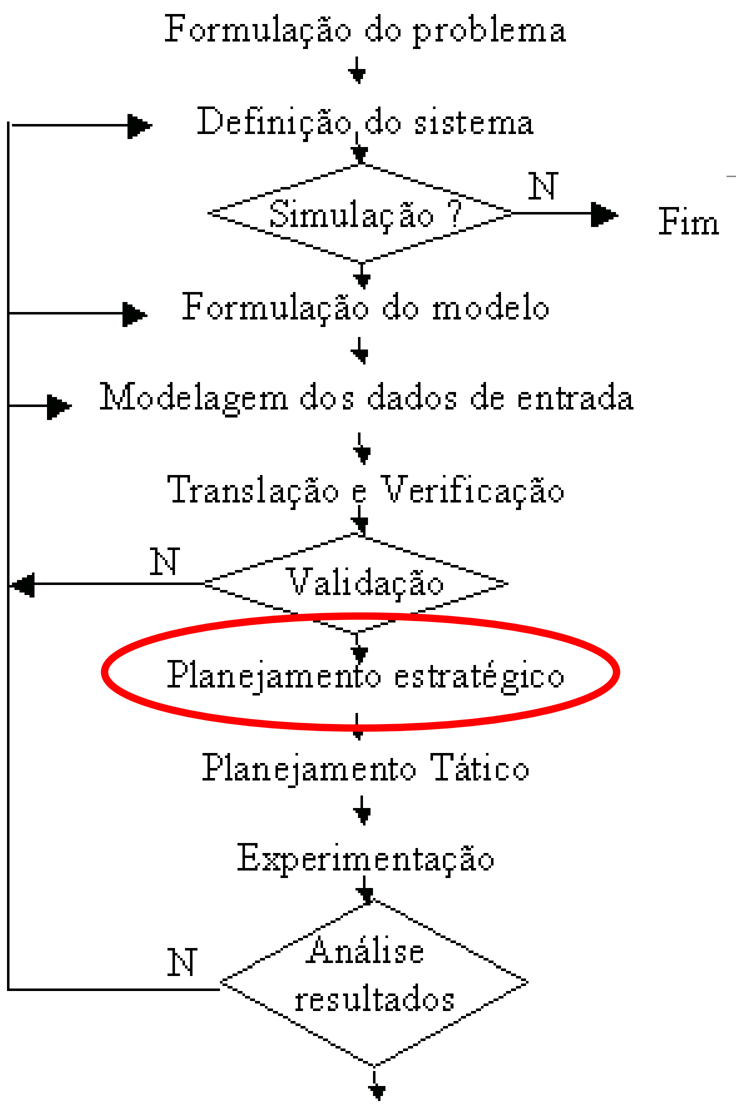
Análise de Resultados

AULA 09

Introdução



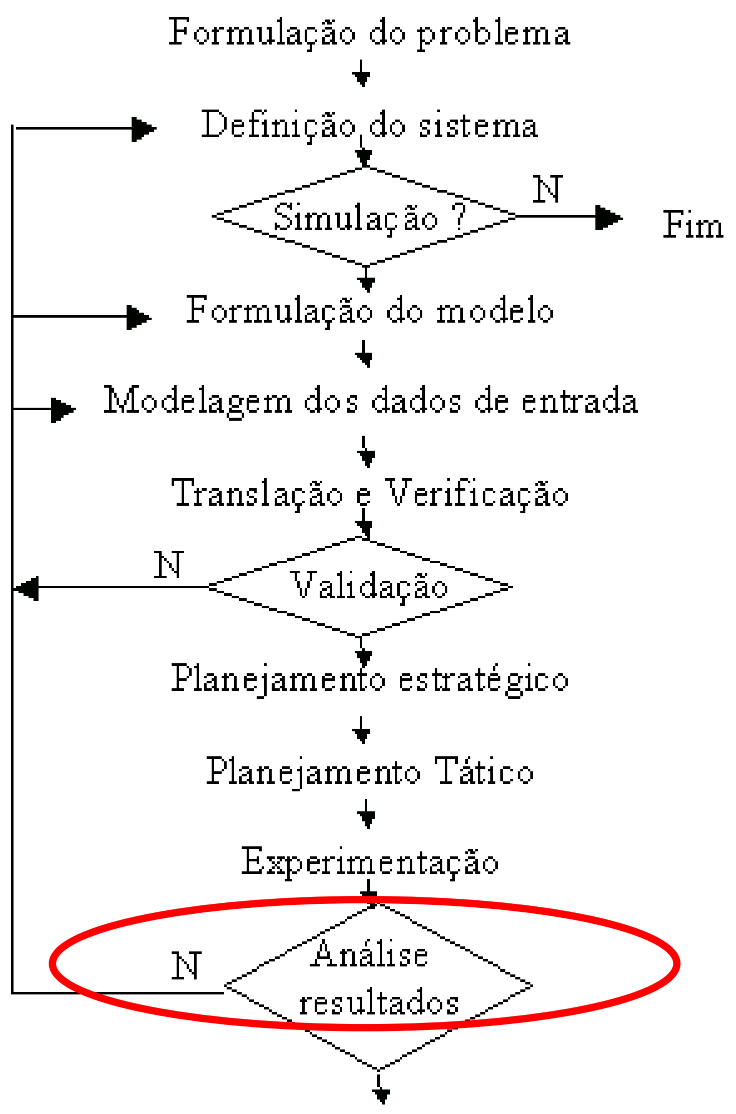
Formulação do problema
Definição do sistema
Simulação ?
N
Fim
Formulação do modelo
Modelagem dos dados de entrada
Translação e Verificação
N
Validação
Planejamento estratégico
Planejamento Tático
Experimentação
N
Análise resultados

