

Planejamento e análise de experimentos

Uma visão panorâmica

Prof. Walmes Zeviani

walmes@ufpr.br

Departamento de Estatística
Universidade Federal do Paraná

Plano de aula

Objetivos

- ▶ Apresentar as **ideias principais** de planejamento de experimentos.
- ▶ Apresentar e discutir a análise de experimentos baseados em **estudos de caso**.

No final, você irá...

- ▶ Compreender a **importância** de planejar um experimento.
- ▶ Conhecer os principais **fundamentos** e delineamentos.
- ▶ Ter uma visão panorâmica de **métodos para análise** de dados.

Por que planejar um experimento?

- ▶ Observar e descrever fenômenos.
- ▶ Isolar efeito e determinar relações e causas.
- ▶ Produzir resultados confiáveis.
- ▶ Otimizar o custo-benefício.
- ▶ A validade de um experimento depende da sua construção e execução, então atenção ao delineamento experimental é importante.



Estudos observacionais vs experimentais

Característica	Observacionais	Experimentais
Controle sob as condições de contorno	não	sim
Controle sob efeito de fatores indesejáveis	não	sim
Controle dos fatores sob investigação	não	sim
Determinar relações causais	não	sim
É o mais comum	sim	não
É o mais barato	sim	não

Definições

Unidade experimental (UE) é o que recebe ou contém o tratamento.

Tratamento é o que o pesquisador administra/aplica às unidades experimentais.

Fator representa a variável independente controlada cujos níveis são definidos pelo pesquisador.

Níveis são as categorias ou valores no domínio de um fator.

Fator ou tratamento? Gera confusão, então tratamento := condição/ponto experimental.

- ▶ Adubação mineral = $\{0, 50, 100, 150 \text{ kg ha}^{-1}\}$.
- ▶ Método de ensino = $\{\text{aula}, \text{aula+quiz}, \text{aula+jogo}, \text{vídeo+aula}\}$.
- ▶ Composição do concreto = $\{2:1, 1:1, 1:2\}$ em areia:pedra.

<http://www.stat.yale.edu/Courses/1997-98/101/expdes.htm>



Figura 1: Exemplos de unidades experimentais em vários contextos aplicados.

Qual é a unidade experimental?

- ▶ Há situações em que não é trivial **definir a UE**.
- ▶ Considere do seguinte experimento como exemplo:
 - ▶ Objetivo: avaliar o efeito do nível de poluição da água sobre lesões em peixes.
 - ▶ Condição: aquários com 50 peixes cada.
 - ▶ Avaliação: 30 dias depois, 10 peixes ao acaso tiveram lesões contadas.
- ▶ Qual é a unidade experimental: peixe ou aquário?

<https://onlinecourses.science.psu.edu/stat502/node/174/>

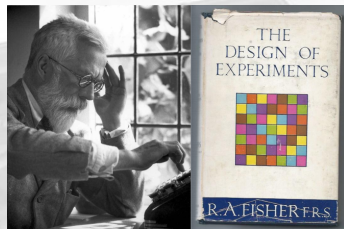
História - as primeiras contribuições

- ▶ Sir Ronald A. Fisher em 1920s.

- ▶ Responsável pela análise dos experimentos na *Rothamsted Agricultural Experimental Station*, Inglaterra.

- ▶ Suas obras tiveram profundo influência para a Estatística.

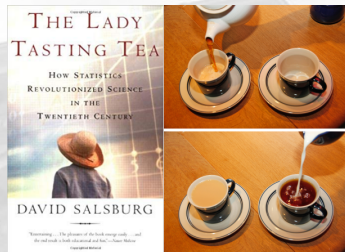
- ▶ R. A. Fisher (1958). *Statistical Methods for Research Workers*. 13th ed. Oliver & Boyd, Edinburgh.
 - ▶ R. A. Fisher (1966). *The Design of Experiments*. 8th ed. Hafner Publishing Company, New York.



- ▶ Princípios da experimentação: **repetição, aleatorização e controle local.**
- ▶ Conceitos de delineamentos experimentais, experimentos fatoriais, análise de variância, etc.

