

**Universidad de Valladolid**



Facultad de Ciencias

*Grado en Estadística*

Curso Académico: 2022-2023

TRABAJO DE FIN DE GRADO

# **Métodos Estadísticos en Psicología de la Educación**

Alumna:  
**Natalia Herrero Arias**

Tutores:  
**Eustasio del Barrio Tellado**  
**Juan Antonio Valdivieso Burón**



# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Estructura del documento . . . . .	2
<b>2. Marco Teórico</b>	<b>3</b>
2.1. Análisis de fiabilidad de las escalas de medida . . . . .	3
2.2. Análisis Factorial Confirmatorio . . . . .	4
2.2.1. Contexto Teórico . . . . .	4
2.2.2. Especificación del Modelo . . . . .	5
2.2.3. Matrices de varianzas/covarianzas implícita del Modelo y no restringida . . . . .	6
2.2.4. Determinación y Extracción de factores . . . . .	7
2.2.5. Rotación de factores . . . . .	8
2.2.6. Identificación del Modelo . . . . .	10
2.2.7. Evaluación de la idoneidad de la muestra . . . . .	10
2.3. Modelos de Ecuaciones Estructuras Generales . . . . .	11
2.3.1. Especificación del Modelo . . . . .	12
2.3.2. Efectos de los Componentes Estructurales . . . . .	13
2.3.3. Matrices de varianzas/covarianzas implícita del Modelo y no restringida . . . . .	14
2.3.4. Identificación del Modelo . . . . .	14
2.4. Relación metodológica entre el análisis factorial confirmatorio y los modelos de ecuaciones estructurales . . . . .	15
2.4.1. Representación del Modelo . . . . .	15
2.4.2. Estimación de los parámetros del Modelo . . . . .	17
2.4.3. Evaluación del ajuste del Modelo . . . . .	19
2.4.4. Modificación del Modelo . . . . .	22
2.4.5. Interpretación de los resultados . . . . .	24
<b>3. Caso de Estudio</b>	<b>25</b>
3.1. Descripción del Caso de Estudio . . . . .	25
3.2. Descripción del conjunto de datos . . . . .	26
3.3. Software Utilizado . . . . .	27
3.4. Preparación de los datos . . . . .	27
3.5. Validación de la escala de medida . . . . .	28
3.6. Análisis factorial confirmatorio . . . . .	28
3.6.1. Exploración de los datos . . . . .	28
3.6.2. Evaluación de la idoneidad de los datos . . . . .	29
3.6.3. Definición del modelo . . . . .	30

3.6.4.	Representación del Modelo . . . . .	34
3.6.5.	Especificación del Modelo . . . . .	36
3.6.6.	Estimación de los parámetros del Modelo . . . . .	36
3.6.7.	Evaluación del ajuste del Modelo . . . . .	38
3.6.8.	Modificación del Modelo . . . . .	39
3.6.9.	Interpretación de los resultados . . . . .	43
3.7.	Modelado de ecuaciones estructurales . . . . .	44

**4. Conclusiones** **49**

# Índice de figuras

2.1.	Diagrama de Relaciones: Ejemplo de un modelo recursivo de AFC . . . . .	16
2.2.	Diagrama de Relaciones: Ejemplo de un modelo no recursivo de AFC . . .	16
2.3.	Diagrama de Relaciones: Ejemplo de un modelo de ecuaciones estructurales.	17
3.1.	Gráfico de sedimentación de la <i>Inteligencia Emocional</i> . . . . .	30
3.2.	Pesos factoriales de la <i>Inteligencia Emocional</i> . . . . .	31
3.3.	Gráfico de sedimentación de las <i>Habilidades Sociales de Prevención del Acoso Escolar</i> . . . . .	31
3.4.	Pesos factoriales de las <i>Habilidades Sociales de Prevención del Acoso Escolar</i> . . . . .	32
3.5.	Gráfico de sedimentación de la <i>Capacidad de Disfrute</i> . . . . .	33
3.6.	Pesos factoriales de la variable <i>Capacidad de Disfrute</i> . . . . .	33
3.7.	Diagrama de Relaciones del modelo propuesto. . . . .	35
3.8.	Diagrama de Relaciones del modelo propuesto con la estimación de los parámetros. . . . .	37
3.9.	Diagrama de Relaciones del modelo modificado con la estimación de los parámetros. . . . .	42
3.10.	Diagrama de Relaciones del modelo de ecuaciones estructurales con la estimación de los parámetros. . . . .	45



# Índice de tablas

3.1.	Índices de ajuste del modelo propuesto . . . . .	38
3.2.	Índices de ajuste del modelo modificado . . . . .	43
3.3.	Índices de validación cruzada esperados . . . . .	43
3.4.	Índices de ajuste del modelo de ecuaciones estructurales . . . . .	46
3.5.	Parámetros estimados de interés del modelo de ecuaciones estructurales .	46



# Resumen

La psicología y la educación precisan de diferentes técnicas estadísticas para su estudio y desarrollo. En el presente Trabajo de Fin de Grado se exponen algunas de ellas como el análisis factorial confirmatorio (AFC), una herramienta estadística utilizada para evaluar la fiabilidad y validez de diferentes escalas de medida, y la consistencia entre los datos y el modelo propuesto, y los modelos de ecuaciones estructurales (SEM, por sus siglas en inglés, *Structural Equation modeling*), una herramienta de análisis multivariante utilizada para estudiar y modelar relaciones entre variables observadas y latentes. La utilización de variables latentes en modelos estadísticos aplicados a estas áreas es fundamental para estudiar factores abstractos y complejos que no pueden medirse directamente. Para ilustrar estos modelos se realiza un caso de estudio en el que se pretende analizar la relación entre la inteligencia emocional, la capacidad de disfrute y las habilidades sociales de prevención del acoso escolar, y si estos influyen o no en el rendimiento académico de los alumnos medido a través de las calificaciones de matemáticas. Para ello se dispone de una muestra de 437 alumnos entre 4º de Educación Primaria y 2º de Educación Secundaria.



# Abstract

Psychology and education require different statistical techniques for their study and development. In this Final Degree Project are presented some of them, such as confirmatory factor analysis (CFA), a statistical tool used to evaluate the reliability and validity of different measurement scales, and the consistency between the data and the proposed model, and structural equation modeling (SEM), a multivariate analysis tool used to study and model relationships between observed and latent variables. The use of latent variables in statistical models applied to these areas is fundamental to study abstract and complex factors that cannot be measured directly. To illustrate these models, a case study is carried out in which the target is to analyze the relationship between emotional intelligence, the capacity for enjoyment and social skills for bullying prevention, and whether or not these influence the academic performance of students measured through mathematics grades. For this purpose a sample of 437 students between 4th grade of Primary Education and 2nd year of Secondary Education is used.



# Capítulo 1

## Introducción

La estadística es una herramienta fundamental en el campo de las ciencias sociales en áreas como la psicología y la educación. Esto es debido a la necesidad de la exploración y el análisis de factores que conciernen al comportamiento humano y a los procesos educativos. Estas áreas requieren el uso de herramientas estadísticas que incluyan variables latentes. Se entiende por variable latente aquella que no puede medirse directamente con instrumentos de medición. Estas variables se utilizan para medir conceptos abstractos y subjetivos que no se pueden observar directamente como la inteligencia que requiere de diferentes test y pruebas para su medición [45].

En el presente Trabajo de Fin de Grado se exponen diferentes técnicas estadísticas que incluyen el uso de variables latentes. El análisis factorial confirmatorio (AFC), una herramienta desarrollada por primera vez en la década de 1960 por el estadístico Karl Jöreskog [18] basada en un modelo que especifica las relaciones entre las variables observadas, aquellas que sí pueden ser medidas u observadas directamente, y las variables latentes. El AFC permite validar <sup>1</sup> y evaluar la fiabilidad <sup>2</sup> de diferentes instrumentos de medida como los test, escalas o cuestionarios, así como evaluar las propiedades psicométricas de diferentes medidas, como las pruebas de aptitud, rendimiento o personalidad. El AFC también se utiliza para evaluar el ajuste entre los datos y el modelo propuesto.

El modelado de ecuaciones estructurales (SEM) es otra de las herramientas cuya construcción incluye el uso de variables latentes. Mueller [28] recogió los principios básicos del modelado de ecuaciones estructurales (SEM), que se puede definir como un método estadístico de análisis multivariante más amplio que el AFC, por lo que se puede decir que el AFC es un antecesor de este tipo de modelos. Estos modelos permiten evaluar las relaciones entre múltiples variables, tanto variables observadas como latentes, permiten analizar patrones, examinar como unas variables influyen en otras y como las relaciones entre las variables van cambiando a lo largo del tiempo, también permiten validar modelos y teorías creados a partir de estas variables.

En general, tanto el análisis factorial confirmatorio como los modelos de ecuaciones estructurales son técnicas estadísticas muy valiosas en la investigación empírica sobre todo en el ámbito de las ciencias sociales donde la medición de ciertos factores requiere de modelos que incluyan variables latentes.

---

<sup>1</sup>Propiedad que indica que el instrumento mide aquello para lo que ha sido diseñado.

<sup>2</sup>Se refiere a la coherencia con la que el instrumento mide aquello que quiere medir.

Para ilustrar el uso de este tipo de modelos se procede a realizar un estudio en el que se analizará la relación entre diferentes factores esenciales para el desarrollo de niños y adolescentes. Estos factores están relacionados con el ámbito de la psicología y la educación y son, la inteligencia emocional, las habilidades sociales de prevención del acoso escolar y la capacidad de disfrute. Se pretende estudiar la influencia que estos factores puedan o tener sobre el rendimiento académico medido a través de la calificación en matemáticas de los alumnos.

## 1.1. Estructura del documento

El presente Trabajo de Fin de Grado se estructura de la siguiente forma:

- **Capítulo 2 - Marco Teórico.** Se presenta el análisis de fiabilidad de las escalas de medida. Se define el análisis factorial confirmatorio y los modelos de ecuaciones estructurales. Se exponen los pasos para la construcción de estos modelos, desde su especificación, identificación, estimación de sus parámetros hasta la evaluación del ajuste del modelo. Se presentan también las diferentes formas de modificación de estos modelos.
- **Capítulo 3 - Caso de Estudio.** Se muestra la aplicación de los modelos expuestos en el Capítulo 2, mediante un caso de estudio, en el que se analiza la relación entre diferentes factores relacionados con la psicología y la educación.
- **Conclusiones.** Se realiza una reflexión sobre el trabajo realizado y se presentan diferentes formas para su mejora y continuación.
- **Anexo A - Covarianza: Definición y propiedades.** Se incluyen las propiedades de la covarianza para el mejor entendimiento del marco teórico expuesto en el Capítulo 2.
- **Anexo B - Cuestionario de Desarrollo Escolar.** Se incorpora como un anexo del documento el cuestionario empleado para la toma de los datos que se utilizan en el caso de estudio.
- **Anexo C - Código empleado.** Se muestra el código empleado para la realización del caso de estudio como un anexo al documento.

## Capítulo 2

# Marco Teórico

### 2.1. Análisis de fiabilidad de las escalas de medida

En el campo de las ciencias sociales es muy común la necesidad de medir variables que no se pueden observar de manera directa. Para poder medir estas variables que denominamos variables latentes se hace uso de diferentes instrumentos de medida como son los test, escalas o cuestionarios. Estos están formados por una serie de preguntas o elementos que servirán para describir esa variable. Debido a la forma en la que se miden estas variables es necesario analizar la consistencia interna entre las preguntas o elementos que forman esos instrumentos de medida. Es decir, como de relacionados están esos elementos o preguntas. Este procedimiento es necesario para demostrar la fiabilidad y validez del instrumento utilizado. Una forma de medir la fiabilidad de estos instrumentos sería tomar dos mediciones independientes a los mismo sujetos y compararlas, pero en el campo de la psicología y la educación es poco común tener la oportunidad de medir de nuevo a los mismos sujetos. Por ello, se utiliza el coeficiente Alpha ( $\alpha$ ), descrito por Cronbach [8]. El Alpha de Cronbach se define como una medida de fiabilidad y consistencia interna, que permite determinar la fiabilidad de un instrumento de medida a través de un conjunto de elementos que se espera que midan la misma variable latente.

Para un instrumento de medida,  $t$ , que contiene  $n$  elementos, denotamos como  $V_i$  a la varianza de cada uno de los elementos individuales, siendo  $i=1, \dots, n$ .  $V_t$  será la varianza total del instrumento de medida. Se denotará como  $C_{ij}$  a la covarianza entre dos elementos  $i$  y  $j$ , siendo  $1 \leq i, j \leq n$ , y por lo tanto  $C_t$  será la covarianza total de todos los pares de elementos del instrumento de medida. Así una forma de expresar el coeficiente  $\alpha$  es de la siguiente forma,

$$\alpha = \frac{n}{n-1} \left( \frac{1 - \sum_i V_i}{V_t} \right) = \frac{n}{n-1} \left( \frac{V_t - \sum_i V_i}{V_t} \right) \quad (2.1)$$

La varianza total de instrumento de medida,  $V_t$  se puede expresar como  $V_t = \sum_t V_i + 2C_t$ , por lo que podemos expresar la ecuación (2.1) como,

$$\alpha = \frac{n}{n-1} \frac{2C_t}{V_t} \quad (2.2)$$

Otra forma de expresar el coeficiente  $\alpha$  es mediante la covarianza entre los pares de elementos  $i$  y  $j$ ,

$$\bar{C}_{ij} = \frac{C_t}{n(n-1)/2}, \quad (2.3)$$

de tal forma que la ecuación (2.2) se puede expresar como,

$$\alpha = \frac{n^2 \bar{C}_{ij}}{V_t}. \quad (2.4)$$

### **Limitaciones del coeficiente $\alpha$**

Un número elevado de elementos aumentan el valor de la varianza total del instrumento de medida, lo que llevaría a obtener un valor sobreestimado de la fiabilidad del instrumento de medida. Por ello, el coeficiente  $\alpha$  está indicado para pruebas o escalas con 20 elementos o menos. Por otro lado, no se deben considerar escalas de un solo elemento ya que no es muy probable que un solo elemento mida una variable latente, lo que haría que la escala fuera poco fiable. También es importante considerar siempre el tamaño de la muestra, ya que al no considerarlo se puede estar sobrestimando el valor de  $\alpha$ , porque al aumentar el número de individuos aumenta la varianza esperada.

### **Interpretación del coeficiente $\alpha$**

El coeficiente  $\alpha$  toma valores entre 0 y 1. Se puede considerar que una prueba es fiable a partir de un  $\alpha \geq 0,7$  siendo preferible valores entre 0.8 y 0.9. Si  $\alpha = 1$ , todos los elementos son iguales, es decir, sus varianzas individuales son iguales y la correlación es perfecta, lo que podría indicar que hay redundancia o duplicidad entre los elementos. Si  $\alpha$  es menor de 0.7 se considera que la prueba no es fiable, y si  $\alpha = 0$  indica que no hay correlación entre los elementos, es decir, que estos son independientes.

## **2.2. Análisis Factorial Confirmatorio**

### **2.2.1. Contexto Teórico**

Se define el análisis factorial como una técnica estadística de análisis multivariante que reduce la dimensionalidad de los datos identificando patrones en un conjunto de variables. Dependiendo del análisis que se quiera llevar a cabo se distinguen dos tipos de análisis factorial, el análisis factorial exploratorio (AFE) y el análisis factorial confirmatorio (AFC).

El análisis factorial exploratorio fue desarrollado por Spearman [41] a principios del siglo XX, se utiliza para explorar la estructura subyacente de un conjunto de datos. Busca encontrar el menor número de variables latentes que expliquen la mayor variabilidad de los datos y genera hipótesis entre las variables observadas y las variables latentes. Gracias a Spearman, varios científicos como Jöreskog [17, 18] o Lawley [23] empezaron a desarrollar una teoría estadística que pudiera representar las relaciones entre variables observadas en términos de ciertas variables latentes. Así en la década de 1950 se definió el análisis factorial confirmatorio (AFC) como una técnica estadística utilizada para confirmar una hipótesis a priori mediante la representación de las relaciones entre variables observadas

en términos de ciertas variables latentes. Esta técnica se basaba en la premisa de que las variables observadas son indicadores imperfectos de ciertas variables latentes. El enfoque del AFC requiere que se teorice una estructura subyacente y se evalúe si los datos observados se ajustan al modelo especificado a priori. El AFC permite agrupar un conjunto de variables observadas indicadoras si estas miden la misma variable latente, es decir, permite evaluar a partir de este conjunto de variables si un conjunto de datos concreto "confirma" lo que se sabe que es su estructura subyacente.

La interpretación del análisis factorial puede ser compleja si el modelo no está completamente definido. Para facilitar la interpretación de los resultados del análisis se utilizan técnicas como la rotación de factores que permite obtener patrones más sencillos y claros y ayuda a reducir la redundancia entre factores.

Para poder realizar un análisis factorial confirmatorio se debe crear un modelo teórico que represente la estructura factorial de las variables observadas, que debe estar basado en un conjunto de hipótesis que deseamos contrastar sobre los factores subyacentes, es decir, sobre lo que llamamos variables latentes.

### 2.2.2. Especificación del Modelo

La especificación del modelo consiste en establecer el número de variables latentes, definir al menos una estructura inicial entre las variables observadas y las variables latentes, es decir, las relaciones hipotéticas entre las variables observadas,  $X_i$ ,  $i=1\dots n$  siendo  $n$  el número de variables observadas del modelo y las variables latentes,  $\xi_j$ ,  $j=1\dots m$  siendo  $m$  el número de variables latentes del modelo. Se deben especificar los errores de medida de las variables observadas,  $\delta_i$ , las cargas factoriales que relacionan linealmente las variables observadas con las variables latentes,  $\lambda_{ij}$ , y la asociación hipotética entre las variables latentes,  $\phi_{jj}$ .

La representación de un modelo de AFC se puede realizar mediante un conjunto de ecuaciones. Mediante estas podemos expresar la relación entre las variables latentes y las observadas de la siguiente manera,

$$X_i = \sum_{j=1}^m \lambda_{ij} \xi_j + \delta_i \quad (2.5)$$

Cada una de las ecuaciones representará una variable observada en función de las variables latentes y un término de error, por lo que habrá una ecuación por cada variable observada del modelo. Otra forma de representar un modelo de AFC es mediante su forma matricial,

$$X = \Lambda \xi + \delta, \quad (2.6)$$

donde  $X$  es un vector de variables observadas,  $\Lambda$ , es una matriz de cargas factoriales de tamaño  $(n \times m)$ ,  $\xi$  es un vector de variables latentes de tamaño  $(n \times 1)$  y  $\delta$  es un vector de los errores de medida asociados a las variables observadas de tamaño  $(n \times 1)$ . Por lo que la matriz anterior es equivalente a,

$$\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m-1} & \lambda_{1m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{n1} & \lambda_{n2} & \cdots & \lambda_{nm-1} & \lambda_{nm} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_m \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_n \end{bmatrix} \quad (2.7)$$

Para completar la especificación de un modelo de AFC se necesita definir otras dos matrices, una matriz que denominaremos  $\Phi$ , para especificar las varianzas y covarianzas de las variables latentes,

$$\Phi = \begin{bmatrix} \sigma_{\xi_1}^2 & & & \\ \sigma_{\xi_2\xi_1} & \sigma_{\xi_2}^2 & & \\ \dots & \dots & \ddots & \\ \sigma_{\xi_m\xi_1} & \sigma_{\xi_m\xi_2} & \dots & \sigma_{\xi_m}^2 \end{bmatrix} \quad (2.8)$$

y por último se define una matriz de varianzas/covarianzas de los errores de medición de las variables observadas,  $\Theta_\delta$ , cuya diagonal contiene las varianzas de los términos de error,

$$\Theta_\delta = \begin{bmatrix} \sigma_{\delta_1}^2 & & \\ 0 & \ddots & \\ 0 & 0 & \sigma_{\delta_n}^2 \end{bmatrix} \quad (2.9)$$

En conclusión, para la especificación de un modelo de AFC completo debemos disponer de tres matrices básicas,  $\Lambda$ ,  $\Phi$ ,  $\Theta_\delta$ .

Al especificar un modelo de AFC debemos comprobar que se cumplen los siguientes supuestos:

- Las medias de las variables observadas y latentes son 0.
- Las relaciones entre las variables observadas y latentes son lineales.
- Los errores de medida de las variables observadas son independientes de las variables latentes y entre sí.

### 2.2.3. Matrices de varianzas/covarianzas implícita del Modelo y no restringida

En un análisis factorial confirmatorio se distingue entre la matriz de varianzas/covarianzas no restringida y la matriz de varianzas/covarianzas implícita del modelo. Cuando no se hace referencia a un modelo de AFC concreto o a un modelo de ecuaciones estructurales (ver sección 2.3), se dice que la matriz de varianzas/covarianzas de las variables observadas a la que denominamos  $\Sigma$  (Sigma), es no restringida. Esto se debe a que no se hipotetiza ninguna estructura a priori. Esta matriz describe la relación entre las variables observadas y latentes sin imponer ninguna restricción en los parámetros. En cambio, si cogemos la ecuación que representa la forma matricial del modelo de AFC, que incluye las restricciones impuestas en el modelo, podemos ver que  $\Sigma = E(XX')$ , esto se debe a que las variables observadas se miden como las desviaciones de sus medias. Bajo los supuestos de que los errores en  $\delta$  tienen media cero y que esos términos de error no se correlacionan con las variables latentes, podemos reescribir  $\Sigma$  de la siguiente manera, utilizando las propiedades de la covarianza del Anexo A,

$$\begin{aligned} \Sigma &= E(XX') \\ &= E[(\Lambda\xi + \delta)(\Lambda\xi + \delta)'] \\ &= E(\Lambda\xi\xi'\Lambda') + E(\Lambda\xi\delta') + E(\delta\xi'\Lambda') + E(\delta\delta') \\ &= \Lambda E(\xi\xi')\Lambda' + \Lambda E(\xi\delta') + E(\delta\xi')\Lambda' + E(\delta\delta') \\ &= \Lambda E(\xi\xi')\Lambda' + E(\delta\delta') \end{aligned} \quad (2.10)$$

donde  $E(\xi\xi') = \Phi$  y  $E(\delta\delta') = \Theta_\delta$ . Así la matriz de varianzas/covarianzas de las variables observadas puede escribirse como,

$$\Sigma = \Lambda\Phi\Lambda' + \Theta_\delta. \quad (2.11)$$

Cuando se hipotetiza que un modelo particular subyace a las variables observadas, la matriz de varianzas/covarianzas puede escribirse en función de las tres matrices básicas del modelo. Denominamos  $\Sigma(\theta)$  a esta matriz de varianzas/covarianzas implícita del modelo, donde  $\theta$ , es el vector que contiene los parámetros a estimar del modelo,  $p$ .

