Um pouco de estatística…

Caixa de Galton

Uma pergunta recorrente

- Quando eu sei se o χ^2 de um ajuste é bom ou ruim?

Letra grega, pronuncia-se chi (quadrado)

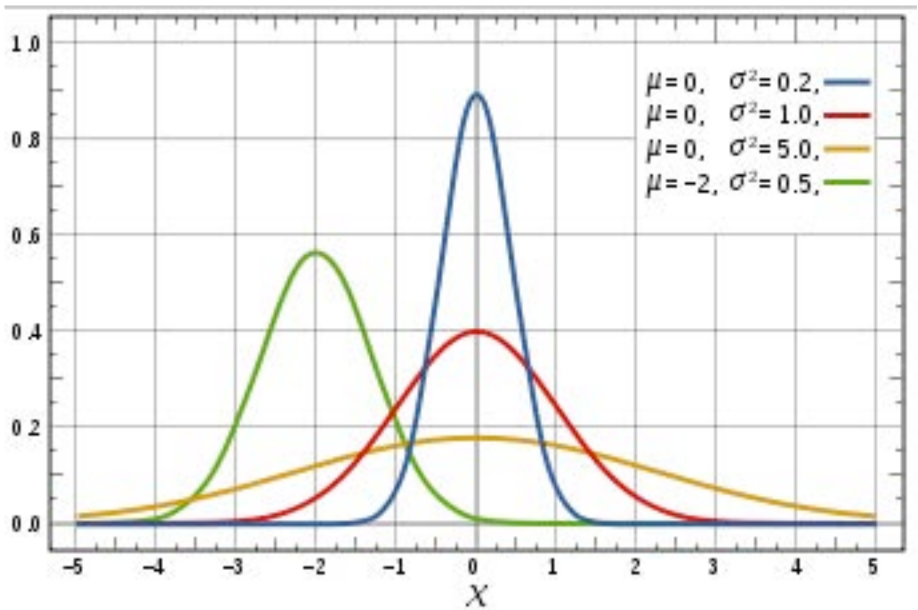
  - $\chi^2_{red} \sim 1$
    - Ordem X de grandeza apenas ou tem algum intervalo numérico definido?

  - E se o χ^2 for muito grande ou muito pequeno?

Funções de densidade de probabilidade (F.D.P.)

- A F.D.P. mais conhecida é a distribuição normal, ou gaussiana.

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



F.D.P. de χ^2 .

- Vamos realizar um experimento virtual, no qual iremos simular a obtenção da amostra $X_{ndf} = \{x_i\}$.

- A partir dessa amostra vamos calcular a variável de interesse (no nosso caso, o χ^2)

- Vamos repetir esse experimento virtual um número muito grande de vezes de modo a obter as F.D.P. das variáveis estudadas.

Para simplificar o problema

- Ao invés de usar a amostra X, vamos fazer uma mudança de variável tal que

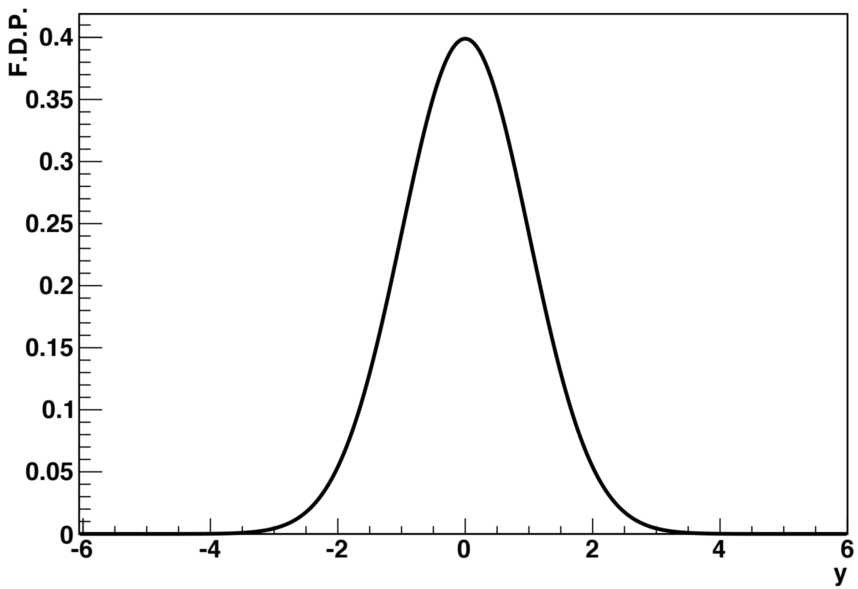
$$Y = \{y_i\} \rightarrow y_i = \frac{x_i - \mu}{\sigma_\mu}$$

- Ou seja, vamos estudar uma amostra de valor verdadeiro 0 e variância 1.
  - .Para F.D.P. com médias e variâncias diferentes, basta uma mudança de escala.

F.D.P. de *y*.

