



**Universidade Estadual do Ceará**  
**Programa de Pós-graduação em Ciência da Computação**  
**Professor: Ismayle de Sousa Santos**

**Aula 17,  
18 e 19**

# **Metodologia Científica para Computação**

## **Experimento e Quase-experimentos**

---



**IsmayleSantos**



**ismayle.santos@uece.br**

# Estratégias Empíricas

- A pesquisa empírica, **também chamada de pesquisa de campo, pode ser entendida como aquela em que é necessária comprovação prática de algo**, especialmente por meio de experimentos ou observação de determinado contexto para coleta de dados em campo
- Dependendo do objetivo da avaliação, sejam técnicas, métodos ou ferramentas, e dependendo das condições para a investigação empírica, existem três principais tipos diferentes de investigações que podem ser realizadas:
  - Survey
  - Estudo de caso
  - Experimento

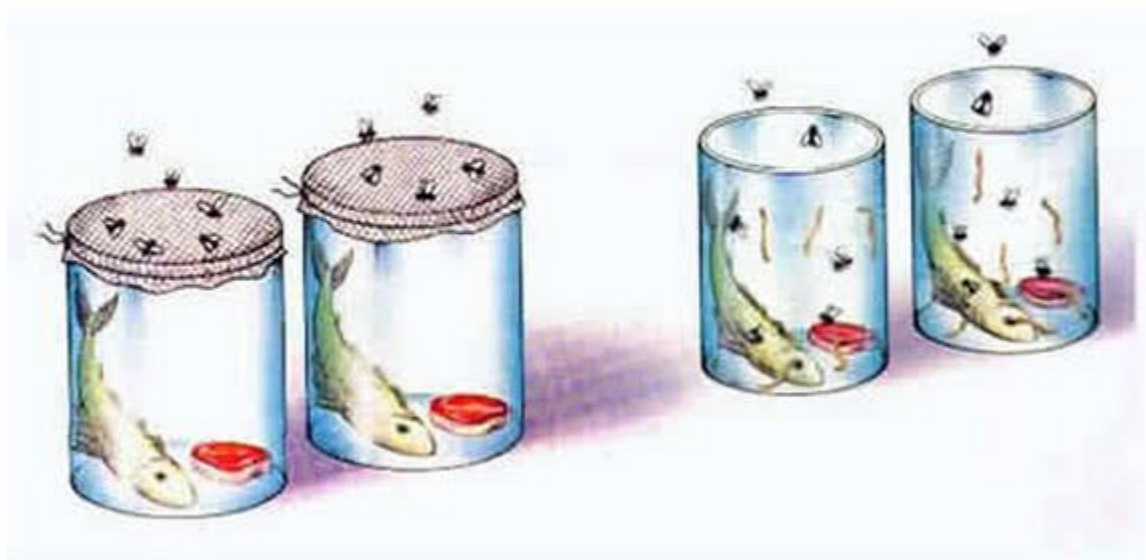
# Estratégias Empíricas

- Por que devemos realizar experimentos e outros estudos empíricos em software Engenharia?
  - As principais razões para a realização de estudos empíricos quantitativos é a oportunidade de **obter resultados objetivos e estatisticamente significativos** em relação ao compreensão, controle, previsão e melhoria do desenvolvimento de software
- Os estudos empíricos são uma entrada importante para a tomada de decisão em uma melhoria buscando organização

# Então o que é Experimentação na Engenharia de Software?

## EXPERIMENTO DE REDI

**Vermes surgem da carne?**



**Para nos guiar nesse processo de experimentação na E.S. vamos usar o livro *Experimentation in Software Engineering* de Claes Wohlin et. al.**

# O que é Experimento em Engenharia de Software?

- Também conhecido como experimento controlado, o experimento é uma **investigação empírica que manipula um fator ou variável do ambiente estudado**
- Os experimentos são apropriados para confirmar as teorias, confirmar o conhecimento convencional, explorar os relacionamentos, avaliar a predição dos modelos, ou validar as medidas



# O que é Experimento em Engenharia de Software?

- Experimentação é o centro do processo científicos
  - Somente experimentos verificam as teorias
  - Somente experimentos podem explorar os fatores críticos e dar luz ao fenômeno novo para que as teorias possam ser formuladas e corrigidas
- Experimentação oferece o modo **sistemático, disciplinado, computável e controlado** para avaliação da atividade humana
- É importante notar que experimentos não provam nada
  - Os experimentos verificam a previsão teórica de encontro à realidade

# O que é Experimento em Engenharia de Software?

- A experimentação parte inicialmente de uma ideia de relação de **causa e efeito**
  - É quando temos uma teoria ou podemos formular uma hipótese onde exista uma relação entre uma construção de causa e efeito
    - Uma hipótese significa que temos uma ideia do que acontece em torno de um problema
- O experimento é criado, por exemplo, **para testar uma teoria ou hipótese**
  - O objetivo principal de um experimento é principalmente avaliar uma hipótese ou relação de causa e efeito

# O que é Experimento em Engenharia de Software?

- Os experimentos são apropriados para investigar diferentes aspectos, tais como:
  - Confirmação de teorias
    - Para testar as teorias existentes
  - Explorar relacionamentos
    - Testar se um determinado relacionamento se mantém
  - Avaliar a precisão dos modelos
    - Para testar se a precisão de certos modelos é o esperado
  - Validar medidas
    - Para garantir que uma medida realmente mede o que é suposto



# O que é Experimento em Engenharia de Software?

- A experimentação **não é simples**
  - Temos que preparar, conduzir e analisar experimentos devidamente
- Uma das principais vantagens de um experimento é o controle de, por exemplo, assuntos, objetos e instrumentação
- Outras vantagens incluem:
  - **A capacidade de realizar análises estatísticas** usando métodos de teste de hipótese e oportunidades para replicação
  - **Reduz incertezas** sobre quais metodologias, linguagens, técnicas e ferramentas são ideais para um contexto
  - **Contribui para que pesquisadores explorem a evolução das mudanças tecnológicas**

# Estratégias para Coleta de Dados Experimentais

- Existem três principais métodos para coleta de dados:
  - **Histórico**
    - É utilizado para coletar os dados experimentais dos projetos que já tenham sido terminados
    - Os dados já existem e é preciso analisá-los
  - **De observação**
    - Coleta os dados relevantes enquanto o projeto está sendo executado
    - O método oferece o controle fraco sobre o processo de desenvolvimento
  - **Controlado**
    - Provê as instâncias múltiplas de uma observação oferecendo a validade estatística dos resultados

# Categorias de Pesquisa

- Categorias gerais de pesquisa:
  - **Estudo qualitativo**
    - Está relacionado à pesquisa sobre os objetos quando os resultados são apresentados em termos naturais
  - **Estudo quantitativo**
    - Pode ser conduzido através um experimento controlado
    - Uma das vantagens do estudo controlado é que os dados quantitativos promovem a comparação e a análise estatística
  - **Benchmarking**
    - É utilizado para a medição do desempenho dos diferentes produtos de software

# Processo do Experimento

- Um experimento deve ser tratado como **um processo da formulação ou verificação de uma teoria** para que este processo ofereça os resultados válidos
  - Todo experimento possui um processo
- O processo da execução de um experimento presume a realização de diferentes atividades
  - O número e a complexidade dessas atividades podem variar de acordo com as características do estudo

# Processo do Experimento

- O objetivo é dar suporte na implantação e condução de um experimento
- A literatura apresenta cinco fases gerais que sempre estão presentes num processo da experimentação
- Mas para compreender esse processo é preciso entender o que são as nomenclaturas envolvidas no processo do experimento:

|                  |                      |
|------------------|----------------------|
| <b>Variáveis</b> | <b>Tratamentos</b>   |
| <b>Objetos</b>   | <b>Participantes</b> |

# O que são Variáveis, Tratamentos, Objetos e Participantes?

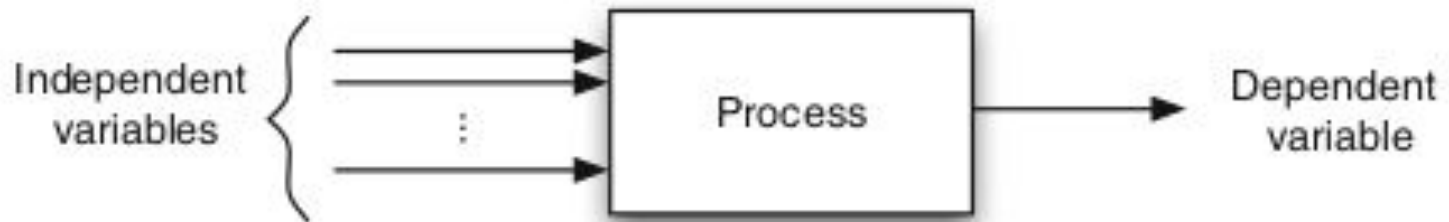


# O que são Variáveis?

- Ao conduzir um experimento formal, estudamos o resultado a partir de mudanças em algumas das variáveis de entrada
- Existem dois tipos de variáveis em um experimento:
  - As variáveis **dependentes** no experimento são aquelas cujos resultados dependem do comportamento das variáveis **independentes**, que são manipuladas e controladas pelo pesquisador
  - Frequentemente há apenas uma variável dependente em um experimento
  - Todas as variáveis em um processo que são manipulados e controlados são chamados de variáveis independentes

# Exemplo de Variáveis

- Em um experimento com o uso de determinado método
  - A variável dependente do experimento pode ser a produtividade
  - Variáveis independentes podem ser o método de desenvolvimento, a experiência do pessoal, uso de ferramenta de suporte e ambiente





# O que são os Tratamentos?

- Um experimento estuda o efeito da mudança de uma ou mais variáveis independentes
  - Essas variáveis são chamadas de **fatores**
- As outras variáveis independentes são controladas em um nível fixo durante o experimento, ou então não podemos dizer se o fator ou outra variável causou o efeito
- Um **tratamento** é um valor particular de um fator

# O que são os Tratamentos?

- Um experimento consiste em um conjunto de testes (às vezes chamados de tentativas), onde cada teste é uma combinação de **tratamento**, **participante** e **objeto**
- Experimentos envolvem mais de um tratamento para comparar os resultados



# O que são os Objetos?

- O objeto é uma ferramenta usada para verificar o relacionamento de causa e efeito em uma teoria
- Por exemplo:
  - Programas a serem desenvolvidos
  - Documento que deve ser revisado com diferentes técnicas de inspeção
- Durante a execução do experimento, os tratamentos são sendo aplicados ao conjunto dos objetos e assim o resultado está sendo avaliado
- Os objetos junto com o sistema de medição e diretrizes da execução do experimento compõem a **instrumentação do experimento**

# O que são os Participantes?

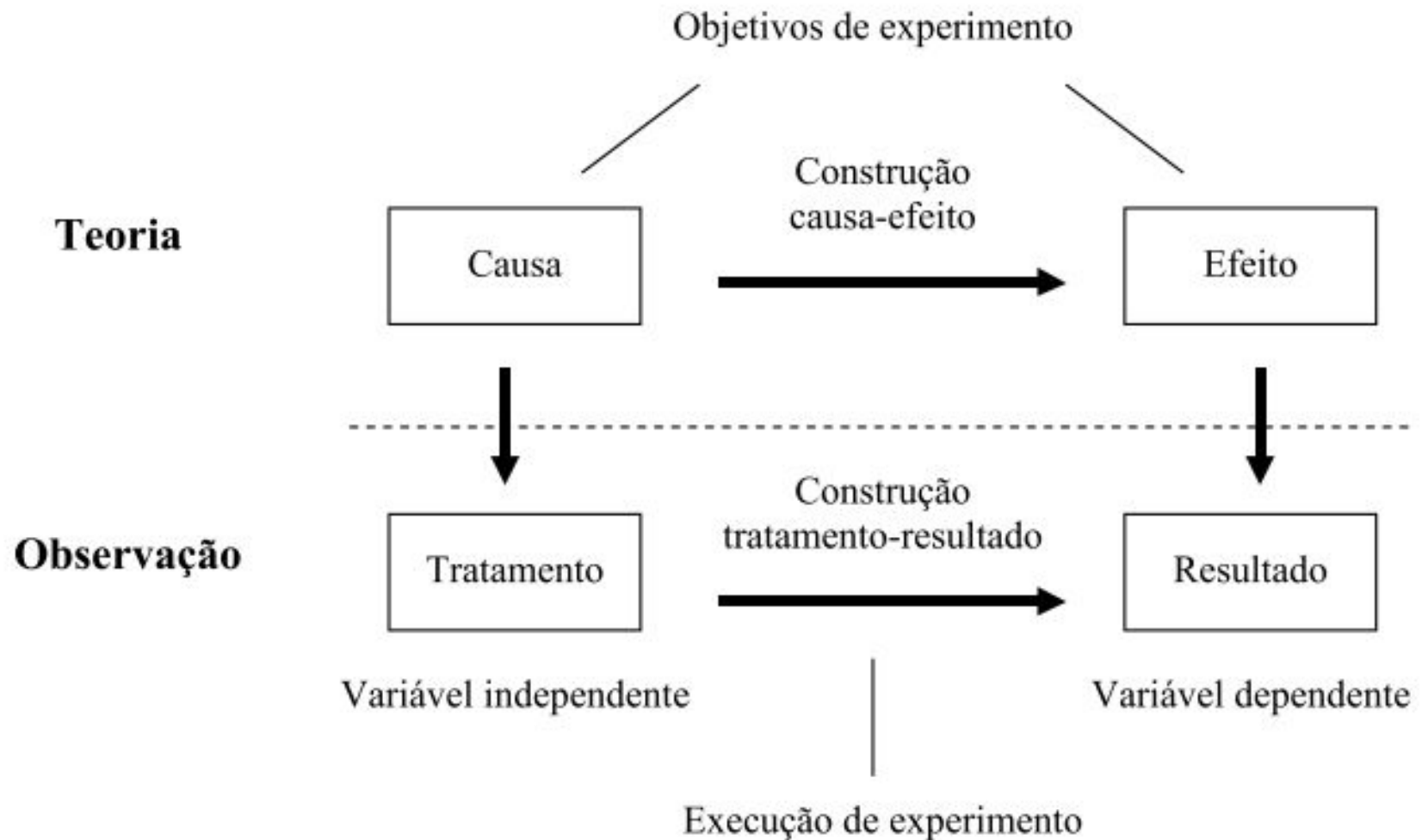
- Os **participantes** também podem ser chamado de sujeitos
  - São as pessoas que aplicam o tratamento