Planejamento estratégico

- Objetivos:
  - Encontrar o conjunto de experimentos que melhor representa a resposta desejada
  - Este conjunto deve ter o menor tamanho possível => minimizar o número de experimentos (*runs*)
  - Obter o máximo de informação com o menor esforço
  - Análise de sensibilidade do modelo aos dados de entrada

Planejamento Tático

- Como será realizado cada um dos experimentos?
- Tipo de Simulação
  - Terminais
  - Não Terminais
- Tipo de Simulação → Método de Análise de Resultados


Formulação do problema
Definição do sistema
Simulação ?
N
Fim
Formulação do modelo
Modelagem dos dados de entrada
Translação e Verificação
N
Validação
Planejamento estratégico
Planejamento Tático
Experimentação
N
Análise resultados

Introdução - Análise de resultados

- Como obter resultados a partir dos experimentos?
- Como realizar inferências e previsões sobre o comportamento e desempenho do sistema?

Introdução ...

- A simulação de um sistema com resultados estocásticos resulta na medida de desempenho que contêm variação randômica.
- A análise apropriada de saída permite obter resultados estatísticos mais relevantes.

Introdução ...

- Para se concluir sobre o desempenho de um sistema é necessário observar o comportamento das variáveis de resposta, na medida da realização dos experimentos.
- Uma vez realizados os experimentos, estima-se o comportamento do sistema real por um processo de inferência a partir do conjunto de resultados obtidos.

Experimentação e Análise de Resultados

- O processo começa com a seleção de variáveis de resposta relacionadas com medidas de desempenho que se deseja aferir no sistema.
- Estas podem ser:
  - elementos contadores de ocorrências,
  - obtidas de alguma expressão incluindo médias, variâncias, etc.

Exemplos

- Alguns exemplos de variáveis de resposta:
  - tempo de sistema,
  - tempo nas filas,
  - número de entidades nas filas,
  - utilização de recursos,
  - vazão, etc.

- Questões que devem ser inicialmente tratadas:
  - Qual a duração apropriada de um rodada de simulação?
  - Como interpretar corretamente os seus resultados?
  - Como analisar corretamente as diferenças obtidas em cada uma das replicações?

- Várias técnicas são utilizadas:
  - Avaliação visual dos dados,
  - Técnicas estatísticas como:
    - médias móveis,
    - correlação e
    - definição de intervalos de confiança.

Análise de Resultados...

- O intervalo de confiança é base da análise de resultados para uma medida de desempenho sendo estimada.
- Um intervalo de confiança compreende um intervalo numérico que possui uma probabilidade de incluir o valor verdadeiro da variável ou medida de desempenho sob análise.
- Exemplo:
  - Intervalo = (60,06 ; 74,74) c/ probabilidade 99%

- O nível de confiança do intervalo

(a probabilidade) é dado por **(1 - α)**.

  - Ele pode assumir valores como 99% ou 90%.

- α é o erro admitido ao se concluir sobre a presença do verdadeiro valor da variável no intervalo calculado.

- Exemplo:
  - Nível de confiança = 95%
  - Erro admitido para α = 0,05 = 1 - 0,05.

Exemplo

Para os dados da tabela deseja-se construir intervalos de confiança de:

- 95% (α = 0,05) e de
- 99% (α = 0,01).

- Os n valores foram obtidos das n replicações realizadas

Número da Replicação	Tempo Médio de Fila
1	63,2
2	69,7
3	67,3
4	64,8
5	72,0

Exemplo

- Como os valores para a variável aleatória são normalmente distribuídos, o semi-intervalo h permite o cálculo do intervalo de confiança

$(1 - \alpha)$ para o verdadeiro valor da média μ que estará centrado em $\bar{X}$

- O semi-intervalo h é calculado por:

$$h = t_{n-1,1-\alpha/2} \frac{S}{\sqrt{n}}$$

Valor tabelado : $t_{n-1,1-\alpha/2}$

S: Variância amostral

n o número de replicações

Exemplo

A partir da tabela, os valores calculados são:

$$\text{Média amostral}: \ \overline{X} = 67,4$$

$$\text{Variância amostral} \ S = 3,57$$

$$\overline{X} = \sum_{i=1}^{n} p_i x_i = \mu$$

$$S = \sum_{i=1}^{n} p_i (x_i - \mu)^2$$

Assim para 95% de confiança

- α = 5%,
- (1 - α/2) = 0,975
- $t_{4 \ , 0.975}$ = 2,78.

