Análise Multivariada
Revisão de Álgebra de Matrizes

Prof. George von Borries
Departamento de Estatística
Universidade de Brasília

2023

Notação

- Seja x_{ij} a medida da j-ésima variável ($j = 1, \ldots, p$) referente a i-ésima observação ($i = 1, \ldots, n$) ou unidade de análise.
- Podemos pensar em qualquer base multivariada como uma matriz $\mathbf{X}_{n \times p}$ ou $\underset{\sim}{X}$, isto é.

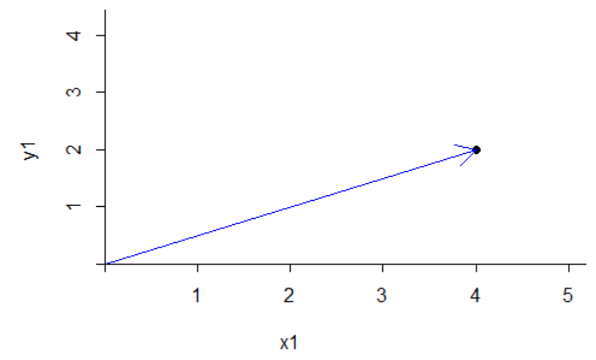
$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1j} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2j} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \ldots & x_{ij} & \ldots & x_{ip} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nj} & \ldots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \\ \vdots \\ \mathbf{X}_i^T \\ \vdots \\ \mathbf{X}_n^T \end{pmatrix}$$

$$= (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_j, \ldots, \mathbf{X}_p)$$

- Forma matemática simplificada e condensada.
- Permite realizar operações com muitas variáveis simultaneamente.
- **Nota:** A matriz será formada por valores númericos. Cuidado quando os números representarem categorias de uma variável!

Vetores
Notação e Definições Básicas

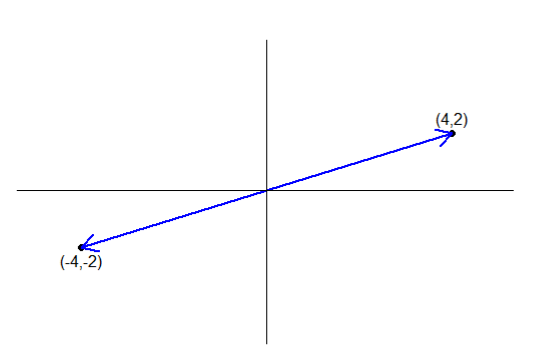
- Um vetor com p elementos identifica um ponto no espaço p-dimensional.
- Todo vetor tem comprimento e direção, i.e., $\mathbf{x}^{\top} = (x_1, x_2) = (4, 2)$



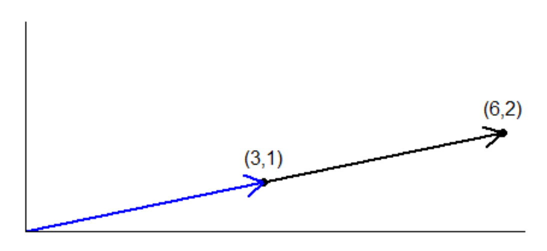
- Igualdade de vetores:

$$\mathbf{a} = \mathbf{b} \iff a_i = b_i \quad \forall \quad i.$$

- (ou) Dois vetores são idênticos se tem a mesma direção e comprimento.
- Multiplicação de um vetor por -1:

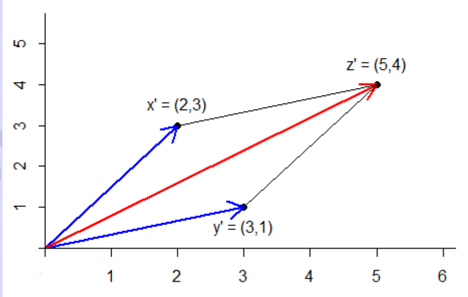


- Suponha que c é um escalar. Então $c\mathbf{x}^\top = (cx_1, \ldots, cx_k)$. Exemplo, seja $c = 2$ e $\mathbf{x}^\top = (3, 1)$,



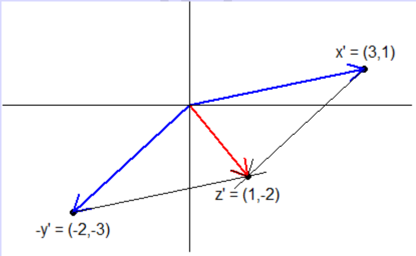
- A soma de dois vetores, cada um com o mesmo número de elementos, é feita elemento por elemento,

$$\mathbf{z} = \mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$



- A subtração de dois vetores, cada um com o mesmo número de elementos, é feita elemento por elemento,

$$\mathbf{z} = \mathbf{x} - \mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} - \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{pmatrix}$$



Definição

Produto Interno *(inner product)* $\langle \, , \rangle$ *é uma função que transforma um par de vetores de um espaço vetorial em um número real tal que: para os vetores* $\mathbf{x}, \mathbf{y}, \mathbf{w}$ *num espaço vetorial* $\mathbf{V}$ *e um escalar* c,

- $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$,
- $\langle \mathbf{x} + \mathbf{y}, \mathbf{w} \rangle = \langle \mathbf{x}, \mathbf{w} \rangle + \langle \mathbf{y}, \mathbf{w} \rangle$,
- $\langle c\mathbf{x}, \mathbf{y} \rangle = c\langle \mathbf{x}, \mathbf{y} \rangle$,
- $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0 \;\; \forall \;\; \mathbf{x}$ *em* $\mathbf{V}$ *e*
- $\langle \mathbf{x}, \mathbf{x} \rangle = 0 \iff \mathbf{x} = \mathbf{0}$.

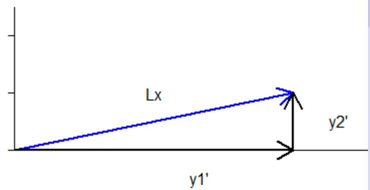
Nota: Seja $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ e $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ quaisquer dois vetores n-dimensionais finitos. O produto interno de $\mathbf{x}$ e $\mathbf{y}$ é o escalar

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n = \sum_{i=1}^{n} x_i y_i .$$

e este produto interno ($\mathbf{V} = \Re^n$) é também chamado de produto escalar (*dot product*).

Comprimento de um vetor: Lembre do Teorema de Pitágoras.

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ x_2 \end{pmatrix} = \mathbf{y}_1 + \mathbf{y}_2$$



$$L_x = \sqrt{y_1^2 + y_2^2} \quad \text{ou} \quad ||\mathbf{x}|| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}.$$

- A extensão para o espaço n dimensional é natural: $L_x = \sqrt{\sum_{j=1}^{n} y_j^2}$.
- Note que $c\mathbf{x} \Rightarrow |c| \cdot ||\mathbf{x}||$.
- If $c = -1$ a direção muda, mas não o comprimento.
- Um vetor de comprimento unitário é dito **vetor normalizado**.

