

# Pesquisa Experimental

**SCC 5921 Metodologia de Pesquisa em Visualização e Imagens**  
**Profa. M. Cristina**  
**[cristina@icmc.usp.br](mailto:cristina@icmc.usp.br)**

Baseado em material preparado e cedido por:

Nemesio Freitas Duarte Filho

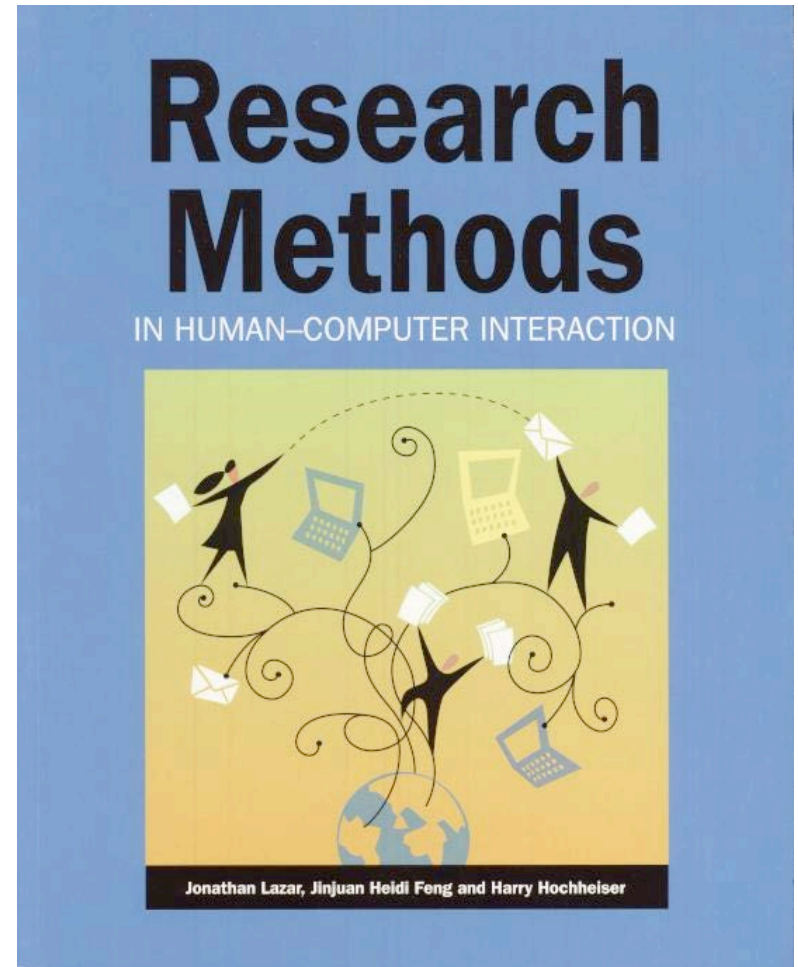
Kleberson Junio do Amaral Serique  
Prof. Dra. Renata Pontin

Prof. José Fernando Rodrigues Jr.

# Leitura recomendada 1

- Capítulo 2:  
Experimental research

(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

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

# Introdução

- A investigação científica e os seus fundamentos são elementos importantes para a Ciência, proporcionando a descoberta de novos conhecimentos em diversas áreas
- Entretanto, existem diferenças significativas nos modos de condução de investigação científica, devido à diversidade de perspectivas em relação ao foco, objetivos, contextos e aspectos a serem pesquisados
- Uma das abordagens possíveis é a empírica
  - pesquisa experimental

# Introdução

- A **pesquisa experimental** busca dar elementos para um pesquisador determinar o efeito causal entre dois fatores
  - P.ex. observou-se que um grupo de adolescentes que jogam um determinado jogo conseguem digitar 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 experimental

# Conceitos básicos da Pesquisa Experimental

A pesquisa experimental consiste, essencialmente, em determinar um objeto de estudo, selecionar as variáveis capazes de influenciá-lo, e definir as formas de controle dessas variáveis e de observação dos efeitos que elas produzem no objeto de estudo

