Análise Multivariada
Distribuição Normal Multivariada

Prof. George von Borries

Departamento de Estatística
Universidade de Brasília

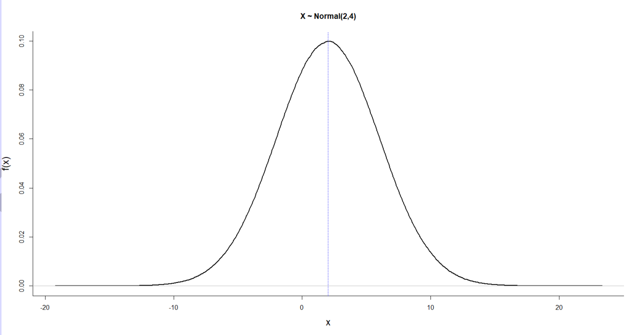
2023

Normal Univariada: $\mathrm{N}(\mu, \sigma^2)$

Definição (Normal Univariada)

$$g(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right)$$

onde $y \in \Re$, $-\infty < \mu < \infty$, $\sigma > 0$.



Normal Multivariada: $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Definição (Normal Multivariada)

A densidade da normal p-dimensional para um vetor aleatório $\mathbf{Y}^T = \{Y_1, Y_2, \ldots, Y_p\}$ tem a forma,

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[\frac{-(\mathbf{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}{2}\right]$$

onde $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^T$, $-\infty < y_i < \infty$, $i = 1, 2, \ldots, p$ *e* $\boldsymbol{\Sigma} > 0$.

- Note que para $p = 1$, a normal multivariada se reduz a normal univariada.
- Para $p = 2$ temos a normal bivariada.

Normal Bivariada: $\mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Neste caso a Normal Multivariada se reduz a,

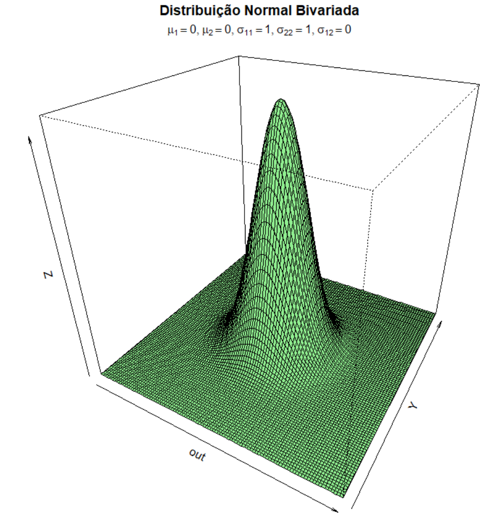
$$f(y_1, y_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}} \times exp\left\{ -\frac{1}{2(1-\rho_{12}^2)} \left[\left(\frac{y_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{y_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 \right.\right.$$
$$\left.\left. -2\rho_{12}\left(\frac{y_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{y_2-\mu_2}{\sqrt{\sigma_{22}}}\right) \right] \right\}$$

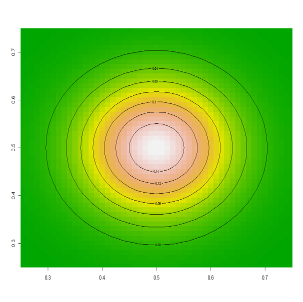
em que
$\mu_1 = \mathrm{E}(Y_1)$, $\mu_2 = \mathrm{E}(Y_2)$, $\sigma_{11} = \mathrm{Var}(Y_1)$, $\sigma_{22} = \mathrm{Var}(Y_2)$, e
$\rho_{12} = \sigma_{12}/(\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}) = \mathrm{Cor}(Y_1, Y_2)$.

Note que a inversa da matriz de variância-covariância,

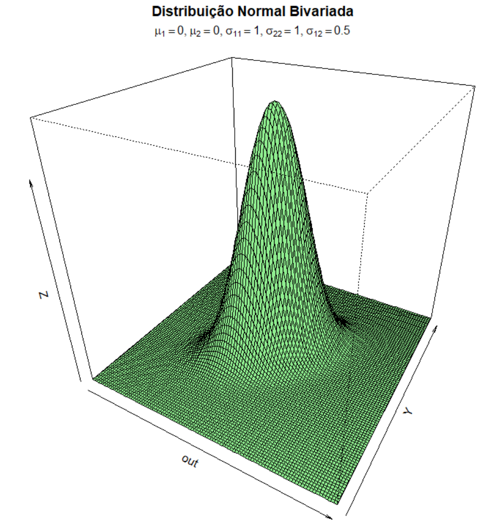
$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \quad \text{é} \quad \boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

