

Utilizando Monte Carlo e Reamostragem em Estimativas

Mauricio Aguiar, TI Métricas



Agenda

- Introdução
- Um Exemplo Simples
- Outro Exemplo
- Reamostragem
- Faça Você Mesmo - Monte Carlo
- Resumo

Introdução

Estimativas

- Estimativas são projeções quantitativas de características dos projetos, tais como:
 - Tamanho do Produto
 - Esforço Requerido
 - Prazo Requerido
 - Qualidade



Introdução

Incerteza e Monte Carlo

- Há um grau de incerteza nos parâmetros de entrada de um modelo de estimativa
- Desejamos avaliar como essa incerteza pode afetar os resultados
- Isso pode ser feito através de simulação (Monte Carlo Simulation)

Introdução

Entradas de um Modelo de Estimativa

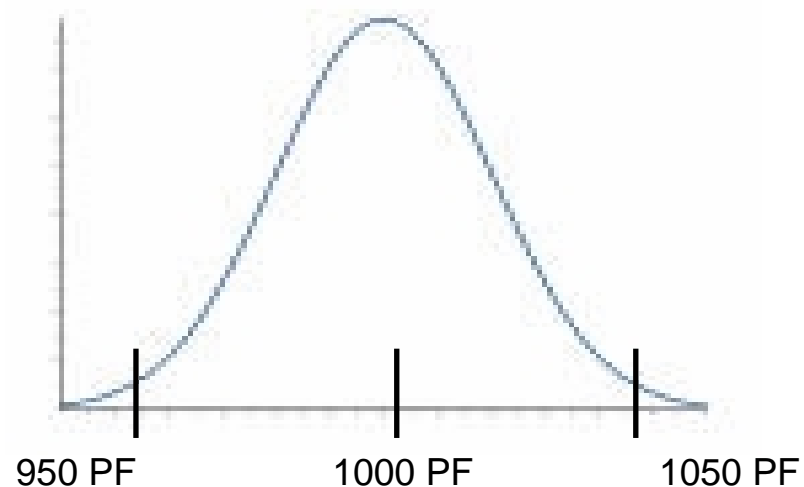
- Tamanho (Pontos de Função, etc.)
- Características do Produto e do Projeto
- Esforço Estimado por Atividade
- etc.

Introdução

Modelando a Incerteza

- Permitir que as entradas variem segundo distribuições estatísticas definidas

Ex.: Tamanho



Distribuição Normal

Um Exemplo Simples

O Problema

- Executar construção e teste unitário para 5 módulos (classes, funções, subrotinas...)

Módulo	Mínimo (d)	Esperado (d)	Máximo (d)	Programador	Qualidade da Estimativa
A	2	4	10	Arnaldo	Baixa
B	4	6	10	Ronaldo	Baixa
C	8	12	16	André	Média
D	3	5	6	Paulo	Alta
E	2	4	6	Nilton	Média
Totais	19	31	48		

