

Universidad de Granada
Departamento de Estadística e I.O



SELECCIÓN DE MODELOS MEDIANTE CRITERIOS DE
INFORMACIÓN EN ANÁLISIS FACTORIAL. ASPECTOS
TEÓRICOS Y COMPUTACIONALES

Tesis Doctoral

Francisco Félix Caballero Díaz
Granada, 2011

Editor: Editorial de la Universidad de Granada
Autor: Francisco Félix Caballero Díaz
D.L.: GR 4504-2011
ISBN: 978-84-694-5742-9

Universidad de Granada
Departamento de Estadística e I.O



SELECCIÓN DE MODELOS MEDIANTE CRITERIOS DE
INFORMACIÓN EN ANÁLISIS FACTORIAL. ASPECTOS
TEÓRICOS Y COMPUTACIONALES

Tesis Doctoral

Francisco Félix Caballero Díaz
Granada, 2011

Francisco Félix Caballero Díaz

**SELECCIÓN DE MODELOS MEDIANTE CRITERIOS DE
INFORMACIÓN EN ANÁLISIS FACTORIAL. ASPECTOS
TEÓRICOS Y COMPUTACIONALES**

**Memoria de Tesis presentada por
Francisco Félix Caballero Díaz
para optar al grado de Doctor
por la Universidad de Granada.**



Fdo. Francisco Félix Caballero Díaz

Vº Bº

Director de Tesis



Fdo. Dr. D. Ramón Gutiérrez Sánchez

Quiero añadir estas líneas para expresar mi agradecimiento a una serie de personas que, de una u otra manera, han sido decisivas en la realización de esta Tesis. Sin ellas, todo esto habría resultado más difícil, o, simplemente, no habría existido.

A mi director, Ramón Gutiérrez Sánchez, por darme la oportunidad de realizar esta Tesis, y por aportarme la confianza y seguridad necesarias para llevarla a cabo. Cuando he necesitado algo, ha estado siempre disponible, al igual que Ramón Gutiérrez Jáimez. A ambos, por los ánimos recibidos, sugerencias aportadas y por su dirección en el trabajo Fin de Máster de este programa de Doctorado.

A mi familia. A mis padres, Félix y Casta. A mi hermano Carlos. Con ellos he vivido el día a día y nunca he echado en falta su apoyo. Si he llegado hasta aquí, es gracias a ellos.

A todos los profesores que he tenido. Gracias especiales a los Departamentos de Estadística e Investigación Operativa de las Universidades de Málaga y Granada. Gracias a César Rodríguez Ortiz. Su Proyecto de Innovación Educativa me permitió conocer mejor el mundo universitario tras la Licenciatura y contactar con gente maravillosa con la que he tenido el gusto de trabajar después.

A todas las personas con las que en algún momento he trabajado, por todo lo que me han permitido aprender. A Maite Anarte, a Teresa Rivas, al Servicio de Endocrinología y Nutrición del Hospital Carlos Haya de Málaga... A Mónica Carreira y Alberto Machado (mucho suerte en sus inminentes Tesis). A Rosa Rondán. Gracias por todas las cosas que hemos compartido juntos. No concibo hacer un resumen de lo vivido estos últimos años sin acordarme de ellos.

Al Departamento de Matemática Aplicada de la Universidad de Málaga. Siempre recordaré el curso 2009/2010, dando clase por primera vez. Por supuesto, a mis compañeros de la Licenciatura en Matemáticas. 5 años son muy largos y suficientes para guardarles cariño. Gracias por estar ahí, especialmente a Antonio Barrera, Karen Delgado, Andrés Furones, Carlos González, Nacho Pallarés. Con Antonio he tenido el gusto de trabajar posteriormente. Es un orgullo para mí tener como amigo a un tipo extraordinario a nivel humano. Le agradezco especialmente su ayuda con el manejo de LaTeX en la realización de esta Tesis.

A los que no he mencionado, pero que han sido igual de importantes, porque me han permitido desconectar del mundo académico cuando lo he necesitado. A Pablo, Carlos A., Mari Carmen, María del Mar, David, José Ángel, Alberto. Echo la vista atrás y me acuerdo de Juanma, Rubén, Carlos M., Marco. El espacio no es ilimitado y puede que me olvide de alguno. Dese por nombrada la gente con la que he compartido risas y buenos momentos.

Finalmente, y de nuevo, gracias a Maite Anarte (investigadora principal del Proyecto SEJ2007-63786, del que he tenido el placer de formar parte) y a la Unidad de Diabetes del Servicio de Endocrinología y Nutrición del Hospital Carlos Haya de Málaga (gracias, Marisol), por permitirme hacer uso de los datos (enorme la labor de Rosa en la recogida) empleados en uno de los ejemplos que ilustran uno de los capítulos de esta Tesis.

Sin vosotros, esto no sería posible. Gracias a todos.

Índice

| | |
|---|-----------|
| Introducción al contexto en que se desarrolla esta Tesis | 7 |
| Objetivos | 11 |
| Estructura de la Tesis | 13 |
| Futuras líneas de investigación | 15 |
| 1. Análisis Factorial | 17 |
| 1.1. Métodos de extracción de factores | 25 |
| 1.2. Selección del modelo adecuado en análisis factorial | 27 |
| 1.3. Contrastes de hipótesis en análisis factorial | 29 |
| 1.3.1. Test asintótico de Bartlett sobre la adecuación de un modelo factorial orto- gonal con m factores comunes | 29 |
| 1.4. Rotación de factores | 35 |
| 1.5. Relación entre análisis factorial y análisis de componentes | 39 |
| 1.5.1. Diferencias entre análisis factorial y análisis de componentes principales | 42 |
| 2. Criterios basados en la información estadística | 45 |
| 2.1. Criterio de Información de Akaike | 49 |
| 2.1.1. Expresión de Akaike para el cálculo del AIC en Análisis Factorial | 53 |
| 2.1.2. Selección de modelos utilizando los pesos de Akaike | 55 |
| 2.2. Criterio de Información de Schwarz | 56 |

| | | |
|-----------|--|-----------|
| 2.2.1. | Una expresión alternativa para el <i>AIC</i> y el <i>BIC</i> | 59 |
| 2.2.2. | Algunas diferencias a considerar entre el <i>AIC</i> y el <i>BIC</i> | 59 |
| 2.2.3. | Extensión de los pesos de Akaike al caso de la selección de modelos mediante el <i>BIC</i> | 61 |
| 2.2.4. | Pesos de Akaike vistos como probabilidades a posteriori del modelo | 63 |
| 2.3. | Criterio de Información de Hannan y Quinn | 63 |
| 2.4. | Criterio de Información de Bozdogan | 64 |
| 2.5. | Otros criterios de información | 66 |
| 2.5.1. | Generalización de los pesos de Akaike a otros criterios de información | 70 |
| 2.6. | Algunas comparaciones efectuadas entre criterios basados en la información estadística | 72 |
| 3. | Selección del modelo factorial adecuado mediante criterios basados en la información estadística | 75 |
| 3.1. | Selección del modelo factorial adecuado mediante criterios expresados en la forma $-2 \ln \mathcal{L}(\theta_j) + k_j(A(n) + C)$ | 77 |
| 3.2. | Relación entre los criterios <i>AIC</i> , <i>BIC</i> y <i>CAIC</i> | 78 |
| 3.2.1. | Ejemplo | 81 |
| 3.3. | Relación entre el criterio <i>RIC</i> de Rissanen y el <i>BIC*</i> de Yang y Yang | 85 |
| 3.4. | Función para el cálculo mediante el programa <i>R</i> de los criterios de información referenciados en este estudio | 86 |
| 3.5. | Relación entre los pesos asociados a <i>BIC</i> y <i>CAIC</i> | 89 |
| 3.5.1. | Función en <i>R</i> para el cálculo de los pesos asociados a los criterios de información referenciados en este estudio | 93 |
| 4. | Uso de criterios de información para la selección del modelo factorial adecuado en ejemplos clásicos de la literatura | 97 |
| 4.1. | Ejemplo de “ <i>stock-prices</i> ” de Johnson y Wichern (1988) | 98 |
| 4.2. | Ejemplo de los 24 tests psicológicos (Harman, 1976) | 104 |
| 4.3. | Ejemplo de las puntuaciones en Decatlón | 111 |

| | |
|--|------------|
| 4.3.1. Análisis factorial de los datos de Decatlón correspondientes al período 2005 - 2009 | 117 |
| 5. Aplicación de la metodología basada en criterios de información para el estudio de la dimensionalidad de un cuestionario | 131 |
| 5.1. Validez factorial del $FH - 15$ | 134 |
| 5.2. Análisis del rendimiento de criterios de información en el estudio de la dimensionali- dad del $FH - 15$ | 138 |
| Conclusiones | 143 |
| A. Aplicación en el programa R de la metodología basada en criterios de información al AFMV | 149 |
| Referencias bibliográficas | 165 |

Introducción al contexto en que se desarrolla esta Tesis

El análisis factorial es un nombre genérico que se da a una clase de métodos estadísticos multivariantes cuyo propósito principal es definir la estructura subyacente en una matriz de datos. Generalmente hablando, aborda el problema de cómo analizar la estructura de las interrelaciones (correlaciones) entre un gran número de variables, con la definición de una serie de dimensiones subyacentes comunes, conocidas como *factores*. El análisis factorial se suele utilizar en la reducción de los datos para identificar un pequeño número de factores, que expliquen la mayor parte de la varianza observada de un número mayor de variables manifiestas.

Una de las razones por las que resulta complicado definir el análisis multivariante es que el término multivariante no se usa de la misma forma en la literatura. En un sentido amplio, el análisis multivariante se refiere a todas las técnicas estadísticas que simultáneamente analizan medidas múltiples recogidas sobre individuos, objetos, etc. Para algunos investigadores, podrían considerarse multivariantes todos los análisis simultáneos de más de dos variables (Tabachnick y Fidell, 1996). Otros autores reservan el término para situaciones en las que el conjunto de variables sigue una distribución normal multivariante (Anderson, 1984) y a veces se añade la condición de que todas las variables deben ser aleatorias e interrelacionadas, de modo que sus efectos no pueden estudiarse por separado (Bernstein, 1987). Puede definirse el análisis multivariante simplemente como un conjunto de métodos que analiza las relaciones entre un número razonablemente amplio de medidas (variables), tomadas sobre cada objeto o unidad de análisis, en una o más muestras simultáneamente (Martínez, 1999).

En 1904, Charles Spearman publicaba un artículo¹ donde trabajaba con las puntuaciones obtenidas por 33 estudiantes en distintos exámenes, observando ciertos efectos sistemáticos en la matriz de correlaciones entre puntuaciones. Este artículo podría ser considerado el origen del análisis factorial (Kendall, 1980).

Las técnicas analíticas de factores pueden lograr sus propósitos desde una perspectiva exploratoria o confirmatoria. Existe una discusión continuada acerca del papel adecuado del análisis factorial (Hair, Anderson, Tatham y Black, 1995). Muchos investigadores lo consideran meramente exploratorio; desde esta perspectiva, las técnicas de análisis factorial extraen lo que proporcionan los datos, y no tienen restricciones sobre la estimación de los factores o el número de factores a retener. Para muchas aplicaciones, resulta adecuado este uso del análisis factorial. No obstante, hay situaciones en las que el investigador tiene unos pensamientos preconcebidos sobre la estructura real de los datos, basados en un apoyo teórico o en investigaciones previas; en esos casos, se requiere un análisis factorial desde un punto de vista confirmatorio. De hecho, resulta útil concebir situaciones puramente exploratorias y situaciones puramente confirmatorias. En un estudio puramente exploratorio, no hay hipótesis a priori sobre los resultados del análisis factorial. Sin embargo, en el caso confirmatorio las hipótesis son desarrolladas para el número de factores, el peso de cada variable en cada factor y la matriz de correlaciones entre factores (Acito, Anderson y Engledow, 1980).

En términos generales, el análisis factorial es una técnica multivariante cuyo objetivo es explicar la correlación presente en la distribución de una variable observable $X = (X_1, X_2, \dots, X_p)$ en términos de un número mínimo de variables aleatorias no observables, llamadas factores. Concretamente, cada variable X_i está generada por una combinación lineal de un número mínimo de variables no observables, llamadas factores comunes (o *variables latentes*) y una variable aleatoria simple llamada *factor específico*. Los factores comunes generan la estructura de covarianza de X y una parte de la varianza de cada X_i , mientras que el factor específico de cada componente X_i explicará el resto de la varianza correspondiente.

Dentro de las hipótesis previas, la primera de ellas sería la de normalidad, aunque ésta es sobre todo necesaria si se realizan pruebas estadísticas de significación. Es deseable que exista un cierto grado

¹El artículo en cuestión es "*General Intelligence*", *objectively determined and measured*, originalmente publicado en American Journal of Psychology, 15 (1904), 201-293.

de multicolinealidad, ya que el objetivo es identificar variables relacionadas. El análisis factorial debe tener suficientes correlaciones altas para poder aplicarse. Si no hay un número sustancial de correlaciones moderadas (por ejemplo, coeficiente de correlación de Pearson $r > 0,30$), entonces probablemente sea inadecuado realizar este análisis. La validez de los resultados y la adecuación del análisis factorial, vienen condicionadas por obtener valores significativos del índice de Kaiser-Meyer-Olkin (KMO²) y el test de esfericidad de Bartlett³, que se obtienen a partir de la matriz de correlaciones.

Uno de los aspectos más importantes a la hora de llevar a cabo un análisis factorial, es la selección del número de factores a extraer. En el análisis factorial con extracción de factores mediante el método de máxima verosimilitud, que es el contexto en que se desarrolla esta Tesis, la metodología tradicionalmente empleada está basada en el contraste de hipótesis. El uso de criterios basados en la información estadística, marca una dirección más reciente en la cuestión, aportando además soluciones a los problemas derivados del uso de otras técnicas.

²El índice KMO es una medida de la adecuación muestral. Si la suma de los coeficientes de correlación al cuadrado entre todos los pares de variables es pequeña, comparada con la suma de los coeficientes de correlación al cuadrado, la medida KMO se aproxima a la unidad. Esta medida oscila entre 0 y 1; valores inferiores a 0,60 cuestionan el empleo del modelo factorial.

³La prueba de esfericidad de Bartlett contrasta si la matriz de correlaciones es igual a la matriz identidad como hipótesis nula, es decir, que las variables estén incorreladas; en caso de aceptarse esta hipótesis, el modelo factorial resultaría inadecuado.

Objetivos

Esta Tesis se desarrolla en el contexto del análisis factorial con extracción de factores mediante el método de máxima verosimilitud. En ella, se afronta el problema de seleccionar el número adecuado de factores que han de componer la solución factorial. Los principales objetivos que se persiguen con la realización de esta Tesis son los siguientes:

1. Investigar acerca de los principales criterios basados en la información estadística (criterios de información) propuestos en la literatura científica.
2. Estudiar el uso de criterios de información como alternativa a la metodología tradicional basada en contrastes de hipótesis (test asintótico de Bartlett), para la selección del número adecuado de factores a extraer en un análisis factorial mediante máxima verosimilitud.
3. Desarrollar algoritmos implementables en el programa *R* (de distribución libre y cada vez más empleado por parte de la comunidad científica) que permitan calcular los valores de los criterios de información expuestos en esta Tesis, para cada uno de los posibles modelos factoriales.
4. Estudiar propiedades teóricas de los criterios de información más comúnmente utilizados, a partir de sus respectivas definiciones.
5. Proporcionar probabilidades de adecuación o pesos asociados a los posibles modelos factoriales, en base a cada uno de los criterios de información estudiados.
6. Aplicar los criterios de información empleados en esta Tesis a ejemplos reales.

7. Analizar el rendimiento de los criterios de información estudiados, en problemas clásicos de la literatura específica de análisis factorial.

Estructura de la Tesis

Cinco capítulos conforman la estructura central de esta Tesis. En secciones previas, se presenta una breve introducción al análisis factorial, campo en el que se desarrolla este trabajo, y objetivos que se pretenden alcanzar con la realización de esta Tesis, así como futuras líneas de investigación.

En el primer capítulo, se exponen los supuestos teóricos en los que se fundamenta el análisis factorial con extracción de factores mediante máxima verosimilitud. Asimismo, se trata el problema de la selección del número adecuado de factores y de la metodología tradicional empleada para tal fin, basada en el contraste de hipótesis. Además, se presenta una sección acerca de los procedimientos de rotación de factores y otra sobre la relación entre el análisis factorial y dos técnicas estadísticas multivariantes: el análisis de componentes principales y el análisis de componentes imagen.

En el segundo capítulo, se plantea el problema de la selección del número adecuado de factores mediante el uso de criterios de selección de modelos, basados en la información estadística. Se presenta una amplia recopilación de estos criterios de información, así como su aplicación al análisis factorial. Para cada criterio, se proponen además pesos asociados, como medidas de evidencia en favor de cada uno de los modelos factoriales candidatos.

En el tercer capítulo, se estudian diversas propiedades teóricas de los criterios de información planteados en el capítulo anterior. Para ilustrar algunas de estas propiedades, se presenta un ejemplo de aplicación. Además, se proporcionan funciones para el cálculo en R de los valores y pesos asociados a cada uno de los posibles modelos factoriales, en base a los diversos criterios de información estudiados en esta Tesis.

En el cuarto capítulo, se presentan tres ejemplos clásicos de la literatura especializada en análisis factorial, y la aplicación de los criterios de información estudiados al problema de la selección del

número adecuado de factores en cada uno de ellos. Considerando como solución factorial adecuada la propuesta por investigadores expertos, se analiza el rendimiento de los criterios de información en cada uno de los problemas.

Por último, en el quinto capítulo, se hace uso de la metodología basada en criterios de información para el estudio de la dimensionalidad en el proceso de validación de un cuestionario recientemente desarrollado y publicado. Se analiza, además, el rendimiento de los criterios de información, considerando como referencia la solución propuesta por los autores del cuestionario.

Finalmente, en la sección de *Conclusiones*, se realiza un breve resumen de los contenidos previamente tratados, y se presentan las principales aportaciones de este trabajo al problema de la selección del modelo factorial adecuado en análisis factorial mediante el método de máxima verosimilitud. En la sección de *Referencias bibliográficas*, se citan todos los documentos referenciados previamente en el texto. Inmediatamente después de la sección de *Conclusiones* y antes de la sección de *Referencias bibliográficas*, se incluye un Anexo en el que se presenta la aplicación práctica, mediante el software correspondiente, de la metodología propuesta, a uno de los ejemplos anteriormente tratados.

Futuras líneas de investigación

La realización de esta Tesis conlleva la apertura de una serie de problemas y líneas de investigación. En el contexto del Análisis Factorial basado en el método de extracción de factores mediante Máxima Verosimilitud (AFMV), algunos de los objetivos que se plantean para ser acometidos con posterioridad a esta Tesis, son los siguientes:

- Desarrollo de un paquete en el programa *R*, que agrupe las funciones implementadas en esta Tesis, con el fin de que la metodología basada en criterios de información para la selección del número adecuado de factores en AFMV esté disponible, de forma libre y fácilmente accesible, para el investigador que la requiera.
- Realización de estudios de simulación para analizar la eficiencia de los criterios de información expuestos en esta Tesis, a la hora de seleccionar el modelo factorial adecuado. Puede resultar especialmente interesante analizar el comportamiento bajo diversos tamaños muestrales, o, por ejemplo, bajo situaciones en que existan ligeras desviaciones del supuesto de normalidad.
- La búsqueda de un modelo factorial que explique la estructura correlacional de las pruebas que componen la disciplina atlética del Decatlón, ha sido objeto de estudio en diferentes análisis a lo largo de las últimas décadas. La diversidad de las soluciones propuestas mantiene abierta la cuestión. En una de las secciones del Capítulo 4 de esta Tesis, ha sido abordado el problema. Puede resultar necesario un análisis desde otra perspectiva, que incluya el uso, por ejemplo, de otros procedimientos de rotación.
- Comparativa de los criterios de información que mejor rendimiento ofrecen, en los conjuntos de datos analizados en esta Tesis, con otros procedimientos de selección de factores que no han

tenido un excesivo protagonismo en la literatura científica, como el MAP (*Minimum Average Partial method*) de Velicer (1976) o el análisis paralelo de Horn (1965), pero que han obtenido buenos resultados en los escasos estudios comparativos publicados.

- Aplicación de la metodología basada en criterios de información al estudio de la dimensionalidad de escalas y cuestionarios, en el ámbito de Ciencias de la Salud. Tal y como se observa en el ejemplo que ilustra el Capítulo 5 de esta Tesis, en tamaños muestrales moderados o suficientemente grandes, la extracción de factores mediante los métodos de componentes principales y máxima verosimilitud, ofrece un resultado similar en el análisis de la validez factorial, por lo que se puede concluir que no existe una desviación fuerte del supuesto de normalidad y el método de máxima verosimilitud puede ser empleado en la extracción de factores, y, por ende, la metodología basada en criterios de información es aplicable al problema de seleccionar el número adecuado de factores o dimensiones subyacentes.

Capítulo 1

Análisis Factorial

El análisis factorial está motivado por el hecho de que las variables medidas pueden estar correlacionadas de tal forma que la estructura correlacional pueda ser reconstruida mediante un número más pequeño de parámetros, los cuales podrían representar la estructura subyacente de forma concisa e interpretable. Un punto importante es decidir qué variables incluir en un modelo factorial, dado que el número de factores y coeficientes depende generalmente de la elección. Si el objetivo del análisis es puramente exploratorio, entonces pueden ser incluidas en el modelo, en un principio, todas las variables originales. Los objetivos de un análisis factorial exploratorio están normalmente limitados a un análisis correlacional de las variables (con la presencia de términos error incorrelados), o la intención puede ser precisamente descubrir qué variables retener como medidas efectivas de las dimensiones de interés.

El modelo factorial general puede ser escrito como

$$Y = \mu + LF + \varepsilon \quad (1.1)$$

o en la siguiente versión reducida de la ecuación (1.1), si las variables originales son centradas,

$$X = LF + \varepsilon, \quad (1.2)$$

donde

- $Y = (y_1, y_2, \dots, y_p)^T$ es el vector de variables aleatorias observadas. Usualmente, estas variables se toman centradas. Sea $E[y_i] = \mu_i$, $\mu = (\mu_1, \mu_2, \dots, \mu_p)$. Así, $X = Y - \mu$ es el vector ($p \times 1$) de variables X_i , $i = 1, \dots, p$, de la población considerada.
- L es una matriz $p \times m$ llamada matriz de cargas factoriales. Sus elementos, l_{ij} , son las cargas de la variable X_i en el factor común F_j .
- F designa un vector ($m \times 1$) constituido por las variables latentes o *factores comunes*, F_i , $i = 1, \dots, m$, $m < p$. F es el vector de factores comunes.
- ε es un vector ($p \times 1$) con componentes ε_i , $i = 1, \dots, p$, que designa los errores asociados a cada variable X_i .

Así, el modelo factorial puede ser expresado en términos algebraicos por la ecuación lineal (Maxwell, 1977)

$$X_i = (l_{i1}F_1 + l_{i2}F_2 + \dots + l_{im}F_m) + \varepsilon_i, \quad i = 1, \dots, p, \quad (1.3)$$

que establece que cada variable X_i es una suma ponderada de m factores, más una variable residual específica para cada una de dichas variables X_i .

Usualmente, las variables objeto de estudio, los factores y las variables residuales, se toman con media cero. Así, $E[X] = E[F] = E[\varepsilon] = 0$.

El término del error aleatorio consiste en errores de medida, junto con efectos individuales únicos asociados con cada variable poblacional. Cuando es considerada una muestra aleatoria, los errores residuales también contienen variación muestral.

Las siguientes hipótesis suelen ser tenidas en cuenta para el modelo factorial dado por la ecuación (1.2):

1. $rg(L) = m < p$.
2. $E[X|F] = LF$.

$$3. E[XX^T] = \Sigma, E[FF^T] = \Omega, E[\varepsilon\varepsilon^T] = \Psi = \begin{bmatrix} \psi_1 & \cdot & \cdot & \cdot & 0 \\ \cdot & \psi_2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \psi_p \end{bmatrix},$$

por lo que los errores se suponen incorrelados. Los factores, sin embargo, están generalmente correlacionados, mientras que Ω no es necesariamente diagonal. Por cuestiones de conveniencia y eficacia computacional, los factores suelen ser supuestos incorrelados y de varianza unidad, por lo que $\Omega = I$ y

$$E[F_i F_j] = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}.$$

Cuando se requiere un sistema no ortogonal, los factores pueden ser rotados a forma oblicua.

$$4. E[F\varepsilon^T] = 0, \text{ por lo que los factores y errores son incorrelados.}$$

Empleando las cuatro hipótesis anteriores, se obtiene, a partir de la ecuación (1.2),

$$\begin{aligned} E[XX^T] &= \Sigma = E[(LF + \varepsilon)(LF + \varepsilon)^T] = E[LF F^T L^T + LF \varepsilon^T + \varepsilon F^T L^T + \varepsilon \varepsilon^T] = \\ &= L \cdot E[FF^T] \cdot L^T + L \cdot E[F\varepsilon^T] + E[\varepsilon F^T] \cdot L^T + E[\varepsilon \varepsilon^T] = L \cdot E[FF^T] \cdot L^T + E[\varepsilon \varepsilon^T] = \\ &= L\Omega L^T + E[\varepsilon \varepsilon^T] = \Gamma + \Psi, \end{aligned} \tag{1.4}$$

donde $\Gamma = L\Omega L^T$ y $\Psi = E[\varepsilon \varepsilon^T]$ son, respectivamente, la matriz de covarianzas real y la matriz de covarianzas de los errores. Así pues, resulta la ecuación estructural de Σ en el modelo de análisis factorial,

$$\Sigma = L\Omega L^T + \Psi. \tag{1.5}$$

En el modelo estadístico factorial supuesto, dada la ecuación (1.5), se estructura la matriz de covarianzas de la población, Σ , en términos del triple $(L; \Omega; \Psi)$. Por tanto, modelizar la población $X = (X_1, \dots, X_p)$ mediante análisis factorial, es equivalente a encontrar las matrices L , Ω y Ψ con las condiciones exigidas en el modelo, de tal forma que la matriz de covarianzas poblacional viene

dada por la ecuación estructural citada.

Por otra parte,

$$E[XF^T] = E[LF F^T + \varepsilon F^T] = L \cdot E[FF^T] + E[\varepsilon F^T] = L \cdot E[FF^T] = L\Omega.$$

Para el caso especial en que $\Omega = I$, la covarianza entre las variables manifiestas y latentes se simplifica a $E[XF^T] = L$.

La matriz L tiene p filas y m columnas. Se tiene también, usando (1.3) y empleando la independencia de los distintos factores, que

$$Var(X_i) = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 + Var(\varepsilon_i) = \sum_{k=1}^m l_{ik}^2 + \psi_i, \quad (1.6)$$

dado que los factores comunes tienen varianza unidad. La parte de la varianza explicada por los factores comunes, llamada $\sum_{k=1}^m l_{ik}^2$, es la comunalidad de la i -ésima variable. La igualdad (1.6) puede interpretarse (Peña, 2002) como una descomposición de la varianza en

Varianza observada = Variabilidad común (comunalidad) + Variabilidad específica,

que es análoga a la descomposición clásica de la variabilidad de los datos, en una parte explicada y otra no explicada, que se realiza en el análisis de la varianza. En el modelo factorial, la parte explicada es debida a los factores, y la no explicada a las perturbaciones o componente aleatorio.

Un caso especialmente importante es el modelo normal de Lawley de Análisis Factorial, caracterizado por la hipótesis de distribución normal multivariante de X (población), F (factores) y ε (errores). Este tipo de análisis factorial fue introducido por Lawley y Maxwell (1971). Sea la distribución condicionada de X dado F , es decir, $X|F$. Considérese que $F \sim \mathcal{N}(0, I)$, de acuerdo con las exigencias del modelo, $E[F] = 0$ y $cov(F) = I$. Entonces, como $X \sim \mathcal{N}(\mu; \Sigma)$, se tiene

$$(X, F) \sim \mathcal{N}_{p+m} \left[\begin{pmatrix} \mu \\ 0 \end{pmatrix}; \begin{pmatrix} cov(X; X) & cov(X; F) \\ cov(F; X) & cov(F; F) \end{pmatrix} = \begin{pmatrix} \Sigma & L \\ L^T & I \end{pmatrix} = \begin{pmatrix} LL^T + \Psi & L \\ L^T & I \end{pmatrix} \right].$$

Por tanto (véase Gutiérrez y González (1991), proposición 1.5), la distribución condicionada será

$$X|F \sim \mathcal{N}_p(LF; \Sigma - LL^T) \equiv \mathcal{N}_p(LF; \Psi).$$

Como Ψ es diagonal, ello es condición necesaria y suficiente para que X y F sean condicionalmente independientes. Por tanto, en este modelo ortogonal normal, el vector F de factores, con sus cargas factoriales dadas por la matriz L , reproduce todas las covarianzas de X . En lo que se refiere a las varianzas de X , tan sólo reproduce una parte de ellas; la parte restante, será reproducida por la matriz Ψ de unicidades.

Esto marca una diferencia sustancial entre el análisis factorial y el análisis de componentes principales, puesto que el análisis de componentes principales explica (a través de sus p componentes principales) las varianzas de la población, pero nada de sus covarianzas.

Se asume que existen m factores comunes, $1 < m < p$, tal que $\Gamma = L\Omega L^T$ y Ψ es una matriz gramiana y diagonal. La matriz de covarianza Σ tiene $\binom{p}{2} + p = \frac{1}{2}p(p+1)$ elementos distintos, el mismo número de ecuaciones que han de ser resueltas. El número de soluciones, sin embargo, es infinito (ver Basilewsky, 1994). Dado que Ω es definida positiva, debe existir una matriz no-singular B ($m \times m$) tal que $\Omega = B^T B$ y

$$\Sigma = L\Omega L^T + \Psi = L(B^T B)L^T + \Psi = (LB^T) \cdot (LB^T)^T + \Psi = L^*(L^*)^T + \Psi. \quad (1.7)$$

Las factorizaciones de Σ dadas por las ecuaciones (1.4) y (1.7), dejan los mismos errores residuales Ψ y además deben representar soluciones factoriales igualmente válidas. Además, en la ecuación (1.7) podría efectuarse la sustitución $L^* = LC$ y $\Omega^* = C^{-1}\Omega(C^T)^{-1}$, que de nuevo proporcionaría un modelo factorial indistinguible del propuesto en la ecuación (1.4). Así, ningún estimador muestral puede distinguir entre las posibles transformaciones, que son infinitas y cada una de ellas de interés potencial.

Esta indeterminación rotacional del modelo factorial requiere restricciones en Ω , la matriz de covarianzas de los factores. La restricción más sencilla y común es tomar $\Omega = I$, es decir, definir los factores como vectores ortogonales unitarios.

Otro supuesto para el modelo factorial ortogonal es que $L^T\Psi^{-1}L \equiv \Delta$, donde Δ es una matriz diagonal ($m \times m$), $\Delta = \text{Diag}(\delta_1, \dots, \delta_m)$, con $\delta_1 > \delta_2 > \dots > \delta_m$. Dicho supuesto es llamado la condición matricial de unicidad, a causa de la multiplicidad de elecciones para la matriz de cargas factoriales, L . Esta condición es utilizada para forzar la convergencia a una solución única de la estimación iterativa máximo verosímil.

Teniendo en cuenta el número de parámetros, bajo el modelo factorial ortogonal, la matriz de cargas factoriales, L , tiene mp parámetros para ser estimados, y la matriz de factores específicos, Ψ , tiene otros p parámetros para ser estimados, para un total de $mp+p$ parámetros. Sin embargo, la condición de unicidad $L^T\Psi^{-1}L = \Delta$ supone $\frac{m(m+1)}{2}$ parámetros distintos y que Δ sea considerada diagonal. Por tanto, esta condición impone $\frac{m(m+1)}{2} - m = \frac{m(m-1)}{2}$ limitaciones adicionales al problema. Así, bajo el modelo factorial ortogonal

$$\Sigma = LL^T + \Psi, \quad (1.8)$$

Σ tiene

$$k = (mp + p) - \frac{1}{2}m(m - 1) \quad (1.9)$$

parámetros libres.

De acuerdo a la ecuación (1.8), la matriz de covarianzas de los datos observados admite una descomposición como suma de dos matrices:

- La primera, LL^T , es una matriz simétrica de rango $m < p$. Esta matriz contiene la parte común al conjunto de las variables y depende de las covarianzas entre las variables y los factores.
- La segunda, Ψ , es diagonal, y contiene la parte específica de cada variable, que es independiente del resto.

La ecuación (1.8) es de vital importancia en el análisis factorial (Chatfield y Collins, 1980). Demuestra que los factores explican exactamente los términos que no están en la diagonal de Σ , dado que

Ψ es diagonal. También establece que encontrar las cargas factoriales es esencialmente equivalente a factorizar la matriz de covarianzas de X en esta forma particular, con la condición añadida de que los elementos diagonales de Ψ deben ser no negativos. Sustituyendo en dicha ecuación la matriz teórica de covarianzas, Σ , por la matriz de covarianzas muestrales, S , el sistema estará identificado si es posible resolverlo de manera única. Para ello, existe una restricción en el número de factores posibles. El número de ecuaciones que se obtienen de (1.8) es igual al conjunto de términos de S , que es $p + \frac{p(p-1)}{2} = \frac{p(p+1)}{2}$. El número de incógnitas en el segundo término es pm , los coeficientes de la matriz L , más los términos de la diagonal de Ψ , menos las restricciones impuestas para identificar la matriz L . Suponiendo que $L^T \Psi^{-1} L$ debe ser diagonal, esto conlleva $\frac{m(m-1)}{2}$ restricciones sobre los términos de L . Para que el sistema esté determinado, debe haber un número de ecuaciones igual o mayor que el de incógnitas. En efecto, si existen menos ecuaciones que incógnitas no es posible encontrar una solución única y el modelo no está identificado. Si el número de ecuaciones es exactamente igual al de incógnitas, existirá una solución única. Si existen más ecuaciones que incógnitas, se puede resolver el sistema en el sentido de los mínimos cuadrados y encontrar unos valores de los parámetros que minimicen los errores de estimación. Así, para que la realización de un análisis factorial tenga sentido, se ha de tener:

$$p + pm - \frac{m(m-1)}{2} \leq \frac{p(p+1)}{2},$$

que supone

$$p + m \leq p^2 - 2pm + m^2,$$

es decir,

$$(p - m)^2 \geq p + m, \tag{1.10}$$

donde, recuérdese, p es el número de variables y m es el número de factores.

La desigualdad (1.10) puede verse también en forma estricta (Sclove, 1987), considerándose entonces como modelos candidatos sólo aquellos, con m factores, que cumplan la condición

$$(p - m)^2 > p + m,$$

es decir, aquellos modelos con un número de grados de libertad estrictamente positivo.

El número de grados de libertad, d , viene dado por el número de ecuaciones implícitas por (1.8) menos el número de parámetros libres, k , dado por (1.9). De esta forma,

$$d = \frac{1}{2}p(p + 1) - (pm + p) - \frac{1}{2}(m^2 - m) = \frac{1}{2}[(p - m)^2 - (p + m)], \quad (1.11)$$

que para una aplicación empírica satisfactoria debe ser un valor estrictamente positivo. Esto sitúa un límite superior para el número de factores m que puede ser obtenido en la práctica, un número que es generalmente algo más pequeño que el número de variables p . Nótese además que la ecuación (1.11) asume que las ecuaciones son linealmente independientes, es decir, Σ debe ser no singular y $\text{rg}(L) = m$.

Cuando Σ , Ω y F son dados en forma numérica, se tiene lo que se conoce como una estructura. Una estructura incluye distribuciones completamente especificadas de los factores, junto con un conjunto de ecuaciones y coeficientes numéricos que relacionan variables observadas y factores comunes. Dada una estructura (Σ, Ω, Ψ) , se puede generar una, y sólo una, distribución de probabilidad de las variables aleatorias observadas. Sin embargo, pueden existir varias estructuras que generen la misma distribución. Dado que estas estructuras son todas equivalentes entre sí, el modelo factorial teórico puede no ser identificado (entiéndase por modelo identificado que exista solución y ésta sea única), dado que la posibilidad de que L y Ψ puedan poseer diferentes valores en estructuras equivalentes conlleva la ausencia de unicidad. Sin embargo, la identificación es a menudo posible, como se demuestra por medio de los siguientes teoremas:

Teorema 1.1 (Reiersol, 1950) Sea Σ una matriz de covarianzas $p \times p$. Una condición necesaria y suficiente para la existencia de m factores comunes es que exista una matriz diagonal Ψ , con elementos no negativos, tal que $\Sigma - \Psi = LL^T$ sea definida positiva y de rango m .

El Teorema 1.1 proporciona una condición necesaria y suficiente para la existencia de $1 \leq m < p$ factores comunes, y puede ser considerado el teorema fundamental del análisis factorial. La condición es que exista una matriz diagonal Ψ con valores mayores que cero, tal que $LL^T = \Gamma$ sea semidefinida positiva de rango m . Así, dada una matriz diagonal definida positiva Ψ , la condición garantiza la existencia matemática de una descomposición $\Sigma = \Gamma + \Psi$, sujeta a las limitaciones pertinentes para la identificación, para $1 \leq m < p$. En la práctica, sin embargo, Ψ es raramente conocida y debe ser estimada a partir de los datos, junto con los coeficientes de L . Este teorema no afirma la existencia de varianzas de error no negativas; simplemente, relaciona la existencia de tales varianzas con la existencia de m factores comunes. Si uno, o más de uno, de los valores estimados $\hat{\psi}_i$ de ψ_i toman un valor negativo, o si la matriz estimada de varianzas residuales $\hat{\Psi}$ no es definida positiva, tal solución es claramente inadmisibile y llamada *impropia*. En la práctica, se pueden encontrar, a veces con sorprendente frecuencia, estos resultados, también llamados *casos de Heywood* (Heywood, 1931) en la literatura psicométrica. Un caso de Heywood puede ser interpretado como una llamada a la posibilidad de que el modelo factorial no sea realmente apropiado.

Teorema 1.2 (ver prueba en Takeuchi, Yanai y Mukherjee, 1982) Supóngase que $X = LF + \varepsilon$ posee una estructura tal que $m = m_0 < p$, donde m_0 es el rango mínimo de la matriz de cargas factoriales, L , y Ψ es no singular. Entonces, el modelo factorial contiene un número infinito de estructuras equivalentes cuando $m = m_0 + 1$.

Así, para cualquier valor de m , existe un número infinito de matrices diagonales no singulares Ψ , con elementos no negativos, que satisface el modelo factorial de la ecuación (1.6). Por otra parte, para determinar una única Ψ se requiere el valor de m , y el valor del rango mínimo, $m = m_0$, es logrado sólo cuando Ψ es única.

1.1. Métodos de extracción de factores

Se pueden emplear diversos métodos para extraer los factores comunes en un modelo de análisis factorial ortogonal. Existen dos métodos que son paradigma de la metodología de estimación, según

se suponga o no normalidad en el modelo. Cuando no se supone explícitamente normalidad de los elementos aleatorios del modelo básico, X , F y ε , la metodología de extracción por componentes principales suele ser la más utilizada. En cambio, cuando se supone normalidad multivariante (modelo factorial de Lawley-Maxwell), el método de estimación por máxima verosimilitud está ligado de forma natural a la situación y es el método idóneo al respecto. Existen otros métodos de estimación; por ejemplo, el método del centroide o el de mínimos cuadrados, que también pueden ser útiles en la práctica en determinados enfoques de análisis factoriales propios de ciertos campos de aplicación (ver, por ejemplo, Basilewsky (1994)). Los diferentes métodos de extracción de factores intentan determinar el número mínimo de factores comunes capaces de reproducir la varianza observada en la matriz de correlaciones inicial. Una comparación entre ellos puede hacerse en función del enfoque utilizado en la extracción, el método de extracción de las comunalidades o el hecho de que las puntuaciones factoriales sean estimadas o calculadas (García, Gil y Rodríguez, 2000).

El método de Máxima Verosimilitud en el modelo de Lawley-Maxwell de análisis factorial

Este método (Lawley y Maxwell, 1971) está basado en el modelo dado por la ecuación (1.2); asumiendo la hipótesis de normalidad multivariante, consiste en aplicar el método de máxima verosimilitud. Puede emplearse con la matriz de correlaciones o con la matriz de covarianzas muestrales. Este método tiene la ventaja de que las estimaciones obtenidas no dependen de la escala de medida de las variables. Además, los estimadores de máxima verosimilitud son eficientes y consistentes, y permiten el contraste estadístico de parámetros; es el caso del test asintótico de Bartlett para la selección del número de factores, aplicable en el análisis factorial basado en máxima verosimilitud (AFMV).