*Trata-se, portanto, de uma pesquisa em que o pesquisador é um agente ativo, e não um observador passivo.*

# Conceitos básicos da Pesquisa Experimental

Uma pesquisa experimental pode ser desenvolvida em qualquer lugar, desde que apresente as seguintes propriedades:

- a) *Manipulação*: o pesquisador deve ser capaz de manipular pelo menos uma das características/propriedades dos elementos sob estudo;
- b) *Controle*: o pesquisador precisa introduzir um ou mais controles na situação experimental, sobretudo criando um “grupo de controle”;
- c) *Atribuição aleatória*: a designação dos elementos para participar dos grupos experimentais e de controle deve ser feita aleatoriamente

# Conceitos básicos da Pesquisa Experimental

## Princípios da pesquisa experimental

Temos uma ideia 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

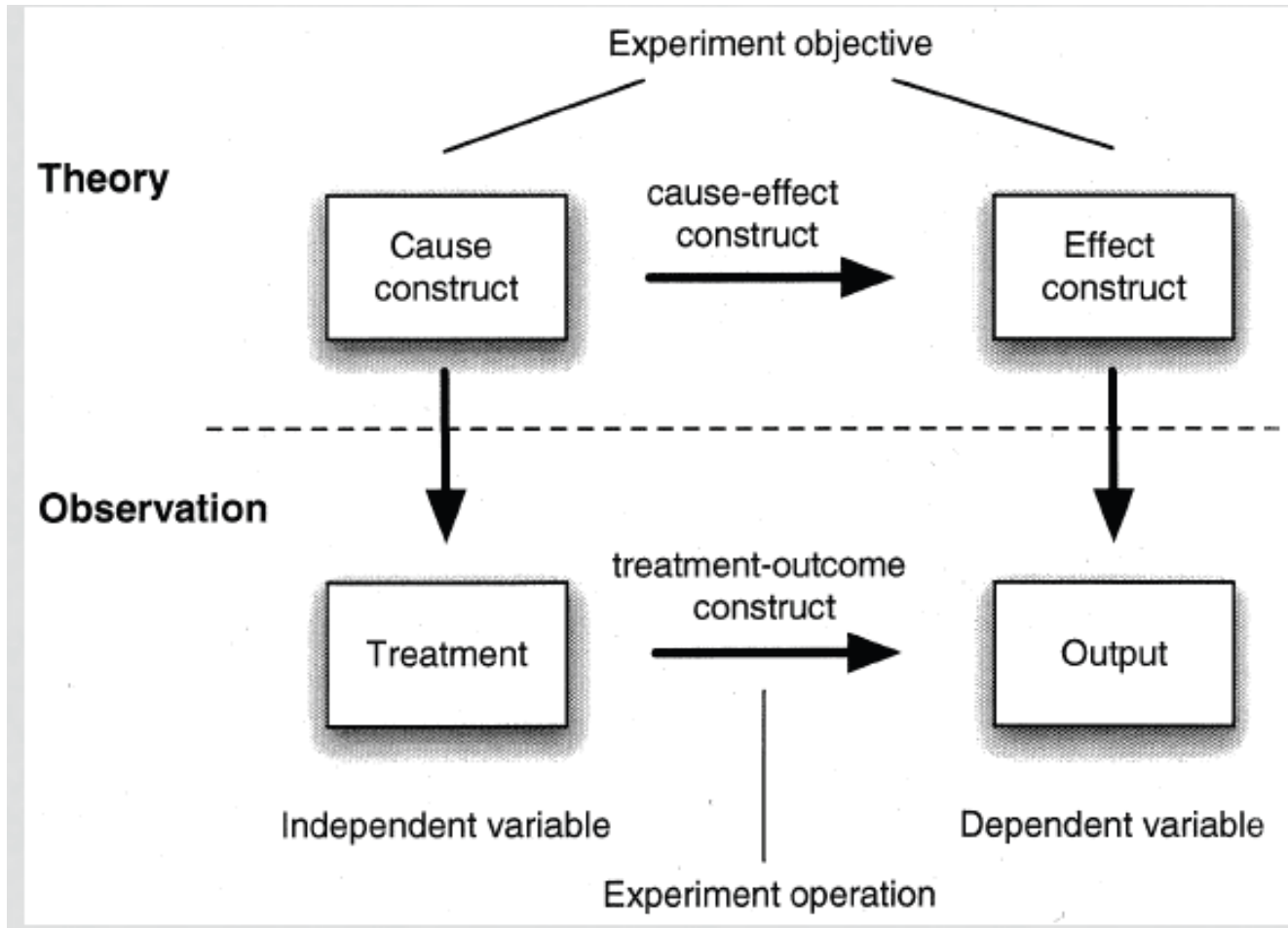
Temos vários tratamentos (variáveis independentes)

