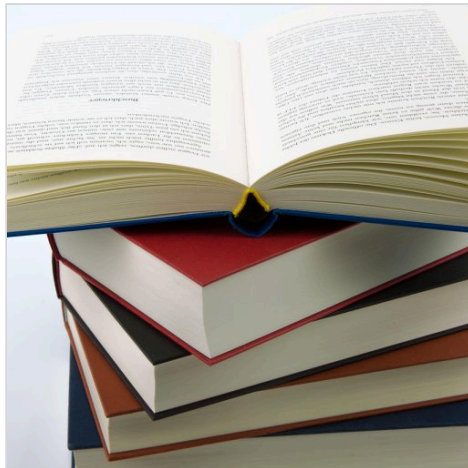


L U M B R E



Bivariante: relación y confusión

Grado: 4

Materia: Análisis Exploratorio de Datos · Análisis de Datos

11/10/2026

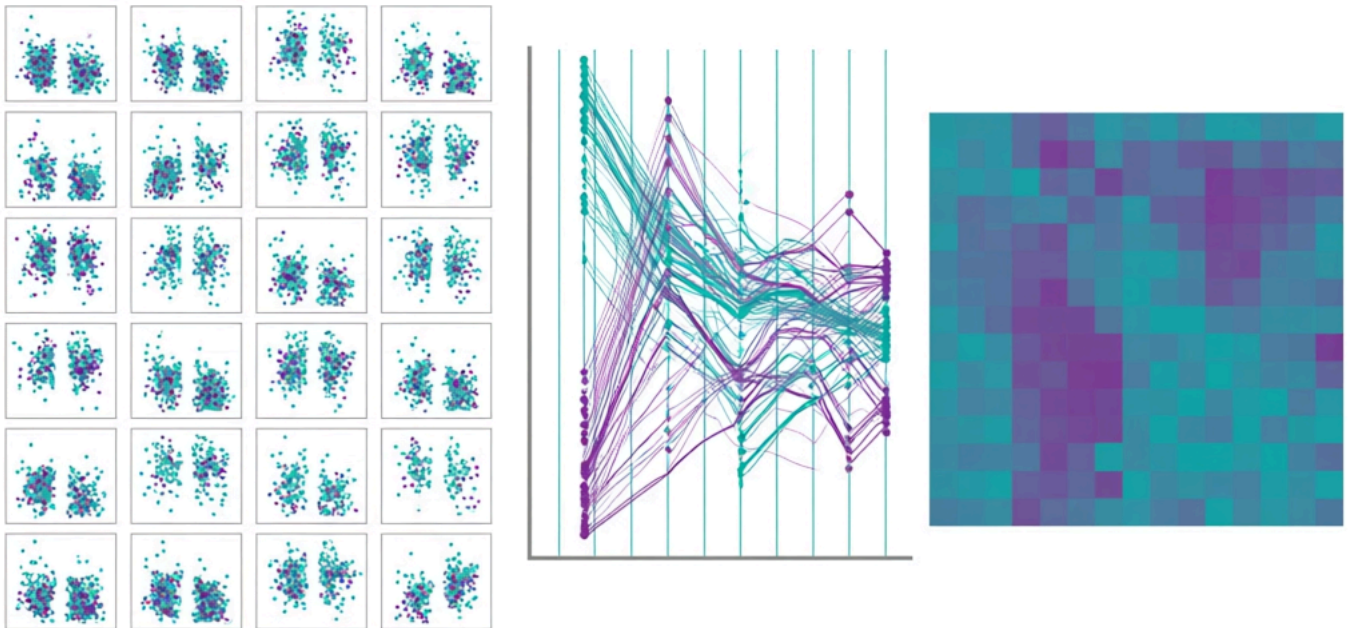
Antes de empezar

El interés real de los datos rara vez vive en una variable sola: está en cómo se mueven juntas. Esta lección enseña a mirar relaciones de a dos, a cuantificarlas sin mentir y a no confundir acompañarse con causar.

QUÉ APRENDERÁS

- Explorar la relación entre pares de variables, elegir la visualización y el estadístico según los tipos, y reconocer la confusión y la correlación espuria.

Bivariante: relación y confusión



Análisis exploratorio de datos: visualización multivariante y detección de patrones.

Mapa conceptual

- **EDA**
 - Univariante
 - Histogramas

- Boxplots
- Atípicos
- Bivalente
 - Scatter
 - Correlación
 - Confusión
- Multivalente
 - Pairs plot
 - Parallel coord
- Modelo
 - Hipótesis
 - Selección

El mapa mental interactivo es interactivo en la versión online.



Juntos no significa por causa

Dos variables pueden moverse juntas por causalidad, por azar, o porque una tercera las mueve a ambas. Explorar relaciones de a dos abre la puerta al hallazgo, pero también a la trampa más antigua del oficio: confundir acompañarse con causarse.

Depende de los tipos

La herramienta bivalente la deciden los tipos del par. Numérica-numérica pide un diagrama de dispersión; categórica-numérica, un boxplot por grupo; categórica-categórica, una tabla de contingencia. Emparejar mal el gráfico con los tipos produce visualizaciones sin significado.

