EL ANÁLISIS FACTORIAL

Por Tevni Grajales G.

A continuación una breve discusión de esta técnica la cual corresponde a técnicas de interdependencia. En primer lugar se especifican los objetivos que se procuran cuando se usa la técnica, después se presentan los supuestos que deben ser considerados para proceder en el uso de la técnica y por último se comenta, de manera suscinta, el proceso usado.

- **Objetivos de la técnica**

  Identificar las dimensiones subyacentes o contenidas en la medición de otras variables observadas de manera directa. Las dimensiones mencionadas, son factores o constructos que no pueden ser medidos por medio de una entidad mesurable simple. Por ejemplo: desde el siglo pasado se habla de la inteligencia; desde entonces se han desarrollado pruebas que pretenden medir el cociente intelectual, asunto respecto al que estamos acostumbrados; pero si nos preguntamos ¿qué es la inteligencia? concluiremos en que se trata de una característica compleja la cual no puede ser observada o medida a partir de una sola observación sino que es necesario observar diferentes conductas (entidad mesurable simple) como destrezas verbales, matemáticas, espaciales, etc. las cuales en su conjunto y de acuerdo con una teoría definida, pueden conducir a una conclusión respecto a la inteligencia (constructo) del sujeto o sujetos observados.

  Otros ejemplos de constructos pueden ser: la satisfacción laboral de los maestros, la inteligencia emocional, grado de creatividad del estudiante, etc.

  Es una técnica estadística que se usa para identificar un relativo pequeño número de factores cada uno de los cuales puede ser usados para representar la relación entre un conjunto de varias variables interrelacionadas.

- **Supuestos que fundamentan la técnica**.
  - Las dimensiones subyacentes o factores pueden usarse para explicar fenómenos complejos.
  - Las correlaciones que se observan entre las variables son el resultado del hecho que tales variables comparten los mismos factores.

El modelo matemático para el análisis factorial parece ser similar a la ecuación de regresión múltiple. Pero se debe recordar que en el caso de la regresión múltiple, ésta considera variables simples las cuáles son predictoras de la variable dependiente (criterio)..

En cambio, en el caso del análisis factorial la variable dependiente se expresa en términos de una combinación lineal de grupos de variables que caracterizan un concepto en particular (factores). Los factores no son variables independientes simples sino que cada uno está constituido por un grupo de variables que caracterizan el concepto que representa el factor. Es por esta causa que se clasifica esta técnica entre las técnicas de interdependencia. (Tanto las variables a un lado de la ecuación como en el otro están interactuando como criterios y predictoras).

Por lo general, los factores que pueden caracterizar a un grupo de variables no se conocen con anticipación sino que llegan a ser determinados por medio del análisis factorial. Estos factores se llama factores comunes dado que todas la variables en observación se llegan a expresar como funciones de ellos.Cuando no se conoce con anticipación los factores que constituyen las variables se dice que procede una análisis exploratorio. Pero en cambio, si el investigador ha elaborado el análisis anticipando (posiblemente apoyado en la teória) la existencia de ciertos número de factores en particular y anticipando qué variables conforman cada uno de los factores, se trata de un análisis confirmatorio. En este caso al realizar el análisis se debe solicitar al procesador estadístico que extraiga el número de factores hipotetizados.

Para efectos del presente tema vamos a proseguir según un análisis exploratorio de factores.

Según el ejemplo al final del capítulo, la satisfacción que dicen sentir los docentes de secundaria en Nuevo León está determinada por tres factores cada uno de los cuales es el resultado de la interrelación de catorce variables simples medidas según una escala Likert.

Estas variables se agrupan según el factor en el cual parecen tener mayor carga, resultando entonces que el factor 1 responde a cinco variables, el factor 2 a otras cinco variables, el factor 3 a cuatro variables; según aparece en el anexo B (al final de este capítulo) Ud. puede identificar cada uno de los factores y las variables que tienen una carga importante en el mismo.

Corresponde al investigador determinar lo que representa o constituye cada uno de estos factores, para lo cual deberá hechar mano de la información existente (marco teórico).

Como se evidencia, los factores pueden ser desconocidos por el investigador hasta que se realiza el análisis factorial. Lo que se procura es explicar la relación entre el conjunto de variables por medio del menor número de factores. Y estos factores deben tener significado, deben ofrecer una solución simple e interpretable que conduzca a nuevas percepciones. Según Garza García "hay que analizar si en este caso las variables que quedan bajo cada factor se pueden agrupar lógicamente bajo una característica común, si no se puede, no sería la mejor solución y se tendría que probar otras soluciones con más o menos factores, hasta que las variables queden agrupadas de manera lógica"

Pasos en el Análisis Factorial

El análisis de la matriz de correlación

Computar una matriz de correlación de todas las variables con el fin de identificar las variables que no parezcan estar correlacionadas con las otras. Se debe evaluar lo apropiado del modelo factorial y determinar lo que se debe hacer con los casos que adolecen de información completa respecto a las variables en estudio.

