

Métodos Numéricos em Física Médica

3^a aula

ESTATÍSTICA II

III. Valores sumário em estatística descritiva.

- Na aula anterior, vimos como distinguir diferentes tipos de variáveis estatísticas, e também como as analisar do ponto de vista da sua frequência, tanto em tabelas como em gráficos.
- Os dados estatísticos de uma dada amostra são muitas vezes numerosos, e variados; e difíceis de interpretar olhando apenas para tabelas e gráficos de frequências.
- Em estatística descritiva, podemos definir uma série de valores que procuram fornecer de forma rápida, o máximo de informação sobre dados de uma determinada amostra.
- Estes valores são chamados de **valores-sumário**, ou **valores de sumário**.



Exemplos:

– média, moda, mediana, desvio padrão, percentil 75, percentil 95, 1º quartil, etc.

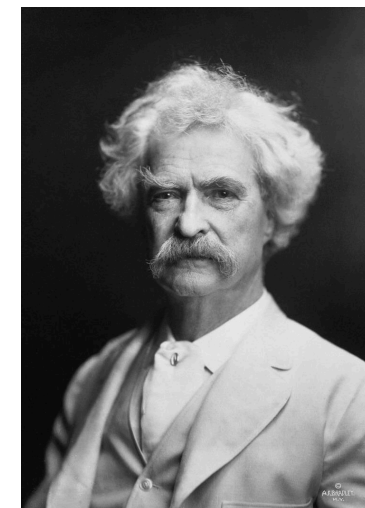
III. Valores sumário em estatística descritiva II



- Os valores sumário devem ser utilizados com cuidado.
- Dependendo do tipo de dados, os valores sumário podem não permitir qualquer tipo de conclusão, ou mesmo levar a interpretações erradas. Se os números não mentem, é seguramente possível mentir com números.

Alguns tipos de erros comuns:


- Viés amostral, ou erro amostral:
 - O centro de estatísticas X quis fazer um estudo sobre os hábitos da utilização da internet de uma população. Fizeram um inquérito *online*.
(solução: amostragens aleatórias)
- Viés de confirmação:
 - Ao analisarem os hábitos de doentes de diabetes, perceberam que estes comiam mais doces que a população geral.
(solução: grupo de controlo)



"Lies, damn lies, and statistics".

III. Valores–sumário em estatística descritiva II

Alguns exemplos da utilização errada de valores sumário:

- *Conjuntos não-lineares:* Considere-se a seguinte amostra de dados: “0,2,9,1000,40000”. A média aritmética desta amostra é: 8202.  *média geométrica*
- *Comparações de percentagens:* No país X a vacinação de covid-19 aumentou de 1 para 2 pessoas. O aumento foi de 100%. No país Y a vacinação aumentou de 100 mil para 110 mil. O aumento foi de 10%.
- *Assumir um ponto como significativo de uma amostra:* A média de idades dos fumadores em Portugal é de 35 anos. O Manuel tem 80 anos. É fumador ou não?

 Resumindo:

A estatística descritiva, em específico os valores–sumário, podem significar muita coisa ou coisa nenhuma. A utilização científica deve ser feita com cuidado, e no sentido para o qual estes valores servem.

III. Medidas de tendência central

Muitos conjuntos de dados estatísticos tendem a aglomerar-se em torno de um valor, muitas vezes um valor central. Os três principais indicadores de tendência central em estatística descritiva são:

- *A moda:* A moda corresponde ao valor com frequência máxima numa amostra. No exemplo dos partos, a moda de partos é de 2 partos (com 14 de frequência).
- *A mediana:* A mediana corresponde ao valor que separa a metade inferior da metade superior de uma amostra. No exemplo da aula anterior, a mediana de partos é de 2. *(no caso de amostras com um número par de dados, a mediana é calculada como a média dos dois valores centrais)*
- *A média aritmética:* A média corresponde ao valor da soma de todos os dados de uma amostra, dividida pelo número de dados.

$$\bar{x} = \langle x \rangle = \frac{\sum_{i=1}^N x_i}{N} = \frac{\sum_{i=1}^k f_i \cdot x_i}{N}$$

