

Clase 12

Iniciación al análisis multivariante



Le Sphinx

CAMPUS

El pack Sphinx dedicado a la enseñanza de métodos y técnicas para encuestas

Objetivos

Objetivos metodológicos

- Identificar situaciones en las que los análisis multivariantes permiten ir más allá en la explotación de los resultados
- Aprender los principios básicos del análisis multivariante
- Poder interpretar correctamente los resultados de algunos análisis multivariantes comunes como la clasificación o el análisis de regresión múltiple

Objetivos Sphinx

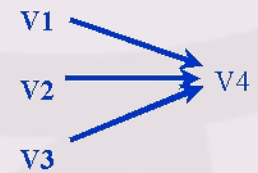
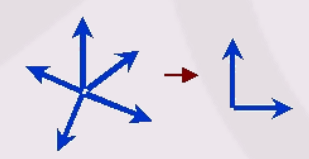
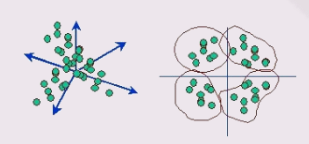
- Saber activar las funciones de análisis multivariante en Sphinx e interpretar los resultados principales

Tabla de contenidos

1. Objetivos e intereses de los análisis multivariantes
2. Análisis de regresión múltiple
3. Análisis factorial de correspondencias
4. Clasificación automática / tipología

1. Objetivos e intereses de los análisis multivariantes

- Los análisis multivariantes cumplen en general tres objetivos distintos, que se presentan en el cuadro que figura a continuación. Aquí daremos una visión general de tres técnicas: una para cada categoría principal.

Objetivo	Principios	Técnicas multivariantes	Ilustraciones
Explicar	Al igual que en el caso de los análisis bivariantes o del análisis cruzado, el objetivo es comprender cómo una variable puede ser influenciada por una serie de otras variables: cuáles tienen más impacto, cuáles no tienen influencia, etc.	- Regresión múltiple - Matriz de importancia y valoración - Árboles de decisión	
Sintetizar	El objetivo es simplificar los datos identificando variables o modalidades que sean similares, que midan lo mismo. Estos grupos de variables "correlacionadas" pueden ser identificados para que los resultados sean más sintéticos y, por lo tanto, más sencillos.	- Análisis factorial en componentes principales - Análisis factorial de correspondencias	
Clasificar	Los métodos de clasificación no se centran en las variables sino en los individuos. Una vez más, con miras a la simplificación, el objetivo es identificar grupos de encuestados con comportamientos similares.	- Análisis de tipologías o clasificación	

2. Análisis de regresión múltiple

- Esta es una extensión del análisis de regresión "simple" descrito en la clase 10. En lugar de tratar sólo dos variables, se tratarán 3, 4 o más. Las variables elegibles deben ser numéricas (numérica abierta o escala).
- El objetivo es comprender cómo se comporta una variable **y** explicada en función de varias variables potencialmente explicativas **x1**, **x2** y **x3**, etc.
- Los resultados del análisis de regresión múltiple indican principalmente:
 - Qué variables candidatas están significativamente correlacionadas con la variable explicada y cuáles no.
 - El coeficiente de correlación múltiple del modelo, que es la calidad general del modelo (cercano a 1) y su capacidad para predecir y en función de x.
 - La constante y especialmente el valor de los coeficientes de regresión para cada variable explicativa seleccionada.
 - De esto podemos deducir el peso relativo de cada variable explicativa en su influencia sobre la variable explicada.

2. Análisis de regresión múltiple

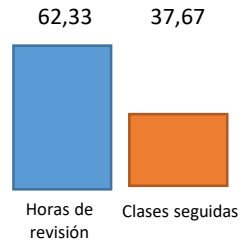
Variables sin influencia

1 variable no contribuye a la explicación (umbral de 5%): **Nota del Grado de Matemáticas**

Variables influyentes

2 variables contribuyen a la explicación (umbral de 5%): **Clases seguidas, horas de revisión**

Influencias de las variables explicativas



Peso relativo de cada variable explicativa

Las variables candidatas correlacionadas significativamente (o no) con la variable explicada

Ecuación del modelo

Nota en estadística = $2,78 + 0,23 + \text{Clases seguidas} + 0,34 * \text{Horas de revisión}$

Ecuación de la línea de regresión múltiple con coeficientes y constante

Influencia de las variables explicativas

	Coefficiente	P-value	Coefficiente estandarizado	Contribución
Const	2,78	0,01	-	-
Clases seguidas	0,23	<0,01	0,39	37,67
Horas de revisión	0,34	<0,01	0,64	62,33

En este ejemplo, podemos predecir muy bien y (nota en estadística) en base a x1 (horas de clases seguidas) y x2 (horas de revisión). Pero no según la nota del grado de matemáticas. El impacto de la revisión es más fuerte que el impacto de la asistencia actual.