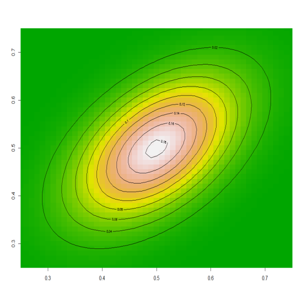
Normal Bivariada



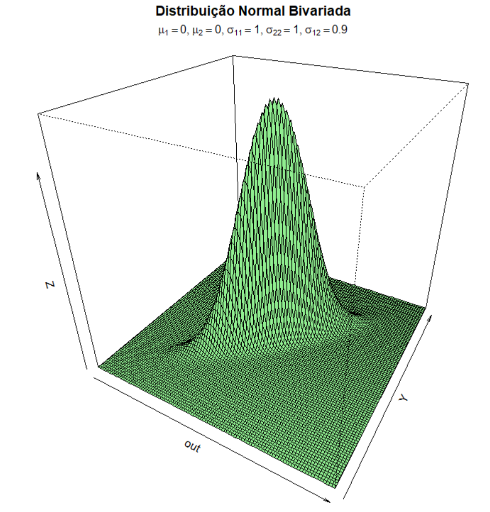


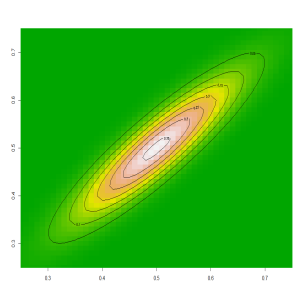
Normal Bivariada



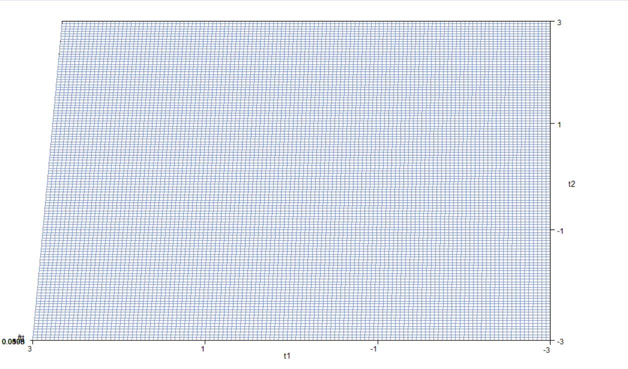


Normal Bivariada





Normal Bivariada



Obs: utilize Adobe Reader DC para ver a animação na figura acima.

Padronização

Resultado 1: Seja $\mathbf{Y} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Um vetor padronizado pode ser obtido de duas formas:

1. $\mathbf{Z} = (\mathbf{M}^\top)^{-1}(\mathbf{Y} - \boldsymbol{\mu})$, com $\boldsymbol{\Sigma} = \mathbf{M}^\top\mathbf{M}$ fatorado pela decomposição de Cholesky ou
2. $\mathbf{Z} = (\boldsymbol{\Sigma}^{1/2})^{-1}(\mathbf{Y} - \boldsymbol{\mu})$, sendo $\boldsymbol{\Sigma}^{1/2}$ a raiz quadrada simétrica de $\boldsymbol{\Sigma}$.

Assim, $\mathbf{Z} \sim \mathrm{N}_p(\mathbf{0}, \mathbf{I}_p)$.

Definição:

$\mathbf{Y}$ tem distribuição normal multivariada $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ se a função geradora de momentos (fgm) é

$$\mathrm{M}_{\mathbf{Y}}(\mathbf{t}) = \exp\left\{\mathbf{t}^{\top}\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^{\top}\boldsymbol{\Sigma}\mathbf{t}\right\}.$$

para todo $\mathbf{t} \in \Re^n$, $\boldsymbol{\Sigma}$ simétrica, positiva semidefinida e $\boldsymbol{\mu} \in \Re^n$.

Nota: Se $\boldsymbol{\Sigma}$ singular, $f(\mathbf{y})$ não é definida, mas a função geradora de momentos continua válida.

Resultado 2:

1. Seja $\mathbf{a}$ é um vetor de constantes. Se $\mathbf{Y} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então $\mathbf{a}^\mathsf{T}\mathbf{Y} \sim \mathrm{N}_p(\mathbf{a}^\mathsf{T}\boldsymbol{\mu}, \mathbf{a}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{a})$.
2. Seja $\mathbf{A}_{(q \times p)}$ uma matriz de constantes de rank $q \leq p$. Se $\mathbf{Y} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então $\mathbf{AY} \sim \mathrm{N}_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\mathsf{T})$.
3. Seja $\mathbf{Y} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Considere $\mathbf{X} = \mathbf{AY} + \mathbf{b}$, com $\mathbf{A}_{(m \times p)}$ e $\mathbf{b} \in \Re^m$. Então $\mathbf{X} \sim \mathrm{N}_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\mathsf{T})$.