- Y segue uma distribuição normal de valor verdadeiro 0 e variância 1, ou seja:

$$p(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$



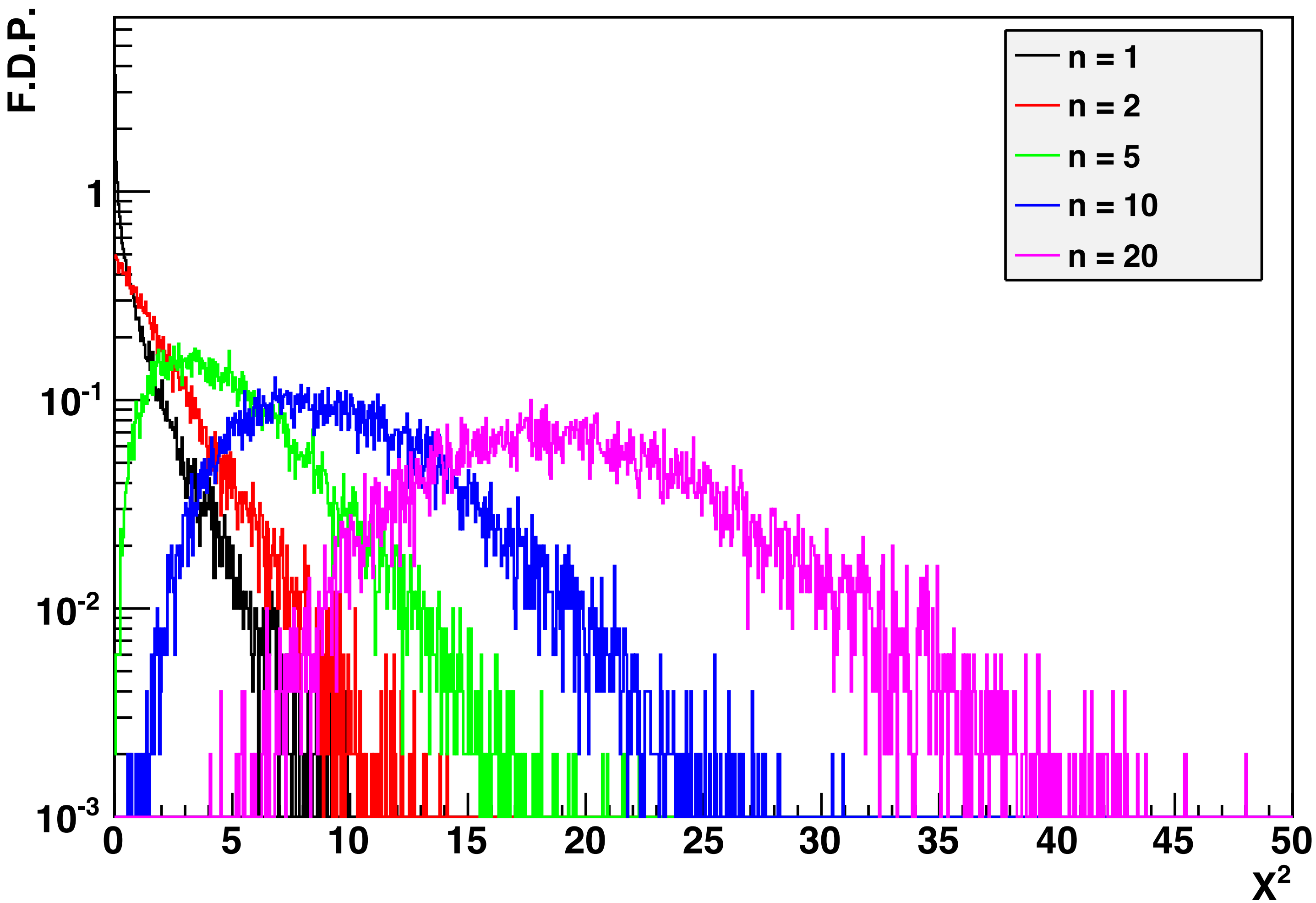
F.D.P. de χ^2 .

- A função χ^2 é definida como:

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma_\mu} \right)^2 = \sum_{i=1}^{n} y_i^2$$

- A F.D.P. é obtida calculando o valor de χ^2 para cada conjunto de dados simulado

F.D.P. de χ^2 .



F.D.P. de χ^2 .

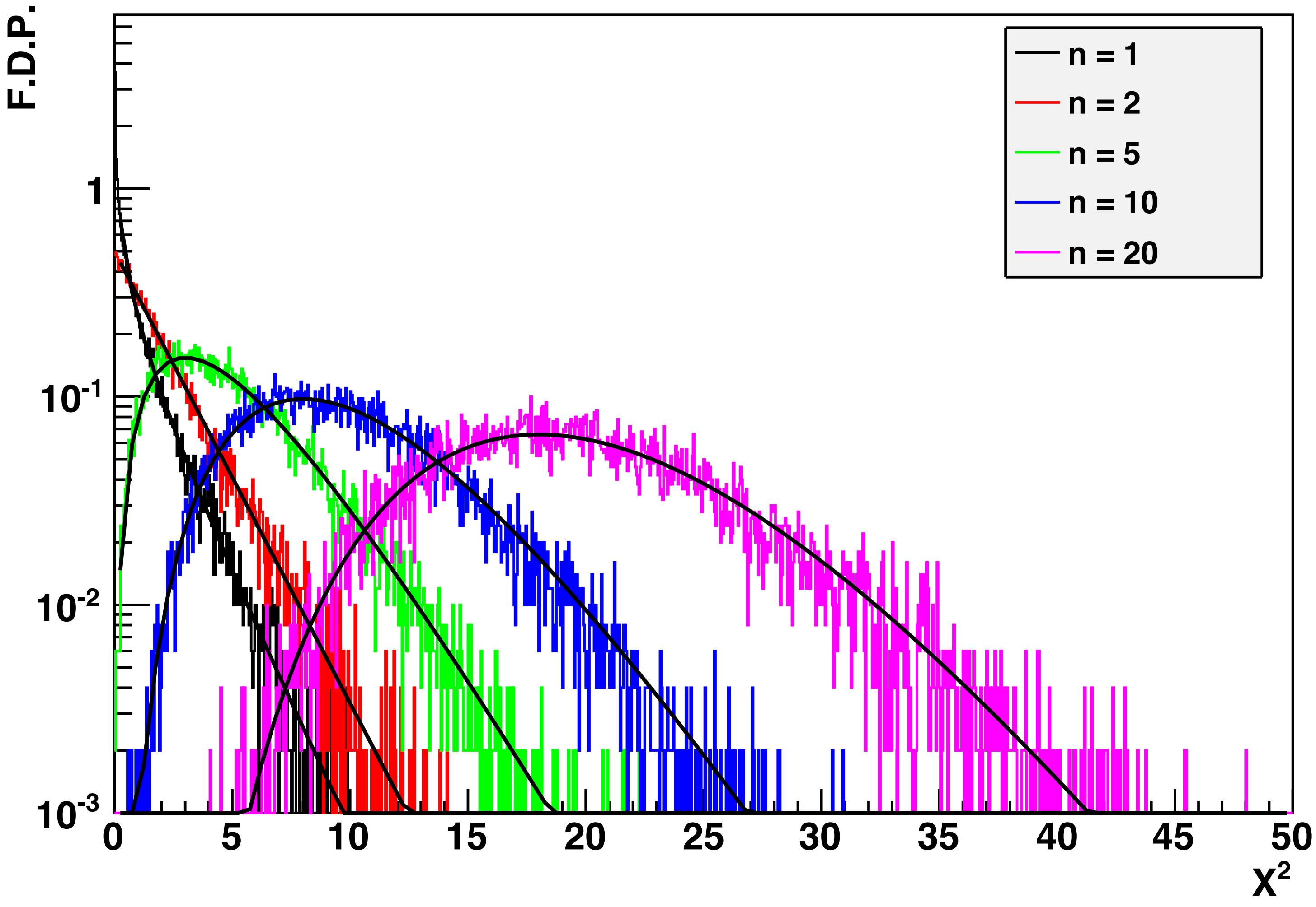
- A F.D.P. não segue mais uma distribuição normal:

