

# *Monte Carlo em Estimativas de Software*

*Mauricio Aguiar*  
*ti MÉTRICAS Ltda*

# Agenda

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

# *Introdução*

# 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

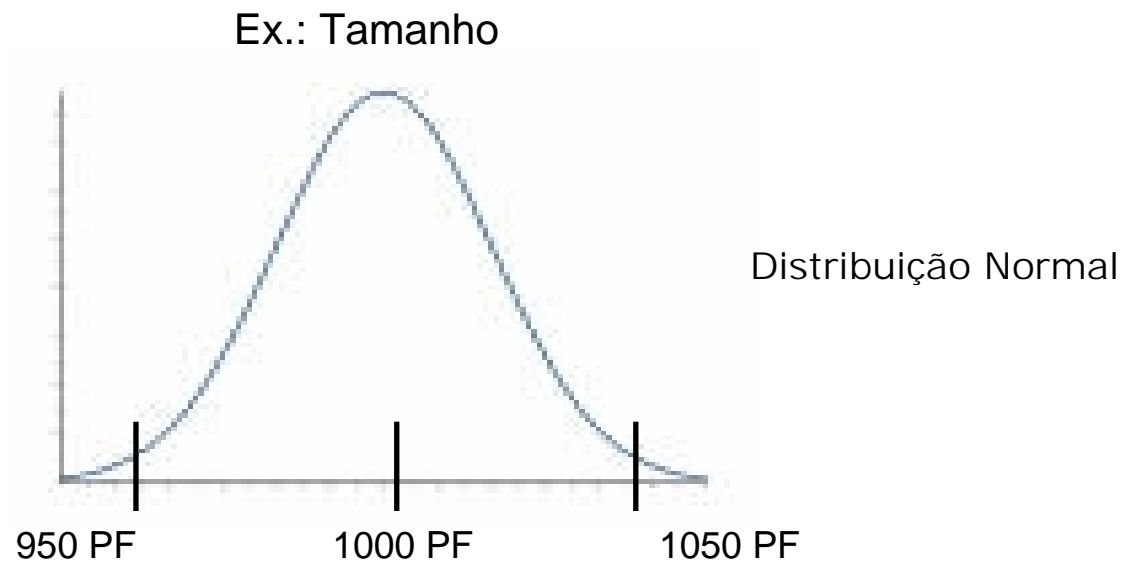
## Entradas de um Modelo de Estimativa

- 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