Assim para 99% de confiança

- α = 1%,
- (1 - α/2) = 0,995
- $t_{4 \ , 0.995}$ = 4,60

Semi-intervalo h é calculado por:

$$h = t_{n-1,1-\alpha/2} \frac{S}{\sqrt{n}}$$

Onde n = 5 repetições

Número de Replicações	Tempo Médio de Fila
1	63,2
2	69,7
3	67,3
4	64,8
5	72,0

Exemplo

- Dessa forma
- h = 4,44 (95% de confiança)
- h = 7,34 (99% de confiança)
- Como o intervalo é simétrico diante da medida os limites serão:

$$\left[\overline{X} - h, \overline{X} + h\right] \qquad \overline{X} = 67,4$$

  - O intervalo com 95% de confiança é (62,6;71,84)
  - O intervalo com 99% de confiança é (60,06; 74,74)

Exemplo

- O tamanho do intervalo depende do nível de confiança desejado.
- Como vimos, o número de replicações n e o desvio padrão S, também são utilizados na obtenção de t e, por consequência, na obtenção do intervalo de confiança.

Exemplo

- A variância da medida de desempenho de um sistema e de seu modelo, é dependente dos parâmetros do sistema.
- Não sendo possível modificar os elementos do modelo e sua natureza estocástica, a liberdade da atuação do analista fica restrita à possibilidade de lidar com:
  - alterações no número de replicações e
  - no nível de confiança.

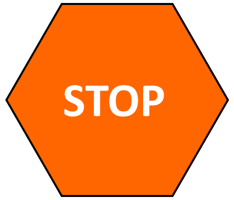
Exemplo

- Se o nível de confiança é fixo (ex: 95%), um grande número de replicações resultará num menor intervalo de confiança (maior precisão).
- Se o tamanho do intervalo de confiança é fixo ex:(62,6;71,84),um grande número de replicações resultará em um maior grau de confiança
- Se o número de replicações é fixo (ex: n=5), um alto nível de confiança resultará em um grande intervalo de confiança.

- Assim, a definição do número de replicações que deverão ser realizadas em um estudo de simulação é um fator relevante na determinação do intervalo de confiança.

- O procedimento de análise de resultados é diferente de acordo se o sistema terminal ou não terminal.

Análise de Resultados em Sistemas Terminais



Terminantes

- Tem condição de início e
- Tem uma definição de um evento que marca o fim da simulação.

O sistema volta para a condição inicial fixada, antes de iniciar nova operação.

Sistemas Não Terminantes

- Sistema cuja duração não é definida;
- O sistema está perpetuamente em operação;
- O objetivo é compreender o comportamento de um estado estável.
- Para estudar esse comportamento, os efeitos das condições iniciais, ou fases transientes, devem ser removidos dos resultados.
- A maioria dos sistemas é não terminal.

Análise de Resultado em Sistemas Terminais

- O objetivo nesses sistemas é compreender o comportamento do sistema em uma duração típica fixa.
- Como as condições iniciais e o tamanho da simulação são fixos, o único fator a controlar é o número de replicações.

- Um procedimento de análise é:
  - simular o número de replicações,
  - computar a amostragem da variância da medida estimada selecionada e
  - determinar **se** o tamanho do intervalo de confiança resultante está nos limites aceitáveis.

Considerações:

Deseja-se estimar o desempenho de um sistema, considerando o valor médio de uma variável de controle com:

- um nível de precisão de ± **r %** e
- um nível de confiança de **100 (1- α)%.**

- O tamanho da amostra piloto pode ser definido de acordo com um procedimento estatístico ou de uma forma intuitiva

- Para uma amostra n o intervalo de confiança 100 (1- α)% da média populacional da variável de interesse é dado por:

$$\left[\overline{X} \pm z\frac{s}{\sqrt{n}}\right]$$

$$\overline{X} = 67{,}4$$

Z : variável Normal padronizada para o nível de confiança x

- A precisão desejada para r % implica em que o intervalo de confiança deve ser:

$$(\overline{x}(1-r/100), \overline{x}(1+r/100))$$

$$\bar{x} \pm z\frac{s}{\sqrt{n}} = \bar{x}\left(1 \pm \frac{r}{100}\right)$$

simplificando-se tem-se:

$$z\frac{s}{\sqrt{n}} = \bar{x}\,\frac{r}{100}$$

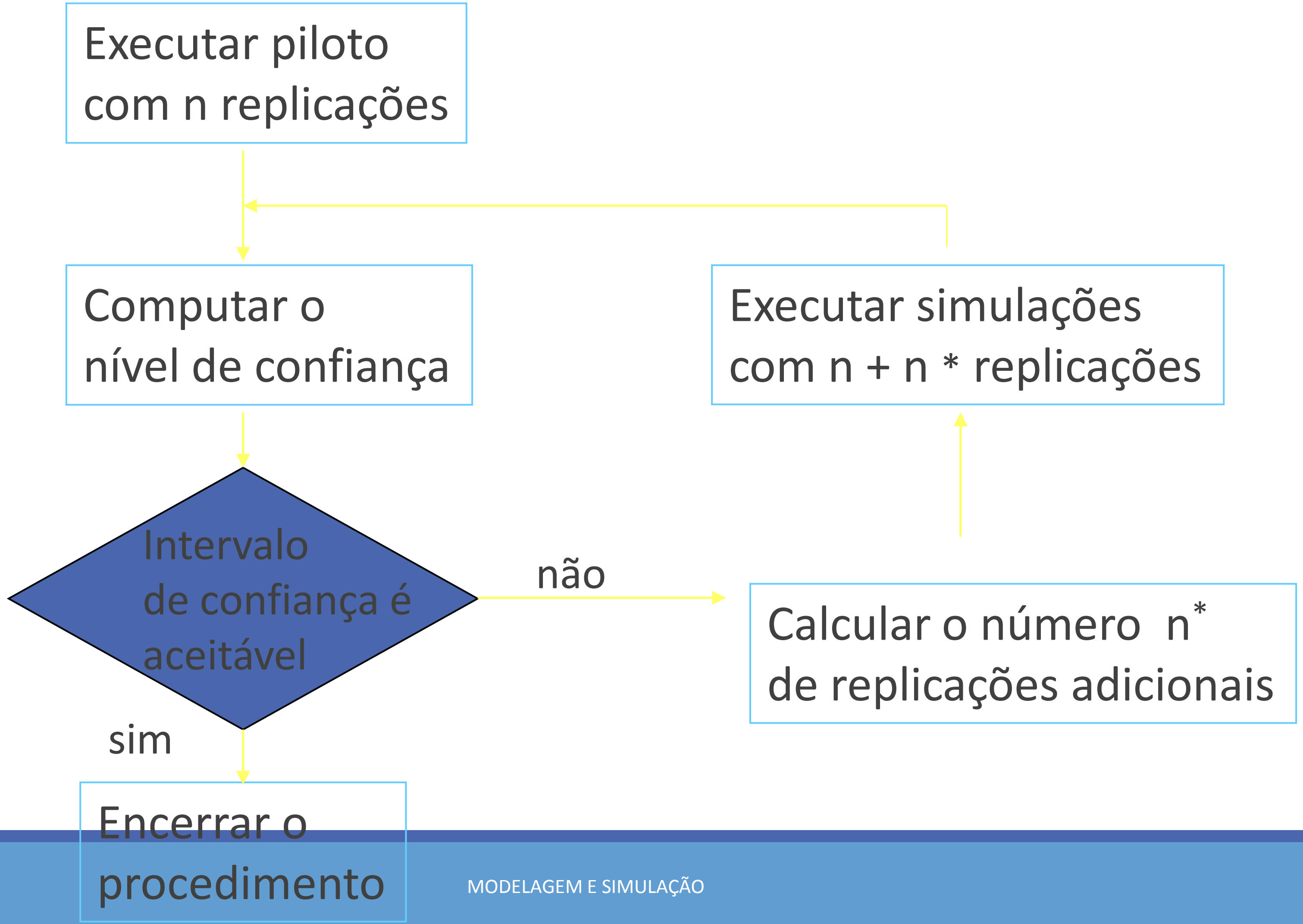
De onde se obtém a fórmula para obter o número de replicações a partir dos parâmetros:

$$\mathbf{n} = \left(\frac{\mathbf{100\,z\,s}}{\mathbf{r\,\bar{x}}}\right)^{2}$$

x: média amostral

s : variância

Análise de Resultado em Sistemas Terminais



Análise de Resultado em Sistemas Não Terminais

Introdução

- Maioria dos sistemas é não terminal
- Não se tem estado inicial pré-definido em um evento caracterizando o encerramento do período de simulação.
- Problemas básicos a contornar:
  1. Descarte das observações que pertencem ao período transiente.
  2. Determinação do Tamanho do Período de Simulação.

     Quão longas devem ser as simulações?

Eliminação de efeitos de fases transientes

- Simulação Longa
  - Suprime efeitos de condições iniciais conduzindo execuções muito longas tal que as condições iniciais tenham um efeito diluído no desempenho,
  - Desvantagens:
    - desperdício de recursos e
    - dificuldade em assegurar que a rotina é suficientemente longa.
  - Este método por si só <u>não</u> <u>é recomendado</u>

Eliminação de efeitos de fases transientes ...

