta3*q2
139 p2 ~ alpha2*p1 + beta2*q1
140
141 q3 ~ delta3*q2 + gamma3*p2
142 q2 ~ delta2*q1 + gamma2*p1
143
144 p1 ~~ p1 #variance
145 p2 ~~ u2*p2
146 p3 ~~ u3*p3
147 q1 ~~ q1 #variance
148 q2 ~~ v2*q2
149 q3 ~~ v3*q3
150
151 p1 ~~ q1 #p1 and q1 covariance
152 p2 ~~ q2 #p2 and q2 covariance
153 p3 ~~ q3 #p2 and q2 covariance'
154
155 fit <- lavaan(clpmModel, data = df,missing="ml")
156 return(fit)
157 }
158
159 # ri-clpm model
160
161 ri_clpm_model <- function(df){
162   # fit RI-CLPM model
163
164   riclpmModel <-
165     '
166     #Note, the data contain x1-3 and y1-3
167     #Latent mean Structure with intercepts
168
169     kappa =~ 1*x1 + 1*x2 + 1*x3
170     omega =~ 1*y1 + 1*y2 + 1*y3
171
172     x1 ~ mu1*1 #intercepts
173     x2 ~ mu2*1
174     x3 ~ mu3*1
175     y1 ~ pi1*1
176     y2 ~ pi2*1
177     y3 ~ pi3*1
178
179     kappa ~~ kappa #variance
180     omega ~~ omega #variance
181     kappa ~~ omega #covariance
182
183     #latent vars for AR and cross-lagged effects
184     p1 =~ 1*x1 #each factor loading set to 1
185     p2 =~ 1*x2
186     p3 =~ 1*x3
187     q1 =~ 1*y1
188     q2 =~ 1*y2
189     q3 =~ 1*y3
190
191     #Later, we may constrain autoregression and cross-lagged
192     #effects to be the same across both lags.
193     p3 ~ alpha3*p2 + beta3*q2
194     p2 ~ alpha2*p1 + beta2*q1
195
196     q3 ~ delta3*q2 + gamma3*p2
197     q2 ~ delta2*q1 + gamma2*p1
198
199     p1 ~~ p1 #variance
200     p2 ~~ u2*p2

```



```

201 p3 ~~ u3*p3
202 q1 ~~ q1 #variance
203 q2 ~~ v2*q2
204 q3 ~~ v3*q3
205
206 p1 ~~ q1 #p1 and q1 covariance
207 p2 ~~ q2 #p2 and q2 covariance
208 p3 ~~ q3 #p2 and q2 covariance'
209
210 fit <- lavaan(riclpmModel, df,missing="ml")
211 return(fit)
212 }
213
214 # compare models
215
216 compare_ld <- function(model_riclpm,model_clpm){
217   return(anova(model_riclpm,model_clpm))
218 }
219 }
220
221
222 #####
223 # 3.- application ###
224 #####
225
226 # load original dataset
227
228 data_folder <- "/home/francisco/Escritorio/Universidad/Magister/Rutinas/R/codigos_tesis"
229 name <- paste(data_folder, '/Wave_Peggy_wiw2w3_v2.sav', sep='')
230 df <- show_data_as_df(name)
231
232
233 # select chilean worker
234
235 df_ch <- chilean_woker(df)
236
237 # def. important columns
238
239 ## Happiness
240
241 # a) PERMA
242
243 positive_emotions <- c("P1L_1","P2L_1","P3L_1","P1L_2","P2L_2","P3L_2","P1L_3","P2L_3","P3L_3")
244 engagement <- c("E1L_1","E2L_1","E3L_1","E1L_2","E2L_2","E3L_2","E1L_3","E2L_3","E3L_3")
245 relationships <- c("R1L_1","R2L_1","R3L_1","R1L_2","R2L_2","R3L_2","R1L_3","R2L_3","R3L_3")
246 meaning <- c("M1L_1","M2L_1","M3L_1","M1L_2","M2L_2","M3L_2","M1L_3","M2L_3","M3L_3")
247 accomplishment <- c("A1L_1","A2L_1","A3L_1","A1L_2","A2L_2","A3L_2","A1L_3","A2L_3","A3L_3")
248
249 PERMA <- c(positive_emotions,engagement,relationships,meaning,accomplishment)
250
251
252 ## work
253
254 # a) BURNOUT
255
256 emotional_exhaustion <- c("EE1_1","EE2_1","EE3_1","EE4_1", "EE5_1" ,
257   "EE1_2","EE2_2","EE3_2","EE4_2", "EE5_2" ,
258   "EE1_3","EE2_3","EE3_3","EE4_3", "EE5_3" )
259
260 depersonalization <- c("DP1_1", "DP2_1", "DP3_1",
261   "DP1_2", "DP2_2", "DP3_2",
262   "DP1_3", "DP2_3", "DP3_3")
263
264 BURNOUT <- c(emotional_exhaustion,depersonalization)
265
266
267 # b) ENGAGEMENT
268
269 vigor <- c("VIG1_1", "VIG2_1", "VIG3_1", "VIG4_1", "VIG5_1", "VIG6_1",
270   "VIG1_2", "VIG2_2", "VIG3_2", "VIG4_2", "VIG5_2", "VIG6_2",
271   "VIG1_3", "VIG2_3", "VIG3_3", "VIG4_3", "VIG5_3", "VIG6_3")
272
273 dedication <- c("DED1_1", "DED2_1", "DED3_1", "DED4_1", "DED5_1",
274   "DED1_2", "DED2_2", "DED3_2", "DED4_2", "DED5_2",

```

