CC0293 - Análise Multivariada

Notas de aula sobre vetores aleatórios

Prof. Gualberto Agamez Montalvo

1 Vetores aleatórios

1.1 Notação e conceitos básicos

Definição 1.1. Seja $\mathbf{y}$ um vetor aleatório de p $(p > 1)$ variáveis medidas em uma unidade amostral (sujeito ou objeto). Se houver n indivíduos na amostra, os n vetores de observações são denotados por

$$\mathbf{Y}_{n\times p} = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n)^\top = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{bmatrix} = [y_{ij}],$$

em que

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{ip} \end{bmatrix},$$

para $i = 1, 2, \cdots, n$ representa a informação ou medições do $i-$ésimo individuo da amostra.

Exemplo 1.1. Em um estudo foi considerada uma amostra de 5 pessoas do Nordeste do Brasil com idades entre 18 e 25 anos, com o objetivo de investigar algumas variáveis antropométricas que podem estar associadas ao estado de saúde e ao perfil físico da população. A Tabela 1 apresenta os dados coletados associados as variáveis de interesse:

- y_1 : altura (em centímetros);
- y_2 : diâmetro da cintura (em centímetros);
- y_3 : diâmetro do pescoço (em centímetros).

i	y_1	y_2	y_3
1	168	82	37
2	175	90	39
3	160	78	35
4	180	85	40
5	165	80	36

Tabela 1: Dados antropométricos da amostra do Nordeste do Brasil.

Neste exemplo temos que $p = 3$ e $n = 5$, portanto, os dados podem ser representado como

$$\mathbf{Y} = \begin{bmatrix} 168 & 82 & 37 \\ 175 & 90 & 39 \\ 160 & 78 & 35 \\ 180 & 85 & 40 \\ 165 & 80 & 36 \end{bmatrix}$$

Definição 1.2. Seja uma amostra aleatória $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n$ em $\mathbb{R}^p$, então a média amostral é definida por

$$\bar{\mathbf{y}}_. = \frac{\sum_{i=1}^n \mathbf{y}_i}{n} = \frac{1}{n}\mathbf{Y}^\top \mathbf{1}_n = \begin{bmatrix} \bar{y}_{.1} \\ \bar{y}_{.2} \\ \vdots \\ \bar{y}_{.p} \end{bmatrix},$$

em que $\bar{y}_{.k}$ representa a média amostral da $k-$ésima variável.

Exemplo 1.2. Continuando com o Exemplo 1.1 temos que

$$\bar{\mathbf{y}}_. = \begin{bmatrix} 168 & 175 & 160 & 180 & 165 \\ 82 & 90 & 78 & 85 & 80 \\ 37 & 39 & 35 & 40 & 36 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 848 \\ 415 \\ 187 \end{bmatrix} = \begin{bmatrix} 169,6 \\ 83,0 \\ 37,4 \end{bmatrix}.$$

No R, temos que:

```
> Y <- matrix(c(168, 82, 37,     # Indivíduo 1
+               175, 90, 39,     # Indivíduo 2
+               160, 78, 35,     # Indivíduo 3
+               180, 85, 40,     # Indivíduo 4
+               165, 80, 36),    # Indivíduo 5
+             nrow = 5, byrow = TRUE)
>
> # Opção 1
> apply(Y, 2, mean)
[1] 169.6  83.0  37.4
>
> # Opção 2
> n <- 5
> t(Y)%*%rep(1/n,n)
      [,1]
[1,] 169.6
[2,]  83.0
[3,]  37.4
>
> # Opção 3
> apply(Y, 2, sum)/n
[1] 169.6  83.0  37.4
```

Definição 1.3. A média de $\mathbf{y}$ sobre todos os valores da população é denominada vetor de médias populacionais ou valor esperado de $\mathbf{y}$. Este é definido como o vetor de valores esperados de cada variável

$$\mathbb{E}(\mathbf{y}) = \mathbb{E}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} \mathbb{E}(y_1) \\ \mathbb{E}(y_2) \\ \vdots \\ \mathbb{E}(y_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu},$$

em que μ_k é a média populacional da $k-$ésima variável.

