

Projeto de Experimentos

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

Baseado em material preparado e cedido por:

Nemesio Freitas Duarte Filho

Kleberson Junio do Amaral Serique

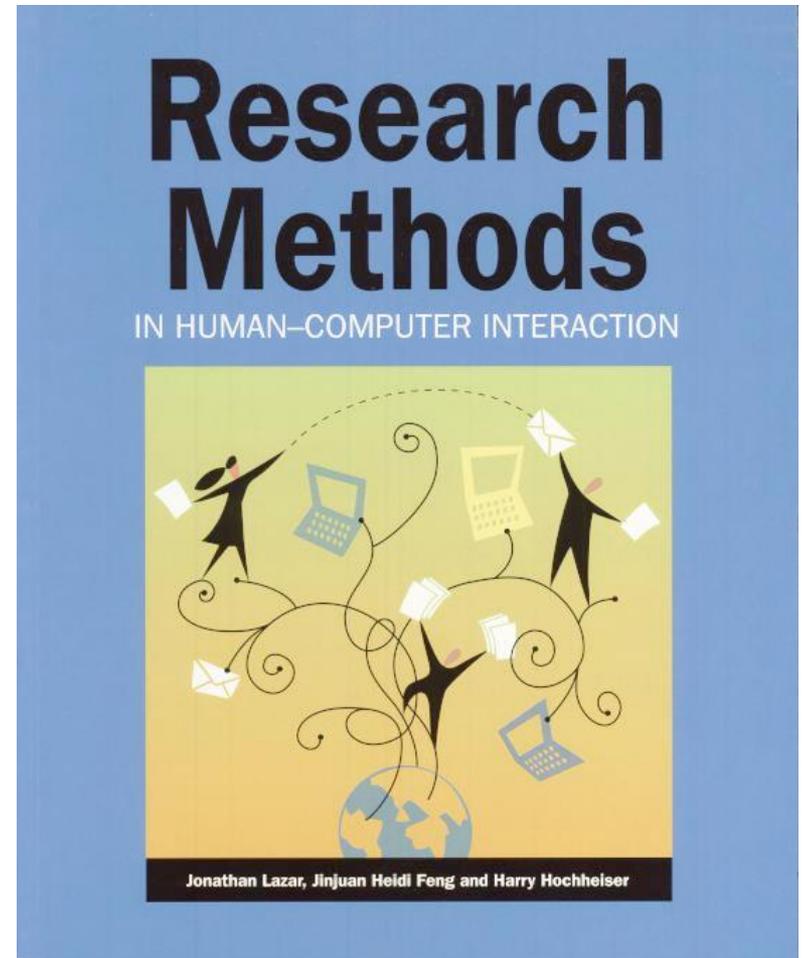
Prof. Dra. Renata Pontin

Prof. José Fernando Rodrigues Jr.

Leitura recomendada 2

- Capitulo 3:
Experimental design

(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)



