Interação Homem-Computador

Avaliação de técnicas de interação e análise de resultados

.inf
INSTITUTO DE INFORMÁTICA
UFRGS

Sumário da aula

- O método experimental
- Variáveis independentes e dependentes
- Tipos de amostras
- Análise de resultados quantitativos

Revisão

O MÉTODO EXPERIMENTAL

Tipos de avaliação

- Cognitive walkthrough (percurso cognitivo)
  - Avaliação feita por experts
  - Cada tarefa é questionada (passo a passo)
- Avaliação heurística
  - Feita por experts em interação, usando guidelines
- Avaliação formativa
  - Usada para refinar widgets, técnicas de interação, metáforas de interação
  - Estudos observacionais com usuários (sessões informais)
  - Questionários e entrevistas (resultados qualitativos)
- Avaliação somativa
  - Usada frequentemente para avaliar um produto finalizado
  - Avaliação de usabilidade baseada em tarefas
  - Experimentação formal (resultados quantitativos)

Experimento controlado: avaliação somativa

- Somativa: mede resultado final
  - Compara diferentes técnicas
  - Muitos usuários, protocolo estrito
  - Variáveis independentes e dependentes
  - Resultados quantitativos
  - Significância estatística

Medindo o desempenho do sistema

- Frame rate médio (fps)
- Latência média (msec)
- Variabilidade do frame rate/latência
- Atraso da rede
- Distorções

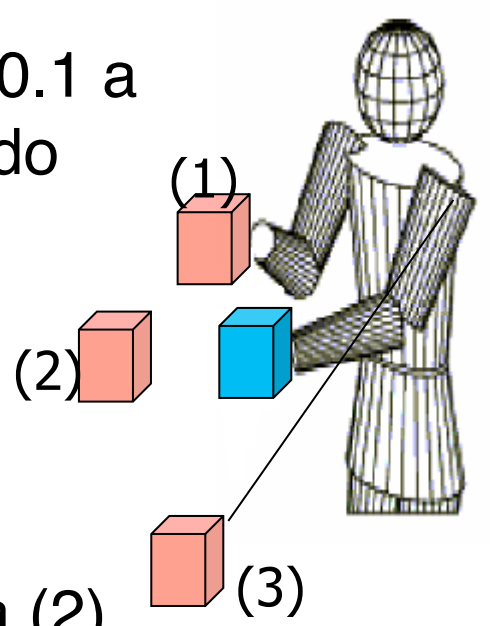
Medindo o desempenho do usuário nas tarefas

- Velocidade/eficiência
- Precisão: número de erros
- Métricas específicas do domínio
  - Educação: aprendizado
  - Treinamento: consciência espacial
  - Design: expressividade

- O que é quantitativo e o que é qualitativo?

Experimento com objeto virtual

- Tarefa
  - Um cubo semi-transparente vermelho deveria ser sobreposto pelo cubo azul opaco preso a mão do sujeito. O objeto foi colocado num ponto inicial variando de 0.1 a 0.6 metros, ou preso à mão dominante do sujeito



- Condições
  - Objetos presos a mão dominante (1)
  - Objetos localizados a uma distância fixa (2)
  - Objetos colocados a uma distância variável em relação à extensão do braço do sujeito (3)

Experimento com objeto virtual

- Critério
  - Rapidez para executar a tarefa
- Sujeitos
  - 7 mulheres e 11 homens
- Resultado
  - A comparação dos tempos mostrou que manipular objetos presos à mão foi significativamente mais rápido que manipular objetos localizados a uma distância fixa e objetos colocados a uma distância variável