Observação: o valor esperado de $\bar{\mathbf{y}}_{.}$ é dado por

$$\mathbb{E}\left(\bar{\mathbf{y}}_{.}\right) = \mathbb{E}\begin{bmatrix} \bar{y}_{.1} \\ \bar{y}_{.2} \\ \vdots \\ \bar{y}_{.p} \end{bmatrix} = \begin{bmatrix} \mathbb{E}(\bar{y}_{.1}) \\ \mathbb{E}(\bar{y}_{.2}) \\ \vdots \\ \mathbb{E}(\bar{y}_{.p}) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu}.$$

Definição 1.4. A matriz de variâncias e covariâncias amostrais é dada por

$$\mathbf{S}_p = \mathbf{S} = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{21} & S_{22} & \dots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \dots & S_{pp} \end{bmatrix},$$

em que o $k-$ésimo elemento da diagonal é dado por

$$S_{kk} = S_k^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_{ik} - \bar{y}_{.k}\right)^2$$

e a covariância amostral entre as variáveis $k-$ésima e $l-$ésima é dada por

$$S_{kl} = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_{ik} - \bar{y}_{.k}\right)\left(y_{il} - \bar{y}_{.l}\right).$$

Exemplo 1.3. Continuando com o Exemplo 1.1 temos que

$$\mathbf{S} = \begin{bmatrix} 63,30 & 30,50 & 16,45 \\ 30,50 & 22,00 & 8,25 \\ 16,45 & 8,25 & 4,30 \end{bmatrix}.$$

No R, temos que:

```
> S <- var(Y); S
      [,1]  [,2]  [,3]
[1,] 63.30 30.50 16.45
[2,] 30.50 22.00  8.25
[3,] 16.45  8.25  4.30
```

Observação: A matriz de covariâncias amostrais pode ser reescrita em termos dos vetores $\mathbf{y}_i$ como

$$\begin{aligned}\mathbf{S} &= \frac{1}{n-1}\left[\sum_{i=1}^{n}\left(\mathbf{y}_i - \bar{\mathbf{y}}_.\right)\left(\mathbf{y}_i - \bar{\mathbf{y}}_.\right)^\top\right] \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n}\left(\mathbf{y}_i\mathbf{y}_i^\top - \bar{\mathbf{y}}_.\mathbf{y}_i^\top - \bar{\mathbf{y}}_.\mathbf{y}_i^\top + \bar{\mathbf{y}}_.\bar{\mathbf{y}}_.^\top\right)\right] \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n}\mathbf{y}_i\mathbf{y}_i^\top - n\bar{\mathbf{y}}_.\bar{\mathbf{y}}_.^\top - n\bar{\mathbf{y}}_.\bar{\mathbf{y}}_.^\top + n\bar{\mathbf{y}}_.\bar{\mathbf{y}}_.^\top\right] \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n}\mathbf{y}_i\mathbf{y}_i^\top - n\bar{\mathbf{y}}_.\bar{\mathbf{y}}_.^\top\right] \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n}\mathbf{y}_i\mathbf{y}_i^\top - \frac{\left(\sum_{i=1}^{n}\mathbf{y}_i\right)\left(\sum_{i=1}^{n}\mathbf{y}_i\right)^\top}{n}\right].\end{aligned}$$

Adicionalmente, a matriz de covariâncias amostrais pode ser reescrita em termos da matriz $\mathbf{Y}$ como

$$\begin{aligned}\mathbf{S} &= \frac{1}{n-1}\mathbf{Y}^\top\mathbf{Y} - \frac{1}{n}\mathbf{Y}^\top\mathbf{J}\mathbf{Y} \\ &= \frac{1}{n-1}\mathbf{Y}^\top\left[\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\right]\mathbf{Y}\end{aligned}$$

em que a matriz $\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n$ é conhecida como matriz de projeção.

