

**UNIVERSIDADE FEDERAL DO PARANÁ  
SETOR DE CIÊNCIAS SOCIAIS APLICADAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM  
CIÊNCIA, GESTÃO E TECNOLOGIA DA INFORMAÇÃO**

**RUBENS JOSÉ ROSA**

**AVALIAÇÃO DA DISTRIBUIÇÃO *BOOTSTRAP* NA ANÁLISE  
DOS RISCOS EM CRONOGRAMAS DE PROJETOS**

**CURITIBA  
2017**

**RUBENS JOSÉ ROSA**

**AVALIAÇÃO DA DISTRIBUIÇÃO *BOOTSTRAP* NA ANÁLISE  
DOS RISCOS EM CRONOGRAMAS DE PROJETOS**

Plano de dissertação apresentado ao Programa de Pós-graduação em Ciência, Gestão e Tecnologia da Informação do Setor de Ciências Sociais Aplicadas da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Ciência, Gestão e Tecnologia da Informação.

Orientador: Prof. Dr. José Simão de Paula Pinto

**CURITIBA  
2017**

UNIVERSIDADE FEDERAL DO PARANÁ. SISTEMA DE BIBLIOTECAS.  
CATALOGAÇÃO NA FONTE

Rosa, Rubens José

Avaliação da distribuição *bootstrap* na análise dos riscos em cronogramas de projetos / Rubens José Rosa. - 2017.  
138 f.

Orientador: José Simão de Paula Pinto.

Dissertação (Mestrado) – Universidade Federal do Paraná.  
Programa de Pós- Graduação em Ciência, Gestão e Tecnologia da Informação, do Setor de Ciências Sociais Aplicadas.

Defesa: Curitiba, 2017

1. Administração de projetos. 2. Bootstrap (Estatística). 3. Avaliação de riscos. I. Pinto, Jose Simão de Paula, 1963-. II. Universidade Federal do Paraná. Setor de Ciências Sociais Aplicadas. Programa de Pós- Graduação em Ciência, Gestão e Tecnologia da Informação. III. Título.

CDD 658.4



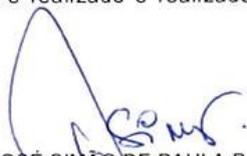
MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DO PARANÁ  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
Setor CIÊNCIAS SOCIAIS APLICADAS  
Programa de Pós-Graduação CIÊNCIA, GESTÃO E TECNOLOGIA DA INFORMAÇÃO

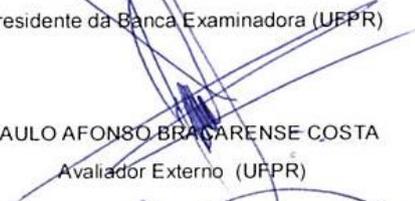
## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIA, GESTÃO E TECNOLOGIA DA INFORMAÇÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **RUBENS JOSÉ ROSA** intitulada: **Avaliação da distribuição bootstrap na análise dos riscos de cronogramas de projeto**, após terem inquirido o aluno e realizado e realizado a avaliação do trabalho, são de parecer pela sua

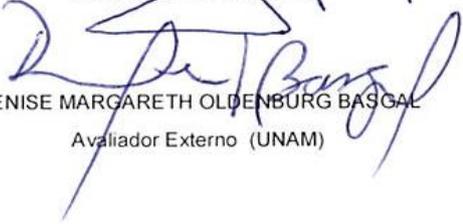
APROVAÇÃO

CURITIBA, 13 de Fevereiro de 2017.

  
JOSÉ SIMÃO DE PAULA PINTO  
Presidente da Banca Examinadora (UFPR)

  
PAULO AFONSO BRACARENSE COSTA  
Avaliador Externo (UFPR)

  
EGON WALTER WILDAUER  
Avaliador Externo (UFPR)

  
DENISE MARGARETH OLDENBURG BASGAL  
Avaliador Externo (UNAM)

Dedico este trabalho  
à minha querida esposa Anai,  
pelo companheirismo e incentivo.  
E aos meus filhos, Braian e Rubiana,  
os quais eu amo incondicionalmente.

