UNIVERSIDAD DE JAÉN

Material del curso "Análisis de datos procedentes de investigaciones mediante programas informáticos"

Manuel Miguel Ramos Álvarez

MATERIAL VIII "MODELO DE ANÁLISIS MULTIVARIADO"

Índice

8. Extensión del Modelo de Análisis a investigaciones Multivariadas 2
8.1. Nuevas variantes del Modelo general en función de la clasificación de las técnicas de análisis aplicables al contexto Multivariado 2
8.1.1. A. Técnicas Explicativas-1 3
8.1.2. B. Técnicas Descriptivas-1 5
8.1.3. Conclusiones 9
8.1.4. Ejercicios 10
8.2. Bases conceptuales del análisis Multivariado 11
8.2.1. Expresión de Regresión lineal en términos matriciales 11
8.2.2. Concepto de Variado 15

8. Extensión del Modelo de Análisis a investigaciones Multivariadas

8.1. Nuevas variantes del Modelo general en función de la clasificación de las técnicas de análisis aplicables al contexto Multivariado

Introducción

Criterio fundamental es el Objetivo de la técnica: explicativas vs descriptivas.

- Las explicativas se emplean en contextos de investigación experimentales. Para contrastar hipótesis relativas a la influencia que nuestras variables independientes sobre las dependientes.
  - Para establecer explicaciones causales, siempre en el contexto experimental.
  - Para establecer meras explicaciones, en el contexto correlacional-covariacional.
- Las descriptivas se usan en contextos no experimentales. Para resumir (reducir) el conjunto de datos de una forma adecuada a nuestros objetivos.
  - Para abstraer los diferentes patrones de respuestas, reduciendo los datos a un número menor de dimensiones.
  - Para obtener una información sobre la estructura interna (las supervariables) de nuestras medidas. Por ejemplo detectar o poner a prueba una estructura.

Diferenciación:

- Multivariadas (en sentido estricto). Analizan simultáneamente más de una variable medida o dependiente,
- Multivariables. Consideran simultáneamente un conjunto de variables predictoras o independientes (como la regresión múltiple).

- La mayoría de las técnicas tiene una variante específica para fines muy concretos y variantes no paramétricas o robustas frente al incumplimiento de los supuestos.

8.1.1. A. Técnicas Explicativas-1

A.1. Análisis de regresión múltiple ("Multiple Regression")

Objetivo / Definición

- Se mide una sola variable dependiente o criterio, y varias independientes o predictores, que no han sido manipulados, sino que, en general, son características que los individuos poseen en cierto grado.
- Cuando lo que se pretende es predecir un criterio desde el conjunto de predictores, ambos métricos

Ejemplo

- Predecir el rendimiento académico a partir de las puntuaciones de los sujetos en un test de inteligencia como el WAIS.

Variante

Análisis de regresión canónica ("Canonical Análisis")

- Permite analizar la relación entre dos conjuntos de variables: es decir donde hay más de un predictor y simultáneamente más de un criterio.

Ejemplo

- Un investigador educativo que está interesado en la relación entre tres medidas de habilidad con cinco medidas de rendimiento.
- Un investigador médico que desea analizar la relación entre varios factores de riesgo y el desarrollo de un grupo de síntomas.

Análisis de regresión canónica no lineal ("OVERALS") mediante Escalamiento Óptimo. Para categóricos y No-Lineal.

A.2. Análisis multivariado de varianza y covarianza (MANOVA y MANCOVA)

Objetivo / Definición

- Determinar si una o más variables independientes –usualmente no métricas- producen efectos diferenciales sobre dos o más variables dependientes –métricas-
- También cuando se ha medido la misma variable dependiente de forma reiterada en el tiempo a los mismos sujetos, esto es, en diseños de medidas repetidas.

MANCOVA

- Como anterior pero se realiza un control estadístico de variables (covariados) que estén relacionadas con las dependientes antes de que se apliquen las independientes a los sujetos.
- MANOVA es una alternativa frente al ANOVA cuando las diferentes medidas tienen mucho en común, ya sea desde el punto de vista conceptual ya desde el estadístico.