En esta fase del procedimiento, se debe recordar que las variables que tienen pequeñas correlaciones entre sí, son aquellas que no comparten factores en común. Por medio de la prueba de esfericidad de Bartlett se prueba la hipótesis que la matriz de correlación es una matriz de identidad, es decir que todos los valores en la diagonal son 1 y todos fuera de la diagonal son 0. Se supone que si las variables no están correlacionadas entre sí, no es posible encontrar en ellas un factor común. Con esta prueba se muestra la probabilidad estadística de que la matriz de correlación tiene correlaciones significativas al menos entre algunas variables. Pero hay que tener en cuenta que si la cantidad de casos incluídos en la observación se muy grande, esta prueba puede identificar correlaciones significativas no importantes. La significatividad debe ser menor a .05 para proceder con la técnica.

Otra forma de observar el grado de correlación entre la variables es por medio de los coeficientes de correlación parcial. Las correlaciones parciales son estimados de la correlación entre los factores únicos y deben ser cercanos a cero cuando se cumplen las suposiciones del análisis factorial. Una matriz que contiene los coeficientes de correlación parciales negativos (La matriz de correlaciones anti-imágen) debe mostrar una proporción muy reducida de coeficientes de correlación altos, a fin de que pueda considerarse el análisis factorial.

Existe una medida de la adecuación de la muestra (MSA) la cual es un índice que compara las magnitudes de los coeficientes de correlación observados y las magnitudes de los coeficientes de la correlación parcial. Si una variable tiene una MSA de 1 significa que puede ser predicha de manera perfecta sin error por las otras variables en el estudio. Es decir, cuantifica el grado de intercorrelación entre las variables y lo apropiado del análisis de factor. Otra prueba de adecuación de la muestra que es conocida como Kaiser-Meyer-Olkin (KMO); y valores pequeños en este índice (cercanos a cero) indican que no es recomendable usar el análisis factorial, siendo que las correlaciones entre pares de variables no son explicadas por la otras variables. En el caso de nuestro ejemplo el valor KMO es de .89167 el cual es meritorio según el criterio de Kaiser.

El cuadrado del coeficiente de correlación múltiple (R^2) entre una variable y todas las demás, es otro indicador de la fortaleza de la asociación lineal entre las variables y es reconocido como Comunalidad. Cuando este coeficiente es pequeño para un variable en particular, es recomendable considerar la posibilidad de eliminarla del conjunto de variables en estudio.

La extracción de factorial

El propósito central del procedimiento es determinar los factores que subyacen en las variables medidas (observadas). Para ello se cuenta con diversos métodos los cuales difieren en el criterio que usan para definir lo que es una buena selección.

El método de Análisis de Componentes Principales, se forma una combinación lineal de las variables observadas. El primer componente principal es la combinación que da cuenta de la mayor cantidad de la varianza en la muestra. El segundo componente principal responde a la siguiente cantidad de varianza inmediatamente inferior a la primera y no está correlacionado con el primero. Así sucesivamente los componentes explican proporciones menores de la varianza de la muestra total.

Con el fin de determinar el número de factores necesarios para representar los datos, resulta muy útil examinar el porcentaje total de la varianza que es explicada por cada uno de ellos. La varianza total es la suma de las varianzas de cada variable. Esto se expresa de manera estandarizada con una media de cero y una desviación estándar de 1. Por lo que el total de la varianza estará determinada por el total de las variables incluidas en el estudio.

El total de la varianza explicada por cada factor se identifica como Eigenvalue y se sugiere que sólo se consideren los factores cuyo Eigenvalue sea superior a 1 (criterio de la raíz latente) siendo que valores menores resultarían en factores inferiores a lo que representa una simple variable la cual tiene una varianza de 1. Este procedimiento es recomendable cuando se trata de entre 20 y 50 variables. Si son menos de 20 variables se puede utilizar el criterio conocido como *scree test criterio*- el cual consiste en utilizar una gráfica en la que aparecen los factores según su Eigenvalue representados por una línea que va cayendo de derecha a izquierda, se supone que en el punto en que la línea deja de caer y toma una tendencia horizontal al eje de las x se ubica un número de factores que podrían ser considerados para ser extraídos.

