Introdução à Análise de dados em FAE e Tecnologias Associadas **(DATA)**

TRABALHO DOS SLIDES:AULA ESTATISTICA

Professores: Dilson Damião, Mauricio Thiel e Eliza Costa *Name:* João Marcos Modesto Ribeiro

EXERCICIO 2 DO SLIDE

Os dados fornecidos para o problema são:

- $N_{\text{Total}} = 2567$;
- $N_{\text{background}} = 1223.5$;
- $L = 25\,\text{fb}^{-1}$.

1 Passo 1: Cálculo da Seção de Choque

Utilizando a fórmula fornecida, temos:

$$\sigma = \frac{2567 - 1223.5}{25} = \frac{1343.5}{25} = 53.74\,\text{fb} \tag{1.1}$$

Assim, a seção de choque calculada é:

$$\sigma = 53.74\,\text{fb}$$

2 Passo 2: Cálculo da Incerteza Estatística

A incerteza estatística pode ser calculada usando a raiz quadrada da soma do número de eventos observados e do número de eventos de fundo:

$$\Delta N = \sqrt{N_{\text{Total}} + N_{\text{background}}} = \sqrt{2567 + 1223.5} = \sqrt{3790.5} \approx 61.58 \tag{2.1}$$

Logo, a incerteza estatística na seção de choque é dada por:

$$\Delta\sigma_{\text{stat}} = \frac{\Delta N}{L} = \frac{61.58}{25} \approx 2.46\,\text{fb} \tag{2.2}$$

3 Passo 3: Cálculo da Incerteza Sistemática

Considerando que a incerteza sistemática na luminosidade integrada é de 10%, temos:

$$\Delta L_{\text{sist}} = 0.1 \times L = 0.1 \times 25 = 2.5\,\text{fb}^{-1} \tag{3.1}$$

Portanto, a incerteza sistemática na seção de choque é:

$$\Delta\sigma_{\text{sist}} = \sigma \times \frac{\Delta L_{\text{sist}}}{L} = 53.74 \times \frac{2.5}{25} = 5.37\,\text{fb} \tag{3.2}$$

4 Resultado Final

O valor final da seção de choque, considerando as incertezas estatísticas e sistemáticas, é:

$$\sigma = 53.74 \pm 2.46\,(\text{estatística}) \pm 5.37\,(\text{sistemática})\,\text{fb} \tag{4.1}$$

EXERCICIO 3 DO SLIDE

O número de eventos observados, a probabilidade de observar n eventos com $s+b$ eventos esperados é:

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

onde n é o número de eventos observados no experimento, s é o número esperado de eventos do sinal, e b é o número esperado de eventos de fundo.

Para o caso da análise atual, a probabilidade de observar zero eventos, dado um modelo que prediz 0.07 eventos de fundo e s eventos de sinal, pode ser escrita como:

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

Esse modelo seria excluído ao nível de confiança de 95

$$P(0; s, 0.07) < 1 - 0.95,$$

ou seja:

$$e^{-(s+0.07)} < 0.05.$$

Aplicando o logaritmo natural para resolver a desigualdade:

$$-(s + 0.07) < \ln(0.05).$$

Calculando o valor de $\ln(0.05)$:

$$\ln(0.05) \approx -2.9957.$$

Portanto, substituindo na desigualdade:

$$-(s + 0.07) < -2.9957,$$

ou seja:

$$s + 0.07 > 2.9957.$$

Finalmente, isolando s:

$$s > 2.9957 - 0.07,$$

$$s > 2.9257.$$

Arredondando para duas casas decimais, temos:

$$s > 2.93.$$

EXERCICIO 4 DO SLIDE

Definições

Para começar, vamos definir a expressão para χ^2. O valor de χ^2 é uma medida de discrepância entre os valores observados y_i e os valores preditos $f(x_i)$ pelo modelo ajustado, ponderados pelas incertezas σ_i:

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

onde:

- y_i são os valores observados;
- $f(x_i)$ é a função ajustada (modelo teórico que desejamos ajustar aos dados);
- σ_i é a incerteza associada a cada valor y_i;
- N é o número total de pontos de dados.

Interpretação da Equação de χ^2

A expressão para χ^2 quantifica a soma dos desvios quadráticos entre os valores observados e os valores preditos, normalizados pelas incertezas σ_i. Isso significa que χ^2 considera não apenas as diferenças absolutas, mas também o quanto essas diferenças se relacionam com as incertezas em cada medida.

Definição de χ^2 Reduzido

O valor de χ^2 sozinho não nos dá uma medida da qualidade do ajuste porque depende diretamente do número de pontos de dados N. Para isso, usamos o χ^2 reduzido, que é definido como:

$$\chi^2_{\text{red}} = \frac{\chi^2}{\text{ndf}}$$

onde $\text{ndf} = N - p$ é o número de graus de liberdade (*number of degrees of freedom*), e p é o número de parâmetros ajustados no modelo.

O Número de Graus de Liberdade (ndf)

O número de graus de liberdade é dado por:

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

onde:

- N é o número total de pontos de dados observados;
- p é o número de parâmetros ajustados pelo modelo teórico.

Os graus de liberdade representam a quantidade de informações independentes nos dados após a realização do ajuste. Cada parâmetro ajustado ”remove”um grau de liberdade dos dados originais.

Demonstração

Para um ajuste adequado, esperamos que o valor de χ^2 seja aproximadamente igual ao número de graus de liberdade:

$$\chi^2 \approx \text{ndf} = N - p$$

ou seja:

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

Isso ocorre porque, quando o modelo se ajusta bem aos dados, a discrepância entre os valores observados e os valores preditos é consistente com as incertezas σ_i. Isso implica que as diferenças $y_i - f(x_i)$ são, em média, da ordem de magnitude das incertezas σ_i.

Justificação em Detalhes

Para entender melhor, podemos reescrever χ^2 como:

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

Essa forma mostra que χ^2 é a soma dos quadrados das diferenças normalizadas pelos desvios padrão σ_i. Se $y_i - f(x_i) \approx \sigma_i$ para todos os pontos i, então cada termo da soma contribuirá aproximadamente com 1 para χ^2. Isso leva a:

$$\chi^2 \approx N.$$

No entanto, como estamos ajustando p parâmetros, cada parâmetro "consome" um grau de liberdade. Assim, o valor esperado de χ^2 é dado por $N - p$.

Valor Reduzido de χ^2

O valor reduzido de χ^2 é então:

$$\chi^2_{\text{red}} = \frac{\chi^2}{\text{ndf}} = \frac{\sum_{i=1}^{N} \frac{(y_i - f(x_i))^2}{\sigma_i^2}}{N - p}.$$

Se o ajuste for bom, então:

$$\chi^2_{\text{red}} \approx \frac{\text{ndf}}{\text{ndf}} = 1.$$

Isso implica que, para a melhor função ajustada, esperamos que o valor reduzido de χ^2 tenda a 1, ou seja, $\chi^2/\text{ndf} \to 1$.