



História da Estatística
intimamente ligada à história da biologia
primeira metade do século XX

ESIB
CENTD – Instituto Butantan

PhD Flávio Lichtenstein

Bioinformatics, Systems Biology, and Biostatistics

Instituto Butantan – CENTD - Bioinformática

Janeiro 2026



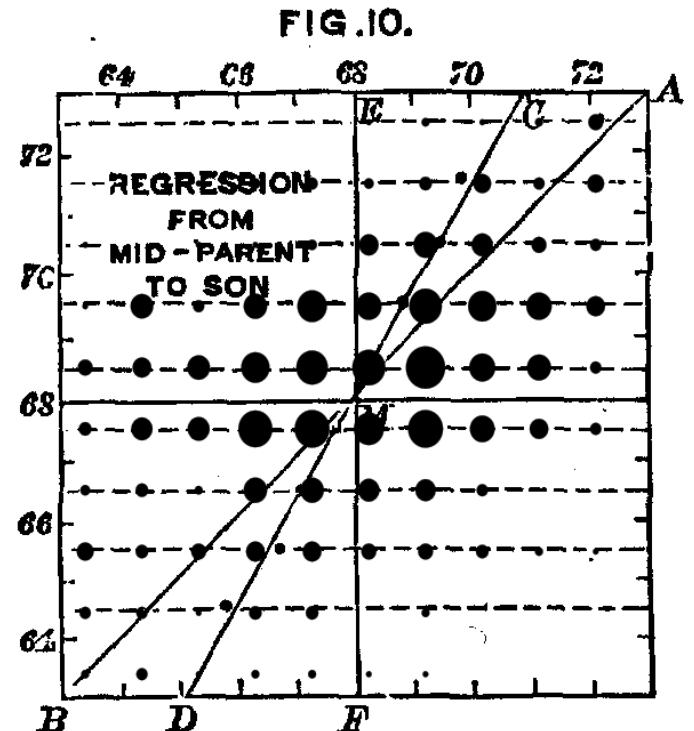
Francis Galton (1822-1911)

Galton e seus estudos

Sir Francis Galton (1822–1911) é um dos pais da estatística. Ele fundou muitos conceitos, como correlação, quartil, percentil e regressão (junto com seu aluno Pearson), que são básicos em estatística

Galton iniciou seus estudos tabelando a altura de pais e filhos. Os dados foram coletados no final do século XIX na Inglaterra. Ele cunhou o termo regressão à mediocridade para descrever o resultado de seu modelo linear (1886).

Analisaremos este estudo nas próximas aulas.



<https://commons.wikimedia.org/wiki/File:Galton-height-regress.png>

Galton e seus estudos

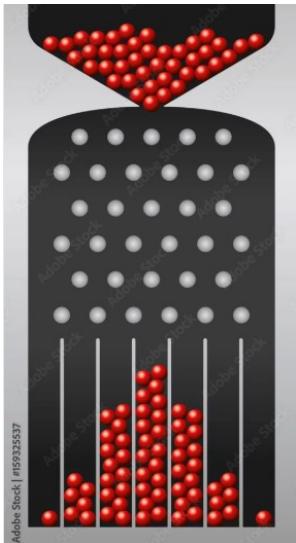
Os estudos de Galton foram pioneiros na aplicação de métodos quantitativos a características humanas, com foco em hereditariedade, inteligência e eugenia. Ele introduziu conceitos como:

- correlação
- regressão à média
- processos de ramificação (Galton-Watson)
- psicometria
- análise de impressões digitais

utilizando pesquisas e dados antropométricos para compreender a variação e influenciar campos como estatística e genética.

Estatística e Correlação:

- Correlação: Desenvolveu o conceito para medir relações entre variáveis.
- Regressão à Média: Observou que características extremas nos pais tendem a ser menos extremas nos filhos (mais próximas da média), crucial para o planejamento experimental.
- Painel de Galton: demonstrou visualmente a distribuição normal e a regressão usando sua tábua de feijões.



