

Pesquisa Experimental

Profa. M. Cristina
cristina@icmc.usp.br

Baseado em material preparado e cedido por:

Nemesio Freitas Duarte Filho

Kleberson Junio do Amaral Serique

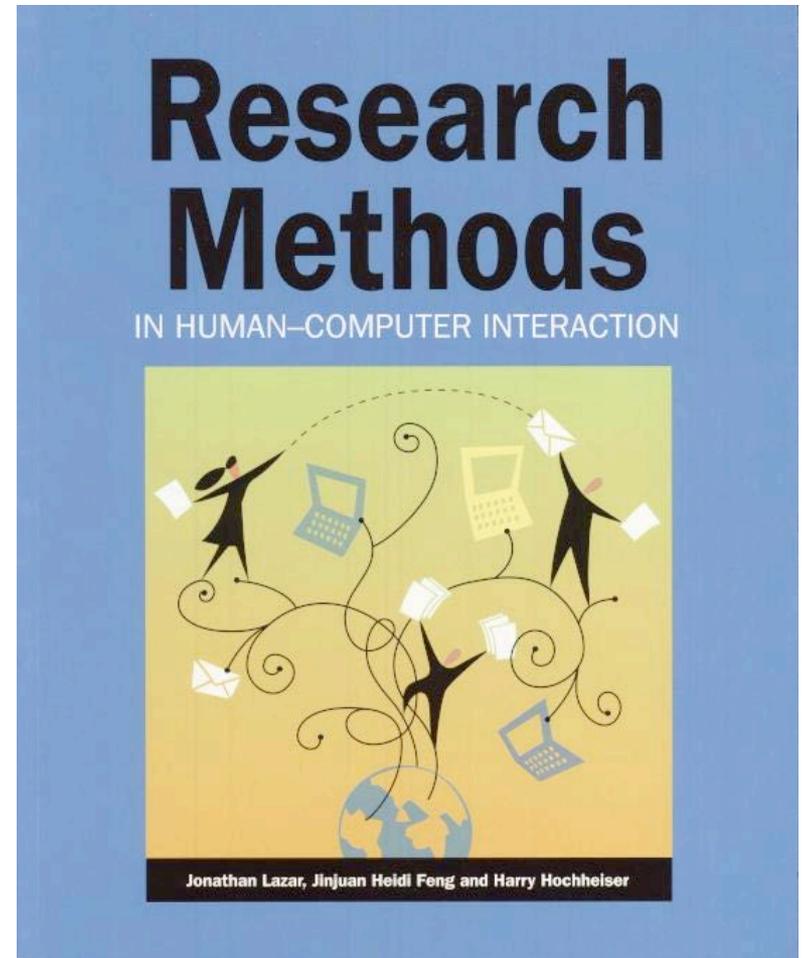
Prof. Dra. Renata Pontin

Prof. José Fernando Rodrigues Jr.

Leitura recomendada

- Capítulos 2, 3, 4:
 - Experimental research
 - Experimental design
 - Statistical tests

(Jonathan Lazar, Jinjuan Heidi Feng, & Harry Hochheiser - Research Methods in Human-Computer Interaction, Wiley, 2010. ISBN 0-470-72337-8, 978-0-470-72337-1)



Roteiro (parte 1)

- Conceitos básicos da pesquisa experimental
- Passos de um experimento
- Variável dependente & independente
- Hipóteses alternativas e nulas
- Tipos de erros
- Limitações da pesquisa experimental

Introdução

- A **pesquisa experimental** busca ajudar um pesquisador a determinar o efeito causal entre dois fatores
 - P.ex. observou-se que um grupo de adolescentes que jogam um certo jogo digitam mais rápido do que outro grupo de adolescentes que não jogam
 - Existe uma relação causal, i.e., jogar esse jogo faz com que os adolescentes digitem mais rápido?
- Para isso, é preciso testar hipóteses que expressam a convicção do pesquisador
 - Uma hipótese é uma afirmação precisa relativa um problema que pode ser verificada (testada) diretamente por meio de uma investigação empírica