#### 2.2.4. Determinación y Extracción de factores

Sabemos que los factores o variables latentes no son observables directamente, dependen de las variables observadas para su medición. La medición de un factor concreto puede requerir un número indeterminado de variables latentes para su correcta descripción. Para determinar el número de factores óptimo que se deben incluir en el modelo hipotetizado a priori se pueden utilizar dos criterios: la regla de Kaiser-Guttman y el gráfico de sedimentación.

Para el mayor entendimiento de estos métodos hay que saber que el análisis factorial confirmatorio y el análisis de componentes principales están muy relacionados. El ACP se define como una técnica estadística que se utiliza para reducir la dimensión de un conjunto de variables, identificando las combinaciones lineales de estas variables que retienen la mayor parte de la información. Esto permite simplificar y resumir la estructura de los datos, identificando los componentes principales que explican la mayor variabilidad en el conjunto de variables [16]. En un ACP las variables latentes son combinaciones lineales de las variables observadas y en un AFC las variables observadas se forman como funciones lineales de las variables latentes. En ambos métodos se está reduciendo la dimensión de los datos.

Para relacionar un modelo de AFC con el ACP primero debemos considerar la matriz  $X$  que contiene una muestra de  $n$  individuos con  $p$  variables observadas, por lo que  $X$  tendrá una dimensión de  $(n \times p)$ . Se asume que esta matriz esta centrada por columnas. La matriz  $X$  por su traspuesta,  $XX^T$ , es una matriz simétrica y semidefinida positiva y se puede expresar como,

$$XX^T = VDVT \quad (2.12)$$

$D$  es una matriz diagonal de dimensión  $(n \times p)$ ,  $D = \text{diag}(d_1, \dots, d_p)$ ,  $d_1 \geq \dots \geq d_p \geq 0$ , siendo  $d_i$  un valor propio de  $XX^T \forall i$ .  $V = [v_1, \dots, v_p]$  son los vectores propios.  $V$  es ortogonal por lo que  $VV^T = V^TV = I$ .  $XX^T$  tiene  $(p - r)$  valores propios iguales a 0 por lo que la ecuación (2.12) puede escribirse como,

$$XX^T = (V_r D_r^{1/2})(D_r^{1/2} V_r^T) \quad (2.13)$$

es decir,

$$X = V_r D_r^{1/2} \quad (2.14)$$

donde  $D_r^{1/2} = \text{diag}(d_1^{1/2}, \dots, d_r^{1/2})$ , esta matriz diagonal tiene una dimensión de  $(n \times r)$ , y  $V_r = [v_1, \dots, v_r]$ . Al seleccionar los valores propios distintos de cero hay que incluir un

término de error por lo que la ecuación quedaría de la siguiente forma.

$$X = D_r^{1/2}V_r + \delta \quad (2.15)$$

Esta ecuación se puede relacionar directamente con la ecuación (2.6),  $D_r^{1/2}$  se puede relacionar con la matriz de cargas factoriales  $\Lambda$ , y  $V_r$  se puede relacionar con el vector  $\xi$  de variables latentes.

Uno de los métodos utilizado para determinar el número de factores es la regla de Kaiser-Guttman que consiste en establecer aquellos factores cuyos valores propios sean mayores a uno. Los valores propios representan la cantidad de varianza en las variables observadas que puede ser explicada por cada factor. El criterio de Kaiser-Guttman tradicional utiliza el análisis de componentes principales (ACP) para la búsqueda de los valores propios, pero dispone también de una variante que utiliza el análisis factorial (AF) para ello. La variante que utiliza AF en ocasiones es menos precisa que el método tradicional [42]. Esta regla presenta ciertas limitaciones ya que puede presentar resultados imprecisos y no establecer el número de factores suficientes que se deben extraer.

El otro método utilizado para la determinación de los factores es el gráfico de sedimentación, que representa la magnitud de los valores propios. En el eje de abscisas se representa el número de factores y en el eje de ordenadas se representan los valores propios. El número de factores se establece como el número de factores que queden por encima de una línea horizontal que indica a partir de qué punto los valores propios explican una cantidad de varianza que no es significativa [33]. Los valores propios se pueden obtener mediante ACP o AF.

Una vez que se ha determinado el número de factores hay que encontrar las cargas factoriales, es decir, establecer las variables observadas que van a ser indicadoras de cada uno de esos factores. Existen diferentes métodos para la extracción de factores, algunos de ellos son los mismos que se pueden utilizar para la estimación de los parámetros del modelo y se rigen por los mismos criterios que veremos en la sección 2.4.2. Estos métodos son: máxima verosimilitud (ML), mínimos cuadrados generalizados (GLS) y mínimos cuadrados no ponderados (ULS). Otro de los métodos de extracción utilizado en el AFC es la factorización de ejes principales, este método extrae aquellos factores que explican la mayor parte de la varianza y no es necesaria la distribución normal de los datos [27].

### 2.2.5. Rotación de factores

El análisis factorial confirmatorio es una herramienta muy útil, pero no está exenta de ciertas limitaciones. Debido a la utilización de factores latentes, se considera que este método es muy subjetivo. Esto es debido a que la interpretación de los factores latentes puede no ser clara ya que una variable observada puede ser indicadora de más de un factor latente, cuando lo ideal es que una variable observada sea indicadora de un solo factor latente. Para poder resolver esta limitación se utiliza una técnica denominada rotación factorial, esta técnica es propia del análisis de componentes principales y se utiliza en el AFC ya que ambos métodos presentan ciertas similitudes. Se utiliza después de la extracción de los factores para facilitar su interpretación.

La ecuación (2.6) presenta un modelo de AFC en su forma matricial. Este modelo presenta ciertas limitaciones una de ellas es que no es un modelo identificable ya que, si definimos una matriz T ortogonal, es decir,  $TT' = T'T = I$ , de tamaño (m x m), si  $\tilde{\xi} = T\xi$  tendremos que,

$$X = \Lambda T' T \xi + \delta = \tilde{\Lambda} \tilde{\xi} + \delta \quad (2.16)$$

Esto quiere decir que un modelo es válido con una matriz de cargas  $\Lambda T$  si T es ortogonal y con una matriz de cargas  $\Lambda$ , es esta ambigüedad la que se corrige con los métodos de rotación ortogonal.

Para crear modelos más sencillos y fáciles de interpretar se utiliza un tipo de rotación factorial llamada rotación ortogonal. Se utiliza cuando buscamos que los factores sean ortogonales entre sí, es decir, que no estén correlacionados. La rotación ortogonal más habitual sobre todo en psicología es la rotación varimax (Kaiser) [20], que busca maximizar la varianza de las cargas al cuadrado de cada factor. La rotación varimax maximiza la expresión [7],

$$\sum_{j=1}^m \left( \frac{1}{n} \sum_{i=1}^n \tilde{\lambda}_{ij}^4 - \left( \frac{1}{n} \sum_{i=1}^n \tilde{\lambda}_{ij}^2 \right)^2 \right) \quad (2.17)$$

Recordemos que m es el número de variables latentes del modelo y n es el número de variables observadas. Se define  $\tilde{\lambda}_{ij}$  como,

$$\tilde{\lambda}_{ij} = \frac{\lambda_{ij}}{h_i} \quad (2.18)$$

siendo  $h_i$  la "comunalidad"<sup>1</sup> de X, su tamaño es una medida de lo bien que el modelo es capaz de representar el comportamiento de la variable X en función de las variables latentes. La "comunalidad" de X, es decir,  $h_i$  puede expresarse como,

$$h_i^2 = \sum_{j=1}^m \lambda_{ij}^2 \quad (2.19)$$

Recordemos que  $\lambda_{ij}$  representa las cargas factoriales.

La rotación varimax no es la única utilizada, hay otros métodos de rotación ortogonal como la rotación quartimax (Neuhaus y Wrigley), que busca maximizar la suma de las cargas factoriales elevadas a potencia 4 [30]. La rotación quartimax maximiza la expresión [7],

$$\sum_{i=1}^n \sum_{j=1}^m \lambda_{ij}^4 \quad (2.20)$$

Por último, se dispone de otro tipo de rotación ortogonal, la rotación equamax. Saunders [38] propuso la rotación equamax como una combinación entre la rotación varimax y la rotación quartimax, por lo que este método busca maximizar tanto la varianza de las cargas al cuadrado de cada factor, como la suma de las cargas elevadas a potencia 4. Este método es menos utilizado debido a que es menos fiable que los otros. Esta rotación maximiza la expresión [7],

$$\sum_{j=1}^m \left( \frac{m}{2n} \sum_{i=1}^n \tilde{\lambda}_{ij}^4 - \left( \frac{m}{2n} \sum_{i=1}^n \tilde{\lambda}_{ij}^2 \right)^2 \right) \quad (2.21)$$

<sup>1</sup>La comunalidad es una medida que indica la proporción de varianza de una variable observada que es explicada por los factores o variables latentes.

Hay otro tipo de rotación factorial que se utiliza para crear modelos más sencillos, esta es la rotación oblicua. Se suele utilizar cuando los resultados obtenidos al realizar una rotación ortogonal no son los que cabría esperar y busca que los factores estén correlacionados. El criterio de rotación oblicua más utilizado es el criterio Oblimin directo propuesto por Hotelling [14]. Este permite establecer relaciones jerárquicas entre los factores a partir del grado de inclinación,  $\delta$ . Un valor de 0 indica soluciones más oblicuas y valores negativos indican soluciones menos oblicuas. Este tipo de rotación puede hacer que la interpretación resulte menos sencilla debido a que las relaciones entre los factores pueden ser complejas.

La elección de un tipo de rotación u otro no solo depende de la correlación o incorrelación de los factores, también depende del tipo de estudio que se esté realizando, de las hipótesis que se quieran probar y del tipo de datos que se estén utilizando.

### 2.2.6. Identificación del Modelo

Para poder estimar los parámetros del modelo que hemos definido primero, se deben identificar los  $p$  parámetros del modelo y asegurarse de que hay suficiente información sobre las variables observadas para poder realizar la estimación. Para ello se deben verificar tres condiciones para identificar el modelo.

1. La primera condición necesaria pero no suficiente es que el número de parámetros que deben estimarse,  $p$ , no puede ser mayor que el número de varianzas/covarianzas no redundantes de las variables observadas. Este número lo denominamos  $c$ , donde  $c = (n)(n+1)/2$ . El cálculo de  $p$  se realiza contando los términos que son distintos de cero en las tres matrices básicas y posteriormente se comprueba que  $p$  sea menor que  $c$  para saber si disponemos de información suficiente para estimar los parámetros.
2. La segunda condición es que cada una de las variables latentes debe tener una única unidad de medida, esto se puede conseguir de dos formas, la primera es definir la escala de la variable latente para que sea igual a la escala de una de sus variables indicadoras observadas o fijar la varianza de la variable latente a 1.
3. Por último, si sólo hay una variable indicadora observada para una variable latente, hay que suponer que la fiabilidad de la variable observada es perfecta, es decir, que mide perfectamente la variable latente. Esto supone fijar su error de medida en la matriz  $\Theta_\delta$  a 0. Si no se pudiera justificar esto, la solución podría ser incluir datos de una variable observada adicional e incluirla en el modelo.

En el caso de que la primera condición no se cumpla, verificar el resto de las condiciones, podría disminuir el número de parámetros  $p$ , lo que podría hacer que se cumpliera que  $c$  fuera mayor que  $p$ . Una vez se han cumplido las condiciones se puede decir que el modelo está identificado y ya es posible estimar todos los parámetros.

### 2.2.7. Evaluación de la idoneidad de la muestra

Para poder realizar un análisis factorial confirmatorio debemos disponer de unos datos que sean aceptables para la realización de este análisis. Se dispone de una serie de test que permiten comprobar la idoneidad de la muestra que deseamos utilizar. Algunos de estos test son el test de esfericidad de Bartlett y el test de Kaiser-Meyer-Olkin (KMO).

### Test de esfericidad de Bartlett

El test de esfericidad de Bartlett se utiliza para comprobar la hipótesis nula de que las variables de la muestra no están correlacionadas entre sí. Una de las condiciones que deben cumplirse es que las variables estén muy correlacionadas entre sí, en caso de que no lo estén los datos podrían no ser adecuados para este tipo de análisis. Se calcula de la siguiente forma,

$$\chi^2 = -(n - 1 - \frac{2p + 5}{6}) \ln|R| \quad (2.22)$$

donde  $n$  es el tamaño de la muestra,  $p$  es el número de variables observadas y  $R$  es la matriz de correlaciones de las variables observadas. El estadístico se distribuye asintóticamente bajo una distribución  $\chi^2$  con  $(p(p - 1))/2$  grados de libertad. Valores altos de este estadístico rechazan la hipótesis nula. Se concluiría que las variables están correlacionadas entre sí. Es un test muy sensible a la normalidad multivariante, por lo que primero hay que comprobar los supuestos de normalidad de la muestra, y en el caso de que no se cumplan el uso de este test puede no ser adecuado [39].

### Test de Kaiser-Meyer Olkin

El test KMO mide si la muestra es adecuada o no para realizar un análisis factorial. Este test se puede utilizar, aunque los datos no sigan los supuestos de normalidad multivariante. Se calcula de la siguiente forma,

$$KMO_j = \frac{\sum_{i \neq j} R_{ij}^2}{\sum_{i \neq j} R_{ij}^2 + \sum_{i \neq j} U_{ij}^2} \quad (2.23)$$

donde  $R_{i \neq j}$  es la matriz de correlaciones de las variables observadas y  $U_{i \neq j}$  es la matriz de correlaciones parciales de las variables observadas. Valores mayores o iguales que 0.75 indican que la muestra es buena para realizar el análisis, valores mayores o iguales que 0.5 indican que la muestra es aceptable y valores menores de 0.5 indican que la muestra no es aceptable para poder realizar el análisis [39].

## 2.3. Modelos de Ecuaciones Estructuras Generales

El modelado de ecuaciones estructurales (SEM) es una técnica estadística de análisis multivariante que permiten modelar y examinar la relación entre múltiples variables, tanto variables observadas como latentes. Estos modelos incluyen el análisis factorial confirmatorio. Los modelos de ecuaciones estructurales constan de dos partes, un modelo de medición que describe la relación entre las variables observadas y latentes, y un modelo estructural que describe la relación entre las variables latentes.

### Tipos de variables

En un modelo de ecuaciones estructurales las variables observadas y latentes se distinguen en dos tipos, exógenas o endógenas:

- Las variables exógenas son aquellas variables que están determinadas por factores externos al modelo.
- Las variables endógenas son aquellas cuyo valor se determina por las relaciones existentes dentro del modelo en el que se encuentre.

### 2.3.1. Especificación del Modelo

Al igual que ocurre con un modelo de AFC, primero partimos de unas hipótesis que deseamos contrastar, para ello establecemos cuales son las variables que serán relevantes en la especificación del modelo, tanto las variables observadas como las latentes y se establecerán las relaciones entre las variables, estos modelos incluyen tanto relaciones directas como indirectas entre variables. Se definirá entonces el modelo estructural y el modelo de medición.

Se define el modelo estructural que representa la relación entre las variables latentes, es decir, la relación entre las variables latentes endógenas  $\eta_r$ ,  $r=1\dots m$ , y las variables latentes exógenas  $\xi_s$ ,  $s=1\dots k$ , y viene determinada por

$$\eta = \eta B + \Gamma \xi + \zeta \quad (2.24)$$

siendo  $m$  y  $k$  el número de variables latentes endógenas y exógenas respectivamente,  $\eta$  es un vector ( $m \times 1$ ) de las variables latentes endógenas,  $\xi$  es un vector ( $k \times 1$ ) de las variables latentes exógenas,  $B$  es una matriz de coeficientes estructurales ( $m \times m$ ) de las variables latentes endógenas a otras variables endógenas,  $\Gamma$  es una matriz de coeficientes estructurales ( $m \times k$ ) de las variables latentes exógenas a endógenas,  $\zeta$  es un vector, ( $m \times 1$ ), de los términos de error asociados a las variables latentes endógenas.

Se define el modelo de medición que relaciona las variables observadas exógenas  $X_j$ ,  $j=1,\dots,n$  con las variables latentes exógenas  $\xi_s$ ,  $s=1,\dots,k$ , y viene determinado por

$$X = \Lambda_X \xi + \delta \quad (2.25)$$

donde  $n$  y  $k$  determinan respectivamente el número de variables exógenas observadas y latentes,  $X$  es un vector ( $n \times 1$ ) de las variables observadas medidas como la desviación de sus medias,  $\Lambda_X$  es una matriz de coeficientes estructurales ( $n \times k$ ) de variables latentes exógenas a sus variables indicadoras,  $\xi$  es un vector ( $k \times 1$ ) de variables latentes exógenas y  $\delta$  es un vector ( $n \times 1$ ) de los términos de error de las variables observadas.

El modelo de medición relativo a la relación de las variables observadas endógenas  $Y_i$ ,  $i=1,\dots,o$  con las variables latentes endógenas,  $\eta_r$ ,  $r=1,\dots,m$ , viene determinado por

$$Y = \Lambda_Y \eta + \epsilon \quad (2.26)$$

donde  $Y$  es un vector de variables observadas endógenas, ( $o \times 1$ ) medidas como la desviación de sus medias,  $\Lambda_Y$ , es una matriz de coeficientes estructurales ( $o \times m$ ) de las variables latentes endógenas a sus indicadores observados,  $\eta$  es un vector ( $m \times 1$ ) de variables latentes endógenas y  $\epsilon$ , es un vector ( $o \times 1$ ) de los términos de error de las variables observadas endógenas.

Para completar la especificación del modelo es necesario determinar la matriz de varianzas/covarianzas  $\Phi$ , de tamaño ( $k \times k$ ), de variables latentes exógenas, la matriz de varianzas/covarianzas  $\Psi$ , ( $m \times m$ ) de los términos de error asociados a las ecuaciones estructurales implícitas del modelo, la matriz de varianzas/covarianzas  $\Theta_\delta$ , ( $n \times n$ ), de los términos de error de las variables observadas exógenas y por último, la matriz de varianzas/covarianzas  $\Theta_\epsilon$ , ( $o \times o$ ), de los términos de error asociados con las variables observadas endógenas. En conclusión, la definición de estas matrices  $B$ ,  $\Gamma$ ,  $\Phi$ ,  $\Psi$ ,  $\Lambda_X$ ,  $\Theta_\delta$ ,  $\Lambda_Y$  y  $\Theta_\epsilon$ ,

determina un modelo general de ecuaciones estructurales.

La especificación de un modelo de ecuaciones estructurales debe cumplir los siguientes supuestos:

1. Las variables latentes endógenas y exógenas tienen media 0,  $[E(\xi) = E(\eta) = 0]$ .
2. Las relaciones estructurales entre las variables latentes endógenas y exógenas son lineales.
3. Los términos de error  $\zeta$  de la ecuación tienen media 0 y una varianza constante entre las observaciones, son independientes y las observaciones no están correlacionadas y por último, son incorrelados respecto a las variables latentes exógenas,  $[E(\xi\zeta') = E(\zeta\zeta') = 0]$ .
4. La matriz  $(I - B)$ , siendo  $I$  la matriz identidad, es invertible.
5. Las medias de las variables observadas exógenas y endógenas son 0,  $E(X) = E(Y) = 0$ .
6. Las relaciones entre las variables indicadora tanto exógenas como endógenas y sus variables latentes son lineales.
7. Los términos de error en las ecuaciones  $\delta$  y  $\epsilon$  tienen media 0,  $[E(\delta) = E(\epsilon) = 0]$  y una varianza constante entre las observaciones.
8. Los elementos de  $\epsilon$ , son incorrelados con las variables latentes tanto exógenas como endógenas y son incorrelados con entre sí,  $[E(\epsilon\delta') = E(\delta\epsilon') = 0]$ .

Hay que saber que no es necesario considerar los supuestos 1 y 5 si se realiza un análisis de estructuras medias, es decir, si queremos evaluar si los parámetros del modelo varían entre diferentes subgrupos de la muestra, y que el supuesto 8 puede relajarse si especificamos los elementos distintos de cero en una matriz  $\Theta_{\delta\epsilon}$  de covarianzas entre  $\delta$  y  $\epsilon$ .

### 2.3.2. Efectos de los Componentes Estructurales

Las variables del modelo pueden tener ciertos efectos en otras variables debido a las relaciones entre ellas, estos efectos presentan ciertas ventajas:

- Se utilizan para comprender la complejidad de las relaciones entre las variables del modelo.
- Evalúan la importancia de las diferentes variables latentes en el modelo.
- También pueden ser útiles para evaluar la robustez del modelo.

Se pueden definir tres tipos de efectos, efecto directo, indirecto y total:

1. El efecto directo (ED) se define como la influencia de una variable sobre otra, es decir, como la relación inmediata que hay entre dos variables. Este efecto se puede producir entre dos variables latentes o entre una variable latente y una variable observada. Por ejemplo, el efecto directo de una variable latente hacia otra, equivale al valor de su asociación hipotética. El efecto directo entre una variable latente hacia una observada equivaldría al valor de la carga factorial entre ellas.

2. El efecto indirecto (EI) se define como la influencia de una variable sobre otra a través de al menos otra variable que sirve de intermediario.
3. El efecto total (ET) entre dos variables se define como la suma de los componentes del efecto directo y del efecto indirecto total.