Tábua de feijões

Biometria e Psicometria:

- **Antropometria:** primeiro laboratório antropométrico para coletar dados sobre a variação fenotípica humana, sendo pioneiro no uso de pesquisas para dados sociais.
- **Psicometria:** Fundou este campo aplicando a mensuração a características psíquicas, estudando as diferenças individuais na mente (seguido por Carl Gustav Jung - psicologia)

Gemini

<https://stock.adobe.com/br/images/galton-board-showing-normal-distribution-generating-gaussian-bell-curve-also-bean-machine-quincunx-or-galton-box-device-to-demonstrate-the-central-limit-theorem-mathematics-illustration-vector/159325537>

Galton e seus estudos

Galton foi um pioneiro que trouxe uma abordagem estatística para a compreensão das características humanas, lançando as bases para a estatística, genética e psicologia modernas.

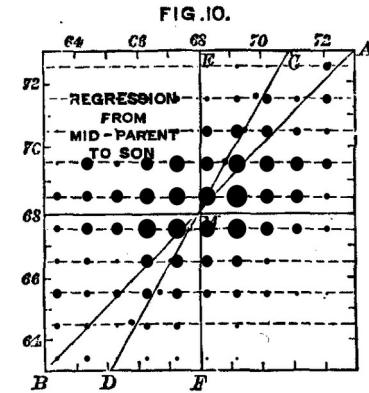
Seu aluno, Pearson, colaborou ainda mais como novas ferramentas matemáticas e um desenvolvimento matemático mais formal.

Galton e seus estudos

O que Galton observou

Galton estudou características herdáveis, como altura. Ele notou que:

1. Pais muito altos tendiam a ter filhos altos, mas menos extremos que eles.
2. Pais muito baixos tendiam a ter filhos baixos, mas também mais próximos da média da população.



Ou seja, os descendentes “regrediam” em direção à média (mediocrity ou média ou “algo comum”).

Por que “mediocridade”? No inglês do século XIX, mediocrity significava simplesmente “valor médio”, não “algo ruim” ou “sem qualidade”.

Portanto, “regressão à mediocridade” queria dizer: valores extremos tendem a ser seguidos por valores menos extremos, mais próximos da média, quando a correlação entre gerações é imperfeita.

Galton e Eugenia

Eugenia:

Promoveu a eugenica (University College London), com o objetivo de melhorar a hereditariedade humana por meio da reprodução seletiva.

As consequências deste trabalho e da sociedade de eugenica inglesa são vistos criticamente como ‘racistas’, pois colaboraram na criação de movimentos eugenistas racistas na Inglaterra, USA (KKK) e Alemanha (nazismo).

Já ao nascer, a eugenia está em crise!

Darwin: homens e macacos têm um ancestral comum

Galton: “Lamarck não funciona”, nossa população não está se tornando mais alta e mais forte.



Desafio

Média de altura:

Brasil: Homens (~1.74 m) Mulheres (1.61 m)

Alemanha: Homens (~1.79 m) Mulheres (1.67 m)

Holanda: Homens (~1.84 m) Mulheres (1.71 m)

Dados prováveis (Gemini)

A média de altura da Holanda, atualmente, é muito alta.

1. Porque homens na Holanda são tão altos?
2. Como as teorias que você estudou até agora explicam este fenômeno?

Karl Pearson (1857-1936)



Pearson with Sir Francis Galton, 1909 or 1910
https://en.wikipedia.org/wiki/Karl_Pearson

Pearson e seus estudos

Karl Pearson foi um dos pioneiros da estatística e colaborador junto a Sir Francis Galton, cujo trabalho em hereditariedade e antropometria lançou as bases para o desenvolvimento da estatística moderna. Enquanto Galton concebeu as ideias iniciais e intuitivas de correlação e regressão, foi Pearson quem desenvolveu os conceitos usando uma matemática rigorosa.

- coeficiente de correlação (r ou ρ Pearson) que mede a força de associação de duas variáveis
- Chi-quadrado (chi-squared test) ~ teste de ajuste perfeito (goodness of fit)
- Regressão múltipla

Pearson fundou o departamento de Estatística Aplicada na University College of London (UCL) e escreveu uma bibliografia sobre Galton.