Método científico

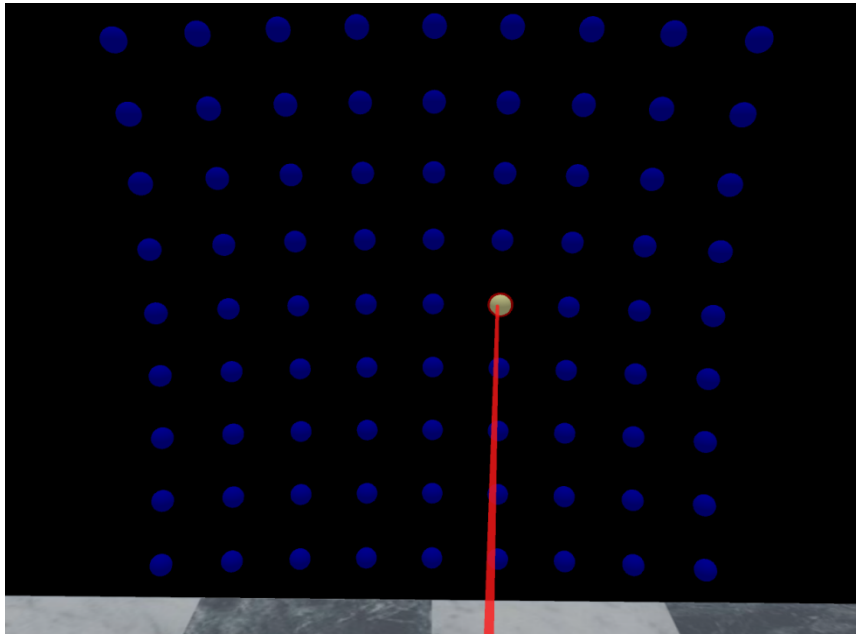
1. Formar hipóteses (hipótese real, hipótese nula)
2. Coletar dados (como planejar a amostragem?)
3. Analisar dados (o que usar?)
4. Aceitar/rejeitar hipóteses

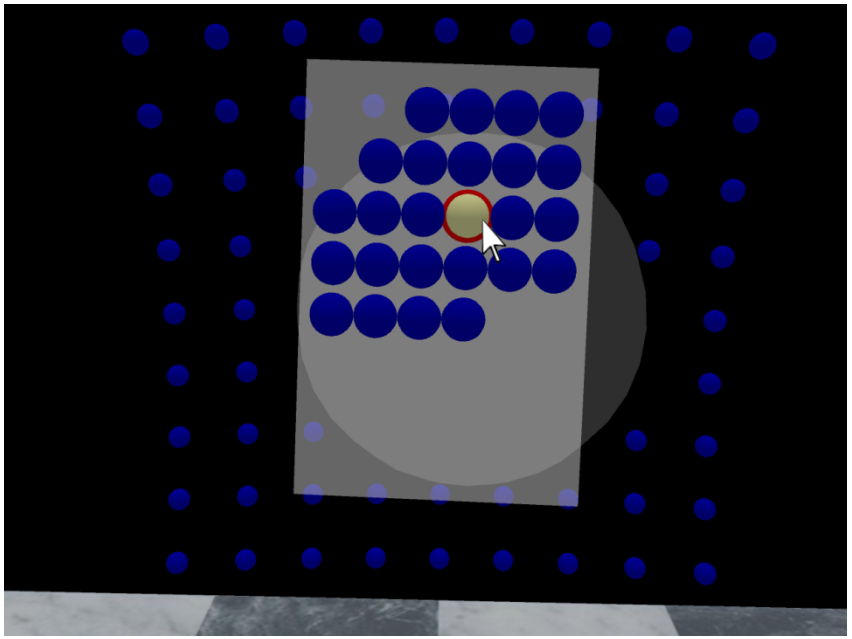
- Como provar uma hipótese?
  - Mais fácil é comprovar o inverso de uma condição, por contra-exemplo
  - “Hipótese nula” = oposto da hipótese
  - Provar a falsidade da hipótese nula
  - Então, a hipótese fica provada como verdadeira

Experimento Empírico

- Questão típica:
  - Qual técnica de interação é melhor?

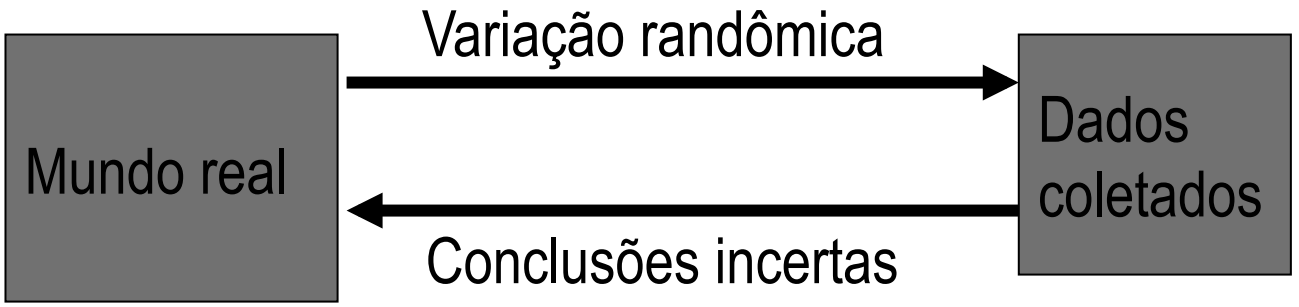
Raio virtual vs. LOP cursor





Causa e efeito

- Meta: determinar causa e efeito
  - Causa = técnica (Raio vs. LOP)
  - Efeito = tempo para completar a tarefa T
- Procedimento:
  - Variar a causa
  - Medir o efeito
- Problema: a causa é uma variação randômica?



