

Regresión lineal múltiple

La regresión múltiple permite a los investigadores examinar el efecto de más de una variable independiente en la respuesta al mismo tiempo. Para algunas preguntas de investigación, la regresión se puede utilizar para examinar en qué medida un conjunto particular de variables independientes puede explicar suficientemente el resultado.

1.1 Espacio vectorial y Proyección

1.1.1 Espacio vectorial

Un espacio vectorial es un conjunto de vectores que es cerrado bajo la adición y la multiplicación por un escalar de vectores finitos. Suponga V es cerrado bajo el vector adición en \mathbb{R}^n : si $u, v \in V$ entonces $u + v \in V$. Por otro lado V es cerrado bajo la multiplicación escalar: Si $a \in \mathbb{R}^1$ y $v \in V$, entonces $av \in V$. Por lo que V es un espacio vectorial en \mathbb{R}^n .

Una base hace posible expresar cada vector del espacio como una tupla única de los elementos del campo, aunque se debe tener precaución cuando un espacio vectorial no tiene una base finita. En álgebra lineal, una base es un conjunto de vectores que, en una combinación lineal, pueden representar cada vector en un espacio vectorial dado, y tales que ningún elemento del conjunto pueda representarse como una combinación lineal de los demás. En otras palabras, una base es un conjunto generador linealmente independiente. Es decir, cualquier vector $x' = (x_1, x_2, \dots, x_n)$ en \mathbb{R}^n puede ser una combinación lineal de e_1, e_2, \dots, e_n . En efecto,

$$x = x_1e_1 + x_2e_2 + x_3e_3 + \dots + x_ne_n.$$

Esta representación es única. Que quiere decir que no existe otra representación tal que

$$x = x_1^*e_1 + x_2^*e_2 + x_3^*e_3 + \dots + x_n^*e_n.$$

Ya que,

$$0 = \mathbf{x} - \mathbf{x} = (x_1 - x_1^*)e_1 + (x_2 - x_2^*)e_2 + \dots + (x_n - x_n^*)e_n$$

Entonces, $x_i = x_i^*$ para todo $i = 1, 2, \dots, n$.

Dado un espacio vectorial V , un subconjunto no vacío W de V que es cerrado bajo la suma y la multiplicación escalar se llama subespacio de V . La intersección de todos los subespacios que contengan un conjunto dado de vectores es llamado SPAN o generador. Si no se puede eliminar ningún vector

sin cambiar el SPAN (generador), se dice que los vectores en este conjunto son linealmente independientes. Un conjunto linealmente independiente cuyo intervalo es V se llama base de V . Un vector generado (SPAN) por dos vectores v y w (ambos no necesariamente independientes) se pueden definir como: $x : x = av + bw$, para todo $(a, b) \in \mathbb{R}^2$. Si un espacio vectorial S está generado por un conjunto de vectores independientes v_1, v_2, \dots, v_p . Es decir, S es el conjunto de vectores

$$\{x : x = a_1v_1 + a_2v_2 + \dots + a_pv_p, \text{ para todo } (a_1, a_2, \dots, a_p) \in \mathbb{R}^p\},$$

Entonces la dimensión de S es p . Los vectores v_1, v_2, \dots, v_p son la base del espacio vectorial S . La dimensión de un espacio vectorial S es el mayor número de un conjunto de vectores independientes en S . Si la dimensión de un espacio lineal S es p , se escribe $\text{Dim}(S) = p$.

1.1.2 Vectores linealmente independientes

Si existe un número finito de vectores distintos v_1, v_2, \dots, v_n en el espacio vectorial V y los escalares a_1, a_2, \dots, a_n no todos cero, tal que

$$a_1v_1 + a_2v_2 + \dots + a_nv_n = 0,$$

entonces los vectores v_1, v_2, \dots, v_n se dice que son linealmente dependientes. Si v_1, v_2, \dots, v_n son dependientes entonces de estos n vectores hay al menos un vector que puede expresarse como una combinación lineal de otros vectores. Tenga en cuenta que el cero a la derecha es el vector cero, no el número cero. Si no existen tales escalares, entonces se dice que los vectores v_1, v_2, \dots, v_n son linealmente independientes. Esta condición se puede reformular de la siguiente manera: Siempre que a_1, a_2, \dots, a_n sean escalares tales que

$$a_1v_1 + a_2v_2 + \dots + a_nv_n = 0,$$

tenemos $a_i = 0$ para $i = 1, 2, \dots, n$. Entonces, v_1, v_2, \dots, v_n son linealmente independientes.