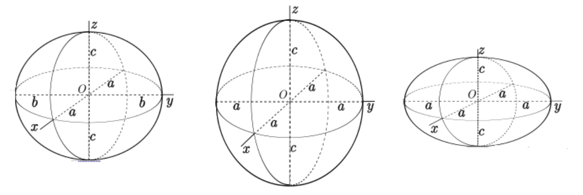
Distância:

- A distância entre dois pontos arbitrários é

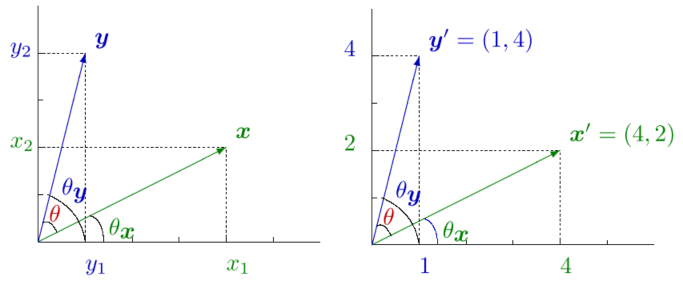
  $P = (x_1, \ldots, x_n)$ e $Q = (y_1, \ldots, y_n)$ é dada por L_{x-y}, i.e.,

  $$d(P, Q) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2}$$

- Note que $||\mathbf{x}|| = d(0, P)$, sendo $P = (x_1, \ldots, x_n)$.
- $d^2(\mathbf{0}, P) = x_1^2 + x_2^2 + \ldots + x_n^2 = c^2$ é a equação de uma esfera no espaço n-dimensional. Espaço Euclidiano não ponderado.
- A distância estatística é ponderada pelo desvio-padrão. Isto gera um elipsoide no espaço n-dimensional (voltaremos a este assunto).



Ângulo entre vetores



Lembre que:

$$\begin{aligned} \cos(\theta) &= \cos(\theta_y - \theta_x) \\ &= \cos(\theta_y)\cos(\theta_x) + \sin(\theta_y)\sin(\theta_x) \\ &= \left(\frac{y_1}{||\mathbf{y}||}\right)\left(\frac{x_1}{||\mathbf{x}||}\right) + \left(\frac{y_2}{||\mathbf{y}||}\right)\left(\frac{x_2}{||\mathbf{x}||}\right) \end{aligned}$$

Produto Interno

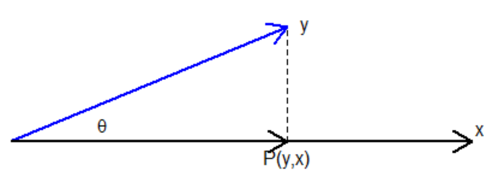
Logo,

$$\cos(\theta) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{||\mathbf{x}|| \cdot ||\mathbf{y}||}$$

Verifique que $\theta = 49.40^o$ acima.

- $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y}$
- $\theta = 90^o \Rightarrow cos(\theta) = 0$ $\iff$ vetores são perpendiculares ($\mathbf{x}^T\mathbf{y} = 0$).
- Podemos generalizar facilmente para n dimensões.

Projeção de um vetor y em x:



$$P(\mathbf{y},\mathbf{x}) = \frac{\langle \mathbf{y},\mathbf{x} \rangle}{||\mathbf{x}||} \frac{\mathbf{x}}{||\mathbf{x}||}$$

com comprimento

$$L_{P(\mathbf{y},\mathbf{x})} = \frac{\langle \mathbf{y},\mathbf{x} \rangle}{||\mathbf{x}||} \frac{1}{||\mathbf{x}||}\mathbf{x} = ||\mathbf{y}|| \frac{\langle \mathbf{y},\mathbf{x} \rangle}{||\mathbf{y}|| \cdot ||\mathbf{x}||} = ||\mathbf{y}|| \cdot |cos(\theta)|$$

- Qual a relação com estimação por mínimos quadrados num problema de regressão?

Matrizes
Notação e Definições Básicas

- Considere uma matriz $\mathbf{A}_{p \times p}$:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{ip} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pj} & \dots & a_{pp} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1^T \\ \mathbf{A}_2^T \\ \vdots \\ \mathbf{A}_j^T \\ \vdots \\ \mathbf{A}_p^T \end{pmatrix}$$

$$= (\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_j, \dots, \mathbf{A}_p)$$

- A matriz $\mathbf{A}$ é denominada **matriz quadrada**.

- A diagonal de uma matriz $\mathbf{A}_{p\times p}$ é o vetor formado por $(a_{11}, a_{22}, \ldots, a_{pp})$.
- Uma matriz quadrada ($\mathbf{D}$) com $a_{ij} = 0 \ \forall \ \ i \neq j$ é chamada de matriz diagonal. Exemplo no `R`: `diag(c(1,2,3),3,3)`.
- Matriz Identidade ($\mathbf{I}$) é uma matriz diagonal com 1 em cada elemento da diagonal. Exemplo no `R`: `diag(c(1,1,1),3,3)`.
- Uma matriz triangular superior (inferior) é uma matriz quadrada com zeros abaixo (acima) da diagonal.

  Exemplo no `R`: veja programa algmatex.R.
- Um vetor e uma matriz de zeros são representados por $\mathbf{o}$ e $\mathbf{O}$, respectivamente.

$$\mathbf{D} = \begin{pmatrix} a_{11} & 0 & \ldots & 0 \\ 0 & a_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & a_{pp} \end{pmatrix} \qquad \mathbf{o} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad \mathbf{O} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Nota: Alguns operadores do `R` podem ser vistos em `http://www.statmethods.net/advstats/matrix.html`.

- Igualdade de matrizes: $\mathbf{A} = \mathbf{B} \iff a_{ij} = b_{ij} \quad \forall \quad i, j$.
- Importante: $\mathbf{AB} = \mathbf{CB} \not\Rightarrow \mathbf{A} = \mathbf{C}$.
- Um vetor de $1's$ é representado por $\mathbf{j}$ e uma matriz quadrada de $1's$ é representada por $\mathbf{J}$.

$$\mathbf{j} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \qquad \mathbf{J} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

- $\mathbf{j}^{\mathsf{T}}\mathbf{j} = n$ e $\mathbf{j}\mathbf{j}^{\mathsf{T}} = \mathbf{J}_{n \times n}$, com $\mathbf{j}_{n \times 1}$.
- $diag(\mathbf{j}^{\mathsf{T}}) = \mathbf{I}$.
- $\mathbf{J}^2 = \mathbf{j}\mathbf{j}^{\mathsf{T}}\mathbf{j}\mathbf{j}^{\mathsf{T}} = n\mathbf{J}$.
- $\mathbf{J}_{n \times p}\, \mathbf{J}_{p \times n} = p\mathbf{J}_{n \times n}$ (considerando $\mathbf{J}$ qualquer matriz de 1's).
- $\mathbf{a}^{\mathsf{T}}\mathbf{j} = \mathbf{j}^{\mathsf{T}}\mathbf{a} = \sum_i a_i$ (escalar), com $\mathbf{a}_{n \times 1}$.
- $\mathbf{j}^{\mathsf{T}}\mathbf{A} \neq \mathbf{A}\mathbf{j}$, com $\mathbf{A}_{n \times p}$.

