

FIP 606


Análise e Visualização de dados

Unidade 2 - Planejamento do experimento

Emerson M. Del Ponte

Departamento de Fitopatologia - UFV

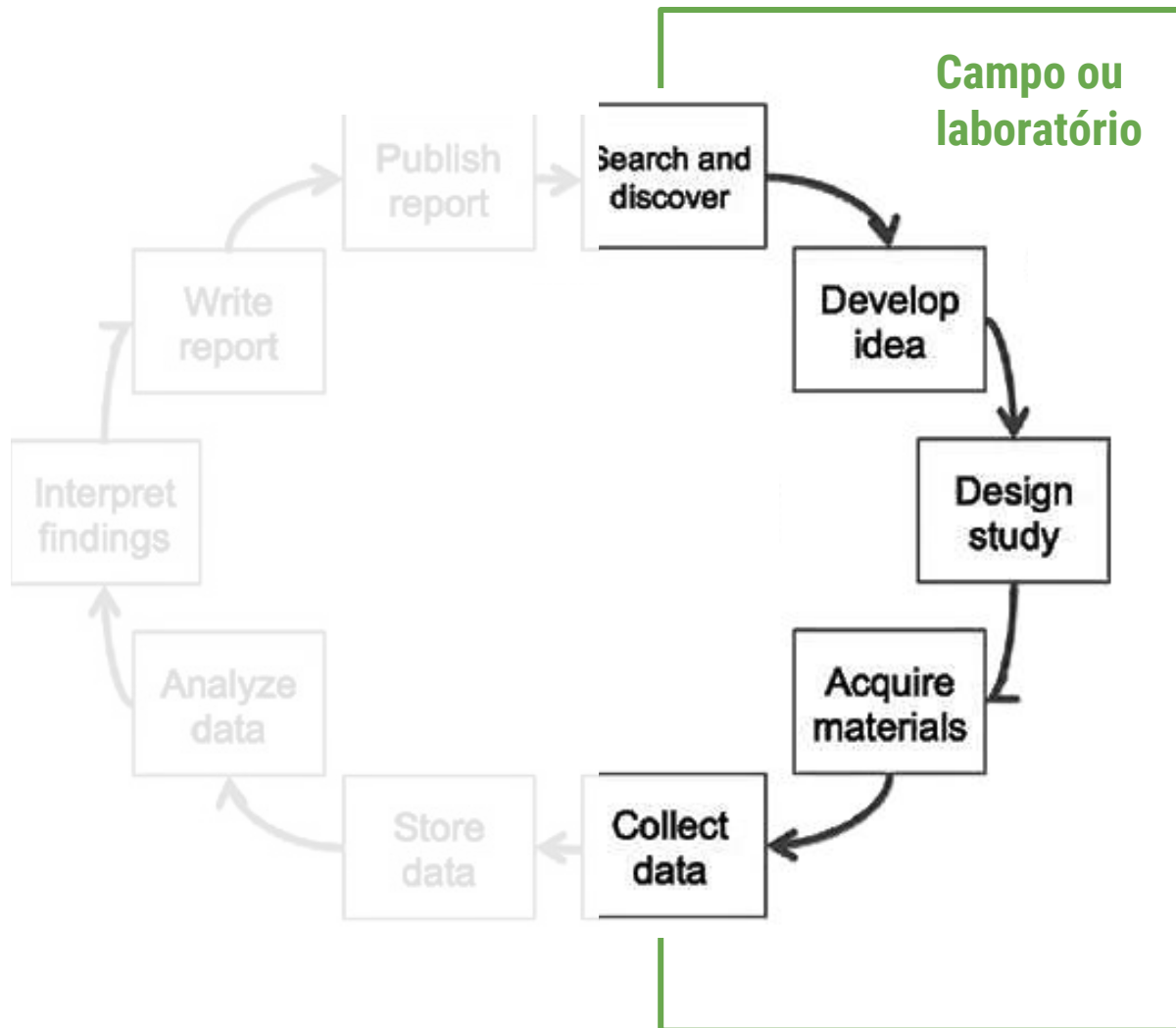
delponte@ufv.br



Planejando o estudo: o que e como
coletar e como organizar os dados



Ciclo de vida da pesquisa científica



Open Science Framework

<https://osf.io/>

Tipos de Pesquisa ou Estudos

Pesquisa observacional
(Levantamento de campo)

Censitário
(População)

Descritivos
(Descrevem a situação)

Amostral
(Parte da população)

