Introdução à Análise de dados em FAE **(30/08/2024)**

Exercicios de estatística para análise de dados em HEP

Professores: Eliza Melo, Dilson Damião e Mauricio Thiel *Name:* Thiago de Andrade Rangel Monteiro

EXERCICIO 1

Para um ajuste linear da reta $y(x) = ax + b$ aos N pares (x_i, y_i), onde as incertezas são distintas, os resíduos são ponderados pela função $\chi^2(a, b)$, que deve ser minimizada. A função χ^2 é dada por:

$$\chi^2(a,b) = \sum_{i=1}^{N} \frac{[y_i - (ax_i + b)]^2}{\sigma_i^2} \tag{0.1}$$

Expandindo o quadrado e reescrevendo a função χ^2:

$$\chi^2(a,b) = \sum_{i=1}^{N} \frac{y_i^2 - 2(ax_i + b)y_i + (ax_i + b)^2}{\sigma_i^2} \tag{0.2}$$

Separando os termos:

$$\chi^2(a,b) = \sum_{i=1}^{N} \left(\frac{y_i^2}{\sigma_i^2} - 2a\frac{x_i y_i}{\sigma_i^2} - 2b\frac{y_i}{\sigma_i^2} + a^2\frac{x_i^2}{\sigma_i^2} + 2ab\frac{x_i}{\sigma_i^2} + b^2\frac{1}{\sigma_i^2} \right) \tag{0.3}$$

Reescrevendo os termos em termos das médias ponderadas:

$$\chi^2(a,b) = \sum_{i=1}^{N} \frac{y_i^2}{\sigma_i^2} - 2a\sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2} - 2b\sum_{i=1}^{N} \frac{y_i}{\sigma_i^2} + a^2\sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} + 2ab\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} + b^2\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \tag{0.4}$$

Para encontrar os valores ótimos de a e b, calculamos as derivadas parciais de χ^2 em relação a a e b, e igualamos a zero:

$$\frac{\partial \chi^2}{\partial a} = -2\sum_{i=1}^{N} \frac{x_i(y_i - ax_i - b)}{\sigma_i^2} = 0 \tag{0.5}$$

$$\frac{\partial \chi^2}{\partial b} = -2\sum_{i=1}^{N} \frac{y_i - ax_i - b}{\sigma_i^2} = 0 \tag{0.6}$$

Rearranjando as equações, obtemos:

$$\sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2} - a\sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} - b\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} = 0 \tag{0.7}$$

$$\sum_{i=1}^{N} \frac{y_i}{\sigma_i^2} - a\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} - b\sum_{i=1}^{N} \frac{1}{\sigma_i^2} = 0 \tag{0.8}$$

Com isso, podemos formular o seguinte sistema de equações:

$$\begin{cases} \bar{x^2} a + \bar{x} b = \bar{xy} \\ \bar{x} a + b = \bar{y} \end{cases} \tag{0.9}$$

Resolvendo esse sistema para a e b, obtemos:

$$\begin{cases} a = \frac{\bar{xy} - \bar{x}\bar{y}}{\bar{x^2} - \bar{x}^2} \\ b = \bar{y} - a\bar{x} \end{cases} \tag{0.10}$$

EXERCICIO 2

Para calcular a seção de choque σ, usamos a seguinte fórmula:

$$\sigma = \frac{N_{\text{Total}} - N_{\text{background}}}{L} \tag{0.11}$$

onde:

- $N_{\text{Total}} = 2567$ é o número total de eventos observados,
- $N_{\text{background}} = 1223.5$ é o número de eventos de fundo esperado,
- $L = 25\,\text{fb}^{-1}$ é a luminosidade integrada.

Substituindo os valores na fórmula (0.11), obtemos:

$$\sigma = \frac{2567 - 1223.5}{25} = \frac{1343.5}{25} = 53.74\,\text{pb} \tag{0.12}$$

Para calcular a incerteza estatística, usamos a distribuição de Poisson. A incerteza estatística é dada por:

$$\sigma_{\text{stat}}^2 = \left(\frac{\sqrt{N_{\text{Total}}}}{L}\right)^2 + \left(\frac{\sqrt{N_{\text{background}}}}{L}\right)^2 \tag{0.13}$$