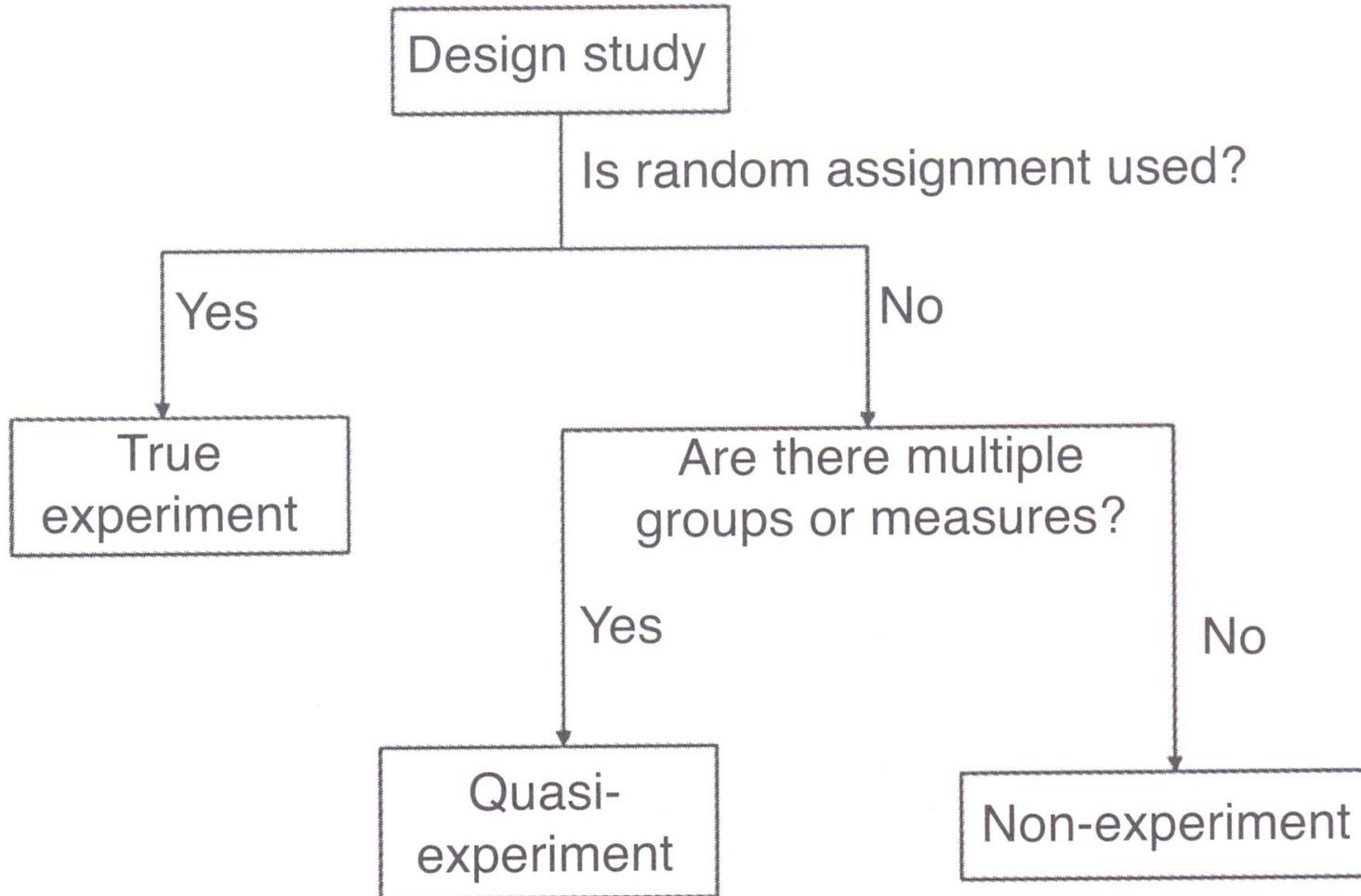
Roteiro

- ✓ O que é preciso ser considerado quando delineamos experimentos?
- Determinar a estrutura básica
 - Investigar uma única variável independente
 - Investigar mais de uma variável independente
- Confiabilidade dos resultados experimentais
- Procedimentos experimentais

Introdução

- Experimentos em HCI
 - Desenvolver e modificar modelos ou tarefas
 - Avaliar diferenças no design de soluções
 - Responder questões críticas:
 - Ex. Qual tecnologia a ser adotada?
- Em visualização
 - Avaliar técnicas, algoritmos, mapeamentos visuais...
- Experimentos bem projetados vs experimentos mal projetados

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 ser considerado quando delineamos experimentos?

- Questões universais para todos os experimentos científicos
 - Hipótese de pesquisa
 - Medição de variáveis dependentes
 - Controle de múltiplas condições
- Questões humanas
 - Efeito de aprendizagem
 - Conhecimento prévio dos participantes
 - Tamanho do *pool* de participantes

O que é preciso ser considerado quando delineamos experimentos?

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

Ex. “Não existe diferença na velocidade de seleção de alvos em uma interface por pessoas utilizando um *mouse*, um *joystick* ou um *trackball* para selecionar ícones de diferentes tamanhos (P, M e G)””

Variável dependente: velocidade

Variáveis independentes: tipo de dispositivo, tamanho dos ícones

Quantas condições?

O que é preciso ser considerado quando delineamos experimentos?

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

- 2 variáveis independentes:
 - Tipos de dispositivos (3)
 - Tamanho dos ícones (3)
- Numero de condições:
 - (9)
- E a variável dependente, qual é?

O que é preciso ser considerado quando delineamos experimentos?

