Modelos de ecuaciones estructurales (Continuación)

Status	In progress
Assign	Sharop

3.5 Modelo de medición

Los modelos de medición son modelos estadísticos utilizados para describir la relación entre una variable observable y las variables latentes subyacentes que influyen en ella. Estos modelos se utilizan en diversas disciplinas, como la psicología, la sociología, la educación y la salud pública, entre otras, para medir constructos abstractos que no pueden medirse directamente. El modelo de medición en SEM es un modelo factorial confirmatorio.

Recordemos…

Un modelo de factor confirmatorio es un tipo de modelo estadístico utilizado para evaluar la adecuación de un conjunto de datos a una teoría o modelo conceptual previamente establecido. En general, se utiliza para investigar la estructura de relaciones entre variables latentes o constructos teóricos subyacentes que no son directamente observables.

Por ejemplo, imaginemos que un psicólogo está interesado en evaluar la estructura subyacente de las habilidades cognitivas de los estudiantes universitarios. El psicólogo podría plantear una hipótesis teórica que sugiere que estas habilidades están compuestas por tres factores subyacentes: habilidades verbales, habilidades numéricas y habilidades espaciales. Para evaluar esta hipótesis, el psicólogo podría recolectar datos mediante una prueba cognitiva y utilizar un modelo de factor confirmatorio para analizar si los datos se ajustan al modelo teórico.

El desarrollo de un modelo de factor confirmatorio implica varias etapas:

1. Especificación del modelo: en esta etapa, se define la estructura del modelo teórico y se establecen las relaciones entre las variables latentes y

observables.

2. Selección de medidas: se seleccionan las medidas o indicadores que se utilizarán para medir los constructos teóricos subyacentes.
3. Estimación de los parámetros del modelo: se utiliza un software estadístico para ajustar el modelo a los datos, estimando los parámetros del modelo (por ejemplo, las cargas factoriales, varianzas y covarianzas).
4. Evaluación del ajuste del modelo: se evalúa la calidad del ajuste del modelo a los datos utilizando varios criterios de ajuste, como el índice de ajuste comparativo (CFI), el índice de bondad de ajuste (GFI) y el error cuadrático medio de aproximación (RMSEA).
5. Interpretación del modelo: se interpreta el modelo ajustado y se extraen conclusiones sobre la estructura subyacente de los constructos teóricos.

En resumen, un modelo de factor confirmatorio es una herramienta estadística que se utiliza para evaluar la adecuación de los datos a una teoría o modelo conceptual previamente establecido.

En estos modelos, la variable observable se mide mediante un conjunto de indicadores o variables observables que se supone que reflejan la variable latente subyacente. Por ejemplo, la ansiedad es una variable latente que no se puede medir directamente, pero se puede medir mediante una serie de indicadores, como el nerviosismo, la inquietud, la preocupación, etc.

Los modelos de medición se basan en la teoría de los tests clásicos, que establece que los indicadores de una variable latente están afectados por el error de medición y por la varianza de la variable latente subyacente. Estos modelos permiten estimar los parámetros de la variable latente subyacente a partir de los datos observados de los indicadores.

Algunos de los modelos de medición más comunes son:

1. El modelo de medición factorial: Este modelo se utiliza para evaluar la estructura factorial de una variable latente y determinar si está compuesta por una o varias dimensiones subyacentes.
2. El modelo de medición de Rasch: Este modelo se utiliza para evaluar la validez y fiabilidad de un conjunto de ítems en una escala de medición.

3. El modelo de medición de respuesta al ítem: Este modelo se utiliza para evaluar la capacidad de una escala de medición para discriminar entre diferentes niveles de la variable latente subyacente.

En resumen, los modelos de medición son herramientas útiles para evaluar la validez y fiabilidad de una escala de medición, y para medir constructos abstractos que no pueden medirse directamente. Estos modelos permiten a los investigadores estimar los parámetros de la variable latente subyacente a partir de los datos observados de los indicadores.

▼ Ejemplos

1. Modelo de medición en psicología

   Supongamos que se está interesado en medir la inteligencia de un grupo de estudiantes utilizando tres pruebas diferentes: una prueba verbal, una prueba de habilidades numéricas y una prueba de razonamiento espacial. Se hipotetiza que la inteligencia es una variable latente que subyace a las tres pruebas, lo que significa que las puntuaciones en cada prueba son indicadores observables de la inteligencia.

   En este modelo, la variable latente de inteligencia se mide indirectamente a través de las tres pruebas. Cada prueba es un indicador observado que está asociado con la variable latente de inteligencia. Además, se incluyen tres errores de medición, uno para cada indicador observado, que capturan la variabilidad no explicada por la variable latente.

   Para analizar este modelo de medición, se pueden utilizar técnicas de análisis factorial confirmatorio (CFA) o de análisis de ecuaciones estructurales (SEM). Los resultados del análisis permitirán evaluar la validez de las pruebas como medidas de la inteligencia y determinar cuál es la estructura subyacente de la inteligencia.

2. Modelo de medición en marketing

   Otro ejemplo de modelo de medición se encuentra en el ámbito del marketing, donde se utilizan para evaluar la relación entre las variables latentes de la marca y la lealtad de los consumidores. Supongamos que se está interesado en medir la calidad percibida de una marca de automóviles utilizando tres dimensiones diferentes: calidad de fabricación, diseño y servicio al cliente. Se hipotetiza que la calidad percibida es una variable latente que subyace a las tres dimensiones, lo que significa que las puntuaciones en cada dimensión son indicadores observables de la calidad percibida.

En este modelo, la variable latente de calidad percibida se mide indirectamente a través de las tres dimensiones. Cada dimensión es un indicador observado que está asociado con la variable latente de calidad percibida. Además, se incluyen tres errores de medición, uno para cada indicador observado, que capturan la variabilidad no explicada por la variable latente.

Para analizar este modelo de medición, se pueden utilizar técnicas de análisis factorial confirmatorio (CFA) o de análisis de ecuaciones estructurales (SEM). Los resultados del análisis permitirán evaluar la validez de las dimensiones como medidas de la calidad percibida de la marca y determinar cuál es la estructura subyacente de la calidad percibida.

▼ Referencias

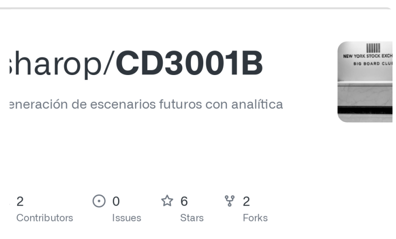
- Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.
- Byrne, B. M. (2016). Structural equation modeling with AMOS: Basic concepts, applications, and programming (3rd ed.). New York: Routledge.
- Kline, R. B. (2016). Principles and practice of structural equation modeling (4th ed.). New York: Guilford Press.