Ejemplo

- Una investigación psicofisiológica en la que se toman diferentes medidas relacionadas con la actividad cardiaca, como la tasa, la presión sanguínea, etc. En un caso como este las variables en realidad miden en última instancia un proceso o actividad común y por lo tanto podría interesarnos plantear el tipo de preguntas propias de la investigación experimental pero tendiendo en cuenta la parte común de todas las variables dependientes que se han medido, a la vez.
- MANCOVA: Además se controlan las diferencias sexuales medidas previamente.
- Se emplean dos materiales didácticos muy diferentes, uno basado en el método tradicional de enseñanza y otro según el crédito europeo, y estamos interesados en las mejoras de rendimiento tanto en física como en matemáticas. Nuestra hipótesis es que ambas medidas conjuntamente se verán afectadas por las diferencias didácticas.

Variante

Componentes de Varianza y Modelo Mixto ("Variante Components and Mixed Model").

- Es una extensión del ANOVA/MANOVA expresamente para analizar diseños especiales: que incluyen efectos aleatorios, con muchos niveles (e.g. con 200 o más niveles) y en general siempre que interese una descomposición ponderada de la variabilidad.

Componentes Principales categórico (CATPCA) mediante Escalamiento Óptimo. Para categóricos y No-Lineal.

A) Técnicas Explicativas-2

A.3. Análisis discriminante ("Discriminant function analysis")

Objetivo / Definición

- Ahora el criterio es dicotómico (puede adoptar únicamente dos valores) o politómico (puede adoptar más de dos valores), pero no métrico. Las predictoras sí son métricas.
- El objetivo es determinar cuáles son las variables predictoras que permiten discriminar entre los grupos (realizados en el criterio). Además, determinar cuáles de las variables predictoras son realmente útiles.

Ejemplo

- Si definimos el rendimiento académico en función de si aprueba o no el curso completo: apto y no apto.
- Los profesionales de los servicios de urgencias podrían estar interesados en discernir sobre aquéllas variables que permiten discriminar a personas con urgencia real de las que no la requieren.
- Para un biólogo podría ser de interés discriminar el conjunto de características que mejor discrimina entre grupos de flores similares.

A.4. Análisis de regresión logística (subtipo del categórico) ("logistic Regression") y de regresión categórica.

Objetivo / Definición

- Comparable al discriminante pero cuando las variables predictoras no son métricas: predecir el grupo al que pertenece un individuo.
- El procedimiento también es similar, se trata de mezclar las variables predictoras con la ponderación adecuada, sin embargo, la mezcla no es lineal como en el análisis discriminante sino que se basa en los modelos del tipo categórico.
- Otra posible aplicación es la predicción, como en regresión lineal, de un criterio a partir de diversos predictores, pero donde el criterio no es una variable métrica.

Ejemplo

- Supongamos que un investigador clínico pretende estudiar la posibilidad de que se sufra una enfermedad cardiaca o no en función de predictores como el consumo de cigarrillos y el grado de estrés que experimentan las personas.

Variantes:

- Otros tipos de regresión con datos categóricos son el tipo **Logit** y **Probit**, ver más abajo en categóricas (B.1.)

A.5. Ecuaciones lineales estructurales ("Structural Equation Modeling")

Objetivo / Definición

- Relaciones de predicción secuenciales en las que se pone a prueba una cadena causal de acontecimientos. También llamado análisis causal (causal modeling, or path analysis).

Ejemplo

- Predecir la capacidad de memoria operativa a partir de un conjunto de tareas que miden cosas como la capacidad para coordinar la realización de tareas simultáneas, de alternar entre tareas o de actualizar el contenido de la información en memoria a medida que va apareciendo información nueva.
- A su vez, podemos intentar predecir la inteligencia general, medida, por ejemplo, mediante las Matrices Progresivas de Raven, a partir de la capacidad de memoria operativa.
- Una variable dependiente se convierte en un predictor en el paso siguiente.

8.1.2. B. Técnicas Descriptivas-1

B.1. Análisis de datos categóricos (Log-Linear Analysis of Frequency Tables)

Objetivo / Definición

- Cuando la investigación que se ha efectuado incluye variables categóricas, estrictamente ya no tendremos medidas de una variable que explicar a partir de otras sino más bien frecuencias conjuntas de aparición de los valores de sendas variables.

