covariación – covariation

Authored by
memjavad

November 26, 2025

Covariación

Primary Disciplinary Field(s): Estadística, Probabilidad, Econometría, Finanzas

1. Definición Central

La covariación, o covarianza (denotada comúnmente como $Cov(X, Y)$ o $sigma_{XY}$), es un concepto fundamental en la estadística y la teoría de la probabilidad que mide el grado en que dos variables aleatorias, X e Y, cambian juntas. En esencia, cuantifica la tendencia de estas variables a desviarse simultáneamente de sus respectivos valores esperados o medias. Formalmente, la covariación se define como el valor esperado del producto de las desviaciones de cada variable respecto a su media. Si E y E representan las medias de X e Y, respectivamente, la covariación se expresa matemáticamente como $Cov(X, Y) = E)(Y - E)]$. Este valor resultante puede ser positivo, negativo o cero, proporcionando una interpretación directa sobre la naturaleza de la relación lineal entre las variables.

Una covariación **positiva** indica que, cuando los valores de la variable X tienden a ser mayores que su media, los valores de la variable Y también tienden a ser mayores que su media, y viceversa. Esto sugiere una relación directa o una tendencia a moverse en la misma dirección. Por el contrario, una covariación **negativa** implica que cuando X se desvía por encima de su media, Y tiende a desviarse por debajo de la suya, lo que señala una relación inversa o una tendencia a moverse en direcciones opuestas. Finalmente, si la covariación es **cero** o cercana a cero, esto sugiere que no existe una relación lineal discernible entre las variables. Es crucial notar que una covariación nula no implica necesariamente la independencia total de las variables; solo indica la ausencia de una dependencia lineal, pudiendo existir complejas relaciones no lineales que la covariación no logra capturar.

El valor de la covariación es expresado en las unidades de las variables multiplicadas, lo que a menudo dificulta su interpretación directa en términos de la fuerza de la relación, a diferencia de su concepto hermano, el coeficiente de correlación. Sin embargo, su papel como medida de la variación conjunta es indispensable para el desarrollo de modelos estadísticos multivariados. La magnitud de la covariación informa sobre la escala de la dispersión conjunta, pero no sobre la intensidad intrínseca de la dependencia. Por ejemplo, una covariación de 100 puede ser fuerte en un contexto de variables pequeñas, pero débil en un contexto de variables con grandes varianzas. Esta dependencia de la escala es una característica definitoria y, a menudo, una limitación de la medida.

2. Etimología y Desarrollo Histórico

El concepto de covariación emerge directamente de los estudios pioneros sobre la regresión y la variación conjunta de características biológicas a finales del siglo XIX. El trabajo seminal fue realizado por Sir Francis Galton, quien buscaba cuantificar la herencia de rasgos. Galton introdujo la idea de la "regresión a la media" y desarrolló métodos gráficos para visualizar la relación entre variables, sentando las bases conceptuales para la medición de la variación conjunta. Aunque Galton no formuló la covariación en su forma matemática moderna, su insistencia en cuantificar cómo se relacionan las desviaciones individuales fue el catalizador.

La formalización matemática de la covariación y, más concretamente, la introducción del término y la fórmula que hoy conocemos, se atribuye principalmente a Karl Pearson. Pearson, influenciado por Galton, desarrolló el coeficiente de correlación producto-momento (a menudo denominado coeficiente de correlación de Pearson), que utiliza la covariación como su numerador. Al estandarizar la covariación dividiéndola por el producto de las desviaciones estándar de las variables, Pearson creó una medida que era independiente de la escala, resolviendo así una de las principales limitaciones de la covariación pura. Este desarrollo, ocurrido alrededor de 1895, estableció la covariación como el momento central mixto de segundo orden en la estadística.

A lo largo del siglo XX, la covariación se consolidó como un pilar en la teoría de la probabilidad. Su importancia creció exponencialmente con el desarrollo de la estadística multivariada y la econometría. La comprensión de la covariación permitió la creación de herramientas complejas como el análisis de componentes principales (PCA), el análisis factorial y, crucialmente, la teoría moderna de portafolio desarrollada por Harry Markowitz, donde la covariación entre los rendimientos de diferentes activos es el elemento clave para la gestión del riesgo. Así, lo que comenzó como una herramienta para la biología se transformó en un instrumento esencial para casi todas las ciencias cuantitativas.

3. Propiedades Matemáticas Fundamentales

La covariación posee varias propiedades matemáticas esenciales que la hacen útil en el álgebra de variables aleatorias y en la simplificación de cálculos complejos. La primera propiedad fundamental es la **simetría**: la covariación de X con Y es idéntica a la covariación de Y con X. Matemáticamente, $Cov(X, Y) = Cov(Y, X)$. Esta propiedad es intuitiva, ya que la relación conjunta no depende del orden en que se midan las variables. Esta simetría es vital al construir matrices de covariación, asegurando que estas matrices sean simétricas y, por lo tanto, posean propiedades algebraicas deseables para el análisis multivariado.