Ejemplo: Para ilustrar el desarrollo completo de una solución a un problema utilizando modelos de medición y Python, vamos a utilizar el ejemplo de un estudio de satisfacción del cliente en un restaurante.

CD3001B/Modelo_de_medicion.ipynb at main · sharop/CD3001B

Generación de escenarios futuros con analítica. Contribute to sharop/CD3001B development by creating an account on GitHub.

https://github.com/sharop/CD3001B/blob/main/SEM_2/Modelo_de_medicion.ipynb



▼ Modelo de senderos (path analysis), efectos mediadores y efectos indirectos

El modelo de senderos, también conocido como análisis de ruta o path analysis, es una técnica estadística utilizada para analizar las relaciones causales entre variables en un modelo teórico. En este modelo, se representan las relaciones causales entre las variables mediante flechas o senderos, y se estiman los coeficientes de regresión que indican el efecto de una variable sobre otra.

Características

- Representación esquemática de un modelo teórico usando una notación estándar.
- Ecuaciones de regresión especificadas entre variables medidas.
- Los 'efectos' de las variables predictoras sobre las variables de criterio/dependientes pueden ser:
  - Directo
  - Indirecto
  - Total

En path analysis no estaremos usando variables latentes, sino variables que están directamente observadas.

Notación

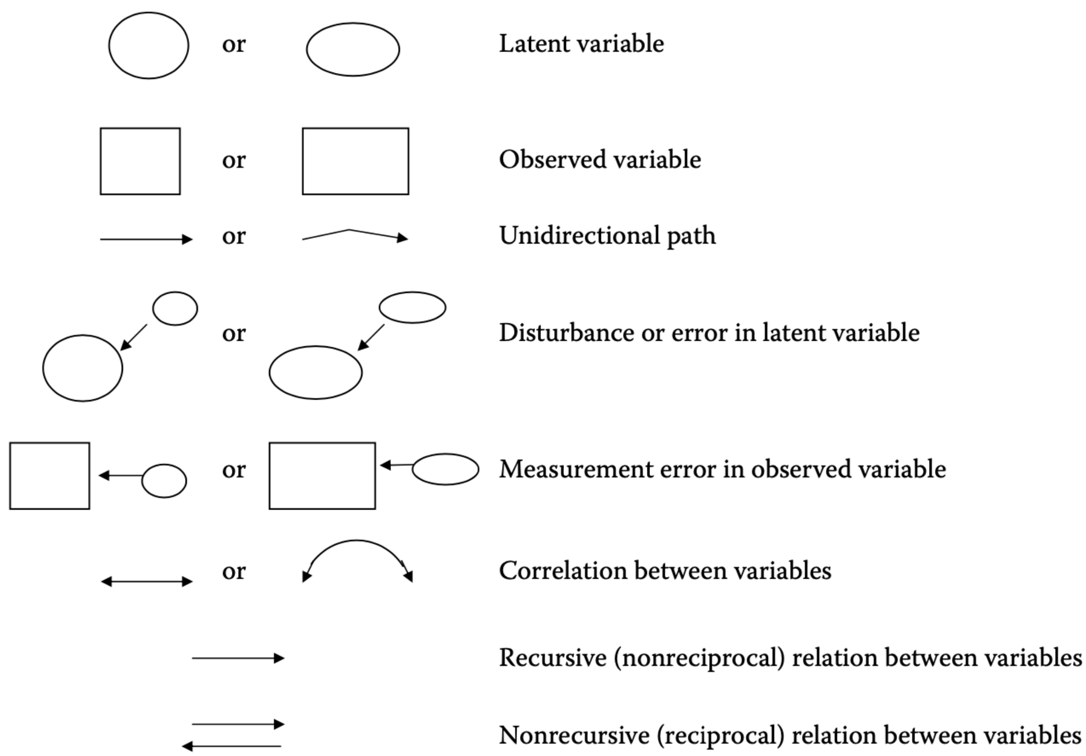


Imagen tomada de “A Beginner’s Guide to Structural Equation Modeling”, Randall E. Schumacker

Ejemplos:

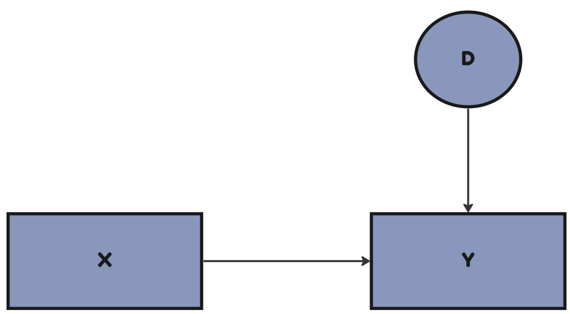


Figura 2

La figura 2, muestra que X tiene un efecto causal en Y y el termino D es el termino de error.

$$Y = \beta X + D$$

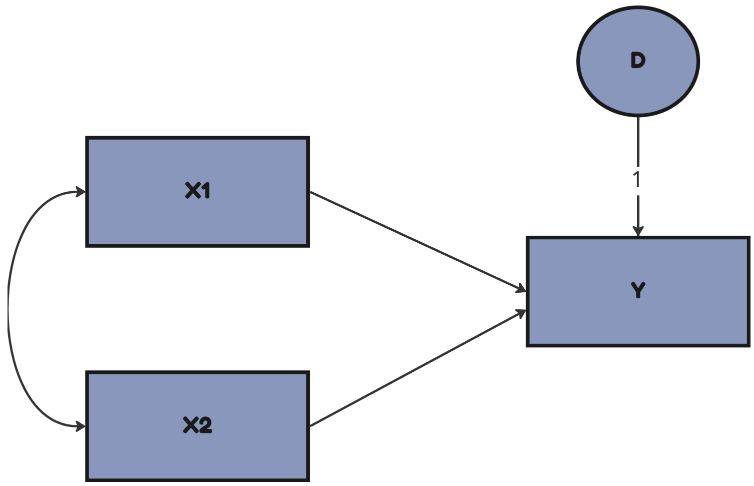


Figura 3

La figura 3 muestra una regresión lineal multiple con dos variable independientes y una variable dependiente Y; con un error etiquetado como D.

Efecto indirectos y efectos mediadores.