```

275         "DED1_3", "DED2_3", "DED3_3", "DED4_3", "DED5_3")
276
277 ENGAGEMENT <- c(vigor,dedication)
278
279 # c) WORK SATISFACTION
280
281 WORK_SATISFACTION <- c("WSG1_1", "WSG1_2", "WSG1_3")
282
283
284 # cronbach alpha
285
286 # PERMA
287
288 positive_emotions_alpha <- cronbach_alpha_wave(df_ch,positive_emotions)
289 engagement_alpha <- cronbach_alpha_wave(df_ch,engagement)
290 relationships_alpha <- cronbach_alpha_wave(df_ch,relationships)
291 meaning_alpha <- cronbach_alpha_wave(df_ch,meaning)
292 accomplishment_alpha <- cronbach_alpha_wave(df_ch,accomplishment)
293 PERMA_alpha <- cronbach_alpha_wave(df_ch,PERMA)
294
295 ## work
296
297 # a) BURNOUT
298
299 emotional_exhaustion_alpha <- cronbach_alpha_wave(df_ch,emotional_exhaustion)
300 depersonalization_alpha <- cronbach_alpha_wave(df_ch,depersonalization)
301 BURNOUT_alpha <- cronbach_alpha_wave(df_ch,BURNOUT)
302
303 # b) ENGAGEMENT
304
305 vigor_alpha <- cronbach_alpha_wave(df_ch,vigor)
306 dedication_alpha <- cronbach_alpha_wave(df_ch,dedication)
307 ENGAGEMENT_alpha <- cronbach_alpha_wave(df_ch,ENGAGEMENT)
308
309 # c) WORK SATISFACTION
310
311 #WORK_SATISFACTION_alpha <- cronbach_alpha_wave(df_ch,WORK_SATISFACTION)
312
313
314 # create df with new variables
315
316 df_perma_mean <- df_mean_nan(df_ch,PERMA,'perma_mean')
317 df_burnout_mean <- df_mean_nan(df_ch,BURNOUT,'burnout_mean')
318 df_engagement_mean <- df_mean_nan(df_ch,ENGAGEMENT,'engagement_mean')
319 df_work_satisfaction_mean <- select_columns(df_ch,WORK_SATISFACTION)
320
321 # create df to working CLPM - RICLPM
322
323 df_model1 <- df_model(df_perma_mean,df_burnout_mean)
324 df_model2 <- df_model(df_perma_mean,df_engagement_mean)
325 df_model3 <- df_model(df_perma_mean,df_work_satisfaction_mean)
326
327
328 # apply models
329
330 # A) PERMA versus BURNOUT
331
332 # clpm model
333 fit1 <- ri_clpm_model(df_model1)
334 summary(fit1, standardized = TRUE)
335
336 # ri-clpm model
337
338 fit2 <- clpm_model(df_model1)
339 summary(fit2, standardized = T)
340
341 # compare models
342
343 c1 <- compare_ld(fit1,fit2)
344 c1
345
346 # B) PERMA versus ENGAGEMENT
347
348 # clpm model

```