(em que N é o tamanho da amostra, e k o número de ocorrências diferentes)
no exemplo da aula anterior, a média é de 2 partos.

| partos | contagem |
|--------|----------|
| 0 | 5 |
| 1 | 6 |
| 2 | 14 |
| 3 | 10 |
| 4 | 3 |
| 5 | 1 |
| 8 | 1 |

Diagram illustrating the distribution of births (partos) and their frequency (contagem). The table shows the following data:

- 0 partos: 5 contagem
- 1 parto: 6 contagem
- 2 partos: 14 contagem (highlighted as the mode and median)
- 3 partos: 10 contagem
- 4 partos: 3 contagem
- 5 partos: 1 contagem
- 8 partos: 1 contagem

Annotations:

- moda** (orange arrow) points to the row with 2 partos and 14 contagem.
- mediana** (green arrow) points to the row with 2 partos and 14 contagem.
- média** (blue arrow) points to the row with 2 partos and 14 contagem.

Exercício

- *Calcular a moda, média e mediana do exercício anterior em python.*



Casa, determinar média e mediana do peso e moda dos pesos da tabela 2.5 em python

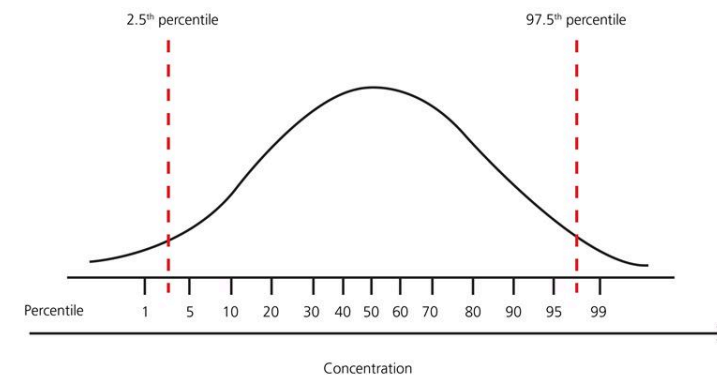
III. Outras medidas de localização.

Para além das medidas de tendência central existem ainda indicadores de localização que não são centrais. Alguns exemplos em baixo:

- *1º, 2º, e 3º Quartis:* valores que separam 25, 50, ou 75% dos valores inferiores, dos 75, 50, ou 25% dos valores superiores. O 2º quartil corresponde à mediana.
- *Decis:* 9 valores que separam 10, 20, 30, 40, 50, 60, 70, 80, 90% dos valores inferiores, dos 90, 80, 70, 60, 50, 40, 30, 20, 10, dos valores superiores. O 5º decil corresponde à mediana.
- *Percentis:* o mesmo que acima mas em percentagens.



Estes valores determinam-se exactamente como a mediana, isto é ordenam-se os valores, e em seguida procuram-se os valores nas casas $n/4$ para o 1º quartil, $3n/4$ para o 3º, $78n/100$ para o 78º percentil, etc. Quando o tamanho da amostra é par, faz-se a média dos dois dados adjacentes.



Exercício

- *Calcular os quartis, decis e percentis do exercício anterior em python.*



Casa, determinar quartis, decis e percentis dos pesos da tabela 2.5 (livro) em python

