

L U M B R E



Matriz de covarianza y correlación

Grado: 4

Materia: Estadística Multivariada · Probabilidad y Estadística

11/10/2026

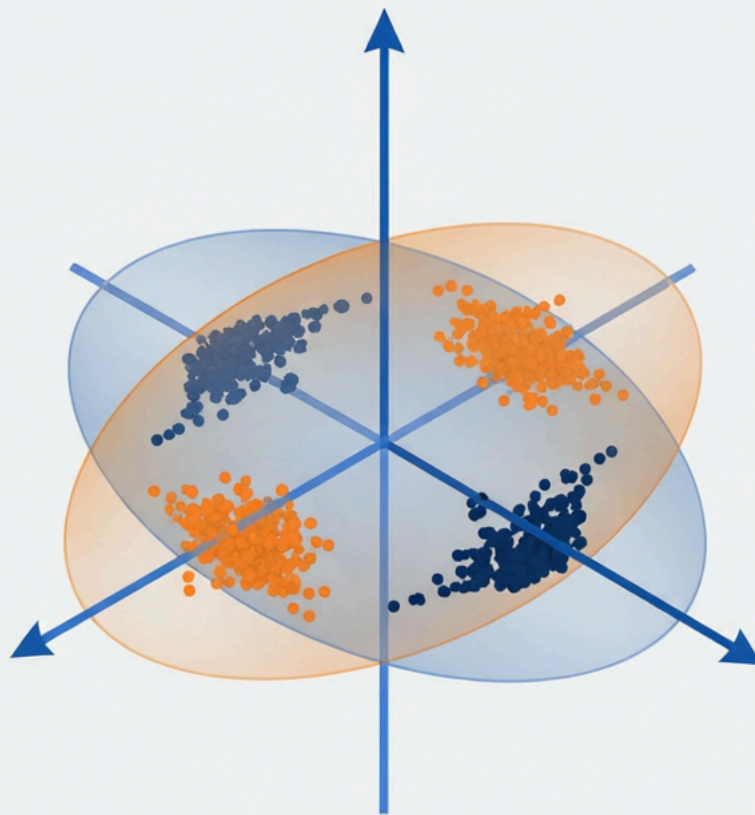
Antes de empezar

En cuanto hay más de una variable, su relación mutua importa tanto como su comportamiento individual. La matriz de covarianza condensa toda esa estructura de segundos momentos: es la piedra angular del análisis multivariante, del componente principal y del modelo gaussiano.

QUÉ APRENDERÁS

- Resumir la variabilidad conjunta de varias variables con las matrices de covarianza y correlación, e interpretar su geometría como base del análisis multivariante.

Matriz de covarianza y correlación



Estadística multivariada: PCA, covarianza y técnicas de reducción de dimensión.

Mapa conceptual

- **EDA**

- Univariante
 - Histogramas
 - Boxplots
 - Atipicos
- Bivariante
 - Scatter
 - Correlacion
 - Confusion
- Multivariante
 - Pairs plot
 - Parallel coord
- Modelo
 - Hipotesis
 - Selecccion

El mapa mental interactivo es interactivo en la versión online.



La forma conjunta de la nube

Una sola variable tiene una varianza. Dos tienen, además de sus varianzas, una covarianza que dice si se mueven juntas. Con p variables, toda esa información cabe en una matriz simétrica: el retrato estadístico de cómo coexisten.

De la varianza a la covarianza

La covarianza mide si dos variables varían a la unísono: positiva si suben juntas, negativa si se contrarrestan, cero si su variación lineal conjunta desaparece. Es el número que falta al estudiar variables de una en una.

La matriz de covarianza

Con p variables, se colocan las varianzas en la diagonal y las covarianzas por pares fuera de ella. El resultado es una matriz cuadrada, simétrica y semidefinida positiva que encapsula toda la estructura de segundo orden de los datos.

Propiedades de la matriz

- Simétrica: la covarianza de X con Y es la de Y con X.
- Semidefinida positiva: ninguna combinación lineal de variables tiene varianza negativa.
- Sus elementos diagonales son las varianzas individuales.
- Determinante y traza resumen volumen y dispersión total.