- Assuma que o trabalho será feito sequencialmente

Um Exemplo Simples

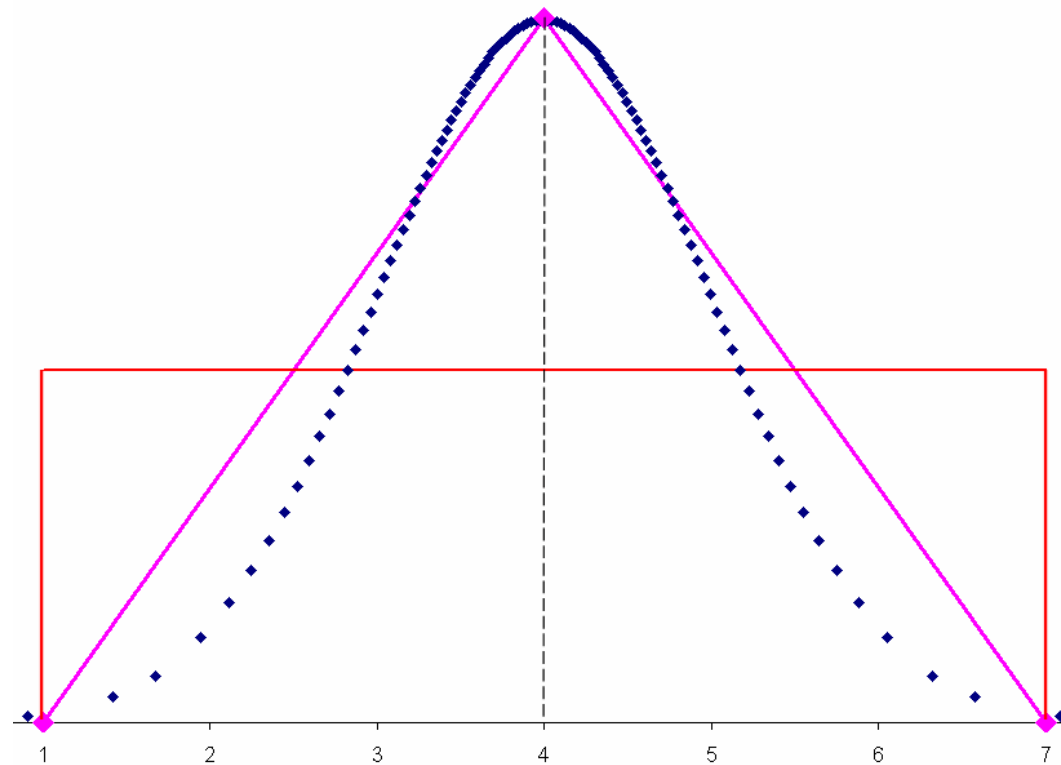
Como Fazer Melhor

- Com 100% de probabilidade de acerto, o prazo seria 48 dias
- Um prazo menor com 90% de probabilidade de acerto seria suficiente
- Qual seria esse prazo?

Um Exemplo Simples

Modelando a Incerteza

Distribuições: Normal, Triangular e Uniforme



Um Exemplo Simples

Monte Carlo

- Simular 10000 vezes a execução da construção e teste unitário dos 5 módulos
 - Variar os prazos individuais conforme as respectivas distribuições
 - Avaliar a variação do prazo total

