Introdução à Análise de dados em FAE **(05/09/2024)**

Exercicios de estatística para análise de dados em HEP

Professores: Eliza Melo, Dilson Damião e Mauricio Thiel *Name:* Jorge Júlio Barreiros Venuto de Siqueira

EXERCÍCIO 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:

$$\chi^2(a, b) = \sum_{i=1}^{N} \frac{[y_i - (ax_i + b)]^2}{\sigma_i^2} \tag{1.1}$$

Expandindo o quadrado, temos:

$$\chi^2(a, b) = \sum_{i=1}^{N} \left(\frac{y_i^2}{\sigma_i^2} - 2a\frac{x_i y_i}{\sigma_i^2} - 2b\frac{y_i}{\sigma_i^2} + a^2\frac{x_i^2}{\sigma_i^2} + 2ab\frac{x_i}{\sigma_i^2} + b^2\frac{1}{\sigma_i^2} \right) \tag{1.2}$$

Reescrevendo os termos em termos das médias ponderadas:

$$\chi^2(a, b) = \sum_{i=1}^{N} \frac{y_i^2}{\sigma_i^2} - 2a\sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2} - 2b\sum_{i=1}^{N} \frac{y_i}{\sigma_i^2} + a^2\sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} + 2ab\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} + b^2\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \tag{1.3}$$

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

$$\frac{\partial \chi^2}{\partial a} = -2\sum_{i=1}^{N} \frac{x_i(y_i - ax_i - b)}{\sigma_i^2} = 0 \tag{1.4}$$

$$\frac{\partial \chi^2}{\partial b} = -2\sum_{i=1}^{N} \frac{(y_i - ax_i - b)}{\sigma_i^2} = 0 \tag{1.5}$$

Rearranjando as equações:

$$\sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2} - a\sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} - b\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} = 0 \tag{1.6}$$

$$\sum_{i=1}^{N} \frac{y_i}{\sigma_i^2} - a\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} - b\sum_{i=1}^{N} \frac{1}{\sigma_i^2} = 0 \tag{1.7}$$

Obtendo assim, um sistema de equações:

$$\begin{cases} \bar{x^2}a + \bar{x}b = \bar{xy} \\ \bar{x}a + b = \bar{y} \end{cases} \tag{1.8}$$

Resolvendo esse sistema para a e b, obtemos:

$$\begin{cases} a = \frac{\bar{xy} - \bar{x}\bar{y}}{\bar{x^2} - \bar{x}^2} \\ b = \bar{y} - a\bar{x} \end{cases} \tag{1.9}$$

EXERCÍCIO 2

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

$$\sigma = \frac{N_{\text{Total}} - N_{\text{background}}}{L} \tag{2.1}$$

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 (2.1), obtemos:

$$\sigma = \frac{2567 - 1223.5}{25} = \frac{1343.5}{25} = 53.74\,\text{fb} \tag{2.2}$$

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

$$\sigma_{\text{stat}}^2 = \left(\frac{\sqrt{N_{\text{Total}}}}{L}\right)^2 + \left(\frac{\sqrt{N_{\text{background}}}}{L}\right)^2 \tag{2.3}$$

Utilizando nossos valores na equação (2.3), obtemos:

$$\sigma_{\text{stat}}^2 = \left(\frac{\sqrt{2567}}{25}\right)^2 + \left(\frac{\sqrt{1223.5}}{25}\right)^2 \tag{2.4}$$

$$\sigma_{\text{stat}} = \sqrt{6.07} \approx 2.46\,\text{fb} \tag{2.5}$$

Para o calculo da incerteza sistemática, ela é dada por uma porcentagem da seção de choque e deve ser propagada diretamente. Para uma incerteza sistemática de 10%, temos:

$$\sigma_{\text{sist}} = 0.10 \times \sigma = 0.10 \times 53.74 = 5.37\,\text{fb} \tag{2.6}$$

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

$$\sigma = 53.74 \pm 2.46_{\text{stat}} \pm 5.37_{\text{sist}}\,\text{fb} \tag{2.7}$$

EXERCICIO 3

Para calcular quantos eventos esperados podem ser excluídos com 95% de confiança (C.L.) em uma análise onde a contagem de eventos segue uma distribuição de Poisson, consideramos que a probabilidade de observar k eventos quando o número esperado de eventos é λ é dada por:

$$P(k;\lambda) = \frac{e^{-\lambda}\lambda^k}{k!} \tag{3.1}$$

Neste caso, temos $k = 0$, então:

$$P(0;\lambda) = e^{-\lambda} \tag{3.2}$$

Para um nível de confiança de 95%, queremos encontrar o valor de λ tal que:

$$P(0;\lambda) \geq 0.05 \tag{3.3}$$

Isto implica que:

$$e^{-\lambda} \geq 0.05 \tag{3.4}$$

Tomando o logaritmo natural de ambos os lados, obtemos:

$$-\lambda \geq \ln(0.05) \tag{3.5}$$

Portanto:

$$\lambda \leq -\ln(0.05) \tag{3.6}$$

Calculando $-\ln(0.05)$:

$$-\ln(0.05) \approx 2.9957 \tag{3.7}$$

Assim, ao nível de confiança de 95%, podemos excluir um número esperado de eventos λ maior que aproximadamente 2.9957 eventos.

Como o número esperado de eventos deve ser um número inteiro, podemos excluir até 3 eventos esperados com 95% de confiança.

Portanto, a conclusão é que até 3 eventos esperados podem ser excluídos com esta análise com 95% de confiança.

EXERCICIO 4

A função χ^2 é definida como:

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

onde:

- y_i são os dados observados,
- $f(x_i)$ é o valor predito pelo modelo para x_i,
- σ_i é a incerteza associada aos dados observados.

ndf (graus de liberdade)

O número de graus de liberdade é dado por:

$$\text{ndf} = N - p$$

onde N é o número total de dados e p é o número de parâmetros ajustados no modelo.

Propriedades da Distribuição χ^2

A distribuição χ^2 com ndf graus de liberdade tem algumas propriedades importantes:

- **Média:** A média da distribuição χ^2 é igual a ndf:

$$E[\chi^2] = \text{ndf}$$

- **Variância:** A variância da distribuição χ^2 é $2 \cdot \text{ndf}$:

$$\text{Var}(\chi^2) = 2 \cdot \text{ndf}$$

Ajuste do Modelo e χ^2/ndf

- **Expectativa de χ^2:** Quando o modelo se ajusta adequadamente aos dados, podemos esperar que:

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

Isso significa que, em um ajuste ideal, χ^2/ndf deve ser próximo de 1.

- **Desvio da Média:** Se χ^2 estiver significativamente maior que ndf, isso indicará que a soma dos quadrados dos resíduos é maior do que o esperado, sugerindo que o modelo não está capturando adequadamente a variabilidade dos dados. Neste caso, $\chi^2/\text{ndf} > 1$.
- **Desvio da Média Inferior:** Se χ^2 for significativamente menor que ndf, isso indicará que as incertezas foram superestimadas ou que o modelo é muito simples (subajuste). Neste caso, $\chi^2/\text{ndf} < 1$.