Indicador de la calidad del modelo

El modelo representa el **85,53%** de la varianza de la variable a explicar

Coefficiente de correlación múltiple: $R=0,92$

P-value de R: $p(R)= <0,01$

Coefficiente de Fisher: $F=50,24$

P-value de F: $p(F)= <0,01$

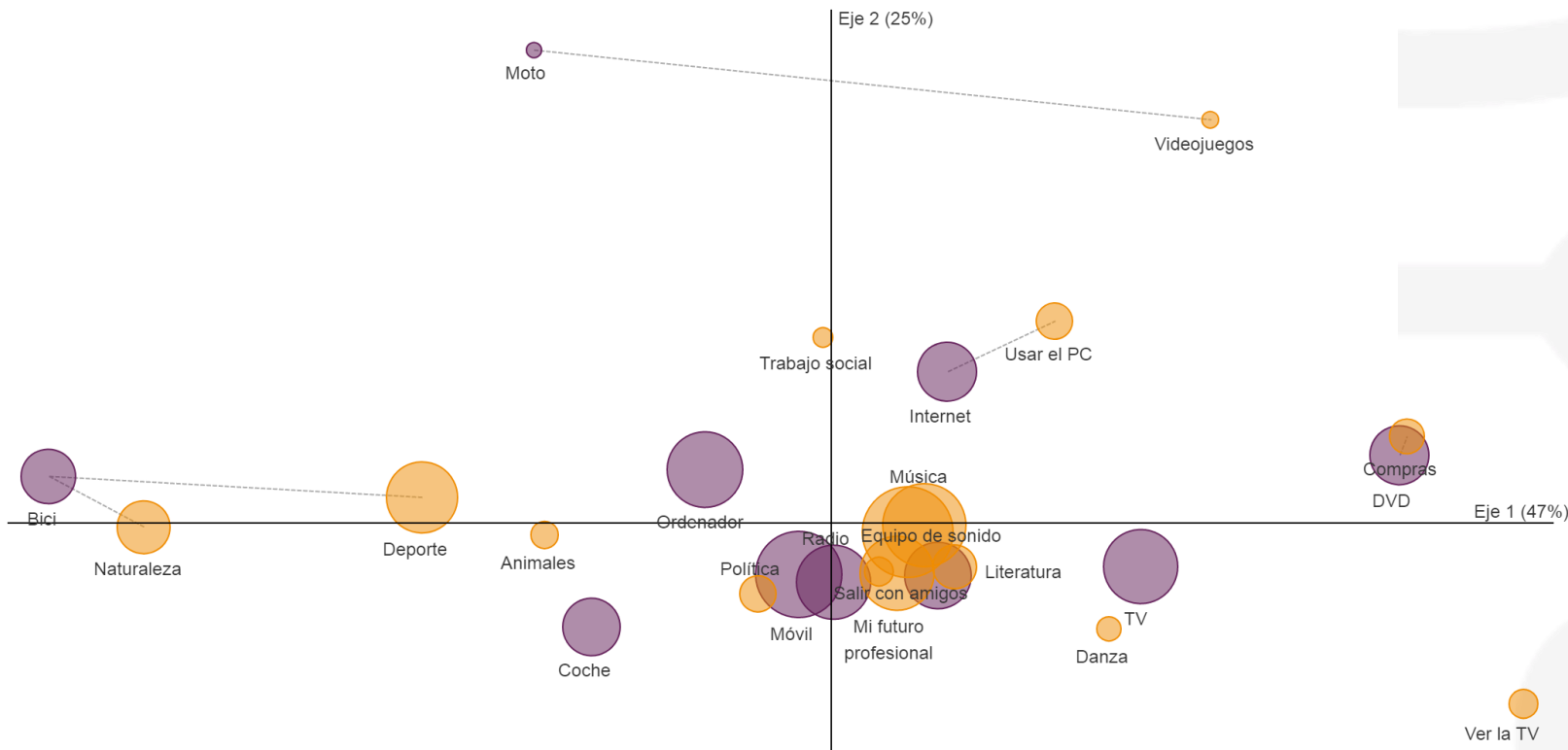
Coefficiente de correlación múltiple del modelo = calidad del modelo