Operações com Matrizes

Suponha $\mathbf{A}_{n\times p}$ e $\mathbf{B}_{n\times p}$.

- $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$.
- $(\mathbf{A}\pm\mathbf{B})^{\top}=\mathbf{A}^{\top}\pm\mathbf{B}^{\top}$.

Suponha $\mathbf{A}_{n\times m}$ e $\mathbf{B}_{k\times p}$ (quando necessário suponha $m=k$).

- $\mathbf{C}=\mathbf{AB}$ está definido se $m=k$. Neste caso dizemos que $\mathbf{A}$ e $\mathbf{B}$ são conformes e $c_{ij}=\sum_{h=1}^{m}a_{ih}b_{hj}$.
- Em geral, $\mathbf{AB}\neq\mathbf{BA}$.
- $\mathbf{A}(\mathbf{B}\pm\mathbf{C})=\mathbf{AB}\pm\mathbf{AC}$.
- $(\mathbf{AB})^{\top}=\mathbf{B}^{\top}\mathbf{A}^{\top}$.
- $\mathbf{A}^2$ está definido somente se $\mathbf{A}$ quadrada, isto é, $n=m$.
- Em geral, $\mathbf{ABC}=\mathbf{A}(\mathbf{BC})=(\mathbf{AB})\mathbf{C}$.

✓ Se $\mathbf{X}_{n\times p}$ e $\mathbf{A}_{n\times n}$, então
$\mathbf{X}^{\top}\mathbf{X}-\mathbf{X}^{\top}\mathbf{AX}=\mathbf{X}^{\top}(\mathbf{X}-\mathbf{AX})=\mathbf{X}^{\top}(\mathbf{I}-\mathbf{A})\mathbf{X}$.

- $(\mathbf{AB})^\mathsf{T} = \mathbf{B}^\mathsf{T}\mathbf{A}$ e $(\mathbf{A}^\mathsf{T})^\mathsf{T} = \mathbf{A}$
- $\mathbf{a}^\mathsf{T}\mathbf{b} = \mathbf{b}^\mathsf{T}\mathbf{a} \Rightarrow (\mathbf{a}-\mathbf{b})^\mathsf{T}(\mathbf{a}-\mathbf{b}) = \mathbf{a}^\mathsf{T}\mathbf{a} - 2\mathbf{a}^\mathsf{T}\mathbf{b} + \mathbf{b}^\mathsf{T}\mathbf{b}$.
- Mas, $(\mathbf{A}-\mathbf{B})^\mathsf{T}(\mathbf{A}-\mathbf{B}) = \mathbf{A}^\mathsf{T}\mathbf{A} - \mathbf{A}^\mathsf{T}\mathbf{B} - \mathbf{B}^\mathsf{T}\mathbf{A} + \mathbf{B}^\mathsf{T}\mathbf{B}$.
- $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$.
- Se $\mathbf{A}$ simétrica, $\mathbf{y}^\mathsf{T}\mathbf{Ay} = \sum_i a_{ii}y_i^2 + \sum_{i\neq j} a_{ij}y_iy_j$
  é denominada de **Forma Quadrática** (um escalar).
- Ainda, $\mathbf{x}^\mathsf{T}\mathbf{Ay} = \sum_{ij} a_{ij}x_iy_j$
  é denominada de **Forma Bilinear** (um escalar).

Partição de Matrizes

$$\mathbf{A} = \left(\begin{array}{rrr|rr} 2 & 1 & 3 & 8 & 4 \\ -3 & 4 & 0 & 2 & 7 \\ 9 & 3 & 6 & 5 & -2 \\ \hline 4 & 8 & 3 & 1 & 6 \end{array}\right) = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

$$\mathbf{A}^{\mathsf{T}} = \left(\begin{array}{rrr|r} 2 & -3 & 9 & 4 \\ 1 & 4 & 3 & 8 \\ 3 & 0 & 6 & 3 \\ \hline 8 & 2 & 5 & 1 \\ 4 & 7 & -2 & 6 \end{array}\right) = \begin{pmatrix} \mathbf{A}_{11}^{\mathsf{T}} & \mathbf{A}_{21}^{\mathsf{T}} \\ \mathbf{A}_{12}^{\mathsf{T}} & \mathbf{A}_{22}^{\mathsf{T}} \end{pmatrix}$$

Partição de Matrizes

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \quad e \quad \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}$$

- Matriz Bloco Diagonal e Bloco Diagonal Superior:

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_{22} \end{pmatrix} \qquad \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{O} & \mathbf{A}_{22} \end{pmatrix}$$

$$\mathbf{A} \pm \mathbf{B} = \begin{pmatrix} \mathbf{A}_{11} \pm \mathbf{B}_{11} & \mathbf{A}_{12} \pm \mathbf{B}_{12} \\ \mathbf{A}_{21} \pm \mathbf{B}_{21} & \mathbf{A}_{22} \pm \mathbf{B}_{22} \end{pmatrix}$$

e

$$\mathbf{AB} = \begin{pmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{pmatrix}$$

(Neste caso as partições de **A** e **B** são chamadas conformes.)

Definição (Dependência e Independência linear)

Um conjunto de vetores é **linearmente dependente** *se existem constantes* $a_1, a_2, \ldots, a_n$ *tais que*

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \ldots + a_n\mathbf{x}_n = \mathbf{0}$$

com pelo menos um $a_i \neq 0, i = 1, \ldots, n$.
Caso contrário o conjunto de vetores é **linearmente independente**.

Definição (Posto ou Rank de uma Matriz)

É o número máximo de linhas (colunas) linearmente independentes. É o tamanho da maior submatriz de $\mathbf{A}$ *que tem determinante não nulo.*

Lembre que $|\mathbf{A}|_{3\times 3} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21} - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{12}a_{21}$.

Veja algumas propriedades de Determinantes em Rencher and Christensen[1].

