Geometría de la Muestra y Muestreo Aleatorio

Varianzas Generalizadas $\mathbf{S}$ y $\mathbf{R}$
Varianza Total de la Muestra

Santiago Alférez

Agosto de 2020

MACC
Universidad del Rosario

Contenidos

¿Cuando la Varianza Muestral Generalizada puede ser Cero?

Varianza muestral generalizada cero

Situación cuando la varianza muestral generalizada es cero

La varianza generalizada es cero cuando, y sólo cuando, al menos un vector desviación yace en el hiper-plano formado por todas las combinaciones lineales de los restantes vectores. Es decir, cuando las columnas de la matriz de desviaciones son linealmente dependientes.

Es decir que, al menos una de las columnas de la matriz de desviaciones:

$$\begin{bmatrix} \mathbf{x}_1' - \overline{\mathbf{x}}' \\ \mathbf{x}_2' - \overline{\mathbf{x}}' \\ \vdots \\ \mathbf{x}_n' - \overline{\mathbf{x}}' \end{bmatrix} = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1p} - \bar{x}_p \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2p} - \bar{x}_p \\ \vdots & & \vdots & \ddots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{np} - \bar{x}_p \end{bmatrix}$$

$$= \underset{(n\times p)}{\mathbf{X}} - \underset{(n\times 1)}{\mathbf{1}} (1 \times p)$$

es una combinación lineal de las otras columnas. Este es el caso cuando, por ejemplo, uno de los vectores de desviación $\mathbf{d}_i' = [x_{1i} - \bar{x}_i, \ldots, x_{ni} - \bar{x}_i]$ se encuentra en el hiper-plano generado por $\mathbf{d}_1, \ldots, \mathbf{d}_{i-1}, \mathbf{d}_{i+1}, \ldots, \mathbf{d}_p$.

Varianza muestral generalizada cero

Ejemplo

Muestre que $|S| = 0$ para

$$\underset{(3\times 3)}{\mathbf{X}} = \begin{bmatrix} 1 & 2 & 5 \\ 4 & 1 & 6 \\ 4 & 0 & 4 \end{bmatrix}$$

Varianza muestral generalizada cero

Ejercicio

La siguiente matrix de datos contiene información sobre algunos scores de prueba, con $x_1 =$ score de la primera prueba, $x_2 =$ score de la segunda prueba y $x_3 =$ score total sobre las dos pruebas:

$$\mathbf{X} = \begin{bmatrix} 12 & 17 & 29 \\ 18 & 20 & 38 \\ 14 & 16 & 30 \\ 20 & 18 & 38 \\ 16 & 19 & 35 \end{bmatrix}$$

a Obtenga la matriz de datos corregida por la media y verifique que las columnas sean linealmente dependientes. Especifique un vector $\mathbf{a}' = [a_1, a_2, a_3]$ que establezca la dependencia lineal.

b Obtenga la matriz de covarianza muestral $\mathbf{S}$ y verifique que la varianza generalizada sea cero. Además, muestre que $\mathbf{Sa} = \mathbf{0}$, por lo que a se puede reescalar para que sea un vector propio correspondiente al valor propio cero.

c Verifique que la tercera columna de la matriz de datos sea la suma de las dos primeras columnas. Es decir, demuestre que hay dependencia lineal, con $a_1 = 1, a_2 = 1$ y $a_3 = -1$.

Varianza muestral generalizada cero

¿Qué sucede cuando se crea una variable que es combinación lineal de otras?

- Una matriz de covarianza es singular, por ejemplo, cuando los datos son puntajes de pruebas y el investigador ha incluido variables que son sumas de las demás.
- Por ejemplo, se pueden sumar las puntuaciones de los exámenes de mitad de período y finales de la clase para obtener la puntuación total. En otro caso, incluir el peso total de varios productos químicos junto con el de cada componente.
- Esta práctica de crear nuevas variables que son sumas de las variables originales y luego incluirlas en el conjunto de datos es una enorme pérdida de tiempo y hay la necesidad de estar alerta para evitar estas consecuencias.

Varianza muestral generalizada cero

¿Qué sucede cuando se crea una variable que es combinación lineal de otras?

