

AGOSTO DE 2024

SIMULAÇÃO COMPUTACIONAL

Aplicações em Transporte Aéreo

Marcelo Xavier Guterres, Ph.d

Mauro Caetano de Souza, Ph.d

Produzido por:

Marcelo Xavier Guterres, Ph.d
Mauro Caetano de Souza, Ph.d

Direitos autorais da MXG.

Aprovado por:

Marcelo Xavier Guterres
Professor Associado II do Instituto Tecnológico de Aeronáutica.

Sumário

1	Simulação Computacional	1
1.1	Conceito de Simulação Computacional	2
1.2	Por que Simular?	4
1.3	Vantagens e desvantagens da Simulação	4
1.4	Áreas de aplicação	6
1.5	Tipos de Simulação	7
1.5.1	Sistemas contínuo e discreto	7
1.5.2	Modelo estático e dinâmico	8
1.5.3	Modelo determinístico ou estocástico	8
1.5.4	Simulação terminante e não terminante	9
2	Um Modelo de Simulação	11
2.1	Um Sistema de Atendimento Simples	12
3	Coleta e Modelagem dos Dados de Entrada	19
3.1	Coleta dos Dados	19
3.2	Outliers ou Valores Discrepantes	22
3.3	Exercício: Questionário sobre Testes de Aderência à Normalidade	25
3.4	Exercício: Análise de Sistema de Filas de Atendimento	26

Lista de Figuras

1.1	Simulação de sistemas contínuo versus discreto	8
1.2	Classificação de Sistema, Modelo e Simulação	10
2.1	Modelo físico de atendimento em um banco.	12
2.2	Representação lógica simplificada de um sistema de atendimento.	13
2.3	Metodologia de desenvolvimento do modelo de simulação.	14
2.4	Exemplo de dados coletados para o modelo de atendimento.	15
2.5	Parâmetros do sistema de atendimento.	16
2.6	Tradução do modelo de simulação para o Arena® (a).	17
2.7	Tradução do modelo de simulação para o Arena® (b).	17
2.8	Tradução do modelo de simulação para o Arena® (c).	18
3.1	Intervalo entre chegadas de pessoas nos balcões de atendimento.	21

Lista de Tabelas

3.1 Medidas estatísticas dos intervalos entre chegadas.	21
---	----

1. Simulação Computacional

Objetivos

1. Definir o conceito de Simulação Computacional;
2. Por que simular;
3. Vantagens e desvantagens da Simulação;
4. Áreas de aplicação;
5. Tipos de Simulação;

Observação

Este material não substitui o livro texto, as referências recomendadas e nem as aulas expositivas. Este material fez transcrições parciais e completas dos materiais listados na bibliografia.

1.1 Conceito de Simulação Computacional

Um **SISTEMA** é todo e qualquer objeto sobre o qual se pretende realizar um determinado estudo, enquanto o seu **MODELO** é uma representação qualquer desse objeto na qual se irá efetivamente executar tal estudo.

A **EXPERIMENTAÇÃO**, como se sabe, é realizada nos mais diversos campos da ciência. Sendo executada sobre o próprio **sistema**, estimulando-o de forma apropriada e registrando depois as alterações observadas nos parâmetros que definem o seu estado. Este método de abordar e de caracterizar sistemas tem vindo a ser usado desde os tempos mais remotos da ciência, permanecendo, ainda hoje, como o **procedimento mais eficaz no que respeita à fiabilidade dos resultados obtidos**.

Na **construção de modelos**, como é óbvio, tal processo introduz, só por si, uma **fonte de incertezas nos resultados**. Uma vez que o **modelo será sempre uma "aproximação" ao sistema real e não o verdadeiro sistema**. Esta é uma questão para a qual o analista deverá estar alertado, pois nem sempre poderão ser revestidos de significado os resultados obtidos através de experiências realizadas sobre um modelo. Tudo depende, como é de esperar, do **grau de precisão** com que esse modelo consegue representar o sistema, e, portanto, também da incerteza admissível associada aos resultados da experiência, ou seja, em última análise, dos objetivos que se pretendem atingir e da sensibilidade do próprio analista.

Taha [3] coloca que a **SIMULAÇÃO** é o segundo melhor procedimento depois da observação do sistema real. Trata de uma **IMITAÇÃO COMPUTADORIZADA** do **COMPORTAMENTO ALEATÓRIO** de um sistema com a finalidade de **ESTIMAR** suas **MEDIDAS DE DESEMPENHO**.

De acordo com Shimizu [2], **SIMULAÇÃO** é essencialmente um trabalho com analogias. É uma modalidade de pesquisa que procura tirar conclusões por meio de exercícios com **MODELOS** que representam a **REALIDADE**.

SIMULAÇÃO é portanto um **PROCESSO** de imitar uma realidade por meio

de modelos. Tais modelos podem conservar ou não, as características **FÍSICAS** e **LÓGICAS** do **SISTEMA REAL** imitado.

Em computação, **SIMULAÇÃO** consiste em empregar formalizações em computadores, tais como expressões matemáticas ou especificações mais ou menos formalizadas, com o propósito de imitar um processo ou operação do mundo real. Desta forma, para ser realizada uma simulação, é necessário construir um **MODELO COMPUTACIONAL** que corresponda à situação real que se deseja simular.

A simulação computacional de sistemas, ou apenas simulação, consiste na utilização de certas técnicas matemáticas, empregadas em computadores, as quais permitem imitar o funcionamento de, praticamente qualquer tipo de operação ou processo do mundo real. Ou seja, é o estudo do comportamento de sistemas reais através do exercício de modelos.

Existem diversas outras definições para a simulação, dentre elas podemos destacar:

a simulação é um processo de projetar um modelo computacional de um sistema real e conduzir experimentos com este modelo com o propósito de entender seu comportamento e/ou avaliar estratégias para sua operação.