$$p(\zeta) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\zeta^{\frac{n}{2}-1}e^{-\frac{\zeta}{2}}$$

- Onde Γ é a função gama, n é o número de graus de liberdade e ζ, o valor de χ^2 .

Letra grega, pronuncia-se “zeta”

F.D.P. de χ^2 .



F.D.P. de χ^2_{red} e σ.

- A função χ^2_{red} é definida como:

$$\chi^2_{\mathrm{red}} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\frac{\mathrm{x_i} - \mu}{\sigma_\mu} \right)^2 = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{y}_\mathrm{i}^2$$

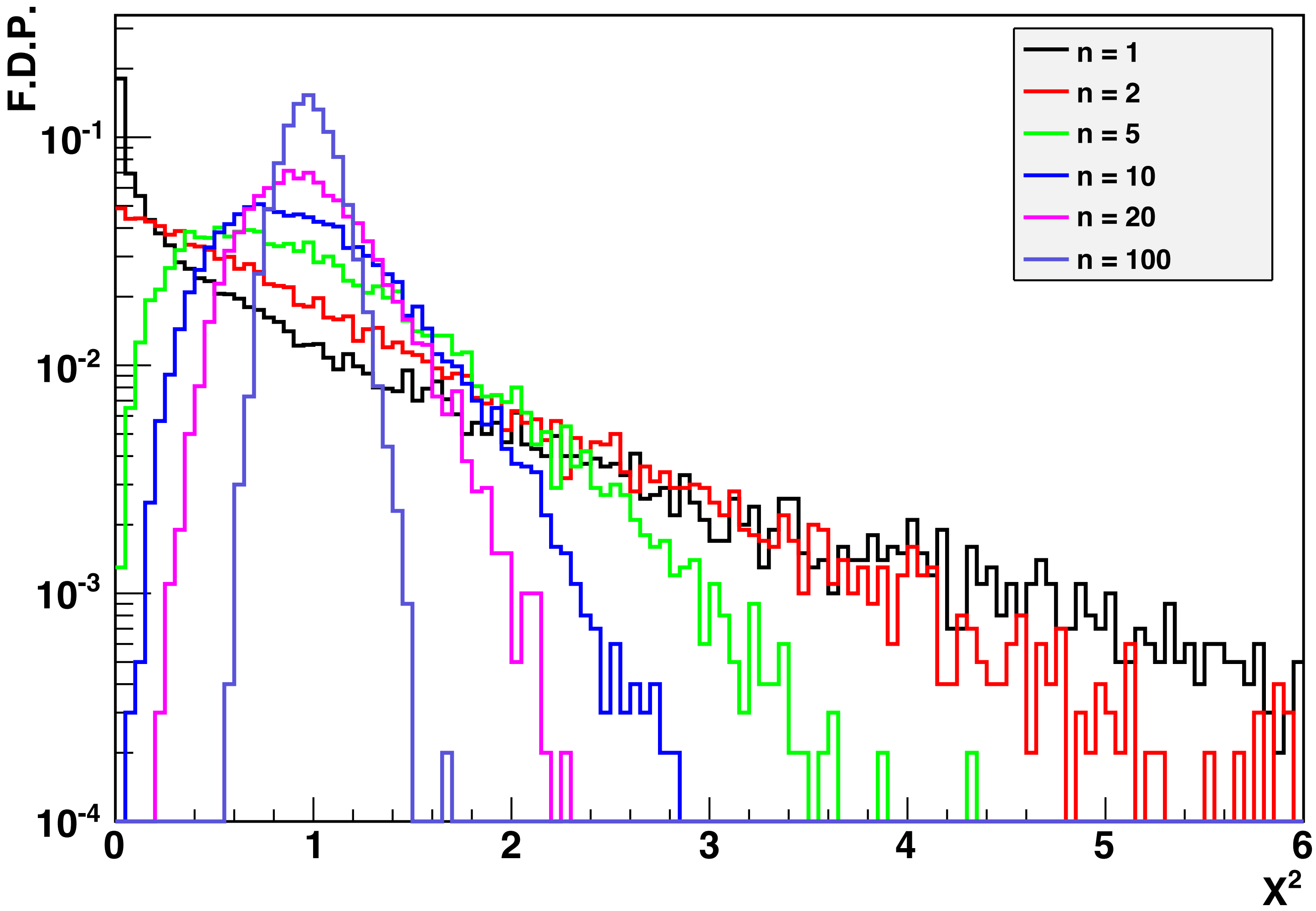
- Por outro lado, a variância de um conjunto de medidas é:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \mu \right)^2 = \frac{1}{n} \sum_{i=1}^{n} y_i^2$$

