

Introdução aos Métodos Estatísticos para Bioinformática

***Profa. Júlia Maria Pavan Soler
pavan@ime.usp.br***

***IBI 5086 – Bioinformática - IME/USP
2º Sem/2019***

Programa

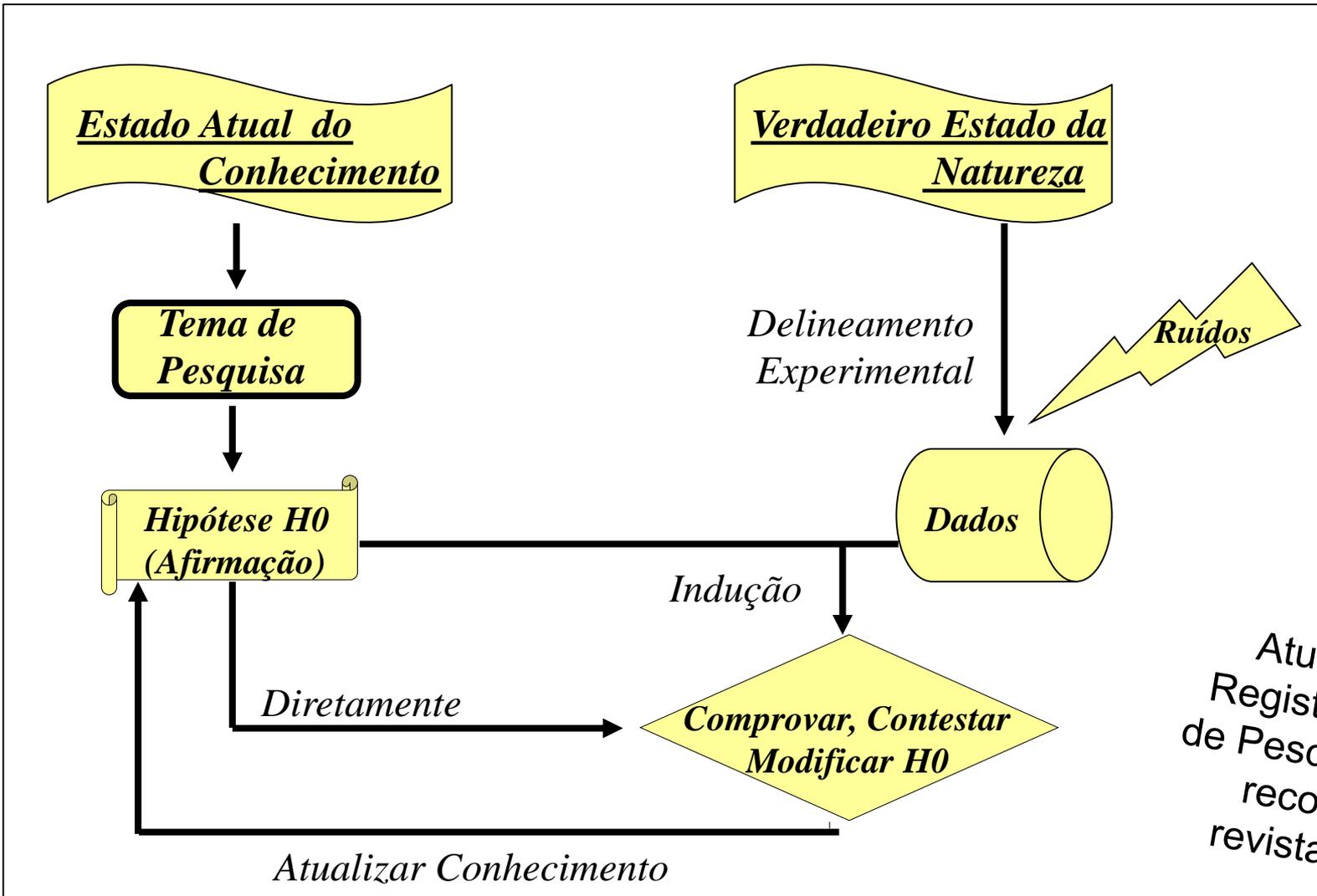
1. Álgebra linear básica: cálculo matricial, determinantes, sistemas lineares, produto interno, norma, ortogonalidade, autovalores e autovetores
2. Variáveis quantitativas: testes de hipóteses para comparação de médias (Testes t e ANOVA), teste de hipóteses para variância, correlação e modelos de regressão
3. Variáveis qualitativas ou categorizadas: tabelas de contingência, testes qui-quadrado, concordância, modelos de regressão logística.
4. Métodos de Análise Multivariada de Dados: agrupamentos, componentes principais, discriminação e classificação, análise de variância e de regressão multivariada, correlação canônica
5. Intervalos de Confiança e Testes baseados em Bootstrap e Permutação
6. Correções para múltiplos testes (Bonferroni e FDR)
7. Simulação de Monte Carlo

Motivação

- **Por que coletamos dados?**
- **Como coletar os dados? Que critérios usar?**
- **O que estamos fazendo com os dados que coletamos?**
- **Os resultados obtidos são válidos e reprodutíveis?**



Estatística e Conhecimento Científico



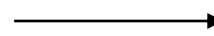
Atualmente o Pré-Registro de Projetos de Pesquisa tem sido recomendado por revistas científicas.

Estatística e Conhecimento Científico



**Como fazer
Ciência?**

Observar fatos



Criar Teorias

Investigação Científica

- Não existe um procedimento único
- Não é uniforme mas é convergente

Assegurar Convergência

- Hipóteses bem elaboradas
- Plano amostral e experimental eficiente
- Análise de dados sensível e legítima

BOAS PRÁTICAS



Para prevenir novas crises

Documento propõe mudanças em práticas de pesquisa a fim de reduzir a publicação de artigos cujos resultados ninguém consegue repetir

A Real Academia de Artes e Ciências da Holanda lançou um documento propondo mudanças em práticas de pesquisa a fim de enfrentar o que se convencionou chamar de “crise da reprodutibilidade” – uma sucessão de casos de artigos científicos que caíram em descrédito por seus resultados não terem sido confirmados em experimentos subsequentes. As recomendações do relatório divulgado em janeiro, intitulado *Estudos de replicação – Melhorando a reprodutibilidade nas ciências empíricas*, buscam aumentar o rigor com que o trabalho científico é realizado e apoiar pesquisadores interessados em verificar resultados obtidos por colegas. Uma das propostas apresentadas consiste em estimular o financiamento de estudos voltados para ratificar outros estudos, seguindo o exemplo da Organização Holandesa para Pesquisa Científica (NWO), que destinou no ano passado € 3 milhões a um programa-piloto para projetos dessa natureza. As sugestões também incluem reforçar o treinamento de cientistas e estudantes em tópicos como desenho de experimentos e análise estatística, e incentivar periódicos científicos a publicar pesquisas que não confirmaram as hipóteses testadas ou então chegaram a resultados nulos.

“O conhecimento só avança se os pesquisadores puderem confiar nos resultados de estudos feitos anteriormente”, escreveu, no prefácio do relatório, a presidente da Real Academia, José van Dijk, pesquisadora de mídia e cultura da Universidade de Utrecht. Na avaliação da entidade, produzir dados fidedignos é essencial para evitar o desperdício de recursos na pesquisa e garantir a confiança do público na ciência. “O relatório conclui que estudos para reproduzir outros estudos devem ser conduzidos de forma mais frequente e sistemática, o que exige um esforço conjunto de agências de fomento, pesquisadores, periódicos e instituições científicas”, afirmou Van Dijk.

O panorama da “crise da reprodutibilidade” apresentado no relatório mostra a relevância do problema. À procura de novos medicamentos contra o câncer, a empresa farmacêutica Amgen tentou confirmar os achados de 53 estudos pré-clínicos publicados que pareciam ter grande potencial. Apenas 11% dos resultados foram corroborados. A Bayer fez um esforço semelhante para tentar validar dados sobre alvos potenciais de novos remédios, obtidos por 67 projetos de pesquisa, e só teve sucesso em 25% dos casos. Uma colaboração internacional para investigar estudos de

psicologia experimental, área que foi palco de escândalos envolvendo manipulações e fraudes, só conseguiu ratificar resultados obtidos em 36 dos 100 artigos avaliados. No final do ano passado, as Academias Nacionais de Ciências, Engenharias e Medicina dos Estados Unidos criaram um comitê de 15 membros para estudar estratégias que previnam a publicação de estudos não confirmáveis – as conclusões devem ser divulgadas em 2019.

Embora a crise se manifeste de modo aberto em medicina, ciências da vida e psicologia, o documento propõe que as demais áreas investiguem a extensão do problema em suas comunidades. “Quando observamos o conhecimento existente sobre causas da irreprodutibilidade, fica claro que muitas delas estão presentes em todas as disciplinas”, disse à revista *Science* Johan Mackenbach, pesquisador da área de saúde pública do Centro Médico Erasmus, em Roterdã, chefe do painel que organizou o relatório. Ele se refere a causas genéricas apontadas pelo documento, como a pressão para que pesquisadores publiquem novidades ou achados de impacto da forma mais rápida possível, caso contrário ficam em desvantagem na competição por financiamento e oportunidades de trabalho.

O relatório enumera 20 diferentes razões para uma pesquisa chegar a resultados não confirmáveis. A maioria está relacionada a questões metodológicas, como falhas no controle de vieses, conclusões baseadas em amostras restritas ou falta de rigor estatístico na interpretação de dados. Na origem do problema, também há vícios na forma de reportar resultados, como selecionar dados favoráveis à hipótese da pesquisa, omitindo os negativos, ou modificar a proposta original de modo a adaptá-la às conclusões obtidas.

Fraudes são o jeito mais extremo de gerar resultados inválidos, mas também existem ameaças à reprodutibilidade que são parte natural da atividade científica, como falhas humanas, erros técnicos inesperados ou mudanças não detectadas nas condições de

20 causas da irreprodutibilidade

Por que resultados de alguns trabalhos científicos não são confirmados por outros estudos

- ◆ Desenho experimental ineficiente associado a controle de vieses falho
- ◆ Amostras de tamanho insuficiente
- ◆ Problemas em testes estatísticos que geram falsos resultados negativos
- ◆ Erro técnico ou humano na execução do estudo, associado a controle de qualidade ineficaz
- ◆ Fraude ou fabricação de dados
- ◆ Falta de rigor na análise estatística
- ◆ Análise estatística equivocada
- ◆ Falta de conhecimento sobre variáveis que influenciam o resultado
- ◆ Falhas do pesquisador em reproduzir os resultados antes da publicação
- ◆ Omissão de resultados nulos ou análise seletiva que faz os nulos parecerem positivos
- ◆ Não compartilhamento de dados ou de detalhes metodológicos
- ◆ Escolha de variáveis que se adequam aos resultados
- ◆ Formulação de hipótese depois que os resultados são conhecidos
- ◆ Discrepância entre os resultados registrados e os publicados
- ◆ Ausência de revisão por pares adequada
- ◆ Ênfase no incentivo a artigos de alto impacto
- ◆ Recompensas exageradas a resultados de pesquisa tidos como disruptivos
- ◆ Sistemas de financiamento à pesquisa demasiadamente competitivos
- ◆ Falta de recompensa para práticas que favoreçam a replicação de estudos
- ◆ Crença de que o rigor no processo de pesquisa dificulta novas descobertas

FONTE REPLICATION STUDIES – IMPROVING REPRODUCIBILITY IN THE EMPIRICAL SCIENCES, 2018

amostras. Nem todos os estudos não ratificados estão equivocados. Há situações em que não é possível alcançar o resultado obtido porque o autor da pesquisa deixou de informar detalhes do experimento essenciais para sua replicação. Para evitar essas situações, diz o relatório, periódicos e agências de fomento devem exigir que pesquisadores disponibilizem em repositórios públicos seus dados brutos de pesquisa e informações sobre as metodologias que adotaram.

Há propostas de caráter prático para prevenir desvios, como determinar que o pesquisador, ao propor seu projeto, registre antecipadamente e de forma transparente sua hipótese, o protocolo de pesquisa e o plano de análise a serem adotados. Esse tipo de precaução já é exigido por agências de fomento, mas vem sendo aprimorado. Em uma iniciativa articulada pela Royal Society, do Reino Unido, vários periódicos já

publicam sistematicamente os chamados relatórios registrados, um tipo de *paper* que apresenta os métodos e planos de análise de uma pesquisa ainda não iniciada, mas que foram avaliados por pares. Mais tarde, as revistas publicam os resultados obtidos, mesmo que sejam nulos.