Ejemplo

- En el contexto médico, supongamos que deseamos describir una muestra amplia en función de su conducta de fumados (Sí vs. No) y del desarrollo de cáncer de pulmón (Sí vs. No).
- Determinar si hay relación entre el tipo de psicopatología que se sufre (esquizofrenia, depresión, etc.) y el área geográfica donde se vive (norte, centro, sur).

Variantes

- La forma más común: análisis de frecuencias de múltiples vías ("Multiway Tables"), cuya extensión es más conocida como análisis log-lineal, pero hay otras variantes como el logit o probit, dependiendo del tipo de función, que son útiles más bien en un contexto explicativo, al igual que regresión logística que ya vimos.
- **Logit-Markov** para analizar datos categóricos dentro de la lógica de un diseño longitudinal con medidas a través del tiempo,
- **Logit-Latente**. Dentro de la lógica de un de Análisis Factorial para reducir las dimensiones.

FUNCIONES DE ENLACE	MODELOS DISTRIBUCIÓN
Log: $\hat{Y}_i = \log(\mu_i)$; $\hat{Y}_i = \beta \log(\mu_i)$	Binomial (binaria ó dicotómica)
Logit: $\hat{Y}_i = \log\left(\frac{\mu_i}{N-\mu_i}\right)$; $\hat{Y}_i = \log\left(\frac{\mu_i}{\mu_j}\right)$	Multinomial (policotómica)
Probit: $\hat{Y}_i = \frac{1}{Z(\mu_i)}$	Poisson (contero de eventos raros)

B) Técnicas Descriptivas-2

B.2. Análisis factorial (Factor Analysis) y de componentes principales (Principal Components).

Objetivo / Definición

- Agrupar a las variables medidas en función de las correlaciones que muestran, de cara a abstraer algunas dimensiones o factores subyacentes. Reducir los datos a partir de las correlaciones para encontrar un número reducido de factores que los expliquen.

Ejemplo

- Diversas tareas para medir la memoria a corto plazo: bloques de Corsi, amplitud de dígitos hacia delante, hacia atrás, amplitud lectora. Corsi y dígitos hacia delante correlacionan entre sí, pero poco con dígitos hacia atrás y amplitud lectora, que a su vez correlacionarán entre sí. Agrupando las variables según la correlación podemos obtener factores, que serán indicadores del proceso que ha dado lugar a la correlación. En nuestro ejemplo los dos factores: memoria a corto plazo (los dígitos hacia delante y Corsi) y ejecutivo central (dígitos hacia atrás y amplitud lectora).
- Muy usual cuando se mide algún rasgo mediante una prueba o test compuesto por muchos ítems y deseamos abstraer las agrupaciones o factores.

Variantes

Componentes principales

- La característica que diferencia al AF del ACP es que en el segundo se asume que se debería de emplear en los cálculos toda la variabilidad contenida en un ítem, mientras que en el AF únicamente se retiene la variabilidad que el ítem tiene en común con los otros.
- Usualmente llevan a resultados similares.
- El primero para detectar la estructura de los datos (como un método de Clasificación) y el segundo para lograr la reducción de los datos.

B.3. Análisis de Correspondencias ("Correspondence Analysis")

Objetivo / Definición

- Su objetivo es también estudiar las dependencias entre las variables y también se desarrolla mediante el cómputo de tablas de contingencia.
- Sin embargo, realiza una transformación de los datos no métricos para convertirlos en datos métricos y obtiene las dimensiones subyacentes a los mismos, así como los mapas que representan la cercanía de los valores de las variables medidas, proporcionando información similar a la del Análisis Factorial.

Ejemplo

- Hábitos de fumadores de diferentes tipos de empleados de una compañía.

	Smoking Category				
Staff Group	**(1) None**	**(2) Light**	**(3) Medium**	**(4) Heavy**	**Row Totals**
(1) Senior Managers	4	2	3	2	11
(2) Junior Managers	4	3	7	4	18
(3) Senior Employees	25	10	12	4	51
(4) Junior Employees	18	24	33	13	88
(5) Secretaries	10	6	7	2	25
Column Totals	61	45	62	25	193

- El objetivo es simplificar los datos, de manera que se puedan abstraer algunas dimensiones básicas para representarlos.