Propriedades:

1. $\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n$ é simétrica e idempotente;

2. $\text{rk}(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n) = \text{tr}(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n) = n - 1.$

No R, temos que:

```
> J <- matrix(1, ncol = n, nrow = n)
> I <- diag(n)
> t(Y)%*%(I - J/n)%*%Y/(n-1)
      [,1]  [,2]  [,3]
[1,] 63.30 30.50 16.45
[2,] 30.50 22.00  8.25
[3,] 16.45  8.25  4.30
```

Definição 1.5. A matriz de covariância populacional associada ao vetor $\mathbf{y}$ é definida por

$$\boldsymbol{\Sigma} = \text{Cov}\left(\mathbf{y}\right) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_p^2 \end{bmatrix}$$

em que σ_k^2 é a variância populacional da $k-$ésima variável e σ_{kl} é a covariância populacional entre as variáveis $k-$ésima e $l-$ésima.

Observação: a matriz de covariância populacional pode ser obtida por

$$\begin{aligned}
\boldsymbol{\Sigma} &= \mathbb{E}\left[\left(\mathbf{y} - \boldsymbol{\mu}\right)\left(\mathbf{y} - \boldsymbol{\mu}\right)^\top\right] \\
&= \mathbb{E}\left(\mathbf{y}\mathbf{y}^\top - \boldsymbol{\mu}\mathbf{y}^\top - \mathbf{y}\boldsymbol{\mu}^\top + \boldsymbol{\mu}\boldsymbol{\mu}^\top\right) \\
&= \mathbb{E}\left(\mathbf{y}\mathbf{y}^\top\right) - \boldsymbol{\mu}\mathbb{E}\left(\mathbf{y}\right)^\top - \mathbb{E}\left(\mathbf{y}\right)\boldsymbol{\mu}^\top + \boldsymbol{\mu}\boldsymbol{\mu}^\top \\
&= \mathbb{E}\left(\mathbf{y}\mathbf{y}^\top\right) - \boldsymbol{\mu}\boldsymbol{\mu}^\top - \boldsymbol{\mu}\boldsymbol{\mu}^\top + \boldsymbol{\mu}\boldsymbol{\mu}^\top \\
&= \mathbb{E}\left(\mathbf{y}\mathbf{y}^\top\right) - \boldsymbol{\mu}\boldsymbol{\mu}^\top.
\end{aligned}$$

Definição 1.6. A matriz de correlações amostrais é dada por

$$\mathbf{R} = \mathbf{D_s}^{-\frac{1}{2}} \mathbf{S} \mathbf{D_s}^{-\frac{1}{2}} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{bmatrix},$$

em que $\mathbf{D_s} = \operatorname{diag}(\mathbf{S}) = \operatorname{diag}(S_{kk}) = \operatorname{diag}(S_{11}, S_{22}, \dots, S_{pp}) = \operatorname{diag}\left(S_1^2, S_2^2, \dots, S_p^2\right) = \operatorname{diag}(S_k^2)$, $\mathbf{D_s}^{-\frac{1}{2}} = \operatorname{diag}\left(\frac{1}{S_1}, \frac{1}{S_2}, \dots, \frac{1}{S_p}\right)$ e

$$r_{kl} = \frac{S_{kl}}{S_k S_l},$$

em que, r_{kl} representa a correlação amostral entre as variáveis k-ésima e l-ésima.