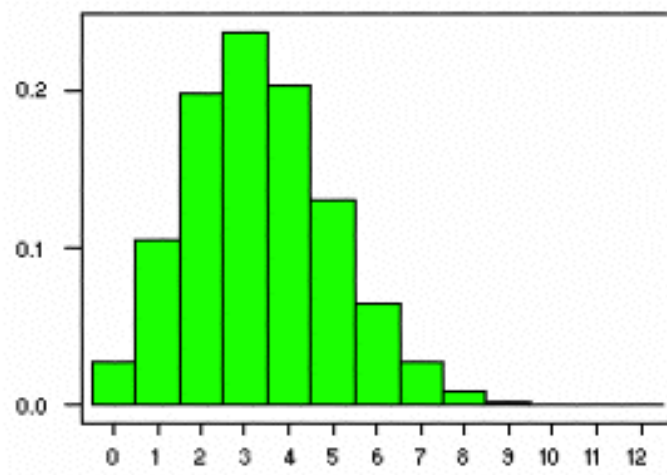
Resumo

- A estatística descritiva utiliza valores sumário que permitem de forma rápida transmitir o máximo de informação sobre um determinado conjunto estatístico.
- Estes valores podem ser valores de tendência central (assumindo que os valores tendem a agregar-se em torno de um ponto central) – a moda, a mediana, e a média.
- Podem também ser valores de localização, os quantis, que informam sobre uma determinada posição num conjunto estatístico – os quartis, os decis, e os percentis (por exemplo).

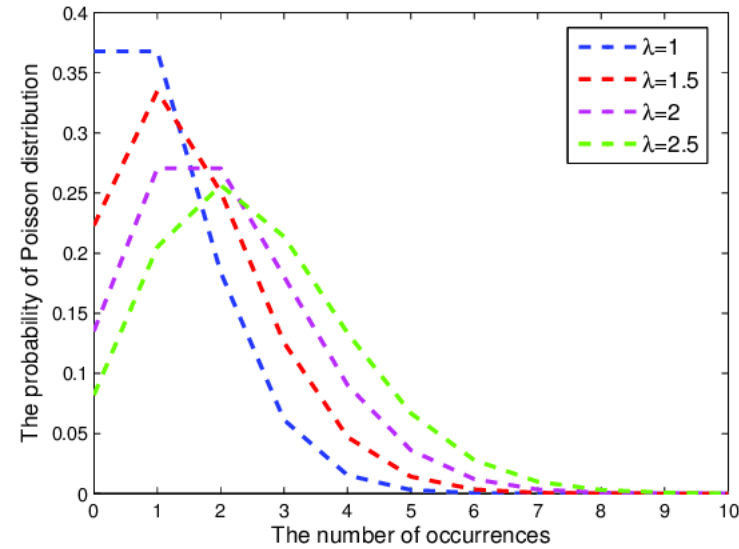


Estes valores devem ser utilizados de forma científica, de maneira a “cumprirem o seu dever”, i.e. o de fornecerem informação relevante sobre um conjunto estatístico. Devem ser evitados todos os tipos de viés e erros.

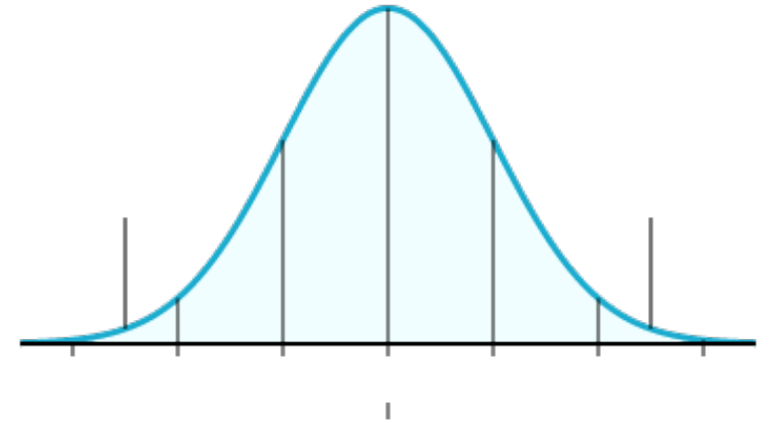
Próxima aula



Distribuição binomial



Distribuição de Poisson



Distribuição Gaussiana

MCNPX – continuação

- Conectar ao servidor lxlabs0 usando ssh tal como fizemos na aula passada.
- Criar pasta com o nome aula 3
- Transferir o ficheiro mcnpconfig
- Correr ./mcnpconfig

MCNPX – breve história

- A utilização de números aleatórios para fins de cálculos numéricos data pelo menos ao século XVIII.
- Os métodos de Monte-Carlo consistem na utilização de números aleatórios e algoritmos de rejeição para realizar cálculos.
- Iremos falar destes conceitos em mais detalhes em aulas posteriores