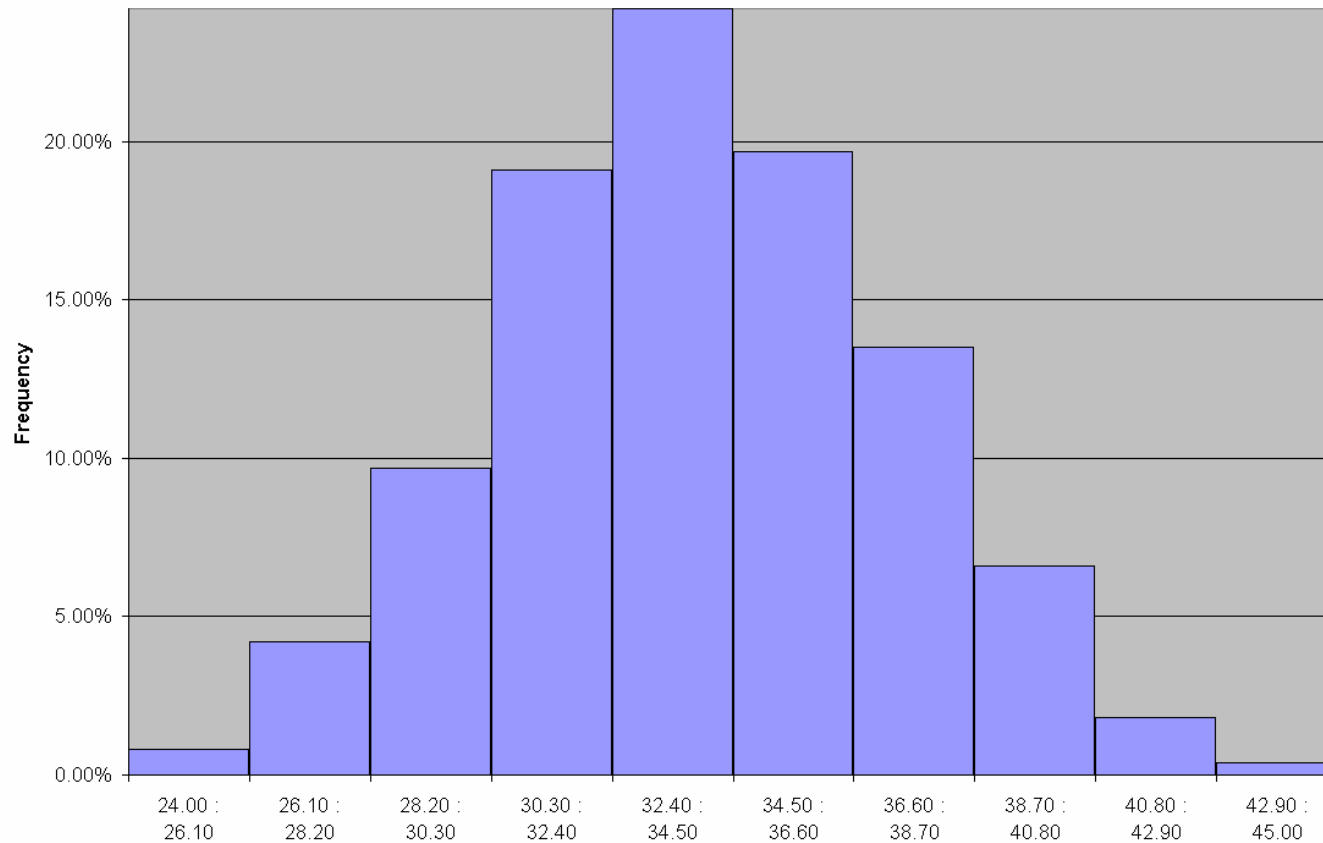
Um Exemplo Simples

Monte Carlo

Módulo	Mínimo (d)	Esperado (d)	Máximo (d)	Simulado	Distribuição
A	2	4	10	3.31064072	Uniforme
B	4	6	10	6.95388124	Uniforme
C	8	12	16	11.1681224	Triangular
D	3	5	6	5.81785033	Normal
E	2	4	6	3.68017662	Triangular
Totais	19	31	48	30.9306713	