## **AGRADECIMENTOS**

Agradeço à minha esposa, Anai Rodrigues Rosa, por estar sempre ao meu lado e me dar seu total apoio durante o mestrado, sem o qual eu não teria conseguido concluí-lo.

Aos meus filhos, Braian Lucas Machado dos Santos e Rubiana Rodrigues Rosa, pela compreensão e tolerância durante minha ausência no período do mestrado.

Ao Professor José Simão de Paula Pinto, pela orientação e paciência na confecção desse trabalho.

À professora Denise Margareth Oldenburg Basgal e professores Paulo Afonso Bracarense Costa e Egon Walter Wildauer, por terem participado da banca de avaliação deste trabalho e pelas valiosas contribuições dadas no momento da qualificação.

Ao Programa de Pós-graduação em Ciência, Gestão e Tecnologia da Informação da Universidade Federal do Paraná, representado pela professora Denise Fukumi Tsunoda, e agradeço também aos professores Edson Ronaldo Guarido Filho, Helena de Fátima Nunes Silva, Deborah Ribeiro Carvalho.

Resumindo, agradeço aos professores das disciplinas que cursei: Egon Wildauer, José Simão, Edson Guarido, Helena Silva, Deborah Carvalho, Denise Tsunoda, pois foram muito importantes para mim neste mestrado e, principalmente, para o aprendizado obtido, que levarei para vida toda.

Agradeço ao meu superintendente, André Luís de Castro David, por ter acreditado que a união Universidade-Empresa traz bons resultados a ambos os segmentos, e me liberou para realizar este mestrado.

Agradeço aos meus colegas de trabalho, que apoiaram meus estudos e, principalmente, à Marini Lucia Farinon, pela revisão e discussões sobre como aliar a teoria à prática, e à Esther Ribeiro Wowk, pela revisão desta dissertação na fase qualificação.

"A qualidade nunca é acidental;  
é sempre o resultado de um  
esforço inteligente."

John Ruskin

## RESUMO

A análise de risco em cronogramas é um fator crítico para o sucesso do planejamento e execução de projetos. Uma das formas de realizá-la é por meio da simulação Monte Carlo, atribuindo distribuições de probabilidades à duração das atividades do cronograma e simulando a execução do projeto várias vezes. O problema é que isso aumenta a carga de trabalho para o gerente do projeto e o obriga a ter conhecimentos em estatística. Para buscar resolver esse problema, esta pesquisa foi elaborada para demonstrar que a distribuição *bootstrap* pode ser utilizada na análise dos riscos em cronogramas de projetos, porque não seria necessário definir outras distribuições de probabilidades nem estimar seus parâmetros para executar a simulação Monte Carlo. Este estudo foi realizado por meio da análise de correlação entre a distribuição *bootstrap* e a distribuição triangular em uma amostra de 4.100 projetos fictícios nas mais diversas estruturas. Os resultados mostram que existe correlação entre as duas distribuições e chegou-se à conclusão de que a distribuição *bootstrap* (*bootstrap-rate*) pode ser utilizada na análise dos riscos em cronograma de projetos.

Palavras-chave: *bootstrap-rate*; simulação Monte Carlo; análise de risco; cronograma; gerenciamento de projetos; *bootstrap*.

## **ABSTRACT**

Risk analysis in schedules is a critical factor for the success of project planning and execution. One way to do this is using the Monte Carlo simulation, assigning probability distributions to the duration of the schedule activities and simulating the execution of the project several times. The problem is that this increases the workload for the project manager and compels him to have knowledge in statistics. In order to solve this problem, this research was designed to demonstrate that the bootstrap distribution can be used in risk analysis in project schedules, because it would not be necessary to define other probability distributions nor to estimate its parameters for Monte Carlo simulation. This study was carried out through a correlation analysis between the bootstrap distribution and the triangular distribution in a sample of 4,100 fictitious projects in the most diverse structures. The results show that there is a correlation between the two distributions and it was concluded that the bootstrap-rate distribution can be used in the analysis of risks in a project schedule.