Exemplo 1.4. Continuando com o Exemplo 1.1 temos que

$$\mathbf{R} \approx \begin{bmatrix} 1,0000 & 0,8173 & 0,9971 \\ 0,8173 & 1,0000 & 0,8482 \\ 0,9971 & 0,8482 & 1,0000 \end{bmatrix}.$$

No R, temos que:

```
> # Opção 1
> round(cor(Y), 4)
       [,1]   [,2]   [,3]
[1,] 1.0000 0.8173 0.9971
[2,] 0.8173 1.0000 0.8482
[3,] 0.9971 0.8482 1.0000
>
> # Opção 2
>
> d <- diag(var(Y))
> Ds <- diag(d); Ds
```

```
     [,1] [,2] [,3]
[1,] 63.3    0  0.0
[2,]  0.0   22  0.0
[3,]  0.0    0  4.3
>
> round( solve(sqrt(Ds))%*%S%*%solve(sqrt(Ds)), 4)
       [,1]   [,2]   [,3]
[1,] 1.0000 0.8173 0.9971
[2,] 0.8173 1.0000 0.8482
[3,] 0.9971 0.8482 1.0000
```

Definição 1.7. A matriz de correlações populacionais é dada por

$$\mathbf{P} = \mathbf{D_\Sigma}^{-\frac{1}{2}}\mathbf{\Sigma}\mathbf{D_\Sigma}^{-\frac{1}{2}} = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{bmatrix},$$

em que $\mathbf{D_\Sigma} = \mathrm{diag}\,(\sigma_k^2)$, $\mathbf{D_\Sigma}^{-\frac{1}{2}} = \mathrm{diag}\left(\frac{1}{\sigma_k}\right)$ e

$$\rho_{kl} = \frac{\sigma_{kl}}{\sigma_k\sigma_l}.$$

Observação: as matrizes de variâncias e covariâncias $\mathbf{S}$ e $\mathbf{\Sigma}$ podem ser obtidas por

$$\mathbf{S} = \mathbf{D_s}^{\frac{1}{2}}\mathbf{R}\mathbf{D_s}^{\frac{1}{2}}$$

e

$$\mathbf{\Sigma} = \mathbf{D_\Sigma}^{\frac{1}{2}}\mathbf{P}\mathbf{D_\Sigma}^{\frac{1}{2}},$$

em que $\mathbf{D_s}^{\frac{1}{2}} = \mathrm{diag}\,(S_k)$ e $\mathbf{D_\Sigma}^{\frac{1}{2}} = \mathrm{diag}\,(\sigma_k)$.

1.2 Vetores de médias e matrizes de covariâncias para subconjuntos de variáveis

Definição 1.8. Considere o vetor $\mathbf{y}$ particionado por

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_q \end{bmatrix},$$

em que $\mathbf{y}_i$ é de tamanho p_i para $i = 1, 2, \cdots, q$ e $p = p_1 + p_2 + \cdots + p_q$. Então, o vetor de médias amostrais e a matriz de covariâncias para uma amostra aleatória são dados por

$$\bar{\mathbf{y}} = \begin{bmatrix} \bar{\mathbf{y}}_1 \\ \bar{\mathbf{y}}_2 \\ \vdots \\ \bar{\mathbf{y}}_q \end{bmatrix}$$

e

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} & \dots & \mathbf{S}_{1q} \\ \mathbf{S}_{21} & \mathbf{S}_{22} & \dots & \mathbf{S}_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{q1} & \mathbf{S}_{q2} & \dots & \mathbf{S}_{qq} \end{bmatrix},$$

em que a submatriz $\mathbf{S}_{kl}$ contem as variâncias e covariâncias das variáveis $\mathbf{y}_k$ e $\mathbf{y}_l$. Os valores populacionais correspondentes são

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{\mu}_q \end{bmatrix}$$

e

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \dots & \boldsymbol{\Sigma}_{1q} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \dots & \boldsymbol{\Sigma}_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{q1} & \boldsymbol{\Sigma}_{q2} & \dots & \boldsymbol{\Sigma}_{qq} \end{bmatrix}.$$