Carga factorial es un coeficientes usado para expresar una variable estandarizada en términos de los factores, e indican el peso que es atribuído a cada factor. Los factores con coeficiente grandes (en valores absolutos) para una variable son factores relacionados estrechamente con dicha variable. Esta información se presenta en la matriz de estructura factorial (factor structure matrix) lo cual contiene las correlaciones simples entre la variables y los factores, pero estas cargas contiene tanto la varianza única entre las variables y los factores más las correlaciones entre los factores. A medida que las correlaciones entre los factores son más grandes, se hace más difícil distinguir cuáles variables tienen mayor carga en cada factor (esta matriz resulta cuando se utiliza la rotación oblicua y no la ortogonal, lo cual supone que los factores están relacionados) La matriz de patron factorial (factor pattern matrix) tiene las cargas que representan la contribución única de cada variabla a cada factor y es la que se utiliza preferentemente al presentar resultados.

Cuando los factores estimados no se han correlacionado entre si (ortogonales), las cargas factoriales son también las correlaciones entre los factores y las variables. Para juzgar la bondad del modelo factorial al describir las variables originales, se puede computar la proporción de la varianza de cada variable que es explicada por el modelo factorial. Siendo que los factores no están correlacionados, el total de la proporción de varianza explicada es simplemente la suma de las proporciones de las varianzas explicadas por cada factor.

Para interpretar estas matrices se identifica la primera variable en el primer factor y se mueve horizontalmente hacia la derecha buscando a ver en cuál factor dicha variable tiene una mayor carga. Entonces se subraya el valor identificado y se procede de la misma manera con la siguiente variable hasta terminar con todas. Al final de proceso se espera tener debidamente identificadas las variables que más cargan en cada factor.

En la fase de extracción de factores se determina un número de factores comunes necesarios para describir los datos. Esta decisión es tomada en base a los eigenvalues y el porcentaje de la varianza total que aporta cada uno de los diferentes factores.

La fase de rotación

La rotación de factores pretende transformar la matriz inicial en una que sea más fácil de interpretar, lo cual es importante siendo que lo que se pretende es identificar factores que sean substancialmente significativos. Esta rotación puede ser ortogonal u oblicua. La rotación es ortogonal cuando los ejes de coordenadas se rotan manteniendo un ángulo de 90 grados entre ellos y eso supone que los factores identificados no se relacionan entre sí. En cambio si los ejes que se rotan conservan entre sí un ángulo diferente a 90 grados se trata de una rotación oblícua y supone cierto grado de relación entre los factores que lleguen a conformarse.

Cuando varios factores tienen una carga grande respecto a varias variables, resulta muy difícil determinar la forma como difieren los factores. La rotación no afecta la bondad de la solución factorial, y aunque la matriz factorial cambia, las comunalidades y los porcentajes de la varianza total explicada no cambian, pero si cambian los porcentajes atribuibles a cada factor si cambian. La rotación redistribuye la varianza explicada por los factores individuales. Así que diferentes métodos de rotación pueden conducir a la identificación de factores diferentes.

Método Varimax: es el más común y trata de minimizar el número de variables que tienen alta carga en un factor. Esto debe fortalecer la interpretabilidad de los factores. Esta es una rotación ortogonal. . Es el más utilizado y ofrece una clara separación entre factores.

Método Quartimax: enfatiza la interpretación simple de las variables siendo que la solución minimiza el número de factores necesarios para explicar la variable. Se enfoca en rotar el factor inicial de tal manera que una variable cargue alto en un factor y tan bajo como sea posible en todos los otros. Con esta rotación se obtiene que muchas variables carguen alto o muy cerca en el mismo factor siendo que trata de simplificar las filas. Este método no ha mostrado mucho éxito en producir estructuras más simples. Resulta en factores generales con alta hacia moderada carga en la mayoría de las variables. Esta es una de sus desventajas. (Parece un análisis de componente principal)

Método Equamax: es una combinación de los dos anteriores, que simplifica los factores y simplifica las variables. No tiene mucha aceptación y es utilizado muy pocas veces.

Para mejor interpretar los resultados

Es necesario agrupar las variables que tienen una carga grande respecto al mismo factor. Para esto se puede requerir del procesador estadístico que organice las variables de modo que aparezcan juntas las variables con carga grande en un factor. También se puede requerir que las cargas pequeñas sean omitidas. (carga grande o pequeña es relativo a cada estudio y contexto teórico).

No olvide que una vez se han agrupado las variables con los factores que más carga comparten, es indispensable un esfuerzo teórico lógico por parte del investigador para encontrar y demostrar significado y sentido a los resultados.

En el Anexo B se presenta un análisis factorial de referencia para el estudio del tema contenido en esta sección. De esta manera se pretende que el lector se relacione con la forma como el paquete estadístico ofrece los resultados.

Anexos disponibles por solicitud a tevgra@umontemorelos.edu.mx
Altius
tgrajales.net

J