Variantes No-Paramétricas o Robustas:

- **Análisis de Correspondencias múltiple** o de Homogeneidad (HOMALS) mediante Escalamiento Óptimo. Es una variante categórica para No Linealidad.

B) Técnicas Descriptivas-3

B.4. Análisis multidimensional ("Multidimensional Scaling")

Objetivo / Definición

- Es una alternativa al Análisis Factorial con la meta de detectar dimensiones subyacentes significativas que permitan explicar las semejanzas (o diferencias) entre los objetos (i.e. sujetos) investigados.
- En el Análisis Factorial, las semejanzas entre objetos (e.g.., variables) se expresan en términos de matrices de correlaciones, mientras que en Multidimensional se puede analizar cualquier tipo de matriz más allá de la de correlaciones.

Ejemplo

- Pedimos a un grupo de sujetos que nos indiquen la semejanza que hay entre varias parejas de conceptos (pongamos psicoanálisis-conductismo, psicoanálisis-yo, etc.). Para analizar el conjunto de semejanzas podríamos convertirlas en distancias en un espacio de múltiples dimensiones. El resultado será un mapa de dos o más dimensiones en el que los conceptos semejantes (psicoanálisis-yo) aparecerán cercanos entre sí y distantes respecto de otros no semejantes (conductismo).
- Imaginemos por un lado las tres ciudades A, B y C, y por otro lado D, E y F; mostrando la distancia en Km. que hay entre las mismas:

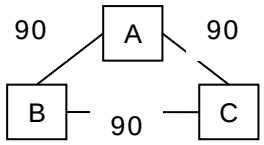
	A	B	C
A	0		
B	90	0	
C	90	90	0

	D	E	F
D	0		
E	90	0	
F	180	90	0

- Todas se alejan en 90 Km. salvo las ciudades D y E. La pregunta es ¿Cómo podemos organizar las 3 ciudades (objetos) en una única dimensión?
- Por ejemplo, en una línea recta tendríamos:

D---90 ---E---90 ---F

- Pero esto no es válido para las otras 3 ciudades. Sin embargo, podríamos organizarlos según dos dimensiones siguiendo una estructura triangular. Dimensión 1 para A-B o A-C y dimensión 2 para B-C



B) Técnicas Descriptivas-4

B.5. Análisis de cluster ("Cluster Analysis") o de conglomerados

Objetivo / Definición

- Para agrupar a individuos (o variables) en diferentes categorías (clusters) en función de su semejanza: clasificar individuos (o variables) semejantes entre sí en las variables que les hemos medido, pero no tenemos un criterio de clasificación a priori. Luego todas las variables tienen el mismo estatus.
- Cuando queremos saber si en nuestra muestra de sujetos podemos diferenciar grupos de individuos diferentes.
- Para organizar datos en estructuras significativas, es decir el desarrollo de taxonomías.
- El que se basa en las medias ("K-Means Cluster Analysis") es preferible para gran cantidad de datos y el del tipo jerárquico ("Hierarchical Cluster Analysis") para pocos datos.

Ejemplo

- En un supermercado, ítems similares, como carnes y verduras, se colocan en localizaciones próximas; mientras que los productos de droguería o de ropa están en lugares alejados de los primeros.
- Para los biólogos será de interés organizar a diferentes especies de animales: los hombres pertenecen a los primates, los mamíferos, "amniotes", vertebrados y animales a su vez.
- En Psicología sobre las taxonomías de categorías o conceptos como la de E. Rosch.

Variantes

- Joining (Tree Clustering),
- Two-way Joining (Block Clustering),
- k-Means Clustering
- Boosting Trees,
- CHAID -Chi-squared Automatic Interaction Detector-
- Classification and Regression Trees (C&RT).

Todas ellas en función del algoritmo de agrupación.