- Identificar a variável dependente e definir como medi-la
 - Ex. velocidade de execução de uma tarefa
 - Ex. velocidade de digitação: palavras digitadas por minuto, palavras corretas digitadas por minuto, etc.
- Como controlar as variáveis independentes
 - Na maior parte dos casos é factível, mas pode ser difícil...
 - Ex. aplicativo de reconhecimento de fala
 - Controle de erro de comando

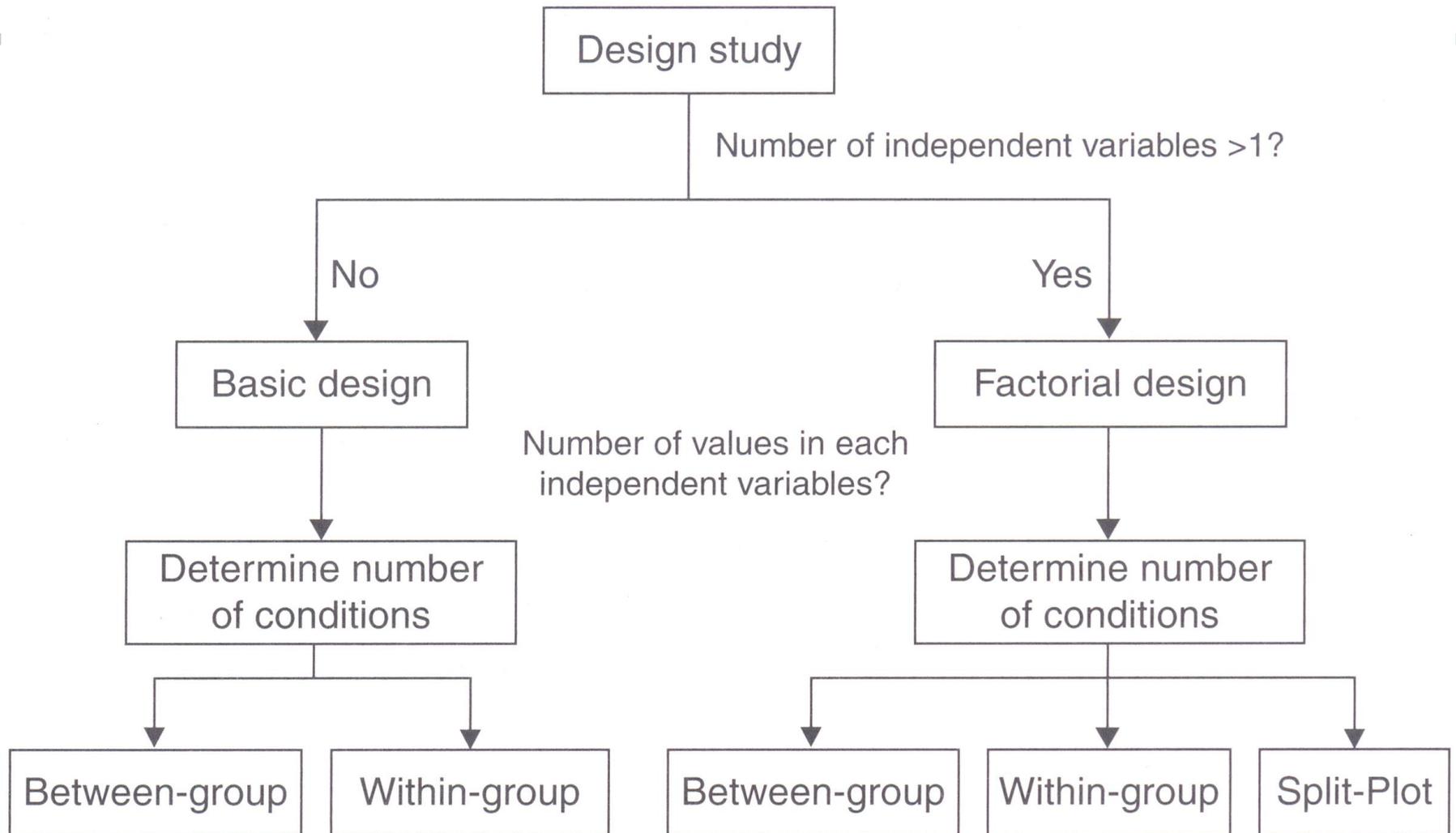
Roteiro

- ✓ O que é preciso ser considerado quando delineamos experimentos?
- ✓ Determinar a estrutura básica
 - Investigar uma única variável independente
 - Investigar mais de uma variável independente
- Confiabilidade dos resultados experimentais
- Procedimentos experimentais

Determinar a estrutura básica

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

Determinar a estrutura básica



Investigar uma única variável independente

- Planejamento é bem simples em relação a múltiplas variáveis
- Exemplos:
 - H_1 : Não há diferenças no tempo de digitação (teclados: QWERTY, DVORAK ou Alfanumérico)
 - H_2 : Não há diferenças entre usuários novatos e experientes no tempo que levam para localizar um item em uma loja virtual
 - H_3 : Não há diferenças na confiança percebida relativa a um agente on-line por clientes localizados nos EUA, na Rússia, na China ou na Nigéria