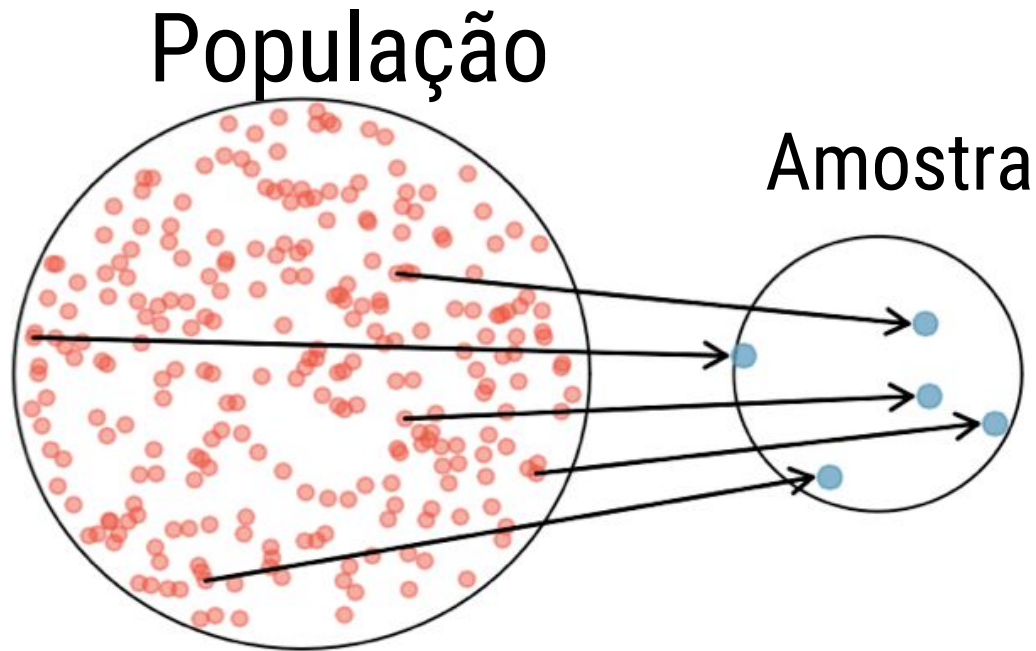
Analíticos
(Tentam explicar a situação)

Pesquisa experimental

Exposição a um **fator**
(controlado) por intervenção do pesquisador

Seguem os **princípios** de casualização, repetição e (às vezes) controle local

Pesquisa Observacional



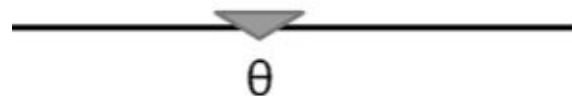
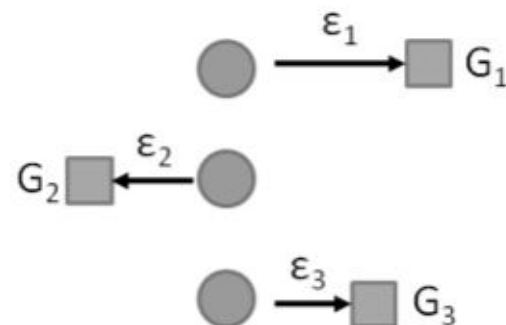
Exemplos de população e amostras?