# *Um Exemplo Simples*



# 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
<b>Totais</b>	<b>19</b>	<b>31</b>	<b>48</b>		

- Assuma que o trabalho será feito sequencialmente

# Um Exemplo Simples

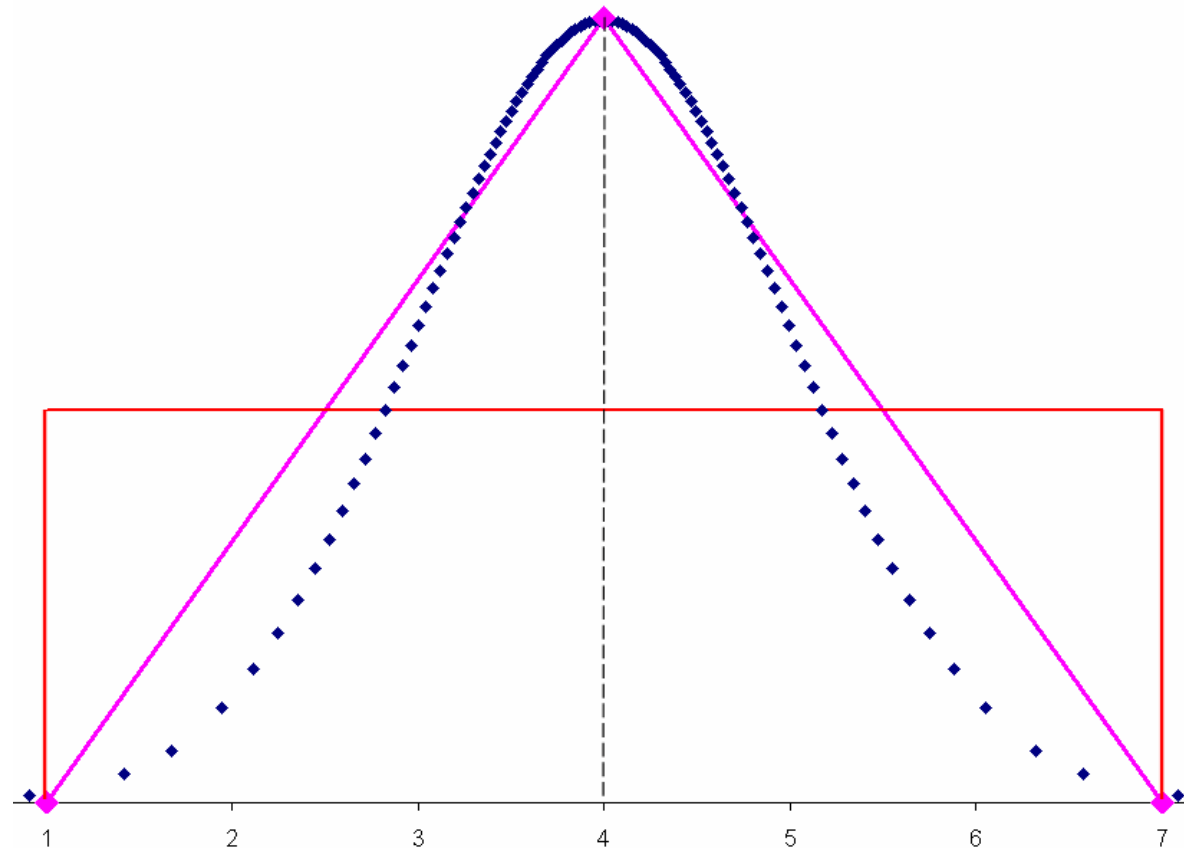
## Como Fazer Melhor

- 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

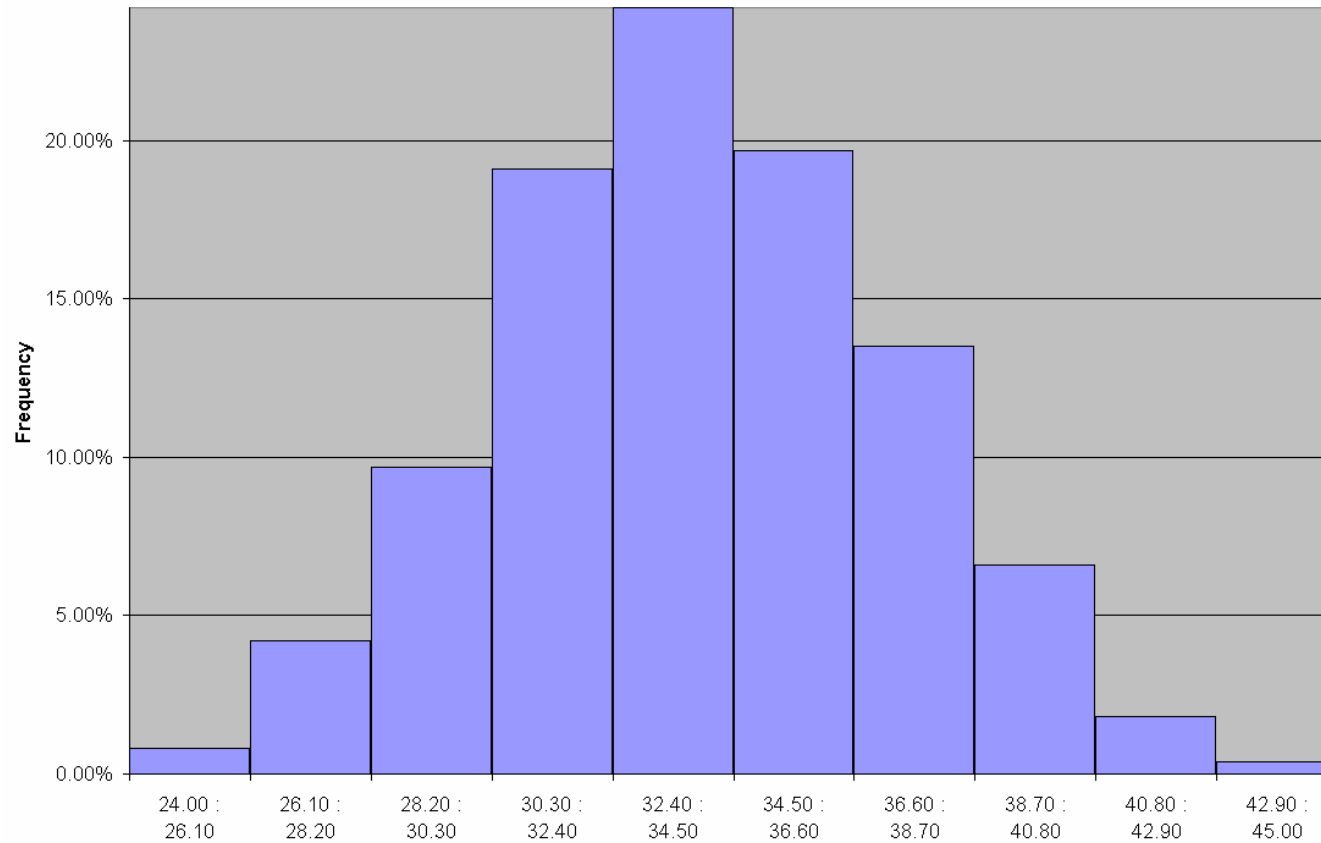
# Um Exemplo Simples

## 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