Otra propiedad crítica es la **linealidad**. La covariación es lineal en sus argumentos. Esto significa que si se considera una transformación lineal de una de las variables, como $aX + b$, donde a y b son constantes, la covariación se ajusta de manera predecible. Específicamente, $Cov(aX + b, Y) = a cdot Cov(X, Y)$. Las constantes aditivas (b) no afectan la covariación, ya que la covariación mide las desviaciones respecto a la media, y sumar una constante simplemente desplaza la media en la misma cantidad. Esta propiedad se extiende para incluir transformaciones lineales en ambas variables: $Cov(aX + b, cY + d) = ac cdot Cov(X, Y)$. Esta linealidad es fundamental en la modelización de regresión y en la manipulación de combinaciones lineales de variables aleatorias.

Una tercera propiedad crucial relaciona la covariación con la varianza. La varianza de una variable es simplemente la covariación de esa variable consigo misma: $Var(X) = Cov(X, X)$. Esta relación unifica los conceptos de variación individual (varianza) y variación conjunta (covariación) dentro del marco de los momentos estadísticos. Además, la covariación juega un papel determinante en el cálculo de la varianza de una suma de variables aleatorias. Para dos variables, la varianza de la suma es $Var(X + Y) = Var(X) + Var(Y) + 2 cdot Cov(X, Y)$. Si las variables son independientes, su covariación es cero, y la fórmula se simplifica a la suma de las varianzas. Esta propiedad es la base de la matriz de covarianzas, que resume todas las varianzas y covariaciones de un conjunto de variables aleatorias, siendo un insumo indispensable para el análisis multivariado.

4. Métodos de Cálculo y Estimación

El método de cálculo de la covariación difiere ligeramente si se trabaja con la población completa (parámetro) o con una muestra (estimador). Para una **población** discreta finita de N pares de datos (x_i, y_i), la covariación poblacional ($sigma_{XY}$) se calcula como el promedio de los productos de las desviaciones: $sigma_{XY} = frac{1}{N} sum_{i=1}^{N} (x_i - mu_X)(y_i - mu_Y)$, donde mu_X y mu_Y son las medias poblacionales. Si las variables son continuas, la suma es reemplazada por una integral y se utiliza la función de densidad de probabilidad conjunta. Un método de cálculo alternativo y a menudo más eficiente, derivado de la definición del valor esperado, es la fórmula computacional: $Cov(X, Y) = E - EE$, donde E es el valor esperado del producto de las variables.

Cuando se trabaja con una **muestra** de datos, se utiliza la covariación muestral (s_{XY}) como un estimador de la covariación poblacional. La fórmula para la covariación muestral es: $s_{XY} = frac{1}{n-1} sum_{i=1}^{n} (x_i - bar{x})(y_i - bar{y})$, donde $bar{x}$ y $bar{y}$ son las medias muestrales y n es el tamaño de la muestra. El uso del divisor $n-1$ en lugar de n es crucial. Este ajuste se realiza para obtener un estimador insesgado de la covariación poblacional. Al usar las medias muestrales ($bar{x}, bar{y}$) en lugar de las verdaderas medias poblacionales (mu_X, mu_Y), se pierde un grado de libertad, y la división por $n-1$ asegura que el estimador muestral no subestime sistemáticamente la verdadera covariación poblacional.

La elección entre el cálculo poblacional y el muestral depende del contexto del análisis. En la práctica estadística y econométrica, casi siempre se trabaja con muestras, haciendo que la estimación insesgada con $n-1$ sea la norma. La precisión en el cálculo de la covariación es vital, ya que cualquier error en esta estimación se propaga a medidas más complejas, como la correlación o la matriz de covarianzas. Los paquetes de software estadístico modernos automatizan estos cálculos, pero la comprensión de la diferencia entre los divisores N (población) y $n-1$ (muestra) es fundamental para la correcta interpretación de los resultados.

5. Relación Crucial con la Correlación

Aunque la covariación y la correlación miden la relación lineal entre dos variables, son fundamentalmente distintas en su interpretación y uso. La covariación, como se mencionó, es sensible a la **escala** de las variables. Si se miden las alturas de personas en metros y luego se vuelven a medir en centímetros, el valor numérico de la covariación aumentará cien veces, aunque la relación subyacente entre las alturas no haya cambiado. Esta dependencia de la unidad de medida hace que sea imposible comparar la fuerza de la relación entre diferentes pares de variables que tienen unidades o magnitudes distintas.

El coeficiente de correlación de Pearson (r), por otro lado, es una versión **normalizada** de la covariación. Se define como $r = frac{Cov(X, Y)}{sigma_X sigma_Y}$, donde $sigma_X$ y $sigma_Y$ son las desviaciones estándar de X e Y. Al dividir la covariación por el producto de las desviaciones estándar, el coeficiente de correlación se convierte en una medida adimensional que siempre oscila entre -1 y $+1$. El valor de r indica la fuerza y la dirección de la relación lineal de manera universal: un valor cercano a $+1$ indica una relación lineal positiva fuerte; un valor cercano a -1 indica una relación lineal negativa fuerte; y un valor cercano a 0 indica una relación lineal débil o nula.