Sea $X = (X_1, X_2, \dots, X_p)^T \sim \mathcal{N}_p(\mu; \Sigma)$, el logaritmo de la función de verosimilitud, log-verosimilitud, puede ser expresado como (ver Basilewsky, 1994)

$$\mathcal{L}(\Sigma) = \ln c - \frac{n}{2} \ln |\Sigma| + \frac{1}{2}(n - p - 1) \cdot \ln |S| + \frac{n}{2} \text{tr}(\Sigma^{-1}S), \quad (1.12)$$

donde

- c es un término constante.
- n es el número de observaciones.
- p es el número de variables observadas.
- Σ es la matriz de covarianzas teóricas poblacionales.
- S es la matriz de covarianzas muestrales.

El máximo de la función dada en (1.12) debe ser, omitiendo funciones constantes de las observaciones, igual a

$$M = -\frac{n}{2} \cdot [\ln |\Sigma| + \text{tr}(S\Sigma^{-1})] = -\frac{n}{2} \cdot [\ln |LL^T + \Psi| + \text{tr}(S(LL^T + \Psi)^{-1})].$$

La estimación de las cargas factoriales se obtiene minimizando la función

$$F(\Sigma) = \ln |\Sigma| + \text{tr}(S\Sigma^{-1}) - \ln |S| - p,$$

con $\Sigma = LL^T + \Psi$. Dicha minimización es equivalente a maximizar la función dada por la expresión (1.12). Un compendio de éste y otros resultados puede verse en Basilewsky (1994, págs. 370-379).

1.2. Selección del modelo adecuado en análisis factorial

La determinación del número adecuado de factores es el primer problema que debe ser resuelto en la selección del modelo factorial. La identificación de una estructura estable de factores se ha realizado tradicionalmente por medio de tests estadísticos basados en la razón de verosimilitudes y, más recientemente, a través de otros métodos como los criterios de información (Costa, 1996).

La selección del modelo se refiere al problema de utilizar los datos para seleccionar un modelo a partir de un conjunto de posibles modelos. Esencialmente, implica el uso de un criterio de selección de modelos para encontrar el modelo que mejor se ajusta a los datos. Dicha selección puede estar basada en contrastes de hipótesis o realizarse a través del uso de criterios de información. El procedimiento de contraste de hipótesis se construye basado en una secuencia de tests de hipótesis. Asumiendo que

el conjunto de modelos candidatos $\{M_i : i = 1, 2, \dots\}$ está ordenado, de forma que M_i es preferible a M_{i+1} , puede contrastarse una secuencia de hipótesis $H_{i0}: M_i$ “retiene el verdadero número de factores” frente a sus respectivas alternativas $H_{i1}: M_{i+1}$ “retiene el verdadero número de factores”, para cada i . Una vez H_{i0} es aceptada, el procedimiento de contraste se detiene y el modelo M_i es seleccionado. Este procedimiento, sin embargo, no penaliza la sobreparametrización (Sclove, 1987), mientras que los criterios de información son criterios de selección que equilibran el ajuste del modelo y su complejidad.

Hay que estar de acuerdo, en primer lugar, en el hecho de que no existen modelos ciertos o verdaderos. Los modelos son, por definición, aproximaciones a la realidad o verdad desconocida. Para ilustrar esto, sirva la frase del estadístico norteamericano George Box: “Esencialmente, todos los modelos son erróneos, pero algunos son útiles”. En el análisis de datos empíricos, se debe encarar la cuestión “¿Qué modelo debería ser usado para aproximar mejor la realidad, dado un conjunto de datos?”. La selección del modelo no ha de estar solamente basada en la bondad del ajuste, sino también en la complejidad; un modelo altamente complejo puede proporcionar un buen ajuste sin que muestre una relación interpretable con el proceso subyacente. La selección del modelo debería de estar, pues, basada en su capacidad para captar las características de la población (Myung, 2000).

Considérese el análisis factorial ortogonal de Lawley-Maxwell, es decir, un modelo factorial en el que sus elementos aleatorios X , F y ε se distribuyen según normales multivariantes y en particular X sigue una distribución $\mathcal{N}_p(\mu; \Sigma)$, con $\Sigma = LL^T + \Psi$. La construcción de tests de hipótesis en dicho contexto se basa, en un enfoque clásico, en la metodología del cociente de verosimilitudes y su comportamiento asintótico en distribución bajo unas condiciones de regularidad conocidas. Esta metodología es viable bajo la hipótesis de normalidad del modelo factorial. De esta manera, se obtiene el test asintótico de Bartlett sobre la existencia de un modelo factorial con m factores comunes ortogonales. Bajo la extracción de factores mediante máxima verosimilitud, pueden obtenerse además una serie de índices adicionales para evaluar la dimensionalidad. Es el caso de los criterios de información, que marcan una dirección más reciente en la cuestión.

En el análisis factorial basado en máxima verosimilitud, se asume generalmente que $F = [F_1, F_2, \dots, F_m]^T$ y $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p]^T$ siguen distribuciones normales m -variante y p -variante, respectivamente. Con-

secuientemente, varios métodos de inferencia estadística basados en la teoría de distribución muestral están disponibles, incluyendo el test basado en la razón de verosimilitudes concerniente al número de factores, el llamado test asintótico de Bartlett sobre la adecuación de un modelo factorial ortogonal con m factores comunes, cuya descripción se presenta en el siguiente apartado.

1.3. Contrastes de hipótesis en análisis factorial

La decisión más importante que debe ser tomada en un análisis factorial es la elección de m , el número de factores comunes (Johnson y Wichern, 1988; pág. 415). La aceptación del supuesto de normalidad, lleva directamente a los tests de hipótesis sobre la adecuación del modelo factorial. Cuando $X \sim \mathcal{N}_p(0; \Sigma)$, la hipótesis de m factores comunes puede ser testada contrastando la hipótesis nula,

$$H_0 : \Sigma = LL^T + \Psi,$$

frente a la hipótesis alternativa

$$H_1 : \Sigma \neq LL^T + \Psi,$$

para determinar si Σ contiene $m > 0$ factores comunes (más una matriz diagonal de varianzas de error) frente a la hipótesis alternativa de que Σ sea una matriz de covarianzas arbitraria.

1.3.1. Test asintótico de Bartlett sobre la adecuación de un modelo factorial ortogonal con m factores comunes

Supóngase que se ha estimado un modelo con m factores. La aceptación de la hipótesis de una población normal lleva directamente a los tests de hipótesis sobre la adecuación del modelo. El contraste de que la descomposición es adecuada puede plantearse como un contraste de razón de verosimilitudes, que recibe el nombre de contraste asintótico de Bartlett-Lawley, o test asintótico de la chi-cuadrado de Bartlett (Bartlett, 1950), en un análisis factorial ortogonal de Lawley-Maxwell. Su desarrollo se muestra de forma resumida en este apartado:

$$H_0 : \Sigma = LL^T + \Psi$$

frente a

$$H_1 : \Sigma \text{ es cualquier otra matriz definida positiva.}$$

Cuando la matriz Σ no tiene ninguna forma especial, el máximo de la función de verosimilitud es proporcional a

$$|\hat{\Sigma}|^{-n/2} \cdot \exp\left(-\frac{np}{2}\right),$$

donde

- $p =$ número de variables.
- $\hat{\Sigma} = \frac{n-1}{n} \cdot S = \frac{A}{n}$.
- $n =$ número de observaciones.
- $S =$ matriz de covarianzas muestrales.

Concretamente, la verosimilitud maximizada bajo la hipótesis alternativa H_1 tiene la forma

$$\mathcal{L}(\bar{X}; \frac{A}{n}) = \exp\left(-\frac{np}{2} \ln(2\pi)\right) \cdot \left|\frac{A}{n}\right|^{-n/2} \cdot \exp\left(-\frac{np}{2}\right).$$

Bajo la hipótesis nula, H_0 , Σ está restringida a tener la forma de la ecuación (1.8). En este caso, el máximo de la verosimilitud (con $\hat{\mu} = \bar{X}$ y $\hat{\Sigma} = \hat{L}\hat{L}^T + \hat{\Psi}$, donde \hat{L} y $\hat{\Psi}$ son los estimadores de máxima verosimilitud de L y Ψ) es proporcional a

$$\begin{aligned} |\hat{\Sigma}| \cdot \exp\left(-\frac{1}{2} \text{tr}(\hat{\Sigma}^{-1}A)\right) &= |\hat{L}\hat{L}^T + \hat{\Psi}|^{-n/2} \cdot \exp\left(-\frac{1}{2} \text{tr}(n\hat{\Sigma}^{-1}\frac{A}{n})\right) = \\ &= |\hat{L}\hat{L}^T + \hat{\Psi}|^{-n/2} \cdot \exp\left(-\frac{n}{2} \text{tr}(\hat{\Sigma}^{-1}\frac{A}{n})\right). \end{aligned}$$

Concretamente, la verosimilitud maximizada es

$$\max_{A.Factorial} \mathcal{L} = \exp\left(-\frac{np}{2} \ln(2\pi)\right) \cdot |\hat{L}\hat{L}^T + \hat{\Psi}|^{-n/2} \cdot \exp\left(-\frac{np}{2}\right).$$

En consecuencia, siguiendo la metodología del “cociente de verosimilitudes (multivariante de Wilks)” y calculados los dos máximos del cociente de verosimilitudes, éste se expresa finalmente así:

$$\frac{\max_{A.Factorial} \mathcal{L}}{\max_{\mu; \Sigma} \mathcal{L}} = \frac{|\frac{A}{n}|^{n/2}}{|\hat{L}\hat{L}^T + \hat{\Psi}|^{n/2}} = \left(\frac{|\frac{A}{n}|}{|\hat{L}\hat{L}^T + \hat{\Psi}|}\right)^{n/2} = W.$$

Por tanto, según el método de construcción de test de hipótesis basado en el cociente de verosimilitudes (procedimiento habitual en hipótesis nulas y alternativas compuestas con estimadores obtenidos bajo máxima verosimilitud), se ha de tender a rechazar la hipótesis nula H_0 (en este caso, la existencia de un análisis factorial con m factores) en la medida en que W sea pequeño. La cuestión es cómo definir “valores pequeños” de W , en términos estadísticos. Habría que conocer para ello la distribución exacta de W , lo cual no es posible. Como es sabido, existe un resultado asintótico para la distribución de W en el contexto aquí supuesto, de normalidad multivariante (test de la χ^2). Para ello, se va a considerar el valor de $-2 \ln W$, es decir:

$$-2 \ln W = -2 \frac{n}{2} \ln \left(\frac{|\frac{A}{n}|}{|\hat{L}\hat{L}^T + \hat{\Psi}|} \right) = -n \ln \left(\frac{|\frac{A}{n}|}{|\hat{L}\hat{L}^T + \hat{\Psi}|} \right) = n \ln \left(\frac{|\hat{L}\hat{L}^T + \hat{\Psi}|}{|\frac{A}{n}|} \right),$$

que es tal que se comporta (Test de Bartlett), bajo la hipótesis nula, así:

$$\left[n - 1 - \frac{2p + 4m + 5}{6} \right] \ln \left(\frac{|\hat{L}\hat{L}^T + \hat{\Psi}|}{|\frac{A}{n}|} \right) \underset{n \rightarrow \infty}{\sim} \chi_{v-v_0}^2,$$

en donde se ha sustituido n por la expresión indicada (aproximación sugerida por Bartlett (1947)) y en donde $v - v_0 = \frac{1}{2}p(p+1) - [p(m+1) - m(m-1)/2] = \frac{1}{2}[(p-m)^2 - p - m] = d$, es decir: “el número de parámetros (elementos) de Σ , $\frac{p(p+1)}{2}$, más el número de restricciones identificadas, $\frac{m(m-1)}{2}$, menos el número de parámetros en L y Ψ , $pm + p = p(m+1)$ ”.

Obsérvese que se ha realizado la sustitución “rechazar H_0 para valores pequeños de W ” por “rechazar H_0 para valores grandes de $-2 \ln W$ ”. Por tanto, el test final obtenido puede formularse en los siguientes términos:

”Se rechaza H_0 , al nivel de significación α , si:

$$\left[n - 1 - \frac{2p + 4m + 5}{6} \right] \ln \left(\frac{|\hat{L}\hat{L}^T + \hat{\Psi}|}{|\frac{A}{n}|} \right) > \chi^2_{\frac{(p-m)^2 - p - m}{2}}(\alpha).$$

En este test aproximado, como los grados de libertad han de ser positivos en una χ^2 , es claro que el número de factores comunes, m , ha de verificar:

$$m < \frac{1}{2} \left[2p + 1 - \sqrt{8p + 1} \right],$$

es decir,

$$(p - m)^2 - (p + m) > 0.$$

Al implementar el test, se está contrastando la adecuación de los m factores del modelo comparando las varianzas generalizadas, $|\hat{L}\hat{L}^T + \hat{\Psi}|$ y $|S_n| = |\frac{A}{n}|$. Si n es grande y m es pequeño en relación a p , la hipótesis H_0 será normalmente rechazada, indicando ello la necesidad de tomar más factores. En muchas ocasiones, el investigador no conoce el valor del número de factores, m , que supone. Entonces, el objetivo es determinar el menor número posible para que el modelo sea consistente con los datos. Así, es usual realizar el contraste para sucesivos valores de m . El investigador arranca con un test tal que el número de factores es un m_0 específico (posiblemente, 0 ó 1). Si la hipótesis es rechazada, procede a comprobar el comportamiento del test con $m_0 + 1$ factores. Continúa de tal forma hasta que la hipótesis se acepte, o hasta que

$$(p - m)^2 - (p + m) \leq 0.$$

En el último caso, se concluye que se ajusta un modelo con factores no triviales. Desgraciadamente, las posibilidades de error bajo ese modelo son desconocidas, incluso de manera asintótica.

Otra forma de expresar el test asintótico de la chi-cuadrado (Lawley-Maxwell, 1971)

Una forma alternativa de expresar el test asintótico de Bartlett dado en la sección anterior, corresponde a Lawley y Maxwell (1971) y se puede ver de forma detallada en Basilewsky (1994, págs. 385-387). En este apartado, se presenta de forma resumida.

Considérese en primer lugar el contraste de independencia completa de las variables (Lawley y Maxwell, 1971), es decir, si uno o más factores comunes están presentes. Bajo la hipótesis nula, H_0 , la función de verosimilitudes es

$$\begin{aligned}\mathcal{L}(\omega) &= c \cdot |\hat{\Sigma}|^{-n/2} \cdot \exp \left[-\frac{n}{2} \cdot \text{tr}(\hat{\Sigma}^{-1}S) \right] = \\ &= c \cdot |\hat{L}\hat{L}^T + \hat{\Psi}|^{-n/2} \cdot \exp \left[\text{tr}(\hat{L}\hat{L}^T + \hat{\Psi}) \cdot S \right],\end{aligned}$$

y bajo la hipótesis alternativa, H_1 , se tiene

$$\mathcal{L}(\Omega) = c \cdot |S|^{-n/2} \cdot \exp \left[-\frac{n}{2} \cdot \text{tr}(S^{-1}S) \right] = c \cdot |S|^{-n/2} \cdot \exp \left(-\frac{np}{2} \right),$$

dado que, bajo H_1 , $\hat{\Sigma} = S$, el estimador muestral. El estadístico basado en la razón de verosimilitudes para contrastar H_0 , es $\lambda = \frac{\mathcal{L}(\omega)}{\mathcal{L}(\Omega)}$, donde $-2 \ln \lambda$ es asintóticamente una chi-cuadrado. Para un tamaño n suficientemente grande,

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

Dado que $\text{diag}(\hat{\Sigma}) = \text{diag}(S)$, el criterio se reduce a

$$\chi^2 \simeq n \cdot \left[\ln |\hat{\Sigma}| + p - \ln |S| - p \right] = n \cdot \ln \left(\frac{|\hat{\Sigma}|}{|S|} \right). \quad (1.13)$$

Puesto que el test para la independencia completa es equivalente a contrastar si $m = 0$, se tiene, a

partir de (1.13), que para muestras grandes

$$\chi^2 = -n \cdot \ln \left(\frac{|S|}{|\hat{\Sigma}|} \right) = n \cdot \ln |R|, \quad (1.14)$$

dado que, cuando $m = 0$, se tiene que $\hat{\Sigma} = \text{diag}(S)$ y la razón de determinantes es igual a $|R|$, el determinante de la matriz de correlaciones. El estadístico dado en (1.14) es válido para un amplio tamaño muestral n . Para tamaños muestrales que no son lo suficientemente amplios, la aproximación chi-cuadrado es mejorable si n es reemplazado por $n - 1 - \frac{2p+4m+5}{6}$, tal como sugiere Bartlett (1947), donde p es el número de variables observadas y m el de factores, siendo $d = \frac{(p-m)^2 - (p+m)}{2}$ el número de grados de libertad, la diferencia entre el número de parámetros en Σ y el número de restricciones lineales impuestas mediante la hipótesis nula. Cuando se acepta H_0 , es decir, cuando el estadístico chi-cuadrado asintótico indica la presencia de al menos un factor común, el test es repetido para valores mayores de m , dado que el objetivo se convierte entonces en estimar el número adecuado de factores comunes, F_1, F_2, \dots, F_m ($1 \leq m < p$).

Teorema 1.3 (Basilewsky, 1994) Sea $X = (X_1, X_2, \dots, X_p)^T$ un vector de p variables aleatorias tal que $X \sim \mathcal{N}_p(\Phi; \Sigma)$ y $\Sigma = LL^T + \Psi$. Entonces,

1. Si existe una única matriz diagonal Ψ con elementos positivos en la diagonal, tal que las m mayores raíces características de $\Sigma^* = \Psi^{-1/2}\Sigma\Psi^{-1/2}$ son distintas y mayores que la unidad, y las restantes $(p - m)$ raíces son iguales a la unidad, entonces L puede ser únicamente definida como $[\Sigma^* - (\eta_i + 1)I] \Psi_i^{-1/2} L_i = 0$.
2. Sea $\Sigma^* = \Psi^{-1/2}\Gamma\Psi^{-1/2} + I$. Entonces, $\eta_i = L_i^T \Psi_i^{-1} L_i$ ($i = 1, 2, \dots, m$) es la i -ésima raíz característica de $\Psi^{-1/2}\Gamma\Psi^{-1/2}$, de forma que $\eta_1 > \eta_2 > \dots > \eta_m$ y $\eta_{m+1} = \eta_{m+2} = \dots = \eta_p = 0$.

En base al Teorema 1.3, las raíces características de $\Sigma^* = \Psi^{-1/2}\Sigma\Psi^{-1/2}$ son iguales a la unidad cuando las de la matriz de correlación reducida $\Psi^{-1/2}\Gamma\Psi^{-1/2}$ son iguales a 0. El test para la existencia de $0 < m < p$ factores comunes es además equivalente a contrastar si las últimas $(p - m)$ raíces características $(\hat{\eta}_i + 1)$ de $\hat{\Psi}^{-1/2}S\hat{\Psi}^{-1/2}$ difieren de la unidad. Así, suponiendo ordenadas las raíces $\hat{\eta}_1 > \hat{\eta}_2 > \dots > \hat{\eta}_m > \dots > \hat{\eta}_p$, el test asintótico de la chi-cuadrado puede expresarse como:

“Se acepta H_0 (existencia de m factores comunes) al nivel de significación α , si

$$\chi^2 = - \left[n - 1 - \frac{1}{6} \cdot (2p + 4m + 5) \right] \cdot \sum_{i=m+1}^p \ln(\eta_i + 1) < \chi_d^2(\alpha), \quad (1.15)$$

con $d = \frac{(p-m)^2 - (p+m)}{2}$ grados de libertad”.

1.4. Rotación de factores

Cuando se emplea el método de máxima verosimilitud en análisis factorial, se ha de tener en cuenta la restricción de que la matriz $L^T \Psi^{-1} L$, donde L es la matriz de cargas factoriales y Ψ la matriz de unicidades, es diagonal. Esta restricción hace que la solución inicial obtenida mediante máxima verosimilitud oculte en la práctica los factores subyacentes, por lo que resulta prácticamente obligatorio realizar una rotación, con el fin de obtener factores rotados que evidencien una interpretación más clara.

Utilizar algún procedimiento de rotación responde a la idea de hacer más fácil la interpretación de los valores que presenta la matriz factorial tras la extracción. La rotación consiste en girar los ejes factoriales. La solución inicial extrae los factores según su importancia, y los siguientes factores van explicando progresivamente menor porcentaje de varianza. Con la rotación se distribuye la varianza en otros factores para lograr un patrón de factores más simple y más significativo. Existen muchos métodos de rotación y de la elección de uno u otro dependerá la solución. Un método frecuentemente utilizado es Varimax (Kaiser, 1958), basado en el hecho de que si se logra aumentar la varianza de las cargas factoriales al cuadrado de cada factor, consiguiendo que algunas de sus cargas factoriales tiendan a acercarse a uno mientras otras se acercan a cero, lo que se obtiene es una pertenencia más clara e inteligible de cada variable a ese factor. La rotación Varimax es ortogonal, lo que significa que los factores permanecen incorrelados en el proceso de rotación. Hay ocasiones en que los factores no necesitan estar incorrelados e incluso pueden estar conceptualmente ligados, lo que requiere una correlación entre los factores. Puede ser adecuado entonces el uso de una rotación oblicua (Hair et al., 1995).

Después de realizar un análisis factorial, suele ser necesario rotar los factores para formular una

solución más interpretable y, por tanto, más adecuada. Thompson (1984, págs. 31-34) demostró que la matriz de estructura factorial sin rotar no representa totalmente la verdadera naturaleza de los factores, y cómo la rotación de factores resuelve este problema. Generalmente, las cargas factoriales proporcionadas por un análisis factorial no son interpretables, a causa de que no reflejan una estructura simple. Gorsuch (1983) da una definición detallada de estructura simple. Colectivamente, un conjunto de variables o ítems exhibe una estructura simple cuando cada factor tiene algunos ítems con cargas elevadas, y el resto de ítems con cargas próximas a cero. La rotación clarifica la estructura factorial, distribuyendo la varianza, algo más equitativamente, entre los factores (Tinsley y Tinsley, 1987).

Según Thurstone (1947), la intención fundamental al realizar una rotación es encontrar una estructura simple. Las propiedades que deberían cumplirse, o, dicho de otra manera, los principios a tener en cuenta al llevar a cabo una rotación de factores, son:

1. Cada variable debería tener al menos una carga nula.
2. Cada factor debería tener un conjunto de variables linealmente independientes cuyas cargas factoriales sean nulas.
3. Para todo par de factores, deberían existir varias variables cuyas cargas sean nulas para un factor, pero no para el otro.
4. Si son extraídos más de 4 factores, para todo par de factores debería existir una gran proporción de variables con cargas nulas en ambos factores.
5. Si son extraídos más de 4 factores, para todo par de factores debería existir sólo un pequeño número de variables con cargas no nulas en ambos factores.

La rotación factorial implica una transformación de las cargas factoriales iniciales, para la obtención de una estructura más simple, que conforme una solución más interpretable. Esta transformación no modifica las relaciones matemáticas subyacentes en los datos, de forma que el ajuste global permanece intacto. Dado que la rotación de factores no modifica el ajuste original de la solución original, ninguno de los métodos disponibles puede ser considerado uniformemente mejor que otro

en todos los casos. Cada método presenta una serie de ventajas e inconvenientes, que aconsejan su aplicación en algunas circunstancias, pero no en otras (Finch, 2006).

Las rotaciones factoriales pueden ser de dos tipos: ortogonales u oblicuas. Las rotaciones ortogonales parten del supuesto de que los factores están incorrelados. Se obtiene una matriz de cargas factoriales rotada, que es realmente una versión transformada de la matriz de cargas factoriales original. En contraste con las rotaciones ortogonales, los métodos de rotación oblicua asumen que los factores están correlacionados unos con otros, y permiten que esta correlación sea tenida en cuenta en la transformación. Las rotaciones oblicuas proporcionan una matriz de estructura factorial y una matriz de configuración. La primera de ellas contiene las correlaciones entre los factores e ítems, mientras que la segunda contiene una medida de las relaciones exclusivas entre los factores individuales y los ítems, excluyendo los efectos superpuestos de otros factores.

De entre las distintas técnicas de rotación, la más conocida y utilizada es la rotación Varimax, propuesta por Kaiser (1958), cuyo objetivo es que los factores tengan pocas saturaciones altas y muchas saturaciones que sean prácticamente nulas, en las variables. Otra rotación ortogonal utilizada, aunque con menor frecuencia que Varimax, es la Cuartimax (Harman, 1976), que pretende que una variable dada esté muy correlacionada con un factor y muy poco correlacionada con el resto de factores. Las rotaciones oblicuas persiguen los mismos objetivos que las ortogonales, y suelen llevarse a cabo cuando las rotaciones ortogonales no ofrecen un rendimiento adecuado. Entre las rotaciones oblicuas más conocidas, destacan la Oblimín directa (Jennrich & Sampson, 1966), la rotación oblicua de Harris y Kaiser (1964) y la rotación Promax (Hendrickson y White, 1964).

Tinsley y Tinsley (1987) avanzaban que, aunque pudiera esperarse que los procedimientos de rotación oblicua presentasen un rendimiento superior a los procedimientos de rotación ortogonal, esa superioridad no ha sido demostrada empíricamente. En su lugar, los numerosos estudios comparativos de procedimientos de rotación, que habían sido llevados a cabo hasta la fecha, apoyaban las tres siguientes conclusiones:

1. Cuando la estructura simple es clara, los procedimientos de rotación estándar conducen a interpretaciones similares.
2. Varimax es, en apariencia, el mejor de los procedimientos de rotación ortogonales, mientras que

Promax y el criterio de Harris - Kaiser parecen ser los más fiables en cuanto a procedimientos de rotación oblicuos se refiere, proporcionando, los tres métodos referidos, similares estructuras factoriales.

3. El uso de una secuencia de rotación Varimax - Promax proporciona una base para determinar si es preferible una solución ortogonal u oblicua. Si las dimensiones subyacentes son incorreladas, o sólo débilmente correlacionadas, el procedimiento Varimax proporcionará una solución comparable a la resultante del procedimiento de rotación oblicua. Esto evita la necesidad de realizar hipótesis a priori, acerca del grado de correlación entre los factores.

Actualmente, no existe consenso sobre qué tipo de rotación utilizar y sobre cuándo es preferible la rotación oblicua a la ortogonal, y viceversa. En relación a la teoría de tests, McDonald (1997) establece que resulta difícil concebir que los factores subyacentes a un test estén realmente incorrelados, sugiriendo que la rotación oblicua resultaría más adecuada. De Vellis (2003) sugiere emplear rotaciones ortogonales cuando las correlaciones entre factores sean menores que 0,15, y utilizar en cualquier otro caso rotaciones oblicuas, teniendo siempre presente que es difícil dar una regla universal para la aplicación de rotaciones oblicuas, a causa de que las relaciones entre los ítems y los factores son tan importantes como las correlaciones entre factores, a la hora de determinar el tipo de rotación que se va a utilizar.

Kieffer (1998) recomienda utilizar tanto la rotación ortogonal como la oblicua, y comparar entonces los resultados de los dos métodos. Si ambos proporcionan resultados similares, el investigador puede confiar en la solución ortogonal; si, al contrario, los resultados difieren considerablemente, la interpretación basada en la rotación oblicua es preferible. Además, Kieffer sugiere que, si el grado de correlación entre los factores decrece, tanto la solución ortogonal como la oblicua tenderán a proporcionar resultados similares, considerando que, en virtud de que las soluciones oblicuas son menos parsimoniosas, una rotación oblicua debería ser empleada sólo si se supone que los factores resultantes van a estar correlacionados.

Finch (2006) obtiene, en una comparación entre los métodos de rotación Varimax y Promax, que, incluso para altas correlaciones interfactoriales, el método ortogonal proporciona una estructura factorial (es decir, qué ítems estarán relacionados con qué factores) igual de adecuada que el método

oblicuo. Por otra parte, observa que Promax ofrece un mejor rendimiento que Varimax en el ejercicio de reconocer correctamente la presencia de una estructura simple en los ítems.

1.5. Relación entre análisis factorial y análisis de componentes

El análisis factorial y el análisis de componentes son dos tipos de procedimientos que comparten un objetivo común: reducir un conjunto de p variables observadas a un conjunto de m variables nuevas ($m < p$). Esta reducción de la dimensionalidad sirve para dos propósitos diferentes. En primer lugar, para describir la relación entre las variables originales y las nuevas variables. En segundo lugar, la matriz de configuración A puede ser interpretada para describir la relación entre las variables originales y las variables nuevas. En segundo lugar, pueden obtenerse unas puntuaciones para las m variables nuevas que reemplacen las puntuaciones observadas de las variables originales. Dichas puntuaciones pueden ser interpretadas o empleadas como la base para un análisis posterior. El problema básico para ambos tipos de procedimientos consiste en la descripción de un conjunto de p variables aleatorias $\eta^T = (Y_1, Y_2, \dots, Y_p)$ en términos de $m \leq p$ variables aleatorias, $\zeta = (X_1, X_2, \dots, X_m)$ y p variables residuales, $\varepsilon^T = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)$. Tanto el análisis factorial como el análisis de componentes, pueden ser expresados como un modelo de la forma

$$\eta = A\zeta + \varepsilon,$$

donde A es la matriz de configuración de p filas y m columnas que predice de manera óptima las p variables de η a partir de las m variables de ζ . En análisis de componentes, ζ debe ser expresable como $\zeta = A^T\eta$, en cuyo caso la matriz de covarianzas de ε no puede ser diagonal y de rango completo. En análisis factorial, la matriz de covarianzas de ε debe ser diagonal y de rango completo. El análisis de componentes abarca un abanico de métodos que incluyen una descomposición propia de una matriz de covarianzas. La versión más empleada es el análisis de componentes principales, donde todas las variables observadas son transformadas a puntuación típica y el análisis de componentes se lleva a cabo sobre la matriz de correlaciones R , matriz cuadrada de orden p .

El análisis de componentes principales puede ser expresado como

$$R = L_C D_C^2 L_C^T,$$

donde D_C^2 es una matriz diagonal de orden m , que contiene los m autovalores mayores, y L_C es la matriz $p \times m$ que contiene los autovectores correspondientes.

Una variante del análisis de componentes principales, que ha recibido considerable atención, es el análisis de componentes imagen (Guttman, 1953; Harris, 1962). El análisis de componentes imagen es un análisis de componentes llevado a cabo sobre la matriz de covarianzas imagen. Puede ser expresado como

$$S^{-1} R S^{-1} = L_I D_I^2 L_I^T,$$

donde L_I y D_I^2 son las matrices de autovectores y autovalores, respectivamente, y $S^2 = \text{Diag}^{-1}(R^{-1})$.

La matriz de configuración de componentes principales puede ser expresada como

$$A_C = L_C D_C$$

y la matriz de configuración de componentes imagen como

$$A_I = S L_I D_I.$$

El análisis factorial puede ser visto como un modelo lineal que relaciona variables manifiestas y latentes, con una importante limitación en la estructura de covarianzas de las variables latentes. Sin embargo, esta versión del análisis factorial ajusta a η si y sólo si la matriz de covarianzas de las p variables que conforman el vector aleatorio η puede escribirse en la forma

$$\Sigma = A A^T + U^2.$$

Considerando que se emplea como método de extracción de factores el método de máxima verosimilitud, si se emplea la condición de diagonalidad $A^T U^{-2} A = \text{matriz diagonal}$, el resultado puede

expresarse como

$$U^{-1}RU^{-1} = L_F D_F^2 L_F^T,$$

donde L_F y D_F^2 son las matrices que contienen, respectivamente, a los autovectores y autovalores, y la correspondiente configuración factorial puede ser representada como

$$A_F = UL_F(D_F^2 - I)^{1/2}.$$

En resumen, considerando la condición de diagonalidad para el análisis factorial basado en máxima verosimilitud (AFMV), los tres métodos citados pueden ser expresados como descomposiciones propias (Velicer, 1977; Velicer y Jackson, 1990). En la Tabla 1.1, aparecen expresadas tales descomposiciones y las matrices de configuración asociadas a cada método.

| Método | Descomposición propia | Matriz de configuración |
|--------|-----------------------------------|-------------------------------|
| ACP | $R = L_C D_C^2 L_C^T$ | $A_C = L_C D_C$ |
| ACI | $S^{-1}RS^{-1} = L_I D_I^2 L_I^T$ | $A_I = S L_I D_I$ |
| AFMV | $U^{-1}RU^{-1} = L_F D_F^2 L_F^T$ | $A_F = UL_F(D_F^2 - I)^{1/2}$ |

Tabla 1.1. Descomposiciones propias y matrices de configuración asociadas al Análisis de Componentes Principales (ACP), Análisis de Componentes Imagen (ACI) y Análisis Factorial basado en el método de extracción mediante Máxima Verosimilitud (AFMV).

Velicer (1977) obtuvo, en una comparación sobre varios conjuntos de datos, que las configuraciones producidas por el AFMV, el análisis de componentes principales y el análisis de componentes imagen, producían configuraciones similares. El tipo de rotación empleado tenía un efecto pequeño en el grado de similaridad. Por otra parte, el análisis de componentes imagen y el AFMV producían generalmente los resultados más similares, observándose las mayores diferencias entre el AFMV y el análisis de componentes principales, aunque estas resultaban mínimas. Generalmente, las diferencias se encontraron en el último factor extraído, probablemente a causa de la extracción de un número elevado de factores.

Una conclusión diferente la obtuvieron Snook y Gorsuch (1989), que, a partir de un estudio de simu-

lación, observaron que el análisis de componentes y el análisis factorial proporcionaban soluciones discrepantes, especialmente cuando el número de variables, p , era pequeño. Además, consideraban que las ponderaciones correspondientes al análisis de componentes resultaban *infladas*. Velicer y Jackson (1990) rebatían estas conclusiones, considerando que las diferencias en el tamaño numérico de las ponderaciones serían las esperables dadas las diferencias algebraicas entre ambos métodos. Por otra parte, consideraban erróneo el uso del término *infladas*, porque daba a entender que las ponderaciones factoriales eran entonces el valor correcto.

1.5.1. Diferencias entre análisis factorial y análisis de componentes principales

El análisis factorial guarda cierta relación con otra técnica estadística multivariante, como es el análisis de componentes principales. Entre ambas técnicas, existen sin embargo notables diferencias. La primera es que la finalidad de las componentes principales es simplificar la estructura de los datos, sin obedecer a un modelo fijado “a priori”, para poder explicar, en pocas componentes, la mayor parte de la información que contienen las variables; por su parte, el análisis factorial pretende estudiar relaciones de dependencia entre las variables, expresadas a través de un modelo de factores comunes y únicos (Cuadras, 1991). Además, mientras el análisis de componentes principales es una técnica descriptiva, apoyada únicamente en las propiedades matemáticas de las matrices de covarianzas o de correlaciones, el análisis factorial presupone un modelo estadístico al que deben adaptarse los datos.

En el caso del análisis de componentes principales, el objetivo es obtener p combinaciones lineales (llamadas componentes principales) Y_i de las variables del vector X , con la condición de maximizar la varianza de manera que

$$Y_1 = w_{11}X_1 + \dots + w_{1p}X_p$$

$$Y_2 = w_{21}X_1 + \dots + w_{2p}X_p$$

$$\vdots$$

$$Y_p = w_{p1}X_1 + \dots + w_{pp}X_p$$

siendo X_i , $i = 1, \dots, p$, las variables observables. Además, los vectores de los coeficientes de las Y_i se consideran unitarios.

Mientras, en el análisis factorial se detecta la existencia de unos factores comunes F_i , $i = 1, \dots, m$, tales que las variables observables de la población, X_i , se expresan así:

$$X_1 = l_{11}F_1 + \dots + l_{1m}F_m + \varepsilon_1$$

$$X_2 = l_{21}F_1 + \dots + l_{2m}F_m + \varepsilon_2$$

$$\vdots$$

$$X_p = l_{p1}F_1 + \dots + l_{pm}F_m + \varepsilon_p$$

donde $m < p$; los factores comunes F_1, F_2, \dots, F_m son no observables y los $\varepsilon_1, \dots, \varepsilon_p$ son factores únicos aleatorios ligados a cada X_i , que recogen el error de observación de cada X_i y otros factores específicos ligados a cada variable X_i al margen de las otras variables.

Por otra parte, el modelo ecuacional lineal del análisis factorial, es formalmente equivalente a un modelo de regresión lineal múltiple multivariante, en el que la población $X = (X_1, X_2, \dots, X_p)$ multivariante se modeliza mediante relaciones lineales en términos de múltiples regresores F_1, F_2, \dots, F_m ; es decir, p -variante en X y múltiple m en los regresores. Todo ello con el vector error $\varepsilon = (\varepsilon_1, \dots, \varepsilon_p)^T$ aleatorio. Pero esta equivalencia formal entre análisis factorial y regresión lineal múltiple multivariante es meramente aparente. Existe una diferencia fundamental en la filosofía de dichos modelos, y es que en el análisis factorial los *regresores* (factores comunes) son aleatorios y no observables, además de tener en cuenta su número, siempre menor que p ($m < p$). No obstante, el análisis factorial está más próximo al modelo de regresión citado que el propio análisis de componentes principales, en lo que al término error se refiere, ya que éste no existe en análisis de componentes principales y sí en regresión. En este sentido, Gorsuch (1983) estableció que la principal diferencia teórica entre análisis de componentes principales y análisis factorial estriba en el hecho de que el análisis factorial incluye explícitamente el término error en el modelo.

Algunos autores concluyen que las diferencias no existen o son mínimas entre los patrones o resultados obtenidos por ambos procedimientos (Velicer, Peacock y Jackson, 1982; Velicer y Jackson,

1990). Sin embargo, otros investigadores advierten diferencias en el resultado tras aplicar como procedimiento el análisis de componentes principales y el análisis factorial (Snook y Gorsuch, 1989; Widaman, 1990). Dichas diferencias, según Snook y Gorsuch (1989), tienden a reducirse para un número creciente de variables por factor y tamaño de carga, ya que, para un gran número de variables y altas cargas factoriales, el análisis de componentes se asemeja al análisis factorial común. Por otra parte, Kim (2008) encuentra que el análisis de componentes principales tiende a incrementar las cargas factoriales, especialmente en estudios con un número pequeño de variables y/o bajas communalidades estimadas. Hay que tener también en cuenta el problema de la sobreextracción. Según Velicer y Jackson (1990), ambos métodos son similares cuando se trabaja con el número correcto de factores, pero pueden encontrarse diferencias cuando existe sobreestimación, muchas veces provocada por el uso de la llamada *regla de autovalores mayores que la unidad* de Kaiser-Guttman (Guttman, 1953; Kaiser, 1960) para la extracción de factores.

Capítulo 2

Criterios basados en la información estadística

Algunos problemas tratados mediante secuencias de tests de hipótesis, podrían ser tratados más expeditivamente mediante la aplicación de criterios de selección de modelos. A menudo, la selección de modelos proporcionará una alternativa viable a los tests de hipótesis (Sclove, 1987). Además, los contrastes de hipótesis incluyen la elección de niveles significativos. Como indica Akaike (1987), los contrastes de hipótesis no penalizan la sobreparametrización, a causa de que un modelo saturado es usado como referencia. Según Bozdogan y Ramírez (1988), la teoría clásica de contraste de hipótesis no es apropiada en el contexto de problemas de selección del modelo factorial, ya que el problema no es contrastar una hipótesis en particular, sino que el escenario es un problema de decisión multiple. El test de Bartlett basado en la razón de verosimilitudes no siempre resulta válido, y el nivel de significación para el test no está ajustado en el proceso secuencial de contraste.

El test asintótico de la chi-cuadrado de Bartlett resulta generalmente adecuado cuando los datos están distribuidos aproximadamente según una normal. Además, dicho test rechazará, en general, la adecuación de un modelo factorial con un número de factores m pequeño si el número de variables, p , y el de observaciones, n , es elevado. Tal y como sugieren Johnson y Wichern (1988, pág. 399), si el tamaño muestral es amplio y el número de factores extraídos es relativamente pequeño respecto del número de variables originales, la hipótesis nula, H_0 , del test de Bartlett, dado en la sección 1.3.1, será usualmente rechazada, llevando este hecho a una retención de más factores comunes, o, lo que es lo mismo, sobreestimación del número de parámetros. Esta es la situación a la que el investigador se enfrenta aún cuando el análisis factorial proporciona una aproximación útil a la solución del

problema. La mayoría de las veces, la elección final de m está basada en una combinación de:

1. La proporción de varianza muestral explicada.
2. El conocimiento del problema por parte del investigador.
3. La razonabilidad e interpretación teórica de los resultados obtenidos.