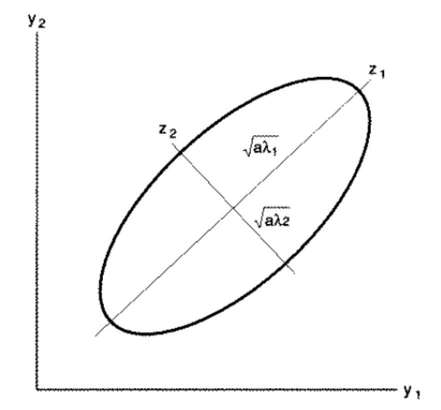
Resultado 3: (Ver prova em J&W, pág. 153)

- Se $\boldsymbol{\Sigma}$ é positiva definida, então $\boldsymbol{\Sigma}^{-1}$ existe, é positiva definida e

$$\boldsymbol{\Sigma}\mathbf{e} = \lambda\mathbf{e} \quad \Rightarrow \quad \boldsymbol{\Sigma}^{-1}\mathbf{e} = \frac{1}{\lambda}\mathbf{e}$$

Resultado 4:

Seja $\mathbf{Y} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Então contornos de densidade constante são elipsóides definidos por $\mathbf{Y}$ tais que $(\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) = a$.
Os elipsóides são centrados em $\boldsymbol{\mu}$ e tem eixos $\pm\sqrt{a\lambda_i}$, $i = 1, \ldots, p$.



Resultado 5:

Se $\mathbf{Y} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então $\underbrace{(\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu})}_{\text{distância estatística}} \sim \chi^2_p$, sendo $|\boldsymbol{\Sigma}| > 0$.

Resultado 6:

O elipsóide de $\mathbf{Y}$ com valores $(\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \leq \chi^2_p(\alpha)$, tem probabilidade $(1 - \alpha)$.

Resultado 7:

Sejam $\mathbf{Y} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{A}$ e $\mathbf{B}$ matrizes apropriadas.
$\mathbf{AY}$ e $\mathbf{BY}$ são independentes se $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^\top = \mathbf{0}$.

Resultado 8: (Ver J&W, páginas 157 e 158)

Todos os subconjuntos de $\mathbf{Y}_{(n\times p)} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ tem distribuição normal.

Não existe perda de generalidade em escrever $\mathbf{Y}$ como

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}^a \\ \mathbf{Y}^b \end{bmatrix}$$

sendo $\mathbf{Y}^a_{(n\times m)}$ e $\mathbf{Y}^b_{(n\times k)}$, $k = p - m$. Da mesma forma teremos,

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix} \text{ e } \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}$$

com as dimensões apropriadas. Assim, $\mathbf{Y}^a \sim \mathrm{N}_m(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_{aa})$ e $\mathbf{Y}^b \sim \mathrm{N}_{p-m}(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_{bb})$.

Importante: Este teorema diz que toda distribuição marginal de $\mathbf{Y}$ é também normal, com média e variâncias associadas ao vetor de variáveis que formam a marginal.

Resultado 9:

Se os vetores $\mathbf{Y}_1$ e $\mathbf{Y}_2$ tem distribuição conjunta Normal Multivariada e $\mathbf{\Sigma}_{12} = \mathbf{0}$, então $\mathbf{Y}_1$ e $\mathbf{Y}_2$ são independentes.

Dica: A prova é feita utilizando a função geradora de momentos e verificando que $\mathrm{M}_{\mathbf{Y}_1,\mathbf{Y}_2}(\mathbf{t}_1, \mathbf{t}_2) = \mathrm{M}_{\mathbf{Y}_1}(\mathbf{t}_1)\mathrm{M}_{\mathbf{Y}_2}(\mathbf{t}_2)$ se, e somente se, $\mathbf{\Sigma}_{12} = \mathbf{0}$. (Note que $\mathbf{\Sigma}_{12} = \mathbf{\Sigma}_{21}$)

Resultado 10:

1. Normal multivariada $\Rightarrow$ marginais com dist. normal.
2. Marginais com dist. normal $\not\Rightarrow$ normal multivariada.
3. Marginiais com dist. não normal $\Rightarrow$ dist. não é normal multivariada.

Resultado 11:

A distribuição $\mathrm{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ é simétrica em torno de $\boldsymbol{\mu}$.