### 2.3.3. Matrices de varianzas/covarianzas implícita del Modelo y no restringida

En la sección 2.2.3, vimos la diferencia entre la matriz de varianzas/covarianzas implícita del modelo y la matriz no restringida para un modelo de AFC. En el caso de los modelos de ecuaciones estructurales, la matriz de varianzas/covarianzas no restringida, se define en función de las 8 matrices básicas, esta matriz,  $\Sigma$ , de variables observadas exógenas  $X_j$ , y de variables observadas endógenas  $Y_i$ , de tamaño  $[1 \times (o + n)]$ , que contiene  $(o \times n)$  variables observadas para un modelo de ecuaciones estructurales, sería,

$$\Sigma = E\left(\begin{bmatrix} Y \\ X \end{bmatrix} \begin{bmatrix} Y \\ X \end{bmatrix}'\right) \quad (2.27)$$

donde,

$$\begin{bmatrix} Y \\ X \end{bmatrix}' = [Y_1 \quad Y_2 \quad \dots \quad Y_o \quad X_1 \quad X_2 \quad \dots \quad X_n]. \quad (2.28)$$

Sustituyendo el lado derecho de las ecuaciones (2.25) y (2.26) por los vectores X e Y respectivamente se obtiene,

$$\begin{aligned} \Sigma &= E\left(\begin{bmatrix} Y \\ X \end{bmatrix} \begin{bmatrix} Y \\ X \end{bmatrix}'\right) = E\left(\begin{bmatrix} YY' & XX' \\ XY' & XX' \end{bmatrix}\right) \\ &= E\left(\begin{bmatrix} (\Lambda_Y \eta + \epsilon)(\Lambda_Y \eta + \epsilon)' & (\Lambda_Y \eta + \epsilon)(\Lambda_x \xi + \delta)' \\ (\Lambda_x \xi + \delta)(\Lambda_Y \eta + \epsilon)' & (\Lambda_x \xi + \delta)(\Lambda_x \xi + \delta)' \end{bmatrix}\right) \end{aligned} \quad (2.29)$$

A partir de la expresión de  $\Sigma$  de la ecuación (2.27) y utilizando las definiciones de las ocho matrices básicas puede demostrarse que la matriz de varianzas/covarianzas implícita en el modelo,  $\Sigma(\theta)$  para un modelo de ecuaciones estructurales involucrando a las variables latentes es,

$$\Sigma(\theta) = \begin{bmatrix} \Lambda_Y(I - B)^{-1}(\Gamma\Phi\Gamma' + \Psi)[(I - B)^{-1}]' \Lambda_Y' + \Theta_\epsilon & \Lambda_Y(I - B)^{-1}\Gamma\Phi\Lambda_X' \\ \Lambda_X\Phi\Gamma'[(I - B)^{-1}]' \Lambda_Y' & \Lambda_X\Phi\Lambda_X' + \Theta_\delta \end{bmatrix} \quad (2.30)$$

### 2.3.4. Identificación del Modelo

Una vez especificado un modelo el siguiente paso sería estimar los parámetros de ese modelo para ello primero hay que identificar esos parámetros y hay que seguir las normas que se exponen a continuación:

1. El número de parámetros p, en las 8 matrices básicas, no debe ser mayor que el número de varianzas y covarianzas disponibles entre las  $(n + o)$  variables observadas, este término se denomina c, siendo  $c = (n + o)(n + o + 1)/2$ , donde n es el número de variables observadas exógenas y o es el número de variables observadas endógenas. Esta condición es necesaria pero no suficiente para la identificación del modelo.
2. Todas las variables latentes endógenas y exógenas tienen asignada una unidad de medida.
3. Si una variable observada exógena o endógena es el único indicador de una variable latente se considera esa variable observada sin error.

## 2.4. Relación metodológica entre el análisis factorial confirmatorio y los modelos de ecuaciones estructurales

Esta sección se enfocará en explorar las metodologías comunes entre el análisis factorial confirmatorio y los modelos de ecuaciones estructurales. Los modelos de ecuaciones estructurales engloban al análisis factorial confirmatorio por lo que coinciden entre otras cosas en la forma en la que se representan este tipo de modelos. Comparten la forma en la que se estiman los parámetros del modelo y el uso de medidas de ajuste para evaluación del ajuste del modelo con los datos utilizados. Por último, tienen en común la forma en la que estos modelos pueden ser modificados para su mejora y búsqueda del mejor modelo.

### 2.4.1. Representación del Modelo

Una forma que tenemos de representar tanto los modelos de AFC como los modelos de ecuaciones estructurales es mediante los diagramas de relaciones. Estos se utilizan para mostrar gráficamente la estructura hipotética a priori de las variables de un modelo, es decir, para representar las relaciones hipotéticas entre las variables observadas y las variables latentes y las relaciones entre las variables latentes.

Para dos variables cualesquiera, X e Y, son posibles cuatro relaciones básicas:

1.  $X \rightarrow Y$ , X podría influir estructuralmente en Y, pero no a la inversa.
2.  $X \leftarrow Y$ , Y podría influir estructuralmente en X, pero no a la inversa.
3.  $X \rightleftharpoons Y$ , X podría influir estructuralmente en Y, e Y podría influir estructuralmente en X.
4.  $X \curvearrowright Y$ , No se hipotetiza ninguna relación estructural entre X e Y pero las variables podrían covariar.

Se entiende por influencia estructural a la relación directa o causal entre dos variables del modelo. Por ejemplo, cuando se expresa que X podría influir estructuralmente en Y, quiere decir que cambios en la variable X producen el mismo cambio en la variable Y. En cambio, si no hay una relación estructural podría darse una correlación entre las variables. Esta correlación puede ser positiva o negativa e indica la relación o asociación entre las variables. Esta asociación no implica que cambios en una variable causen un cambio en otra.

Las flechas representan una relación entre dos variables indicando la fuerza y la dirección entre ellas. Las flechas rectas unidireccionales representan relaciones estructurales de una variable a otra, mientras que las flechas con forma de curva bidireccionales representan la covarianza entre variables sin una relación estructural entre ellas, las flechas que apuntan a variables que no se encuentran en el modelo denotan la presencia de errores de medición en las variables observadas. La representación de las variables observadas se realiza mediante cuadrados o rectángulos y mediante círculos o elipses para las variables latentes. Según el diagrama de relaciones se pueden distinguir dos tipos de modelos, los modelos recursivos que incluyen las relaciones 1, 2 o 4 que son los que se exponen en este documento, y los no recursivos que pueden incluir relaciones estructurales bidireccionales.

A continuación, se muestra un ejemplo de un diagrama de relaciones para un modelo recursivo de AFC y un ejemplo de un diagrama de relaciones para un modelo no recursivo. En estos modelos se incluyen cuatro variables observadas con sus términos de error y dos variables latentes, se incluyen también las cargas factoriales entre las variables observadas y latentes y la asociación hipotética entre las dos variables latentes.

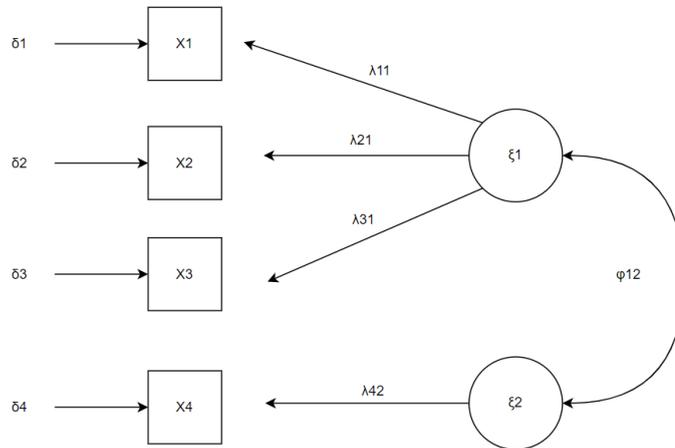


Figura 2.1: Diagrama de Relaciones: Ejemplo de un modelo recursivo de AFC

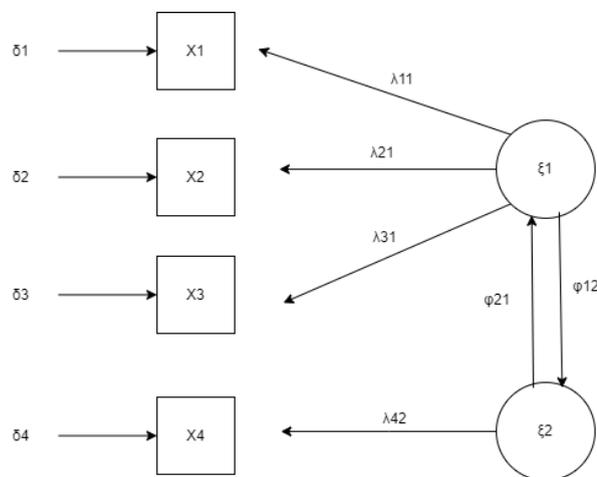


Figura 2.2: Diagrama de Relaciones: Ejemplo de un modelo no recursivo de AFC

Por último, se muestra un ejemplo de un diagrama para un modelo de ecuaciones estructurales, en el que se incluyen dos variables latentes exógenas, una variable latente endógena con su correspondiente error de medición, cuatro variables observadas exógenas con sus correspondientes errores de medición y dos variables observadas endógenas con sus errores de medición, se incluyen también las cargas factoriales y la asociación hipotética entre las variables latentes.

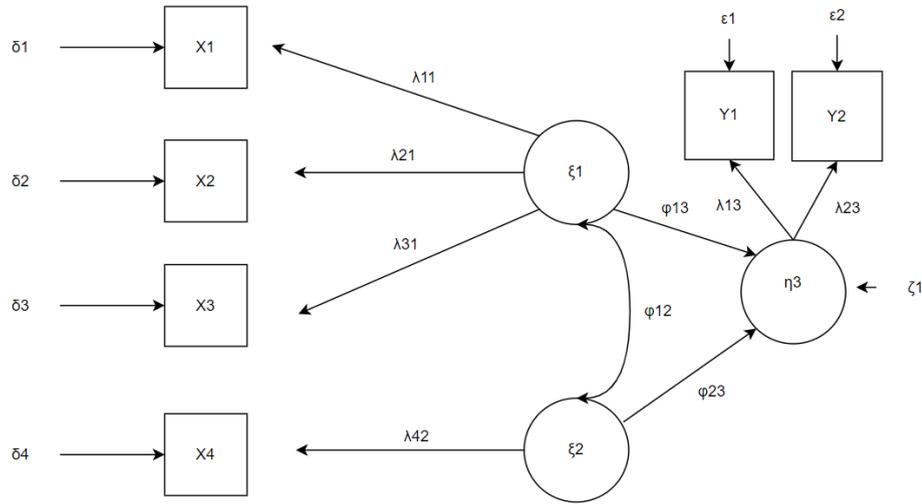


Figura 2.3: Diagrama de Relaciones: Ejemplo de un modelo de ecuaciones estructurales.

#### 2.4.2. Estimación de los parámetros del Modelo

La tarea principal de la estimación de los parámetros tanto de un modelo de AFC como de los modelos de ecuaciones estructurales es encontrar estimaciones de las matrices básicas que definen cada uno de los modelos. En el caso de un modelo de AFC sería la estimación de las tres matrices básicas  $\hat{\Gamma}$ ,  $\hat{\Phi}$ ,  $\hat{\Theta}_\delta$  y en el caso de los modelos de ecuaciones estructurales la estimación de las ocho matrices básicas  $\hat{B}$ ,  $\hat{\Gamma}$ ,  $\hat{\Phi}$ ,  $\hat{\Psi}$ ,  $\hat{\Lambda}_X$ ,  $\hat{\Theta}_\delta$ ,  $\hat{\Lambda}_Y$  y  $\hat{\Theta}_\epsilon$ . Estas estimaciones se buscan para que la matriz de varianzas/covarianzas implícita del modelo  $\Sigma(\hat{\Theta})$ , sea lo más parecida posible a la matriz no restringida,  $\Sigma$ , este proceso de estimación es un proceso iterativo.  $\hat{\Theta}$  es un vector que contiene una estimación de los  $p$  parámetros del modelo.

La estimación de los parámetros para un modelo de AFC es equivalente a la de un modelo de ecuaciones estructurales. A continuación, se expone como realizar la estimación para los parámetros de un modelo de ecuaciones estructurales. Para ello, disponemos de estimaciones iniciales de las 8 matrices básicas,  $\bar{B}$ ,  $\bar{\Gamma}$ ,  $\bar{\Phi}$ ,  $\bar{\Psi}$ ,  $\bar{\Lambda}_X$ ,  $\bar{\Theta}_\delta$ ,  $\bar{\Lambda}_Y$  y  $\bar{\Theta}_\epsilon$  y suponemos también que  $\Sigma(\bar{\Theta})$ , es lo suficiente parecida a  $\Sigma$ . Entonces estas estimaciones se consideran las mejores estimaciones posibles y el proceso iterativo termina. En el caso de que esto no ocurra se necesita encontrar nuevas estimaciones de las 8 matrices básicas,  $\bar{B}$ ,  $\bar{\Gamma}$ ,  $\bar{\Phi}$ ,  $\bar{\Psi}$ ,  $\bar{\Lambda}_X$ ,  $\bar{\Theta}_\delta$ ,  $\bar{\Lambda}_Y$  y  $\bar{\Theta}_\epsilon$  y por lo tanto  $\Sigma(\bar{\Theta})$  es más parecida a  $\Sigma$  que  $\Sigma(\bar{\Theta})$ . El proceso de mejora de las estimaciones de los parámetros a partir de las obtenidas durante la iteración anterior se repite hasta que se consigue una aproximación satisfactoria de la matriz de varianzas/covarianzas implícita en el modelo y la no restringida. Para comprobar si realmente esas matrices son parecidas se utiliza una función de ajuste, que depende del método de estimación utilizado,  $F[S, \Sigma(\theta)]$ . Esta función de ajuste tiene como parámetros  $S$ , que es la matriz de varianzas/covarianzas muestral de las variables observadas, y  $\Sigma(\theta)$ . Esta función de ajuste debe satisfacer las siguientes condiciones generales:

- Los valores de  $F[S, \Sigma(\theta)]$  son escalares positivos.
- $F[S, \Sigma(\theta)] \geq 0$

- $F[S, \Sigma(\theta)] = 0$  si y solo si  $S = \Sigma(\theta)$ .

Es decir, la función de ajuste se minimiza iterativamente, esto implica que el vector  $\tilde{\theta}$  contenga mejores estimaciones de los elementos en las matrices básicas que  $\bar{\theta}$ . La estimación final corresponde al valor mínimo de la función de ajuste, que se utiliza para comprobar el ajuste de los datos al modelo hipotetizado.

Los métodos de estimación se utilizan para encontrar los valores de los parámetros del modelo más compatibles con los datos observados. Estos minimizan una función de ajuste  $F$  y proporcionan para un modelo correctamente especificado y un conjunto de datos suficientemente grande, estimaciones coherentes de los parámetros, errores estándar y estadísticos de prueba adecuados, que pueden utilizarse para realizar pruebas de significancia de los parámetros del modelo y determinar la precisión de las estimaciones. Se conoce que  $(n - 1)$  veces el valor mínimo de la función de ajuste se distribuye como un estadístico  $\chi^2$  aproximado con  $df = (c - p)$ , donde  $c$  denota el número de elementos no redundantes en  $\Sigma$ , y  $p$  es el número de parámetros libres en el vector  $\theta$ . A continuación, se exponen algunos de los métodos de estimación más utilizados.

### **Método de Máxima Verosimilitud**

El método más utilizado es el de Máxima Verosimilitud (ML). Este método depende del supuesto de que las variables observadas siguen una distribución normal multivariante y depende de la escala de medida de los elementos. Este método proporciona las estimaciones de los parámetros que maximizan la probabilidad de observar los datos reales. La función de ajuste del método de Máxima Verosimilitud es,

$$F_{ML}[S, \Sigma(\theta)] = \ln|\Sigma(\theta)| - \ln|S| + \text{tr}[S\Sigma(\theta)^{-1}] - (n + o) \quad (2.31)$$

### **Método de Mínimos Cuadrados Generalizados**

Otro de los métodos utilizados es el de Mínimos Cuadrados Generalizados (GLS). Este método también trabaja sobre el supuesto de normalidad de los datos, minimiza la diferencia entre los valores observados de la muestra,  $S$  y sus correspondientes valores predichos en  $\Sigma(\theta)$ . La función de ajuste para el método de Mínimos Cuadrados Generalizados se puede representar como,

$$F_{GLS}[S, \Sigma(\theta)] = \frac{1}{2} \text{tr}[S^{-1}(S - \Sigma(\theta))]^2 \quad (2.32)$$

### **Método de Mínimos Cuadrados No Ponderados**

Por último, se presenta el método de Mínimos Cuadrados No Ponderados (ULS), este método no requiere que los datos sigan una distribución concreta. Cuando los datos no siguen una distribución normal multivariante este es el método más utilizado. Su función de ajuste minimiza la suma de cuadrados de cada elemento de  $(S - \Sigma(\theta))$  :

$$F_{ULS}[S, \Sigma(\theta)] = \frac{1}{2} \text{tr}[(S - \Sigma(\theta))^2] \quad (2.33)$$

### 2.4.3. Evaluación del ajuste del Modelo

Una vez que se han estimado los parámetros del modelo, debemos comprobar que los datos se ajustan al modelo hipotetizado, es decir, hay que comprobar si realmente el modelo especificado hace una representación viable de las relaciones entre las variables observadas y las variables latentes. Hay que tener en cuenta que un modelo se puede rechazar basándose en un desajuste observado entre los datos y el modelo, pero no se puede confirmar que una estructura concreta sea el modelo correcto, aunque el ajuste sea aceptable. Antes de comprobar que los datos se ajustan bien al modelo se deberían de tener en cuenta las siguientes consideraciones:

1. Si fuera posible se deberían de formular varios modelos.
2. Se debería comparar los resultados del ajuste con los de estudios anteriores del mismo modelo o similares siempre que sea posible.
3. Además de las medidas que evalúan el ajuste del modelo hipotético en su conjunto, habría que considerar medidas de ajuste de varios componentes de un modelo.
4. Se deberían utilizar diferentes medidas de ajuste.

Para poder comprobar que los datos se ajustan al modelo hipotetizado a priori disponemos de varias medidas de ajuste entre los datos y el modelo, entre las medidas más utilizadas están el estadístico chi-cuadrado, los índices de bondad de ajuste, los índices de ajuste normalizado y no-normalizado, los índices de bondad de ajuste basados en el principio de parsimonia y el error cuadrático medio aproximado, entre otros. En esta sección explicaré con más detalle estos índices:

#### El estadístico Chi-Cuadrado

Como acabo de mencionar una de las medidas de las que disponemos para comprobar el ajuste entre los datos y el modelo es el estadístico chi-cuadrado. Esta prueba evalúa si la matriz de varianzas/covarianzas no restringida de las variables observadas  $\Sigma$  es igual o no a la matriz de varianzas/covarianzas implícita en el modelo,  $\Sigma(\theta)$ , es decir, prueba la hipótesis nula,  $H_0: \Sigma = \Sigma(\theta)$ . Si no rechazamos  $H_0$ , entonces el modelo especificado es correcto y las varianzas y covarianzas de las variables observadas en  $\Sigma$  pueden escribirse como una combinación de los parámetros implícitos del modelo que se da en  $\Sigma(\theta)$ . En la práctica debemos estimar los elementos de  $\Sigma$  mediante los elementos de la matriz de varianzas/covarianzas de la muestra,  $S$ , y los elementos del vector  $\theta$  ya que tampoco los conocemos. Sabemos que se dispone de varios métodos de estimación de los parámetros del modelo, todos ellos utilizan una función de ajuste,  $F$ , que depende del método utilizado. Bajo  $H_0$ , sabemos que  $(n-1)$  veces el valor mínimo de la función de ajuste,

$$\chi^2 = (q - 1)F[S, \Sigma(\hat{\theta})] \quad (2.34)$$

se distribuye asintóticamente como una distribución  $\chi^2$  con  $(c-p)$  grados de libertad, donde  $q$  indica el tamaño de la muestra,  $c = (n)(n+1)/2$  es el número de varianzas y covarianzas no redundantes de las variables observadas y  $p$  el número total de parámetros a estimar. Entonces si se cumple  $H_0$  en la población, si se cumplen ciertos supuestos de distribución y si el tamaño de la muestra es lo suficientemente grande, la hipótesis del ajuste de los datos al modelo puede evaluarse mediante el estadístico chi-cuadrado. Cuando se cumplen

esas condiciones el estadístico chi-cuadrado es perfectamente aplicable para la evaluación del modelo, pero puede que este método no sea el óptimo debido a que presenta ciertas deficiencias:

- El método  $\chi^2$  depende de que las variables observadas sigan una distribución normal multivariante y de que el tamaño de la muestra sea suficientemente grande.
- Los modelos más complejos suelen conseguir mejores resultados de ajuste que los más sencillos, es decir, que si se pueden estimar más parámetros, se imponen menos restricciones a la matriz de varianzas/covarianzas, lo que reduce el valor de la función de ajuste,  $F[S, \Sigma(\theta)]$  y por lo tanto el valor del estadístico  $\chi^2$ .
- A medida que aumenta el tamaño de la muestra, también lo hace el estadístico de la ecuación (2.34), lo que hace que se puedan rechazar modelos plausibles basándose en un estadístico  $\chi^2$  significativo, aunque la diferencia entre  $S$  y  $\Sigma(\hat{\theta})$  sea mínima y sin importancia.

Debido a estas limitaciones es preferible comparar de manera informal la magnitud de un valor  $\chi^2$  observado con la media de su distribución de muestra subyacente [ $E(\chi^2) = df$ ], es decir, que un valor chi-cuadrado más bien pequeño es indicativo de un buen ajuste y un valor más bien grande es indicativo de un mal ajuste entre los datos y el modelo. Un p-valor menor de 0.05 también indica un buen ajuste del modelo y si se obtiene un valor del chi-cuadrado entre los grados de libertad del modelo ( $\chi^2/g.l$ ) menor de 3 también es indicativo de un buen ajuste del modelo.

### Índice de bondad de ajuste e índice de bondad de ajuste modificado

Otra forma de comprobar si los datos se ajustan al modelo es determinar la cantidad de información de la varianza/covarianza observada que puede ser explicada por el modelo hipotetizado, esto se hace mediante el índice de bondad de ajuste (GFI) y el índice de bondad de ajuste modificado (AGFI) de Jöreskog y Sörbom [19]. El valor de ambos índices está comprendido entre 0 y 1. Valores mayores de estos índices, es decir, cercanos a 1 indican un mejor ajuste entre los datos y el modelo. El índice de bondad de ajuste puede expresarse como,

$$GFI = 1 - \frac{F[S, \Sigma(\hat{\theta})]}{F[S, \Sigma(\theta)]} \quad (2.35)$$

donde el numerador es el mínimo valor de la función de ajuste,  $F$ , para el modelo hipotetizado, y el denominador es el mínimo valor de  $F$  cuando no se hipotetiza ningún modelo.

El índice de bondad de ajuste ajustado puede expresarse como,

$$AGFI = 1 - \frac{c}{df_h} \frac{F[S, \Sigma(\hat{\theta})]}{F[S, \Sigma(\theta)]} = 1 - \frac{c}{df_h} (1 - GFI) \quad (2.36)$$

donde  $c$  es el número de varianzas/covarianzas no redundantes de las variables observadas y  $df_h = c - p$  son los grados de libertad para el modelo hipotetizado. El AGFI representa una versión reducida del GFI, que intenta penalizar los modelos más complejos.

Para concluir, hay que tener en cuenta que, si aumenta el número de parámetros libres, los grados de libertad disminuyen y disminuye el AGFI, en cambio si los grados de libertad se acercan al valor de  $c$ , el AFGI se aproxima al valor de GFI.

## Índice de ajuste normalizado y no normalizado

El índice de ajuste normalizado (NFI) propuesto por Bentler y Bonett [3], y el índice de ajuste no normalizado (NNFI), también conocido como índice de Tucker-Lewis (TLI) propuesto por Tucker y Lewis [43], son medidas que evalúan el ajuste de los datos al modelo comparando un modelo hipotético con otro modelo de referencia, pero más restrictivo, es decir, con menos parámetros a estimar.