Siempre que un vector $\mathbf{a}$, diferente de cero, satisfaga una de las siguientes tres condiciones, entonces satisface el resto de condiciones:

1. $\mathbf{Sa} = \mathbf{0}$: $\mathbf{a}$ es un vector propio (escalado) de $\mathbf{S}$ con valor propio 0.
2. $\mathbf{a}'\left(\mathbf{x}_j - \overline{\mathbf{x}}\right) = 0$ para todo j: La combinación lineal de los datos corregidos por la media, usando a, es cero.
3. $\mathbf{x}_j = c$ para todo j $j\,(c = \mathbf{a}'\overline{\mathbf{x}})$: La combinación lineal de los datos originales, usando $\mathbf{a}$, es una constante.

Varianza muestral generalizada cero

¿Qué sucede cuando se crea una variable que es combinación lineal de otras?

- $|\mathbf{S}| = 0$ significa que las mediciones de algunas variables deben eliminarse del estudio en lo que respecta a los cálculos matemáticos.
- La correspondiente matriz de datos reducida conducirá entonces a una matriz de covarianza de rango completo y una varianza generalizada distinta de cero.
- La pregunta de qué variables eliminar en casos degenerados no es fácil de responder. Cuando hay una opción, se deben retener las mediciones de una variable (presuntamente) causal en lugar de las de una variable secundaria.

Varianza muestral generalizada cero

Condición para que S tenga rango reducido

Si $n \leq p$, es decir, (tamaño de la muestra) $\leq$ (número de variables), entonces $|\mathbf{S}| = 0$ para todas las muestras.

Condiciones para que S tenga rango completo y rango reducido

Sean los vectores de dimensión $p \times 1$ $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, donde $\mathbf{x}_j'$ es la j-ésima fila de la matriz de datos $\mathbf{X}$, realizaciones de los vectores aleatorios independientes $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$. Entonces

1. Si la combinación lineal $\mathbf{a}'\mathbf{X}_j$ tiene una varianza positiva para cada vector constante $\mathbf{a} \neq \mathbf{0}$, entonces, siempre que $p < n$, S tendrá rango completo con probabilidad 1 y $|\mathbf{S} > 0$
2. Si, con probabilidad $1, \mathbf{a}'\mathbf{X}_j$ es una constante (por ejemplo, c) para todos los j, entonces $|\mathbf{S}| = 0$.

Varianza Generalizada Determinada por $|\mathbf{R}|$

Varianza generalizada $|\mathrm{R}|$

Consideraciones

- La varianza de la muestra generalizada se ve afectada indebidamente por la variabilidad de las mediciones en una sola variable.
- Por ejemplo, suponga que s_{ii} es grande o muy pequeña. Entonces, geométricamente, el vector de desviación correspondiente $\mathbf{d}_i = (\mathbf{y}_i - \bar{x}_i\mathbf{1})$, será muy largo o muy corto, y afectará de forma importante el volumen.
- En consecuencia, a veces es útil escalar todos los vectores de desviación para que tengan la misma longitud.

Varianza generalizada $|\mathrm{R}|$

Varianza muestral generalizada de variables estandarizadas

- Escalar los vectores residuales equivale a reemplazar cada observación original x_{jk} por su valor estandarizado $(x_{jk} - \bar{x}_k)/\sqrt{s_{kk}}$.
- Así, la matriz de covarianza muestral de las variables estandarizadas es entonces $\mathbf{R}$, la matriz de correlación muestral de las variables originales.

$$\left(\begin{array}{c} \text{Varianza muestral generalizada} \\ \text{de las variables estandarizadas} \end{array} \right) = |\mathbf{R}|$$