B.6. Análisis de supervivencia ("Survival / Failure Time Analysis")

Objetivo / Definición

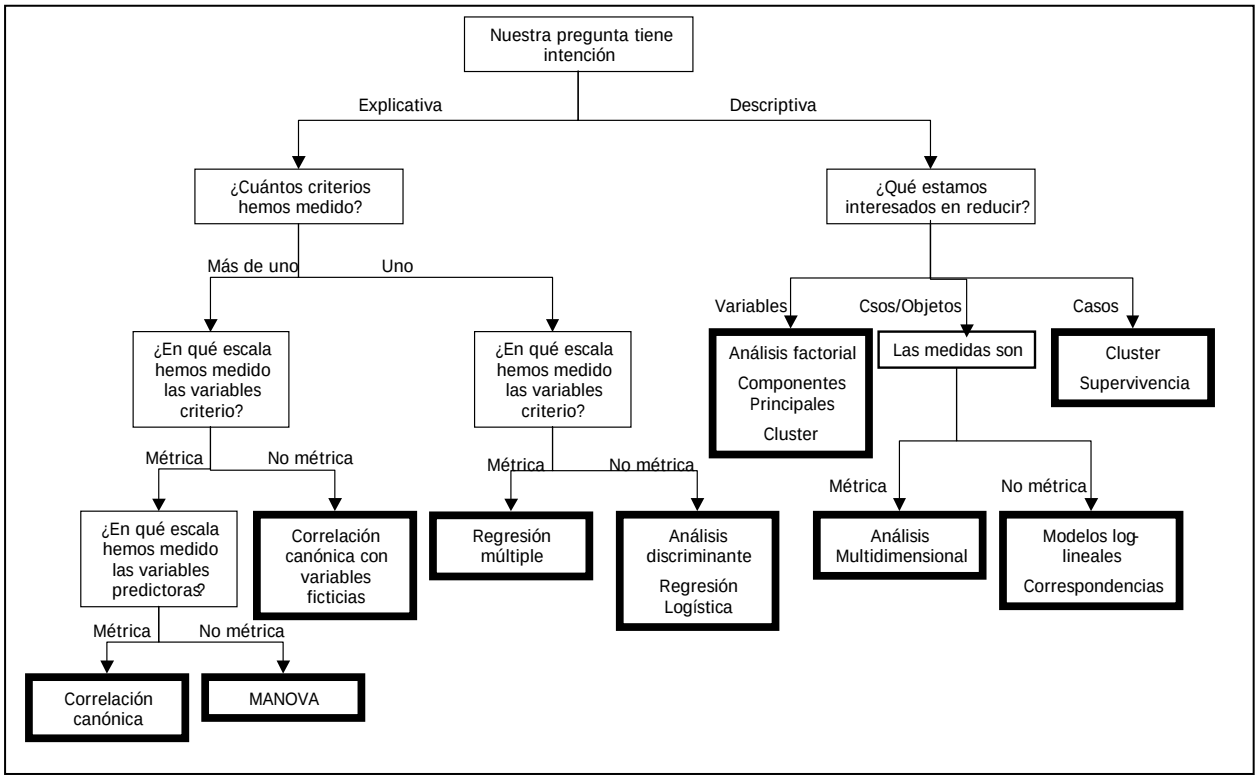
- Determinar cuánto tiempo es necesario que transcurra para que suceda algo (muerte, cura, logro de empleo, etc.).
- Describir la proporción de casos que en diferentes momentos del tiempo permanecen en el estudio.
- En determinados contextos investigación: medicina, biología o economía; como por ejemplo para estudiar envejecimiento, rehabilitación, empleo, etc.

Ejemplo

- ¿Cuántos pacientes siguen en rehabilitación después de transcurrido cierto tiempo? A menudo se emplea más de un grupo, por ejemplo, uno en rehabilitación y otro sin ella. El análisis presenta las curvas de supervivencia para cada grupo, pero también es posible realizar pruebas estadísticas para comparar los grupos.
- Investigador de un hospital que estudia la efectividad de un nuevo tratamiento para una enfermedad terminal. La variable de mayor interés en este contexto es el número de días que los pacientes logran sobrevivir.

8.1.3. Conclusiones

La elección de la técnica de análisis (**Cuadro 1.1, tomado de** Catena, A.; Ramos, M.M. y Trujillo, H. (2003). *Análisis Multivariado. Una manual para investigadores.*Madrid: Biblioteca Nueva).



