Métodos Quantitativos para Ciência da Computação Experimental
Aula #4

Jussara Almeida
DCC-UFMG
2017

Apresentacao derivada dos slides originais de Virgilio Almeida

"Measurements are not to provide numbers, but insights"

Apresentacao derivada dos slides originais de Virgilio Almeida

Metodologia de Comparação de Sistemas Experimentais

- Comparando quantitativamente sistemas experimentais:
  - Algoritmos, protótipos, modelos, etc
- Significado de uma amostra
- Intervalos de confiança
- Tomando decisões e comparando alternativas
- Considerações especiais sobre intervalo de confiança
- Tamanho das amostras

Apresentacao derivada dos slides originais de Virgilio Almeida

Estimando os Intervalos de Confiança

- Duas fórmulas para intervalo de confiança
  - Acima de 30 amostras de qualquer distribuição: distribuição-*z (Normal)*
  - Pequenas amostras de populações normalmente distribuídas: distribuiçao-*t (Student)*

- Um erro comum: usar distribuição-*t para populações não normalmente distribuídas.*

Intervalo de Confiança da Média da Amostra

- Chave: Teorema *Central* do *Limite*
  - As médias de amostras são distribuidas pela Normal.
  - Desde que sejam independentes
  - Média das medias converge para a média da população μ
  - Desvio padrão (*erro padrão) é* $\sigma/\sqrt{n}$

A Distribuição-z

- Intervalo em cada lado da média:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \qquad \bar{x} \pm z_{1-\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

- O nível de significância α *é pequeno* para níveis maiores de confiança (100*(1- α)%)
- Existem tabelas para a variável *z*!

A Distribuição t

- Fórmula quase a mesma:

$$\bar{x} \pm t_{\left[1-\alpha/2;n-1\right]}\left(\frac{s}{\sqrt{n}}\right)$$

- Usável para populações normalmente distribuídas!
- Funciona para pequenas amostras
- Similar a Normal (bell-shaped, porem mais espalhada) e depende do tamanho da amostra n.

Como calcular Intervalo de Confiança

- Você pode calcular intervalos de confiança para diferentes estatísticas

$$\hat{y} \pm z_{1-\alpha/2} s_y \quad \text{ou} \quad \hat{y} \pm t_{1-\alpha/2,df} s_y$$

onde df são os graus de liberdade na estimativa ŷ

Isto é só pra você saber que as fórmulas dadas nos slides anteriores podem ser aplicadas apenas se você estiver trabalhando com médias

- Erro comum: usar distribuição t para população não – Normal
  - Tipicamente aproximação está ok (Teorema do Limite Central)

Tomando decisões sobre os dados experimentais

- Sumarizar o erro na média da amostra
  (ou qualquer outra estatística obtida a partir dela)
  - Confiança = 1- α
  - Precisão = 100% - metade do intervalo / média
- Prover elementos para saber se a amostra é significativa (estatisticamente)
- Permitir comparações à luz dos erros

Apresentacao derivada dos slides originais de Virgilio Almeida

Testando a Média Zero

- A média da população é significativamente não-zero?
- Se o intervalo de confiança inclui 0, a resposta é *não!*
- Pode-se testar para qualquer valor:
  - Suponha IC de 90% para precisao média de método A:

    0.875 ± 0.12 (0.755, 0.995)
  - A precisão pode ser 0.96?
    - Sim, com 90% de confiança, já que o IC contém 0.96
  - Qual erro máximo na minha estimativa da precisão média?
    - Com 90% de confiança, o erro máximo é

$$\frac{0.995 - 0.875}{0.875} = 13.2\%$$

Apresentacao derivada dos slides originais de Virgilio Almeida

Comparando Alternativas

- Num projeto de pesquisa, geralmente, procura-se o melhor sistema, o melhor algoritmo:
  - Exemplos:
    - Determinar o sistema que apresenta a melhor relação QoS-preço, onde QoS é medido experimentalmente.
    - Mostrar que um algoritmo *Y* executa mais rápido que outros existentes e sejam similares funcionalmente.