Conceitos básicos da Pesquisa Experimental

Princípios da pesquisa experimental:

Temos uma idéia de relação causa-efeito

Acreditamos que existe uma relação entre a construção da causa e a construção do efeito

Temos formulações de hipóteses a serem testadas

Temos vários tratamentos (variáveis independentes)

Executamos o experimento e observamos a saída (variáveis dependentes)

Se o experimento foi corretamente elaborado, podemos formular conclusões a respeito da relação causa-efeito para a hipótese estabelecida

Hipótese nula vs Hipóteses alternativas

Um experimento normalmente tem ao menos uma hipótese nula e uma hipótese alternativa

Hipótese nula (H_0) – declara que não existe diferença entre os tratamentos experimentais, i.e., não será possível observar um padrão de causa-efeito no experimento

Hipótese alternativa (H_1) – declara o oposto, i.e., será possível observar um padrão de causa-efeito no experimento

As hipóteses devem ser mutuamente exclusivas, i.e., se uma é falsa a outra é verdadeira, e vice-versa.

Hipótese nula vs Hipóteses alternativas

O objetivo de um experimento é achar evidência estatística que permita **rejeitar a hipótese nula**, para fortalecer a hipótese alternativa

Alguns experimentos podem investigar múltiplos pares de hipóteses nulas/alternativas

mas quanto mais hipóteses serão testadas, mais fatores precisarão ser controlados e mais variáveis precisarão ser medidas – aumentando os riscos

Hipótese nula vs Hipóteses alternativas

H0 é a hipótese que queremos **REJEITAR** com a maior significância (certeza) possível.

Exemplo: “Uma técnica nova de inspeção encontra, na média (μ), o mesmo número de falhas (#F) do que a técnica antiga”

$$H_0: \mu_{\#F_antiga} = \mu_{\#F_nova}$$

Hipótese nula vs Hipóteses alternativas

Hipótese Alternativa (H1) .

Exemplo: “Uma técnica nova de inspeção encontra, na média (μ), mais falhas (#F) que a técnica antiga”.

$$H_1: \mu_{\#F_antiga} < \mu_{\#F_nova}$$

Hipótese nula vs Hipóteses alternativas

A hipótese nula H_0 representa *a circunstância que está sendo testada, e o objetivo dos testes de hipóteses é sempre tentar rejeitar a hipótese nula.*

A hipótese alternativa H_1 representa o que se deseja provar ou estabelecer, sendo formulada para contradizer a hipótese nula.

Uma hipótese bem formulada coloca claramente quais são as **variáveis dependentes** e as **variáveis independentes** do estudo.

Variáveis dependentes e independentes

Independentes

Todas aquelas que são manipuladas ou controladas

Identificar as v.i. não é fácil e, normalmente, exige *conhecimento do domínio*

Dependentes

As que são observadas (medidas) para ver os efeitos das manipulações nas variáveis independentes

Normalmente temos apenas uma variável dependente, que na maioria das vezes *não é diretamente mensurável*

Variáveis dependentes e independentes

- Ex.
 - H0: *There is no difference between the pull-down menu and the pop-up menu in the time spent locating pages*
 - Independent variable: type of menu (pull-down ou pop-up)
 - + individual factors?
 - Dependent variable: time to locate a page
- H0: Uma técnica nova de inspeção não encontra, na média (μ), mais falhas (#F) que a técnica antiga
 - Variável independente: número de falhas encontradas
 - Variável dependente: a técnica inspeção

Componentes do experimento

- Identificada a hipótese de pesquisa, o design do experimento envolve 3 componentes: **tratamento, unidades, método de atribuição**
- **Tratamentos:** as diferentes técnicas, dispositivos, procedimentos a serem comparados
- **Unidades:** os objetos aos quais os tratamentos serão aplicados
- **Método de atribuição:** definição de como unidades serão alocadas aos diferentes tratamentos (deve garantir que seja aleatório) (*randomization*)

Testes de significância

- Porque são necessários?
- Porque não somente comparar duas médias de uma medida de erro, p.ex., e concluir que o tratamento associado ao menor valor tem melhor desempenho do que o outro?
- Porque o experimento observou uma “amostra” da população alvo
 - Até que ponto podemos confiar no resultado obtido observando o comportamento dessa amostra?
 - Que certeza temos de que esse resultado generaliza para a população como um todo?