Desta maneira, podemos entender a simulação como um processo amplo que engloba não apenas a construção do modelo, mas todo o método experimental que se segue, buscando:

- Descrever o comportamento do sistema;
- Construir teorias e hipóteses considerando as observações efetuadas;
- Usar o modelo para prever o comportamento futuro, isto é, os efeitos produzidos por alterações no sistema ou nos métodos empregados em sua operação;

1.2 Por que Simular?

Imitar! Construir modelos que apresentam o maior número possível de características de sistemas reais, e então experimentar, para depois aprender [2].

Cada vez mais procuramos resolver os nossos problemas por meio de uma analogia com a realidade. Muitas vezes somos obrigados a isso, pela impossibilidade de testar, diretamente no sistema real, as nossas técnicas e hipóteses de resolução [2].

Seria o caso, por exemplo, dos nossos "ancestrais- os símios - testados como primeiros astronautas [2].

Outras vezes a limitação está no tempo, como as experiências feitas no campos da genética aplicada, de resultados bastantes lentos, deixando sempre muito a desejar [2].

Nos problemas de gerência, onde o "tempo-dinheiro" assume proporções "alarmantes", as decisões de negócios reclamam um teste prévio em simuladores os mais avançados possíveis, para que se possa avaliar as suas consequências dentro de um tempo e confiabilidade admissíveis [2].

ESTUDOS DE SISTEMAS - SIMULAÇÃO

Simulação não é uma técnica de otimização: estima-se medidas de performance de um sistema modelado

1.3 Vantagens e desvantagens da Simulação

As vantagens principais da simulação são, de acordo com Santos [1]:

- Novas políticas, procedimentos operacionais, regras de negócio, fluxos de informação, etc..., podem ser estudadas sem se alterar o mundo real.

- Novos equipamentos, layouts, sistemas de transporte, etc..., podem ser testados sem se comprometer recursos na sua aquisição.
- Hipóteses sobre como e porque certos fenômenos ocorrem podem ser testados visando verificar sua praticabilidade.
- O tempo pode ser comprimido ou expandido permitindo acelerar ou retardar o fenômeno sob investigação.
- Pode-se entender melhor sob a interação das variáveis do sistema.
- Pode-se entender melhor a participação das variáveis na performance do sistema.
- Um modelo de simulação pode ajudar a entender como um sistema funciona como um todo, em relação a como se pensa que o sistema opera individualmente.
- Questões do tipo "e se..." podem ser respondidas. Isto é extremamente útil na fase de design de um projeto.

Com mais sofisticação tem-se:

- Estudar as interações internas de um sistema complexo, ou de um sub-sistema dentro de um sistema complexo;
- Realizar alterações nas informações, na organização e no ambiente do sistema para observar seus efeitos;
- Experimentar novos projetos ou novos procedimentos antes de implementá-las, e assim estar preparado para o que puder acontecer.
- Identificar as variáveis mais importantes de um sistema e como elas interagem através do estudo dos sinais de entrada e das saídas resultantes.
- Verificar soluções analíticas, sendo neste caso utilizado como um instrumento de validação.
- Adquirir maior conhecimento sobre o modelo de simulação e sobre o processo de desenvolvimento do modelo para melhorias do sistema.

As desvantagens a serem consideradas são, de acordo com Santos [1]:

- A construção de Modelos de Simulação requer treinamento especial. É uma arte que é aprendida com tempo e experiência. Além disto se 2 modelos são construídos por 2 profissionais competentes, eles terão semelhanças, mas será altamente improvável que sejam iguais.
- Os resultados de uma Simulação podem ser difíceis de interpretar. Como a maioria das saídas de uma simulação são variáveis aleatórias (elas estão normalmente baseadas em entradas aleatórias), é difícil determinar se uma observação é o resultado do relacionamento entre as variáveis do sistema ou consequência da própria aleatoriedade.
- A construção e análise de Modelos de Simulação pode consumir muito tempo e, como consequência, muito dinheiro. Economizar por sua vez pode levar a modelos incompletos.
- A Simulação é usada em muitos casos onde uma solução analítica é possível. A simulação não dá resultados exatos.

1.4 Áreas de aplicação

Santos [1] enumera algumas possibilidades:

- Simulação das operações de uma companhia aérea para testar alterações em seus procedimentos operacionais.
- Simulação da passagem do tráfego em um cruzamento muito grande, onde novos sinais estão para ser instalados.
- Simulação de uma siderúrgica para avaliar alterações nos seus procedimentos operacionais.
- Simulação da economia de um setor de um país para prever o efeito de mudanças econômicas.
- Simulação de batalhas militares visando avaliar o desempenho de armas estratégicas.
- Simulação de sistemas de distribuição e controle de estoque, para melhorar o funcionamento destes sistemas.
- Simulação de uma linha de produção em determinada indústria, para avaliar efeitos de mudanças previstas no processo produtivo.

1.5 Tipos de Simulação

1.5.1 Sistemas contínuo e discreto

MODELOS CONTÍNUOS: Tratam de sistemas cujo comportamento muda continuamente ao longo do tempo. Esses modelos costumam usar equações diferenciais para descrever as interações entre os diferentes elementos do sistema. Um exemplo típico trata da dinâmica da população mundial.

Exemplos clássicos de simulação de sistemas contínuos - é de um avião voando - e a passagem de água por uma barragem.

MODELOS DISCRETOS: Os sistemas cujos estados se possam considerar discretos, ou seja, que variem de forma discreta ao longo do intervalo de valores da variável independente.

O estado do sistema muda somente no instante que ocorre um evento, para todos os demais instantes de tempo, nada muda no sistema. Para isto, utiliza-se os conceitos de Filas.

A simulação de um sistema discreto caracteriza-se por eventos onde as mudanças ocorrem de maneira descontínua, ou seja, sofrem mudanças bruscas.

Um exemplo seria a simulação de uma agência bancária onde entre a chegada (ou a saída) de clientes, o estado do sistema não se altera.

A Fig. 1.1 ilustra essa diferença entre os tipos de simulação de modelos discreto e contínuo.