## Em resumo ...

- As **variáveis independentes** referem-se à entrada do processo de experimentação
- Essas variáveis também se chamam "**fatores**" e apresentam a causa que afeta o resultado do processo de experimentação
- O próprio valor de um fator se chama "**tratamento**"
- As **variáveis dependentes** referem-se à saída do processo de experimentação
- Essas variáveis apresentam o efeito que é causado pelos fatores do experimento
- O próprio valor de uma variável dependente se chama "**resultado**"

# Princípios do Experimento

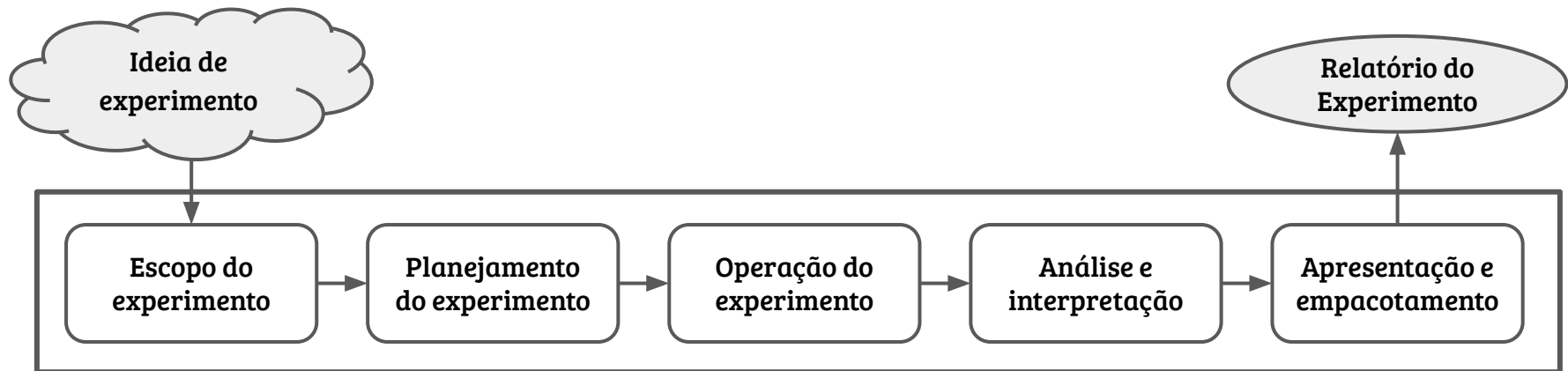


# Processo do Experimento



# Processo do Experimento

- Um processo fornece etapas que apoiam um experimento
- Os processos são importantes, pois podem ser usados como listas de verificação e diretrizes de **o que fazer** e **como fazer**





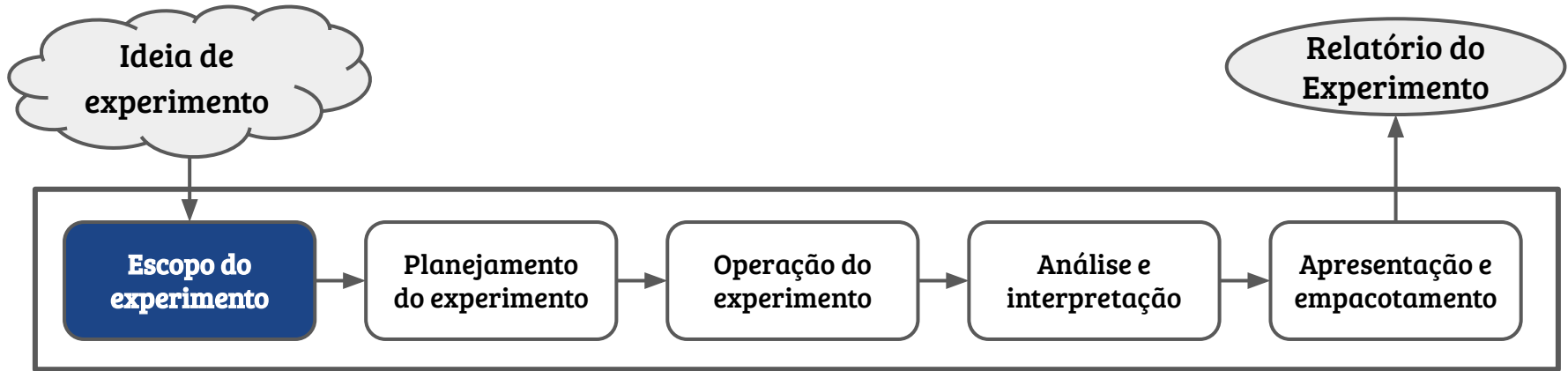
# Processo do Experimento

- O processo de experimento é dividido nas seguintes atividades:
  - **Escopo do experimento** é a primeira etapa, onde definimos o experimento em termos de problema, objetivo e metodologia
  - O **planejamento do experimento** é onde o projeto do experimento é determinado, a instrumentação é considerada e as ameaças ao experimento são avaliadas
  - A **operação do experimento** decorre do design e é onde as medições são coletadas e posteriormente, analisadas, avaliadas e interpretadas na etapa de **análise e interpretação**

# Processo do Experimento

- Finalmente, os resultados são apresentados e empacotados na atividade de **apresentação e empacotamento**
- O processo não deve ser um modelo em cascata "verdadeiro"; **não é assumido que uma atividade é necessariamente concluída antes que a próxima atividade seja iniciada**
  - A principal exceção é quando a operação do experimento foi iniciada, então não é possível voltar para o escopo e planejamento do experimento
    - Razão: os participantes já estão influenciados pelo experimento
- A ordem de atividades no processo indica principalmente a ordem de início das atividades

# Processo do Experimento



# Escopo do Experimento

- A primeira atividade é o escopo
  - O objetivo é formulado a partir do problema a ser resolvido
- Para elaborar o escopo, foi sugerido uma estrutura:
  - Objeto de estudo (o que é estudado?)
  - Objetivo (qual é a intenção?)
  - Foco na qualidade (qual efeito é estudado?)
  - Perspectiva (visão de quem?)
  - Contexto (onde o estudo é realizado?)



# Escopo do Experimento

- O **escopo** do experimento é determinado pelos objetivos
- Ao definir o objetivo do experimento conforme o modelo abaixo, você poderá garantir que aspectos importantes de um experimento sejam definidos antes que o planejamento e a execução ocorram
  - Analisar <Objeto (s) de estudo>
  - com o propósito de <Finalidade/Objetivo>
  - com relação ao seu <foco na qualidade>
  - do ponto de vista da <Perspectiva>
  - no contexto de <Contexto>

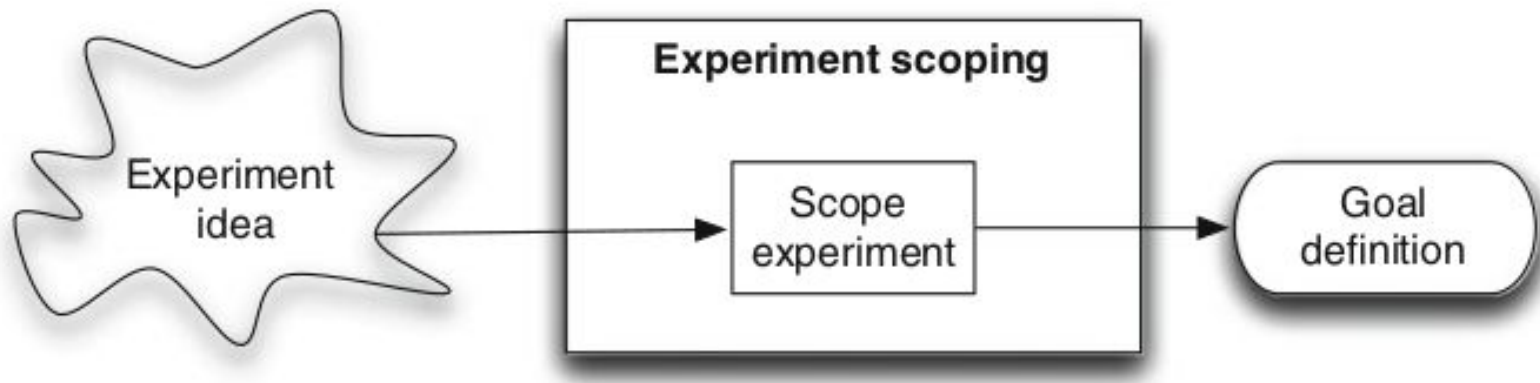
# Escopo do Experimento

- O **objeto** de estudo é a entidade estudada no experimento
  - O objeto de estudo pode ser produtos, processos, recursos, modelos, métricas ou teorias
  - Um bom exemplo, é o produto final, o processo de desenvolvimento ou inspeção
- O **objetivo** define qual é a intenção do experimento
  - Pode ser para avaliar o impacto de duas técnicas diferentes, ou para caracterizar a aprendizagem curva de uma organização

# Escopo do Experimento

- O **foco da qualidade** de um experimento pode ser eficácia, custo, confiabilidade, etc
- A **perspectiva** informa o ponto de vista do qual os resultados do experimento são interpretados
  - Ex.: desenvolvedores, gerente de projeto, cliente e pesquisador
- O **contexto** é o "ambiente" em que o experimento é executado
  - Define quais os participantes e quais artefatos de software (objetos) são usados no experimento
    - Os **participantes** podem ser caracterizados pela experiência, tamanho da equipe, carga de trabalho, etc
    - Os **objetos** podem ser caracterizados por tamanho, complexidade, prioridade, domínio do aplicativo etc...

# Classificação do Contexto do Experimento



**Fig. 7.1** Scoping phase overview

**Table 7.1** Experiment context classification

|            |               | # Objects                      |                              |
|------------|---------------|--------------------------------|------------------------------|
|            |               | One                            | More than one                |
| # Subjects | One           | Single object study            | Multi-object variation study |
| per object | More than one | Multi-test within object study | Blocked subject-object study |



# Escopo do Experimento

- Os estudos de **único objeto** são realizados com um único participante e um único objeto
- **Estudo de objetos com vários testes** são conduzidos com um único participante em um conjunto de objetos
- **Estudos da variação de objetos** examinam um único objeto em um conjunto de participantes
- Os estudos **agrupados por participantes e objeto** examinam um conjunto de participantes e um conjunto de objetos

# Escopo do Experimento

- Todos esses tipos de experimento podem ser executados em um experimento ou quase-experimento
  - Em um **quase-experimento**, há uma falta de randomização de sujeitos ou objetos
    - O estudo de objeto único é um quase-experimento se o único sujeito e objeto não são selecionados ao acaso
    - Será um experimento se o sujeito e o objeto forem escolhidos por acaso

# Exemplo de Escopo do Experimento

**Table 7.3** Goal definition framework

| Object of study | Purpose      | Quality focus   | Perspective       | Context  |
|-----------------|--------------|-----------------|-------------------|----------|
| Product         | Characterize | Effectiveness   | Developer         | Subjects |
| Process         | Monitor      | Cost            | Modifier          | Objects  |
| Model           | Evaluate     | Reliability     | Maintainer        |          |
| Metric          | Predict      | Maintainability | Project manager   |          |
| Theory          | Control      | Portability     | Corporate manager |          |
|                 | Change       |                 | Customer          |          |
|                 |              |                 | User              |          |
|                 |              |                 | Researcher        |          |