O relatório é explícito sobre a importância de divulgar estudos com resultados nulos. Propõe que agências de fomento incentivem os pesquisadores a reportar tais conclusões e as revistas a publicá-las. “Instituições de pesquisa, agências de fomento e periódicos devem oferecer aos pesquisadores incentivos para a condução de estudos rigorosos e a produção de resultados de pesquisa reproduzíveis, em vez de recompensá-los principalmente por publicações de ‘alto impacto’, estudos ‘inovadores’ e afirmações infladas”, afirma o documento. ■

20 causas da irreprodutibilidade

Por que resultados de alguns trabalhos científicos não são confirmados por outros estudos

- ◆ Desenho experimental ineficiente associado a controle de vieses falho
- ◆ Amostras de tamanho insuficiente
- ◆ Problemas em testes estatísticos que geram falsos resultados negativos
- ◆ Erro técnico ou humano na execução do estudo, associado a controle de qualidade ineficaz
- ◆ Fraude ou fabricação de dados
- ◆ Falta de rigor na análise estatística
- ◆ Análise estatística equivocada
- ◆ Falta de conhecimento sobre variáveis que influenciam o resultado
- ◆ Falhas do pesquisador em reproduzir os resultados antes da publicação
- ◆ Omissão de resultados nulos ou análise seletiva que faz os nulos parecerem positivos
- ◆ Não compartilhamento de dados ou de detalhes metodológicos
- ◆ Escolha de variáveis que se adequam aos resultados
- ◆ Formulação de hipótese depois que os resultados são conhecidos
- ◆ Discrepância entre os resultados registrados e os publicados
- ◆ Ausência de revisão por pares adequada
- ◆ Ênfase no incentivo a artigos de alto impacto
- ◆ Recompensas exageradas a resultados de pesquisa tidos como disruptivos
- ◆ Sistemas de financiamento à pesquisa demasiadamente competitivos
- ◆ Falta de recompensa para práticas que favoreçam a replicação de estudos
- ◆ Crença de que o rigor no processo de pesquisa dificulta novas descobertas

FONTE REPLICATION STUDIES – IMPROVING REPRODUCIBILITY IN THE EMPIRICAL SCIENCES, 2018

**Leia também o artigo da BBC-News (Feb/2019):
Machine learning 'causing science crisis'**

Pesquisas nas Áreas Factuais

Estudos Observacionais

Não há intervenção, nenhum fator sendo controlado

- Casos de uma síndrome no município
- Distribuição de um CNV

Variação	Frequência
-2	1
-1	5
0	34
1	12
2	3

Estudos Experimentais

Há intervenção, há fatores sendo controlados

- Efeito do tratamento na intensidade de expressão gênica (y)

Dose	Baixa		Média		Alta	
Trat.	T1	T2	T1	T2	T1	T2
	—	—		—	—	—
	—	—	y	—	—	—
	—	—		—	—	—

Coleta de Dados

Estudos Observacionais

- Prospectivos
- Retrospectivos
- Transversais

Em geral, os dados desses estudos podem ser dispostos em Tabelas de Contingência para estudos de associação. Ex. GWAS

Estudos Experimentais

- Estrutura dos Tratamentos:
fatores e seus níveis
- Estrutura das Unidades Experimentais
aleatorização e réplicas
- Blocagem e Covariáveis

Em geral esses dados são analisados por meio de modelos de Análise de Variância (ANOVA) ou Regressão. Ex. Experimentos de Microarrays

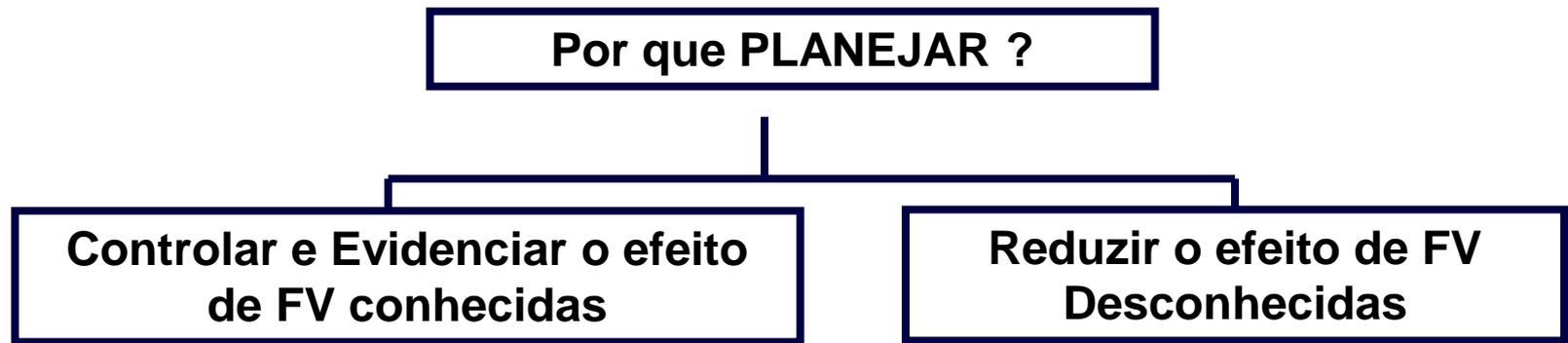


Estudos experimentais como os Ensaio Clínicos Controlados e Aleatorizados são o padrão ouro dos estudos de causalidade (indústria Farmacêutica). Porém, atualmente, o desafio é extrair relações causais de Estudos Observacionais (Teoria de Grafos, Modelos de Equações Estruturais, Aleatorização Mendeliana, Variáveis Instrumentais)

Planejamento de Experimentos

- Experimentos na Agricultura - Fisher (1920s): Aleatorização, replicação, bloco
- Experimentos Industriais - Box e Wilson (1950s), Box (1999): Modelos de superfície de respostas
- Experimentos para o Controle da Qualidade – Taguchi (1980s): processos robustos, resistentes a variações transmitidas de componentes, Modelos fatoriais fracionais, Gráficos de Controle
- Planejamento de Experimentos na área da Genômica: Experimentos em Microarrays (Bioconductor-R); Experimentos em Proteômica (Espectrometria de Massas, MSstats-R, MixOmics)
- ...

Planejamento de Experimentos



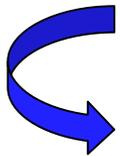
modelo estatístico

$$Y = FV \text{ conhecidas} + e$$

Variável resposta de interesse

Fontes de variação conhecidas (preditores de interesse e var. de controle)

Erro (fontes de variação desconhecidas)



Experimentos Controlados e aleatorizados representam o “padrão ouro” para a realização de Inferências Causais
Exemplo: Ensaios Clínicos da indústria Farmacêutica (ANVISA – Agência reguladora)

Modelo Estatístico

$$Y = FV \text{ conhecidas} + e$$

Princípio da Análise: Avaliar a “importância” de possíveis fontes de variação (FV conhecidas e desconhecidas) que atuam sobre uma resposta de interesse.

A Estatística é uma lição de humildade, pois temos que compreender o significado do termo “erro”.

Planejamento de Experimentos

- População Alvo do estudo
 - População de Interesse
 - Unidades amostrais (como foram coletadas?)
- Estrutura de Tratamento e Variáveis de Controle (X)
 - Fatores e seus níveis
 - Unidades Experimentais - Aleatorização
 - Blocagem e Covariáveis
- Variável resposta de interesse
 - Unidades de mensuração
 - Réplicas

Classificação de Variáveis

Modelo Estatístico

$$Y = FV \text{ conhecidas} + e$$

Componente Fixo

Componente Aleatório

Y: var. resposta (dependente, aleatória)

X: var. preditora ou fator de interesse (independente, fixada)

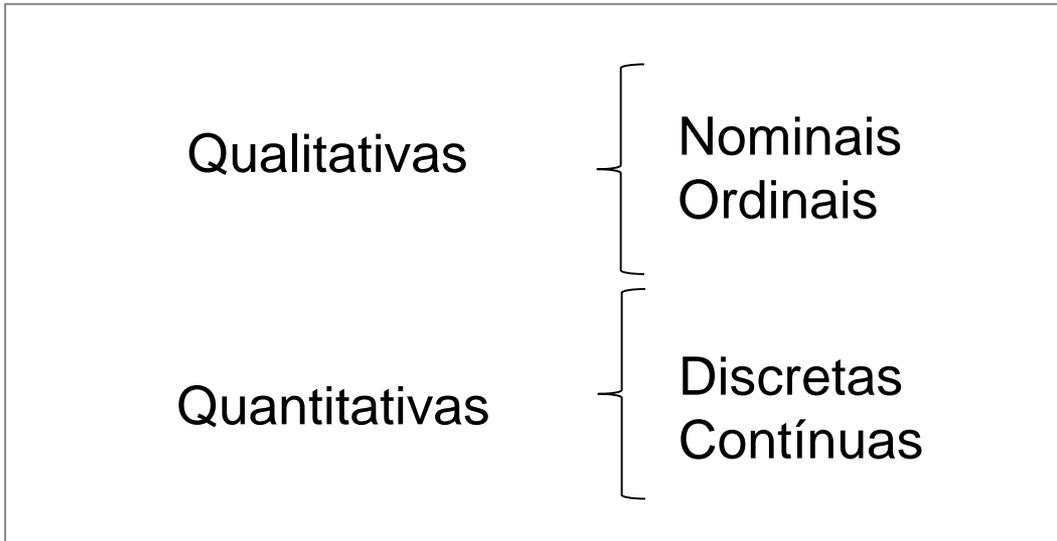
f : função associando X a Y, em geral, linear e assumida conhecida, mas pode ser não linear ou desconhecida

e : variável aleatória (em geral, envolve somente o termo de erro, mas pode ser decomposto em efeitos aleatórios de interesse no estudo além do erro)



Discutiremos diferentes situações ao longo do curso!

Classificação de Variáveis no Modelo Estatístico



*Modelos ANOVA e de Regressão
Abordagens diferentes mas
unificadas via a teoria de
modelos lineares!*

Modelos ANOVA: Y é variável resposta quantitativa univariada e X variável independente qualitativa (especificação de efeitos de grupos bem como covariáveis)

Modelos de Regressão: Y é variável resposta quantitativa univariada e X variável independente quantitativa (variáveis preditoras bem como covariáveis)

Princípios do Planejamento de Experimentos e da Análise de Dados

O que se quer **EXTRAIR** dos dados?

→ estimativas confiáveis, com a maior precisão possível, do efeito das variáveis preditoras X (tratamentos) sobre Y

Como planejar um experimento para a coleta dos dados?

→ Evitar erros sistemáticos (vícios de seleção, desgaste de aparelho, diferentes fornecedores, etc.), controlar fontes de variação conhecidas mas que não são de interesse primário.

Contudo, **NÃO** é possível controlar flutuações (aleatórias) nos dados por fontes desconhecidas

→ **Aleatorização e Cegamento** são técnicas para evitar vícios sistemáticos e “balancear” o efeito do erro entre os grupos de tratamentos (gerar grupos comparáveis, com características semelhantes)

→ A posteriori: **eliminar o efeito de fontes desconhecidas, sistemáticas, por meio de ajuste de modelos e obter resíduos (respostas filtradas, corrigidas)**

Estratégia
a priori!

Princípios do Planejamento de Experimentos e da Análise de Dados

Como extrair **INFORMAÇÃO** dos dados?

Como planejar um experimento para garantir qualidade dos dados?



Análise dos Dados

$$Y = f(X) + e$$

Adotar Modelos que estruturam apropriadamente os dados e levem em conta o planejamento adotado na sua coleta. Os modelos devem respeitar a escala da variável resposta (Y) e características do planejamento (variáveis preditoras e de controle, independência ou não entre as observações, etc.). Os pressupostos devem ser validados por meio de análises de diagnóstico ou outros métodos (cross-validation)

Em geral: Média de Y em função de X e a Covariância de Y em função do erro

$$E(Y) = E[f(X)]; \quad Cov(Y) = Cov(e)$$

componente fixo



Modela a fonte de
variação devido a X

componente aleatório



Modela a fonte de
variação devido ao "erro"

Princípios do Planejamento de Experimentos e da Análise de Dados

Como tomar decisão sobre hipóteses científicas?

- Pré-Registrar o Plano de Pesquisa
- Coletar os Dados
- Realizar a análise estatística dos dados (Intervalos de Confiança x Intervalos de Compatibilidade com os Dados)
- Realizar a análise factual profunda e detalhada dos resultados, em coerência com o Plano Inicial
- Extrair conclusões dessas duas análises

Acompanhe as discussões sobre "decisões baseadas em significância estatística!"

Motivação: Experimentos Comparativos

Objetivo: Introduzir e discutir alguns conceitos em análise exploratória de dados e análise inferencial a partir de um problema de “Comparação de duas Populações”.

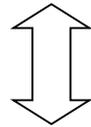
- Delineamento experimental e estrutura dos dados
- Medidas resumo e Representação Gráfica
- Testes Clássicos (Testes “t” e testes não-paramétricos)
- Uso de Dados Externos na construção de testes
- Testes de aleatorização

*Temas a serem
abordados*

Comparando Duas Populações

Hipóteses Científicas do Estudo

Hipóteses Estatística (Parâmetros-População)



Delineamento Experimental – Estrutura dos dados

Estatísticas e Distribuições Amostrais

Testes de Significância Estatística

Métodos Clássicos x Uso de Dados Externos x Permutações

Tomada de Decisão (resultante das análises estatística e factual)

Dados de um Experimento “Industrial”

(Box, Hunter and Hunter, 1978)

Unidade amostral	Tratamento	Resposta
1	A	89,7
2	A	81,4
3	A	84,5
4	A	84,8
5	A	87,3
6	A	79,7
7	A	85,1
8	A	81,7
9	A	83,7
10	A	84,5
11	B	84,7
12	B	86,1
13	B	83,2
14	B	91,9
15	B	86,3
16	B	79,3
17	B	82,6
18	B	89,1
19	B	83,7
20	B	88,5

Objetivo: Comparar dois métodos/tratamentos (A e B)

1. Discuta (exaustivamente) a estrutura dos dados!
2. Há evidência amostral de que o tratamento B é melhor que o A?
3. Que medida amostral você adotaria para investigar o efeito do tratamento?