Estatísticas

- **Meta:**
  - Provar que o efeito medido não é resultado de uma variação randômica no ambiente
- **Hipótese:**
  - Por exemplo, a causa do efeito é a tecnica de interação (Raio virtual ≠ LOP cursor)
- **Hipótese nula :**
  - A técnica de interação não tem efeito nenhum (ou tanto faz usar raio virtual ou LOP cursor: raio virtual = LOP cursor)
  - Então: causa é variação randômica
- **Estatísticas:**
  - Se a hipótese nula é verdadeira, o efeito medido ocorre com probabilidade < 5%
  - Mas se o efeito ocorreu, efeito medido >> variação randômica
- **Então:**
  - Hipótese nula provavelmente é falsa
  - Hipótese provavelmente é verdadeira

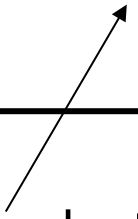
Variáveis do experimento

- Variáveis independentes (o que se altera), e “tratamentos” (os valores das variáveis independentes):
  - Técnica de interação
    - Raio virtual, LOP cursor
  - Tamanho do objeto
    - 1cm, 2cm e 3cm
  - Densidade de objetos
    - Grid de 9x9, 18x18 e 27x27
- Variáveis dependentes (o que é medido)
  - Tempo para completar a tarefa
  - Número de erros

Exemplo: experimento com design 2 x 3

Var ind. 2: Tipo de tarefa

Var ind. 1: Técnica	Tarefa 1	Tarefa 2	Tarefa 3
Raio virtual			
LOP cursor			



Var. dependente (tempo de completude da tarefa)

- *n* usuários por célula

Composição da amostra em grupos

- “With-in subjects” (medidas repetidas)
  - Todos os usuários executam todos os tratamentos
  - Eliminar o efeito de ordem de execução
  - Grupo 1: 5 usuários, Raio virtual e depois LOP cursor
  - Grupo 2: 4 usuários, LOP cursor e depois Raio virtual
  - Total: 9 usuários, 9 por célula
  - (mais usuários por vir)

Procedimento

- Para cada um dos *n* usuários:
  - *Pre-survey* de caracterização
  - Instruções técnica 1
    - Não definir o objetivo do experimento
  - Treinamento prévio da técnica 1
  - Execução real com tomada de tempo da técnica 1
  - Questinário sobre a técnica 1
  - Instruções técnica 2
    - Não definir o objetivo do experimento
  - Treinamento prévio da técnica 2
  - Execução real com tomada de tempo da técnica 2
  - Questinário sobre a técnica técnica 2
  - *Post-survey*: medidas subjetivas para comparação

