



Introdução à bioestatística utilizando Python

Porque mais um curso de estatística???

PhD Flávio Lichtenstein

Bioinformatics, Systems Biology, and Biostatistics

Instituto Butantan – CENTD - Bioinformática

Janeiro 2026



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

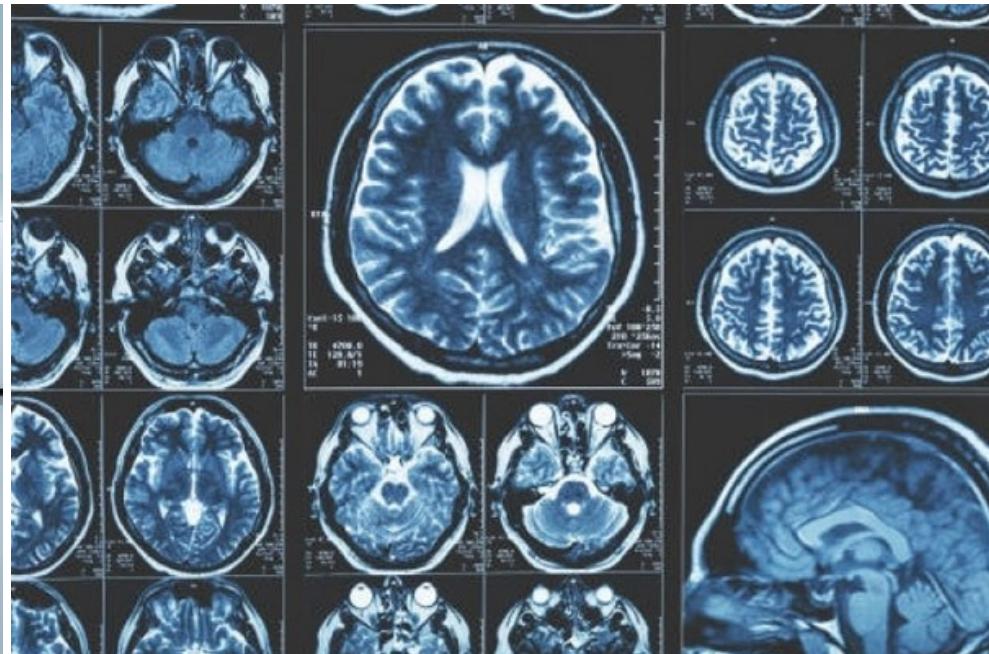
**Aqui vamos fazer uma forte associação
entre Estatística e Ciência Biológica**

**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?

Quantificação é o processo de transformar dados qualitativos em quantitativos, permitindo a análise estatística e a tomada de decisões baseadas em evidências.

Os dados qualitativos são caracterizados por serem descrevíveis por palavras ou frases, enquanto os dados quantitativos são expressos por números.

Alguns exemplos de quantificação incluem:

- Quantificação de opiniões:** Transformar respostas qualitativas em escala numérica, como "muito satisfeita", "satisfatória" e "insatisfatória".
- Quantificação de comportamentos:** Transformar descrições de comportamentos em códigos numéricos, como "não fumante", "fumante ocasional" e "fumante regular".
- Quantificação de sentimentos:** Transformar descrições de sentimentos em escala numérica, como "muito feliz", "feliz" e "descontente".
- Quantificação de preferências:** Transformar descrições de preferências em escala numérica, como "não prefiro", "prefiro" e "muito prefero".

É importante notar que a quantificação não significa que os dados qualitativos perdem sua essência ou significado. Ela simplesmente transforma a forma de representação dos dados para facilitar a análise estatística e a tomada de decisões.