A correlação de Pearson mede a associação entre duas variáveis, porém não mede a causa. Algo que só foi elucidado por Judea Pearl (a partir dos anos 1980)

Redescoberta dos estudos de Mendel (1965-1900)

Redescoberta dos estudos de Mendel

A redescoberta do trabalho de Gregor Mendel ocorreu em 1900, 35 anos após a primeira publicação. Ou seja, o artigo original de Mendel, de 1865, "Experimentos sobre Hibridização de Plantas", não foi lido por importantes pesquisadores durante sua vida, e foi redescoberto de forma independentemente por três botânicos que trabalhavam em problemas semelhantes de hereditariedade.

Os três cientistas aos quais são creditados a redescoberta dos estudos de Mendel, são:

- Hugo de Vries (Holanda): Conduziu experimentos com várias plantas, incluindo a prímula (*Oenothera lamarckiana*), e chegou à proporção de 3:1 antes de citar o trabalho de Mendel em suas publicações subsequentes.
- Carl Correns (Alemanha): Trabalhando com milho e ervilhas, ele percebeu a importância das proporções de Mendel e é creditado por formular as "Leis Mendelianas" de segregação e distribuição independente, como as conhecemos hoje.
- Erich von Tschermak (Áustria): Um desenvolvedor de novas variedades de plantas que também observou a proporção 3:1 nas ervilhas.

Redescoberta dos estudos de Mendel

Por que o trabalho de Mendel foi inicialmente ignorado?

1. **Abordagem matemática:** Mendel usou estatística e probabilidade para explicar fenômenos biológicos, um método que era amplamente desconhecido e pouco aceito para os botânicos do século XIX.
2. **Foco em Darwin:** A comunidade científica da década de 1860 estava preocupada com a teoria da evolução de Charles Darwin, que se concentrava na **variação contínua** em vez dos "**fatores discretos**" descritos por Mendel.
3. **Publicação em revista desconhecida:** Seu trabalho foi publicado num periódico de uma sociedade local de história natural em Brno, que tinha distribuição limitada.
4. Mendel não fez parte da sociedade científica 'formal'

Redescoberta dos estudos de Mendel

Impacto da Redescoberta

- 1. Nascimento da Genética:** A redescoberta transformou a biologia de uma ciência descritiva para uma experimental. Wilhelm Johannsen introduz o termo “gene” que era um “fator” para Mendel e o termo “genética” foi cunhado posteriormente, em 1905, por William Bateson, que se tornou um dos principais defensores das teorias de Mendel.
- 2. Teoria Cromossômica:** No início do século XX, cientistas como Walter Sutton e Theodor Boveri associaram os “fatores” de Mendel aos cromossomos, fornecendo uma base física para a hereditariedade (teoria que a informação hereditária) que nos anos 1990 foram complementadas com a descoberta da epigenética.
- 3. Síntese Moderna:** O legado de Mendel permanece como a base da genética moderna, recentemente evidenciado pela conclusão, em 2025, do mapeamento dos sete genes específicos que Mendel estudou originalmente no genoma da ervilha (estudar: Extended Evolutionary Synthesis)

Síntese moderna: simplificada

Síntese moderna

1. Não existem grandes saltos (de Vries, saltacionistas)
2. As mutações são pequenas alterações nos genes e podem alterar a distribuição fenotípica (variância das características medidas).
3. A herança mendeliana conserva as mutações (Punnett e Equilíbrio de Hardy-Weinberg)
4. A seleção natural atua sobre estas pequenas variações mendelianas.

Ronald Fisher: estatística moderna

O legado de Ronald Fisher: pai da estatística moderna



Sir Ronald Fisher (1889-1962) é o pai da estatística moderna, revolucionando a área ao estabelecer as bases para o planejamento experimental (ANOVA, randomização), teste de hipóteses, estimativa de máxima verossimilhança e desenvolver distribuições estatísticas importantes, como a distribuição F, tornando a estatística uma disciplina matemática rigorosa, e não apenas uma ferramenta descritiva. Seu trabalho na Estação Experimental de Rothamsted aplicou a teoria estatística a dados do mundo real, influenciando a agricultura, a genética (Síntese Moderna) e inúmeros outros campos.