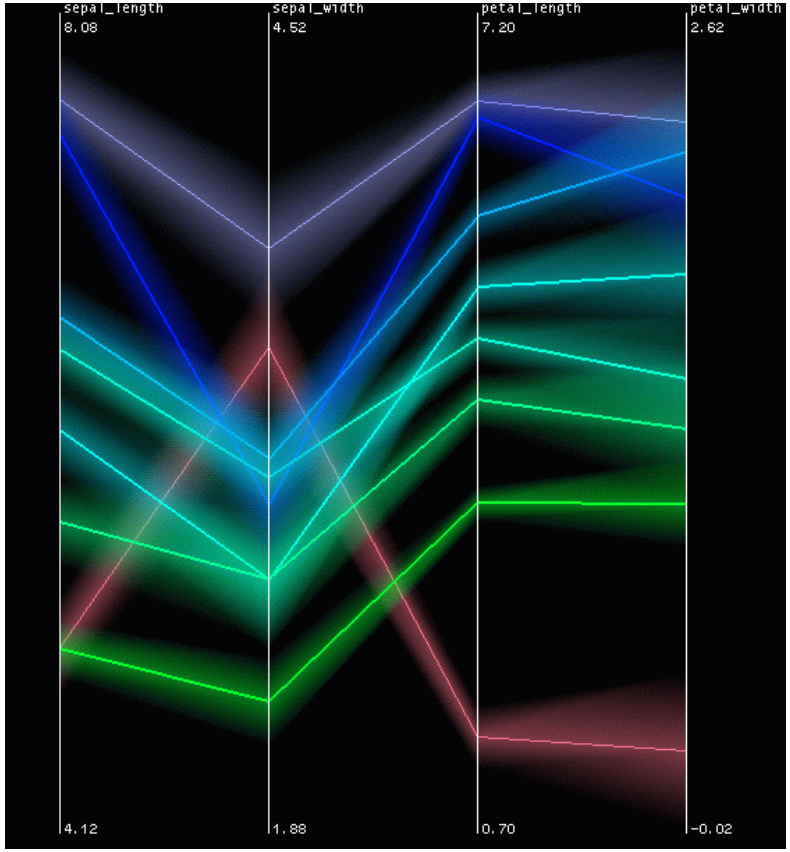
Dados

- Variáveis dependentes medidas
- Planilha

Usuário	LOP cursor			Raio virtual		
	tarefa1	tarefa2	tarefa3	tarefa1	tarefa2	tarefa3

Primeiro passo: ver dados brutos

- Observar fatos interessantes
  - Identificar padrões
  - Identificar *outliers*
- Conclusões qualitativas
- Determinar estatísticas
- Determinar futuros experimentos