[1] Rencher, A.C. e Christensen, W.F. *Methods of Multivariate Analysis*, 2012, Terceira Edição - Capítulo 2

Posto ou Rank de uma Matriz

- Uma matriz é de rank completo se $\text{rank}(\mathbf{A}) = \min(n, p)$ para $\mathbf{A}_{n\times p}$.
- Matriz de rank completo $\Rightarrow$ linhas e/ou colunas são linearmente independentes.
- No R: função `rankMatrix` do pacote `Matrix`.
- Se dependência linear:
  - $\mathbf{A}_{n\times p}$ e $\text{rank}(\mathbf{A}) < \min(n, p)$.
  - $\mathbf{Ac} = 0$ mesmo que $\mathbf{A} \neq \mathbf{0}$ e $\mathbf{c} \neq \mathbf{0}$.
  - Possibilidade de $\mathbf{AB} = \mathbf{CB}$ aonde $\mathbf{A} \neq \mathbf{C}$.
  - $|\mathbf{B}| \neq 0 \Rightarrow \text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{A})$.
- Por definição $\text{rank}(\mathbf{AB}) \leq \min(\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B}))$.

Note que se $\mathbf{A}_{n\times p}$ e $\text{rank}(\mathbf{A}) = n < p$, a matriz é dita linearmente independente, mas as colunas (variáveis) de $\mathbf{A}_{n\times p}$ são linearmente dependentes.

Inversa de Matrizes

- Uma matriz quadrada $\mathbf{A}$ de rank completo é chamada de **não singular** e tem inversa única $\mathbf{A}^{-1}$.

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

- No R: `solve(A)`.
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
- Se $\mathbf{B}$ não singular, $\mathbf{AB} = \mathbf{CB} \Rightarrow \mathbf{A} = \mathbf{C}$. ($\checkmark$)
- $\mathbf{A}^{\top -1} = (\mathbf{A}^{-1})^{\top}$.

Definição (Matriz Positiva Definida)

Uma matriz é chamada positiva definida (p.d.) se, e somente se

- $\mathbf{A} = \mathbf{A}^T$ *(simétrica)*
- $\mathbf{y}^T\mathbf{A}\mathbf{y} > 0$ *para todo vetor* $\mathbf{y} \in E_n$ *tal que* $\mathbf{y} \neq 0$.

Definição (Matriz Positiva Semidefinida)

Uma matriz é chamada positiva semidefinida (p.s.d.) se, e somente se

- $\mathbf{A} = \mathbf{A}^T$
- $\mathbf{y}^T\mathbf{A}\mathbf{y} \geq 0$ *para todo vetor* $\mathbf{y} \in E_n$ *tal que* $\mathbf{y} \neq 0$.

1. Para $\mathbf{A}$ p.d. $a_{ii} > 0$ para todo i.
2. Para $\mathbf{A}$ p.s.d. $a_{ii} \geq 0$ para todo i.
3. Lembre que $\mathbf{y}^T\mathbf{A}\mathbf{y}$ é uma forma quadrática (um escalar).

Matriz p.d. e Fatoração de Matrizes

1. Se $\mathbf{A} = \mathbf{B}^\top\mathbf{B}$ com $r(\mathbf{B}_{n\times p}) = p < n$, então $\mathbf{B}^\top\mathbf{B}$ é p.d. ($\checkmark$)
2. Se $\mathbf{A} = \mathbf{B}^\top\mathbf{B}$ com $r(\mathbf{B}_{n\times p}) = k < p < n$, então $\mathbf{B}^\top\mathbf{B}$ é p.s.d.
3. Podemos fatorar uma matriz $\mathbf{A}$ p.d. em $\mathbf{T}^\top\mathbf{T}$, sendo $\mathbf{T}$ uma matriz triangular superior e não singular utilizando

Decomposição de Cholesky:

Se $\mathbf{A} = (a_{ij})$ e $\mathbf{T} = (t_{ij})$ matrizes $n \times n$,

$$
\begin{aligned}
t_{11} &= \sqrt{a_{11}}, \qquad t_{1j} = \frac{a_{1j}}{t_{11}} && 2 \le j \le n \\
t_{ii} &= \sqrt{a_{ii} - \sum_{k=1}^{i-1} t_{ki}^2} && 2 \le i \le n \\
t_{ij} &= \frac{a_{ij} - \sum_{k=1}^{i-1} t_{ki} t_{kj}}{t_{ii}} && 2 \le i \le j < n \\
tij &= 0 && 1 \le j < i \le n
\end{aligned}
$$

Exemplo R

```
> A <- matrix(c(4,12,-16,12,37,-43,-16,-43,98),3,3)
> T <- chol(A)

> A
     [,1] [,2] [,3]
[1,]    4   12  -16
[2,]   12   37  -43
[3,]  -16  -43   98

> T
     [,1] [,2] [,3]
[1,]    2    6   -8
[2,]    0    1    5
[3,]    0    0    3

> t(T) %*% T
     [,1] [,2] [,3]
[1,]    4   12  -16
[2,]   12   37  -43
[3,]  -16  -43   98
```

Vetores e Matrizes Ortogonais

- Sejam $\mathbf{a}_{n\times 1}$ e $\mathbf{b}_{n\times 1}$. Se $\mathbf{a}^{\mathsf{T}}\mathbf{b} = a_1 b_1 + \ldots + a_n b_n = 0$ então $\mathbf{a}$ e $\mathbf{b}$ são ortogonais (perpendiculares), ou $\langle \mathbf{a}, \mathbf{b} \rangle = 0$.
- Se $\mathbf{a}\mathbf{a}^{\mathsf{T}} = 1$ o vetor é unitário (normalizado).
- Um conjunto de vetores $\{\mathbf{v}_i\}$ em $\Re^n$ é Ortonormal se

$$\mathbf{v}_i^{\mathsf{T}}\mathbf{v}_j = \delta_{ij} = \begin{cases} 0, i \neq j \\ 1, i = j \end{cases}$$

  (δ é conhecido como Kroenecker Delta.)

