Tratamento Estatístico de dados em Física Experimental

Prof. Zwinglio Guimarães

2º semestre de 2016

Tópico 8 - Testes estatísticos

O método dos mínimos quadrados (revisão)

- O método dos mínimos quadrados consiste em determinar os parâmetros como àqueles que minimizam a seguinte somatória:

$$Q(\vec{a}) = \sum_{i=1}^{N} \left(\frac{y_i - G(x_i, \vec{a})}{\sigma_i} \right)^2$$

onde a função modelo, $G(x_i, \vec{a})$, descreve a relação entre o valor esperado do i-ésimo dado experimental com os parâmetros a serem estimados. Por exemplo, no caso do ajuste de uma reta $G(x, \vec{a}) = a_1 + a_2 x$

Ajuste de funções lineares nos parâmetros pelo MMQ (revisão)

- Escrevendo a função modelo como:

$$G(x,\vec{a}) = a_1 g_1(x) + a_2 g_2(x) + \cdots + a_M g_M(x)$$

- O sistema linear de equações do MMQ pode ser escrito de forma matricial: $\overline{D} = \mathbf{M}\overline{\tilde{A}}$

$$D_l = \sum_{i=1}^{N} \frac{y_i g_l(x_i)}{\sigma_i^2} \qquad M_{l,c} = \sum_{i=1}^{N} \frac{g_l(x_i) g_c(x_i)}{\sigma_i^2}$$

cuja solução é: $\overline{\tilde{A}} = \begin{bmatrix} \tilde{a}_1 \\ \tilde{a}_2 \\ \vdots \\ \tilde{a}_M \end{bmatrix} = (\mathbf{M}^{-1})\overline{D}$ com $\mathrm{V}_{\tilde{A}} = \mathbf{M}^{-1}$

Exemplos de ajuste de funções lineares nos parâmetros pelo MMQ

$$G(x,\vec{a}) = a_1 g_1(x) + a_2 g_2(x) + \cdots + a_M g_M(x)$$

- Posição em função do tempo de um corpo em queda livre partindo do repouso:

$$G = a_1 + a_2 t^2,$$

o que implica $g_1 = 1$ e $g_2 = t^2$

- Tensão de um sinal senoidal de frequência conhecida medida em um osciloscópio (sem nível DC):

$$G = a_1 \cos(2\pi f t) + a_2 \sin(2\pi f t),$$

o que implica $g_1 = \cos(2\pi f t)$ e $g_2 = \sin(2\pi f t)$

O teste de χ^2

- O teste de χ^2 avalia se a dispersão dos pontos ao redor da função ajustada é consistente com as incertezas dos dados. Para dados estatisticamente independentes:

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - G(x_i, \vec{\tilde{a}})}{\sigma_i} \right]^2$$

- O valor esperado do χ^2 é $\langle \chi^2 \rangle = N - P$, onde N é o número de dados e P o número de parâmetros.

- O χ^2_{Red} é definido como $\chi^2_{Red} = \frac{\chi^2}{\nu}$, onde $\nu = N - P$ é o número de graus de liberdade.