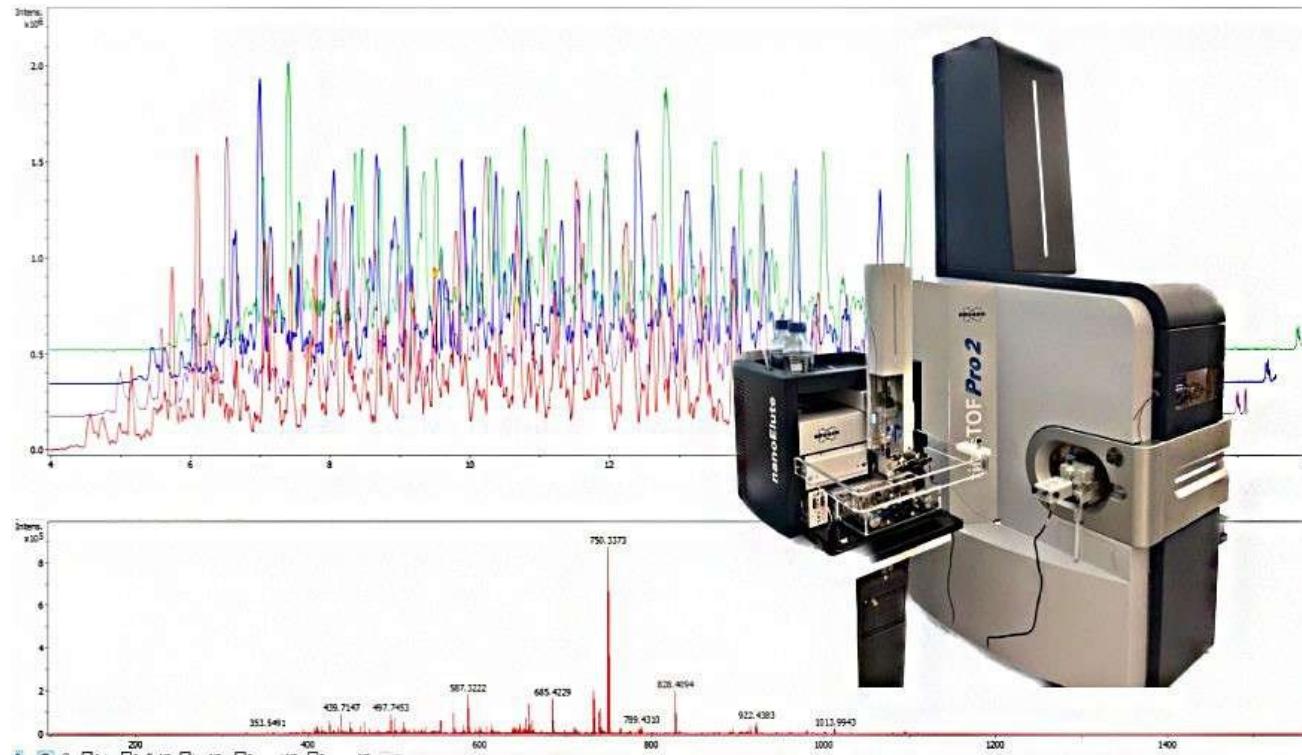
Pressão Sanguínea

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



Batimentos cardíacos
Eletrocardiograma (ECG)

Espectrômetro de Massas Proteômica

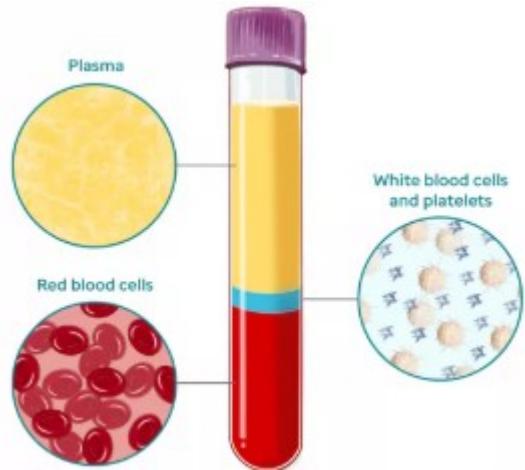


Fusion Orbitrap ETD mass spectrometer.



Contagem de Hemácias

Proporções: RBC + WBC + Plasma

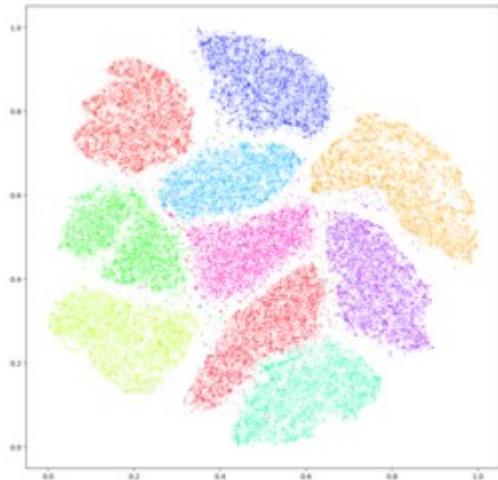


Contagem de todos os componentes da série vermelha (eritrograma) e branca (leucograma)



Ressonância Nuclear Magnética (NMR)

Mobilidade do dipolo da água em um organismo: densidade de matéria.



Clusterização por t-SNE (ou UMAP)

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

Tranciptômica (transcrição de mRNA)

LFC: \log_2 fold-change - \log_2 da variação transcrecional 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 reproduutíveis.

Como funciona?

Como funciona?

- Multiplex (analizador 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 reproduzibilidade experimental:

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

Falta de reproduzibilidade 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 reproduzibilidade computacional:

- Códigos fontes têm que ser de livre acesso (depositados no GitHub, Kaggle ou Zenodo)
- Códigos fontes e respectivos dados têm que ser reproduzíveis
 - Python, R, C, etc
 - Docker
- Dados similares de outros laboratórios têm que ser reproduzíveis 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 Flávio Lichtenstein

Bioinformatics, Systems Biology, and Biostatistics

Instituto Butantan – CENTD - Bioinformática

Janeiro 2026