Una base de un espacio vectorial V se define como un subconjunto de vectores en V que son linealmente independientes y estos vectores generan el espacio V . En consecuencia, si (v_1, v_2, \dots, v_n) es una lista de vectores en V , entonces estos vectores forman una base si y solo si todo vector $x \in V$ puede expresarse de forma única mediante una combinación lineal de v_1, v_2, \dots, v_p . Es decir,

$$x = a_1v_1 + a_2v_2 + \dots + a_nv_n, \text{ para cualquier } x \in V.$$

El número de vectores base en V se denomina dimensión del espacio lineal V . Tenga en cuenta que un espacio vectorial puede tener más de una base, pero el número de vectores que forman la base del espacio vectorial V siempre es fijo. Es decir, la dimensión del espacio vectorial V es fija pero habrá más de una base. De hecho, si la dimensión del espacio vectorial V es n , entonces cualquier n vector linealmente independiente en V forma su base.

1.1.3 Producto punto y Proyección

Si $\mathbf{x} = (x_1, x_2, \dots, x_n)$ e $\mathbf{y} = (y_1, y_2, \dots, y_n)$ dos vectores en un espacio Euclidiano vectorial \mathbb{R}^n . El producto punto de los vectores \mathbf{x} e \mathbf{y} se define como

$$\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2 + \dots + x_ny_n.$$

Dos vectores se llaman ortogonales si su producto vectorial es 0. Si θ es el ángulo entre dos vectores, el coseno del ángulo se define por

$$\cos(\theta) = \frac{\mathbf{x} \cdot \mathbf{y}}{|\mathbf{x}||\mathbf{y}|} = \frac{x_1y_1 + x_2y_2 + \dots + x_ny_n}{\sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \sqrt{y_1^2 + y_2^2 + \dots + y_n^2}}. \quad (1.1)$$

Dos vectores ortogonales forman 90° ; Es decir, son perpendiculares. Una aplicación importante del producto punto es la proyección. La proyección de un vector \mathbf{y} sobre otro vector \mathbf{x} forma un nuevo

vector que tiene la misma dirección que el vector \mathbf{x} y la longitud $|\mathbf{y}| \cos(\theta)$, donde $|\mathbf{y}|$ denota la longitud del vector \mathbf{y} y θ es el ángulo entre los vectores. Escribimos la proyección como $P_{\mathbf{x}}\mathbf{y}$. La proyección vectorial puede ser expresado como

$$\begin{aligned} P_{\mathbf{x}}\mathbf{y} &= |\mathbf{y}| \cos(\theta) \frac{\mathbf{x}}{|\mathbf{x}|} \\ &= |\mathbf{y}| \frac{\mathbf{x} \cdot \mathbf{y}}{|\mathbf{x}| |\mathbf{y}|} \frac{\mathbf{x}}{|\mathbf{x}|} \\ &= \frac{x_1 y_1 + x_2 y_2 + \cdots + x_n y_n}{x_1^2 + x_2^2 + \cdots + x_n^2} \mathbf{x} \\ &= \lambda \mathbf{x}. \end{aligned}$$

Donde λ es un escalar,

$$\lambda = \frac{x_1 y_1 + x_2 y_2 + \cdots + x_n y_n}{x_1^2 + x_2^2 + \cdots + x_n^2} = \frac{\mathbf{x} \cdot \mathbf{y}}{\mathbf{x} \cdot \mathbf{x}}. \quad (1.2)$$

Así, la proyección de \mathbf{y} sobre el vector \mathbf{x} es un vector \mathbf{x} que multiplica un escalar λ donde λ es el $\cos(\theta)$ y θ es el ángulo entre los vectores \mathbf{x} e \mathbf{y} .