Unidade amostral

Tratamento

resposta

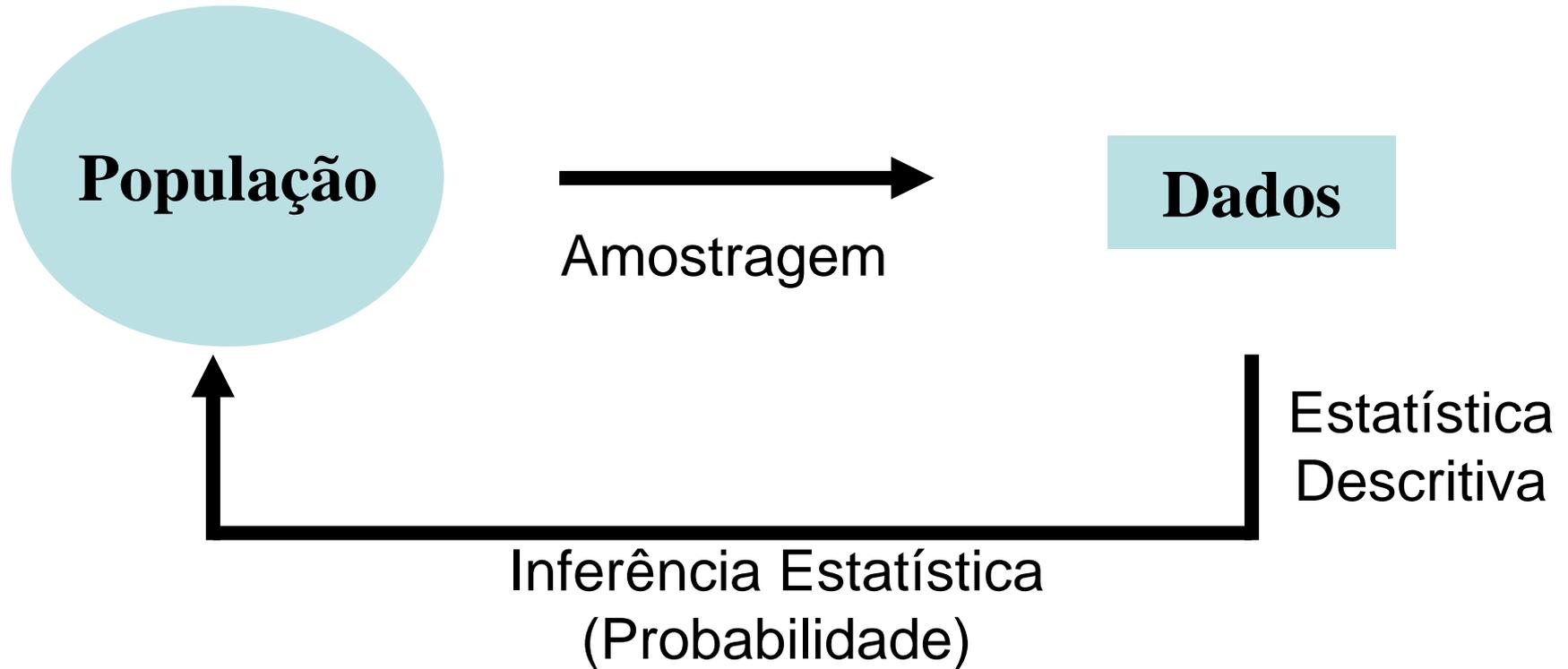
1	A	89,7
2	A	81,4
3	A	84,5
4	A	84,8
5	A	87,3
6	A	79,7
7	A	85,1
8	A	81,7
9	A	83,7
10	A	84,5
11	B	84,7
12	B	86,1
13	B	83,2
14	B	91,9
15	B	86,3
16	B	79,3
17	B	82,6
18	B	89,1
19	B	83,7
20	B	88,5

Estrutura dos Dados

- Tamanho amostral
- Variáveis Resposta (dependente)
- Variáveis Explicativas (independentes)
- Variáveis de Controle e Confundidores
- Unidades amostrais (População Alvo):
 - observações independentes?
- Aleatorização: atribuição aleatória de tratamentos às unidades experimentais
 - irrestrita (completamente aleatorizada)
 - restrita (uso de blocagem)
- ...

→ Box, Hunter and Hunter, 1978: unidade amostral corresponde à ordem no tempo em que as respostas foram coletadas (!!)

Raciocínio Estatístico (Clássico)



As pesquisas que estamos conduzindo
seguem esse racional?
E, necessariamente, precisam seguir?

Estatísticas Descritivas

Trat = A		Trat = B	
	estatísticas		estatísticas
tamanho amostral	10	tamanho amostral	10
mínimo	79.7000	mínimo	79.3000
quartil.25%	82.2000	quartil.25%	83.3250
mediana	84.5000	mediana	85.4000
média	84.2400	média	85.5400
quartil.75%	85.0250	quartil.75%	87.9500
máximo	89.7000	máximo	91.9000
variância	8.4204	variância	13.3249
desvio padrão	2.9018	desvio padrão	3.6503
erro padrão	0.9176	erro padrão	1.1543
coef. variação	3.4447	coef. variação	4.2674
curtose	-0.8722	curtose	-1.0249
assimetria	0.2425	assimetria	0.0841



Diferença entre Médias = 1,3

Estatísticas Descritivas

Método	A	B
n	10	10
Média	84,24	85,54
Dif	1,3	
Desvio Padrão	2,9018	3,650327
Erro Padrão	0,91763	1,154335

É uma diferença importante!

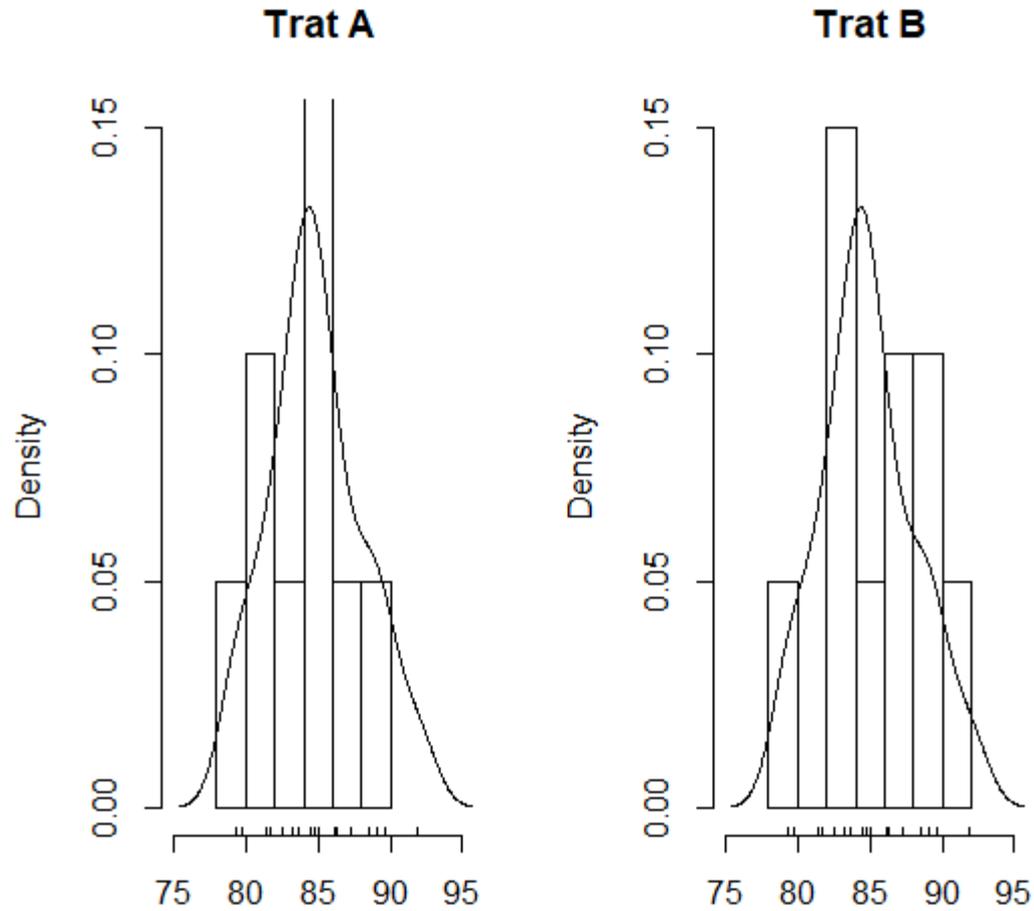
Interprete

- amostras balanceadas
- \bar{Y} , s^2 : estimadores relevantes ao problema sob estudo?
- $\bar{Y}_B - \bar{Y}_A = 1,3$ é uma evidência amostral informativa para a tomada de decisão? (método B é “melhor” que o A?)

Significância estatística e significância factual/clínica

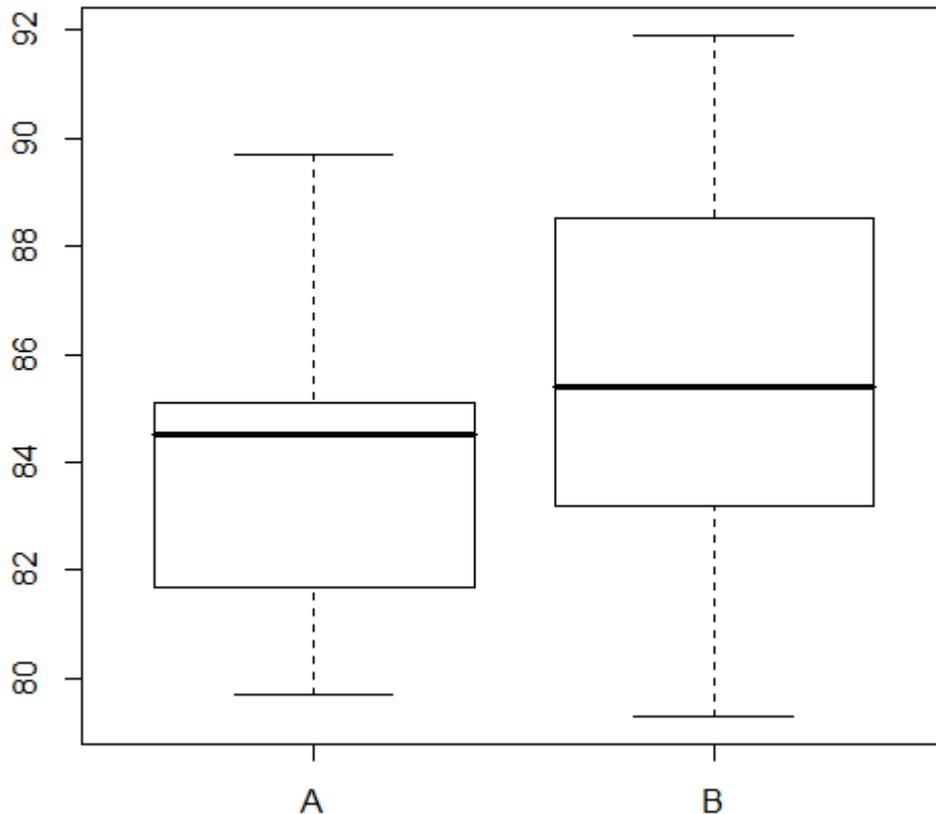
Como avaliar a significância de uma estimativa?

Histograma das Observações



Desenho Esquemático das Observações

Boxplot de acordo com Tratamento



Representação esquemática dos 5 pontos
Min, Q1, Q2, Q3, Max
⇒ e dos limites para diagnóstico de
observações atípicas

Outras medidas:

$$triMedia = \frac{1}{4} Q1 + \frac{1}{2} Q2 + \frac{1}{4} Q3$$

Observações atípicas (outliers)
ocorrem fora do intervalo

$$(Q1 - 1,5IQ ; Q3 + 1,5IQ)$$

$$IQ = Q3 - Q1$$

Revisar como
construir o boxplot!

Intervalo de Dispersão de Observações

Sob: $X_i \stackrel{iid}{\sim} N(\mu; \sigma^2)$

Intervalos usados (descritivamente) para diagnóstico de observações outliers.

$$P(X_i \leq Q1 - 1,5IQ) = 0,0034883 = P(X_i \geq Q3 + 1,5IQ)$$

$$P(Q1 - 1,5IQ \leq X_i \leq Q3 + 1,5IQ) = 0.9930234$$

$$P(\mu - \sigma \leq X_i \leq \mu + \sigma) = 1 - 2 * 0,1586553 = 0,6826894$$

$$P(\mu - 2\sigma \leq X_i \leq \mu + 2\sigma) = 1 - 2 * 0,02275013 = 0,9544997$$

$$P(\mu - 3\sigma \leq X_i \leq \mu + 3\sigma) = 1 - 2 * 0,001349898 = 0,9973002$$

...

$$P(\mu - 6\sigma \leq X_i \leq \mu + 6\sigma) = 1 - 2 * 9.865876e-10 \cong 1$$

Estrutura dos Dados

u.a.	Tratamento	Resposta
1	A	89,7
2	A	81,4
3	A	84,5
4	A	84,8
5	A	87,3
6	A	79,7
7	A	85,1
8	A	81,7
9	A	83,7
10	A	84,5
11	B	84,7
12	B	86,1
13	B	83,2
14	B	91,9
15	B	86,3
16	B	79,3
17	B	82,6
18	B	89,1
19	B	83,7
20	B	88,5

Esquema de geração dos dados:

- As unidades amostrais correspondem a observações ordenadas no tempo (funcionamento da máquina)

Suposição de independência NÃO está satisfeita!