Investigar uma única variável independente

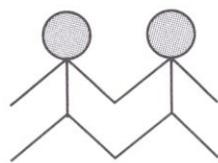
- Número de condições possíveis está relacionando a quantidade de valores assumidos pelas variáveis independentes.
- H_1 : 3 condições
- H_2 : 2 condições
- H_3 : 4 condições

Investigar uma única variável independente

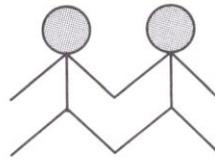
- Desenho entre-grupos (*between-group design*)
- Desenho intra-grupos (*within-group design*)
- Escolha é um passo crítico, que impacta
 - Qualidade dos dados
 - Os métodos estatísticos que devem ser adotados

Between-group design ou Within-group design

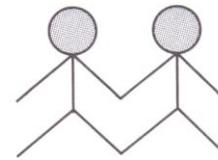
- *Between-group design* (desenho entre-grupos)
 - Cada participante é exposto a uma condição
 - Grupos de participantes == Número de Condições



QWERTY
Keyboard



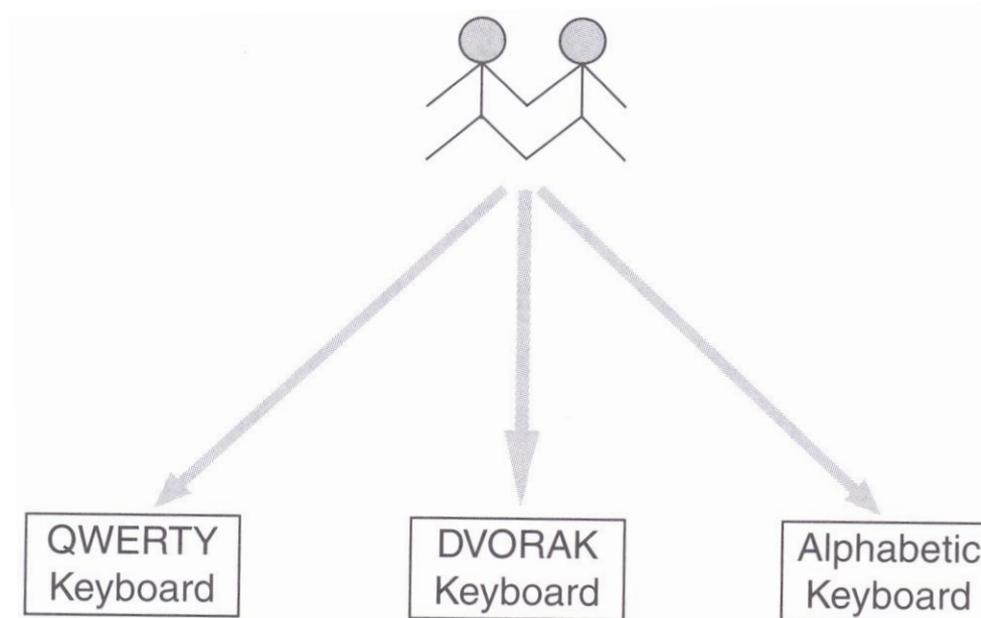
DVORAK
Keyboard



Alphabetic
Keyboard

Between-group design ou Within-group design

- *Within-group design* (desenho intra-grupos)
 - Cada participante é exposto a todas as condições
 - Um único grupo de participantes



Vantagens e desvantagens do *Between-group design*

- Vantagens
 - Sem efeito de aprendizagem
 - Tempo exigido do participante é menor
 - Fadiga e confusão mental podem ser melhor controlados
- Desvantagens
 - Compara o desempenho de grupos de pessoas diferentes
 - Diferenças individuais podem afetar o desempenho
 - Alto nível de ruído nos dados, mais difícil detectar diferenças significativas
 - É preciso contar com mais participantes para minimizar esses problemas...