Sclove, Johnson y Wichern, y otros autores, coinciden con Tucker y Lewis (1973). El problema que plantea el uso del estadístico basado en la razón de verosimilitudes, y otros estadísticos similares, incluye la forma del procedimiento de decisión. La hipótesis estadística es que el modelo factorial con un número especificado de factores comunes, se aplica estrictamente a una población. El rechazo de esta hipótesis sucederá con más frecuencia para una amplia muestra de sujetos, a cualquiera de los niveles de confianza usuales. Así, para tamaños muestrales elevados, el estadístico basado en la razón de verosimilitudes, y por tanto el test asintótico de la chi-cuadrado de Bartlett, podría indicar que un modelo factorial, que en realidad es adecuado, no representa exactamente las inter-relaciones entre las variables en una población.

Tucker y Lewis (1973) proponían como alternativa un coeficiente de fiabilidad, que mide la adecuación de un modelo con m factores para representar las covarianzas de las variables en una población; la ausencia de ajuste indicaría que las relaciones entre las variables son más complejas que las que puedan ser representadas por m factores comunes.

En las últimas décadas, se ha investigado la utilización de criterios de información basados en la entropía para estimar el número de factores explicatorios que pueden extraerse de una matriz de covarianzas. La razón principal que motiva la utilización de estos criterios es que la estimación clásica de factores por máxima verosimilitud tiende frecuentemente a elegir un número mayor de factores, es decir, en general, a elegir un modelo con exceso de dimensiones que pueden ser interpretadas. También se presenta este problema en el contexto de las series temporales. Asimismo, otros inconvenientes, de los tests basados en máxima verosimilitud dentro del análisis factorial, están relacionados con su carácter asintótico y el tamaño muestral adecuado.

Los criterios de información pueden aplicarse a modelos con parámetros estimados mediante métodos de máxima verosimilitud. No son de uso exclusivo para el análisis factorial y su aplicación se extiende

a otros procedimientos donde seleccionar el modelo adecuado sea el objetivo. Es el caso de modelos de regresión o modelos de series temporales, por citar algunos ejemplos.

Supóngase que los datos $x_n = \{x_1, x_2, \dots, x_n\}$ son generados a partir de una distribución desconocida $G(x)$ con función de densidad $g(x)$; el objetivo del modelado estadístico es entonces construir un modelo razonable basado en los datos observados. En situaciones prácticas, es difícil estimar la densidad correcta $g(x)$ de manera precisa, a partir de un número finito de observaciones. Generalmente, se hace uso de un modelo aproximado que consiste en una familia de densidades de probabilidad $\{f(x|\theta) \cdot \theta \in \Theta\}$, donde $\theta = (\theta_1, \theta_2, \dots, \theta_p)^T$ es el vector p -dimensional de parámetros desconocidos y Θ es un subconjunto abierto de \mathbb{R}^p . El modelo paramétrico es estimado encontrando el estimador adecuado, $\hat{\theta}$, del vector de parámetros desconocidos θ y reemplazando θ en $f(x|\theta)$ por el estimador $\hat{\theta}$. A esta densidad de probabilidad $f(x|\hat{\theta})$ se le llama modelo estadístico.

El problema entonces radica en cómo evaluar la bondad del modelo estadístico, cuando éste es usado para predecir una observación futura z a partir de la densidad $g(z)$. Para medir como se aproxima $f(z|\hat{\theta})$ a la densidad $g(z)$, se utiliza la Información de Kullback-Leibler (Kullback y Leibler, 1951)

$$I\{g(z); f(z|\hat{\theta})\} = E_{G(z)} \left[\ln \frac{g(Z)}{f(Z|\hat{\theta})} \right] = E_{G(z)}[\ln g(Z)] - E_{G(z)}[\ln f(Z|\hat{\theta})], \quad (2.1)$$

donde la esperanza es tomada sobre la distribución de z , sujeta a los datos observados x_n . La información de Kullback-Leibler siempre toma un valor positivo, a menos que $f(z|\hat{\theta})$ sea igual a $g(z)$ en casi todo punto. Se elige el modelo que minimiza el valor de $I\{g(z); f(z|\hat{\theta})\}$ de entre los modelos estadísticos candidatos. Es suficiente considerar sólo el segundo término de la derecha de la expresión (2.1), dado que el primer término, $E_{G(z)}[\ln g(Z)]$, es una constante que depende sólo de la verdadera distribución $g(z)$. El término

$$\eta(G; \hat{\theta}) \equiv E_{G(z)}[\ln f(Z|\hat{\theta})] = \int \ln f(z|\hat{\theta}) dG(z) \quad (2.2)$$

está condicionado a los datos observados x_n a través del estimador $\hat{\theta} = \hat{\theta}(x_n)$, y depende de la verdadera distribución, desconocida, G . Dicha expresión es la esperanza del logaritmo neperiano de la verosimilitud (log-verosimilitud). Cuanto más grande sea su valor para un modelo estadístico,

más pequeño será el valor de su Información de Kullback-Leibler y mejor, en definitiva, el modelo. El término de la log-verosimilitud esperada contiene la distribución desconocida $g(z)$. Además, no se puede directamente evaluar la log-verosimilitud, pero puede ser estimada a partir de las observaciones. La expresión

$$\eta(\hat{G}; \hat{\theta}) = \int \ln f(z|\hat{\theta}) d\hat{G}(z) = \frac{1}{n} \sum_{\alpha=1}^n \ln f(x_\alpha|\hat{\theta}) \equiv \mathcal{L}_n(x_n|\hat{\theta}) \quad (2.3)$$

es obtenida reemplazando la distribución desconocida G en $\eta(G; \hat{\theta})$ por la distribución empírica, \hat{G} , definida asignando una ponderación $\frac{1}{n}$ a cada observación x_α . Llamamos a (2.3) log-verosimilitud, la cual sobreestima por lo general a la log-verosimilitud esperada, $\eta(G; \hat{\theta})$; es decir, $\eta(G; \hat{\theta}) < \mathcal{L}_n(x_n|\hat{\theta})$. El sesgo de la log-verosimilitud al estimar la log-verosimilitud esperada, es definido por

$$b(G) = E_{G(x)} \left[\mathcal{L}_n(x_n|\hat{\theta}) - \int \ln f(z|\hat{\theta}) dG(z) \right], \quad (2.4)$$

donde la esperanza es tomada sobre la unión de las distribuciones de x_n , dada por $\prod_{\alpha=1}^n dG(x_\alpha)$. Si el sesgo puede ser estimado mediante un procedimiento apropiado, entonces un criterio de información puede ser definido (Bozdogan, 2000; Konishi y Kitagawa, 2003; Kitagawa y Konishi, 2010) como una log-verosimilitud corregida para el sesgo,

$$IC(x_n; \hat{\theta}) = -2n\{\mathcal{L}_n(x_n|\hat{\theta}) - \hat{b}(G)\} = -2 \sum_{\alpha=1}^n \ln f(x_\alpha|\hat{\theta}) + 2n\hat{b}(G), \quad (2.5)$$

donde $\hat{b}(G)$ es un estimador del sesgo $b(G)$ dado en la ecuación (2.4).

El término de corrección del sesgo, $\hat{b}(G)$, suele venir dado como un sesgo asintótico en (2.5). De acuerdo a los supuestos considerados en la estimación del modelo y la relación entre el modelo especificado y la verdadera función de densidad, el sesgo asintótico toma diferentes formas, y, como consecuencia, se obtienen diversos criterios de información, como se verá en los siguientes subapartados.

El objetivo, al construir criterios de información, es estimar la cantidad (2.2) a partir de los datos

observados, de la forma más adecuada posible. Si existe un estimador $\bar{\eta}(\hat{G}; \hat{\theta})$ tal que

$$E_{G(x)} [\bar{\eta}(\hat{G}; \hat{\theta}) - \eta(G; \hat{\theta})] = 0,$$

entonces $\bar{\eta}(\hat{G}; \hat{\theta})$ es un estimador insesgado de la log-verosimilitud $\eta(G; \hat{\theta})$.

En resumen, los criterios de información fueron construidos como estimadores aproximadamente insesgados de la log-verosimilitud esperada $\eta(G; \hat{\theta}) \equiv E_{G(z)}[\ln f(Z|\hat{\theta})]$, o, equivalentemente, de la discrepancia de la Información de Kullback-Leibler entre la verdadera distribución $g(z)$ y un modelo estadístico $f(z|\hat{\theta})$ desde un punto de vista predictivo.

Un aspecto importante de los criterios de información es que éstos están asociados con el método de máxima verosimilitud, el cual, en modelos causales, está basado en el supuesto de que las variables observadas siguen una distribución normal multivariante. Teniendo esto en cuenta, Homburg (1991) realizó una comparativa (en situaciones en las cuales estaba justificada la suposición de que las variables seguían una distribución multinormal) entre el método de validación cruzada¹, que no está asociado con un método de estimación específico, y quizás los dos criterios de información más utilizados: el Criterio de Información de Akaike (*AIC*) y el Criterio de Información de Schwarz (*BIC*), los cuales serán definidos en los siguientes subapartados. Homburg (1991) encontró que las muestras debían contener al menos 200 casos cuando los criterios de información eran aplicados, y que la fiabilidad de los resultados mejoraba conforme aumentaba el tamaño muestral. Por contra, el método de validación cruzada fue encontrado apropiado sólo para tamaños muestrales medios, entre 300 y 500 sujetos, lo que en el contexto del estudio hablaba de un mejor rendimiento de los criterios basados en la información estadística.

2.1. Criterio de Información de Akaike

El criterio básico de esta línea de criterios basados en la información estadística, es el criterio de información de Akaike, que fue inicialmente desarrollado en series temporales (Akaike, 1974) y

¹Tradicionalmente, el método de validación cruzada ha sido un método ampliamente utilizado en comparación de modelos. El investigador divide la muestra en dos submuestras, estima los parámetros en una y realiza la validación en otra. Una descripción detallada puede encontrarse en el artículo de Charles I. Mosier *Problems and design of cross-validation*, publicado en *Educational and Psychological Measurement*, 11 (1951), 5-11.

que posteriormente ha sido propuesto para su utilización en el análisis factorial (ver, por ejemplo, Akaike, 1983, 1987; Bozdogan, 1987; Sclove, 1987). Su idea clave es la de penalizar un exceso de parámetros ajustados, algo que no hace el test asintótico de la chi-cuadrado. El *AIC* (*Akaike Information Criterion*) es un estimador muestral de $E[\ln f(X | \theta)]$, esperanza de la log-verosimilitud (o negantropía), que viene dado por la expresión general

$$AIC(k) = -2 \ln \mathcal{L}[\hat{\theta}(k)] + 2k, \quad (2.6)$$

en donde $\mathcal{L}[\theta(k)]$ es la función de verosimilitud de las observaciones, $\hat{\theta}(k)$ es la estimación máximo verosímil del vector de parámetros θ (en este caso, $\theta = (\mu, L, \Psi)$) y k es el número de parámetros independientes estimados dentro del modelo, mientras “ln” denota al logaritmo neperiano.

Para el modelo factorial ortogonal, $\Sigma = LL^T + \Psi$, la primera componente del *AIC* está dada por

$$-2 \ln \mathcal{L}(\hat{\mu}, \hat{L}, \hat{\Psi}) = np \ln(2\pi) + n \ln |\hat{L}\hat{L}^T + \hat{\Psi}| + np.$$

Teniendo en cuenta el número de parámetros, bajo el modelo factorial ortogonal, la matriz de cargas factoriales L tiene mp parámetros para ser estimados y la matriz de factores específicos, Ψ , tiene otros p parámetros para ser estimados, para un total de $(mp + p)$ parámetros. Sin embargo, la condición de unicidad $L^T \Psi^{-1} L \equiv \Delta$ supone $\frac{m(m+1)}{2}$ parámetros distintos y que Δ sea considerada diagonal. Por tanto, esta condición impone $\frac{m(m+1)}{2} - m = \frac{m(m-1)}{2}$ limitaciones adicionales al problema. Efectivamente, bajo el modelo factorial ortogonal $\Sigma = LL^T + \Psi$, Σ tiene

$$k = (mp + p) - \frac{1}{2}m(m-1)$$

parámetros libres.

Así, el *AIC* bajo este modelo, queda dado por

$$AIC(m) = np \ln(2\pi) + n \ln |\hat{L}\hat{L}^T + \hat{\Psi}| + np + 2[(mp + p) - \frac{1}{2}m(m-1)], \quad (2.7)$$

donde

- n = Tamaño muestral total.
- p = Número de variables originales.
- \hat{L} = Matriz de cargas factoriales estimada mediante máxima verosimilitud.
- $\hat{\Psi}$ = Matriz de factores específicos estimada mediante máxima verosimilitud.
- m = Número de factores, $m = 1, 2, \dots, M$.
- $\ln |\hat{L}\hat{L}^T + \hat{\Psi}|$ es el logaritmo neperiano del determinante de la matriz $\hat{L}\hat{L}^T + \hat{\Psi}$.

Recuérdese que, para que la realización de un análisis factorial tenga sentido, ha de tenerse que el número de parámetros libres a estimar dentro del modelo, $p + pm - \frac{m(m-1)}{2}$, sea menor que $\frac{p(p+1)}{2}$, el número de términos de la matriz de covarianzas. Haciendo uso de esta condición, se obtiene la expresión

$$(p + m) < (p - m)^2,$$

donde p es el número de variables originales y m el número de factores extraídos. Esta expresión es equivalente a

$$m < \frac{1}{2} [2p + 1 - \sqrt{8p + 1}],$$

obteniéndose así un límite superior para el número de factores a extraer, m . De esta forma, el número de factores, m , no puede exceder al mayor entero M tal que

$$m \leq M < \frac{1}{2} [2p + 1 - \sqrt{8p + 1}].$$

Akaike (1977) introdujo un principio de construcción de modelos estadísticos, el principio de maximización de la entropía, que considera cualquier actividad estadística como un intento de maximizar la entropía esperada de la estimación resultante de la distribución de una observación futura. El principio está caracterizado por la introducción del criterio de entropía y el punto de vista predictivo. La entropía de una distribución $f(\cdot)$ con respecto a otra distribución $g(\cdot)$ es definida como

$$B(f, g) = - \int \frac{f(x)}{g(x)} \ln \left[\frac{f(x)}{g(x)} \right] g(x) dx.$$

La entropía es un criterio natural del ajuste de $g(\cdot)$ a $f(\cdot)$. Cuanto mayor es la entropía, mejor es la aproximación de $f(\cdot)$ mediante $g(\cdot)$. Dado que la entropía puede ser expresada como

$$B(f, g) = E[\ln g(x)] - E[\ln f(x)],$$

un mayor valor de la log-verosimilitud esperada, $E[\ln g(x)]$, lleva aparejado un mayor valor de la entropía.

La idea básica subyacente al uso del *AIC* para la selección de modelos es la maximización de la log-verosimilitud esperada de un modelo determinado, mediante el método de máxima verosimilitud, lo que puede ser visto como la aplicación del principio de maximización de la entropía. La negantropía, $-B(f, g)$, no es otra cosa que la Información de Kullback-Leibler (1951), $I(f, g)$. Es por ello que este criterio es conocido como *criterio de información*, más que como *criterio de entropía*. El objetivo de la selección de modelos mediante el *AIC* es estimar la pérdida de información cuando la distribución de probabilidad f , asociada con el modelo verdadero, es aproximada mediante la distribución de probabilidad g , asociada con el modelo que va a ser evaluado. Una medida para la discrepancia entre el modelo real y el modelo aproximado viene dada por la cantidad $I(f, g)$, que es igual al valor negativo de la entropía (ver Bozdogan, 1987).

La ecuación (2.6), obtenida por Akaike para series temporales, puede ser adaptada al caso de la metodología máximo verosímil del análisis factorial exploratorio. En base al criterio *AIC*, es seleccionado el modelo con m factores tal que $AIC(m)$ alcance el valor mínimo de entre los modelos factoriales candidatos. En efecto, se utiliza la ecuación (2.7) para $m = 1$ y se va variando el número de factores posibles, seleccionándose el modelo factorial con el m tal que para dicho modelo se minimice $AIC(m)$.

En definitiva, el primer término del *AIC* puede ser interpretado como una medida de la bondad del ajuste, mientras el segundo término es una penalización, creciente conforme aumenta el número de parámetros, de acuerdo al Principio de Parsimonia. Posteriormente, se selecciona el modelo factorial adecuado, a partir del valor mínimo del *AIC*.

El *AIC* enfatiza la bondad del modelo. Como puntualizaba Takane (1987), el *AIC* no pretende identificar el modelo verdadero. Que un modelo sea el que mejor se ajusta a los datos, no quiere

decir que sea el modelo real o verdadero. Más bien, significa que el modelo es el mejor de entre los modelos candidatos, en el sentido de que proporciona la aproximación más cercana a la realidad o al verdadero modelo. El modelo que mejor se ajusta a los datos, podría cambiar en función del tamaño muestral, dado que, con un tamaño muestral mayor, los parámetros de un modelo pueden ser estimados de manera más fiable; por ejemplo, en el contexto del análisis factorial, una cuestión apropiada para plantearse podría ser, en lugar de cuál es el número correcto de factores, cuántos factores pueden ser extraídos adecuadamente dado el conjunto de datos.

Algunas de las ventajas del *AIC* que lo hacen tan utilizado en la práctica, son su simplicidad (no requiere acudir a ninguna tabla para observar el valor correspondiente) y facilidad para ser implementado, y el hecho de que no existe el problema de especificar subjetivamente un nivel de significación arbitrario para contrastar dos modelos.

2.1.1. Expresión de Akaike para el cálculo del *AIC* en Análisis Factorial

Con el objetivo de extender al análisis factorial este criterio concebido originalmente para series temporales, Akaike (1983, 1987) proporciona una sencilla expresión, basada en la relación entre el estadístico chi-cuadrado y el valor de *AIC*, de acuerdo a la propia definición de Akaike (1974).

La clave del procedimiento está en el comportamiento asintótico del estadístico del test basado en la razón de verosimilitudes. El estadístico chi-cuadrado es definido como

$$\chi^2 = (-2) \text{máx log } \mathcal{L}(H) - (-2) \text{máx log } \mathcal{L}(H_\infty),$$

donde $\text{máx log } \mathcal{L}(H)$ denota el máximo de la log-verosimilitud bajo la hipótesis H , y H_∞ denota el modelo saturado o sin restricciones.

Se tiene entonces

$$AIC(H) = (-2) \text{máx log } \mathcal{L}(H) + 2 \dim \theta,$$

donde $\dim \theta$ denota la dimensión del vector θ de parámetros desconocidos o parámetros libres estimados dentro del modelo, siendo $\dim \theta = p(m+1) - \frac{1}{2}m(m-1)$, con p el número de variables

originales y m el número de factores extraídos. Por otra parte,

$$AIC(H_\infty) = (-2) \text{máx log } \mathcal{L}(H_\infty) + \frac{p(p+1)}{2},$$

donde $\frac{p(p+1)}{2}$ es el número de parámetros libres estimados dentro del modelo factorial sin restricciones.

Se tiene que

$$\left[p(m+1) - \frac{1}{2}m(m-1) \right] - \left[\frac{p(p+1)}{2} \right] = -\frac{1}{2}((p-m)^2 - (p+m)).$$

Sea $d = \frac{1}{2}((p-m)^2 - (p+m))$,

$$\begin{aligned} AIC(H) - AIC(H_\infty) &= (-2) \text{máx log } \mathcal{L}(H) - (-2) \text{máx log } \mathcal{L}(H_\infty) - 2d = \\ &= \chi_d^2 - 2d. \end{aligned}$$

Akaike propone omitir el término $AIC(H_\infty)$, que es constante, y definir $AIC(H)$ de una forma más simple:

$$AIC(H) = \chi_d^2 - 2d,$$

donde $d = \frac{1}{2}((p-m)^2 - (p+m))$, siendo p el número de variables originales y m el de factores extraídos.

Sin embargo, a pesar de la sencillez de esta expresión y de la facilidad para su implementación, téngase en cuenta que la expresión

$$\chi^2 = (-2) \text{máx log } \mathcal{L}(H) - (-2) \text{máx log } \mathcal{L}(H_\infty)$$

es cierta, en general, para amplios tamaños muestrales (ver sección 1.3.1).

Dicho lo anterior, en esta Tesis se opta por calcular, y se implementan para ello los algoritmos correspondientes, el AIC en base al procedimiento teórico dado por Bozdogan y Ramírez (1987, 1988) y Sclove (1987).

2.1.2. Selección de modelos utilizando los pesos de Akaike

Wagenmakers y Farrell (2004) demuestran que los valores del AIC para cada modelo pueden ser transformados a las llamadas ponderaciones o pesos de Akaike (Akaike, 1978, 1979; Bozdogan, 1987; Burnham y Anderson, 2002), los cuales pueden ser interpretados como probabilidades condicionales para cada modelo.

En primer lugar se calcula, para cada modelo, la diferencia en el valor AIC con respecto al AIC del candidato a mejor modelo (el que minimiza el valor AIC), es decir,

$$\Delta_i(AIC) = AIC_i - \min(AIC), \quad (2.8)$$

donde i representa al correspondiente modelo.

Dado que el AIC es definido para ser un estimador asintóticamente insesgado de menos dos veces la log-verosimilitud de un modelo determinado mediante las estimaciones de máxima verosimilitud de los parámetros, Akaike (1978, 1979) sugiere que $\exp\left\{-\frac{1}{2}AIC\right\}$ será asintóticamente una definición razonable para la verosimilitud de un modelo especificado por los parámetros determinados mediante el método de máxima verosimilitud. De esta forma, a partir de las diferencias en el AIC dadas por (2.8), se puede obtener una estimación de la verosimilitud relativa al modelo M_i , puesto que

$$\mathcal{L}(M_i|\text{datos}) \propto \exp\left\{-\frac{1}{2}\Delta_i(AIC)\right\},$$

donde \propto es la notación que indica la proporcionalidad entre los dos extremos de la expresión.

Finalmente, las verosimilitudes relativas a cada modelo son normalizadas (es decir, divididas por la suma de las verosimilitudes de todos los modelos) para obtener los pesos de Akaike, $w_i(AIC)$ (ver Burnham y Anderson, 2002), cuyo valor viene dado por

$$w_i(AIC) = \frac{\exp\left\{-\frac{1}{2}\Delta_i(AIC)\right\}}{\sum_{k=1}^K \exp\left\{-\frac{1}{2}\Delta_k(AIC)\right\}}, \quad (2.9)$$

donde M_i , $i = 1, \dots, K$, son los modelos candidatos, de forma que $\sum_{i=1}^K w_i(AIC) = 1$.

Los pesos $w_i(AIC)$ dados por la expresión (2.9) pueden ser interpretados como la probabilidad de que M_i sea el mejor modelo (es decir, el que minimiza la discrepancia de Kullback-Leibler), dados los datos y el conjunto de modelos candidatos, siendo por tanto estos pesos medidas de la evidencia en favor de un modelo concreto (Burnham y Anderson, 2001).

El nivel de evidencia en favor de un modelo respecto a otro, es obtenido mediante la división de sus respectivos pesos de Akaike. La razón de evidencia de los pesos de Akaike de un modelo A sobre un modelo B puede ser calculada mediante la expresión (Wagenmakers y Farrell, 2004)

$$\frac{w_A(AIC)}{w_B(AIC)} = \exp(\ln \mathcal{L}_A - \ln \mathcal{L}_B + k_B - k_A) = \frac{\mathcal{L}_A}{\mathcal{L}_B} \exp(k_B - k_A),$$

donde \mathcal{L}_A y \mathcal{L}_B son, respectivamente las verosimilitudes de los modelos A y B , mientras que k_A y k_B son, respectivamente, el número de parámetros libres estimados dentro de los modelos A y B . Así, dados dos modelos, A y B , con $w_A(AIC) = r_1$ y $w_B(AIC) = r_2$, y $r_1 > r_2$ ², se dice que la evidencia de que el modelo A sea mejor que B es $\frac{r_1}{r_2}$ veces favorable al modelo A , en términos de la discrepancia de Kullback-Leibler.

El procedimiento de los pesos de Akaike (1978) fue aplicado, originariamente, a series temporales, pero, es extensible a otros ámbitos, como análisis de regresión o análisis factorial. En este último caso, sirva como ejemplo el trabajo de Schomaker y Heumann (2011), acerca de la elección del modelo factorial adecuado para explicar la dimensionalidad de la disciplina atlética de Decatlón.

2.2. Criterio de Información de Schwarz

Obsérvese que el término de penalización del AIC en (2.6), $2k$, no depende del tamaño de muestral n de la población considerada. Esto conduce al hecho de que un mismo número de parámetros comunes es seleccionado mediante el AIC , tanto para muestras pequeñas como para muestras grandes. Es por ello que el AIC no es un estimador consistente del número adecuado (o correcto) de factores comunes, aunque habría que decir también que la consistencia es una propiedad asintótica y cualquier problema real tendrá un tamaño muestral finito n .

²El mejor modelo de acuerdo al procedimiento de minimización del AIC , es también el que presenta un mayor peso. Es decir, dados dos modelos, A y B , si $AIC_A < AIC_B$, entonces $w_A(AIC) > w_B(AIC)$.

En el contexto de procedimientos basados en la verosimilitud, Schwarz (1978) sugirió que el *AIC* podría no ser asintóticamente justificable, y presentó un criterio de información alternativo a partir de un enfoque bayesiano, el *BIC* (*Bayesian Information Criterion*). Con este criterio, se penaliza el número de parámetros con $\ln n$, en lugar de 2. Así,

$$BIC(k) = -2 \ln \mathcal{L}[\hat{\theta}(k)] + (\ln n)k, \quad (2.10)$$

donde $L[\theta(k)]$ es la función de verosimilitud de las observaciones, $\hat{\theta}(k)$ es la estimación máximo verosímil del vector de parámetros $\theta = (\mu, L, \Psi)$ y k es el número de parámetros independientes estimados dentro del modelo, mientras que n es el tamaño muestral.

El criterio es obtenido analizando el comportamiento de la probabilidad *a posteriori* del j -ésimo modelo cuando n tiende a infinito, bajo la hipótesis de algunas distribuciones *a priori* de los parámetros.

$$\ln f(X|M_j) = \mathcal{L}_j(\hat{\theta}_j|X) + \ln P(\hat{\theta}_j|M_j) + \frac{k_j}{2} \ln(2\pi) - \frac{k_j}{2} \ln n + \frac{1}{2} \ln |R_j|,$$

donde

- $f(X|M_j)$ es la verosimilitud marginal de los datos para el modelo M_j .
- $\mathcal{L}_j(\hat{\theta}_j|X)$ es la verosimilitud del modelo M_j , donde $\hat{\theta}_j$ es el estimador máximo verosímil del parámetro θ_j .
- $P(\hat{\theta}_j|M_j)$ es la probabilidad *a priori* para los parámetros.
- k_j es el número de parámetros libres estimados dentro del modelo M_j .
- $|R_j|$ es el determinante de $R_j = n \cdot S_j$, donde S_j es la matriz de covarianzas del estimador $\hat{\theta}_j$.

Para un n suficientemente grande,

$$\ln f(X|M_j) \simeq \mathcal{L}_j(\hat{\theta}_j|X) - \frac{k_j}{2} \ln n.$$

Esta expresión fue obtenida por vez primera por Schwarz (1978), que propuso escoger el modelo que conduzca a un valor máximo de esta cantidad. Una forma equivalente de este criterio es calcular

para cada modelo la cantidad

$$BIC(M_j) = -2 \ln \mathcal{L}_j(\hat{\theta}_j|X) + k_j \cdot \ln n,$$

y seleccionar aquel modelo para el cual esta cantidad sea mínima. De esta manera, este criterio pondera la desviación del modelo, medida por $-2 \ln \mathcal{L}_j(\hat{\theta}_j|X)$, con el número de parámetros. Si introducimos más parámetros en el modelo, mejorará el ajuste, con lo que aumentará el soporte o disminuirá la desviación, y este efecto quedará compensado por el aumento del número de parámetros que aparece en $k_j \cdot \ln n$.

De esta forma, los términos principales del desarrollo asintótico conforman un criterio válido para muestras grandes, más allá del contexto bayesiano, dado que ellos no dependen de la distribución a priori.

Este criterio, a diferencia del *AIC*, considera el número n de observaciones en el término de penalización, y es además menos favorable a la inclusión de factores. El valor óptimo de k es calculado minimizando $BIC(k)$ en la ecuación (2.10). Como $k = (mp + p) - \frac{1}{2}m(m - 1)$, el valor del *BIC* asociado a un modelo con m factores es

$$BIC(m) = np \ln(2\pi) + n \ln |\hat{L}\hat{L}^T + \hat{\Psi}| + np + (\ln n)[(mp + p) - \frac{1}{2}m(m - 1)], \quad (2.11)$$

donde

- n = Tamaño muestral total.
- p = Número de variables originales.
- \hat{L} = Matriz de cargas factoriales estimada mediante máxima verosimilitud.
- $\hat{\Psi}$ = Matriz de factores específicos estimada mediante máxima verosimilitud.
- m = Número de factores, $m = 1, 2, \dots, M$.
- $\ln |\hat{L}\hat{L}^T + \hat{\Psi}|$ es el logaritmo neperiano del determinante de la matriz $\hat{L}\hat{L}^T + \hat{\Psi}$.

El procedimiento de selección de un modelo factorial en base al *BIC* será el siguiente: se calcula el *BIC* para $m = 1, 2, \dots, M$, y se selecciona el modelo con un número de factores, m , tal que se minimice el valor del *BIC*.

2.2.1. Una expresión alternativa para el *AIC* y el *BIC*

Basilewsky (1994, pág. 388) da una expresión alternativa para los criterios de información de Akaike y Schwarz. Para m factores comunes, se tiene

$$\mathcal{L}(m) = \frac{n}{2} \cdot \left(\sum_{i=m+1}^p \ln \hat{\theta}_i \right),$$

donde $\hat{\theta}_i = (\hat{\eta}_i + 1)$, $i = m + 1, \dots, p$, representa las $(p - m)$ raíces características más pequeñas de $\hat{\Psi}^{-1/2} S \hat{\Psi}^{-1/2}$ (ver Teorema 1.3). Teniendo en cuenta el número de parámetros libres del modelo, $k = (mp + p) - \frac{1}{2}m(m - 1)$, donde p es el número de variables observadas y m el número de factores, la ecuación (2.6) puede ser expresada como

$$AIC(m) = (-2) \cdot \left(\frac{n}{2} \cdot \left(\sum_{i=m+1}^p \ln \hat{\theta}_i \right) \right) + [2p(m + 1) - m(m - 1)].$$

El *BIC* puede ser expresado como

$$BIC(m) = \left(-\frac{n}{2} \cdot \left(\sum_{i=m+1}^p \ln \hat{\theta}_i \right) \right) + \frac{k}{2} \cdot \ln n.$$

Nótese que k no es el número de grados de libertad como en el caso del estadístico chi-cuadrado sucedía con d , sino que corresponde al número de parámetros libres del modelo. Téngase en cuenta que esta forma de selección del modelo factorial adecuado, a partir de los criterios de información estadística, no es en realidad un test estadístico.

2.2.2. Algunas diferencias a considerar entre el *AIC* y el *BIC*

Aparte de las diferencias propias de la definición entre el *AIC* y el *BIC*, que son probablemente los criterios comúnmente más usados y a partir de los cuales se derivan otros criterios de información

existentes, han de apuntarse otros motivos diferenciales y algunos aspectos a tener en cuenta a la hora de efectuar una comparación entre ambos criterios.

Cualitativamente, el procedimiento de Schwarz y el de Akaike dan *una formulación matemática del principio de parsimonia en la construcción de modelos*. Cuantitativamente, dado que el procedimiento de Schwarz difiere del de Akaike sólo en que la dimensión es multiplicada por $\ln n$, el *BIC* se inclina más que el *AIC* hacia modelos de dimensión más baja cuando el número de observaciones es $n \geq 8$. Para un número elevado de observaciones, existen diferencias notables entre el *AIC* y el *BIC* (Schwarz, 1978).

Otra diferencia entre estos dos criterios, el *AIC* y el *BIC*, que son además los comúnmente más usados, se explica por su distinto objetivo (Peña, 2002). El criterio *BIC* trata de seleccionar el modelo correcto, con máxima probabilidad a posteriori, y puede demostrarse que es un criterio consistente, de manera que la probabilidad de seleccionar el modelo correcto tiende a 1 cuando aumenta el tamaño muestral. El criterio *AIC* no pretende seleccionar el modelo correcto, ya que admite que el modelo verdadero puede no estar entre los estimados, y trata de obtener el modelo que proporcione mejores predicciones entre los existentes. Puede demostrarse que, suponiendo que el modelo verdadero puede aproximarse arbitrariamente bien con los estimados, al aumentar el tamaño muestral, el criterio *AIC* es eficiente, escogiendo el modelo que proporciona (en promedio) mejores predicciones. Sin embargo, en muestras pequeñas o medianas, el criterio *AIC* tiende a seleccionar modelos con más parámetros de los necesarios (Peña, 2002).

En definitiva, el *BIC* asume que el modelo real está incluido en el conjunto de modelos candidatos, y mide el grado de certeza de que un cierto modelo sea el verdadero modelo generador de los datos. El *AIC* no asume que ninguno de los modelos candidatos sea necesariamente el verdadero, calculando para cada modelo la información o discrepancia de Kullback-Leibler, que es una medida de la distancia entre la densidad de probabilidad generada por el modelo y la realidad.

Stone (1979) desarrolló un estudio comparativo entre el *BIC* y el *AIC*, observando que dicha comparación es sensible al tipo de análisis asintótico adoptado, y cuestionó el realismo de investigaciones asintóticas que mantienen el modelo fijado, supuestamente verdadero, conforme $n \rightarrow \infty$. En la práctica, la definición de modelo verdadero sería más refinada cuando $n \rightarrow \infty$, mediante la introducción

de modelos más complejos que incorporen un mayor número de parámetros.

Akaike (1981) sugirió que el uso del procedimiento de Schwarz en situaciones prácticas, con muestras finitas, podría resultar problemático, y que el uso de ese tipo de procedimientos sería sólo factible bajo la convicción de las distribuciones *a priori* asumidas de los parámetros. Eso sólo es posible cuando se tienen claramente definidas las distribuciones propias *a priori*, una situación donde la aplicación sencilla de los procedimientos bayesianos es todo lo necesario para resolver el problema de selección de modelos. Schwarz no desarrolló caracterización alguna de su procedimiento para un tamaño muestral finito. Así, Akaike (1981) argumentó que cualquiera que pretenda utilizar dicho procedimiento debe desarrollar su propia justificación para cada aplicación particular. Dicha consideración sería también aplicable a aquellos criterios de información pertenecientes a la familia *BIC*, o, lo que es lo mismo, aquellos criterios cuya definición está basada en la estructura del Criterio de Información de Schwarz, algunos de los cuales serán expuestos a continuación.

Wagenmakers y Farrell (2004) puntualizan que realizar una comparación formal en términos de rendimiento entre *AIC* y *BIC* es muy difícil, a causa de que ambos criterios direccionan cuestiones diferentes. La mayoría de simulaciones que demuestran que el *BIC* se comporta mejor que el *AIC*, asumen que el modelo verdadero está en el conjunto de candidatos y es de dimensionalidad baja. En contraste con esto, la mayoría de simulaciones que favorecen al *AIC* sobre el *BIC*, asumen que dicho modelo es de dimensionalidad infinita, de ahí que el modelo verdadero no esté en el conjunto de candidatos.

2.2.3. Extensión de los pesos de Akaike al caso de la selección de modelos mediante el *BIC*

Burnham y Anderson (2002) presentan una extensión del procedimiento propuesto por Akaike (1978), basado en ponderaciones o pesos, para la selección del modelo adecuado. En el caso de que el criterio de información empleado sea el *BIC*, los valores dados por la expresión (2.11) pueden ser transformados en pesos del modelo o *pesos de Schwarz* (Wagenmakers y Farrell, 2004). Dichos pesos pueden ser obtenidos reemplazando en la expresión (2.9) los valores del *AIC* por los valores del *BIC*. Pueden ser interpretados como probabilidades del modelo, dados los datos, el conjunto

de modelos y las probabilidades a priori $\left(\frac{1}{K}\right)$ de cada modelo, siendo K el número de modelos candidatos. Se definen las diferencias en el valor del BIC , para cada posible modelo i , como

$$\Delta_i BIC = BIC_i - \text{mín}(BIC),$$

donde $\text{mín}(BIC)$ es el valor mínimo para el BIC , considerando los K modelos candidatos. En base al BIC , la probabilidad a posteriori de que un modelo g_i sea el modelo (en este caso, modelo factorial) adecuado, viene dada por

$$P(g_i) = \frac{\exp\left\{-\frac{1}{2}\Delta_i(BIC)\right\}}{\sum_{k=1}^K \exp\left\{-\frac{1}{2}\Delta_k(BIC)\right\}}.$$

Así, se puede definir el peso de Schwarz asociado a un valor BIC como

$$w_i(BIC) = \frac{\exp\left\{-\frac{1}{2}\Delta_i(BIC)\right\}}{\sum_{k=1}^K \exp\left\{-\frac{1}{2}\Delta_k(BIC)\right\}},$$

Análogamente al caso de los pesos de Akaike, el modelo que tenga un menor valor BIC asociado, será el que tenga una mayor ponderación. Así, estas ponderaciones permiten establecer una evidencia de adecuación de un modelo respecto a otro. La razón de evidencia en favor de un modelo A sobre un modelo B (siendo A y B modelos incluidos en el conjunto de modelos candidatos) puede ser calculada de la siguiente forma:

$$\frac{w_A(BIC)}{w_B(BIC)} = \frac{\mathcal{L}_A}{\mathcal{L}_B} \cdot n^{\frac{1}{2}(k_B - k_A)},$$

donde

- n es el tamaño muestral.
- \mathcal{L}_A y \mathcal{L}_B , las verosimilitudes, respectivamente, de los modelos A y B .
- k_A y k_B , respectivamente, el número de parámetros libres estimados dentro de los modelos A y B .

2.2.4. Pesos de Akaike vistos como probabilidades a posteriori del modelo

Sea p_i la probabilidad a priori asociada al modelo g_i . Entonces, la probabilidad a posteriori del modelo viene dada por la expresión

$$P(g_i|\text{datos}) = \frac{\exp\left\{-\frac{1}{2}\Delta_i(BIC)\right\} p_i}{\sum_{k=1}^K \exp\left\{-\frac{1}{2}\Delta_k(BIC)\right\} p_k}.$$

Para obtener los pesos de Akaike, se utiliza el modelo a priori

$$p_i = B \cdot \exp\left(\frac{1}{2}\Delta_i BIC\right) \cdot \exp\left(-\frac{1}{2}\Delta_i AIC\right),$$

donde B es un término constante (ver Burnham y Anderson, 2002; pág. 303). Así,

$$\exp\left(-\frac{1}{2}\Delta_i BIC\right) \cdot \exp\left(\frac{1}{2}\Delta_i BIC\right) \cdot \exp\left(-\frac{1}{2}\Delta_i AIC\right) = \exp\left(-\frac{1}{2}\Delta_i AIC\right).$$

Teniendo en cuenta esta distribución de probabilidad a priori de los modelos, se obtiene

$$P(g_i|\text{datos}) = \frac{\exp\left\{-\frac{1}{2}\Delta_i(BIC)\right\} p_i}{\sum_{k=1}^K \exp\left\{-\frac{1}{2}\Delta_k(BIC)\right\} p_k} = \frac{\exp\left\{-\frac{1}{2}\Delta_i(AIC)\right\}}{\sum_{k=1}^K \exp\left\{-\frac{1}{2}\Delta_k(AIC)\right\}} = w_i,$$

que es el peso de Akaike para el modelo g_i .