Teste de hipótese usando o χ^2

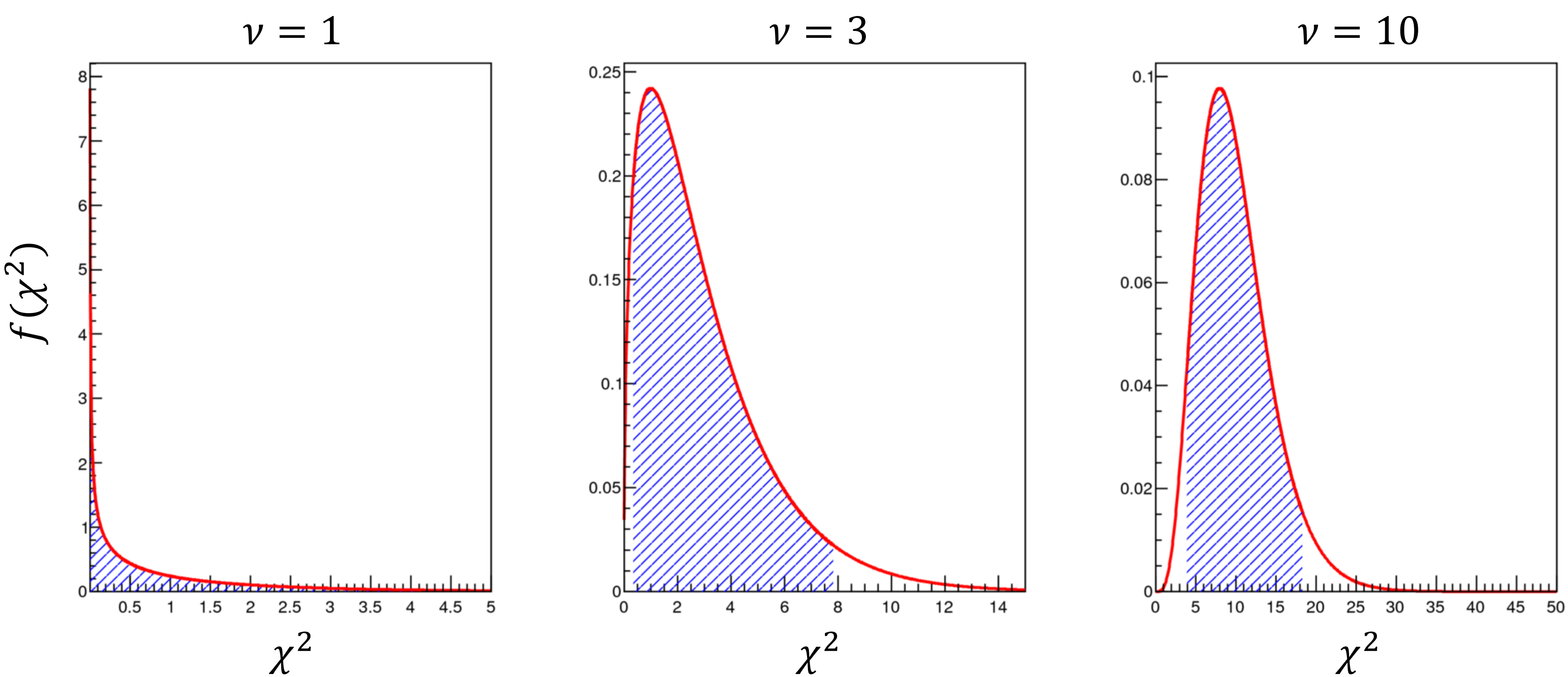
- Se a função ajustada for adequada (isto é, se o gráfico de resíduos não tiver estrutura clara), e a função densidade de probabilidade dos dados forem gaussianas com incertezas conhecidas, a função densidade de probabilidade do χ^2 é dada por:

$$f(\chi^2) = \frac{1}{2\,\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\chi^2}{2}\right)^{\left(\frac{\nu}{2}\right)-1} e^{\frac{-\chi^2}{2}}$$

onde $\Gamma(n)$ é a função gama, definida para n inteiro ou semi-inteiro. Se n for inteiro, $\Gamma(n) = (n-1)!$, mas se n for semi-inteiro, $\Gamma(n) = (n-1)(n-2)\ldots\left(\frac{1}{2}\right)\sqrt{\pi}$.

Teste de hipótese usando o χ^2 (parte II)

- A função densidade de probabilidade de χ^2 tem uma assimetria positiva bastante pronunciada quando o número de graus de liberdade, $\nu = N - P$, é pequeno:



Os testes z e t – intervalos de confiança (I)

- Se os dados seguem uma função densidade de probabilidade gaussiana com desvio-padrão conhecido, σ_x, então a variável aleatória z, definida como

$$z = \frac{x - x_0}{\sigma_x},$$

segue uma função densidade de probabilidade gaussiana com valor médio verdadeiro 0 e desvio-padrão 1.

- Os intervalos de confiança para z mais conhecidos são:

$$P(|z| \leq 1) \cong 68{,}27\%$$

$$P(|z| \leq 2) \cong 95{,}45\%$$

$$P(|z| \leq 3) \cong 99{,}73\%$$

Os testes z e t – intervalos de confiança (II)

- Quando a variável x seguir uma função densidade de probabilidade gaussiana, porém com desvio-padrão desconhecido, o teste é feito com a variável aleatória t,

$$t = \frac{x - x_0}{\tilde{\sigma}_x},$$

onde $\tilde{\sigma}_x$ é o desvio-padrão estimado de x, que segue a função densidade de probabilidade *t de Student*.