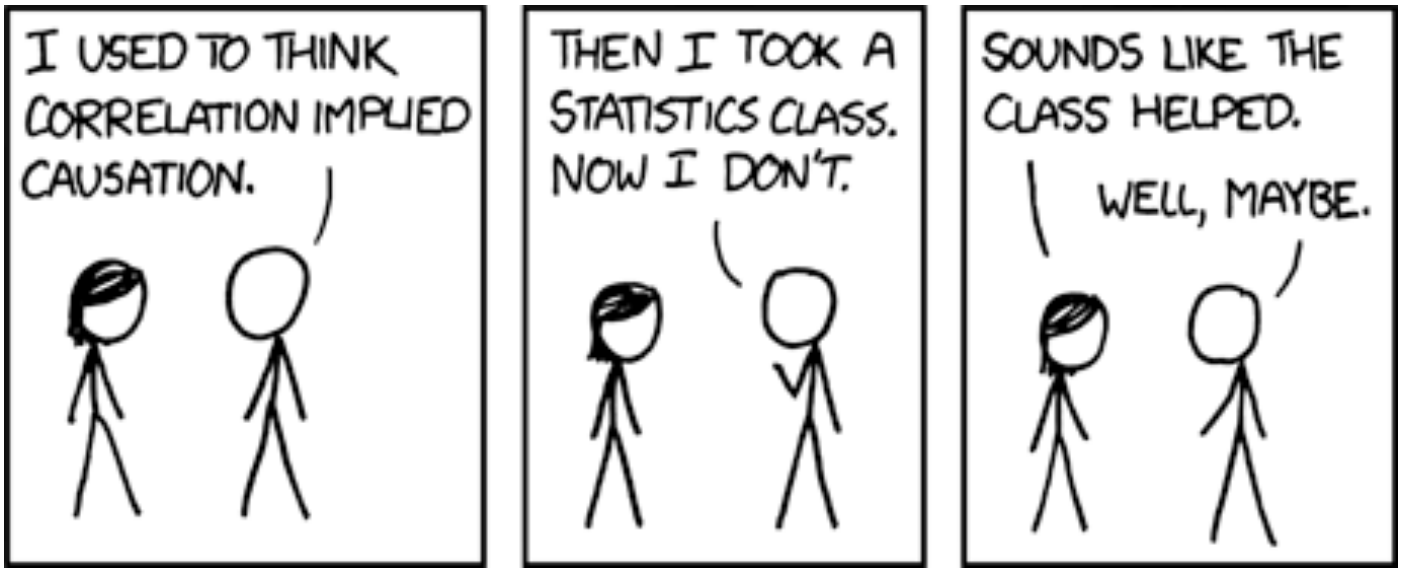
Segundo passo: estatísticas

Var ind. 2: tipo de tarefa

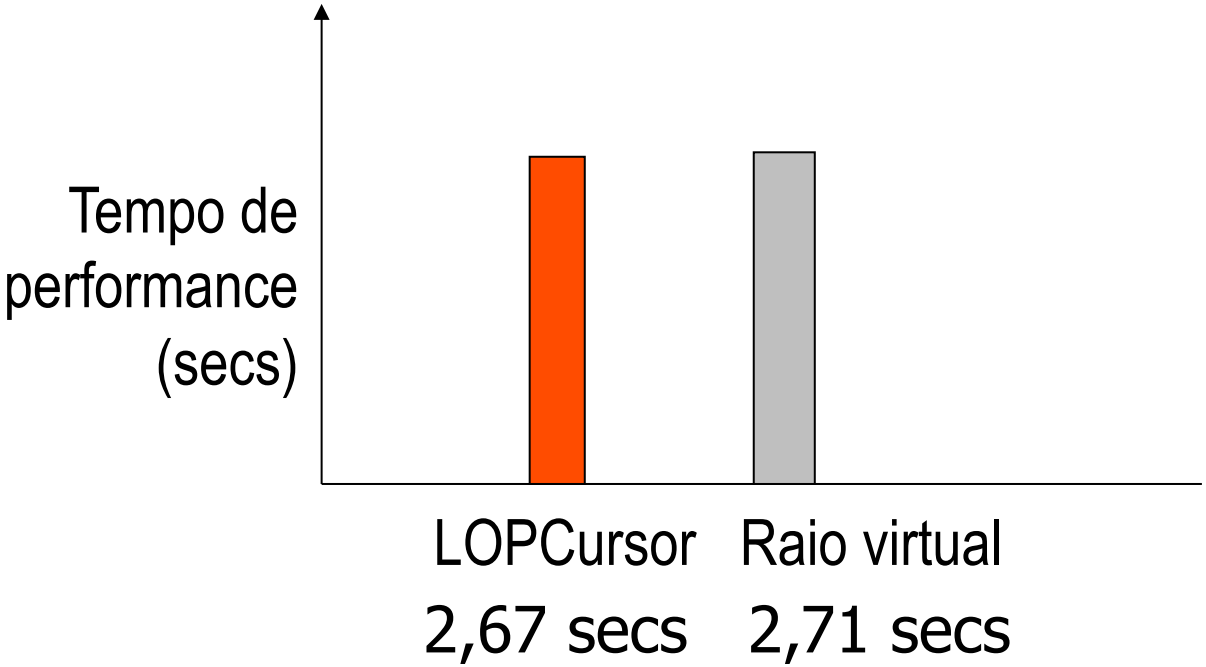
Var ind. 1: técnica	Tarefa1	Tarefa2	Tarefa3
Raio virtual	37.2	54.5	103.7
LOP cursor	29.8	53.2	145.4