- MCNP - Monte Carlo N-Particle transport code foi desenvolvido em Los Alamos por Ulam, Von Neumann, Richtmeier e Metropolis para simular a difusão de neutrões em material fissionável nos anos 40. À época, o computador existente, ENIAC, poderia realizar o cálculo do transporte de 100 neutrões, cada um atravessando 100 colisões em cerca de 5 horas.
- Isto deu origem a vários códigos ao longo das décadas seguintes, em que se assistiu também ao aumento da capacidade computacional (lei de Moore), que foram recebendo diferentes nomes MCS, MCN, MCP, e MCG
- Atualmente, o MCNP/MCNPX utiliza bases de dados de secções eficazes de interação para neutrões, eletrões, fótons, prótons, etc. a diferentes energias para a realização do cálculo de transporte ao longo de geometrias virtuais definidas pelo utilizador.

<https://mcnp.lanl.gov/>

Exemplo

Titulo

c celulas

1 1 -1 -1 imp:p=1

2 0 1 imp:p=0

c superficies

1 s 0 0 0 5 \$ esfera raio 5 cm

c dados

mode p

m1 1001 2 8016 1

sdef par=2 pos=0 0 0 erg=1

nps 1000

mcnpx i=<input> ip → visualização gráfica

mcnpx i=<input> o=<output> → correr simulação

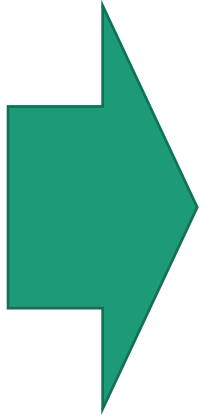
Notas: - 80 colunas por linha

- c significa comment

- \$ significa comentário
após linha

- Não utilizar TAB

Exercício 1



Copiar o código escrito na página anterior e correr o código em modo visualização e em modo de correr.

Ferramenta de visualização

- Na ferramenta de visualização pode utilizar vários comandos:
 - ✓ `px, py, pz` → visualizar nos planos definidos desta forma.
 - ✓ `cursor` → permite escolher uma parte do visualizador em detalhe a partir do cursor.
 - ✓ `factor` → permite fazer zoom in e out pelo factor referido.
 - ✓ `scales` → 0,1,2 permite desenhar escalas no desenho.
 - ✓ `label` → 0,1... permite colocar labels nas superfícies, células e nos materiais (separados por vírgulas)
 - ✓ `help,?,options` → lista dos comandos disponíveis.


Geometria – superfícies e células

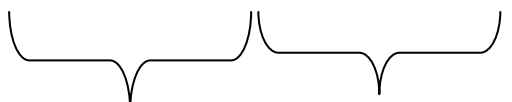
CELL E SURFACE CARDS

- Superfícies e células são numeradas de 1 a 9999
 - sólidos geométricos simples são definidos por comandos que definem superfícies
 - Planos (px, py, pz)
 - Esferas (s, s0)
 - Cilindros (cx, cy, cz)
 - Cones (kx, ky, kx)
 - etc.
 - Células são definidas através de operações booleanas sobre as superfícies definidas.
 - TODO o espaço tem de ficar definido

EXEMPLO geral de célula:

| | | | | |
|---|---|-----|----|-----------|
| 1 | 1 | -15 | | |
| 2 | 0 | | -1 | imp:p,e=1 |
| | | | 1 | imp:p,e=0 |

 material dens

 superfícies dados como a importancia

EXEMPLOS Superfícies

1 px 5 (plano perpendicular ao eixo dos x=5 cm)
2 so 10 (esfera centrada no centro (0,0,0) com r=10 cm)
3 cy 20 (cilindro ao longo do eixo dos y com 20 cm de raio)

