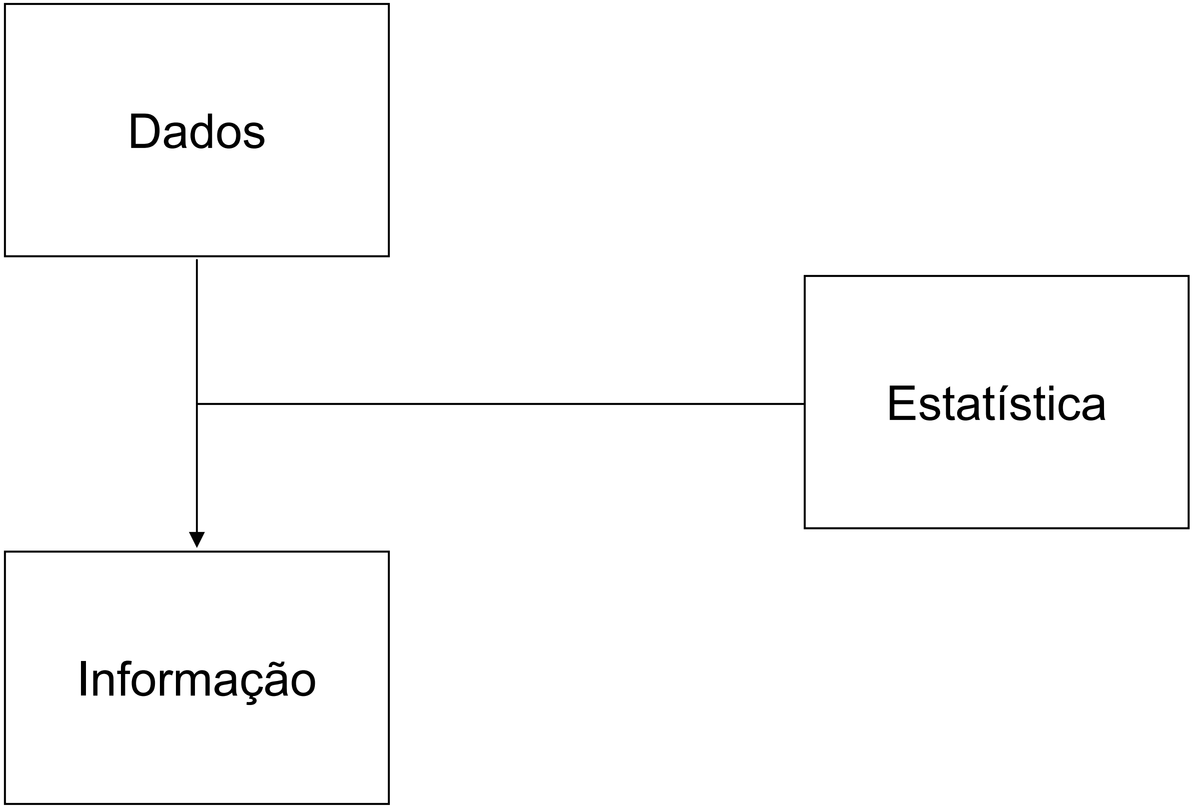
Prof. Doutor Tiago Mexia

A Estatística

O Que é a Estatística?



O Que é a Estatística?