Los efectos mediadores y efectos indirectos son conceptos relacionados con el modelo de senderos. Un efecto mediador se produce cuando una variable intermedia (también conocida como variable mediadora) explica la relación causal entre dos variables. Por ejemplo, si se observa que la relación entre el nivel de educación y el ingreso está mediada por la experiencia laboral, se dice que la experiencia laboral es un efecto mediador.

Por otro lado, un efecto indirecto se produce cuando el efecto de una variable sobre otra variable se produce a través de una cadena de efectos mediadores. Por ejemplo, si se observa que el nivel de educación influye en el ingreso a través de la experiencia laboral y la competencia técnica, se dice que la competencia técnica es un efecto indirecto.

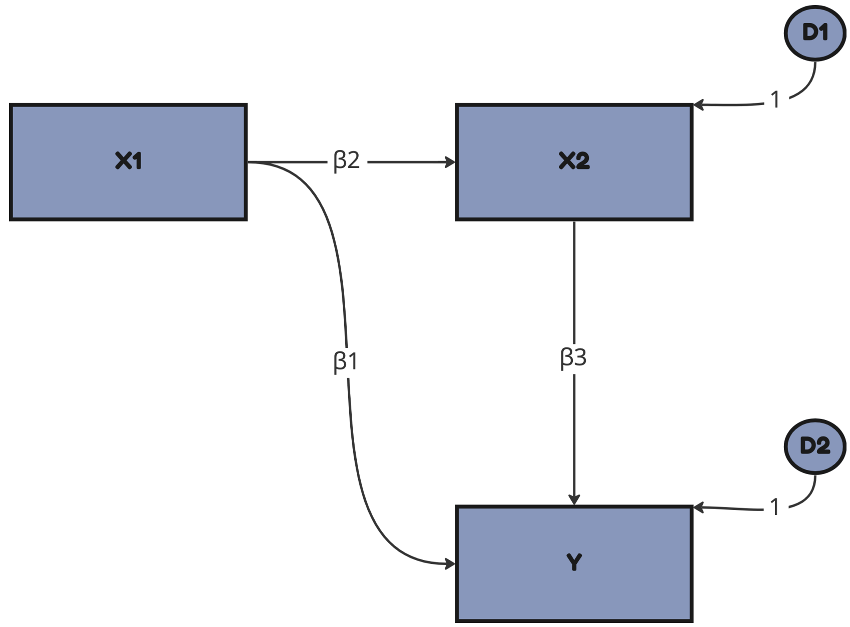


Figura 4

En la figura 4 se muestra que X1 tiene efectos directos con X2 y también tiene efectos directos en Y, así que ahora tienen efectos indirectos de X1 en Y a travez de X2. Y los componentes Beta son los coeficientes de regresión. entonces en la figura anterior tenemos que:

$\beta 1$ = Efecto directo de $X1$ en Y

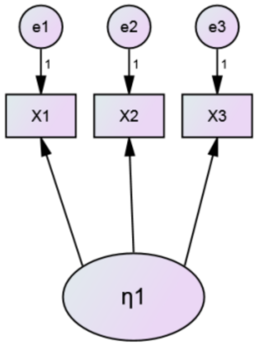
$\beta 2$ = Efecto directo de $X1$ en $X2$

$\beta 3$ = Efecto directo de $X2$ en Y

$\beta 2 * \beta 3$=efecto indirecto de $X1$ en Y

$\beta 1 + (\beta 2 * \beta 3)$= Efecto total de $X1$ en Y

¿Como podemos leer los diagramas?

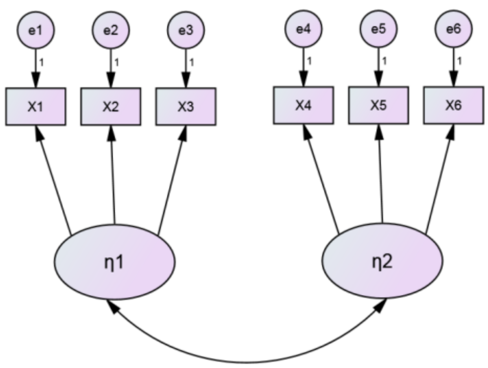


Con tres varianzas de error

Causada/medida por tres variables observadas

Una variable latente

Figura 5



de 2 variables latentes, cada una medida por 3 variable observadas.

Correlación

Figura 6

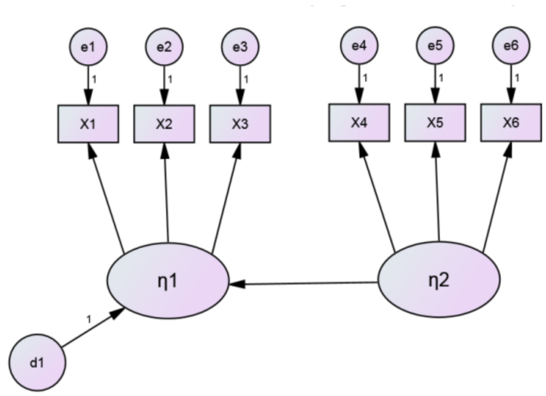


Figura 7

Variables endogenas y exogenas.

Cuando trabajamos con diagramas, es necesario distinguir entre dos tipos de variables.

- Endogenas (dependiente)
  - Causada por variables en el sistema.
- Exogenas (independiente)
  - Causada por variables fuera del sistema.
- En modelos estructurales una variable puede ser un predictor y un resultado.