Para los modelos mencionados denominamos  $F_h$  al valor mínimo de la función de ajuste  $F$  para el modelo hipotético, y denotamos como  $df_h$  los grados de libertad del modelo, denominamos  $F_i$  al valor mínimo de la función de ajuste  $F$  para el modelo de referencia con  $df_i$  grados de libertad y así el índice de ajuste normalizado se define como:

$$NFI = \frac{\chi_i^2 - \chi_h^2}{\chi_i^2} = \frac{F_i - F_h}{F_i} = 1 - \frac{F_h}{F_i} \quad (2.37)$$

Este índice está normalizado entre 0 y 1, y valores mayores de este indican un mejor ajuste entre los datos y el modelo. Una de las desventajas que tiene este índice es que se ve afectado por el tamaño de la muestra y el índice podría no ser nunca 1 aunque el modelo hipotetizado sea correcto por eso se propuso el índice de ajuste no normalizado o índice de Tucker-Lewis,

$$NNFI = \frac{(\chi_i^2/df_i) - (\chi_h^2/df_h)}{\chi_i^2/df_i - 1} = \frac{(F_i/df_i) - (F_h/df_h)}{(F_i/df_i) - (1/(n-1))} \quad (2.38)$$

Como este índice no está normalizado puede presentar valores fuera del intervalo (0,1) y tiene en cuenta los grados de libertad del modelo hipotetizado y del modelo de referencia. También tiene alguna desventaja como que los modelos más restrictivos tienden a valores del índice más altos y los modelos menos restrictivos tienden a valores del índice más bajos. Valores del índice superiores a 0.9 indican un buen ajuste del modelo.

Debido a las desventajas que estos índices presentan, Bentler [2] propuso dos índices diseñados para resolver las desventajas que presentan el NFI y el NNFI, estos son el índice de ajuste comparativo (CFI), para solucionar el problema del NFI sobre el efecto del tamaño de la muestra, y el índice de ajuste no normado (FI) que se diseñó para reducir el número de veces que los valores del NNFI se sitúan fuera del intervalo (0,1), ambos índices tienen un parámetro de población asociado conocido. El CFI se puede expresar como,

$$CFI = 1 - \frac{l_1}{l_2} \quad (2.39)$$

donde  $l_1 = \max(l_h, 0)$  y  $l_2 = \max(l_h, l_i, 0)$ , y el FI se puede representar como,

$$FI = 1 - \frac{l_h}{l_i} \quad (2.40)$$

siendo  $l_h = [(n-1)F_h - df_h]$  y  $l_i = [(n-1)F_i - df_i]$ .

Como ocurre con el NFI, los valores de CFI se encuentran entre 0 y 1, y valores mayores a 0.9 indican un buen ajuste del modelo. Los valores de FI pueden estar fuera del intervalo (0,1).

## Índices de ajuste basados en el principio de parsimonia

Los índices de ajuste basados en el principio de parsimonia<sup>2</sup>, tienen en cuenta la complejidad del modelo hipotetizado a la hora de evaluar el ajuste de los datos al modelo. James *et al.* [15] y Mulaik *et al.* [29] propusieron el índice de bondad de ajuste de parsimonia (PGFI) y el índice de ajuste normalizado de parsimonia (PNFI), en los que se combina la bondad de ajuste del modelo con la parsimonia del modelo, ya que los índices como el GFI y el NFI no tenían en cuenta el principio de parsimonia.

El PGFI se define como,

$$PGFI = \frac{df_h}{df_n} GFI \quad (2.41)$$

y el PNFI se define como,

$$PNFI = \frac{df_h}{df_n} NFI \quad (2.42)$$

donde  $df_h$  representan los grados de libertad asociados con el modelo hipotético y  $df_n = c$  son los grados de libertad cuando el modelo no ha sido hipotetizado, es decir,  $c$  es el número de varianzas y covarianzas no redundantes de las variables observadas, y  $df_i$  son los grados de libertad del modelo de referencia. Estos índices toman valores entre 0 y 1, y valores próximos a 1 indican un buen ajuste del modelo.

## Error cuadrático medio aproximado

El error cuadrático medio aproximado o RMSEA [40] es uno de los índices de ajuste propuesto más recientemente, analiza la diferencia entre la matriz de correlaciones observada y estimada, se puede definir como:

$$RMSEA = \sqrt{\frac{\chi^2 - df}{(N - 1)df}} \quad (2.43)$$

donde  $N$  es el tamaño de la muestra y  $df$  son los grados de libertad del modelo.

Se considera un buen ajuste del modelo si el valor del índice es menor de 0.05, se considera un ajuste aceptable si el valor del índice es menor de 0.08, y con valores superiores a 0.08 se considera que el ajuste del modelo no es aceptable.

### 2.4.4. Modificación del Modelo

Una vez realizadas las comprobaciones sobre si los datos ajustan bien al modelo, podemos encontrarnos dos opciones, la primera es que efectivamente el ajuste sea "correcto". Como ya he comentado, esto no quiere decir que este modelo sea el verdadero ya que puede darse el caso de que esos mismos datos ajusten igual o mejor a otro modelo completamente diferente y que ese modelo sea el verdadero. La otra opción que nos podemos encontrar es que se den incoherencias entre los datos y el modelo. En el caso de que esto suceda tenemos diferentes formas de proceder:

---

<sup>2</sup>El principio de parsimonia o la Navaja de Ockham es un principio metodológico y filosófico que prioriza las explicaciones más sencillas frente a las demás [12].

1. Al encontrarse incoherencias entre los datos y el modelo, se podría intentar desarrollar un nuevo modelo que sea coherente con los datos y evaluar la viabilidad de este, es decir, modificar el modelo hipotetizado a priori.
2. Si se han especificado varios modelos a priori para los mismos conjuntos de datos y al evaluar el ajuste con los datos, todos ellos presentan incoherencias con los datos, se podría escoger como el modelo "correcto" aquel que presente menos incoherencias, siempre que los resultados del ajuste se consideren aceptables.

La forma más común de proceder es mediante la primera opción, es decir, corregir y modificar el modelo hipotetizado a priori para generar un nuevo modelo. Hay que tener en cuenta que todas las modificaciones "post hoc" deben tener sentido y ser teóricamente justificables. Siguiendo este procedimiento, Kaplan [21] utilizó el término error de especificación externo para definir el error de haber omitido variables importantes en el modelo y el término error de especificación interno para definir el error de haber omitido relaciones importantes dentro del modelo. Existen algunas medidas que pueden ayudar a identificar los errores de especificación internos, pero estas medidas por sí solas no proporcionan una justificación razonable para modificar el modelo hipotetizado a priori.

Algunas de las medidas que podemos utilizar son los índices de modificación (IM), como el multiplicador de Lagrange o el contraste de Wald, junto con estadísticas de cambio esperado de parámetros (EPC), para cada parámetro fijo de un modelo. El IM es una estimación de la disminución de la medida de mal ajuste del  $\chi^2$  si se liberara el parámetro, y el EPC es un valor estimado de ese parámetro, estas medidas ayudan a decidir si se libera o fija un parámetro concreto. Un parámetro fijo es un valor constante asignado a la relación entre dos variables del modelo. Liberar un parámetro implica que se este se estima libremente.

Saris, Satorra y Sörbom [37] y Kaplan [21] sugirieron que,

1. Un parámetro fijo con un IM y un EPC grandes puede liberarse.
2. Un parámetro fijo con un IM grande pero un EPC pequeño debería permanecer fijo.
3. Un parámetro fijo con un IM y un EPC pequeños debe permanecer fijo.
4. Si un parámetro fijo está asociado a un IM pequeño pero un EPC grande, la situación no está clara y es necesario conocer en más detalle los errores de especificación.

Jöreskog y Sörbom [19] definieron una posible forma de modificar el modelo que implica la utilización del estadístico chi-cuadrado, concluyeron que un  $\chi^2$  grande debería tomarse como una indicación de que ciertos parámetros deberían liberarse para mejorar el ajuste entre los datos y el modelo. En cambio, un  $\chi^2$  muy pequeño podría indicar que el modelo está sobreajustado, en el sentido de que ciertos parámetros podrían fijarse en 0 sin aumentar significativamente el valor  $\chi^2$ .

Hay que tener en cuenta que modificar el modelo no quiere decir que resulte un modelo con los mejores resultados e incluso pueden llegar a generarse varios modelos con resultados parecidos que deban compararse. Cuando tenemos el nuevo modelo, este se evalúa sobre el mismo conjunto de datos lo que suele llevar a que los resultados sean mejores, pero esto se puede deber a que solo se utiliza un conjunto de datos para evaluar el modelo.

Una forma de paliar las posibles consecuencias de la modificación del modelo sería validar mediante validación cruzada la estructura o estructuras que se hayan modificado con una muestra nueva e independiente. Si no se dispone de una muestra de este tipo, pero la muestra inicial es suficientemente grande, Cudeck y Browne [4] sugirieron que se reprodujeran los procedimientos de validación cruzada disponibles para regresiones lineales. Esto es, dividir la muestra aleatoriamente en submuestras de calibración y validación y calcular un índice de validación cruzada (CVI) midiendo la distancia entre la matriz de varianzas/covarianzas sin restricciones obtenida de la muestra de validación  $S_v$  y la matriz de varianzas/covarianzas implícita del modelo de la muestra de calibración  $\Sigma(\hat{\theta}_c)$ . Este índice puede expresarse como,

$$CVI = F[S_v, \Sigma(\hat{\theta}_c)], \quad (2.44)$$

donde  $F$  es una función de ajuste del método de estimación utilizado. Cuanto menor sea el valor del CVI mejor será la validez predictiva estimada del modelo, en el caso de haber considerado varios modelos a priori, el modelo con el menor CVI será el que consideremos como el mejor, pero si no se dispone de diferentes modelos para comparar, se podría comparar el CVI del modelo generado con el modelo saturado<sup>3</sup> o con el modelo independiente<sup>4</sup>.

Si por alguna razón no es posible o razonable dividir la muestra en submuestras de calibración y validación, Cudeck y Browne [4] sugirieron utilizar una estimación de muestra única del índice de validación cruzada esperado (ECVI).

$$ECVI = F[S_v, \Sigma(\hat{\theta}_c)] + \frac{2p}{q-1}, \quad (2.45)$$

donde  $F[S_v, \Sigma(\hat{\theta}_c)]$  es el valor mínimo de la función de ajuste de la estructura hipotetizada,  $p$  es el número de parámetros y  $q$  es el tamaño de la muestra. El modelo con menor ECVI será el que se considere el mejor.

#### 2.4.5. Interpretación de los resultados

Una vez hemos obtenido el modelo más "correcto" para los datos de los que se dispone, el último paso es interpretar los resultados obtenidos al construir y desarrollar un modelo de AFC o un modelo de ecuaciones estructurales. Debemos comprender los resultados obtenidos, es decir, analizar los valores de los parámetros, como las asociaciones hipotéticas entre las variables latentes e interpretar la relación entre las variables observadas y latentes. Todo ello para evaluar si los resultados del modelo son consistentes con las hipótesis iniciales del estudio que queremos realizar.

---

<sup>3</sup>Modelo sin restricciones en la que se estiman tantos parámetros como el número de varianzas y covarianzas de las variables observadas, es decir,  $df = 0$ .

<sup>4</sup>Modelo en el que las variables observadas no están correlacionadas.

## Capítulo 3

# Caso de Estudio

### 3.1. Descripción del Caso de Estudio

Es bien conocido que la inteligencia y la capacidad de trabajo son claves para el éxito en el rendimiento escolar. Debido a los crecientes estudios que relacionan la influencia de distintos factores personales y contextuales sobre el rendimiento escolar, para ilustrar la utilidad de los modelos presentados en el Capítulo 2 se procederá a realizar un estudio en el que se explorarán las relaciones entre diversos aspectos personales de niños y adolescentes y se determinará la influencia que puedan tener estos aspectos sobre su rendimiento escolar en el área de las matemáticas [5]. Hay tres factores cuya relación nos interesa estudiar debido a su posible influencia en la calificación en matemáticas. Estos son, la inteligencia emocional, las habilidades sociales de prevención del acoso escolar y la gaudibilidad o capacidad de disfrute.

El término inteligencia emocional fue popularizado por Goleman [13] que lo definió como la capacidad que tienen las personas para reconocer sus emociones y las de los demás y la capacidad que tienen de gestionarlas. Las emociones son experiencias subjetivas por lo que no es posible su medición de manera directa. Por ello, existen herramientas como test o cuestionarios que permiten medir la inteligencia emocional. Sin embargo, hay que tener en cuenta que todos ellos van a estar sujetos a cierto grado de subjetividad. A pesar del carácter subjetivo de este factor, es necesario conocer las emociones y competencias sociales y fomentar estas competencias para el desarrollo personal en los niños y adolescentes [11].

Otro factor que se incluye en el estudio son las habilidades sociales de prevención del acoso escolar. Estas se refieren a la percepción que tienen niños y adolescentes sobre el acoso escolar o *bullying*. Olweus [31] describió el acoso escolar como la situación en la que un niño o adolescente sufre acciones negativas, es decir, daño o malestar hacia su persona por parte de uno o más compañeros. Como ocurre con las emociones el acoso escolar y las habilidades sociales para prevenirlo no pueden medirse de forma directa. El acoso escolar está determinado por las interacciones sociales, y las habilidades sociales varían según la situación y las personas involucradas. También existen diferentes métodos para su medición y en todos ellos hay que tener en cuenta el carácter subjetivo de este factor a la hora de su evaluación e interpretación.

El último factor sería la gaudibilidad o capacidad de disfrute. Se ha relacionado la baja capacidad de disfrute directamente con diferentes aspectos como son el abuso, tanto físico como mental, la negligencia, etc. Fue Ribot [35] quien definió el término anhedonia como la incapacidad de experimentar disfrute en la vida cotidiana, esta incapacidad de disfrute se puede considerar un factor de riesgo para niños y adolescentes. El disfrute es una experiencia personal por ello, medir la satisfacción y placer que puede experimentar una persona conllevará cierto grado de subjetividad. Al igual que ocurre con los otros factores se dispone de métodos para medir la capacidad de disfrute como los cuestionarios que siempre estará influenciados por la perspectiva y subjetividad de cada persona.

Los factores que describo son claves en el desarrollo personal y social de los niños y adolescentes y aunque la medición de estos factores sea subjetiva esto no quiere decir que sea menos válida. El objetivo final del estudio es determinar, siempre que exista una correlación positiva entre estos factores y el rendimiento académico que se mide de forma directa a través de las calificaciones en matemáticas, la importancia de fomentar las habilidades sociales, y la inclusión de programas en los centros escolares sobre educación emocional para enseñar a los alumnos a manejar sus emociones.

Las hipótesis iniciales que planteo para el estudio de estos factores son:

1. Hay una relación lineal significativa entre la *Inteligencia Emocional* y las *Habilidades Sociales de Prevención del Acoso Escolar*.
2. Hay una relación lineal significativa entre la *Inteligencia Emocional* y la *Capacidad de Disfrute*.
3. Hay una relación lineal significativa entre la *Capacidad de Disfrute* y las *Habilidades Sociales de Prevención del Acoso Escolar*.
4. La *Inteligencia Emocional* influye positivamente en el *Rendimiento Académico*.
5. La *Capacidad de Disfrute* influye positivamente en el *Rendimiento Académico*.
6. Las *Habilidades Sociales de Prevención del Acoso Escolar* influyen positivamente en el *Rendimiento Académico*.

### **3.2. Descripción del conjunto de datos**

Para llevar a cabo el estudio dispongo de un conjunto de datos que contiene una muestra de 437 alumnos de 7 centros educativos de la provincia de Valladolid, pertenecientes a los cursos 4º, 5º, y 6º de Educación Primaria y 1º y 2º de Educación Secundaria. Este conjunto de datos tiene información sobre las respuestas a un *Cuestionario de Desarrollo Escolar* creado por Valdivieso (ver Anexo B) a partir de 3 cuestionarios, que miden los diferentes factores que se van a analizar:

1. El Test de Inteligencia Emocional para niños creado a partir de los componentes que expone Goleman sobre inteligencia emocional [6]. Consta de 23 preguntas para medir la *Inteligencia Emocional*, con 3 opciones de respuesta, 1, 2 o 3.
2. El Cuestionario de Acoso entre Iguales (CAI) [25], que mide las *Habilidades Sociales de Prevención de Acoso Escolar* y contiene 39 preguntas, con 7 opciones de respuesta, que van desde 1 a 7.

3. El Cuestionario EGNA que mide la *Gaudibilidad* o *Capacidad de disfrute* [32], que contiene 15 preguntas, con 7 opciones de respuesta, que van desde 1 a 7.

Junto con las preguntas seleccionadas de estos tres cuestionarios el *Cuestionario de Desarrollo Escolar* contiene preguntas sobre el centro educativo y la localidad a la que pertenece, el género, con tres opciones de respuesta, 'Masculino', 'Femenino' y 'Prefiero no decirlo', la edad del alumno, un identificador para cada alumno y la calificación media en matemáticas medida en una escala de 1 a 10, a través de la cual se mide el *Rendimiento Académico*.

### 3.3. Software Utilizado

Para llevar a cabo este estudio se utilizará R, un entorno de programación libre y gratuito que permite la realización de análisis estadísticos y gráficos. En concreto utilizaré la interfaz de R llamada *Rstudio*, es decir, un entorno que me va a permitir el uso de R de forma más cómoda e interactiva [44].

R dispone de diversos paquetes o librerías, es decir, de colecciones de funciones, datos y código para facilitar el uso de este software. Uno de los paquetes de los que dispone R es el paquete *lavaan* [36], este es un paquete libre y gratuito y proporciona una colección de herramientas que pueden utilizarse para el análisis factorial, modelos de ecuaciones estructurales, longitudinales, multinivel, modelos de datos perdidos, entre otros. Se utilizarán también paquetes complementarios a *lavaan* como el paquete *Psych* para realizar procedimientos sobre psicometría [48], el paquete *MVN* [22] cuya función entre otras es la realización de test de normalidad multivariante, el paquete *EFAtools* [42], para la realización de test como el de Kaiser-Guttman, el paquete *parameters* [24] para la extracción de parámetros en diversos modelos estadísticos, y el paquete *readxl*[46], que permite la lectura de archivos tipo '.xlsx', formato en el que se encuentran los datos de los que dispongo.

### 3.4. Preparación de los datos

Antes de la realización de cualquier procedimiento estadístico hay que preparar los datos para garantizar que sean adecuados para el análisis estadístico que voy a realizar. Primero voy a seleccionar las variables de interés para este caso de estudio, es decir, todas aquellas que se utilizan para medir cada uno de los factores que se quieren analizar.

```
1 datos=datos[,7:84]
```

Una vez que se han seleccionado las variables que son relevantes, debo comprobar que estas no presenten errores ni valores ausentes. Debido a la forma en la que se han tomado los datos no debería de presentarse ningún valor ausente. Al realizar la comprobación obtenemos *FALSE* como resultado, esto nos indica que no se ha encontrado ningún valor ausente.

```
1 >any(is.na(datos))  
2 [1] FALSE
```

Por último, para facilitar la realización de los procedimientos en *Rstudio* y su futura interpretación, renombraré las variables de interés. Las variables relativas al Test de Inteligencia Emocional, es decir, las que comienzan por *CIEMP*, pasarán a llamarse  $c_i$ , siendo

$i=1,\dots,23$ , ya que hay 23 preguntas relativas a la inteligencia emocional. Las variables relativas al cuestionario de acoso entre iguales, es decir, aquellas que comienzan por *CAI*, pasarán a llamarse  $d_j$ , siendo  $j=1,\dots,39$ . Las variables relativas al cuestionario EGNA se denominarán  $e_k$ , siendo  $k=1,\dots,15$ . Por último, la variable *Calificación media en Matemáticas* pasará a llamarse *CM*. En el Anexo B se puede ver de forma detallada los elementos que forman el *Cuestionario de Desarrollo Escolar* junto con sus equivalencias para su uso en *Rstudio*.

### 3.5. Validación de la escala de medida

Una vez que los datos ya están preparados, el siguiente paso de este caso de estudio es determinar la consistencia interna de cada una de las escalas de medidas de cada factor. En la sección 3.2 se plantea a través de qué elementos del *Cuestionario de Desarrollo Escolar* se quiere medir cada uno de los factores. Para determinar la consistencia interna de estas escalas utilizaré el Alpha de Cronbach, excepto para el *Rendimiento Académico* ya que no se puede calcular este coeficiente para una escala de medida que conste de un solo elemento. Para la comprobación de la consistencia interna del resto de escalas utilizaré la función *alpha()* del paquete *Psych*. En la sección 2.1 se determina que una escala se considerará fiable si presenta un valor del Alpha de Cronbach de 0.7 o superior.

Para la escala de medida de la *Inteligencia Emocional* obtengo un valor del Alpha de Cronbach de 0.88, por lo que se puede decir que la escala de medida es muy buena. Para la escala de medida de las *Habilidades Sociales de Prevención del Acoso Escolar* obtengo un valor de 0.97, por lo que se puede decir que la escala de medida es excelente. Y, por último, para la escala de medida de la *Capacidad de Disfrute* se obtiene un valor de 0.93 por lo que también se puede decir que la escala de medida es excelente.

### 3.6. Análisis factorial confirmatorio

Voy a llevar a cabo un análisis factorial confirmatorio. El primer paso será definir un modelo teórico en el que se represente la relación entre los factores descritos en la sección 3.1, que en el modelo propuesto serán los factores o variables latentes, y los elementos que miden cada uno de estos factores, que son las variables observadas. El AFC se utilizará para confirmar la estructura subyacente propuesta, es decir, la relación entre las variables observadas y los factores latentes. Por último, se comprobará la adecuación de los datos de los que dispongo al modelo propuesto. Para ello seguiré los pasos descritos en el Capítulo 2 de este documento.

#### 3.6.1. Exploración de los datos

La realización de algunos procedimientos relativos al AFC requiere conocer si los datos de los que dispongo siguen una distribución normal multivariante. Para llevar a cabo este análisis utilizaré la función *mvn()* del paquete *MVN*. Esta función tiene como entrada los datos y la prueba de Mardia, un test basado en la curtosis y la asimetría que comprueba si un conjunto de variables sigue una distribución normal multivariante [26]. Como se puede ver en la siguiente salida de *Rstudio* los datos no siguen una distribución normal multivariante. Realizo un test para averiguar si los datos siguen una distribución normal

univariante mediante el test de Shapiro-wilks [34]. Se debe saber que el cumplimiento del supuesto de normalidad univariante no implica que los datos sigan una distribución normal multivariante, aunque en el caso contrario si se cumple. Al realizar el test nos encontramos con el mismo resultado, los datos no siguen una distribución normal univariante.

```

1 > testmvm <- mvn(data = datos, mvnTest = "mardia")
2 > testmvm$multivariateNormality
3      Test      Statistic p value Result
4 1 Mardia Skewness 212422.731002983      0      NO
5 2 Mardia Kurtosis 222.11053568072      0      NO

```

Las variables observadas de las que dispongo son de tipo cuantitativo discreto, por lo que es de esperar que los datos no sigan una distribución normal multivariante ni univariante, ya que son distribuciones de tipo continuo. Aun así realizaré una estandarización de los datos para que tengan media 0 y desviación típica 1, mediante la función *scale()* del paquete *scales*[47].