O movimento de Síntese Evolucionária ou Síntese Moderna foi iniciado por ele, Sewall Wright e JBS Haldane que fundaram os estudos matemáticos de genética populacional e permitiram a unificação das teorias de Darwin e Mendel.

Como não há perfeição, Ronald Fisher também era simpatizante da Sociedade de Eugenia Inglesa assim como Galton e Pearson. Por ser fumante, como uma visão equivocada, ele foi contra a teoria de que o fumo causa câncer de pulmão, estudos dos anos 1950 na briga contra a indústria de cigarros. Na época, foi demonstrado graficamente esta associação, mas a causalidade ficou em aberto.

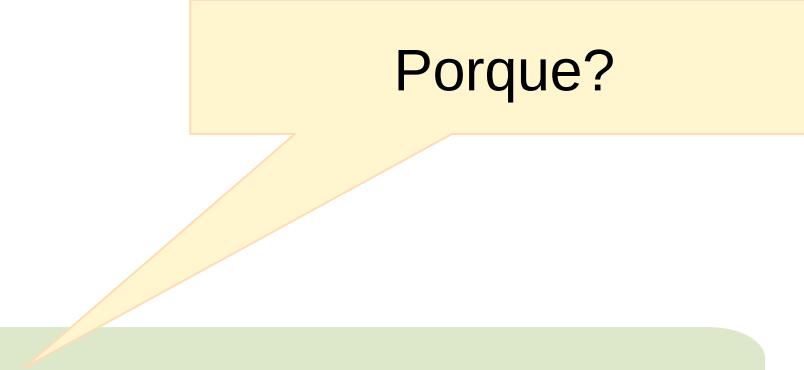
O legado de Ronald Fisher: pai da estatística moderna

O grande pioneiro da Estatística Moderna

1. Planejamento Experimental: Desenvolveu princípios como randomização, replicação e blocos, demonstrados de forma notável no famoso problema da "Senhora Provando Chá", garantindo conclusões válidas a partir de experimentos.
2. Analisou as teorias de Darwin e Mendel sob a ótica de Variância.
3. Criou a Análise de Variância (ANOVA) para particionar a variação, um pilar fundamental para a comparação de médias em experimentos.
4. Teste de Hipóteses e Inferência: Introduziu o conceito de p-value (p-valor), o teste de significância e o conceito de hipótese nula, formando o núcleo da inferência estatística.
5. Estimação de Máxima Verossimilhança (EMV): Forneceu um método poderoso para estimar parâmetros de modelos estatísticos.
6. Distribuições Estatísticas: Desenvolveu a distribuição F e esclareceu o uso da distribuição Qui-quadrado, vital para testes de hipóteses.
7. É autor de “The Genetic Theory of Natural Selection” (1930) e "Statistical Method for Research Workers" (1934).

Sem estatística não há ciência

Porque?



Sem estatística não há ciência

Sem estatística não há ciência

Porque a ciência tem que ser quantitativa:

- Fazer hipóteses
- Desenhar experimentos
- Definir variáveis
- Medir as variáveis
- Comparar resultados: testes de hipótese

Sem estatística não há ciência

Porque a ciência tem que ser quantitativa:

- Fazer hipóteses
- Desenhar experimentos
- Definir variáveis
- Medir as variáveis
- Comparar resultados: testes de hipótese