Substituindo os valores na fórmula (0.13), obtemos:

$$\sigma_{\text{stat}}^2 = \left(\frac{\sqrt{2567}}{25}\right)^2 + \left(\frac{\sqrt{1223.5}}{25}\right)^2 \tag{0.14}$$

$$\sigma_{\text{stat}}^2 = \left(\frac{50.67}{25}\right)^2 + \left(\frac{34.99}{25}\right)^2 \tag{0.15}$$

$$\sigma_{\text{stat}}^2 = (2.0268)^2 + (1.3996)^2 \tag{0.16}$$

$$\sigma_{\text{stat}}^2 \approx 4.11 + 1.96 = 6.07 \tag{0.17}$$

$$\sigma_{\text{stat}} = \sqrt{6.07} \approx 2.46\,\text{pb} \tag{0.18}$$

A incerteza sistemática é dada por uma porcentagem da seção de choque e deve ser propagada diretamente. Para uma incerteza sistemática de 10%:

$$\sigma_{\text{sist}} = 0.10 \times \sigma = 0.10 \times 53.74 = 5.37\,\text{pb} \tag{0.19}$$

Finalmente, o valor da seção de choque com suas incertezas associadas é:

$$\sigma = 53.74 \pm 2.46_{\text{stat}} \pm 5.37_{\text{sist}}\,\text{pb} \tag{0.20}$$

EXERCICIO 3

Para determinar o número máximo de eventos esperados que podemos excluir com 95% de Confiança Limitada (C.L.), dado que observamos zero eventos e um fundo esperado de 0.07 eventos, usamos a distribuição de Poisson.

A função de massa de probabilidade (PMF) da distribuição de Poisson é dada por:

$$P(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{0.21}$$

onde k é o número de eventos observados (neste caso, $k = 0$) e λ é o número de eventos esperados.

Queremos encontrar o valor máximo de λ para o qual a probabilidade cumulativa de observar zero eventos é menor ou igual a 95% C.L. A probabilidade cumulativa de observar zero eventos é:

$$P(X \le 0) = P(X = 0) = e^{-\lambda} \tag{0.22}$$

Para 95% de C.L., queremos:

$$e^{-\lambda} \leq 0.05 \tag{0.23}$$

Resolvendo para λ:

$$-\lambda \leq \ln(0.05) \tag{0.24}$$

$$\lambda \geq -\ln(0.05) \tag{0.25}$$

$$\lambda \geq \ln\left(\frac{1}{0.05}\right) \tag{0.26}$$

$$\lambda \geq \ln(20) \approx 2.9957 \tag{0.27}$$

Portanto, com 95% de C.L., o número máximo de eventos esperados que podemos excluir é aproximadamente 2.996 eventos.

EXERCICIO 4

Para mostrar que a melhor função de ajuste é aquela em que $\chi^2/\mathrm{ndf} \to 1$, consideramos a distribuição χ^2 da estatística de ajuste.

Para um modelo que se ajusta bem aos dados, a estatística χ^2 deve seguir uma distribuição χ^2 com ndf graus de liberdade.

Propriedades da Distribuição χ^2

1. **Esperança e Variância da Distribuição χ^2:**

Para uma distribuição χ^2 com k graus de liberdade:

$$\mathrm{E}[\chi^2] = k \tag{0.28}$$

$$\mathrm{Var}[\chi^2] = 2k \tag{0.29}$$

2. **Valor Esperado de χ^2/ndf:**

Se o modelo está ajustado adequadamente, a expectativa de χ^2 é igual ao número de graus de liberdade:

$$\mathrm{E}\left[\frac{\chi^2}{\mathrm{ndf}}\right] = \frac{\mathrm{E}[\chi^2]}{\mathrm{ndf}} = \frac{\mathrm{ndf}}{\mathrm{ndf}} = 1 \tag{0.30}$$

Portanto, o valor reduzido de χ^2 deve ser 1 para um ajuste adequado.

3. **Distribuição Assintótica e χ^2/ndf:**

À medida que o número de dados n aumenta e o número de parâmetros ajustados p é fixo, a distribuição de χ^2 se aproxima de uma distribuição normal com média ndf e variância 2ndf. Assim, a variável:

$$\frac{\chi^2 - \mathrm{ndf}}{\sqrt{2\mathrm{ndf}}} \tag{0.31}$$

segue uma distribuição normal padrão ($N(0,1)$).

Portanto, o valor reduzido de χ^2 deve estar próximo de 1 para um ajuste adequado.

Interpretação dos Valores

- $\frac{\chi^2}{\mathrm{ndf}} > 1$: Sugere que o modelo não está se ajustando bem aos dados (ou as incertezas são subestimadas).
- $\frac{\chi^2}{\mathrm{ndf}} < 1$: Pode indicar que o modelo está superajustando os dados (ou as incertezas foram superestimadas).