- Métodos diferentes para observações pareadas (com par) e não pareadas (sem par).
  - Pareadas se o i-*ésimo* teste em cada sistema foi o mesmo
  - *Não pareadas,* caso contrário

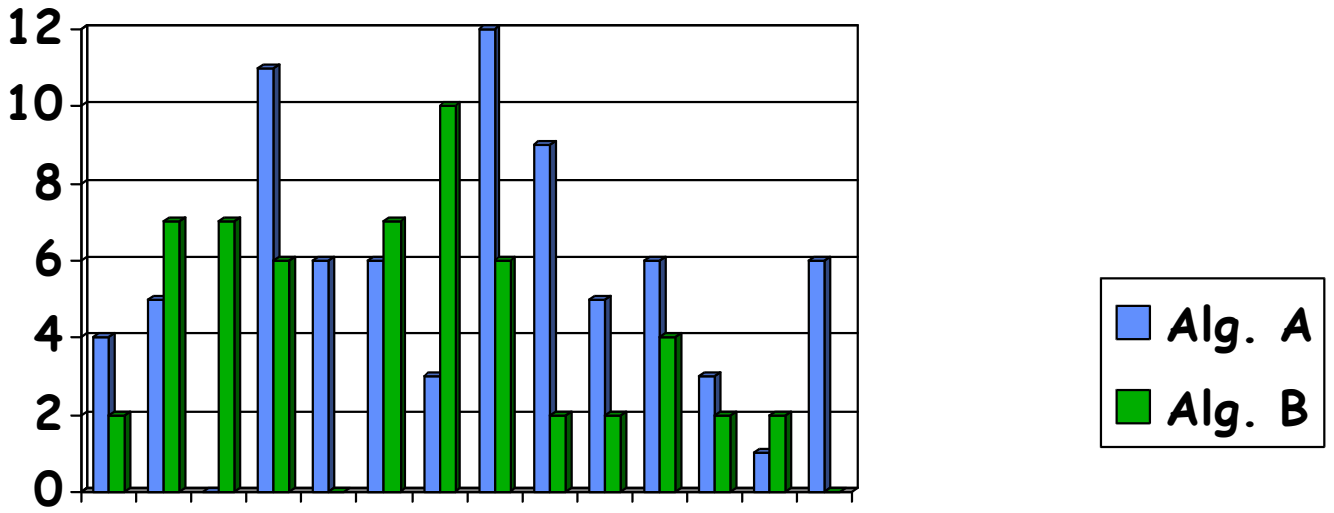
Comparando Observações Pareadas: método

1. Tratar o problema como uma amostra de n pares
2. Para cada teste: calcule as diferenças dos resultados
3. Calcule o intervalo de confiança para a diferença media
4. Se o intervalo inclui 0 (zero), os objetos de comparação (ex.: sistemas, algoritmos, etc) não são diferentes com a dada confianca
5. Se o intervalo não inclui zero, o sinal da diferença indica qual dos objetos é melhor, baseado nos dados experimentais.

Apresentacao derivada dos slides originais de Virgilio Almeida

Exemplo: Comparando Observações Pareadas

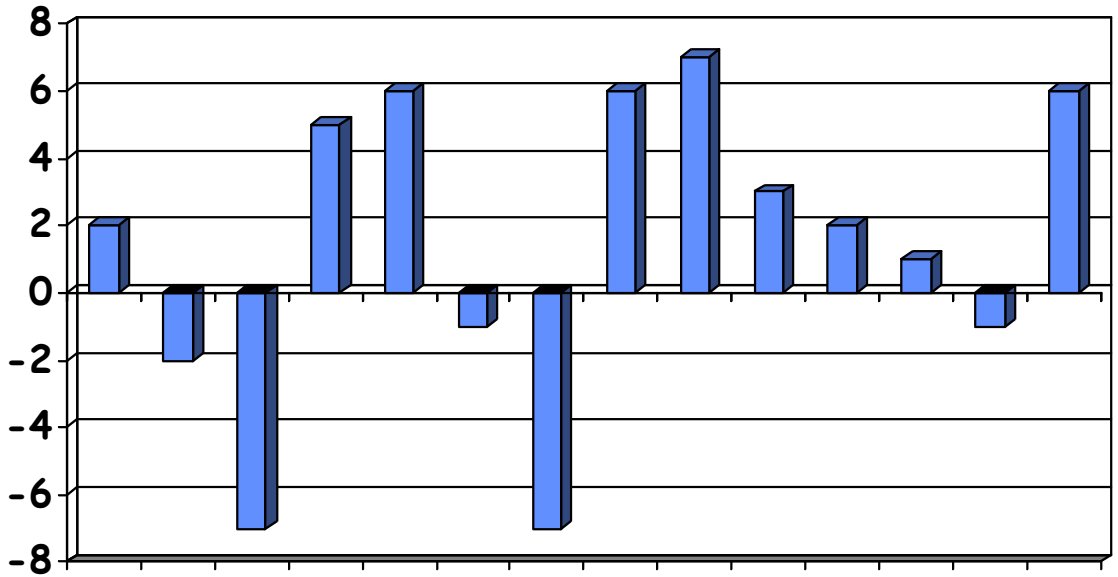
- Considere dois metodos de busca A e B que sao avaliados em funcao do # de documentos relevantes (em um total de 100) que cada um retorna
- Num teste com várias consultas, o algoritmo A retorna mais documentos relevantes que o B?
- Amostra de testes com 14 consultas:



Alg. A	4	5	0	11	6	6	3	12	9	5	6	3	1	6
Alg. B	2	7	7	6	0	7	10	6	2	2	4	2	2	0

Apresentacao derivada dos slides originais de Virgilio Almeida

Exemplo: Comparando observações pareadas



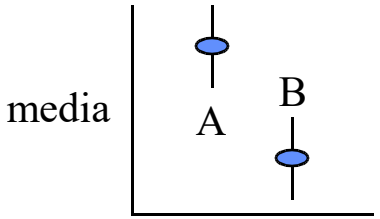
- Diferenças entre algoritmos A-B: 2 -2 -7 5 6 -1 -7 6 7 3 2 1 -1 6
- Média 1.4, intervalo de 90% $\Rightarrow$ (-0.75, 3.6)
  - Não se pode rejeitar a hipótese que a diferença é 0 e que portanto os algoritmos tem desempenho similar.
  - Intervalo de 70% é (0.10, 2.76), A tem desempenho melhor que B

Apresentacao derivada dos slides originais de Virgilio Almeida

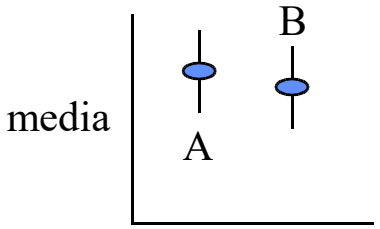
Comparando Observações Não Pareadas
o número de experimentos comuns não precisa ser o mesmo!

- Considere as médias das amostras para cada uma das alternativas, A e B, $\overline{x_a}$ e $\overline{x_b}$

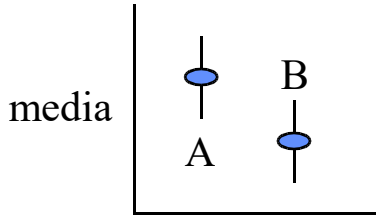
- Comece com os intervalos de confiança
  - Se não houver sobreposição:
    - Algoritmos são diferentes e a maior média é melhor (pelas métricas usadas)



  - Se houver sobreposição e cada IC contem a outra média:
    - Algoritmos não são diferentes neste nivel



  - Se houver sobreposição e uma média não está no outro IC
    - Tem de fazer o teste-***t***



O teste-t (1)

1 Compute as médias das amostras $\bar{x}_a$ e $\bar{x}_b$

2. Compute os desvio-padrões s_a e s_b

3. Compute a diferença das médias = $\bar{x}_a - \bar{x}_b$

4. Compute o desvio padrão das diferenças:

$$s = \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}$$

O teste-t (2)

5. Compute os graus efetivos de liberdade:

$$\nu = \frac{\left(s_a^2/n_a + s_b^2/n_b\right)^2}{\frac{1}{n_a - 1}\left(\frac{s_a^2}{n_a}\right)^2 + \frac{1}{n_b - 1}\left(\frac{s_b^2}{n_b}\right)^2} - 2$$

6. Compute o intervalo de confiança:

$$\left(\bar{x}_a - \bar{x}_b\right) \mp t_{[1-\alpha/2;\nu]} s$$

7. Se o intervalo inclui zero, não há diferença

Exemplo

O tempo de processamento necessario para executar uma tarefa foi medido em dois sistemas.

