Introdução à Análise de Dados FAE
Trabalho 2

Diogo Caffonso
Matrícula: 202010115711

Exercício 1:

Para o método dos mínimos quadrados, existe uma função $S(a,b)$, sendo a e b os coeficientes angular e linear da reta proposta, respectivamente, tal que esta deva ser minimizada para garantir a melhor reta que se adéque aos pontos. Para quando se tratam de dados tais que os erros para y_i variam, isto é, para o MMQ ponderado, a função $S(a,b)$ é dada por:

$$S(a,b) = \sum_{i=1}^{N} \left[\frac{y_i - (ax_i + b)}{\sigma_i} \right]^2 > 0$$

Expandindo a equação, pode-se ver que:

$$\frac{1}{\sigma^2} = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \quad \& \quad w_i = \left(\frac{\sigma}{\sigma_i} \right)^2 \rightarrow \sum_{i=1}^{N} w_i = 1$$

Visto que σ^2 e $\overline{x^2}$ são positivos, a e b são dados pelas mesmas expressões que no caso de um ajuste linear não ponderado, isto é:

$$\begin{cases} a = \frac{\sigma_{xy}}{\sigma_x^2} = r\frac{\sigma_y}{\sigma_x} \\ b = \overline{y} - a\overline{x} \end{cases}$$

Reescrevendo os parâmetros, temos:

$$\begin{cases} a = \frac{\sigma^2}{\sigma_x^2} \sum_{i=1}^{N} \frac{(x_i - \overline{x})}{\sigma_i^2} y_i \\ b = \sum_{i=1}^{N} \frac{\sigma^2}{\sigma_x^2} y_i - a\overline{x} \end{cases}$$

Agora, de acordo com a fórmula de propagação de erros, as incertezas, desta forma, são dadas por:

$$\begin{cases} \sigma_a^2 = \frac{\sigma^2}{\sigma_x^2} \\ \\ \sigma_b^2 = \overline{x^2}\sigma_a^2 \end{cases}$$

E, para minimizar a função $S(a, b)$, devem ser tiradas suas derivadas parciais em relação a a e b e igualar a zero, ou seja:

$$\begin{cases} \frac{\partial S}{\partial a} = -2\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} \left(y_i - ax_i - b\right) = 0 \\ \frac{\partial S}{\partial b} = -2\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left(y_i - ax_i - b\right) = 0 \end{cases}$$

Este sistema pode ser reescrito da seguinte forma:

$$\begin{cases} \left(\sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2}\right) a + \left(\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}\right) b = \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2} \\ \left(\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}\right) a + \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\right) b = \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2} \end{cases}$$

e pode ser simplificado para:

$$\begin{cases} \overline{x^2}a + \overline{x}b = \overline{xy} \\ \overline{x}a + b = \overline{y} \end{cases}$$

que, por fim, entrega as seguintes soluções para a e para b:

$$a = \frac{\sigma_{xy}}{\sigma_x^2} \quad \& \quad b = \overline{y} - a\overline{x}$$

com incertezas:

$$\sigma_a = \frac{\sigma}{\sigma_x} \quad \& \quad \sigma_b = \sigma_a\sqrt{\overline{x^2}}$$

Exercício 2:

Sabemos que:

$$\sigma = \frac{N_{\mathrm{T}} - N_{\mathrm{B}}}{\mathcal{L}}$$

Onde:

$N_{\text{total}} = N_T = 2567,0$ eventos
$N_{\text{background}} = N_B = 1223,5$ eventos
$\mathcal{L} = 25\ \text{fb}^{-1}$

Para a incerteza estatística $\Delta\sigma_{est}$, faz-se:

$$\Delta\sigma_{est} = \frac{\sqrt{\Delta N_T^2 + \Delta N_B^2}}{\mathcal{L}}$$

em que, para a distribuição de Poisson: $\Delta N_T = \sqrt{N_T}$ e $\Delta N_B = \sqrt{N_B}$.

Agora, para a incerteza sistemática $\Delta\sigma_{sis}$, basta fazer:

$$\Delta\sigma_{sis} = 10\%\sigma$$

que leva ao resultado:

$$\sigma = \left(53,74 \pm 2,46_{(\text{estatística})} \pm 5,37_{(\text{sistemática})}\right)\ \text{fb}$$

Exercício 3:

A distribuição de Poisson pode ser dada pela forma:

$$P_{(n;s,b)} = \frac{(s+b)^n e^{-(s+b)}}{n!}$$

Do problema, temos que $n = 0$ e $b = 0,07$. Assim, a equação se reduz a:

$$P_{(n=0;s,b=0,07)} = e^{-(s+0,07)}$$

Agora, para calcular até quantos eventos esperados podemos excluir com 95% de C.L, fazemos:

$$P_{(n=0;s,b=0,07)} = e^{-(s+0,07)} = 100\% - 95\% = 0,05$$

e, resolvendo para s, encontramos:

$$s \simeq 2,93$$

Ou seja, são previstos mais que 2,93 eventos.

Exercício 4:

Para demonstrar que χ^2/ndf tender a 1 indica um bom fit, parte-se de:

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

para incertezas iguais:

$$\chi^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - f(x_i))^2$$

Para quando $n \to \infty$, σ^2 pode ser aproximado por:

$$\sigma^2 \simeq \frac{1}{n} \sum_{i=1}^{n} (y_i - f_v(x_i))^2 \to n \simeq \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - f_v(x_i))^2$$

em que $f_v(x_i)$ é a função verdadeira. Por simplicidade, $f(x_i) = f$. Então:

$$n \simeq \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - f_v + f - f)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} [(y_i - f) + (f - f_v)]^2 =$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - f)^2 + \frac{1}{\sigma^2} \sum_{i=1}^{n} (f - f_v)^2 + \frac{2}{\sigma^2} \sum_{i=1}^{n} (y_i - f)(f - f_v)$$

Pode-se ver que o primeiro termo é igual a χ^2. O terceiro termo se anula para quando $n \to \infty$. Para o segundo termo, temos que

$(f - f_v)^2 \simeq \sigma_f^2$, que é a variância para a função ajustada, que também pode ser escrita da forma:

$$\sigma_f^2 = f_1^2 \sigma_{a_1}^2 + \ldots + f_p^2 \sigma_{a_p}^2 = \sum_{k=1}^{p} f_k^2 \sigma_{a_k}^2 = \sum_{k=1}^{p} f_k^2 m_{kk}$$

assim, podemos ver que:

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (f - f_v)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} \sigma_f^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} \sum_{k=1}^{p} f_k^2 m_{kk} =$$

$$= \sum_{k=1}^{p} \sum_{i=1}^{n} \frac{f_k^2}{\sigma^2} m_{kk} = \sum_{k=1}^{p} M_{kk} m_{kk}$$

e, já que os parâmetros a_k são independentes, $M_{kk} m_{kk} = 1$. Ou seja:

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (f - f_v)^2 = \sum_{k=1}^{p} 1 = p$$

Assim, substituindo os três termos de volta, obtemos:

$$n \simeq \chi^2 + p + 0 \rightarrow \chi^2 \simeq n - p = ndof \rightarrow \chi^2 \simeq ndof \rightarrow$$

$$\rightarrow \frac{\chi^2}{ndof} \simeq 1$$

que mostra que, para quando χ^2/ndf tende a 1, o fit de uma dada função f se mostra satisfatório.