# Escopo do Experimento

## 1.3 Objetivo do estudo:

**Analisar** as competências oferecidas aos alunos

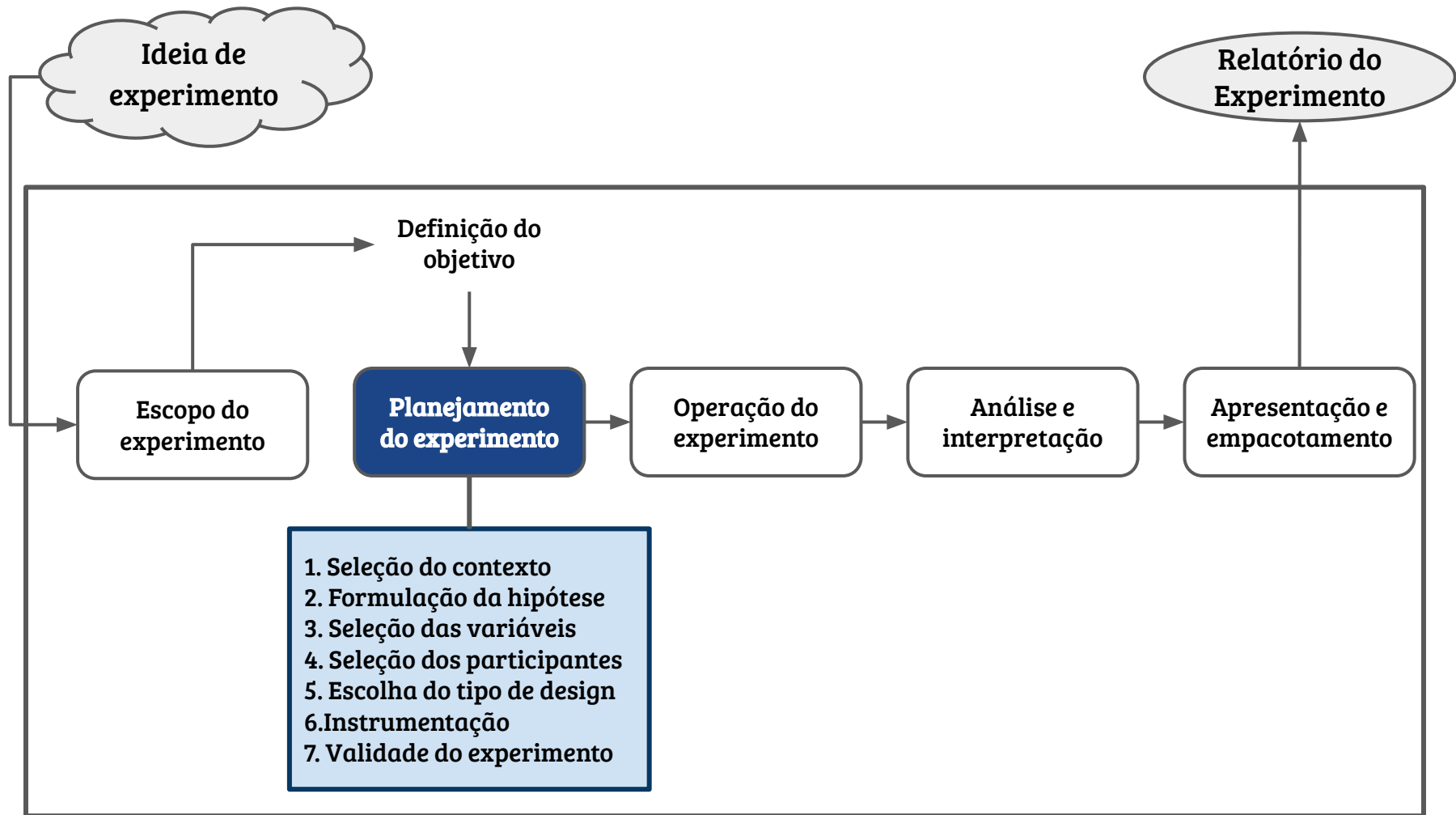
**Com o propósito de** caracterizar

**Com respeito à** interseção com as competências do currículo mínimo da SBC

**Do ponto de vista** do desenvolvedor de software pessoal

**No contexto de** alunos de curso básico de ES.

# Processo do Experimento



# Planejamento do Experimento

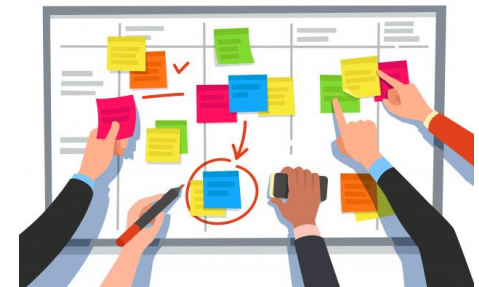
- A atividade de planejamento é onde a base para o experimento é delimitado
  - Na escrita do planejamento do experimento, os verbos devem estar (quase) todos no futuro do presente
  - O documento pode ser alterado, a partir de uma execução-piloto
- O planejamento deve incluir **pessoal** e o **ambiente**
  - Por exemplo, se o experimento é executado em uma universidade com alunos
- Também é onde a hipótese do experimento é declarada formalmente, incluindo uma hipótese nula e alternativa
- No planejamento **são determinadas as variáveis independentes (entradas) e as variáveis dependentes (saídas)**

# Planejamento do Experimento

- Uma questão importante em relação às variáveis são os valores que elas realmente podem assumir
- Isso também inclui a determinação da **escala de medição**, o que coloca restrições ao método que mais tarde podemos aplicar para análise estatística
- Além disso, o experimento é projetado, o que inclui a escolha de um adequado de **projeto de experimentos** incluindo, por exemplo, randomização de sujeitos

# Planejamento do Experimento

- Um problema intimamente relacionado com o design é o preparo para a instrumentação do experimento
- Devemos identificar e preparar objetos adequados, desenvolver diretrizes e definir procedimentos de medição
- Como parte do planejamento, é importante considerar a questão **da validade dos resultados** que podemos esperar
- A validade pode ser dividida em quatro classes principais:
  - Validade interna
  - Validade externa
  - Validade de construção
  - Validade de conclusão





# Planejamento do Experimento

- **Validade interna**
  - Está relacionada com a validade dentro do ambiente executado e a confiabilidade dos resultados. Ameaçam a relação causal.
- **Validade externa**
  - É o quão gerais são os resultados, pois as vezes, afirmamos que os resultados de um experimento são válidos fora do ambiente no qual o experimento foi executado

# Planejamento do Experimento

- **Validade de construção**
  - É uma questão de julgar se o tratamento reflete a construção da causa e o resultado fornece uma imagem verdadeira do efeito
- **Validade de conclusão**
  - É relacionada a habilidade de chegar a uma conclusão correta a respeito dos relacionamentos entre o tratamento e os resultados do experimento

# Seleção do Contexto

- **Seleção do ambiente em que o experimento vai ser executado**
- Para obter os resultados mais generalizáveis em um experimento, o escopo deve ser executado em projetos grandes de software reais, com equipe profissional
- O contexto do experimento pode ser caracterizado de acordo com quatro dimensões:
  - Off-line x On-line (i.e., “real”)
  - Aluno x Profissional
  - Problemas simulados vs. Problemas reais
  - Específico vs. Geral

# Exemplo de Seleção do Contexto

## 2.3 Seleção do contexto

O contexto pode ser caracterizado conforme quatro dimensões:

- o processo: on-line / off-line;
- os participantes: alunos / profissionais;
- realidade: o problema real / modelado;
- generalidade: específico / geral.

Nosso estudo supõe o processo off-line porque os alunos não estão sendo entrevistados durante todo o tempo do curso, mas em um certo instante. Os participantes são os alunos que estão realizando o curso. O estudo é modelado porque as competências dos alunos não são caracterizadas durante a resolução do problema real, mas utilizando as notas subjetivas. As competências dos alunos do certo curso são comparadas com as competências listadas no currículo da SBC, então, o contexto possui o caráter específico.

## Seleção do Contexto

- Uma situação comum em um experimento é que **algo existente é comparado a algo novo**
  - Ex.: um método de inspeção existente é comparado a um novo
- Existem dois problemas relacionados a este tipo de pesquisa experimental:
  - Qual é o método existente? Ele pode ter sido aplicado por algum período de tempo, mas raramente é bem documentado e não há aplicação consistente do método
  - Aprender um novo método pode influenciar a forma como o antigo é aplicado

***Essas e outras questões têm que ser levadas em conta durante o planejamento do experimento!***

## Formulação da Hipótese

- É uma teoria ou suposição que pode explicar um determinado comportamento ou fenômeno de interesse da pesquisa indicando a existência de um relacionamento esperado entre uma variável independente e uma outra dependente na pesquisa
- Tipos de hipóteses:
  - **Nula ( $H_0$ ):** declara que não há nenhum relacionamento estatisticamente significativo entre a causa e efeito
  - **Alternativa ( $H_1$ ):** contraditória à primeira
- O **objetivo do experimento** é rejeitar a hipótese nula a favor de uma ou algumas hipóteses alternativas

## Formulação da Hipótese

- Uma hipótese nula,  $H_0$ , **afirma que não há tendências reais**
  - As únicas razões para diferenças nas observações são coincidências
  - Esta é a hipótese de que o experimentador deseja rejeitar com a maior significância possível
- Existem vários testes estatísticos diferentes descritos na literatura que podem ser usados para avaliar o resultado de um experimento
  - Eles são todos baseados nas hipóteses formuladas antes que os testes estatísticos sejam escolhidos e realizados

## Formulação da Hipótese

- O teste de hipóteses envolve diferentes tipos de riscos
- Esses riscos são referidos na literatura como:
  - Type-I-Error (falso positivo)
    - Ocorre quando um teste estatístico indica um padrão ou relacionamento, mesmo se não houver um relacionamento real
    - A probabilidade de cometer um erro do tipo I pode ser expressa com  $P(\text{type-I-error}) = P(\text{reject } H_0 \mid H_0 \text{ true})$
    - Ex.:
      - O erro tipo I é a probabilidade de rejeitar  $H_0$ , embora os dois métodos, em média, encontrem o mesmo número de falhas



## Formulação da Hipótese

- Type-II-Error (falso negativo)
  - Ocorre um erro do tipo II quando um teste estatístico não indicou um padrão ou relacionamento, mesmo se que haja um relacionamento real
  - A probabilidade de cometer um erro do tipo II pode ser expressa como:
$$P(\text{type-II-error}) = P(\text{not reject } H_0 \mid H_0 \text{ false})$$
  - Ex.:
    - O erro tipo II é a probabilidade de não rejeitar  $H_0$ , embora os dois métodos em média tenham médias diferentes

## Formulação da Hipótese

- O tamanho do erro durante a verificação das hipóteses depende da **potência do teste estatístico**
- A potência do teste implica a probabilidade de que o teste vai encontrar o relacionamento se a hipótese nula é falsa
- A potência pode ser avaliada como:

$$\text{Potência} = P(H_0 \text{ é rejeitada} | H_0 \text{ é falsa}) = 1 - P(\text{type-II-error})$$

- O teste estatístico com a maior potência deve ser escolhido
  - Nota:
    - **O poder estatístico é influenciado pelo tamanho da amostra**

## Exemplo de Formulação da Hipótese

- **Hipótese nula**
  - Um novo método de inspeção encontra em média o mesmo número de falhas que o antigo.
  - $H_0: \mu_{old} = \mu_{new}$
- **Hipótese alternativa**
  - Um novo método de inspeção encontra em média mais falhas que o antigo.
  - $H_A: \mu_{old} < \mu_{new}$

# Exemplo de Formulação da Hipótese

## 2. PLANEJAMENTO

### 2.1 Definição das Hipóteses

**Hipótese nula (H0):** As competências oferecidas para os alunos do curso básico de ES são similares às competências de ES que o currículo mínimo da SBC considera fundamental para o desenvolvimento de software

$C_a$  – competências oferecidas para os alunos do curso básico de ES;

$C_c$  – competências de ES do currículo mínimo da SBC.

$$H0: C_c - (C_a \cap C_c) = \emptyset$$

**Hipótese alternativa (H1)** A lista de competências oferecidas para os alunos do curso básico de ES é diferente da lista de competências de ES que o currículo mínimo da SBC considera fundamental para o desenvolvimento de software.

$C_a$  – competências oferecidas para os alunos do curso básico de ES;

$C_c$  – competências de ES do currículo mínimo da SBC.

$$H1: C_c - (C_a \cap C_c) \neq \emptyset$$

## Seleção das Variáveis

- Antes que qualquer projeto possa começar, temos que escolher as variáveis
- Escolher as variáveis certas não é fácil e geralmente requer conhecimento do assunto
- As escolhas das variáveis independentes e dependentes geralmente são feitas simultaneamente ou em ordem reversa
- A escolha de variáveis independente **também inclui a escolha das escalas de medição, o intervalo para as variáveis e os níveis específicos em que os testes serão feitos**

## Seleção das Variáveis

- **Variáveis independentes** (ou fatores, quando controladas)
  - Referem-se à entrada do processo de experimentação, podendo ser controladas durante este processo
  - Exige conhecimento do domínio
  - Representam a causa que afeta o resultado do processo de experimentação
    - Quando é possível seu controle, os valores são chamados de "tratamentos"
    - Demais variáveis independentes são mantidas em um valor fixo

# Seleção das Variáveis

- **Variáveis dependentes**
  - Referem-se à saída do processo de experimentação, sendo afetadas durante o processo de experimentação
  - As hipóteses devem ser refinadas quando as variáveis dependentes forem escolhidas
  - Há, em geral, apenas uma variável dependente e, portanto, deve ser derivada diretamente da hipótese
  - Representam o efeito da combinação dos valores das variáveis independentes (incluindo os fatores)
    - Seus possíveis valores são chamados de "resultados"

## Exemplo de Seleção das Variáveis

- Em um experimento focado em comparar o impacto da POO e a abordagem orientada a funções na produtividade
  - Variável dependente
    - Produtividade
  - Variável independentes
    - Método de Desenvolvimento (**FATOR**)
      - **Tratamento 1:** POO
      - **Tratamento 2:** abordagem orientada a funções
    - Experiência dos participantes
    - Ferramenta de Suporte



# Exemplo de Seleção das Variáveis

## 2.5 Variáveis

Variável independente:

A lista de competências de ES do currículo mínimo da SBC.

Variáveis dependentes:

1. A similaridade entre competências oferecidas para os alunos do curso básico de ES e competências de ES do currículo mínimo da SBC.

Pode receber os valores:

Igual, quando todas as competências têm o valor  $PUA = \{1, X, X\}$  (métricas 5-8);

Diferente, quando todas as competências têm o valor  $PUA = \{0, X, X\}$  (métricas 1-4).