- Inicialização Apropriada:
  - Apronta o sistema antes dele iniciar a simulação;
  - Tenta fazer com que as condições inicias sejam semelhantes à de estados estáveis.
  - Exemplo: o número de entidades nas filas poderá ser fixado considerando análises anteriores.

Eliminação de efeitos de fases transientes ...

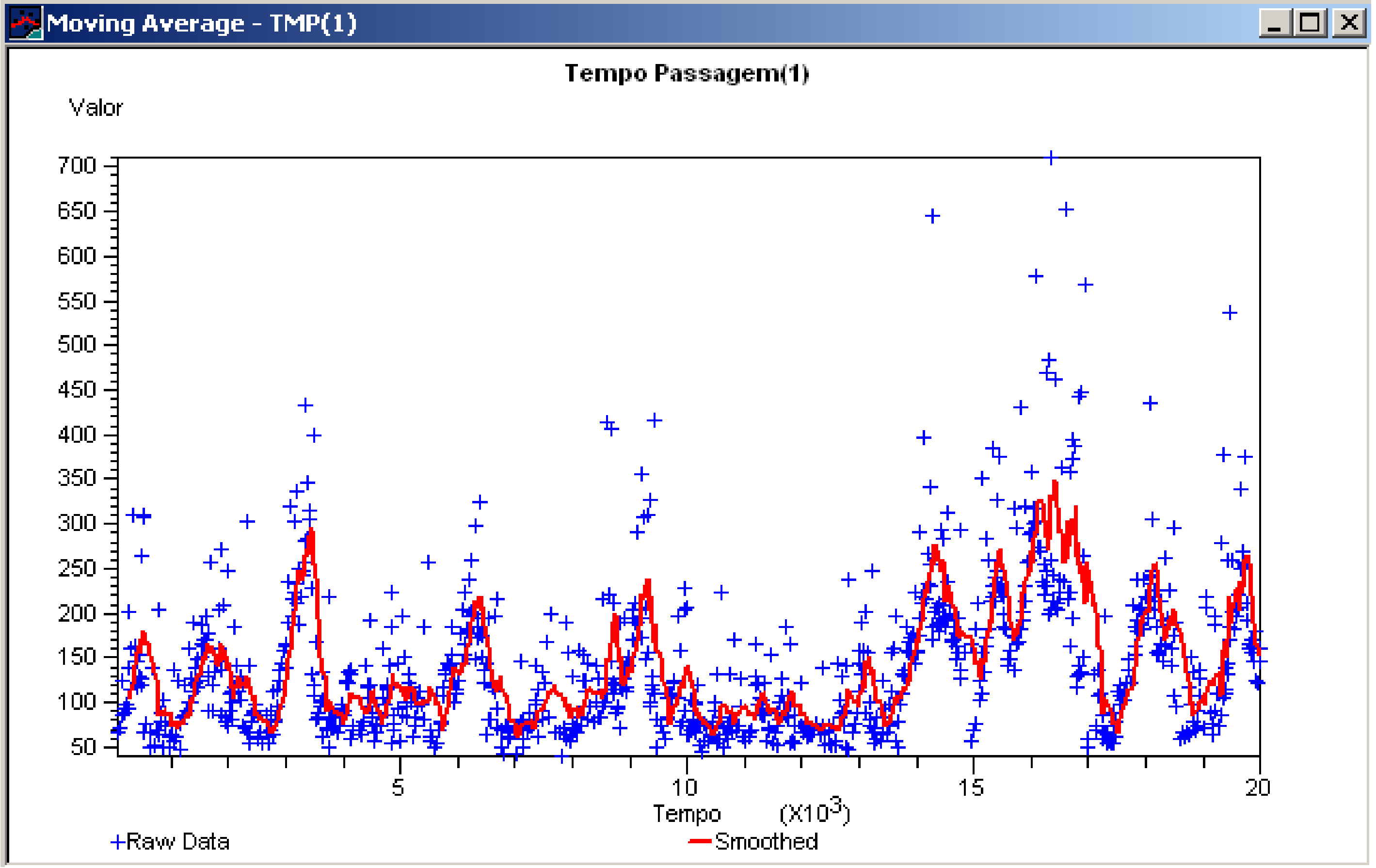
- Truncagem:
  - Exclui a fase inicial transiente que é influenciada por condições iniciais.
  - Os dados são coletados apenas quando a fase transiente acaba.
  - Embora de fácil implementação nem sempre os resultados são corretos.

Eliminação de efeitos de fases transientes ...

- Observação visual
  - Método mais fácil.
  - A partir da construção de um gráfico que aponta o comportamento da variável de reposta ao longo do tempo, procura-se observar, de forma aproximada, em que momento as respostas passam a ter uma conduta mais estabilizada.