Debido a los resultados obtenidos, el estadístico chi-cuadrado como medida de ajuste de los datos al modelo no será extremadamente fiable para determinar el correcto ajuste del modelo. Se descarta el uso del método de máxima verosimilitud que puede ser utilizado tanto para la estimación de los parámetros del modelo como para la extracción de factores. Por último, no sería adecuado el uso del test de esfericidad de Bartlett para evaluar la idoneidad de los datos.

### 3.6.2. Evaluación de la idoneidad de los datos

Para poder realizar un análisis factorial confirmatorio hay que verificar que los datos sean idóneos para realizar dicho análisis. Para poder realizar esta comprobación se dispone de dos pruebas, el test de esfericidad de Bartlett y el test de Kaiser-Meyer-Olkin (KMO). Como se especifica en la sección 3.6.1 no es adecuado el uso del test de esfericidad de Bartlett, pero si lo es el test KMO. Para ello utilizo la función *KMO()* del paquete *psych*. En la sección 2.2.7 se indica que valores mayores a 0.5 me indican que los datos son aceptables para este análisis.

```

1 Kaiser-Meyer-Olkin factor adequacy
2 Call: KMO(r = datos)
3 Overall MSA = 0.93
4 MSA for each item =
5   CM   c1   c2   c3   c4   c5   c6   c7   c8   c9   c10  c11  c12  c13  c14  c15  c16
6 0.84 0.92 0.89 0.90 0.94 0.76 0.85 0.70 0.88 0.81 0.86 0.88 0.88 0.91 0.74 0.87 0.80
7 c17  c18  c19  c20  c21  c22  c23  d1   d2   d3   d4   d5   d6   d7   d8   d9  d10
8 0.79 0.90 0.86 0.79 0.88 0.91 0.91 0.95 0.97 0.97 0.96 0.96 0.96 0.95 0.93 0.95 0.95
9 d11  d12  d13  d14  d15  d16  d17  d18  d19  d20  d21  d22  d23  d24  d25  d26  d27
10 0.97 0.94 0.96 0.93 0.97 0.95 0.96 0.97 0.95 0.94 0.93 0.96 0.94 0.96 0.95 0.95 0.96
11 d28  d29  d30  d31  d32  d33  d34  d35  d36  d37  d38  d39  e1   e2   e3   e4   e5
12 0.96 0.93 0.94 0.94 0.96 0.94 0.88 0.93 0.89 0.96 0.96 0.95 0.84 0.86 0.86 0.81 0.90
13 e6   e7   e8   e9   e10  e11  e12  e13  e14  e15
14 0.87 0.91 0.81 0.91 0.91 0.91 0.92 0.90 0.88 0.89

```

En esta salida de *Rstudio*, podemos ver que el valor del test KMO es de 0.93, por lo que se considera que los datos son muy adecuados para realizar este tipo de análisis. También se muestra el valor del test para cada elemento en el que se observa que todos los valores son mayores de 0.7, lo que corrobora la idoneidad de la muestra.

### 3.6.3. Definición del modelo

Una vez realizadas las comprobaciones iniciales debo definir un modelo para poder ser evaluado mediante el análisis factorial confirmatorio. En este modelo debo incluir los 4 factores que se miden a través del *Cuestionario de Acoso Escolar*, estos son: la *Inteligencia Emocional*, las *Habilidades Sociales de Prevención del Acoso Escolar*, la *Capacidad de Disfrute* y el *Rendimiento Académico*. Se deben incluir también los elementos que miden cada uno de los factores, es decir, las variables observadas. Por último, se establecen las relaciones entre los factores o variables latentes y las variables observadas. El primer paso es determinar el número de variables latentes en los que se va a dividir cada uno de los factores. Para ello, utilizaré la prueba de Kaiser-Guttman y el gráfico de sedimentación. El *Rendimiento Académico* se mide directamente a través de la *Calificación media en Matemáticas* por lo que la realización de estos test no es necesaria. El siguiente paso es determinar cuáles son las variables observadas que conforman cada uno de los factores o variables latentes. Este proceso se conoce como extracción de factores donde se buscará extraer los factores que expliquen la mayor cantidad de varianza. Este proceso lo realizaré mediante el método de mínimos cuadrados no ponderados (ULS) ya que lo considero el más apropiado debido a que los datos no siguen una distribución normal multivariante y este método no requiere que los datos sigan una distribución concreta. Aplicaré la rotación varimax para maximizar los pesos de cada factor y así simplificar la estructura de los factores y obtener una solución más sencilla y fácil de interpretar.

#### Inteligencia Emocional

##### Prueba de Kaiser-Guttman:

```
1 -- Number of factors suggested by Kaiser-Guttman criterion -----  
2 ( ) With PCA-determined eigenvalues: 5  
3 ( ) With EFA-determined eigenvalues: 3
```

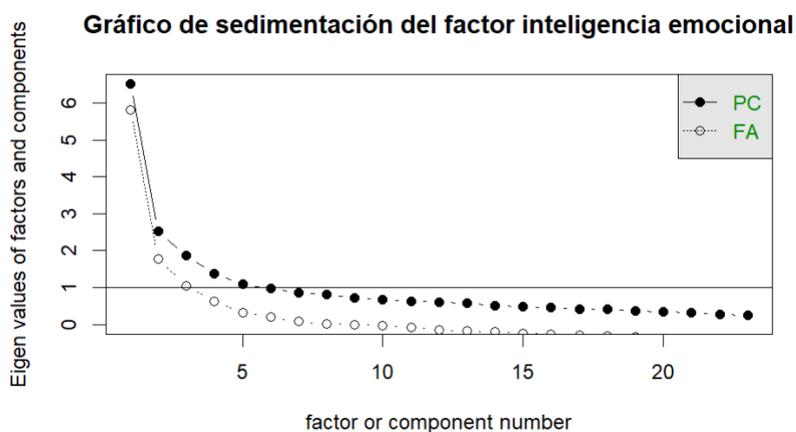


Figura 3.1: Gráfico de sedimentación de la *Inteligencia Emocional*

Tanto el gráfico de sedimentación como el test de Kaiser-Guttman proporcionan el número de variables latentes en los que se debería dividir el factor *Inteligencia Emocional*. Ambos utilizan análisis de componentes principales (ACP) y análisis factorial (AF) para la búsqueda de valores propios. Mediante ACP se obtienen 5 variables latentes y mediante el AF se obtienen 3. Compruebo con cuál de los dos métodos se obtiene la mayor cantidad

de varianza. La varianza obtenida mediante la extracción de 3 factores es del 40.10%, en cambio, mediante la extracción de 5 factores se obtiene una varianza del 47.84%. El resultado indica un aumento de la varianza al establecer 5 factores y debido a que la formación de 5 factores es consistente, establezco 5 como el número de variables latentes en las que se puede dividir el factor *Inteligencia Emocional*. A continuación, se muestra la extracción de factores en la que se identifican que variables observadas conforman cada variable latente.

Variable	ULS1	ULS4	ULS3	ULS2	ULS5	Complexity	Uniqueness
c23	0.75					1.30	0.35
c13	0.70					1.37	0.42
c18	0.66					1.24	0.51
c1	0.63					1.49	0.52
c12	0.48					1.92	0.67
c3	0.35					2.78	0.71
c10		0.76				1.33	0.34
c15		0.73				1.35	0.38
c22		0.58				2.05	0.49
c4		0.55				1.95	0.54
c19		0.31				3.28	0.75
c2		0.31				2.57	0.79
c14			0.81			1.04	0.34
c5			0.63			1.08	0.58
c9			0.59			1.55	0.55
c17			0.46			1.77	0.71
c16				0.85		1.13	0.23
c20				0.80		1.19	0.31
c8				0.61		1.68	0.50
c7					0.59	1.04	0.65
c6					0.57	1.71	0.55
c21					0.54	2.13	0.52
c11					0.48	2.17	0.59

The 5 latent factors (varimax rotation) accounted for 47.84% of the total variance of the original data (ULS1 = 12.66%, ULS4 = 11.05%, ULS3 = 9.11%, ULS2 = 8.60%, ULS5 = 6.42%).

Figura 3.2: Pesos factoriales de la *Inteligencia Emocional*

## Habilidades Sociales de Prevención del Acoso Escolar

*Prueba de Kaiser-Guttman:*

```

1 -- Number of factors suggested by Kaiser-Guttman criterion -----
2 ( ) With CPA-determined eigenvalues: 3
3 ( ) With EFA-determined eigenvalues: 3

```

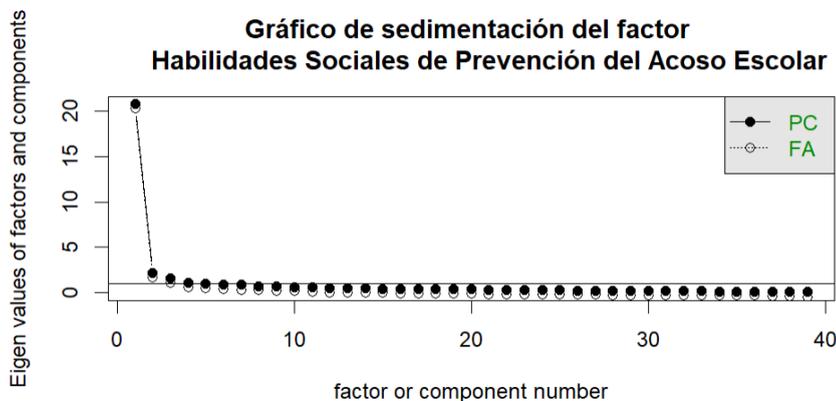


Figura 3.3: Gráfico de sedimentación de las *Habilidades Sociales de Prevención del Acoso Escolar*

Mediante la prueba de Kaiser-Guttman se obtiene 3 como el número de variables latentes en los que se debería dividir las *Habilidades Sociales de Prevención del Acoso Escolar*. Se obtiene el mismo resultado tanto si la búsqueda de valores propios se realiza mediante análisis de componentes principales como si se realiza mediante análisis factorial. El gráfico de sedimentación establece que el número de factores en los que se debería dividir las *Habilidades Sociales de Prevención del Acoso Escolar* son 3, tanto si se realiza la búsqueda de valores propios mediante ACP o AF. El número de variables latentes se establece entonces como 3 ya que mediante ambos procedimientos se obtiene el mismo resultado. A continuación, se realiza la extracción de factores para determinar las variables observadas que conforman cada variable latente. La varianza obtenida es del 60.04 %.

Variable	ULS1	ULS2	ULS3	Complexity	Uniqueness
d13	0.75			1.65	0.25
d3	0.74			1.54	0.30
d5	0.69			1.74	0.35
d1	0.69			1.40	0.43
d18	0.68			1.93	0.31
d7	0.67			2.11	0.30
d2	0.66			1.89	0.33
d15	0.63			1.93	0.39
d11	0.63			1.89	0.41
d9	0.59			1.96	0.47
d38	0.57			2.60	0.32
d12	0.56			1.98	0.46
d24	0.56			2.07	0.42
d22	0.54			2.59	0.40
d16	0.51			2.38	0.53
d20	0.50			2.20	0.55
d37	0.44			2.99	0.45
d29		0.77		1.43	0.28
d27		0.77		1.38	0.30
d30		0.73		1.36	0.37
d39		0.69		1.78	0.34
d4		0.67		1.89	0.35
d32		0.61		2.32	0.35
d35		0.60		2.37	0.38
d10		0.58		2.48	0.35
d33		0.58		2.63	0.34
d19		0.56		2.21	0.39
d25		0.56		2.13	0.42
d26		0.49		2.28	0.54
d28		0.49		2.98	0.36
d31			0.64	2.29	0.30
d14			0.63	2.08	0.36
d21			0.54	2.47	0.40
d23			0.53	2.83	0.35
d17			0.52	2.86	0.35
d34			0.51	1.97	0.59
d8			0.50	1.30	0.71
d36			0.49	2.13	0.54
d6			0.44	2.77	0.58

The 3 latent factors (varimax rotation) accounted for 60.04% of the total variance of the original data (ULS1 = 24.11%, ULS2 = 21.31%, ULS3 = 14.62%).

Figura 3.4: Pesos factoriales de las *Habilidades Sociales de Prevención del Acoso Escolar*

## Capacidad de Disfrute

### Prueba de Kaiser-Guttman:

```

1 -- Number of factors suggested by Kaiser-Guttman criterion -----
2 ( ) With CPA-determined eigenvalues: 2
3 ( ) With EFA-determined eigenvalues: 1

```

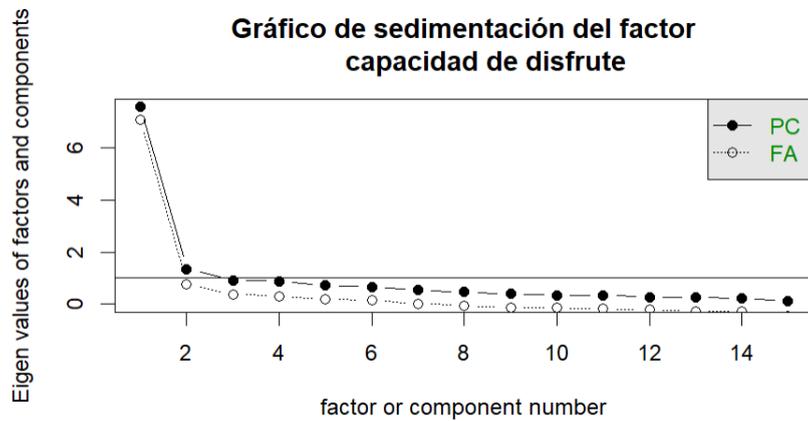


Figura 3.5: Gráfico de sedimentación de la *Capacidad de Disfrute*

El número de factores que se debería establecer según el gráfico de sedimentación son, 2 si la búsqueda de valores propios se hace mediante ACP y 1 si se hace mediante AF. El test de Kaiser-Guttman me indica el mismo resultado. Obtengo la varianza mediante 1 o 2 factores para comprobar cuál es el número de factores que se debería establecer. Con 1 factor se obtiene una varianza del 47.19% y mediante 2 factores se obtiene una varianza del 54.60%. Debido a que es consistente la formación de dos factores y al aumento de la varianza establezco el número de factores en que se puede dividir la *Capacidad de Disfrute* en 2. A continuación, se realiza la extracción de factores para determinar las variables observadas que forman cada variable latente.

# Rotated loadings from Factor Analysis (varimax-rotation)

Variable	ULS1	ULS2	Complexity	Uniqueness
e11	0.77		1.19	0.36
e6	0.72		1.28	0.41
e14	0.72		1.18	0.44
e7	0.69		1.21	0.47
e15	0.65		1.63	0.43
e9	0.63		1.56	0.48
e13	0.60		1.82	0.46
e2	0.58		1.24	0.62
e1	0.56		1.09	0.67
e5	0.56		1.93	0.48
e3	0.56		1.30	0.64
e10	0.53		1.81	0.57
e8		0.92	1.11	0.10
e4		0.85	1.15	0.23
e12		0.54	2.00	0.44

The 2 latent factors (varimax rotation) accounted for 54.60% of the total variance of the original data (ULS1 = 34.60%, ULS2 = 19.99%).

Figura 3.6: Pesos factoriales de la variable *Capacidad de Disfrute*

### 3.6.4. Representación del Modelo

El modelo propuesto constará del factor *Inteligencia Emocional* y las 5 variables latentes en que se divide, el factor *Habilidades Sociales de Prevención del Acoso Escolar* y las 3 variables latentes en las que se divide, el factor *Capacidad de Disfrute* y las 2 variables latentes en que se divide y las variables observadas indicadoras de cada variable latente según se indica en la extracción de factores. Incluirá también una variable latente unidimensional para el *Rendimiento Académico*, ya que este es medido directamente mediante la calificación en matemáticas.

La representación del modelo se realizará mediante un diagrama de relaciones (ver figura 3.7). Para facilitar la comprensión del diagrama de relaciones, denominaré *IE* al factor *Inteligencia Emocional*, cada una de los variables latentes en las que se divide este factor *Inteligencia Emocional* se denominarán *IE1*, *IE2*, *IE3*, *IE4* y *IE5*. Se denotará como *AE* al factor *Habilidades Sociales de Prevención del Acoso Escolar*, cada una de las variables latentes en las que se divide este factor se denominarán *AE1*, *AE2* y *AE3*. El factor *Capacidad de disfrute* pasará a ser *CD*, cada una de las variables latentes en las que se divide serán *CD1* y *CD2*. Por último, el factor *Rendimiento Académico* será *RA*. En la sección 3.4 se indica que las variables observadas que miden la *Inteligencia Emocional* se denotan como  $c_i$ , las variables observadas que miden las *Habilidades Sociales de Prevención del Acoso Escolar* se denotan como  $d_j$  y las variables observadas que miden la *Capacidad de Disfrute* se denotan como  $e_k$ .

En la representación del modelo se incluirán también los términos de error asociados a cada una de las variables observadas  $\delta_l$ , siendo  $l$ ,  $c_i$ ,  $d_j$ ,  $d_k$  o CM. El término de error asociado a la variable observada *Calificación media en Matemáticas* se fija en cero debido a que es la única variable observada indicadora del *Rendimiento Académico*. Se incluye también en el modelo el término de error asociado a cada una de las variables latentes  $\zeta_m$ , siendo  $m$  igual a *IE*, *IE1*, *IE2*, *IE3*, *IE4*, *IE5*, *AE*, *AE1*, *AE2*, *AE3*, *CD*, *CD1*, *CD2* o *RA*, se representarán también las cargas factoriales que relacionan las variables latentes y las variables observadas  $\lambda_{lm}$ , y las cargas factoriales que relacionan los factores con las variables latentes en que se divide cada uno, y la asociación hipotética entre variables latentes  $\phi_{mm}$ .

A continuación, se muestra el diagrama de relaciones para el modelo propuesto, la función *semPath()* del paquete *semPlot* [10], permite la visualización de los diagramas de relaciones creados mediante el paquete *lavaan*. La gran cantidad de variables que forman este modelo hace que la visualización de este diagrama sea incomprensible por lo que todos los diagramas presentes en este Trabajo de Fin de Grado no se podrán realizar mediante el software R. La realización de estos diagramas se hará mediante *draw.io* [1] un software gratuito que permite la creación de múltiples tipos de diagramas.

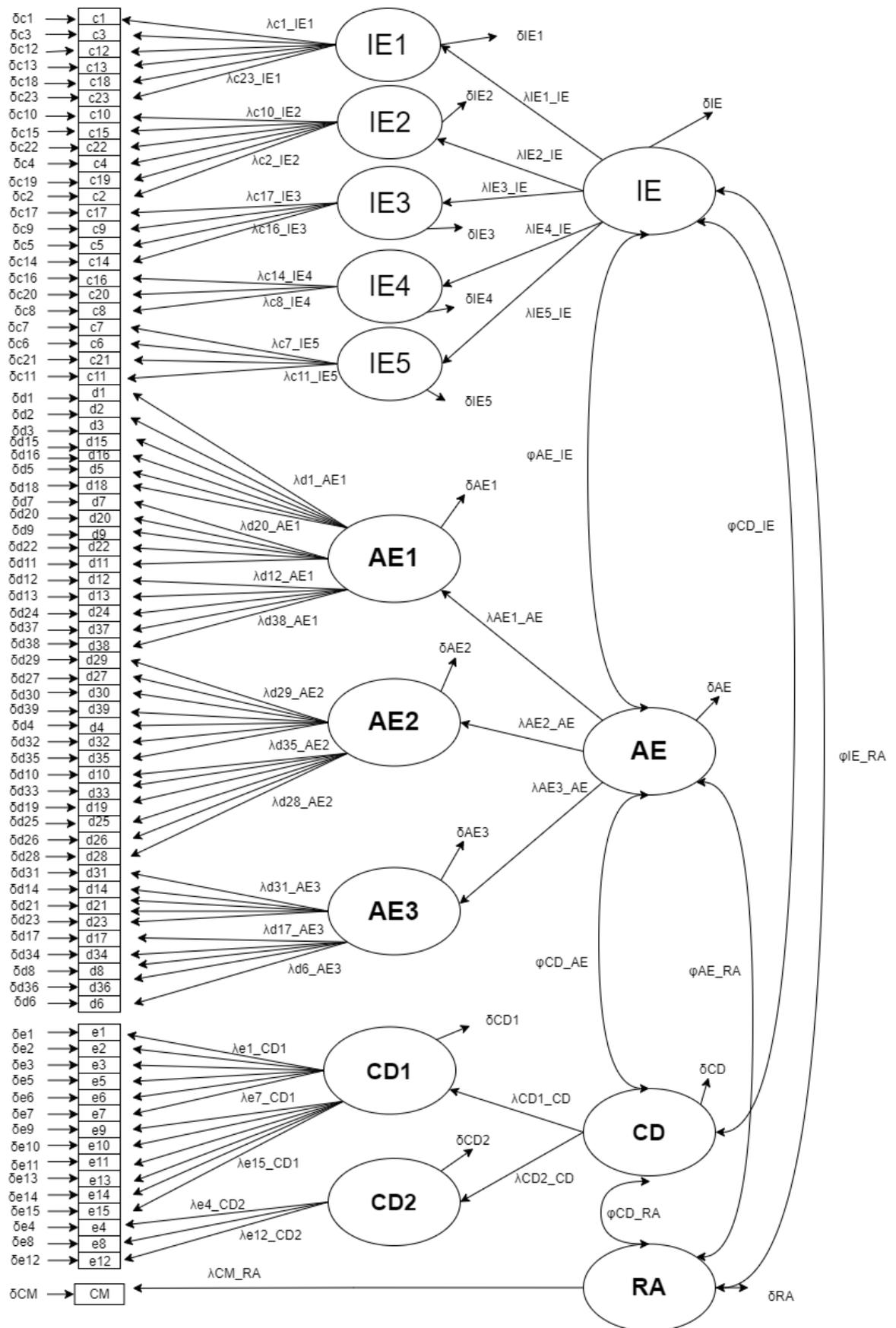


Figura 3.7: Diagrama de Relaciones del modelo propuesto.