3. Análisis factorial de correspondencias

- El AFC se refiere a las respuestas a preguntas cerradas (o las llamadas variables nominales). Se interesa principalmente por la **correspondencia** entre las **modalidades** de las diferentes variables estudiadas.
- Propone combinaciones de modalidades frecuentemente asociadas (los "**factores**") que pueden interpretarse como amplias dimensiones estructurantes del tema objeto de estudio.
- La principal restitución de un AFC es un mapa en el que se pueden visualizar las **asociaciones** (y las oposiciones) entre modalidades según su **proximidad** (y su distancia) en el mapa.
- Los **ejes** horizontales y verticales -dependiendo de las modalidades que se posicionen en ellos- sirven para interpretar las principales dimensiones estructurantes del sujeto estudiado.
- El AFC puede realizarse a partir del análisis de una pregunta cerrada múltiple, un cruce o un análisis multivariante (en el que se indican las variables que deben considerarse).

Ejemplo

Cruce: ¿Qué mobiliario es tuyo? / ¿En cuál de los siguientes temas está interesado? Puede elegir hasta 5 de ellos.



Modalidades en el centro = compartido por todos
La música es una actividad compartida por todos los encuestados en esta encuesta.

Eje horizontal = dimensión principal que estructura las respuestas
El materialismo a la derecha contra la naturaleza a la izquierda

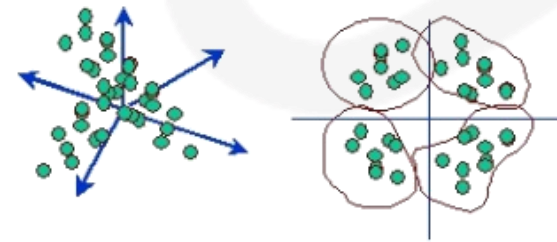
Modalidades cercanas = frecuentemente asociadas
Los que aman la naturaleza frecuentemente tienen una bicicleta

Modalidades lejanas = raramente asociadas
La gente a la que le gusta ver la televisión rara vez tiene una bicicleta

El mapa de correspondencias proyecta **71%** de la información, dividida en **47%** horizontalmente (F1) y **25%** verticalmente (F2). La proximidad o la distancia entre elementos visualiza las asociaciones sobre o infrarrepresentadas.

4. Clasificación automática/tipología

- Identificar grupos de personas con respuestas similares en un conjunto definido de variables. Con el fin de segmentar en marketing, por ejemplo.
- Existen muchos métodos, como las clasificaciones jerárquicas (ascendentes o descendentes) o el método de los centros móviles, que detallaremos a continuación.
- Método = proceso iterativo iniciado al azar que produce sucesivamente varias soluciones manteniendo siempre la mejor (según criterios de diferenciación y homogeneidad de los grupos).
- El programa (utilizando el método de k means) indica entonces las clases que ha identificado, su número y sus valores en las variables utilizadas, lo cual permite su interpretación.
- La clase a la que pertenece cada individuo se puede entonces guardar y cruzar con otras variables de la encuesta.



Ejemplo



La clasificación agrupa los individuos en clases homogéneas desde el punto de vista de las variables consideradas. El cálculo se inicializa de forma aleatoria. El algoritmo reasigna los individuos para mejorar progresivamente la homogeneidad de las clases.

Encuesta: Jóvenes - Los jóvenes y el consumo. La clasificación kmeans realizada en las variables que miden la importancia de los factores de influencia para la compra.

- precio, calidad, moda, respeto al medio ambiente y ética, marca, practicidad, imagen, padres, amigos, novio/novia, hermano/hermana, publicidad, experto externo, personaje conocido.