Contudo:

- Suponha, inicialmente, que houve a atribuição aleatória dos tratamentos às unidades amostrais.

Como os dados podem ser analisados?

Como construir um teste válido de comparação entre medias nessas situações?

Em qual distribuição a diferença observada entre os tratamentos deve ser avaliada?

Comparação de Duas Populações

Em Box et al. (1978) a significância da diferença amostral encontrada entre os dois métodos, A e B, é avaliada nos seguintes casos:

1. Sob “premissas clássicas”:

Teste t de Student

Teste Não Paramétrico (Mann-Whitney)

2. Uso de distribuição externa que respeita a série temporal que gerou os dados:

Adotando um esquema de Amostras Adjacentes Sequenciais ou

Amostras Adjacentes Disjuntas



3. Vamos discutir também Testes de Aleatorização (sob esquemas de aleatorização dos tratamentos às unidades experimentais)

Testes Clássicos de Hipóteses Estatísticas

	Decisão	
	Rej H_0	N Rej H_0
H_0 V	<i>Erro Tipo I</i> (α)	<i>Nenhum Erro</i>
H_0 F	<i>Nenhum Erro</i>	<i>Erro Tipo II</i> (β)

$\alpha = P(\text{erro tipo I})$: nível de significância

$\beta = P(\text{erro tipo II})$

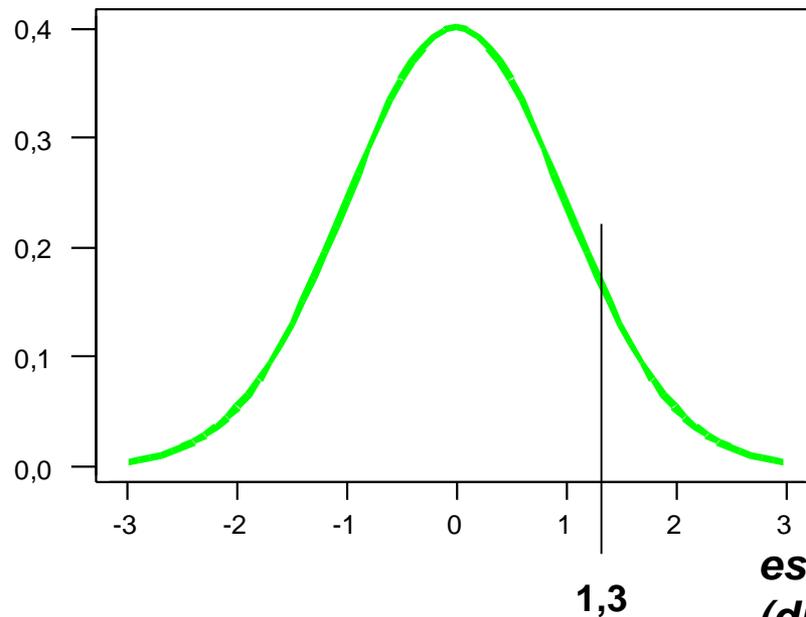
$1 - \beta$: poder do teste

valor - p: nível descritivo

Revise esses
conceitos!

Significância Estatística de um Resultado Amostral

Finalidade: Adotar uma “métrica” (uma distribuição) que permita tomar decisão sobre uma evidência amostral.



Sob premissas clássicas (normalidade, independência e homocedasticidade) e sob $H_0: \mu_1 = \mu_2$ poderíamos adotar a Distribuição “t” com (n_1+n_2-2) graus de liberdade! Mas, este NÃO é o caso dos dados do experimento industrial!

**estatística de teste
(diferença entre médias)**

⇒ Sob uma suposta distribuição (satisfazendo premissas clássicas e sob H_0), avaliar se o resultado amostral é raro ou muito provável.

Análise do Experimento Industrial

Formule as hipóteses.
Sob quais suposições o teste é válido?

Método	p
Testes Clássicos	
Teste "t" unicaudal	0,195
Mann Whitney	0,248

$$t = \frac{\bar{d}}{s_{\bar{d}}} = \frac{\bar{d}}{s_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1,3}{1,47} = 0,88 \sim t_{(n_1+n_2-2)}$$

$$IC\ 95\% (\mu_d = \mu_1 - \mu_2) = \left(\bar{d} \mp t_{(n_1+n_2-2);2,5\%} s_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

Compare os resultados e as suposições adotadas em cada caso.

Teste “t” – Experimentos Comparativos

$$H_0 : \mu_d = \mu_1 - \mu_2 = 0$$

Variâncias Homogêneas

Suposição: normalidade, observações independentes e variância homogênea
São válidas ao problema sob estudo?

$$\bar{d} = \bar{y}_1 - \bar{y}_2 \sim N\left(\mu_d, \sigma^2\left(1/n_1 + 1/n_2\right)\right)$$

$$t = \frac{\bar{d} - \mu_d}{s_{\bar{d}}} \sim t_{(n_1+n_2-2)}$$

$$s_{\bar{d}}^2 = s^2\left(1/n_1 + 1/n_2\right)$$

$$s^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

Análise Exploratória

Testes não-paramétricos são baseados nos postos.

A	B	A_ord	B_ord	A_pos	B_pos
89,7	84,7	79,7	79,3	19	11
81,4	86,1	81,4	82,6	3	14
84,5	83,2	81,7	83,2	9,5	6
84,8	91,9	83,7	83,7	12	20
87,3	86,3	84,5	84,7	16	15
79,7	79,3	84,5	86,1	2	1
85,1	82,6	84,8	86,3	13	5
81,7	89,1	85,1	88,5	4	18
83,7	83,7	87,3	89,1	7,5	7,5
84,5	88,5	89,7	91,9	9,5	17

Dados
originais



Dados
ordenados



Postos

⇒ Testes baseados nos dados transformados em postos são robustos (mas supõem observações independentes).

> t.test(yield ~ metodo)

Welch Two Sample t-test

data: yield by metodo

t = 0.8816, df = 17.129, p-value = 0.3902

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-1.809417 4.409417

sample estimates:

mean in group A mean in group B

85.54

84.24

Teste bicaudal (sob normalidade, independências e heterocedasticidade)

Qual teste foi realizado (H_0 e H_1)? Quais suposições foram adotadas?

Qual é a conclusão? Interprete o IC95%.

```
> t.test(yield ~ metodo, alternative=c("less"))
```

Welch Two Sample t-test

data: yield by metodo

t = -0.8816, df = 17.129, p-value = 0.1951

alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:

-Inf 1.264168

sample estimates:

mean in group A mean in group B

84.24

85.54

Teste unicaudal (sob
normalidade,
independências e
heterocedasticidade)

Qual teste foi realizado (H_0 e H_1)? Quais suposições foram adotadas?

Qual é a conclusão?

```
> var.test(yield ~ metodo)
```

F test to compare two variances

data: yield by metodo

F = 0.6319, num df = 9, denom df = 9, **p-value = 0.5049**

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.1569634 2.5441608

sample estimates:

ratio of variances

0.6319336

Qual teste foi realizado (H_0 e H_1)? Quais suposições foram adotadas?

Qual é a conclusão?

```
> t.test(yield ~ metodo, alternative=c("less"), var.equal=TRUE)
```

Two Sample t-test

data: yield by metodo

t = -0.8816, df = 18, p-value = 0.1948

alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:

-Inf 1.257102

sample estimates:

mean in group A mean in group B

84.24

85.54

Teste unicaudal (sob
normalidade,
independências e
homocedasticidade)

Qual teste foi realizado (H_0 e H_1)? Quais suposições foram adotadas?

Qual é a conclusão?

> wilcox.test(yield ~ metodo)

Wilcoxon rank sum test with continuity correction

data: yield by metodo

$W = 40.5$, p-value = 0.496

alternative hypothesis: true mu is not equal to 0

Warning message:

Cannot compute exact p-value with ties in: wilcox.test.default(x = c(89.7, 81.4, 84.5, 84.8, 87.3, 79.7,

Qual teste foi realizado (H_0 e H_1)? Quais suposições foram adotadas?

Qual é a conclusão?

```
> wilcox.test(yield ~ metodo, alternative="less")
```

Wilcoxon rank sum test with continuity correction

data: yield by metodo

W = 40.5, p-value = 0.248

alternative hypothesis: true mu is less than 0

Warning message:

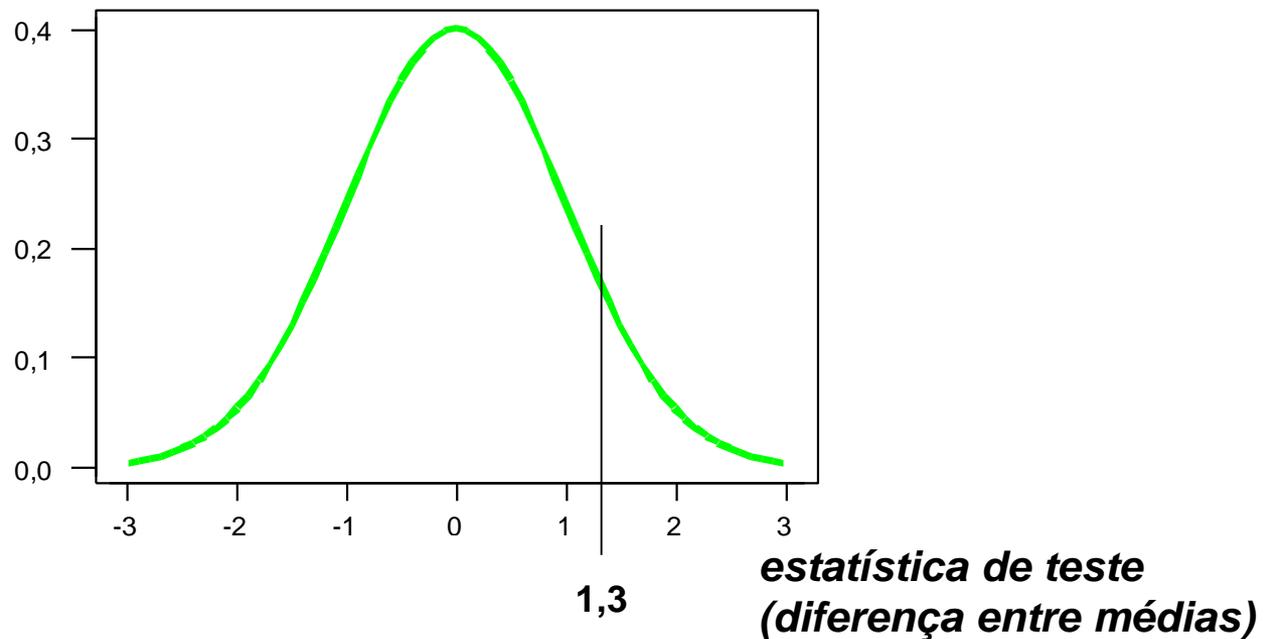
Cannot compute exact p-value with ties in: wilcox.test.default(x = c(89.7, 81.4, 84.5, 84.8, 87.3, 79.7,

Qual teste foi realizado (H_0 e H_1)? Quais suposições foram adotadas?

Qual é a conclusão?

Significância Estatística de um Resultado Amostral

Finalidade: Adotar uma “métrica” (uma distribuição) que permita tomar decisão sobre a evidência amostral ser rara ou muito provável sob H_0 .



⇒ MAS, para dados que não satisfazem as premissas clássicas, análises alternativas devem ser usadas, tais como: métodos não paramétricos, distribuições generalizadas, além de testes baseados em distribuições externas (referência) ou testes de permutação

Distribuição de Referência

Construção de uma distribuição que seja de referência e válida para o cálculo de Níveis Descritivos (avaliação da evidência amostral)

⇒ Estudaremos diferentes cenários que levam em conta a estrutura de série temporal dos dados!