- Observação visual ....
  - Em alguns sistemas a flutuação da resposta é muito acentuada dificultando a observação.
    - Nesse caso, é aconselhável desenhar gráficos com médias móveis.

- A média móvel é construída calculando-se:
  - a média aritmética das k mais recentes observações em cada ponto do conjunto de dados.
  - O valor de k é selecionado pelo analista e poderá variar de acordo com a variável sob investigação.


Moving Average - TMP(1)
Tempo Passagem(1)
Valor
700
650
600
550
500
450
400
350
300
250
200
150
100
50
5
10
15
20
Tempo (X10^3)
Raw Data
Smoothed

Sistemas Não Terminais

Determinação do tamanho do período de simulação

Nos sistemas terminais ...

- Estimava-se a variância da média da medida de desempenho sob observação,
- gerava-se uma amostra a partir dos valores médios da variável de interesse, obtidos a partir das diversas replicações realizadas.
- Nestas replicações o estado inicial do sistema permanece sempre o mesmo, alterando-se apenas sementes geradoras de números aleatórios.

- Dessa forma garante-se que cada valor gerado por cada uma das replicações sejam independentes (o que é exigido pelo formulário aplicado).

- A mesma abordagem pode ser aplicada a sistemas não terminais.

A) Método de Replicações independentes

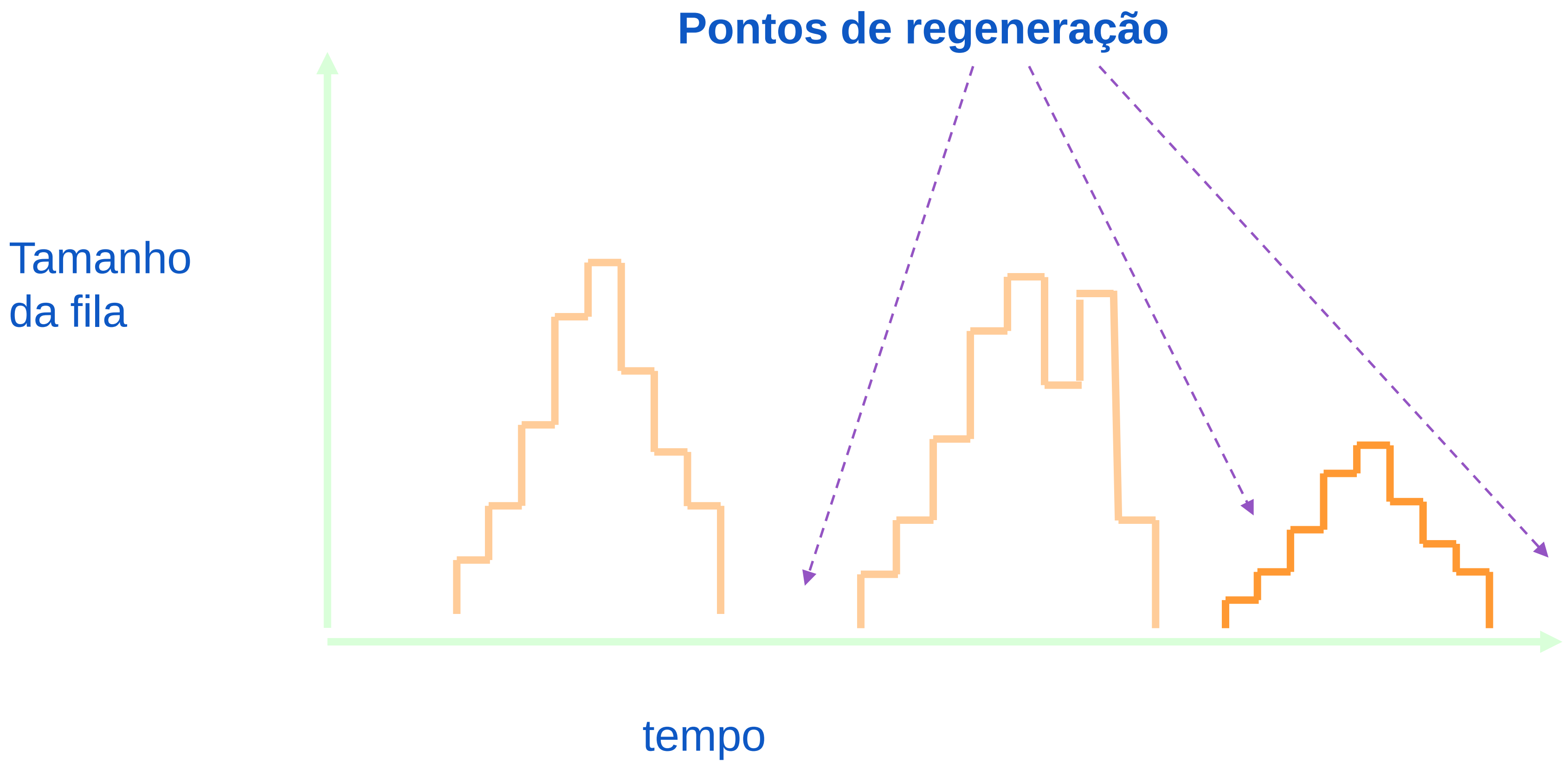
- Conduz m replicações de tamanho $= n + n_0$, onde n_0 é o tamanho do período transiente.
- As primeiras n_0 observações de cada replicação são descartadas.
- Os seguintes passos devem ser seguidos:
  - Calcula-se a média de cada replicação
  - Calcula-se a média geral de todas as replicações (média das médias)
  - Calcula-se a variância das médias das replicações.
  - Obtém-se o intervalo de confiança.