Se identifican 3 grupos:

- Clase C: fuerte en precio, calidad, practicidad = los Racionales
- Clase B: fuerte en la moda, imagen e influencia de los familiares = los Sociales
- Clase A: débil en todos los criterios = los autónomos

Valores medios de las variables para cada clase. (en verde, significativamente por encima de la media, en rojo por debajo - en el sentido de la prueba t al 0,05%)

Detalle de las clases

Valores medios de las variables para cada clase. (en verde, significativamente superior a la media, en rojo, inferior, para un test de significación de 0,05%)

Clase	Padres	Amigos	Pareja	Hermano	Publicidad	Expertos	Famoso	Precio	Calidad	Moda	Respeto_medioambiente	Marca
A	1,444	1,604	1,404	0,749	0,829	0,842	0,762	5,401	6,698	1,912	0,842	1,618
B	2,523	3,341	3,491	1,406	1,705	1,532	0,703	5,841	7,535	5,530	1,221	3,744
C	2,145	1,972	2,258	1,298	0,580	1,912	0,216	6,808	7,777	0,761	1,765	2,803
Total	2,074	2,302	2,413	1,183	0,999	1,504	0,516	6,126	7,410	2,565	1,346	2,778

Conceptos clave

- **Análisis multivariantes:** conjunto de análisis que tienen en cuenta más de dos variables simultáneamente y que se utilizan en el procesamiento de encuestas para sintetizar información o determinar las variables más influyentes.
- **Clasificación o análisis tipológico:** técnica multivariante que consiste en agrupar -según los cálculos de proximidad- a los individuos en grupos de encuestados lo más homogéneos posible.
- **Regresión múltiple:** análisis multivariante que consiste en evaluar la influencia de varias variables explicativas candidatas sobre una variable explicada.

Ejercicios

Ejercicio 12a: clasificación turística

A partir de la encuesta "Turismo", en todas las preguntas relacionadas con las expectativas de vacaciones (Q12 - Agua y sol hasta Q21 - Estar con la familia), realizar una clasificación automática dividiendo la población turística en 4 clases. Interpreta y nombra las cuatro clases propuestas.

Ejercicio 12b: influencia de la edad y la duración de la estancia en los gastos

Utilizando el análisis de regresión múltiple, determina si el gasto total de los turistas está influenciado por la duración de la estancia y/o la edad. Si se verifican ambas influencias, ¿cuál es más fuerte?

Más recursos

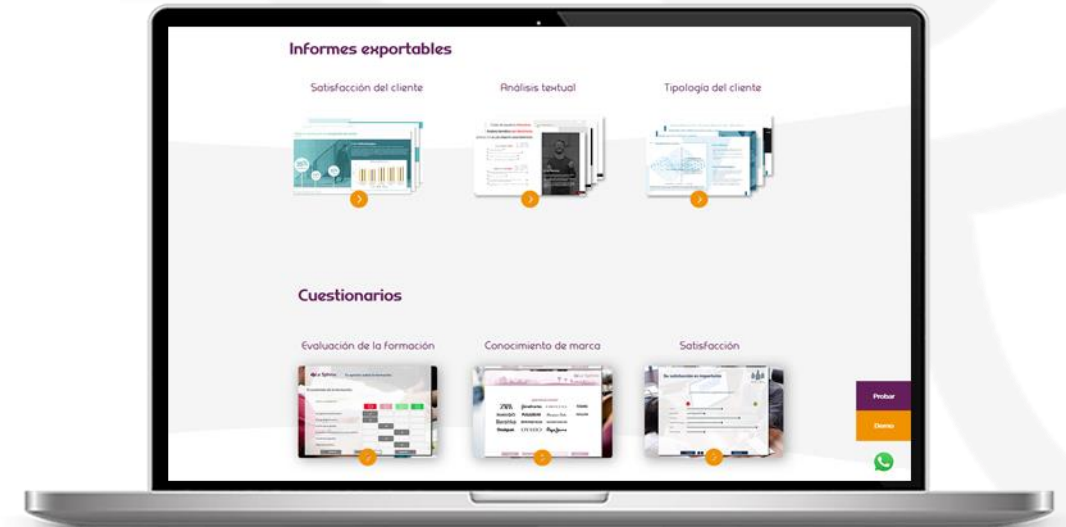
Tutoriales

www.lesphinx.es/videos



Ejemplos

www.lesphinx.es/ejemplos



Descargar una licencia

www.lesphinx.es/probar-gratis