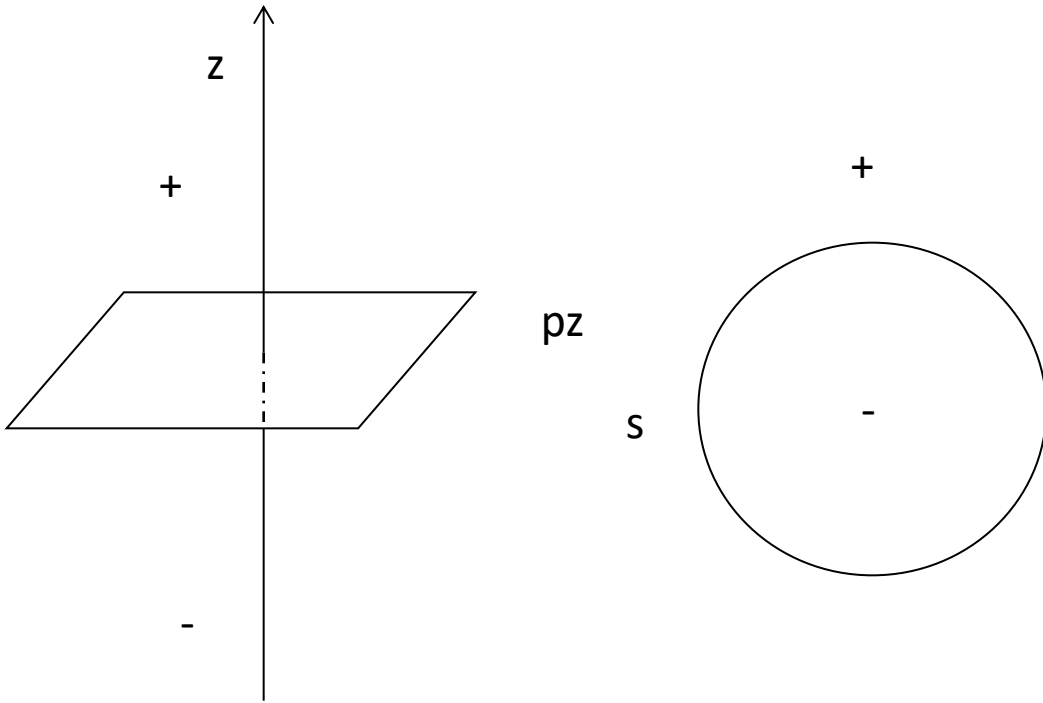
Geometria

| Mnemonic | Type | Description | Equation | Card Entries |
|----------|----------|--------------------|---|---|
| P | Plane | General | $Ax + By + Cz - D = 0$ | A B C D |
| PX | | Normal to x-axis | $x - D = 0$ | D |
| PY | | Normal to y-axis | $y - D = 0$ | D |
| PZ | | Normal to z-axis | $z - D = 0$ | D |
| SO | Sphere | Centered at Origin | $x^2 + y^2 + z^2 - R^2 = 0$ | R |
| S | | General | $(x - \bar{x}) + (y - \bar{y})^2 + (z - \bar{z})^2 - R^2 = 0$ | $\bar{x} \ \bar{y} \ \bar{z} \ R$ |
| SX | | Centered on x-axis | $(x - \bar{x})^2 + y^2 + z^2 - R^2 = 0$ | $\bar{x} \ R$ |
| SY | | Centered on y-axis | $x^2 + (y - \bar{y})^2 + z^2 - R^2 = 0$ | $\bar{y} \ R$ |
| SZ | | Centered on z-axis | $x^2 + y^2 + (z - \bar{z})^2 - R^2 = 0$ | $\bar{z} \ R$ |
| C/X | Cylinder | Parallel to x-axis | $(y - \bar{y})^2 + (z - \bar{z})^2 - R^2 = 0$ | $\bar{y} \ \bar{z} \ R$ |
| C/Y | | Parallel to y-axis | $(x - \bar{x})^2 + (z - \bar{z})^2 - R^2 = 0$ | $\bar{x} \ \bar{z} \ R$ |
| C/Z | | Parallel to z-axis | $(x - \bar{x})^2 + (y - \bar{y})^2 - R^2 = 0$ | $\bar{x} \ \bar{y} \ R$ |
| CX | | On x-axis | $y^2 + z^2 - R^2 = 0$ | R |
| CY | | On y-axis | $x^2 + z^2 - R^2 = 0$ | R |
| CZ | | On z-axis | $x^2 + y^2 - R^2 = 0$ | R |
| K/X | Cone | Parallel to x-axis | $\sqrt{(y - \bar{y})^2 + (z - \bar{z})^2} - t(x - \bar{x}) = 0$ | $\bar{x} \ \bar{y} \ \bar{z} \ t^2 \pm 1$ |
| K/Y | | Parallel to y-axis | $\sqrt{(x - \bar{x})^2 + (z - \bar{z})^2} - t(y - \bar{y}) = 0$ | $\bar{x} \ \bar{y} \ \bar{z} \ t^2 \pm 1$ |
| K/Z | | Parallel to z-axis | $\sqrt{(x - \bar{x})^2 + (y - \bar{y})^2} - t(z - \bar{z}) = 0$ | $\bar{x} \ \bar{y} \ \bar{z} \ t^2 \pm 1$ |
| KX | | On x-axis | $\sqrt{y^2 + z^2} - t(x - \bar{x}) = 0$ | $\bar{x} \ t^2 \pm 1$ |