8.1.4. Ejercicios.

1. Intente ubicar las diferentes variantes analíticas en los diferentes menús que define el programa SPSS o Statistica.
2. Decida cuál es la técnica más adecuada para analizar los datos de las diferentes investigaciones proporcionadas por los compañeros.

Volver Principio

8.2. Bases conceptuales del análisis Multivariado

8.2.1. Expresión de Regresión lineal en términos matriciales

- Frecuentemente, lo que se denomina Modelo Lineal General corresponde a todo el planteamiento analítico basado en el Modelo Lineal pero expresado en términos matriciales, lo que le confiere una gran potencia de desarrollo computacional.

Ecuación Lineal:

$$Y_i = \beta_0 + \beta_1 X1_i + \beta_2 X2_i + \ldots + \beta_{p-1} XP-1_i + \beta_p XP_i + \varepsilon_i$$

Ecuación Lineal Generalizada:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_i \end{bmatrix} = \begin{bmatrix} 1 & X1_1 & X2_1 & \ldots & XP_1 \\ 1 & X1_2 & X2_2 & \ldots & XP_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X1_i & X2_i & \ldots & XP_i \end{bmatrix} \bullet \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \end{bmatrix}$$

$$Y_{Nx1} = X_{Nxh} \bullet \beta_{hx1} + \varepsilon_{Nx1}$$

- Es decir, por un lado, las puntuaciones en la variable dependiente,
- por otro lado las puntuaciones en el predictor,
- en tercer lugar los parámetros que habrá que estimar mediante Mínimos Cuadrados y
- finalmente una matriz para los errores o residuales.

Ampliación: Álgebra matricial y cálculo de determinantes

A) ÁLGEBRA MATRICIAL

Se llama **matriz** de orden "p x q" a un conjunto de elementos organizados según una determinada estructura: en forma de tabla bidimensional con "p" filas y "q" columnas, que contiene por lo tanto un total de "pxq" elementos

$$A=\begin{bmatrix}1 & 5 & 6\\ -1 & 0 & 5\end{bmatrix}$$ De orden 2x3

El elemento a_{12}=5 (primera fila y la segunda columna)

Tipos de matrices:

- Un **escalar:** una matriz que contiene un elemento único $k=4$.
- Una matriz **columna** o vector columna: de orden "p x 1".
- Una matriz **fila** o vector fila: de orden "1 x q".
- La matriz **nula** es aquella cuyos elementos son todos nulos.
- Una matriz **cuadrada:** igual número de filas que de columnas: "p = q". Partes:
  - Diagonal principal: desde el vértice superior-izquierdo al inferior-derecho
  - Diagonal segundaria en el sentido opuesto.
  - Las diagonales delimitan un triángulo superior e inferior.
  - La **traza:** suma de los elementos en la diagonal principal.
- Matriz **simétrica**. Matriz cuadrada en la que se cumple: $a_{ij}=a_{ji}$.
- Matriz **diagonal**. Matriz cuadrada y simétrica y en la que los elementos fuera de la diagonal principal son todos nulos.
- **Matriz escalar**. Matriz diagonal con elementos diagonales todos idénticos entre sí.
- **Identidad** (I) Matriz escalar con los elementos de la diagonal todos iguales a la unidad.

Operaciones con matrices.

Igualdad. Siempre que tengan el mismo orden y sus elementos coincidan uno a uno.

Suma. $C=A+B=\begin{bmatrix}3 & 4 & 2\\ 1 & 0 & 5\end{bmatrix}+\begin{bmatrix}2 & 1 & 0\\ -1 & 0 & 10\end{bmatrix}=\begin{bmatrix}3+2 & 4+1 & 2+0\\ 1-1 & 0+0 & 5+10\end{bmatrix}=\begin{bmatrix}5 & 5 & 2\\ 0 & 0 & 15\end{bmatrix}$

Producto con un escalar: $C=3A=3\begin{bmatrix}3 & 4 & 2\\ 1 & 0 & 5\end{bmatrix}=\begin{bmatrix}3(3) & 3(4) & 3(2)\\ 3(1) & 3(0) & 3(5)\end{bmatrix}=\begin{bmatrix}9 & 12 & 6\\ 3 & 0 & 15\end{bmatrix}$

