

Introdução à bioestatística utilizando Python

**Porque mais um curso de estatística???**

PhD Flavio Lichtenstein

Bioinformatics, Systems Biology, and Biostatistics

Instituto Butantan – CENTD - Bioinformática

Janeiro 2026



**Porque a maioria dos cursos de estatística  
dissociam Estatística e Ciência.**

**Aqui vamos fazer uma forte associação  
entre Estatística e Ciência Biológica**

**Ciência tem que ser quantitativa (Galileo),  
logo não existiria sem matemática,  
geometria e estatística.**

# Porque a Biologia tem que ser quantitativa apesar de sua complexidade + biologia descritiva





# Porque praticamente não existem aparelhos que medem efeitos físico-químicos sem Matemática, Geometria e Estatística



**Quantificar tudo!**

**E conhecer as incertezas  
mean(SSD)**

mean: media amostral

SSD: desvio padrão amostral

Todas as mensurações têm incertezas

O mundo exato não existe

Se você esqueceu de colocar o desvio  
padrão amostral: você errou!



O que aprendemos a quantificar?





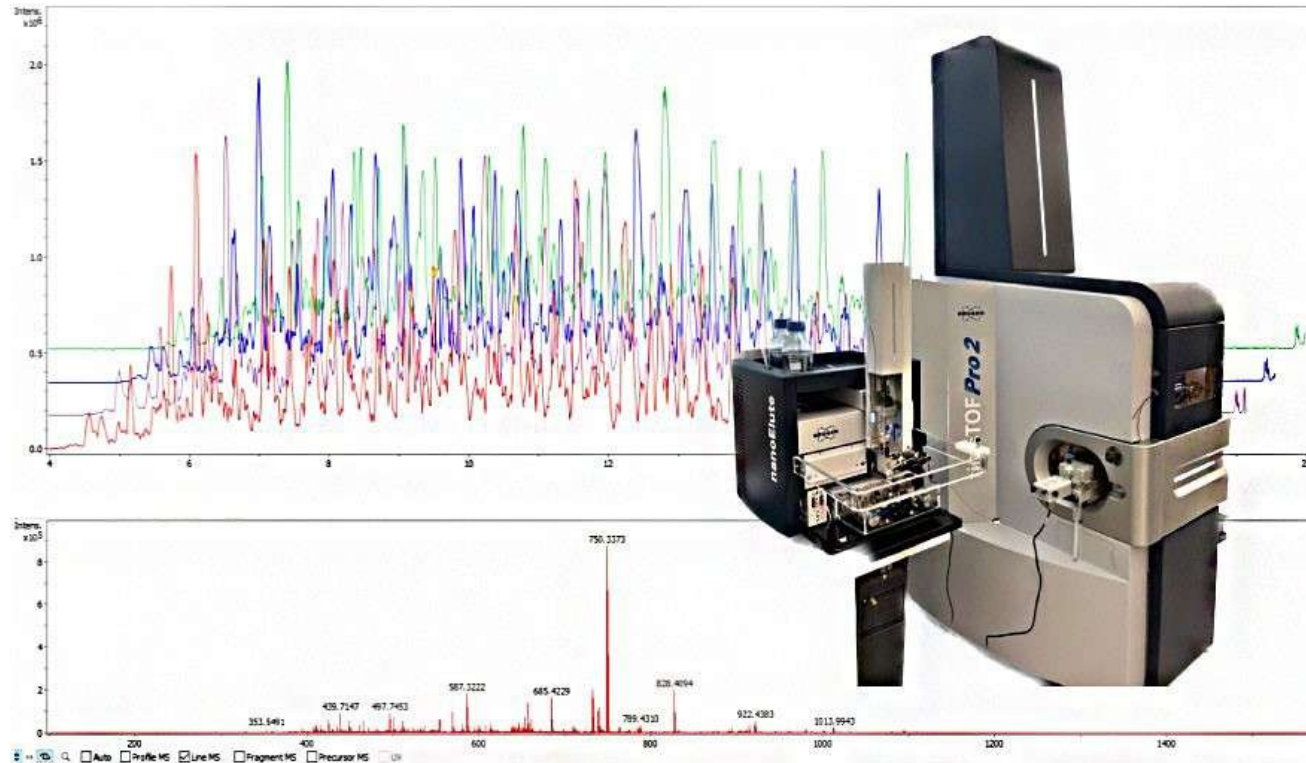
## Pressão Sanguínea

Stephen Hales medindo a pressão sanguínea em um cavalo (1705)



Batimentos cardíacos  
Eletrocardiograma (ECG)

# Espectômetro de Massas Proteômica



Fusion Orbitrap ETD mass spectrometer.