1.3 Distâncias entre vetores

Definição 1.9. Considere dois vetores $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^p$. A distância euclidiana quadrática é dada por

$$d_E^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^\top (\mathbf{x} - \mathbf{y}) = \sum_{k=1}^{p} (x_k - y_k)^2.$$

Exemplo 1.5. Considere os vetores:

$$\mathbf{x} = (4, 9, 5, 1)^\top \quad \text{e} \quad \mathbf{y} = (6, 8, 0, 4)^\top.$$

```
> x <- c(4, 9, 5, 1)
> y <- c(6, 8, 0, 4)

> t(x-y)%*%(x-y)
     [,1]
[1,]   39

> dist(rbind(x, y), method = "euclidean")^2
   x
y 39
```

Observação: a função *dist* tem outras distâncias programadas

1. maximum;
2. manhattan;
3. canberra;
4. binary;
5. minkowski.

Definição 1.10. Considere dois vetores $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^p$ e $\mathbf{S}$ a matriz de covariâncias amostrais. A distância euclidiana padronizada quadrática é dada por

$$d_{EP}^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^\top \mathbf{D}^{-1}(\mathbf{x} - \mathbf{y}) = \sum_{k=1}^{p} \frac{(x_k - y_k)^2}{S_{kk}},$$

em que $\mathbf{D} = \mathrm{diag}(\mathbf{S})$.

Definição 1.11. Considere dois vetores $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^p$ e $\mathbf{S}$ a matriz de covariâncias amostrais. A distância generalizada de Mahalanobis é dada por

$$d_{GM}^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^\top \mathbf{S}^{-1}(\mathbf{x} - \mathbf{y}).$$

Definição 1.12. Considere dois vetores $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^p$ e uma matriz $\mathbf{A}$ positiva definidas. A expressão geral para a distância quadrática é dada por

$$d_A^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^\top \mathbf{A}(\mathbf{x} - \mathbf{y}),$$

em que $\mathbf{A}$ é chamada de métrica.

1.4 Combinações lineares de variáveis

Definição 1.13. Considere as constantes $a_1, a_2, \cdots, a_p$. Uma combinação linear de elementos do vetor $\mathbf{y}$ é dada por

$$z = a_1 y_1 + a_2 y_2 + \cdots + a_p y_p = \mathbf{a}^\top \mathbf{y},$$

em que $\mathbf{a} = (a_1, a_2, \cdots, a_p)^\top$.

Definição 1.14. Considere as constantes $a_1, a_2, \cdots, a_p$ e uma amostra aleatória de vetor $\mathbf{y}_i$ com $i = 1, 2, \cdots, n$. A média e variância amostrais de z são dadas por

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = \mathbf{a}^\top \bar{\mathbf{y}}.$$

e

$$s_z^2 = \frac{\sum_{i=1}^n (z_i - \bar{z})^2}{n-1} = \mathbf{a}^\top \mathbf{S}\mathbf{a},$$

em que $\mathbf{a} = (a_1, a_2, \cdots, a_p)^\top$.

Definição 1.15. Considere $z = \mathbf{a}^\top \mathbf{y}$ e $w = \mathbf{b}^\top \mathbf{y}$, então a covariância e a correlação amostrais de z e w são dadas por

$$s_{z,w} = \frac{\sum_{i=1}^n (z_i - \bar{z})(w_i - \bar{w})}{n-1} = \mathbf{a}^\top \mathbf{S}\mathbf{b},$$

e

$$r_{z,w} = \frac{s_{zw}}{\sqrt{s_z^2 s_w^2}} = \frac{\mathbf{a}^\top \mathbf{S}\mathbf{b}}{\sqrt{(\mathbf{a}^\top \mathbf{S}\mathbf{a})(\mathbf{b}^\top \mathbf{S}\mathbf{b})}},$$

respectivamente.