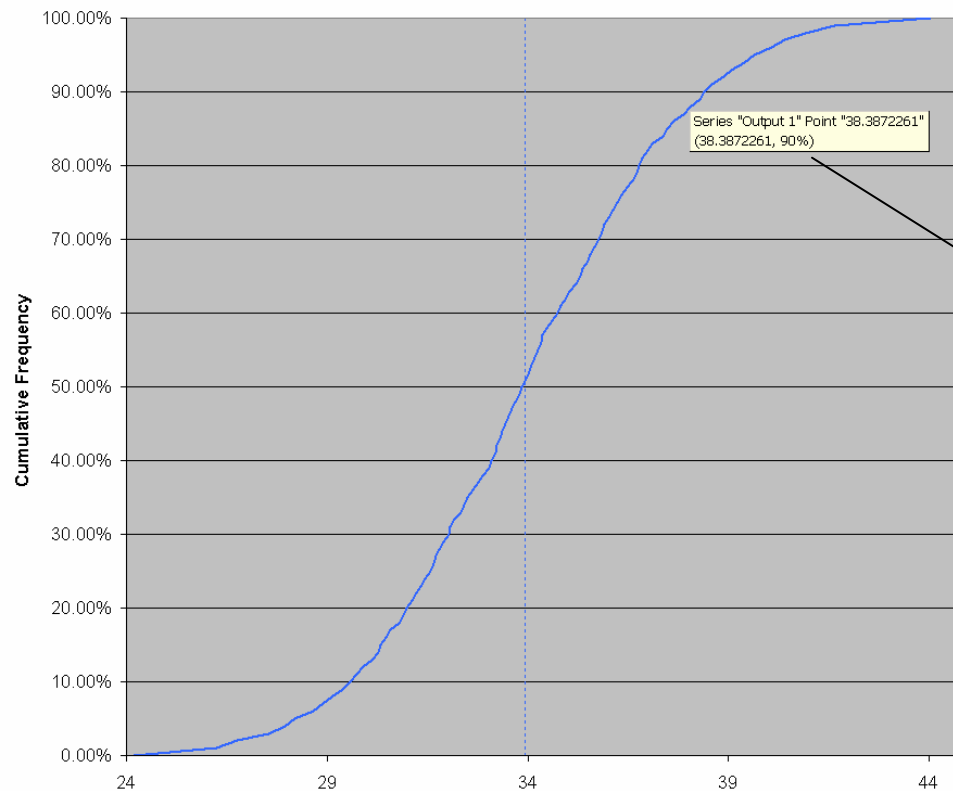
Um Exemplo Simples

Histograma do Prazo Simulado



Um Exemplo Simples

Freqüência Acumulada do Prazo Simulado



38,4 – 90%

Prazo menor ou igual a 38 dias com 90% de probabilidade de acerto

Um Exemplo Simples

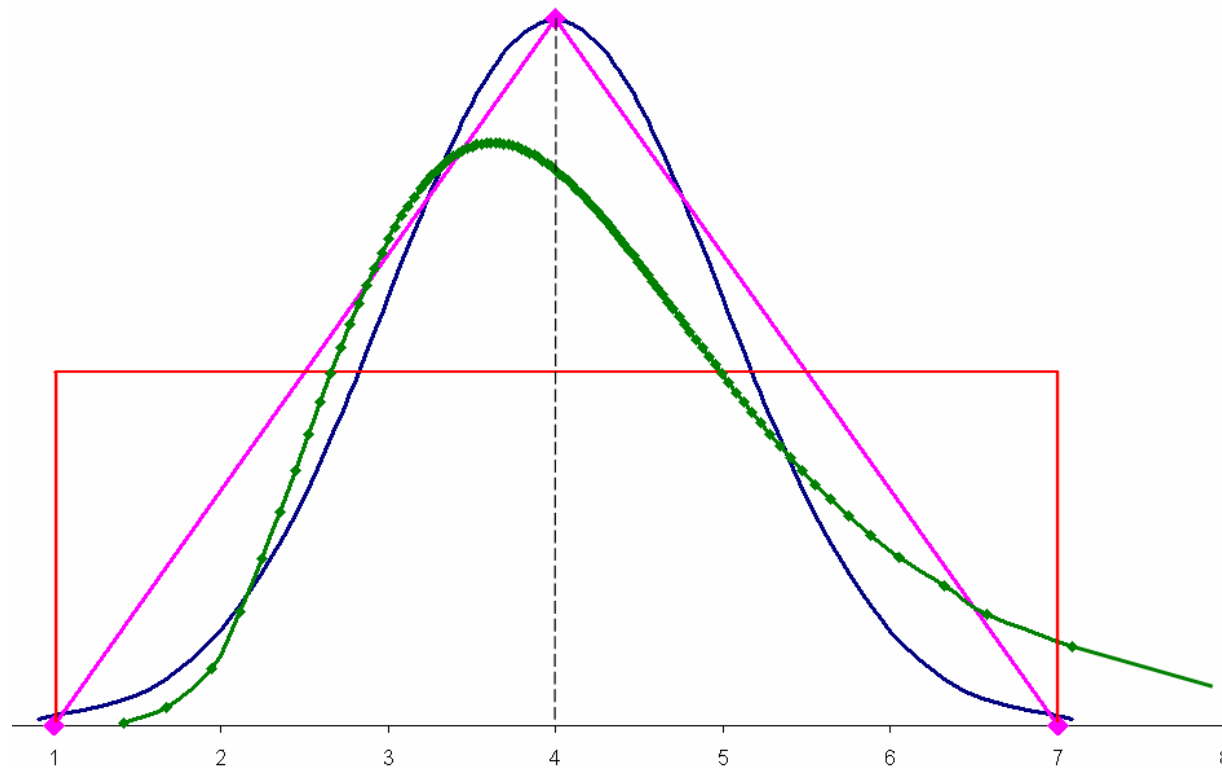
Questionamento

- As distribuições utilizadas correspondem à realidade?
- A probabilidade de terminar antes é a mesma de terminar depois?

Um Exemplo Simples

Questionamento

Distribuições: Normal, Triangular, Uniforme e Lognormal



Um Exemplo Simples

Questionamento

- Simular 10000 vezes a execução da construção e teste unitário dos 5 módulos, utilizando a distribuição lognormal somente no caso anteriormente “normal”