- Para sistemas não terminais alguns problemas se contrapõe ao uso de replicações para geração da amostra e cálculo do intervalo de confiança:
  - critério de parada: replicações longas e inúmeras podem inviabilizar o estudo
  - se a fase transiente for muito grande a quantidade de dados descartados e gerados é muito grande, implicando em desperdício.

B) Método de Regeneração

- A variável sob investigação volta várias vezes ao seu estado inicial.
- O sistema renasce de forma totalmente independente da sua trajetória ou comportamento anterior.
- Os momentos em que o sistema reinicia uma nova fase são chamados de pontos de regeneração.
- O tempo decorrido entre dois pontos sucessivos de regeneração é chamado de ciclo de regeneração.

- Para melhor entender:
  - Considere uma fila diante de um terminal de atendimento eletrônico de um banco que opera entre 8:00 e 20:00h.
  - Uma vez que inicialmente a fila do terminal se encontra vazia ao inicio das operações diárias , é possível que o comportamento da variável tamanho de fila ao longo do período de operação do terminal possa ser parecida com o gráfico a seguir:



Pontos de regeneração
Tamanho
da fila
tempo

Exemplo

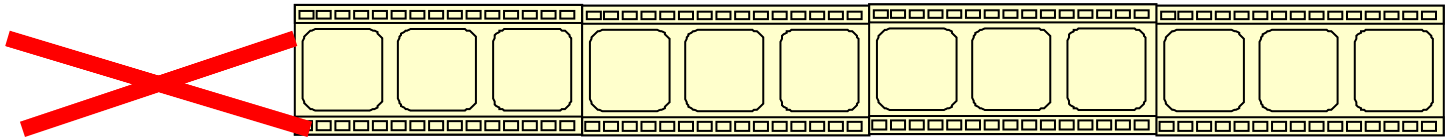
- O tempo de espera de um cliente na fila do terminal depende da demanda deste recurso por parte de clientes anteriores.
- Porém, o tempo no sistema de um cliente que chega para ser servido e encontra o sistema sem filas não depende mais do ocorrido com os clientes que passaram antes pelo sistema.
- Na realidade, inicia-se uma nova fase no sistema.

- Os ciclos são independentes.
- Dessa forma a média do tamanho da fila no segundo ciclo é independente da média do primeiro ciclo.
- O argumento da independência também é válido para as outras variáveis.

- Na medida do aumento do número de filas, os pontos de regeneração tornam-se raros e os ciclos cada vez mais longos.
  - Ex sistema com 2 filas só regenera quando as duas estão vazias
- Considera-se que sistemas com uma memória regenerativa muito longa tornam-se não regenerativos.
  - Computar a variância destes sistemas usando os ciclos de regeneração é mais complexo que fazer uso de outros métodos

C) Método de Loteamento

- Realiza uma simulação longa seguida do descarte das observações iniciais e da divisão das observações restantes numa série de lotes ou de sub-amostras.
- Dada uma simulação longa de N+ n_0 observações, n_0 é o número de observações que pertencem ao período transiente e são descartadas.
- As N observações remanescentes são divididas em m = (N/n) lotes de n observações cada uma.



Loteamento

- Para a definição do intervalo de confiança os seguintes passos devem ser seguidos
  - Calcular a média de cada replicação para cada um dos **m** lotes individualmente
  - Calcular a média geral de todas as replicações (a média das médias dos lotes)
  - Calcular a variância das médias das replicações.
  - O intervalo de confiança para a média da variável de desempenho é então obtido.

- No Loteamento:
  - o desperdício dos dados é muito menor pois apenas um único descarte de n_0 observações é realizado.
  - O tamanho n do lote deverá ser longo o suficientemente para que a correlação entre as médias de cada lote seja muito pequena.
    - Isto permite a formulação acima que pressupõe a independência entre as médias de cada lote.
    - Uma forma de encontrar o tamanho correto para n é calcular a covariância entre as médias de lotes sucessivos.