Observação: A média e variância populacionais associados a z são dadas por

$$\mathbb{E}(z) = \mathbb{E}(\mathbf{a}^\top \mathbf{y}) = \mathbf{a}^\top \mathbb{E}(\mathbf{y}) = \mathbf{a}^\top \boldsymbol{\mu}$$

e

$$\sigma_z^2 = \mathrm{Var}(\mathbf{a}^\top \mathbf{y}) = \mathbf{a}^\top \mathrm{Var}(\mathbf{y})\mathbf{a} = \mathbf{a}^\top \boldsymbol{\Sigma}\mathbf{a}.$$

Observação: A covariância e correlação populacionais associados a z e w são dadas por

$$\sigma_{zw} = \mathrm{Cov}(z, w) = \mathbf{a}^\top \boldsymbol{\Sigma}\mathbf{b}$$

e

$$\rho_{zw} = \frac{\sigma_{zw}}{\sigma_z \sigma_w} = \frac{\mathbf{a}^\top \boldsymbol{\Sigma}\mathbf{b}}{\sqrt{(\mathbf{a}^\top \boldsymbol{\Sigma}\mathbf{a})(\mathbf{b}^\top \boldsymbol{\Sigma}\mathbf{b})}}.$$

Definição 1.16. Considere q transformações lineares, tais que

$$
\begin{aligned}
z_1 &= a_{11}y_1 + a_{12}y_2 + \cdots + a_{1p}y_p = \mathbf{a}_1^\top \mathbf{y}\\
z_2 &= a_{21}y_1 + a_{22}y_2 + \cdots + a_{2p}y_p = \mathbf{a}_2^\top \mathbf{y}\\
&\vdots\\
z_q &= a_{q1}y_1 + a_{q2}y_2 + \cdots + a_{qp}y_p = \mathbf{a}_q^\top \mathbf{y}
\end{aligned}
$$

A forma matricial é dada por

$$
\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_q \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^\top \mathbf{y} \\ \mathbf{a}_2^\top \mathbf{y} \\ \vdots \\ \mathbf{a}_q^\top \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \\ \vdots \\ \mathbf{a}_q^\top \end{bmatrix} \mathbf{y} = \mathbf{A}\mathbf{y},
$$

em que $\mathbf{A}_{q\times p}$.

Definição 1.17. O vetor de médias e a matriz de covariâncias amostrais do vetor $\mathbf{z}$ são dados por

$$
\bar{\mathbf{z}} = \begin{bmatrix} \mathbf{a}_1^\top \bar{\mathbf{y}} \\ \mathbf{a}_2^\top \bar{\mathbf{y}} \\ \vdots \\ \mathbf{a}_q^\top \bar{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \\ \vdots \\ \mathbf{a}_q^\top \end{bmatrix} \bar{\mathbf{y}} = \mathbf{A}\bar{\mathbf{y}}
$$

e

$$
\begin{aligned}
\mathbf{S_z} &= \begin{bmatrix} \mathbf{a_1^\top S_y a_1} & \mathbf{a_1^\top S_y a_2} & \dots & \mathbf{a_1^\top S_y a_q} \\ \mathbf{a_2^\top S_y a_1} & \mathbf{a_2^\top S_y a_2} & \dots & \mathbf{a_2^\top S_y a_q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a_q^\top S_y a_1} & \mathbf{a_q^\top S_y a_2} & \dots & \mathbf{a_q^\top S_y a_q} \end{bmatrix} = \begin{bmatrix} \mathbf{a_1^\top} \\ \mathbf{a_2^\top} \\ \vdots \\ \mathbf{a_q^\top} \end{bmatrix} [\mathbf{S_y a_1}, \mathbf{S_y a_2}, \cdots, \mathbf{S_y a_q}] \\
&= \begin{bmatrix} \mathbf{a_1^\top} \\ \mathbf{a_2^\top} \\ \vdots \\ \mathbf{a_q^\top} \end{bmatrix} \mathbf{S_y}[\mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_q}] = \mathbf{A S_y A^\top}
\end{aligned}
$$