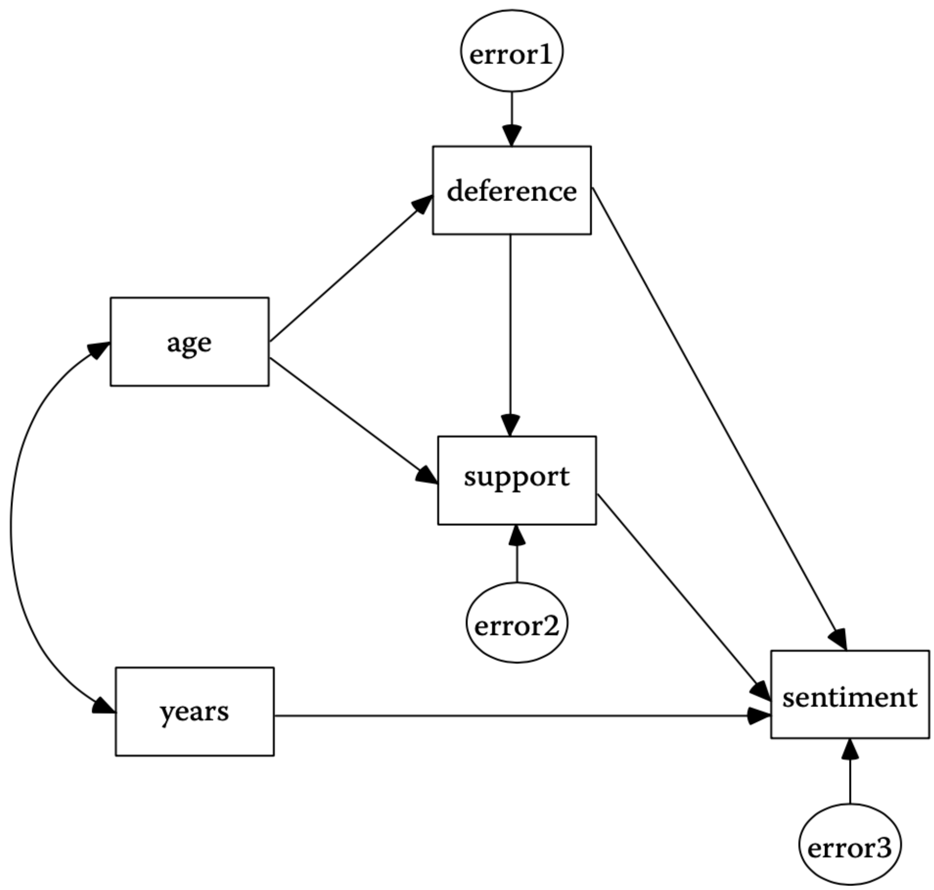
En la figura 6 tenemos 2 variables latentes exogenas, son exogenas por que no hay un camino dirigido, apuntando a alguna de ellas. Ninguna de ellas tiene un termino de error. Es solo una correlación lo único que se observa.

en la figura 7, se observa $\eta 1$ y $\eta 2$, en este caso $\eta 1$ es endogena y $\eta 2$ es exogena. $\eta 2$ no tiene ningún camino directo hacia ella y no tiene un termino de error asociado. Mientras que $\eta 1$ tiene un termino de error asociado y tiene el camino que $\eta 2$ apunta hacia ella $\eta 1$.

Ejemplo

Un modelo de senderos que se utiliza en un ejemplo de datos recolectados por McDonald y Clelland en 1984 sobre la actitud de los trabajadores textiles del sur de Estados Unidos hacia los sindicatos. El modelo consiste en cinco variables observadas y dos variables independientes (años trabajados y edad) y tres variables dependientes (deferencia, apoyo a la actividad sindical y sentimiento hacia los sindicatos). La muestra n, es de 173 trabajadores.

El diagrama del modelo propuesto es el siguiente:



Las lineas directas desde una variable observada a otra variable observada denota **efectos directos,** una influencia directa de una variable a otra. Por ejemplo, se realiza la hipótesis que la edad tiene una influencia directa sobre el apoyo, lo que significa que la edad del trabajador puede influir en un aumento (o disminución) en el apoyo.

Una línea curva y de doble flecha entre dos variables observadas independientes indica covarianza; es decir, están correlacionadas. En este ejemplo, se especifica que la edad y los años se correlacionan.

Finalmente, cada variable dependiente tiene un término de error, indicado por un círculo alrededor del término de error que apunta hacia la variable dependiente correspondiente. Tomemos como ejemplo "deference", parte de la varianza en las puntuaciones de "deference" se explicará o se predecirá por la edad, mientras que otra parte no lo hará. La varianza no explicada se convierte en el término de error, que indica otras posibles influencias en "deference" que no se contemplan en el modelo de ruta especificado.

La matriz de varianza-covarianza original, la matriz de varianza-covarianza del modelo implícito, la matriz residual y la matriz residual estandarizada se presentan en la tabla siguiente:

Original, Reproduced, Residual, and Standardized Residual Covariance Matrices for the Initial Union Sentiment Model

Original Matrix

Variable	Deference	Support	Sentiment	Years	Age
Deference	14.610				
Support	–5.250	11.017			
Sentiment	–8.057	11.087	31.971		
Years	–0.482	0.677	1.559	1.021	
Age	–18.857	17.861	28.250	7.139	215.662

Reproduced Matrix

Variable	Deference	Support	Sentiment	Years	Age
Deference	14.610				
Support	–1.562	11.017			
Sentiment	–5.045	10.210	30.534		
Years	–0.624	0.591	1.517	1.021	
Age	–18.857	17.861	25.427	7.139	215.662

Residual Matrix

Variable	Deference	Support	Sentiment	Years	Age
Deference	0.000				
Support	–3.688	0.000			
Sentiment	–3.012	0.877	1.437		
Years	0.142	0.086	0.042	0.000	
Age	0.000	0.000	2.823	0.000	0.000

Standardized Residual Matrix

Variable	Deference	Support	Sentiment	Years	Age
Deference	0.000				
Support	–4.325	0.000			
Sentiment	–3.991	3.385	3.196		
Years	0.581	0.409	0.225	0.000	
Age	0.000	0.000	0.715	0.000	0.000

El objetivo del modelo es examinar las relaciones directas e indirectas entre las variables y comprender mejor la actitud de los trabajadores textiles hacia los sindicatos.

▼ Identificación del modelo

Es crucial resolver la identificación del problema, antes de la identificación del modelo.

Para el problema de identificación, hacemos la siguiente pregunta: **¿Se puede encontrar un conjunto único de estimaciones de parámetros en función de los datos de muestra contenidos en la matriz de covarianza de muestra S y el modelo teórico implícito por la matriz de covarianza de población Σ?**

Para el modelo de “sentimiento sindical”, por ejemplo, nos gustaría saber si se identifica el camino entre la edad y la deferencia; un ejemplo de un parámetro a estimar.

En el modelo de sentimiento sindical, algunos parámetros están fijos y otros son libres. Un ejemplo de un parámetro fijo es que no hay un camino o relación directa entre la edad y el sentimiento. Un ejemplo de un parámetro libre es que hay un camino o relación directa entre la edad y la deferencia.

Al determinar la identificación, primero se debe considerar la condición de orden. Aquí, el número de parámetros libres a estimar debe ser menor o igual al número de valores distintos en la matriz S. En nuestro modelo de ruta, especificamos lo siguiente:

- 6 coeficientes de ruta
- 3 varianzas de error de ecuación
- 1 correlación entre las variables independientes
- 2 varianzas de las variables independientes.