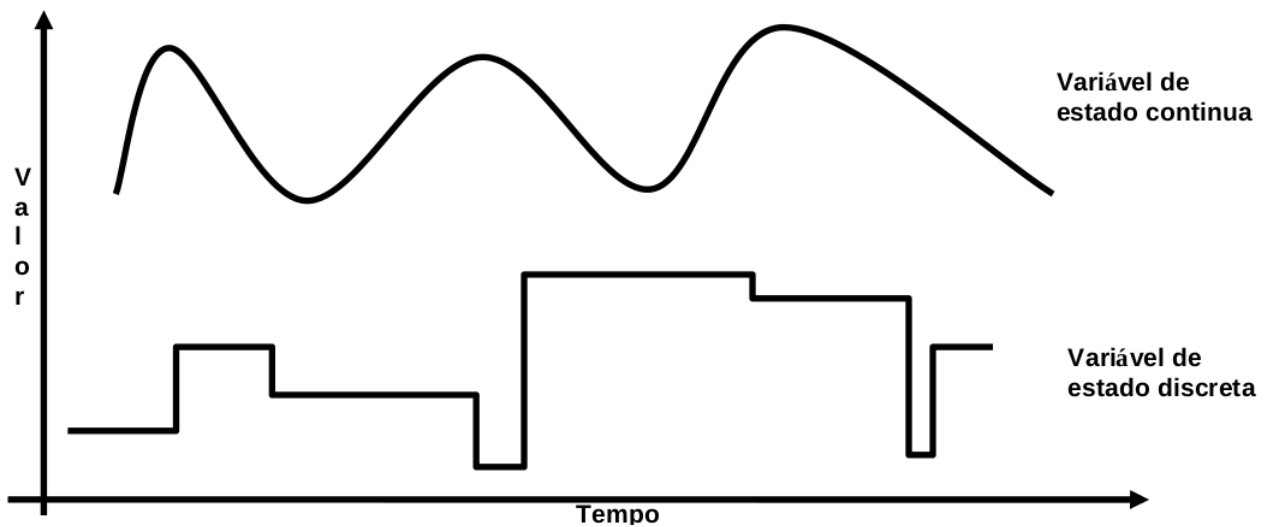


Figura 1.1: Simulação de sistemas contínuo versus discreto

1.5.2 Modelo estático e dinâmico

O MODELO DE SIMULAÇÃO ESTÁTICA é a representação de um sistema em um dado momento. É um modelo onde a passagem do tempo é irrelevante.

O MODELO DE SIMULAÇÃO DINÂMICA é a representação de um sistema no decorrer do tempo.

Esse tipo de simulação é apropriado para a análise de sistemas de manufatura e serviços que sofrem influência do tempo. Um exemplo é a simulação das atividades ocorridas em um banco ao longo de suas oito horas de funcionamento diárias.

1.5.3 Modelo determinístico ou estocástico

Modelos de simulação que não contém nenhuma variável aleatória são classificados como determinísticos, ou seja, para um conjunto conhecido de dados de entrada teremos um único conjunto de resultados de saída.

Um modelo estocástico de simulação que tem uma ou mais variáveis aleatórias como entrada. Onde, estas entradas aleatórias levam a saídas ale-

tórias que podem somente ser consideradas como estimativas das características verdadeiras de um modelo.

Assim, por exemplo, a simulação (estocástica) do funcionamento de uma agência bancária envolve variáveis aleatórias como o intervalo entre chegadas e a duração dos serviços prestados. Logo, medidas como o número médio de clientes esperando e o tempo médio de espera de um cliente, devem ser tratadas como estimativas estatísticas das medidas reais do sistema.

1.5.4 Simulação terminante e não terminante

A Simulação é terminante quando o objetivo é estudar um sistema num dado intervalo de tempo, ou seja, conhecer seu comportamento ao longo deste intervalo, sendo definidas as datas de início e término da simulação.

Exemplos são simulações realizadas em postos de cobrança de pedágio onde se deseja definir o número de postos em funcionamento de acordo com a hora do dia; assim, o intervalo de interesse está entre zero e vinte e quatro horas.

A simulação pode ser classificada como não terminante quando o objetivo é estudar o sistema a partir do momento em que este atinge um estado estável, alcançado após um período de aquecimento, onde se determina e elimina as tendências iniciais.

Presumindo que a simulação poderia continuar infinitamente sem nenhuma mudança estatística no comportamento. Assim, uma simulação do comportamento das pás de uma turbina pode ser considerada não terminante, desde que o interesse seja estudar as características de seu escoamento em condições estáveis, após um período de aquecimento.

SISTEMA	MODELO		SIMULAÇÃO
DISCRETO: Variáveis envolvidas assumem valores finitos ou infinitos numeráveis.	DETERMINISTICO: Variáveis assumem valores determinados.	ESTÁTICO: Estuda o sistema sem levar em conta sua variabilidade com o tempo.	TERMINANTE: Há interesse em se estudar o sistema num dado intervalo de tempo.
CONTÍNUO: Variáveis mudam constantemente com o tempo.	ESTOCÁSTICO: Variáveis assumem valores diversos segundo uma determinada distribuição de probabilidades.	DINÂMICO: Representa o sistema a qualquer tempo.	NÃO TERMINANTE: Há o interesse em estudar o sistema a partir de um determinado estado estável, podendo o estudo, prolongar-se indefinidamente.

Figura 1.2: Classificação de Sistema, Modelo e Simulação

2. Um Modelo de Simulação

Objetivos

1. Compreender e aplicar os conceitos fundamentais da simulação computacional em sistemas de atendimento, incluindo a identificação de componentes físicos e lógicos, e a representação formal de modelos de simulação.
2. Desenvolver habilidades na modelagem de sistemas de atendimento, abrangendo a definição de objetivos, escopo, variáveis de entrada, coleta de dados e análise estatística para parametrização do modelo.
3. Implementar e analisar modelos de simulação utilizando software especializado (como o Arena®), interpretar os resultados obtidos e avaliar a eficiência e adequação de sistemas de atendimento com base em métricas como tempo de espera, ocupação de recursos e tamanho de filas.