Keywords: bootstrap-rate; Monte Carlo simulation; schedule; risk analysis; project management; bootstrap.

## LISTA DE FIGURAS

Figura 1 – Estratégias de pesquisa utilizadas .....	20
Figura 2 – Modelo do Corpo de Conhecimento em Gestão da Informação .....	27
Figura 3 – As bases da gestão da informação para negócios .....	30
Figura 4 – Ciclo de vida da informação .....	33
Figura 5 – Ciclo de vida: dos dados x de desenvolvimento de sistemas.....	34
Figura 6 – Processos do gerenciamento do tempo do projeto .....	37
Figura 7 – Representação do relacionamento entre as atividades $i$ e $j$ .....	39
Figura 8 – Rede para o projeto exemplo P1, baseado na Tabela 1 .....	40
Figura 9 – Rede para o projeto exemplo P2, baseado na Tabela 2.....	41
Figura 10 – Níveis progressivo e regressivo do projeto exemplo P1.....	48
Figura 11 – Níveis progressivo e regressivo do projeto exemplo P2.....	49
Figura 12 – Taxonomia da Ignorância .....	54
Figura 13 – Custo de mudanças x riscos e incerteza.....	67
Figura 14 – Processos do gerenciamento dos riscos do projeto .....	68
Figura 15 – Perspectivas da identificação de riscos.....	70
Figura 16 – Passos da análise de riscos dos cronogramas .....	71
Figura 17 – Indicadores topológicos e índices de sensibilidades.....	81
Figura 18 – Variáveis do estudo .....	82
Figura 19 – Diagrama de rede do projeto P0210 .....	99
Figura 20 – Gráfico de Gantt do projeto P0210.....	100
Figura 21 – Distribuição da duração total do projeto P0210.....	101
Figura 22 – Índice de significância das atividades do projeto P0210 .....	102
Figura 23 – Gráfico dos índices de sensibilidade para as atividades .....	104
Figura 24 – Proporção de igualdade entre $DTP \cong DTPs$ .....	105
Figura 25 – Percentual de igualdade das distribuições da $DTPs$ .....	107
Figura 26 – Percentual de projetos em relação ao valor-p da correlação entre os índices de sensibilidade .....	108
Figura 27 – Índice de correlação para os índices de sensibilidade do cronograma	109
Figura 28 – Esquema geral da simulação <i>bootstrap</i> .....	122
Figura 29 – Resultado da busca nas bases <i>Science Direct</i> e <i>Web of Science</i> .....	127
Figura 30 – Quantidade de artigos publicados por ano.....	128
Figura 31 – Relacionamento das palavras-chaves com periódicos em comum .....	128

## LISTA DE QUADROS

Quadro 1 - Relação de artigos por base de dados.....	21
Quadro 2 - Necessidade de informações de alguns grupos de pessoas .....	31
Quadro 3 – Funções do gerenciamento de dados .....	34
Quadro 4 – Características topológicas básicas da rede do cronograma .....	47
Quadro 5 – Características topológicas básicas dos projetos P1 e P2 .....	47
Quadro 6 - Formulações para o nível progressivo e nível regressivo .....	48
Quadro 7 – Nível progressivo e regressivo para o projeto P1 .....	49
Quadro 8 – Nível progressivo e regressivo para o projeto P2 .....	49
Quadro 9 – Nome dos indicadores topológicos em português e em inglês.....	50
Quadro 10 – Relação entre níveis de mensurações .....	58
Quadro 11 – Vantagens e desvantagens da utilização da simulação Monte Carlo ...	60
Quadro 12 – Resumo da distribuição binomial.....	64
Quadro 13 – Resumo da distribuição uniforme .....	64
Quadro 14 – Resumo da distribuição normal .....	65
Quadro 15 – Resumo das distribuições triangular e PERT .....	66
Quadro 16 – Diferenças entre a análise qualitativa e quantitativa dos riscos .....	70
Quadro 17 – Índices de sensibilidade em português e em inglês .....	72
Quadro 18 – Definição dos grupos de projetos .....	81
Quadro 19 – Especificação do Modelo de Simulação de Cronogramas de Projetos	93
Quadro 20 – Valores utilizados na distribuição triangular.....	94
Quadro 21 – Sorteio das durações das atividades.....	94
Quadro 22 – Etapas do protocolo da pesquisa .....	95
Quadro 23 – Coleta dos dados de testes .....	96
Quadro 24 – Índices de significância de ordenados por importância .....	103
Quadro 25 - Análise de correlação entre as distribuições .....	106
Quadro 26 – Perguntas a serem respondidas.....	121
Quadro 27 - Metadados das variáveis coletadas por periódicos.....	125
Quadro 28 – Relação de termos e seus periódicos.....	129
Quadro 29 – Comentários sobre alguns artigos .....	131
Quadro 30 – Programas de computador desenvolvidos nesta pesquisa .....	135
Quadro 31 – Algoritmo para ordenação topológica .....	136
Quadro 32 – Cálculo da ida da rede do projeto.....	137
Quadro 33 – Cálculo da volta da rede do projeto .....	138