Pesquisa Experimental

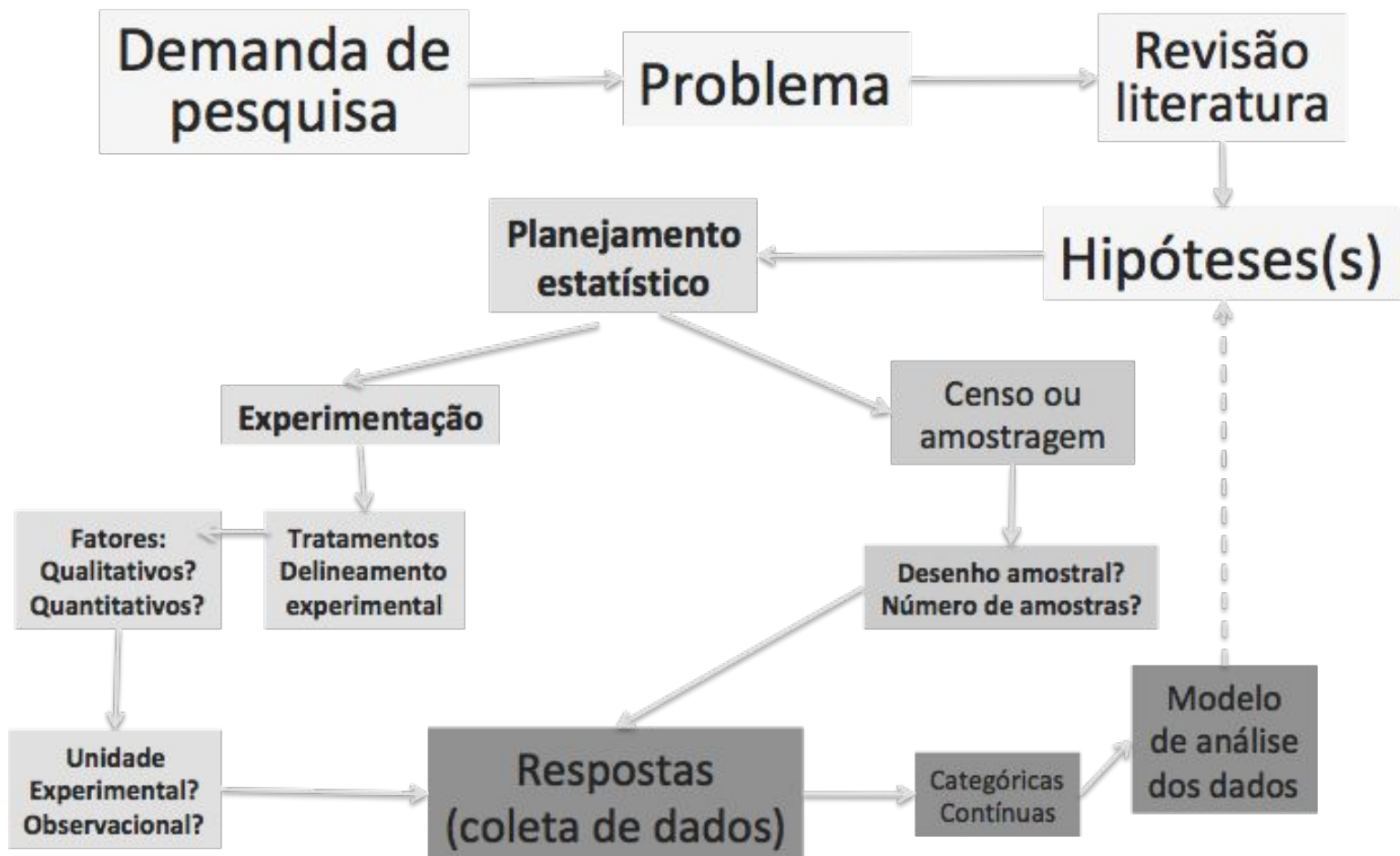
O fenômeno observado tem as causas de sua variação conhecidas