Uma senhora toma chá

- ▶ Aconteceu com Fisher e Muriel Bristol.
- ▶ A senhora declarou saber discriminar a bebida quanto a ordem em que chá e leite eram misturados.
- ▶ Experimento proposto: 8 xícaras, 4 de cada tipo servidas, aleatoriamente.
- ▶ Reposta: número de xícaras corretamente classificadas.
- ▶ Número de arranjos:
$$\frac{8!}{4!(8-4)!} = 70.$$
- ▶ Probabilidade ao acaso de acertar corretamente:
$$1/70 = 0.01429.$$



Princípio: Repetição

- ▶ Repetição é ter mais de uma UE de cada ponto experimental.
- ▶ Permite estimar o erro experimental (variação aleatória).
- ▶ A incerteza sobre efeitos diminui com mais repetições.
- ▶ **Cuidado:** pseudo-repetição *vs* repetição genuína.
- ▶ **Repetição genuína:** as repetições de ponto experimental capturam toda a variação que afeta a variável resposta (aquários).



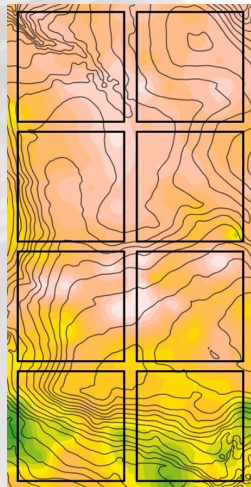
- ▶ **Pseudo-repetição:** capturam menos variação do que as genuínas (peixes).

Portanto, tem-se que ter repetições

- ▶ Para ter resultados de respaldo.
- ▶ Para mitigar o efeito de fatores não controlados.
- ▶ Para estimar o erro experimental.
- ▶ Para aumentar a precisão dos efeitos estimados.

Princípio: Aleatorização

- ▶ Problema: as **condições** nem sempre são **homogêneas**.
- ▶ Podem haver **causas identificáveis** de variação.
- ▶ Deseja-se estimativas não tendenciosas dos efeitos e erro experimental.
- ▶ Não tendenciosas: livre da **influência sistemática** de fatores não controlados.
- ▶ Permite mitigar efeito de origem serial: descalibração de processo/instrumento, mudança condições de contorno/operação.
- ▶ Tem situações em que a aleatorização é complicada: sistemas de cultivo/uso da terra.



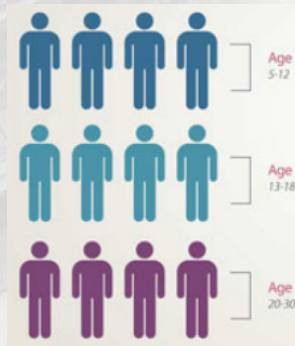
- ▶ As UE podem ser ou estar em condições:
 - ▶ Homogêneas: quaisquer diferenças são devido ao acaso.
 - ▶ Heterogêneas: diferenças são explicadas por uma causa identificável ...
 - ▶ ... nas próprias unidades experimentais (passado).
 - ▶ ... nas condições de contorno as quais estarão sujeitas (futuro).
- ▶ A causa de variação identificável pode ter níveis:
 - ▶ Contínuos:
 - ▶ Peso/idade inicial dos leitões em estudo de ganho de peso.
 - ▶ Tempo de uso de um medicamento por pacientes.
 - ▶ Quantidade de horas semanais em atividades esportivas.
 - ▶ Nível de fertilidade das parcelas de uma estação experimental.
 - ▶ Número de escovações diárias de pacientes odontológicos.
 - ▶ Categóricos:
 - ▶ Sexo dos animais em um estudo metabólico.
 - ▶ Classe de vigor de mudas de árvores frutíferas.
 - ▶ Grau de escolaridade de respondentes de um questionário.
 - ▶ Fabricantes de cimento para um estudo de resistência de concreto.
 - ▶ Região geográfica de origem de isolados de um patógeno vegetal.

Portanto, tem-se que aleatorizar

- ▶ Para ter resultados de respaldo.
- ▶ Para balancear o efeito de fatores não controlados.
- ▶ Tornar os resultados não viciados ou comprometidos por fatores sistemáticos.