Dica: Substituir $\mathbf{y}$ por $\boldsymbol{\mu} + \mathbf{a}$ e verificar que $f(\boldsymbol{\mu} + \mathbf{a}) = f(\boldsymbol{\mu} - \mathbf{a})$.

Resultado 12: Distribuição Condicional

Sejam $\mathbf{Y}_1$ e $\mathbf{Y}_2$ com distribuição normal multivariada, tal que $\boldsymbol{\Sigma}_{12} \neq \mathbf{0}$. Então, $f(\mathbf{Y_1}|\mathbf{Y_2} = \mathbf{y}_2) \sim \mathrm{N}_m(\boldsymbol{\mu}_{1.2}, \boldsymbol{\Sigma}_{12})$ com

$$\boldsymbol{\mu}_{1.2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_2)$$

e

$$\boldsymbol{\Sigma}_{1.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.$$

Resultado 13: Ver J&W, páginas 168 a 173

Os estimadores de Máxima Verossilhança de $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ são $\hat{\boldsymbol{\mu}} = \bar{\mathbf{Y}}$ e $\hat{\boldsymbol{\Sigma}} = \frac{n-1}{n}\,\mathbf{S}$, com

$$\mathbf{S} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{Y}_j - \bar{\mathbf{Y}})(\mathbf{Y}_j - \bar{\mathbf{Y}})^{\mathsf{T}}.$$

As estatísticas $\bar{\mathbf{Y}}$ e $\mathbf{S}$ são suficientes.

Resultado 14: Versão multivariada do Teorema de Cochran

Seja $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ uma amostra aleatória e tamanho n de uma distribuição normal p-variada, com média $\boldsymbol{\mu}$ e matriz de variância-covariância $\boldsymbol{\Sigma}$. Então,

(a) $\bar{\mathbf{Y}} \sim \mathrm{N}_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$.

(b) $(n-1)\mathbf{S}$ tem distribuição de Wishart com $(n-1)$ graus de liberdade.

(c) $\bar{\mathbf{Y}}$ e $\mathbf{S}$ são independentes.

- A distribuição de Wishart é análoga a χ^2 no caso univariado.
- $\mathbf{W}_m(.\,|\boldsymbol{\Sigma})$ é a distrib. de $\sum_{j=1}^{m} \mathbf{Z}_j \mathbf{Z}_j^T$ onde $\mathbf{Z}_j \overset{iid}{\sim} \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$.
- Esta distribuição é importante em estudos de regressão multivariada e análise de variância multivariada (MANOVA).

Resultado 14: Teorema Central do Limite

Seja $\sqrt{n}(\bar{\mathbf{Y}} - \boldsymbol{\mu}) \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$. Assim, $\bar{\mathbf{Y}}$ é aproximadamente $\mathrm{N}_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$.

Segue que $(\mathbf{Y} - \boldsymbol{\mu})^\top \mathbf{S}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$ é aproximadamente χ^2_p para $(n - p)$ suficientemente grande.

Nota: Este resultado apresenta problemas em situações em que $n \approx p$ e $n < p$.

Investigando Normalidade Multivariada

Referências:

- C.J. Mecklin e D.J. Mundfrom (2004). An appraisal and bibliography of tests for multivariate normality. *International Statistical Review*, 72:123-138.

  Mecklin e Mundfrom apresentam uma excelente revisão de procedimentos para teste de bondade de ajustamento para a normal multivariada.

- R.A. Johnson e D.W. Wichern (2007). *Applied multivariate statistical analysis*. Prentice Hall. Sexta Edição. Para abordagem geral, ver seção 4.6.

- A.C. Rencher e W.F. Christensen (2012). *Methods of Multivariate Analysis*. Wiley, Terceira Edição. Ver seção 4.4.2.

Estratégias gráficas:

- Avaliar se as marginais de **Y** aparentam normalidade através de gráficos de probabilidade normal.
- Avaliar se pares de observações aparentam a forma elíptica esperada para populações normais.
- Verificar se existem observações atípicas.

Rencher e Christensen (2012) sugerem o seguinte procedimento:

1. Calcular as distâncias $\mathrm{D}_i^2 = (\mathbf{y}_i - \bar{\mathbf{y}})^\top \mathbf{S}^{-1}(\mathbf{y}_i - \bar{\mathbf{y}}), \quad i = 1, \ldots, n.$
2. Calcular $\mathrm{u}_i = n\mathrm{D}_i^2/(n-1)^2$
3. Calcular $\mathrm{v}_i = (i - \alpha)/(n - \alpha - \beta + 1)$, sendo $\alpha = (p-2)/2p$ e $\beta = (n-p-3)/(2(n-p-1))$. Quantis de uma $\mathrm{Beta}(\alpha, \beta)$.
4. Plotar os pares $(\mathrm{u}_{(i)}, \mathrm{v}_i)), \mathrm{u}_{(1)} \leq \mathrm{u}_{(2)}, \leq \ldots \leq \mathrm{u}_{(n)}$.
5. Padrão não linear sugere não normalidade.