<b>Totais</b>	<b>19</b>	<b>31</b>	<b>48</b>	30.9306713	

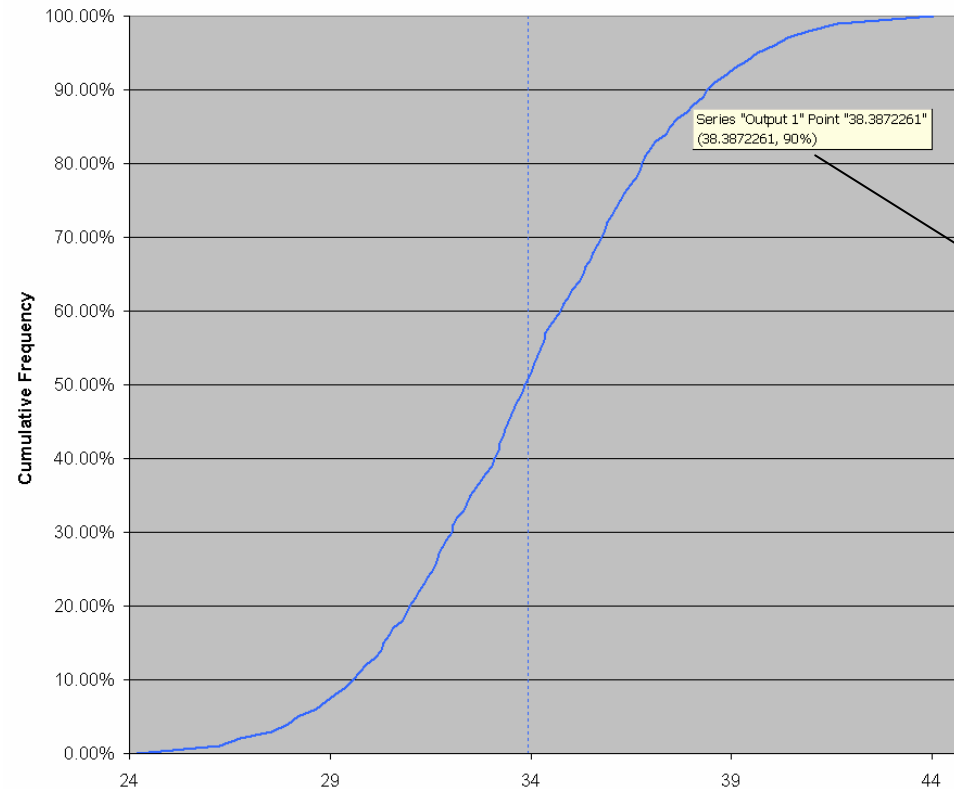
# Um Exemplo Simples

## Histograma do Prazo Simulado



# Um Exemplo Simples

## Frequencia Acumulada do Prazo Simulado



**38,4 – 90%**

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

# Um Exemplo Simples

## Questionamento

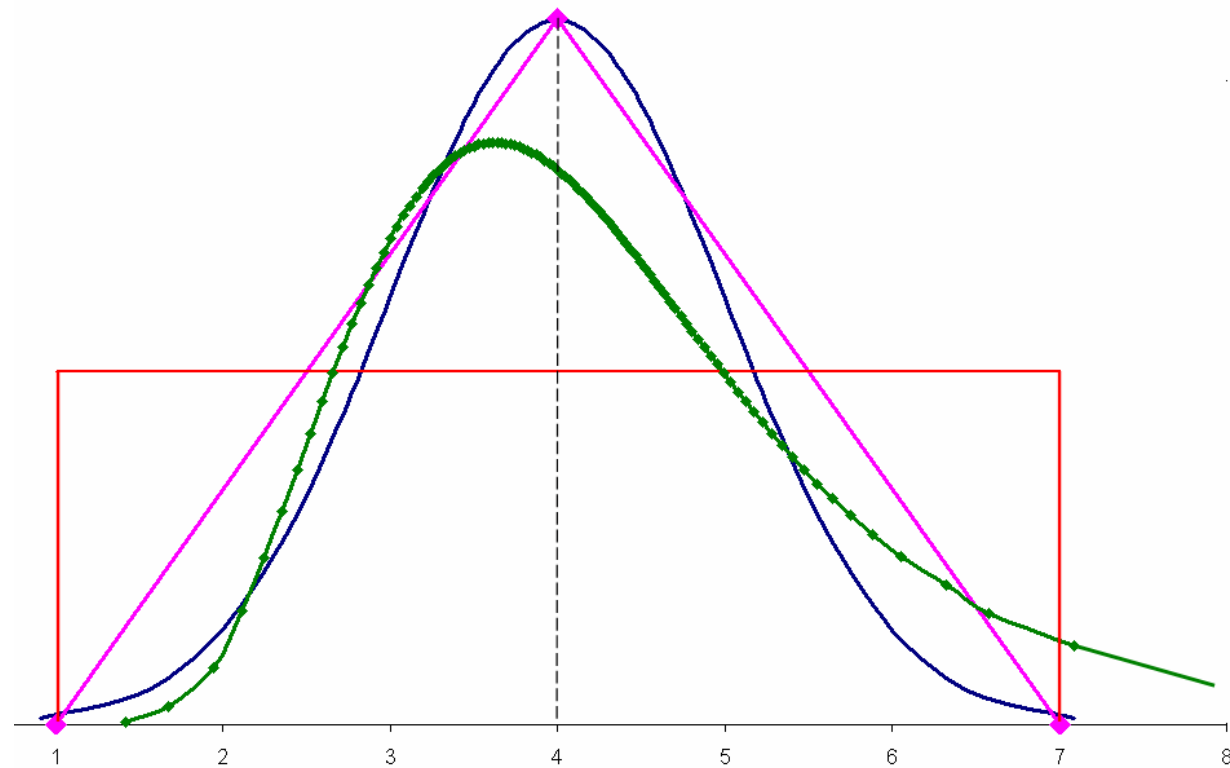