Vantagens e desvantagem do *Between-group design*

- Para evitar dados que são ruído:
 - Número grande de participantes (tamanho da amostra maior)
- Número de participantes em cada condição (m)
- Número de condições (n)
- Número total de participantes necessários ($m \times n$)
- Pode ficar inviável
 - Exemplo: $n = 4$ $m = 16$ total = 64 participantes

Vantagens e desvantagem do *Within-group design*

- Vantagens
 - Pode trabalhar com amostra de tamanho menor
 - Diferenças individuais podem ser isoladas
 - Mais favorável para observar diferenças esperadas
 - Se $n = 4$ e $m = 16$ iremos precisar de **16** e não de 64 participantes
 - Menor custo (tempo, esforço) para realizar o experimento
- Desvantagens
 - Possíveis efeitos de aprendizagem
 - Fadiga dos participantes

Comparação 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

Escolha do desenho mais apropriado

- Decisão muito difícil e diretamente relacionada à qualidade dos resultados do experimento
- *Between-group design*
 - Quando investigamos tarefas simples
 - É difícil estimar as diferenças individuais (DI): as DIs tendem a ser pequenas quando envolvem tarefas simples e pouco esforço cognitivo
 - **H2:** Não há diferença entre usuários novatos e experientes no tempo gasto para localizar um item em uma loja virtual
 - **H3:** Não há diferença na confiança percebida em um agente on-line entre clientes localizados nos EUA, Rússia, China ou Nigéria

Escolha da abordagem de design mais apropriada

- *Within-group design*
 - Mais adequado se o experimento envolve tarefas mais complexas: leitura, compreensão, recuperação da informação e resolução de problemas
 - Tarefas que se acredita serem muito afetadas por diferenças individuais, porque esse desenho é menos suscetível a elas
 - Menos suscetível aos efeitos de aprendizagem
 - Menor esforço de recrutamento de participantes

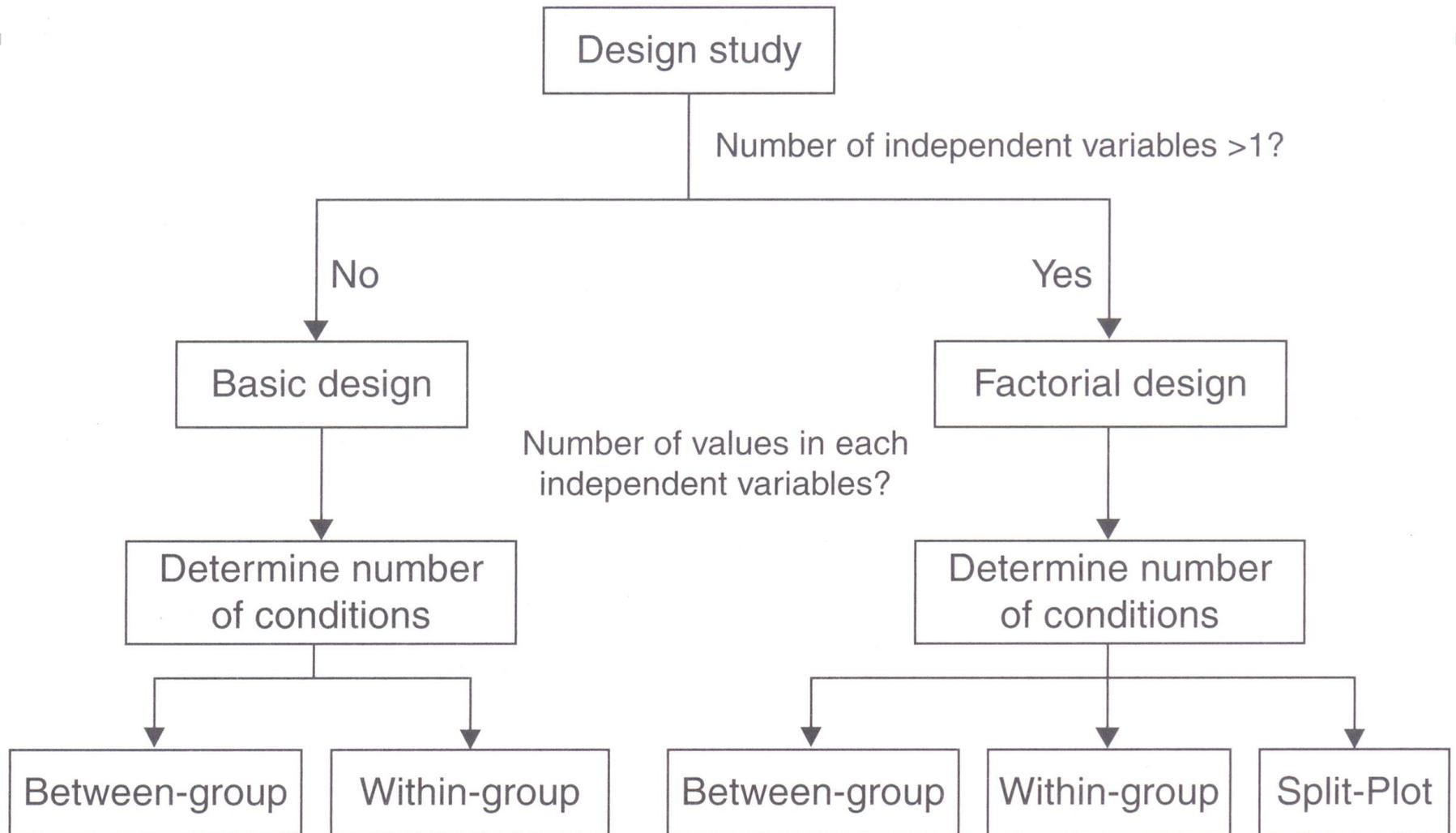
Escolha da abordagem de design mais apropriada