2.3. Criterio de Información de Hannan y Quinn

El criterio de información de Hannan y Quinn, notado por $HQIC$ (*Hannan-Quinn Information Criterion*), puede ser considerado una variante del BIC con una pequeña penalización de la magnitud del tamaño muestral. Así, Hannan y Quinn (1979) sugirieron este criterio de información, inicialmente para seleccionar el orden de una autorregresión.

$$HQIC(k) = -2 \ln \mathcal{L}[\hat{\theta}(k)] + 2 \ln((\ln n))k, \quad (2.12)$$

donde $L[\theta(k)]$ representa la función de verosimilitud de las observaciones, $\hat{\theta}(k)$ es la estimación

máximo verosímil del vector de parámetros $\theta = (\mu, L, \Psi)$ y k es el número de parámetros independientes estimados dentro del modelo, mientras que n es el tamaño muestral. El valor óptimo de k es calculado minimizando $HQIC(k)$ en la ecuación (2.12). Así, en el modelo factorial con m factores, el $HQIC$ viene dado por

$$HQIC(m) = np \ln(2\pi) + n \ln |\hat{L}\hat{L}^T + \hat{\Psi}| + np + (2 \ln(\ln n))[(mp + p) - \frac{1}{2}m(m - 1)],$$

donde

- n = Tamaño muestral total.
- p = Número de variables originales.
- \hat{L} = Matriz de cargas factoriales estimada mediante máxima verosimilitud.
- $\hat{\Psi}$ = Matriz de factores específicos estimada mediante máxima verosimilitud.
- m = Número de factores, $m = 1, 2, \dots, M$.
- $\ln |\hat{L}\hat{L}^T + \hat{\Psi}|$ es el logaritmo neperiano del determinante de la matriz $\hat{L}\hat{L}^T + \hat{\Psi}$.

Se selecciona el modelo con m factores, $m = 1, \dots, M$, de forma que se minimice el valor de $HQIC(m)$.

2.4. Criterio de Información de Bozdogan

Woodrooffe (1982) mostró que el AIC no es consistente, es decir, que en general el AIC no seleccionará el modelo correcto cuando n tienda a infinito. Bozdogan (1987), utilizando argumentos similares, derivó una versión consistente del AIC ($CAIC$) a partir de la medida de información de Kullback-Leibler; esta versión hace al AIC asintóticamente consistente y penaliza el exceso de parámetros de forma rigurosa, con el fin de seleccionar el modelo más simple de entre los modelos adecuados. Bozdogan y Ramírez (1987) subrayaban el hecho de que, introduciendo procedimientos de selección de modelos como el AIC y el $CAIC$, no es necesario llevar a cabo los clásicos tests de hipótesis para determinar el número adecuado de factores en un análisis factorial: no se precisa elegir

subjetivamente el número de factores basado en un nivel de significación arbitrariamente elegido. El *CAIC* es definido de la siguiente forma:

$$CAIC(m) = np \ln(2\pi) + n \ln |\hat{L}\hat{L}^T + \hat{\Psi}| + np + (\ln(n) + 1)[(mp + p) - \frac{1}{2}m(m - 1)],$$

donde

- n = Tamaño muestral total.
- p = Número de variables originales.
- \hat{L} = Matriz de cargas factoriales estimada mediante máxima verosimilitud.
- $\hat{\Psi}$ = Matriz de factores específicos estimada mediante máxima verosimilitud.
- m = Número de factores, $m = 1, 2, \dots, M$.
- $\ln |\hat{L}\hat{L}^T + \hat{\Psi}|$ es el logaritmo neperiano del determinante de la matriz $\hat{L}\hat{L}^T + \hat{\Psi}$.

Hay que recordar que, para un número fijo de p variables originales, el número de factores, m , no puede exceder al mayor entero M tal que

$$m \leq M < \frac{1}{2} [2p + 1 - \sqrt{8p + 1}],$$

o, dicho de otra forma, para todo número de factores m se ha de verificar la condición

$$(p - m)^2 > p + m.$$

El *CAIC* favorece la elección de modelos de menor dimensión. El *AIC* elige modelos parsimoniosos, pero no es tan riguroso como el *CAIC*. En este sentido, Bozdogan y Ramírez (1988) proponían la siguiente estrategia combinada: utilizar el *AIC* para determinar el límite superior, y el *CAIC* para determinar el límite inferior del número de factores a seleccionar, lo que proporciona un intervalo o rango de elección del número de factores.

2.5. Otros criterios de información

Los criterios de información utilizados en selección de modelos, tienen en cuenta simultáneamente la bondad de ajuste (verosimilitud) de un modelo y el número de parámetros empleados para lograr ese ajuste. El criterio toma la forma de una función de verosimilitud penalizada, es decir, la log-verosimilitud con signo negativo más un término de penalización, que aumenta conforme lo hace el número de parámetros.

El parentesco entre los distintos criterios de información puede ser visto en la ecuación de Sclove (1987)

$$-2 \cdot \ln \mathcal{L}(\theta_j) + k_j \cdot A(n) + B(j, n), \quad (2.13)$$

donde

- n es el tamaño muestral total.
- $\ln \mathcal{L}(\theta_j)$ es el logaritmo neperiano de la verosimilitud maximizado bajo el modelo j .
- k_j es el número de parámetros libres estimados dentro del modelo j .
- $A(n)$ es el coste de ajustar un parámetro adicional.
- $B(j, n)$ es un término adicional que depende del criterio y del modelo j .

Yang y Yang (2007) modificaron la ecuación (2.13) a la forma

$$-2 \cdot \ln \mathcal{L}(\theta_j) + k_j \cdot (A(n) + C) + B(j, n),$$

siendo C un término constante, para incluir en ella algunos de los principales criterios encontrados en la literatura científica, clasificándolos posteriormente en función de su pertenencia a la familia *AIC* o a la familia *BIC* (Tabla 2.1).

| | | $A(n)$ | C | $B(j, n)$ |
|---------------|---------------------------------|----------------|--------------|-----------------------------------|
| Familia AIC | AIC (Akaike, 1974) | 0 | 2 | 0 |
| | $AIC3$ (Andrews y Currim, 2003) | 0 | 3 | 0 |
| | $CAIC$ (Bozdogan, 1987) | $\ln n$ | 1 | 0 |
| | $CAIC^*$ (Yang y Yang, 2007) | $\ln(n+2)$ | $1-\ln(24)$ | 0 |
| | AIC_c (Sugiura, 1978) | 0 | 2 | $\frac{2(k_j+1)(k_j+2)}{n-k_j-2}$ |
| Familia BIC | BIC (Schwarz, 1978) | $\ln n$ | 0 | 0 |
| | $DBIC$ (Draper, 1995) | $\ln n$ | $-\ln(2\pi)$ | 0 |
| | BIC^* (Yang y Yang, 2007) | $\ln(n+2)$ | $-\ln(24)$ | 0 |
| | RIC (Rissanen, 1978) | $\ln(n+2)$ | $-\ln(24)$ | $2\ln(k_j+1)$ |
| | $HQIC$ (Hannan y Quinn, 1979) | $2\ln(\ln(n))$ | 0 | 0 |

Tabla 2.1. Criterios de Información expresados en la forma $-2 \cdot \ln \mathcal{L}(\theta_j) + k_j \cdot (A(n) + C) + B(j, n)$.

Hurvich y Tsai (1989) utilizan una corrección al AIC destinada a modelos de regresión y de series temporales, el criterio AIC corregido, AIC_c , que fue originalmente propuesto por Sugiura (1978) con vistas a una reducción del sesgo del AIC . Hurvich y Tsai (1989, 1991) encontraron que el AIC_c no sólo provocaba una reducción considerable del sesgo, sino que también mejoraba altamente la selección del modelo adecuado en tamaños muestrales pequeños. El AIC_c es de uso particular cuando el tamaño muestral es pequeño, o cuando el número de parámetros ajustados es una fracción entre moderada y grande del tamaño muestral. Más recientemente, Andrews y Currim (2003) consideraron un factor de penalización igual a 3 en lugar del tradicional de 2, y promovieron una nueva variante del AIC ($AIC3$). Dicha variante ya fue sugerida por Sclove (1987), añadiendo que el AIC favorecía a modelos con un mayor número de parámetros, de tal forma que si un modelo con un número elevado de parámetros fuese mejor de acuerdo al $AIC3$, entonces también lo sería de acuerdo al AIC ; además, si un modelo con pocos parámetros fuese mejor respecto al AIC , entonces también lo sería conforme al $AIC3$.

En lo que al AIC_c se refiere, su uso en el análisis factorial habría que llevarlo a cabo con cierta cautela. Es recomendable utilizarlo sólo cuando se verifique que $n > k_j + 2$, donde n es el tamaño

muestral y k_j el número de parámetros libres estimados dentro del modelo j . Concretamente, en el caso $n = k_j + 2$, el término adicional $B(j, n) = \frac{2(k_j+1)(k_j+2)}{n-k_j-2}$ no sería finito. Además, para tamaños muestrales n tales que $n < k_j + 2$, el término adicional puede cambiar de signo y provocar fluctuaciones considerables en el valor del criterio, que hagan inviable extraer conclusiones. Es por ello que, si se hace uso en análisis factorial del criterio AIC_c que se puede escribir en función del AIC como

$$AIC_c = AIC + \frac{2(k_j + 1)(k_j + 2)}{n - k_j - 2},$$

sea aconsejable tener, para cada uno de los modelos candidatos, que $n > k_j + 2$. Por otra parte, algunos autores (Burnham y Anderson, 2002, pág. 445) recomiendan el uso de esta corrección al AIC cuando además se cumple la condición $\frac{n}{k_j} < 40$. En caso contrario, para tamaños muestrales n elevados, el valor de AIC_c , por definición, es muy similar al de AIC .

Draper (1995) sugirió que el BIC fuese modificado a

$$DBIC = -2 \cdot \ln \mathcal{L}(\theta_j) + k_j \cdot (\ln n - \ln(2\pi)).$$

Aunque el término añadido al $DBIC$, $-k_j \cdot 2 \ln(2\pi)$ no es asintóticamente importante cuando $n \rightarrow \infty$, podría tener un pequeño efecto, no demasiado preocupante, para tamaños muestrales moderadamente bajos. Draper (1995) concluyó que el $DBIC$ mejoraba el comportamiento del modelo seleccionado en problemas reales, y consideró que el término $-k_j \cdot 2 \ln(2\pi)$ debía ser incluido en el BIC , a causa de la mejora en el rendimiento que la inclusión de este término proporcionaría para muestras finitas.

Hay que mencionar también las modificaciones realizadas por Yang y Yang (2007) sobre los criterios $CAIC$ y BIC (respectivamente, $CAIC^*$ y BIC^*), proponiendo un ajuste de acuerdo al tamaño muestral, de forma que n fuese ajustado por $\frac{n+2}{24}$ para modelos con tamaños muestrales limitados. Dicho ajuste fue inicialmente propuesto por Rissanen (1978) para su criterio, RIC , el cual incluye un término adicional específico para cada modelo, $B(j, n) = 2 \ln(k_j + 1)$, donde k_j es el número de parámetros libres estimados dentro del modelo j . Así, en un modelo con m factores, el valor del

criterio de información de Rissanen, RIC , viene dado por

$$RIC(m) = np \ln(2\pi) + n \ln |\hat{L}\hat{L}^T + \hat{\Psi}| + np + \left(\ln \left(\frac{n+2}{24} \right) \right) [(mp+p) - \frac{1}{2}m(m-1)] + 2 \ln((mp+p) - \frac{1}{2}m(m-1) + 1),$$

donde

- n = Tamaño muestral total.
- p = Número de variables originales.
- \hat{L} = Matriz de cargas factoriales estimada mediante máxima verosimilitud.
- $\hat{\Psi}$ = Matriz de factores específicos estimada mediante máxima verosimilitud.
- m = Número de factores, $m = 1, 2, \dots, M$.
- $\ln |\hat{L}\hat{L}^T + \hat{\Psi}|$ es el logaritmo neperiano del determinante de la matriz $\hat{L}\hat{L}^T + \hat{\Psi}$.

En el contexto del análisis factorial, dado el número de factores, m , el término correspondiente a la primera componente de cada uno de los criterios de información expuestos en la Tabla 2.1, que es igual a menos dos veces el logaritmo de la función de verosimilitud maximizada, viene dado por la expresión

$$-2 \ln \mathcal{L}_m(\hat{\mu}, \hat{L}, \hat{\Psi}) = np \ln(2\pi) + n \ln |\hat{L}\hat{L}^T + \hat{\Psi}| + np, \quad (2.14)$$

donde

- n = Tamaño muestral total.
- p = Número de variables originales.
- \hat{L} = Matriz de cargas factoriales estimada mediante máxima verosimilitud.
- $\hat{\Psi}$ = Matriz de factores específicos estimada mediante máxima verosimilitud.
- m = Número de factores, $m = 1, 2, \dots, M$.

- $\ln |\hat{L}\hat{L}^T + \hat{\Psi}|$ es el logaritmo neperiano del determinante de la matriz $\hat{L}\hat{L}^T + \hat{\Psi}$.

Conforme aumenta el número de factores, lo hace también el término $-2 \ln \mathcal{L}_m(\hat{\mu}, \hat{L}, \hat{\Psi})$ dado en (2.14), donde $\mathcal{L}_m(\hat{\mu}, \hat{L}, \hat{\Psi})$ es la función de verosimilitud bajo el modelo factorial con m factores. La expresión (2.14) es una medida del mal ajuste, inadecuación o sesgo, cuando son utilizados los estimadores de máxima verosimilitud de los parámetros del modelo. Para compensar el sesgo en (2.14), los criterios de información citados tienen una segunda componente, que depende del número de parámetros y/o del tamaño muestral, $k_j \cdot (A(n) + C) + B(j, n)$, donde k_j es el número de parámetros libres estimados dentro del modelo j , $A(n) + C$ es el coste de ajustar un parámetro adicional (siendo C un término constante) y $B(j, n)$ es un término adicional que depende del criterio y el modelo j . Esta segunda componente es llamada la componente de penalización o medida de complejidad; su misión es equilibrar el riesgo debido al sesgo cuando es seleccionada una dimensionalidad más baja, y el riesgo debido al aumento de varianza cuando una dimensionalidad más alta es seleccionada.

2.5.1. Generalización de los pesos de Akaike a otros criterios de información

Buckland, Burnham y Augustin (1997) definen el procedimiento de los pesos de Akaike, dado en la sección 2.1.2, para criterios de información dados en la forma

$$I = -2 \ln \mathcal{L}(\theta_j) + q,$$

donde $\ln \mathcal{L}(\theta_j)$ es el logaritmo neperiano de la verosimilitud maximizado bajo el modelo j y q es una penalización que es una función dependiente del número de parámetros y/o el número de observaciones.

Supuesto que hay K modelos candidatos a ser el modelo adecuado, se define para el modelo k , $k = 1, \dots, K$, $I_k = -2 \ln \mathcal{L}(\theta_k) + q_k$. El modelo para el que se obtenga el valor I_k más pequeño, será, de acuerdo al procedimiento basado en el uso de criterios de información, el más adecuado. Si se

compara el modelo i con el modelo j , se tiene

$$\frac{\mathcal{L}_i \exp\left(-\frac{q_i}{2}\right)}{\mathcal{L}_j \exp\left(-\frac{q_j}{2}\right)} = \frac{\exp\left(-\frac{I_i}{2}\right)}{\exp\left(-\frac{I_j}{2}\right)}.$$

El peso asociado al modelo k es:

$$w_k = \frac{\exp\left(-\frac{I_k}{2}\right)}{\sum_{i=1}^K \exp\left(-\frac{I_i}{2}\right)}, \quad k = 1, \dots, K.$$

De esta forma, y en el caso del estudio que nos ocupa, el proceso de selección de modelos utilizando los pesos de Akaike, puede ser generalizado a criterios de información, IC , dados en la forma $-2 \cdot \ln \mathcal{L}(\theta_j) + k_j \cdot (A(n) + C) + B(j, n)$, donde

- n es el tamaño muestral total.
- $\ln \mathcal{L}(\theta_j)$ es el logaritmo neperiano de la verosimilitud maximizado bajo el modelo j .
- k_j es el número de parámetros libres estimados dentro del modelo j . En el caso del análisis factorial, recuérdese que $k_j = (mp + p) - \frac{1}{2}m(m - 1)$, donde m es el número de factores extraídos para el modelo j y p el número de variables originales.
- $A(n) + C$, siendo $A(n)$ una función de n y C un término constante, es el coste de ajustar un parámetro adicional.
- $B(j, n)$ es un término adicional que depende del criterio y del modelo j .

Así, la expresión dada en (2.9) puede extenderse, en general, al criterio IC correspondiente, obteniéndose

$$w_i(IC) = \frac{\exp\left\{-\frac{1}{2}\Delta_i(IC)\right\}}{\sum_{k=1}^K \exp\left\{-\frac{1}{2}\Delta_k(IC)\right\}}, \quad (2.15)$$

donde M_i , $i = 1, \dots, K$, son los modelos candidatos y $\Delta_i(IC)$ es una transformación³ del valor IC_i , igual a la diferencia entre el valor IC asociado al modelo M_i y el mínimo valor IC encontrado entre

³Esta transformación es útil a la hora de realizar los cálculos correspondientes. Cuando se manejan valores elevados para el criterio de información IC correspondiente, el valor $\exp\left\{-\frac{1}{2}IC_i\right\}$ puede ser redondeado a 0 por algunos procesadores de cálculo. Este eventual problema se soluciona mediante la aplicación de la expresión (2.15).

los modelos candidatos, es decir,

$$\Delta_i(IC) = IC_i - \text{mín}(IC).$$

El valor dado por (2.15) puede ser interpretado, por tanto, análogamente al caso de los pesos de Akaike, como la evidencia en favor de un modelo determinado, en base al criterio de información IC . De hecho, Burnham y Anderson (2002, pág. 297) realizaban la extensión del procedimiento de los pesos de Akaike al otro criterio de información mayoritariamente utilizado, el BIC , como se trata en la sección 2.2.3. Dicho procedimiento es extensible a los criterios de información reseñados en este trabajo.

Mediante la metodología clásica del contraste de hipótesis para elegir el modelo factorial adecuado (test asintótico de Bartlett), no se puede conocer a ciencia cierta cómo mejora un modelo a otro. A partir de los pesos, asociados a los criterios de información dados en la forma $-2 \cdot \ln \mathcal{L}(\theta_j) + k_j \cdot (A(n) + C) + B(j, n)$, se pueden obtener medidas de evidencias de acierto en la elección de un modelo, o de cómo un modelo mejora en rendimiento (en base al criterio de información utilizado) a otro. Así, dados dos modelos, A y B , con $w_A(IC) = r_1$ y $w_B(IC) = r_2$, y $r_1 > r_2$, se dice que la evidencia de que el modelo A sea más adecuado que el modelo B es $\frac{r_1}{r_2}$ veces favorable al modelo A .

2.6. Algunas comparaciones efectuadas entre criterios basados en la información estadística

A continuación, se resumen algunas de las comparaciones que se pueden encontrar en la literatura científica, realizadas entre criterios basados en la información estadística, en diversos campos, que tienen en común la aplicación de procedimientos basados en el método de máxima verosimilitud. Dichas comparaciones tienen como objetivo evaluar el rendimiento de algunos criterios de información, bajo determinadas circunstancias (como puede ser por ejemplo el número de observaciones).

Así, Yang y Yang (2007) evaluaron el rendimiento de los criterios que aparecen en la Tabla 2.1 en la separación de clases latentes, concluyendo que la mayoría de criterios mejoraba su fiabilidad conforme aumentaba el tamaño muestral, sin que dicha mejoría se apreciase para el AIC y el AIC_c ,

los cuales incluso veían disminuida su proporción de aciertos cuando el tamaño muestral se hacía más grande. Por otra parte, observaron que las modificaciones basadas en el tamaño muestral, $CAIC^*$ y BIC^* , mejoraban considerablemente el rendimiento del $CAIC$ y BIC originales.

Costa (1996), en un estudio de simulación que compara diversos métodos para elegir el número de factores en un modelo factorial, evaluó el comportamiento de los criterios AIC , BIC y $HQIC$ (y diversas modificaciones para AIC y $HQIC$) en función del número de observaciones, llegando a la conclusión de que las modificaciones no mejoraban a los originales, y observando además que, de los 3 criterios, era en general el AIC el que presentaba unos mejores resultados. Por otro lado, el BIC infraestimaba sensiblemente el número de factores, particularmente cuando el número de observaciones era muy grande. Cuando el número de observaciones aumentaba, el criterio de Schwarz no sobreestimaba el número de factores, como sí hacían AIC y $HQIC$.

Rust, Simester, Brodie y Nilikant (1995), en el contexto de modelos de regresión lineal, investigaron el comportamiento de una serie de criterios de selección de modelos, destacando entre ellos el AIC y el BIC . Tanto uno como otro se comportaron adecuadamente para seleccionar el mejor modelo, pero parecía ser el criterio de Schwarz el más adecuado, atendiendo a la consistencia. Concluyeron además que el hecho de utilizar más de un criterio de selección de modelos no tiene por qué ser una garantía, y puede resultar innecesario, dado que todos los criterios de selección de modelos que ofrecen un rendimiento óptimo están altamente intercorrelacionados.

Ichikawa (1988) consideró que el procedimiento de minimizar el AIC para seleccionar el modelo adecuado requiere algunas modificaciones en el caso del análisis factorial, dado que a menudo el proceso iterativo no converge (e incluso en el caso de que sea convergente la solución puede ser impropia). Para ello, propuso seleccionar el modelo adecuado en base al valor mínimo del AIC , pero considerando sólo soluciones propias (una solución es llamada impropia si no es definida positiva). Los resultados de la simulación mediante el método de Monte Carlo, indicaron que la capacidad del AIC para seleccionar el modelo correcto aumentaba conforme lo hacía el tamaño muestral. Sin embargo, para tamaños muestrales mayores, el AIC exhibía una ligera tendencia hacia la selección de modelos sobreajustados.

Markon y Krueger (2004), en el contexto de los modelos de comportamiento genético, y empleando

modelos de ecuaciones estructurales, concluyeron que el *AIC* resultaba más adecuado en muestras pequeñas que el *BIC* y el *DBIC*, mientras que estos dos últimos eran preferibles en muestras grandes, especialmente en casos donde la distribución observada no era normal multivariante. Además, aunque el comportamiento del *DBIC* y el *BIC* fue bastante similar, el *DBIC* era recomendable, puesto que generalmente convergía más rápido a su distribución límite.

En lo que respecta al campo del análisis factorial, poco se ha escrito en la literatura científica acerca del uso de criterios de información estadística como herramienta de selección del número adecuado de factores a considerar dentro de un modelo. Hoy en día, en el análisis factorial basado en máxima verosimilitud, sigue siendo el test asintótico de la chi-cuadrado de Bartlett el procedimiento más empleado a la hora de determinar el número de factores a extraer, a pesar de los problemas que puede conllevar su aplicación, sobre todo en situaciones prácticas en las que se produzcan fuertes desviaciones del supuesto de normalidad. O, por poner otro ejemplo, cuando el tamaño muestral es grande, y el número de factores, m , es pequeño en relación con el número de variables originales, p . En ese caso, si los datos no siguen una distribución normal multivariante, el test asintótico de la chi-cuadrado conduce generalmente a rechazar H_0 . Ese es un problema frecuente en contrastes de hipótesis con muestras grandes, donde se tendería a rechazar H_0 . Así, según Peña (2002), es necesario, a la hora de decidir el número de factores, diferenciar entre significatividad práctica y significatividad estadística, como sucede en todo contraste de hipótesis. Tal vez, en esos casos, sea más aconsejable el uso de un procedimiento alternativo cuya decisión final no esté basada en un nivel de significación.

Capítulo 3

Selección del modelo factorial adecuado mediante criterios basados en la información estadística

De acuerdo a Bozdogan y Ramírez (1988), la teoría clásica de contraste de hipótesis no es apropiada en el contexto de problemas basados en la determinación del número de factores a extraer en un análisis factorial realizado sobre una serie de variables, dado que el problema no es contrastar una hipótesis en particular, sino más bien un problema de decisión múltiple. El test usual basado en la razón de verosimilitudes no siempre es válido, y el nivel de significación para el test no está ajustado en el proceso secuencial de contraste.

El uso de criterios de selección de modelos basados en la información estadística, es una herramienta válida a la hora de seleccionar el número de factores adecuado en un análisis factorial mediante el método de máxima verosimilitud, y también en otros problemas de selección de modelos del ámbito del análisis de datos multivariante, tal como apunta Sclove (1987). En el caso del análisis factorial, el modelo factorial ortogonal puede ser escrito en la forma de la ecuación (1.2), es decir, como

$$X = LF + \varepsilon,$$

donde X es el vector aleatorio compuesto por las p variables observadas, F es el vector de los m factores incorrelados ($m < p$), tomados de forma que tengan varianza unidad, y ε es el vector de residuos, las varianzas de las cuales son llamadas unicidades. Las variables que conforman el vector ε son incorreladas entre sí y con las variables del vector F . La matriz L es una matriz de dimensión

$p \times m$, llamada matriz de cargas factoriales. Sus elementos, l_{ij} , son las cargas de la variable X_i en el factor común F_j .

Llegados a este punto, un aspecto importante es la elección del número de factores, m . De acuerdo a este modelo, la matriz de covarianzas de X , Σ , toma la forma

$$\Sigma = LL^T + \Psi,$$

donde Ψ , la matriz de covarianzas de ε , es diagonal, siendo sus elementos diagonales las unicidades (esto puede ser visto con más detalle en el primer capítulo de esta Tesis). Bajo el modelo factorial ortogonal, la matriz de cargas factoriales, L , tiene mp parámetros para ser estimados, y la matriz de factores específicos, Ψ , tiene otros p parámetros para ser estimados, para un total de $mp + p$ parámetros. Sin embargo, la condición de unicidad $L^T\Psi^{-1}L = \Delta$ (otro de los supuestos en los que se basa el modelo factorial ortogonal, utilizado para forzar la convergencia a una solución única de la estimación iterativa máximo verosímil) supone $\frac{m(m+1)}{2}$ parámetros distintos y que Δ sea considerada diagonal. Por tanto, esta condición impone $\frac{m(m+1)}{2} - m = \frac{m(m-1)}{2}$ limitaciones adicionales al problema. Así, el número de parámetros libres a estimar dentro del modelo factorial, es $mp + p - \frac{m(m-1)}{2}$. Para que la realización de un análisis factorial tenga sentido, el número de factores, m , debe estar limitado de forma que la cantidad $mp + p - \frac{m(m-1)}{2}$ sea menor que $\frac{p(p+1)}{2}$, el número de términos de la matriz de covarianzas. Haciendo uso de la condición

$$mp + p - \frac{m(m-1)}{2} < \frac{p(p+1)}{2},$$

se tiene

$$p + m < p^2 - 2pm + m^2,$$

es decir,

$$(p - m)^2 > p + m.$$

Esta expresión proporciona un límite superior para el número de factores. El número de factores, m , no podrá exceder al mayor entero M que verifique

$$M < \frac{1}{2} [2p + 1 - \sqrt{8p + 1}].$$

La elección del número de factores (1, 2, ..., o M) será un problema de decisión múltiple (una elección entre modelos alternativos), para el cual los criterios de selección de modelos, basados en la información estadística, pueden ser razonablemente aplicados.

El procedimiento de decisión del número de factores a extraer se basa, en este caso, en minimizar el valor del criterio de información empleado para el propósito del investigador. Sea IC dicho criterio de información, se calcula el valor de $IC(m)$, donde m es el número de factores que varía en el rango de posibles factores a extraer, $m = 1, \dots, M$. El número natural m para el cual se minimice $IC(m)$, será el número adecuado de factores a extraer, de acuerdo a este procedimiento.

3.1. Selección del modelo factorial adecuado mediante criterios expresados en la forma $-2 \ln \mathcal{L}(\theta_j) + k_j(A(n) + C)$

En general, para cualquier criterio de información, IC , expresado en la forma

$$-2 \ln \mathcal{L}(\theta_j) + k_j(A(n) + C),$$

donde $\mathcal{L}(\theta_j)$ es la función de verosimilitud asociada al modelo j , k_j es el número de parámetros libres estimados dentro del modelo j , $A(n)$ es una función de n , el tamaño muestral, y C un término constante, se tiene

$$\begin{aligned} IC(m+1) &= -2 \ln \mathcal{L}_{m+1}(\theta_j) + (A(n) + C)[(m+1)p + p - \frac{1}{2}(m+1)m], \\ IC(m) &= -2 \ln \mathcal{L}_m(\theta_j) + (A(n) + C)[mp + p - \frac{1}{2}m(m-1)]. \end{aligned}$$

Así,

$$IC(m+1) - IC(m) = t + (A(n) + C)(p - m), \tag{3.1}$$

donde $t = -2 \ln \mathcal{L}_{m+1}(\theta_j) + 2 \ln \mathcal{L}_m(\theta_j)$.

La ecuación (3.1) puede reescribirse como

$$IC(m+1) - IC(m) = n \ln \frac{|\hat{\Sigma}_{m+1}|}{|\hat{\Sigma}_m|} + (A(n) + C)(p - m),$$

puesto que

$$\begin{aligned} t &= -2 \ln \mathcal{L}_{m+1}(\theta_j) + 2 \ln \mathcal{L}_m(\theta_j) = n \ln |\hat{\Sigma}_{m+1}| - n \ln |\hat{\Sigma}_m| = \\ &= n(\ln |\hat{\Sigma}_{m+1}| - \ln |\hat{\Sigma}_m|) = n \ln \frac{|\hat{\Sigma}_{m+1}|}{|\hat{\Sigma}_m|}, \end{aligned}$$

donde $\hat{\Sigma}_{m+1}$ es el estimador de máxima verosimilitud de la matriz de covarianzas asociada al modelo factorial con $m + 1$ factores extraídos, y $\hat{\Sigma}_m$ es el estimador de máxima verosimilitud de la matriz de covarianzas asociada al modelo factorial con m factores extraídos.

En definitiva, en la comparación entre un modelo con m factores y otro con $m + 1$, en el caso en que

$$IC(m+1) - IC(m) > 0,$$

se elegiría como más adecuado, en base al criterio IC , el modelo factorial con m factores.

3.2. Relación entre los criterios AIC , BIC y $CAIC$

Probablemente, el criterio de información de Akaike (AIC ; Akaike, 1974, 1987) y el criterio de información de Schwarz (BIC ; Schwarz, 1978) hayan sido los criterios de información estadística más comúnmente utilizados. Woodroffe (1982) mostró que el AIC no es consistente, es decir, que en general el AIC no seleccionará el modelo correcto cuando n tienda a infinito. Bozdogan (1987), utilizando argumentos similares, derivó una versión consistente del AIC ($CAIC$) a partir de la medida de información de Kullback-Leibler; esta versión hace al AIC asintóticamente consistente y penaliza el exceso de parámetros de forma rigurosa, con el fin de seleccionar el modelo más simple de entre los modelos adecuados. Bozdogan y Ramírez (1987) subrayaban el hecho de que, introduciendo procedimientos de selección de modelos como el AIC y el $CAIC$, y esto sería extensible al BIC

y demás criterios de información estadística, no es necesario llevar a cabo los clásicos tests de hipótesis para determinar el número adecuado de factores en un análisis factorial: no se precisa elegir subjetivamente el número de factores basado en un nivel de significación arbitrariamente elegido.

Cualitativamente, el procedimiento de Schwarz y el de Akaike dan *una formulación matemática del principio de parsimonia en la construcción de modelos*. Cuantitativamente, dado que el procedimiento de Schwarz difiere del de Akaike sólo en que la dimensión es multiplicada por $\ln n$, el *BIC* se inclina más que el *AIC* hacia modelos de dimensionalidad más baja cuando el número de observaciones es $n \geq 8$. Esto se demuestra a continuación:

Sea m el número de factores extraídos, $m = 1, \dots, M$. Sea $\mathcal{L}_m(\theta_j)$ la función de verosimilitud asociada al modelo j con m factores extraídos, mientras que $\mathcal{L}_{(m+1)}(\theta_j)$ es la función de verosimilitud asociada al modelo con $m + 1$ factores extraídos. Sea $k = (mp + p) - \frac{1}{2}m(m - 1)$ el número de parámetros libres estimados dentro de un modelo factorial con m factores extraídos.

$$\begin{aligned} AIC(m + 1) &= -2 \ln \mathcal{L}_{(m+1)}(\theta_j) + 2[(m + 1)p + p] - \frac{1}{2}(m + 1)m, \\ AIC(m) &= -2 \ln \mathcal{L}_m(\theta_j) + 2[(mp + p) - \frac{1}{2}m(m - 1)], \\ BIC(m + 1) &= -2 \ln \mathcal{L}_{(m+1)}(\theta_j) + (\ln n)[(m + 1)p + p] - \frac{1}{2}(m + 1)m, \\ BIC(m) &= -2 \ln \mathcal{L}_m(\theta_j) + (\ln n)[(mp + p) - \frac{1}{2}m(m - 1)]. \end{aligned}$$

Sea $t = -2 \ln \mathcal{L}_{(m+1)}(\theta_j) + 2 \ln \mathcal{L}_m(\theta_j)$,

$$AIC(m + 1) - AIC(m) = t + 2(p - m), \quad (3.2)$$

$$BIC(m + 1) - BIC(m) = t + (\ln n)(p - m). \quad (3.3)$$

Así, la estrategia de selección del modelo adecuado consistiría en realizar sucesivas comparaciones entre un modelo con m factores y otro con $m + 1$ factores, siendo $m = 1, \dots, M - 1$, donde M sería el número máximo posible de factores extraídos. Si se tiene que $AIC(m + 1) - AIC(m) > 0$, $AIC(m) < AIC(m + 1)$, y, por tanto, resulta más adecuado en base al criterio *AIC* el modelo con

m factores extraídos. Procedimiento análogo para el caso del BIC .

Dadas las ecuaciones (3.2) y (3.3), y como el número de variables observadas, p , es mayor que m , se tiene que (puesto que, para $n \geq 8$, $\ln n > 2$) para $n \geq 8$, $n \in \mathbb{N}$,

$$BIC(m+1) - BIC(m) > AIC(m+1) - AIC(m).$$

Así, si $n \geq 8$, en una comparación entre un modelo con m factores y otro con $m+1$, si se tiene que $AIC(m+1) - AIC(m) > 0$, y por tanto es elegido en base al AIC el modelo con m factores, entonces $BIC(m+1) - BIC(m)$ también sería mayor que 0 y en base al BIC sería preferible también el modelo con m factores en lugar del modelo con $m+1$. Por tanto, para $n \geq 8$, el BIC favorece la elección de modelos de menor dimensionalidad con respecto al AIC . Al contrario sucedería cuando $n < 8$, pero quizás no resultaría muy aconsejable extraer conclusiones de un experimento con un número tan reducido de observaciones.

En el caso concreto de

$$BIC(m+1) - BIC(m) > 0 > AIC(m+1) - AIC(m),$$

el uso de ambos criterios daría lugar a diferentes conclusiones sobre si el número de factores extraídos debería ser igual a m o igual a $m+1$.

Particularizando la ecuación (3.1), en el caso de los criterios de información AIC , BIC y $CAIC$, se tendría

$$AIC(m+1) - AIC(m) = t + 2(p - m),$$

$$BIC(m+1) - BIC(m) = t + (\ln n)(p - m),$$

$$CAIC(m+1) - CAIC(m) = t + (\ln n + 1)(p - m),$$

siendo $t = -2 \ln \mathcal{L}_{(m+1)}(\theta_j) + 2 \ln \mathcal{L}_m(\theta_j) = n \ln \frac{|\hat{\Sigma}_{m+1}|}{|\hat{\Sigma}_m|}$, donde $\hat{\Sigma}_{m+1}$ es el estimador de máxima verosimilitud de la matriz de covarianzas asociada al modelo factorial con $m+1$ factores extraídos, y $\hat{\Sigma}_m$ es el estimador de máxima verosimilitud de la matriz de covarianzas asociada al modelo

factorial con m factores extraídos. Así, para todo $n \in \mathbb{N}$,

$$CAIC(m+1) - CAIC(m) > BIC(m+1) - BIC(m). \quad (3.4)$$

Para $n \geq 3$, $\ln(n) + 1 \geq 2$ y, por tanto,

$$CAIC(m+1) - CAIC(m) > AIC(m+1) - AIC(m). \quad (3.5)$$

En relación al AIC y al BIC , el $CAIC$ impone una penalización más severa para la sobreparametrización. De las ecuaciones (3.4) y (3.5), se concluye que el $CAIC$ favorece la elección de modelos de menor dimensionalidad, en comparación con el AIC y el BIC . De hecho, Bozdogan y Ramírez (1988) propusieron una estrategia combinada que proporcionase un rango para la extracción de factores, utilizando el AIC para determinar el límite superior, y el $CAIC$ para determinar el límite inferior.

3.2.1. Ejemplo

Se presenta a continuación un ejemplo de aplicación de los criterios de información expuestos en la sección 3.2, en el contexto de Educación e Investigación y Desarrollo (I+D). Educación e I+D son campos íntimamente relacionados. Diversas investigaciones sugieren que los sujetos con mayor nivel educacional tienen una mayor capacidad para resolver problemas y aprender eficientemente los contenidos de disciplinas específicas. Así, en algunos estudios se ha hipotetizado que individuos con un mayor nivel educativo estarían capacitados para lograr mejores resultados en I+D que aquellos individuos con menor nivel educativo. Nelson y Phelps (1966) ya sugirieron una conexión entre el nivel de estudios y la capacidad para innovar, además de una relación entre educación y crecimiento académico. Lee, Wong y Chen (2005) obtienen que el nivel educativo tiene un impacto positivo en patentes obtenidas y artículos publicados. Griliches (1987) aborda la relación entre patentes y gasto en I+D. Kirankabes (2010) obtiene resultados en consonancia con una relación positiva entre el porcentaje del Producto Interior Bruto (PIB) destinado a I+D y el número de patentes obtenidas. Se consideran en el ejemplo propuesto cinco variables relativas a Educación e I+D:

- *LifeLong Learning*: Población con edad entre 25 y 64 años que sigue recibiendo formación académica o cursos de perfeccionamiento.
- *Abandono de estudios*: Porcentaje de jóvenes con edad entre 18 y 24 años que abandonaron el sistema educativo antes de completar el nivel más bajo de educación secundaria.
- *Desempleo entre titulados superiores*: Tasa de desempleo de la población con edad entre 25 y 64 años que ha realizado estudios superiores.
- *Gasto en I+D*: Porcentaje del PIB destinado a gasto en I+D.
- *Patentes*: Patentes por millón de habitantes.

Los datos objeto de estudio pertenecen a los 27 países que actualmente conforman la Unión Europea. Son correspondientes al año 2009 y han sido extraídos de la base de datos EUROSTAT.

Teniendo en cuenta que el número de factores, m , no puede exceder al mayor entero M tal que

$$m \leq M < \frac{1}{2} \left[2p + 1 - \sqrt{8p + 1} \right],$$

donde p es el número de variables originales ($p = 5$ en este caso), el límite superior de factores a considerar en este ejemplo sería $M = 2$. Así, en este caso, se trata de decidir si resulta más adecuado el modelo con dos factores o el modelo con un solo factor. Ello se hará, en este ejemplo, en base a los criterios de información *AIC*, *BIC* y *CAIC*.

Para el cálculo de los valores correspondientes del *AIC*, *BIC* y *CAIC*, en cada situación, se define en la versión 2.8.1 (el procedimiento sería análogo para cualquier otra versión) del programa *R* (*R Development Core Team*, 2008) una función para el cálculo de estos criterios de información. La función definida hace uso de la función *factanal*, que realiza un análisis factorial basado en el procedimiento de extracción de factores mediante máxima verosimilitud y proporciona el resultado de la aplicación del test asintótico de Bartlett. La función que se define en este estudio toma como elemento de entrada la base de datos correspondiente, compuesta por las variables sobre las que se pretende realizar el análisis factorial, para devolver los valores de los criterios de información definidos, correspondientes a las posibles soluciones factoriales.

```

InformationCriteria<-function(data){
n<-nrow(data)
p<-ncol(data)
for(m in 1:(trunc(1/2*(2*p+1-sqrt(8*p+1))))){
  fit<-factanal(data,factors=m)
  s<-tcrossprod(fit$loadings)+diag(fit$uniquenesses)
  k<-(m*p+p)-1/2*m*(m-1)
  d<-1/2*((p-m)^2-(p+m))
  logl<-(-1/2)*(n*p*(log(2*pi)+1)+n*log(det(s)))
  aic<--2*logl+2*k
  bic<--2*logl+(log(n)*k)
  caic<--2*logl+(log(n)+1)*k
  chi_squared<-fit$STATISTIC
  p_value<-fit$PVAL
  print(c(sample_size=n, variables=p, factors=m, free_parameters=k, log_likelihood=logl,
AIC=aic, BIC=bic, CAIC=caic,df=d, chi_squared=chi_squared, p_value<-p_value))
  }
}

```

De acuerdo al test asintótico de Bartlett, serían aceptables ambas soluciones a un nivel de confianza del 95 %. En base al valor mínimo de *AIC*, *BIC* y *CAIC*, sería elegida la solución correspondiente al modelo factorial compuesto por un único factor (ver Tabla 3.1).

| <i>m</i> | <i>k</i> | <i>AIC(m)</i> | <i>BIC(m)</i> | <i>CAIC(m)</i> | <i>d</i> | <i>p</i> |
|----------|----------|-----------------|-----------------|-----------------|----------|----------|
| 1 | 10 | 331.0096 | 343.9680 | 353.9680 | 5 | 0.4187 |
| 2 | 14 | 333.3357 | 351.4773 | 365.4773 | 1 | 0.7117 |

Tabla 3.1. *AIC*, *BIC*, *CAIC* y *p*-valor asociado al estadístico χ^2 para el conjunto de datos de los 27 países pertenecientes a la UE; *m* = Número de factores; *k* = Número de parámetros libres estimados dentro del modelo; *d* = Grados de libertad; *p* = *p*-valor asociado al estadístico χ^2 .