- As distribuições utilizadas correspondem à realidade?
- A probabilidade de terminar antes é a mesma de terminar depois?



# 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”

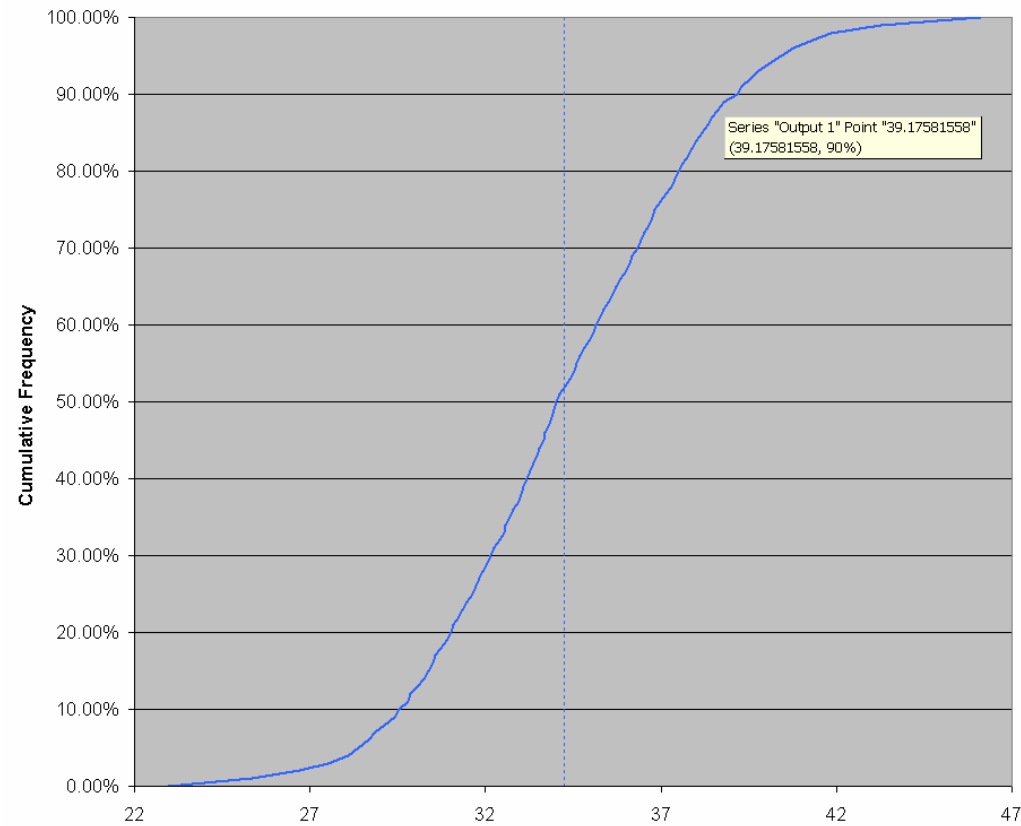
# Um Exemplo Simples

## 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	<b>Lognormal</b>
E	2	4	6	4.86700713	Triangular
<b>Totais</b>	<b>19</b>	<b>31</b>	<b>48</b>	32.9261401	