Por lo tanto, mientras la covariación es esencialmente un componente bruto que indica la dirección de la variación conjunta, la correlación es la métrica estandarizada que permite evaluar la **intensidad** de esa relación, independientemente de las unidades de medida. En la práctica, la covariación es crucial para el análisis interno (por ejemplo, dentro de un modelo financiero donde todas las unidades son monetarias), mientras que la correlación es preferida para la comparación y la interpretación general de la fuerza de la relación en el ámbito académico y científico. Ambos conceptos son inseparables: la covariación es la base, y la correlación es su versión interpretable y estandarizada.

6. Aplicaciones en Campos Especializados

La covariación es una herramienta analítica indispensable en numerosos campos cuantitativos, siendo las finanzas y el análisis multivariado sus aplicaciones más prominentes. En la **teoría moderna de portafolio** (MPT), desarrollada por Markowitz, la covariación entre los rendimientos de diferentes activos es el factor clave para la diversificación y la gestión del riesgo. Los inversores buscan activos cuyos rendimientos tengan una covariación baja o, idealmente, negativa. Una covariación negativa implica que cuando un activo tiene un rendimiento bajo, el otro tiende a tener un rendimiento alto, lo que reduce la volatilidad general del portafolio. El riesgo de un portafolio no es simplemente la suma de los riesgos individuales de los activos, sino que depende fundamentalmente de la matriz de covarianzas que relaciona todos los pares de activos.

En la **estadística multivariada**, la matriz de covarianzas es el concepto central. Esta matriz cuadrada y simétrica contiene las varianzas de cada variable en su diagonal principal y las covariaciones entre cada par de variables fuera de la diagonal. Esta matriz es el punto de partida para técnicas avanzadas como el Análisis de Componentes Principales (PCA), donde la meta es transformar las variables originales correlacionadas en un nuevo conjunto de variables no correlacionadas (componentes principales). La matriz de covarianzas también es crucial en el análisis discriminante, en modelos de ecuaciones estructurales y en la estimación de modelos de regresión lineal múltiple, donde se utiliza para evaluar la multicolinealidad.

En **econometría y series de tiempo**, el concepto de covariación se extiende a la autocovariación, que mide la relación lineal entre una variable y sus propios valores pasados en diferentes momentos del tiempo. Esto es fundamental para identificar patrones de dependencia temporal y para la construcción de modelos ARIMA. Asimismo, en el ámbito de la genética y la epidemiología, la covariación se utiliza para estimar la heredabilidad de rasgos, analizando cómo varían conjuntamente las características entre individuos genéticamente relacionados, proporcionando una medida de la influencia genética y ambiental compartida.

7. Limitaciones y Debates

A pesar de su utilidad fundamental, la covariación presenta varias limitaciones inherentes que han generado debates sobre su aplicación aislada. La crítica más significativa es su ya mencionada **dependencia de la escala**. Dado que el valor numérico de la covariación cambia con las unidades de medida, es imposible establecer umbrales universales para determinar si una covariación es "grande" o "pequeña". Esto obliga a los analistas a recurrir casi siempre a la correlación estandarizada para la interpretación de la fuerza de la relación, relegando la covariación a un rol de ingrediente matemático más que de métrica interpretativa final.

Otra limitación crítica es que la covariación solo mide la **relación lineal**. Si la relación entre dos variables es fuertemente no lineal (por ejemplo, una relación cuadrática o exponencial), la covariación puede ser cero o cercana a cero, sugiriendo incorrectamente que no existe una relación significativa. Este fallo subraya que la covariación no es una prueba de independencia. Si $Cov(X, Y) = 0$, las variables son incorrelacionadas, pero no necesariamente independientes. La independencia de X e Y implica que $Cov(X, Y) = 0$, pero el recíproco no es cierto, salvo en el caso especial de variables que siguen una distribución normal conjunta.

Finalmente, la covariación es **sensible a los valores atípicos** (outliers). Debido a que se calcula utilizando el producto de las desviaciones respecto a la media, un solo par de datos extremos puede influir desproporcionadamente en el valor de la covariación, distorsionando la percepción de la relación subyacente entre la mayoría de los datos. Este problema ha llevado al desarrollo de medidas robustas de asociación, como la correlación de Spearman o la correlación de Kendall, que se basan en rangos en lugar de en los valores absolutos de las desviaciones, mitigando el impacto de los valores extremos en el análisis de la dependencia.

Further Reading

Covarianza (Wikipedia en español)
Karl Pearson (Wikipedia en español)
Matriz de covarianzas (Wikipedia en español)