Obsérvese que $n = 27$ y que por tanto se verifica

$$CAIC(2) - CAIC(1) > BIC(2) - BIC(1) > AIC(2) - AIC(1).$$

Considérense a continuación los mismos datos, pero relativos en este caso a los 15 países que conformaban la antigua Europa de los Quince (es decir, los países que formaron parte de la Unión Europea entre el 1 de Enero de 1995 y el 1 de Mayo de 2004). La solución con un factor y la solución con dos factores vuelven a ser ambas aceptadas a un nivel de confianza del 95 %, mediante la aplicación del test asintótico de Bartlett.

Sin embargo, en esta ocasión existen diferencias en la elección del número de factores en base a la utilización de los criterios de información estadística AIC , BIC y $CAIC$. Así, en este caso el AIC seleccionaría la solución con dos factores, mientras que tanto el BIC como el $CAIC$ mantendrían la estructura con un solo factor (ver Tabla 3.2), y, en principio, la elección en base a ellos no se vería afectada por la reducción del tamaño muestral.

| m | k | $AIC(m)$ | $BIC(m)$ | $CAIC(m)$ | d | p |
|-----|-----|-----------------|-----------------|-----------------|-----|--------|
| 1 | 10 | 188.2364 | 195.3169 | 205.3169 | 5 | 0.1511 |
| 2 | 14 | 186.7206 | 196.6333 | 210.6333 | 1 | 0.2952 |

Tabla 3.2. AIC , BIC , $CAIC$ y p -valor asociado al estadístico χ^2 para el conjunto de datos de los 15 países pertenecientes a la antigua UE; m = Número de factores; k = Número de parámetros libres estimados dentro del modelo; d = Grados de libertad; p = p -valor asociado al estadístico χ^2 .

Así pues, la reducción del tamaño muestral considerado no conlleva una variación en la decisión del número de factores a extraer en base al $CAIC$ y al BIC , aunque dicha decisión sí se vería modificada si fuese el AIC el criterio de información considerado. En cualquier caso, el objetivo de este ejemplo no es otro que el de observar como afecta la definición de los criterios de información empleados a la elección del número de factores. Para evaluar el rendimiento de los criterios utilizados en este ejemplo, sobre todo en el caso del BIC y el $CAIC$, podría ser útil la comparación de los valores obtenidos con los valores correspondientes a versiones de ambos criterios, las cuales ajusten el valor del criterio en función del tamaño muestral (Draper, 1995; Yang y Yang, 2007).

3.3. Relación entre el criterio *RIC* de Rissanen y el *BIC** de Yang y Yang

Tanto el criterio *RIC* de Rissanen (1978) como el criterio *BIC** de Yang y Yang (2007) comparten la estructura del criterio de información bayesiano, *BIC*, de Schwarz (1978). Es por ello que son clasificados como pertenecientes a la familia *BIC* en la Tabla 2.1, aunque no conviene perder de vista que Rissanen derivó el *RIC* de forma independiente al *BIC* de Schwarz (ambos criterios son publicados en fechas próximas entre sí). El *BIC** propuesto por Yang y Yang tiene una definición similar al *RIC*, compartiendo ambos el ajuste de acuerdo al tamaño muestral y omitiéndose en el caso del *BIC** el término adicional $B(j, n)$ propuesto por Rissanen para cada modelo j específicamente considerado (ver Tabla 2.1). Ambos criterios están, por definición, notablemente relacionados. Véase a continuación como varía su comportamiento en un proceso de decisión entre un modelo factorial con m factores y otro con $m + 1$. Para ello, recuérdese en primer lugar la definición de ambos criterios:

$$\begin{aligned}
 BIC^*(m) &= np \ln(2\pi) + n \ln |\hat{L}\hat{L}^T + \hat{\Psi}| + np + \left(\ln \left(\frac{n+2}{24} \right) \right) [(mp+p) - \frac{1}{2}m(m-1)], \\
 RIC(m) &= np \ln(2\pi) + n \ln |\hat{L}\hat{L}^T + \hat{\Psi}| + np + \left(\ln \left(\frac{n+2}{24} \right) \right) [(mp+p) - \frac{1}{2}m(m-1)] + \\
 &\quad + 2 \ln \left((mp+p) - \frac{1}{2}m(m-1) + 1 \right),
 \end{aligned}$$

donde

- n = Tamaño muestral total.
- p = Número de variables originales.
- \hat{L} = Matriz de cargas factoriales estimada mediante máxima verosimilitud.
- $\hat{\Psi}$ = Matriz de factores específicos estimada mediante máxima verosimilitud.
- m = Número de factores, $m = 1, 2, \dots, M$.
- $\ln |\hat{L}\hat{L}^T + \hat{\Psi}|$ es el logaritmo neperiano del determinante de la matriz $\hat{L}\hat{L}^T + \hat{\Psi}$.

Particularizando la ecuación (3.1) al caso del BIC^* , se tiene

$$BIC^*(m+1) - BIC^*(m) = n \ln \left(\frac{|\hat{\Sigma}_{m+1}|}{|\hat{\Sigma}_m|} \right) + \ln \left(\frac{n+2}{24} \right) (p-m),$$

donde $\hat{\Sigma}_{m+1}$ es el estimador de máxima verosimilitud de la matriz de covarianzas asociada al modelo factorial con $m+1$ factores extraídos, y $\hat{\Sigma}_m$ es el estimador de máxima verosimilitud de la matriz de covarianzas asociada al modelo factorial con m factores extraídos.

En el caso del RIC , se tiene que

$$\begin{aligned} RIC(m+1) - RIC(m) &= BIC^*(m+1) - BIC^*(m) + 2 \ln \left(1 + \frac{p-m}{mp + p - \frac{1}{2}m(m-1) + 1} \right) = \\ &= n \ln \left(\frac{|\hat{\Sigma}_{m+1}|}{|\hat{\Sigma}_m|} \right) + \ln \left(\frac{n+2}{24} \right) (p-m) + 2 \ln \left(1 + \frac{p-m}{mp + p - \frac{1}{2}m(m-1) + 1} \right). \end{aligned}$$

Se tiene entonces que, sea cual sea el tamaño muestral n ,

$$RIC(m+1) - RIC(m) > BIC^*(m+1) - BIC^*(m).$$

En efecto, el RIC impone una penalización más severa para la sobreparametrización que el BIC^* , mediante el término $B(j, n) = 2 \ln(k+1)$, donde k es el número de parámetros libres estimados dentro del modelo j . Así pues, el RIC favorece la elección de modelos de menor dimensión en relación con el BIC^* , aunque ambos criterios están muy relacionados entre sí de acuerdo a la definición, y sólo en circunstancias muy especiales el uso de uno u otro criterio daría lugar a la elección de modelos distintos en base a ambos criterios, con m y $m+1$ factores.

3.4. Función para el cálculo mediante el programa R de los criterios de información referenciados en este estudio

A continuación, se amplía la función dada en el apartado 3.2.1, de modo que la nueva función ofrezca como salida los valores correspondientes a cada uno de los 10 criterios de información dados en la Tabla 2.1. Asimismo, la función definida proporciona los resultados correspondientes a la aplicación

del test asintótico de Bartlett, para cada solución factorial. Esta nueva función vuelve a hacer uso de la función *factanal*, que realiza un análisis factorial basado en el procedimiento de extracción de factores mediante máxima verosimilitud. La nueva función definida puede ser ejecutada como un *script* en *R*. Además, esta función devuelve también como salida el número de factores que serían seleccionados en base a cada uno de los criterios de información objeto de estudio, seleccionando para cada criterio *IC* el número de factores, *m*, que minimiza el valor de la función *IC(m)* en el conjunto de modelos candidatos.

```
InformationCriteria<-function(data){
n<-nrow(data)
p<-ncol(data)
AIC<-c(NULL)
BIC<-c(NULL)
HQIC<-c(NULL)
AIC3<-c(NULL)
CAIC<-c(NULL)
CAIC_Yang<-c(NULL)
AICc<-c(NULL)
DBIC<-c(NULL)
BIC_Yang<-c(NULL)
RIC<-c(NULL)
for(m in 1:(trunc(1/2*(2*p+1-sqrt(8*p+1))))){
  fit<-factanal(data,factors=m)
  s<-tcrossprod(fit$loadings)+diag(fit$uniquenesses)
  k<-(m*p+p)-1/2*m*(m-1)
  d<-1/2*((p-m)^2-(p+m))
  logl<-(-1/2)*(n*p*(log(2*pi)+1)+n*log(det(s)))
  aic<--2*logl+2*k
  AIC<-rbind(AIC,aic)
}
```

```

bic $\leftarrow$ --2*logl+(log(n)*k)
BIC $\leftarrow$ rbind(BIC,bic)
hqic $\leftarrow$ --2*logl+2*(log(log(n)))*k
HQIC $\leftarrow$ rbind(HQIC,hqic)
aic3 $\leftarrow$ --2*logl+3*k
AIC3 $\leftarrow$ rbind(AIC3,aic3)
caic $\leftarrow$ --2*logl+(log(n)+1)*k
CAIC $\leftarrow$ rbind(CAIC,caic)
caicYang $\leftarrow$ --2*logl+(log(n+2)+(1-log(24)))*k
CAIC_Yang $\leftarrow$ rbind(CAIC_Yang,caicYang)
aicc $\leftarrow$ --2*logl+2*k+(2*(k+1)*(k+2)/(n-k-2))
AICc $\leftarrow$ rbind(AICc,aicc)
dbic $\leftarrow$ --2*logl+(log(n)-log(2*pi))*k
DBIC $\leftarrow$ rbind(DBIC,dbic)
bicYang $\leftarrow$ --2*logl+(log(n+2)-log(24))*k
BIC_Yang $\leftarrow$ rbind(BIC_Yang,bicYang)
ric $\leftarrow$ --2*logl+log((n+2)/24)*k+2*log(k+1)
RIC $\leftarrow$ rbind(RIC,ric)
chi_squared $\leftarrow$ fit$STATISTIC
p_value $\leftarrow$ fit$PVAL
print(c(sample_size=n, variables=p, factors=m, free_parameters=k, log_likelihood=logl,
AIC=aic, BIC=bic, HQIC=hqic, AIC3=aic3, CAIC=caic, CAIC_Yang=caicYang, AICc=aicc,
DBIC=dbic, BIC_Yang=bicYang, RIC=ric))
print(c(sample_size=n, variables=p, factors=m, df=d, chi_squared=chi_squared,
p_value $\leftarrow$ p_value))
}
print(c(number_of_factors_AIC=(1:length(AIC))[AIC==min(AIC)]))
print(c(number_of_factors_BIC=(1:length(BIC))[BIC==min(BIC)]))

```

```

print(c(number_of_factors_HQIC=(1:length(HQIC))[HQIC==min(HQIC)]))
print(c(number_of_factors_AIC3=(1:length(AIC3))[AIC3==min(AIC3)]))
print(c(number_of_factors_CAIC=(1:length(CAIC))[CAIC==min(CAIC)]))
print(c(number_of_factors_CAIC_Yang=(1:length(CAIC_Yang))[CAIC_Yang==min(CAIC_Yang)]))
print(c(number_of_factors_AICc=(1:length(AICc))[AICc==min(AICc)]))
print(c(number_of_factors_DBIC=(1:length(DBIC))[DBIC==min(DBIC)]))
print(c(number_of_factors_BIC_Yang=(1:length(BIC_Yang))[BIC_Yang==min(BIC_Yang)]))
print(c(number_of_factors_RIC=(1:length(RIC))[RIC==min(RIC)]))
}

```

3.5. Relación entre los pesos asociados a *BIC* y *CAIC*

Recuérdese que para cada uno de los criterios de información presentados en este trabajo (ver sección 2.5, Tabla 2.1) puede obtenerse una evidencia en favor de un modelo factorial, respecto del resto de modelos factoriales candidatos. Son los llamados pesos asociados a cada modelo, dependientes del criterio de información que se utilice para la selección del modelo factorial. Tal y como se expone en la sección 2.5.1, para criterios de información *IC*, dados en la forma

$$-2 \cdot \ln \mathcal{L}(\theta_j) + k_j \cdot (A(n) + C) + B(j, n), \quad (3.6)$$

donde

- n es el tamaño muestral total.
- $\ln \mathcal{L}(\theta_j)$ es el logaritmo neperiano de la verosimilitud maximizado bajo el modelo j .
- k_j es el número de parámetros libres estimados dentro del modelo j . En el caso del análisis factorial, recuérdese que $k_j = (mp + p) - \frac{1}{2}m(m - 1)$, donde m es el número de factores extraídos para el modelo j y p el número de variables originales.
- $A(n) + C$, siendo $A(n)$ una función de n y C un término constante, es el coste de ajustar un parámetro adicional.

- $B(j, n)$ es un término adicional que depende del criterio y del modelo j ,

se puede obtener una probabilidad, llamada ponderación o peso asociado al criterio de información IC , que representa la evidencia en favor de un modelo compuesto por i factores. Este peso asociado al modelo con i factores viene dado por

$$w_i(IC) = \frac{\exp\left\{-\frac{1}{2}\Delta_i(IC)\right\}}{\sum_{k=1}^M \exp\left\{-\frac{1}{2}\Delta_k(IC)\right\}}, \quad (3.7)$$

donde $\Delta_i(IC)$ es la diferencia entre el valor del criterio IC asociado al modelo factorial con i factores y el mínimo valor IC encontrado entre los modelos candidatos, es decir,

$$\Delta_i(IC) = IC_i - \text{mín}(IC). \quad (3.8)$$

El rango de i , que representa el número de factores de los modelos candidatos, varía entre 1 y M , donde M es el mayor entero tal que $M \leq \frac{1}{2} [2p + 1 - \sqrt{8p + 1}]$ (esta desigualdad también se puede ver en forma estricta si se considera que el número de grados de libertad ha de ser estrictamente mayor que cero).

Tanto el BIC de Schwarz como el $CAIC$ de Bozdogan, son criterios de información asintóticamente consistentes, y, por la estructura implícita en sus respectivas definiciones, ambos criterios son proclives a la elección de modelos de baja dimensión. En la comparación exclusiva entre estos dos criterios de información, el $CAIC$ penaliza más que el BIC la inclusión de parámetros, favoreciendo por tanto la elección de modelos de menor dimensión. Este hecho se traduce en una mayor evidencia en favor de modelos de baja dimensión, cuando el $CAIC$ es el criterio de información utilizado.

Supóngase que $CAIC$ y BIC coinciden en seleccionar como modelo factorial más adecuado el modelo con m factores. Sea $j \in \mathbb{N}^+$, $j \leq M - m$, donde M es el máximo número posible de factores a extraer, se tiene (el procedimiento es análogo al de obtención de la expresión dada en la ecuación

(3.4))

$$\begin{aligned} CAIC(m+j) - CAIC(m+(j-1)) &> BIC(m+j) - BIC(m+(j-1)), \\ &\vdots \\ CAIC(m+1) - CAIC(m) &> BIC(m+1) - BIC(m). \end{aligned}$$

Esto implica

$$CAIC(m+j) - CAIC(m) > BIC(m+j) - BIC(m). \quad (3.9)$$

Por otra parte, para todo criterio de información IC expresado en la forma (3.6), se tiene (de forma análoga a lo observado en la sección 3.1)

$$\begin{aligned} IC(m-1) &= -2 \ln \mathcal{L}_{m-1}(\theta_j) + (A(n) + C)[(m-1)p + p - \frac{1}{2}(m-1)(m-2)], \\ IC(m) &= -2 \ln \mathcal{L}_m(\theta_j) + (A(n) + C)[mp + p - \frac{1}{2}m(m-1)]. \end{aligned}$$

Restando ambos términos,

$$IC(m-1) - IC(m) = t + (A(n) + C)((m-1) - p), \quad (3.10)$$

donde $t = -2 \ln \mathcal{L}_{(m-1)}(\theta_j) + 2 \ln \mathcal{L}_m(\theta_j)$.

Particularizando la ecuación (3.10) para los criterios BIC y $CAIC$, se obtiene

$$\begin{aligned} BIC(m-1) - BIC(m) &= t + (\ln n)((m-1) - p), \\ CAIC(m-1) - CAIC(m) &= t + (\ln n + 1)((m-1) - p), \end{aligned}$$

siendo $t = -2 \ln \mathcal{L}_{(m-1)}(\theta_j) + 2 \ln \mathcal{L}_m(\theta_j)$.

De manera similar a la expresión (3.9), se tiene que, $\forall j \in \mathbb{N}^+, j < m$,

$$CAIC(m-j) - CAIC(m) < BIC(m-j) - BIC(m).$$

Haciendo uso de las expresiones (3.7) y (3.8) y atendiendo a los resultados anteriores, supóngase que tanto BIC como $CAIC$ escogen como modelo más adecuado un modelo compuesto por m factores comunes. Se tendría

$$\begin{aligned}\Delta_m(CAIC) &= 0 = \Delta_m(BIC), \\ \Delta_{m+j}(CAIC) &> \Delta_{m+j}(BIC) \quad \forall j \in \mathbb{N}^+, j \leq M - m, \\ \Delta_{m-j}(CAIC) &< \Delta_{m-j}(BIC) \quad \forall j \in \mathbb{N}^+, j < m.\end{aligned}$$

Concretamente, en el caso en que BIC y $CAIC$ seleccionen ambos una solución compuesta por un solo factor (algo que es bastante probable cuando el número de variables originales es relativamente pequeño), se tiene

$$\begin{aligned}\Delta_1(CAIC) &= 0 = \Delta_1(BIC), \\ \Delta_{m+j}(CAIC) &> \Delta_{m+j}(BIC) \quad \forall j \in \mathbb{N}^+, j \leq M - 1.\end{aligned}$$

De esta forma,

$$\exp\left\{-\frac{1}{2}\Delta_1(CAIC)\right\} = 1 = \exp\left\{-\frac{1}{2}\Delta_1(BIC)\right\},$$

mientras que, dado que $\Delta_{m+j}(CAIC) > \Delta_{m+j}(BIC) \forall j \in \mathbb{N}^+, j \leq M - 1$,

$$\sum_{k=1}^M \exp\left\{-\frac{1}{2}\Delta_k(CAIC)\right\} < \sum_{k=1}^M \exp\left\{-\frac{1}{2}\Delta_k(BIC)\right\}.$$

Así, $w_1(CAIC) > w_1(BIC)$, por lo que, en el caso de que ambos criterios seleccionasen un único factor, la evidencia en favor de dicho modelo sería superior en el caso del $CAIC$, que es más riguroso que el BIC a la hora de penalizar la inclusión de parámetros en el modelo.

3.5.1. Función en *R* para el cálculo de los pesos asociados a los criterios de información referenciados en este estudio

En la sección 3.4, se presentaba un algoritmo de cálculo de los valores de los criterios de información dados en la Tabla 2.1, para los posibles modelos candidatos, con el objetivo de la implementación de este algoritmo en *R*. La función proporcionaba además como salida el número de factores que serían seleccionados en base a cada uno de los criterios de información. Sin embargo, en ocasiones puede ser conveniente analizar la evidencia en favor de un modelo con respecto a otro, y estudiar con detenimiento los casos en que un criterio asigna una probabilidad similar de ser el modelo adecuado a dos modelos con un número de factores distinto. La información acerca de la evidencia en favor de un modelo la proporcionan los pesos expuestos en la sección 2.5.1. A tal efecto, se diseña la siguiente función, ejecutable como *script* en *R*, que devuelve el peso asociado a cada uno de los modelos candidatos, para cada uno de los criterios de información objeto de estudio.

```
weightsIC<-function(data){
n<-nrow(data)
p<-ncol(data)
AIC<-c(NULL)
BIC<-c(NULL)
HQIC<-c(NULL)
AIC3<-c(NULL)
CAIC<-c(NULL)
CAIC_Yang<-c(NULL)
AICc<-c(NULL)
DBIC<-c(NULL)
BIC_Yang<-c(NULL)
RIC<-c(NULL)
for(m in 1:(trunc(1/2*(2*p+1-sqrt(8*p+1))))){
  fit<-factanal(data,factors=m)
  s<-tcrossprod(fit$loadings)+diag(fit$uniquenesses)
```

```

k<-(m*p+p)-1/2*m*(m-1)
d<-1/2*((p-m)^2-(p+m))
logl<-(-1/2)*(n*p*(log(2*pi)+1)+n*log(det(s)))
aic<--2*logl+2*k
AIC<-rbind(AIC,aic)
bic<--2*logl+(log(n)*k)
BIC<-rbind(BIC,bic)
hqic<--2*logl+2*(log(log(n)))*k
HQIC<-rbind(HQIC,hqic)
aic3<--2*logl+3*k
AIC3<-rbind(AIC3,aic3)
caic<--2*logl+(log(n)+1)*k
CAIC<-rbind(CAIC,caic)
caicYang<--2*logl+(log(n+2)+(1-log(24)))*k
CAIC_Yang<-rbind(CAIC_Yang,caicYang)
aicc<--2*logl+2*k+(2*(k+1)*(k+2)/(n-k-2))
AICc<-rbind(AICc,aicc)
dbic<--2*logl+(log(n)-log(2*pi))*k
DBIC<-rbind(DBIC,dbic)
bicYang<--2*logl+(log(n+2)-log(24))*k
BIC_Yang<-rbind(BIC_Yang,bicYang)
ric<--2*logl+log((n+2)/24)*k+2*log(k+1)
RIC<-rbind(RIC,ric)
}

diffAIC<-AIC-min(AIC)
diffBIC<-BIC-min(BIC)
diffHQIC<-HQIC-min(HQIC)
diffAIC3<-AIC3-min(AIC3)

```

```

diffCAIC<-CAIC-min(CAIC)
diffCAIC_Yang<-CAIC_Yang-min(CAIC_Yang)
diffAICc<-AICc-min(AICc)
diffDBIC<-DBIC-min(DBIC)
diffBIC_Yang<-BIC_Yang-min(BIC_Yang)
diffRIC<-RIC-min(RIC)
wAIC<-(exp(-diffAIC/2))/(sum(exp(-diffAIC/2)))
wBIC<-(exp(-diffBIC/2))/(sum(exp(-diffBIC/2)))
wHQIC<-(exp(-diffHQIC/2))/(sum(exp(-diffHQIC/2)))
wAIC3<-(exp(-diffAIC3/2))/(sum(exp(-diffAIC3/2)))
wCAIC<-(exp(-diffCAIC/2))/(sum(exp(-diffCAIC/2)))
wCAIC_Yang<-(exp(-diffCAIC_Yang/2))/(sum(exp(-diffCAIC_Yang/2)))
wAICc<-(exp(-diffAICc/2))/(sum(exp(-diffAICc/2)))
wDBIC<-(exp(-diffDBIC/2))/(sum(exp(-diffDBIC/2)))
wBIC_Yang<-(exp(-diffBIC_Yang/2))/(sum(exp(-diffBIC_Yang/2)))
wRIC<-(exp(-diffRIC/2))/(sum(exp(-diffRIC/2)))
for(m in 1:(trunc(1/2*(2*p+1-sqrt(8*p+1))))){
print(c(factors = m, weightAIC = wAIC[m], weightBIC = wBIC[m],
weightHQIC = wHQIC[m], weightAIC3 = wAIC3[m], weightCAIC = wCAIC[m],
weightCAIC_Yang = wCAIC_Yang[m], weightAICc = wAICc[m], weightDBIC = wDBIC[m],
weightBIC_Yang = wBIC_Yang[m], weightRIC = wRIC[m]))
}
}

```


Capítulo 4

Uso de criterios de información para la selección del modelo factorial adecuado en ejemplos clásicos de la literatura

En este capítulo, se consideran tres ejemplos clásicos de aplicación de un análisis factorial basado en la metodología de extracción de factores mediante máxima verosimilitud. Por un lado, un ejemplo de cotizaciones de acciones de la bolsa neoyorquina, el ejemplo de “*stock-prices*” propuesto por Johnson y Wichern (1988); por otra parte, se analiza el conjunto de datos presentado por Harman (1976), compuesto por observaciones en 24 tests psicológicos cumplimentados por niños de una escuela de Chicago. En tercer lugar, se afronta otro de los problemas clásicos tratados en la literatura del análisis factorial: estudiar la dimensionalidad de la disciplina de Decatlón. Johnson y Wichern (1988), y Basilewsky (1994), por citar un par de ejemplos, presentan en sus textos sendos análisis factoriales, llevados a cabo con diversas metodologías, donde estudian la estructura correlacional y la dimensionalidad del Decatlón, a partir de unos datos previos correspondientes a Linden (1977), basados en 8 certámenes olímpicos. En este trabajo, se presentan datos actuales, correspondientes a competiciones celebradas entre 2005 y 2009. Se comparan los resultados obtenidos con los obtenidos en estudios previos por otros autores.

Finalmente, en cada uno de los tres problemas de selección del modelo factorial considerados, se analiza el rendimiento de los criterios de información presentados en la Tabla 2.1, a partir de las soluciones propuestas por los autores que plantearon y han estudiado los respectivos problemas.

4.1. Ejemplo de “*stock-prices*” de Johnson y Wichern (1988)

Johnson y Wichern (1988, ver conjunto de datos analizado en págs. 374-375) consideran la tasa de rendimiento semanal para las acciones de cinco compañías (*Allied Chemical, DuPont, Union Carbide, Exxon y Texaco*) de la bolsa de Nueva York, estudiando los valores desde Enero de 1975 hasta Diciembre de 1976. La tasa de rendimiento semanal se define como el cociente de la diferencia del precio (cotización) al cierre en el viernes actual respecto del precio al cierre el viernes anterior, entre el precio al cierre el viernes anterior, es decir,

$$TR_i = \frac{c_i - c_{i-1}}{c_{i-1}} = \frac{c_i}{c_{i-1}} - 1,$$

donde c_i es la cotización de la acción correspondiente al viernes de la semana i y TR_i representa la tasa de rendimiento correspondiente a la semana i . En este caso, i toma valores entre 1 y 100, puesto que 100 son las semanas sucesivas consideradas.

Las observaciones en las 100 semanas sucesivas, aparentemente, han de estar distribuidas independientemente, pero las tasas de rendimiento están correlacionadas, puesto que las tendencias de las acciones se mueven conforme a las condiciones económicas generales.

El problema es 5-dimensional, con las variables aleatorias $X_i, i = 1, \dots, 5$, que notan, respectivamente, el índice de rendimiento semanal de las 5 empresas citadas. $(x_1, \dots, x_5)_\alpha^T, \alpha = 1, \dots, 100$, representa las 100 observaciones de la tasa de rendimiento semanal para las acciones de *Allied Chemical, DuPont, Union Carbide, Exxon y Texaco*, respectivamente. Las tres primeras (*Allied Chemical, DuPont y Union Carbide*) son empresas pertenecientes al sector industrial químico, mientras que las dos últimas (*Exxon y Texaco*) pertenecen al sector petrolífero. Es por ello que una solución con dos factores parece, en apariencia, lo más apropiado.

Tomando $m = 1$ y $m = 2$ factores, Johnson y Wichern analizan las posibles soluciones factoriales, mediante el método de componentes principales y mediante el método de máxima verosimilitud, para la extracción de factores. Mediante la metodología de extracción basada en componentes principales, llegan a la conclusión de que un modelo con 2 factores resulta más adecuado para la explicación de la estructura factorial del problema. Se puede observar (Tabla 4.1) que existe un primer factor que

está correlacionado con todas las variables (*factor de mercado*) y un segundo factor que contrapone las industrias químicas (*Allied Chemical, DuPont y Union Carbide*) frente a las petrolíferas (*Exxon y Texaco*). Este segundo factor podría ser considerado un *factor de industria*. El uso de un procedimiento de rotación facilitará la interpretación de esta estructura factorial. El objetivo de la rotación es que cada factor tenga algunos pesos altos y el resto cercanos a cero, y que cada variable esté saturada en un solo factor, lo cual ayudará al investigador a interpretar mejor la solución factorial obtenida inicialmente.

| Variable | Solución con un factor | | Solución con dos factores | | |
|---|------------------------|---|---------------------------|--------|---|
| | F_1 | Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$ | F_1 | F_2 | Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$ |
| Allied Chemical | 0,783 | 0,39 | 0,783 | -0,217 | 0,34 |
| DuPont | 0,773 | 0,40 | 0,733 | -0,458 | 0,20 |
| Union Carbide | 0,794 | 0,37 | 0,794 | -0,234 | 0,31 |
| Exxon | 0,713 | 0,49 | 0,713 | 0,472 | 0,27 |
| Texaco | 0,712 | 0,49 | 0,712 | 0,524 | 0,22 |
| Proporción acumulada del total de la varianza explicada | 0,571 | | 0,571 | 0,733 | |

Tabla 4.1. Análisis factorial sobre los datos de “*stock-prices*”, realizado mediante el método de Componentes principales.

Johnson y Wichern asumen que la solución con dos factores resulta la más adecuada (tanto por interpretabilidad teórica como por el aumento considerable en la variabilidad total explicada que conlleva la inclusión de un segundo factor en el modelo). Emplean para $m = 2$ el método de extracción de factores basado en máxima verosimilitud. En la Tabla 4.2, se muestran las cargas factoriales estimadas, comunales, varianzas específicas, proporción de la varianza muestral to-

tal explicada por cada factor¹ y cargas factoriales tras la aplicación de la rotación Varimax. Las comunalidades correspondientes a la factorización por máxima verosimilitud, vienen dadas por $\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2$, donde \hat{l}_{ij} es la carga factorial estimada para la variable i en el factor j . Así, por ejemplo, $\hat{h}_1^2 = \hat{l}_{11}^2 + \hat{l}_{12}^2 = (0,684)^2 + (0,189)^2 = 0,50$.

| Variable | Solución con dos factores | | | Rotación Varimax | |
|---|---------------------------|--------|---|------------------|-------|
| | F_1 | F_2 | Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$ | F_1 | F_2 |
| Allied Chemical | 0,684 | 0,189 | 0,50 | 0,601 | 0,376 |
| DuPont | 0,694 | 0,517 | 0,25 | 0,850 | 0,163 |
| Union Carbide | 0,681 | 0,248 | 0,47 | 0,643 | 0,334 |
| Exxon | 0,621 | -0,073 | 0,61 | 0,366 | 0,506 |
| Texaco | 0,792 | -0,442 | 0,18 | 0,208 | 0,884 |
| Proporción acumulada del total de la varianza explicada | 0,485 | 0,598 | | 0,335 | 0,598 |

Tabla 4.2. Análisis factorial sobre los datos de “*stock-prices*”, realizado mediante el método de Máxima verosimilitud, con rotación Varimax.

En la solución factorial sin rotar correspondiente a la aplicación del método de máxima verosimilitud, se puede apreciar la existencia de un primer factor, que se puede identificar como *factor de mercado*, y otro segundo factor que, en este caso, no parece tan fácil de interpretar, ya que puede, o bien marcar diferencias entre las industrias petrolíferas y químicas, o bien marcar diferencias específicamente entre las empresas *DuPont* y *Texaco*. Precisamente por eso tiene sentido realizar una rotación. Los

¹Obsérvese, con la ayuda de las Tablas 4.1 y 4.2, que la proporción acumulada del total de la variabilidad explicada por dos factores, es mayor en el método de componentes principales (0,733) que en el de máxima verosimilitud (0,598). No resulta una sorpresa, dado que el porcentaje de variabilidad total explicada suele ser favorable en componentes principales. Este hecho no es nada extraño, pues en el análisis factorial basado en componentes principales, se prima la explicación de las varianzas sobre la explicación de todas las varianzas-covarianzas, que es un objetivo más propio del análisis factorial basado en máxima verosimilitud.

factores rotados mediante la rotación ortogonal Varimax, indican que las empresas químicas (*Allied Chemical*, *DuPont* y *Union Carbide*) son las que presentan una mayor ponderación o carga en el primer factor, mientras que las empresas petrolíferas (*Exxon* y *Texaco*) presentan las mayores cargas en el segundo factor. Por tanto, los dos factores rotados marcan diferencias entre ambos tipos de industrias. Es difícil, no obstante, asignar un nombre a ambos factores. El primer factor podría representar las fuerzas económicas que causan movimientos conjuntos en la cotización de las empresas químicas, mientras que el segundo factor podría representar las condiciones económicas que afectan a las empresas petrolíferas.

De este modo, se puede aceptar que la solución compuesta por dos factores es la más adecuada, tal y como concluyen Johnson y Wichern. Mediante el uso de criterios de información estadística, se afronta desde otra perspectiva el problema de la selección del número adecuado de factores.

Existe prácticamente unanimidad entre los criterios de información planteados en este trabajo (ver Tabla 2.1 en la sección 2.5) a la hora de seleccionar el modelo factorial adecuado para los datos de *stock-prices*. A excepción de los criterios *BIC* y *CAIC*, el resto de criterios de información selecciona como modelo factorial más adecuado, de entre los posibles candidatos (los modelos con $m = 1$ y $m = 2$ factores), el modelo que corresponde a la solución compuesta por dos factores. Se presentan además los pesos asociados a cada modelo en función del criterio empleado (Tabla 4.3). Recuérdese que, en cada caso y para cada uno de los criterios de información empleados, el peso asociado a un modelo puede interpretarse como la evidencia en favor de dicho modelo o la probabilidad de que dicho modelo sea el mejor, dados los datos y el conjunto de modelos candidatos.

| m | AIC | BIC | $HQIC$ | $AIC3$ | $CAIC$ | $CAIC^*$ | AIC_c | $DBIC$ | BIC^* | RIC | p |
|-----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-------|
| 1 | 1290,74 | 1316,79 | 1301,29 | 1300,74 | 1326,79 | 1295,21 | 1293,74 | 1298,42 | 1285,21 | 1290,01 | <0,01 |
| 2 | 1283,17 | 1319,64 | 1297,93 | 1297,17 | 1333,64 | 1289,42 | 1288,88 | 1293,91 | 1275,42 | 1280,84 | 0,449 |
| | $w_i(AIC)$ | $w_i(BIC)$ | $w_i(HQIC)$ | $w_i(AIC3)$ | $w_i(CAIC)$ | $w_i(CAIC^*)$ | $w_i(AIC_c)$ | $w_i(DBIC)$ | $w_i(BIC^*)$ | $w_i(RIC)$ | |
| 1 | 0,022 | 0,806 | 0,157 | 0,144 | 0,968 | 0,052 | 0,081 | 0,095 | 0,007 | 0,010 | |
| 2 | 0,978 | 0,194 | 0,843 | 0,856 | 0,032 | 0,948 | 0,919 | 0,905 | 0,993 | 0,990 | |

Tabla 4.3. Valores de los criterios de información y pesos asociados en el ejemplo de “*stock-prices*”. m = Número de factores; p = p -valor asociado al estadístico χ^2 ; $w_i(IC)$ =Peso asociado al modelo con i factores en base al criterio IC .

A la vista de los resultados expuestos en la Tabla 4.3, se observa en efecto que *BIC* y *CAIC*, al contrario que el resto de criterios, y en contraposición a las propias conclusiones e interpretación teórica realizada por Johnson y Wichern, seleccionan la solución compuesta por un solo factor. Por otra parte, mediante la aplicación del test asintótico de la chi-cuadrado de Bartlett, se aceptaría que el modelo con dos factores resulta una solución adecuada, mientras que se rechazaría la solución compuesta por un solo factor, a un nivel de confianza, en ambos casos, igual al 95 %.

En el caso del *BIC* y *CAIC*, aunque ambos seleccionan el mismo número de factores, $m = 1$, la evidencia en favor de un solo factor, aún siendo bastante elevada en ambos casos, es mayor en el caso del *CAIC* respecto del *BIC* (96,8% frente a 80,6%), en consonancia con el hecho reflejado en la sección 3.2, donde se veía que el *CAIC* es aún más proclive a la elección de modelos de menor dimensión, y tratado específicamente en la sección 3.5 con la presentación, precisamente, del caso en que los dos criterios seleccionaban un modelo con un único factor. Igualmente, en el caso del resto de criterios en base a los cuales se seleccionaría la solución compuesta por $m = 2$ factores, los pesos en favor de la elección de dicho modelo son en todos los casos elevados, con probabilidades siempre superiores a 0,80. Los mayores pesos se registran en ese caso para dos criterios con definición bastante similar, el *BIC** de Yang y Yang, y el *RIC* de Rissanen, cuya relación se estudió en la sección 3.3.

De la aplicación de estos criterios de información para la selección del modelo factorial adecuado en el ejemplo de *stock-prices*, se puede extraer como conclusión que los criterios de información de Schwarz y Bozdogan, *BIC* y *CAIC*, a partir de su propia definición, muestran cierta predisposición a seleccionar como modelo más adecuado aquel con un menor número de parámetros, en este caso factores. *BIC* y *CAIC* son criterios asintóticamente consistentes, y penalizan el exceso de parámetros de forma rigurosa, infraestimando a veces el número de factores en tamaños muestrales limitados.

En estas situaciones, podría resultar conveniente el uso de las versiones para *BIC* y *CAIC* propuestas por Yang y Yang, ajustadas de acuerdo al tamaño muestral, y de la versión de Draper para el *BIC*, *DBIC*. Estos nuevos criterios de información, que son realmente modificaciones basadas en el tamaño muestral, pueden suponer una mejora en el rendimiento del *BIC* y *CAIC* originales, en

tamaños muestrales pequeños o moderados. En cualquier caso, para la elección definitiva del modelo factorial adecuado, la interpretabilidad teórica de la solución obtenida sigue siendo el aspecto más relevante al que se enfrenta el investigador.

4.2. Ejemplo de los 24 tests psicológicos (Harman, 1976)

El siguiente ejemplo, que es un clásico en la literatura concerniente al análisis factorial, es atribuido generalmente a Harman (1976, ver matriz de correlaciones en pág. 124), que realiza un análisis factorial (considerando diversos métodos de extracción de factores y diversos procedimientos de rotación) sobre 24 variables que representan 24 tests psicológicos. Los datos corresponden a 145 niños de séptimo y octavo curso, que estudian en una escuela de Chicago. Los datos empleados fueron extraídos de los datos originales de Holzinger y Swineford (1939), que presentaban las puntuaciones obtenidas por 301 niños de séptimo y octavo curso en 26 tests. Los niños pertenecían a dos escuelas de Chicago: *Pasteur* y *Grant-White*. El conjunto de datos original² ha sido ampliamente citado en textos clásicos de análisis factorial, aunque casi siempre utilizando sólo una parte del conjunto de datos original. Es el caso de Harman (1976), quien emplea exclusivamente las puntuaciones obtenidas por los niños pertenecientes a la escuela *Grant-White* ($n = 145$) en los 24 tests que aparecen en la Tabla 4.4. Los nombres de las variables aparecen en el idioma original, en la notación dada por Harman. Con diversos fines, este conjunto de datos ha sido empleado en sucesivos estudios (ver, por ejemplo, Jöreskog (1978) y Akaike (1987)).