Si \mathbf{x} e \mathbf{y} son dos vectores en \mathbb{R}^n . Consideremos la diferencia vector entre el vector \mathbf{e} , $\mathbf{e} = \lambda \mathbf{x} - \mathbf{y}$, y $\lambda = \mathbf{x} \cdot \mathbf{y} / \mathbf{x} \cdot \mathbf{x}$. El vector \mathbf{e} es perpendicular para el vector \mathbf{x} donde $\lambda = (\mathbf{x} \cdot \mathbf{y}) / (\mathbf{x} \cdot \mathbf{x})$. Para ver esto, simplemente calculamos el producto escalar de \mathbf{e} y \mathbf{x} :

$$\mathbf{e} \cdot \mathbf{x} = (\lambda \mathbf{x} - \mathbf{y}) \cdot \mathbf{x} = \lambda \mathbf{x} \cdot \mathbf{x} - \mathbf{x} \cdot \mathbf{y} = \left(\frac{\mathbf{x} \cdot \mathbf{y}}{\mathbf{x} \cdot \mathbf{x}} \right) \mathbf{x} \cdot \mathbf{x} - \mathbf{x} \cdot \mathbf{y} = 0.$$

Por lo tanto, el ángulo entre \mathbf{e} y \mathbf{x} es 90° . Es decir, estos vectores son perpendiculares. Además, dado que \mathbf{e} es perpendicular a \mathbf{x} , es el vector con la distancia más corta entre todos los vectores que comienzan desde el final de \mathbf{y} , y terminan en cualquier punto de \mathbf{x} .

Si un espacio vectorial tiene una base y la longitud de los vectores base es una unidad, entonces esta base es una base ortonormal. Cualquier base dividida por su longitud forma una base ortonormal. Si S es un subespacio p -dimensional de un espacio vectorial V , entonces es posible proyectar vectores en V sobre S . Si el subespacio S tiene una base ortonormal (w_1, w_2, \dots, w_p) , para cualquier vector \mathbf{y} en V , la proyección de \mathbf{y} sobre el subespacio S es

$$P_S \mathbf{y} = \sum_{i=1}^p (\mathbf{y} \cdot w_i) w_i. \quad (1.3)$$

Sean dos subespacios S y T del espacio vectorial V y la unión $S \cup T = V$. Si para cualquier vector $\mathbf{x} \in S$ y cualquier $\mathbf{y} \in T$, tal que el producto punto $\mathbf{x} \cdot \mathbf{y} = 0$, entonces se dice que los dos espacios vectoriales S y T son ortogonales. O podemos decir que T es el espacio ortogonal de S , denotado por

$$T = S^\perp.$$

Así, para un espacio vectorial V , si S es un subespacio vectorial en V , entonces $V = S \cup S^\perp$. Cualquier vector $\mathbf{y} \in V$ puede ser escrito de manera única como $\mathbf{y}_S + \mathbf{y}_S^\perp$, donde $\mathbf{y}_S \in S$ e \mathbf{y}_S^\perp es en S^\perp , el subespacio ortogonal de S .

Una proyección de un vector sobre un espacio lineal S es en realidad una transformación lineal del vector y se puede representar mediante una matriz de proyección multiplicada por el vector. Una matriz de proyección P es una matriz cuadrada de $n \times n$ que da la proyección de \mathbb{R}^n al subespacio S . Las columnas de P son las proyecciones de los vectores base estándar, y S es la imagen de P . Para la matriz de proyección tenemos los siguientes teoremas.

Teorema 1.1 Una matriz cuadrada P es una matriz de proyección si y solo si es idempotente. Es decir, $P^2 = P$. ■

Teorema 1.2 Sea $U = (u_1, u_2, \dots, u_k)$ una base ortonormal para un subespacio W del espacio lineal V . La matriz UU' es una matriz proyección de V sobre W . Es decir, para cualquier valor $v \in V$ la proyección de v sobre W es $\text{Proj}_W v = UU'v$. ■

La matriz UU' es llamada matriz proyección par el subespacio W . No depende de la elección de la base ortonormal. Si no partimos de una base ortonormal de W , podemos construir la matriz de proyección. Esto se puede resumir en el siguiente teorema.

Lema 1.1 Suponga que A es una matriz $a \times k$ donde sus columnas son linealmente independientes. Entonces, AA' es invertible.