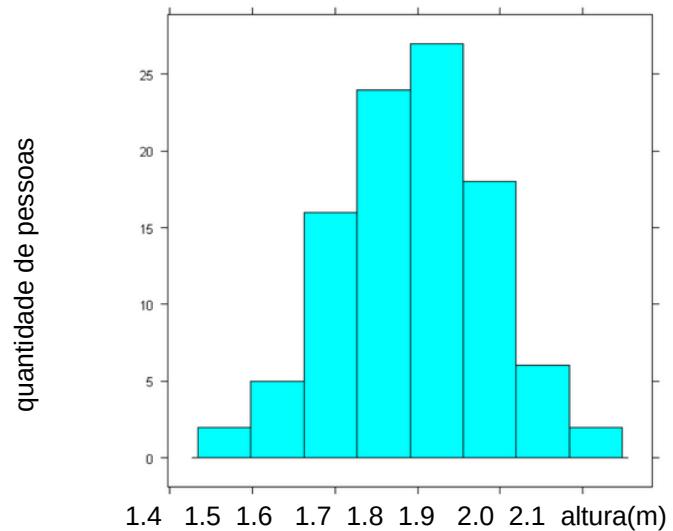
E responder:

- Qual teste estatístico usar?
- Há efeito? Qual seu tamanho?
- Aumentou? Diminuiu?

Uma Análise Estatística se inicia com contagens

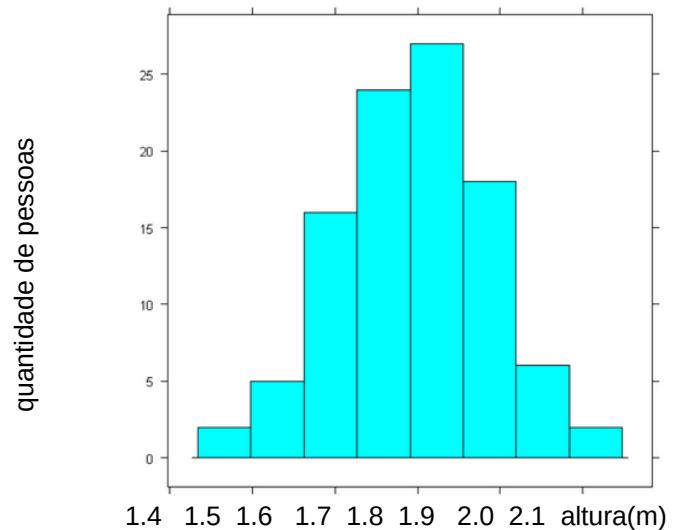
Uma Análise Estatística se inicia com contagens

Calcula-se frequências e distribuições



Uma Análise Estatística se inicia com contagens

Calcula-se frequências e distribuições



Freq e Distribuição
controle

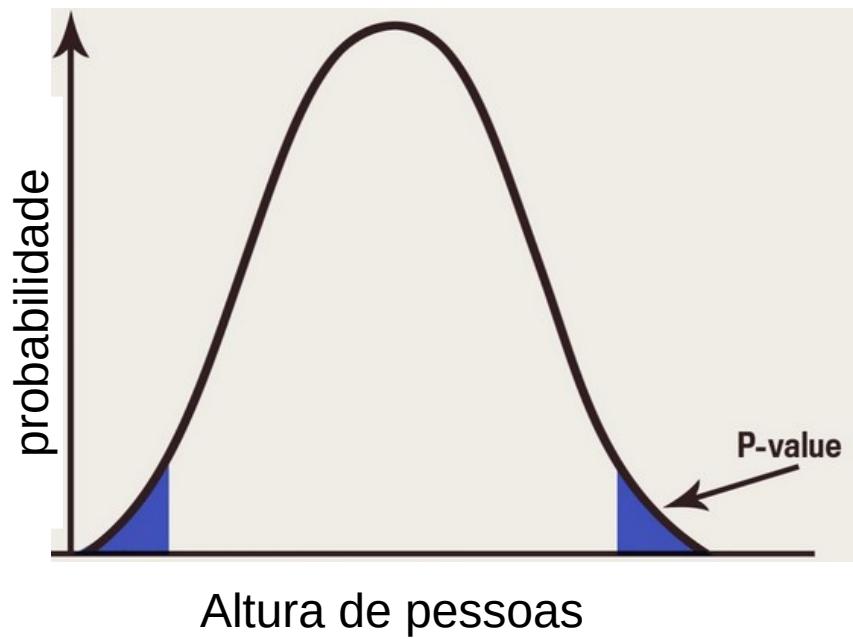
Freq e Distribuição
casos (perturbações)