## LISTA DE TABELAS

Tabela 1 – Lista de atividades com sucessoras imediatas para o projeto P1.....	40
Tabela 2 - Lista de atividades com sucessoras imediatas para o projeto P2 .....	40
Tabela 3 – Lista de atividades com durações para o projeto P1 .....	42
Tabela 4 – Lista de atividades com durações para o projeto P2 .....	42
Tabela 5 – Caminhos para o projeto P1 .....	43
Tabela 6 – Caminhos para o projeto P2 .....	43
Tabela 7 – Cálculo da folga para o projeto P1 .....	45
Tabela 8 – Cálculo da folga para o projeto P2.....	46

## LISTA DE SIGLAS

B-1	– Distribuição <i>bootstrap</i> padrão
BCa	– <i>Bias corrected accelerated</i>
BIM	– <i>Business Information Management</i>
B-R	– Distribuição <i>bootstrap-rate</i>
CA	– Indicador de comprimento dos arcos
CPM	– Método do caminho crítico
DA	– Indicador de distribuição das atividades
DMBOK	– <i>DAMA Guide to the Data Management Body of Knowledge</i>
DTP	– Duração total do projeto
FT	– Indicador de folga topológica
GIE	– Gestão da informação empresarial
IC	– Índice de criticidade
ICR	– Índice de crucialidade
IMBOK	– <i>Information Management Body of Knowledge</i>
IS	– Índice de significância
ISC	– Índice de sensibilidade do cronograma
K-S	– Teste de Kolmogorov-Smirnov
MDP	– Método do Diagrama de Precedência
PDM	– <i>Precedence Diagram Method</i>
PERT	– <i>Program Evaluation and Review Technique</i>
PMI	– <i>Project Management Institute</i>
RAKID	– <i>Result, Action, Knowledge, Information, and Data</i>
SP	– Indicador sequencial/paralelo
SRA	– <i>Schedule Risk Analysis</i>
TCL	– Teorema central do limite
TF-IDF	– <i>Term frequency-inverse document frequency</i>
TI	– Tecnologia da Informação
TR	– Distribuição triangular