Optimização

Delineamento

Dados

- Maximizar a informação para custo dados

ou

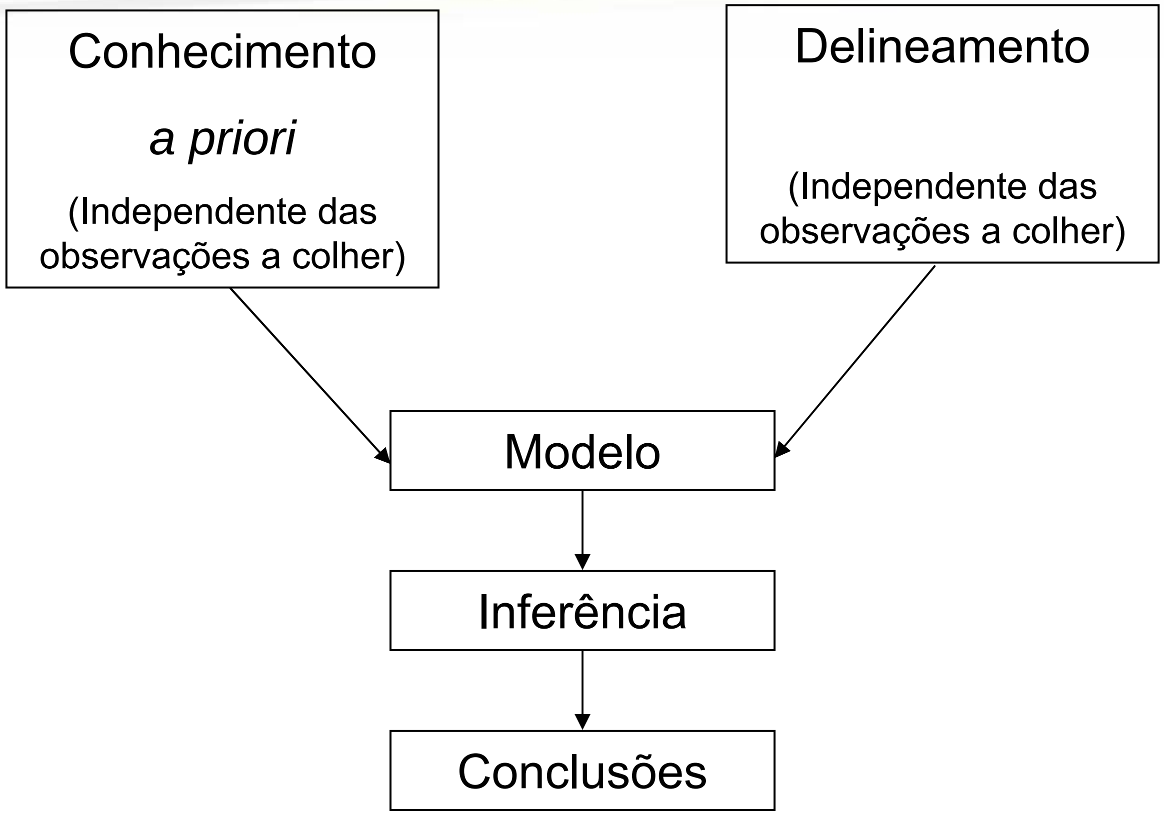
- Minimizar o custo para “informação dada”

LOGO

O que é a Estatística?

LOGO

Inferência Estatística



Teoria da Probabilidade

- "Palavras" ou Linguagem Estatística

$$\underline{\mathbf{x}}_i \sim F(\underline{\mathbf{x}}|\underline{\boldsymbol{\theta}}_i) = \mathbb{P}[\underline{\mathbf{x}}_i \leq \underline{\mathbf{x}}], \quad i = 1, \ldots, n$$

$$\underline{\mathbf{x}}_i \sim f(\underline{\mathbf{x}}|\underline{\boldsymbol{\theta}}_i) = \frac{\partial F(\underline{\mathbf{x}}|\underline{\boldsymbol{\theta}}_i)}{\prod_{i=1}^{k} \partial \underline{\mathbf{x}}_i}, \quad i = 1, \ldots, n$$

(modelo paramétrico contínuo)

- Conceito de Independência

$$\underline{\mathbf{x}}_1 (i) \ldots (i) \underline{\mathbf{x}}_n$$

As observações não herdam informação $\longrightarrow$ ruína na roleta

Meta-Teorema de Fischer(1918)

Selecção Natural

vs

Fisher

Genética

Teoria Integrada da Selecção Natural

Evolução de Populações

(panmixia)

LOGO

Meta-Teorema de Fischer(1918)

Sistemas Poligénicos

Características Biométricas (peso, altura, *etc*)

↓

Contribuições Individuais dos pares de alelos

pequenas
e
aditivas

↓

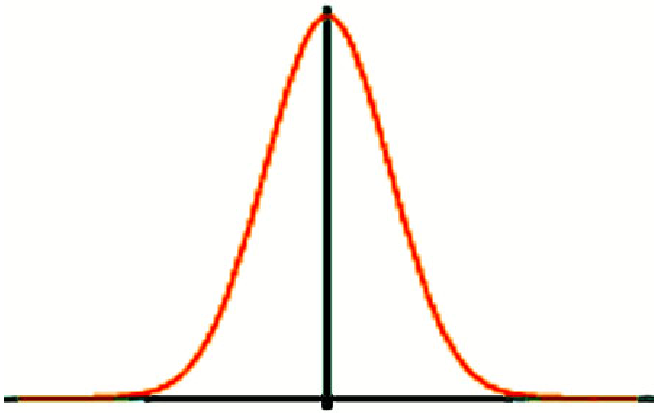
Teorema de Fischer

LOGO

Meta-Teorema de Fischer(1918)