Uma Análise Estatística se inicia com contagens

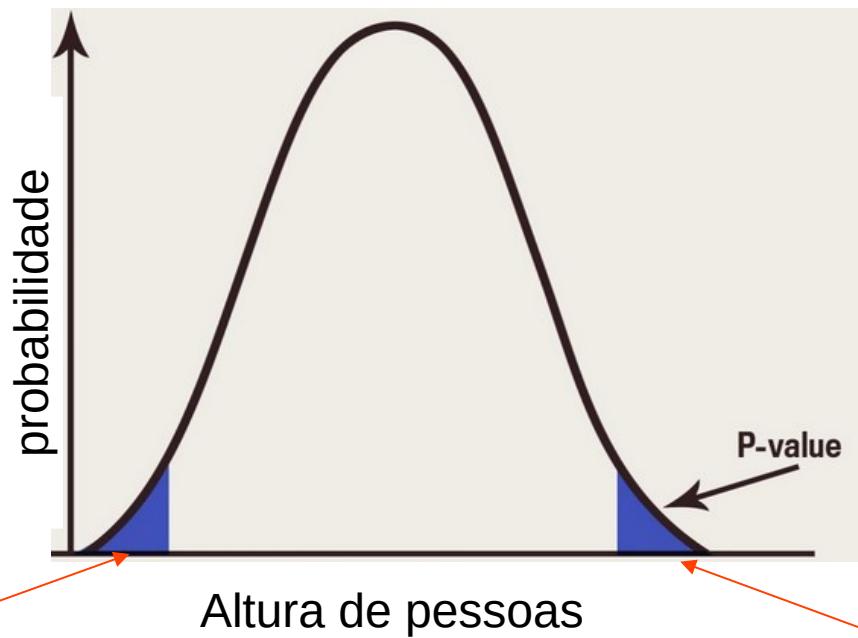
Calcula-se frequências e distribuições

Compara-se as distribuições
com testes de hipóteses

p-valor bem pequeno é algo raro!



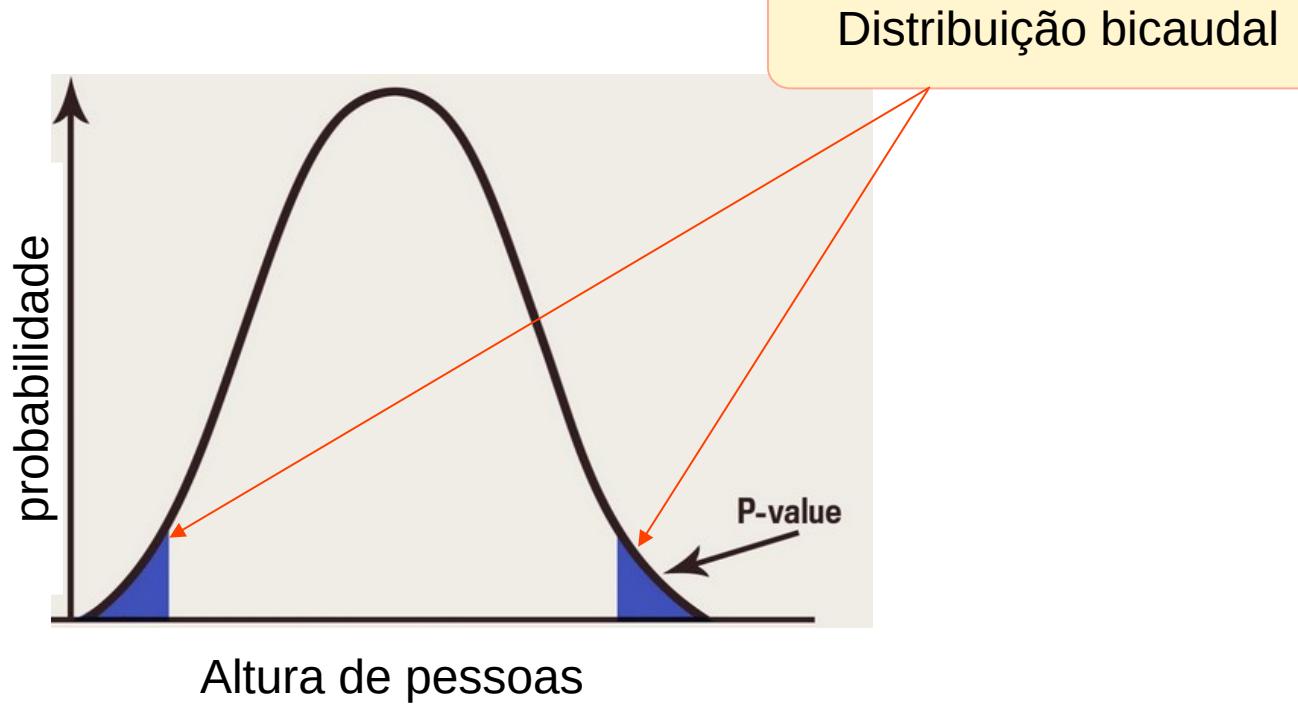
p-valor bem pequeno é algo raro!