Executamos o experimento e observamos a saída (variáveis dependentes)

Se o experimento foi corretamente elaborado (i) é possível formular conclusões a respeito da relação causa-efeito para a hipótese estabelecida; (ii) essas conclusões são generalizáveis para outros elementos observados nas mesmas condições.



# Conceitos básicos da Pesquisa Experimental



# Passos de um Experimento

## **1-Definição do Contexto**

O escopo do experimento é estabelecido em termos do problema existente, objetivos e metas

## **2-Planejamento**

Faz-se o projeto (“desenho”) do experimento, a instrumentação é preparada e as potenciais ameaças à validade do experimento são avaliadas

## **3-Execução**

Faz-se as medições => dados

## **4-Análise e Interpretação**

Dados coletados são analisados com suporte estatístico

## **5-Apresentação e Empacotamento**

Os resultados são apresentados e empacotados

# Hipóteses de Pesquisa

A definição de um experimento é formalizada por meio de hipóteses

Uma hipótese deve ser declarada formalmente os dados coletados durante a execução experimental permitirão verificar a hipótese (rejeitar/aceitar)

Se a hipótese pode ser rejeitada/aceita então

As conclusões são, **em princípio**, generalizáveis para outras situações equivalente, levando em consideração alguns riscos

# Hipótese nula vs Hipóteses alternativas

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

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

**Hipótese alternativa ( $H_a$ )** – declara o oposto, i.e., que 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 do experimento é *rejeitar a hipótese nula* ( $H_0$ ), que contradiz a nossa crença de que existe uma relação causa-efeito em jogo
- Ao observar evidências que permitem rejeitar a hipótese nula, estamos *fortalecendo a hipótese alternativa*, de que a relação causa-efeito existe

# Hipótese nula vs Hipóteses alternativas

$H_0$  é a hipótese que queremos **REJEITAR** com a maior significância (certeza) possível

Exemplo: “Uma nova técnica de inspeção de software encontra, na média ( $\mu$ ), o mesmo número de falhas ( $\#F$ ) do que uma técnica já existente”

$$H_0: \mu_{\#F\_antiga} = \mu_{\#F\_nova}$$

# Hipótese nula vs Hipóteses alternativas

## Hipótese Alternativa ( $H_1$ )

Exemplo: “A nova técnica de inspeção encontra, na média ( $\mu$ ), mais falhas ( $\#F$ ) do que a técnica já existente”.

$$H_1: \mu_{\#F\_antiga} < \mu_{\#F\_nova}$$

# 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

porém, quanto mais hipóteses serão testadas, mais fatores precisarão ser controlados e mais variáveis precisarão ser medidas, aumentando os riscos relativos à validade do experimento



# Exemplos

- $H_0$ : *there is no difference in user satisfaction rating between the pull-down menu and the pop-up menu*
- $H_a$ : *there is a difference in user satisfaction rating between the pull-down menu and the pop-up menu*

# 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  $H_a$  representa a premissa que se deseja verificar ou estabelecer (hipótese de pesquisa), sendo formulada para contradizer a hipótese nula.

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

# Variáveis dependentes e independentes

Ao conduzir um experimento, definimos um **tratamento experimental** que nos permite observar (ou medir) **variáveis de saída**, manipulando (ou controlando) algumas **variáveis de entrada** do processo.