- Teorema

  As características biométricas distribuem-se normalmente



LOGO

Modelos Normais

$$\underline{\mathbf{x}}_1, \ldots, \underline{\mathbf{x}}_n \begin{cases} \text{independentes} \\ \underline{\mathbf{x}}_i \sim \mathcal{N}\left(\underline{\boldsymbol{\mu}}_i, \mathbf{V}_i\right) \end{cases}$$

$$\underline{\mathbf{x}} = [\underline{\mathbf{x}}_1, \ldots, \underline{\mathbf{x}}_n]' \sim \mathcal{N}\left(\underline{\boldsymbol{\mu}}, \mathbf{V}\right) \begin{cases} \underline{\boldsymbol{\mu}} = \left[\underline{\boldsymbol{\mu}}_1, \ldots, \underline{\mu}_n\right]' \\ \begin{bmatrix} \mathbf{V}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{V}_n \end{bmatrix} \end{cases}$$

Modelos Normais

- Aplicabilidade: Metateorema de Fischer, *etc*
- Maleabilidade: “Empilhamento”

$$\underline{\mathbf{x}} \sim \mathcal{N}\left(\underline{\boldsymbol{\mu}}, \mathbf{V}\right)$$

$$\mathbf{A}\underline{\mathbf{x}} + \underline{\mathbf{b}} \sim \mathcal{N}\left(\mathbf{A}\underline{\boldsymbol{\mu}} + \underline{\mathbf{b}}, \mathbf{A}\mathbf{V}\mathbf{A}'\right)$$

$$\underline{\mathbf{x}}'\mathbf{V}^{+}\underline{\mathbf{x}} \sim \chi^{2}_{g,v} \quad \delta^{+} = \underline{\boldsymbol{\mu}}'\mathbf{V}^{+}\underline{\boldsymbol{\mu}}$$

<u>Teoria muito desenvolvida</u>

ex: testes de normalidade

LOGO

Normalidade Aproximada

$$\begin{cases} X_1 = \mu_1 + e_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \\ X_2 = \mu_2 + e_2 \sim \mathcal{N}(\mu_2, \sigma_2^2) \end{cases} \qquad \begin{cases} \frac{X_1 X_2}{\mu_1\mu_2 + \mu_1 e_1 + \mu_2 e_2} \underset{q.c.}{\longrightarrow} 1, \mu_1 \to \infty, \mu_2 \to \infty \\ \mu_1\mu_2 + \mu_1 e_1 + \mu_2 e_2 \sim \mathcal{N}(\mu_1\mu_2, \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2) \end{cases}$$

$$\downarrow$$

$$Z = \frac{X_1 X_2}{\mu_1\mu_2 + \mu_1 e_1 + \mu_2 e_2}, \text{ aproximadamente normal}$$

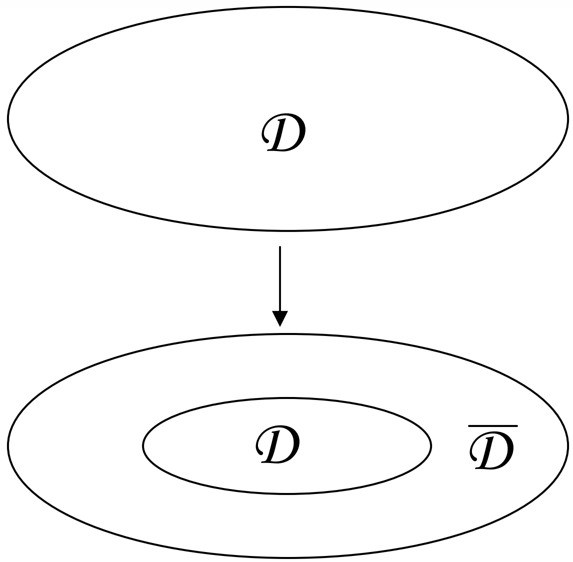
$$CV(X_l) = \frac{\sigma_l}{\mu_l} \to 1, \mu_l \to \infty, l = 1,2$$