Diagrama de dispersión

El scatter es la reina de la exploración bivalente: cada punto es una observación con sus dos valores. Muestra dirección, forma, fuerza y anomalías de la relación a la vez, sin resumirla. Ningún número sustituye a mirarlo.

La trampa de resumir demasiado pronto

Las cuatro nubes de Anscombe comparten casi idénticas medias, varianzas y correlaciones, pero sus dispersiones cuentan historias radicalmente distintas: una relación lineal, una curva, una dominada por un punto. Lección: el estadístico resume, el gráfico revela.

Correlación: qué mide

El coeficiente de correlación mide la fuerza y dirección de la relación LINEAL entre dos numéricas, entre menos uno y uno. Nada más. Uno cercano a cero no significa «sin relación»: puede haber una relación curva fortísima que la correlación lineal no capta.

Correlación y sus límites

- Solo captura dependencia lineal: una U perfecta da correlación casi cero.
- Es sensible a atípicos: un punto lejano puede dispararla o hundirla.
- Asume una relación uniforme: mezclar subgrupos la distorsiona.
- Nunca implica causalidad, por fuerte que sea.

Correlación de rangos

Cuando la relación es monótona pero no lineal, o hay atípicos, las correlaciones basadas en rangos (Spearman, Kendall) capturan mejor el acompañamiento sin exigir linealidad ni estar a merced de los extremos. Conviene mirarlas junto a la lineal.

Categórica frente a numérica

Para ver cómo se reparte una numérica según los niveles de una categoría, el boxplot por grupo compara medianas y dispersiones de un vistazo. Detecta diferencias entre niveles que la media global escondería al promediarlo todo.

Tablas de contingencia

Entre dos categóricas, la tabla de frecuencias cruzadas muestra cuántas observaciones caen en cada combinación. De ella salen las medidas de asociación y los gráficos de mosaico o barras agrupadas que revelan si una categoría se reparte distinto según la otra.

Simpson: la relación que se invierte

La paradoja de Simpson muestra cómo una relación presente en cada subgrupo puede desaparecer o invertirse al agrupar, por culpa de una variable de confusión desequilibrada. Es la prueba visual de que mirar sin estratificar puede mentir.

Variables de confusión

Una confusora influye a la vez en las dos variables estudiadas, creando una asociación aparente entre ellas que no es directa. Heladas y ahogados se correlacionan porque el calor aumenta ambas. Detectar confusores exige pensamiento sobre el mecanismo, no solo cálculo.

Estratificar la mirada

La defensa contra la confusión en la exploración es segmentar: estudiar la relación dentro de cada nivel del posible confusor. Si la asociación persiste en todos los estratos, es más creíble; si aparece solo al agrupar, probablemente era espuria.

Correlaciones espurias

Con suficientes variables y datos, siempre habrá pares altamente correlacionados por puro azar o por tendencias comunes (dos series que crecen con el tiempo se «correlacionan» sin relación alguna). El número alto no exonera de pensar.

No linealidad

Explorar no es solo calcular r : hay que buscar curvas, umbrales, interacciones. Un diagrama de dispersión con una línea de ajuste suave (una local) revela estructuras que un único coeficiente aplanaría y ocultaría.

Errores frecuentes

- Concluir «no hay relación» porque la correlación lineal sale cercana a cero.
- Atribuir causalidad a una correlación bivariante sin considerar confusores.
- Dejarse llevar por un único atípico que fabrica una correlación inexistente.

1. Una correlación de Pearson próxima a cero entre dos variables garantiza que:

- ☐ No existe relación alguna entre ellas
- ☐ No existe relación lineal, pero podría haber una no lineal fuerte
- ☐ Son independientes causalmente
- ☐ alguna es errónea

2. Correlación implica causalidad.

- ☐ Verdadero
- ☐ Falso

Matriz de correlaciones

Con muchas variables, la matriz de correlaciones y su heatmap resumen todos los pares de golpe. Útil para detectar bloques de variables redundantes, pero peligrosa si se lee sin los gráficos: esconde no linealidades y confusión.

Hacia lo multivariante

Las relaciones de a dos son un paso, no el destino. Muchas estructuras solo aparecen al considerar tres o más variables a la vez: interacciones, efectos condicionados, redundancias. La exploración bivariante prepara el terreno, no lo agota.

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. Bivariante



4. Ordena el análisis honesto de una asociación bivariante:

Elementos:

- mira el diagrama de dispersión
- cuantifica la fuerza con el estadístico adecuado
- revisa dentro de posibles confusores
- no concluyas causalidad sin evidencia

Graficar	. . .
Medir	. . .
Estratificar	. . .
Cautela	. . .

Anscombes quartet https://en.wikipedia.org/wiki/Anscombe%27s_quartet Los cuatro conjuntos con igual resumen y relaciones completamente distintas.

VIDEO

Video complementario

Recurso audiovisual para reforzar los conceptos.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=M7tX2MJGEY>

5. Hallas alta correlación entre ventas de helados y ahogamientos. ¿Cuál es el confusor más probable y qué harías con los datos para confirmarlo?

Referencia curricular

Grado:	4
Materia:	Análisis Exploratorio de Datos · Análisis de Datos

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-anex-003-bivariante-relacion-y-confusion>

Material educativo generado desde Lumbre · lumbre.lat