Similar, quando não se cumprem as condições de “Igual” e “Diferente”. O grau de similaridade pode ser avaliado como:

$$\{1, X, X\} / (\{0, X, X\} + \{1, X, X\}) * 100\%$$

2. A utilidade de competências similares. Mostra a parte útil das competências oferecidas pelo curso básico de ES:

Parte útil:  $\{1, 1, X\} / \{1, X, X\} * 100\%$

Parte inútil:  $\{1, 0, X\} / \{1, X, X\} * 100\%$

## Seleção dos Participantes

- A seleção é altamente relacionada com a generalização dos resultados
- Devem ser representativos da população
- A amostra pode ser probabilística ou não- probabilística (indica se a probabilidade de escolher um sujeito é conhecida ou não)
- O tamanho da amostra impacta os resultados quando se generaliza
  - Se há grande variabilidade na população, uma amostra maior é necessária
  - A forma como se analisará os dados coletados influencia a escolha do tamanho da amostra

# Seleção do Participante

- Probabilísticas

- Amostragem aleatória simples
  - Os indivíduos são selecionados de uma lista da população aleatoriamente
- Amostragem sistemática
  - O primeiro sujeito é selecionado da lista aleatoriamente e, em seguida, cada  $n$ : pessoa é selecionada da lista
- Amostragem aleatória estratificada
  - A população é dividida em vários grupos e a amostragem aleatória simples é então aplicada aos grupos

# Seleção do Participante

- Não-probabilísticas
  - Amostragem de conveniência (convenience sampling)
    - As pessoas mais próximas e mais convenientes são selecionadas como sujeitos
  - Amostragem de cota
    - Obtém indivíduos de vários elementos de uma população

# Exemplo de Seleção do Participante

## 2.4 Seleção dos indivíduos

Como participantes para o estudo se propõe utilizar os alunos de pós-graduação, graduação ou alunos de 2o grau da área de Engenharia de Software. Assume-se que esses indivíduos estão disponíveis para o estudo e a maioria deles desenvolve software pessoal.

Seria conveniente utilizar para o estudo os alunos que estão realizando alguns de cursos básicos na área de Engenharia de Software. Nesse caso, dependendo do tamanho da turma, é possível usar uma das técnicas (probabilística ou não-probabilística) para escolha dos indivíduos.

Supõe-se propor aos participantes o questionário que tem como objetivo caracterizar sua formação do ponto de vista acadêmico, experiência, tipo de curso entre outros para analisar os dados e reduzir o viés.

## Escolha do Tipo de Design

- Um experimento consiste em uma série de testes dos tratamentos
- No experimento, a série de testes deve ser cuidadosamente planejada
  - Um design de um experimento descreve como os testes são organizados e executados
- O design e a análise estatística estão intimamente relacionados
- A escolha do design afeta a análise e vice-versa

## Princípios Gerais do Design

- Ao projetar um experimento, muitos aspectos devem ser considerados e por isso foi criado os princípios gerais
- Os princípios gerais de design são **randomização**, **bloqueio** e **balanceamento**
- Muitos design de experimentos utilizam uma combinação destes princípios gerais
  - Exemplo
    - Uma empresa realizará um experimento para investigar o efeito sobre a confiabilidade de um programa ao usar design orientado a objetos em vez do padrão princípio de design da empresa. O experimento usará o programa A como experimento objeto

# Randomização

- Um dos princípios de design mais importantes é a randomização
  - Alguns autores também o chamam de **aleatorização**
- Todos os métodos estatísticos usados para analisar os dados requerem que as observações ser de variáveis aleatórias independentes
  - Aplica-se à escolha dos objetos, sujeitos e em que ordem os testes serão executados
- É usado para “levar para a média” o efeito de um fator que pode estar presente em caso contrário
  - Exemplos
    - Que programa escolher para inspeção?
    - Que analistas da empresa escolher para participar?



## Bloqueio

- Se o efeito do fator for conhecido e controlável, podemos usar uma técnica de design chamada bloqueio
  - Alguns autores também o chamam de grupamento
- Usualmente há algum fator (variável independente) que provavelmente tem um efeito sobre o resultado, mas não estamos interessados nesse efeito
- Em um grupo, o efeito indesejado do tratamento é o mesmo e podemos usar o efeito do tratamento dentro do grupo
  - Exemplo
    - Os sujeitos escolhidos têm diferentes experiências: alguns usaram OO antes e outros não, por isso dividimos os sujeitos em dois grupos

## Balanceamento

- Se atribuirmos os tratamentos de forma que cada tratamento tenha um número igual de assuntos, temos um design equilibrado
- O equilíbrio é desejável porque simplificam e fortalece a análise estatística dos dados, mas não é necessária
- Exemplo
  - O experimento usa um design equilibrado, o que significa que existe o mesmo número de pessoas em cada grupo (bloco)



## Tipos de Projeto (design) Padrão

- **Um fator com dois tratamentos**
  - Em experimentos deste tipo, quer-se comparar os dois tratamentos um com o outro
  - O mais comum é comparar as médias das variáveis dependentes para cada tratamento
  - **Exemplo:** Um novo método de projeto produz software com mais qualidade do que o método usado anteriormente
    - Fator: O método de projeto
    - Tratamentos: O método novo e o velho.
    - Variável dependente: Número de falhas encontradas durante o desenvolvimento

## 5.1

# Tipos de Projeto (design) Padrão

- Um fator com dois tratamentos
  - Possíveis Projetos de Experimento
    - Completely randomized design
      - Cada participante usa 1 tratamento em 1 objeto
    - Paired comparison design (crossover design)
      - Cada participante usa os 2 tratamentos em 1 objeto

| Subjects | Treatment 1 | Treatment 2 |
|----------|-------------|-------------|
| 1        | X           |             |
| 2        |             | X           |
| 3        |             | X           |
| 4        | X           |             |
| 5        |             | X           |
| 6        | X           |             |

Completely randomized design

| Subjects | Treatment 1 | Treatment 2 |
|----------|-------------|-------------|
| 1        | 2           | 1           |
| 2        | 1           | 2           |
| 3        | 2           | 1           |
| 4        | 2           | 1           |
| 5        | 1           | 2           |
| 6        | 1           | 2           |

Paired comparison design

## Tipos de Projeto (design) Padrão

- **Um fator com mais de dois tratamentos**
  - O objetivo é comparar um tratamento com os outros
  - **Exemplo:** Pesquisar a qualidade do software quando se usa diferentes linguagens de programação
    - Fator: linguagens de programação
    - Tratamentos: C, C++ e Java

## 5.1

# Tipos de Projeto (design) Padrão

- Um fator com mais de dois tratamentos
  - Exemplos de design
    - Completely randomized design
    - Randomized complete block design

| Subjects | Treatment 1 | Treatment 2 | Treatment 3 |
|----------|-------------|-------------|-------------|
| 1        |             | X           |             |
| 2        |             |             | X           |
| 3        | X           |             |             |
| 4        | X           |             |             |
| 5        |             | X           |             |
| 6        |             |             |             |

| Subjects | Treatment 1 | Treatment 2 | Treatment 3 |
|----------|-------------|-------------|-------------|
| 1        | 1           | 3           | 2           |
| 2        | 3           | 1           | 2           |
| 3        | 2           | 3           | 1           |
| 4        | 2           | 1           | 3           |
| 5        | 3           | 2           | 1           |
| 6        | 1           | 2           | 3           |

## Tipos de Projeto (design) Padrão

- Dois fatores

- O experimento fica mais complexo quando aumentamos de um fator para dois
- A única hipótese para os experimentos com um fator será dividida em três hipóteses:
  - Uma hipótese para o efeito de um dos fatores
  - Uma para o outro
  - Um para a interação entre os dois fatores
- Usamos as seguintes notações:

|                    |  |
|--------------------|--|
| $\tau_i$           | The effect of treatment $i$ on factor A.                       |
| $\beta_j$          | The effect of treatment $j$ on factor B.                       |
| $(\tau\beta)_{ij}$ | The effect of the interaction between $\tau_i$ and $\beta_j$ . |

## 5.1

## Tipos de Projeto (design) Padrão

- **Dois fatores com dois tratamentos**
  - Atribui-se aleatoriamente os sujeitos a cada combinação de tratamento
  - **Exemplo:** Um experimento para analisar a facilidade de entendimento de um documento de projeto usando projeto estruturado e projeto OO e dois documentos de requisitos, um “bom” e um “ruim”
  - 2\*2 Factorial Design

---

|          |              | Factor A     |              |
|----------|--------------|--------------|--------------|
|          |              | Treatment A1 | Treatment A2 |
| Factor B | Treatment B1 | Subject 4, 6 | Subject 1, 7 |
|          | Treatment B2 | Subject 2, 3 | Subject 5, 8 |

---



## Tipos de Projeto (design) Padrão

- Mais do que dois fatores, cada um com dois tratamentos
  - Em muitos casos, o experimento deve considerar mais do que dois fatores
  - O efeito na variável dependente pode, portanto, ser dependente, não apenas em cada fator separadamente, mas também nas interações entre os fatores
  - Este tipo de design é chamados de experimentos fatoriais

## 5.1

## Tipos de Projeto (design) Padrão

- Mais do que dois fatores, cada um com dois tratamentos
  - Exemplo

**Table 8.7** Example of a  $2^3$  factorial design

| Factor A | Factor B | Factor C | Subjects |
|----------|----------|----------|----------|
| A1       | B1       | C1       | 2, 3     |
| A2       | B1       | C1       | 1, 13    |
| A1       | B2       | C1       | 5, 6     |
| A2       | B2       | C1       | 10, 16   |
| A1       | B1       | C2       | 7, 15    |
| A2       | B1       | C2       | 8, 11    |
| A1       | B2       | C2       | 4, 9     |
| A2       | B2       | C2       | 12, 14   |

## Instrumentação

- Um dos passos mais trabalhosos de um experimento é o desenvolvimento da instrumentação necessária à sua realização
- Os instrumentos são escolhidos no planejamento, antes da execução eles devem ser criados para o experimento específico
- Os instrumentos para um experimento são de três tipos, **objetos, diretrizes e instrumentos de medição**
- O objetivo geral da instrumentação é fornecer meios para realizar o experimento e monitorá-lo, sem afetar o controle deste experimento

## Instrumentação

- Os resultados do experimento devem ser os mesmos, independentemente de como foi realizada a instrumentação
  - Se a instrumentação afetar o resultado do experimento, os resultados são inválidos
- Objetos de experimento podem ser, por exemplo, especificações ou documentos de código
- Ao planejar um experimento, é importante escolher objetos que são apropriado
  - Por exemplo, em um experimento de inspeção, o número de falhas deve ser conhecido nos objetos de inspeção. Isso pode ser conseguido semeando falhas ou por usando um documento com um número conhecido de falhas

## Instrumentação

- São necessárias diretrizes para guiar os participantes do experimento
- Diretrizes incluem descrições de processos e listas de verificação
- Se métodos diferentes são comparados no experimento, as diretrizes para os métodos devem ser preparadas para o experimento
- Além das orientações, os participantes também precisam de treinamento nos métodos a serem usados



## Instrumentação

- As medições em um experimento são realizadas por meio da coleta de dados
  - Documento de consentimento
  - Formulários de caracterização de participantes
  - Formulários de coleta de dados
  - Formulários de feedback
- Os dados são geralmente coletados por meio de formulários manuais ou em entrevistas
- Para validar as perguntas dos formulários é necessário que algumas pessoas com experiência semelhante e habilidades participem do experimento

# Exemplo de Instrumentação

## 2.2 Descrição da instrumentação

Para cada competência que o currículo mínimo da SBC considera fundamental para o desenvolvimento de software oferecer escolha:

| Presença da competência (P)  | Utilidade da competência (U)  | Adequação do nível de detalhamento da competência (A)   |
|--|---|---|
| 1. Não é oferecida pelo curso e não gostaria de receber.<br>2. Não oferecida, mas gostaria de receber.<br>3. Oferecida, parcialmente.<br>4. Oferecida. | 1. Não é útil.<br>2. Provavelmente é útil, mas ainda não apliquei.<br>3. É útil e já apliquei em diferentes projetos. | 1. O detalhamento deve ser aumentado.<br>2. O detalhamento não precisa ser modificado.<br>3. O detalhamento deve ser diminuído. |

Para cada competência aplicar teste estatístico Chi-2 para definir:

se pode considerar que essa competência é fornecida;

se pode considerar que essa competência é útil.

se pode considerar que o detalhamento da competência não precisa de modificação.