- A largura dos intervalos de confiança para t são sempre maiores do que os correspondente intervalos para z.
- Essa diferença diminui quanto maior for o número de graus de liberdade, ν, usados para estimar $\tilde{\sigma}_x$.

Os testes z e t – intervalos de confiança (III)

$$t = \frac{x - x_0}{\tilde{\sigma}_x}$$

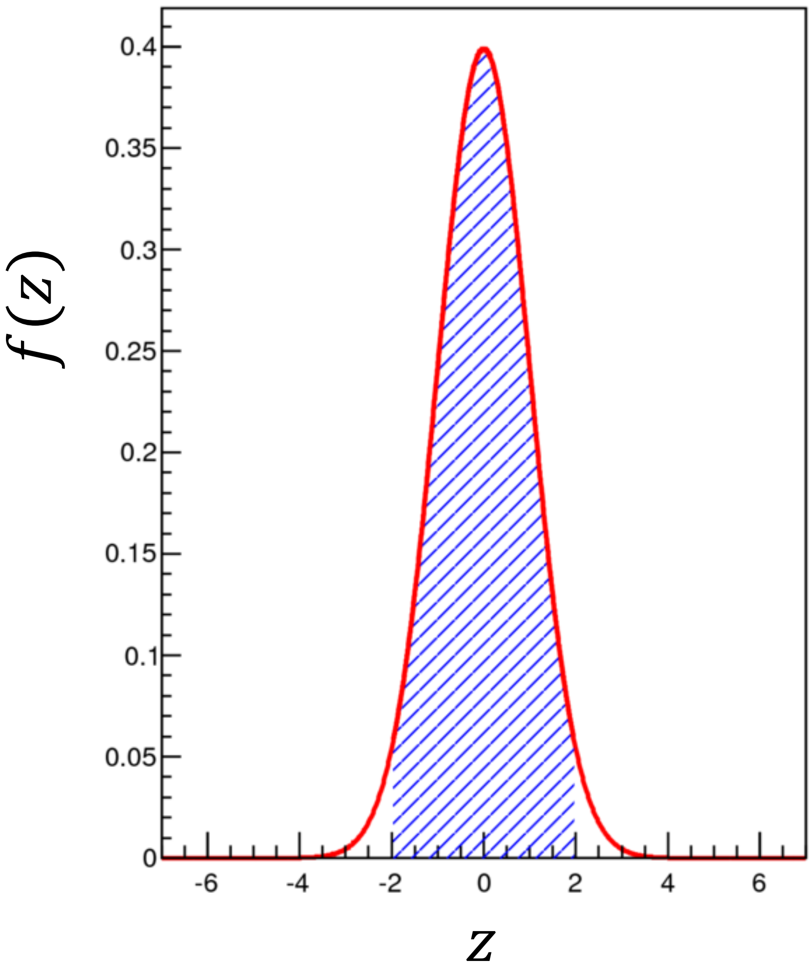
Valores críticos de *t* que definem intervalos de confiança de 68,3%, 95,5% e 99,7%, para alguns casos de número de graus de liberdade, ν, usados para estimar o desvio-padrão, $\tilde{\sigma}_x$.