La correlación estandariza

La covarianza depende de las unidades, lo que dificulta compararla entre pares. Al dividir por las desviaciones típicas, la correlación queda acotada entre menos uno y uno, libre de escala, lista para comparar la fuerza de acompañamiento entre cualquier pareja.

Matriz de correlación

Es la versión tipificada de la de covarianza: unos en la diagonal, correlaciones fuera. Su heatmap asociado es la primera radiografía de un conjunto multivariante: delata bloques de variables redundantes, relaciones fuertes y candidatos a agruparse.

Geometría de la nube

La matriz de covarianza describe la forma de la nube de puntos: sus ejes principales y cuánto se extiende en cada dirección. Una matriz diagonal indica variables sin relación; correlaciones altas, una nube alargada y oblicua.

Autovalores y direcciones

Descomponer la matriz en autovalores y autovectores revela las direcciones ortogonales de máxima variabilidad y cuánta varianza acumula cada una. Es el corazón del análisis de componentes principales, que heredará esta geometría.

Traza y determinante

La traza, suma de varianzas, mide la dispersión total. El determinante, volumen de la elipsoide de dispersión, se anula si las variables son linealmente dependientes: un determinante cercano a cero advierte de colinealidad o redundancia.

Covarianza muestral frente a poblacional

La estimación a partir de datos corrige con n menos uno en el denominador para ser insesgada. Con pocas observaciones o muchas variables, la estimación se vuelve inestable y conviene encogerla o estructurarla.

Estimación en alta dimensión

Cuando el número de variables se acerca o supera al de observaciones, la matriz muestral es singular o casi, y no estimable con fiabilidad. Aquí aparecen la regularización de covarianzas y los estimadores encogidos, esenciales en genómica y finanzas.

Dependencia no lineal

Covarianza y correlación capturan solo acompañamiento lineal. Dos variables pueden tener una relación determinista fuerte y correlación cero si es curva. La matriz de segundos momentos es un resumen parcial, no la verdad completa de la dependencia.

Errores frecuentes

- Comparar covarianzas entre pares con unidades distintas y concluir falsas jerarquías de fuerza.
- Confiar en correlación cero como prueba de independencia cuando hay relación no lineal.
- Invertir una matriz muestral casi singular sin diagnosticar la colinealidad.

1. La matriz de covarianza tiene determinante prácticamente nulo. Esto señala sobre todo:

- ☐ Variables con gran dispersión
- ☐ Dependencia lineal entre variables (redundancia o colinealidad)
- ☐ Ausencia total de datos
- ☐ Distribución normal

2. La matriz de correlación tiene unos en toda su diagonal.

- ☐ Verdadero
- ☐ Falso

Blanquear los datos

Multiplicar los datos por una transformación derivada de la matriz de covarianza puede «blanquearlos»: volver su matriz de covarianza la identidad, eliminando correlaciones y homogeneizando escalas. Útil como preprocesado antes de ciertos modelos.

Puente al multivariante

Covar, correlacionar, descomponer en autovalores: esta secuencia alimenta el componente principal, el análisis factorial, los modelos gaussianos y la detección de

anomalías. Dominar la matriz de covarianza es dominar el idioma de la dependencia multivariante.

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. Covarianza



4. Ordena el análisis de la estructura conjunta:

Elementos:

- formar la matriz de covarianzas
- tipificar a la matriz de correlaciones
- descomponer en autovalores y autovectores
- interpretar ejes, volumen y redundancia

Covar	. . .
Correl	. . .
Autoval	. . .
Leer	. . .

VIDEO

Video complementario

Recurso audiovisual para reforzar los conceptos.



Escanea para verlo en Lumbre

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

Dependence and covariance <https://www.jstatsoft.org/Recursos> sobre matrices de covarianza y su estimación.

5. Inspeccionas el heatmap de correlaciones de un dataset y ves un bloque de ocho variables casi perfectamente correlacionadas. ¿Qué sospechas y qué harías antes de modelar?

Referencia curricular

Grado:	4
Materia:	Estadística Multivariada · Probabilidad y Estadística

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



<https://univ.lumbre.lat/object/cdia-emul-001-matriz-de-covarianza-y-correlacion>

Material educativo generado desde Lumbre · lumbre.lat