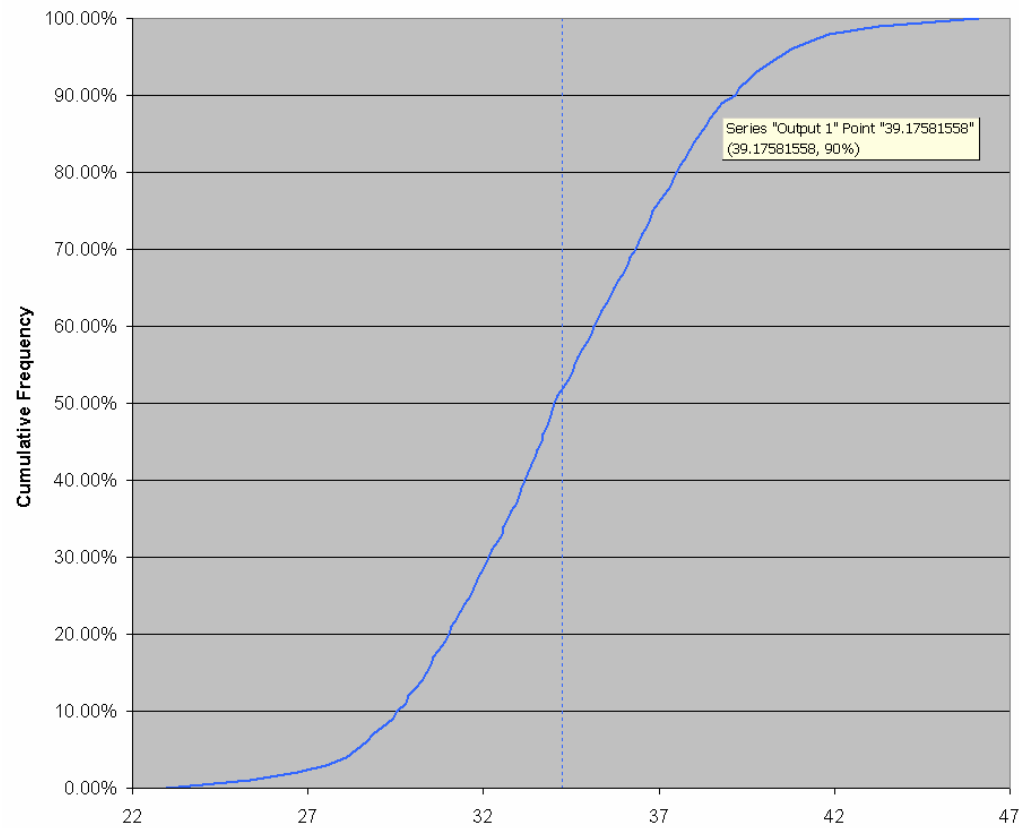
Um Exemplo Simples

Questionamento

Módulo	Mínimo (d)	Esperado (d)	Máximo (d)	Simulado	Distribuição
A	2	4	10	9.21425821	Uniforme
B	4	6	10	5.52832972	Uniforme
C	8	12	16	9.84395835	Triangular
D	3	5	6	3.47258674	Lognormal
E	2	4	6	4.86700713	Triangular
Totais	19	31	48	32.9261401	

Um Exemplo Simples

Questionamento



**O prazo agora
é 39 dias**

Outro Exemplo

Produtividade de 9 Projetos

Projeto	Tamanho (PF)	Esforço	Produtividade
P1	98	1910	19,5
P2	184	2760	15,0
P3	212	2010	9,5
P4	196	1620	8,3
P5	261	1855	7,1
P6	257	1980	7,7
P7	430	7830	18,2
P8	190	1740	9,2
P9	310	4890	15,8
Média			12,2

Produtividade em Horas/Pontos de Função

Nota: Dados fictícios

Outro Exemplo

Estimando o Erro

$$((\text{Estimado} - \text{Real})/\text{Real}) * 100$$

Projeto	Tamanho (PF)	Esforço	Produtividade	Estimativa I	Erro % ABS
P1	98	1910	19,5	1200	37,2%
P2	184	2760	15,0	2253	18,4%
P3	212	2010	9,5	2596	29,1%
P4	196	1620	8,3	2400	48,1%
P5	261	1855	7,1	3195	72,3%
P6	257	1980	7,7	3147	58,9%
P7	430	7830	18,2	5265	32,8%
P8	190	1740	9,2	2326	33,7%
P9	310	4890	15,8	3795	22,4%
		Média	12,2	MRE	39,2%

Mean Relative Error

Outro Exemplo

Dados de Esforço

- Muitas vezes a qualidade dos dados de esforço é questionável
- Como o erro nos dados de esforço afetaria o erro de estimativa?