En un primer momento, y tras realizar un análisis mediante la metodología de factores principales, Harman sugiere que una solución compuesta por 5 factores podría proporcionar un modelo factorial adecuado para la descripción de la estructura correlacional de las 24 variables consideradas, en base a la aplicación del *scree test* (Cattell, 1976) para la determinación del número de factores. No obstante, a la vista de la variabilidad explicada por cada factor, una solución compuesta por 4 factores podría ser tenida en cuenta, puesto que la simplicidad de dicha solución compensaría la ligera pérdida de varianza explicada. Así, desde el punto de vista práctico y en aras de encontrar

²Los datos originales de Holzinger y Swineford son accesibles a partir del paquete MBESS, utilizando el programa de distribución libre *R* (*R Development Core Team*, 2008). Información acerca de la sintaxis y el contenido de los datos, puede obtenerse en la siguiente dirección web: <http://rss.acs.unt.edu/Rdoc/library/MBESS/html/HS.data.html>

un modelo parsimonioso, una solución con 4 factores resultaría del todo adecuada (Harman, 1976; pág. 195). Harman etiqueta los 4 factores de la siguiente forma, tras realizar una rotación Varimax sobre la solución inicial obtenida mediante el método del centroide para la extracción de factores: *Razonamiento verbal, Velocidad, Deducción y Memoria*. Una solución bastante similar a la obtenida por Jöreskog (1978), aplicando el método de máxima verosimilitud y posteriormente una rotación Varimax. Aunque, mediante la aplicación del test asintótico de Bartlett, la idoneidad de la solución con 4 factores era rechazada a un nivel de confianza del 95 %, Jöreskog aceptaba que el modelo con 4 factores constituía un ajuste razonable desde el punto de vista teórico. La solución, fruto de la distinta metodología empleada, difería ligeramente de la de Harman, variando sobre todo la composición del tercer factor. En su caso, Jöreskog etiquetaba a los factores de la siguiente forma: *Razonamiento verbal, Velocidad, Visualización espacial y Memoria*.

En la Tabla 4.4, se presentan las cargas factoriales asociadas a la solución con 4 factores, tras aplicar el método de máxima verosimilitud y posteriormente una rotación (Varimax), con el fin de obtener una interpretación más clara de la solución. En la Tabla 4.5, se puede apreciar, empleando la misma metodología, la solución rotada para 5 factores. En efecto, la variabilidad explicada apenas aumenta con la inclusión de un quinto factor, por lo que la solución con 4 factores parece la más apropiada desde el punto de vista práctico, en consonancia con las interpretaciones realizadas por Harman y Jöreskog. Al haberse empleado la misma metodología que utilizó Jöreskog, los 4 factores que conforman la solución dada en la Tabla 4.4, pueden ser etiquetados como: *Razonamiento verbal* (F_1), *Visualización espacial* (F_2), *Velocidad* (F_3) y *Memoria* (F_4). Aunque aparece resaltado en negrita el factor donde cada test presenta una mayor carga, en algunos casos se puede considerar que determinados tests están asociados con diversos factores a la vez. Sobre todo, es el caso de los tests que miden *Deducción* (nº 20) y *Razonamiento de problemas* (nº 22), con cargas por encima de 0,30 en los factores *Razonamiento verbal*, *Visualización espacial* y *Memoria*. Otro ejemplo es el test nº 24, que evalúa *Problemas de Aritmética* y satura en los factores *Razonamiento verbal*, *Velocidad* y *Memoria*. En cada caso, las cargas son superiores a 0,30. Recuérdese que las cargas de una variable en un factor pueden ser consideradas como una correlación entre dicha variable y dicho factor. En el ámbito de las ciencias sociales, correlaciones superiores a 0,30 merecen ser consideradas.

| Variable | Método de Máxima Verosimilitud con rotación Varimax | | | | |
|---|---|--------------|--------------|--------------|---|
| | F_1 | F_2 | F_3 | F_4 | Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$ |
| 1. Visual Perception | 0,167 | 0,693 | 0,178 | 0,151 | 0,44 |
| 2. Cubes | 0,113 | 0,442 | 0,098 | 0,086 | 0,77 |
| 3. Paper Form Board | 0,139 | 0,561 | -0,020 | 0,119 | 0,65 |
| 4. Flags | 0,234 | 0,523 | 0,098 | 0,096 | 0,65 |
| 5. General Information | 0,742 | 0,185 | 0,212 | 0,136 | 0,35 |
| 6. Paragraph Comprehension | 0,764 | 0,206 | 0,073 | 0,224 | 0,32 |
| 7. Sentence Completion | 0,796 | 0,200 | 0,161 | 0,095 | 0,21 |
| 8. Word Classification | 0,573 | 0,322 | 0,254 | 0,137 | 0,48 |
| 9. Word meaning | 0,814 | 0,185 | 0,044 | 0,221 | 0,25 |
| 10. Addition | 0,170 | -0,130 | 0,833 | 0,175 | 0,23 |
| 11. Code | 0,191 | 0,134 | 0,498 | 0,346 | 0,58 |
| 12. Counting Dots | 0,026 | 0,217 | 0,716 | 0,093 | 0,43 |
| 13. Straight-Curved Capitals | 0,190 | 0,450 | 0,544 | 0,059 | 0,46 |
| 14. Word Recognition | 0,186 | 0,060 | 0,069 | 0,545 | 0,66 |
| 15. Number Recognition | 0,104 | 0,116 | 0,072 | 0,511 | 0,71 |
| 16. Figure Recognition | 0,072 | 0,415 | 0,080 | 0,500 | 0,57 |
| 17. Object-Number | 0,151 | 0,047 | 0,215 | 0,616 | 0,55 |
| 18. Number-Figure | 0,033 | 0,287 | 0,333 | 0,463 | 0,59 |
| 19. Figure-Word | 0,157 | 0,218 | 0,158 | 0,379 | 0,76 |
| 20. Deduction | 0,381 | 0,402 | 0,105 | 0,321 | 0,58 |
| 21. Numerical Puzzles | 0,176 | 0,375 | 0,433 | 0,235 | 0,59 |
| 22. Problem Reasoning | 0,373 | 0,396 | 0,113 | 0,327 | 0,59 |
| 23. Series Completion | 0,371 | 0,486 | 0,235 | 0,258 | 0,50 |
| 24. Arithmetic Problems | 0,365 | 0,146 | 0,448 | 0,325 | 0,50 |
| Proporción acumulada del total de la varianza explicada | 0,153 | 0,271 | 0,381 | 0,479 | |

Tabla 4.4. Análisis factorial sobre los 24 tests tratados por Harman, realizado mediante el método de Máxima verosimilitud, con rotación Varimax; solución con 4 factores.

| Variable | Método de Máxima Verosimilitud con rotación Varimax | | | | | Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$ |
|---|---|--------------|--------------|--------------|--------------|---|
| | F_1 | F_2 | F_3 | F_4 | F_5 | |
| 1. Visual Perception | 0,175 | 0,655 | 0,125 | 0,181 | 0,207 | 0,45 |
| 2. Cubes | 0,108 | 0,442 | 0,087 | 0,095 | 0,002 | 0,78 |
| 3. Paper Form Board | 0,134 | 0,559 | -0,047 | 0,112 | 0,093 | 0,65 |
| 4. Flags | 0,230 | 0,533 | 0,090 | 0,081 | 0,012 | 0,65 |
| 5. General Information | 0,738 | 0,189 | 0,192 | 0,149 | 0,055 | 0,36 |
| 6. Paragraph Comprehension | 0,772 | 0,187 | 0,032 | 0,248 | 0,124 | 0,29 |
| 7. Sentence Completion | 0,798 | 0,214 | 0,143 | 0,088 | 0,050 | 0,29 |
| 8. Word Classification | 0,571 | 0,343 | 0,239 | 0,128 | 0,042 | 0,48 |
| 9. Word meaning | 0,808 | 0,202 | 0,033 | 0,219 | 0,007 | 0,26 |
| 10. Addition | 0,181 | -0,108 | 0,845 | 0,180 | 0,026 | 0,21 |
| 11. Code | 0,195 | 0,066 | 0,423 | 0,436 | 0,418 | 0,41 |
| 12. Counting Dots | 0,030 | 0,232 | 0,694 | 0,102 | 0,129 | 0,44 |
| 13. Straight-Curved Capitals | 0,186 | 0,433 | 0,479 | 0,077 | 0,538 | 0,25 |
| 14. Word Recognition | 0,185 | 0,061 | 0,044 | 0,552 | 0,080 | 0,65 |
| 15. Number Recognition | 0,104 | 0,122 | 0,059 | 0,509 | -0,003 | 0,71 |
| 16. Figure Recognition | 0,070 | 0,406 | 0,056 | 0,509 | 0,054 | 0,57 |
| 17. Object-Number | 0,154 | 0,072 | 0,210 | 0,595 | -0,027 | 0,57 |
| 18. Number-Figure | 0,032 | 0,300 | 0,322 | 0,458 | 0,004 | 0,60 |
| 19. Figure-Word | 0,156 | 0,221 | 0,144 | 0,378 | 0,045 | 0,76 |
| 20. Deduction | 0,373 | 0,461 | 0,127 | 0,293 | -0,194 | 0,51 |
| 21. Numerical Puzzles | 0,172 | 0,398 | 0,431 | 0,238 | -0,001 | 0,57 |
| 22. Problem Reasoning | 0,364 | 0,423 | 0,114 | 0,320 | -0,069 | 0,57 |
| 23. Series Completion | 0,362 | 0,542 | 0,248 | 0,231 | -0,115 | 0,45 |
| 24. Arithmetic Problems | 0,368 | 0,179 | 0,495 | 0,321 | -0,068 | 0,48 |
| Proporción acumulada del total de la varianza explicada | 0,152 | 0,275 | 0,377 | 0,477 | 0,503 | |

Tabla 4.5. Análisis factorial sobre los 24 tests tratados por Harman, realizado mediante el método de Máxima verosimilitud, con rotación Varimax; solución con 5 factores.

A diferencia del ejemplo de *stock-prices* tratado en la sección anterior, en este caso el número de posibles factores, m , varía en un rango más amplio, puesto que el número de variables originales, p , es mayor. Dado que 24 es el número de variables consideradas, y atendiendo al cumplimiento de la condición

$$(p - m)^2 > p + m,$$

el número de factores que pueden constituir la solución al problema oscila, a priori, entre 1 y 17. Aplicando al problema de selección del número adecuado de factores los criterios de información presentados en este estudio, se obtendría (ver Tabla 4.6) un rango de posibles soluciones que varía entre 3 y 5 factores, en función del criterio empleado. Esto ofrecería diversas ideas a un investigador que se enfrentase por vez primera al problema, o que no tenga una base teórica consistente para asegurar la adecuación de una solución concreta.

En este caso, se conocen y se han estudiado algunas propiedades de ciertos criterios de información (como la selección de un menor número de parámetros en el caso del *BIC* y el *CAIC*). Además, en base a diversas investigaciones previas, las ya citadas de Harman (1976) y Jöreskog (1978), la solución con 4 factores parece la más adecuada para explicar la estructura correlacional del problema. Así, se puede analizar el rendimiento en este ejemplo de los criterios considerados.

| m | AIC | BIC | $HQIC$ | $AIC3$ | $CAIC$ | $CAIC^*$ | $DBIC$ | BIC^* | RIC | p |
|-----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-------|
| 2 | 8814,26 | 9025,60 | 8900,13 | 8885,26 | 9096,60 | 8871,93 | 8895,11 | 8800,93 | 8809,49 | <0,01 |
| 3 | 8723,73 | 9000,57 | 8836,22 | 8816,73 | 9093,57 | 8799,28 | 8829,65 | 8706,28 | 8715,37 | <0,01 |
| 4 | 8688,86 | 9028,21 | 8826,75 | 8802,86 | 9142,21 | 8781,47 | 8818,69 | 8667,47 | 8676,96 | 0,037 |
| 5 | 8688,98 | 9087,87 | 8851,06 | 8822,98 | 9221,87 | 8797,84 | 8841,59 | 8663,84 | 8673,65 | 0,156 |
| 6 | 8698,13 | 9153,57 | 8883,19 | 8851,13 | 9306,57 | 8822,43 | 8872,38 | 8669,43 | 8679,50 | 0,264 |
| | $w_i(AIC)$ | $w_i(BIC)$ | $w_i(HQIC)$ | $w_i(AIC3)$ | $w_i(CAIC)$ | $w_i(CAIC^*)$ | $w_i(DBIC)$ | $w_i(BIC^*)$ | $w_i(RIC)$ | |
| 2 | 0,000 | 0,000 | 0,000 | 0,000 | 0,180 | 0,000 | 0,000 | 0,000 | 0,000 | |
| 3 | 0,000 | 1,000 | 0,009 | 0,001 | 0,820 | 0,000 | 0,004 | 0,000 | 0,000 | |
| 4 | 0,513 | 0,000 | 0,991 | 0,999 | 0,000 | 1,000 | 0,996 | 0,132 | 0,153 | |
| 5 | 0,482 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,808 | 0,797 | |
| 6 | 0,005 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,050 | 0,043 | |

Tabla 4.6. Valores de los criterios de información y pesos asociados en el ejemplo de Harman de los 24 tests. m = Número de factores; p = p -valor asociado al estadístico χ^2 ; $w_i(IC)$ =Peso asociado al modelo con i factores en base al criterio IC .

Tanto el *BIC* de Schwarz como el *CAIC* de Bozdogan, vuelven a seleccionar un modelo con un menor número de parámetros (ver Tabla 4.6) del que en apariencia es el más adecuado, de acuerdo a la relación entre parsimonia y variabilidad explicada. Por el contrario, son el *RIC* de Rissanen y el *BIC** de Yang y Yang, los criterios que sobreparametrizan el modelo.

Aceptando como modelo más adecuado el compuesto por 4 factores, el resto de criterios ofrecen un buen rendimiento en este ejemplo. Este rendimiento es notable en el caso de *CAIC** y *DBIC*, es decir, los ajustes para tamaños muestrales limitados propuestos por Yang y Yang, y por Draper, respectivamente, para los criterios *CAIC* y *BIC*. Los pesos asociados son en ambos casos elevados, y se puede concluir, en ambos casos, que seleccionan claramente el modelo con 4 factores. Algo similar sucede con otros dos criterios: el *HQIC* y el *AIC3*. Mención especial merece el caso del *AIC* original de Akaike, que, si bien selecciona el modelo que se considera correcto (en el sentido de “más adecuado”), lo hace con escasa diferencia, en lo que a probabilidad (peso) se refiere, con respecto a la solución compuesta por 5 factores. Es por ello que su rendimiento no es tan óptimo en este caso y merecería prácticamente la misma consideración una solución compuesta por 4 factores que una solución compuesta por 5 factores (51,3% frente a 48,2%), en el caso de que el *AIC* fuese el criterio elegido para realizar la selección.

Recuérdese, tal y como se expuso en la sección 2.5, que el *AIC_c*, propuesto por Sugiura y estudiado por Hurvich y Tsai, puede ser empleado en análisis factorial si se cumple la condición $n > k_j + 2$, donde k_j es el número de parámetros libres estimados dentro del modelo j . Es recomendable que, para cada modelo candidato j , el tamaño muestral n sea suficientemente mayor que $k_j + 2$, para evitar grandes fluctuaciones en los valores del criterio, las cuales pueden afectar a su rendimiento. En este ejemplo, existen modelos entre los candidatos para los que no se cumple lo anterior, por lo que se omite hacer uso de este criterio en este caso.

En la Tabla 4.6, se muestran los valores de cada criterio considerado, y su correspondiente peso asociado, para los modelos con 2, 3, 4, 5 y 6 factores. Los criterios de información empleados en este ejemplo, sugieren soluciones compuestas por un número de factores que en ningún caso es inferior a 3 ni superior a 5. Para hacer más sencilla la visualización de la solución y economizar espacio, se omite el valor asociado a los respectivos criterios en los modelos restantes. No obstante, el cálculo

de los pesos asociados a cada criterio ha sido realizado en base a los 17 modelos candidatos, razón por la cual, en algunos casos, la suma de la correspondiente columna de los pesos es ligeramente inferior a 1.

Obsérvese que, en base al test asintótico de Bartlett, la hipótesis de una solución factorial compuesta por 4 factores sería rechazada a un nivel de confianza del 95 % ($\chi^2(186) = 221,78; p = 0,037$), mientras que a ese mismo nivel de confianza sí sería aceptada una solución con 5 factores ($\chi^2(166) = 184,41; p = 0,156$). ¿Significa esto que la solución compuesta por 5 factores resulta más adecuada? No necesariamente; tal y como se expone al principio del Capítulo 2 de esta Tesis, bajo circunstancias en que el tamaño muestral es amplio y el número de factores extraídos es relativamente pequeño respecto al número de variables originales, el test de Bartlett tiende a retener más factores de los necesarios. Tucker y Lewis (1973) estudian también el ejemplo planteado por Harman³ y estiman, a partir de su metodología propuesta (basada en un coeficiente de fiabilidad), que, en efecto, esta es una de las situaciones en las que el rendimiento del test de Bartlett no es óptimo y el modelo más adecuado es el de 4 factores, dadas las dificultades de interpretación que plantea una solución a este ejemplo compuesta por 5 factores, incluso después de un procedimiento de rotación. Téngase en cuenta que, si el tamaño muestral es amplio y el número de factores extraídos es relativamente pequeño respecto del número de variables originales, la hipótesis nula, H_0 , del test de Bartlett, dado en la sección 1.3.1, será usualmente rechazada, llevando este hecho a una retención de más factores comunes, o, lo que es lo mismo, sobreestimación del número de parámetros.

4.3. Ejemplo de las puntuaciones en Decatlón

El tercer y último ejemplo de este capítulo trata sobre una disciplina atlética, el Decatlón, que ha sido objeto de estudio en numerosos trabajos, con el fin determinar la estructura correlacional de las 10 pruebas que conforman esta disciplina. El Decatlón masculino tiene carácter olímpico desde 1912, mientras que el Decatlón femenino aún no ha tenido un campeonato oficial.

En el Decatlón, cada atleta compite en 10 eventos que tienen lugar a lo largo de 2 días consecutivos.

³En el caso de Tucker y Lewis, la referencia a los datos expuestos por Harman es la correspondiente a la segunda edición del texto de Harman, *Modern factor analysis*, publicada en 1967.

El primer día tienen lugar las pruebas de *100 metros*, *Salto de longitud*, *Lanzamiento de peso*, *Salto de altura* y *400 metros*; el segundo día se celebran las pruebas restantes: *110 metros vallas*, *Lanzamiento de disco*, *Salto con pértiga*, *Lanzamiento de jabalina* y *1500 metros*.

Las carreras de *100 metros*, *110 metros vallas*, *400 metros* y *1500 metros*, son *eventos de pista*; las pruebas restantes, *eventos de campo*. Basadas en los tiempos (en el caso de *eventos de pista*) y en las distancias (en el caso de *eventos de campo*), puntuaciones en cada prueba son asignadas a cada atleta participante. Por ejemplo, en la carrera de *1500 metros*, a un mejor tiempo le corresponderá una mayor puntuación; en la prueba de *Salto de longitud*, un mejor salto será recompensado con una mayor puntuación en esa prueba. La puntuación final en la competición será la suma de las puntuaciones obtenidas en cada prueba.

Todo atleta que no tome una salida en un evento de pista, o no haga un intento en un evento de campo, será excluido de la competición, considerándose que la abandonado, y no apareciendo, por tanto, en la clasificación final.

En definitiva, las marcas obtenidas en las distintas pruebas, son valoradas con arreglo a unas tablas de puntuación de la Federación Internacional de Atletismo (IAAF; *International Amateur Athletic Federation*). Estas tablas han sufrido diversas modificaciones a lo largo de la historia. En la actualidad, se emplean unas tablas vigentes desde 1984, que fueron actualizadas en 1998. El ganador de la competición será aquel atleta que consiga una mayor puntuación. En caso de empate, vencerá el atleta que en mayor número de pruebas haya conseguido el mayor número de puntos; si esto no bastase para deshacer el empate, resultaría vencedor el atleta que mayor puntuación (o segunda mayor puntuación, etc.) hubiese conseguido en una prueba cualquiera.

Linden (1977) llevó a cabo un análisis factorial sobre las puntuaciones en Decatlón, considerando los 8 certámenes olímpicos celebrados desde la Segunda Guerra Mundial. El tamaño muestral es $n = 160$, y la matriz de correlaciones muestrales aparece reproducida en los textos de Johnson y Wichern (1988, pág. 396) y Basilewsky (1994, pág. 408). En ambos textos se analiza este ejemplo mediante la metodología de extracción de factores basada en componentes principales y en máxima verosimilitud. Johnson y Wichern, a la vista de los cuatro primeros autovalores de la matriz de correlaciones, sugieren una solución factorial con $m = 3$ o $m = 4$ factores. Después, tras una

interpretación de las cargas factoriales, llegan a la conclusión de que una solución con 4 factores es la más adecuada. La rotación fue necesaria porque la interpretación de los factores subyacentes que conformaban la solución inicial no era del todo evidente. Después de emplear la rotación Varimax, las soluciones rotadas son similares para el caso de extracción por componentes principales y para el caso de extracción por máxima verosimilitud (ver Tabla 4.7), cambiando sólo el orden de aparición en la solución de algunos factores. Los factores resultantes son etiquetados, siguiendo a Linden, de la siguiente forma: *Fuerza explosiva de brazos*, *Fuerza explosiva de piernas*, *Velocidad* y *Resistencia*. *Peso*, *Disco* y *Jabalina* son las pruebas con mayor influencia en el factor *Fuerza explosiva de brazos*, mientras que *Salto de longitud*, *Salto de altura*, *110 metros vallas* y *Salto con pértiga*, son las que presentan mayor ponderación en *Fuerza explosiva de piernas*. En el factor *Velocidad*, son *100 metros*, *400 metros*, y hay que considerar también el peso de *Salto de longitud*, las pruebas con mayor carga factorial. Finalmente, la carrera de *1500 metros*, que es además el último evento en disputarse de la competición, es la principal carga del último factor, *Resistencia*.

| Variable | Componentes principales | | | | Máxima verosimilitud | | | | | |
|---|-------------------------|--------------|--------------|--------------|--|--------------|--------------|--------------|--------------|--|
| | F_1 | F_2 | F_3 | F_4 | Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$ | F_1 | F_2 | F_3 | F_4 | Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$ |
| 100 metros | 0,884 | 0,136 | 0,156 | -0,113 | 0,16 | 0,167 | 0,857 | 0,246 | -0,138 | 0,16 |
| Salto de longitud | 0,631 | 0,194 | 0,515 | -0,006 | 0,30 | 0,240 | 0,477 | 0,580 | 0,011 | 0,38 |
| Lanzamiento de peso | 0,245 | 0,825 | 0,223 | -0,148 | 0,19 | 0,966 | 0,154 | 0,200 | -0,058 | 0,00 |
| Salto de altura | 0,239 | 0,150 | 0,750 | 0,076 | 0,35 | 0,242 | 0,173 | 0,632 | 0,113 | 0,50 |
| 400 metros | 0,797 | 0,075 | 0,102 | 0,468 | 0,13 | 0,055 | 0,709 | 0,236 | 0,330 | 0,33 |
| 110 metros vallas | 0,404 | 0,153 | 0,635 | -0,170 | 0,38 | 0,205 | 0,261 | 0,589 | -0,071 | 0,54 |
| Lanzamiento de disco | 0,186 | 0,814 | 0,147 | -0,079 | 0,28 | 0,697 | 0,133 | 0,180 | -0,009 | 0,46 |
| Salto con pértiga | -0,036 | 0,176 | 0,762 | 0,217 | 0,34 | 0,137 | 0,078 | 0,513 | 0,116 | 0,70 |
| Lanzamiento de jabalina | -0,048 | 0,735 | 0,110 | 0,141 | 0,43 | 0,416 | 0,019 | 0,175 | 0,002 | 0,80 |
| 1500 metros | 0,045 | -0,041 | 0,112 | 0,934 | 0,11 | -0,055 | 0,056 | 0,113 | 0,990 | 0,00 |
| Proporción acumulada del total de la varianza explicada | 0,21 | 0,42 | 0,61 | 0,73 | | 0,18 | 0,34 | 0,50 | 0,61 | |

Tabla 4.7. Solución con 4 factores, rotada mediante Varimax, para los métodos de Componentes principales y Máxima verosimilitud, en el ejemplo de Johnson y Wichern correspondiente a puntuaciones de Decatlón de Linden (1977).

Basilewsky obtiene idénticos resultados. En su caso, proporciona además una solución rotada de forma oblicua, que define unos patrones similares a los que proporciona la rotación Varimax. Basilewsky menciona el hecho de que la prueba de *Salto de longitud* correlaciona con el factor *Velocidad* y también con el factor *Fuerza explosiva de piernas*, dado que esta prueba requiere de ambas capacidades físicas.

Cox y Dunn (2002) analizan datos correspondientes a las puntuaciones obtenidas por los decatletas en los campeonatos mundiales de Atletismo celebrados entre 1991 y 1999. Aparte de estudiar la dimensionalidad, cuestionan si el Decatlón favorece a un determinado tipo de atletas, concluyendo que los atletas más hábiles en eventos de campo tienen ventaja sobre los atletas más hábiles en eventos de pista. En lo que al estudio de la dimensionalidad se refiere, hacen uso de una metodología distinta, basada en este caso en el análisis cluster, que les lleva a concluir que las disciplinas pueden ser agrupadas en 3 clusters. El primero de ellos, compuesto por pruebas que necesitan de la velocidad como habilidad para un buen rendimiento (*100 metros*, *400 metros* y *110 metros vallas* y *Salto de longitud*); el segundo cluster, lo conformarían *Lanzamiento de peso*, *Lanzamiento de disco*, *Lanzamiento de jabalina* y *Salto con pértiga*, mientras que en el tercer cluster se encontrarían las pruebas de *Salto de altura* y *1500 metros*. Aunque apoyado por los análisis realizados sobre los datos, es cuestionable si ese tercer cluster tiene sentido desde el punto de vista teórico.

Muy recientemente, el análisis de datos de la disciplina atlética de Decatlón y la búsqueda de una estructura correlacional que permita reducir la dimensionalidad, ha sido objeto de estudio por parte de Schomaker y Heumann (2011) y Ertel (2011).

Schomaker y Heumann consideran las marcas (y no las puntuaciones) en las diferentes pruebas de los 30 atletas que finalizaron la competición en los Juegos Olímpicos de Atenas 2004. Emplean el método de máxima verosimilitud y aplican el criterio de información de Akaike, *AIC*, para la selección del modelo adecuado, llegando a la conclusión de que una solución compuesta por 3 factores sería la más adecuada. La composición de los factores es, etiquetada de acuerdo a Schomaker y Heumann, la siguiente:

- Factor 1 (*Componente atlética y de velocidad*): *100 metros*, *400 metros*, *110 metros vallas* y *Salto de longitud*.

- Factor 2 (*Componente técnica y de fuerza*): *Lanzamiento de peso, Salto de altura y Lanzamiento de disco.*
- Factor 3 (*Resistencia y estatus especial del último evento*): *1500 metros.*

Obsérvese que en la solución propuesta por Schomaker y Heumann no aparecen las pruebas de *Lanzamiento de jabalina* y *Salto con pértiga*, las cuales no tienen un peso significativo en ninguno de los 3 factores extraídos. Los autores lo justifican desde el punto de vista de que ambos eventos son los más sofisticados dentro de un concurso de Decatlón.

Ertel (2011) también obtiene una solución basada en 3 factores, aunque en este caso la metodología empleada es totalmente distinta. El método de extracción de factores utilizado es el de componentes principales, seleccionando el número de factores en base a la regla de autovalores mayores que uno de Kaiser-Guttman y el *scree test* de Cattell. Ertel utiliza los datos de Zarnowski (1989), que corresponden a las puntuaciones de 233 decatletas participantes en 11 Juegos Olímpicos (1948-1988), como ejemplo de aplicación de un nuevo método de rotación desarrollado por Ertel, Varimin, basado en la estructura compleja. De acuerdo a esta metodología, Ertel sugiere 3 factores. El primero de ellos se podría considerar un factor de tamaño (*Energía atlética general*). El segundo contrapone la velocidad frente a la resistencia, mientras que el tercero es un factor de localización del gasto de energía (parte superior del cuerpo frente a parte inferior). Véanse en la Tabla 4.8 las cargas factoriales correspondientes a la solución propuesta por Ertel.

| Variable | Rotación Varimin a la solución inicial | | | |
|---|--|-------|-------|---|
| | F_1 | F_2 | F_3 | Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$ |
| 100 metros | 0,49 | 0,61 | 0,42 | 0,21 |
| Salto de longitud | 0,66 | 0,47 | 0,30 | 0,25 |
| Lanzamiento de peso | 0,69 | 0,51 | -0,37 | 0,14 |
| Salto de altura | 0,74 | 0,18 | 0,23 | 0,36 |
| 400 metros | 0,75 | 0,06 | 0,46 | 0,22 |
| 110 metros vallas | 0,65 | 0,47 | 0,30 | 0,25 |
| Lanzamiento de disco | 0,66 | 0,43 | -0,42 | 0,21 |
| Salto con pértiga | 0,82 | 0,16 | 0,10 | 0,33 |
| Lanzamiento de jabalina | 0,73 | 0,11 | -0,35 | 0,29 |
| 1500 metros | 0,65 | -0,65 | 0,23 | 0,11 |
| Proporción acumulada del total de la varianza explicada | 0,472 | 0,644 | 0,756 | |

Tabla 4.8. Solución de 3 factores propuesta por Ertel (2011) para las puntuaciones de decatlón recogidas por Zarnowski (1989), obtenida tras aplicación del método de Componentes principales y la rotación Varimin.

4.3.1. Análisis factorial de los datos de Decatlón correspondientes al período 2005 - 2009

En esta Tesis, se afronta el problema de estudiar la dimensionalidad del Decatlón desde una perspectiva actual. Se han considerado los datos correspondientes a competiciones celebradas a nivel mundial en el lustro 2005-2009. Los datos⁴ corresponden a los campeonatos mundiales disputados en Berlín (en 2009), Osaka (en 2007) y Helsinki (en 2005), y a los Juegos Olímpicos celebrados en

⁴Los datos para este ejemplo han sido obtenidos a partir del archivo de competiciones de la IAAF, disponible en la dirección web <http://www.iaaf.org/history>

Pekín en 2008. En los casos en que un atleta había competido en uno o más de estos campeonatos, se ha tenido en cuenta la puntuación obtenida en la última competición que finalizó, debido a la correlación potencial entre puntuaciones sucesivas para atletas que participaron en más de una competición. Se consideraron en cada caso atletas que habían finalizado la competición, es decir, atletas que presentaban puntuaciones en cada una de las 10 pruebas. Así, el conjunto de datos analizado consta de un total de 69 observaciones, correspondientes a cada una de las 10 pruebas que conforman el Decatlón.

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------------------------|--------|--------|--------|--------|--------|-------|--------|-------|-------|----|
| 1. 100 metros | 1 | | | | | | | | | |
| 2. Salto de longitud | 0,537 | 1 | | | | | | | | |
| 3. Lanzamiento de peso | 0,257 | 0,052 | 1 | | | | | | | |
| 4. Salto de altura | 0,049 | 0,334 | 0,227 | 1 | | | | | | |
| 5. 400 metros | 0,555 | 0,337 | -0,104 | -0,036 | 1 | | | | | |
| 6. 110 metros vallas | 0,300 | 0,284 | 0,156 | 0,318 | 0,299 | 1 | | | | |
| 7. Lanzamiento de disco | 0,088 | -0,086 | 0,704 | 0,190 | -0,232 | 0,044 | 1 | | | |
| 8. Salto con pértiga | 0,281 | 0,366 | 0,063 | 0,258 | 0,188 | 0,196 | 0,002 | 1 | | |
| 9. Lanzamiento de jabalina | 0,184 | 0,154 | 0,324 | 0,255 | 0,112 | 0,103 | 0,312 | 0,142 | 1 | |
| 10. 1500 metros | -0,116 | 0,039 | -0,278 | 0,043 | 0,279 | 0,061 | -0,440 | 0,128 | 0,114 | 1 |

Tabla 4.9. Matriz de correlaciones de las 10 pruebas de Decatlón para 69 atletas en el período 2005-2009.

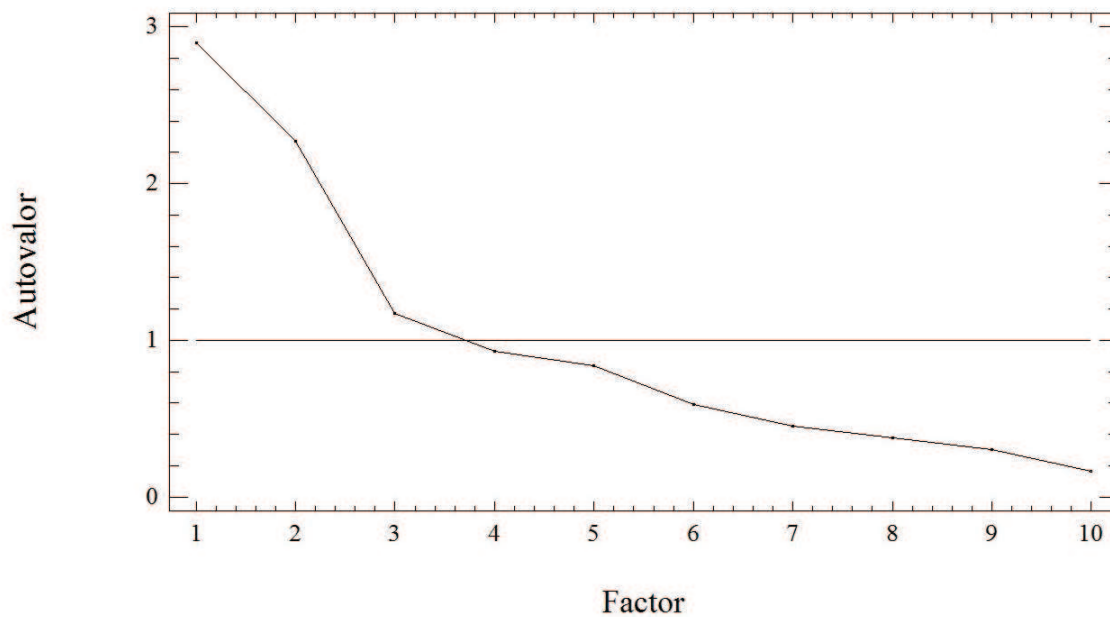


Figura 4.1: Gráfico de sedimentación correspondiente a los autovalores de la matriz de correlaciones dada en la Tabla 4.9

La matriz de correlaciones de los datos puede observarse en la Tabla 4.9. El método empleado para la extracción de factores es el de máxima verosimilitud y se realiza una rotación Varimax para facilitar la interpretación de la solución factorial. De acuerdo a la regla de autovalores mayores que uno y al *scree test* (ver gráfico de sedimentación en la Figura 4.1), 3 sería el número de factores seleccionado. Hay que decir, sin embargo, que el cuarto autovalor ($\lambda_4 = 0,993$) es un valor muy próximo a 1, por lo que podría ser conveniente, y merece al menos ser considerado, un modelo con 4 factores. Las soluciones rotadas para 3 y 4 factores, pueden ser observadas, respectivamente, en las Tablas 4.10 y 4.11.

| | Rotación Varimax a la solución inicial | | | |
|--|--|--------------|--------------|--|
| Variable | F_1 | F_2 | F_3 | Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$ |
| 100 metros | 0,253 | 0,962 | 0,101 | 0,00 |
| Salto de longitud | -0,074 | 0,530 | 0,455 | 0,51 |
| Lanzamiento de peso | 0,739 | 0,047 | 0,241 | 0,39 |
| Salto de altura | 0,078 | -0,046 | 0,729 | 0,46 |
| 400 metros | -0,179 | 0,615 | 0,094 | 0,58 |
| 110 metros vallas | 0,024 | 0,263 | 0,406 | 0,77 |
| Lanzamiento de disco | 0,911 | -0,166 | 0,159 | 0,12 |
| Salto con pértiga | -0,032 | 0,259 | 0,398 | 0,77 |
| Lanzamiento de jabalina | 0,280 | 0,081 | 0,351 | 0,79 |
| 1500 metros | -0,505 | -0,006 | 0,176 | 0,71 |
| Proporción acumulada del total de la varianza explicada | 0,182 | 0,358 | 0,490 | |

Tabla 4.10. Solución compuesta por 3 factores para los datos de Decatlón del período 2005-2009, obtenida empleando el método de Máxima verosimilitud y la rotación Varimax.

| Variable | Rotación Varimax a la solución inicial | | | | Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$ |
|--|--|--------------|--------------|--------------|--|
| | F_1 | F_2 | F_3 | F_4 | |
| 100 metros | 0,926 | 0,197 | 0,145 | -0,141 | 0,06 |
| Salto de longitud | 0,502 | -0,048 | 0,512 | -0,041 | 0,48 |
| Lanzamiento de peso | 0,068 | 0,792 | 0,105 | -0,141 | 0,34 |
| Salto de altura | -0,110 | 0,195 | 0,789 | 0,019 | 0,33 |
| 400 metros | 0,645 | -0,115 | 0,078 | 0,223 | 0,51 |
| 110 metros vallas | 0,254 | 0,096 | 0,386 | 0,034 | 0,78 |
| Lanzamiento de disco | -0,131 | 0,849 | 0,022 | -0,277 | 0,18 |
| Salto con pértiga | 0,251 | 0,046 | 0,371 | 0,094 | 0,79 |
| Lanzamiento de jabalina | 0,104 | 0,422 | 0,221 | 0,176 | 0,73 |
| 1500 metros | 0,051 | -0,194 | 0,086 | 0,975 | 0,00 |
| Proporción acumulada del total de la varianza explicada | 0,170 | 0,337 | 0,463 | 0,580 | |

Tabla 4.11. Solución compuesta por 4 factores para los datos de Decatlón del período 2005-2009, obtenida empleando el método de Máxima verosimilitud y la rotación Varimax.

Atendiendo a las cargas en los distintos factores, la solución con 4 factores parece más sencilla de interpretar, con unas saturaciones más definidas. Es apreciable el aumento de variabilidad explicada con la inclusión de un cuarto factor (58 % frente al 49 % de la solución con 3 factores). El primer factor está influenciado principalmente por las carreras de *100* y *400 metros*; la prueba de *Salto de longitud* muestra también un peso considerable en este factor. En el segundo factor, las pruebas de *Lanzamiento de disco*, *Lanzamiento de peso* y *Lanzamiento de jabalina*, son las que presentan mayor ponderación. Las pruebas de *Salto de altura*, otra vez *Salto de longitud*, *110 metros vallas* y *Salto con pértiga*, son las que presentan una mayor correlación con el tercer factor. Por último, la

carrera de *1500 metros* es la principal carga, la única considerable prácticamente, del cuarto factor. Ambas soluciones son factibles desde el punto de vista del test asintótico de Bartlett, que aceptaría la adecuación de ambas soluciones a un nivel de confianza del 95 %. Sin embargo, la solución compuesta por 4 factores es más sencilla de interpretar teóricamente y además coincide con la propuesta por Linden (1977), y corroborada por Johnson y Wichern (1988) y Basilewsky (1994). Así, los cuatro factores descritos en el párrafo anterior, pueden ser etiquetados, respectivamente, como *Velocidad*, *Fuerza explosiva de brazos*, *Fuerza explosiva de piernas* y *Resistencia*.