Tratamentos

Erro aleatório



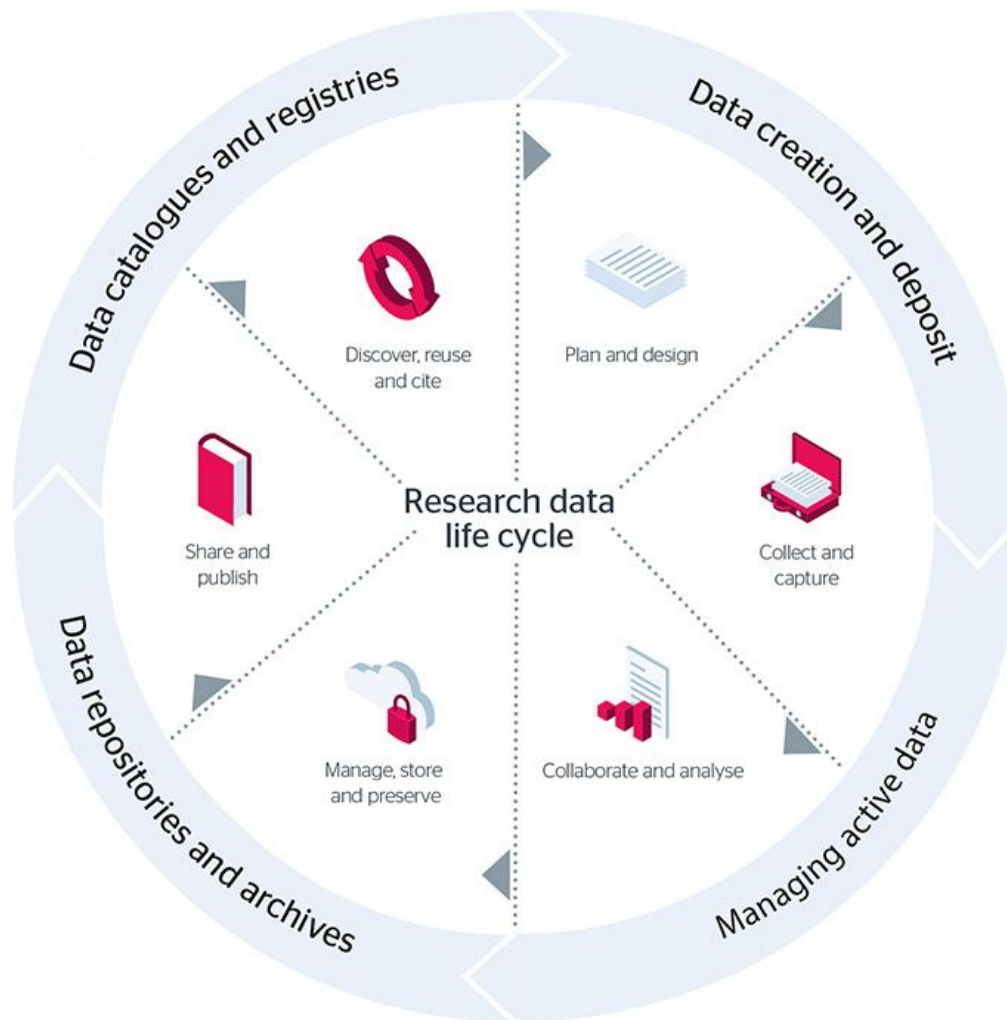
Passos e decisões



Plano, execução e análise experimental

1. Define o problema (científico)
2. Define observações (variáveis ou características)
3. Define as unidades experimentais e observacionais
4. Define os tratamentos
5. Define o delineamento experimental
6. Coleta os dados: variáveis principais e suplementares
7. Confere, organiza, documenta e prepara os dados
8. Conduz análise exploratória (tabelas e gráficos)
9. Ajusta o modelo estatístico (inferencial)
10. Interpreta os resultados
11. Prepara o relatório da análise

"Life cycle" dos dados



How and why you should manage your research data: a guide for researchers

<https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data>

Os Dados

FAIR

Findable, **A**ccessible, **I**nteroperable and **R**eusable

(Wilkinson et al., 2016)

Por que compartilhar os dados?

- Permitir que novas análises sejam feitas com os dados, ou mesmo combinando com outros conjuntos de dados (metanálise).
- Agências de fomento estarão em breve solicitando!

Os dados

Boas práticas

- **Documentação:** dados mais fáceis de entender
- **Formatação:** uso em programas de computador
- **Distribuição:** repositórios e licença aberta

Por que muitos não compartilham?

- Receio de perder a "corrida" pela publicação
- Desconhecimento de como compartilhar
- Postura contrária ao compartilhamento
- Percepção de que é tecnicamente difícil..
- .. e de que toma tempo



Descrição dos dados

- O que são, quando, como e por quem foram coletados
- Como encontrar e acessar os dados (links para repositórios)
- Comentário sobre a utilidade dos dados
- Alertas sobre possíveis problemas ou inconsistências
- Descrição de limitações nos dados ou problemas específicos
- Informação de como conferir se os dados foram importados OK
- Número total de linhas e colunas
- Nome das variáveis, unidades, etc.

Como organizar os dados?

Dados brutos

Os dados podem ser modificados de sua forma original de como foram coletados e anotados. Valores são sumarizados, unidades convertidas, nomes alterados, ou índices calculados. É importante que os dados originais sejam mantidos para se saber o quanto o conjunto foi alterado

Há um formato ideal?

- Dados tabulares de formato retangular são preferidos
- Formato **"Tidy"**: cada linha representa uma observação/sujeito (registro) e cada coluna representa uma variável
- Não duplique colunas com a mesma informação (há exceções)
- Não coloque fórmulas nas células
- Células não são coloridas ou mescladas
- Nomeie as variáveis de maneira simples (trat, rep, peso, altura)
- Datas no formato YYYY-MM-DD
- Não deixe espaços vazios nas células
- Dado faltante: célula vazia ou adicione "NA"
- Exportar para arquivo texto (.csv é o mais comum)
- Escolha um nome curto e descritivo (sem espaços!)