Princípio: Controle local

- ▶ Medidas que visam reduzir/controlar a interferência de fatores externos.
- ▶ Reduzem o erro experimental (aleatório) com o controle das fontes sistemáticas de variação.
- ▶ **Blocagem:** agrupamento de UE uniformes para realização da aleatorização.
- ▶ **Bordadura:** isolar efeitos de borda.
- ▶ **Ensaio cegos** (simples, duplo): sem conhecimento do tratamento.
- ▶ **Questão chave:** quando/em que/como fazer controle local?



Princípio: Controle local



Figura 2: Uma horta onde os cultivos são experimentos conduzidos por alunos.

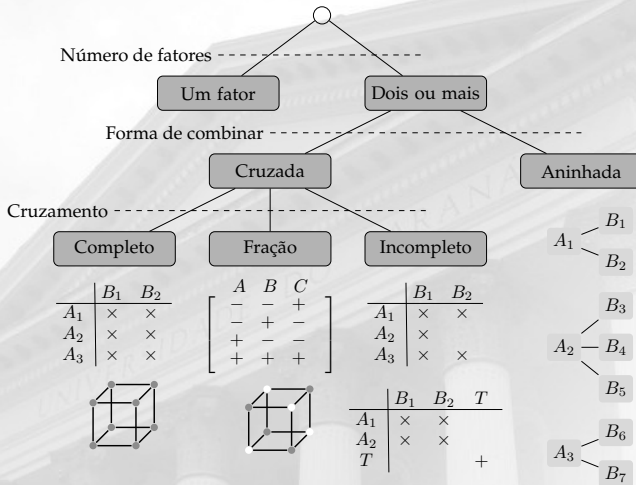
Portanto, tem-se que fazer controle local

- ▶ Para ter resultados de respaldo.
- ▶ Para controlar fontes de variação sistemáticas.
- ▶ Reduzir o erro experimental e assim aumentar precisão.

Tipos de fatores

- ▶ São os fatores que o pesquisador deseja estudar.
- ▶ Fontes de variação controladas no experimento.
- ▶ Podem ser métricos ou categóricos.
- ▶ Métricos ou quantitativos: suporte numérico contínuo ou discreto.
- ▶ Categóricos ou qualitativos: suporte em níveis nominais ou ordinais.
- ▶ Podem ser estudados vários fatores simultaneamente.

Combinação de fatores



Delineamentos

Descrevem as condições nas quais foram feitas a(s) aleatorização(ões).

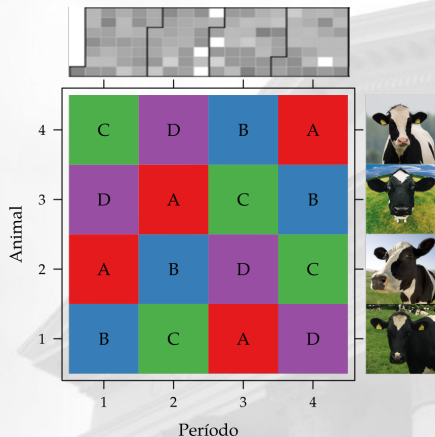
Quanto a blocagem

- ▶ Delineamento inteiramente casualizado.
- ▶ Delineamento de blocos completos casualizados.
- ▶ Delineamento de blocos incompletos balanceados.
- ▶ Delineamento de quadrado latino.
- ▶ Outros.

Quanto às camadas de aleatorização ou tamanho de parcela

- ▶ Um estrato → um erro experimental.
- ▶ Mais de um estrato → mais de um erro experimental.
 - ▶ Parcela subdividida, subsubdividida, etc.
 - ▶ Ensaios em faixas.
- ▶ Séries longitudinais → dependência.
 - ▶ Efeito do tempo/épocas, efeito da profundidade, safra/coorte.