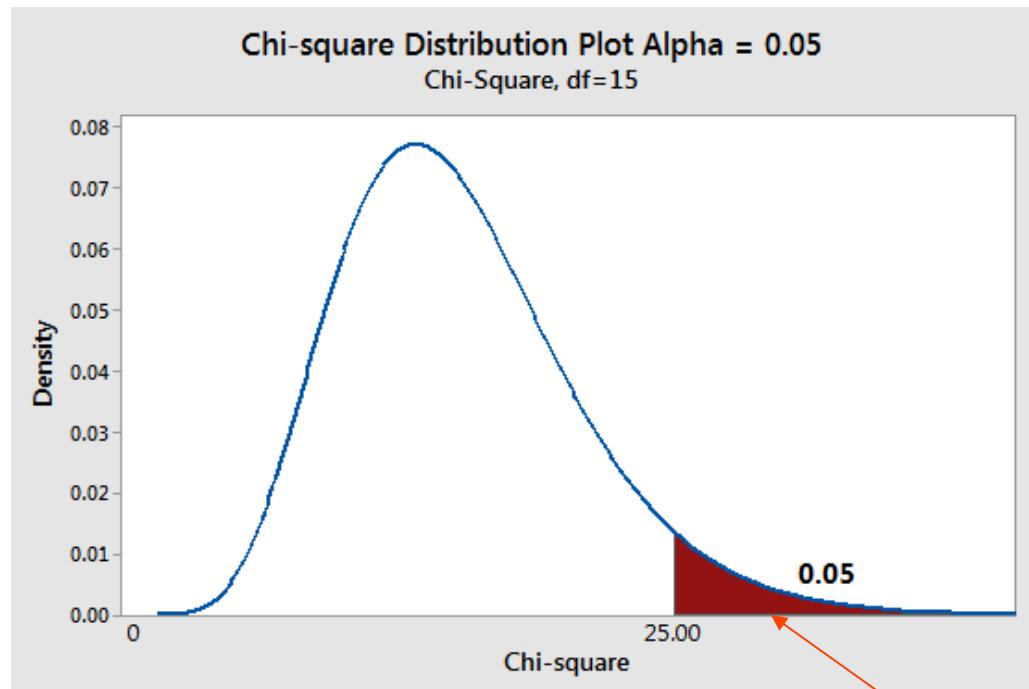
Pessoas muito baixas
estão no limite inferior
área abaixo de 2.5%

Pessoas muito altas
estão no limite superior
área acima de 97.5%

Distribuição bicaudal (two-tailed distribution)