85,5	84,5	80,5	79,5	84,8	81,1
81,7	82,4	86,1	86,7	86,6	85,6
80,6	86,7	82,6	80,5	83,5	86,6
84,7	83	85,4	91,7	78,1	80
88,2	81,8	84,7	81,6	88,8	86,6
84,9	89,3	82,8	83,9	81,9	83,3
81,8	79,3	81,9	85,6	83,3	83,1
84,9	82,7	83,6	84,8	80	82,3
85,2	88	86,8	78,4	87,2	86,7
81,9	79,6	84	89,9	83,3	80,2
89,4	87,8	84,2	85	86,6	
79	83,6	82,8	86,2	79,5	
81,4	79,5	83	83	84,1	
84,8	83,3	82	85,4	82,2	
85,9	88,4	84,7	84,4	90,8	
88	86,6	84,4	84,5	86,5	
80,3	84,6	88,9	86,2	79,7	
82,6	79,7	82,4	85,6	81	
83,5	86	83	83,2	87,2	
80,2	84,2	85	85,7	81,6	
85,2	83	82,2	83,5	84,4	
87,2	84,8	81,6	80,1	84,4	
83,5	83,6	86,2	82,2	82,2	
84,3	81,8	85,4	88,6	88,9	
82,9	85,9	82,1	82	80,9	
84,7	88,2	81,4	85	85,1	
82,9	83,5	85	85,2	87,1	
81,5	87,2	85,8	85,3	84	
83,4	83,7	84,2	84,3	76,5	
87,7	87,3	83,5	82,3	82,7	
81,8	83	86,5	89,7	85,1	
79,6	90,5	85	84,8	83,3	
85,8	80,7	80,4	83,1	90,4	
77,9	83,1	85,7	80,6	81	
89,7	86,5	86,7	87,4	80,3	
85,4	90	86,7	86,8	79,8	
86,3	77,5	82,3	83,5	89	
80,7	84,7	86,4	86,2	83,7	
83,8	84,6	82,5	84,1	80,9	
90,5	87,2	82	82,3	87,3	

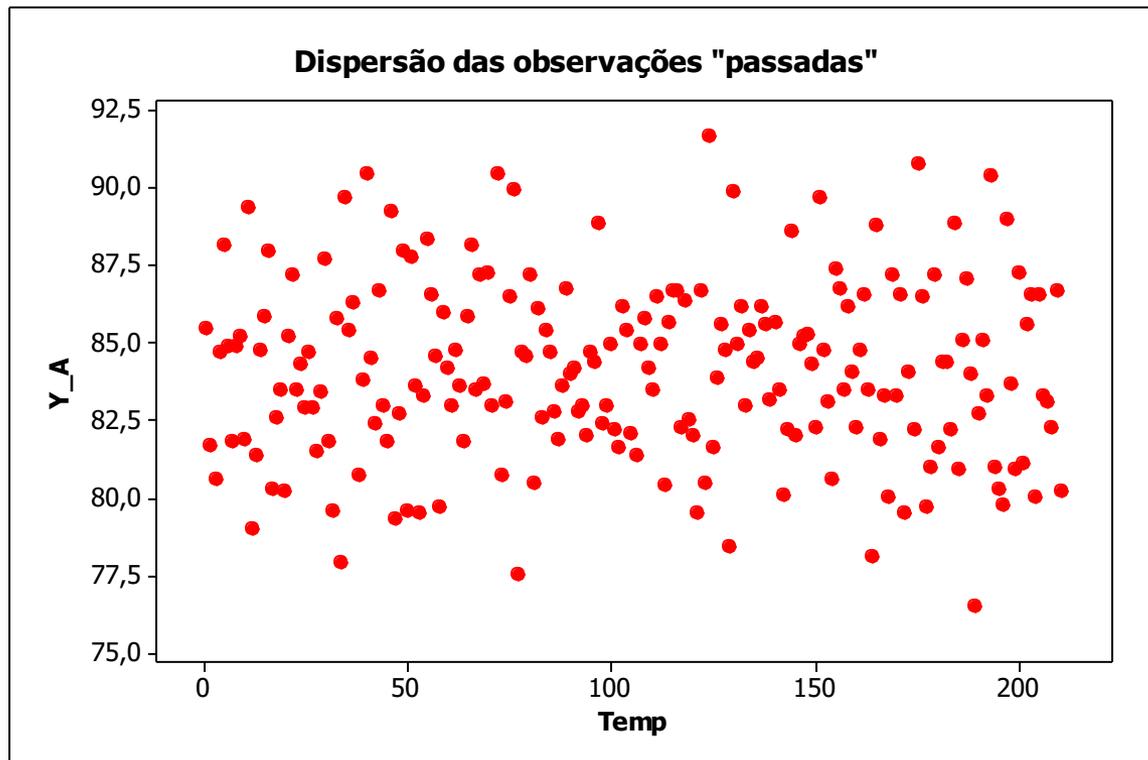
Dados Externos

Informação adicional oferecida pelo pesquisador: Série passada de 210 observações consecutivas das respostas de produção sob o Método A.

Como esta informação adicional pode ser utilizada, apropriadamente, na tomada de decisão sobre se a produção sob o Método B é maior que aquela sob o Método A?

Dados Externos

Série Passada

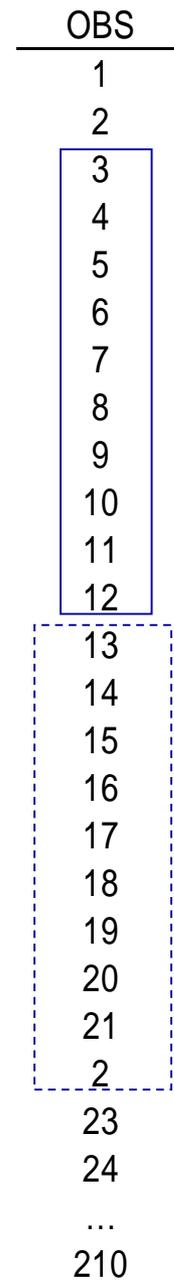
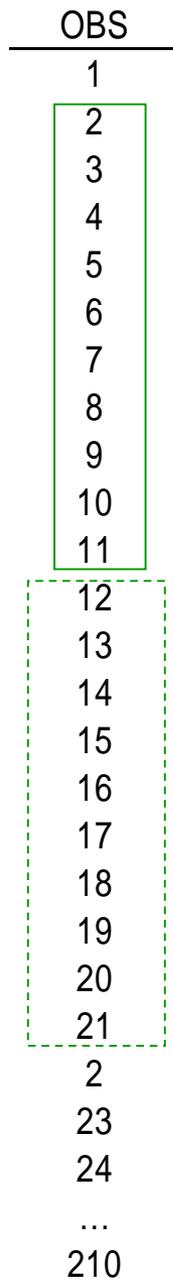
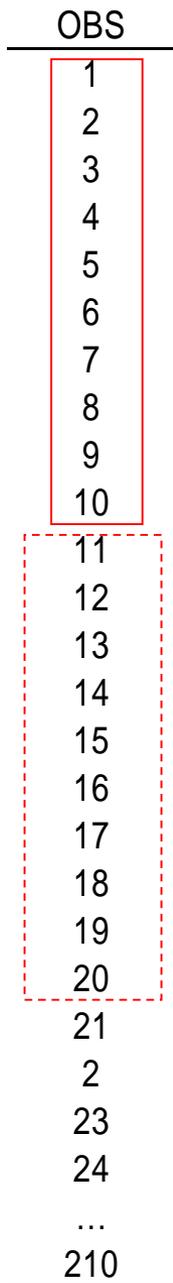


Construção de uma Distribuição de Referência

Cálculo de Níveis Descritivos

- Considere o conjunto de 210 observações do Método A
- Particione (sequencialmente) as observações em conjuntos adjacentes de tamanho 10
- Calcule as diferenças entre as médias dos conjuntos adjacentes para todas as possíveis sequências
- Obtenha assim um conjunto de 191 diferenças, as quais são possíveis de ocorrer sob o método A
- Verifique se a diferença efetivamente observada, 1,3, é um valor pouco ou muito provável de ocorrer sob essa distribuição

⇒ veja esquema a seguir!



Obter uma lista de referência de 191 diferenças entre médias de conjuntos adjacentes de 10 obs.

...

Obs	Média										
85,5		84,5	-84,42	80,5	-84,53	79,5	-83,72	84,8	-84,36	81,1	-83,68
81,7		82,4	-84,7	86,1	-84,09	86,7	-83,89	86,6	-84,54	85,6	-83,91
80,6		86,7	-84,79	82,6	-84,28	80,5	-83,9	83,5	-84,58	86,6	-83,53
84,7		83	-85,3	85,4	-84,51	91,7	-84,5	78,1	-84,33	80	-83,43
88,2		81,8	-84,51	84,7	-84,33	81,6	-83,99	88,8	-84,47	86,6	-84,06
84,9		89,3	-84,9	82,8	-83,61	83,9	-83,71	81,9	-83,98	83,3	-84,41
81,8		79,3	-84,2	81,9	-84,05	85,6	-84,04	83,3	-83,96	83,1	-83,82
84,9		82,7	-84,4	83,6	-83,94	84,8	-83,88	80	-83,34	82,3	-83,68
85,2		88	-84,82	86,8	-84,16	78,4	-83,47	87,2	-83,65	86,7	-84,26
81,9	-83,94	79,6	-83,73	84	-83,84	89,9	-84,26	83,3	-83,75	80,2	-83,55
89,4	-84,33	87,8	-84,06	84,2	-84,21	85	-84,81	86,6	-83,93		
79	-84,06	83,6	-84,18	82,8	-83,88	86,2	-84,76	79,5	-83,22		
81,4	-84,14	79,5	-83,46	83	-83,92	83	-85,01	84,1	-83,28		
84,8	-84,15	83,3	-83,49	82	-83,58	85,4	-84,38	82,2	-83,69		
85,9	-83,92	88,4	-84,15	84,7	-83,58	84,4	-84,66	90,8	-83,89		
88	-84,23	86,6	-83,88	84,4	-83,74	84,5	-84,72	86,5	-84,35		
80,3	-84,08	84,6	-84,41	88,9	-84,44	86,2	-84,78	79,7	-83,99		
82,6	-83,85	79,7	-84,11	82,4	-84,32	85,6	-84,86	81	-84,09		
83,5	-83,68	86	-83,91	83	-83,94	83,2	-85,34	87,2	-84,09		
80,2	-83,51	84,2	-84,37	85	-84,04	85,7	-84,92	81,6	-83,92		
85,2	-83,09	83	-83,89	82,2	-83,84	83,5	-84,77	84,4	-83,7		
87,2	-83,91	84,8	-84,01	81,6	-83,72	80,1	-84,16	84,4	-84,19		
83,5	-84,12	83,6	-84,42	86,2	-84,04	82,2	-84,08	82,2	-84		
84,3	-84,07	81,8	-84,27	85,4	-84,38	88,6	-84,4	88,9	-84,67		
82,9	-83,77	85,9	-84,02	82,1	-84,12	82	-84,16	80,9	-83,68		
84,7	-83,44	88,2	-84,18	81,4	-83,82	85	-84,21	85,1	-83,54		
82,9	-83,7	83,5	-84,07	85	-83,43	85,2	-84,11	87,1	-84,28		
81,5	-83,59	87,2	-84,82	85,8	-83,77	85,3	-84,08	84	-84,58		
83,4	-83,58	83,7	-84,59	84,2	-83,89	84,3	-84,19	76,5	-83,51		
87,7	-84,33	87,3	-84,9	83,5	-83,74	82,3	-83,85	82,7	-83,62		
81,8	-83,99	83	-84,9	86,5	-84,17	89,7	-84,47	85,1	-83,69		
79,6	-83,23	90,5	-85,47	85	-84,51	84,8	-84,94	83,3	-83,58		
85,8	-83,46	80,7	-85,18	80,4	-83,93	83,1	-85,03	90,4	-84,4		
77,9	-82,82	83,1	-85,31	85,7	-83,96	80,6	-84,23	81	-83,61		
89,7	-83,5	86,5	-85,37	86,7	-84,42	87,4	-84,77	80,3	-83,55		
85,4	-83,57	90	-85,55	86,7	-84,95	86,8	-84,95	79,8	-83,02		
86,3	-83,91	77,5	-84,95	82,3	-84,68	83,5	-84,78	89	-83,21		
80,7	-83,83	84,7	-84,7	86,4	-84,74	86,2	-84,87	83,7	-83,18		
83,8	-83,87	84,6	-84,79	82,5	-84,57	84,1	-84,85	80,9	-83,62		
90,5	-84,15	87,2	-84,78	82	-84,42	82,3	-84,85	87,3	-84,08		

Cálculo das médias de conjuntos de 10 observações consecutivas.