Demostración.- Considere la transformación $A : \mathbb{R}^k \rightarrow \mathbb{R}^a$ determinado por A . Ya que las columnas de A son linealmente independientes, esta transformación es uno a uno. Además, el espacio nulo de A' es ortogonal al espacio columna de A . Por lo tanto, A' es uno a uno en el espacio columna de A y como resultado, $A'A$ es una transformación uno a uno $\mathbb{R}^k \rightarrow \mathbb{R}^k$. Por el teorema de matriz invertida, $A'A$ es invertible. ■

Teorema 1.3 Sea $A = (a_1, a_2, \dots, a_k)$ cualquier base para un subespacio W de V . La matriz $A(A'A)^{-1}A'$ es una matriz proyección de V sobre W . Es decir, para cualquier vector $v \in V$ la proyección de v sobre W es

$$\text{Proj}_W v = A(A'A)^{-1}A'v. \quad (1.4)$$

Demostración.- Derivemos ahora la matriz de proyección para el espacio de columnas de A . Obsérvese que cualquier elemento del espacio de columnas de A es una combinación lineal de las columnas de A , es decir, $x_1a_1 + x_2a_2 + \dots + x_ka_k$. Si escribimos

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix},$$

entonces, tenemos

$$x_1a_1 + x_2a_2 + \dots + x_ka_k = Ax.$$

Ahora, para cualquier vector $v \in \mathbb{R}^n$, denotemos la proyección de v sobre W por x_p .

$$\text{Proj}_W v = Ax_p.$$

La matriz de proyección se puede encontrar calculando x_p . La proyección del vector v sobre W se caracteriza por el hecho de que $v - \text{Proj}_W v$ es ortogonal a cualquier vector w en W . Así tenemos

$$w \cdot (v - \text{Proj}_W v) = 0$$

para todo w in W . Ya que $w = Ax$, tenemos

$$Ax(v - Ax_p) = 0$$

para todo $x \in \mathbb{R}^n$. Escribir este producto escalar en términos de matrices da como resultado

$$(A\mathbf{x})'(v - A\mathbf{x}_p) = 0$$

el cual es equivalente a

$$(\mathbf{x}'A')(v - A\mathbf{x}_p) = 0.$$

Convirtiendo a producto punto, tenemos

$$\mathbf{x} \cdot A'(v - A\mathbf{x}_p) = 0.$$

De donde,

$$A'v = A'A\mathbf{x}_p.$$

Luego, ya que $A'A$ es invertible, se tiene

$$(A'A)^{-1}A'v = \mathbf{x}_p.$$

Después, ya que $A\mathbf{x}_p$ es la proyección deseada,

$$A(A'A)^{-1}v = A\mathbf{x}_p = \text{Proj}_w u$$

Por lo tanto, concluimos que la matriz proyección para W es $A(A'A)^{-1}A'$. ■

La matriz proyección es muy útil en las discusiones posteriores del modelo de regresión $Y\mathbf{X}\beta + \epsilon$. Una matriz cuadrada $P = \mathbf{X}(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}'$, se construye utilizando la matriz de diseño. Se puede verificar fácilmente que P es una matriz idempotente:

$$P^2 = \mathbf{X}(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}' = P.$$

Así, $P = \mathbf{X}(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}'$ es una matriz proyección. Además, si definimos una matriz como $I - P = I - \mathbf{X}(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}'$. Es fácil ver que $I - P$ es también idempotente. De hecho,

$$(I - P)^2 = I - 2P + P^2 = I - 2P + P = I - P.$$

Por lo tanto, $I - P = I - \mathbf{X}(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}'$ es una matriz proyección. En las secciones siguientes veremos cómo se utilizan estas matrices de proyección para obtener el mejor estimador lineal insesgado (BLUE) para el modelo de regresión lineal y cómo se utilizan en el diagnóstico del modelo de regresión.

1.2 Forma matricial de regresión lineal múltiple

Definición 1.1 Un modelo lineal es definido como un modelo que es lineal en los parámetros de regresión, es decir, lineal en los β'_i s.