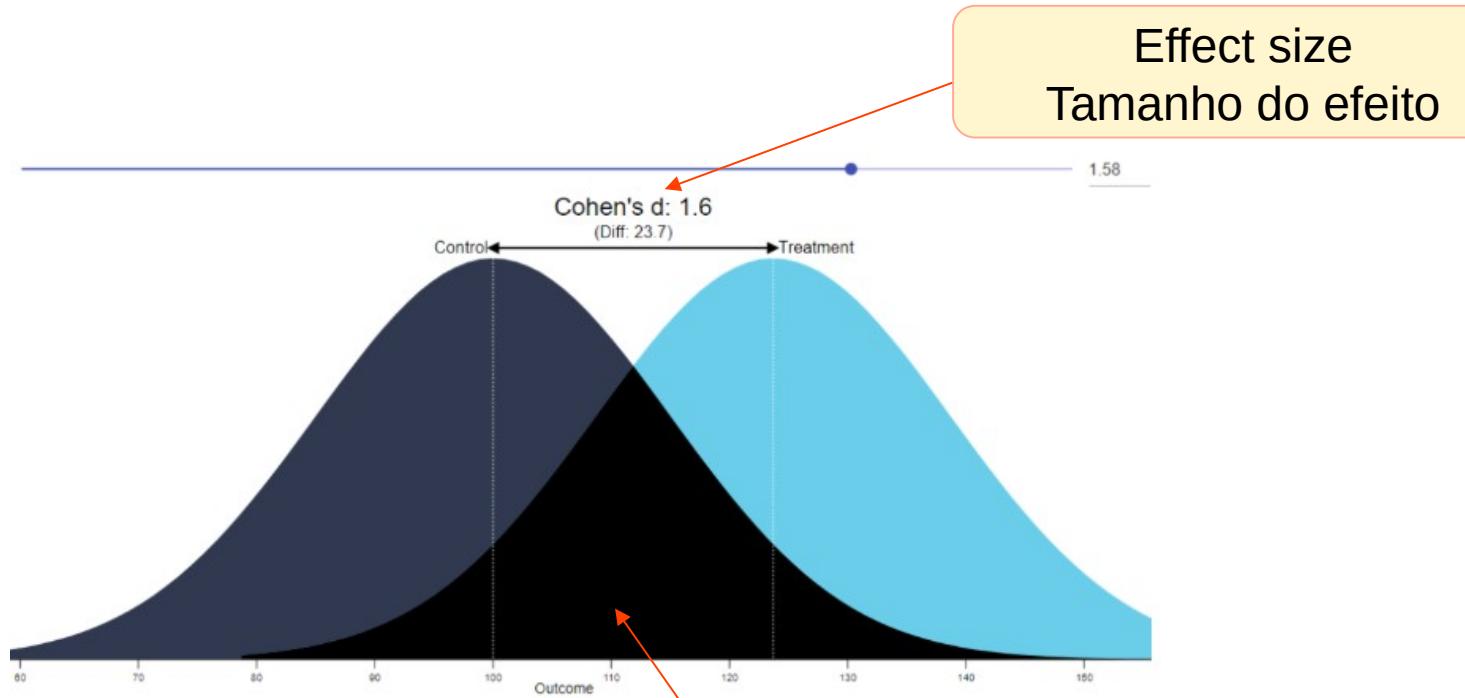


Distribuição moncaudal (one-tailed distribution)



Distribuição moncaudal

Tamanho de Efeito: é diferente?



As duas distribuições
são diferentes?

Bioestatística com Python

Agenda

✓ Frequência e Distribuição

✓ Distribuições discretas:

- Jogos de Azar: experimento de Bernoulli: jogando 1 moeda

- Binomial: jogando várias moedas:

- Canais iônicos

- Poisson:

- Mutações

✓ Distribuições contínuas: normal, z, t

✓ Experimento de Ciclo-Celular com dados de Citômetro de Fluxo

✓ Testes de Hipóteses

Obrigado Perguntas?



PhD Flávio Lichtenstein

Bioinformatics & Systems Biology Lab
Molecular Biology Lab
Development and Innovation Center (CDI)

flavio.lichtenstein@butantan.gov.br



centre of
excellence
in new target
discovery