Testes de significância

- Em termos técnicos, um teste de significância é um processo em que uma hipótese nula (H_0) é contrastada com uma hipótese alternativa (H_1) para determinar a probabilidade de que a hipótese nula seja verdadeira

Um teste de significância envolve dois tipos de erro em potencial:

O teste corre o risco de rejeitar uma hipótese verdadeira

O teste corre o risco de não rejeitar uma hipótese falsa

Erros: type I and type II

Ao testarmos uma hipótese nula, chegamos a uma conclusão: rejeitá-la, ou não rejeitá-la

Entretanto, devemos lembrar que tais conclusões ora são corretas, ora são incorretas (mesmo quando fazemos tudo corretamente!)

Este é o preço a ser pago por estarmos trabalhando em uma situação em que a variabilidade é inerente !!!

Erros: type I and type II

Type-I-error

Ocorre quando um teste estatístico indica um padrão/relacionamento quando tal padrão/relacionamento não existe (**falso positivo**)

A probabilidade de cometer um erro desse tipo pode ser expressa como:

$$P(\text{type-I-error}) = P(\text{rejeitar } H_0 \mid H_0 \text{ é verdadeira})$$

No exemplo da técnica de inspeção, type-I-error é a probabilidade de rejeitar H_0 mesmo que as 2 técnicas, na média (μ), encontrem o mesmo número de falhas (#F)

Erros: type I and type II

Type-II-error

Ocorre quando um teste estatístico não indica um padrão quando tal padrão/relacionamento de fato existe (**falso negativo**)

A probabilidade de cometer um erro desse tipo pode ser expressa como:

$$P(\text{type-II-error}) = P(\text{n\~{a}o rejeitar } H_0 \mid H_0 \text{ \textit{\'e} falsa})$$

No exemplo de hipóteses apresentado, type-II-error é a probabilidade de não rejeitar H_0 mesmo que as 2 técnicas, na média, possuam médias (μ) do número de falhas (#F) encontradas diferentes (e portanto H_0 deveria ter sido rejeitada)

Controle de riscos: type I and type II

O grau de risco de incorrer em erros do tipo I ou II depende de diferentes fatores

Probabilidade de incorrer em um erro do tipo I: α , ou nível de significância, ou p-value

Probabilidade de incorrer em um erro do tipo II: β

O Poder do Teste Estatístico (P) é dado por $(1 - \beta)$: sua probabilidade de rejeitar a hipótese nula quando ela é de fato falsa.

Controle de riscos: type I and type II

Exemplo:

A eficácia de certa vacina após um ano é de 25% (isto é, o efeito imunológico se prolonga por mais de um ano em apenas 25% das pessoas que a tomam). Desenvolve-se uma nova vacina, mais cara, e deseja-se saber se esta é, de fato, melhor.

Sendo “p” a proporção de imunizados por mais de um ano com a nova vacina...

Quais hipóteses podem ser formuladas?

Que erros poderemos cometer?

Controle de riscos: type I and type II

Exemplo:

Hipótese nula $H_0: p \leq 0,25$

Hipótese alternativa $H_1: p > 0,25$

Erro tipo I: aprovar a vacina quando, na realidade, ela não tem nenhum efeito superior ao da vacina em uso (falso positivo)

Erro tipo II: rejeitar a nova vacina quando ela é, de fato, melhor que a vacina em uso (falso negativo)

Seleção dos participantes

A seleção dos participantes afeta diretamente a capacidade de generalização dos resultados de um experimento

Para tanto, a seleção deve ser representativa da população alvo

Seleção de participantes = amostra de uma população

A amostragem pode ser probabilística ou não-probabilística

Amostragem probabilística: *a probabilidade da seleção de cada participante é conhecida*

Amostragem não-probabilística: *a probabilidade da seleção de cada participante não é conhecida*

Seleção dos participantes

O **tamanho da amostra tem impacto sobre** a generalização dos resultados de um experimento

Quanto maior a amostra, menor é a chance de errar ao generalizar os resultados

Princípios gerais para escolher o tamanho da amostra:

Se existir uma ampla variabilidade na população, uma amostra de tamanho maior é necessária;

A análise dos dados pode influenciar a escolha do tamanho da amostra.