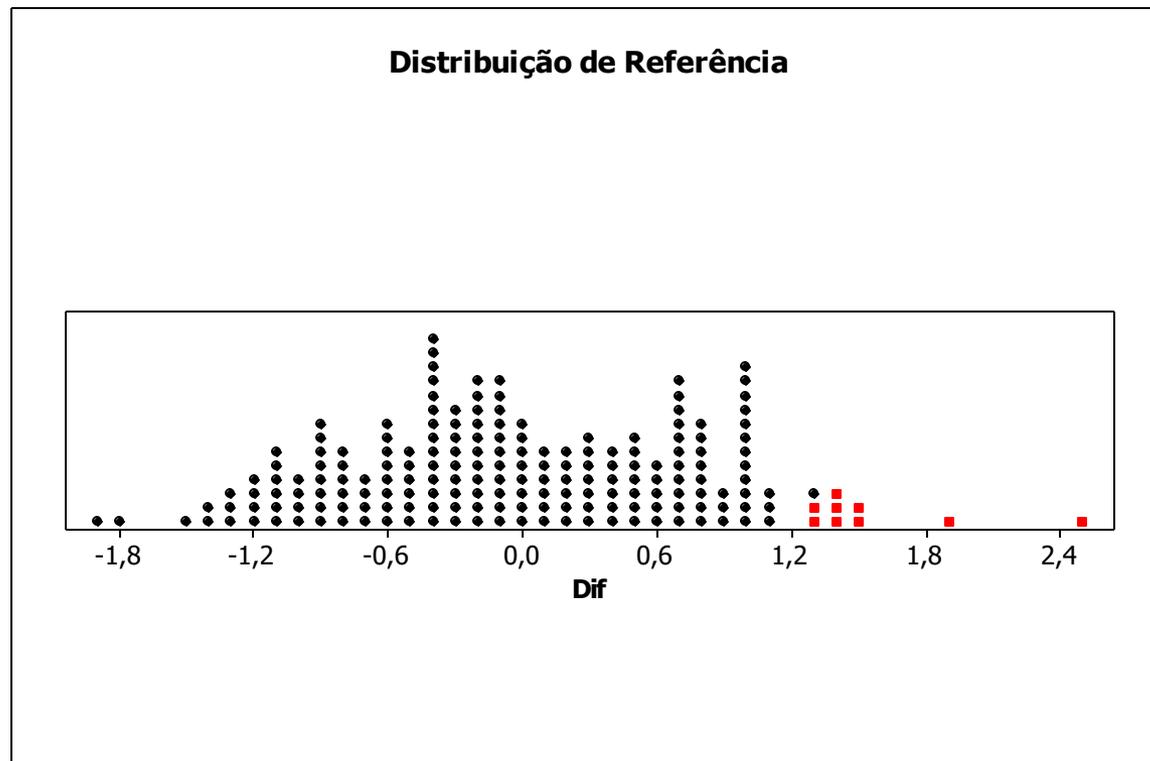
-	-0,36	-0,32	1,09	-0,43
-	-0,52	-0,21	0,87	-1,32
-	-1,33	-0,36	1,11	-1,3
-	-1,81	-0,93	-0,12	-0,64
-	-0,36	-0,75	0,67	-0,58
-	-1,02	0,13	1,01	0,37
-	0,21	0,39	0,74	0,03
-	-0,29	0,38	0,98	0,75
-	-0,91	-0,22	1,87	0,44
-0,43	0,64	0,2	0,66	0,17
-1,24	-0,17	-0,37	-0,04	-0,23
-0,15	-0,17	-0,16	-0,6	0,97
-0,02	0,96	0,12	-0,93	0,72
-0,08	0,78	0,8	0,02	0,98
-0,15	-0,13	0,54	-0,5	-0,21
-0,79	0,3	0,08	-0,51	-0,81
-0,38	-0,34	-1,01	-0,67	0,29
-0,26	0,71	-0,55	-0,78	0,49
-0,1	0,68	-0,05	-1,15	-0,58
0,82	0,53	-0,3	-1,07	-0,3
0,9	1,01	0,33	-0,3	-0,01
-0,68	1,46	0,79	0,78	-0,61
-0,66	0,76	-0,11	0,95	0,4
-1,25	1,04	-0,42	-0,17	-1,06
-0,27	1,35	0,3	0,61	-0,13
0,13	1,37	1,31	0,74	-0,52
0,21	0,88	1,25	0,67	-1,07
0,24	-0,12	0,97	0,79	-1,4
0,29	0,2	0,68	0,66	0,11
-0,18	-0,12	0,68	1	0,46
0,43	-0,37	-0,45	-0,11	-0,01
1,47	-1,38	-0,62	-0,4	0,33
1,33	-0,9	-0,03	-0,45	-0,87
2,48	-0,8	0,54	0,1	-0,18
1,01	-1,04	-0,43	-0,3	0,51
1,33	-1,94	-1,24	-0,97	1,39
0,29	-0,9	-0,64	-0,82	0,61
0,57	-0,76	-0,86	-1,53	0,5
0,95	-0,63	-1,1	-1,2	0,64
-0,42	-0,94	-0,16	-1,1	-0,53

Lista de referência de 191 diferenças entre médias de conjuntos adjacentes de 10 observações

⇒ Construir uma distribuição de referência para avaliação da significância da diferença efetivamente observada (=1,3).

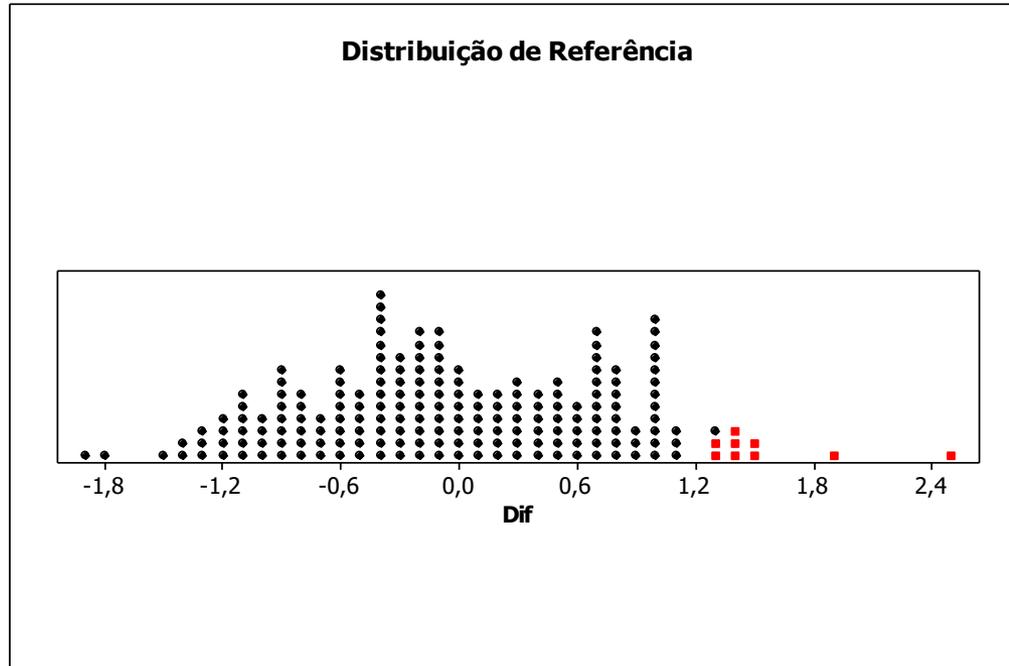
Distribuição de Referência

Diagrama de Pontos das 191 diferenças entre médias de conjuntos adjacentes de 10 observações. Indicação dos pontos maiores ou iguais ao valor da diferença observada de 1,3.



Qual é o nível de significância do valor 1,3?

Distribuição de Referência



A diferença observada entre as médias das respostas de produção dos dois Métodos é significativa ao nível de $9/191=0,047=4,7\%$

⇒ há evidência amostral para concluir que a média de produção via o Método B é significativamente maior que aquela via o Método A (a um nível de significância de 5%).

Análise do Experimento Industrial

Método	p	
Testes Clássicos		
Teste "t"	0,195	→ $t = \frac{\bar{d}}{s_{\bar{d}}} = \frac{\bar{d}}{s_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1,3}{1,47} = 0,88 \sim t_{(n_1+n_2-2)}$
Mann Whitney	0,248	
Distr. Referência		
Conjuntos Adjacentes	0,047	→ $\bar{d} = 1,3 \sim \text{distr. referência}$

Compare os resultados e as suposições adotadas em cada caso.

Experimentos Comparativos

Observações Correlacionadas

$$\bar{Y} = \frac{y_1 + \dots + y_n}{n} = \alpha_1 y_1 + \dots + \alpha_n y_n \quad \alpha_j = 1/n$$

$$E(\bar{Y}) = \sum_{j=1}^n \alpha_j \mu_j \quad V(\bar{Y}) = \sum_{j=1}^n \alpha_j^2 \sigma_j^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n \alpha_i \alpha_j \sigma_i \sigma_j \rho_{ij}$$

$$\Rightarrow V(\bar{Y}) = \frac{\sigma^2}{n} \left[1 + \frac{2(n-1)}{n} \rho_1 \right] \quad \begin{array}{l} \text{modelo de autocorrelação} \\ \text{de ordem 1} \end{array}$$

Testes “t” (sob premissas clássicas) não são válidos neste caso
 \Rightarrow este teste não é robusto para fugas da suposição de independência entre as observações.

Experimentos Comparativos

Observações Correlacionadas

(Y_1, Y_2, \dots, Y_n) : amostra de n observações da população P

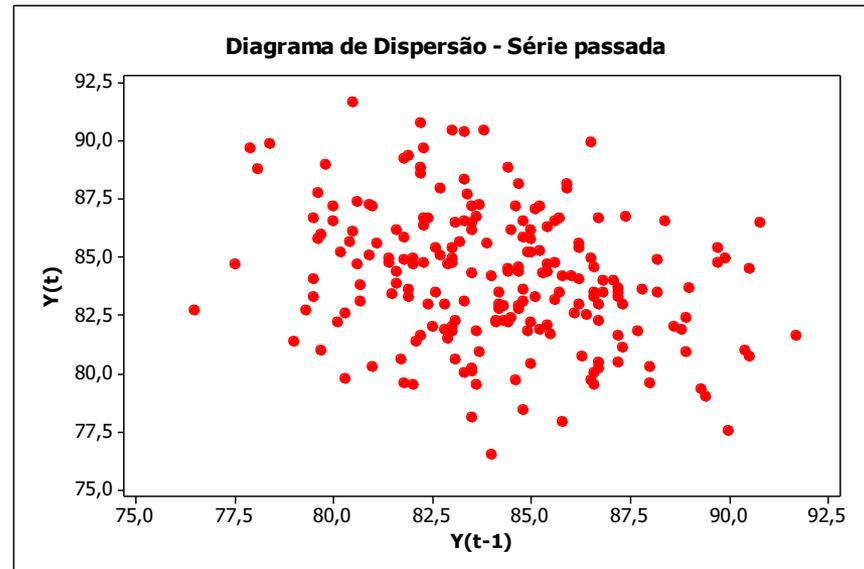
$$\Rightarrow V(\bar{Y})^{iid} = \frac{\sigma^2}{n}$$

$$\Rightarrow V(\bar{Y})^{autocorr_1} = \frac{\sigma^2}{n} \left[1 + \frac{2(n-1)}{n} \rho_1 \right]$$

$$c = \sqrt{1 + \frac{2(n-1)}{n} \rho_1}$$

Fator de correção no erro padrão devido à estrutura de autocorrelação de ordem 1

Análise do Experimento Industrial



$$\rho_1 = -0,29 \Rightarrow c = \sqrt{1 + \frac{2(n-1)}{n} \rho_1} = \sqrt{1 + \frac{2(10-1)}{10} (-0,29)} = \sqrt{0,48} = 0,70$$

Ao incorporar a estrutura de autocorrelação de ordem 1 na análise (via os dados da distribuição externa) houve uma redução no erro padrão de 30% \Rightarrow ganho em precisão \Rightarrow significância do valor amostral 1,3 passou de $p=0,195$ para 0,047

Construção de uma Distribuição de Referência Externa Baseada na Distribuição “t”

⇒ Uma outra alternativa de uso dos dados da série passada de observações

OBS	
1	
2	
...	
10	
11	
12	
...	
20	
21	→
22	
23	
...	
32	
33	
34	
...	
42	
43	→
44	
...	
54	
55	
...	
64	
65	→
...	
66	
...	
210	