**Resultado:** N competências com valores (P;U;A)

# Validade do Experimento

- Todo os experimento têm riscos à validade de seus resultados
  - Quão válidos são eles?
- O projetista do experimento tem a obrigação de identificar e discutir esses riscos
- As principais categorias de riscos são:
  - Validade da conclusão
  - Validade interna
  - Validade externa
  - Validade da construção





## **Riscos: Validade da Conclusão**

- A validade da conclusão está preocupada com a relação entre o tratamento e o resultado
- Trata-se da habilidade de chegar a uma conclusão correta a respeito dos relacionamentos entre o tratamento e o resultado do experimento
- A validade da conclusão é algumas vezes referida como conclusão estatística válida e tem sua contrapartida na confiabilidade para a análise qualitativa
- Ameaças à validade da conclusão estão preocupadas com questões que afetam a capacidade de tirar a conclusão correta sobre as relações entre o tratamento e o resultado de um experimento

## Riscos: Validade Interna

- Se for observada uma relação entre o tratamento e o resultado, devemos ter certeza de que é uma relação causal, e que não é resultado de um fator sobre o qual não temos controle ou não medimos
  - O tratamento causa o resultado (o efeito)
- Ameaças à validade interna dizem respeito a questões que podem indicar uma relação causal, embora não haja nenhum
- Fatores que impactam na validade interna podem fazer o experimento mostrar um comportamento que não é devido ao tratamento, mas ao fator perturbador

## Riscos: Validade Externa

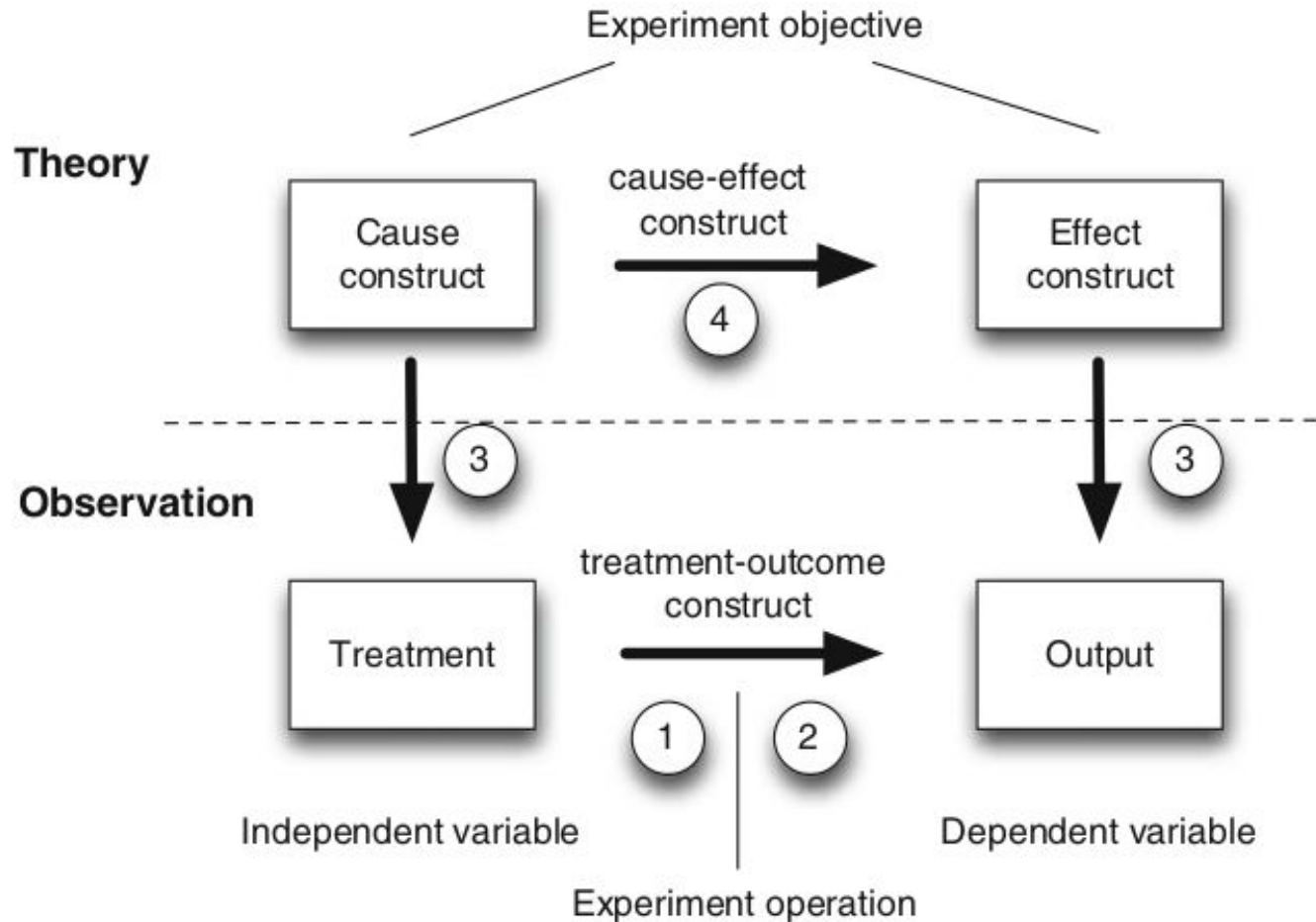
- A validade externa está relacionada à generalização
- Se há uma relação causal entre a construção da causa e o efeito
  - O resultado do estudo pode ser generalizado fora do escopo de nosso estudo?
- Ameaças à validade externa dizem respeito à capacidade de generalizar os resultados do experimento fora do ambiente da experiência
- A validade externa não é só afetada pelo design do experimento escolhido, como também pelos objetos e participantes
- Há três riscos principais: ter participantes errados como sujeitos, conduzir o experimento no ambiente errado e executar o experimento com um tempo que afeta os resultados

## Riscos: Validade da Construção

- Esta validade está preocupada com a relação entre a teoria e observação
- Se a relação entre causa e efeito é causal, devemos garantir duas coisas:
  1. Que o tratamento reflete bem a construção da causa
    - Esquerda da Fig. 8.2
  2. Que o resultado reflete a construção do efeito
    - Direita da Fig. 8.2



# Riscos: Validade da Construção



**Fig. 8.2** Experiment principles (Adapted from Trochim [171])

## Riscos: Validade da Construção

- Ameaças à validade de construção referem-se à extensão em que o cenário do experimento realmente reflete o construto em estudo
  - Por exemplo, o número de disciplinas ofertados no curso de computação pode ser uma medida ruim da experiência dos alunos em linguagem de programação, ou seja, tem validade de construção pobre
  - O número de anos de prático pode ser uma medida melhor



## Validade do Experimento

- A avaliação tem o objetivo de verificar se o projeto experimental é consistente e pode realmente ser executado
- Na avaliação de validade, cada um dos os itens são verificados para ver se há alguma ameaça
- Se houver alguma ameaça, eles têm que ser abordado ou aceito, já que às vezes alguma ameaça à validade tem de ser aceita
- Pode até ser impossível realizar um experimento sem certas ameaças e portanto, eles devem ser aceitos e, em seguida, tratados ao interpretar os resultados

## Validade do Experimento

- Esta etapa deve avaliar o projeto experimental e seus artefatos
- Os pontos mais críticos a serem avaliados em um experimento são:
  - O design e o processo subjacente
  - Tempo destinado a cada atividade
  - Treinamento nas técnicas e artefatos a serem utilizados





# Exemplo de Validade do Experimento

## 2.7 Validade

**Validade interna:** como mencionado na parte “Seleção dos indivíduos” para o estudo se propõe a utilizar os alunos da área de Engenharia de Software, que geralmente costumam desenvolver software pessoal. Assim, assume-se que eles são representativos para a população dos desenvolvedores de software pessoal.

Além disso, para redução da influência dos fatores que não são interesse do nosso estudo e, portanto, para aumento da validade interna do estudo supõe-se utilizar os dados do questionário para divisão dos participantes em grupos conforme a suas características individuais.

**Validade de conclusão:** para receber os valores da presença, utilidade e conformidade o teste binomial será utilizado. A verificação de hipótese será feita por meio de simples demonstração de presença ou não de competências nas listas que representam os variáveis independentes.

**Validade de construção:** esse estudo está caracterizado pela conformidade das competências listadas no currículo mínimo da SBC com as competências reais necessárias para desenvolvimento de software pessoal. O currículo mínimo da SBC representa a lista de competências que os especialistas na área de computação devem possuir para mostrar o desempenho adequado do ponto de vista da SBC. As competências, que têm o maior relacionamento com desenvolvimento de software pessoal do ponto de

## Exemplo de Validade do Experimento

vista dos pesquisadores, foram escolhidas do conjunto total de competências do currículo mínimo da SBC.

**Validade externa:** como foi mencionado nas partes “Seleção dos indivíduos” e “Validade interna” os participantes do estudo em geral podem ser considerados representativos para a população dos desenvolvedores de software pessoal. Para avaliação do nível de envolvimento no processo de desenvolvimento de software pessoal os dados do questionário conforme a experiência dos participantes podem ser analisados.

Os materiais utilizados no estudo podem ser considerados representativos e “em tempo” para o problema sob análise, porque se compõem das competências do currículo mínimo atual da SBC relacionadas ao desenvolvimento de software.

As características temporárias não devem ser o problema, porque os materiais dão a possibilidade de conduzir o estudo durante a meia-hora. Além disso, o estudo será realizado quando os participantes estiverem realizando algum curso e, geralmente, estiverem envolvidos no desenvolvimento de software.

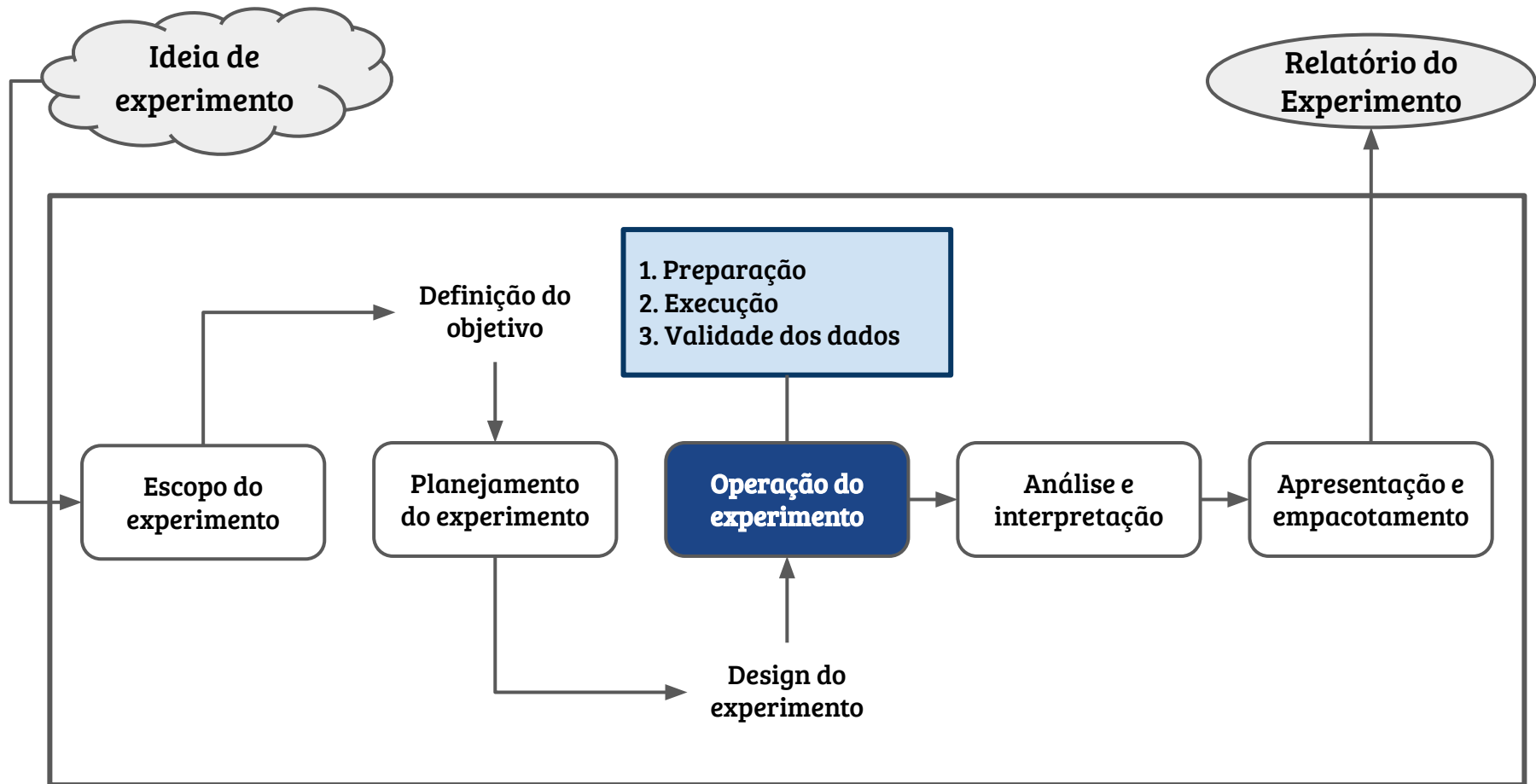
# Planejamento do Experimento

- A avaliação tem o objetivo de verificar se o projeto experimental é consistente e pode realmente ser executado
- Esta etapa deve avaliar o projeto experimental e seus artefatos
- Os pontos mais críticos a serem avaliados em um experimento são o design e o processo subjacente:
  - Tempo destinado a cada atividade
  - Treinamento nas técnicas e artefatos a serem utilizados

# Planejamento do Experimento

- Se possível, execute um **estudo-piloto**
  - Execute uma versão simplificada do experimento para avaliar seu projeto (design) e artefatos
    - Número de participantes
    - Volume de tarefas a serem executadas
    - Tamanho dos artefatos utilizados
  - **Resultados do estudo piloto não devem ser usados na análise de dados e teste de hipóteses**
  - O estudo piloto deve ser usado para avaliar a razoabilidade do projeto experimental, seus artefatos e processos
  - Redesenhe o experimento e melhore os artefatos conforme necessário

# Processo do Experimento



# Operação do Experimento

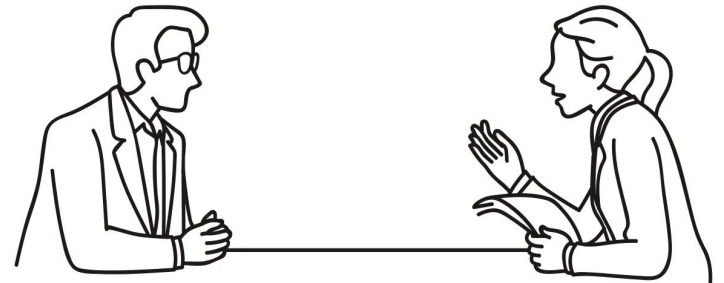
- A operação do experimento diz respeito a execução do experimento e coleta de dados
- Na fase operacional de um experimento, os tratamentos são aplicados aos sujeitos
  - Isso significa que é neste momento onde o pesquisador realmente encontra os participantes/voluntários
  - Antes da execução do experimento pode ser realizado um briefing antes que os participantes se comprometam em participar do experimento e depois do experimento quando os resultados do experimento podem ser apresentados