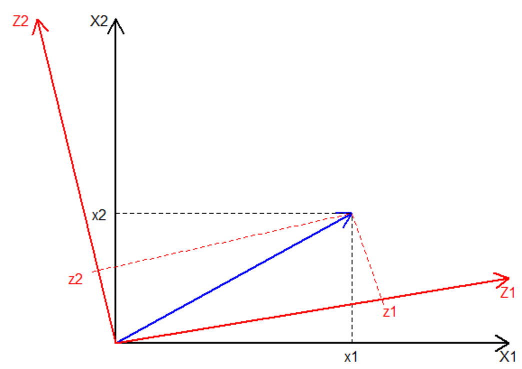
- Uma matriz quadrada $\mathbf{P}$ é ortogonal se $\mathbf{P}^{\mathsf{T}}\mathbf{P} = \mathbf{P}\mathbf{P}^{\mathsf{T}} = \mathbf{I}$, de forma que $\mathbf{P}^{-1} = \mathbf{P}^{\mathsf{T}}$.
- A multiplicação de um vetor por uma matriz ortogonal tem o efeito de rotação dos eixos, isto é, seja $\mathbf{P}$ ortogonal a $\mathbf{X}$ e $\mathbf{Z} = \mathbf{P}\mathbf{X}$, então,

$$\mathbf{Z}^{\mathsf{T}}\mathbf{Z} = (\mathbf{P}\mathbf{X})^{\mathsf{T}}(\mathbf{P}\mathbf{X}) = \mathbf{X}^{\mathsf{T}}\mathbf{P}^{\mathsf{T}}\mathbf{P}\mathbf{X} = \mathbf{X}^{\mathsf{T}}\mathbf{I}\mathbf{X} = \mathbf{X}^{\mathsf{T}}\mathbf{X}$$

  Mas o que significa $\mathbf{Z}^{\mathsf{T}}\mathbf{Z} = \mathbf{X}^{\mathsf{T}}\mathbf{X}$?

Vetores e Matrizes Ortogonais

Mas o que significa $\mathbf{Z}^\top\mathbf{Z} = \mathbf{X}^\top\mathbf{X}$?
A distância de **Z** a origem é a mesma distância de **X** a origem!



Qual a vantagem de fazer este tipo de rotação?

Vetores e Matrizes Ortogonais

Processo de Ortogonalização de Gram-Schmidt:

- Permite a construção de uma base ortonormal a partir de uma base arbitrária.
- Seja $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ uma base em $\Re^n$. O processo produz uma base ortonormal em $\{\mathbf{Z}_1, \ldots, \mathbf{Z}_n\}$ (em $\Re^n$).
- O `R` possui o comando `gramSchmidt` no pacote `pracma`.
- O comando do `R` produz, a partir de uma matriz $\mathbf{X}$, duas matrizes $\mathbf{Q}$ e $\mathbf{R}$, tais que $\mathbf{Q}$ é ortonormal, $\mathbf{R}$ é triangular superior, e $\mathbf{X} = \mathbf{Q} * \mathbf{R}$.

```
> library(pracma)
>
> A = matrix(c(4,0,0,2,3,1,0,-1),4,2,byrow=TRUE)
> A
     [,1] [,2]
[1,]    4    0
[2,]    0    2
[3,]    3    1
[4,]    0   -1
>
> gs <- gramSchmidt(A)
> gs
$Q
     [,1]       [,2]
[1,]  0.8 -0.2021165
[2,]  0.0  0.8421519
[3,]  0.6  0.2694886
[4,]  0.0 -0.4210760

$R
     [,1]     [,2]
[1,]    5 0.600000
[2,]    0 2.374868

> round(t(gs$Q)%*%gs$Q,2)
     [,1] [,2]
[1,]    1    0
[2,]    0    1
>
> round(gs$Q%*%gs$R,2)
     [,1] [,2]
[1,]    4    0
[2,]    0    2
[3,]    3    1
[4,]    0   -1
```

Autovalores e Autovetores

Definição

Uma matriz quadrada $\mathbf{A}$ *tem autovalor* λ *e autovetor correspondente* $\mathbf{x} \neq 0$, *se* $\mathbf{Ax} = \lambda\mathbf{x}$.

- A Equação Característica $|\mathbf{A} - \lambda\mathbf{I}| = 0$ permite encontrar λ's. ($\checkmark$)
- Os autovetores normalizados são representados por $\mathbf{e}$.
- Seja $\mathbf{A}_{k \times k}$. Então $\mathbf{A}$ tem k pares de autovalores e autovetores, $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_k, \mathbf{e}_k)$.
- $\mathbf{e}_1^T\mathbf{e}_1 = \ldots = \mathbf{e}_k^T\mathbf{e}_k = 1$ (mutuamente perpendiculares)
- $\lambda's \neq$ implicam em autovetores únicos.
- O `R` apresenta resultados para vetores ortogonais e normalizados.

Autovalores e Autovetores

- $\mathbf{A}_{k\times k}$ com autovalor λ e autovetor $\mathbf{x}$. Autovalor de $(\mathbf{I}+\mathbf{A})$? $(\checkmark)$
- $tr(A) = \sum_{i=1}^{k} a_{ii} = \sum_{i=1}^{k} \lambda_i$, mas $a_{ii} \neq \lambda_i$.
- $|A| = \prod_{i=1}^{k} \lambda_i$.
- Se $\mathbf{A}$ p.d. então $\lambda_i > 0 \;\forall\; k$.
- Se $\mathbf{A}$ p.s.d. então $\lambda_i = 0$ para algum k.
- $r(\mathbf{A}) =$ número de $\lambda_{i's}$ diferentes de zero.

Decomposição Espectral

Seja $\mathbf{A}_{d\times d}$ uma matriz quadrada e $\mathbf{X}$ a matriz de autovetores normalizados de $\mathbf{A}$.

- $\mathbf{X} = \left[\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(\mathrm{d})}\right]$;
- $\mathbf{A}\mathbf{X}_{(j)} = \lambda_j \mathbf{X}_{(j)}$, sendo λ_j o j-ésimo autovalor de $\mathbf{A}$;
- $\mathbf{X}\mathbf{X}^\mathsf{T} = \mathbf{X}^\mathsf{T}\mathbf{X} = \mathbf{I}$

Então,

$$\begin{aligned}\mathbf{A} &= \mathbf{A}\mathbf{X}\mathbf{X}^\mathsf{T} = \mathbf{A}(\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(\mathrm{d})})\mathbf{X}^\mathsf{T} = (\mathbf{A}\mathbf{X}_{(1)}, \ldots, \mathbf{A}\mathbf{X}_{(\mathrm{d})})\mathbf{X}^\mathsf{T} \\ &= (\lambda_1\mathbf{X}_{(1)}, \ldots, \lambda_\mathrm{d}\mathbf{X}_{(\mathrm{d})})\mathbf{X}^\mathsf{T} = \mathbf{X}\mathbf{D}\mathbf{X}^\mathsf{T} \quad \text{sendo} \quad \mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_\mathrm{d})\end{aligned}$$

Dizemos que $\mathbf{X}\mathbf{D}\mathbf{X}^\mathsf{T}$ é a Decomposição Espectral de $\mathbf{A}$.
Esta fatoração da matriz $\mathbf{A}$ é uma transformação linear expressa por rotações e dimensionamento (*scaling*), efetuadas por $\mathbf{X}$ e $\mathbf{D}$ (respectivamente).
Note ainda que,