# Um Exemplo Simples

## Questionamento



**O prazo agora  
é 39 dias**

## *Outro Exemplo*

# 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
<b>Média</b>			<b>12,2</b>

**Produtividade em Horas/Pontos de Função**

**Nota: Dados fictícios**

# Outro Exemplo

## Estimando o Erro

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

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%
<b>Média</b>			<b>12,2</b>	<b>:</b>	<b>39,2%</b>

$$\text{MMRE} = \text{Mean MRE}$$

# 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

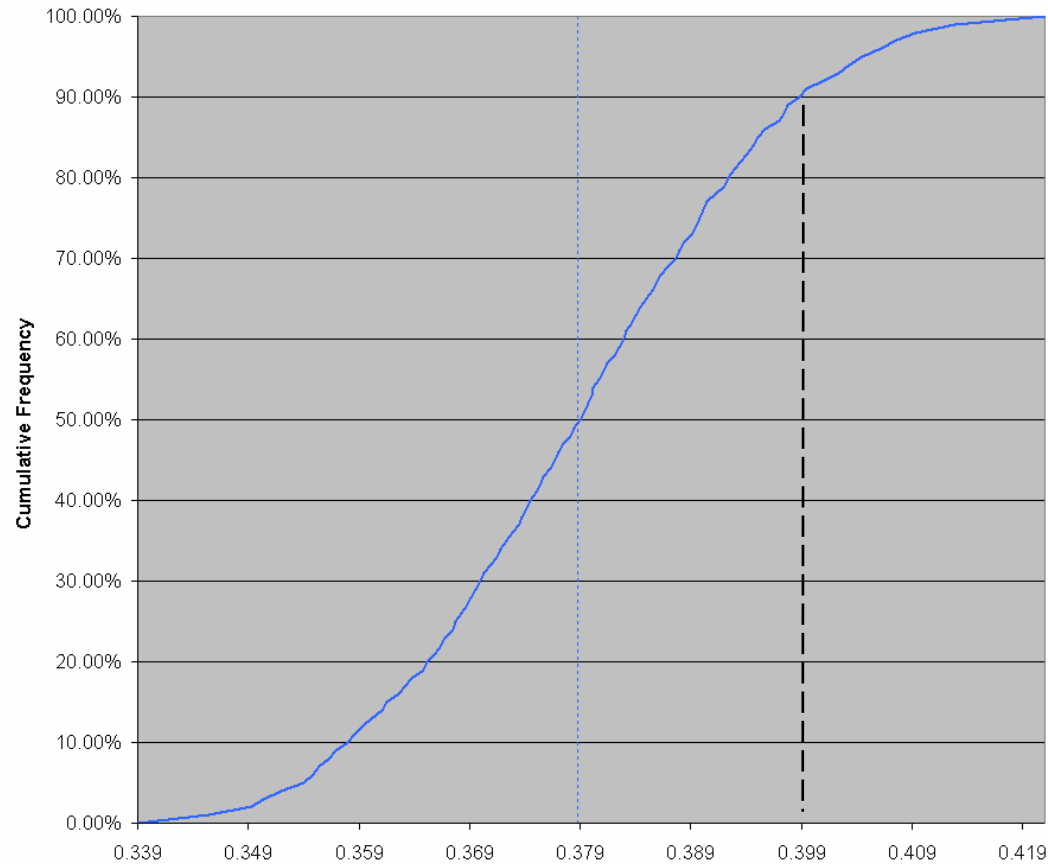
$$\text{MRE} = (\text{ABS}(\text{Estimado} - \text{Real})/\text{Real}) * 100$$

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%
							<b>Média</b>	<b>12.6</b>		<b>40.8%</b>