Observação

Este material não substitui o livro-texto, as referências recomendadas, ou as aulas expositivas. Ele contém transcrições parciais e completas dos materiais listados na bibliografia e deve ser utilizado como complemento ao estudo. Em síntese, esse material é uma compilação de diferentes materiais de outros autores.

2.1 Um Sistema de Atendimento Simples

Como exemplo inicial, ilustramos a **formação física de uma fila em um caixa de banco**, conforme apresentado na Figura 2.1.

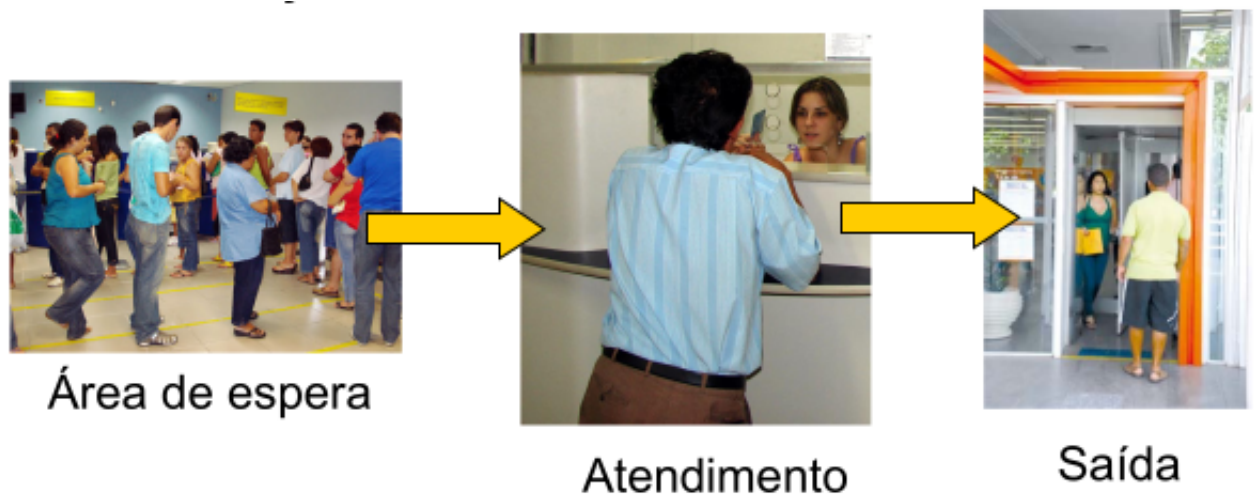


Figura 2.1: Modelo físico de atendimento em um banco.

Os componentes físicos representados na Figura 2.1 são:

- **Área de espera:** Espaço designado para os clientes aguardarem antes de serem atendidos. É essencial para acomodar o fluxo de clientes, evitando a superlotação e garantindo conforto enquanto esperam.
- **Atendimento:** O processo onde o cliente interage diretamente com o caixa ou atendente para realizar suas transações ou solicitações. A eficiência deste processo é fundamental para a satisfação do cliente e para a eficácia operacional do sistema.
- **Saída:** O ponto final do processo de atendimento, onde o cliente deixa o local após completar sua transação. A saída deve ser organizada para evitar congestionamentos e garantir a fluidez no ambiente.

Por outro lado, o **modelo de simulação** deve ser representado de maneira mais formal, conforme ilustrado na Figura 2.2.

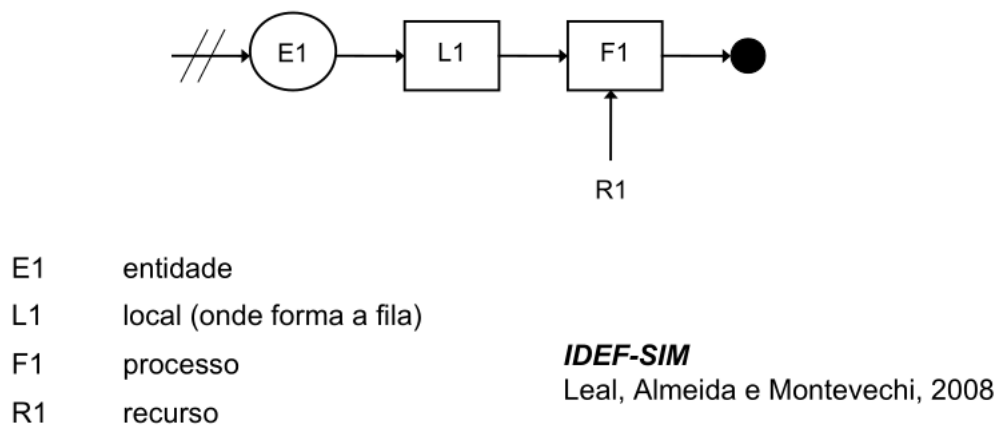


Figura 2.2: Representação lógica simplificada de um sistema de atendimento.

- **E1: Entidade:** Refere-se ao cliente que chega ao sistema de atendimento. Em uma simulação de eventos discretos, a entidade é o objeto principal que flui pelo sistema, desencadeando eventos como chegada, início e fim de atendimento. O comportamento das entidades é uma das principais preocupações do modelo de simulação, pois define como os recursos do sistema serão utilizados.
- **L1: Local:** Representa o ponto no sistema onde as entidades formam uma fila, aguardando para serem atendidas. Em termos de modelagem, o local é o espaço onde a fila se acumula e onde a lógica de fila (como a política de atendimento, por exemplo, FIFO - *First In, First Out*) é aplicada. O gerenciamento eficaz da fila é crucial para evitar congestionamentos e melhorar o fluxo geral do sistema.
- **F1: Processo:** Este é o estágio no qual a entidade é atendida. No contexto de uma simulação de eventos discretos, o processo é onde o recurso (por exemplo, o caixa) interage com a entidade (o cliente) para realizar a transação ou serviço necessário. A duração do processo pode ser modelada com diferentes distribuições de probabilidade, dependendo da natureza do serviço.
- **R1: Recurso:** Refere-se ao elemento do sistema que realiza o processo, como o atendente ou caixa. Em uma simulação de eventos discretos, o recurso é limitado e deve ser alocado adequadamente às entidades que aguardam na fila. A disponibilidade e eficiência do recurso influenciam diretamente o tempo de espera e a taxa de ocupação do sistema.