1.3 Forma cuadrática de variables aleatorias

Definición 1.2 Sea $y' = (y_1, y_2, \dots, y_n)$, n variables reales y a_{ij} $n \times n$ números reales, donde $i, j = 1, 2, \dots, n$. Una forma cuadrática de y_1, y_2, \dots, y_n es definida como

$$f(y_1, y_2, \dots, y_n) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} y_i y_j = y' A y.$$

Esta forma cuadrática puede ser escrita en la forma matricial: $y' A y$, donde A es una matriz $n \times n$, $A = (a_{ij})_{n \times n}$.

En el entorno clásico, los parámetros de un modelo de regresión lineal se estiman minimizando la suma de los residuos al cuadrado:

$$\begin{aligned} \mathbf{b} &= (b_0, b_1, \dots, b_k) \\ &= \arg \min_{(\beta_0, \beta_1, \dots, \beta_k)} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki})]^2. \end{aligned}$$

Este residuo al cuadrado es en realidad una forma cuadrática. Por lo tanto, es importante discutir algunas propiedades generales de esta forma cuadrática que se usarán en las discusiones subsiguientes.

1.4 Matrices idempotentes

Definición 1.3 Una matriz simétrica $n \times n$, A es idempotente si $A^2 = A$.

Sea $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ un vector k -dimensional y A es una matriz $k \times k$. $\alpha' A \alpha$ es una forma cuadrática de $\alpha_1, \alpha_2, \dots, \alpha_k$. Cuando A es una matriz idempotente, la forma cuadrática correspondiente tiene propiedades particulares. La forma cuadrática con matrices idempotentes se usa ampliamente en el análisis de regresión lineal. Ahora discutimos las propiedades de la matriz idempotente.

Teorema 1.4 Sea $A_{n \times n}$ una matriz idempotente de rango p , entonces los valores propios de A son 0 o 1.

Demostración.- Sea λ_i y v_i el valor propio y el correspondiente vector propio normalizado de la matriz A , respectivamente. Entonces, tenemos $A v_i = \lambda_i v_i$, y $v_i' A v_i = \lambda_i v_i' v_i = \lambda_i$. Por otro lado, ya que $A^2 = A$, podemos escribir

$$\lambda_i = v_i' A v_i = v_i' A^2 v_i = v_i' A' v_i = (A v_i)' A v_i = (\lambda_i v_i)' (\lambda_i v_i) = \lambda_i^2.$$

Dado que, tenemos $\lambda_i(\lambda_i - 1) = 0$, se tiene $\lambda_i = 1$ o $\lambda_i = 0$. Esto completa la demostración. ■

Es fácil saber que p valores propios de A son 1 y $n - p$ valores propios de A son cero. Por lo tanto, el rango de una matriz idempotente A es la suma de sus valores propios distintos de cero.

Definición 1.4 Sea $A = (a_{ij})_{n \times n}$ una matriz $n \times n$, la traza de A se define como la suma de los elementos diagonales; es decir,

$$tr(A) = a_{11} + a_{22} + \dots + a_{nn}.$$

Si A es una matriz simétrica. Entonces, la suma de todos los elementos cuadrados de A puede ser expresado mediante $\text{tr}(A^2)$. Es decir, $\sum_{i,j} A_{ij}^2 = \text{tr}(A^2)$. Es fácil verificar que $\text{tr}(AB) = \text{tr}(BA)$ para cualesquieras dos matrices A y B . El siguiente teorema da la relación entre el rango de la matriz A y la traza de A cuando A es una matriz idempotente.

Teorema 1.5 Si A es una matriz idempotente, entonces $\text{tr}(A) = \text{rango}(A) = p$.

Demostración.- Si el rango de una matriz A idempotente $n \times n$ es p , entonces A tiene p valores propios de 1 y $n - p$ valores propios de 0. Por lo tanto, podemos escribir $\text{rango}(A) = \sum_{i=1}^n \lambda_i = p$. Ya que $A^2 = A$, los valores propios de la matriz idempotente A es 1 o 0. De la teoría de matrices existe una matriz ortogonal V tal que

$$VV = \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix}.$$

Así, tenemos

$$\text{tr}(V'AV) = \text{tr}(VV'A) = \text{tr}(A) = \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix} = p = \text{rango}(A).$$

Aquí, usamos el simple hecho: $\text{tr}(AB) = \text{tr}(BA)$ para cualesquier matrices $A_{n \times n}$ y $B_{n \times n}$. ■

Una forma cuadrática de un vector aleatorio $\mathbf{y}' = (y_1, y_2, \dots, y_n)$ se puede escribir en una forma matricial $\mathbf{y}'A\mathbf{y}$, donde A es una matriz $n \times n$. Es de interés encontrar la esperanza y la varianza de $\mathbf{y}'A\mathbf{y}$. El siguiente teorema da el valor esperado de $\mathbf{y}'A\mathbf{y}$ cuando los componentes de \mathbf{y} son independientes.