```

349 fit3 <- ri_clpm_model(df_model2)
350 summary(fit3, standardized = T)
351
352 # ri-clpm model
353
354 fit4 <- clpm_model(df_model2)
355 summary(fit4, standardized = T)
356
357 # compare models
358
359 c2 <- compare_ld(fit3, fit4)
360 c2
361
362 # C) PERMA versus WORK SATISFACTION
363
364 # clpm model
365 fit5 <- ri_clpm_model(df_model3)
366 summary(fit5, standardized = T)
367
368 # ri-clpm model
369
370 fit6 <- clpm_model(df_model3)
371 summary(fit6, standardized = T)
372
373 # compare models
374
375 c3 <- compare_ld(fit5, fit6)
376 c3
377
378
379 # graficos
380
381 library('ggplot2')
382
383 # a ) conteo
384
385 name_list <- c('WORK_SATISFACTION', 'BURNOUT', 'ENGAGEMENT', 'PERMA')
386 quantity_list <- c('1', '11', '8', '15')
387 type_list <- c('desempeno laboral', 'desempeno laboral', 'desempeno laboral', 'bienestar personal')
388
389 df_g1 <- data.frame(name_list, quantity_list, type_list)
390 colnames(df_g1) <- c('nombre', 'cantidad', 'tipo')
391 df_g1$nombre <- factor(df_g1$nombre, levels = df_g1$nombre)
392
393 p1 <- ggplot(data=df_g1, aes(x=nombre, y=cantidad, fill=tipo)) +
394   geom_bar(stat="identity") +
395   coord_flip()
396 p1
397
398
399 # b) box-plot
400
401 # florecimiento
402
403 grup_1 <- c(df_model1[, 'x1'], df_model1[, 'x2'], df_model1[, 'x2'])
404 name_1 <- c(rep('Y1', length(df_model1[, 'x1'])),
405            rep('Y2', length(df_model1[, 'x2'])),
406            rep('Y3', length(df_model1[, 'x3'])))
407
408 df_bp1 <- data.frame(grup_1, name_1)
409 p1 <- ggplot(df_bp1, aes(x=name_1, y=grup_1)) +
410   geom_boxplot() +
411   labs(x = "grupo", y = "valor")
412 p1
413
414 # agotamiento
415
416 grup_2 <- c(df_model1[, 'y1'], df_model1[, 'y2'], df_model1[, 'y2'])
417 name_2 <- c(rep('Y1', length(df_model1[, 'y1'])),
418            rep('Y2', length(df_model1[, 'y2'])),
419            rep('Y3', length(df_model1[, 'y3'])))
420
421 df_bp2 <- data.frame(grup_2, name_2)
422 p2 <- ggplot(df_bp2, aes(x=name_2, y=grup_2)) +

```