- $\mathbf{X}^\mathsf{T}\mathbf{A}\mathbf{X} = \mathbf{X}^\mathsf{T}\mathbf{X}\mathbf{D}\mathbf{X}^\mathsf{T}\mathbf{X} = \mathbf{D} \Rightarrow$ transformação de $\mathbf{A}$ em matriz diagonal.
- Se $\mathbf{A}$ p.d. e $\mathbf{D}^{1/2} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_\mathrm{d}})$, então $\mathbf{A}^{1/2} = \mathbf{X}\mathbf{D}^{1/2}\mathbf{X}^\mathsf{T}$.
- $\mathbf{I} = (\mathbf{X}\mathbf{D}^{-1}\mathbf{X}^\mathsf{T})(\mathbf{X}\mathbf{D}\mathbf{X}^\mathsf{T}) \;\Rightarrow\; \mathbf{A}^{-1} = \mathbf{X}\mathbf{D}^{-1}\mathbf{X}^\mathsf{T} = \sum_{j=1}^{\mathrm{d}} \frac{1}{\lambda_j}\mathbf{X}_{(j)}\mathbf{X}_{(j)}^\mathsf{T}$.
- Ainda, $(\mathbf{A}^{1/2})^{-1} = \mathbf{X}\mathbf{D}^{-1/2}\mathbf{X}^\mathsf{T}$. (Verifique)

Definição

Uma matriz quadrada $\mathbf{A}$ *é idempotente quando* $\mathbf{A} = \mathbf{A}^2$.

- Sejam $\mathbf{A}$ e $\mathbf{B}$ matrizes $n \times k$, tais que $n \geq k$ e $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{B}) = k$. Uma matriz idempotente $\mathbf{D}$ pode ser construída como $\mathbf{D} = \mathbf{B}(\mathbf{A}^\top\mathbf{B})^{-1}\mathbf{A}^\top$. ($\checkmark$)

Teorema (1)

Uma matriz idempotente $\mathbf{A}$ *é sempre singular, exceto para a matriz identidade* $\mathbf{I}$. ($\checkmark$)

Teorema (2)

Sejam $\mathbf{A}$ *e* $\mathbf{B}$ *matrizes idempotentes. Então,*

1. $\mathbf{A} + \mathbf{B}$ *é idempotente somente quando* $\mathbf{AB} = \mathbf{BA} = 0$. ($\checkmark$)
2. $\mathbf{C} = \mathbf{AB}$ *é idempotente somente quando* $\mathbf{AB} = \mathbf{BA}$. ($\checkmark$)
3. $\mathbf{I} - \mathbf{A}$ *é idempotente.* ($\checkmark$)

Projeção de Matrizes (✓)

Teorema (3)

Seja **P** *uma matriz de projeção. Então,*

1. **P** *é associada a uma transformação linear.*
2. **P** *é idempotente.*

Teorema (4)

Uma matriz **P** *projeta vetores ortogonalmente em um subespaço se, e somente se,* **P** *é uma matriz idempotente e simétrica.*

Teorema (5)

Seja $\mathbf{X}_{(n \times k)}$ *tal que* $r(\mathbf{X}) = k < n$. *Então,* $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ *é idempotente e simétrica e consequentemente, uma matriz projeção ortogonal.*

Traço de uma Matriz

O traço de uma matriz $\mathbf{A}_{p\times p}$, $tr(\mathbf{A})$ é a soma dos elementos da diagonal.

Sejam $\mathbf{A}_{p\times p}$, $\mathbf{B}_{p\times p}$, $\mathbf{C}_{p\times n}$, $\mathbf{D}_{n\times p}$ e β um escalar.

- $tr(\beta) = \beta$.
- $tr(\mathbf{A} \pm \mathbf{B}) = tr(\mathbf{A}) \pm tr(\mathbf{B})$.
- $tr(\beta\mathbf{A}) = \beta\ tr(\mathbf{A})$.
- $tr(\mathbf{CD}) = tr(\mathbf{DC}) = \sum_i \sum_j c_{ij} d_{ij}$
  $\Rightarrow tr(\mathbf{CC}^\top) = tr(\mathbf{C}^\top\mathbf{C}) = \sum_{i=1}^{n} \sum_{j=1}^{p} c_{ij}^2$.

Sistemas de Equações Lineares

$\mathbf{y} = \mathbf{A}\mathbf{x}$ representa um sistema de transformações lineares, i.e.

$$
\begin{aligned}
y_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1p}x_p \\
y_2 &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2p}x_p \\
\vdots &= \vdots \\
y_p &= a_{p1}x_1 + a_{p2}x_2 + \ldots + a_{pp}x_p
\end{aligned}
$$

$\mathbf{y} = \mathbf{A}\mathbf{x}$ transforma um vetor $\mathbf{x} \in E_p$ em $\mathbf{y} \in E_p$. Possibilidades:

1. Exatamente uma solução existe. $\mathbf{A}$ não singular e $\mathbf{A}^{-1}$ existe. A transformação representa uma mudança de coordenadas.
2. Infinitas soluções existem. $\mathbf{A}^{-1}$ não existe, mas $\mathbf{A}^{-}$ existe.
3. Nenhuma solução é possível. Inversa não existe.

Sistemas de Equações Lineares

O caso 2 pode ser resolvido com matrizes não quadradas.

Inversa Generalizada: Seja $\mathbf{A}$ uma matriz $m \times n$. Se $\mathbf{A}^-$ existe, então

- $\mathbf{A}\mathbf{A}^-$ e $\mathbf{A}^-\mathbf{A}$ são simétricas;
- $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$;
- $\mathbf{A}^-\mathbf{A}\mathbf{A}^- = \mathbf{A}^-$.

- Para cada matriz $\mathbf{A}$ existe uma única inversa generalizada. ($\checkmark$)

Inversa Condicional: Seja $\mathbf{A}\mathbf{A}^c\mathbf{A} = \mathbf{A}$. Então $\mathbf{A}^c$ é chamada de *inversa condicional*.
- $\mathbf{A}^{-1}$ é também $\mathbf{A}^-$ que é também $\mathbf{A}^c$.
- Em geral, a literatura (e também este curso) trata $\mathbf{A}^c$ de inversa generalizada.
- No pacote `MASS` do `R` o comando é `ginv(A)`.

Vetores Aleatórios, Amostragem e Medidas Resumo

Amostragem

- Uma amostragem é feita para aprendizado sobre um fenômeno de interesse.
- Para cada uma das n observações da amostra são obtidas medidas de p variáveis.
- As medições das p variáveis para uma única observação são geralmente correlacionadas ou dependentes.
- As medições de diferentes observações devem ser independentes.