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>16</b>
1.1 PROBLEMA DE PESQUISA.....	17
1.2 OBJETIVOS.....	19
1.3 JUSTIFICATIVA .....	19
1.4 ESTRUTURA DA DISSERTAÇÃO .....	22
<b>2 REFERENCIAL TEÓRICO</b> .....	<b>23</b>
2.1 GESTÃO DA INFORMAÇÃO EMPRESARIAL .....	23
2.1.1 Gerenciamento de projetos .....	23
2.1.2 Corpo de conhecimento em gestão da informação.....	26
2.1.3 Gestão da informação para negócios.....	29
2.1.4 Corpo de conhecimento em gestão de dados.....	32
2.1.5 Relacionamento entre gestão de projetos e gestão da informação .....	35
2.2 GERENCIAMENTO DO TEMPO EM PROJETOS.....	36
2.2.1 Planejamento do gerenciamento do cronograma e definição das atividades .....	37
2.2.2 Sequenciamento das atividades .....	38
2.2.3 Estimação dos recursos e das durações das atividades.....	41
2.2.4 Desenvolvimento do cronograma.....	42
2.2.5 Indicadores topológicos.....	46
2.3 A RELEVÂNCIA DA TAXONOMIA DA IGNORÂNCIA NA MEDIÇÃO DA INCERTEZA .....	53
2.3.1 Níveis de mensuração.....	56
2.3.2 Simulação .....	58
2.3.3 Distribuições de probabilidades .....	62
2.4 GERENCIAMENTO DOS RISCOS.....	66
2.4.1 Processos do gerenciamento dos riscos.....	67
2.4.2 Análise quantitativa dos riscos .....	69
2.4.3 Índices de sensibilidade do tempo .....	71
2.5 DISTRIBUIÇÃO <i>BOOTSTRAP</i> .....	75
2.5.1 Caracterização da distribuição <i>bootstrap</i> .....	76
2.5.2 Fontes de variação da distribuição <i>bootstrap</i> .....	77

2.5.3	Áreas de aplicação da distribuição <i>bootstrap</i> .....	79
<b>3</b>	<b>MATERIAL E MÉTODOS.....</b>	<b>80</b>
3.1	CLASSIFICAÇÃO DA PESQUISA .....	80
3.2	POPULAÇÃO .....	80
3.3	GRUPOS DE MÉTODOS .....	82
3.4	ANÁLISE DE DADOS.....	83
3.5	TESTES ESTATÍSTICOS.....	85
3.5.1	Teorema central do limite .....	86
3.5.2	Teste de Kolmogorov-Smirnov .....	87
3.5.3	Testes de correlação .....	88
3.5.4	Protocolos de análise .....	89
3.6	ANÁLISE DE RISCOS COM A DISTRUBIÇÃO <i>BOOTSTRAP-RATE</i> .....	91
3.7	DEFINIÇÃO DO INSTRUMENTO DE COLETA DE DADOS .....	92
3.7.1	Simulação utilizando a distribuição triangular .....	93
3.7.2	Simulação utilizando a distribuição <i>bootstrap</i> .....	94
3.7.3	Coleta e tratamento dos dados .....	95
3.8	PRÉ-TESTE DO INSTRUMENTO DE COLETA DE DADOS.....	95
3.8.1	Fase um .....	96
3.8.2	Fase dois.....	97
<b>4</b>	<b>APRESENTAÇÃO E ANÁLISE DOS RESULTADOS.....</b>	<b>98</b>
4.1	APRESENTAÇÃO DE UM PROJETO .....	98
4.2	IGUALDADE DA DURAÇÃO PLANEJADA.....	105
4.3	IGUALDADE DAS DISTRIBUIÇÕES DA DURAÇÃO TOTAL DO PROJETO .....	107
4.4	CORRELAÇÃO DOS ÍNDICES DE SENSIBILIDADE .....	108
4.5	DISCUSSÃO.....	110
4.5.1	Comparação com outros trabalhos .....	110
4.5.2	Comparação entre as distribuições .....	111
<b>5</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>113</b>
5.1	LIMITAÇÕES .....	114

5.2 SUGESTÕES PARA TRABALHOS FUTUROS .....	114
5.3 CONTRIBUIÇÕES DA PESQUISA.....	115
<b>REFERÊNCIAS.....</b>	<b>117</b>
<b>APÊNDICES .....</b>	<b>120</b>
<b>APÊNDICE A - ÁREAS DE APLICAÇÃO DA DISTRIBUIÇÃO <i>BOOTSTRAP</i> .....</b>	<b>121</b>
<b>1 Introdução.....</b>	<b>121</b>
<b>2 Referencial teórico .....</b>	<b>121</b>
<b>3 Método.....</b>	<b>124</b>
<b>4 Apresentação dos resultados .....</b>	<b>126</b>
<b>5 Análise dos resultados .....</b>	<b>133</b>
<b>6 Conclusão .....</b>	<b>134</b>
<b>Referências .....</b>	<b>134</b>
<b>APÊNDICE B - ALGORITMOS DESENVOLVIDOS .....</b>	<b>135</b>

