Introdução à Análise de dados em Física de Altas Energias

Exercícios de Estatística

Professores: Dilson de Jesus Damião, Mauricio Thiel e Eliza Melo

Aluno: Thiago Henrique de Sousa

(08/10/2024)

EXERCÍCIO 1

A função S é dada por:

$$S(a,b) = \sum_{i=1}^{N} \frac{[y_i - (ax_i + b)]^2}{\sigma_i^2} \tag{1}$$

Para ajustar de forma linear a reta $y(x) = ax+b$ aos N pares (x_i, y_i), onde as incertezas de y são diferentes em cada evento, devemos minimizar os resíduos, ponderados pela função $S(a,b)$. Expandindo o quadrado, temos:

$$S(a,b) = \sum_{i=1}^{N} \left(\frac{y_i^2}{\sigma_i^2} - 2a\frac{x_i y_i}{\sigma_i^2} - 2b\frac{y_i}{\sigma_i^2} + a^2\frac{x_i^2}{\sigma_i^2} + 2ab\frac{x_i}{\sigma_i^2} + b^2\frac{1}{\sigma_i^2} \right) \tag{2}$$

Reescrevendo os termos em termos das médias ponderadas:

$$S(a,b) = \sum_{i=1}^{N} \frac{y_i^2}{\sigma_i^2} - 2a\sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2} - 2b\sum_{i=1}^{N} \frac{y_i}{\sigma_i^2} + a^2\sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} + 2ab\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} + b^2\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \tag{3}$$

Para encontrar os valores de a e b, é calculado as seguintes derivadas parciais e igualando a zero: $\frac{\partial S}{\partial a}$ e $\frac{\partial S}{\partial b}$.

$$\frac{\partial S}{\partial a} = -2\sum_{i=1}^{N} \frac{x_i(y_i - ax_i - b)}{\sigma_i^2} = 0 \tag{4}$$

$$\frac{\partial S}{\partial b} = -2\sum_{i=1}^{N} \frac{(y_i - ax_i - b)}{\sigma_i^2} = 0 \tag{5}$$

Rearrumando as equações:

$$\sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2} - a\sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} - b\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} = 0 \tag{6}$$

$$\sum_{i=1}^{N} \frac{y_i}{\sigma_i^2} - a\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} - b\sum_{i=1}^{N} \frac{1}{\sigma_i^2} = 0 \tag{7}$$

Com isso, obtemos o seguinte sistema de equações:

$$
\begin{cases}
\bar{x^2}a + \bar{x}b = x\bar{y} \\
\bar{x}^2 a + b = \bar{y}
\end{cases}
\tag{8}
$$

Resolvendo esse sistema para a e b, temos:

$$
\begin{cases}
a = \frac{\bar{x}\bar{y} - \bar{x}\bar{y}}{\bar{x}^2 - \bar{x}^2} \\
b = \bar{y} - a\bar{x}
\end{cases}
\tag{9}
$$

EXERCÍCIO 2

O cálculo da seção de choque, σ, é dado por:

$$\sigma = \frac{N_{Total} - N_{Background}}{\mathcal{L}} \tag{10}$$

onde, a questão fornece os seguintes dados:

1. $N_{Total} = 2567$;
2. $N_{Background} = 1223.5$;
3. $\mathcal{L} = 25fb^{-1}$.

Sendo N_{Total} todos os eventos observados, $N_{Background}$ é o número de eventos de fundo esperados e $\mathcal{L}$ é a luminosidade integrada. Substituindo esses valores fornecidos na equação 10, temos:

$$\sigma = \frac{2567 - 1223.5}{25} = \frac{1343.5}{25} \rightarrow \sigma = 53.74fb \tag{11}$$

O cálculo da incerteza estatística é dado pela distribuição de Poisson, usando a fórmula:

$$\sigma_{stat}^2 = \left(\frac{\sqrt{N_{Total}}}{\mathcal{L}}\right)^2 + \left(\frac{\sqrt{N_{Background}}}{\mathcal{L}}\right)^2 \tag{12}$$

Substituindo os valores do problema, temos:

$$\sigma_{stat}^2 = \left(\frac{\sqrt{2567}}{25}\right)^2 + \left(\frac{\sqrt{1223.5}}{25}\right)^2 \tag{13}$$

$$\sigma_{stat} = \sqrt{6.07} \simeq 2.46fb \tag{14}$$

O problema diz que a incerteza sistemática deve ser de 10%, logo:

$$\sigma_{sist} = 0.10 \times \sigma = 0.10 \times 53.74 = 5.37fb \tag{15}$$

A seção de choque do sinal e suas incertezas associadas é:

$$\sigma = 53.74 \pm 2.46_{stats} \pm 5.37_{sist}$$

EXERCÍCIO 3

O cálculo dos eventos esperados, após todos os cortes, utilizando uma pdf de Poisson com 95% de confiança (C.L.), é dado a partir da fórmula:

$$P(n;x) = \frac{e^{-x}x^{n}}{n!} \tag{16}$$

Para um número n de eventos esperados. Como temos n=0, a nossa equação fica:

$$P(0;x) = \frac{e^{-x}x^{0}}{0!} = e^{-x} \tag{17}$$

Para um nível de confiança de 95%, temos:

$$P(0;x) \geq 0.05 \rightarrow e^{-x} \geq 0.05 \tag{18}$$

Fazendo o logaritmo natural de ambos os lados da equação, temos:

$$-x \geq ln(0.05) \rightarrow x \leq -ln(0.05) \tag{19}$$

$$-ln(0.05) \simeq 2.9957 \tag{20}$$

Podemos concluir que para um nível de confiança de 95% o número de eventos esperados é $x \geq 2.9957$. O número de eventos esperados deve ser de um número inteiro, logo, podemos assumir que teremos 3 eventos esperados.

EXERCÍCIO 4

A função de χ^2 é:

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - f(x_i))^2}{\sigma_i^2} \tag{21}$$

onde o y_i são os dados observados, $f(x_i)$ é o valor que vou predito pelo modelo e σ_i é a incerteza associada a y_i.

O número de graus de liberdade (ndf) é dado por: $ndf = N - p$, onde N é o número total de dados e p o parâmetro de ajuste. A média e a variância dessa distribuição são dadas por, respectivamente:

$$Med[\chi^2] = ndf \tag{22}$$

$$Var(\chi^2) = 2ndf \tag{23}$$

Para um ajuste ideal, os dados devem apresentar:

$$Mdf[\chi^2/ndf] = \frac{Med[\chi^2]}{ndf} = \frac{ndf}{ndf} = 1 \tag{24}$$

Isso mostra que para um ajuste perfeito $\chi^2/ndf \rightarrow 1$. Se esse valor for maior, ou muito maior, que 1 significa que a soma dos quadrados dos resíduos é maior que o esperado, ou seja, o ajuste não está feito de forma correta para os dados. Se for menor, ou muito menor, significa que o ajuste é muito simples para os dados analisados ou que não consideraram as incertezas com a importância necessária.