# Operação do Experimento

- A etapa de **operação** realiza a preparação onde os participantes são escolhidos e/ou formulários etc. são preparados, e a execução é aplicada de acordo com diferentes tratamentos e os dados são coletados e validados
- A operação consiste em três etapas:
  - **Preparação**
    - A preocupação está em preparar os assuntos, bem como o material necessário para a execução do experimento, por exemplo, formulários de coleta de dados
    - É também a fase em que os participantes devem ser informados sobre a intenção do experimento, pois devemos ter o consentimento do participante

# Operação do Experimento

- **Execução**
  - A principal preocupação é garantir que o experimento seja conduzido de acordo com o plano e design do experimento, o que inclui a coleta de dados
- **Validação de dados**
  - É onde nos certificamos de que os dados realmente coletados estão corretos e que fornecem uma imagem válida do experimento





# Preparação

- Antes que o experimento seja realmente executado, existem alguns preparativos que devem ser feitos
- Quanto melhor essa preparação for realizada, mais fácil será a execução do experimento
- Existem dois aspectos importantes na preparação:
  - Selecionar e informar **os participantes do experimento**
  - **Preparar materiais**, como formulários e ferramentas



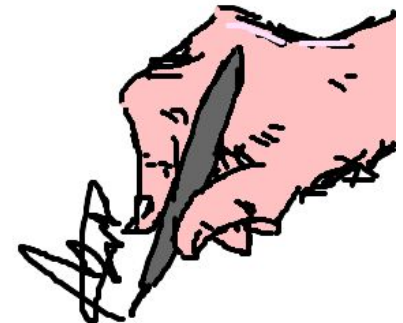
# Preparação

- Antes que um experimento possa ser iniciado, as pessoas que desejam ser participantes de um experimento têm que ser encontradas
- É essencial que as pessoas **estejam motivadas e dispostas a participar** ao longo de todo o experimento
- Quando as pessoas certas são encontradas é necessário convencê-las a participar do experimento
- **Vários aspectos éticos** devem ser considerados quando as pessoas estão participando do experimento



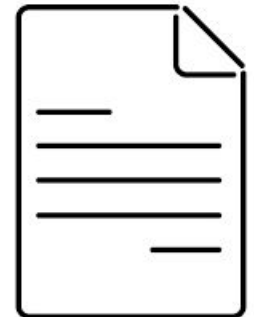
# Preparação

- **Obtenha consentimento dos participantes**
  - Os participantes devem concordar com os objetivos da pesquisa
  - Se os participantes não sabem a intenção da pesquisa ou a pesquisa não cumpre o que eles pensaram que deveriam fazer quando concordaram em participar, há um risco deles não realizem o experimento de acordo com os objetivos e habilidades pessoal
  - Isso pode fazer com que os dados se tornem inválidos



# Preparação

- Como obter o consentimento dos participantes de forma correta?
  - É importante descrever como o resultado do experimento será usado e publicado
  - Deve deixar claro, para os participantes, que eles são livres para se retirar do experimento
  - Se os participantes forem afetados de alguma forma pelo experimento isso afetará a validade da pesquisa!



# Preparação

- **Resultados sensíveis**
  - Se os resultados obtidos no experimento são sensíveis para os participantes, é importante assegurar aos participantes que **os resultados de seus desempenho no experimento serão mantidos em sigilo**
  - As vezes é difícil julgar se o resultado é sensível ou não, mas geralmente pode-se dizer que se o resultado teria um significado para os participantes fora do experimento, ele está de alguma forma sensível
  - Por exemplo,
    - se o experimento mede a produtividade de um programador, o resultado indicaria o quão hábil o programador é e este resultado seria sensível

# Preparação

- Incentivos

- Uma maneira de atrair pessoas para um experimento é oferecer algum tipo de incentivo
- O valor dele, entretanto, não deve ser muito grande, pois isso pode causar atrair as pessoas apenas para receber o incentivo
  - Isso não motivaria que as pessoas participem seriamente do experimento



# Preparação

- **Apresentação do Experimento**
  - Significa **revelar todos os detalhes do experimento** tão abertamente quanto possível para os participantes
  - A única **exceção** para não haver divulgação dos resultados; é se a pesquisa tratar de aspectos que são insignificante para os participantes e que não afetam a sua vontade de participar do experimento



# Preparação

- Antes que o experimento possa ser executado, **todos os instrumentos de experimento devem estar prontos**, tais como os objetos do experimento, diretrizes para o experimento, formulários e ferramentas de medição
- Os instrumentos necessários são determinados pelo design do experimento e o método que será utilizado para a coleta de dados
  - Quando for elaborar formulários, eles devem ser pessoais ou os participantes devem preencher anonimamente
    - Formulários pessoais ajudam na realização de diferentes perguntas para diferentes participantes
    - O anonimato significa que não há nenhuma possibilidade de entrar em contato com o participante se algo for preenchido de forma pouco clara



## Execução

- O experimento pode ser executado de várias maneiras diferentes
- Alguns experimentos, podem ser realizados quando todos os participantes estiverem reunidos
  - A vantagem disso é que o resultado da coleta de dados **pode ser obtido diretamente na reunião** e não há necessidade de entrar em contato com os participantes e, posteriormente, solicitar seus respectivos resultados
  - Outra vantagem é que o **pesquisador está presente** durante a reunião e se houver dúvidas sobre a pesquisa ele pode retirá-las imediatamente

## Execução

- Alguns experimentos são executados durante um período de tempo muito longo, e pode ser impossível para o pesquisador participar de todos os detalhes do experimento e a coleta de dados
- Este é, **por exemplo**, o caso quando o experimento é realizada em relação a um ou vários grandes projetos, onde diferentes métodos para o desenvolvimento de software são avaliados

# Execução

- Coleta de dados
  - Os dados podem ser coletados manualmente pelos participantes que preenchem os formulários, em entrevistas ou automaticamente por ferramentas
  - Uma vantagem do uso de **formulários**, é que não requer tanto esforço para o pesquisador
    - Uma desvantagem é que não há possibilidade do pesquisador tirar dúvidas diretamente sobre inconsistências, incertezas e falhas nos formulários etc.
  - Uma vantagem das **entrevistas** é que o pesquisador tem a possibilidade de se comunicar melhor com os participantes, mas, claro, que requer mais esforço do pesquisador

# Execução

- **Ambiente do Experimento**
  - Se um experimento é realizado dentro de um projeto de desenvolvimento, o experimento não deve afetar o projeto mais do que o necessário
  - A razão para realizar o experimento dentro do projeto poderia ser ver os efeitos de diferentes tratamentos em um ambiente como o do projeto
    - Se o ambiente do projeto é alterado demais por causa do experimento, o efeito será perdido

## Validade dos Dados

- Depois de coletado os dados é **preciso verificar se eles foram coletado corretamente**
  - Aspectos como falta de compreensão sobre o preenchimento dos formulários afetam a coleta de dados
  - Outra fonte de erro é que **alguns participantes podem não ter participado de um experimento de forma séria** e alguns dados, portanto, devem ser removidos antes do análise
- É importante revisar se o experimento foi realmente conduzido no maneira que foi planejado
  - Um mal entendido no tratamento dos dados pode torná-los inválidos

## Validade dos Dados

- Uma forma de verificar se os participantes não compreenderam mal as intenções do pesquisador é aplicação de um **seminário**, ou de alguma outra forma apresentar os resultados da coleta de dados
  - Isso dará aos participantes a possibilidade de refletir sobre os resultados com os quais eles não concordam



# Exemplo de Operação do Experimento

- Na preparação primeiramente devem ser identificados os participantes
  - Exemplo, **alunos de Ph.D. e M.Sc.** são convidados como participantes, depois de ter sido validado como um potencial participante
- É importante convencê-los a participar e firmar o compromisso de participar do experimento até o final
- Após ter um compromisso inicial, o consentimento deve ser assegurado aos participantes
- Recomenda-se que os **formulários de consentimento** sejam usados

# Exemplo de Operação do Experimento

- A atribuição dos participantes ao tratamento deve ser feita usando um procedimento de **randomização**
- Se o projeto inclui um **fator de bloqueio** (tipo de aluno), os participantes devem ser divididos de acordo com esse fator e, em seguida, atribuídos aleatoriamente a aos tratamentos dentro de cada grupo de bloqueio
- Se um projeto **balanceado** for escolhido, a seleção deve terminar no mesmo número de participantes para cada grupo



# Exemplo de Operação do Experimento

- A próxima etapa é garantir que a **infraestrutura** necessária esteja pronta
  - Isso inclui ter uma sala adequada reservada, por exemplo, fornecendo distância suficiente entre os participantes
- Cópias de todos os documentos e formulários devem estar disponíveis para todos os participantes
- Deve ser dado um tempo para o recolhimento dos formulários
  - É necessário um relógio na sala
  - Não se pode assumir que todos têm acesso ao seu próprio relógio

# Exemplo de Operação do Experimento

- Durante a execução, é importante garantir que os participantes estejam adequadamente distribuídas pela sala
- Se possível, o experimento pode ser executado com todos os participantes ao mesmo tempo
  - Isso também significa que é fácil fornecer suporte para qualquer dúvida que possa surgir
- Os dados devem ser coletados via formulários à mão ou por meio de um computador, a preparação deve ser feita em conformidade

# Exemplo de Operação do Experimento

- Finalmente na etapa de validação, os dados precisam ser validados
- Pode ser o caso de que um ou vários participantes saiam do experimento muito cedo e os formulários preenchidos devem ser verificado cuidadosamente para garantir que foram respondidos corretamente
- Além disso, deve-se verificar se todos compreenderam como preencher o formulário de forma correta
  - Se este não for o caso, pode ser que os dados de um ou vários participantes devam ser removidos

# Exemplo de Operação do Experimento

## 3. OPERAÇÃO

### 3.1 Questionário do Perfil do Participante

Marque um X sobre a opção que melhor representar sua formação.

FORMAÇÃO

Instituição.

☐ Pública

☐ Particular

Tipo de curso.

☐ Engenharia.

☐ Informática / Ciência da computação.

☐ Matemática

☐ Outros.

Acadêmica:

☐ Ensino Médio Técnico

☐ Universitária

☐ Pós-graduação

# Exemplo de Operação do Experimento

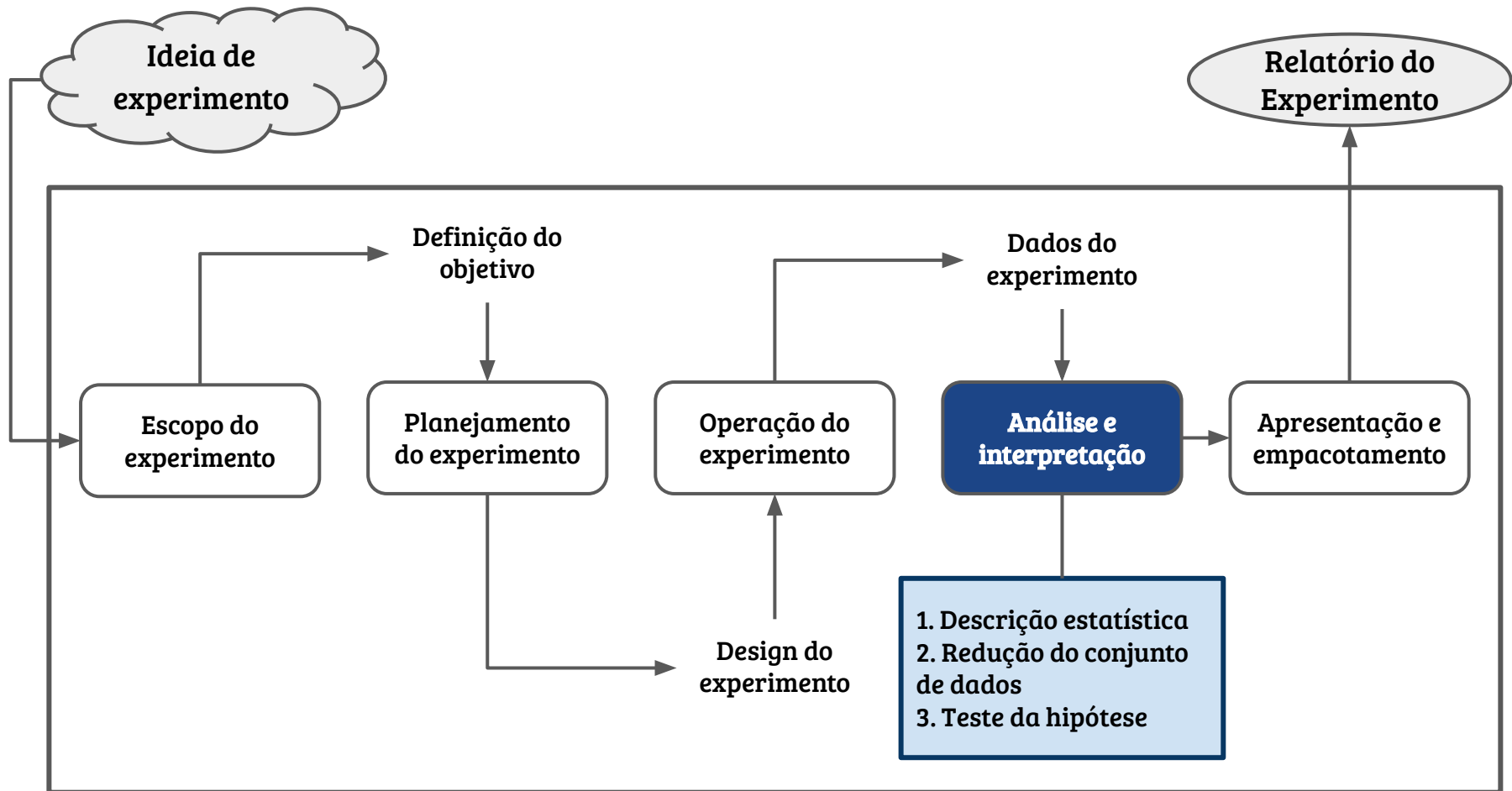
## 3.2 Questionário de Competências

Sob o ponto de vista de *desenvolvimento de software pessoal* e considerando o tipo de curso que você indicou acima, por favor, avalie e marque as colunas correspondentes segundo as escalas abaixo, a presença, utilidade e adequação quanto ao detalhamento que lhe foi apresentado durante o curso, das competências listadas no questionário:

| Presença |  | Utilidade |   | Adequação do nível de detalhamento |  |
|----------|--|-----------|---|------------------------------------|--|
| 1        | Não oferecida durante os cursos e não gostaria de receber. | 1         | Não se demonstrou útil.                       | 1                                  | O detalhamento deve ser aumentado.         |
| 2        | Não oferecida, mas gostaria de receber.                    | 2         | Provavelmente é útil, mas ainda não apliquei. | 2                                  | O detalhamento não precisa ser modificado. |
| 3        | Oferecida, parcialmente.                                   | 3         | É útil e apliquei em muitos projetos.         | 3                                  | O detalhamento deve ser diminuído.         |
| 4        | Oferecida.   |           |   |                                    |  |

| N | Competência                    | Descrição  | Presença |   |   |   | Utilidade |   |   | Adequação |   |   |
|---|--------------------------------|--|----------|---|---|---|-----------|---|---|-----------|---|---|
|   |                                |  | 1        | 2 | 3 | 4 | 1         | 2 | 3 | 1         | 2 | 3 |
|   | <b>Processo</b>                |  |          |   |   |   |           |   |   |           |   |   |
| 1 | <u>Qualidade do produto</u>    | é capaz de avaliar a qualidade de um produto de software utilizando-se de normas apropriadas; é capaz de empregar os modelos de medição. |          |   |   |   |           |   |   |           |   |   |
| 2 | <u>Interface homem-máquina</u> | é capaz de avaliar e propor mudanças na interface de um produto de software, sob o ângulo da usabilidade.                                |          |   |   |   |           |   |   |           |   |   |

# Processo do Experimento



# Análise e Interpretação

- A análise e interpretação está dividida em três etapas:
  1. Na **descrição estatística**, os dados são caracterizados por meio de estatística descritiva, para visualizarmos tendência central, dispersão, etc.
  2. Na **redução do conjunto de dados**, pontos de dados anormais ou falsos são excluídos, reduzindo assim o conjunto de dados a um conjunto de pontos de dados válidos
  3. No **teste de hipótese**, os dados são analisados por teste de hipóteses, onde as hipóteses do experimento são avaliados estatisticamente, em um determinado nível de significância

# Descrição Estatística

- A estatística descritiva pode ser usada para descrever e apresentar graficamente aspectos interessantes do conjunto de dados
  - Que incluem medidas que indicam, por exemplo, **onde em alguma escala os dados estão posicionados e quanto concentrado ou espalhado é o conjunto de dados**
- O objetivo da estatística descritiva **é visualizar como o conjunto de dados é distribuído**
- Estatísticas descritivas podem ser usadas antes de realizar o teste de hipótese, a fim de melhorar a compreensão da natureza do dados e para identificar pontos de dados anormais ou falsos (os chamados outliers)
- **Depende da escala dos dados**



# Descrição Estatística

- **Escala Nominal**

- Usa números como rótulos para identificar e classificar objetivos, indivíduos ou objetos
  - Exemplo: Estado Civil, Matrícula de Automóvel

- **Escala Ordinal**

- Estabelece uma ordenação de dados quantitativos sem estabelecer o grau de variação entre eles.
- A ordem das variáveis é relativa
  - Exemplo: Respostas do tipo 'muito satisfeito' até 'muito insatisfeito'

# Descrição Estatística

- **Escala de intervalo**
  - A escala de intervalo é definida como uma escala de medição de dados quantitativos na qual a diferença entre duas variáveis é medida
    - Exemplo: °C

Considerando sua experiência com a gente, você recomendaria nossos serviços aos seus amigos?

|            |   |   |   |   |   |   |   |   |   |          |
|------------|---|---|---|---|---|---|---|---|---|----------|
| 0          | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10       |
| Improvável |   |   |   |   |   |   |   |   |   | Provável |

# Descrição Estatística

- **Escala de razão**
  - Essa escala é caracterizada por um ponto zero absoluto, o que significa que **não há valor numérico negativo**
    - Ou seja, 40 é 2x maior que 20
    - Exemplo: Idade, preço

Quantos cafés você toma por dia?

- ☐ 1-2
- ☐ 2-3
- ☐ 3-4
- ☐ Mais de 4

Qual o peso do seu cachorro?

- ☐ Até 2kg
- ☐ 2kg-6kg
- ☐ 6kg-10kg
- ☐ Mais de 10kg

# Descrição Estatística

- **Medidas de tendência central**
  - Medidas de tendência central, como **média, mediana e moda**, indicam um “meio” de um conjunto de dados
- Ao descrever as medidas de tendência central, assumimos que temos  $n$  pontos de dados  $x_1 \dots x_n$  mostrados de alguma variável estocástica

# Descrição Estatística

- **Medidas de tendência central**

- A **média** (aritmética), denotado  $\bar{x}$ , é calculado como:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

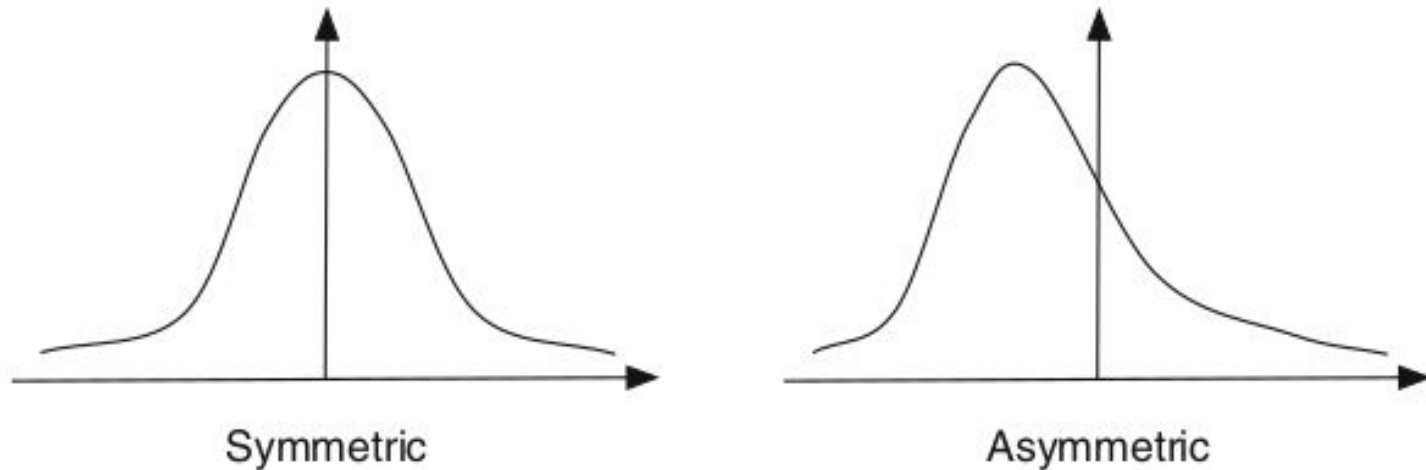
- O valor médio é significativo para as escalas de intervalo e razão
  - Por exemplo, podemos calcular a **média** para o conjunto de dados (1, 1, 2, 4) que resulta em  $\bar{x} = 2.0$

# Descrição Estatística

- **Medidas de tendência central**
  - A **mediana** é o valor do meio
    - Podemos calcular a **mediana** para o conjunto de dados (1, 1, 2, 4) que resulta em  $\tilde{x} = 1.5$
  - A **moda** representa a amostra de ocorrência mais comum
    - A moda é calculada contando o número de amostras para cada valor único e selecionando o valor com a contagem mais alta
      - O valor da moda é significativo para as escalas nominal, ordinal, intervalo e razão
      - Por exemplo, podemos calcular a moda para o conjunto de dados (1, 1, 2, 4) que resulta na moda de valor 1

# Descrição Estatística

- Medidas de tendência central



**Fig. 10.2** A symmetric distribution has the same values of mean, median, and mode, while they may differ if the distribution is asymmetric

# Descrição Estatística

- **Medidas de dispersão**

- As medidas de tendência central não transmitem informações sobre a dispersão do conjunto de dados
- Para calcular a dispersão é necessário medir o nível de variação da central, ou seja, **para ver se os dados estão espalhados ou concentrados**
- A variância é a média da distância quadrada da média da amostra
- Desvio padrão é a raiz quadrada da variância
  - quanto menor o desvio padrão, mais homogênea é a amostra (i.e., indica que os dados estão próximos da média )



# Descrição Estatística

- **Medidas de dependência**

- Quando o conjunto de dados consiste em amostras relacionadas em pares  $(x_i, y_i)$  de duas variáveis estocásticas,  $X$  e  $Y$ , é interessante examinar a dependência entre estas variáveis
- Uma medida de dependência é um parâmetro associado a um par de variáveis aleatórias que codifica em seu valor a intensidade da dependência estatística entre as variáveis
- Uma medida de dependência diz quanta informação se obtém a respeito de uma variável quando o valor de outra variável é conhecido
- Exemplo:
  - **Coeficiente de Correlação**

# Descrição Estatística

A distribuição normal tem destaque na ciência da computação, pois muitas variáveis típicas dos processos da área apresentam distribuição normal. Mas vale ressaltar que existem muitas outras métricas estatísticas que podem ser usadas.

**Não se limite!**



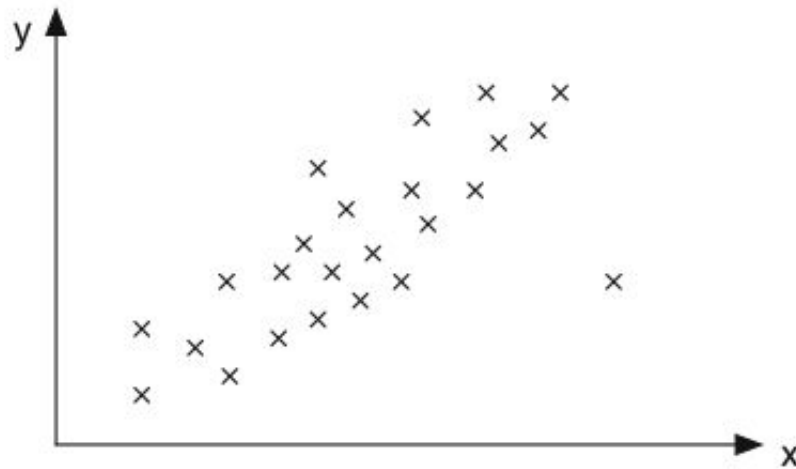
# Descrição Estatística

- **Visualização gráfica**
  - Os dados estatísticos precisam ser ilustrados para haver maior compreensão e os gráficos são capazes de fornecer uma boa visão geral do conjunto de dados
    - O **gráfico de dispersão** (scatter plot) é bom para avaliar dependências entre variáveis
    - O **gráfico de caixa** (box plot) é bom para visualizar a dispersão e distorção das amostras

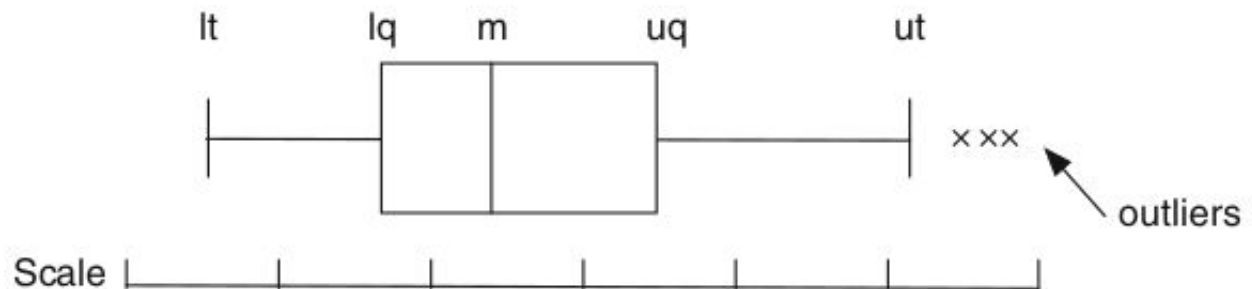
# Descrição Estatística

- O **histograma** pode ser usado para dar uma visão geral da densidade de distribuição do amostras de uma variável
- Um **gráfico de pizza**, ilustra a frequência relativa dos dados, através de um número específico de classes distintas, com ângulos proporcionais à frequência relativa

# Exemplo de Visualização Gráfica



**Fig. 10.3** A scatter plot



**Fig. 10.4** A box plot

## Redução do Conjunto de Dados

- Os resultados da fase de análise e interpretação oferecem as conclusões sobre a possibilidade da rejeição da hipótese nula usando a estatística descritiva, a redução do conjunto de dados, e a verificação das hipóteses
- Os aspectos mais importantes são:
  - Eliminar dados fora da distribuição normal (outliers)
  - Escolher o teste estatístico apropriado
  - Explicar os resultados considerando os aspectos da validade
  - Realizar a análise custo-benefício
  - Interpretar corretamente os resultados negativos