Varianza generalizada $|\mathrm{R}|$

Interpretación geométrica

- Dado que todos los vectores resultantes $\left[(x_{1k}-\bar{x}_k)/\sqrt{s_{kk}},(x_{2k}-\bar{x}_k)/\sqrt{s_{kk}},\ldots,(x_{nk}-\bar{x}_k)/\sqrt{s_{kk}}\right]=(\mathbf{y}_k-\bar{x}_k\mathbf{1})'/\sqrt{s_{kk}}$ tienen una longitud $\sqrt{n-1}$, la varianza muestral generalizada de las variables estandarizadas será grande cuando estos vectores sean casi perpendiculares y serán pequeños cuando dos o más de esos vectores casi se encuentren en la misma dirección.
- Sabemos que el coseno del ángulo entre los vectores de desviación es el coeficiente de correlación, por lo tanto el coseno del ángulo θ_{ik} entre los vectores $(\mathbf{y}_i-\bar{x}_i\mathbf{1})/\sqrt{s_{ii}}$ y $(\mathbf{y}_k-\bar{x}_k\mathbf{1})/\sqrt{s_{kk}}$ es el coeficiente de correlación muestral r_{ik}.
- Por lo tanto, se puede afirmar que $|\mathbf{R}|$ es grande cuando todos los r_{ik} son casi cero y es pequeño cuando uno o más de los r_{ik} son casi $+1$ o -1.

Varianza generalizada $|\mathrm{R}|$

Vectores de desviación de las variables estándarizadas

$$\frac{(\mathbf{y}_i - \bar{x}_i\mathbf{1})}{\sqrt{s_{ii}}} = \begin{bmatrix} \frac{x_{1i}-\bar{x}_i}{\sqrt{s_{ii}}} \\ \frac{x_{2i}-\bar{x}_i}{\sqrt{s_{ii}}} \\ \vdots \\ \frac{x_{ni}-\bar{x}_i}{\sqrt{s_{ii}}} \end{bmatrix}, \quad i = 1, 2, \ldots, p$$

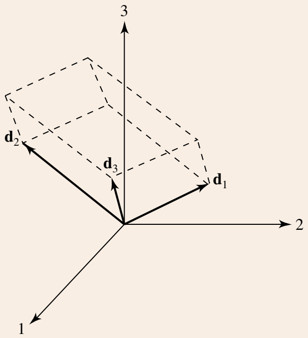
Interpretación geométrica

- Los i-ésimos vectores de desviación se encuentran en la dirección de $\mathbf{d}_i$, con una longitud al cuadrado de $n-1$.
- El volumen generado en el espacio p por los vectores de desviación se relaciona a la varianza muestral generalizada:

$$\left(\begin{array}{l}\text{Varianza de muestra generalizada} \\ \text{de las variables estandarizadas}\end{array}\right) = |\mathbf{R}| = (n-1)^{-p}(\text{volumen})^2$$

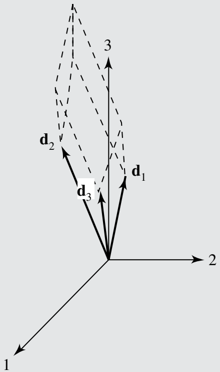
Varianza muestral generalizada $|S|$

Varianza generalizada grande



- La varianza generalizada aumenta si la longitud de cualquier $\mathbf{d}_i$ se incrementa.
- Se incrementa si los vectores residuales se mueven hasta que están a ángulos rectos unos de otros.

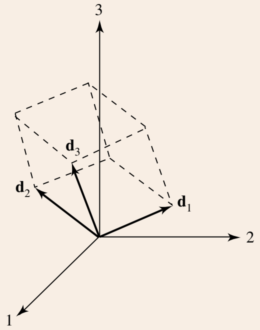
Varianza generalizada pequeña



- El volumen disminuye si uno de los $s_i i$ es pequeño o si uno de los vectores de desviación se encuentra cerca al (hiper) plano formado por los otros.

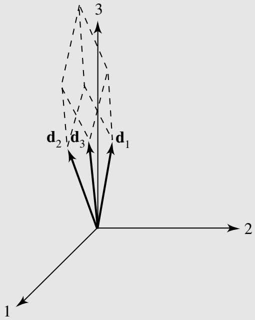
Varianza muestral generalizada $|\mathrm{R}|$

Varianza generalizada grande



- La varianza generalizada aumenta si la longitud de cualquier $\mathbf{d}_i$ se incrementa.
- Se incrementa si los vectores residuales se mueven hasta que están a ángulos rectos unos de otros.

Varianza generalizada pequeña



- El volumen disminuye si uno de los $s_i i$ es pequeño o si uno de los vectores de desviación se encuentra cerca al (hiper) plano formado por los otros.