Uso dos Dados Externos para Cálculo do Fator de Escala da Estatística t

- Obter k=10 diferenças entre médias de conjuntos adjacentes de 10 observações para todas as possíveis sequências “disjuntas” de 20 observações
- Pule uma observação para tornar os conjuntos aproximadamente independentes
- Usar estes valores de referência para calcular o desvio padrão das diferenças entre médias
- Adotar o valor obtido como o fator de escala na estatística a seguir:

$$t = \frac{\bar{d}}{s_{\bar{d}}} \sim t_{(k-1)}$$

fator de escala

$$\bar{d} = \bar{Y}_B - \bar{Y}_A$$

$$s_{\bar{d}}^2 = \frac{1}{k-1} \sum_{j=1}^k (\bar{d}_j - \mu_d)^2$$

Obs	Média										
85,5		84,5	-84,42	80,5	-84,53	79,5	-83,72	84,8	-84,36	81,1	-83,68
81,7		82,4	-84,7	86,1	-84,09	86,7	-83,89	86,6	-84,54	85,6	-83,91
80,6		86,7	-84,79	82,6	-84,28	80,5	-83,9	83,5	-84,58	86,6	-83,53
84,7		83	-85,3	85,4	-84,51	91,7	-84,5	78,1	-84,33	80	-83,43
88,2		81,8	-84,51	84,7	-84,33	81,6	-83,99	88,8	-84,47	86,6	-84,06
84,9		89,3	-84,9	82,8	-83,61	83,9	-83,71	81,9	-83,98	83,3	-84,41
81,8		79,3	-84,2	81,9	-84,05	85,6	-84,04	83,3	-83,96	83,1	-83,82
84,9		82,7	-84,4	83,6	-83,94	84,8	-83,88	80	-83,34	82,3	-83,68
85,2		88	-84,82	86,8	-84,16	78,4	-83,47	87,2	-83,65	86,7	-84,26
81,9	-83,94	79,6	-83,73	84	-83,84	89,9	-84,26	83,3	-83,75	80,2	-83,55
89,4	-84,33	87,8	-84,06	84,2	-84,21	85	-84,81	86,6	-83,93		
79	-84,06	83,6	-84,18	82,8	-83,88	86,2	-84,76	79,5	-83,22		
81,4	-84,14	79,5	-83,46	83	-83,92	83	-85,01	84,1	-83,28		
84,8	-84,15	83,3	-83,49	82	-83,58	85,4	-84,38	82,2	-83,69		
85,9	-83,92	88,4	-84,15	84,7	-83,58	84,4	-84,66	90,8	-83,89		
88	-84,23	86,6	-83,88	84,4	-83,74	84,5	-84,72	86,5	-84,35		
80,3	-84,08	84,6	-84,41	88,9	-84,44	86,2	-84,78	79,7	-83,99		
82,6	-83,85	79,7	-84,11	82,4	-84,32	85,6	-84,86	81	-84,09		
83,5	-83,68	86	-83,91	83	-83,94	83,2	-85,34	87,2	-84,09		
80,2	-83,51	84,2	-84,37	85	-84,04	85,7	-84,92	81,6	-83,92		
85,2	-83,09	83	-83,89	82,2	-83,84	83,5	-84,77	84,4	-83,7		
87,2	-83,91	84,8	-84,01	81,6	-83,72	80,1	-84,16	84,4	-84,19		
83,5	-84,12	83,6	-84,42	86,2	-84,04	82,2	-84,08	82,2	-84		
84,3	-84,07	81,8	-84,27	85,4	-84,38	88,6	-84,4	88,9	-84,67		
82,9	-83,77	85,9	-84,02	82,1	-84,12	82	-84,16	80,9	-83,68		
84,7	-83,44	88,2	-84,18	81,4	-83,82	85	-84,21	85,1	-83,54		
82,9	-83,7	83,5	-84,07	85	-83,43	85,2	-84,11	87,1	-84,28		
81,5	-83,59	87,2	-84,82	85,8	-83,77	85,3	-84,08	84	-84,58		
83,4	-83,58	83,7	-84,59	84,2	-83,89	84,3	-84,19	76,5	-83,51		
87,7	-84,33	87,3	-84,9	83,5	-83,74	82,3	-83,85	82,7	-83,62		
81,8	-83,99	83	-84,9	86,5	-84,17	89,7	-84,47	85,1	-83,69		
79,6	-83,23	90,5	-85,47	85	-84,51	84,8	-84,94	83,3	-83,58		
85,8	-83,46	80,7	-85,18	80,4	-83,93	83,1	-85,03	90,4	-84,4		
77,9	-82,82	83,1	-85,31	85,7	-83,96	80,6	-84,23	81	-83,61		
89,7	-83,5	86,5	-85,37	86,7	-84,42	87,4	-84,77	80,3	-83,55		
85,4	-83,57	90	-85,55	86,7	-84,95	86,8	-84,95	79,8	-83,02		
86,3	-83,91	77,5	-84,95	82,3	-84,68	83,5	-84,78	89	-83,21		
80,7	-83,83	84,7	-84,7	86,4	-84,74	86,2	-84,87	83,7	-83,18		
83,8	-83,87	84,6	-84,79	82,5	-84,57	84,1	-84,85	80,9	-83,62		
90,5	-84,15	87,2	-84,78	82	-84,42	82,3	-84,85	87,3	-84,08		

Cálculo das 10 diferenças entre médias de conjuntos “independentes” de observações consecutivas.

Análise do Experimento Industrial

	Média_A	Média_B	Dif(B-A)	
	83,94	83,51	-0,43	$\Rightarrow s_{\bar{d}}^2$
	83,99	84,42	0,43	
	84,18	84,01	-0,17	
	85,18	84,28	-0,9	
	83,58	84,38	0,8	
	84,42	83,99	-0,43	
	84,72	84,21	-0,51	
	84,78	83,96	-0,82	
	84,09	84,58	0,49	
	83,62	84,26	0,64	
Experimento	84,24	85,54	1,3	$\Rightarrow \bar{d} = \bar{Y}_B - \bar{Y}_A$

$$s_{\bar{d}}^2 = \frac{(-0,43 - 0)^2 + (0,43 - 0)^2 + \dots + (0,64 - 0)^2}{10 - 1} = 0,36$$

$$t = \frac{\bar{d}}{s_{\bar{d}}} = \frac{1,3}{\sqrt{0,36}} = 2,17 \sim t_{(9)} \Rightarrow p = 0,029$$

Análise do Experimento Industrial

Box, Hunter and Hunter, 1978.

Método	p
Testes Clássicos	
Teste "t"	0,195
Mann Whitney	0,248
Distr. Referência	
Conjuntos Adjacentes	0,047
Conj. Adj. Disjuntos	0,029

Um componente da tomada de decisão na pesquisa científica é a significância estatística. Esta deve ser formulada respeitando a estrutura dos dados!

Pontos de Discussão:

- Utilidade e limitação de cada metodologia de teste
- Outras alternativas

Testes de Aleatorização ou Permutação

Gary Oehlert (2010) – Capítulo 2

A	B
89,7	84,7
81,4	86,1
84,5	83,2
84,8	91,9
87,3	86,3
79,7	79,3
85,1	82,6
81,7	89,1
83,7	83,7
84,5	88,5

MAS, como construir um teste de aleatorização sob o esquema de observações correlacionadas serialmente?

Suponha agora que os tratamentos A e B foram aleatorizados às n=20 observações:

Sob um Delineamento Completamente Aleatorizado (DCA) e sob a hipótese H_0 (inexistência de efeito de tratamento) é possível construir uma distribuição de referência para a estatística observada “t” com base nos possíveis resultados amostrais que poderiam ter ocorrido: da lista de 20 observações amostrais selecionar aleatoriamente 10 para comporem A e as restantes para B. Neste caso, há C_{10}^{20} possibilidades.

Calcule a estatística t para cada possível amostra gerada. Com esses valores construa a distribuição de referência e calcule o p-valor para a estatística t efetivamente observada.

Como mostrado em Oehlert (2010) este valor-p é próximo àquele calculado sob premissas clássicas. Assim a aleatorização além de proteger contra vícios experimentais permite a obtenção de níveis descritivos válidos mesmo sem a hipótese de normalidade.

Auditoria

Desconfie e seja crítico com:

- Procedimentos adotados na construção de um teste de significância estatística

Amostragem aleatória

Independência entre observações

Tipo da variável de interesse

Forma da distribuição dos dados

Homocedasticidade, etc.

Validação e Análises de Diagnóstico

Auditoria

Desconfie e seja crítico com:

- Resultados que não podem ser explicados por fatos
- Resultados que destoam de resultados de outras pesquisas
- Variações muito grandes entre períodos amostrais curtos
- Entenda significância estatística e significância biológica
- ...

Referência Bibliográfica

- Box, G.E.; Hunter, W.G. and Hunter, J.S. (1978). Statistics for Experimenters. An Introduction to Designs, Data Analysis and Model Building. John Wiley & Sons.
- Conover, W.J. (1999). Practical Nonparametric Statistics, 3.ed. John Wiley & Sons, Inc. 584p.
- Oehlert, GW. A First Course in Design and Analysis of Experiments. Univ. of Minnesota, Licensed by Creative Commons, 2010.
- Tukey, J.W. (1977). Exploratory Data Analysis. Addison-Wesley.

Método Estatístico (Clássico)

**Análise
Descritiva**

**Teoria de
Probabilidade**



**Análise
Inferencial**



- **O Experimento foi Realizado!**

“análise exploratória dos dados”



- **Se o Experimento for Realizado!**

“que dados são gerados sob diferentes modelos probabilísticos”



- **Qual a probabilidade dos dados terem sido gerados sob um específico modelo ?**

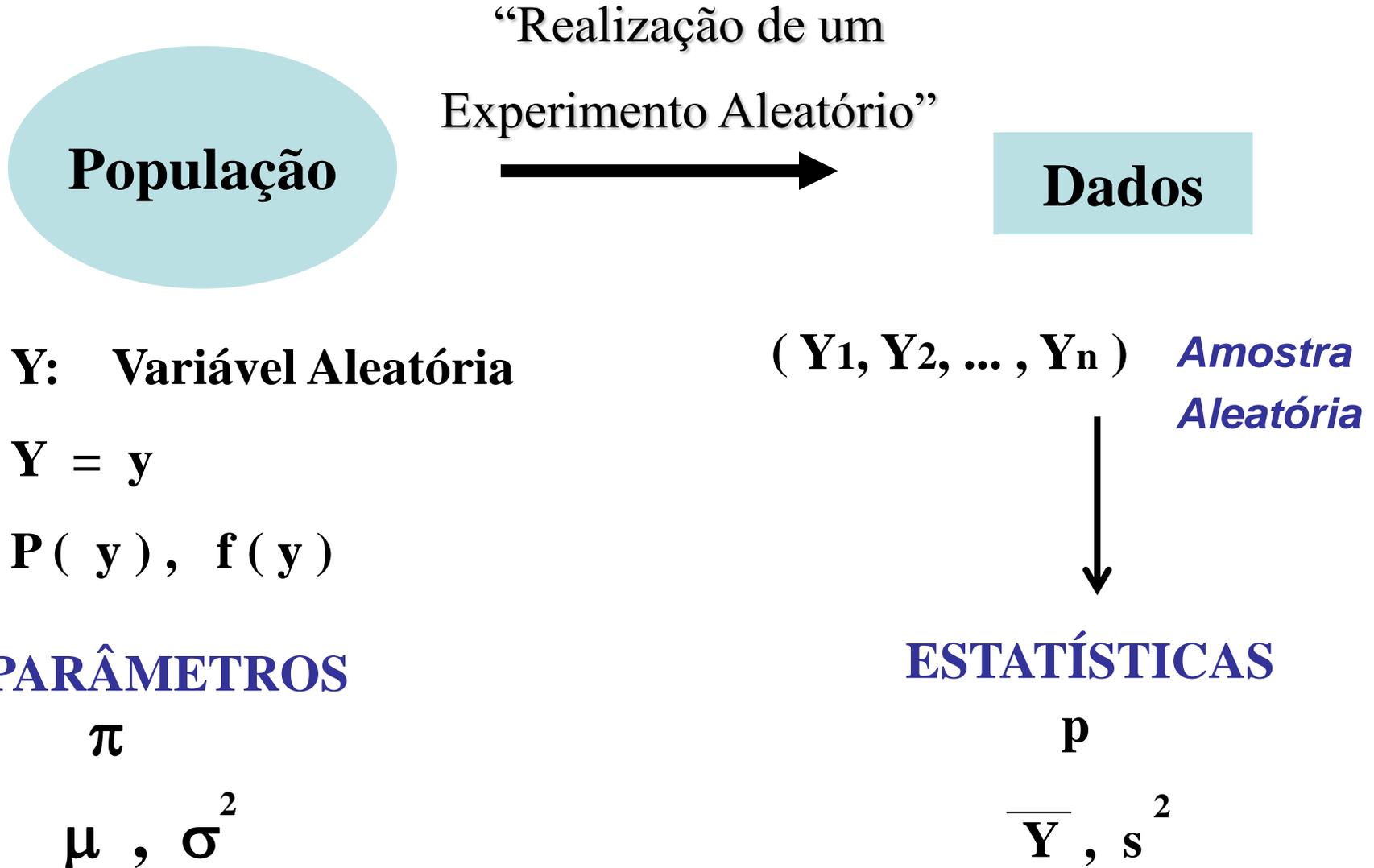


**Testes
Estimação**



Segundo Tukey, com base nos resultados uma nova análise descritiva deve ser realizada

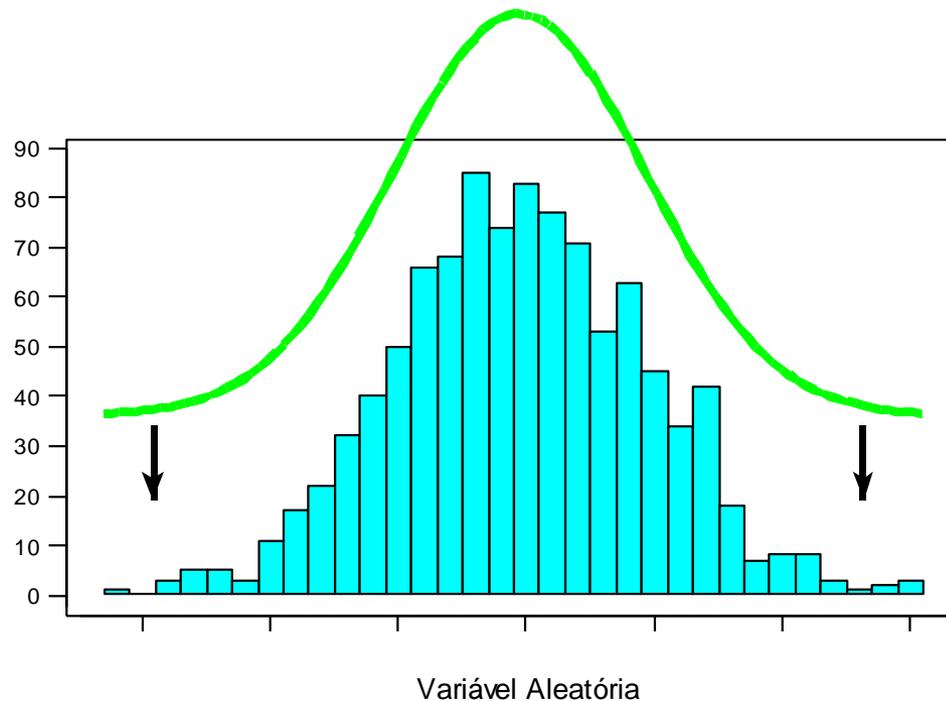
Formulação Estatística (Clássica)