Limitações da Pesquisa Experimental

As pesquisas experimentais constituem um valioso procedimento disponível aos cientistas para testar hipóteses que estabelecem relações de causa e efeito entre as variáveis

Em virtude de suas possibilidades de controle, os experimentos oferecem garantia muito maior do que qualquer outro delineamento de que a variável independente causa efeitos na variável dependente

A despeito de suas vantagens, a pesquisa experimental apresenta várias limitações

Limitações da Pesquisa Experimental

Primeiramente, existem muitas variáveis cuja manipulação experimental pode ser muito difícil ou mesmo impossível

Uma série de características humanas, tais como idade, sexo ou histórico familiar não podem ser conferidas às pessoas de forma aleatória

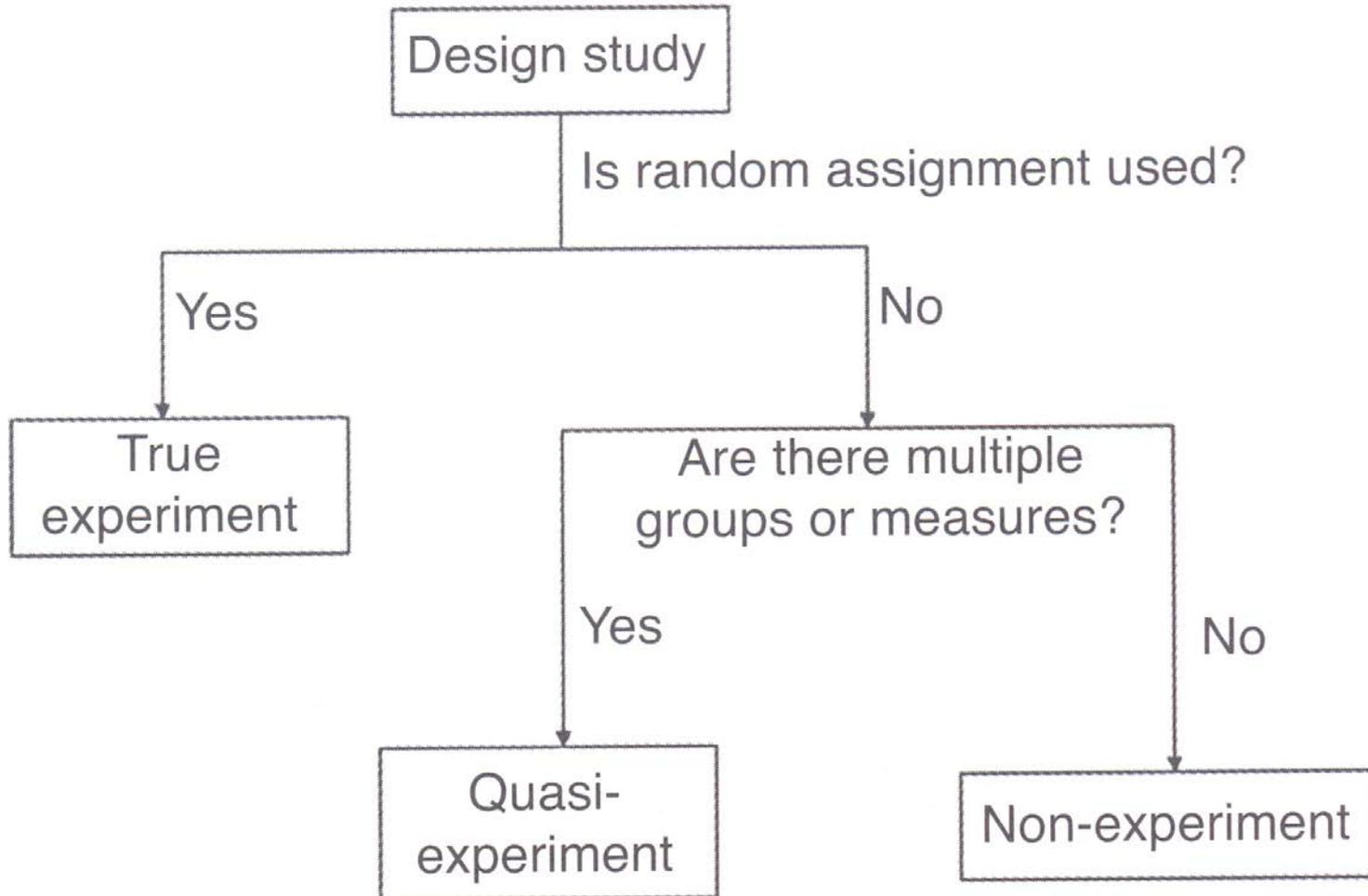
Outra limitação é que muitas variáveis que poderiam ser tecnicamente manipuladas estão sujeitas as considerações de ordem ética que proíbem sua manipulação