Delineamentos



- ▶ 7 blocos de tamanho 3.
- ▶ Fator de 7 níveis com 3 repetições.

bloco 1	1	5	6
bloco 2	1	4	3
bloco 3	7	2	1
bloco 4	5	7	4
bloco 5	7	3	6
bloco 6	4	6	2
bloco 7	3	5	2

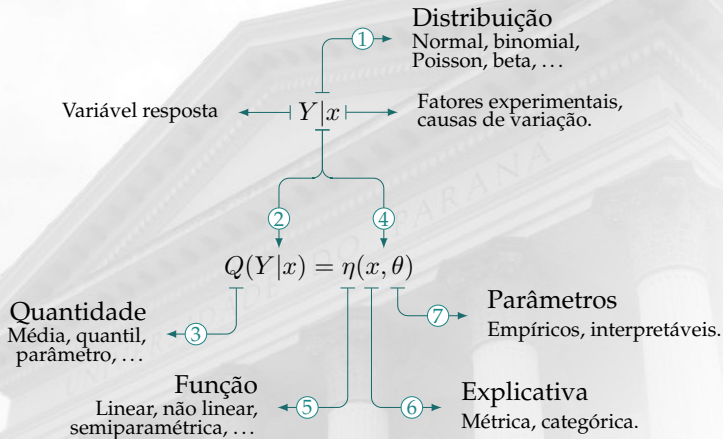
Análise exploratória

- ▶ Inspeção dos dados antes da análise → conformidade.
- ▶ Definição do modelo estatístico:

modelo = características do delineamento
+ propriedades dos dados
+ suposições
+ hipóteses.

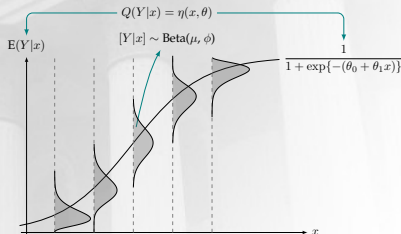
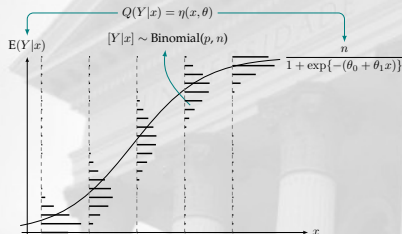
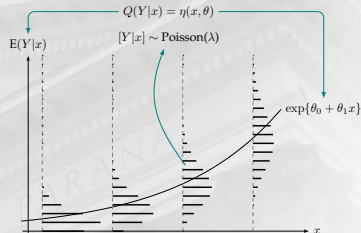
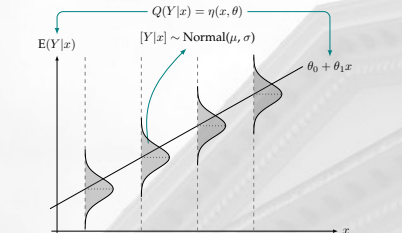
- ▶ Conformidade com as suposições do modelo.
- ▶ Comunicação dos resultados:
 - ▶ Representação dos efeitos dos fatores.
 - ▶ Descrição do modelo ajustado.
 - ▶ Evidência para as hipóteses, etc.

Especificação do modelo



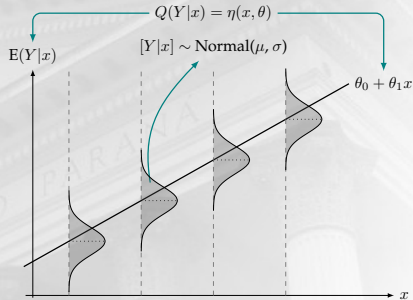
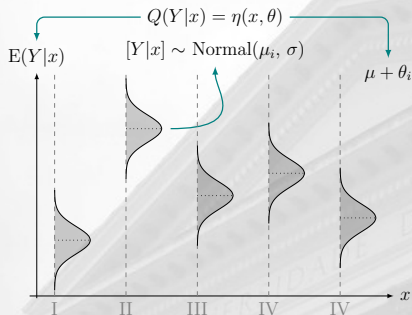
Especificação do modelo

Distribuição da variável resposta



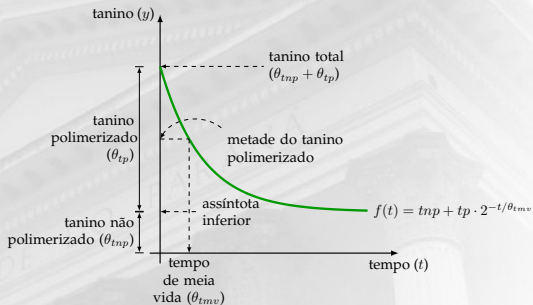
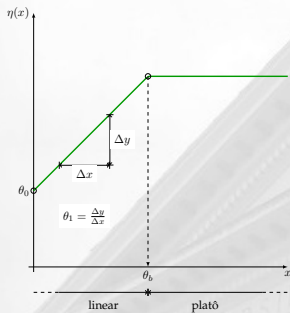
Representação do efeito