**Variáveis independentes:** aquelas que são manipuladas ou controladas (ou seja, as variáveis de entrada, que podem `causar` efeitos na variável de saída)

**Variáveis dependentes:** aquelas que queremos observar para verificar os efeitos das manipulações aplicadas às variáveis de entrada (ou seja, as variáveis de saída observadas)

# Variáveis dependentes e independentes

## **Independentes**

Todas aquelas que são manipuladas ou controladas

Identificar as v.i. nem sempre é trivial e, normalmente, exige *conhecimento do domínio*

## **Dependentes**

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

Normalmente tem-se apenas uma variável dependente, que muitas vezes *não é diretamente mensurável*

# Variáveis dependentes e independentes

- Ex.
  - $H_0$ : *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
- $H_0$ : Uma técnica nova de inspeção não encontra, na média ( $\mu$ ), mais falhas (#F) do que a técnica antiga
  - Variável independente: a técnica de inspeção
  - Variável dependente: número de falhas encontradas

# Variáveis dependentes e independentes

## Variáveis independentes

São aquelas que *podemos*, em princípio, *controlar e alterar*

Têm algum *efeito* esperado *sobre as variáveis dependentes*

Identificar as variáveis independentes que podem afetar as variáveis dependentes pode não ser trivial. Normalmente, exige um bom *conhecimento do domínio*

## Variáveis dependentes

Sofrem o *efeito dos tratamentos* (o qual pode ser medido)

Normalmente, *define-se uma única variável dependente* derivada diretamente da hipótese

Na maioria das vezes essa variável *não é mensurável diretamente*

# Variáveis dependentes e independentes

**Exemplo:** estudar os efeitos de um novo método de desenvolvimento na produtividade dos desenvolvedores de software, p.ex. um método OO a ser introduzido em substituição a um método funcional

## **Variável dependente**

Produtividade

## **Variáveis independentes**

Método de desenvolvimento

Experiência profissional

Suporte de ferramentas

Ambiente de trabalho

# Exemplos (HCI)

- V.I. típicas:
  - Relacionadas a tecnologias ou dispositivos, p. ex. mouse vs joystick, tipo de menu, técnica A ou B, ...
  - Relacionadas às pessoas, p.ex. idade, gênero, nível educacional, profissão, experiência profissional, limitações físicas, ...
  - Relacionadas ao ambiente, p.ex. nível de ruído, nível de iluminação, ...
- V.D. típicas:
  - Medidas de eficiência, de precisão, de satisfação, de esforço cognitivo, ...



# Componentes do experimento

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

# *Aleatorização*

- A alocação dos diferentes tratamentos experimentais às unidades deve ser **totalmente aleatória**
  - Requisito **essencial para a validade** do estudo experimental
- É a atribuição aleatória dos participantes aos diferentes tratamentos experimentais que potencialmente anula a influência de variáveis ocultas (não consideradas)
- Pode ser necessário “randomizar” outros fatores ou condições, p.ex., a ordem de execução de tarefas...

# Teste de hipótese

- Testes estatísticos da hipótese (teste de significância). Porque são necessários?
- Ex.  $H_0$ : *There is no difference between the pull-down menu and the pop-up menu in the time spent locating pages*
- Porque não podemos simplesmente comparar, p.ex., duas médias da medida de tempo (ou de uma medida de erro) e concluir que o tratamento que produziu o menor valor tem melhor desempenho do que o outro?

# Teste de hipótese

- Testes estatísticos da hipótese (teste de significância). Porque são necessários?
- Porque o experimento observou o comportamento de uma “amostra” da população alvo
  - Até que ponto podemos confiar na conclusão obtida observando o comportamento dessa amostra em particular?
  - Que certeza temos de que esse resultado generaliza para a população como um todo?
  - Até que ponto podemos confiar que o efeito observado não foi causado por outros fatores não considerados? (fatores ocultos)

# 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 ( $H_a$ ) para determinar, em função dos dados observados, se a hipótese nula pode ser rejeitada, e com que **grau de confiança**
  - resultado **positivo**: hipótese nula é **rejeitada**
  - resultado **negativo**: hipótese nula **não pode ser rejeitada**
- O resultado tem uma certa probabilidade de estar correto, mas pode estar errado...