Varianza generalizada

$|\mathbf{S}| = (n-1)^{-p}$ (volumen) ^{2} $\qquad$ $|\mathbf{R}| = (n-1)^{-p}(\text{volumen})^2$

Relación entre $|\mathbf{S}|$ y $|\mathbf{R}|$

Las cantidades $|\mathbf{S}|$ y $|\mathbf{R}|$ están conectadas por la relación,

$$|\mathbf{S}| = (s_{11}s_{22}\cdots s_{pp})\,|\mathbf{R}|$$

Entonces,

$$(n-1)^p|\mathbf{S}| = (n-1)^p\,(s_{11}s_{22}\cdots s_{pp})\,|\mathbf{R}|$$

Ejemplo

Para las siguiente matriz de covarianzas determine las varianzas muestrales generalizadas $|\mathbf{S}|$ y $|\mathbf{R}|$. Compruebe su relación mediante $|\mathbf{S}| = (s_{11}s_{22}\cdots s_{pp})\,|\mathbf{R}|$.

$$\underset{(3\times 3)}{\mathbf{S}} = \begin{bmatrix} 4 & 3 & 1 \\ 3 & 9 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

Varianza Total de la Muestra

Varianza generalizada

Varianza total de la muestra

La varianza total de la muestra se define como la suma de los elementos de la diagonal de la matriz de varianza-covarianza muestral $\mathbf{S}$:

$$\text{Varianza total de la muestra} = s_{11} + s_{22} + \cdots + s_{pp}$$

Varianza generalizada

Ejercicio

Con la siguiente matriz

$$\mathbf{X} = \begin{bmatrix} -1 & 3 & -2 \\ 2 & 4 & 2 \\ 5 & 2 & 3 \end{bmatrix}$$

a Calcule la matriz de desviaciones (residuos), $\mathbf{X} - \mathbf{1}\overline{\mathbf{x}}'$. ¿Es esta matriz de rango completo? Explique.

b Determine $\mathbf{S}$ y calcule la varianza muestral generalizada $|\mathbf{S}|$. Interprete este último geométricamente.

c Con los resultados de (b), calcule la varianza total de la muestra.

Media, Covarianza y Correlación como Operaciones Matriciales

Media y Covarianza como Operaciones Matriciales

$$\bar{x}_i = (x_{1i} \cdot 1 + x_{2i} \cdot 1 + \cdots + x_{ni} \cdot 1)/n = \mathbf{y}_i'\mathbf{1}/n$$

Media como operación matricial

$$\overline{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{y}_1'\mathbf{1}}{n} \\ \frac{\mathbf{y}_2'\mathbf{1}}{n} \\ \vdots \\ \frac{\mathbf{y}_p'\mathbf{1}}{n} \end{bmatrix} = \frac{1}{n} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

o,

$$\overline{\mathbf{x}} = \frac{1}{n}\mathbf{X}'\mathbf{1}$$

Media y Covarianza como Operaciones Matriciales

$$\overline{\mathbf{x}} = \frac{1}{n}\mathbf{X}'\mathbf{1}$$

Media como operación matricial

Si trasponemos la expresión anterior y la multiplicamos por la izquierda por $\mathbf{1}$:

$$\mathbf{1}\overline{\mathbf{x}}' = \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X} = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_p \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_p \end{bmatrix}$$

Si restamos este resultado a $\mathbf{X}$ se produce la matriz de desviaciones de $n \times p$ (o residuos):

$$\mathbf{X} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X} = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1p} - \bar{x}_p \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2p} - \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{np} - \bar{x}_p \end{bmatrix}$$

$$\mathbf{X} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X} = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1p} - \bar{x}_p \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2p} - \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{np} - \bar{x}_p \end{bmatrix}$$

Covarianza como operación matricial

La matriz anterior transpuesta multiplicada por ella misma representa la suma de cuadrados y productos cruzados que es igual a $(n-1)\mathbf{S}$:

$$(n-1)\mathbf{S} = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{21} - \bar{x}_1 & \cdots & x_{n1} - \bar{x}_1 \\ x_{12} - \bar{x}_2 & x_{22} - \bar{x}_2 & \cdots & x_{n2} - \bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} - \bar{x}_p & x_{2p} - \bar{x}_p & \cdots & x_{np} - \bar{x}_p \end{bmatrix}$$

$$\times \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1p} - \bar{x}_p \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2p} - \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{np} - \bar{x}_p \end{bmatrix}$$