Fator categórico vs métrico



Representação do efeito

Modelos não lineares



Exemplos de análise

Pontos relevantes

- ▶ Análise exploratória.
- ▶ Especificação e ajuste do modelo.
- ▶ Verificação dos pressupostos.
- ▶ Testes de comparação múltipla.
- ▶ Uso de regressão.
- ▶ Apresentação dos resultados.

Exemplos de análise

- ▶ Tutoriais do pacote **labestData**.
- ▶ Relatórios de análise do pacote **EACS**.
- ▶ Transformação de dados: **controle de pulgões**.
- ▶ Análise de dados de contagem: **vagens e grãos em soja**.
- ▶ Regressão múltipla: **modelo hipsométrico**.
- ▶ Regressão por categorias: **preço do HB20**.
- ▶ Regressão não linear: **adubação potássica em soja**.
- ▶ Regressão local: **temperatura corporal em suínos**.
- ▶ Regressão quantílica: **fronteira de suficiência em soja**.

Recomendações

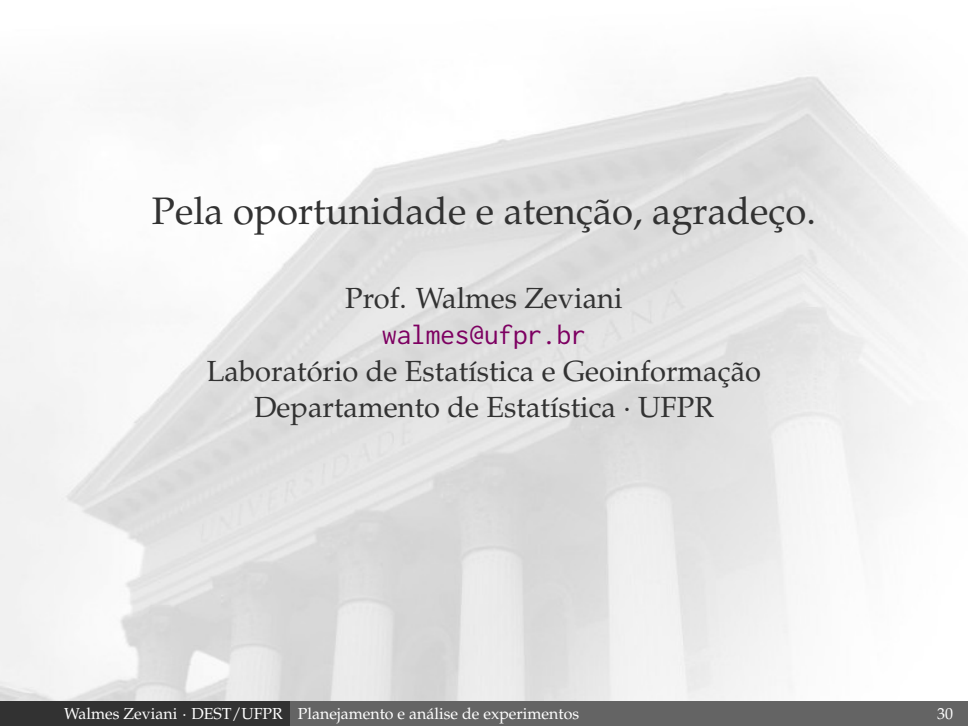
“To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.”

Ronald Fisher

“If your experiment needs statistics, you ought to have done a better experiment.”

Lord Ernest Rutherford

- ▶ Não procure o estatístico apenas no final.
- ▶ Experimentos bem planejados devem ser extremamente fáceis de analisar.



Pela oportunidade e atenção, agradeço.

Prof. Walmes Zeviani

walmes@ufpr.br

Laboratório de Estatística e Geoinformação

Departamento de Estatística · UFPR