Os tempos no sistema A foram:
{5.36, 16.57, 0.62, 1.41, 0.64, 7.26}.

Os tempos no sistema B foram:
{19.12, 3.52, 3.38, 2.50, 3.60, 1.74}.

Os dois sistemas sao significativamente diferentes?

Exemplo

Sistema A:

$$\text{media } \overline{x_A} = 5.31$$

$$\text{variancia } s_A^2 = 37.92$$

$$n_A = 6$$

Sistema B:

$$\text{media } \overline{x_A} = 5.64$$

$$\text{variancia } s_A^2 = 44.11$$

$$n_A = 6$$

Exemplo

Diferenca das medias $\overline{x_A} - \overline{x_B} = -0.33$

Desvio Padrao da diferenca media $s = 3.698$

Graus de liberdade $v = 7.943$

0.95 − *quantil* da va t com 8 graus de liberdade

$t_{[0.95,12]} = 1.86$

IC de 90% para a diferenca das medias = (-7.21, 6.54)

IC inclui 0, logo os dois sistemas NAO sao diferentes neste nivel de confianca

Apresentacao derivada dos slides originais de Virgilio Almeida

Outros Testes

- Teste t assume a população segue distribuição Normal
- Caso esta premissa não seja válida:
  - Testes não paramétricos (distribution-free tests)
  - Importante principalmente para amostras pequenas
  - Desvios da normalidade causam grandes distorções nos resultados

Apresentacao derivada dos slides originais de Virgilio Almeida

Outros Testes

- Mann-Whitney-Wilcoxon test (Wilcoxon Rank-Sum test)
  - Dados são amostras aleatórias das populações analisadas
    - Amostras são independentes
    - Dados em uma amostra são independentes
  - Explora a "soma" dos rankings dos dados das amostras
    - Calcula estatística U, cuja distribuição é conhecida e tabelada (como a distribuição t)
    - Procedimento equivalente ao t-test (mas não explora médias)
  - Quase tão eficiente quanto t-test se população é Normal

- Wilcoxon signed rank test (pareado)

Apresentacao derivada dos slides originais de Virgilio Almeida

Outros Testes

- Comparação de n > 2 sistemas
  - Correção de Bonferroni : controle da taxa de erro
  - Testa cada par de sistemas com nível de significância α/k, onde k é o número de comparações sendo feitas
    - Maior confiança na comparação de cada par
    - k = C (n,2)

Comparando Proporções

- Se n_1 de n experimentos dão um certo resultado, então pode-se dizer que a proporção das amostras é dada por:

$$p = \frac{n_1}{n}$$

- Exemplos:
  - A precisão do algoritmo A de recuperação de informação foi superior a precisão de B em 55 dos 100 casos analisados. Com 90% de confiança pode-se dizer que A supera B em precisão?
  - 5000 “samples” coletados, em 1000, o percentual de “system time” foi inferior a 20%. Com 95% de confiança, qual o intervalo de confiança para a porcentagem das vezes em que o sistema operacional gasta menos de 20% dos recursos?

Comparando Proporções

- Se n_1 de n experimentos dão um certo resultado, então o intervalo de confiança (IC) para a proporção:

$$IC \rightarrow p \pm z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

- A fórmula acima é baseada numa aproximação da distribuição binomial por uma normal que e valida somente se $np>10$

Apresentacao derivada dos slides originais de Virgilio Almeida

Exemplo

- Um experimento foi repetido em dois sistemas 40 vezes. O sistema A foi superior ao B em 26 repeticoes. Podemos dizer que, com confianca de 99%, o sistema A e superior ao sistema B (na maioria daz vezes)? E com uma confianca de 90%?

Exemplo

$$\text{proporcao } p = \frac{26}{40} = 0.65$$

desvio padrao da estimativa de proporcao:

$$s = \sqrt{\frac{p(1-p)}{n}} = 0.075$$

$$\text{IC de } 99\% = 0.65 \pm (2.576)0.075 = (0.46, 0.84)$$

IC inclui 0.5, logo A nao e superior a B neste nivel de confianca

$$\text{IC de } 90\% = 0.65 \pm (1.645)0.075 = (0.53, 0.77)$$

IC nao inclui 0.5, logo A e superior a B neste nivel de confianca

Considerações Especiais

1. Selecionar um intervalo de confiança para trabalhar
2. Teste de Hipótese
3. Intervalos de confiança de um único lado

A Seleção um Intervalo de Confiança

- Depende do custo de se estar errado!!!
  - Produção de um *paper* científico
  - Demonstração de um novo algoritmo experimentalmente
  - Geração de um produto
- Os níveis de confiança entre 90% e 95% são os valores comuns para *papers* cientificos (em Computacao)
- Em geral, use o maior valor que lhe permita estabelecer conclusões sólidas num processo experimental!
- Mas é melhor ser consistente durante todo o paper que se está trabalhando.