Por lo tanto, hay un total de 12 parámetros libres que deseamos estimar. El número de valores distintos en la matriz S es igual a:

Donde p es el número de variables observadas en la matriz. Por lo tanto, el número de valores distintos en la matriz de muestra S, 15, es de hecho mayor que el número de parámetros libres, 12. Sin embargo, esto es solo una condición necesaria y no garantiza que el modelo esté identificado. Según la condición de orden, el modelo también está sobredeterminado porque hay más valores en S que

parámetros a estimar.
Aunque la condición de orden es fácil de evaluar, otras condiciones suficientes no lo son, por ejemplo, la condición de rango. Las condiciones suficientes nos obligan a determinar algebraicamente si cada parámetro en el modelo se puede estimar a partir de la matriz de covarianza S.

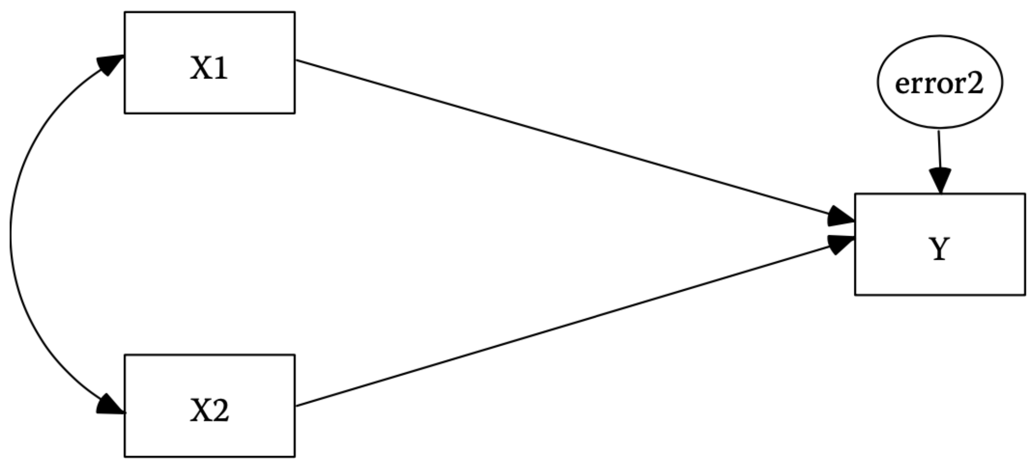
▼ Estimación del modelo

Una vez que se ha abordado el problema de identificación, el siguiente paso es estimar los parámetros del modelo especificado. El enfoque tradicional de la estimación en análisis de trayectorias es la descomposición de la matriz de correlación, que consiste en que si se incluyen todos los posibles efectos unidireccionales o recursivos en un modelo de trayectorias, se puede reproducir completamente la matriz de correlación observada.

Se pueden estimar los parámetros mediante diferentes procedimientos de estimación, como la máxima verosimilitud (ML), mínimos cuadrados generalizados (GLS) y mínimos cuadrados no ponderados (ULS), así como estimaciones estandarizadas (los coeficientes de trayectoria), y existen distintas funciones de estimación de información completa y limitada.

En el ejemplo del modelo de sentimiento sindical, se realiza una modificación del modelo inicialmente especificado y se ejecuta una estimación máxima verosimilitud. Todos los parámetros estimados son significativamente diferentes de cero, lo que indica que el modelo está ajustando bien a los datos.

Por ejemplo tomamos el siguiente modelo:



(b)

Aquí hay dos efectos directos, de X1 a Y y de X2 a Y. También hay efectos indirectos debido a la correlación entre X1 y X2. En otras palabras, X1 influye indirectamente en Y a través de X2, y también X2 afecta indirectamente a Y a través de X1. Las correlaciones entre estas tres variables se pueden descomponer de la siguiente manera:

1. $r_{12} = p_{12}$

   (CO)

2. $r_{Y1} = p_{Y1} + p_{12}p_{Y2}$

   (DE) (IE)

3. $r_{Y2} = p_{Y2} + p_{12}p_{Y1}$

   (DE) (IE)

donde los valores r son las correlaciones observadas reales y los valores p son los coeficientes de trayectoria (estimaciones estandarizadas). Así, en la ecuación (1), la correlación entre X1 y X2 es simplemente una función de la trayectoria o relación de correlación (CO) entre X1 y X2. En la ecuación (2), la correlación entre X1 e Y es una función de (a) el efecto directo (DE) de X1 sobre Y y (b) el efecto indirecto (IE) de X1 sobre Y a través de X2 [el producto de la trayectoria o correlación entre X1 y X2 (p12) y el efecto directo desde X2 hasta Y (pY2)]. La ecuación (3) es similar a la ecuación (2) excepto que X1 y X2 están invertidos; hay tanto un efecto directo como un efecto indirecto.

En el ejemplo, vemos el proceso de estimación en acción. Evaluamos este modelo inicial y esperamos, a través del proceso de modificación del modelo, obtener el modelo como se especificó originalmente El modelo intencionalmente mal especificado se ejecutó.
Las estimaciones de máxima verosimilitud para el modelo inicial se muestran en la primera columna de la Tabla siguiente. Todas las estimaciones de parámetros son significativamente diferentes de cero, $p < .05$.

La edad tiene un efecto directo tanto en deferencia como en soporte; la deferencia tiene un efecto directo sobre el sentimiento; los años tienen un efecto directo sobre el sentimiento; y el soporte tiene un efecto directo sobre el sentimiento. También hay numerosos efectos indirectos en el modelo de ruta, como el efecto indirecto de la edad en el sentimiento a través del soporte. La edad y los años también tienen una covarianza significativa, lo que indica que una o más variables comunes no medidas influyen en ambas variables.

Maximum Likelihood Estimates[a] and Selected Fit Indices for the Initial and Final Union Sentiment Models

Paths	**Initial Model**	**Final Model**
Age → deference	−.09	−.09
Age → support	0.08	0.06
Deference → support	—	−.28
Years → sentiment	0.86	0.86
Deference → sentiment	−.22	−.22
Support → sentiment	0.85	0.85
Equation error variances		
Deference	12.96	12.96
Support	9.54	8.49
Sentiment	19.45	19.45
Independent variables		
Variance (age)	215.66	215.66
Variance (years)	1.02	1.02
Covariance (age, years)	7.14	7.14
Selected fit indices		
χ^2	19.96	1.25
df	4	3
p value	.00	.74
RMSEA	.15	.00
SRMR	.087	.015
GFI	.96	1.00

[a] All estimates significantly different from zero ($p < .05$).