MMRE = Mean MRE

# Outro Exemplo

## Dados de Esforço



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

**O máximo é  
42%.**

# *Reamostragem*

# 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

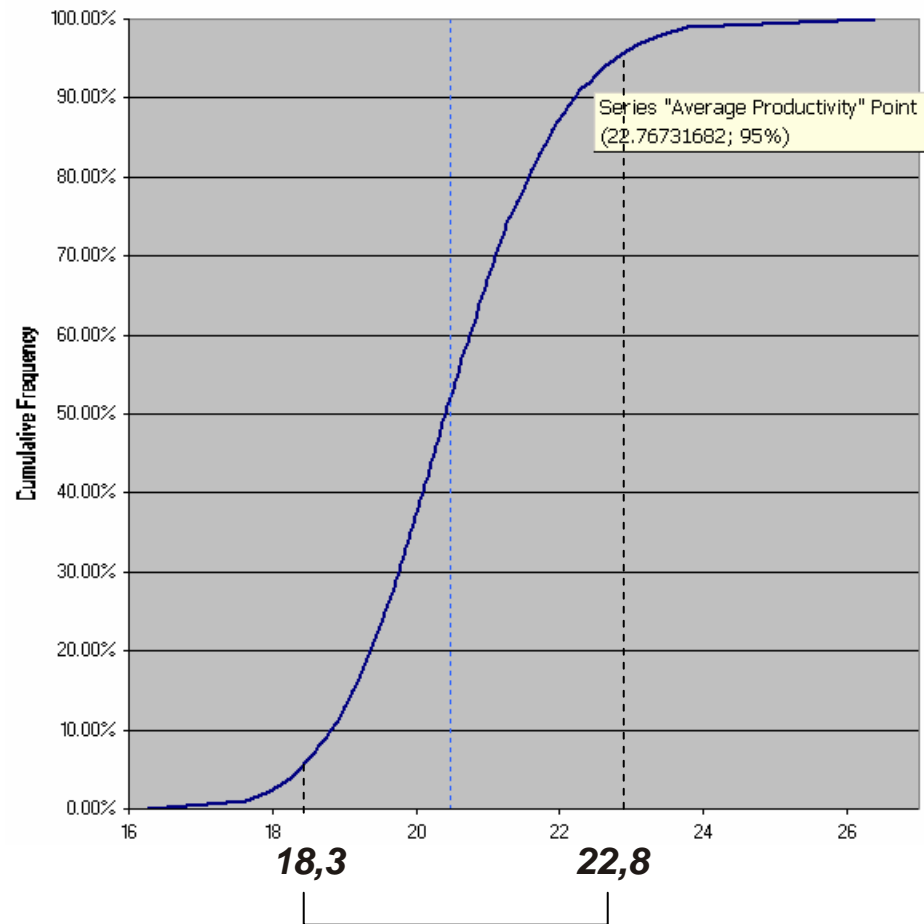
# Reamostragem

## 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**
- Foram extraídas 10000 amostras dos 615 projetos (cada amostra com 615 projetos, podendo haver repetições)

# Reamostragem

## Intervalo de Confiança - Percentil



*Faça Você Mesmo*

# Faça Você Mesmo

Simulando a Distribuição Triangular com o Excel

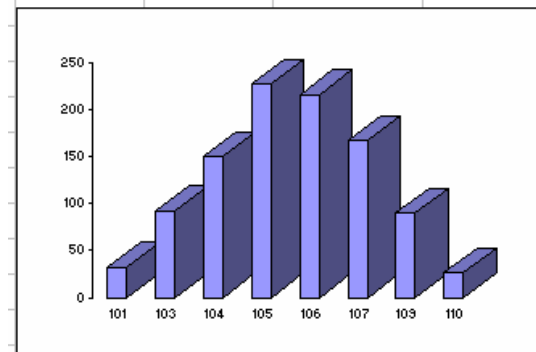
	A	B	C	D
1	<b>Simulação da Distribuição Triangular</b>			
2				
3	<b>Iterações:</b>	1000	Simular	
4				
5	<b>Mínimo</b>	100		
6	<b>Esperado</b>	105		
7	<b>Máximo</b>	110		
8				
9	<b>Val Aleat.</b>	0.119205		
10	<b>Triangular</b>	102.4414		
11				



# Faça Você Mesmo

Simulando a Distribuição Triangular com o Excel

<b>Mínimo</b>	100.078	<b>Média</b>	<b>104.968</b>
<b>Máximo</b>	109.9004	<b>Mediana</b>	<b>104.987</b>
<b>Amplitude</b>	9.8224		
<b>Freqs</b>	<b>Faixas</b>	<b>%</b>	
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%
<b>1000</b>		<b>100.00%</b>	



# Faça Você Mesmo

## Simulando a Distribuição Triangular com o Excel

```
Function TriDist(ByVal prob As Single, ByVal opt As Single, ByVal expect As  
'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 com o Excel