Observação: O vetor de médias e a matriz de covariâncias populacionais associados ao vetor $\mathbf{z}$ são dados por

$$\mathbb{E}(\mathbf{z}) = \mathbb{E}(\mathbf{A}\mathbf{y}) = \mathbf{A}\mathbb{E}(\mathbf{y}) = \mathbf{A}\boldsymbol{\mu}$$

e

$$\boldsymbol{\Sigma}_z = \mathrm{Cov}(\mathbf{A}\mathbf{y}) = \mathbf{A}\mathrm{Cov}(\mathbf{y})\mathbf{A}^\top = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top.$$

Observação: no caso geral em que $\mathbf{z} = \mathbf{A}\mathbf{y} \pm \mathbf{b}$, o vetor de médias e a matriz de covariâncias populacionais são dados por

$$\mathbb{E}(\mathbf{z}) = \mathbf{A}\mathbb{E}(\mathbf{y}) \pm \mathbf{b} = \mathbf{A}\boldsymbol{\mu} \pm \mathbf{b}$$

e

$$\boldsymbol{\Sigma}_z = \mathrm{Cov}(\mathbf{A}\mathbf{y} \pm \mathbf{b}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top.$$

1.5 Variâncias generalizadas

A matriz de variâncias e covariâncias amostrais contém p variâncias e $\frac{p(p-1)}{2}$ covariâncias (usualmente todas são diferentes). Quando p é grande é difícil obter conclusões gerais a partir destes parâmetros. Portanto, é desejável atribuir um único valor (resumo) para expressar a variação representada por $\mathbf{S}$. Duas medidas usuais para resolver este problema são o determinante e o traço da matriz de covariâncias.

Definição 1.18. A variância generalizada amostral (VGA) para as variáveis não padronizadas é definida como o determinante da matriz de covariância amostral

$$\mathrm{VGA} = |\mathbf{S}| = \prod_{k=1}^{p} \lambda_k,$$

em que λ_k para $k = 1, 2, \cdots, p$, são os autovalores de $\mathbf{S}$.

Observação: se $\lambda_k = 0$ para algum $k = 1, 2, \cdots, p$, então $|\mathbf{S}| = 0$. Um autovalor zero indica uma redundância no relacionamento linear das variáveis, isto é, existe dependência linear. Uma solução para este dilema é remover uma ou mais variáveis.

Definição 1.19. A variância generalizada amostral para as variáveis padronizadas é definida como o determinante da matriz de correlação amostral.

Observação: a relação entre $|\mathbf{S}|$ e $|\mathbf{R}|$ é dada por

$$\begin{aligned}|\mathbf{R}| &= |\mathbf{D_s}^{-\frac{1}{2}}\mathbf{S}\mathbf{D_s}^{-\frac{1}{2}}| = |\mathbf{D_s}^{-\frac{1}{2}}||\mathbf{S}||\mathbf{D_s}^{-\frac{1}{2}}| \\ &= \prod_{k=1}^{p}\frac{1}{s_k} \times |\mathbf{S}| \times \prod_{k=1}^{p}\frac{1}{s_k} \\ &= |\mathbf{S}|\prod_{k=1}^{p}\frac{1}{s_k^2},\end{aligned}$$

e

$$|\mathbf{S}| = |\mathbf{R}|\prod_{k=1}^{p}s_k^2.$$

Definição 1.20. A variância total amostral (VTA) para as variáveis não padronizadas é definida como o traço da matriz de covariância amostral

$$\text{VTA} = \text{tr}(\mathbf{S}) = \sum_{k=1}^{p}s_k^2 = \sum_{k=1}^{p}\lambda_k.$$

Atividade

Pesquisar:

1. Interpretação geométrica da variância generalizada;
2. Matriz raiz quadrada;
3. Decomposição em valor singular.