### 3.6.5. Especificación del Modelo

Una vez definido y representado el modelo propuesto procedo a especificar el modelo utilizando el software R, para especificar el modelo el paquete *lavaan* cuenta con una sintaxis específica. El paquete dispone de una serie de operadores que nos indican las distintas relaciones entre variables que se pueden representar. El símbolo, ' $\sim$ ', permite representar las variables observadas indicadoras que conforman una variable latente. Mediante el símbolo ' $\sim$ ' se pueden representar las relaciones de regresión entre variables y mediante los símbolos ' $\sim\sim$ ' se pueden representar las covarianzas residuales o correlaciones entre errores. Por último, mediante el símbolo ' $\sim 1$ ' se define la intersección (*intercept*) de una variable (latente u observada) como 1. La especificación del modelo propuesto es la que se muestra a continuación.

```
1 modelo_afc <- ' IE1 =~ c1+c3+c12+c13+c18+c23
2                 IE2 =~ c10+c15+c22+c4+c19+c2
3                 IE3 =~ c17+c9+c5+c14
4                 IE4 =~ c16+c20+c8
5                 IE5 =~ c7+c6+c21+c11
6                 AE1 =~ d13+d3+d5+d1+d18+d7+d2+d15+d11+d9+d38+d12+d24+d22+d16+d20+d37
7                 AE2 =~ d29+d27+d30+d39+d4+d32+d35+d10+d33+d19+d25+d26+d28
8                 AE3 =~ d31+d14+d21+d23+d17+d34+d8+d36+d6
9                 CD1 =~ e1+e2+e3+e5+e6+e7+e9+e11+e13+e14+e15+e10
10                CD2 =~ e8+e4+e12
11                RA =~ CM
12                IE =~ IE1+IE2+IE3+IE4+IE5
13                AE =~ AE1+AE2+AE3
14                CD =~ CD1+CD2 '
```

### 3.6.6. Estimación de los parámetros del Modelo

La estimación de los parámetros del modelo requiere primero la identificación de los p parámetros que queremos estimar. Estos los podemos ver representados en el diagrama de relaciones de la figura 3.7 y son, los 77 errores correspondientes a cada una de las variables observadas, no se incluye el error asociado a la variable observada del rendimiento académico ya que se fija en 0, os 14 errores asociados a cada una de las variables latentes, las 78 cargas factoriales entre las variables observadas y latentes, las 10 cargas factoriales entre los factores y las variables latentes en que se dividen y las 6 asociaciones hipotéticas que se dan entre las variables latentes. Para poder estimar los p parámetros se tiene que cumplir la condición de que p sea menor que el número de varianzas/covarianzas no redundantes de las variables observadas, que es c, que se calcula como  $(n)(n+1)/2$ , donde  $n=78$ , que es el número de variables observadas, por lo que p será igual que 185, y se cumple que sí es menor que  $(78)(78+1)/2 = 3081$ .

El siguiente paso es ajustar el modelo para ello utilizaré la función *cfa()* del paquete *lavaan*, esta función tiene como entradas el modelo propuesto y los datos.

```
1 fit_afc <- cfa(model=modelo_afc, data=datos, estimator='ULS')
```

El método de estimación utilizado es el método de mínimos cuadrados no ponderados (ULS) ya que no requiere que los datos sigan ninguna distribución en particular, y los datos de los que disponemos no siguen una distribución normal multivariante. Por último, la estimación de los parámetros se obtendrá mediante la función *coef()* del paquete *lavaan*. A continuación, se muestra el diagrama de relaciones del modelo propuesto con los parámetros estimados.

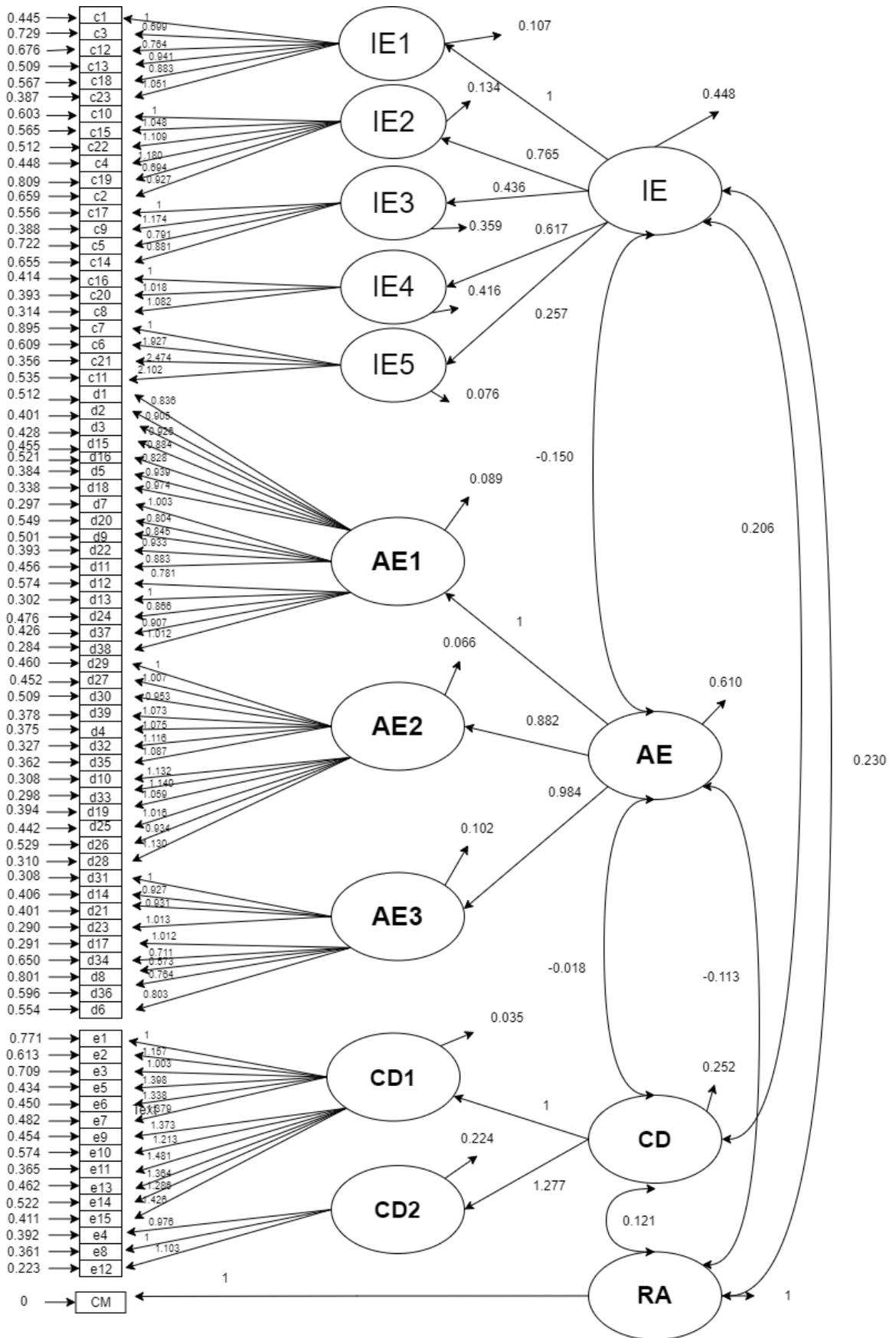


Figura 3.8: Diagrama de Relaciones del modelo propuesto con la estimación de los parámetros.

### 3.6.7. Evaluación del ajuste del Modelo

Para facilitar la comprobación del ajuste del modelo, con la función *fitMeasures()* del paquete *lavaan*, obtengo una tabla con los índices de ajustes más utilizados que han sido presentados en la sección 2.4.3. Estos son: el índice de bondad de ajuste y el índice de bondad de ajuste modificado, los índices de bondad de ajuste normalizado y no normalizado, el error cuadrático medio aproximado y el índice de ajuste comparativo, obtendré también el estadístico chi-Cuadrado, aunque su interpretación no será suficiente para obtener una conclusión debido a la falta de normalidad multivariante en los datos.

```
1 # Índices de ajuste
2 fitMeasures( fit_afc, c( 'gfi','agfi', 'nfi','nnfi','tli','rmsea','cfi' , 'chisq' ) )
```

GFI	AGFI	NFI	TLI	RMSEA	CFI	$\chi^2$
0.970	0.968	0.961	0.984	0.038	0.985	4706.438

Cuadro 3.1: Índices de ajuste del modelo propuesto

- Para los índices de bondad de ajuste (GFI) y bondad de ajuste modificado (AGFI), valores cercanos a 1 indican un buen ajuste de los datos y el modelo, los valores obtenidos son superiores a 0.9, según este criterio, el ajuste entre los datos y el modelo es correcto.
- Para el índice de bondad ajuste normalizado (NFI) el criterio es el mismo que para los índices anteriores, en este caso el valor obtenido es también superior a 0.9 por lo que se llegaría a la misma conclusión, el ajuste entre los datos y el modelo es correcto.
- El índice de bondad de ajuste no normalizado (NNFI) o índice de Tucker-Lewis (TLI), establece que un modelo está bien ajustado si el valor del índice es mayor de 0.9, por lo que sí se puede decir que el ajuste del modelo es correcto.
- Mediante el error cuadrático medio aproximado (RMSEA) podemos saber si el ajuste de un modelo es correcto si este valor es próximo a 0 y menor de 0.05, este índice también nos indica que el ajuste es correcto, ya que el valor es de 0.038.
- Un valor próximo a 1 del índice comparativo (CFI), nos indica que el ajuste de los datos y el modelo es bueno, se obtiene la misma conclusión obtenida con los índices anteriores.
- Por último, en el caso del estadístico chi-cuadrado, valores altos de este, indican un mal ajuste de los datos y el modelo, aparentemente obtenemos un valor del estadístico elevado, comprobaremos si el criterio  $\chi^2/g.l < 3$  se cumple o si el p-valor es menor de 0.05. Obtenemos que 4706.438 entre los grados de libertad 2910, obtenidos mediante la función *summary()*, se cumple que es menor de 3, por lo que consideramos que el ajuste es correcto.

Se obtienen unos valores de los índices de ajuste que concluirían que el modelo es correcto, pero se obtiene también un valor del chi-cuadrado algo elevado aunque cumple los criterios para que el modelo sea correcto, debido a que la interpretación del chi-cuadrado no es extremadamente fiable ya que los datos no siguen una distribución normal multivariante, se considera este modelo como correcto. Se podría encontrar un modelo mejor por ejemplo un modelo en el que el valor del chi-cuadrado no fuera tan elevado, por ello voy a proceder a su modificación para intentar encontrar un modelo mejor.

### 3.6.8. Modificación del Modelo

Obtengo mediante la función *modindices()* también del paquete *lavaan*, los índices de modificación (IM) y las estadísticas de cambio esperado de parámetros (EPC). Los índices que se obtienen nos devuelven aquellos parámetros fijados en cero. Estos nos darán una orientación de las relaciones entre variables que disminuirán el estadístico chi-cuadrado, y en qué medida lo disminuirán. Las relaciones que nos pueden sugerir pueden ser relativas a correlaciones de errores de los elementos o variables observadas o a que una variable observada sea indicadora de un factor latente diferente al que hemos propuesto a priori. Como se expone en la sección 2.4.4, para poder liberar un parámetro fijo, es decir, para incluir alguna de esas relaciones en el modelo para que ese parámetro sea estimado se debe cumplir que el índice de modificación sea elevado y el estadístico de cambio esperado de parámetro se superior o igual a 0.15 [9]. El estadístico  $\chi^2$  depende del valor obtenido en el análisis del modelo para determinar a partir de que valor se considera elevado.

Obtengo los primeros 400 índices de modificación, ya que considero que el IM es suficientemente grande como para que se cumpla el primer criterio para poder liberar un parámetro. Entre ellos solo aquellos que sugieren una correlación de errores entre variables observadas. No voy a considerar aquellos que me indican que una variable observada puede ser indicadora de un factor diferente, ya que he obtenido que la consistencia interna de las escalas de medida de los factores es excelente. A continuación, se exponen todos los parámetros fijos que cumplen las condiciones para poder ser liberados. Para concluir que un parámetro fijo puede ser liberado no solo se deben cumplir las condiciones que he expuesto, si no también se debe verificar que hay una similitud entre las variables observadas y que estas sean indicadoras del mismo factor. Se puede ver que todos los parámetros fijos expuestos son correlaciones de los errores entre variables observadas que pertenecen al mismo factor.

CIEMP. 16. Siento que soy muy valioso.	~~	CIEMP.20. Siento que soy importante.	Los elementos son muy similares por lo que se puede añadir al modelo la correlación de los errores entre c16 y c20.
CIEMP.05. Soy bastante tranquilo.	~~	CIEMP.14. Casi siempre estoy muy relajado.	Los elementos son muy similares por lo que se puede añadir al modelo la correlación de los errores entre c5 y c14.
EGNA.08. Tengo mucha imaginación.	~~	EGNA.04. Soy una persona muy imaginativa.	Los elementos son muy similares por lo que se puede añadir al modelo la correlación de los errores entre e3 y e4.
CAI.34. No quieren hacer trabajos conmigo.	~~	CAI.36. No se quieren sentar conmigo.	Los elementos son muy similares por lo que se puede añadir al modelo la correlación de los errores entre d34 y d36.
CIEMP.10. Me gusta devolver favores.	~~	CIEMP.15. Siempre que puedo, devuelvo los favores que recibo.	Los elementos son muy similares por lo que se puede añadir al modelo la correlación de los errores entre c10 y c15.
CIEMP.07. Si veo llorar a un niño/a me dan ganas de llorar a mí también.	~~	CIEMP.06. Me pongo muy mal si veo que alguien se lastima.	Los elementos son muy similares por lo que se puede añadir al modelo la correlación de los errores entre c7 y c6.
CIEMP.10. Me gusta devolver favores.	~~	CIEMP.11. Cuando alguien está solo y aburrido me dan ganas de acercarme y jugar con él.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.

CIEMP.22. Me gusta agradecerle a la gente.	~~	CIEMP.11. Cuando alguien está solo y aburrido me dan ganas de acercarme y jugar con él.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CAI.29. Me hacen fotos o me graban con el móvil mientras se meten conmigo.	~~	CAI.30. Se burlan colgando fotos o vídeos sobre mí en internet.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CAI.29. Me hacen fotos o me graban con el móvil mientras se meten conmigo.	~~	CAI.27. Me envían mensajes al móvil o correos electrónicos para insultarme o amenazarme.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
EGNA.03. Fácilmente me río.	~~	EGNA.06. A lo largo del día hay muchas cosas que me hacen reír.	Los elementos son muy similares por lo que se puede añadir al modelo la correlación de los errores entre e3 y e6.
CIEMP.22. Me gusta agradecerle a la gente.	~~	CIEMP.06. Me pongo muy mal si veo que alguien se lastima.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CIEMP.10. Me gusta devolver favores.	~~	CIEMP.06. Me pongo muy mal si veo que alguien se lastima.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CIEMP.10. Me gusta devolver favores.	~~	CIEMP.21. Si alguien está llorando me dan ganas de abrazarlo o consolarlo.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CIEMP.15. Siempre que puedo, devuelvo los favores que recibo.	~~	CIEMP.11. Cuando alguien está solo y aburrido me dan ganas de acercarme y jugar con él.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CIEMP.22. Me gusta agradecerle a la gente	~~	CIEMP.21. Si alguien está llorando me dan ganas de abrazarlo o consolarlo.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CIEMP.19. Me quedo tranquilo aunque no pueda hacer lo que me gusta.	~~	CIEMP.17. Aunque tenga problemas, igual mantengo la calma.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CIEMP.03. La mayor parte de los días me siento en paz	~~	CIEMP.14. Casi siempre estoy muy relajado.	Los elementos son muy similares por lo que se puede añadir al modelo la correlación de los errores entre c3 y c14.
CAI.27. Me envían mensajes al móvil o correos electrónicos para insultarme o amenazarme.	~~	CAI.30. Se burlan colgando fotos o vídeos sobre mí en internet.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
EGNA.06. A lo largo del día hay muchas cosas que me hacen reír.	~~	EGNA.14. Acostumbro a reír varias veces al día.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CIEMP.04. Valoro cuando los demás me ayudan.	~~	CIEMP.16. Siento que soy muy valioso.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CAI.12. Me acusan de cosas que no he dicho o hecho.	~~	CAI.14. Les dicen a otros que no estén o que no hablen conmigo.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CIEMP.15. Siempre que puedo, devuelvo los favores que recibo.	~~	CIEMP.21. Si alguien está llorando me dan ganas de abrazarlo o consolarlo.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CAI.24. Me ponen la zancadilla.	~~	CAI.19. Me rompen cosas.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CAI.03. Me pegan patadas.	~~	CAI.02. Me pegan puñetazos.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CAI.24. Me ponen la zancadilla.	~~	CAI.25. Me pellizcan.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CIEMP.15. Siempre que puedo, devuelvo los favores que recibo.	~~	CIEMP.06. Me pongo muy mal si veo que alguien se lastima	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CIEMP.03. La mayor parte de los días me siento en paz.	~~	CIEMP.05. Soy bastante tranquilo	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.

CIEMP.04. Valoro cuando los demás me ayudan.	~~	CIEMP.21. Si alguien está llorando me dan ganas de abrazarlo o consolarlo.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
EGNA.03. Fácilmente me río	~~	EGNA.14. Acostumbro a reír varias veces al día.	Los elementos son muy similares por lo que se puede añadir al modelo la correlación de los errores entre e3 y e14.
CAI.24. Me ponen la zancadilla.	~~	CAI.25. Me pellizcan.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.
CAI.20. Me roban cosas.	~~	CAI.19. Me rompen cosas.	Los elementos no son lo bastante parecidos para liberar el parámetro fijo.

La especificación del nuevo modelo es la que se presenta a continuación, en la que se incluyen las correlaciones de los errores entre variables observadas que cumplen las condiciones. Para indicar estas correlaciones de los errores se utiliza el símbolo '~~' como se especifica en el apartado 3.6.5. Si un error de una variable esta correlacionado con más de un error se combinan en la misma expresión mediante el símbolo '+'.

```

1 modelo_afc_mod <- ' IE1 =~ c1+c3+c12+c13+c18+c23
2     IE2 =~ c10+c22+c4+c19+c2
3     IE3 =~ c17+c9+c5
4     IE4 =~ c16+c8
5     IE5 =~ c7+c21+c11
6     AE1 =~ d13+d3+d5+d1+d18+d7+d15+d11+d9+d38+d12+d24+d22+d16+d20+d37
7     AE2 =~ d29+d27+d30+d39+d4+d32+d35+d10+d33+d19+d25+d26+d28
8     AE3 =~ d31+d14+d21+d23+d17+d34+d8+d6
9     CD1 =~ e1+e2+e3+e5+e7+e9+e11+e13+e15+e10
10    CD2 =~ e8+e12
11    RA =~ CM
12    IE =~ IE1+IE2+IE3+IE4+IE5
13    AE =~ AE1+AE2+AE3
14    CD =~ CD1+CD2
15    c16 =~ c20+c4
16    c14 =~ c5+c3
17    e14 =~ e8+e3+e6
18    d34 =~ d36
19    c10 =~ c15
20    c7 =~ c6
21    d3 =~ d2
22    e3 =~ e6'
```

### Estimación de los parámetros del modelo modificado

La estimación de los parámetros del modelo se realiza mediante el mismo método de estimación utilizado en el modelo propuesto a priori, el método de mínimos cuadrados no ponderados (ULS). Antes de estimar los parámetros del modelo modificado debo identificar los  $p$  parámetros de este modelo. Estos parámetros son los 185 descritos en la sección 3.6.6 junto con las 12 correlaciones de los errores entre variables observadas incluidas en este nuevo modelo. En total se identifican  $p=185+12=197$ . Se debe verificar que  $p=197$  sea menor que  $c = (n)(n + 1)/2$ ,  $n$  es el número de variables observadas, en este modelo hay 78, por lo que para poder estimar los parámetros del modelo,  $p=197$  debe ser menor de  $(78)(78 + 1)/2 = 3081$ . Esta regla se cumple por lo que se procede a estimar los parámetros del modelo. Para ilustrar el modelo modificado realizo un diagrama de relaciones, que permite ver de forma gráfica las modificaciones respecto al modelo anterior junto con la estimación de los parámetros.



### Ajuste del modelo modificado

Obtengo los índices de bondad de ajuste del modelo modificado para verificar si el ajuste de los datos y el modelo es correcto.

GFI	AGFI	NFI	TLI	RMSEA	CFI	$\chi^2$
0.971	0.970	0.963	0.987	0.035	0.987	4422.351

Cuadro 3.2: Índices de ajuste del modelo modificado

Se puede ver que todos los índices han mejorado ligeramente respecto al modelo propuesto a priori. Los índices GFI, AGFI, NFI son muy próximos a 1, lo que indica que el ajuste es muy bueno. En cuanto a los índices TLI y CFI que indican que el ajuste es correcto si el valor es mayor de 0.9, vemos que esta condición se sigue cumpliendo. El RMSEA es menor de 0.05 por lo que también se considera un buen ajuste. La modificación del modelo se ha realizado para disminuir el estadístico  $\chi^2$  y encontrar un modelo mejor para este conjunto de datos, vemos que el estadístico chi-cuadrado ha disminuido levemente, aunque como sabemos esta medida no es muy robusta debido a que los datos no siguen una distribución normal multivariante. Al obtener los grados de libertad del modelo mediante la función *summary()*, 2898, se cumple la condición de que el valor del índice chi-cuadrado entre los grados de libertad del modelo sea menor de 3 ( $4422,351/2898 = 1,52$ ).

A priori este modelo modificado se puede considerar algo mejor que el modelo propuesto, ya que obtenemos unos valores mejores de los índices de ajuste. Una forma de comprobar que realmente el modelo modificado es más razonable que el propuesto a priori es mediante el índice de validación cruzada esperado (ECVI). El modelo con el que se obtenga un valor menor del ECVI será el modelo que consideremos mejor. Este índice se obtiene mediante la función *fitMeasures()*, al igual que el resto de índices. En la siguiente tabla se puede ver que el valor del índice de validación cruzada esperado es ligeramente inferior en el caso del modelo modificado, por lo que concluyo que el mejor de los dos es el modelo modificado y este es el modelo que escojo como el modelo correcto.

	ECVI
Modelo propuesto a priori	11.6
Modelo modificado	10.9

Cuadro 3.3: Índices de validación cruzada esperados

### 3.6.9. Interpretación de los resultados

He definido un modelo en el que he incluido los factores que deseo analizar junto con sus variables observadas indicadoras. Después de modificar el modelo y concluir que este es el mejor de los dos para estos datos puedo confirmar que la estructura factorial definida se ajusta bien a los datos de los que disponemos para la realización de este caso de estudio. Este buen ajuste del modelo me indica que las escalas de medida utilizadas para medir los factores latentes son válidas y confiables. Aunque haya escogido el modelo modificado como el mejor de los dos hay que recordar que puede haber un modelo mejor para estos datos.