Distribuição de Probabilidades

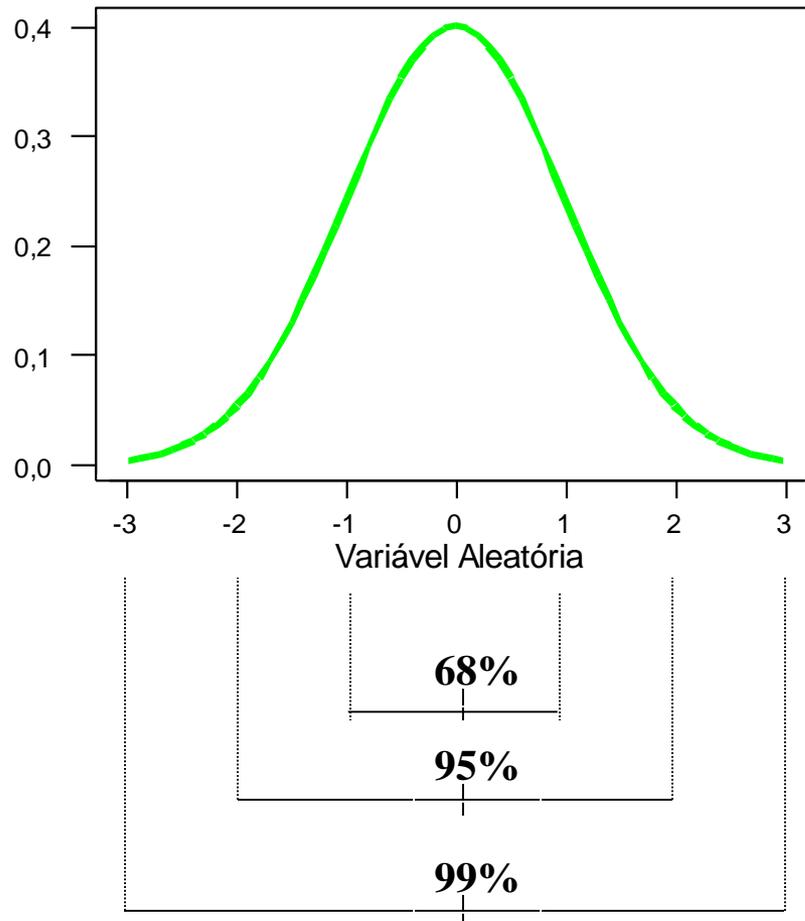
População hipotética de resultados representados por uma distribuição

Modelo teórico adotado para análise de situações empíricas sujeitas a incertezas



Distribuição Normal

$$Y \sim N(\mu, \sigma^2)$$



- Contínua $[-\infty, +\infty]$
- Simétrica
- Forma de sino
- Caracterizada por dois parâmetros:
 - Média (μ)
 - Desvio Padrão (σ)
- Normal Padrão
 - $\mu = 0$ $\sigma = 1$
- Maior entropia
- Efeito Limite Central
- Robustez

Teorema Limite Central

Y: Variável Aleatória

Momentos Finitos: $E(Y) = \mu$ $V(Y) = \sigma^2$

(Y_1, Y_2, \dots, Y_n) é amostra aleatória $\longrightarrow \bar{Y}$

$$\bar{Y} \stackrel{n \rightarrow \infty}{\sim} N(\mu ; \sigma^2 / n)$$

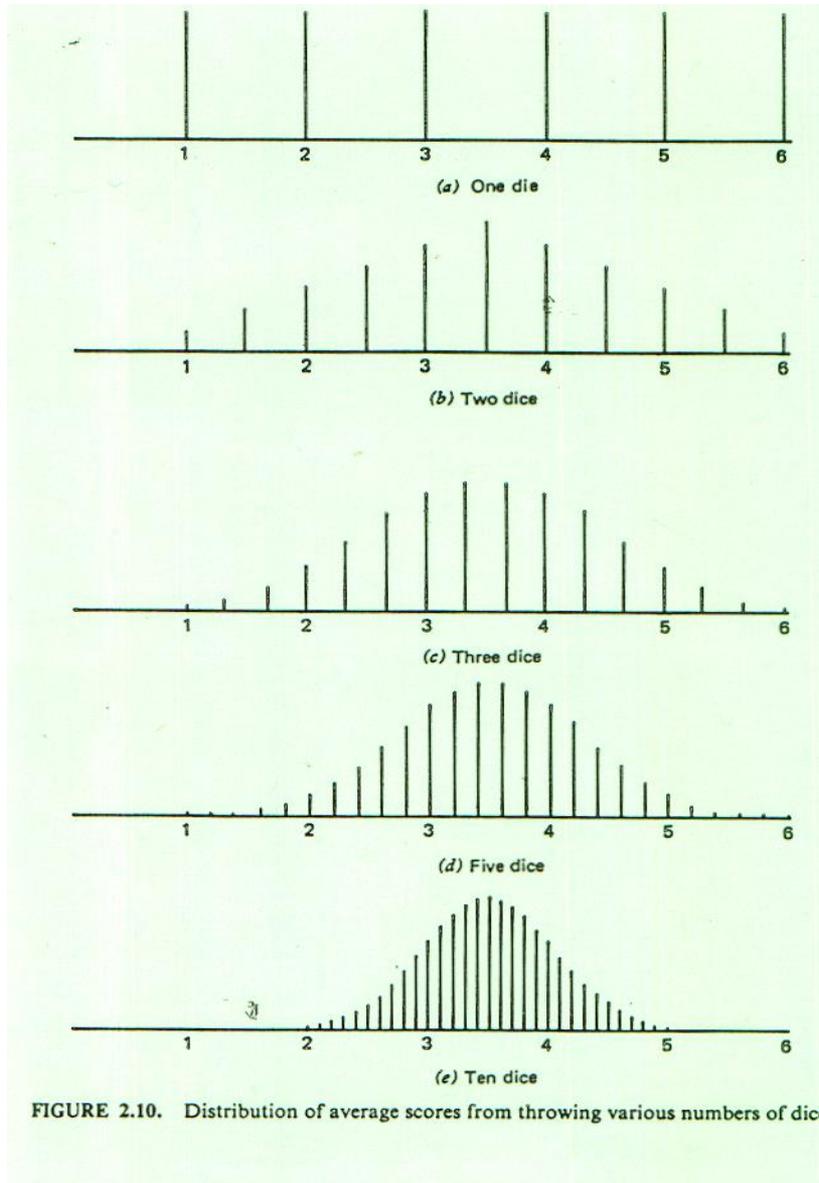
Distribuição amostral
da Média



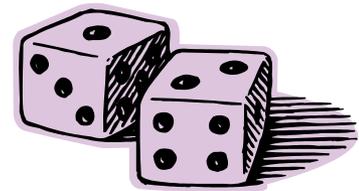
Precisão das Estatísticas

Robustez (não vale para s^2)

Distribuição Amostral da Média



Considere que n dados são lançados e que é calculada a média das faces obtidas.



Distribuição Amostral da Média

$$\bar{Y} = \bar{Y}_n \sim N(\mu ; \sigma^2 / n)$$

$$Z = \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}} \sim N(0 ; 1)$$

$$t = \frac{\bar{Y} - \mu}{s / \sqrt{n}} \sim t_{(n-1)}$$



$$(\bar{Y} - z * \sigma / \sqrt{n} ; \bar{Y} + z * \sigma / \sqrt{n}) \quad (\bar{Y} - t * s / \sqrt{n} ; \bar{Y} + t * s / \sqrt{n})$$

Distribuição t de Student

$$t = \frac{Y - \mu}{s} \sim t_{(v)}$$

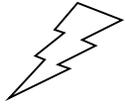
- $Y \sim N(\mu, \sigma^2)$
- $s \perp Y$
- s^2 tem v graus de liberdade e é calculada de observações *iid* normal com variância σ^2

$$(Y_1, Y_2, \dots, Y_n) \rightarrow \bar{Y} \Rightarrow t = \frac{\bar{Y} - \mu}{s/\sqrt{n}} \sim t_{(n-1)}$$

amostra aleatória da $N(\mu, \sigma^2)$

Teste t – Experimentos Comparativos

$$N(\mu_1; \sigma^2) \quad N(\mu_2; \sigma^2)$$



T₁

T₂

Y₁₁

Y₂₁

...

...

Y_{1n1}

Y_{2n2}

n₁

n₂

\bar{y}_1

\bar{y}_2

S₁

S₂

População



Amostra

- ✓ **Normalidade**
- ✓ **Variância constante**
- ✓ **Independência**

Teste t – Experimentos Comparativos

$$H_0 : \mu_d = \mu_1 - \mu_2 = 0$$

Variâncias Homogêneas

$$\bar{d} = \bar{y}_1 - \bar{y}_2 \quad \sim \quad N\left(\mu_d, \sigma^2\left(1/n_1 + 1/n_2\right)\right)$$

$$t = \frac{\bar{d} - \mu_d}{s_{\bar{d}}} \quad \sim \quad t_{(n_1+n_2-2)}$$
$$s_{\bar{d}}^2 = s^2\left(1/n_1 + 1/n_2\right)$$
$$s^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

Teste t – Experimentos Comparativos

$$H_0 : \mu_d = \mu_1 - \mu_2 = 0$$

Variâncias Homogêneas - σ conhecido

$$\bar{d} = \bar{y}_1 - \bar{y}_2 \sim N\left(\mu_d, \sigma^2\left(1/n_1 + 1/n_2\right)\right)$$

$$z = \frac{\bar{d} - \mu_d}{\sigma \sqrt{1/n_1 + 1/n_2}} \sim N(0,1)$$

Note que para o experimento “industrial”, σ pode ser calculado da distribuição de referência com 209 graus de liberdade ($\sigma = 2,88$)

Teste t – Experimentos Comparativos

$$H_0 : \mu_d = \mu_1 - \mu_2 = 0$$

Variâncias Heterogêneas

$$\bar{d} = \bar{y}_1 - \bar{y}_2 \quad \sim \quad N\left(\mu_d, \sigma_1^2 / n_1 + \sigma_2^2 / n_2\right)$$

$$t = \frac{\bar{d} - \mu_d}{s_{\bar{d}}} \quad \sim \quad t_{\nu} \quad s_{\bar{d}}^2 = s_1^2 / n_1 + s_2^2 / n_2 \quad \nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Experimentos Comparativos

Observações Correlacionadas

$$\bar{Y} = \frac{y_1 + \dots + y_n}{n} = \alpha_1 y_1 + \dots + \alpha_n y_n \quad \alpha_j = 1/n$$

$$E(\bar{Y}) = \sum_{j=1}^n \alpha_j \mu_j \quad V(\bar{Y}) = \sum_{j=1}^n \alpha_j^2 \sigma_j^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n \alpha_i \alpha_j \sigma_i \sigma_j \rho_{ij}$$

$$\Rightarrow V(\bar{Y}) = \frac{\sigma^2}{n} \left[1 + \frac{2(n-1)}{n} \rho_1 \right] \quad \text{modelo de autocorrelação de ordem 1}$$

Testes t não são válidos neste caso!

Teste de Mann Whitney

A	B	A_P	B_P
89,7	84,7	19	11
81,4	86,1	3	14
84,5	83,2	9,5	6
84,8	91,9	12	20
87,3	86,3	16	15
79,7	79,3	2	1
85,1	82,6	13	5
81,7	89,1	4	18
83,7	83,7	7,5	7,5
84,5	88,5	9,5	17
		95,5	114,5

$\Rightarrow W = 95,5$

Testes não paramétricos são robustos para fugas da normalidade, mas fortemente baseados na suposição de amostragem aleatória ou independência entre as observações \Rightarrow cálculo do nível de significância obtido a partir das amostras geradas por permutação dos dados efetivamente observados

Testes Não Paramétricos

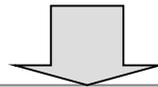
Distribuição de Referência

Situação Experimental: $n=20$ \longrightarrow $n_1 = 10$ $n_2 = 10$

Dados Originais: Y_1 Y_2 ... Y_{20}

Dados organizados em Postos: $1 = \min$... $20 = \max$

Gerar todas as possíveis ocorrências amostrais dos $n=20$ postos sujeitos à partição $n_1=10$ e $n_2=10$



Avaliar a significância da amostra efetivamente observada no conjunto das amostras geradas sob aleatorização