- Ou seja, essas grandezas são muito similares e seguem a mesma F.D.P.

F.D.P. De χ^2_{red} e σ.

Note que quanto maior o número de graus de liberdade, mais estreita é a distribuição.



F.D.P. de χ^2_{red} e σ.

- χ^2_{red} e σ são importantes em testes de significância.
  - A função χ^2_{red} é calculada quando se faz um ajuste de curvas. Como avaliar se o ajuste é bom?
  - Em uma medida estatística, como saber se a variância que estou obtendo é representativa?

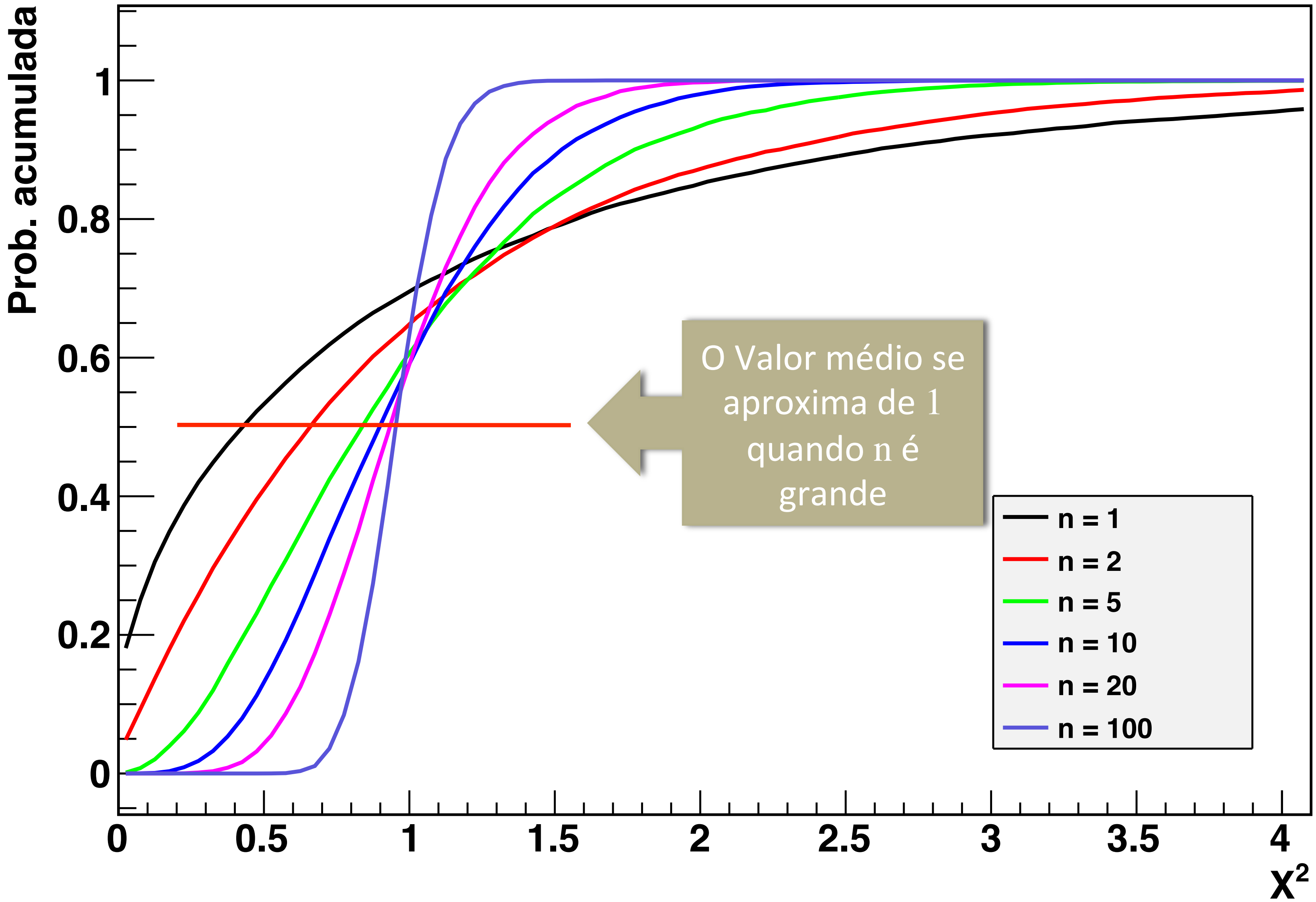
Probabilidade acumulada

- Podemos definir como probabilidade acumulada a grandeza:

$$P(x) = \int_{-\infty}^{x} p(t)dt \rightarrow \begin{cases} \text{Probabilidade da variável } t \\ \text{assumir um valor menor ou} \\ \text{igual a } x. \end{cases}$$

- Essa grandeza é particularmente útil para definir intervalos de confiança
  - Ex: qual o intervalo de 95% de confiança para a distribuição de χ^2_{red} de um ajuste com 5 graus de liberdade?