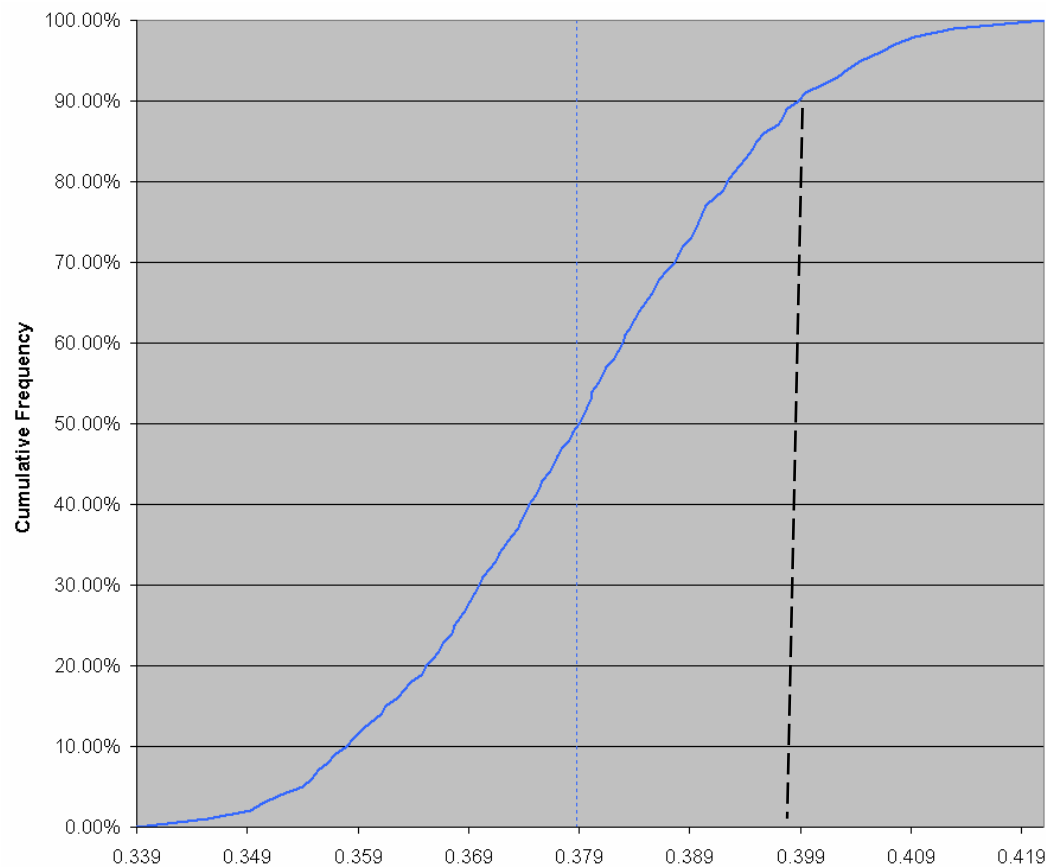
Outro Exemplo

Dados de Esforço

Proj	PF	Qual.	Mín.	Esp.	Máx.	Distribuição	Esforço	Prod.	Est.	Erro % ABS
P1	98	Baixa	1600	1910	2100	Uniforme	2094	21.4	1238	35.2%
P2	184	Baixa	2500	2760	3000	Uniforme	2621	14.2	2325	15.8%
P3	212	Alta	1950	2010	2100	Normal	2013	9.5	2679	33.3%
P4	196	Média	1500	1620	1800	Triangular	1671	8.5	2477	52.9%
P5	261	Baixa	1500	1855	2200	Uniforme	2122	8.1	3298	77.8%
P6	257	Baixa	1500	1980	2400	Uniforme	1681	6.5	3248	64.0%
P7	430	Baixa	5000	7830	9000	Uniforme	8872	20.6	5434	30.6%
P8	190	Alta	1700	1740	1800	Normal	1707	9.0	2401	38.0%
P9	310	Média	4600	4890	4900	Triangular	4899	15.8	3917	19.9%
							Média	12.6	MRE	40.8%

Outro Exemplo

Dados de Esforço



**O erro
é inferior
a 40% com
90% de
probabilidade.**

**O máximo é
42%.**

Reamostragem

A Idéia

- Monte Carlo exige que façamos suposições sobre as distribuições
- A Reamostragem basea-se na replicação de uma amostra (podendo haver repetições)
- As estatísticas baseadas em reamostragem aproximam-se dos valores reais, conforme cresce o número de amostras
- A Reamostragem independe de suposições sobre as distribuições

Reamostragem

Um Exemplo

- Construir um intervalo de confiança a 90% para a produtividade do COBOL, com base no banco de dados ISBSG V10
- Os dados: 615 projetos, produtividade média **20.5 H/PF**, produtividade mediana **11.9 H/PF**