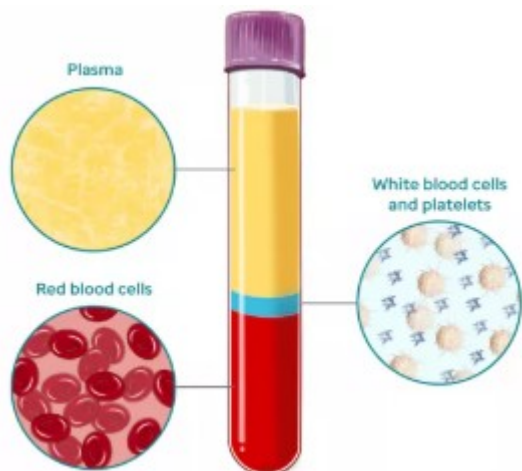




## Contagem de Hemácias

Proporções: RBC + WBC + Plasma

Contagem de todos os componentes da série vermelha (eritrograma) e branca (leucograma)

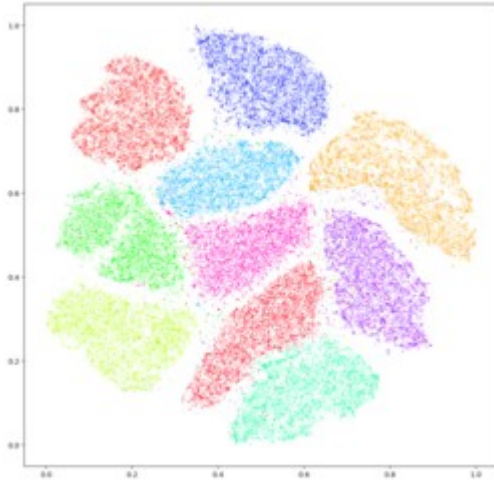


## Ressonância Nuclear Magnética (NMR)



Núcleos como os de H e C apresentam spin (propriedade quântica relativa a momento angular - prótons e nêutrons podem girar no seu próprio eixo), logo funcionam como magnetos. Uma vez alinhado a um campo magnético externo (B) eles se alinham e no relaxamento emitem radiação que pode ser interpretada de acordo com a matéria existente nos núcleos vizinhos.

<https://www.youtube.com/watch?v=cqOewFZPgOM>



Clusterização por t-SNE (ou UMAP)

Quantificação de tipos celulares e seus principais genes diferencialmente expressos (LFC)

Transcriptômica (transcrição de mRNA)

LFC:  $\log_2$  fold-change -  $\log_2$  da variação transcricional relativa a um controle.



# Como medir?

Como quantificar?

- Vírions, Bactérias e Virulência
- Eficácia de uma vacina
- miRNA (micro RNA)
- Biodiversidade na floresta Amazônica

Para cada uma destas perguntas os cientistas criaram métodos, os validaram e publicaram, gerando conhecimento, patentes, ciência (conhecimento oficial) e tecnologia reprodutíveis.

# Como funciona?

Como funciona?

- Multiplex (analisador de proteínas)
- Citômetro de Fluxo
- Microscópio confocal
- Microscópio eletrônico de varredura
- High-Content Screening
- Tomógrafo
- Ressonância Magnética

**A ciência tem que ser quantitativa (Galileo),  
logo não existiria sem matemática e  
estatística.**

**Palavras chaves**



**A ciência tem que ser quantitativa (Galileo),  
logo não existiria sem matemática e  
estatística.**

**Palavras chaves**

**Quantificar e Medir  
Reprodutibilidade**

**Significância e Tamanho do Efeito**

# Problemas e Soluções

Falta de reprodutibilidade experimental:

- Declaração explícita de todos materiais e recursos utilizados
- Declaração explícita de todos os métodos utilizados
- Materiais e Métodos validados

Falta de reprodutibilidade resultados:

- Todos os dados medidos e inferidos têm que ser apresentados (supplemental material)
- Caso os dados sejam de pacientes, estes têm que ser anonimizados
- Métodos estatísticos tem que ser declarados
- Resultados têm que ser apresentados com:
  - Media e desvio padrão: mean(SSD) - nunca use mean( $\pm$  SSD), NUNCA!
  - Número amostral
  - Teste de hipótese
    - p-valor (só p-valor NÃO PODE; é insuficiente)
    - intervalo de confiança
- Inferências só podem ser aceitas se seguirem os testes de hipóteses

# Problemas e Soluções

Falta de reprodutibilidade computacional:

- Códigos fontes têm que ser de livre acesso (depositados no GitHub, Kaggle ou Zenodo)
- Códigos fontes e respectivos dados têm que ser reprodutíveis
  - Python, R, C, etc
  - Docker
- Dados similares de outros laboratórios têm que ser reprodutíveis com os códigos fontes apresentados

Utilização de Inteligência Artificial (IA):

- IA não pode ser um referência científica
- Nada impede do Pesquisador usar IA + PubMed ou Web of Science:
  - Caso o pesquisador utilize IA, tem que mostrar que todas as anotações apresentadas pela IA são reais, válidas e constam em referências validadas por pares
- IA não pode fazer hipóteses, e sim o pesquisador
- IA não pode fazer inferência, e sim o pesquisador

## **Porque mais um curso de estatística???**

Em caso de dúvidas escreva no chat para o moderador

Algumas perguntas serão respondidas e depositadas no Q&A  
Possivelmente, devido ao grande número de alunos não conseguiremos  
responder a todas as perguntas, mas tentaremos.

PhD Flavio Lichtenstein

Bioinformatics, Systems Biology, and Biostatistics

Instituto Butantan – CENTD - Bioinformática

Janeiro 2026