# Testes de significância

- Ex.

*“On average, participants performed significantly better ( $F(1, 25)=20,83, p < 0.01$ )... in condition A than in condition B...”*

*“A t test showed that there was a significant difference in the number of lines of text entered ( $t(11) = 6.28, p < 0.001$ ) with more entered in the tactile condition. ... ”*

# Erros: tipo I e tipo II

Ao testarmos uma hipótese nula, chegamos a uma conclusão: rejeitá-la (positivo), ou não rejeitá-la (negativo)

Entretanto, devemos lembrar que a conclusão pode estar correta, ou pode estar incorreta (mesmo tudo tendo sido feito corretamente!)

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

# Testes de significância

- O teste da hipótese resulta em 4 cenários possíveis:
  1. Hipótese nula é rejeitada (corretamente) (positivo)
  2. Hipótese nula não é rejeitada (corretamente) (negativo)
  3. Hipótese nula rejeitada (incorretamente) (falso positivo)
  4. Hipótese nula não rejeitada (incorretamente) (falso negativo)
- 3 e 4: risco de tomar uma decisão errada...



# Testes de significância

Existem vários testes estatísticos de hipótese

Todos estão baseados na premissa de que as hipóteses são formuladas antes dos testes estatísticos serem escolhidos e realizados

O teste de hipótese envolve dois tipos de risco:

Ou o teste rejeita uma hipótese nula verdadeira (falso positivo)

Ou o teste não rejeita uma hipótese nula falsa (falso negativo)

# Erros: tipo I e tipo II

## Erro Tipo-I

Ocorre quando um teste estatístico rejeita a hipótese nula, sendo que ela é verdadeira e não deveria ser rejeitada (**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, um erro tipo-I é dado pela probabilidade de observar dados que me levem a **rejeitar  $H_0$**  *sendo que na verdade as duas técnicas encontram, na média ( $\mu$ ), o mesmo número de falhas ( $\#F$ )*

# Erros: tipo I e tipo II

## Erro Tipo-II

Ocorre quando um teste estatístico não rejeita a hipótese nula quando ela é falsa, e deveria ter sido rejeitada (**falso negativo**)

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

$$P(\text{type-II-error}) = P(\text{não rejeitar } H_0 \mid H_0 \text{ é falsa})$$

No exemplo apresentado, o erro tipo-II é dado pela probabilidade de observar dados que me levem a **não rejeitar  $H_0$**  *sendo que na verdade as duas técnicas, na média, possuem médias ( $\mu$ ) diferentes do número de falhas ( $\#F$ ) encontradas* (e, portanto,  $H_0$  deveria ter sido rejeitada)

# Erros: tipo I e tipo II

		O Verdadeiro Estado da Natureza	
		A hipótese nula é verdadeira.	A hipótese nula é falsa.
Decisão	Decidimos rejeitar a hipótese nula.	<b>Erro tipo I</b> (rejeição de uma hipótese nula verdadeira)	Decisão correta
	Não rejeitamos a hipótese nula.	Decisão correta	<b>Erro tipo II</b> (Não rejeição de uma hipótese nula falsa)

# Erros: tipo I e tipo II

		Jury decision	
		Not guilty	Guilty
Reality	Not guilty	✓	Type I error
	Guilty	Type II error	✓

Table 2.3 Type I and Type II errors in the judicial case.

# Erros: tipo I e tipo II

		Study conclusion	
		No difference	Touchscreen ATM is easier to use
Reality	No difference	✓	Type I error
	Touchscreen ATM is easier to use	Type II error	✓

**Table 2.4** Type I and Type II errors in a hypothetical HCI experiment.

# Risco do experimento

- Qual o risco de se tomar uma decisão errada?
- Como minimizar esse risco?