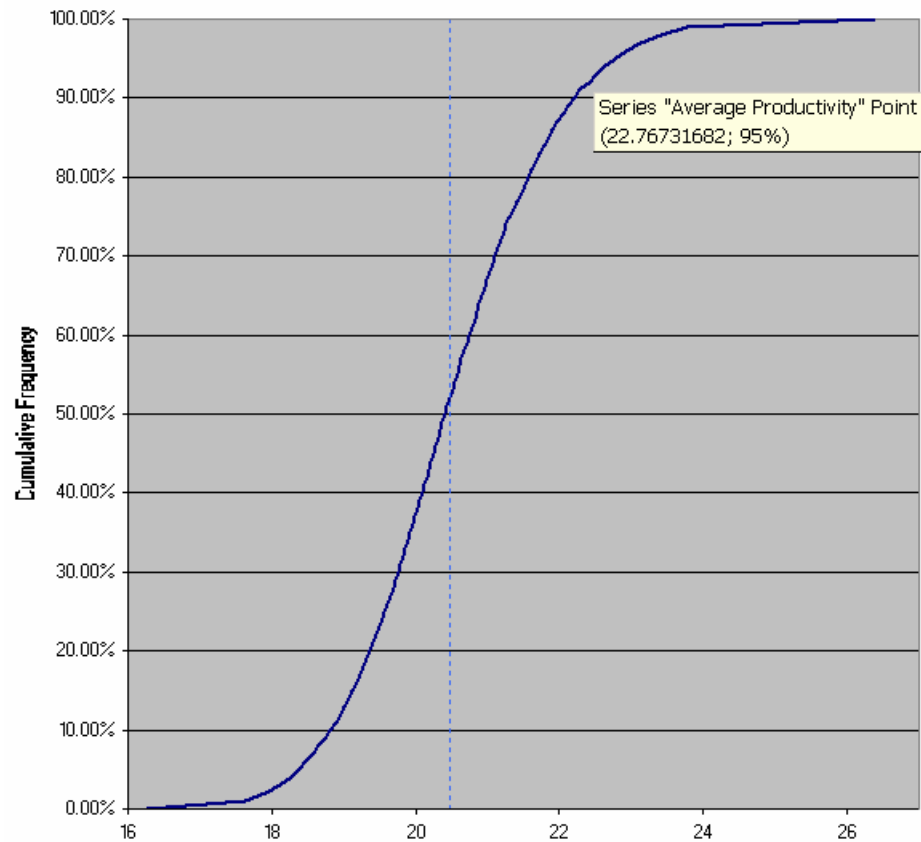
Reamostragem

Distribuição Amostral: Média, Mediana

	Average Productivity	Median Productivity
Average	20.47295952	11.70256901
Std Dev	1.32359835	0.76645153
Std Err	0.013235984	0.007664515
Max	26.36926842	14.30000019
Min	16.26861763	9.399999619

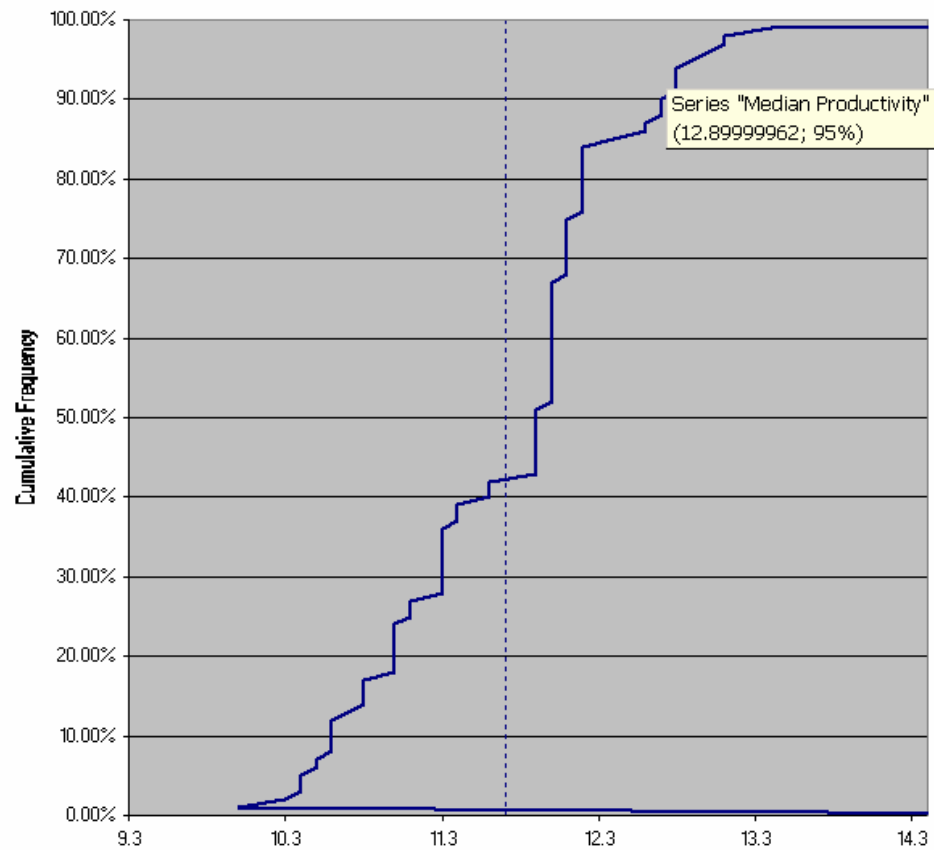
Reamostragem

Distribuição Amostral: Média



Reamostragem

Distribuição Amostral: Mediana



Faça Você Mesmo

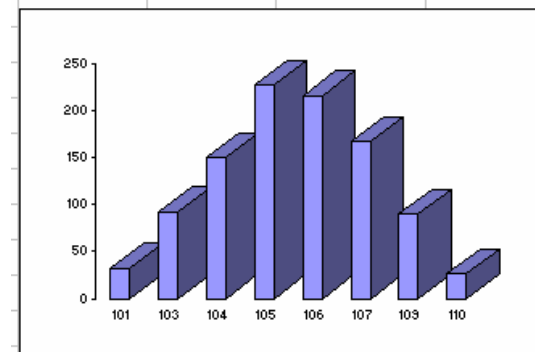
Simulando a Distribuição Triangular

	A	B	C	D	
1	Simulação da Distribuição Triangular				
2					
3	Iterações:	1000	<input type="button" value="Simular"/>		
4					
5	Mínimo	100			
6	Esperado	105			
7	Máximo	110			
8					
9	Val Aleat.	0.119205			
10	Triangular	102.4414			
11					