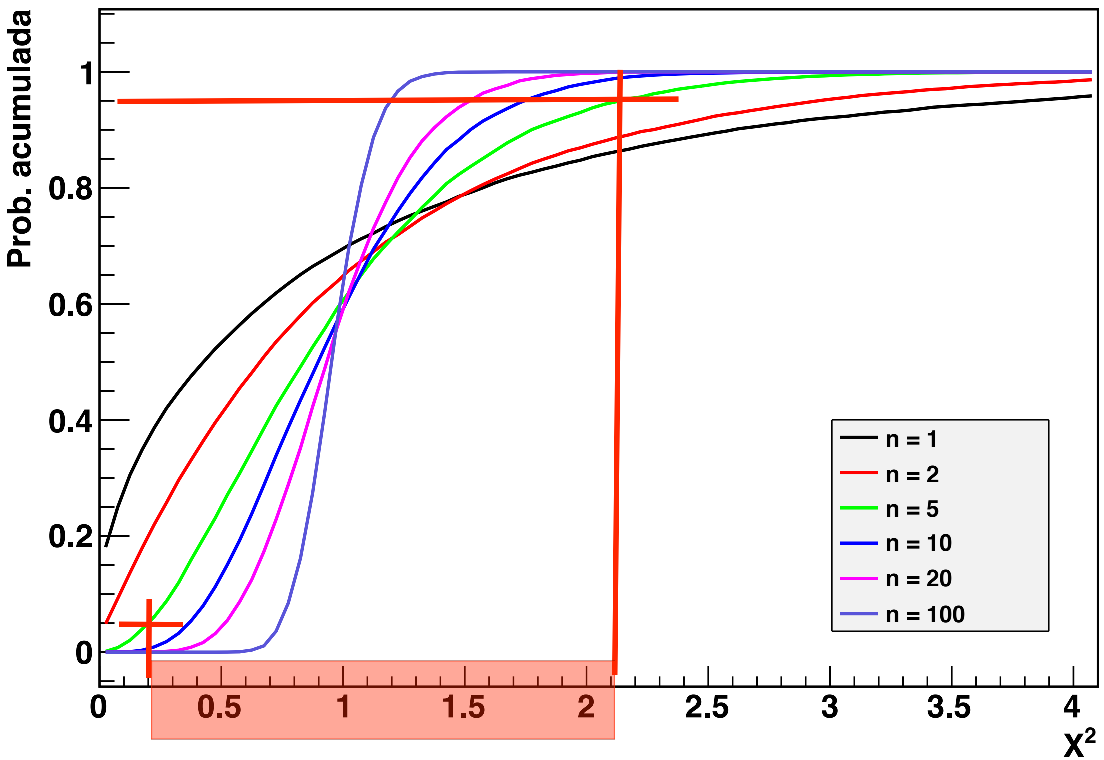
Probabilidade acumulada de χ^2_{red} (ou σ)



Probabilidade acumulada de χ^2_{red} (ou σ)

- Se eu faço um ajuste de **5 graus de liberdade**, qual o intervalo esperado de χ^2_{red} com 90% de confiança?

$$0.2 < \chi^2_{\mathrm{red}} < 2.1$$



Moral da história

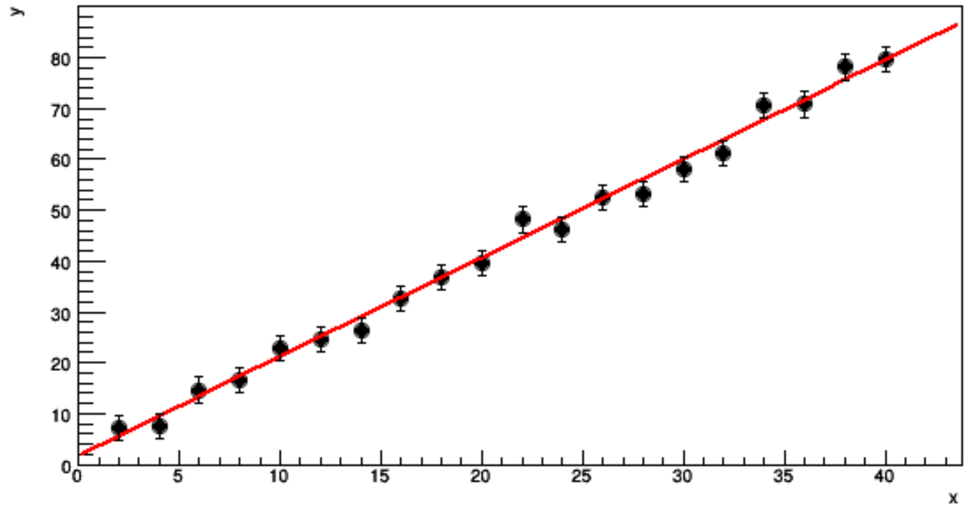
- Testes de chi-quadrado e análise de resíduos (análise mesmo, não apenas fazer um gráfico) constituem ferramentas poderosas na validação de resultados experimentais.