Tratamento de sistemas não terminais

Executar uma simulação longa piloto

Determinar o período transiente

Período transiente curto?

não

Adotar técnica de loteamento

Determinar correlação

Determinar tamanho de lote mínimo

Executar longa simulação para gerar lotes necessários à criação da amostra piloto

sim

Adotar técnica de múltiplas replicações independentes, eliminando o período transiente de cada uma delas

Seguir com procedimento semelhante adotado para sistemas terminais

Média

Média $\mu = E(x)$

$$= \sum_{i=1}^{n} p_i\, x_i$$ **Para variáveis discretas**

$$= \int_{-\infty}^{+\infty} x f(x)\, dx$$ **Para variáveis contínuas**

Soma de todos os valores possíveis, ponderada pela probabilidade de ocorrência de cada um dos valores.

Variância

A quantidade *(x-u)*^{2} representa a distância quadrática entre *x* e a sua média.

A variância de *x* é o valor esperado desta quantidade:

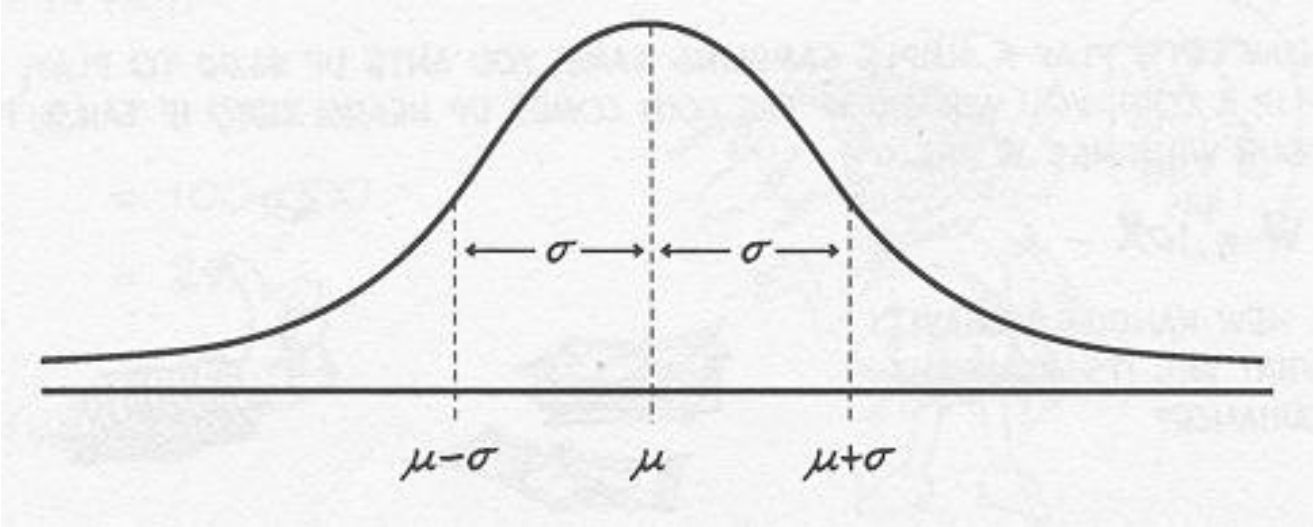
$$Var(x) = E[(x-\mu)^2] = \sum_{i=1}^{n} p_i (x_i - \mu)^2$$

$$= \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx$$

Desvio Padrão

A variância é normalmente denotada por s^2.

A raiz quadrada da variância é chamada de desvio padrão e é denotado por s.



Ex: Número de vendas realizadas em um período de 12 dias

15	12	18	10
12	13	13	15
14	16	20	22

$$\bar{X} = \frac{15+12+21+18+10+12+13+13+15+14+16+20+22}{12} = 15$$

$$\text{Variância} = \frac{1(0+9+9+25+9+4+4+0+1+1+25+29)}{12} = 11{,}3$$

Desvio padrão= σ = raiz quadrada de 11,3 =3,4

x_i	$\bar{x}-x_i$	$(\bar{x}-x_i)^2$
15	0	0
12	3	9
18	-3	9
10	5	25
12	3	9
13	2	4
13	2	4
15	0	0
14	1	1
16	-1	1
20	-5	25
22	-7	49

Coeficiente de Variação

$$\text{C.O.V.} = \frac{\text{desvio padrão}}{\text{média}} = \frac{\sigma}{\mu}$$

Covariância

Dadas duas v.a.s *x* e *y* com médias m_x e m_y, a covariância delas é dada por:

$$Cov(x, y) = s_{xy}^{2} = E[(x - m_x)(y - m_y)]$$

$$= E(xy) - E(x)E(y)$$

Para variáveis independentes a covariância é zero, dado que $E(xy) = E(x)E(y)$

Apesar da independência sempre implicar em covariância zero, o contrário nem sempre é verdade.

Input e Output Analyzer

PRÓXIMA AULA