- Seja

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1^\top \\ \mathbf{X}_2^\top \\ \vdots \\ \mathbf{X}_n^\top \end{pmatrix}$$

em que $\mathbf{X}_i^\top = (\mathrm{X}_{i1}, \ldots, \mathrm{X}_{ip})$ é um vetor contendo p medições.

As distâncias entre n pontos no espaço $p-$dimensional é determinada pela função de probabilidade conjunta de dos $\mathbf{X}_j$,

$$f(\mathbf{x}_i) = f(\mathrm{x}_{i1}, \mathrm{x}_{i2}, \ldots, \mathrm{x}_{ip})$$

- Um vetor aletório é um vetor com variáveis aleatórias.

- Média Populacional

$$\mathrm{E}(\mathbf{X}) = \mathrm{E}\begin{pmatrix} \mathrm{X}_1^\top \\ \mathrm{X}_2^\top \\ \vdots \\ \mathrm{X}_p^\top \end{pmatrix} = \begin{pmatrix} \mathrm{E}(\mathrm{X}_1^\top) \\ \mathrm{E}(\mathrm{X}_2^\top) \\ \vdots \\ \mathrm{E}(\mathrm{X}_p^\top) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} = \boldsymbol{\mu}$$

- (Matriz de) Média amostral

$$\bar{\mathbf{X}} = \begin{pmatrix} \bar{\mathbf{X}}_1 \\ \bar{\mathbf{X}}_2 \\ \vdots \\ \bar{\mathbf{X}}_p \end{pmatrix} = \frac{1}{n}\mathbf{X}^\top \mathbf{j}$$

- $\bar{\mathbf{X}}$ é um estimador não viesado de $\boldsymbol{\mu}$, i.e., $\mathrm{E}(\bar{\mathbf{X}}) = \boldsymbol{\mu}$.
- $\bar{\mathbf{X}}$ tem matriz de covariância $(1/n)\boldsymbol{\Sigma}$.

- (Matriz de) Variância-Covariância populacional (Cov($\mathbf{X}$))

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} = E[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^\top] = E(\mathbf{X}\mathbf{X}^\top)-\boldsymbol{\mu}\boldsymbol{\mu}^\top$$

- A Equação Característica de $\boldsymbol{\Sigma}$ é $\boldsymbol{\Sigma\Gamma} = \boldsymbol{\Gamma\Lambda}$, em que $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_p)$ e $\boldsymbol{\Gamma}$ é a matriz de autovetores normalizados de $\boldsymbol{\Sigma}$. Então,

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma\Lambda\Gamma}^\top = \boldsymbol{\Gamma\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Gamma}^\top = \boldsymbol{\Gamma\Lambda}^{1/2}\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma\Lambda}^{1/2}\boldsymbol{\Gamma}^\top = \mathbf{C}\mathbf{C}^\top.$$

Qualquer matriz da forma $\mathbf{C}\mathbf{C}^\top$ é postiva semidefinida.

- (Matriz de) Variância-Covariância amostral

$$\mathbf{S} = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{pmatrix} = (s_{jk})$$

Diferentes formas de calcular **S**:

1. $s_{jj} = s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_i)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_{ij}^2 - n\bar{x}_j^2\right)$ e

   $s_{jk} = \frac{1}{n-1}\sum_{t=1}^{n}(x_{tj} - \bar{x}_j)(x_{tk} - \bar{x}_k) = \frac{1}{n-1}\left(\sum_{t=1}^{n} x_{tj}x_{tk} - n\bar{x}_j\bar{x}_k\right)$

2. $\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^{\top} = \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^{\top} - n\bar{\mathbf{X}}\bar{\mathbf{X}}^{\top}\right)$

3. $\mathbf{S} = \frac{1}{n-1}\mathbf{X}^{\top}\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\mathbf{X}.$

Nota: A matriz $\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)$ é idempotente e centraliza a matrix $\mathbf{X}$. Se $\mathbf{X}_c = \left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\mathbf{X}$, então $\mathbf{S} = \frac{1}{n-1}\mathbf{X}_c^{\top}\mathbf{X}_c$ é pelo menos positiva semidefinida .

- A Matriz de Correlação **R** é obtida de **S** fazendo $\mathrm{r}_{jk} = \frac{\mathrm{s}_{jk}}{\sqrt{\mathrm{s}_{jj}\mathrm{s}_{kk}}}$. Definindo a matriz diagonal $\mathbf{D}_s = [\mathrm{diag}(\mathbf{S})]$ podemos obter as seguintes relações:
  1. $\mathbf{R} = \mathbf{D}_s^{-1/2}\mathbf{S}\mathbf{D}_s^{-1/2}$ e
  2. $\mathbf{S} = \mathbf{D}_s^{1/2}\mathbf{R}\mathbf{D}_s^{1/2}$.
- A matriz de correlação populacional (**P**) é obtida de forma equivalente fazendo $\rho_{jk} = \frac{\sigma_{jk}}{\sqrt{\sigma_{jj}\sigma_{kk}}}$

- O estimador de máxima-verossimilhança de $\boldsymbol{\Sigma}$ é

$$\hat{\boldsymbol{\Sigma}} = \mathbf{S}_n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^{\top}.$$

Note que

$$\mathrm{E}(\mathbf{S}_n) = \frac{n-1}{n}\boldsymbol{\Sigma} = \boldsymbol{\Sigma} - \frac{1}{n}\boldsymbol{\Sigma}$$

Então

$$\mathbf{S} = \frac{n}{n-1}\mathbf{S}_n = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^{\top}$$

é um estimador não viesado de $\boldsymbol{\Sigma}$.

Se $\mathbf{X}$ tem vetor de médias $\boldsymbol{\mu}$ e matriz de variância-covariância $\boldsymbol{\Sigma}$,

- O elemento (i, j) de $\mathbf{S}$ é um estimador não viesado de σ_{ij}.
- $\sqrt{\mathrm{s}_{ij}}$ **não** é um estimador não viesado de $\sqrt{\sigma_{ij}}$.
- $\sqrt{\mathrm{r}_{ij}}$ **não** é um estimador não viesado de $\sqrt{\rho_{ij}}$.
- Para "grandes" amostras o viés é pequeno.
- Se os dados seguem uma distribuição normal multivariada (N_p), então substituindo os respectivos EMVs na equação da N_p irá produzir boas estimativas da distribuição populacional.
- Para dados esparsos e/ou superdimensionados, o problema é obter um bom estimador $\mathbf{S}$. Nestes casos $\bar{\mathbf{X}}$ e $\mathbf{S}$ não são bons estimadores.

Independência Estatística

- Independência estatística implica covariância nula.
- Covariância nula não implica indepedência estatística.
- Covariância nula e dist. normal multivariada implicam independência estatística.