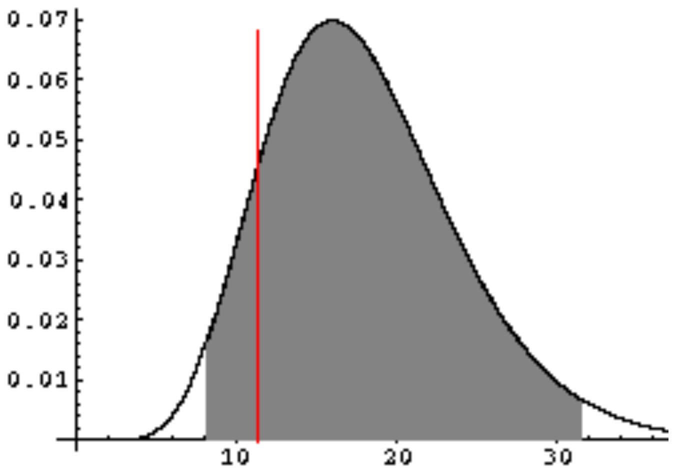
- Cálculo de intervalo de confiança para χ^2, teste-z, teste-t, etc.
  - No WebROOT -> Calculadoras

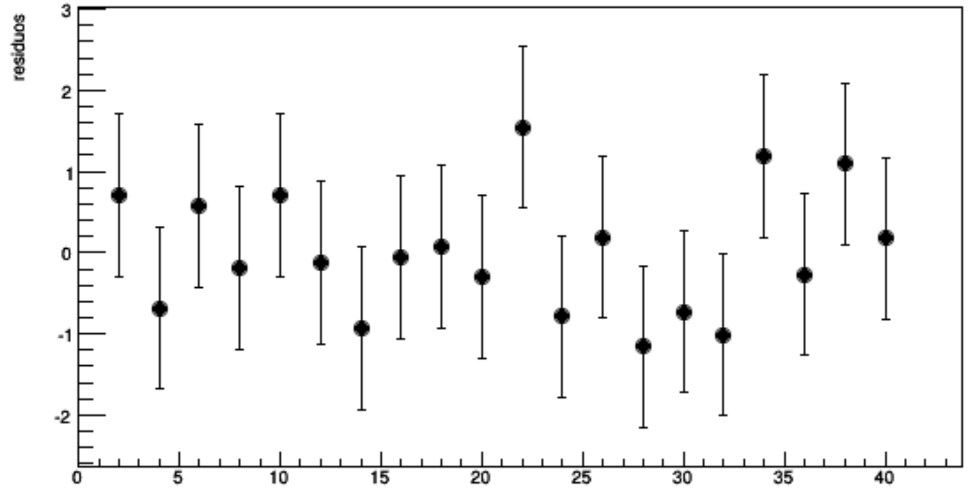
- E se o valor de χ^2 estiver fora do intervalo de significância?
  - Em geral:
    - Incertezas super/subestimadas
    - Modelo teórico não se aplica aos dados