Teorema 1.6 Sea $\mathbf{y}' = (y_1, y_2, \dots, y_n)$ un vector aleatorio $n \times 1$ con media $\mu' = (\mu_1, \mu_2, \dots, \mu_n)$ y varianza σ^2 para cada componente. Además, se supone que y_1, y_2, \dots, y_n son independientes. Sea A una matriz $n \times n$, $\mathbf{y}'A\mathbf{y}$ es una forma cuadrática de variables aleatorias. La esperanza de esta forma cuadrática viene dada por

$$E(\mathbf{y}'A\mathbf{y}) = \sigma^2 \text{tr}(A) + \mu'A\mu.$$

Demostración.- Primero observemos que

$$\mathbf{y}'A\mathbf{y} = (\mathbf{y} - \mu)'A(\mathbf{y} - \mu) + 2\mu'A(\mathbf{y} - \mu) + \mu'A\mu.$$

De donde podemos escribir,

$$\begin{aligned} E(\mathbf{y}'A\mathbf{y}) &= E[(\mathbf{y} - \mu)'A(\mathbf{y} - \mu)] + 2E[\mu'A(\mathbf{y} - \mu)] + \mu'A\mu \\ &= E\left[\sum_{i,j=1}^n a_{ij}(y_i - \mu_i)(y_j - \mu_j)\right] + 2\mu'A E(\mathbf{y} - \mu) + \mu'A\mu \\ &= \sum_{i=1}^n a_{ii}E(y_i - \mu_i)^2 + \mu'A\mu \\ &= \sigma^2 \text{tr}(A) + \mu'A\mu. \end{aligned}$$

■

Ahora veremos la varianza de la forma cuadrática de $\mathbf{y}'A\mathbf{y}$.

Teorema 1.7 Sea \mathbf{y} un vector aleatorio $n \times 1$ con media $\mu' = (\mu_1, \mu_2, \dots, \mu_n)$ y varianza σ^2 para cada componente. Se asume que y_1, y_2, \dots, y_n son independientes. Sea A una matriz simétrica $n \times n$, $E(y_i - \mu_i)^4 = u_i^{(4)}$,

$E(y_i - \mu_i)^3 = u_i^{(3)}$, y $a' = (a_{11}, a_{22}, \dots, x_{nn})$. La varianza de la forma cuadrática $\mathbf{y}A\mathbf{y}$ es dada por

$$\begin{aligned}\text{Var}(\mathbf{y}'A\mathbf{y}) &= \left(\mu^{(4)} - 3\sigma^2\right)a'a + \sigma^4 \left(2\text{tr}(A^2) + [\text{tr}(A)]^2\right) \\ &= 4\sigma^2\mu'A^2\mu + 4\mu^{(3)}a'A\mu.\end{aligned}$$

Demostración.- Sea $Z = \mathbf{y} - \mu$, $A(A_1, A_2, \dots, A_n)$ y $b = (b_1, b_2, \dots, b_n) = \mu'(A_1, A_2, \dots, A_n) = \mu'A$, de donde podemos escribir

$$\begin{aligned}\mathbf{y}'A\mathbf{y} &= (\mathbf{y} - \mu)A(\mathbf{y} - \mu) + 2\mu'A(\mathbf{y} - \mu) + \mu'A\mu \\ &= Z'AZ + 2b'Z + \mu'A\mu\end{aligned}$$

Así,

$$\text{Var}(\mathbf{y}'A\mathbf{y}) = \text{Var}(Z'AZ) + 4\text{Var}(bZ) + 4\text{Cov}(Z'AZ, bZ)$$