Combinações Lineares

- $\mathrm{E}(\mathbf{X}+\mathbf{Y}) = \mathrm{E}(\mathbf{X}) + \mathrm{E}(\mathbf{Y})$.
- Se $\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{b}$, então
  $\mathrm{E}(\mathbf{Y}) = \mathbf{B}\,\mathrm{E}(\mathbf{X}) + \mathbf{b}$ e $\mathrm{Var}(\mathbf{Y}) = \mathbf{B}\,\mathrm{Cov}(\mathbf{X})\,\mathbf{B}^\top$
- $\mathrm{Var}(\mathbf{a}^\top\mathbf{X}) = \mathbf{a}^\top\,\mathrm{Cov}(\mathbf{X})\,\mathbf{a}$.
- $\mathrm{Cov}(\mathbf{A}^\top\mathbf{X}, \mathbf{B}^\top\mathbf{X}) = \mathbf{A}\,\mathrm{Cov}(\mathbf{X})\,\mathbf{B}^\top$.
- Esperança de uma forma quadrática (Cuidado!) Seja $\mathbf{X}$ um vetor com média $\boldsymbol{\mu}$ e matriz de variância-covariância $\mathrm{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$. Pode-se mostrar que

$$\mathrm{E}(\mathbf{X}^\top\mathbf{A}\mathbf{X}) = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}\mathbf{A}\boldsymbol{\mu}$$

  Provar.

Estimação da Variabilidade da Amostra

- O volume de um elipsóide centrado em $\mathbf{x}$ é obtido por

$$(\mathbf{x} - \boldsymbol{\mu})^\intercal \mathbf{S}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \mathrm{k}$$

  para todo k positivo no espaço Euclideano p-dimensional. Logo o volume de um elipsóide é proporcional a $|\mathbf{S}|$. O que isto significa?

- Duas medidas de gerais de variância são definidas em estudos multivariados:
  1. Variância amostral generalizada $\rightarrow |\mathbf{S}|$ (determinante de $\mathbf{S}$) e
  2. Variância amostral total $\rightarrow \mathrm{tr}\ \mathbf{S} = \lambda_1 + \ldots + \lambda_p$.

Variância Amostral Generalizada (GSV)

$$\mathbf{S} = \begin{pmatrix} 3 & -1 & 2 \\ -1 & 4 & 3 \\ 2 & 3 & 9 \end{pmatrix} \Rightarrow \mathrm{GSV} = 44$$

- GSV está relacionada ao volume do paralelepípedo definido por p vetores, i.e.,

$$\mathrm{GSV} = |\mathbf{S}| = (n-p)^{-p}(\mathrm{volume})^2$$

- No espaço $p-$multidimensional consideramos a dispersão de n pontos em torno da média amostral $\bar{\mathbf{X}}$, i.e.,

$$\mathbf{X} - \mathbf{j}\bar{\mathbf{X}} = \begin{pmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \dots & x_{1p} - \bar{x}_p \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \dots & x_{2p} - \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \dots & x_{np} - \bar{x}_p \end{pmatrix}$$

- O volume do ellipsoid

$$(\mathbf{x}-\bar{\mathbf{x}})^{\top}\mathbf{S}^{-1}(\mathbf{x}-\bar{\mathbf{x}}) \leq \mathrm{k}^2$$

  é igual a $c|\mathbf{S}|^{1/2}\,\mathrm{k}^p$, em que c é uma constante.

- Então GSV é proporcional ao volume do ellipsoid que representa as distâncias das observações em relação ao vetor de médias.
- $\mathrm{GSV} = 0$ indica que pelo menos uma das colunas da matriz de desvios em relação a média é combinação linear das demais.

Problemas com (GSV)

- Se $n \leq p$ então $\mathrm{GSV} = 0$ para todas as amostras.
- GSV representa variância e covariância. Logo podemos ter diferentes padrões de variabilidade e associação (estrutura de correlação) mas com o mesmo valor de GSV.


```
> s1
     [,1] [,2]
[1,]    3    0
[2,]    0    3

> s2
     [,1] [,2]
[1,]    5    4
[2,]    4    5

> s3
     [,1] [,2]
[1,]    5   -4
[2,]   -4    5

> det(s1)
[1] 9
> det(s2)
[1] 9
> det(s3)
[1] 9
```

Possíveis soluções:

- Padronizar as variáveis, obtendo **R** e calcular GSV sobre esta matriz $(\mathrm{GSV}_{\mathbf{R}} = |\mathbf{R}|)$. Neste caso a medida irá refletir apenas a estrutura de covariância.

```
> r1
      [,1] [,2]
[1,]   1.0  0.8
[2,]   0.8  1.0

> r2
      [,1] [,2]
[1,]   1.0 -0.8
[2,]  -0.8  1.0

> r3
      [,1] [,2]
[1,]   1.0  0.5
[2,]   0.5  1.0

> det(r1)
[1] 0.36

> det(r2)
[1] 0.36

> det(r3)
[1] 0.75
```

- Quando os vetores são perpendiculares, $\mathrm{GSV}_{\mathbf{R}}$ é max e igual a 1.
- Quando os vetores estão na mesma direção, $\mathrm{GSV}_{\mathbf{R}}$ é min e igual a 0.

```
> r4
     [,1] [,2]
[1,]    1    0
[2,]    0    1

> r5
     [,1] [,2]
[1,]    1    1
[2,]    1    1

> det(r4)
[1] 1

> det(r5)
[1] 0
```

Possíveis soluções:

- Utilizar uma medida que reflete apenas a variância em **S**.
- A variância amostral total é

$$\sum_{i=1}^{p} s_{ii} = \text{tr}\,\mathbf{S} = \lambda_1 + \ldots + \lambda_p$$

```
> s1
     [,1] [,2]
[1,]    3    0
[2,]    0    3

> s2
     [,1] [,2]
[1,]    5    4
[2,]    4    5

> s3
     [,1] [,2]
[1,]    5   -4
[2,]   -4    5

> sum(diag(s1))
[1] 6

> sum(diag(s2))
[1] 10

> sum(diag(s2))
[1] 10
```

Referências

- Johnson, R.A. e Wichern, D.W. (2007) *Applied Multivariate Statistical Analysis*, Sexta Edição. Prentice Hall.
- Rencher, A.C. e Christensen, W.F. (2012) *Methods of Multivariate Analysis*, Terceira Edição. Wiley.

As seguintes referências são mais completas.

- Yoshida, R. (2021) Linear Algebra and Its Applications with R. CRC Press.
- Basilevsky, A. (2005) Applied Matrix Algebra in the Statistical Sciences. Dover.
- Harville, D.A. (1997) Matrix Algebra from a Statistician's Perspective. Springer.
- Graybill, F.A. (1969) Matrices with Applications in Statistics. Duxbury.