Simulações: $CV(X_l) \leq \frac{1}{30} \to Z, \text{ aproximadamente normal}, l = 1,2$

Generalizações: Polinómios de baixo grau em variáveis normais independentes com baixos coeficientes de v

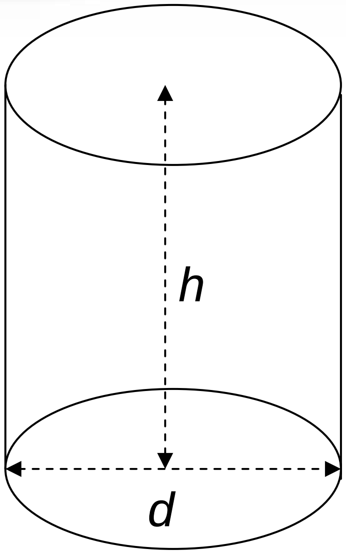
LOGO

Normalidade Aproximada



Modelos Normais $\longrightarrow$ Soluções Óptimas para Inferência

Um Primeiro Exemplo



$$V = \frac{\pi^2}{4} d^2 h$$

$$X \sim \mathcal{N}(d, \sigma_X^2) \ \ (i) \ \ Y \sim \mathcal{N}(h, \sigma_Y^2)$$

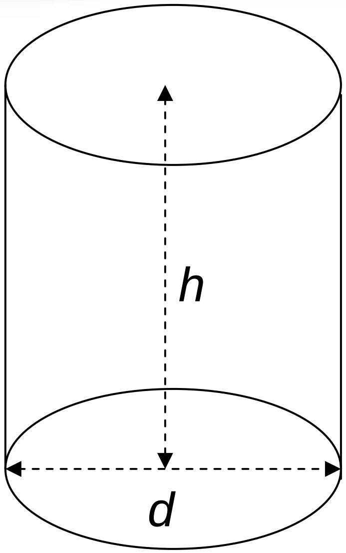
$$\widetilde{V} = \frac{\pi^2}{4} X^2 Y$$

$$X^2 \sim \sigma_X^2 \chi^2_{1,\delta_X} \ \ (i) \ \ Y \sim \mathcal{N}(h, \sigma_Y^2), \ \text{com}\, \delta_X = \frac{d^2}{\sigma_X^2}$$

$$\begin{cases} X = d + \varepsilon_X; \ \ \varepsilon_X \sim \mathcal{N}(0, \sigma_X^2) \\ Y = h + \varepsilon_Y; \ \ \varepsilon_Y \sim \mathcal{N}(0, \sigma_Y^2) \end{cases}$$

$$X^2 = d^2 + 2d\varepsilon_X + \varepsilon_X^2$$

Um Primeiro Exemplo (Continuação)



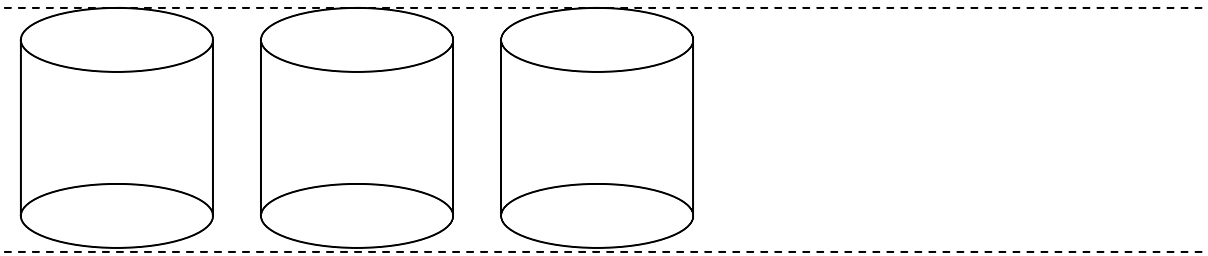
$$X^2Y = d^2h + d^2\varepsilon_Y + 2dh\varepsilon_X + 2dh\varepsilon_X\varepsilon_Y + h\varepsilon_X^2 + \varepsilon_X^2\varepsilon_Y^2$$

$$\approx d^2h + d^2\varepsilon_Y + 2dh\varepsilon_X$$

$$\sim \mathcal{N}(d^2h, d^4\sigma_Y^2 + 4d^2h^2\sigma_X^2)$$

$$\widetilde{V} \approx V^* = \frac{\pi^2}{4}(d^2h + d^2\varepsilon_Y + 2dh\varepsilon_X)$$