- *Within-group design*
- Controlar
 - Efeitos de aprendizagem (randomizar a sequencia das condições, treinamento dos participantes)
 - Fadiga (tarefas modestas, 60-90 minutos, interrupções)
 - Outros problemas relacionados

Roteiro

- ✓ O que é preciso ser considerado quando delineamos experimentos?
- ✓ Determinar a estrutura básica
 - ✓ Investigar uma única variável independente
 - ✓ Investigar mais de uma variável independente
- Confiabilidade dos resultados experimentais
- Procedimentos experimentais

Determinar a estrutura básica



Investigar mais de uma variável independente

- *Factorial design*
 - Investigação envolve 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)

Split-plot design

- abordagem híbrida quando trata múltiplas variáveis independentes: uma variável investigada no *design between-group*, outras no *design within-group*
- Exemplo: variáveis independentes: idade e uso do GPS

	20 to 40 years old	41 to 60 years old	Above 60
Driving without GPS assistance	1	2	3
Driving with GPS assistance	4	5	6

- #Condições = $2 * 3 = 6$
- Idade: 3 grupos de idades: *between-group*, 1 grupo p/ cada faixa etária
- Com/Sem GPS: *within-group*, cada participante dirige nas duas condições

Efeitos de Interação

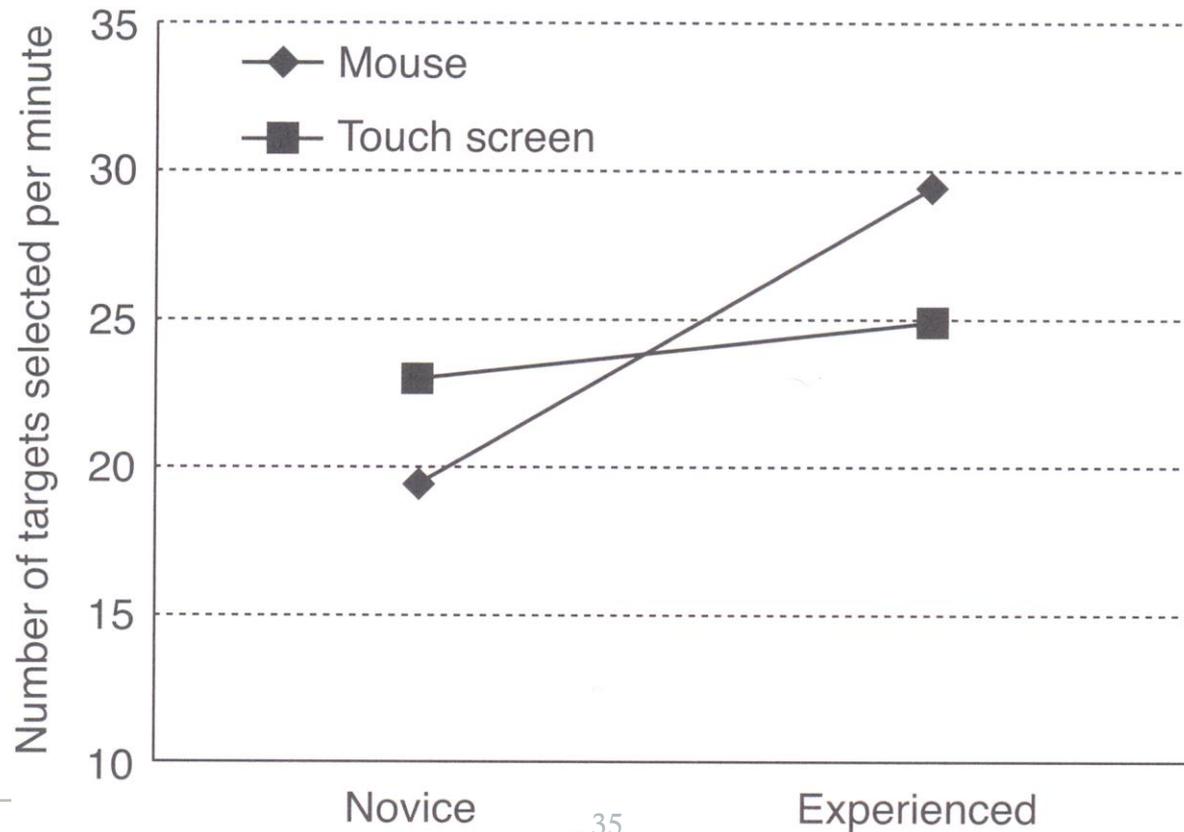
- *Factorial design* – vantagem é que permite analisar os efeitos de interação entre duas ou mais variáveis independentes
- *Interaction effect*: “the differing effect of one independent variable on the dependent variable, depending on the particular level of another independent variable” (Cozby 1997)