## 1 INTRODUÇÃO

As organizações sofrem pressões competitivas que as obrigam a um contínuo processo de adaptação às condições ambientais. Diante das incertezas e das mudanças, o planejamento estratégico tem importância fundamental para que elas possam aumentar sua competitividade. "O planejamento estratégico é uma técnica corporativa que proporciona a análise ambiental de uma empresa, identificando suas oportunidades, ameaças, pontos fortes e fracos para que saia de seu estado atual (missão) e chegue ao estado esperado (visão)." (VALLE et al., 2007, p. 41-42) e, para que isso ocorra, são necessárias ações que normalmente são implementadas por meio de projetos.

Os projetos são instrumentos de mudanças e inovações porque "... são a principal força motriz de muitas organizações [...] e podem ser entendidos como os esforços de transformação da sociedade" e, além disso, "envolvem mudanças – a criação de algo novo ou diferente – e possuem um início e um término" (WEBSTER JR; KNUTSON, 2009, p. 1).

O *Project Management Institute* (PMI) conceitua projeto como sendo "... um esforço temporário empreendido para criar um produto, serviço ou resultado exclusivo." (PMI, 2013, p. 5). Essa característica, de ser temporário, indica que o cronograma é um item importante, pois é por meio dele que se calcula o início e o fim do projeto.

As atividades do cronograma levam certo tempo para serem executadas e utilizam recursos das organizações. A duração total do projeto é calculada a partir do tempo de execução das atividades e da relação de precedência entre elas, chegando-se a uma data prevista de finalização. A partir dessa data, as organizações podem estabelecer contratos e prever multas em caso de descumprimento. Desse modo, o cronograma torna-se um elemento fundamental no projeto e a duração das atividades deve ser a mais realista possível.

Existem outros elementos que podem afetar um projeto, como escopo, custo, qualidade, recursos humanos, aquisições e partes interessadas e todos eles estão relacionados entre si e com o cronograma (PMI, 2013). Além disso, esses elementos devem ser executados em uma determinada ordem, durante certo período de tempo. Assim sendo, o foco desta pesquisa é no cronograma e na rede

do projeto, que justamente trata da ordem e da estimativa do período de tempo para sua execução.

### 1.1 PROBLEMA DE PESQUISA

O cronograma é um plano de quando as atividades irão começar e finalizar, caracterizando-se como um elemento para o cálculo da duração total do projeto, ou seja, é uma previsão de algo que ocorrerá no futuro. Como o futuro é incerto, o cronograma pode sofrer variações inesperadas e, em consequência disto, seu resultado não é certo. Como as empresas assumem compromissos baseados nele, há riscos associados. Os riscos e as incertezas são maiores no início do projeto e diminuem na medida em que o projeto avança (PMI, 2013). Isso indica que os riscos do cronograma devem ser avaliados desde o início do projeto para poder se preparar para eles, caso aconteçam.

Para o PMI (2013), há três formas de se avaliar os riscos de um projeto quantitativamente, as quais são:

- a) análise de sensibilidade: determina qual risco, ou neste caso, quais atividades têm maior impacto no projeto, ajudando na compreensão de como elas influenciam na duração do projeto;
- b) análise do valor monetário esperado: expressos em valores monetários, é calculado por meio da multiplicação da probabilidade de um risco pelo seu impacto;
- c) modelagem e simulação: modela o projeto por meio da atribuição de distribuição de probabilidade para a duração das atividades e simula sua execução várias vezes.

Tavares (1999) afirma que existem redes de cronogramas com diferentes graus de complexidade e que esse conhecimento pode ser uma contribuição útil para o gerenciamento do cronograma. Uma das formas de identificar o grau de complexidade é por meio da simulação, que vem tornando-se uma questão central no gerenciamento de projetos.