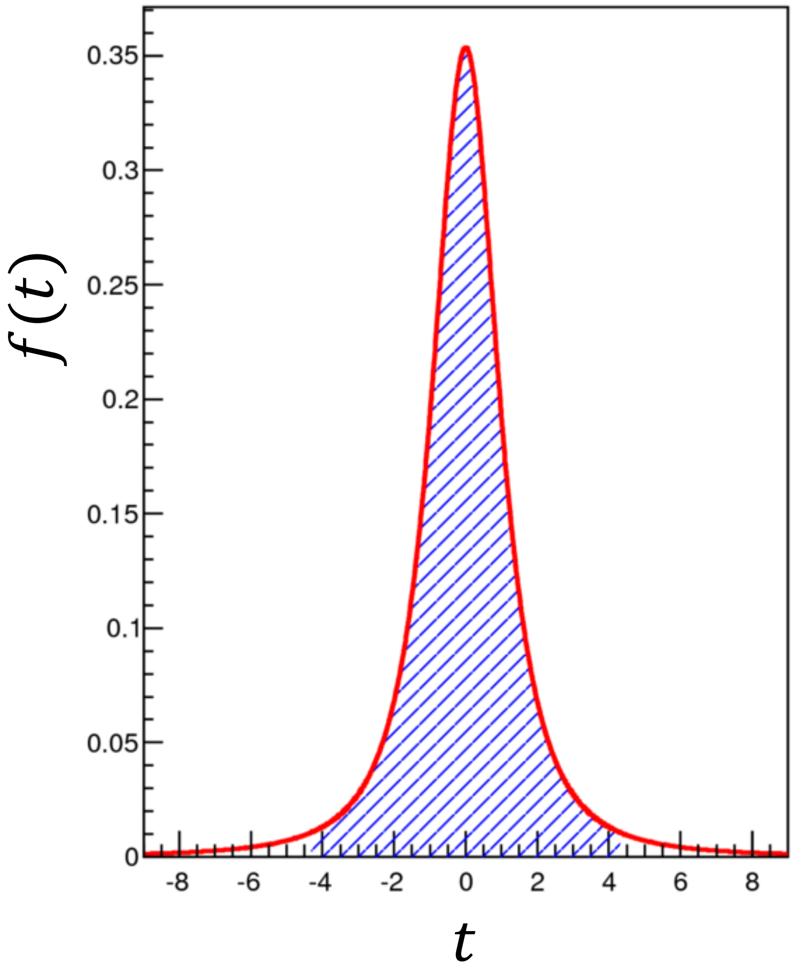
ν	t_1 $\alpha_1 = 68{,}27\ \%$	t_2 $\alpha_2 = 95{,}45\ \%$	t_3 $\alpha_3 = 99{,}73\ \%$
1	1,84	14,0	235,8
2	1,32	4,53	19,21
3	1,20	3,31	9,22
5	1,11	2,65	5,51
10	1,053	2,28	3,96
20	1,026	2,13	3,42
100	1,005	2,03	3,08
∞	1	2	3

As funções densidade de probabilidade de z e t
(com intervalo de confiança de 95%)

$$z = \frac{x - x_0}{\sigma_x}$$

$$t = \frac{x - x_0}{\tilde{\sigma}_x}, \text{ com } \boldsymbol{\nu} = \mathbf{2}$$