Algumas questões relevantes sobre o sistema que está sendo modelado incluem:

- A área de espera é adequada?
- O tempo de serviço é aceitável?
- O número de caixas é suficiente?

Para abordar as questões acima, a metodologia para resolução do problema deve seguir as etapas apresentadas na Figura 2.3.

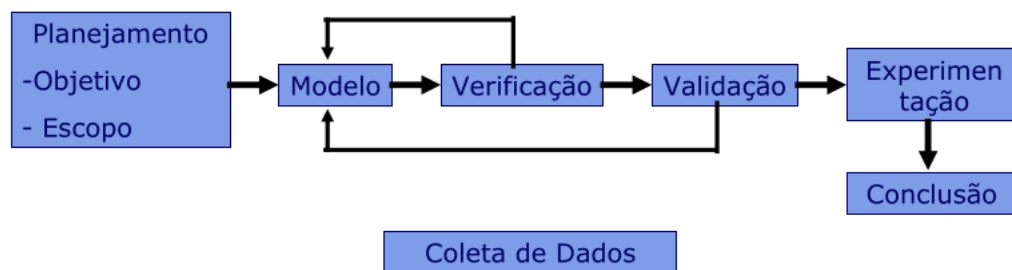


Figura 2.3: Metodologia de desenvolvimento do modelo de simulação.

No que se refere aos **objetivos** e ao **escopo** do sistema em análise, estes poderiam ser, por exemplo:

- Avaliar se a área de espera é adequada.
- Determinar se o tempo de serviço é aceitável.
- Verificar se o número de caixas é suficiente.

Com a finalidade de responder às seguintes **questões**:

- Qual o número de clientes na fila de espera?
- Qual o tempo de permanência do cliente no banco?
- Qual a taxa de ocupação dos caixas?

As **variáveis de entrada** mínimas do modelo computacional seriam:

- Tempo Entre Chegadas (TEC): *Time Between Arrivals*;
- Tempo de Serviço (TS): *Time of Service*;
- Tempo de operação: 6 horas.

Após a definição das variáveis de entrada e do tempo de operação do sistema, deve-se proceder com a coleta de dados. A Figura 2.4 apresenta uma possível forma de organização desses dados.

Cliente	Tempo entre Chegada TEC	Tempo de Chegada Relógio	Tempo Serviço (TS)	Tempo do início do serviço no relógio	Tempo do fim do serviço no relógio
1	0,00	9:23:47	1,08	9:23:50	9:24:55
2	2,68	9:26:28	1,83	9:26:40	9:28:30
3	3,77	9:30:14	1,12	9:30:45	9:31:52
4	7,50	9:37:44	1,00	9:37:55	9:38:55
5	2,68	9:40:25	1,20	9:40:31	9:41:43
6	14,10	9:54:31	0,95	9:54:38	9:55:35
7	1,60	9:56:07	1,82	9:56:20	9:58:09
8	2,38	9:58:30	0,45	9:58:40	9:59:07
9	1,50	10:00:00	0,25	10:00:05	10:00:20
...
50	4,82	12:32:59	0,25	12:33:10	12:33:35

Figura 2.4: Exemplo de dados coletados para o modelo de atendimento.

Após a análise estatística dos dados coletados, os parâmetros estatísticos do sistema são obtidos. Esses parâmetros são então utilizados para formalizar a modelagem, calibrando e ajustando os valores matemáticos do sistema de atendimento do caixa de banco. A Figura 2.5 ilustra os parâmetros ajustados.

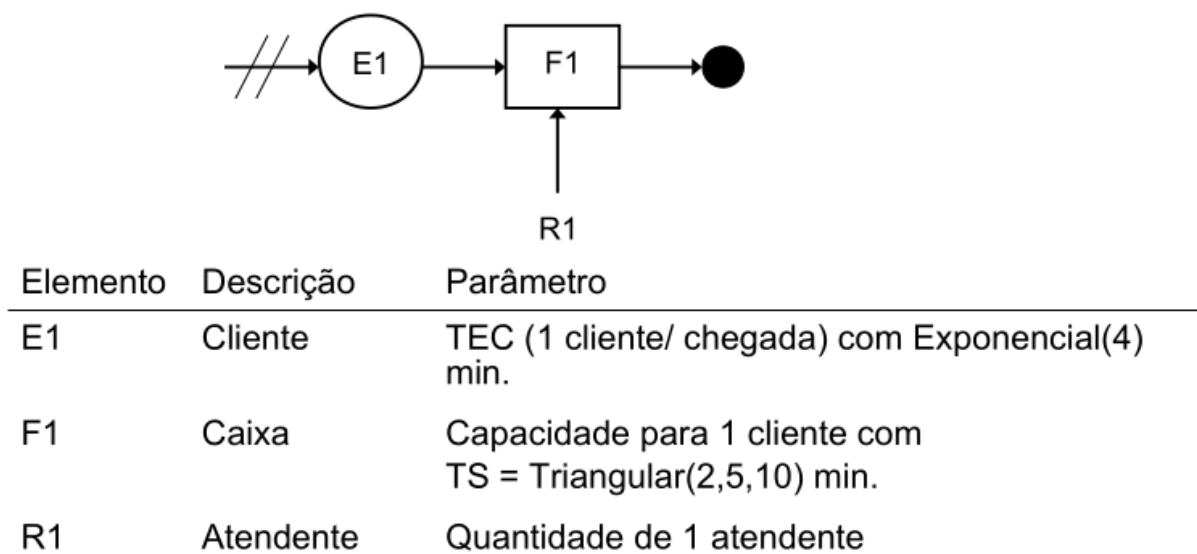


Figura 2.5: Parâmetros do sistema de atendimento.