Ex: Um bom ajuste

- $\chi^2 = 11.45$
- NDf = 18
- 95% CL = 8.2 < chi2 < 32

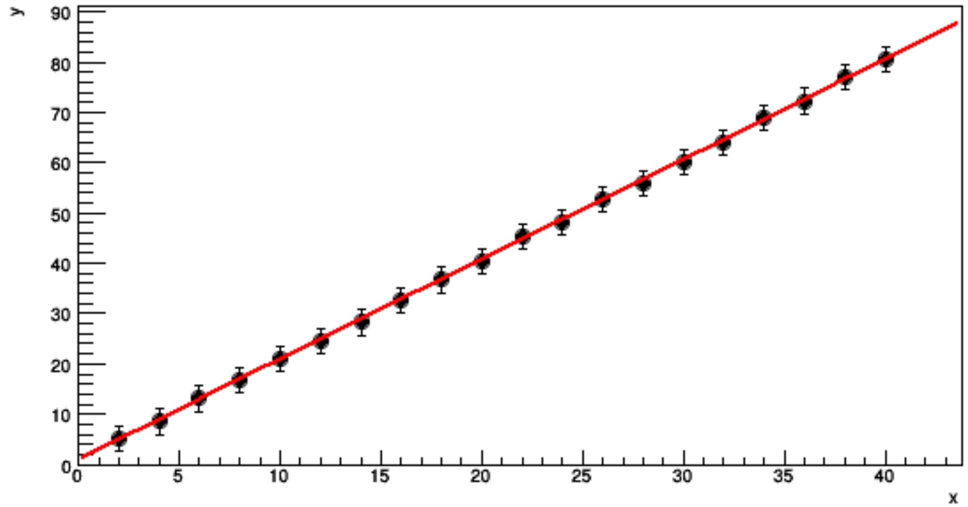


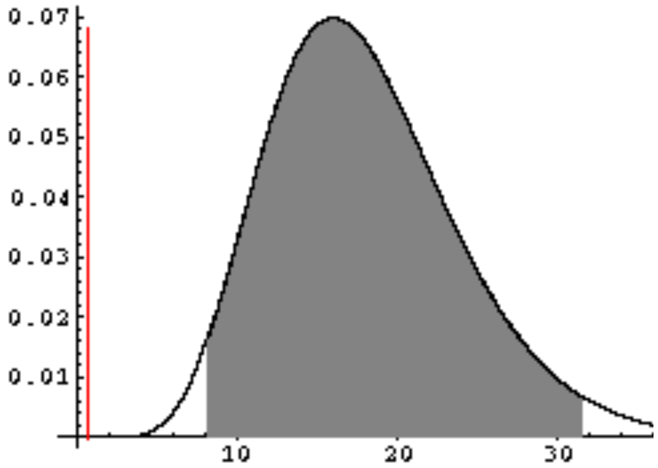


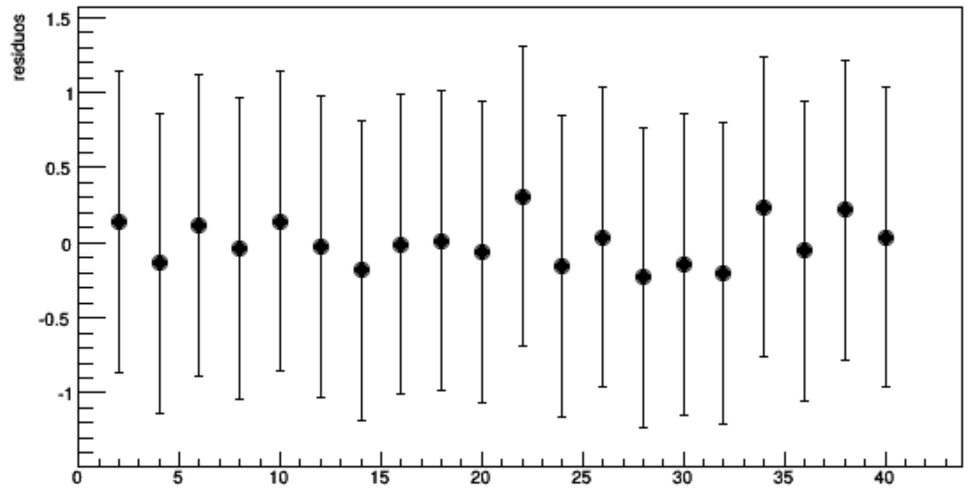


Ex: Incertezas superestimadas

- $\chi^2 = 0.45$
- NDf = 18
- 95% CL = 8.2 < chi2 < 32

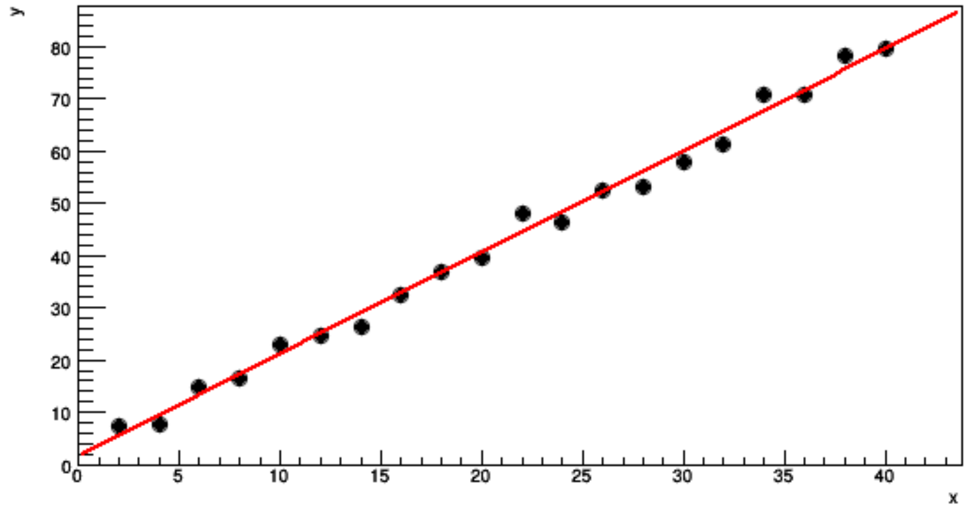


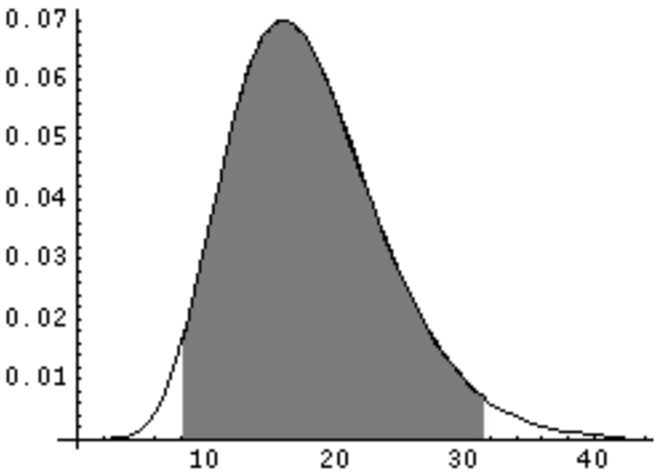


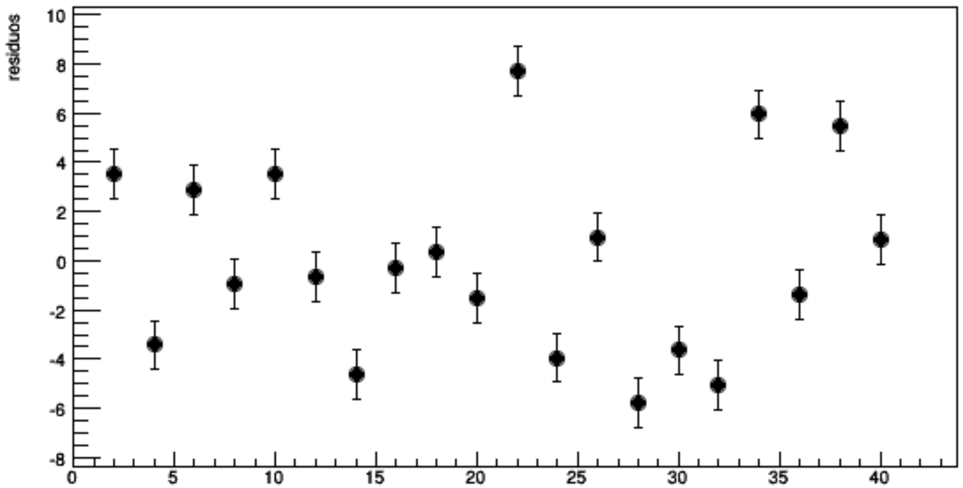


Ex: Incertezas subestimadas

- $\chi^2 = 286$
- NDf = 18
- 95% CL = 8.2 < chi2 < 32

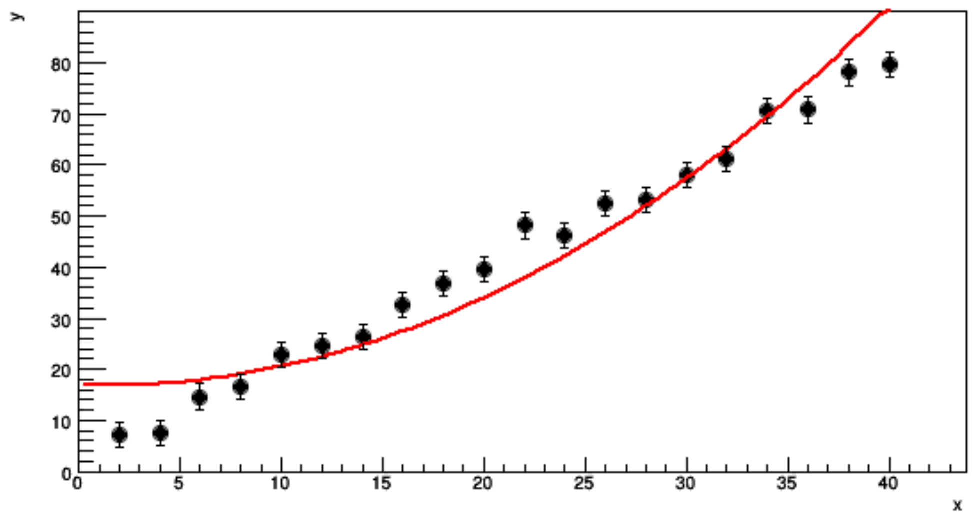


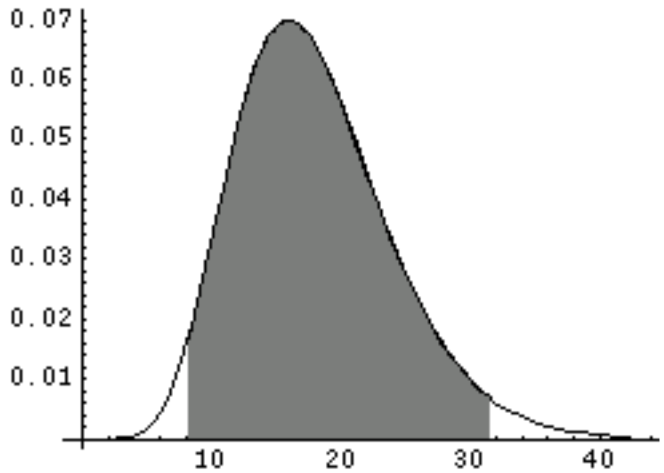


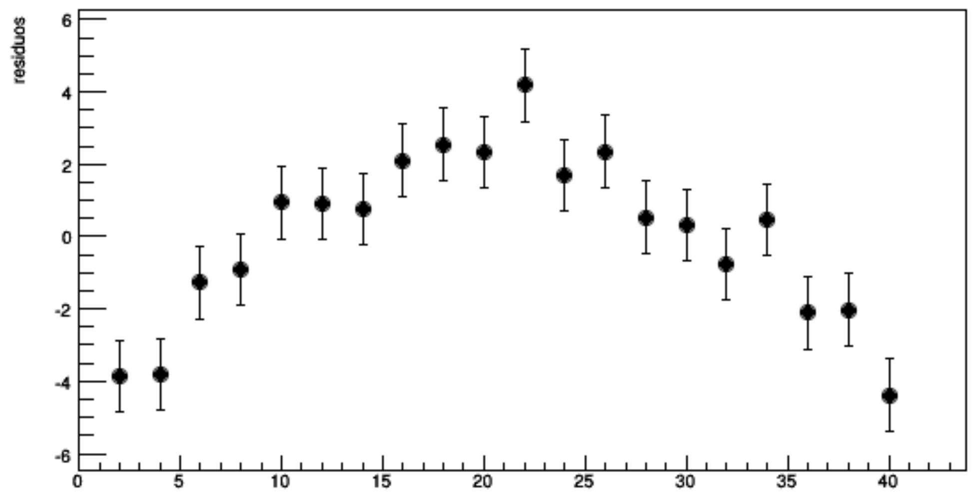


Ex: Função incorreta

- χ^2 = 105
- NDf = 18
- 95% CL = 8.2 < chi2 < 32







Alguns comentários importantes

- Relatório a ser entregue após a semana santa
  - Ver data no calendário no site da disciplina
  - Upload no site de reservas em arquivo PDF

- O que vamos observar nos relatórios
  - Formato apropriado para textos científicos
    - Fiquem atentos a referenciamento, legendas, algarismos significativos, etc. → grande problema semestre passado
    - NÃO MAIS DO QUE 10 PÁGINAS → exercitem o poder de síntese de vocês
  - Conteúdo
    - Análise estatística consistente → não basta calcular incertezas, χ^2, etc. → Interprete a qualidade dos seus dados corretamente
    - Comparação com os resultados dos colegas
      - Como seu resultado se insere no contexto da turma?
    - Conclusões consistentes com objetivos iniciais.