Calculando cada termino por separado, se tiene

$$\begin{aligned}(ZA'Z)^2 &= \sum_i \sum_j a_{ij}a_{lm}Z_iZ_jZ_lZ_m \\ E(ZA'Z)^2 &= \sum_i \sum_j \sum_l \sum_m a_{ij}a_{lm} E(Z_iZ_jZ_lZ_m)\end{aligned}$$

Notemos que

$$E(Z_iZ_jZ_lZ_m) = \begin{cases} \mu^{(4)}, & \text{si } i = j = l = m; \\ \sigma^4, & \text{si } i = j, l = k \text{ o } i = l, j = k, \text{ o } i = k, j = l; \\ 0, & \text{los demás.} \end{cases}$$

De donde se tiene

$$\begin{aligned}E(Z'AZ)^2 &= \sum_i \sum_j \sum_l \sum_m a_{ij}a_{lm} E(Z_iZ_jZ_lZ_m) \\ &= \mu^{(4)} \sum_{i=1}^n a_i^2 + \sigma^4 \left(\sum_{i \neq j} a_{ii}a_{kk} + \sum_{i \neq j} a_{ij}^2 + \sum_{i \neq j} a_{ij}a_{ji} \right).\end{aligned}$$

Ya que A es simétrica. $a_{ij} = a_{ji}$, tenemos

$$\begin{aligned}\sum a_{ij}^2 + \sum_{i \neq j} a_{ij}a_{ji} &= 2 \sum_{i \neq j} a_{ij}^2 \\ &= 2 \sum_{i,j} a_{ij}^2 - 2 \sum_{i=j} a_{ij}^2 \\ &= 2\text{tr}(A^2) - 2 \sum_{i=1}^n a_{ii}^2 \\ &= 2\text{tr}(A^2) - 2a'a.\end{aligned}$$

Y

$$\begin{aligned}
\sum_{i \neq k} a_{ii} a_{kk} &= \sum_{i,k} a_{ii} a_{kk} - \sum_{i=k} a_{ii} a_{kk} \\
&= [\text{tr}(A)]^2 - \sum_{i=1}^n a_{ii}^2 \\
&= [\text{tr}(A)]^2 - a' a.
\end{aligned}$$

Luego,

$$E(Z'AZ)^2 = (\mu^{(4)} - 3\sigma^4) a' a + \sigma^4 \{2\text{tr}(A^2) + [\text{tr}(A)]^2\}.$$

Para $\text{Var}(bZ)$, entonces

$$\text{Var}(bZ) = b \text{Var}(Z) b' = b b' \sigma^2 = (\mu' A) (\mu' A) \sigma^2 = \mu' A^2 \mu \sigma^2.$$

Para calcular $\text{Cov}(Z'AZ, bZ)$, notemos que $E(Z) = 0$, por lo que

$$\begin{aligned}
\text{Cov}(Z'AZ, bZ) &= \text{Cov}\left(\sum_{i,j} a_{ij} Z_i Z_j, \sum_k b_k Z_k\right) \\
&= \sum_{i,j,k} \text{Cov}(Z_i Z_j, Z_k) \\
&= \sum_{i,j,k} a_{ij} b_k E\{[Z_i Z_j - E(Z_i Z_j)] Z_k\} \\
&= \sum_{i,j,k} a_{ij} b_k [E(Z_i Z_j Z_k) - E(Z_k)] \\
&= \sum_{i,j,k} a_{ij} b_k [E(Z_i Z_j Z_k)] \quad (\text{ya que } E(Z_k) = 0).
\end{aligned}$$

Es fácil saber que

$$E(Z_i Z_j Z_k) = \begin{cases} \mu^{(3)}, & \text{si } i = j = k; \\ 0, & \text{los demás.} \end{cases}$$

Por lo tanto,

$$\begin{aligned}
\text{Cov}(Z'AZ, bZ) &= \sum_{i=1}^n a_{ii} b_i \mu^{(3)} \\
&= \mu^{(4)} \sum_{i=1}^n a_{ii} \mu' A_i \mu^{(3)} \\
&= \sum_{i=1}^n a_{ii} A_i' \mu \mu^{(3)} \\
&= a' A \mu \mu^{(3)}
\end{aligned}$$

La combinación de los resultados anteriores completa la demostración. ■

1.5 Distribución normal multivariante