## Redução do Conjunto de Dados

- Se os dados sobre os quais os métodos estatísticos aplicados não representam o que achamos que representa, então as conclusões que tiramos dos resultados dos métodos não estão corretos
- Existem diferentes métodos para identificar outliers
  - scatter plots
  - box plot
- Note que essa etapa tem relação com a **validação dos dados**
  - **Um outlier pode significar alguém que não participou seriamente do experimento**

## Teste da Hipótese

- O objetivo do teste de hipótese é ver se é possível rejeitar um determinada hipótese nula ( $H_0$ ), com base em uma amostra de alguma distribuição estatística
- Configurar  $H_0$ , significa formular a distribuição e atribuir um valor ao parâmetro, que será testado



## Teste da Hipótese Tipo t

- Para testar  $H_0$ , uma unidade de teste,  $t$ , é definida e uma área crítica,  $C$ , é fornecida também, a qual é uma parte da área sobre a qual  $t$  varia
- Isso significa que o teste de significância pode ser formulado como:

If  $t \in C$ , reject  $H_0$

If  $t \notin C$ , do not reject  $H_0$

- A hipótese nula deve, portanto, ser formulada negativamente, ou seja, a intenção de o teste é rejeitar a hipótese
- Se a hipótese nula não for rejeitada, nada pode ser dito sobre o resultado, enquanto se a hipótese for rejeitada, pode-se afirmar que a hipótese é falsa com um dado significado ( $\Delta$ )

## Exemplo de Teste da Hipótese Tipo t

- Se um pesquisador observa um veículo e quer mostrar que o veículo não é um carro
  - O pesquisador sabe que todos os carros têm quatro rodas, mas também que há outros veículos com quatro rodas
- Um exemplo muito simples de uma hipótese nula pode ser formulada como “ $H_0$ : o veículo observado é um carro”
- A unidade de teste  $t$  é o número de rodas e a área crítica é  $C = 1, 2, 3, 5, 6, \dots$
- O teste é if  $t \leq 3$  ou  $t \geq 5$ , rejeite  $H_0$ , senão, não rejeitar  $H_0$
- Se for observado que  $t = 4$ , significa que a hipótese não pode ser rejeitada e nenhuma conclusão pode ser tirada
- Isso ocorre porque pode haver outros veículos além de carros com quatro rodas

## Testes Paramétricos e Não Paramétricos

- Os testes podem ser classificados em testes paramétricos e testes não paramétricos
  - **Testes Paramétrico** são baseados em um modelo que envolve uma distribuição específica
  - Os testes paramétricos também exigem que os parâmetros possam ser medidos pelo menos em uma escala de intervalo
  - Se os parâmetros não podem ser medido em pelo menos uma escala de intervalo, isso significa que os testes paramétricos não pode ser usado

### 3

## Testes Paramétricos e Não Paramétricos

- Os **testes não paramétricos** não fazem o mesmo tipo de suposições sobre o distribuição de parâmetros como os testes paramétricos
- Apenas suposições muito gerais são feitas para derivar testes não paramétricos
- Testes não paramétricos são mais gerais do que testes paramétricos
  - Isso significa que os testes não paramétricos podem ser usados em vez de testes paramétricos, mas o contrário não é possível
- Com relação à escolha de teste paramétrico ou não paramétrico, existem dois fatores a serem considerados:

1. Aplicabilidade

2. Poder

# Testes Paramétricos e Não Paramétricos

- **Aplicabilidade**

- Quais são as suposições feitas pelos diferentes testes?
- É importante que as suposições sobre as distribuições de parâmetros e as suposições relativas às escalas são realistas

## aplicabilidade



Qualidade do que ocasiona um efeito;  
característica do que se consegue aplicar,  
empregar, colocar em prática: aplicabilidade da  
lei.

Característica ou particularidade do que é  
aplicável: aplicabilidade das normas.

[www.dicionario.info](http://www.dicionario.info)

## Testes Paramétricos e Não Paramétricos

- Poder

- O poder dos testes paramétricos é geralmente maior do que para testes não paramétricos
- Os testes paramétricos requerem menos dados e, portanto, experimentos menores, do que o teste não paramétrico se as suposições forem verdadeiras
- Simulações têm mostrado que métodos paramétricos, como como o teste t são bastante robustos a desvios das pré-condições (escala de intervalo), desde que esses desvios não sejam muito grandes

## Visão Geral dos Testes

| Teste         | Descrição   |
|---------------|---|
| t-test        | Um dos testes paramétricos mais usados. O teste é usado para comparação entre duas amostras Ou seja, o design é um fator com dois tratamentos |
| Mann-Whitney  | Esta é uma alternativa não paramétrica para o teste t   |
| F-test        | Este é um teste paramétrico que pode ser usado para comparar distribuições de duas amostras   |
| Paired t-test | Um teste t para uma comparação de design pareado  |
| Wilcoxon      | Esta é uma alternativa não paramétrica para o teste t pareado   |
| Sign test     | É uma alternativa não paramétrica ao teste t pareado. O sign test é uma alternativa mais simples ao teste de Wilcoxon                         |

## Visão Geral dos Testes

| Teste          | Descrição  |
|----------------|--|
| ANOVA          | <p>ANalysis Of VAriance, uma família de testes paramétricos que podem ser usado para projetos com mais de dois níveis de fator. ANOVA pode ser usados nos seguintes projetos:</p> <ul style="list-style-type: none"><li>• Um fator com mais de dois níveis,</li><li>• Um fator e variável de bloqueio,</li><li>• Design fatorial e</li><li>• Design aninhado</li></ul> |
| Kruskal-Wallis | <p>Esta é uma alternativa não paramétrica de teste para ANOVA no caso de um fator com mais de dois tratamentos</p>   |
| Chi-2          | <p>Esta é uma família de testes não paramétricos que podem ser usados quando os dados estão na forma de frequências</p>  |



## Visão Geral dos Testes

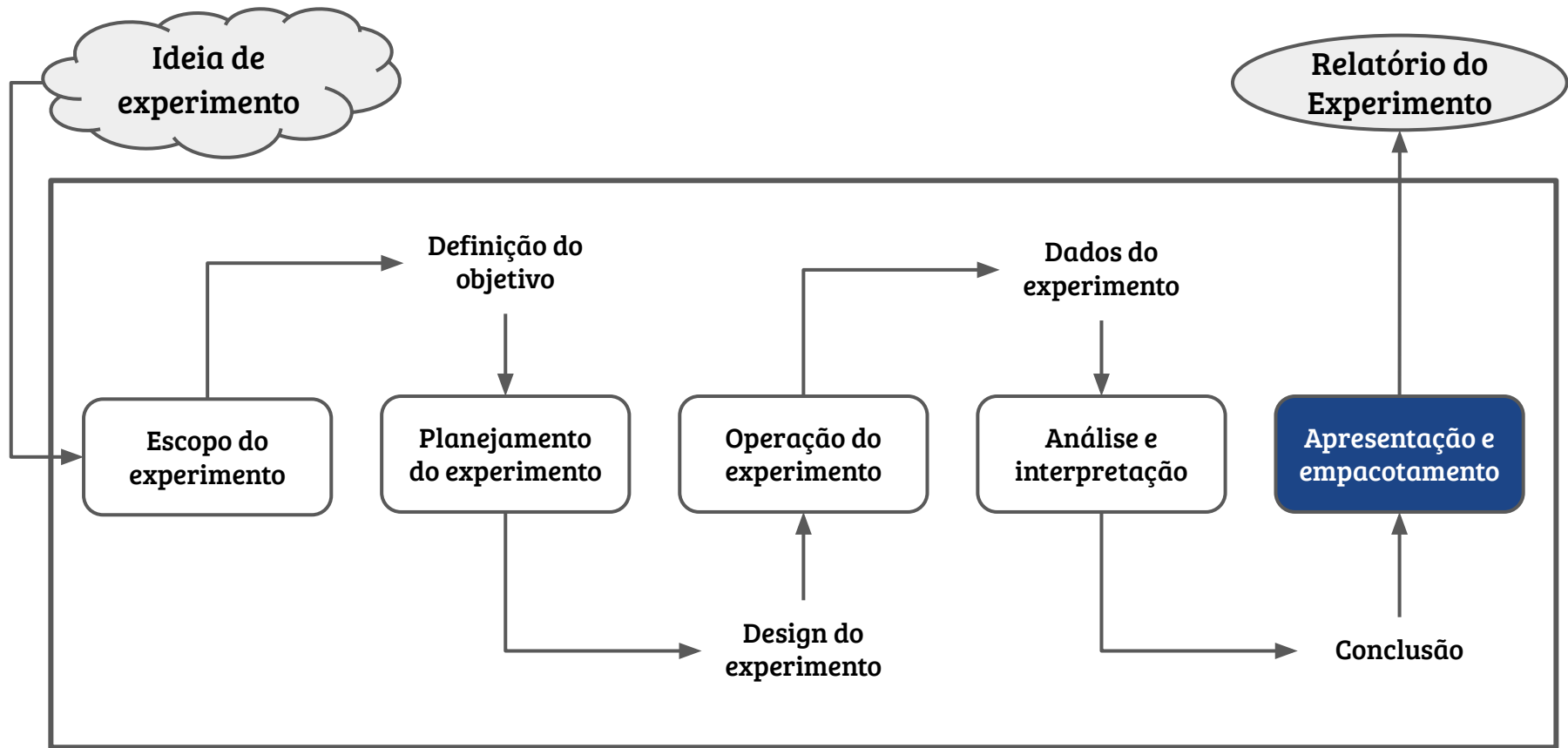
- A visão geral torna possível uso dos testes baseados nas descrições
- Os diferentes testes podem ser classificados conforme tipo de design e parâmetros na tabela 10.3

**Table 10.3** Overview of parametric/non-parametric tests for different designs

| Design   | Parametric         | Non-parametric        |
|--|--------------------|-----------------------|
| One factor, one treatment                                |                    | Chi-2, Binomial test  |
| One factor, two treatments, completely randomized design | t-test, F-test     | Mann-Whitney, Chi-2   |
| One factor, two treatments, paired comparison            | Paired t-test      | Wilcoxon, Sign test   |
| One factor, more than two treatments                     | ANOVA              | Kruskal-Wallis, Chi-2 |
| More than one factor                                     | ANOVA <sup>a</sup> |                       |

<sup>a</sup> This test is not described in this book. Refer instead to, for example, Marascuilo and Serlin [119] and Montgomery [125]

# Processo do Experimento



# Apresentação e Empacotamento

- Esta etapa está preocupada em apresentar e empacotar as conclusões que incluem principalmente a **documentação dos resultados**
  - Pode ser feito por meio de um artigo de pesquisa para publicação e um pacote para fins de replicação ou como parte da base de experiência
- Um experimento nunca fornecerá a resposta final a uma pergunta, portanto, **é importante facilitar a replicação do experimento**
  - Uma documentação completa é um pré-requisito para atingir esse objetivo
- Independentemente, devemos documentar e apresentar os resultados de maneira adequada

# Empacotamento

- A característica mais importantes do experimento é a necessidade da sua **repetição**
  - Com a repetição os pesquisadores adquirem o conhecimento adicional a respeito dos conceitos estudados, e recebem os resultados que são iguais ou diferentes dos resultados do experimento original
- O pré-requisito necessário para a repetição do experimento é o seu empacotamento propriamente realizado



# Empacotamento

- O empacotamento padronizado dos dados experimentais pode servir como base para a criação das bibliotecas de experimentação
  - Como organizo essas informações?
  - Através de um banco de dados com a informação empírica organizada
    - Dessa forma será possível abrir a possibilidade de armazenar os artefatos diferentes desde as idéias ou hipóteses até os resultados e experiências finais

# Empacotamento

- A organização do experimento
  - A organização do experimento inclui o conjunto total das informações a respeito do design do experimento, a preparação dos participantes, a instrumentação, as diretrizes para a execução do experimento, etc...
- Os artefatos do experimento
  - Os artefatos incluem a descrição da instrumentação usada para a coleta dos resultados
  - Dependendo do tipo e dos objetivos do experimento, o papel do artefato do experimento pode ser cumprido pela documentação do projeto do desenvolvimento de software, código fonte, módulos executáveis, ou algo outro que possa ajudar a coletar a informação

# **Empacotamento**

- Os resultados do experimento
  - Os resultados incluem a descrição detalhada dos resultados recebidos
  - Os resultados são apresentados como: dados puros, dados refinados sem outliers, e dados analisados (depois da aplicação da estatística descritiva e os testes estatísticos)
  - Os dados analisados podem ser utilizados para verificar as hipóteses e fazer as conclusões a respeito do atendimento dos objetivos
  - Os resultados devem incluir a informação geral sobre os resultados do experimento, a informação sobre o processo de experimentação, a lista dos problemas e as questões que devem ser resolvidas nos próximos estudos

# Apresentação e Empacotamento

- O relatório técnico oferece discussão sobre os conceitos básicos da organização de estudos experimentais na área de Engenharia de Software
- Experimentação oferece o modo sistemático, disciplinado, computável e controlado para avaliação da atividade humana
- Novas invenções e sugestões não devem ser apenas sugeridas, publicadas ou apresentadas para venda, mas pelo menos devem ser comparadas com as existentes





# *Obrigado!*

## *Por hoje é só pessoal...*

# Dúvidas?



**IsmayleSantos**



**ismayle.santos@uece.br**



**@IsmayleSantos**