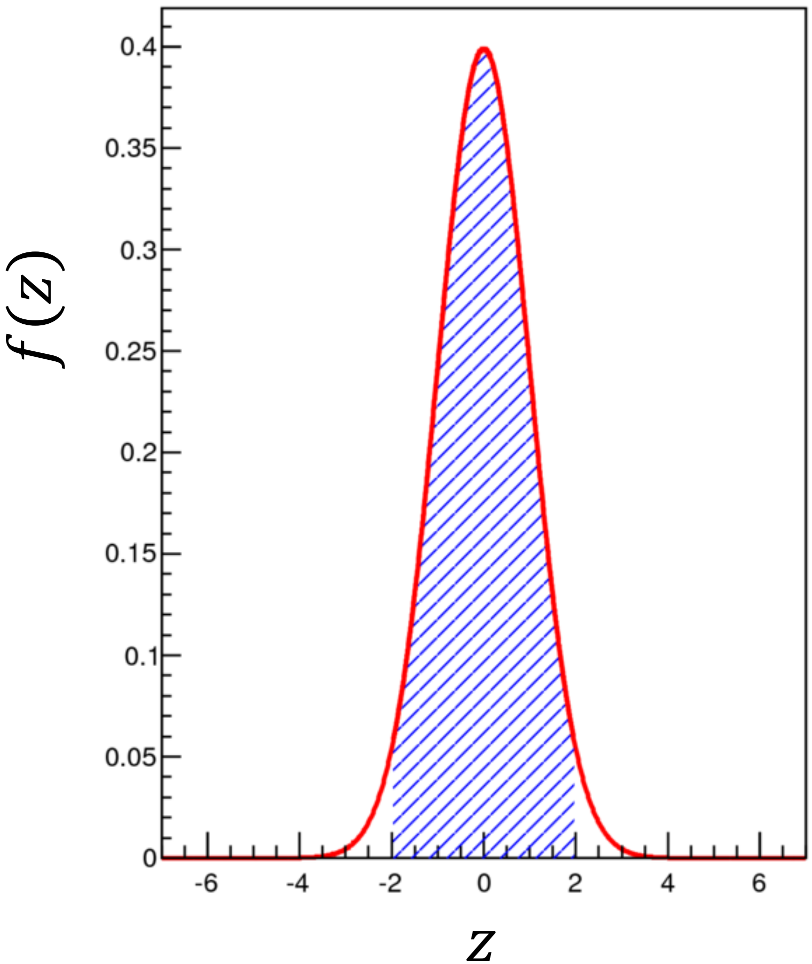


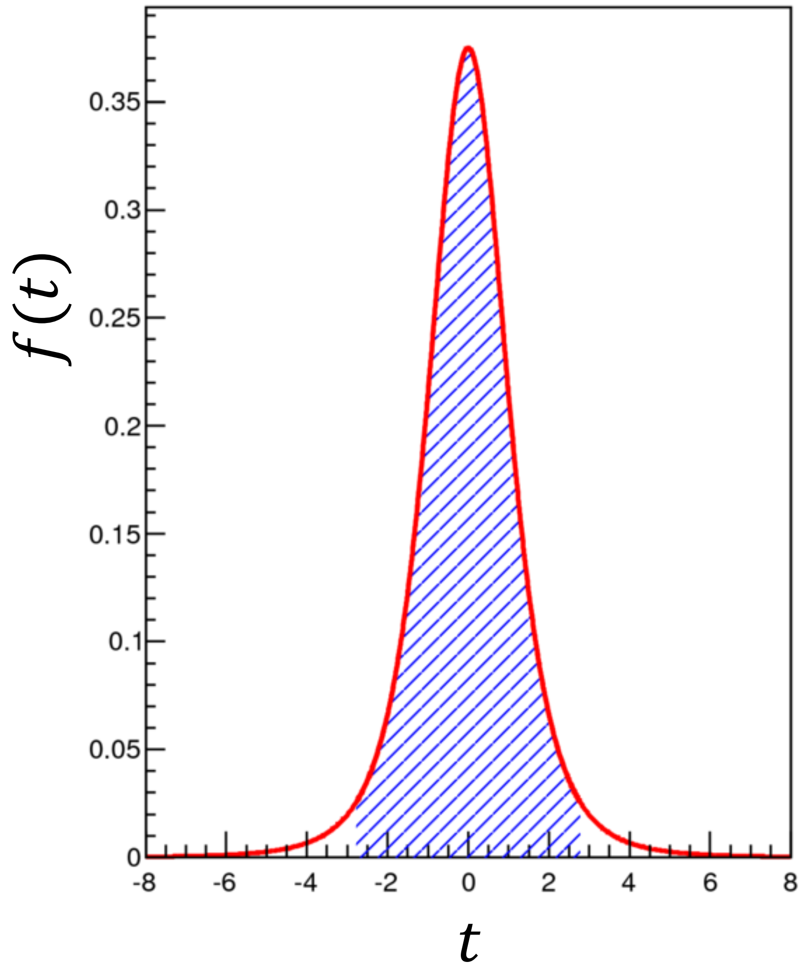


As funções densidade de probabilidade de z e t
(com intervalo de confiança de 95%)

$$z = \frac{x - x_0}{\sigma_x}$$

$$t = \frac{x - x_0}{\tilde{\sigma}_x}, \text{ com } \boldsymbol{\nu} = \mathbf{4}$$





As funções densidade de probabilidade de z e t (com intervalo de confiança de 95%)

$$z = \frac{x - x_0}{\sigma_x}$$

$$t = \frac{x - x_0}{\tilde{\sigma}_x}, \text{ com } \boldsymbol{\nu} = \mathbf{10}$$