>

▼ Pruebas del modelo

Un importante resultado de cualquier análisis de senderos es el ajuste del modelo especificado. Si el ajuste del modelo de senderos es bueno, entonces el modelo especificado ha sido respaldado por los datos de muestra. Si el ajuste del modelo de senderos no es tan bueno, entonces el modelo especificado no ha sido respaldado por los datos de muestra, y el investigador generalmente intenta modificar el modelo de senderos para lograr un mejor ajuste

Para el ejemplo que se ha estado siguiendo, incluimos algunos índices de ajuste de modelo en la parte inferior de la Tabla anterior. Para el modelo de senderos inicial, la estadística χ^2, técnicamente una medida de mala calidad de ajuste, es igual a 19.96, con cuatro grados de libertad, y $p < 0.01$. Como el valor p es muy pequeño y el valor χ^2 no se acerca al número de grados de libertad, entonces según esta medida de ajuste, el modelo de senderos inicial está mal especificado. El error de aproximación de raíz cuadrada media (RMSEA) es igual a 0.15, algo por debajo del nivel aceptable para esta medida de ajuste (RMSEA < 0.08 o 0.05). El residuo de raíz cuadrada media estandarizado (SRMR) es 0.087, también por debajo del nivel usualmente aceptable de ajuste (SRMR < 0.08 o 0.05). Finalmente, el índice de bondad de ajuste (GFI) es 0.96 para el modelo inicial, lo cual es un nivel aceptable para esta medida de ajuste (GFI > 0.95). En este conjunto particular de índices de ajuste de modelo, la conclusión es que el ajuste de los datos al modelo se acerca a un nivel razonable, pero que algunas modificaciones del modelo podrían permitirnos lograr un mejor ajuste del modelo entre la matriz de varianza-covarianza de la muestra S y la matriz de varianza-covarianza implícita (reproducida) Σ del modelo. La modificación del modelo se considera en la siguiente sección.

▼ Modificación del modelo

El último paso en el modelado de ecuaciones estructurales es la modificación del modelo. En otras palabras, si el ajuste del modelo es menos que satisfactorio, el investigador típicamente realiza una búsqueda de especificación para encontrar un modelo de ajuste mejorado. Como se describe en los capítulos 4 y 5, se pueden utilizar varios procedimientos diferentes para ayudar en esta búsqueda. Se pueden eliminar parámetros que no son significativamente diferentes de cero y/o incluir parámetros adicionales para llegar a un modelo modificado. Para la eliminación de parámetros, el procedimiento más comúnmente utilizado en LISREL es comparar la estadística t para cada parámetro con un valor t tabulado (por ejemplo, $t > 1,96$) para determinar la significación estadística.

Para la inclusión de parámetros adicionales, las técnicas más comúnmente utilizadas en LISREL son (a) el índice de modificación (MI) (el valor esperado que disminuiría c2 si se incluyera dicho parámetro; valores grandes indican parámetros potencialmente útiles), y (b) la estadística de cambio esperado del parámetro (EPC) (el valor aproximado del nuevo parámetro si se agrega al modelo).

Además, un examen de la matriz de residuos, o la matriz de residuos estandarizados más útil, a menudo da pistas sobre qué covarianza o correlaciones originales no están bien representadas por el modelo. Recuerde que la matriz de residuos es la diferencia entre la matriz de varianza-covarianza observada S y la matriz de varianza-covarianza modelo implícita (reproducida) Σ. Los residuos grandes indican valores que no están bien representados por el modelo. Los residuos estandarizados son como puntuaciones z en el sentido de que valores grandes (mayores que 1,96 o 2,58) indican que una relación en particular no está bien representada por el modelo de la ruta (Tabla A.1).

Para el ejemplo inicial de la unión de sentimientos, se presentan las matrices de covarianza residuales, reproducidas, originales y estandarizadas en la Tabla 7.1. Aquí vemos que el residuo estandarizado más grande es entre deferencia y apoyo (-4,325). Las estadísticas t no sugieren la eliminación de ningún parámetro existente del modelo de ruta inicial porque cada parámetro es estadísticamente diferente de cero. Con respecto a la posible inclusión de nuevos parámetros, el índice de modificación más grande es para la ruta de deferencia a apoyo (MI = 18,9). Para esa ruta potencial, el valor estimado, o cambio esperado del parámetro (EPC), es -0,28.

En conjunto, estas estadísticas indican que hay algo **mal especificado** entre deferencia y apoyo que no es capturado por el modelo inicial. Específicamente, se recomienda agregar un camino desde la deferencia hasta el apoyo. Este es precisamente el camino del modelo de camino original que eliminamos intencionalmente del modelo inicial. Por lo tanto, la búsqueda de especificación tuvo éxito en la obtención del modelo original. Las estimaciones de ML y los índices de ajuste seleccionados para el modelo final, donde se incluye este camino, se muestran en la segunda columna de la Tabla previa.

Todos los parámetros incluidos son significativamente diferentes de cero ($p < 0.05$), todos los índices de ajuste indican ahora un nivel aceptable de ajuste y no se indican índices de modificación adicionales para más cambios recomendados. Por lo tanto, consideramos este como el modelo de camino final para el ejemplo de sentimiento de la unión.

▼ Ejemplo

Para este ejemplo, utilizaremos un conjunto de datos de la Encuesta Nacional de Salud y Nutrición de los Estados Unidos (NHANES) para investigar la relación entre la actividad física y la obesidad.

Primero, instalamos la librería `semopy` mediante el siguiente comando:

```
pip install semopy
```

Luego, cargamos los datos del archivo CSV utilizando la librería `pandas`:

```
import pandas as pd

data = pd.read_csv("https://raw.githubusercontent.com/GTPB/PSLS20/master/data/NHANES.
csv")
```

A continuación, definimos las variables que vamos a utilizar en nuestro modelo y creamos una matriz de correlación para evaluar la relación entre ellas:

```
import numpy as np

# Definimos las variables
x1 = data['age']
x2 = data['gender']
x3 = data['physical_activity']
y = data['obesity']

# Creamos la matriz de correlación
corr = np.corrcoef([x1, x2, x3, y])
```

Después, definimos el modelo de senderos utilizando la notación estándar y los coeficientes de regresión a partir de la matriz de correlación:

```
from semopy import Model

# Definimos el modelo utilizando la notación estándar
model = Model('''
    age -> obesity
    gender -> obesity
    physical_activity -> obesity
    age -> physical_activity
''')

# Establecemos los coeficientes de regresión a partir de la matriz de correlación
model.x[0, 3] = corr[0, 3]
model.x[1, 3] = corr[1, 3]
model.x[2, 3] = corr[2, 3]
model.x[0, 2] = corr[0, 2]
```