Media, y covarianza como operaciones matriciales

$$(n-1)\mathbf{S} = (\mathbf{X} - \tfrac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X})'(\mathbf{X} - \tfrac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X}) =$$

$$\begin{bmatrix} x_{11}-\bar{x}_1 & x_{21}-\bar{x}_1 & \cdots & x_{n1}-\bar{x}_1 \\ x_{12}-\bar{x}_2 & x_{22}-\bar{x}_2 & \cdots & x_{n2}-\bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p}-\bar{x}_p & x_{2p}-\bar{x}_p & \cdots & x_{np}-\bar{x}_p \end{bmatrix} \times \begin{bmatrix} x_{11}-\bar{x}_1 & x_{12}-\bar{x}_2 & \cdots & x_{1p}-\bar{x}_p \\ x_{21}-\bar{x}_1 & x_{22}-\bar{x}_2 & \cdots & x_{2p}-\bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1}-\bar{x}_1 & x_{n2}-\bar{x}_2 & \cdots & x_{np}-\bar{x}_p \end{bmatrix}$$

Covarianza como operación matricial

Dado que

$$\left(\mathbf{I} - \tfrac{1}{n}\mathbf{1}\mathbf{1}'\right)'\left(\mathbf{I} - \tfrac{1}{n}\mathbf{1}\mathbf{1}'\right) = \mathbf{I} - \tfrac{1}{n}\mathbf{1}\mathbf{1}' - \tfrac{1}{n}\mathbf{1}\mathbf{1}' + \tfrac{1}{n^2}\mathbf{1}\mathbf{1}'\mathbf{1}\mathbf{1}' = \mathbf{I} - \tfrac{1}{n}\mathbf{1}\mathbf{1}',$$

la expresión anterior (bloque azul) se transforma en:

$$(n-1)\mathbf{S} = \left(\mathbf{X} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X}\right)'\left(\mathbf{X} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X}\right) = \mathbf{X}'\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{X}$$

Media, Covarianza y Correlación como Operaciones Matriciales

Media y covarianzas como operaciones matriciales

$$\overline{\mathbf{x}} = \frac{1}{n}\mathbf{X}'\mathbf{1}$$
$$\mathbf{S} = \frac{1}{n-1}\mathbf{X}'\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{X}$$

- El resultado para $\mathbf{S}_n$ es similiar, excepto que $1/n$ reemplaza a $1/(n-1)$ como el primer factor.
- Las expresiones anteriores muestran como las operaciones matriciales sobre la matriz de datos $\mathbf{X}$ llevan a $\overline{\mathbf{x}}$ y $\mathbf{S}$.

Media, Covarianza y Correlación como Operaciones Matriciales

Matriz de desviación estándar muestral

$$\underset{(p\times p)}{\mathbf{D}^{1/2}} = \begin{bmatrix} \sqrt{s_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{s_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{s_{pp}} \end{bmatrix} \qquad \underset{(p\times p)}{\mathbf{D}^{-1/2}} = \begin{bmatrix} \frac{1}{\sqrt{s_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{s_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{s_{pp}}} \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} \frac{s_{11}}{\sqrt{s_{11}}\sqrt{s_{11}}} & \frac{s_{12}}{\sqrt{s_{11}}\sqrt{s_{22}}} & \cdots & \frac{s_{1p}}{\sqrt{s_{11}}\sqrt{s_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{s_{1p}}{\sqrt{s_{11}}\sqrt{s_{pp}}} & \frac{s_{2p}}{\sqrt{s_{22}}\sqrt{s_{pp}}} & \cdots & \frac{s_{pp}}{\sqrt{s_{pp}}\sqrt{s_{pp}}} \end{bmatrix} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & 1 \end{bmatrix}$$

Entonces,

$$\mathbf{R} = \mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2}$$

Multiplicando la expresión anterior a la izquierda y a la derecha por $\mathbf{D}^{1/2}$ y dado que $\mathbf{D}^{-1/2}\mathbf{D}^{1/2} = \mathbf{D}^{1/2}\mathbf{D}^{-1/2} = \mathbf{I}$:

$$\mathbf{S} = \mathbf{D}^{1/2}\mathbf{R}\mathbf{D}^{1/2}$$

Medidas de la Muestra para Combinaciones Lineales de Variables

Medidas de la muestra para combinaciones lineales de variables

Una combinación lineal de p variables

$$\mathbf{c}'\mathbf{X} = c_1 X_1 + c_2 X_2 + \cdots + c_p X_p$$

cuyo valor observado en el j-ésimo ítem es:

$$\mathbf{c}'\mathbf{x}_j = c_1 x_{j1} + c_2 x_{j2} + \cdots + c_p x_{jp}, \quad j = 1, 2, \ldots, n$$