Em seguida, deve-se escolher um ambiente de simulação comercial, como por exemplo, o Arena® da Rockwell Corporation. Alguns detalhes do Arena® incluem:

- União da linguagem SIMAN com o CINEMA (animação);
- Simulação discreta com modelagem por fluxo de entidades.

As Figuras 2.6 a 2.8 mostram a tradução do modelo de simulação para o Arena®.

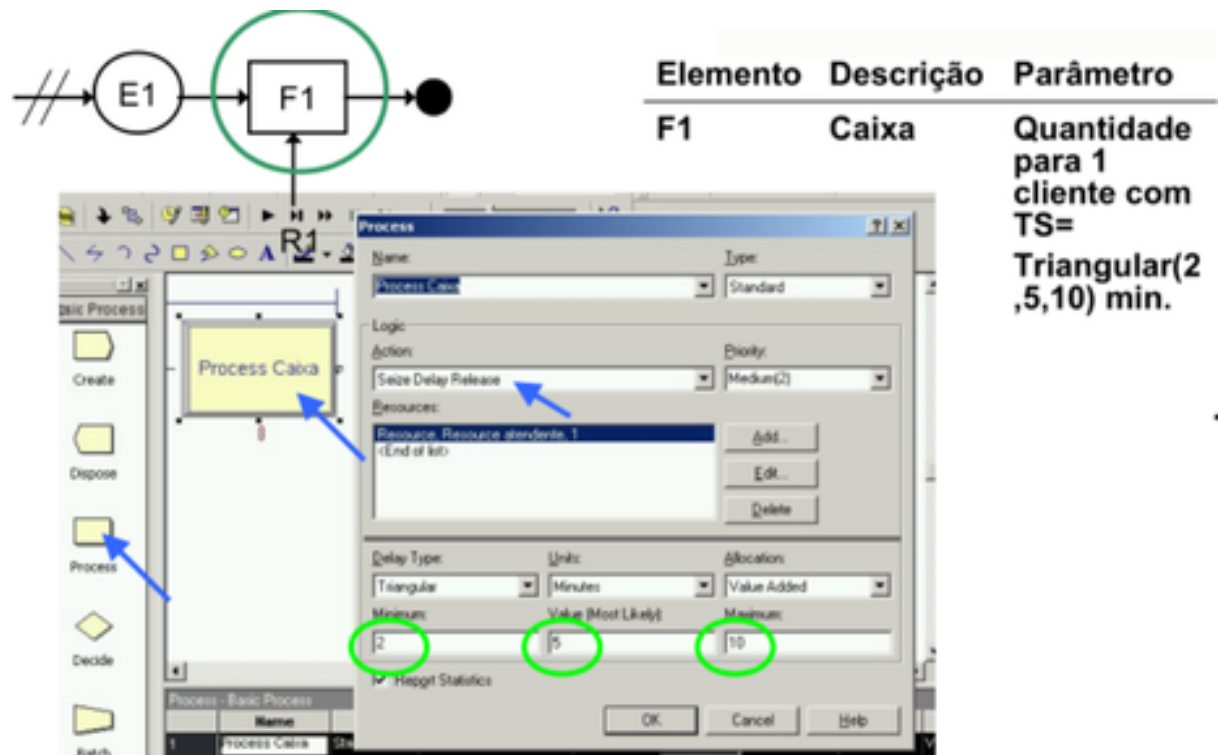


Figura 2.6: Tradução do modelo de simulação para o Arena® (a).

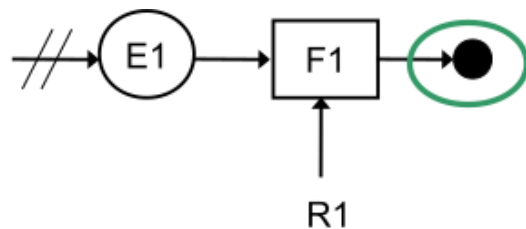


Figura 2.7: Tradução do modelo de simulação para o Arena® (b).

- Configurar experimento

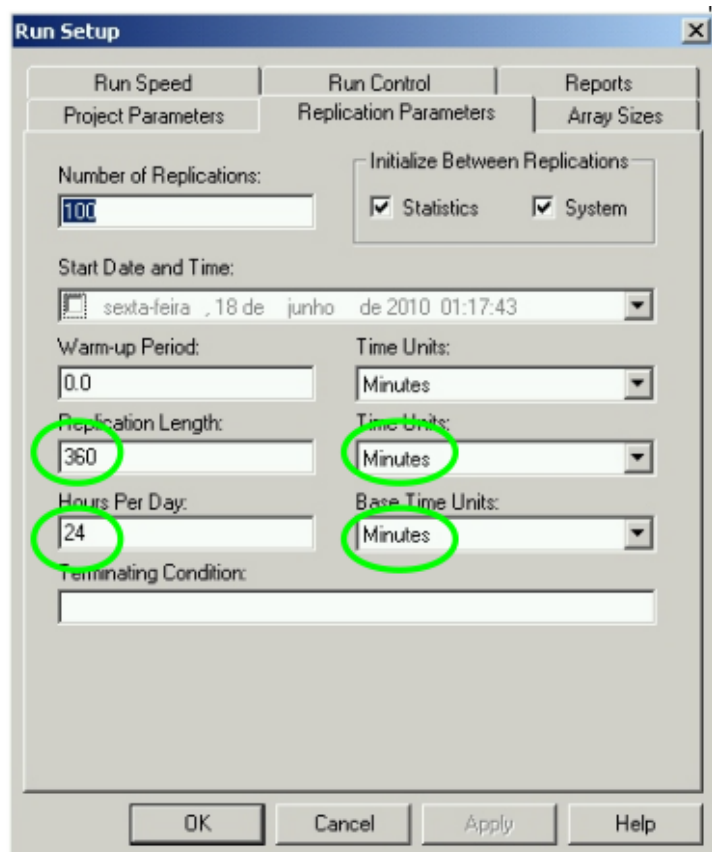


Figura 2.8: Tradução do modelo de simulação para o Arena® (c).