Parece lógico desde el punto de vista teórico sugerir la adecuación de la solución compuesta por 4 factores. A partir de la utilización de criterios de información estadística, se pueden obtener evidencias en favor de una u otra solución. Además, suponiendo adecuada la solución con 4 factores, se puede analizar el rendimiento de los criterios presentados. En la Tabla 4.12, se presentan los correspondientes resultados.

| m | AIC | BIC | $HQIC$ | AIC_3 | $CAIC$ | $CAIC^*$ | AIC_c | $DBIC$ | BIC^* | RIC | p |
|-----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-------|
| 1 | 1921,55 | 1966,23 | 1939,28 | 1941,55 | 1986,23 | 1923,25 | 1941,21 | 1929,48 | 1903,25 | 1909,33 | <0,01 |
| 2 | 1858,52 | 1923,31 | 1884,22 | 1887,52 | 1952,31 | 1860,97 | 1907,47 | 1870,01 | 1831,97 | 1838,78 | 0,017 |
| 3 | 1846,15 | 1928,81 | 1878,95 | 1883,15 | 1965,81 | 1849,28 | 1944,95 | 1860,81 | 1812,28 | 1819,56 | 0,475 |
| 4 | 1845,72 | 1944,02 | 1884,72 | 1889,72 | 1988,02 | 1849,44 | 2025,72 | 1863,15 | 1805,44 | 1813,06 | 0,943 |
| 5 | 1853,92 | 1965,62 | 1898,23 | 1903,92 | 2015,62 | 1858,15 | 2165,92 | 1873,73 | 1808,15 | 1816,01 | 0,928 |
| | $w_i(AIC)$ | $w_i(BIC)$ | $w_i(HQIC)$ | $w_i(AIC_3)$ | $w_i(CAIC)$ | $w_i(CAIC^*)$ | $w_i(AIC_c)$ | $w_i(DBIC)$ | $w_i(BIC^*)$ | $w_i(RIC)$ | |
| 1 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | |
| 2 | 0,001 | 0,940 | 0,063 | 0,098 | 0,999 | 0,001 | 1,000 | 0,008 | 0,000 | 0,000 | |
| 3 | 0,442 | 0,060 | 0,887 | 0,870 | 0,001 | 0,516 | 0,000 | 0,757 | 0,025 | 0,031 | |
| 4 | 0,548 | 0,000 | 0,049 | 0,033 | 0,000 | 0,476 | 0,000 | 0,235 | 0,774 | 0,789 | |
| 5 | 0,009 | 0,000 | 0,000 | 0,000 | 0,000 | 0,006 | 0,000 | 0,001 | 0,200 | 0,180 | |

Tabla 4.12. Valores de los criterios de información y pesos asociados en el ejemplo de las puntuaciones en las 10 pruebas del Decatlón. m = Número de factores; p = p -valor asociado al estadístico χ^2 ; $w_i(IC)$ =Peso asociado al modelo con i factores en base al criterio IC .

En este ejemplo de las puntuaciones en las pruebas de Decatlón, con datos correspondientes al período 2005-2009, las decisiones acerca del número adecuado de factores a extraer, varían en función del criterio de información empleado. Nuevamente, los criterios *BIC* y *CAIC* seleccionan un modelo con un número reducido de factores, 2, en este caso. También son 2 los factores seleccionados en base al *AIC_c*, aplicable en este ejemplo, aunque con una solución difícil de justificar. El *BIC** de Yang y Yang, y el *RIC* de Rissanen, son los que seleccionan, de entre los criterios considerados, el modelo con más factores, 4, que es además el modelo factorial que, a priori, en base a la interpretación teórica, parece más adecuado. El *HQIC*, el *AIC3* y el *DBIC*, seleccionan como modelo más adecuado el compuesto por 3 factores. En el caso del *DBIC*, existiría una probabilidad, moderadamente baja, en torno al 23,5%, de seleccionar la solución con 4 factores (ver peso asociado al modelo con 4 factores en la Tabla 4.12).

Consideración aparte merecen los casos del *AIC* y del *CAIC** de Yang y Yang. Si el investigador se centra exclusivamente en hallar el número de factores que minimiza el valor del criterio de información, no hay discusión posible. En base al *AIC*, se seleccionarían 4 factores, y, en base al *CAIC**, 3. Un análisis más exhaustivo de la situación, considerando los pesos asociados a cada modelo, demuestra que el rendimiento de ambos criterios no es tan diferente (como hace suponer el simple proceso de minimización) en este ejemplo. El *CAIC** seleccionaría por escaso margen un modelo con 3 factores (51,60% frente al 47,64% asociado al modelo con 4 factores), mientras que el *AIC* haría algo similar con el modelo compuesto por 4 factores (54,83% frente al 44,17% asociado al modelo con 3 factores).

Es momento ahora de volver a la Tabla 4.11; obsérvese el valor de las comunalidades. La décima variable, la prueba de *1500 metros*, presenta una comunalidad $h_{10}^2 = 1,00$ (lo que es equivalente a una varianza específica $\psi_{10} = 0,00$), indicando lo que parece ser un caso de Heywood, es decir, una solución impropia cuando se ajusta el modelo con 4 factores. Si se observa la solución con 3 factores mostrada en la Tabla 4.10, también se encontraba un caso de Heywood, en este caso en la primera variable, la prueba de *100 metros*.

Retornando a la solución con 4 factores, que es, de las obtenidas en este estudio, la que está sostenida por una mejor argumentación teórica (dado que coincide con la solución que tiene más arraigo en

la literatura especializada), es cierto que la prueba de *1500 metros* presenta unas características especiales: no sólo por las capacidades atléticas que requiere la prueba en sí, sino también por ser el último evento del Decatlón, donde los atletas pueden tener opciones de ganar una medalla o incluso no tener opciones de mejorar puestos en la competición, con las variaciones en motivación que ambos hechos conllevan con respecto a otras pruebas. Surgen entonces una serie de interrogantes: ¿Es adecuada la metodología empleada para el análisis de la dimensionalidad de las 10 pruebas de Decatlón? Si se realiza un análisis factorial sobre las 9 variables restantes (excluyendo *1500 metros*), ¿la estructura correlacional define 3 factores, que son los que acompañaban al de *Resistencia* en la Tabla 4.11?

La respuesta a esta segunda cuestión viene dada en la Tabla 4.13. Si se extraen 3 factores para explicar la estructura correlacional de esas 9 variables, los 3 factores pueden ser etiquetados como: *Velocidad* (influenciado principalmente por las variables *100 metros*, *400 metros* y *Salto de longitud*), *Fuerza explosiva de brazos* (con las cargas principales de *Lanzamiento de disco*, *Lanzamiento de peso* y *Lanzamiento de jabalina*) y *Fuerza explosiva de piernas* (donde la mayor asociación con este factor corresponde a las pruebas de *Salto de altura*, también *Salto de longitud*, *110 metros vallas* y *Salto con pértiga*).

| | Rotación Varimax a la solución inicial | | | |
|---|--|--------------|--------------|---|
| Variable | F_1 | F_2 | F_3 | Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$ |
| 100 metros | 0,912 | 0,239 | 0,094 | 0,10 |
| Salto de longitud | 0,540 | -0,021 | 0,469 | 0,49 |
| Lanzamiento de peso | 0,056 | 0,819 | 0,095 | 0,32 |
| Salto de altura | -0,082 | 0,191 | 0,816 | 0,29 |
| 400 metros | 0,644 | -0,164 | 0,071 | 0,55 |
| 110 metros vallas | 0,276 | 0,091 | 0,378 | 0,77 |
| Lanzamiento de disco | -0,135 | 0,869 | 0,017 | 0,23 |
| Salto con pértiga | 0,270 | 0,027 | 0,356 | 0,80 |
| Lanzamiento de jabalina | 0,096 | 0,357 | 0,235 | 0,81 |
| Proporción acumulada del total de la varianza explicada | 0,192 | 0,379 | 0,516 | |

Tabla 4.13. Solución compuesta por 3 factores para los datos de 9 variables de Decatlón (se omite la prueba de *1500 metros*) del período 2005-2009, obtenida empleando el método de Máxima verosimilitud y la rotación Varimax.

Se puede decir entonces, después de abordar las dudas metodológicas, que, dado que estas 9 variables mantienen la estructura correlacional dada en la Tabla 4.11, añadiendo un factor influenciado principalmente por la prueba de 1500 metros, se tendría la solución de 4 factores (*Velocidad, Fuerza explosiva de brazos, Fuerza explosiva de piernas y Resistencia*) para explicar la estructura correlacional del Decatlón.

¿Cómo podría afectar el posible caso de Heywood a la selección del número de factores adecuado mediante criterios de información, dada en la Tabla 4.12? A continuación, se muestran en la Tabla 4.14 los valores de los criterios de información y los pesos asociados, para la selección del número de factores adecuado para explicar la variabilidad de las 9 pruebas de Decatlón (omitiendo la prueba

de *1500 metros*). Acorde con la argumentación teórica, 3 debería ser el número adecuado de factores para explicar la estructura correlacional de esas 9 variables. En este caso, el número de factores para los modelos candidatos varía en un rango entre 1 y 4.

| m | AIC | BIC | $HQIC$ | $AIC3$ | $CAIC$ | $CAIC^*$ | AIC_c | $DBIC$ | BIC^* | RIC | p |
|-----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-------|
| 1 | 1724,59 | 1764,80 | 1740,54 | 1742,59 | 1782,80 | 1726,11 | 1740,10 | 1731,72 | 1708,11 | 1714,00 | <0,01 |
| 2 | 1670,09 | 1728,18 | 1693,14 | 1696,09 | 1754,18 | 1672,29 | 1706,97 | 1680,40 | 1646,29 | 1652,89 | 0,092 |
| 3 | 1659,93 | 1733,65 | 1689,18 | 1692,93 | 1766,65 | 1662,72 | 1729,93 | 1673,00 | 1629,72 | 1636,77 | 0,939 |
| 4 | 1667,11 | 1754,24 | 1701,68 | 1706,11 | 1793,24 | 1670,41 | 1784,26 | 1682,57 | 1631,41 | 1638,79 | 0,979 |
| | $w_i(AIC)$ | $w_i(BIC)$ | $w_i(HQIC)$ | $w_i(AIC3)$ | $w_i(CAIC)$ | $w_i(CAIC^*)$ | $w_i(AIC_c)$ | $w_i(DBIC)$ | $w_i(BIC^*)$ | $w_i(RIC)$ | |
| 1 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | |
| 2 | 0,006 | 0,939 | 0,121 | 0,170 | 0,998 | 0,008 | 1,000 | 0,024 | 0,000 | 0,000 | |
| 3 | 0,967 | 0,061 | 0,877 | 0,829 | 0,002 | 0,971 | 0,000 | 0,968 | 0,700 | 0,733 | |
| 4 | 0,027 | 0,000 | 0,002 | 0,001 | 0,000 | 0,021 | 0,000 | 0,008 | 0,300 | 0,267 | |

Tabla 4.14. Valores de los criterios de información y pesos asociados para el análisis de las puntuaciones en 9 pruebas del Decatlón (omitiendo la prueba de 1500 metros). m = Número de factores; p = p -valor asociado al estadístico χ^2 ; $w_i(IC)$ =Peso asociado al modelo con i factores en base al criterio IC .

Como se puede apreciar en la Tabla 4.14, el número de factores mayoritariamente seleccionado es 3. BIC , $CAIC$ y AIC_c continúan seleccionando un número de factores menor que el deseado. Obsérvese que $CAIC^*$, $DBIC$, $HQIC$ y AIC_3 , vuelven a seleccionar 3 factores, que en este caso, en que el análisis se realiza sobre 9 variables (tras eliminar del análisis la variable *1500 metros* ante la posibilidad de una solución impropia), es el número de factores, en apariencia adecuado, para explicar la estructura correlacional de esas 9 variables. AIC , BIC^* y RIC muestran un buen rendimiento en este ejemplo, y ahora seleccionan 3 factores, si bien en el caso del BIC^* y del RIC la evidencia en favor de la solución con 3 factores no es tan elevada (ver pesos asociados en la Tabla 4.14).

Capítulo 5

Aplicación de la metodología basada en criterios de información para el estudio de la dimensionalidad de un cuestionario

En este capítulo, se presenta un ejemplo de cómo la metodología basada en criterios de información puede ser aplicada al estudio de la dimensionalidad de un cuestionario, resolviendo el problema del número adecuado de factores a extraer. Una de las etapas clave para llevar a cabo con éxito el proceso de validación de un instrumento de medida (un cuestionario, una escala) es el análisis de la validez factorial (también llamada *validez de constructo*), que incluye el estudio de la dimensionalidad y de la estructura correlacional.

El instrumento *FH – 15* (*Fear of Hypoglycemia 15-item scale*) ha sido desarrollado y validado por Anarte *et al.* (2011) para evaluar miedo a las hipoglucemias en pacientes con diabetes tipo 1. Está compuesto por 15 ítems negativos (ver Tabla 5.1) medidos en una escala Likert de rango 1 - 5. El instrumento posee una adecuada consistencia interna y temporal, así como validez factorial, convergente, discriminante y de criterio, que lo hacen útil en la práctica clínica. La población donde se llevó a cabo la validación del cuestionario *FH – 15*, estuvo compuesta por 229 pacientes con diabetes tipo 1 de la Unidad de Diabetes del Servicio de Endocrinología y Nutrición del Hospital Carlos Haya de Málaga. La matriz de correlaciones muestrales entre los 15 ítems que conforman el cuestionario, puede ser observada en la Tabla 5.2.

- FH1. ¿Con qué frecuencia teme no reconocer los síntomas de una hipoglucemia?
- FH2. ¿Con qué frecuencia tiene miedo de no saber qué hacer ante una hipoglucemia?
- FH3. ¿Con qué frecuencia tiene miedo de que le ocurra una hipoglucemia en el trabajo?
- FH4. ¿Con qué frecuencia tiene miedo de que le ocurra una hipoglucemia fuera del ámbito hospitalario/sanitario?
- FH5. ¿Con qué frecuencia tiene miedo de que le ocurra una hipoglucemia estando solo?
- FH6. ¿Con qué frecuencia evita situaciones sociales (reuniones, salidas, etc.) por miedo a padecer una hipoglucemia?
- FH7. ¿Con qué frecuencia deja de hacer cosas que antes hacía por miedo a padecer una hipoglucemia?
- FH8. ¿Con qué frecuencia sufre hipoglucemias que le impiden conducir o usar una máquina?
- FH9. ¿Con qué frecuencia sufre hipoglucemias que le incapacitan para trabajar?
- FH10. ¿Con qué frecuencia sufre hipoglucemias que interfieren en sus actividades de tiempo libre?
- FH11. ¿Con qué frecuencia sufre hipoglucemias que interfieren en su vida familiar?
- FH12. ¿Con qué frecuencia sufre hipoglucemias que interfieren en su vida social?
- FH13. ¿Con qué frecuencia le preocupa si perderá el conocimiento por una hipoglucemia?
- FH14. ¿Con qué frecuencia teme quedarse dormido por miedo a sufrir una hipoglucemia durante la noche?
- FH15. ¿Con qué frecuencia tiene miedo de iniciar un viaje o unas vacaciones por miedo a sufrir una hipoglucemia?

Tabla 5.1. Relación de ítems que conforman el cuestionario *FH* – 15. Las opciones de respuesta a cada ítem son: 1 (Nunca), 2 (Casi nunca), 3 (Algunas veces), 4 (Casi siempre), 5 (Todos los días).

| Ítem | FH1 | FH2 | FH3 | FH4 | FH5 | FH6 | FH7 | FH8 | FH9 | FH10 | FH11 | FH12 | FH13 | FH14 | FH15 |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| FH1 | 1 | | | | | | | | | | | | | | |
| FH2 | 0,346 | 1 | | | | | | | | | | | | | |
| FH3 | 0,493 | 0,332 | 1 | | | | | | | | | | | | |
| FH4 | 0,405 | 0,460 | 0,384 | 1 | | | | | | | | | | | |
| FH5 | 0,540 | 0,489 | 0,516 | 0,530 | 1 | | | | | | | | | | |
| FH6 | 0,213 | 0,315 | 0,352 | 0,320 | 0,210 | 1 | | | | | | | | | |
| FH7 | 0,220 | 0,332 | 0,347 | 0,340 | 0,308 | 0,546 | 1 | | | | | | | | |
| FH8 | 0,254 | 0,243 | 0,224 | 0,371 | 0,233 | 0,235 | 0,281 | 1 | | | | | | | |
| FH9 | 0,270 | 0,269 | 0,145 | 0,278 | 0,220 | 0,183 | 0,313 | 0,368 | 1 | | | | | | |
| FH10 | 0,308 | 0,354 | 0,410 | 0,305 | 0,327 | 0,437 | 0,524 | 0,366 | 0,415 | 1 | | | | | |
| FH11 | 0,304 | 0,357 | 0,275 | 0,256 | 0,273 | 0,278 | 0,353 | 0,341 | 0,460 | 0,448 | 1 | | | | |
| FH12 | 0,353 | 0,403 | 0,399 | 0,359 | 0,412 | 0,452 | 0,484 | 0,405 | 0,431 | 0,623 | 0,640 | 1 | | | |
| FH13 | 0,460 | 0,342 | 0,404 | 0,297 | 0,538 | 0,169 | 0,293 | 0,234 | 0,230 | 0,293 | 0,245 | 0,323 | 1 | | |
| FH14 | 0,439 | 0,399 | 0,378 | 0,339 | 0,400 | 0,287 | 0,302 | 0,269 | 0,216 | 0,288 | 0,312 | 0,320 | 0,489 | 1 | |
| FH15 | 0,266 | 0,358 | 0,383 | 0,383 | 0,315 | 0,555 | 0,586 | 0,346 | 0,254 | 0,427 | 0,274 | 0,493 | 0,283 | 0,321 | 1 |

Tabla 5.2. Matriz de correlaciones muestrales entre los ítems del $FH - 15$ ($n = 229$).

Los autores del instrumento llevaron a cabo un análisis factorial mediante componentes principales para estudiar la validez factorial. En este capítulo se llevará a cabo empleando el método de máxima verosimilitud. Si las soluciones factoriales y las comunales son similares en ambos casos, podrá argumentarse que el supuesto de normalidad no afecta, en este caso, a la solución factorial, y que la aplicación de componentes principales o máxima verosimilitud da lugar a resultados sustancialmente idénticos. En ese caso, se puede optar por emplear el método de máxima verosimilitud para la extracción de factores, siendo aplicable entonces la metodología basada en criterios de información. Así, se podrá analizar el rendimiento de estos últimos en base a la solución que resulta en apariencia más adecuada.

5.1. Validez factorial del $FH - 15$

Anarte *et al.* (2011) realizaron un análisis factorial basado en el método de componentes principales, con objeto de estudiar la estructura correlacional y las posibles dimensiones subyacentes del cuestionario. Se empleó la rotación oblicua Promax para factores no ortogonales, permitiendo así la existencia de correlaciones entre los factores obtenidos. Se consideró adecuada la solución con 3 factores, en base a la variabilidad explicada (cerca al 60%) y la interpretación teórica. Los 3 factores obtenidos fueron etiquetados como *Miedo*, *Evitación* e *Interferencia*, que son respectivamente los factores F_1 , F_2 y F_3 de la solución dada en la Tabla 5.3. La rotación Promax resultó adecuada y los factores estuvieron claramente correlacionados (Tabla 5.4).

| | Rotación Promax a la solución inicial | | | |
|--|---------------------------------------|--------------|--------------|--|
| Ítem | F_1 | F_2 | F_3 | Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$ |
| FH1 | 0,811 | -0,184 | 0,087 | 0,40 |
| FH2 | 0,500 | 0,134 | 0,123 | 0,57 |
| FH3 | 0,637 | 0,293 | -0,179 | 0,46 |
| FH4 | 0,546 | 0,161 | 0,046 | 0,55 |
| FH5 | 0,878 | -0,057 | -0,049 | 0,31 |
| FH6 | -0,082 | 0,938 | -0,123 | 0,29 |
| FH7 | -0,040 | 0,793 | 0,075 | 0,33 |
| FH8 | 0,057 | 0,019 | 0,596 | 0,59 |
| FH9 | -0,047 | -0,186 | 0,904 | 0,36 |
| FH10 | -0,002 | 0,413 | 0,466 | 0,41 |
| FH11 | 0,007 | -0,048 | 0,823 | 0,36 |
| FH12 | 0,054 | 0,308 | 0,591 | 0,31 |
| FH13 | 0,794 | -0,128 | 0,008 | 0,45 |
| FH14 | 0,658 | 0,027 | 0,019 | 0,53 |
| FH15 | 0,056 | 0,821 | -0,058 | 0,33 |
| Proporción acumulada del total de la varianza explicada | 0,399 | 0,504 | 0,583 | |

Tabla 5.3. Matriz de configuración correspondiente al análisis factorial realizado sobre los ítems del $FH - 15$, mediante el método de Componentes principales y rotación Promax.

| Factor | Miedo | Evitación | Interferencia |
|---------------|-------|-----------|---------------|
| Miedo | 1 | | |
| Evitación | 0,509 | 1 | |
| Interferencia | 0,495 | 0,538 | 1 |

Tabla 5.4. Matriz de correlaciones entre los 3 factores correspondientes a la solución dada en la Tabla 5.3.

A continuación, se lleva a cabo en esta Tesis el mismo procedimiento, cambiando el método de extracción de factores por el de máxima verosimilitud. Se mantiene el uso de la rotación Promax, pues se estima que los factores resultantes van a estar correlacionados. La solución compuesta por 3 factores (Tabla 5.5) es esencialmente la misma que la obtenida mediante el método de componentes principales. Cambia el orden en que algunos son extraídos. En este caso, los factores F_1 , F_2 y F_3 de la Tabla 5.5 son *Miedo*, *Interferencia* y *Evitación*, explicando conjuntamente algo más del 48 % de la variabilidad total (el porcentaje explicado se reduce al emplear máxima verosimilitud), y están claramente correlacionados (Tabla 5.6).

| | Rotación Promax a la solución inicial | | | |
|--|---------------------------------------|--------------|--------------|--|
| Ítem | F_1 | F_2 | F_3 | Varianzas específicas $\hat{\psi}_i = 1 - \hat{h}_i^2$ |
| FH1 | 0,729 | 0,080 | -0,148 | 0,51 |
| FH2 | 0,450 | 0,128 | 0,107 | 0,63 |
| FH3 | 0,568 | -0,064 | 0,202 | 0,56 |
| FH4 | 0,531 | -0,028 | 0,176 | 0,60 |
| FH5 | 0,884 | -0,047 | -0,090 | 0,34 |
| FH6 | -0,078 | -0,051 | 0,810 | 0,46 |
| FH7 | -0,017 | 0,065 | 0,719 | 0,43 |
| FH8 | 0,097 | 0,355 | 0,111 | 0,75 |
| FH9 | 0,010 | 0,622 | -0,069 | 0,66 |
| FH10 | 0,017 | 0,452 | 0,327 | 0,48 |
| FH11 | -0,031 | 0,895 | -0,151 | 0,38 |
| FH12 | 0,030 | 0,687 | 0,185 | 0,30 |
| FH13 | 0,693 | -0,001 | -0,068 | 0,57 |
| FH14 | 0,528 | 0,039 | 0,064 | 0,65 |
| FH15 | 0,064 | -0,084 | 0,786 | 0,40 |
| Proporción acumulada del total de la varianza explicada | 0,366 | 0,439 | 0,485 | |

Tabla 5.5. Matriz de configuración correspondiente al análisis factorial realizado sobre los ítems del $FH - 15$, mediante el método de Máxima verosimilitud y rotación Promax.

| Factor | Miedo | Interferencia | Evitación |
|---------------|-------|---------------|-----------|
| Miedo | 1 | | |
| Interferencia | 0,590 | 1 | |
| Evitación | 0,581 | 0,654 | 1 |

Tabla 5.6. Matriz de correlaciones entre los 3 factores correspondientes a la solución dada en la Tabla 5.5.

Los resultados obtenidos, en base al método de componentes principales y al de máxima verosimilitud, son muy similares, idénticos en lo que a estructura factorial se refiere, como puede verse en las matrices de configuración y comunalidades de las Tablas 5.3 y 5.5. Esta similaridad de los resultados obtenidos parece confirmar (ver Johnson y Wichern, 1988; pág. 415) que, en efecto, 3 es el número de dimensiones dentro del cuestionario $FH-15$. Además, como es usual en la práctica del análisis factorial, el hecho de que mediante ambos métodos se obtenga una solución similar, indica, indirectamente, un cierto grado de normalidad en la población de la que se ha extraído la muestra utilizada. Así, se puede emplear la metodología de extracción de factores por máxima verosimilitud y utilizar entonces criterios de información para seleccionar el número adecuado de factores, para posteriormente (considerando que, en este caso, 3 es el número adecuado de factores) analizar su rendimiento.

5.2. Análisis del rendimiento de criterios de información en el estudio de la dimensionalidad del $FH-15$

Considerando que 3 es el número adecuado de factores para explicar la estructura correlacional de los ítems del $FH-15$, o, lo que es lo mismo, el número de dimensiones subyacentes del cuestionario, se analiza en esta sección el rendimiento de los criterios de información expuestos en este trabajo. Para la aplicación de criterios de información al problema de la selección del número adecuado de factores, es necesario que se verifiquen los supuestos que permiten el uso de procedimientos basados en máxima verosimilitud. Tal y como se ha visto en la sección anterior, el tamaño de la muestra ($n = 229$) y la similaridad de las soluciones factoriales obtenidas mediante los métodos

de componentes principales y máxima verosimilitud, permiten suponer que no existe una desviación excesiva de la normalidad en los datos. Así, un análisis factorial realizado con extracción de factores mediante el método de máxima verosimilitud es factible, y, la metodología basada en criterios de información, adecuada.

En la Tabla 5.7, se presentan los valores de cada uno de los criterios de información estudiados para cada uno de los posibles modelos factoriales. Recuérdese que, en este caso, en que 15 son las variables originales, el número posible de factores m que conforman la solución al problema, oscila entre 1 y 9, pues en un análisis factorial el número de factores a extraer no puede exceder del mayor entero M tal que

$$m \leq M < \frac{1}{2} [2p + 1 - \sqrt{8p + 1}],$$

donde p es el número de variables originales.

| m | AIC | BIC | $HQIC$ | AIC_3 | $CAIC$ | $CAIC^*$ | AIC_c | $DBIC$ | BIC^* | RIC | p |
|-----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-------|
| 2 | 8645,83 | 8786,91 | 8696,78 | 8679,83 | 8830,91 | 8691,46 | 8658,45 | 8706,04 | 8647,46 | 8655,07 | <0,01 |
| 3 | 8565,65 | 8761,37 | 8644,61 | 8622,65 | 8818,37 | 8637,72 | 8605,91 | 8656,61 | 8580,72 | 8588,84 | <0,01 |
| 4 | 8560,78 | 8797,71 | 8656,36 | 8629,78 | 8866,71 | 8648,02 | 8623,69 | 8670,90 | 8579,02 | 8587,52 | 0,043 |
| 5 | 8560,57 | 8835,27 | 8671,39 | 8640,57 | 8915,27 | 8661,72 | 8650,94 | 8688,24 | 8581,72 | 8590,51 | 0,178 |
| 6 | 8564,09 | 8873,13 | 8688,77 | 8654,09 | 8963,13 | 8677,89 | 8686,31 | 8707,72 | 8587,89 | 8596,91 | 0,357 |
| 7 | 8568,52 | 8908,46 | 8705,66 | 8667,52 | 9007,46 | 8693,69 | 8726,33 | 8726,51 | 8594,69 | 8603,90 | 0,569 |
| | $w_i(AIC)$ | $w_i(BIC)$ | $w_i(HQIC)$ | $w_i(AIC_3)$ | $w_i(CAIC)$ | $w_i(CAIC^*)$ | $w_i(AIC_c)$ | $w_i(DBIC)$ | $w_i(BIC^*)$ | $w_i(RIC)$ | |
| 2 | 0,000 | 0,000 | 0,000 | 0,000 | 0,002 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | |
| 3 | 0,036 | 1,000 | 0,997 | 0,972 | 0,998 | 0,994 | 1,000 | 0,999 | 0,252 | 0,295 | |
| 4 | 0,414 | 0,000 | 0,003 | 0,027 | 0,000 | 0,006 | 0,000 | 0,001 | 0,588 | 0,571 | |
| 5 | 0,461 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,153 | 0,128 | |
| 6 | 0,079 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,007 | 0,005 | |
| 7 | 0,009 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | |

Tabla 5.7. Valores de los criterios de información y pesos asociados en el ejemplo del $FH - 15$. $m =$ Número de factores; $p =$ p -valor asociado al estadístico χ^2 ; $w_i(IC)$ =Peso asociado al modelo con i factores en base al criterio IC .

Obsérvese en la Tabla 5.7 que el test asintótico de Bartlett rechazaría a un nivel de confianza del 95 % cualquier solución con menos de 5 factores, sobreparametrizando de esta forma el número adecuado de factores.

En lo que a los criterios de información respecta, la mayoría de criterios expuestos en esta Tesis selecciona el número adecuado de factores, 3. En relación con los ejemplos tratados en el capítulo anterior, cabe destacar que el *BIC* y el *CAIC* no infraestiman en este caso (donde el tamaño muestral, $n = 229$, es más elevado) el número de factores, y seleccionan la solución con 3 factores, al igual que *DBIC* y *CAIC**, que son versiones del *BIC* y el *CAIC*, ajustadas al tamaño muestral. El *HQIC*, el *AIC3* y el *AIC_c* también seleccionan un modelo con 3 factores. El *AIC_c* es aplicable en este ejemplo y muestra un buen rendimiento, superior al del *AIC* original, que, para este tamaño muestral, sobreestima claramente el número de factores.

El valor mínimo del *AIC*, en este ejemplo, se alcanza para 5 factores. La evidencia sería parecida para el modelo con 5 y para el modelo con 4 factores (46,1 % frente a un 41,4 %), pero, en cualquier caso, el número de factores extraídos en base a este criterio sería superior al adecuado. El *BIC** y el *RIC* sobreestiman también el número de factores, seleccionando una solución con 4. En ambos casos, existe cierta evidencia, aunque moderadamente baja, en favor de la solución con 3 factores: 25,2 % si se utiliza el *BIC** y 29,5 % si es el *RIC* el criterio empleado.

En la Tabla 5.7, se muestran los valores de los criterios y los pesos asociados para modelos con un número de factores entre 2 y 7. Se omiten el resto de modelos, dado que la evidencia en favor de dichos modelos, respecto a cada uno de los criterios de información estudiados, sería inferior al 0,05 %. Hay que notar, asimismo, que en el caso de todos los criterios que seleccionan el modelo con 3 factores, la evidencia en favor de dicho modelo es notablemente elevada, siendo superior al 99 % en todos los casos (con la excepción del *AIC3*, en cuyo caso es de un 97,2 %).

Conclusiones

Una de las decisiones más importantes que hay que tomar en un análisis factorial, es la elección del número de factores comunes, m . En palabras de Johnson y Wichern (1988, pág. 415), la elección del método para la extracción (máxima verosimilitud, componentes principales, factorización de ejes principales, etc.) y el procedimiento de rotación, son decisiones menos cruciales. De hecho, los análisis factoriales más satisfactorios son aquellos donde diversos métodos de extracción y rotación dan lugar, sustancialmente, a la misma solución, confirmando la estructura factorial.

La elección del número de factores adecuado en un análisis factorial mediante el método de máxima verosimilitud (AFMV), se ha efectuado, tradicionalmente, empleando el test asintótico de la chi-cuadrado de Bartlett (Bartlett, 1950). Sin embargo, el test de Bartlett, que tiene en cuenta el estadístico basado en la razón de verosimilitudes, no siempre resulta válido, y el nivel de significación no es ajustado en el proceso secuencial de contraste.

El estadístico basado en la razón de verosimilitudes se aproxima asintóticamente a una distribución chi-cuadrado cuando el vector de variables originales sigue una distribución normal multivariante, por lo que Basilewsky (1994, págs. 385-387) considera que su aplicación requiere tamaños muestrales moderadamente amplios. El propio Bartlett (1950) sugiere $n - m \geq 50$, donde n es el número de variables originales y m el número de factores extraídos. Geweke y Singleton (1980) consideran que este estadístico puede ser más fiable en muestras pequeñas (considerando al menos 30 observaciones) de lo que se piensa, siempre y cuando sean satisfechas las condiciones usuales de regularidad; sin embargo, si la matriz de cargas factoriales no es de rango completo por columnas, la teoría asintótica no es adecuada.

Además, el test de Bartlett tiende, en general, a rechazar la adecuación del modelo (rechazar la

hipótesis nula) con un número m de factores, cuando el tamaño muestral es grande y m es un valor relativamente pequeño respecto a p , el número de variables originales.

De acuerdo a Sclove (1987), Akaike (1987) y Bozdogan y Ramírez (1988), la teoría clásica de contraste de hipótesis no es apropiada en el contexto de la selección del modelo adecuado en análisis factorial, dado que no se trata de contrastar una hipótesis particular, sino de afrontar un problema de decisión múltiple. Además, este tipo de procedimientos no penalizan la sobreparametrización.

De hecho, en los problemas objeto de estudio de esta Tesis, el test asintótico de Bartlett se comporta de forma inconsistente. Obsérvese el caso del estudio de la dimensionalidad del cuestionario $FH-15$, dado en el Capítulo 5, donde el test de Bartlett no selecciona soluciones con menos de 5 factores a un nivel de confianza del 95 %, sobreparametrizando claramente el número de factores considerado adecuado, 3. El tamaño muestral en ese caso es $n = 229$. En el ejemplo de Harman del Capítulo 4, con un tamaño muestral $n = 145$, rechaza la solución con 4 factores, que es la aceptada en la literatura y que proporciona un mayor equilibrio entre complejidad y variabilidad explicada. En definitiva, en ambos casos en que el tamaño muestral es moderado, sobreparametriza la solución. Además, no es un procedimiento de decisión múltiple; simplemente acepta o rechaza una hipótesis concreta (con unos niveles de significación que permanecen fijos). Así, en casos como el del ejemplo de aplicación dado en el Capítulo 3, en que se aceptan en base al test de Bartlett las dos soluciones posibles, resulta necesario el uso de otra metodología para extraer el número adecuado de factores. Otros métodos para la selección del número de factores, frecuentemente utilizados e independientes del método empleado para la extracción de factores, son la regla de autovalores mayores que uno (Guttman, 1953; Kaiser, 1960) y el *scree test* de Cattell (1966), que se basa en la representación gráfica de los autovalores de la matriz de correlaciones, de mayor a menor, y en el punto del gráfico de sedimentación donde el descenso empieza a suavizarse. Zwick y Velicer (1986) apoyan el uso de otras metodologías, como el análisis paralelo de Horn (1965), que compara los autovalores de la matriz de correlaciones con los obtenidos a partir de variables normales incorreladas generadas aleatoriamente, o el *MAP* de Velicer (*Minimum Average Partial method*; 1976), este último originalmente propuesto en análisis de componentes principales y análisis de componentes imagen, y basado en el análisis de matrices de correlaciones parciales.

La utilización de criterios basados en Teoría de la Información, supone un enfoque distinto y más reciente a la cuestión de seleccionar el modelo adecuado en un análisis factorial llevado a cabo mediante el método de máxima verosimilitud. Estos criterios no son de uso exclusivo en el análisis factorial y su aplicación se extiende a otro tipo de procedimientos donde seleccionar el modelo adecuado sea el objetivo. De hecho, este tipo de criterios fueron diseñados originalmente, en general, para seleccionar modelos de regresión o series temporales.

Mientras que procedimientos basados en contrastes de hipótesis no penalizan el exceso de parámetros porque usualmente es empleado un modelo saturado como referencia, los criterios de información equilibran el ajuste del modelo y su complejidad, tomando la forma de una función de verosimilitud penalizada, que es el valor negativo del logaritmo de la función de verosimilitud más un término de penalización, que aumenta conforme lo hace el número de parámetros.

La selección del modelo factorial adecuado mediante criterios de información, no se encuentra entre las opciones de análisis de los paquetes estadísticos más utilizados. Es un suceso poco frecuente encontrar trabajos de AFMV donde se haga uso de criterios de información para la selección del modelo, y, cuando se encuentran estos trabajos, suelen emplear los que son los criterios de información más conocidos: generalmente, el *AIC* de Akaike, y, a veces, el *BIC* de Schwarz o el *CAIC* de Bozdogan.

El *CAIC* y el *BIC* son empleados, usualmente, bajo la creencia de que proporcionan un mejor rendimiento que el *AIC*. Sin obviar el hecho constatado por Woodroffe (1982) de que el *AIC* no es consistente (es decir, en general el *AIC* no seleccionará el modelo correcto cuando el tamaño muestral n tienda a infinito) y no penaliza el exceso de parámetros, lo cierto es que el *BIC* y el *CAIC* (que son consistentes, sí, pero asintóticamente) pueden dar una solución errónea al problema cuando se trabaja con tamaños muestrales limitados: en su afán por penalizar el exceso de parámetros, pueden seleccionar un número de parámetros (factores, en el contexto del AFMV en que se desarrolla esta Tesis) inferior al deseado.

Así, en problemas con tamaños muestrales limitados, una herramienta adecuada puede ser el uso de versiones del *CAIC* y el *BIC* para muestras finitas. A la vista de los resultados obtenidos en los distintos ejemplos tratados en esta Tesis y la comparación con soluciones propuestas por otros

autores (teniendo en cuenta siempre la argumentación teórica), parece recomendable la utilización en problemas reales, con tamaños muestrales moderados, del $CAIC^*$ de Yang y Yang y del $DBIC$ de Draper, ajustes en base al tamaño muestral realizados respectivamente sobre el $CAIC$ y el BIC originales. Tanto el $CAIC^*$ como el $DBIC$ son aconsejables para muestras finitas.

Yang y Yang proponen también (al igual que para el $CAIC$) una versión ajustada al tamaño muestral para el BIC , BIC^* . Dicho ajuste es muy similar, en su definición, al criterio de información de Rissanen, RIC . Ambos criterios tienen una definición tan parecida que, salvo circunstancias excepcionales, será seleccionado el mismo modelo en base a ambos, tal y como se expone en la sección 3.3 de esta Tesis. Concretamente, BIC^* y RIC tienden a seleccionar un número elevado de factores, pudiendo a veces sobreparametrizar la solución.

En cuanto al resto de criterios de información analizados en esta Tesis, rara vez han sido aplicados a la selección del modelo adecuado en análisis factorial. El AIC_c , que fue definido originalmente por Sugiura, es una corrección del AIC para tamaños muestrales pequeños, siendo aplicable en análisis factorial siempre y cuando el tamaño muestral, n , sea mayor que $k_j + 2$, para cada modelo candidato con un número de factores j , siendo k_j el número de parámetros libres a estimar dentro de dicho modelo. Téngase en cuenta que, asintóticamente, el AIC_c y el AIC proporcionan los mismos valores. El rendimiento del AIC_c en los problemas estudiados donde era factible su aplicación, resulta dispar, aunque en el caso del ejemplo del cuestionario $FH - 15$ (con un tamaño muestral $n = 229$) proporcionaba un rendimiento claramente superior al AIC original. Obsérvese que, en este ejemplo real, que es (de los ejemplos presentados en este trabajo) el que presenta un tamaño muestral más elevado, el AIC sobreestima de forma contundente el número de factores adecuado. Otra de las versiones del AIC , el $AIC3$ de Andrews y Currim, con una penalización mayor para el número de parámetros extraídos, ofrece buenos resultados en los ejemplos tratados, similares a los que son obtenidos empleando el $HQIC$ de Hannan y Quinn.

Es especialmente importante evaluar el peso de la evidencia en favor del mejor modelo de acuerdo al procedimiento de minimización. Los valores brutos de un determinado criterio no permiten discernir cual es el peso de la evidencia en favor de un modelo con i factores frente a otro modelo con j factores. Mediante los pesos asociados, se pueden obtener evidencias en favor de cada uno de los modelos

candidatos. En la sección 2.5.1, se presenta una extensión de los pesos de Akaike y Schwarz para los criterios AIC y BIC al resto de criterios de información expuestos en esta Tesis.

Una de las causas, y quizás sea la principal, por las que un investigador que realiza un AFMV sigue empleando (sean cuales sean las circunstancias) para la selección del modelo factorial adecuado una metodología basada en la regla de autovalores mayores que uno, el *scree test* o el test asintótico de Bartlett, a pesar de las críticas que estos métodos han recibido en la literatura especializada, es su fácil disponibilidad en cualquier software; mientras que otras herramientas, que supuestamente ofrecen mejores resultados, no son de fácil acceso. Entonces, el investigador, ante la disyuntiva entre implementar un nuevo algoritmo para el cálculo de un nuevo procedimiento (que no es de utilización frecuente y es por lo general desconocido) y emplear la metodología más tradicional, opta por la segunda opción; pero esta opción no siempre resulta la más adecuada y su utilización puede dar lugar a conclusiones erróneas.

El desarrollo de software de distribución libre permite la implementación, por parte de particulares, de nuevos procedimientos. El programa R es uno de los máximos exponentes de este tipo de software y uno de los paquetes estadísticos más empleados, en los últimos tiempos, por la comunidad científica. En esta Tesis, se proporcionan distintos algoritmos, que pueden ser ejecutados como *script* en R , para el cálculo de los valores de los criterios de información en cada uno de los modelos candidatos (incluyendo la selección del modelo adecuado en base a cada uno de los criterios y proporcionando también el resultado de la aplicación del test de Bartlett en cada caso), y también para los pesos asociados a cada uno de los modelos, permitiendo así obtener probabilidades de adecuación de un modelo con m factores extraídos. Todo ello está disponible para cada uno de los criterios de información estudiados en esta Tesis.