Geração da Normal Multivariada

Referências:

- Rizzo, M.L. (2008). *Statistical computing with R*, Chapman & Hall/CRC.
- L. Han. *Generating multivariate normal data by using PROC IML*. Technical report, University of Georgia, Athens, GA.

- O `R` apresenta algumas funções para geração, como `mvrnorm` do pacote `MASS` e `rmvnorm` do pacote `mvtnorm`. A função `rmvnorm` possui os métodos eigen, svd e chol para geração da normal multivariada.

1 Geração por Decomposição Espectral

- $\mathbf{\Sigma}$ positiva definida $\Rightarrow \mathbf{\Sigma}^{1/2}$ positiva definida.
- $\mathbf{\Sigma}^{1/2} = \mathbf{\Gamma}^\top \mathbf{\Lambda}^{1/2} \mathbf{\Gamma}$.
- A decomposição espectral será $\mathbf{\Sigma} = \mathbf{\Gamma}^\top \mathbf{\Lambda} \mathbf{\Gamma}$.
- Se fizermos $\mathbf{Q} = \mathbf{\Sigma}^{1/2}$, então a fatorização de $\mathbf{\Sigma}$ é $\mathbf{\Sigma} = \mathbf{Q}^\top \mathbf{Q}$.

  Nota: Lembre que $\mathbf{\Gamma}\mathbf{\Gamma}^\top = \mathbf{I}$ e $\mathbf{\Gamma}^\top = \mathbf{\Gamma}^{-1}$.


```
# 1. POR DECOMPOSIÇÃO ESPECTRAL

rmvn.eigen <-
  function(n, mu, Sigma) {
    p <- length(mu)
    ev <- eigen(Sigma, symmetric = TRUE)
    lambda <- ev$values
    V <- ev$vectors
    R <- V %*% diag(sqrt(lambda)) %*% t(V)
    Z <- matrix(rnorm(n*p), nrow = n, ncol = p)
    X <- Z %*% R + matrix(mu, n, p, byrow = TRUE)
    X
  }
```

2 Geração por Decomposição em Valores Singulares

- $\boldsymbol{\Sigma} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathsf{T}}$.
- $\mathbf{U}$: autovetores de $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{T}}$.
- $\mathbf{S}$: diagonal da raiz dos autovalores de $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{T}}$.
- $\mathbf{V}$: autovetores de $\boldsymbol{\Sigma}^{\mathsf{T}}\boldsymbol{\Sigma}$.

  Como $\boldsymbol{\Sigma}$ simétrica e positiva definida,
  $\mathbf{U} = \mathbf{V}$ e $\boldsymbol{\Sigma}^{1/2} = \mathbf{U}\mathbf{D}^{1/2}\mathbf{V}^{\mathsf{T}} \equiv$ método espectral, porém menos eficiente.

```
  # 2. POR DECOMPOSIÇÃO EM VALORES SINGULARES

rmvn.svd <-
  function(n, mu, Sigma) {
    p <- length(mu)
    S <- svd(Sigma)
    R <- S$u %*% diag(sqrt(S$d)) %*% t(S$v)
    Z <- matrix(rnorm(n*p), nrow=n, ncol=p)
    X <- Z %*% R + matrix(mu, n, p, byrow=TRUE)
    X
    }
```

3. **Geração por Decomposição de Cholesky**

   - $\mathbf{\Sigma} = \mathbf{Q}^{\mathsf{T}}\mathbf{Q}$, sendo $\mathbf{Q}$ triangular superior.

```
    # POR DECOMPOSIÇÃO DE CHOLESKY

rmvn.cholesky <-
  function(n, mu, Sigma) {
    p <- length(mu)
    Q <- chol(Sigma)
    Z <- matrix(rnorm(n*p), nrow=n, ncol=p)
    X <- Z %*% Q + matrix(mu, n, p, byrow=TRUE)
    X
  }
```

De forma geral,

- Fazer a decomposição de $\mathbf{\Sigma}$ (chame de $\mathbf{D}$)
- Gerar uma matriz $\mathbf{Z}_{n \times p}$ de valores da $\mathrm{N}(0,1)$
- Gerar $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{Z} * \mathbf{D}$

Veja programa `geranp.R`.