Os resultados finais após as rodadas de simulação foram:

- Número de clientes na fila de espera: 14;
- Tempo médio do cliente na fila: 55 minutos;
- Taxa de ocupação do caixa: 99%.

3. Coleta e Modelagem dos Dados de Entrada

Objetivos

1. Compreender os processos de coleta e modelagem de dados de entrada
2. Identificar e tratar outliers em conjuntos de dados
3. Aplicar técnicas estatísticas para análise de dados discretos

Observação

Este material não substitui o livro-texto, as referências recomendadas ou as aulas expositivas. Ele contém transcrições parciais e completas dos materiais listados na bibliografia e deve ser utilizado como complemento ao estudo. Em síntese, esse material é uma compilação de diferentes materiais de outros autores.

3.1 Coleta dos Dados

O processo de Modelagem dos Dados de Entrada é realizado em três etapas distintas:

1. Coleta
2. Tratamento

3. Inferência

Para garantir a qualidade e representatividade dos dados, alguns procedimentos iniciais devem ser estabelecidos:

- Escolha adequada da variável de estudo;
- O tamanho da amostra deve estar entre 100 e 200 observações. Amostras com menos de 100 observações podem comprometer a identificação do melhor modelo probabilístico, enquanto amostras com mais de 200 observações geralmente não trazem ganhos significativos ao estudo;
- Coletar e registrar as observações na mesma ordem em que o fenômeno ocorre, permitindo a análise de correlação;
- Se houver suspeita de que os dados variam em função do horário ou do dia da coleta, a coleta deve ser realizada em diferentes horários e dias. Na modelagem de dados, toda suspeita deve ser comprovada ou descartada estatisticamente.

Para ilustrar, consideremos um exemplo do setor de aviação. Um gerente de companhia aérea está preocupado com as filas formadas nos balcões de atendimento durante um dos turnos de operação. As variáveis de estudo adequadas para coleta de dados seriam:

- Os tempos de atendimento nos balcões;
- Os tempos entre chegadas sucessivas de clientes nos caixas de atendimento;
- *O número de clientes em fila* (não é uma variável de entrada, pois é um resultado).

A Figura 3.1 apresenta os intervalos entre chegadas de pessoas nos balcões de atendimento (100 medidas) em minutos.

11	5	2	0	9	9	1	5	5	1
1	3	3	3	7	4	12	8	7	5
5	2	6	1	11	1	2	4	4	2
2	1	3	9	0	10	3	3	4	5
1	5	18	4	22	8	3	0	4	4
8	9	2	3	12	1	3	1	11	9
7	5	14	7	7	28	1	3	3	4
2	11	13	2	0	1	6	12	8	12
15	0	6	7	19	1	1	9	12	4
1	5	3	17	10	15	43	2	9	11
6	1	13	13	19	10	9	20	17	24
19	2	27	5	20	5	10	8	728	8
2	3	1	1	4	3	6	13	12	12
10	9	1	1	3	9	9	4	6	3
0	3	6	3	27	3	18	4	4	7
6	0	2	2	8	4	5	1	3	1
4	18	1	0	16	20	2	2	9	3
2	12	28	0	7	3	18	12	2	1
3	2	8	3	19	12	5	4	0	3
6	0	5	0	3	7	0	8	5	8

Figura 3.1: Intervalo entre chegadas de pessoas nos balcões de atendimento.

Após o uso das ferramentas da **Estatística Descritiva**, os resultados das **Medidas de Posição** e das respectivas **Medidas de Dispersão** são apresentados na Tabela 3.1.

Medidas de posição	
Média	10,44
Mediana	5
Moda	3
Mínimo	0
Máximo	728
Medidas de dispersão	
Amplitude	728
Desvio padrão	51,42
Variância da amostra	2.643,81
Coefficiente de Variação	493%
Coeficiente Assimetria	13,80

Tabela 3.1: Medidas estatísticas dos intervalos entre chegadas.

Uma questão importante a ser avaliada é: o valor 728 é um outlier?

3.2 Outliers ou Valores Discrepantes

Na análise de dados, é importante identificar e tratar adequadamente os *outliers*. As possíveis razões para a existência de outliers incluem:

- **Erro na coleta de dados:** Este é o tipo mais comum de outlier, especialmente quando o levantamento de dados é feito manualmente.
- **Eventos raros:** Situações atípicas podem ocorrer durante a coleta de dados, resultando em valores extremos.

Alguns exemplos de eventos raros que podem gerar outliers são:

- Uma temperatura negativa registrada no verão da cidade do Rio de Janeiro.
- Um tempo de execução de um operador ser excepcionalmente curto em relação aos melhores desempenhos obtidos naquela tarefa.
- Um tempo de viagem de um caminhão de entregas na cidade de São Paulo, durante o horário de pico, ser significativamente menor do que fora deste horário.

A identificação e tratamento adequado de outliers são fundamentais para a modelagem precisa de sistemas discretos. Técnicas estatísticas, como o teste de Grubbs ou o método do intervalo interquartil, podem ser aplicadas para detectar valores discrepantes. Uma vez identificados, é necessário investigar a causa do outlier e decidir se ele deve ser mantido, ajustado ou removido do conjunto de dados, sempre considerando o impacto dessa decisão na análise global do sistema em estudo.

Teste de Grubbs

O teste de Grubbs é um método estatístico utilizado para detectar **outliers** (valores atípicos) em um conjunto de dados. Esse teste é particularmente útil quando você suspeita que um único valor em um conjunto de dados possa ser um outlier e deseja confirmar se ele deve ou não ser excluído da análise.

Quando usar o Teste de Grubbs?

O teste de Grubbs é aplicável a conjuntos de dados que seguem uma distribuição aproximadamente normal e é mais eficaz quando você tem um único outlier para testar. Ele pode ser usado tanto para detectar o menor quanto o maior valor como um outlier.

Como funciona o Teste de Grubbs?

O teste de Grubbs calcula uma estatística chamada G , que compara a diferença entre o valor suspeito de ser um outlier e a média do conjunto de dados, dividida pelo desvio padrão do conjunto de dados.