Var. dep: média de performance dos usuários

Atenção: correlação ≠ casualidade

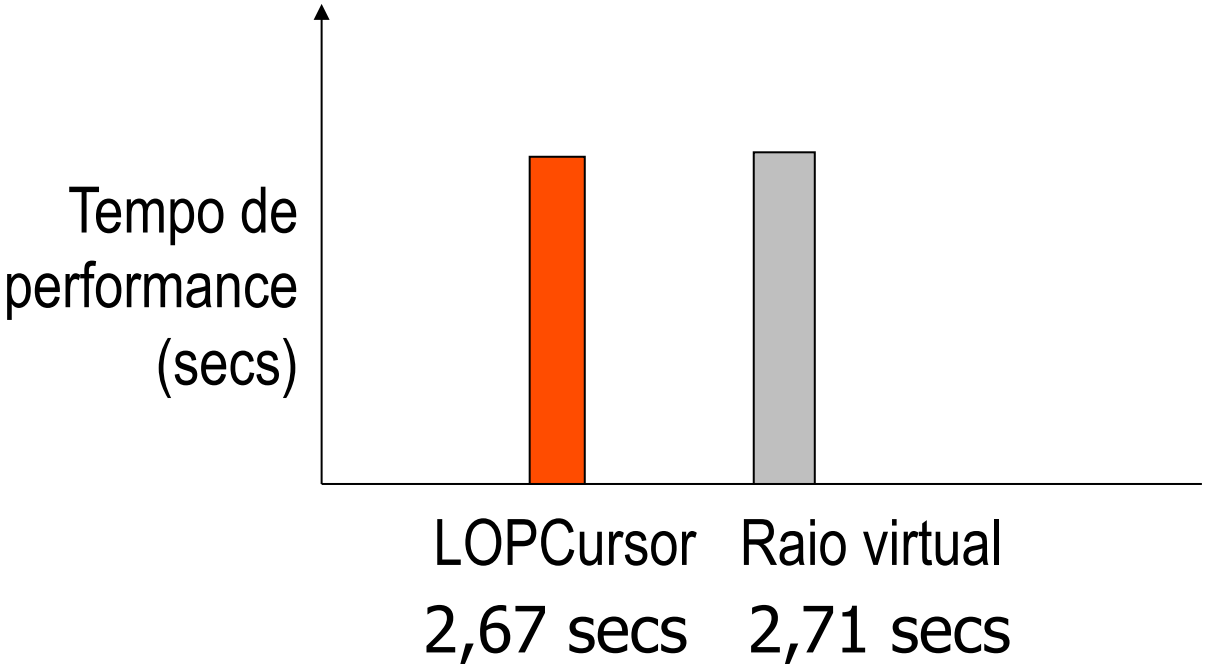


Raio virtual melhor que LOP cursor?



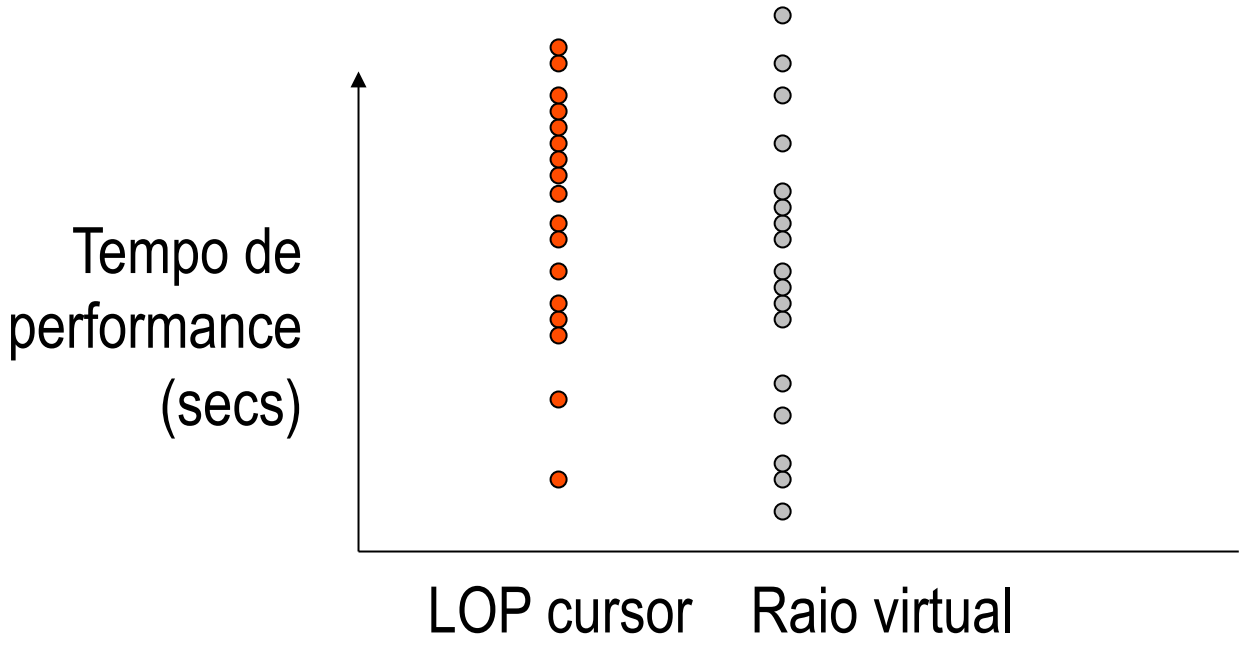
- Raio virtual
  - 18,77% de seleções erradas erros
- LOPCursor
  - 1,75% de seleções erradas

Raio virtual melhor que LOP cursor?



- Problema com médias: há perdas
  - Comparação de somente 2 números
  - O que fazer com os 40 valores?

A realidade ….



- É necessário comparar todos os dados.

Análise quantitativa

- Como comparar médias e ver se são estatisticamente diferentes?
- Qual o “melhor” tratamento?