Não se pode, por exemplo, submeter pessoas a atividades estressantes com vistas a verificar alterações em sua saúde física ou mental

Roteiro (parte 2)

- ✓ O que é preciso considerar ao delinear um experimento?
- Determinar a estrutura básica
- Investigar uma única variável independente
- Investigar mais de uma variável independente
- Confiabilidade dos resultados experimentais

Introdução



Experimentos verdadeiros

- Baseado no teste e validação da hipótese de pesquisa
- Ao menos duas condições ou grupos (treatment group e control group)
- Variável dependente → medidas quantitativas
- Resultados analisados → testes estatísticos de significância
- Projetado e conduzido buscando remover qualquer tendenciosidade (bias)
- Replicável com amostras distintas, em outros momentos, locais, por outros experimentadores

O que é preciso considerar ao delinear experimentos?

- Número de valores que as variáveis independentes podem assumir determina a quantidade de condições:

“Não existe diferença na velocidade de seleção de alvos por pessoas utilizando um mouse, um joystick ou um trackball para seleccionar ícones em diferentes tamanhos”

Variável dependente: velocidade

Variáveis independentes: tipo de dispositivo, tamanho dos ícones

O que é preciso considerar ao delinear experimentos?

“Não existe diferença na velocidade de seleção de alvos ao se usar um mouse, um joystick ou um trackball para selecionar ícones em diferentes tamanhos (P, M e G)”

- Variáveis independentes:
 - Tipos de dispositivos (3)
 - Tamanho dos ícones (3)
- Numero de condições:
 - (9)

Determinar a estrutura básica

- Baseado na hipótese de pesquisa
- Estimativas de tempo e custo
- Estrutura básica
 - Quantas variáveis independentes serão investigadas?
 - Quantos valores diferentes cada variável independente pode assumir?

Investigar uma única variável independente

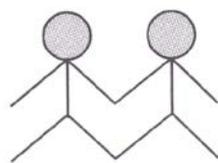
- Planejamento é bem simples em relação a múltiplas variáveis
- Exemplos:
 - H_{01} : Não há diferenças no tempo de digitação (Teclados: QWERTY, DOVORAK ou Alfanumérico)
 - H_{02} : Não há diferenças entre usuários novatos e experientes no tempo que levam para localizar um item em uma loja virtual
 - H_{03} : Não há diferenças na confiança percebida em um agente on-line entre clientes dos EUA, Rússia, China e Nigéria

Investigar uma única variável independente

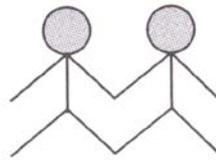
- Entre grupos (between-group design)
- Intra grupos (within-group design)
- Passo crítico que impacta
 - Qualidade dos dados
 - Os métodos estatísticos que devem ser usados