Faça Você Mesmo

Simulando a Distribuição Triangular

Mínimo	100.078	Média	104.968
Máximo	109.9004	Mediana	104.987
Amplitude	9.8224		
Freqs	Faixas	%	
32	101.3058	3.20%	3.20%
92	102.5336	9.20%	12.40%
150	103.7614	15.00%	27.40%
227	104.9892	22.70%	50.10%
216	106.217	21.60%	71.70%
167	107.4448	16.70%	88.40%
90	108.6726	9.00%	97.40%
26	109.9004	2.60%	100.00%
1000		100.00%	



Faça Você Mesmo

Simulando a Distribuição Triangular

Trabalhando com Excel VBA

```
Function TriDist(ByVal prob As Single, ByVal opt As Single, ByVal expect As Single) As Single
    'Esta função retorna um valor segundo a distribuição triangular com
    'opt = valor mínimo
    'expect = valor esperado
    'pess = valor máximo

    Dim x, d As Single
    d = pess - opt
    x = (expect - opt) / d
    If prob <= x Then TriDist = opt + (((prob * x) ^ 0.5) * d)
    If prob > x Then TriDist = pess - (((1 - prob) * (1 - x)) ^ 0.5) * d

End Function
```

Faça Você Mesmo

Simulando a Distribuição Triangular

```
Sub Simular()  
  
    Dim intI, intNumIteracoes As Integer  
  
    intNumIteracoes = Sheets("Main").Range("B3").Value  
    If intNumIteracoes <= 0 Then  
        MsgBox ("Número de iterações deve ser > 0")  
        Exit Sub  
    End If  
    If intNumIteracoes > 10000 Then  
        MsgBox ("Número máximo de iterações = 10000")  
        Exit Sub  
    End If  
  
    'Resultados: Usar Colunas A, B, C, D, E  
    Sheets("Dados").Select  
    Sheets("Dados").Range("A27").Select  
    ActiveWindow.FreezePanes = True  
    Sheets("Dados").Range("A1").Select  
    Sheets("Dados").Range("A1:E10000").Clear  
  
    For intI = 1 To intNumIteracoes  
        Application.StatusBar = "Processando Iteração... " & intI  
        'Distr Triangular  
        Sheets("Main").Range("B10").Copy  
        Sheets("Dados").Cells(intI, 1).PasteSpecial Paste:=xlValues, _  
            Operation:=xlNone, _  
            SkipBlanks:=False, _  
            Transpose:=False  
  
    Next intI  
    ActiveWindow.FreezePanes = False  
    Application.StatusBar = False  
  
End Sub
```

Resumo

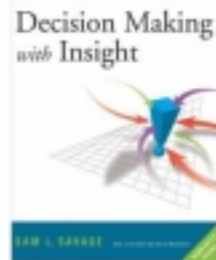
O Que Vimos

- Monte Carlo ajuda a identificar a variação nos resultados em função da incerteza nas entradas
- A escolha das distribuições estatísticas é muito importante em Monte Carlo
- A Reamostragem funciona mesmo quando a distribuição é desconhecida
- É possível fazer muita coisa com Excel e VBA
- Há ferramentas profissionais para Monte Carlo (Crystal Ball, @Risk, XLSim, etc.)

Resumo

Referências

- Savage, Sam L., Decision Making with Insight
 - XLSim 2.0



Decision Making with Insight (with Insight.xla 2.0 and CD-ROM) (Paperback)

by [Sam L. Savage](#) (Author)

[Reviews](#) (2 customer reviews)

List Price: \$73.96

Price: **\$60.42** & this item ships for **FREE** with **Super Saver Shipping**. [Details](#)

You Save: **\$13.53 (18%)**

- Mooney, C.Z. and Duval, R.D, Bootstrapping: A Nonparametric Approach to Statistical Inference, SAGE, 1993
- Davidson, A.C. and Hinkley, D.V., Bootstrap Methods and Their Application, Cambridge University Press, 1997



Agradecemos a sua Participação

Mauricio Aguiar

ti MÉTRICAS

mauricio@metricas.com.br

www.metricas.com.br