| Mnemonic | Type | Description | |
|----------|--|--|--|
| SQ | Ellipsoid Hyperboloid Paraboloid | Axis not parallel to x-, y-, or z-axis | $A(x - \bar{x})^2 + 2D(x - \bar{x}) + G = 0$ |
| GQ | Cylinder Cone Ellipsoid Hyperboloid Paraboloid | Axes not parallel to x-, y-, or z-axis | $Ax^2 + By^2 + Gx + Hy + I = 0$ |
| TX | Elliptical or Circular Torus. | Axis is parallel to x-, y-, or z-axis | $(x - \bar{x})^2 / B^2 -$ |
| TY | | | $(y - \bar{y})^2 / B^2 -$ |
| TZ | | | $(z - \bar{z})^2 / B^2 +$ |

Geometria – células

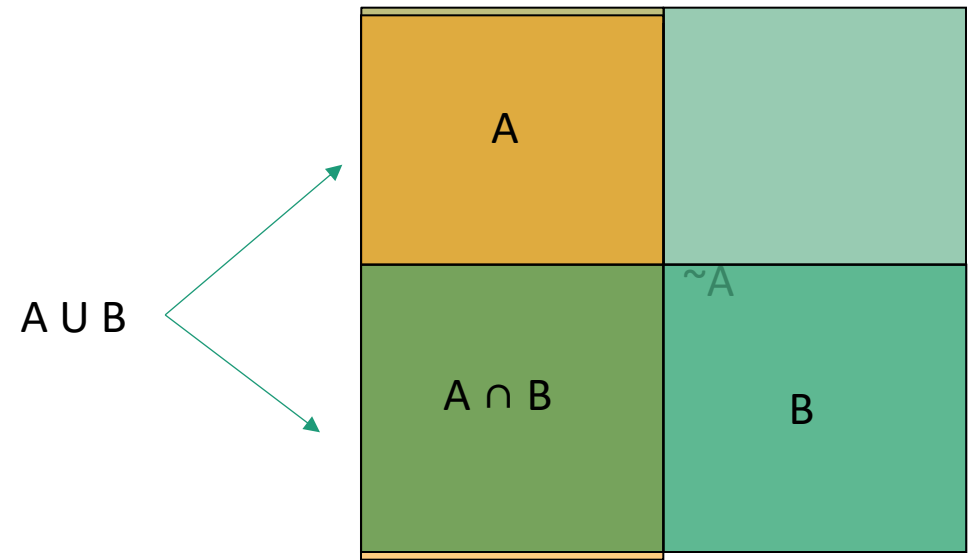
SINAIS



Explicação dos sinais – e + na definição das células

OPERAÇÕES BOOLEANAS

- o Intersecção (e) -> é um espaço “ “
- o Reunião (ou) -> é o sinal dois pontos “:”
- o Complementar (negação) -> é o sinal “#”



Exercício 2



Desenhar uma esfera no vácuo no ponto $(0,26,0)$ com $r=3$ cm e visualizar a estrutura.

titulo: ejercicio 2

C cell cards

1 0 -1 imp:p=1

2 0 1 imp:p=0

C surface cards

1 s 0 26 0 3

(1 sy 26 3)

C data cards

mode p