Um Primeiro Exemplo (Continuação)



$$\tilde{V} = \frac{\pi^2}{4}\left(\sum_{i=1}^{n} X_i^2 + nd^2\varepsilon_{\overline{Y}} + 2nd^2\sum_{i=1}^{n}\varepsilon_{X_i}\right)\overline{Y}$$

$$\begin{cases} \sum_{i=1}^{n} X_i^2 \sim \sigma_X^2 \chi^2_{n,\delta_X}; \quad \delta_X = \frac{n}{\sigma_X^2} \\ \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \sim \mathcal{N}\left(h, \frac{\sigma_Y^2}{n}\right) \end{cases}$$

$$\tilde{V} \approx V^* = \frac{\pi^2}{4}(nd^2h)$$

$$\varepsilon_{\overline{Y}} \sim \mathcal{N}\left(0, \frac{\sigma_Y^2}{n}\right) \qquad \sum_{i=1}^{n}\varepsilon_{X_i} \sim \mathcal{N}(0, n\sigma_X^2)$$

LOGO

Um Primeiro Exemplo (Continuação)

Quando é que se pode “admitir a normalidade”?

- Para polinómios até ao 4º grau pode admitir-se se:

$$CV \leq \frac{1}{30}$$

A utilização de técnicas padronizadas, permite em geral que este limite seja “folgadamente satisfeito”

$\downarrow$

Inferência Simplificada $\longrightarrow$ Comparação *Var(v)* para 2 processos

Teste *F*

Informação Metodológica

É a informação que se tem sobre os métodos de medida utilizados para:

Obter as observações

$$\mathbf{e}_1,\ldots,\mathbf{e}_n \; iid \; \sim \mathcal{N}(0,\sigma^2\mathbf{M})$$

$$\mathbf{M} = \begin{bmatrix} m_{1,1} & \cdots & m_{1,k} \\ \vdots & \ddots & \vdots \\ m_{k,1} & \cdots & m_{k,k} \end{bmatrix}$$

$$\sigma^2 m_{j,j} \leq \overline{\sigma}^2, \quad j = 1,\ldots,k$$

$$\mathbf{x}_i = \boldsymbol{\mu}_i + \mathbf{e}_i, \quad i = 1,\ldots,n$$

$$\mathbf{y} = g(\mathbf{x}_1,\ldots,\mathbf{x}_n), \quad \text{aproximadamente normal}$$

Informação Metodológica $\longrightarrow$ $\begin{cases} \text{Modelos Normais} \\ \text{Inferência Simplificada} \end{cases}$

Validação do Modelo

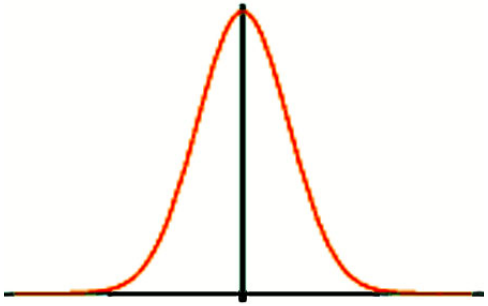
Erro de 3ª Espécie $\longrightarrow$ Escolha do Modelo Errado

$$X_1, \ldots, X_n \text{ iid}$$

Modelo Proposto

$$X_1, \ldots, X_n \text{ iid} \sim \mathcal{N}(x|\lambda;\sigma^2)$$

$$\text{Estimador} : \lambda_n^* = \frac{1}{n}\overline{X}_n$$



Validação do Modelo

Erro

- A densidade embora simétrica não tem valor médio

$$\lambda_n^* \nrightarrow \lambda$$

Procedimento

1) Testar a normalidade da amostra

2) Em caso de dúvida usar a mediana como estimador

LOGO

Validação do Modelo

Em geral: Validar o modelo

Obtenção de consequências testáveis

$$\mathbf{y} = g(\mathbf{x}_1, \ldots, \mathbf{x}_n) \underbrace{\sim F}_{\text{Modelo}}$$

Testando-se efectivamente se **y** tem distribuição F

Ex: testes de Normalidade...

Obrigado!

Prof. Doutor João Tiago Mexia