Between-group design ou Within-group design

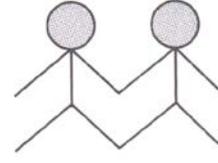
- Between-group design
 - Cada participante é exposto a uma condição
 - Grupos de participantes == Número de Condições



QWERTY
Keyboard



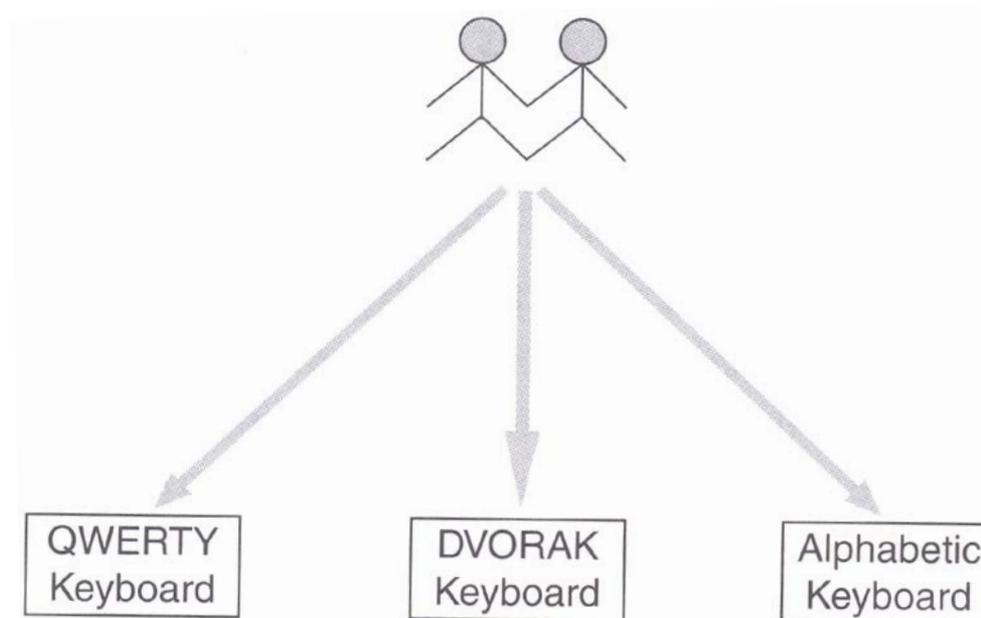
DVORAK
Keyboard



Alphabetic
Keyboard

Between-group design ou Within-group design

- Within-group design
 - Cada participante é exposto a todas as condições
 - Apenas um grupo de participantes



Comparaçãõ entre Between-group design e Within-group design

Type of experiment design

Between-group design

Within-group design

Advantages

Cleaner
Avoids learning effect
Better control of confounding factors,
such as fatigue