- Teste t de Student
  - Comparar 1 var. dependente obtida de 2 tratamentos de 1 variável independente
    - Comparar resultados de duas amostras

Análise quantitativa: teste t

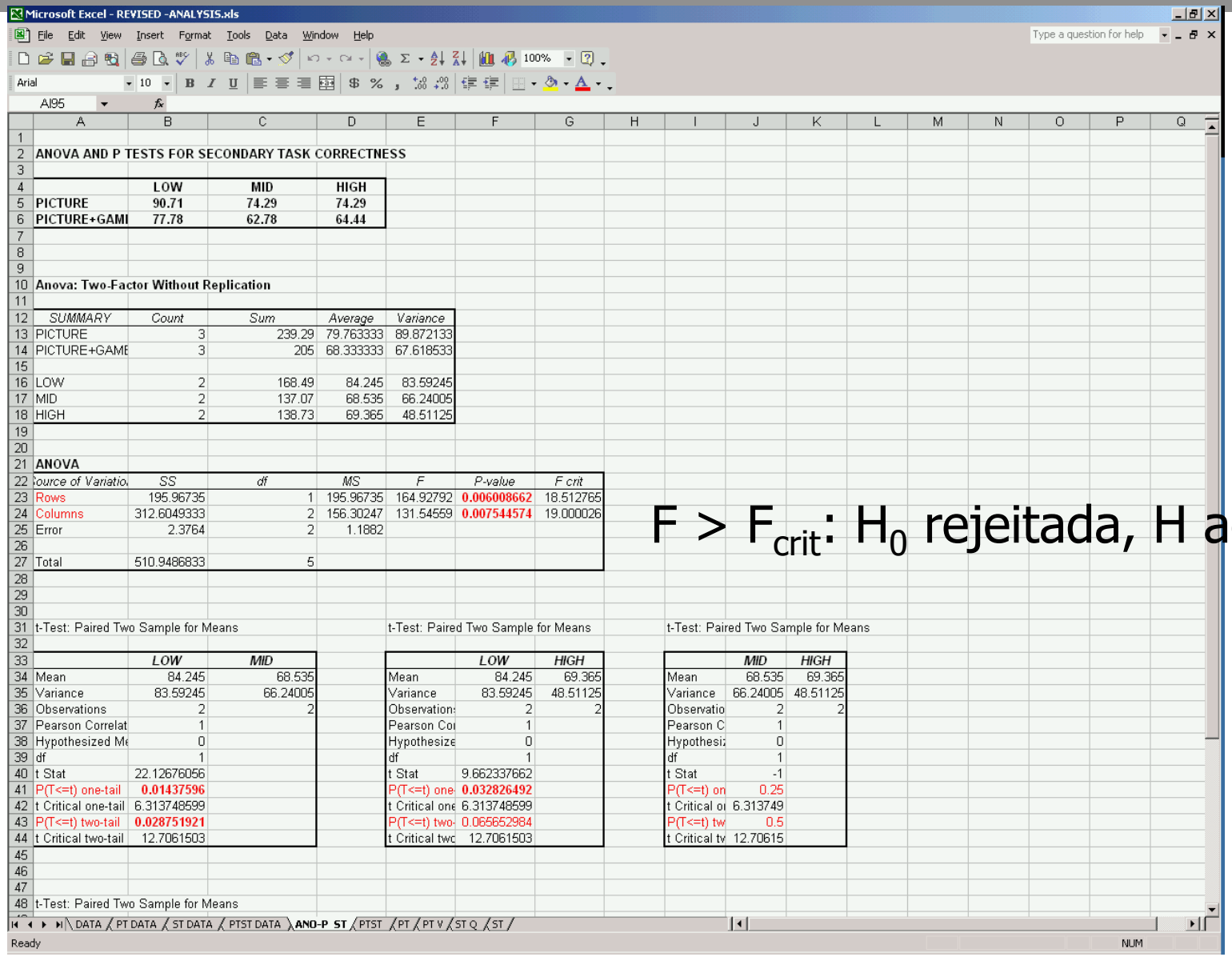
- Resultado
  - p = probabilidade de que a diferença entre os resultados dos diferentes tratamentos seja randômica (hipótese nula)
  - Nível de significância estatística
    - Valor típico: $p < 0.05$
    - Confiança na hipótese = 1 – p, ou seja, 95 % de chance da hipótese ser verdadeira

