Universidade do Estado do Rio de Janeiro

Instituto de Física

Departamento de Física Nuclear e Altas Energias

Atividade 1

Aluno: Pedro Henrique e Bruno Kron

Professor(a): Mauricio Thiel, Eliza Melo e Dilson Damião

Outubro

2024

Exercício 1

De acordo com o livro *Estimativas e Erros em Experimentos de Física*, para ajustar a reta $y(x) = ax + b$ a um conjunto de N pares de dados (x_i, y_i), onde as incertezas entre os pares são distintas, os resíduos são ponderados pela seguinte expressão:

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

Ao expandir a expressão em quadrados, temos:

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

Depois, ao reescrever em termos das médias ponderadas, temos:

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

Para encontrar os valores de a e b que minimizam χ^2, derivamos parcialmente em relação a a e b e igualamos a zero:

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

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

Isso nos leva ao sistema de equações:

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

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

Resolvendo o sistema, encontramos:

$$a = \frac{\sum_{i=1}^{N} \frac{x_iy_i}{\sigma_i^2} - \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} \right)^2} \tag{8}$$

$$b = \frac{\sum_{i=1}^{N} \frac{y_i}{\sigma_i^2} - a \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \tag{9}$$

Exercício 2

O exercício informa que a seção de choque é calculado pela equação:

$$\sigma = \frac{N_{total} - N_{background}}{\mathcal{L}} \tag{10}$$

Onde $N_{total} = 2567$,$N_{background} = 1223,5$ e $\mathcal{L} = 25fb^{-1}$. Calculando a seção de choque temos:

$$\sigma = \frac{2567 - 1223,5}{25} = 53,74fb$$

Agora podemos calcular o erro estatístico do sistema; Por se tratar de uma distribuição de Poisson temos que o erro é dado por:

$$\sigma^2 = \left(\frac{\sqrt{N_{total}}}{\mathcal{L}}\right)^2 + \left(\frac{\sqrt{N_{background}}}{\mathcal{L}}\right)^2 \tag{11}$$

$$\sigma_{estatistico} = \sqrt{\left(\frac{\sqrt{2567}}{25}\right)^2 + \left(\frac{\sqrt{1223,5}}{25}\right)^2} = 2,46fb$$

O exercício informa que o erro sistemático é de 10%, então podemos demostra-lo como:

$$\sigma_{sistematico} = \sigma \times 10\% = 5,37fb$$

Concluímos que a seção de choque do sinal com sua incertezas é:

$$\sigma = (53,74 \pm 2,46 \pm 5,37)fb \tag{12}$$

Exercício 3

No exercício, a seleção de eventos para a busca de acoplamentos anômalos previu um número de eventos de fundo de $b = 0.07$ após todos os cortes. Diferentes modelos de acoplamentos anômalos previram uma variedade de números de eventos a serem observados, variando entre 0.09 e 35 eventos. Considerando que a contagem de eventos se dá através de uma pdf de Poisson, a probabilidade de observar n eventos é dada pela expressão:

$$P(n; s, b) = \frac{(s+b)^n e^{-(s+b)}}{n!} \tag{13}$$

No caso específico em que $n = 0$, a probabilidade torna-se:

$$P(0; s, 0.07) = e^{-(s+0.07)} \tag{14}$$

Para excluir um modelo com um nível de confiança de 95%, é necessário que a probabilidade de observar zero eventos seja inferior a 5%, logo:

$$P(0; s, 0.07) < 0.05 \tag{15}$$

Substituindo na expressão (14), temos:

$$e^{-(s+0.07)} < 0.05 \tag{16}$$

Aplicando o logaritmo natural em ambos os lados, obtemos:

$$-(s + 0.07) < \ln(0.05) \tag{17}$$

Ajustando a expressão, teremos:

$$s + 0.07 > -\ln(0.05) \tag{18}$$

Calculando $-\ln(0.05)$:

$$-\ln(0.05) \approx 2.9957 \tag{19}$$

Sendo assim, a desigualdade será reescrita como:

$$s + 0.07 > 2.9957 \tag{20}$$

Isolando s, obtemos:

$$s > 2.9957 - 0.07 \Rightarrow s > 2.9257 \tag{21}$$

Concluímos que os parâmetros de acoplamentos anômalos que preveem um número esperado de eventos superior a $s > 2.93$ podem ser excluídos com um nível de confiança de 95%. Levando em consideração o intervalo de eventos previstos pelos modelos, que varia de 0.09 a 35 eventos, todos os modelos que preveem até 2.93 eventos permanecem como potenciais candidatos, enquanto aqueles que preveem um número de eventos maior que 2.93 são inviáveis de acordo com a análise realizada.

Referência: Tese "Study of WW Central Exclusive Production in the semileptonic channel with tagged protons at CMS detector" cap. 4.4.

Exercício 4

A estatística χ^2 é uma medida utilizada para avaliar a qualidade de um ajuste, definida por:

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

onde:

- y_i é o valor observado no experimento.
- $f(x_i)$ é o valor previsto pela função de ajuste no ponto x_i.
- σ_i representa a incerteza associada ao valor y_i.
- n é o número total de dados.

Essa métrica χ^2 fornece a soma das discrepâncias quadradas normalizadas pelas incertezas. Quanto melhor o ajuste, menor será o valor de χ^2, uma vez que as discrepâncias serão menores.

O número de graus de liberdade ndf é calculado como:

$$\text{ndf} = n - p,$$

onde:

- n é o número total de pontos de dados.
- p é o número de parâmetros ajustados no modelo.

O número de graus de liberdade representa o número de observações que restam para medir a qualidade do ajuste após considerar os parâmetros ajustados.

A quantidade χ^2/ndf é usada para avaliar a qualidade do ajuste de maneira mais robusta. Se as incertezas σ_i forem estimadas corretamente e o modelo ajustado for adequado, espera-se que $\chi^2/\text{ndf} \approx 1$.

- **Se χ^2/ndf > 1:** Um valor de χ^2/ndf significativamente maior que 1 sugere que o modelo não descreve bem os dados. As discrepâncias são maiores que o esperado, o que pode ocorrer devido a:
  - Um modelo inadequado para os dados.
  - Subestimação das incertezas experimentais.
- **Se χ^2/ndf < 1:** Um valor de χ^2/ndf menor que 1 pode indicar que:
  - As incertezas foram superestimadas.
  - O modelo pode estar sobreajustando os dados, capturando até mesmo flutuações aleatórias (overfitting).

A condição $\chi^2/\text{ndf} \approx 1$ indica que a função ajustada é apropriada, ou seja, descreve os dados dentro das incertezas estatísticas esperadas. Isso significa que as discrepâncias observadas são consistentes com as flutuações devido às incertezas dos dados. Portanto, a melhor função para um ajuste ocorre quando $\chi^2/\text{ndf} \to 1$, pois isso garante que o modelo é compatível com os dados e que as incertezas foram corretamente estimadas.