Apresentacao derivada dos slides originais de Virgilio Almeida

Teste de Hipótese

- A *null hypothesis* (H_0) é comum em estatísticas e tratamento de dados experimentais:
  - Pode ser confuso em negativas duplas
  - Provê menos informação que intervalos de confiança
  - Em geral mais dificil de interpretar/entender

- Deve-se entender que rejeitar a hipótese nula implica que o resultado é significativo.

Apresentacao derivada dos slides originais de Virgilio Almeida

Intervalos de Confiança de um-Lado

- Intervalos de dois lados testam se a média está fora ou dentro de uma variação definida pelos dois lados do intervalo
  - Ex: Com 90% de confiança, a média da população pode ser valor alvo?

- Teste de intervalos de um único lado são úteis somente quando se está interessado em um limite.
  - Ex.: Com 90% de confiança, média da população é menor/maior que um certo valor alvo?

Intervalos de Confiança de um-lado

Limite inferior

$$\left(\bar{x} - t_{[1-\alpha;n-1]} \frac{s}{\sqrt{n}}, \infty \right)$$

Limite superior

$$\left(-\infty, \bar{x} + t_{[1-\alpha;n-1]} \frac{s}{\sqrt{n}}, \right)$$

Apresentacao derivada dos slides originais de Virgilio Almeida

Intervalos de Confiança de um-lado

Limite inferior

$$\left(\bar{x} - t_{[1-\alpha;n-1]} \frac{s}{\sqrt{n}}, \infty \right)$$

Limite superior

$$\left(-\infty, \bar{x} + t_{[1-\alpha;n-1]} \frac{s}{\sqrt{n}}, \right)$$

Atenção!!!

Apresentacao derivada dos slides originais de Virgilio Almeida

Intervalos de Confiança de um-lado Exemplo

- Tempo entre quedas foi medido em dois sistemas A e B. Os valores de media e desvio padrao obtidos estao listados abaixo. O sistema A e mais suceptivel a falhas do que o sistema B

Sistema	Numero	Media	Desvio Padrao
A	972	124.10	198.20
B	153	141.47	226.11

Exemplo

- Solucao: obter IC para diferenca media usando procedimento de analise das observacoes nao pareadas

Diferenca das medias $\overline{x_A} - \overline{x_B} = 124.1 - 141.47 = -17.37$

Desvio Padrao da diferenca media

$$s = \sqrt{\left(\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}\right)} = \sqrt{\left(\frac{198.2^2}{972} + \frac{226.11^2}{153}\right)} = 19.35$$

Graus de liberdade

$$v = \frac{(s_A^2/n_A + s_B^2/n_B)^2}{\frac{1}{n_A - 1}(s_A^2/n_A)^2 + \frac{1}{n_B - 1}(s_B^2/n_B)^2} - 2 = 188.55$$

Exemplo

- Como os graus de liberdade sao mais que 30, podemos usar a normal unitaria ao inves da distribuicao t. Alem disto, como e um intervalo de um unico lado, usamos $z_{0.90} = 1.28$ para um IC de 90%:

$$(-\infty,\ -17.37 + 1.28*19.35) = (-\infty, 7.402)$$

Como o IC contem valores positivos, nao podemos dizer que A e mais suceptivel a falhas do que B.

ICs: 1 lado ou 2 lados?