```
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.FreezPanes = 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.FreezPanes = False  
    Application.StatusBar = False  
  
End Sub
```

# Faça Você Mesmo

## A Linguagem/Ambiente "R"



- "R" é uma linguagem e um ambiente, específicos para cálculos e gráficos estatísticos
- "R" é software gratuito ("GNU General Public License")
- "R" é uma implementação "free" da linguagem "S" desenvolvida nos Laboratórios Bell (produto comercial S-Plus)
- Existem vários livros, artigos e documentos descrevendo a linguagem "S" e o "R"
- Conheça e baixe o "R" em [www.r-project.org](http://www.r-project.org)

# Faça Você Mesmo

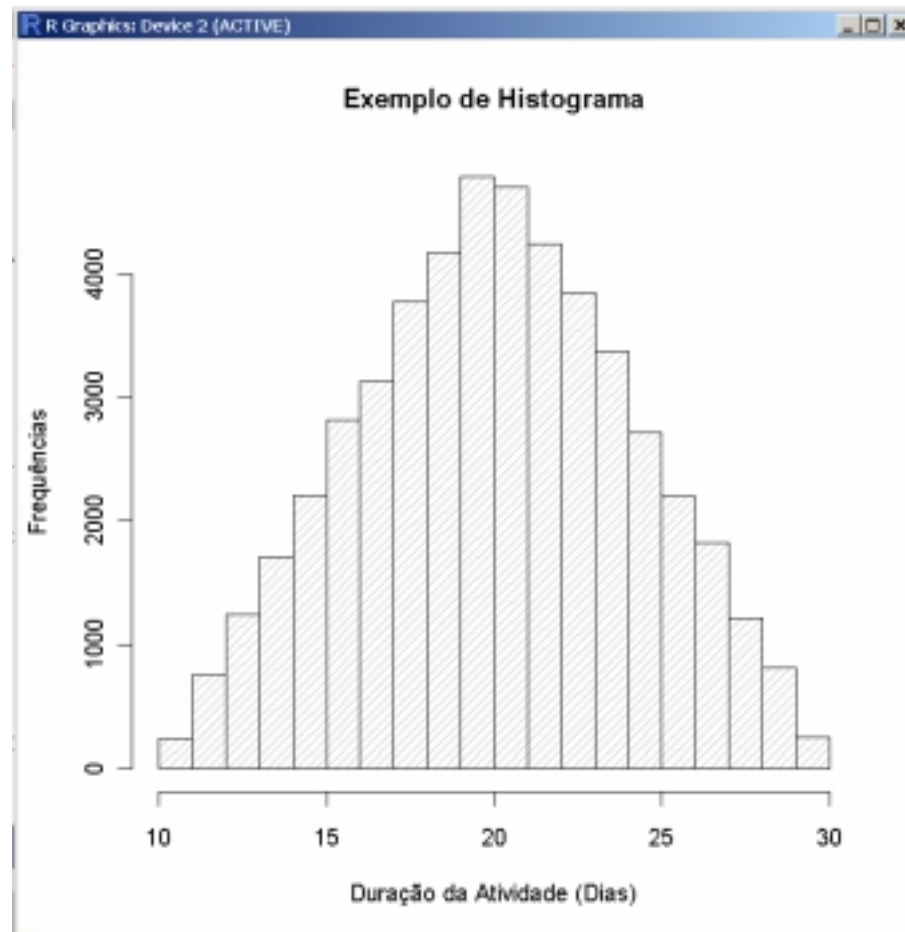
## Simulando a Distribuição Triangular com o R

```
Untitled - R Editor
###
### Distribuição Triangular
###
rm(list=ls())
tridist <- function(aleat, minimo, esperado, maximo) {
# Esta função retorna um valor segundo a distribuição triangular com
# aleat = número aleatório entre 0 e 1
# minimo = valor mínimo
# esperado = valor esperado
# maximo = valor máximo