Observando, en cada uno de los problemas planteados, la selección del modelo en base a cada criterio de información y el posterior análisis de los pesos asociados, se pueden extraer las siguientes conclusiones generales:

- Los criterios de información $CAIC^*$, $DBIC$, $AIC3$ y $HQIC$ fueron los que, en general, ofrecieron un mejor rendimiento, siendo además mejores los pesos, o evidencias, en el caso del $CAIC^*$ y el $DBIC$.

- El *BIC* y el *CAIC* originales suelen infraestimar el número de factores adecuado en tamaños muestrales limitados. En el ejemplo del *FH* – 15 (con un tamaño muestral $n = 229$) dado en el Capítulo 5, sí ofrecen un buen rendimiento; curiosamente, en ese mismo ejemplo, el *AIC* extrae más factores de los convenientes. Estos resultados sugieren que, a partir de tamaños muestrales de ese orden, el *BIC* y el *CAIC* pueden ofrecer un buen rendimiento, mientras que la aplicación del *AIC* original puede plantear problemas.
- El AIC_c es aplicable bajo determinadas condiciones y no siempre selecciona el modelo adecuado, aunque en ocasiones puede ofrecer un rendimiento mejor que el *AIC*; de hecho, Burnham y Anderson (2002, pág. 405) recomendaban su uso frente al *AIC* cuando $\frac{n}{k_j} < 40$ (téngase en cuenta que en AFMV ha de tenerse también $n > k_j + 2$ para cada modelo con j factores).
- El BIC^* y el *RIC* no ofrecen, en los ejemplos tratados, un buen rendimiento, extrayendo más factores de los, en apariencia, necesarios.
- La metodología basada en criterios de información para seleccionar el número adecuado de factores en AFMV, resulta una buena alternativa a otras metodologías más tradicionales, como el test asintótico de Bartlett. La utilización de algunos criterios de información concretos mejora sustancialmente los resultados obtenidos mediante otros procedimientos.

Así, si hubiese que recomendar la utilización de criterios de información concretos en AFMV, a la vista de los resultados obtenidos, se recomendaría la utilización, sobre todo en tamaños muestrales limitados, pero generalmente en problemas reales, de dos versiones, ajustadas para el tamaño muestral, del *CAIC* y el *BIC*, respectivamente el $CAIC^*$ de Yang y Yang, y el *DBIC* de Draper. Los originales *CAIC* y *BIC*, frecuentemente utilizados en la literatura, seleccionan el modelo adecuado a partir de ciertos tamaños muestrales. Para dichos tamaños muestrales, el $CAIC^*$ y el *DBIC* también se comportan, en apariencia, adecuadamente. Téngase en cuenta que, asintóticamente, tienden a los valores de los criterios originales.

Anexo A

Aplicación en el programa *R* de la metodología basada en criterios de información al AFMV

A continuación, se presenta un ejemplo de cómo sería la aplicación en la práctica (empleando software estadístico) de la metodología basada en criterios de información estadística al problema de seleccionar el número adecuado de factores en un análisis factorial mediante máxima verosimilitud (AFMV). Para ello, se emplean el programa *R* (información detallada acerca del uso de este software de distribución libre, puede ser encontrada en Crawley (2007, 2009)) y las funciones propuestas en esta Tesis para su implementación en él, dadas en el Capítulo 3.

Se considera uno de los ejemplos clásicos tratados en el Capítulo 4, concretamente el de “*stock-prices*” de Johnson y Wichern (1988, datos en págs. 374-375). Dicho ejemplo medía la tasa de rendimiento semanal para las acciones de cinco compañías (*Allied Chemical, DuPont, Union Carbide, Exxon y Texaco*) de la bolsa de Nueva York, desde Enero de 1975 hasta Diciembre de 1976. Los datos pueden verse en la Tabla A.1.

| Semana | <i>Allied Chemical</i> | <i>DuPont</i> | <i>Union Carbide</i> | <i>Exxon</i> | <i>Texaco</i> |
|--------|------------------------|---------------|----------------------|--------------|---------------|
| 1 | 0 | 0 | 0 | 0,039473 | 0 |
| 2 | 0,027027 | -0,044855 | -0,00303 | -0,014466 | 0,043478 |
| 3 | 0,122807 | 0,060773 | 0,088146 | 0,086238 | 0,078124 |
| 4 | 0,057031 | 0,029948 | 0,066808 | 0,013513 | 0,019512 |
| 5 | 0,06367 | -0,003793 | -0,039788 | -0,018644 | -0,024154 |
| 6 | 0,003521 | 0,050761 | 0,082873 | 0,074265 | 0,049504 |
| 7 | -0,045614 | -0,033007 | 0,002551 | -0,009646 | -0,028301 |
| 8 | 0,058823 | 0,041719 | 0,081425 | -0,01461 | 0,014563 |
| 9 | 0 | -0,019417 | 0,002353 | 0,001647 | -0,028708 |
| 10 | 0,006944 | -0,02599 | 0,007042 | -0,041118 | -0,02463 |
| 11 | 0,010345 | 0,006353 | 0,083916 | 0,010291 | 0 |
| 12 | -0,030717 | 0,020202 | -0,04086 | -0,039049 | -0,050505 |
| 13 | -0,003521 | 0,118812 | 0,089686 | 0,06007 | 0,021276 |
| 14 | 0,060071 | 0,079646 | 0,028807 | 0,036666 | 0,026041 |
| 15 | -0,003333 | -0,001025 | 0,028 | 0,028938 | -0,010152 |
| 16 | 0,055596 | 0,092182 | 0,042759 | 0,059375 | -0,015812 |
| 17 | 0,051282 | -0,007519 | -0,041431 | -0,016269 | 0,05851 |
| 18 | -0,060976 | -0,043561 | 0,023576 | 0,004566 | -0,015075 |
| 19 | -0,035714 | 0,01817 | -0,021113 | -0,007575 | -0,010204 |
| 20 | 0 | -0,021569 | -0,007843 | 0,088549 | 0,082474 |
| 21 | -0,006734 | -0,01503 | -0,086956 | -0,021037 | -0,019047 |
| 22 | 0 | -0,017294 | 0,017316 | 0,054441 | 0,03398 |
| 23 | 0,030508 | 0,047619 | 0,055319 | -0,008152 | 0,032863 |
| 24 | 0,023026 | 0,012846 | -0,002016 | 0,013698 | -0,031518 |
| 25 | -0,061093 | -0,043902 | -0,042424 | -0,029729 | -0,014084 |
| 26 | 0,041096 | 0,016326 | 0,048523 | 0,018105 | 0,071428 |
| 27 | -0,013158 | -0,004016 | -0,038229 | -0,042407 | -0,048888 |
| 28 | 0,003333 | -0,008065 | -0,014992 | 0 | -0,028037 |
| 29 | -0,056479 | -0,014228 | -0,038627 | -0,005714 | -0,019607 |
| 30 | 0,051899 | 0,018557 | 0,066964 | 0,020302 | -0,015 |
| 31 | -0,013559 | -0,029352 | 0,012552 | -0,008571 | -0,010152 |
| 32 | -0,037801 | 0,003252 | -0,012397 | -0,020172 | -0,025641 |
| 33 | -0,021429 | 0,031446 | 0,039749 | 0,016176 | 0,005263 |

| Semana | <i>Allied Chemical</i> | <i>DuPont</i> | <i>Union Carbide</i> | <i>Exxon</i> | <i>Texaco</i> |
|--------|------------------------|---------------|----------------------|--------------|---------------|
| 34 | -0,014599 | -0,02439 | -0,01006 | 0,004341 | -0,005235 |
| 35 | -0,014815 | -0,020833 | -0,091463 | -0,007204 | -0,015789 |
| 36 | 0,011278 | -0,017021 | 0,064877 | 0,065312 | 0,026737 |
| 37 | -0,096654 | -0,075758 | -0,073529 | -0,053133 | -0,026041 |
| 38 | 0,020576 | 0,058548 | 0,018141 | 0,063309 | 0,016042 |
| 39 | 0,08871 | 0,04646 | 0,022272 | 0,004059 | 0 |
| 40 | 0,007407 | 0,019027 | 0,045752 | -0,008086 | 0,052631 |
| 41 | -0,022059 | 0,002075 | -0,017272 | -0,021739 | -0,045 |
| 42 | -0,031579 | 0,010352 | 0,012848 | -0,013888 | 0,010695 |
| 43 | 0,03937 | 0,054303 | -0,014799 | 0,011428 | -0,005291 |
| 44 | 0,015151 | 0,029154 | -0,021459 | -0,009887 | -0,021276 |
| 45 | 0 | -0,010466 | 0,035088 | -0,014265 | 0,038043 |
| 46 | -0,037313 | -0,024039 | -0,019068 | -0,024602 | -0,010471 |
| 47 | 0,015504 | -0,027586 | 0,006479 | 0,022255 | -0,026455 |
| 48 | 0,034351 | 0,024316 | 0,034335 | 0,020319 | 0,005434 |
| 49 | -0,0369 | 0,011869 | 0,014523 | 0,007112 | 0,016216 |
| 50 | 0,068965 | 0,014663 | 0,01636 | 0,038135 | 0,063829 |
| 51 | 0,089606 | 0,079961 | 0,102616 | 0,002721 | 0,02 |
| 52 | 0 | 0,016949 | 0,029197 | 0,002713 | 0,004901 |
| 53 | 0,05921 | 0,077193 | 0,019504 | -0,012178 | 0,039024 |
| 54 | 0,02795 | 0,009772 | 0 | 0 | -0,000265 |
| 55 | -0,004196 | 0,014516 | -0,031696 | -0,004445 | -0,014354 |
| 56 | 0,018405 | -0,0469 | 0,061594 | -0,043235 | -0,029126 |
| 57 | 0,069277 | 0,056888 | 0,040956 | 0,040816 | 0,02 |
| 58 | -0,016901 | -0,018268 | -0,008197 | -0,005602 | -0,019607 |
| 59 | -0,017192 | -0,001618 | -0,001653 | -0,016901 | 0,005 |
| 60 | -0,040816 | -0,035656 | 0 | 0,014326 | 0,004975 |
| 61 | -0,018237 | -0,003361 | -0,028146 | 0,03531 | 0,014851 |
| 62 | -0,003096 | -0,021922 | -0,027257 | 0,005457 | 0,039024 |
| 63 | 0,018634 | 0,025862 | -0,017513 | 0,018995 | -0,004694 |
| 64 | -0,057827 | -0,018487 | 0 | -0,023968 | -0,037735 |
| 65 | 0,087379 | 0,049657 | 0,033868 | 0,047748 | 0,039215 |
| 66 | 0 | -0,011419 | -0,010345 | -0,005208 | 0,028301 |

| Semana | <i>Allied Chemical</i> | <i>DuPont</i> | <i>Union Carbide</i> | <i>Exxon</i> | <i>Texaco</i> |
|--------|------------------------|---------------|----------------------|--------------|---------------|
| 67 | -0,019367 | -0,011551 | -0,022817 | 0,007853 | 0,013761 |
| 68 | -0,046012 | 0,035893 | 0,044964 | 0,040612 | 0,004608 |
| 69 | -0,07717 | -0,004029 | -0,003442 | 0,003797 | -0,027522 |
| 70 | 0,034843 | -0,008157 | -0,018998 | 0,008827 | -0,014151 |
| 71 | -0,006734 | -0,019737 | -0,026408 | 0,023749 | 0,014354 |
| 72 | -0,023729 | -0,019295 | -0,03255 | -0,001221 | 0,023584 |
| 73 | 0,065972 | 0,024807 | 0,057944 | 0,020782 | 0,004608 |
| 74 | 0 | -0,036728 | -0,014134 | -0,007185 | 0,004587 |
| 75 | -0,052117 | -0,058925 | -0,069892 | 0,00965 | 0,009132 |
| 76 | 0,054983 | -0,003683 | 0,026975 | -0,002389 | 0,009049 |
| 77 | -0,003257 | -0,009242 | -0,022514 | 0,005988 | -0,013452 |
| 78 | 0,022876 | 0,033582 | 0,001919 | 0,02619 | 0,004545 |
| 79 | -0,003195 | -0,005415 | -0,003831 | -0,013921 | -0,000165 |
| 80 | 0,04359 | -0,014519 | -0,013358 | 0,021176 | 0,013824 |
| 81 | -0,009317 | 0,013812 | 0,021654 | -0,014927 | -0,00909 |
| 82 | -0,056426 | -0,005557 | -0,003854 | -0,023696 | -0,018348 |
| 83 | 0,003322 | -0,041475 | -0,029014 | -0,002427 | -0,004672 |
| 84 | 0,016556 | 0,017308 | 0,033864 | 0,034063 | 0,009389 |
| 85 | -0,009772 | -0,016068 | -0,003854 | 0,014117 | 0,013953 |
| 86 | 0,026316 | -0,01633 | -0,009671 | 0,032482 | 0,027522 |
| 87 | 0,009615 | 0,009766 | 0,017578 | 0,016247 | 0,017857 |
| 88 | -0,047619 | -0,027079 | -0,051823 | -0,045468 | -0,021929 |
| 89 | -0,026667 | -0,06163 | -0,05668 | -0,013452 | -0,040358 |
| 90 | 0,010274 | 0,023305 | 0,034335 | -0,018181 | -0,004672 |
| 91 | -0,044068 | 0,020704 | -0,006224 | -0,018518 | 0,004694 |
| 92 | 0,039007 | 0,03854 | 0,024988 | -0,028301 | 0,03271 |
| 93 | -0,039457 | -0,029297 | -0,065844 | -0,015837 | -0,045758 |
| 94 | 0,039568 | 0,024145 | -0,006608 | 0,028423 | -0,009661 |
| 95 | -0,031142 | -0,007941 | 0,01108 | 0,007537 | 0,014634 |
| 96 | 0 | -0,02008 | -0,006579 | 0,029925 | -0,004807 |
| 97 | 0,021429 | 0,04918 | 0,006622 | -0,002421 | 0,028985 |
| 98 | 0,045454 | 0,046375 | 0,074561 | 0,014563 | 0,018779 |
| 99 | 0,050167 | 0,03638 | 0,004082 | -0,011961 | 0,009216 |
| 100 | 0,019108 | -0,033303 | 0,008362 | 0,033898 | 0,004566 |

Tabla A.1. Tabla de los “*stock-prices*” (índice de rendimiento semanal).

El primer paso es abrir el programa *R* y cargar el fichero correspondiente a los datos. *R* puede leer archivos de texto o ficheros importados de otros programas estadísticos, como SPSS (*Statistical Package for the Social Sciences*). En este caso, el fichero con los datos dados en la Tabla A.1, ha sido originalmente creado como archivo de texto.

```
> stock<-read.table("stockprices.txt",header=TRUE)
> attach(stock)
```

Una vez cargados los datos, se procede a calcular el valor de cada uno de los 10 criterios de información expuestos en esta Tesis, para cada uno de los modelos factoriales posibles. Para ello, se emplea la función *InformationCriteria* dada en la sección 3.4 del Capítulo 3. Dicha función proporciona la decisión final acerca del número de factores a extraer, en base a cada uno de los criterios de información estudiados. La función *InformationCriteria* requiere el uso de la función *factanal*¹, incluida dentro del paquete *stats*; *factanal* realiza un análisis factorial sobre una matriz de datos o de covarianzas mediante el método de máxima verosimilitud.

La función *InformationCriteria* se puede cargar en *R* mediante la siguiente instrucción, a realizar en el menú de la consola: Archivo → Nuevo script. En la ventana del *script*, se inserta la función a implementar:

```
InformationCriteria<-function(data){
n<-nrow(data)
p<-ncol(data)
AIC<-c(NULL)
BIC<-c(NULL)
HQIC<-c(NULL)
AIC3<-c(NULL)
CAIC<-c(NULL)
CAIC_Yang<-c(NULL)
AICc<-c(NULL)
```

¹Información acerca de la función *factanal*, puede obtenerse en la siguiente dirección web: <http://stat.ethz.ch/R-manual/R-patched/library/stats/html/factanal.html>


```

DBIC<-c(NULL)
BIC_Yang<-c(NULL)
RIC<-c(NULL)
for(m in 1:(trunc(1/2*(2*p+1-sqrt(8*p+1))))){
  fit<-factanal(data,factors=m)
  s<-tcrossprod(fit$loadings)+diag(fit$uniquenesses)
  k<-(m*p+p)-1/2*m*(m-1)
  d<-1/2*((p-m)^2-(p+m))
  logl<-(-1/2)*(n*p*(log(2*pi)+1)+n*log(det(s)))
  aic<--2*logl+2*k
  AIC<-rbind(AIC,aic)
  bic<--2*logl+(log(n)*k)
  BIC<-rbind(BIC,bic)
  hqic<--2*logl+2*(log(log(n)))*k
  HQIC<-rbind(HQIC,hqic)
  aic3<--2*logl+3*k
  AIC3<-rbind(AIC3,aic3)
  caic<--2*logl+(log(n)+1)*k
  CAIC<-rbind(CAIC,caic)
  caicYang<--2*logl+(log(n+2)+(1-log(24)))*k
  CAIC_Yang<-rbind(CAIC_Yang,caicYang)
  aicc<--2*logl+2*k+(2*(k+1)*(k+2)/(n-k-2))
  AICc<-rbind(AICc,aicc)
  dbic<--2*logl+(log(n)-log(2*pi))*k
  DBIC<-rbind(DBIC,dbic)
  bicYang<--2*logl+(log(n+2)-log(24))*k
  BIC_Yang<-rbind(BIC_Yang,bicYang)
  ric<--2*logl+log((n+2)/24)*k+2*log(k+1)
}

```

```

RIC<-rbind(RIC,ric)

chi_squared<-fit$STATISTIC

p_value<-fit$PVAL

print(c(sample_size=n, variables=p, factors=m, free_parameters=k, log_likelihood=logl,
AIC=aic, BIC=bic, HQIC=hqic, AIC3=aic3, CAIC=caic, CAIC_Yang=caicYang, AICc=aicc,
DBIC=dbic, BIC_Yang=bicYang, RIC=ric))

print(c(sample_size=n, variables=p, factors=m, df=d, chi_squared=chi_squared,
p_value<-p_value))
}

print(c(number_of_factors_AIC=(1:length(AIC))[AIC==min(AIC)]))
print(c(number_of_factors_BIC=(1:length(BIC))[BIC==min(BIC)]))
print(c(number_of_factors_HQIC=(1:length(HQIC))[HQIC==min(HQIC)]))
print(c(number_of_factors_AIC3=(1:length(AIC3))[AIC3==min(AIC3)]))
print(c(number_of_factors_CAIC=(1:length(CAIC))[CAIC==min(CAIC)]))
print(c(number_of_factors_CAIC_Yang=(1:length(CAIC_Yang))[CAIC_Yang==min(CAIC_Yang)]))
print(c(number_of_factors_AICc=(1:length(AICc))[AICc==min(AICc)]))
print(c(number_of_factors_DBIC=(1:length(DBIC))[DBIC==min(DBIC)]))
print(c(number_of_factors_BIC_Yang=(1:length(BIC_Yang))[BIC_Yang==min(BIC_Yang)]))
print(c(number_of_factors_RIC=(1:length(RIC))[RIC==min(RIC)]))
}

```

Una vez escrito el algoritmo, el siguiente paso es pulsar en el Menú “*Editar*” y marcar la opción “*Ejecutar todo*”. A partir de ese instante, el programa puede hacer uso de la función *InformationCriteria* y aplicarla al conjunto de datos considerado de la siguiente forma:

```
> InformationCriteria(stock)
```

La salida del programa viene dada en la Figura A.1. La función definida proporciona el valor de cada criterio para los posibles modelos factoriales (en este caso, con 1 y 2 factores). Además, con fines comparativos, proporciona el resultado de la aplicación del test de Bartlett para ambos

casos. Se puede observar que, en base a la mayoría de criterios de información, sería seleccionado el modelo con 2 factores, que es en apariencia el más adecuado (ver sección 4.1). Sólo el *BIC* y el *CAIC*, con tendencia a infraestimar el número de factores en ciertos tamaños muestrales limitados, seleccionarían el modelo con un único factor.

```

> InformacionCriteriosIAStock()
sample_size  variables  factors free_parameters  log_likelihood  AIC  AIC3  HQIC  CAIC  AICC  DBIC  BIC_Yang
100.0000    5.0000    1.0000    10.0000    -685.2427  1290.4853  1300.4853  1301.0289  1326.5370  1289.4853  1298.1593  1284.5545
RIC
1289.7503
sample_size  variables  factors free_parameters  log_likelihood  df  ch1_squared_objective  AIC3  HQIC  CAIC  AICC  DBIC  BIC_Yang
1.000000e+02  5.000000e+00  1.000000e+00  1.000000e+00  -685.2427  5.000000e+00  1290.4853  1300.4853  1301.0289  1326.5370  1289.4853  1298.1593  1284.5545
100.0000    5.0000    2.0000    20.0000    -627.4621  1282.8223  1296.8223  1297.6939  1333.3946  1288.6866  1293.6644  1275.1791
RIC
1280.5582
sample_size  variables  factors free_parameters  log_likelihood  df  ch1_squared_objective  AIC3  HQIC  CAIC  AICC  DBIC  BIC_Yang
100.000000    5.000000    2.000000    2.000000    -627.4621  1.000000    1282.8223  1296.8223  1297.6939  1333.3946  1288.6866  1293.6644  1275.1791
number_of_factors_AIC  2
number_of_factors_BIC  2
number_of_factors_HQIC  1
number_of_factors_AIC3  2
number_of_factors_CAIC  1
number_of_factors_CAIC_Yang  2
number_of_factors_AICC  2
number_of_factors_DBIC  2
number_of_factors_BIC_Yang  2
number_of_factors_RIC  2

```

Figura A.1: Valores de los criterios de información y resultado de la aplicación del test de Bartlett, en el ejemplo de "stock-prices".

Además, se pueden obtener evidencias acerca de la adecuación de las posibles soluciones factoriales, para cada uno de los criterios estudiados. A tal fin, se ha definido la función *weightsIC*, dada en la sección 3.5 del Capítulo 3. Para su implementación en *R*, el procedimiento es similar al realizado anteriormente con la función *InformationCriteria*. Se abre un nuevo *script* y se inserta en él la función *weightsIC*.

```
weightsIC<-function(data){
n<-nrow(data)
p<-ncol(data)
AIC<-c(NULL)
BIC<-c(NULL)
HQIC<-c(NULL)
AIC3<-c(NULL)
CAIC<-c(NULL)
CAIC_Yang<-c(NULL)
AICc<-c(NULL)
DBIC<-c(NULL)
BIC_Yang<-c(NULL)
RIC<-c(NULL)
for(m in 1:(trunc(1/2*(2*p+1-sqrt(8*p+1))))){
  fit<-factanal(data,factors=m)
  s<-tcrossprod(fit$loadings)+diag(fit$uniquenesses)
  k<-(m*p+p)-1/2*m*(m-1)
  d<-1/2*((p-m)^2-(p+m))
  logl<-(-1/2)*(n*p*(log(2*pi)+1)+n*log(det(s)))
  aic<--2*logl+2*k
  AIC<-rbind(AIC,aic)
  bic<--2*logl+(log(n)*k)
  BIC<-rbind(BIC,bic)
}
```

```

hqic<--2*logl+2*(log(log(n)))*k
HQIC<-rbind(HQIC,hqic)
aic3<--2*logl+3*k
AIC3<-rbind(AIC3,aic3)
caic<--2*logl+(log(n)+1)*k
CAIC<-rbind(CAIC,caic)
caicYang<--2*logl+(log(n+2)+(1-log(24)))*k
CAIC_Yang<-rbind(CAIC_Yang,caicYang)
aicc<--2*logl+2*k+(2*(k+1)*(k+2)/(n-k-2))
AICc<-rbind(AICc,aicc)
dbic<--2*logl+(log(n)-log(2*pi))*k
DBIC<-rbind(DBIC,dbic)
bicYang<--2*logl+(log(n+2)-log(24))*k
BIC_Yang<-rbind(BIC_Yang,bicYang)
ric<--2*logl+log((n+2)/24)*k+2*log(k+1)
RIC<-rbind(RIC,ric)
}

diffAIC<-AIC-min(AIC)
diffBIC<-BIC-min(BIC)
diffHQIC<-HQIC-min(HQIC)
diffAIC3<-AIC3-min(AIC3)
diffCAIC<-CAIC-min(CAIC)
diffCAIC_Yang<-CAIC_Yang-min(CAIC_Yang)
diffAICc<-AICc-min(AICc)
diffDBIC<-DBIC-min(DBIC)
diffBIC_Yang<-BIC_Yang-min(BIC_Yang)
diffRIC<-RIC-min(RIC)
wAIC<-(exp(-diffAIC/2))/(sum(exp(-diffAIC/2)))

```

```

wBIC<-(exp(-diffBIC/2))/(sum(exp(-diffBIC/2)))
wHQIC<-(exp(-diffHQIC/2))/(sum(exp(-diffHQIC/2)))
wAIC3<-(exp(-diffAIC3/2))/(sum(exp(-diffAIC3/2)))
wCAIC<-(exp(-diffCAIC/2))/(sum(exp(-diffCAIC/2)))
wCAIC_Yang<-(exp(-diffCAIC_Yang/2))/(sum(exp(-diffCAIC_Yang/2)))
wAICc<-(exp(-diffAICc/2))/(sum(exp(-diffAICc/2)))
wDBIC<-(exp(-diffDBIC/2))/(sum(exp(-diffDBIC/2)))
wBIC_Yang<-(exp(-diffBIC_Yang/2))/(sum(exp(-diffBIC_Yang/2)))
wRIC<-(exp(-diffRIC/2))/(sum(exp(-diffRIC/2)))
for(m in 1:(trunc(1/2*(2*p+1-sqrt(8*p+1))))) {
print(c(factors = m, weightAIC = wAIC[m], weightBIC = wBIC[m],
weightHQIC = wHQIC[m], weightAIC3 = wAIC3[m], weightCAIC = wCAIC[m],
weightCAIC_Yang = wCAIC_Yang[m], weightAICc = wAICc[m], weightDBIC = wDBIC[m],
weightBIC_Yang = wBIC_Yang[m], weightRIC = wRIC[m]))
}
}

```

Una vez escrito el código, se selecciona en el menú Editar → Ejecutar todo. Este algoritmo vuelve a utilizar la función *factanal*, incluida dentro del paquete *stats*. Si éste ha sido ya cargado, se puede ejecutar la función, aplicándola al conjunto de datos considerado.

```
> weightsIC(stock)
```

La salida del programa se muestra en la Figura A.2. Así, por ejemplo, el *CAIC** de Yang y Yang (2007) y el *DBIC* de Draper (1995), proporcionarían evidencias del 94,7% y del 90,4%, respectivamente, en favor de la solución compuesta por 2 factores.

```

> weightsIC(stock)
factors      weightAIC      weightBIC      weightHQIC      weightAIC3      weightAICC      weightDBIC      weightBIC_Yang      weightRIC
1.000000000  0.022279912  0.806714650  0.158048635  0.144113332  0.968592657  0.052764957  0.081331265  0.095612728  0.007482334  0.010175497
2.000000000  0.97772009  0.19328535  0.84195136  0.85588667  0.03140734  0.94723504  0.91866874  0.90438727  0.99251767  0.98982450

```

Figura A.2: Pesos asociados a cada uno de los posibles modelos factoriales, en base a los respectivos criterios de información. Ejemplo de "stock-prices".

Por último, se puede hacer uso de la función *factanal* citada para visualizar la composición de la solución con dos factores. Empleando la rotación Varimax, esta sería la instrucción dada al programa para obtener la solución con dos factores correspondiente a un AFMV sobre los datos de "stock-prices":

```
> factanal(stock,factors=2,rotation="varimax")
```

Call:

```
factanal(x = stock, factors = 2, rotation = "varimax")
```

Uniquenesses:

| AlliedChemical | DuPont | UnionCarbide | Exxon | Texaco |
|----------------|--------|--------------|-------|--------|
| 0.495 | 0.252 | 0.474 | 0.609 | 0.178 |

Loadings:

| | Factor1 | Factor2 |
|----------------|---------|---------|
| AlliedChemical | 0.602 | 0.377 |
| DuPont | 0.849 | 0.165 |
| UnionCarbide | 0.642 | 0.336 |
| Exxon | 0.365 | 0.508 |
| Texaco | 0.207 | 0.883 |

| | Factor1 | Factor2 |
|----------------|---------|---------|
| SS loadings | 1.672 | 1.320 |
| Proportion Var | 0.334 | 0.264 |
| Cumulative Var | 0.334 | 0.599 |

Test of the hypothesis that 2 factors are sufficient.

The chi square statistic is 0.56 on 1 degree of freedom.

The p-value is 0.453

En estos resultados, se pueden observar las unicidades, las cargas factoriales correspondientes a la solución rotada y la proporción de varianza explicada. Además, la función *factanal* proporciona también el resultado de la aplicación del test de Bartlett sobre la adecuación del número de factores extraídos.

Referencias bibliográficas

- Acito, F., Anderson, R. D. y Engledow, J. L. (1980). A simulation study of methods for hypothesis testing in factor analysis. *Journal of Consumer Research*, 7, 2, 141-150.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions On Automatic Control*, 19, 716-723.
- Akaike, H. (1977). On entropy maximization principle. In P. R. Krishnaiah (Ed.), *Applications of Statistics* (pp. 27-41). Amsterdam, Holland: North-Holland Publishing Company.
- Akaike, H. (1978). On the likelihood of a time series model. *The Statistician*, 27, 217-235.
- Akaike, H. (1979). A bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika*, 66, 237-242.
- Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics*, 16, 3-14.
- Akaike, H. (1983). Information measures and model selection. *International Statistical Institute*, 44, 277-291.
- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 3, 317-332.
- Anarte, M. T., Caballero, F. F., Ruiz de Adana, M. S., Rondán, R. M., Carreira, M., Domínguez, M., Machado, A., Gonzalo, M., Tapia, M. J., Valdés, S., González, S. y Soriguer, F. C. (2011). Development of a New Fear of Hypoglycemia Scale: FH-15. *Psychological Assessment*. Advance online publication. doi: 10.1037/a0021927.
- Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*. New York: Wiley.
- Andrews, R. L. y Currim, I. S. (2003). A comparison of segment retention criteria for finite mixture logit models. *Journal of Marketing Research*, XL, 235-243.

- Bartlett, M. S. (1947). Multivariate analysis. *Journal of the Royal Statistical Society, Supplement 9B*, 176-197.
- Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, *3*, 77-85.
- Basilewsky, A. (1994). *Statistical factor analysis and related methods: theory and applications*. New York: John Wiley and Sons.
- Bernstein, I. H. (1987). *Applied multivariate analysis*. New York: Springer.
- Bozdogan, H. (1987). Model selection and Akaike's Information Criteria (AIC): the general theory and its analytic extensions. *Psychometrika*, *52*, **3**, 345-370.
- Bozdogan, H. (2000). Akaike's information criterion and recent developments in information complexity. *Journal of Mathematical Psychology*, *44*, 62-91.
- Bozdogan, H. y Ramírez, D. E. (1987). An expert model selection approach to determine the "best" pattern structure in factor analysis models. In H. Bozdogan & A. K. Gupta (Eds.), *Multivariate statistical modeling and data analysis* (pp. 35-60). Dordrecht, Holland: D. Reidel Publishing Company.
- Bozdogan, H. y Ramírez, D. E. (1988). FACAIC: Model selection algorithm for the orthogonal factor model using AIC and CAIC. *Psychometrika*, *53*, **3**, 407-415.
- Buckland, S. T., Burnham, K. P. y Augustin, N. H. (1997). Model selection: an integral part of inference. *Biometrics*, *53*, 603-618.
- Burnham, K. P. y Anderson, D. R. (2001). Kullback-Leibler information as a basis for strong inference in ecological studies. *Wildlife Research*, *28*, 111-119.
- Burnham, K. P. y Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer-Verlag.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245-276.
- Chatfield, C. y Collins, A. J. (1980). *Introduction to multivariate analysis*. London: Chapman and Hall.
- Costa, M. (1996). Factor analysis and information criteria. *Qüestió*, *20*, **3**, 409-425.

- Cox, T. F. y Dunn, R. T. (2002). An analysis of Decathlon data. *The Statistician*, 51, 2, 179-187.
- Crawley, M. J. (2007). *The R book*. Chichester: John Wiley and Sons.
- Crawley, M. J. (2009). *Statistics: an introduction using R*. Chichester : John Wiley and Sons.
- Cuadras, C. M. (1991). *Métodos de análisis multivariante*. Barcelona: PPU.
- De Vellis, R. F. (2003). *Scale development: theory and applications*. Thousand Oaks, CA: Sage.
- Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society, Series B*, 57, 45-97.
- Ertel, S. (2011). Exploratory factor analysis revealing complex structure. *Personality and Individual Differences*, 50, 196-200.
- Finch, H. (2006). Comparison of the performance of Varimax and Promax rotations: factor structure recovery for dichotomus items. *Journal of Educational Measurement*, 43, 1, 39-52.
- García, E., Gil, J. y Rodríguez, G. (2000). *Análisis factorial*. Cuadernos de Estadística. Madrid: La Muralla.
- Geweke, J. F. y Singleton, K. J. (1980). Interpreting the likelihood ratio statistic in factor models when sample size is small. *Journal of the American Statistical Association*, 75, 369, 133-137.
- Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale, NJ: Earlbaum.
- Griliches, Z. (1984). *R & D, patents and productivity*. Chicago: The University of Chicago Press.
- Gutiérrez, R. y González, A. (1991). *Introducción al análisis multivariante; Volumen 1*. Universidad de Granada.
- Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, 18, 277-296.
- Hannan, E. J. y Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B*, 41, 2, 190-195.
- Hair, J. F., Anderson, R. E., Tatham, R. L. y Black, W. C. (1995). *Multivariate data analysis*. 4th edition. Englewood Cliffs, NJ: Prentice Hall.
- Harman, H. H. (1976). *Modern factor analysis*. Third edition. Chicago: The University of Chicago Press.
- Harris, C. W. (1962). Some Rao-Guttman relationships. *Psychometrika*, 27, 247-263.

- Harris, C. W. y Kaiser, H. F. (1964). Oblique factor analytic solutions by orthogonal transformations. *Psychometrika*, *29*, 347-362.
- Hendrickson, A. E. y White, P. O. (1964). Promax: a quick method to oblique simple structure. *British Journal of Statistical Psychology*, *17*, 65-70.
- Heywood, H. B. (1931). On finite sequences of real numbers. *Journal of the Royal Society, Series A*, *134*, 486-501.
- Holzinger, K. J. y Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution. *Supplementary Educational Monographs*, number 48. Chicago: University of Chicago, Department of Education.
- Homburg, C. (1991). Cross-validation and information criteria in causal modeling. *Journal of Marketing Research*, *XXVIII*, 137-144.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179-185.
- Hurvich, C. M. y Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*, 297-307.
- Hurvich, C. M. y Tsai, C. L. (1991). Bias of the corrected AIC criterion for underfitted regression and time series models. *Biometrika*, *78*, 499-509.
- Ichikawa, M. (1988). Empirical assessments of AIC procedure for model selection in factor analysis. *Behaviormetrika*, *24*, 33-40.
- Jennrich, R. I. y Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika*, *31*, 313-323.
- Johnson, R. A. y Wichern, D. W. (1988). *Applied multivariate statistical analysis*. 2nd edition. Englewood Cliffs: Prentice Hall.
- Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, *43*, 433-477.
- Kaiser, H. F. (1958). The Varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187-200.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141-151.

- Kendall, M. G. (1980). *Multivariate analysis*. 2nd edition. London: Charles Griffin.
- Kieffer, K. M. (1998). *Orthogonal versus oblique rotation: a review of the literature regarding the pros and cons*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, New Orleans, LA.
- Kim, H. J. (2008). Common factor analysis versus principal component analysis: choice for symptom cluster research. *Asian Nursing Research*, 2, 1, 17-24.
- Kirankabes, M. C. (2010). Relationship between gross Domestic Expenditure on R & D and patent applications. A panel data analysis for EU countries and some other selected countries. *Middle Eastern Finance and Economics*, 8, 161-171.
- Kitagawa, G. y Konishi, S. (2010). Bias and variance reduction techniques for bootstrap information criteria. *Annals of the Institute of Statistical Mathematics*, 62, 209-234.
- Konishi, S. y Kitagawa, G. (2003). Asymptotic theory for information criteria in model selection – functional approach. *Journal of Statistical Planning and Inference*, 114, 45-61.
- Kullback, S. y Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22, 79-86.
- Lawley, D. N. y Maxwell, A. E. (1971). *Factor analysis as a statistical method*. 2nd edition. London: Butterworth.
- Lee, S. H., Wong, P. K. y Chong, C. L. (2005). Human and social capital explanations for R&D outcomes. *IEEE Transactions on Engineering Management*, 52, 1, 59-68.
- Linden, M. (1977). A factor analytic study of Olympic Decathlon data. *Research Quarterly*, 48, 3, 562-568.
- Markon, K. E. y Krueger, R. F. (2004). An empirical comparison of information – theoretic selection criteria for multivariate behavior genetic models. *Behavior Genetics*, 34, 6, 593-610.
- Martínez, R. (1999). *El análisis multivariante en la investigación científica*. Cuadernos de Estadística. Madrid: La Muralla.
- McDonald, R. P. (1997). *Test theory: a unified treatment*. Mahwah, NJ: Erlbaum.
- Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, 44, 190-204.

- Nelson, R. R. y Phelps, E. S. (1966). Investment in humans, technological diffusion and economic growth. *The American Economic Review*, 56, 2, 69-75.
- Peña, D. (2002). *Análisis de datos multivariantes*. Madrid: McGraw-Hill.
- R Development Core Team (2008). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org>
- Reiersol, O. (1950). On the identifiability of parameters in Thurstone's multiple factor analysis. *Psychometrika*, 15, 121-159.
- Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14, 465-471.
- Rust, R. T., Simester, D., Brodie, R. J. y Nilikant, V. (1995). Model selection criteria: An investigation of relative accuracy, posterior probabilities, and combinations of criteria. *Management Science*, 41, 322-332.
- Schomaker, M. y Heumann, C. (2011). Model averaging in factor analysis: an analysis of olympic decathlon data. *Journal of Quantitative Analysis in Sports*, Vol. 7: Iss. 1, Article 4.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 2, 461-464.
- Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52, 3, 333-343.
- Snook, S. C. y Gorsuch, R. L. (1989). Principal component analysis versus common factor analysis: a Monte Carlo study. *Psychological Bulletin*, 106, 148-154.
- Stone, M. (1979). Comments on model selection criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society, Series B*, 41, 2, 276-278.
- Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics*, A7, 13-26.
- Tabachnick, B. G. y Fidell, L. S. (1996). *Using multivariate statistics*. New York: Harper Collins.
- Takane, Y. (1987). Introduction to special section. *Psychometrika*, 52, 315.
- Takeuchi, K., Yanai, H. y Mukherjee, B. N. (1982). *The foundations of multivariate analysis*. New Delhi: Wiley Eastern Ltd.
- Thompson, B. (1984). *Canonical correlation analysis: uses and interpretation*. Thousand Oaks, CA: Sage.

- Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Tinsley, H. E. A. y Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology, 34*, 414-424.
- Tucker, L. R. y Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38, 1*, 1-10.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*, 321-327.
- Velicer, W. F. (1977). An empirical comparison of the similarity of principal component, image, and factor patterns. *Multivariate Behavioral Research, 12*, 3-22.
- Velicer, W. F. y Jackson, D. N. (1990). Component analysis versus common factor analysis: some issues in selecting an appropriate procedure. *Multivariate Behavioral Research, 25, 1*, 1-28.
- Velicer, W. F., Peacock, A. C. y Jackson, D. N. (1982). A comparison of component and factor patterns: a Monte Carlo approach. *Multivariate Behavioral Research, 17*, 371-388.
- Wagenmakers, E. J. y Farrell, S. (2004). *Psychonomic Bulletin & Review, 11, 1*, 192-196.
- Widaman, K. F. (1990). Bias in pattern loadings represented by common factor analysis and component analysis. *Multivariate Behavioral Research, 25, 1*, 89-95.
- Woodroffe, M. (1982). On model selection and the arc-sine laws. *Annals of Statistics, 10*, 1182-1194.
- Yang, C.-C. y Yang C.-C. (2007). Separating latent classes by information criteria. *Journal of Classification, 24, 2*, 183-203.
- Zarnowski, F. (1989). *The decathlon. A colorful history of track and field's most challenging events*. Champaign, IL: Leisure press.
- Zwick, W. R. y Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99, 3*, 432-442.