- Se usarmos ICs de 2 lados, podemos dizer:
  - Tenho 90% confianca de que a media (proporcao) esta entre os dois extremos
  - Para tal, usamos $z_{0.95}$ ou $t_{0.95,df}$

- Se usarmos ICs de 1 lado, podemos dizer:
  - Tenho 90% de confianca de que a media (proporcao) e no maximo (no minimo) o extremo superior (inferior)
  - Para tal, usamos $z_{0.90}$ ou $t_{0.90,df}$

Apresentacao derivada dos slides originais de Virgilio Almeida

Tamanho das Amostras

- Amostras maiores levam a intervalos mais estreitos
  - Obtem-se menores valores de t à medida que n cresce
  - $\sqrt{n}$ nas formulas
- Coleta de amostras pode ser um processo caro!
  - Qual o mínimo que se pode querer num experimento?
- Comece com um pequeno número de medições preliminares para estimar a variância.

Apresentacao derivada dos slides originais de Virgilio Almeida

Escolha do Tamanho da Amostra

- Suponha que queremos determinar o intervalo de confiança para $\bar{x}$ com uma certa largura

$$(c_1, c_2) = ((1-e)\bar{x}, (1+e)\bar{x})$$

$$c_1 = (1-e)\bar{x} = \bar{x} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}$$

$$c_2 = (1+e)\bar{x} = \bar{x} + z_{1-\alpha/2}\frac{s}{\sqrt{n}}$$

$$n = \left(\frac{z_{1-\alpha/2}s}{e\bar{x}}\right)^2$$

Escolhendo Tamanho da Amostra

- Para obter um erro percentual ± *r* %:

$$n = \left(\frac{100zs}{r\bar{x}} \right)^2$$

- Para uma proporção $p = n_1/n$:

$$n = z^2 \frac{p(1-p)}{r^2}$$

Escolha do Tamanho da Amostra: Exemplo 1

- Cinco execuções de um *query* levaram 22.5, 19.8, 21.1, 26.7, 20.2 seconds
- Quantas execuções devem ser executadas para obter ± 5% em um IC com nível de confiança de 90%?
- $\bar{x}$ = 22.1, s = 2.8, $t_{0.95;4}$ = 2.132

$$n = \left(\frac{(100)(2.132)(2.8)}{(5)(22.1)} \right)^2 = 5.4^2 = 29.2$$

Escolha do Tamanho da Amostra: Exemplo 2

- Suponha que o tempo médio para gravar um arquivo é 7,94 seg com desvio padrão de 2,14. Aproximadamente, quantas medidas serão requeridas se nós desejamos um IC de 90% e que a média esteja dentro de um intervalo com erro de 3.5%.

Escolha do Tamanho da Amostra: Exemplo 2

- Suponha que o tempo médio para gravar um arquivo é 7,94 seg com desvio padrão de 2,14. Aproximadamente, quantas medidas serão requeridas se nós desejamos um IC de 90% e que a média esteja dentro de um intervalo com erro de 3.5%.

∀ $\alpha = 0.10$, $1-\alpha/2 = 0.95$ $e = 0.035$

$$n = \frac{(z_{1-\alpha/2}s)^2}{e\bar{x}} = \left(\frac{1.645(2.14)}{0.035(7.94)}\right)^2 = 160.46$$

$$n = 161$$

Escolha do Tamanho da Amostra: Exemplo 3

- Dois algoritmos para transmissao de pacotes foram analisados. Medicoes preliminares mostraram que o algoritmo A perde 0.5% dos pacotes e o algoritmo B perde 0.6%. Quantos pacotes precisamos observar para podermos dizer com confianca de 95% que o algoritmo A e melhor que o algoritmo B?

Escolha do Tamanho da Amostra: Exemplo 3

- Dois algoritmos para transmissao de pacotes foram analisados. Medicoes preliminares mostraram que o algoritmo A perde 0.5% dos pacotes e o algoritmo B perde 0.6%. Quantos pacotes precisamos observar para podermos dizer com confianca de 95% que o algoritmo A e melhor que o algoritmo B?

- IC de 95% para % de pacotes perdidos por A:

$$0.005 \pm 1.960\sqrt{\frac{0.005(1-0.005)}{n}}$$

- IC de 95% para % de pacotes perdidos por B:

$$0.006 \pm 1.960\sqrt{\frac{0.006(1-0.006)}{n}}$$

Apresentacao derivada dos slides originais de Virgilio Almeida

Escolha do Tamanho da Amostra: Exemplo 3

- Para podermos dizer que algoritmo A e melhor que algoritmo B, com 95% de confianca, o limite superior do intervalo de A tem que ser menor que o limite inferior do intervalo de B

$$0.005 + 1.960\sqrt{\frac{0.005(1-0.005)}{n}} < 0.006 - 1.960\sqrt{\frac{0.006(1-0.006)}{n}}$$

$$n > 84340$$

Temos que observar 85000 pacotes

Exercício de Revisao 1

- Considere que seu trabalho é comparar o desempenho de dois algoritmos (A e B) de computação gráfica, que usam métodos diferentes para geração de faces humanas realistas.