Análise quantitativa: ANOVA

- ANOVA: *Analysis of Variance*
  - Comparar 1 var. dependente obtida de n tratamentos com m variáveis independentes.
    - Comparar resultados de n amostragens para 1 variável dependente

- Resultado
  - p = probabilidade de que a diferença entre os resultados dos diferentes tratamentos seja randômica (hipótese nula)
  - Nível de significância estatística
    - Valor típico: $p < 0.05$
    - Confiança na hipótese = 1 – p, ou seja, 95 % de chance da hipótese ser verdadeira

Excel



Microsoft Excel - REVISED -ANALYSIS.xls

File Edit View Insert Format Tools Data Window Help

Arial 10 — A195

ANOVA AND P TESTS FOR SECONDARY TASK CORRECTNESS

	LOW	MID	HIGH
PICTURE	90.71	74.29	74.29
PICTURE+GAMI	77.78	62.78	64.44

Anova: Two-Factor Without Replication

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
PICTURE	3	239.29	79.763333	89.872133
PICTURE+GAME	3	205	68.333333	67.618533
LOW	2	168.49	84.245	83.59245
MID	2	137.07	68.535	66.24005
HIGH	2	138.73	69.365	48.51125

ANOVA

ource of Variatio	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	195.96735	1	195.96735	164.92792	**0.006008662**	18.512765
Columns	312.6049333	2	156.30247	131.54559	**0.007544574**	19.000026
Error	2.3764	2	1.1882			
Total	510.9486833	5				

t-Test: Paired Two Sample for Means

	LOW	*MID*
Mean	84.245	68.535
Variance	83.59245	66.24005
Observations	2	2
Pearson Correlat	1	
Hypothesized Me	0	
df	1	
t Stat	22.12676056	
P(T<=t) one-tail	**0.01437596**	
t Critical one-tail	6.313748599	
P(T<=t) two-tail	**0.028751921**	
t Critical two-tail	12.7061503	

t-Test: Paired Two Sample for Means

	LOW	*HIGH*
Mean	84.245	69.365
Variance	83.59245	48.51125
Observations	2	2
Pearson Cor	1	
Hypothesize	0	
df	1	
t Stat	9.662337662	
P(T<=t) one-	**0.032826492**	
t Critical one	6.313748599	
P(T<=t) two-	0.065652984	
t Critical two	12.7061503	

t-Test: Paired Two Sample for Means

	MID	*HIGH*
Mean	68.535	69.365
Variance	66.24005	48.51125
Observatio	2	2
Pearson C	1	
Hypothesi	0	
df	1	
t Stat	-1	
P(T<=t) on	0.25	
t Critical o	6.313749	
P(T<=t) tw	0.5	
t Critical tv	12.70615	

t-Test: Paired Two Sample for Means

DATA / PT DATA / ST DATA / PTST DATA / ANO-P ST / PTST / PT / PT V / ST Q / ST

Ready NUM

$F > F_{crit}$: H_0 rejeitada, H aceita

$t > t_{obtido}$ combinado com P(t): H_0 rejeitada, H aceita

Quando $p < 0.05$

- Encontrada diferença significante estatisticamente
- As médias determinam o que é melhor
- Conclusão:
  - Causa = técnica de interação (e.g. Raio virtual; ≠ LOP cursor)
  - “A técnica de interação tem efeito sobre a performance do usuário na tarefa T …”
  - 95% confiança de que Raio Virtual é melhor que LOP Cursor, por exemplo
  - 5% chance de estar errado

Quando p > 0.05

- Quer dizer que não há diferença?
  - Quer dizer que a técnica de interação não tem efeito sobre a performance na tarefa T?
  - Quer dizer que Raio virtual = LOP cursor?
- Errado:
  - Apenas não foi detectada diferença!
  - Efeito real não “venceu” a variação randômica
  - Fornece indícios de que as técnicas são “iguais”, mas não prova
  - Ou seja, não se encontrou nada ☹
- Por quê?
  - Número insuficiente de usuários?
  - Tarefas mal especificadas?

Fim

- Leitura interessante:
  - http://norvig.com/experiment-design.html
- Teste de hipóteses:
  - http://www.mat.ufrgs.br/~viali/exatas/material/laminas/THipoteses_1.pdf
  - http://www.mat.ufrgs.br/~viali/exatas/material/laminas/THipoteses_2.pdf