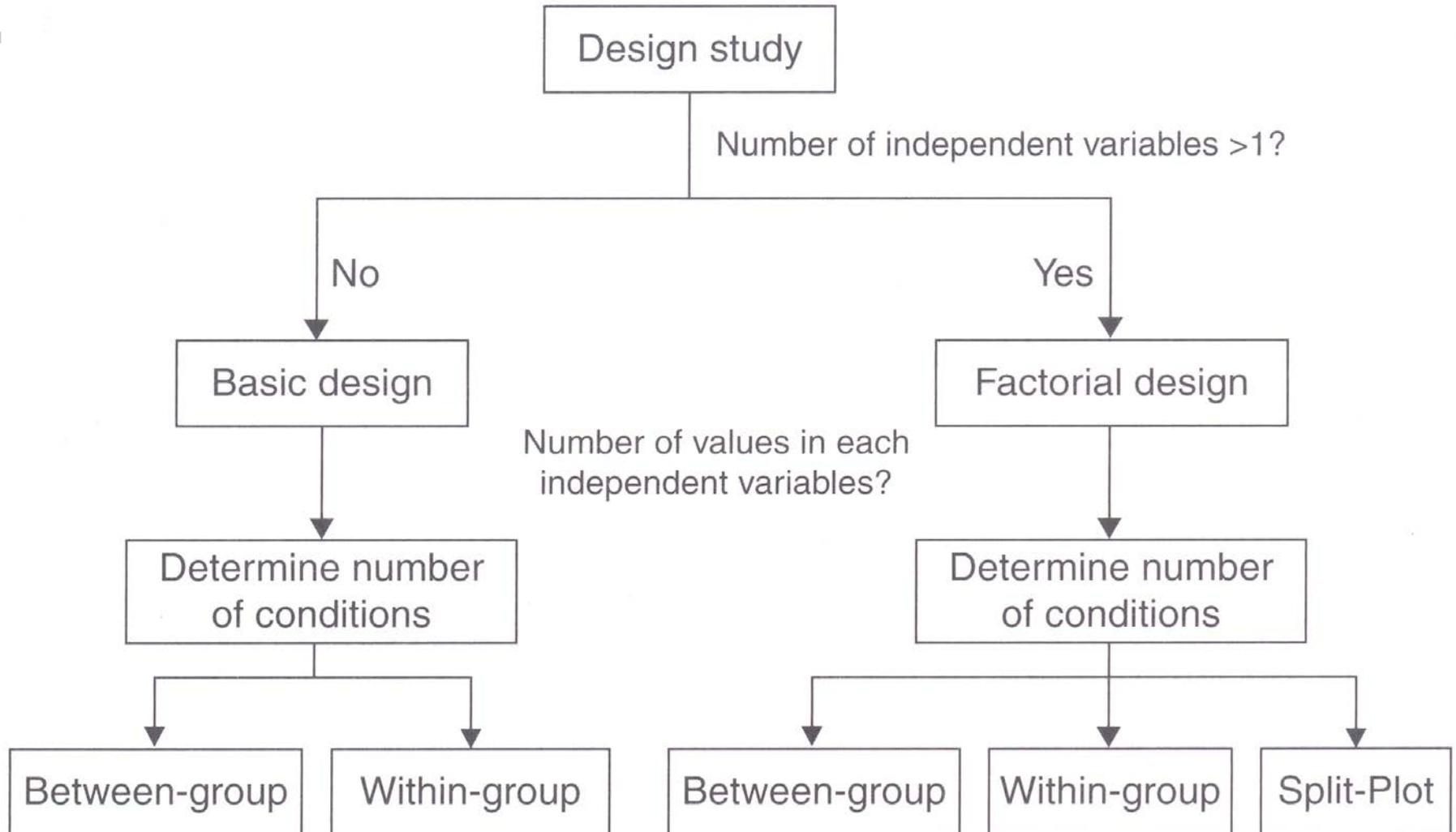
Smaller sample size
Effective isolation of individual
differences
More powerful tests

Limitations

Larger sample size
Large impact of individual differences
Harder to get statistically significant
results

Hard to control learning effect
Large impact of fatigue

Determinar a estrutura básica



Investigar mais de uma variável independente

- Factorial Design
 - Mais de uma variável independente ou fator
 - Fórmula para o número de condições:

$$C = \prod_{a=1}^n V_a$$

- C = número de condições
- V = número de valores assumidos por cada variável (levels)

Investigar mais de uma variável independente

- Exemplo:
 - QWERTY, DVORAK e Alfanumérico
 - Tempo de digitação
 - **Efeito de tarefas diferentes (Composição e Transcrição)**
 - Variável tipo de teclado: 3 valores
 - Variável tipo de tarefa: 2 valores
 - Numero de condições = $3 \times 2 = 6$

Investigar mais de uma variável independente

- Exemplo: (resultado)