La media muestral para n observaciones

$$\begin{aligned} \text{Media muestral} &= \frac{(\mathbf{c}'\mathbf{x}_1 + \mathbf{c}'\mathbf{x}_2 + \cdots + \mathbf{c}'\mathbf{x}_n)}{n} \\ &= \mathbf{c}'\left(\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_n\right)\frac{1}{n} = \mathbf{c}'\overline{\mathbf{x}} \end{aligned}$$

Medidas de la muestra para combinaciones lineales de variables

Varianza muestral para n observaciones de una combinación lineal de p variables

Dado que $\left(\mathbf{c}'\mathbf{x}_j - \mathbf{c}'\overline{\mathbf{x}}\right)^2 = \left(\mathbf{c}'\left(\mathbf{x}_j - \overline{\mathbf{x}}\right)\right)^2 = \mathbf{c}'\left(\mathbf{x}_j - \overline{\mathbf{x}}\right)\left(\mathbf{x}_j - \overline{\mathbf{x}}\right)'\mathbf{c}$:

$$\text{Varianza muestral } = \frac{\left(\mathbf{c}'\mathbf{x}_1 - \mathbf{c}'\overline{\mathbf{x}}\right)^2 + \left(\mathbf{c}'\mathbf{x}_2 - \mathbf{c}'\overline{\mathbf{x}}\right)^2 + \cdots + \left(\mathbf{c}'\mathbf{x}_n - \mathbf{c}'\overline{\mathbf{x}}\right)^2}{n-1}$$

$$= \frac{\mathbf{c}'\left(\mathbf{x}_1 - \overline{\mathbf{x}}\right)\left(\mathbf{x}_1 - \overline{\mathbf{x}}\right)'\mathbf{c} + \mathbf{c}'\left(\mathbf{x}_2 - \overline{\mathbf{x}}\right)\left(\mathbf{x}_2 - \overline{\mathbf{x}}\right)'\mathbf{c} + \cdots + \mathbf{c}'\left(\mathbf{x}_n - \overline{\mathbf{x}}\right)\left(\mathbf{x}_n - \overline{\mathbf{x}}\right)'\mathbf{c}}{n-1}$$

$$= \mathbf{c}'\left[\frac{\left(\mathbf{x}_1 - \overline{\mathbf{x}}\right)\left(\mathbf{x}_1 - \overline{\mathbf{x}}\right)' + \left(\mathbf{x}_2 - \overline{\mathbf{x}}\right)\left(\mathbf{x}_2 - \overline{\mathbf{x}}\right)' + \cdots + \left(\mathbf{x}_n - \overline{\mathbf{x}}\right)\left(\mathbf{x}_n - \overline{\mathbf{x}}\right)'}{n-1}\right]\mathbf{c}$$

Varianza muestral de $\mathbf{c}'\mathbf{X} = \mathbf{c}'\mathbf{S}\mathbf{c}$

Medidas de la muestra para combinaciones lineales de variables

Una segunda combinación lineal de p variables

$$\mathbf{b}'\mathbf{X} = b_1 X_1 + b_2 X_2 + \cdots + b_p X_p$$

cuyo valor observado en el j-ésimo ítem es:

$$\mathbf{b}'\mathbf{x}_j = b_1 x_{j1} + b_2 x_{j2} + \cdots + b_p x_{jp}, \quad j = 1, 2, \ldots, n$$

La media y varianza muestrales para una segunda combinación lineal

Media muestral de $\mathbf{b}'\mathbf{X} = \mathbf{b}'\overline{\mathbf{x}}$

Varianza muestral de $\mathbf{b}'\mathbf{X} = \mathbf{b}'\mathbf{S}\mathbf{b}$

Medidas de la muestra para combinaciones lineales de variables

Covarianza muestral de pares de observaciones obtenidas mediante $\mathbf{b}'\mathbf{X}$ y $\mathbf{c}'\mathbf{X}$