  d <- maximo - minimo
  x <- (esperado - minimo) / d
  if (aleat <= x) minimo + (((aleat * x) ^ 0.5) * d)
  else if (aleat > x) maximo - (((1 - aleat) * (1 - x)) ^ 0.5) * d
}
###
### Gerar as observações (da distr triangular)
###
nobs <- 50000
mybreaks <- seq(10,30,1)
vtriang <- numeric(nobs)
for (i in 1:nobs) {
  vtriang[i] <- tridist(runif(1, min=0, max=1), 10, 20, 30)
}
###
hist(vtriang, freq = 1, breaks = mybreaks, density = 20, angle = 45,
     col = "grey", border = "black", main = "Exemplo de Histograma",
     xlab = "Duração da Atividade (Dias)", ylab = "Frequências")
```

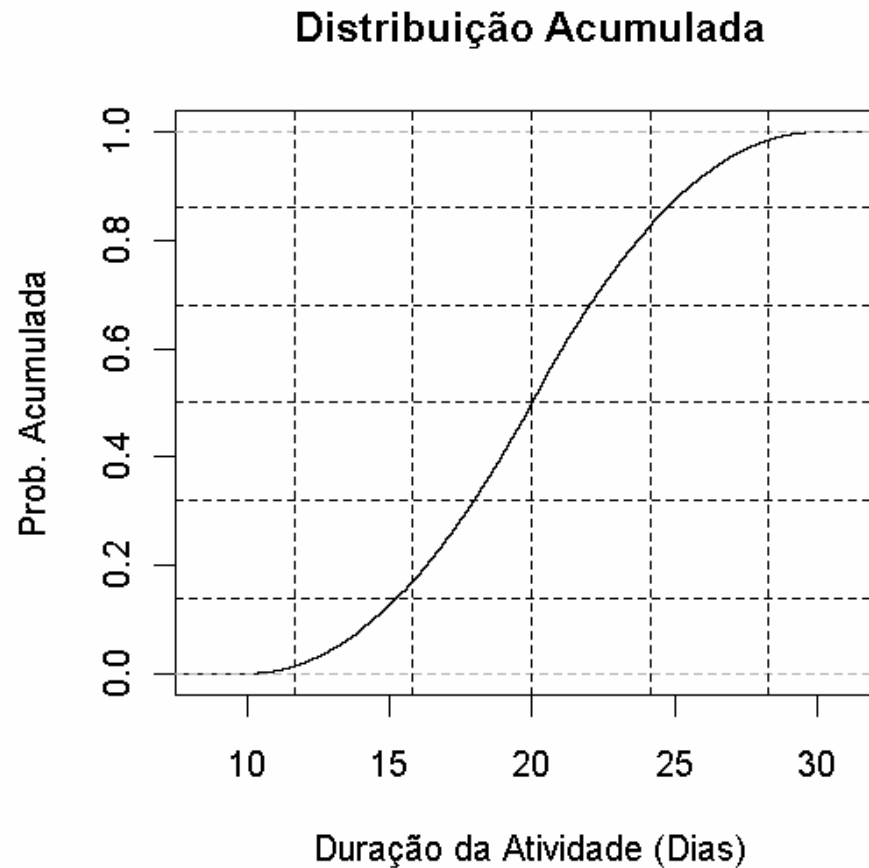
# Faça Você Mesmo

## Simulando a Distribuição Triangular com o R



# Faça Você Mesmo

Simulando a Distribuição Triangular com o R



# *Resumo*



# 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/VBA e mais ainda com o R
- Existem ferramentas profissionais para Monte Carlo (Crystal Ball, @Risk, XLSim, etc.)

# Resumo

## Referências



**Decision Making with Insight (with Insight.xla 2.0 and CD-ROM)** by Sam L. Savage (**Paperback** - Jan 14, 2003)

Buy new: ~~\$85.95~~ **\$77.35** [31 Used & new](#) from **\$22.19**

In Stock

Eligible for **FREE** Super Saver Shipping.

★★★★★ (5)



**An Introduction to the Bootstrap (Monographs on Statistics and Applied Probability)** by Bradley Efron and R.J. Tibshirani (**Hardcover** - May 15, 1994)

Buy new: ~~\$104.95~~ **\$83.96** [25 Used & new](#) from **\$70.84**

Get it by **Tuesday, Feb 17** if you order in the next **12 hours** and choose one-day shipping.

Eligible for **FREE** Super Saver Shipping.

★★★★★ (9)



**Bootstrap Methods and Their Application (Cambridge Series in Statistical and Probabilistic Mathematics, No 1)** by A. C. Davison and D. V. Hinkley (**Paperback** - Oct 28, 1997)

Buy new: ~~\$70.00~~ **\$63.00** [13 Used & new](#) from **\$47.25**

In Stock

Eligible for **FREE** Super Saver Shipping.

★★★★★ (7)



**Statistics: An Introduction using R** by Michael J. Crawley (**Paperback** - May 6, 2005)

Buy new: ~~\$50.00~~ **\$43.80** [54 Used & new](#) from **\$30.00**

Get it by **Tuesday, Feb 17** if you order in the next **11 hours** and choose one-day shipping.

Eligible for **FREE** Super Saver Shipping.

★★★★☆ (7)

# Agradecemos a sua participação



[info@metricas.com.br](mailto:info@metricas.com.br)