```

423     geom_boxplot() +
424     labs(x = "grupo", y = "valor")
425 p2
426
427
428 # compromiso de trabajo
429
430 grup_3 <- c(df_model2[, 'y1'], df_model2[, 'y2'], df_model2[, 'y2'])
431 name_3 <- c(rep('Z1', length(df_model1[, 'y1'])),
432            rep('Z2', length(df_model1[, 'y2'])),
433            rep('Z3', length(df_model1[, 'y3'])))
434
435 df_bp3 <- data.frame(grup_3, name_3)
436 p3 <- ggplot(df_bp3, aes(x=name_3, y=grup_3)) +
437     geom_boxplot() +
438     labs(x = "grupo", y = "valor")
439 p3
440
441 # satisfacion laboral
442
443 grup_4 <- c(df_model3[, 'y1'], df_model3[, 'y2'], df_model3[, 'y2'])
444 name_4 <- c(rep('W1', length(df_model1[, 'y1'])),
445            rep('W2', length(df_model1[, 'y2'])),
446            rep('W3', length(df_model1[, 'y3'])))
447
448
449 df_bp4 <- data.frame(grup_4, name_4)
450 p4 <- ggplot(df_bp4, aes(x=name_4, y=grup_4)) +
451     geom_boxplot() +
452     labs(x = "grupo", y = "valor")
453 p4
454
455 # c) matriz de correlaciones
456
457 library(ggcorrplot)
458
459 corr <- round(cor(df_model1[complete.cases(df_model1), ]), 1)
460 ggcorrplot(corr, hc.order = TRUE, outline.col = "white", lab = TRUE)
461
462 corr2 <- round(cor(df_model2[complete.cases(df_model2), ]), 1)
463 ggcorrplot(corr2, hc.order = TRUE, outline.col = "white", lab = TRUE)
464
465 corr3 <- round(cor(df_model3[complete.cases(df_model3), ]), 1)
466 ggcorrplot(corr3, hc.order = TRUE, outline.col = "white", lab = TRUE)
467
468
469 #####

```

LISTING 1: Rutinas para ajustes de los modelos CLPM Y RI-CLPM.

Anexo C

C.3. Cuestionario realizado a los distintos trabajadores

C.3.1. Preguntas relacionadas con el florecimiento

Variable	Enfoque	Pregunta
perma life	emociones positivas	P1: ¿En general, con qué frecuencia se siente alegre?
perma life	emociones positivas	P2: ¿En general, con qué frecuencia se siente positivo(a)?
perma life	emociones positivas	P3: ¿En general, en qué medida se siente usted contento(a)?
perma life	compromiso	E1: ¿Con qué frecuencia logra usted sentirse concentrado/inmerso en lo que está haciendo?
perma life	compromiso	E2: ¿En general, en qué medida se siente entusiasmado(a) e interesado(a) en las cosas?
perma life	compromiso	E3: ¿Qué tan seguido se le pasa el tiempo "volando" haciendo cosas que disfruta?
perma life	relaciones	R1: ¿En qué medida usted recibe ayuda y apoyo de otros(as) personas cuando la necesita?

Variable	Enfoque	Pregunta
perma life	relaciones	R2: ¿En qué medida se siente usted querido(a)?
perma life	relaciones	R3: ¿Qué tan satisfecho(a) está con sus relaciones personales?
perma life	sentido	M1: ¿En general, en qué medida lleva una vida con propósito y significado?
perma life	sentido	M2: ¿En general, en qué medida usted siente que lo que hace en la vida es valioso y vale la pena?
perma life	sentido	M3: ¿En qué medida usted generalmente siente que tiene un sentido de dirección en su vida?
perma life	logro	A1: ¿Qué parte del tiempo siente que está avanzando hacia el logro de sus metas?
perma life	logro	A2: ¿Con qué frecuencia alcanza las metas importantes que se ha propuesto para usted mismo(a)?
perma life	logro	A3: ¿Qué tan seguido es usted capaz de manejar sus responsabilidades?

CUADRO 1: Preguntas relacionadas con perma life.

C.3.2. Preguntas relacionadas con el agotamiento

Variable	Enfoque	Pregunta
burnout	agotamiento emocional	EE1: ¿Estoy emocionalmente agotado(a) por mi trabajo?
burnout	agotamiento emocional	EE2: ¿Estoy agotado(a) al final de un día de trabajo?
burnout	agotamiento emocional	EE3: ¿Estoy cansado(a) cuando me levanto por la mañana y tengo que afrontar otro día en mi puesto de trabajo?
burnout	agotamiento emocional	EE4: ¿Trabajar todo el día es una tensión para mí?
burnout	agotamiento emocional	EE5: ¿Estoy “desgastado(a)” por mi trabajo?
burnout	despersonalización	DP1: ¿He perdido interés por mi trabajo desde que empecé en este puesto?
burnout	despersonalización	DP2: ¿He perdido entusiasmo por mi trabajo?
burnout	despersonalización	DP3: ¿Quiero simplemente hacer mi trabajo y no ser molestado(a)?
burnout	despersonalización	DP4: ¿Me he vuelto más escéptico(a) respecto a la utilidad de mi trabajo?
burnout	despersonalización	DP5: ¿Dudo de la trascendencia y valor de mi trabajo?

CUADRO 2: Preguntas relacionadas con agotamiento.

C.3.3. Preguntas relacionadas con el compromiso

Variable	Enfoque	Pregunta
engagement	vigor	VIG1: ¿En mi trabajo me siento lleno(a) de energía?
engagement	vigor	VIG2: ¿Soy fuerte y vigoroso(a) en mi trabajo?
engagement	vigor	VIG3: ¿Cuando me levanto por las mañanas tengo ganas de ir a trabajar?
engagement	vigor	VIG4: ¿Puedo continuar trabajando durante largos períodos de tiempo?
engagement	vigor	VIG5: ¿Soy muy persistente en mi trabajo?
engagement	vigor	VIG6: ¿Incluso cuando las cosas no van bien, continuo trabajando?
engagement	dedicación	DED1: ¿Mi trabajo está lleno de significado y propósito?
engagement	dedicación	DED2: ¿Estoy entusiasmado(a) con mi trabajo?
engagement	dedicación	DED3: ¿Mi trabajo me inspira?
engagement	dedicación	DED4: ¿Estoy orgulloso del trabajo que hago?
engagement	dedicación	DED5: ¿Mi trabajo es desafiante?

CUADRO 3: Preguntas relacionadas con el compromiso.