- São sistema complexos cuja execução leva tempos longos para geração das faces. O sistema A foi testado 8 vezes e o sistema B apenas 5, onde em cada experimento utilizou-se o mesmo padrão de resultado a obter.

- Os tempos de teste dos algoritmos estão na tabela a seguir. Com base nesses resultados, pede-se que se determine qual algoritmo teve melhor desempenho.

Apresentacao derivada dos slides originais de Virgilio Almeida

Exercício de Revisao 1

Experimento	Algoritmo A (seg)	Algoritmo B (seg)
1	1011	894
2	998	963
3	1113	1098
4	1008	982
5	1100	1046
6	1039	-
7	1003	-
8	1098	-

Apresentacao derivada dos slides originais de Virgilio Almeida

Exercicios de Revisao 2

Considere que num conjunto de servidores de uma máquina de busca, um servidor tem a probabilidade de falhar no período noturno igual 0.25 (i.e., a probabilidade de qualquer servidor ter falhado ao amanhecer é 25%). Para dois servidores, desenhe os gráficos da pmf e CDF da variável aleatória X, onde $P[X=x]$ = P[x servidores falharam]. Determine a média, a variância e o coeficiente de variacao de X . Assuma que as falhas são independentes e identicamente distribuídas.

Repita o processo para n servidores

Exercicios de Revisao 3

Considere um “switch” de ***N*** portas de entrada e ***N*** portas de saída (NxN). O sistema opera com o tempo dividido em intervalos (“time slots”) . Um pacote chega em qualquer “time slot” numa porta de entrada com probabilidade ***p***, independente de outros “time slots” e das outras portas de entrada. Assuma uma probabilidade de roteamento uniforme (i.e., um pacote que chegou em uma dada porta de entrada vai para qualquer porta de saída com probabilidade igual – igualmente provavel). Qual é a probabilidade de exatamente ***n*** (n < N) pacotes irem para uma porta de saída qualquer num “time slot”.

Exercicios de Revisao 4

Considere um sistema constituido de *n tarefas sequenciais.* Cada tarefa X_i (i=1..n) executa em um tempo exponencial com media de 10*i segundos.

Se o sistema termina execucao somente quando todas as tarefas terminarem, qual a media e o desvio padrao do tempo de execucao do sistema?

Se o sistema for revisado de tal maneira que as tarefas executem em paralelo, qual a probabilidade do sistema executar por mais de z minutos? Calcule a probabilidade para z=1 minuto e n = 3?

Dica:

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

Exercicios de Revisao 5

O numero de operacoes de I/O realizadas por um conjunto de programas foi medido e obteve-se: {23, 33, 14, 15, 42, 28, 33, 45, 23, 34, 39, 21, 36, 23, 34}. Responda:

a) Quais sao o 10^{th} e o 90^{th} percentis da amostra?

b) Qual o numero medio de operacoes de I/O realizadas por um programa?

c) Qual e o IC de 90% para este numero? Se voce assumir que o numero medio de operacoes de I/O realizadas pelos programas de mesma classe e igual a media da amostra, qual o maior erro que voce pode incorrer, assumindo uma confianca de 90%?

d) Qual e a porcentagem de programas que fazem no maximo 25 operacoes de I/O? Voce pode dizer, com 90% de confianca, que menos que 50% dos programas realizam no maximo 25 operacoes de I/O? E menos que 60%?

Utilize a formula de IC para proporcao dada em sala mesmo embora $np < 10$

e) Suponha que o numero de operacoes medido corresponde ao numero de I/Os realizados por segundo por cada programa. Suponha ainda que o mix de programas acima deve executar (em paralelo) em um determinado sistema. O sistema A consegue suportar, em media, no maximo 450 operacoes de I/O por segundo. O sistema A ira suportar a execucao simultanea do mix de programas acima?

Apresentacao derivada dos slides originais de Virgilio Almeida