	QWERTY	DVORAK	Alphabetic
Composition	1	2	3
Transcription	4	5	6

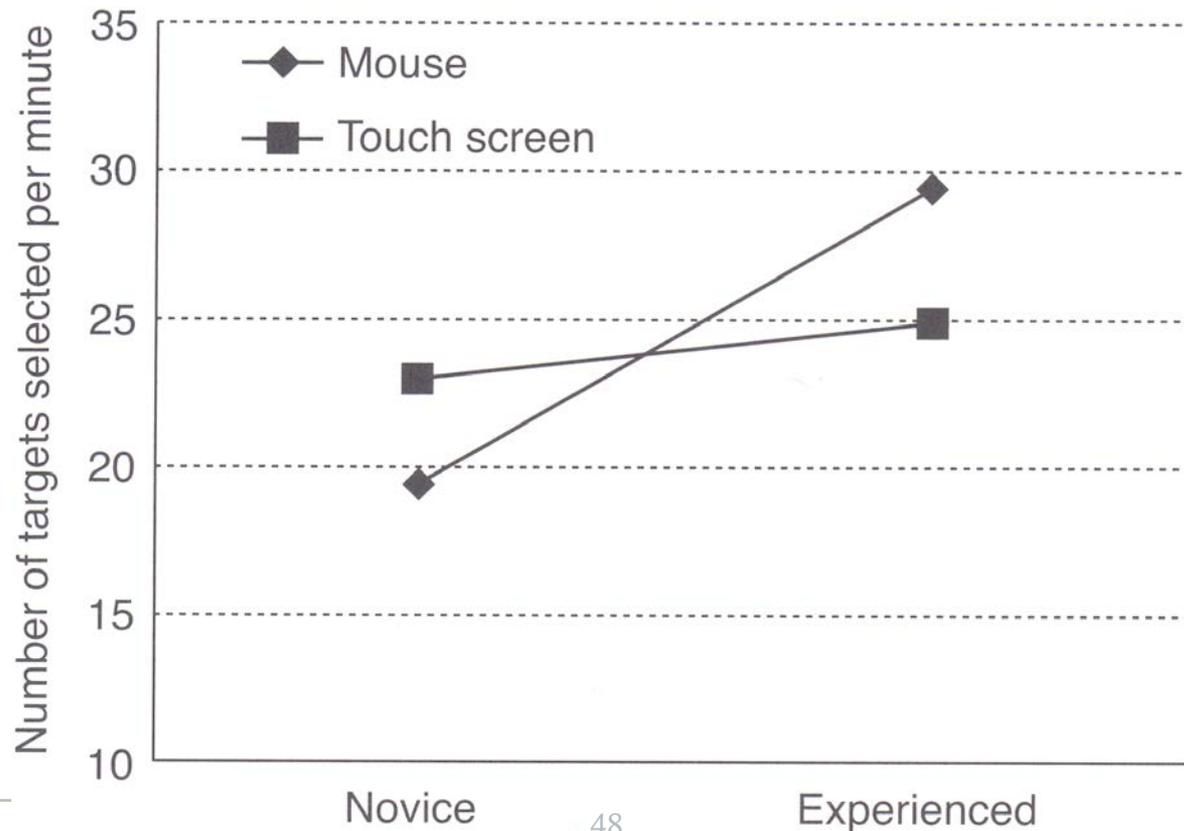
- Within-group design (1 grupo)
- Between-group design (6 grupos)

Efeitos de Interação

- Factorial design – permite analisar os efeitos de interação entre 2 ou mais variáveis independentes
- Variáveis independentes X e Y
- Variável dependente Z (X e Y)
- Exemplo:
 - Tipos de dispositivos (mouse ou touch screen)
 - Tipos de usuários (novatos e experientes)
 - Impacto na eficiência de tarefas de seleção

Efeitos de Interação

- Exemplo:



Confiabilidade do resultados experimentais

- Confiabilidade de um experimento – pode ser replicado em outro local, outro momento, outros grupos de pesquisa, etc.
 - Erros aleatórios
 - Erros casuais” ou “ruído” – não correlacionados com o valor real
 - Erros sistemáticos
 - “biases” – viéses, tendências, polarizações,...
- HCI & “hardsciences”
 - Comportamento humano e interação social
 - Alto grau de flutuação (erros)
 - Menos replicável