Covarianza muestral

$$=\frac{(\mathbf{b}'\mathbf{x}_1-\mathbf{b}'\overline{\mathbf{x}})(\mathbf{c}'\mathbf{x}_1-\mathbf{c}'\overline{\mathbf{x}})+(\mathbf{b}'\mathbf{x}_2-\mathbf{b}'\overline{\mathbf{x}})(\mathbf{c}'\mathbf{x}_2-\mathbf{c}'\overline{\mathbf{x}})+\cdots+(\mathbf{b}'\mathbf{x}_n-\mathbf{b}'\overline{\mathbf{x}})(\mathbf{c}'\mathbf{x}_n-\mathbf{c}'\overline{\mathbf{x}})}{n-1}$$

$$=\frac{\mathbf{b}'(\mathbf{x}_1-\overline{\mathbf{x}})(\mathbf{x}_1-\overline{\mathbf{x}})'\mathbf{c}+\mathbf{b}'(\mathbf{x}_2-\overline{\mathbf{x}})(\mathbf{x}_2-\overline{\mathbf{x}})'\mathbf{c}+\cdots+\mathbf{b}'(\mathbf{x}_n-\overline{\mathbf{x}})(\mathbf{x}_n-\overline{\mathbf{x}})'\mathbf{c}}{n-1}$$

$$=\mathbf{b}'\left[\frac{(\mathbf{x}_1-\overline{\mathbf{x}})(\mathbf{x}_1-\overline{\mathbf{x}})'+(\mathbf{x}_2-\overline{\mathbf{x}})(\mathbf{x}_2-\overline{\mathbf{x}})'+\cdots+(\mathbf{x}_n-\overline{\mathbf{x}})(\mathbf{x}_n-\overline{\mathbf{x}})'}{n-1}\right]\mathbf{c}$$

Covarianza muestral de $\mathbf{b}'\mathbf{X}$ y $\mathbf{c}'\mathbf{X}=\mathbf{b}'\mathbf{S}\mathbf{c}$

Resumen de medidas muestrales de combinación lineal de variables

Las combinaciones lineales

$$\mathbf{b}'\mathbf{X} = b_1X_1 + b_2X_2 + \cdots + b_pX_p$$
$$\mathbf{c}'\mathbf{X} = c_1X_1 + c_2X_2 + \cdots + c_pX_p$$

tienen medias, varianzas y covarianzas muestrales que están relacionadas a $\overline{\mathbf{x}}$ y $\mathbf{S}$ mediante:

$$\text{Media muestral de } \mathbf{b}'\mathbf{X} = \mathbf{b}'\overline{\mathbf{x}}$$
$$\text{Media muestral de } \mathbf{c}'\mathbf{X} = \mathbf{c}'\overline{\mathbf{x}}$$
$$\text{Varianza muestral de } \mathbf{b}'\mathbf{X} = \mathbf{b}'\mathbf{S}\mathbf{b}$$
$$\text{Varianza muestral de } \mathbf{c}'\mathbf{X} = \mathbf{c}'\mathbf{S}\mathbf{c}$$
$$\text{Covarianza muestral de } \mathbf{b}'\mathbf{X} \text{ y } \mathbf{c}'\mathbf{X} = \mathbf{b}'\mathbf{S}\mathbf{c}$$

Medidas de la muestra para combinaciones lineales de variables

q combinaciones lineales

Consideremos q combinaciones lineales

$$a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ip}X_p, \quad i = 1, 2, \ldots, q$$

$$\begin{bmatrix} a_{11}X_1 & + & a_{12}X_2 & + \cdots + & a_{1p}X_p \\ a_{21}X_1 & + & a_{22}X_2 & + \cdots + & a_{2p}X_p \\ \vdots & & \vdots & \vdots & \vdots \\ a_{q1}X_1 & + & a_{q2}X_2 & + \cdots + & a_{qp}X_p \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{q1} & a_{q2} & \cdots & a_{qp} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \mathbf{AX}$$

Media y Covarianza de q combinaciones lineales

Las q combinaciones lineales $\mathbf{AX}$ en la expresión anterior tienen vector media $\mathbf{A\bar{x}}$ y covarianza muestral $\mathbf{ASA}'$