### 3.7. Modelado de ecuaciones estructurales

Una vez he realizado el análisis factorial confirmatorio y habiendo obtenido una estructura subyacente válida y un ajuste entre los datos y el modelo correcto procedo a la realización de un modelo de ecuaciones estructurales para contrastar las hipótesis definidas en la sección 3.1, partiendo del modelo seleccionado como el mejor, es decir, del modelo modificado. En este modelo de ecuaciones estructurales se deben incluir las relaciones necesarias para contrastar las hipótesis iniciales por ello se debe ampliar el modelo propuesto añadiendo el efecto directo de los factores *Inteligencia Emocional*, *Habilidades Sociales de Prevención del Acoso Escolar* y *Capacidad de Disfrute* hacia el *Rendimiento Académico*, para poder analizar la posible influencia de estos factores hacia el *Rendimiento Académico*. En un modelo de ecuaciones estructurales se distingue entre variables exógenas y endógenas, por ello en este modelo, el *Rendimiento Académico*, es una variable latente endógena por el efecto directo que recibe del resto de factores que se incluyen en el modelo y por lo tanto su variable indicadora, *Calificación Media en Matemáticas* se considerará una variable observada endógena. El resto de las variables del modelo son exógenas porque no están determinadas por elementos internos del modelo.

La especificación de un modelo de ecuaciones estructurales requiere dos partes el modelo de medida, esto es lo que se considera el modelo modificado realizado en la sección 3.6.8, y el modelo estructural, que incluye las relaciones entre las variables latentes. Se debe incluir en la especificación del modelo la siguiente relación que indica el efecto directo entre los factores y el *Rendimiento Académico*.

```
1 #modelo estructural
2 RA - IE+AE+CD
```

La estimación de los parámetros del modelo se realiza mediante el método de mínimos cuadrados no ponderados (ULS), que ha sido el método utilizado a lo largo del caso de estudio. Primero se deben establecer el número de parámetros a estimar, estos son los 77 parámetros relativos a los errores de medida de las variables observadas exógenas, no se incluye el error de medida para la variable observada endógena porque no se considera que haya un error debido a que la variable es única. También se deben estimar las 78 cargas factoriales entre las variables latentes y observadas, las 10 cargas factoriales entre los factores y las variables en las que se dividen, los 14 errores asociados a las variables latentes, las 3 asociaciones hipotéticas entre las variables latentes exógenas y las 3 asociaciones hipotéticas entre las variables latentes exógenas y el *Rendimiento Académico*. Por último, se deben estimar las 12 correlaciones de los errores entre las variables observadas, en total el número de parámetros a estimar es 197. Este número debe ser menor que el número de varianzas/covarianzas no redundantes del modelo, es decir, menor que  $c$  siendo  $c = (n + o)(n + o + 1)/2$ , siendo  $n$  el número de variables observadas exógenas que son 77 y  $o$  el número de variables observadas endógenas que es 1, por lo que  $c = (77 + 1)(77 + 1 + 1)/2 = (78)(79)/2 = 3081$ , por lo que se cumple que  $p=197$  sea menor que  $c$ .

La representación del modelo se realiza mediante un diagrama de relaciones, en el que se pueden ver las estimaciones de los parámetros. La estimación de los parámetros es muy similar a la obtenida en el modelo modificado, el cambio más significativo se puede ver en los parámetros que indican la relación entre los factores y el *Rendimiento Académico*.

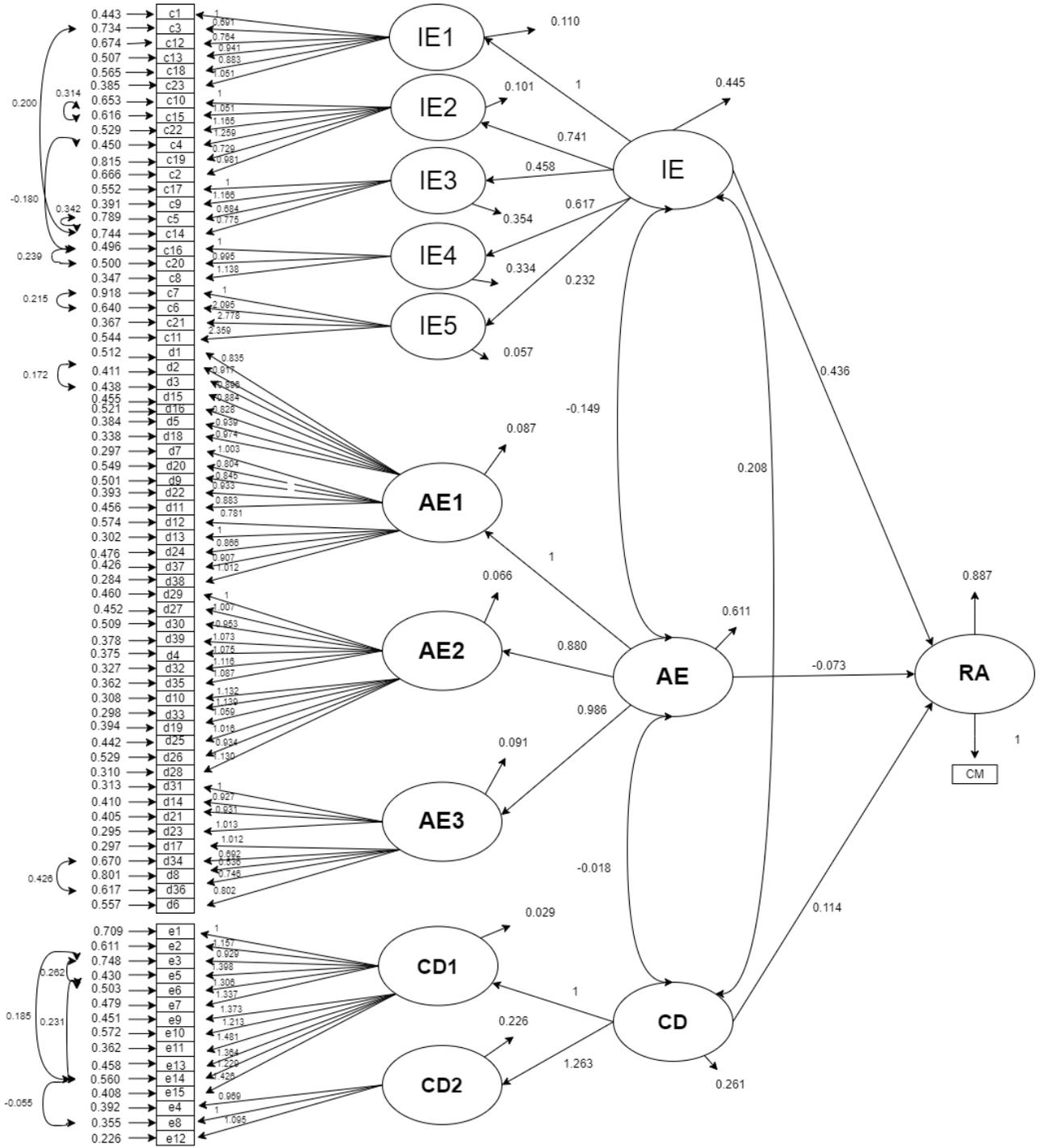


Figura 3.10: Diagrama de Relaciones del modelo de ecuaciones estructurales con la estimación de los parámetros.

La evaluación del ajuste del modelo debería ser muy similar a la obtenida en el modelo modificado aun así, compruebo los índices de bondad de ajuste para asegurarme de que no se ha producido ningún error y el ajuste entre los datos y el modelo sigue siendo correcto. Como se puede observar en la siguiente tabla los índices mantienen su valor respecto al modelo de AFC modificado. En conclusión, el ajuste de los datos y el modelo es correcto.

GFI	AGFI	NFI	TLI	RMSEA	CFI	$\chi^2$
0.971	0.970	0.963	0.987	0.035	0.987	4422.351

Cuadro 3.4: Índices de ajuste del modelo de ecuaciones estructurales

Una vez que he establecido que este modelo es correcto puedo proceder a verificar las hipótesis iniciales. La estimación de los parámetros que son relevantes para contrastar estas hipótesis se pueden ver en el diagrama 3.10 sin embargo, para facilitar su interpretación en la siguiente tabla se muestran estas estimaciones junto con sus p-valores obtenidos mediante la función *summary()*. La interpretación de estos p-valores es clave para confirmar o rechazar estas hipótesis ya que un p-valor menor de 0.05 indica que el valor de la estimación de ese parámetro es significativo y en el caso contrario indica que el valor de la estimación no es significativo. En la columna 'Estimate' se pueden ver las asociaciones hipotéticas entre las variables latentes que se muestran en el diagrama de relaciones y en la última columna vemos el p-valor que nos indica si el valor de la asociación es significativo o no.

	Estimate	Std.Err	z-value	P(> z )
<b>Regressions</b>				
RA ~IE	0.436	0.067	6.505	0.000
RA ~AE	-0.073	0.021	-3.491	0.000
RA ~CD	0.114	0.077	1.499	0.134
<b>Correlations</b>				
IE ~AE	-0.149	0.005	-30.960	0.000
IE ~CD	0.208	0.009	24.099	0.000
AE ~CD	-0.019	0.002	-10.207	0.000

Cuadro 3.5: Parámetros estimados de interés del modelo de ecuaciones estructurales

La primera hipótesis que debo verificar es si existe una relación lineal significativa entre la *Inteligencia Emocional* y las *Habilidades Sociales de Prevención del Acoso Escolar*. En la tabla se puede ver que el valor de la asociación hipotética entre estas dos variables es de -0.149, es decir, esto indica que hay una correlación negativa entre ambos factores. Esto significa que valores altos de cualquiera de los dos factores indican una mayor tendencia a encontrar valores bajos en el otro factor, y en el caso contrario ocurriría lo mismo. Aunque esto no implica que una variable cause directamente cambios en la otra. Por ejemplo, valores altos de la *Inteligencia Emocional* implicaría una mayor tendencia a encontrar valores bajos en las *Habilidades Sociales de Prevención del Acoso Escolar*. Esta tendencia sería muy leve debido a que el valor de la estimación es muy próximo a cero. La interpretación del parámetro sería la expuesta, pero hay que verificar que la correlación entre ambos factores es significativa para validar su interpretación. El p-valor es 0, es menor de 0.05 y se puede concluir que la correlación es significativa. El valor de la correlación es cercano a cero, esto implica que la relación lineal entre ambos factores es muy débil. Por ello, no puedo afirmar que exista una relación lineal significativa entre ambos factores.

La segunda hipótesis consiste en verificar si hay una relación lineal significativa entre la *Capacidad de disfrute* y las *Habilidades Sociales de Prevención del Acoso Escolar*. El valor de la asociación hipotética entre ambas variables es de -0.019, es decir, la correlación entre ambos factores es negativa. Una correlación negativa indica que hay una mayor tendencia a encontrar valores bajos de un factor si el otro presenta valores altos, en el caso contrario ocurre lo mismo. Como sabemos esto no quiere decir que un factor cause directamente cambios en el otro. Para comprobar la significancia de esta correlación, compruebo su p-valor que en este caso tiene un valor de 0, este es menor de 0.05, por lo que la correlación sí es significativa. Al igual que ocurre con la primera hipótesis la correlación es muy cercana a cero, esto quiere decir que la relación entre los factores es muy débil incluso prácticamente nula y no podemos afirmar que haya una relación lineal significativa entre ambos factores.

En cuanto a la tercera hipótesis debo verificar si existe una relación lineal significativa entre la *Inteligencia Emocional* y la *Capacidad de disfrute*. El valor de la asociación hipotética entre ellos es de 0.208, es decir, hay una correlación positiva entre ambos factores. En este caso la interpretación sería que valores altos de un factor indican una mayor tendencia a encontrar valores altos del otro factor y valores bajos indican una mayor tendencia a encontrar valores bajos del otro factor. Por ejemplo, valores bajos de la *Inteligencia Emocional* indicaría una mayor tendencia a encontrar valores bajos de la *Capacidad de disfrute*. Esta correlación es significativa ya que el p-valor obtenido es 0, que es menor de 0.05. El valor de la correlación no es muy elevado lo que sugiere que la relación lineal entre ambos factores es muy débil. Por lo que tampoco puedo afirmar que exista una relación lineal significativa entre ambos factores.

El resto de las hipótesis que debo contrastar tienen que ver con la influencia que tienen estos factores en el *Rendimiento Académico*. La *Inteligencia Emocional* provoca un efecto positivo en el *Rendimiento Académico*, es decir, que un mayor nivel de *Inteligencia Emocional* implicaría un *Rendimiento Académico* mejor, y una disminución del nivel de *Inteligencia Emocional* implicaría un *Rendimiento Académico* más bajo, ya que la asociación hipotética entre ambas variables tiene un valor de 0.436. Este valor indicaría una influencia moderada entre ambos factores. El p-valor es 0, es decir, la asociación entre ambos factores es significativa. En conclusión, puedo afirmar que la *Inteligencia Emocional* influye positivamente en el *Rendimiento Académico*, aunque esta influencia sea moderada. En el caso de la *Capacidad de Disfrute* se obtiene una estimación de la asociación hipotética con el *Rendimiento Académico* de 0.114. El p-valor de esta asociación es de 0.134 que es mayor de 0.05 por lo que no se considera esta asociación significativa y por lo tanto no hay evidencia a favor de la hipótesis inicial. Por último, las *Habilidades Sociales de Prevención del Acoso Escolar* y el *Rendimiento Académico* tienen una asociación cuyo valor es -0.073. El p-valor es 0 por lo que sí se considera esta asociación significativa, aunque muy débil debido a que el valor de la asociación es prácticamente 0. Como el valor de la asociación obtenido es negativo esto quiere decir que hay una relación inversa entre los factores, un aumento de uno de los factores quiere decir que hay una disminución en el otro y viceversa. Por ejemplo, el aumento en las *Habilidades Sociales de Prevención del Acoso Escolar* implicaría una disminución en el *Rendimiento Académico*. Esta relación inversa supone que no podemos verificar la hipótesis de que exista una influencia positiva de las *Habilidades Sociales de Prevención del Acoso Escolar* sobre el *Rendimiento Académico*.



## Capítulo 4

# Conclusiones

En el presente Trabajo de Fin de Grado se han presentado dos herramientas estadísticas muy utilizadas en el campo de las ciencias sociales, especialmente en psicología y educación. Estas son, los modelos de ecuaciones estructurales (SEM), una técnica estadística muy potente para modelar y analizar relaciones complejas entre múltiples variables, y el análisis factorial confirmatorio (AFC), un antecesor de los modelos de ecuaciones estructurales que permite evaluar la validez y fiabilidad de las escalas de medida y permite confirmar si los datos utilizados se ajustan bien a un modelo propuesto a priori. Se han expuesto los diferentes pasos para la construcción de estos modelos, desde su definición, pasando por su especificación, la identificación y estimación de parámetros, hasta la evaluación del ajuste del modelo. Se han presentado también las diferentes formas de modificar el modelo creado en el caso de que fuera necesario.

Para ilustrar este tipo de modelos se ha realizado un caso de estudio cuya finalidad es analizar las relaciones entre la *Inteligencia Emocional*, las *Habilidades Sociales de Prevención del Acoso Escolar* y la *Capacidad de Disfrute* en niños y adolescentes y si estas influyen o no en el *Rendimiento Académico*. Este tipo de factores son muy importantes en niños y adolescentes por lo que cabe esperar que presenten un efecto positivo sobre el *Rendimiento Académico*. Una buena *Inteligencia Emocional* hace que los niños presenten características como la empatía, felicidad, positividad, los resultados nos indican que el *Rendimiento Académico* se ve mejorado si la *Inteligencia Emocional* está bien desarrollada. En cuanto a la *Capacidad de Disfrute* no hay evidencia de que influya en el *Rendimiento Académico* y las *Habilidades Sociales de Prevención del Acoso Escolar* influyen en el *Rendimiento Académico*, pero de forma negativa.

En relación con los resultados obtenidos llegamos a la conclusión de que fomentar el desarrollo de la *Inteligencia Emocional* en las aulas va a ser beneficioso para los alumnos que verán una mejoría en su *Rendimiento Académico*. En cuanto a las *Habilidades Sociales de Prevención del Acoso Escolar* y la *Capacidad de Disfrute*, el resultado obtenido no es el esperado. Los resultados no indican que habilidades como la comunicación, el saber expresarse, el respeto por los demás, etc, van a producir un efecto negativo sobre *Rendimiento Académico*, y que capacidades como la alegría, la satisfacción, el bienestar, etc, simplemente no van a tener ningún efecto sobre el *Rendimiento Académico*. Desarrollar estas habilidades en las aulas con el fin de ver una mejoría en el *Rendimiento Académico* no sería adecuado.

Por qué se puede haber producido este resultado que a priori es contradictorio con lo que cabría esperar. Si bien este tipo de modelos son muy útiles presentan ciertas limitaciones. Por definición los modelos de ecuaciones estructurales son una técnica de análisis multivariante, es decir, los datos utilizados deberían seguir una distribución normal multivariante, por ello la utilización de variables de tipo cuantitativo continuo sería la mejor forma de manejar este tipo de modelos, pero esto no es posible debido a los instrumentos de medida como los cuestionarios utilizados para medir estos factores. El tipo de datos utilizados suele contener variables cuantitativas discretas, incluso en algunos casos variables categóricas que no pueden seguir una distribución normal multivariante. Aunque la aplicación de estos modelos sigue siendo posible, el uso de este tipo de variables no es el más adecuado ya que los resultados no serán tan robustos. Otra de las limitaciones de este tipo de modelos es el tamaño de la muestra, una muestra pequeña puede dar lugar a resultados poco representativos de la población, por ello un tamaño de muestra adecuado es primordial para obtener unos resultados que sean confiables. Una mayor muestra de alumnos podría dar lugar a resultados más robustos y fiables. La medición de las variables latentes es otra limitación que presentan estos modelos ya que puede suponer una gran dificultad ya que al ser variables no observadas pueden omitirse variables importantes indicadoras a la hora de realizar el modelo. Aunque se ha obtenido que las escalas de medida de los tres factores eran prácticamente perfectas puede que se hayan omitido ciertas variables que mejorarían la medición de estos factores. Por último, una de las principales limitaciones es la búsqueda del modelo óptimo. Un modelo puede ajustarse bien a unos datos, pero no ser el mejor modelo, por eso es interesante modificar el modelo o crear diferentes modelos y contrastarlos entre ellos, esto hace que la búsqueda del modelo óptimo sea un proceso costoso. Debido a estas limitaciones que se presentan en este tipo de modelos y en el modelo de este trabajo en particular, mejorar y encontrar el modelo óptimo son algunas de las propuestas descritas sería parte de un trabajo futuro que se podría realizar.

Para la realización de este caso de estudio se ha utilizado el software R, pero este también presenta ciertas limitaciones, las herramientas utilizadas para la realización de modelos de ecuaciones estructurales están preparados para modelos sencillos. Estas herramientas presentan problemas para modelos con más complejidad, por ejemplo, paquetes como *semPlot* no pueden producir diagramas de relaciones adecuados con modelos complejos.

En cuanto a los modelos de análisis factorial confirmatorio y los modelos de ecuaciones estructurales en general todavía hay mucho trabajo que se puede realizar para su extensión y mejora. Entre los trabajos futuros que se pueden pensar se encuentra desarrollo de este tipo de modelos junto a otras herramientas como la minería de datos o el aprendizaje automático para crear modelos más precisos y sencillos. Se podrían explorar nuevos métodos de estimación de parámetros que fueran más robustos. Y su uso podría extenderse fuera del campo de las ciencias sociales y utilizar este tipo de modelos en otros campos como el de las ciencias de la salud, en áreas como la biología o la medicina. Por último, una de las limitaciones mencionadas es la forma en la que se miden las variables latentes. En el campo de las ciencias sociales donde hay una gran cantidad de conceptos abstractos que se miden a través de cuestionarios, escalas, etc, las variables de las que vamos a disponer serán en la mayor parte de los casos cuantitativas discretas, por lo que utilizar modelos de este tipo no va a ser adecuado aunque sea posible su uso. Por ello, considero que se debería centrar la investigación en la búsqueda de modelos más adecuados para los estudios que requieran del uso de variables que sean de tipo cuantitativo discreto o categóricas.

# Bibliografía

- [1] Gaudenz Alder and David Benson. 'draw.io'. <https://app.diagrams.net/>, 2000.
- [2] Peter M. Bentler. Comparative fit indexes in structural models. *Psychological bulletin*, 107(2):238, 1990.
- [3] Peter M. Bentler and Douglas G. Bonett. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, 88(3):588, 1980.
- [4] Michael W. Browne and Robert Cudeck. Single sample cross-validation indices for covariance structures. *Multivariate behavioral research*, 24(4):445–455, 1989.
- [5] Gamal Cerda, Carlos Pérez, Paz Elipe, José A. Casas, and Rosario Del Rey. Convivencia escolar y su relación con el rendimiento académico en alumnado de educación primaria. *Revista de psicodidáctica*, 24(1):46–52, 2019.
- [6] Ruben D. Chiriboga Zambrano and Jenny E. Franco Muñoz. Validación de un test de inteligencia emocional en niños de diez años de edad. *Med. fam.(Caracas)*, pages 13–23, 2001.
- [7] Charles B. Crawford and George A. Ferguson. A general rotation criterion and its use in orthogonal rotation. *Psychometrika*, 35(3):321–332, 1970.
- [8] Lee J. Cronbach. Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3):297–334, 1951.
- [9] Sergio Dominguez-Lara and César Merino-Soto. Evaluación de las malas especificaciones en modelos de ecuaciones estructurales. *Revista Argentina de Ciencias del Comportamiento*, 10(2):19–24, 2018.
- [10] Sacha Epskamp. *semPlot: Path Diagrams and Visual Analysis of Various SEM Packages*, 2022. R package version 1.1.6.
- [11] Pablo Fernández-Berrocal and Desiree Ruiz Aranda. La inteligencia emocional en la educación. *Electronic Journal of Research in Education Psychology*, 6(15):421–436, 2008.
- [12] Carlos Alfonso López García and Noemí Baniandrés García. El principio de parsimonia en la ciencia cognitiva actual: Riesgos y soluciones. *Cognitiva*, 7(2):28–30, 2013.
- [13] Daniel Goleman. *La inteligencia emocional*. J. Vergara Editor, 1995.
- [14] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

- [15] Lawrence R. James, Stanley A. Mulaik, and Jeanne M. Brett. *Causal analysis: Assumptions, models, and data*. Beverly Hills (Calif.): Sage, 1983., 1983.
- [16] Ian T. Jolliffe. *Principal component analysis for special types of data*. Springer, 2002.
- [17] Karl G. Jöreskog. Some contributions to maximum likelihood factor analysis. *psychometrika*, 32(4):443–482, 1967.
- [18] Karl G. Jöreskog. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2):183–202, 1969.
- [19] Karl G. Jöreskog and Dag Sörbom. *Analysis of linear structural relationships by maximum likelihood and least squares methods*. University of Uppsala: Uppsala, 1981.
- [20] Henry F. Kaiser. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200, 1958.
- [21] David Kaplan. Evaluating and modifying covariance structure models: A review and recommendation. *Multivariate Behavioral Research*, 25(2):137–155, 1990.
- [22] Selcuk Korkmaz, Dincer Goksuluk, and Gokmen Zararsiz. Mvn: An r package for assessing multivariate normality. *The R Journal*, 6(2):151–162, 2014.
- [23] D. N. Lawley. Vi.—the estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh*, 60(1):64–82, 1940.
- [24] Daniel Lüdtke, Mattan S. Ben-Shachar, Indrajeet Patil, and Dominique Makowski. Extracting, computing and exploring the parameters of statistical models using R. *Journal of Open Source Software*, 5(53):2445, 2020.
- [25] Ana M. Magaz, Paloma Chorot, Miguel A. Santed, Rosa M. Valiente, and Bonifacio Sandín. Evaluación del bullying como victimización: Estructura, fiabilidad y validez del cuestionario de acoso entre iguales (cai). *Revista de psicopatología y psicología clínica*, 21(2):77–95, 2016.
- [26] Kanti V. Mardia. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3):519–530, 1970.
- [27] Irini Mavrou. Análisis factorial exploratorio: cuestiones conceptuales y metodológicas. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, (19):71–80, 2015.
- [28] Ralph O. Mueller. *Basic principles of structural equation modeling : an introduction to LISREL and EQS / Ralph O. Mueller*. Springer Texts in Statistics. Springer, New York, New York, 1st ed. 1996. edition, 1996.
- [29] Stanley A. Mulaik, Larry R. James, Judith Van Alstine, Nathan Bennett, Sherri Lind, and C. Dean Stilwell. Evaluation of goodness-of-fit indices for structural equation models. *Psychological bulletin*, 105(3):430, 1989.
- [30] Jack O. Neuhaus and Charles Wrigley. The quartimax method: An analytic approach to orthogonal simple structure 1. *British Journal of Statistical Psychology*, 7(2):81–91, 1954.