Efeitos de Interação

- *Factorial design* – analisar efeitos de interação entre duas 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 ou experientes)
 - Impacto na eficiência de tarefas de seleção

Efeitos de Interação

- Exemplo:



Roteiro

- ✓ O que é preciso ser considerado quando delineamos experimentos?
- ✓ Determinar a estrutura básica
- ✓ Investigar uma única variável independente
- ✓ Investigar mais de uma variável independente
- ✓ Confiabilidade do resultados experimentais
- Procedimentos experimentais

Confiabilidade do resultados experimentais

- Confiabilidade de um experimento – de que ele pode ser replicado em outro local, outro momento, outros grupos de pesquisa, etc. (com resultados equivalentes)
- HCI & “*hard sciences*”
 - Comportamento humano e interação social
 - Alto grau de flutuação (erros)
 - Menos replicável

Erros randômicos

- Observações de uma secretária digitando textos
- Valor medido: 50 palavras / minuto
- 5 sessões:
 - Sessão 1: 46 palavras / minuto
 - Sessão 2: 52 palavras / minuto
 - Sessão 3: 47 palavras / minuto
 - Sessão 4: 51 palavras / minuto
 - Sessão 5: 53 palavras / minuto
- Valores observados = valor real + erro aleatório

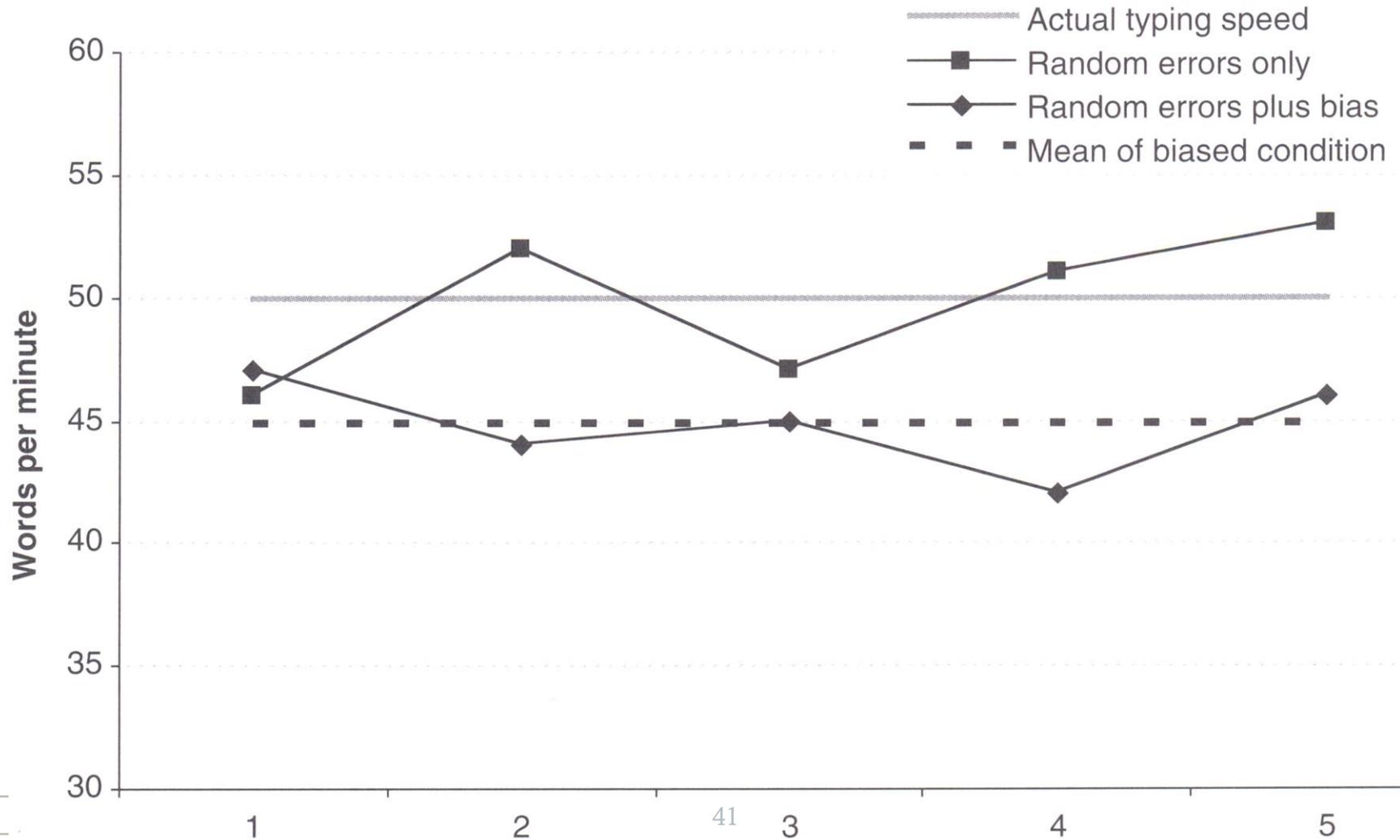
Erros randômicos

- “Erros casuais” ou “ruído” – não correlacionados com o valor real
- Variam em torno do valor real, nas duas direções contrárias
- Como tratar
 - Ampliando o tamanho da amostra
 - 100% de confiança: ($>$ a taxa de amostragem) \Rightarrow valor observado \sim valor real

Erros sistemáticos

- “biases” – viés, tendências, polarizações,...
- Tendem a “puxar” os valores observados para uma mesma direção
- Ex. (cansaço e nervosismo)
 - Sessão 1: 47 palavras / minuto
 - Sessão 2: 44 palavras / minuto
 - Sessão 3: 45 palavras / minuto
 - Sessão 4: 42 palavras / minuto
 - Sessão 5: 46 palavras / minuto