A fórmula para o cálculo da estatística G é:

$$G = \frac{|X_i - \bar{X}|}{s} \quad (3.1)$$

Onde:

- X_i é o valor suspeito de ser um outlier.
- \bar{X} é a média amostral dos dados.
- s é o desvio padrão amostral.

Passos do Teste de Grubbs

1. **Calcule a média** \bar{X} e o desvio padrão s dos dados.
2. **Calcule a estatística** G para o valor suspeito de ser um outlier usando a fórmula acima.
3. **Determine o valor crítico** G_{crit} a partir da tabela de distribuição t de Student (considerando o nível de significância desejado e o tamanho da amostra).
4. **Compare G com G_{crit} :**
 - Se G for maior que G_{crit} , o valor é considerado um outlier.
 - Se G for menor ou igual a G_{crit} , o valor não é considerado um outlier.

Exemplo

Suponha que você tenha o seguinte conjunto de dados: 10, 12, 12, 13, 12, 14, 19. Suspeitamos que 19 seja um outlier.

1. **Calcule a média:** $\bar{X} = 13,14$
2. **Calcule o desvio padrão:** $s = 2,64$
3. **Calcule G para 19:**

$$G = \frac{|19 - 13,14|}{2,64} = 2,22 \quad (3.2)$$

4. **Compare com o valor crítico:** Para um nível de significância de 0,05 e 7 amostras, o valor crítico G_{crit} é aproximadamente 2,02.
5. Como $G = 2,22$ é maior que $G_{crit} = 2,02$, podemos concluir que 19 é um outlier e pode ser considerado para exclusão.

Importância do Teste de Grubbs

O teste de Grubbs é importante porque permite que estatísticos e analistas de dados tratem outliers de maneira objetiva, ajudando a manter a

integridade dos resultados estatísticos e evitando que os outliers influenciem indevidamente a análise. No entanto, deve ser aplicado com cuidado e com uma compreensão clara do contexto dos dados.

3.3 Exercício: Questionário sobre Testes de Adequência à Normalidade

Leia atentamente os seguintes artigos:

1. **"Utilização dos testes estatísticos de Kolmogorov-Smirnov e Shapiro-Wilk para verificação da normalidade para materiais de pavimentação"** de Lopes et al.
2. **"Normalidade de variáveis: métodos de verificação e comparação de alguns testes não-paramétricos por simulação"** de Torman et al.

Em seguida, responda às questões abaixo:

1. Por que é importante verificar a normalidade dos dados antes de realizar análises estatísticas?
2. Quais são os principais métodos gráficos para avaliar visualmente a normalidade de uma variável? Quais são as vantagens e desvantagens desses métodos?
3. Cite pelo menos 5 testes estatísticos de aderência à normalidade mencionados nos artigos.
4. De acordo com os resultados da simulação realizada no segundo artigo, quais testes de normalidade tiveram o melhor desempenho geral? E quais tiveram o pior?
5. Como o tamanho da amostra afeta o desempenho dos testes de normalidade, segundo os resultados apresentados?
6. Para amostras pequenas ($n \leq 10$), qual é a recomendação dos autores em relação aos testes de normalidade?

7. Por que a distribuição t-student foi a mais difícil de ser identificada como não-normal pelos testes?
8. Quais são as implicações de usar técnicas estatísticas paramétricas (como o teste t) quando os dados não seguem uma distribuição normal?
9. Como os resultados desses estudos podem influenciar a escolha de métodos estatísticos em pesquisas na área de pavimentação e transportes?
10. Baseado nos artigos, qual seria sua recomendação para verificar a normalidade de dados em um estudo prático? Justifique sua resposta.

3.4 Exercício: Análise de Sistema de Filas de Atendimento

Você recebeu um conjunto de dados simulados de um sistema de filas de atendimento para um período de 4 horas (das 09:00 às 13:00). Os dados estão no arquivo *dados_fila_atendimento.csv* que contém informações sobre o horário de chegada dos clientes, o tempo entre chegadas sucessivas e o tempo de serviço para cada cliente.

Instruções

1. Carregue os dados do arquivo CSV no ambiente de análise de sua preferência.
2. Para as variáveis "tempo entre chegadas" e "tempo de serviço":
 - (a) Calcule as medidas de posição: média, mediana, moda, mínimo e máximo.
 - (b) Calcule as medidas de dispersão: amplitude, variância, desvio padrão e coeficiente de variação.
 - (c) Crie histogramas para visualizar a distribuição.
 - (d) Crie box plots para identificar outliers.

3. Analise a aderência a distribuições teóricas:
 - (a) Para o "tempo entre chegadas", teste a distribuição exponencial.
 - (b) Para o "tempo de serviço", teste as distribuições normal, lognormal e gamma.
 - (c) Utilize testes estatísticos e gráficos (Q-Q plots) para avaliar a aderência.
4. Responda:
 - (a) Qual a melhor distribuição para o tempo entre chegadas? Justifique.
 - (b) Qual a melhor distribuição para o tempo de serviço? Justifique.
 - (c) Existem outliers? Como tratá-los?
 - (d) Qual é a taxa média de chegada (λ) e de atendimento (μ)?
 - (e) O sistema está em estado estacionário? (Compare λ e μ .)
5. Crie um gráfico de dispersão dos tempos de chegada e de serviço. Há correlação?
6. Quais recomendações para melhorar a eficiência do sistema de atendimento?

Bônus

7. Implemente uma simulação do sistema de filas e compare com os dados originais.
8. Calcule o tempo médio no sistema e o número médio de clientes na fila.

Referências Bibliográficas

- [1] Mauricio Pereira dos Santos. *Introdução a simulação discreta*. Rio de Janeiro, 1 edition, 1999.
- [2] Tamio Shimizu. *Simulação em computador digital*. Edgard blucher Ltda, São Paulo, 1 edition, 1975.
- [3] Hamdy Taha. *Pesquisa operacional: uma visão geral*. Pearson Education do Brasil, São Paulo, 8 edition, 2008.