- [31] Dan Olweus. Hakkyclinger och översittare. *Forskning om skolmobbing. Estocolmo: Almqvist y Wicksell*, 1973.
- [32] Ferran Padrós-Blázquez, María Patricia Martínez-Medina, Joan Guardia-Olmos, and Ariel Graff-Guerrero. Propiedades psicométricas de la escala de gaudibilidad (moduladores de disfrute) para niños y adolescentes (egna). *Anales de Psicología/Annals of Psychology*, 37(1):69–76, 2021.
- [33] Edgardo R. Pérez and Leonardo Adrián Medrano. Análisis factorial exploratorio: bases conceptuales y metodológicas. *Revista Argentina de Ciencias del Comportamiento (RACC)*, 2(1):58–66, 2010.
- [34] Jaime C. Porras. Comparación de pruebas de normalidad multivariada. In *Anales Científicos*, volume 77, pages 141–146. Universidad Nacional Agraria La Molina, 2016.
- [35] Théodule-Armand Ribot. The psychology of the emotions. *Londres: W. Scott Pub. Co.*, 1897.
- [36] Yves Rosseel. lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2):1–36, 2012.
- [37] Willem E. Saris, Albert Satorra, and Dag Sörbom. The detection and correction of specification errors in structural equation models. *Sociological methodology*, pages 105–129, 1987.
- [38] David R. Saunders. Integrating the implementation of quartimax, varimax, oblimax, and related rotational methods. *Psychological Reports*, 1962.
- [39] Noora Shrestha. Factor analysis as a tool for survey analysis. *American Journal of Applied Mathematics and Statistics*, 9(1):4–11, 2021.
- [40] Tracy D. Smith and Bradley F. McMillan. A primer of model fit indices in structural equation modeling. 2001.
- [41] Charles Spearman. General intelligence, objectively determined and measured. *American Journal of Psychology*, 15(2):201–293, 1904.
- [42] Markus D. Steiner and Silvia Grieder. Efatools: An r package with fast and flexible implementations of exploratory factor analysis tools. *Journal of Open Source Software*, 5(53):2521, 2020.
- [43] Ledyard R. Tucker and Charles Lewis. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1):1–10, 1973.
- [44] John Verzani. *Getting started with RStudio*. 'O'Reilly Media, Inc.', 2011.
- [45] David Wechsler et al. Wais-iii: test de inteligencia para adultos de wechsler. 2002.
- [46] Hadley Wickham and Jennifer Bryan. *readxl: Read Excel Files*, 2023. R package version 1.4.2.
- [47] Hadley Wickham and Dana Seidel. *scales: Scale Functions for Visualization*, 2022. R package version 1.2.1.

- [48] William Revelle. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois, 2023. R package version 2.3.3.

# Anexo A

## Covarianza: Definición y propiedades

Se define la covarianza [28] entre dos variables  $X$  e  $Y$  como el valor esperado de los productos de las desviaciones de las variables respecto a sus medias, se puede expresar de la siguiente forma,

$$COV(X, Y) = E([X - E(X)][Y - E(Y)]) = \frac{\sum_{k=1}^N (X_k - \mu_x)(Y_k - \mu_y)}{N} \quad (4.1)$$

donde  $\mu_x$  es la media de la variable  $X$  y  $\mu_y$  es la media de la variable  $Y$ .

El calculo de la covarianza tiene algunas propiedades que hay que tener en cuenta,

1.  $Cov(X, c) = 0$ , siendo  $c$  una constante.
2.  $Cov(X, X) = Var(X)$ , la covarianza de una variable consigo misma es igual a la varianza de esa variable
3.  $Cov(X, Y) = Cov(Y, X)$ , la covarianza entre dos pares de variables es igual independientemente de su orden.
4.  $Cov(b \cdot X, c \cdot Y) = c \cdot b \cdot Cov(X, Y)$  la covarianza de dos variables multiplicadas por dos constantes es igual a la covarianza de esas dos variables por la multiplicación de esas constantes.
5.  $Cov(b+X, c+Y) = Cov(X, Y)$ , la covarianza no se ve afectada por la suma de constantes a las variables.
6.  $Cov(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$ , la covarianza es igual a la esperanza de la multiplicación de las dos variables menos el producto de la esperanza de cada variable.

# Anexo B

## Cuestionario de Desarrollo Escolar

Este cuestionario ha sido desarrollado por Juan Antonio Valdivieso (2022). A continuación, se presentan una serie de cuestiones que ayudan a evaluar el desarrollo escolar y personal del alumnado. Para facilitar su interpretación en *Rstudio* se muestra entre paréntesis la equivalencia de cada elemento de interés del cuestionario para el estudio realizado en este software.

ÍTEM	RESPUESTA
ID del alumno/a	
Localidad	
Denominación del Centro Educativo	
Genero	<input type="radio"/> Masculino <input type="radio"/> Femenino <input type="radio"/> Prefiero no decirlo
Edad	
Calificación en matemáticas	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9 <input type="radio"/> 10
CIEMP. 01. Soy una persona alegre ( $c_1$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
CIEMP. 02. Estoy agradecido con varias personas, por lo que hacen por mí. ( $c_2$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
CIEMP. 03. La mayor parte de los días me siento en paz. ( $c_3$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
CIEMP. 04. Valoro cuando los demás me ayudan. ( $c_4$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
CIEMP. 05. Soy bastante tranquilo. ( $c_5$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
CIEMP. 06. Me pongo muy mal si veo que alguien se lastima. ( $c_6$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
CIEMP. 07. Si veo llorar a un niño/a me dan ganas de llorar a mí también. ( $c_7$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
CIEMP. 08. Me quiero mucho a mí mismo. ( $c_8$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
CIEMP. 09. Soluciono mis problemas con mucha tranquilidad. ( $c_9$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3
CIEMP. 10. Me gusta devolver favores. ( $c_{10}$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3

CIEMP. 11. Cuando alguien está solo y aburrido me dan ganas de acercarme y jugar con él. ( $c_{11}$ )	○ 1 ○ 2 ○ 3
CIEMP. 12. Me divierto mucho con las cosas que hago. ( $c_{12}$ )	○ 1 ○ 2 ○ 3
CIEMP. 13. Soy muy feliz. ( $c_{13}$ )	○ 1 ○ 2 ○ 3
CIEMP. 14. Casi siempre estoy muy relajado. ( $c_{14}$ )	○ 1 ○ 2 ○ 3
CIEMP. 15. Siempre que puedo, devuelvo los favores que recibo. ( $c_{15}$ )	○ 1 ○ 2 ○ 3
CIEMP. 16. Siento que soy muy valioso. ( $c_{16}$ )	○ 1 ○ 2 ○ 3
CIEMP. 17. Aunque tenga problemas, igual mantengo la calma. ( $c_{17}$ )	○ 1 ○ 2 ○ 3
CIEMP. 18. Casi siempre me lo paso bien. ( $c_{18}$ )	○ 1 ○ 2 ○ 3
CIEMP. 19. Me quedo tranquilo aunque no pueda hacer lo que me gusta. ( $c_{19}$ )	○ 1 ○ 2 ○ 3
CIEMP. 20. Siento que soy importante. ( $c_{20}$ )	○ 1 ○ 2 ○ 3
CIEMP. 21. Si alguien está llorando me dan ganas de abrazarlo o consolarlo. ( $c_{21}$ )	○ 1 ○ 2 ○ 3
CIEMP. 22. Me gusta agradecerle a la gente. ( $c_{22}$ )	○ 1 ○ 2 ○ 3
CIEMP. 23. Casi siempre estoy contento/a. ( $c_{23}$ )	○ 1 ○ 2 ○ 3
CAI.01. Me pegan collejas. ( $d_1$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.02. Me pegan puñetazos. ( $d_2$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.03. Me pegan patadas. ( $d_3$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.04. Me pegan con objetos (ej. tijeras, piedras...). ( $d_4$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.05. Me insultan. ( $d_5$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.06. Me impiden jugar con ellos. ( $d_6$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.07. Me amenazan con que me pegan. ( $d_7$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.08. Me dejan de hablar. ( $d_8$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.09. Me llaman con un nombre que no me gusta ('cuatro ojos', 'bola', 'enano', 'palillo', etc.). ( $d_9$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

CAI.10. Me amenazan con objetos (ej.: tijeras, piedras, etc.). ( $d_{10}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.11. Me mandan callar. ( $d_{11}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.12. Me acusan de cosas que no he dicho o hecho. ( $d_{12}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.13. Me empujan. ( $d_{13}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.14. Les dicen a otros que no estén o que no hablen conmigo. ( $d_{14}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.15. Me dicen que se van a 'chivar de mi'. ( $d_{15}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.16. Me esconden cosas. ( $d_{16}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.17. Les prohíben a otros que jueguen conmigo. ( $d_{17}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.18. Me gritan. ( $d_{18}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.19. Me rompen cosas. ( $d_{19}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.20. Me roban cosas. ( $d_{20}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.21. Me impiden jugar con otros niños. ( $d_{21}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.22. Se ríen de mí cuando me equivoco. ( $d_{22}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.23. Me ocultan a dónde van. ( $d_{23}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.24. Me ponen la zancadilla. ( $d_{24}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.25. Me pellizcan. ( $d_{25}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.26. Me tiran del pelo. ( $d_{26}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.27. Me envían mensajes al móvil o correos electrónicos para insultarme o amenazarme. ( $d_{27}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.28. Cuando hablo con otros niños/as, se meten conmigo. ( $d_{28}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.29. Me hacen fotos o me graban con el móvil mientras se meten conmigo. ( $d_{29}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
CAI.30. Se burlan colgando fotos o vídeos sobre mí en internet. ( $d_{30}$ )	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

CAI.31. Me mienten para que no vaya con ellos. ( $d_{31}$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
CAI.32. Me tiran cosas (por ejemplo, pelotas de papel, cosas de clase, piedras...). ( $d_{32}$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
CAI.33. Me amenazan para que diga cosas que no quiero. ( $d_{33}$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
CAI.34. No quieren hacer trabajos conmigo. ( $d_{34}$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
CAI.35. Me amenazan con contar cosas de mí o de mi familia. ( $d_{35}$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
CAI.36. No se quieren sentar conmigo. ( $d_{36}$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
CAI.37. Hacen cosas para molestarte (se ponen en medio, no me dejan pasar, me despeinan, cada vez que pasan por mi lado me tocan, etc.). ( $d_{37}$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
CAI.38. Me hacen gestos para meterse conmigo o asustarme. ( $d_{38}$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
CAI.39. Me amenazan para que haga cosas que no quiero (darles dinero, dejarles copiar mis deberes, hacer bromas pesadas y otras cosas). ( $d_{39}$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
EGNA.01. Disfruto mucho cuando estoy con algunos de mis compañeros/as. ( $e_1$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
EGNA.02. Creo que puedo disfrutar incluso cuando estoy con gente aburrida. ( $e_2$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
EGNA.03. Fácilmente me río. ( $e_3$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
EGNA.04. Soy una persona muy imaginativa. ( $e_4$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
EGNA.05. Me intereso fácilmente en la mayoría de las cosas que hago. ( $e_5$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
EGNA.06. A lo largo del día hay muchas cosas que me hacen reír. ( $e_6$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
EGNA.07. Casi siempre que estoy con los/las compañeros/as del colegio/instituto me divierto ( $e_7$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7
EGNA.08. Tengo mucha imaginación. ( $e_8$ )	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7

EGNA.09. Creo que puedo parármelo bien incluso cuando estoy en un lugar aparentemente aburrido. ( $e_9$ )       1  2  3  4  5  6  7

---

EGNA.10. Algunas veces me intereso por algo con mucha intensidad. ( $e_{10}$ )       1  2  3  4  5  6  7

---

EGNA.11. Me lo paso bien con la mayoría de los/las compañeros/as. ( $e_{11}$ )       1  2  3  4  5  6  7

---

EGNA.12. Si me lo propongo me imagino historias con facilidad. ( $e_{12}$ )       1  2  3  4  5  6  7

---

EGNA.13. Me intereso frecuentemente por varios temas. ( $e_{13}$ )       1  2  3  4  5  6  7

---

EGNA.14. Acostumbro a reír varias veces al día. ( $e_{14}$ )       1  2  3  4  5  6  7

---

EGNA.15. Si me lo propongo soy capaz de disfrutar de una situación aparentemente aburrida. ( $e_{15}$ )       1  2  3  4  5  6  7

---

# Anexo C

## Código empleado

```
1 #Librería readxl, lectura de los datos
2 library(readxl)
3 #lectura y visualización de datos#
4 CDESC_PS_2_respuestas <- read_excel("C:/Users/nataf/OneDrive/Desktop/TFG ESTADISTICA/
  DATOS/CDESC-PS 2 (respuestas).xlsx")
5 datos=CDESC_PS_2_respuestas
6
7 #Preparación de los datos#
8 #Selección de las variables de interes
9 datos=datos[,7:84]
10 #Búsqueda de valores ausentes
11 any(is.na(datos))
12
13 #Validación de las escalas de medida#
14 #variables observadas indicadoras de cada variable latente
15 var_ra=datos[,1]
16 var_ie=datos[,2:24]
17 var_ae=datos[,25:63]
18 var_cd=datos[,64:78]
19 library(psych)
20 #Alpha de Cronbach para la escala de medida de la Inteligencia Emocional
21 alpha_ie=alpha(var_ie)
22 alpha_ie
23 alpha_ae=alpha(var_ae)
24 alpha_ae
25 alpha_cd=alpha(var_cd)
26 alpha_cd
27
28 #Exploración de los datos#
29 #Test de normalidad multivariante
30 library(MVN)
31 testmvn <- mvn(data = datos, mvnTest = "mardia")
32 testmvn$multivariateNormality
33 testn <- mvn(data = datos, univariateTest = "SW", desc = TRUE)
34 testn$univariateNormality
35 #Estandarización de los datos
36 library(scales)
37 datos=scale(datos)
38 datos=as.data.frame(datos)
39
40 #Evaluación de la idoneidad de los datos#
41 #Librería para realizar el test de Kaiser-Meyer-Olkin
42 library(EFAtools)
43 KMO(datos)
44
45 #Determinación y extracción de factores#
46
47 #1.Inteligencia Emocional
48 #Gráfico de sedimentación
49 scree(var_ie, main="Gráfico de sedimentación del factor inteligencia emocional")
50 #Test de Kaiser-Guttman
51 KGC(var_ie)
52 #Extracción de factores
53 #Librería que me permite realizar la extracción de factores
54 library(parameters)
55 ie <- psych::fa(var_ie, nfactor = 5,rotate = "varimax", fm="uls") %>%
56   model_parameters(sort = TRUE, threshold = "max")
57 ie
58
59
```

```

60 #2.Habilidades Sociales de Prevención del Acoso Escolar
61 #Gráfico de sedimentación
62 scree(var_ae, main="Gráfico de sedimentación del factor
63 Habilidades Sociales de Prevención del Acoso Escolar ")
64 #Test de Kaiser-Guttman
65 KGC(var_ae)
66 #Extracción de factores
67 ae <- psych::fa(var_ae, nfactor = 3,rotate = "varimax", fm="uls") %>%
68   model_parameters(sort = TRUE, threshold = "max")
69 ae
70
71 #3.Capacidad de Disfrute
72 #Gráfico de sedimentación
73 scree(var_cd, main="Gráfico de sedimentación del factor capacidad de disfrute")
74 #Test de Kaiser-Guttman
75 KGC(var_cd)
76 #Extracción de factores
77 cd <- psych::fa(var_cd, nfactor = 2,rotate = "varimax", fm="uls") %>%
78   model_parameters(sort = TRUE, threshold = "max")
79 cd
80
81
82 #ANÁLISIS FACTORIAL CONFIRMATORIO#
83 #Especificación del modelo
84 library(lavaan)
85 modelo_afc<-'IE1 =~ c1+c3+c12+c13+c18+c23
86           IE2=~c10+c15+c22+c4+c19+c2
87           IE3=~c17+c9+c5+c14
88           IE4=~c16+c20+c8
89           IE5=~c7+c6+c21+c11
90           AE1 =~ d13+d3+d5+d1+d18+d7+d2+d15+d11+d9+d38+d12+d24+d22+d16+d20+d37
91           AE2 =~ d29+d27+d30+d39+d4+d32+d35+d10+d33+d19+d25+d26+d28
92           AE3 =~ d31+d14+d21+d23+d17+d34+d8+d36+d6
93           CD1=~ e1+e2+e3+e5+e6+e7+e9+e11+e13+e14+e15+e10
94           CD2=~e8+e4+e12
95           RA =~ CM
96           IE=~IE1+IE2+IE3+IE4+IE5
97           AE=~AE1+AE2+AE3
98           CD=~CD1+CD2
99 '
100
101 #Ajuste del modelo
102 fit_afc=cfa(model=modelo_afc, data=datos,estimator='ULS')
103 #Índices de bondad de ajuste
104 fitMeasures( fit_afc,c( 'gfi','agfi', 'nfi','tli','rmsea','cfi','chisq' ) )
105 #Resumen
106 summary(fit_afc)
107 #Estimación de los parámetros
108 coef(fit_afc)
109 #Índices de modificación
110 m=as.data.frame(modindices(fit_afc, sort = TRUE, maximum.number = 400))
111 m2=m[m$op=="~",]
112
113 #Modelo modificado
114 modelo_afc_mod <-'IE1 =~ c1+c3+c12+c13+c18+c23
115           IE2=~c10+c15+c22+c4+c19+c2
116           IE3=~c17+c9+c5+c14
117           IE4=~c16+c20+c8
118           IE5=~c7+c6+c21+c11
119           AE1 =~ d13+d3+d5+d1+d18+d7+d2+d15+d11+d9+d38+d12+d24+d22+d16+d20+d37
120           AE2 =~ d29+d27+d30+d39+d4+d32+d35+d10+d33+d19+d25+d26+d28
121           AE3 =~ d31+d14+d21+d23+d17+d34+d8+d36+d6
122           CD1=~ e1+e2+e3+e5+e6+e7+e9+e11+e13+e14+e15+e10
123           CD2=~e8+e4+e12
124           RA =~ CM
125           IE=~IE1+IE2+IE3+IE4+IE5
126           AE=~AE1+AE2+AE3
127           CD=~CD1+CD2
128           #correlaciones de los errores
129           c16~~c20+c4

```

```

130         c14~~c5+c3
131         e14~~e8+e3+e6
132         d34~~d36
133         c10~~c15
134         c7~~c6
135         d3~~d2
136         e3~~e6
137     '
138
139 #Ajuste del modelo modificado
140 fit_afc_mod <- cfa(model=modelo_afc_mod, data=datos, estimator='ULS')
141 #Índices de ajuste del modelo modificado
142 fitMeasures( fit_afc_mod, c( 'gfi', 'agfi', 'nfi', 'tli', 'rmsea', 'cfi', 'chisq' ) )
143 #Resumen del modelo modificado
144 summary(fit_afc_mod)
145 #Estimación de los parámetros del modelo modificado
146 coef(fit_afc_mod)
147
148 #Comparación de modelos: índice de validación cruzada esperado
149 ecvi_mod_propuesto=fitmeasures(fit_afc, c('ecvi'))
150 ecvi_mod_propuesto
151 ecvi_mod_modificado=fitmeasures(fit_afc_mod, c('ecvi'))
152 ecvi_mod_modificado
153
154 #MODELO DE ECUACIONES ESTRUCTURALES#
155 modelo_sem<- '#modelo de medida
156         IE1 =~ c1+c3+c12+c13+c18+c23
157         IE2 =~ c10+c15+c22+c4+c19+c2
158         IE3 =~ c17+c9+c5+c14
159         IE4 =~ c16+c20+c8
160         IE5 =~ c7+c6+c21+c11
161         AE1 =~ d13+d3+d5+d1+d18+d7+d2+d15+d11+d9+d38+d12+d24+d22+d16+d20+d37
162         AE2 =~ d29+d27+d30+d39+d4+d32+d35+d10+d33+d19+d25+d26+d28
163         AE3 =~ d31+d14+d21+d23+d17+d34+d8+d36+d6
164         CD1 =~ e1+e2+e3+e5+e6+e7+e9+e11+e13+e14+e15+e10
165         CD2 =~ e8+e4+e12
166         RA =~ CM
167         IE =~ IE1+IE2+IE3+IE4+IE5
168         AE =~ AE1+AE2+AE3
169         CD =~ CD1+CD2
170         #correlaciones de los errores
171         c16~~c20+c4
172         c14~~c5+c3
173         e14~~e8+e3+e6
174         d34~~d36
175         c10~~c15
176         c7~~c6
177         d3~~d2
178         e3~~e6
179         #modelo estructural
180         RA~IE+AE+CD
181     '
182
183 #Ajuste del modelo
184 fit_sem=sem(model=modelo_sem, data=datos, estimator='ULS', check.gradient = FALSE)
185 #Índices de bondad de ajuste
186 fitMeasures( fit_sem, c( 'gfi', 'agfi', 'nfi', 'tli', 'rmsea', 'cfi', 'chisq' ) )
187 #Resumen del modelo
188 summary(fit_sem)
189 #Estimación de los parámetros
190 coef(fit_sem)

```