Erros sistemáticos



Erros sistemáticos

- Verdadeiro inimigo da pesquisa experimental
- 5 maiores fontes:
 - Erros de medição
 - Procedimentos experimentais inadequados (fadiga, efeito do aprendizado, problemas nas instruções, ... estudos piloto podem ajudar a detectar e evitar)
 - Participantes
 - Comportamento do experimentador
 - O ambiente em que o experimento é executado

Roteiro

- ✓ O que é preciso ser considerado quando delineamos experimentos?
- ✓ Determinar a estrutura básica
- ✓ Investigar uma única variável independente
- ✓ Investigar mais de uma variável independente
- ✓ Confiabilidade do resultados experimentais
- ✓ Procedimentos experimentais

Procedimentos experimentais

- Experimentos → Responder um infinidade de questões
- HCI é similar a outros campos (comportamento humano)
- Erros e “viés” não podem ser totalmente eliminados (100%)
- Cuidados ao analisar os dados e fazer afirmações a partir dos resultados de um experimento

Procedimentos experimentais

- Ciclo de vida de um experimento de HCI:
 1. Identificar uma hipótese de pesquisa
 2. Especificar o design do estudo
 3. Estudo piloto para testes
 4. Recrutar participantes
 5. Executar a coleta de dados reais
 6. Analisar os dados
 7. Reportar os resultados

Procedimentos experimentais

- Passos típicos em sessão experimental:
 1. Sistemas e dispositivos estão OK
 2. Cumprimentar os participantes
 3. Apresentar a finalidade do estudo e os procedimentos
 4. Obter o consentimento dos participantes
 5. Atribuir os participantes a condição experimental de acordo com o método aleatório predefinido
 6. Treinamento das tarefas

Procedimentos experimentais

- Passos típicos em sessão experimental
 7. Participantes na tarefa real
 8. Participantes respondem questionários (se for o caso)
 9. Sessão de esclarecimentos
 10. Pagamento (se for o caso)