Amostras aleatórias

Patrícia de Siqueira Ramos

UNIFAL-MG, *campus* Varginha

12 de Abril de 2018

2.1 Amostras aleatórias

2.1 Amostras aleatórias

- Seja uma amostra aleatória (a.a.) de tamanho n e, para cada unidade amostral (u.a.), foram observados os valores de p variáveis (ou seja, tem-se n vetores aleatórios i.i.d. - independentes identicamente distribuídos):

$$\begin{array}{cccc} u.a.1 & u.a.2 & \dots & u.a.n \\ \square & \square & \dots & \square \\ \mathbf{X}_{1\cdot} & \mathbf{X}_{2\cdot} & \dots & \mathbf{X}_{n\cdot} \end{array}$$

em que $\mathbf{X}_{i\cdot}$ é o vetor de observações para a i-ésima unidade amostral:

$$\mathbf{X}_{i\cdot} = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{bmatrix}$$

Juntos, os vetores $\mathbf{X}_{1\cdot}, \mathbf{X}_{2\cdot}, \ldots, \mathbf{X}_{n\cdot}$ formam a matriz de dados:

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \ldots & X_{1p} \\ X_{21} & X_{22} & \ldots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \ldots & X_{np} \end{bmatrix},$$

em que X_{ij} é a j-ésima variável medida da i-ésima unidade amostral

a) Vetor de médias amostrais

a) Vetor de médias amostrais

- O vetor μ será estimado pelo vetor de médias amostrais $\bar{\mathbf{X}}$.

$$\bar{\mathbf{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_1 \end{bmatrix} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_{i\cdot} = \frac{1}{n}(\mathbf{X}_{1\cdot} + \mathbf{X}_{2\cdot} + \cdots + \mathbf{X}_{n\cdot}),$$

em que $\bar{X}_j$ é a média amostral da j-ésima variável, $j = 1, \ldots, p$ e $\mathbf{X}_{i\cdot}$ é o vetor de observações para a i-ésima unidade amostral:

$$\mathbf{X}_{i\cdot} = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{bmatrix}$$

b) Matriz de covariâncias amostrais

b) Matriz de covariâncias amostrais

- A matriz de covariâncias $\mathbf{\Sigma}$ será estimada por:

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

sendo $S_{ij} = S_{ji}$, $i \neq j$.

b) Matriz de covariâncias amostrais

- Variância amostral da i-ésima variável: $S_{ii} = S_i^2 = \sum_{l=1}^{n} \frac{(X_{li} - \bar{X}_i)^2}{n-1}$
- Covariância amostral entre a i-ésima e a j-ésima variáveis:
$S_{ij} = \sum_{l=1}^{n} \frac{(X_{li} - \bar{X}_i)(X_{lj} - \bar{X}_j)}{n-1}$.
- Lembrar que a variância é a covariância de uma variável com ela mesma

b) Matriz de covariâncias amostrais

- Variância amostral da i-ésima variável: $S_{ii} = S_i^2 = \sum_{l=1}^{n} \frac{(X_{li} - \bar{X}_i)^2}{n-1}$
- Covariância amostral entre a i-ésima e a j-ésima variáveis: $S_{ij} = \sum_{l=1}^{n} \frac{(X_{li} - \bar{X}_i)(X_{lj} - \bar{X}_j)}{n-1}$.
- Lembrar que a variância é a covariância de uma variável com ela mesma
- Matricialmente, $\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{X}_{i\cdot} - \bar{\mathbf{X}})(\mathbf{X}_{i\cdot} - \bar{\mathbf{X}})^T$.

c) Matriz de soma de quadrados e produtos

- Outra matriz que pode aparecer em técnicas multivariadas:

$$\mathbf{W}_{p\times p} = \begin{bmatrix} W_{11} & W_{12} & \dots & W_{1p} \\ W_{21} & W_{22} & \dots & W_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ W_{p1} & W_{p2} & \dots & W_{pp} \end{bmatrix} = (n-1)\mathbf{S}.$$

d) Matriz de correlações amostrais

- A matriz $\boldsymbol{\rho}$ será estimada por:

$$\mathbf{R}_{p \times p} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{bmatrix},$$

em que

$$r_{ij} = \frac{S_{ij}}{\sqrt{S_{ii} S_{jj}}}$$

é o coeficiente de correlação amostral entre as variáveis i e j.

d) Matriz de correlações amostrais

- Matricialmente, definindo $\mathbf{D} = diag(\mathbf{S}) = diag(S_{ii})$ e $\mathbf{D}^{-1/2} = diag(1/\sqrt{S_{ii}})$,

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

Exercício

Obtenha $\bar{\mathbf{X}}$, $\mathbf{S}$, $\mathbf{W}$ e $\mathbf{R}$ de todas as maneiras apresentadas nesta aula considerando a seguinte amostra aleatória ($n = 4$, $p = 3$):

$$\mathbf{X}_{1\cdot} = \begin{bmatrix} 7 \\ 3 \\ 9 \end{bmatrix}, \mathbf{X}_{2\cdot} = \begin{bmatrix} 4 \\ 6 \\ 11 \end{bmatrix}, \mathbf{X}_{3\cdot} = \begin{bmatrix} 4 \\ 2 \\ 5 \end{bmatrix}, \mathbf{X}_{4\cdot} = \begin{bmatrix} 5 \\ 5 \\ 7 \end{bmatrix}.$$

(Monte a matriz de dados $\mathbf{X}$ antes)