Producto de dos matrices.

$$\underbrace{\underset{px\underbrace{q}_{Iguales}}{A}\quad\underset{q\,xr}{B}}_{Orden\ de\ C:\,pxr}$$

$$C=A\times B=\begin{bmatrix}3 & 4 & 2\\ 1 & 0 & 5\end{bmatrix}\times\begin{bmatrix}2 & 4\\ 0 & 1\\ -3 & -2\end{bmatrix}=\begin{bmatrix}\begin{pmatrix}3 & 4 & 2\end{pmatrix}\begin{pmatrix}2\\0\\-3\end{pmatrix} & \begin{pmatrix}3 & 4 & 2\end{pmatrix}\begin{pmatrix}4\\1\\-2\end{pmatrix}\\ \begin{pmatrix}1 & 0 & 5\end{pmatrix}\begin{pmatrix}2\\0\\-3\end{pmatrix} & \begin{pmatrix}1 & 0 & 5\end{pmatrix}\begin{pmatrix}4\\1\\-2\end{pmatrix}\end{bmatrix}=$$

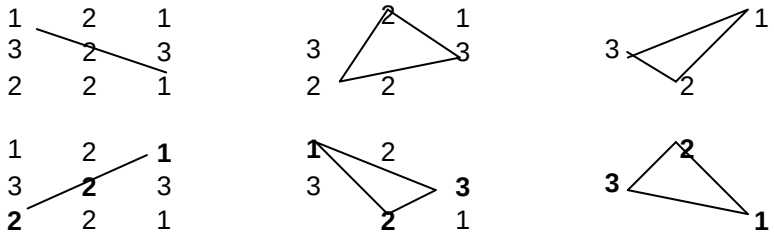
$$C=\begin{bmatrix}3(2)+4(0)+2(-3) & 3(4)+4(1)+2(-2)\\ 1(2)+0(0)+5(-3) & 1(4)+0(1)+5(-2)\end{bmatrix}=\begin{bmatrix}0 & 12\\ -13 & -6\end{bmatrix}$$

Función **= *MMULT()***.

B) DETERMINANTE DE UNA MATRIZ

- Es un valor escalar que condensa la información algebraica de la matriz, tomando todas las permutaciones de los elementos que la forman de manera que varíe simultáneamente la información en las dos dimensiones (filas y columnas) de la misma.

1	2	1
3	2	3
2	2	1



$$|A| = [(1*2*1)+(2*2*3)+(3*2*1)] - [(2*2*1)+(1*2*3)+(3*2*1)] = 20-16 = 4$$

Función para calcular el determinante ***=MDETERM()***.

C) MATRIZ ADJUNTA E INVERSIÓN DE MATRICES

- Es la matriz construida a partir de los correspondientes adjuntos (cofactores) de los elementos de la matriz original.
- Por ejemplo:

1	2	1
3	2	3
2	2	1

- Para ele elemento a_{11}=1 en el vértice superior derecho, el menor complementario vendrá dado por el determinante:

$$\begin{vmatrix} 2 & 3 \\ 2 & 1 \end{vmatrix} = -4$$

- La inversa de una matriz dada se define: $A^{-1} = \frac{\|A\|'}{|A|}$
- Es decir, la traspuesta de la matriz adjunta dividida por el determinante de la matriz.
- Función para calcular la matriz inversa es ***=MINVERSA()***.

1	2	1		-1	0	1
3	2	3		0,75	-0,25	0
2	2	1		0,5	0,5	-1

D) EL CONCEPTO DE AUTOVECTOR Y AUTOVALOR

- Para una matriz cuadrada A_n, un escalar λ y un vector fila x_{nx1} y dada la relación ó ecuación:

$$Ax = \lambda x$$

- Si existe un vector **x** diferente de cero que verifique dicha ecuación; éste es un **autovector** de A y λ es un **autovalor** de A.
  - Autovalor o raíz característica, valor propio, valor latente o eigenvalor
  - Autovectores o vectores característicos, vectores propios, vectores latentes o eigenvectores.
- La ecuación es equivalente a:

$$(A - \lambda I)x = 0$$

- Que admite una solución no nula si y solo si se puede resolver el determinante del término a la izquierda, es decir sólo cuando $|A - \lambda I| = 0$. Lo que equivale a resolver:

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix} = 0$$

- Una ecuación de grado n en lambda y se denomina **ecuación característica**, que admite "n" raíces (no necesariamente todas distintas) cada una de las cuales será autovalor de A.
- Si λ_i es uno de los autovalores, entonces sustituyendo dicho valor en la ecuación de partida $(A - \lambda I)x = 0$, obtendremos el autovector x_i correspondiente.