# Controle de riscos: erros tipo I e tipo 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 (falso positivo, i.e., erroneamente rejeitar a hipótese nula): *alpha error*, ou grau de significância do teste estatístico
  - Probabilidade de incorrer em um erro do tipo II (falso negativo, i.e., erroneamente aceitar a hipótese nula): *beta error*, ou grau de confiança do teste estatístico



# Controle de riscos: erros tipo I e tipo 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 (falso positivo, i.e., erroneamente rejeitar a hipótese nula): *alpha error*.
  - Probabilidade de incorrer em um erro do tipo II (falso negativo, i.e., erroneamente não rejeitar a hipótese nula): *beta error*.
- Quanto menores essas probabilidades, melhor (em princípio)

# Controle de riscos: erros tipo I e tipo II

- *alpha*: **nível de significância do teste estatístico**
- $(1 - beta)$ : **poder estatístico (confiabilidade) do teste de hipótese**: a probabilidade do teste tomar a decisão de rejeitar a hipótese nula (esperada), quando ela é de fato falsa.

# Erros: tipo I e tipo II

		O Verdadeiro Estado da Natureza	
		A hipótese nula é verdadeira.	A hipótese nula é falsa.
Decisão	Decidimos rejeitar a hipótese nula.	<b>Erro tipo I</b> (rejeição de uma hipótese nula verdadeira) <i>alpha</i>	Decisão correta <i>1 - beta</i>
	Não rejeitamos a hipótese nula.	Decisão correta <i>1 - alpha</i>	<b>Erro tipo II</b> (Não rejeição de uma hipótese nula falsa) <i>beta</i>

# Controle de riscos: erros tipo I e tipo II

Para tanto, ao realizar um experimento devemos escolher um teste com o maior poder estatístico *possível*

$$P = (\text{rejeitar } H_0 \mid H_0 \text{ é falsa}) = 1 - P(\text{type-II-error})$$

Entretanto, *alpha* e *beta* são relacionados: um valor muito baixo para *alpha* reduz a chance de incorrer em erros do tipo I, mas aumenta a chance de incorrer em erros do Tipo II...

# Controle de riscos: erros tipo I e tipo 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 (mais eficaz)*

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

Quais hipóteses devem ser formuladas?

Que erros pode-se cometer?

# Controle de riscos: erros tipo I e tipo II

Exemplo:

*Hipótese nula  $H_0: r \leq 0,25$*

*Hipótese alternativa  $H_a: r > 0,25$*

**Erro tipo I (FP):** rejeitar  $H_0$ , i.e., aprovar a nova vacina quando, na realidade, ela não tem um efeito superior ao da vacina em uso

**Erro tipo II (FN):** aceitar  $H_0$ , i.e., não aprovar a nova vacina quando ela é, de fato, mais eficaz do que a vacina em uso

# 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, é necessário uma amostra de tamanho maior;

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.

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

# Limitações da Pesquisa Experimental

Na prática, pode ser muito difícil construir uma hipótese bem definida e testável para avaliar uma técnica de HCI, ou de Visualização...

Também pode ser inviável fazer um controle estrito dos fatores que podem influenciar a variável dependente – i.e., garantir que esses fatores se mantêm idênticos nas diferentes condições experimentais

p. ex. estudo sobre como jovens e idosos interagem com um dispositivo... muitos *confounding factors* em potencial!

# Limitações da Pesquisa Experimental

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

Outra limitação consiste no fato de que muitas variáveis que poderiam ser tecnicamente manipuladas estão sujeitas as considerações de ordem ética que proíbem sua manipulação.

p.ex., não se pode submeter pessoas a atividades estressantes com vistas a verificar alterações em sua saúde física ou mental

# Task 7 (next class)

- Bring **two** examples of papers (in HCI, Imaging or Visualization) that verify research hypothesis
  - Describe the null and alternative hypotheses in each case
  - Identify the independent variable(s) and the dependent variable
  - Identify the experimental design adopted
  - Identify and explain the choice of statistical significance test applied
- Enviar por email:
  - Tarefa 7 Metodologia Fulano de Tal