En el modelo definido, `age`, `gender`, y `physical_activity` son las variables predictoras y `obesity` es la variable de criterio o dependiente. Además, en el modelo se establecen las siguientes relaciones:

- `age` tiene un efecto directo sobre `obesity`
- `gender` tiene un efecto directo sobre `obesity`
- `physical_activity` tiene un efecto directo sobre `obesity`
- `age` tiene un efecto directo sobre `physical_activity`

Finalmente, ajustamos el modelo y evaluamos los resultados:

```
from semopy import OptimalScaler, Model, sem

# Escalamos los datos óptimamente
scaler = OptimalScaler()
data_scaled = scaler.fit_transform(data)

# Ajustamos el modelo utilizando SEM
fit = sem(model, data_scaled)

# Evaluamos los resultados
print(fit.summary())
```

Para poder ajustar el modelo, primero escalamos los datos óptimamente utilizando la clase `OptimalScaler`. Luego, ajustamos el modelo utilizando el método `sem` de la librería `semopy`. Finalmente, evaluamos los resultados mediante el método `summary` que nos muestra una tabla que contiene los coeficientes de regresión, los errores estándar, los valores t y los valores p correspondientes para cada variable.

Es importante mencionar que el modelo de senderos puede ser modificado y ajustado según sea necesario, y que `semopy` ofrece varias funciones y métodos para hacerlo.

En conclusión, este ejemplo detallado te muestra cómo realizar un análisis de senderos en Python utilizando la librería `semopy`. Espero que te haya sido útil y que puedas aplicar esta técnica en tus propios proyectos de análisis de datos. Si tienes alguna duda o comentario, no dudes en hacerlo saber. ¡Buena suerte!

▼ Ejemplos

Los modelos de senderos, también conocidos como modelos de ecuaciones estructurales o path analysis en inglés, son una técnica estadística que permite examinar las relaciones entre variables en un modelo causal. En estos modelos, se representan gráficamente las relaciones entre variables y se estima la magnitud y dirección de las influencias entre ellas. A continuación se presentan dos ejemplos detallados de modelos de senderos:

Ejemplo 1: Modelo de senderos para la relación entre estrés, ansiedad y depresión

Imaginemos que queremos investigar la relación entre el estrés, la ansiedad y la depresión. Se sabe que el estrés puede tener un impacto directo sobre la ansiedad y la depresión, pero también puede influir indirectamente a través de la ansiedad. En este caso, podemos construir un modelo de senderos para evaluar estas relaciones, como se muestra en la siguiente figura:

```
markdownCopy code
Estrés → Ansiedad → Depresión
      ↘_________________↙
```

En este modelo, se establecen las relaciones directas entre el estrés y la ansiedad, y entre la ansiedad y la depresión, así como la relación indirecta entre el estrés y la depresión a través de la ansiedad. La flecha que apunta desde el estrés hacia la ansiedad indica que el estrés tiene un impacto directo sobre la ansiedad, mientras que la flecha que apunta desde la ansiedad hacia la depresión indica que la ansiedad tiene un impacto directo sobre la depresión. La flecha que apunta desde el estrés hacia la depresión a través de la ansiedad indica que el estrés también tiene un impacto indirecto sobre la depresión a través de la ansiedad.

Una vez que se ha construido el modelo de senderos, se pueden estimar los coeficientes de regresión que representan la magnitud y dirección de las relaciones entre las variables. Estos coeficientes pueden ayudar a responder preguntas como "¿Cuál es el impacto directo del estrés sobre la depresión?" y "¿En qué medida la ansiedad media la relación entre el estrés y la depresión?".

Ejemplo 2: Modelo de senderos para la relación entre variables socioeconómicas y consumo de tabaco

Otro ejemplo de modelo de senderos puede ser para examinar la relación entre variables socioeconómicas y el consumo de tabaco en jóvenes. En este caso, se puede construir un modelo de senderos como el siguiente:

```
Edad → Género → Nivel socioeconómico → Consumo de tabaco
```

En este modelo, se establecen las relaciones directas entre la edad, el género y el nivel socioeconómico sobre el consumo de tabaco. La flecha que apunta desde la edad hacia el consumo de tabaco indica que la edad tiene un impacto directo sobre el consumo de tabaco, mientras que la flecha que apunta desde el género hacia el consumo de tabaco indica que el género también tiene un impacto directo sobre el consumo de tabaco. La flecha que apunta desde el nivel socioeconómico hacia el consumo de tabaco indica que el nivel socioeconómico tiene un impacto directo sobre el consumo de tabaco.

Ejemplo 3: Modelo de senderos en ecología

Otro ejemplo de modelo de senderos se encuentra en el ámbito de la ecología, donde se utilizan para explorar las relaciones entre las variables ambientales y la abundancia de especies. Supongamos que se está interesado en estudiar las relaciones entre la temperatura del agua, la calidad del agua y la abundancia de una especie de pez en un río.

En este modelo, la temperatura del agua y la calidad del agua son variables exógenas que influyen directamente en la abundancia de la especie de pez. Además, se incluye un efecto mediador de la calidad del agua en la relación entre la temperatura del agua y la abundancia de la especie de pez.

Para analizar este modelo de senderos, se pueden utilizar técnicas de análisis de regresión múltiple o de análisis de varianza (ANOVA) para evaluar la significatividad y la fuerza de las relaciones entre las variables. Los resultados del análisis permitirán determinar cuáles son las variables ambientales más importantes para explicar la variabilidad en la abundancia de la especie de pez.

▼ Referencias

- Kenny, D. A. (1979). Mediation. En J. P. Robinson, L. C. Wrightsman y P. R. Hampson (Eds.), Measurement of personality and social psychological attitudes (pp. 143-166). San Diego, CA: Academic Press.
- Baron, R. M., y Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51(6), 1173-1182.
- Hayes, A. F. (2013). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. Nueva York, NY: Guilford Press.
- Preacher, K. J., y Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behavior Research Methods, 40(3), 879-891.

Algunos videos que pueden resultar útiles para entender el modelo de senderos y los efectos mediadores y efectos indirectos son:

- "Path Analysis: A Brief Introduction" por Michael Furr (**https://www.youtube.com/watch?v=3qfcfJjNqy0**)
- "Mediation Analysis" por Kristopher Preacher (**https://www.youtube.com/watch?v=9SCAGexvT8k**)
- "Indirect Effects in Structural Equation Modeling" por Kenny y McCoach (**https://www.youtube.com/watch?v=zsYhB2V2kFw**)

▼ Indicadores de ajuste

Los indicadores de ajuste son medidas utilizadas para evaluar qué tan bien un modelo estadístico se ajusta a los datos observados. Estos indicadores permiten comparar diferentes modelos y determinar cuál es el más adecuado para explicar los datos.