Ejemplo: Para $A = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix}$

$$\begin{vmatrix} 2-\lambda & 1 \\ 3 & 4-\lambda \end{vmatrix} = (2-\lambda)\bullet(4-\lambda) - 3\bullet 1 =$$

$$8 - 2\lambda - 4\lambda + \lambda^2 - 3 = \lambda^2 - 6\lambda + 5 =$$

$$(\lambda - 5)(\lambda - 1) = 0$$

De donde obtendríamos los autovalores: $\lambda_1 = 5$ y $\lambda_2 = 1$

A continuación los sustituimos en la ecuación para obtener los autovectores.

- Para la primera solución, $\lambda_1 = 5$

$$\begin{bmatrix} 2-5 & 1 \\ 3 & 4-5 \end{bmatrix}\bullet\begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = \begin{bmatrix} -3 & 1 \\ 3 & -1 \end{bmatrix}\bullet\begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = 0$$

y Para el autovalor $\lambda_1 = 5$ el autovector $x_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

$$\left\{\begin{matrix} -3x_{11} + x_{21} = 0 \\ 3x_{11} - x_{21} = 0 \end{matrix}\right\} \rightarrow x_{21} = 3x_{11}$$

- Para la segunda solución, $\lambda_2 = 1$

$$\begin{bmatrix} 2-1 & 1 \\ 3 & 4-1 \end{bmatrix}\bullet\begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix}\bullet\begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = 0$$

y Para el autovalor $\lambda_2 = 1$ el autovector $x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

$$\left\{\begin{matrix} x_{11} + x_{21} = 0 \\ 3x_{11} + 3x_{21} = 0 \end{matrix}\right\} \rightarrow x_{21} = -x_{11}$$

8.2.2. Concepto de Variado

- Suele haber una mezcla o combinación de variables, casi siempre lineal.

- Este variado o supervariable no se ha medido directamente, pero informa sobre el significado subyacente del conjunto de variables.

$$\hat{Y}_i = \gamma_1 Y1_i + \gamma_2 Y2_i + ... + \gamma_c YC_i$$

- La magnitud del peso nos permite decidir sobre la pertinencia de las variables. Algunas puede que no aportan nada significativo (peso entorno a cero) y sería recomendable suprimirlas.

A) MATRIZ DE VARIANZAS-COVARIANZAS Y SUPUESTO DE SIMETRÍA

	A	B	C	D	E	F	G	H	I
1		**Matriz Datos del Diseño**							
2			a_1	a_2	a_3				
3		1	1	4	4				
4		2	6	4	2				
5		3	3	6	7				
6		4	1	8	8				
7		5	6	9	7				
8		6	1	2	5				
9		Var	6	7,1	5,1				
10									
11		Covar1-2	1,833						
12									
13		Covar1-3		-1,167					
14									
15		Covar2-3		3,583					
16									
17		**Estructura Matriz Varianzas-Covarianzas**							
18			a_1	a_2	a_3				
19		a_1	6	1,833	-1,167	Triángulos: Covarianzas			
20		a_2	1,833	7,1	3,583				
21		a_3	-1,167	3,583	5,1				
22						Diagonal: Varianzas			
23									
24		**Implicaciones supuesto Simetria**							
25			Var	Var	Cov				
26		1-2	6	7,1	1,833	**9,433**			
27		1-3	6	5,1	-1,167	**13,43**	Varianzas Diferencias		
28		2-3	7,1	5,1	3,583	**5,033**			

B) LÓGICA DE ESTIMACIÓN DEL VARIADO

Acceso al fichero de Excel