Algunos de los indicadores de ajuste más comunes son:

- Chi-cuadrado (χ^2): mide la discrepancia entre los valores observados y esperados en el modelo. Un valor pequeño de χ^2 indica que el modelo se ajusta bien a los datos.
- Índice de ajuste comparativo (CFI): compara el modelo propuesto con un modelo nulo que no tiene variables predictoras. Valores cercanos a 1 indican un buen ajuste del modelo.
- Raíz del error cuadrático medio (RMSEA): mide la cantidad de error en el modelo en relación con el grado de libertad. Valores cercanos a 0.05 indican un buen ajuste del modelo.
- Índice de bondad de ajuste (GFI): mide el grado de correlación entre los valores observados y los valores estimados en el modelo. Valores cercanos a 1 indican un buen ajuste del modelo.
- Estadístico Akaike (AIC): mide la calidad del modelo en relación con el número de parámetros estimados. Valores más pequeños de AIC indican un mejor ajuste del modelo.

Algunas referencias para consultar sobre los indicadores de ajuste son:

- Kline, R. B. (2016). Principles and Practice of Structural Equation Modeling (4ª ed.). Nueva York, NY: Guilford Press.
- Hu, L. T., y Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6(1), 1-55.
- Hair, J. F., Black, W. C., Babin, B. J., y Anderson, R. E. (2010). Multivariate Data Analysis (7ª ed.). Upper Saddle River, NJ: Prentice Hall.

Algunos ejemplos de cómo se utilizan los indicadores de ajuste en modelos de regresión y análisis factorial se pueden encontrar en:

- "Using goodness-of-fit indices in regression analysis" por William P. Fisher (**https://www.youtube.com/watch?v=kXhB1V7HmrA**)
- "Factor Analysis - Confirmatory Factor Analysis - Part 5 - Fit Indices" por Damián Pérez Valdés (**https://www.youtube.com/watch?v=3bAdM_vT6sc**)

También se pueden encontrar tutoriales y explicaciones más detalladas sobre los diferentes indicadores de ajuste en línea, como por ejemplo en la página web de la Enciclopedia de Métodos de Investigación en Ciencias Sociales (**https://www.socialresearchmethods.net/kb/fit.php**) o en el libro electrónico "A Guide to Structural Equation Modeling" de G. David Garson (**https://statisticalassociates.com/ebook/sem-guide/**).

▼ Ejemplos y material de apoyo

▼ Ejemplo de análisis factorial en Python

El análisis factorial es una técnica estadística utilizada para describir la variabilidad de un conjunto de variables observadas en términos de un número menor de variables latentes subyacentes. Estas variables latentes se denominan factores y se utilizan para explicar la covarianza entre las variables observadas. En Python, se puede realizar un análisis factorial utilizando la librería `factor_analyzer`. A continuación, te presentamos un ejemplo de cómo realizar un análisis factorial con esta librería:

```
import pandas as pd
from factor_analyzer import FactorAnalyzer

# Cargamos los datos
data = pd.read_csv('datos.csv')

# Creamos un objeto FactorAnalyzer y ajustamos el modelo
fa = FactorAnalyzer(n_factors=3, method='ml', rotation='varimax')
fa.fit(data)

# Obtenemos los resultados
loadings = fa.loadings_
communality = fa.get_communalities()
eigenvalues = fa.get_eigenvalues()

# Imprimimos los resultados
print('Loadings:')
print(loadings)
print('Communalities:')
print(communality)
print('Eigenvalues:')
print(eigenvalues)
```

En este ejemplo, cargamos los datos desde un archivo CSV utilizando la librería `pandas`, creamos un objeto `FactorAnalyzer` y ajustamos el modelo utilizando el método de máxima verosimilitud (`ml`) y la rotación varimax. Luego, obtenemos los resultados del análisis, incluyendo los loadings, los communalities y los eigenvalues, y los imprimimos en la consola.

▼ Ejemplo de análisis de componentes principales en Python

El análisis de componentes principales (PCA) es una técnica estadística utilizada para reducir la complejidad de un conjunto de variables observadas mediante la identificación de un número menor de componentes principales que explican la mayor parte de la varianza de los datos. En Python, se puede realizar un análisis de componentes principales utilizando la librería `sklearn`. A continuación, te presentamos un ejemplo de cómo realizar un análisis de componentes principales con esta librería:

```
import pandas as pd
from sklearn.decomposition import PCA

# Cargamos los datos
data = pd.read_csv('datos.csv')

# Creamos un objeto PCA y ajustamos el modelo
pca = PCA(n_components=3)
pca.fit(data)

# Obtenemos los resultados
loadings = pca.components_
variance = pca.explained_variance_ratio_
scores = pca.transform(data)

# Imprimimos los resultados
print('Loadings:')
print(loadings)
print('Variance explained:')
print(variance)
print('Scores:')
print(scores)
```

En este ejemplo, cargamos los datos desde un archivo CSV utilizando la librería `pandas`, creamos un objeto `PCA` y ajustamos el modelo utilizando tres componentes principales. Luego, obtenemos los resultados del análisis, incluyendo los loadings, la varianza explicada y los scores, y los imprimimos en la consola.

▼ MLE

Estimador de Maxima Verosimilitud (MLE, Maximun Likelihood Estimator)

- El estimador de máxima verosimilitud (MLE) es un método utilizado para encontrar los valores de los parámetros del modelo que maximizan la verosimilitud de los datos observados.
- La verosimilitud es una medida de la probabilidad de que los datos observados se ajusten a un modelo estadístico específico. En términos simples, la verosimilitud es una medida de qué tan probable es que los datos observados sean generados por un modelo en particular.

- El log(L) de un modelo puede ser usado para probar el ajuste contra mas/menos una linea base restrictiva.

El logaritmo de la verosimilitud (log(L)) se utiliza para evaluar el ajuste de un modelo a los datos en comparación con una línea base restrictiva. En términos simples, se compara la verosimilitud del modelo ajustado con la verosimilitud de un modelo nulo o lineal que no tiene en cuenta ninguna relación entre las variables. Si el logaritmo de la verosimilitud del modelo ajustado es significativamente mayor que el del modelo nulo, se concluye que el modelo ajustado proporciona un mejor ajuste a los datos que el modelo nulo. Por lo tanto, el logaritmo de la verosimilitud se utiliza como una medida de la bondad de ajuste del modelo y se utiliza para comparar diferentes modelos entre sí.