Existem vários métodos de análise de riscos de cronogramas (VOSE, 2008; HULETT, 2009; VANHOUCHE, 2015) e, para este último autor, eles são divididos em quatro etapas:

- a) criar a linha de base do cronograma: planejamento tradicional já feito pelo gerente do projeto;

- b) identificar as incertezas do cronograma: atribuir uma distribuição de probabilidade para a duração das atividades;
- c) simular a execução do cronograma: utilizando a técnica Monte Carlo;
- d) analisar os resultados da simulação: por meio de métodos estatísticos e métricas de qualidade.

Assim, para identificar as incertezas do cronograma do projeto, são atribuídas distribuições de probabilidades para a duração das atividades. Para isso, o gerente do projeto precisa ter conhecimentos de métodos estatísticos, englobando distribuições de probabilidades e seus parâmetros (VANHOUCKE, 2013).

Além disso, a identificação da distribuição de probabilidade e dos seus parâmetros aumenta a carga de trabalho da equipe de gerenciamento do projeto, dificultando a realização da análise de risco do cronograma e, muitas vezes, inviabilizando-a. Como exemplo, se a duração de uma atividade for caracterizada pela distribuição triangular, é necessária a identificação de três estimativas: a pessimista, a otimista e a mais provável (PMI, 2013), o que multiplica a carga de trabalho por três, pois não é só necessário uma estimativa, mas três.

Como os gestores de projetos precisam conhecer as distribuições estatísticas e saber definir seus parâmetros, Vanhoucke (2013) definiu três níveis de conhecimento em estatística para contornar essa questão: especialista, conhecimento básico e iniciante. Baseado nessa classificação, o autor orienta aos gestores de risco definir a incerteza das atividades com as seguintes opções:

- especialista: escolher quais distribuições utilizar e atribuir valores aos seus parâmetros;
- conhecimento básico: escolher a distribuição a partir de uma lista, com valores padrões para os parâmetros;
- iniciante: é definida uma distribuição padrão.

A distribuição *bootstrap* pode ser uma possibilidade para reduzir o trabalho dos gestores, visto que ela não necessita de parâmetros. Ela gera reamostras a partir de uma amostra (EFRON, 1979; MOORE et al., 2006) e, nesse caso, as estimativas das durações das atividades serão consideradas como amostra.

A partir dessas considerações, a pergunta que se pretende responder nesta pesquisa é:

- A distribuição *bootstrap* pode ser utilizada na análise de riscos em cronograma de projetos?

Para Vanhoucke (2009), o resultado da análise de riscos dos cronogramas são seis indicadores de sensibilidade das atividades: índice de criticidade, índice de significância, índice de sensibilidade do cronograma e três formas do índice de crucialidade: de Pearson, de Spearman e de Kendall. Assim, para responder à pergunta de pesquisa e para minimizar os problemas apresentados, pretende-se nesta pesquisa demonstrar a seguinte hipótese:

- A distribuição *bootstrap* pode ser utilizada na análise de riscos em cronograma de projetos.

## 1.2 OBJETIVOS

Para comprovar a hipótese de pesquisa, foi definido o objetivo geral da pesquisa:

- Demonstrar que a distribuição *bootstrap* pode ser utilizada na análise dos riscos em cronograma de projetos.

E para que esse objetivo possa ser atendido, ele foi dividido nos seguintes objetivos específicos:

- a) caracterizar a gestão da informação em gerenciamento de projetos;
- b) discutir como fazer a análise dos riscos em cronograma de projetos;
- c) avaliar formas de se fazer a análise de riscos em cronograma utilizando a distribuição *bootstrap*;
- d) comparar os resultados da análise de risco da distribuição *bootstrap* com os da distribuição triangular.

## 1.3 JUSTIFICATIVA

Essa dissertação surgiu a partir de observações em cronograma de projetos como profissional da área de gerenciamento de projetos. A partir da perspectiva de analista de projetos em um escritório de projetos corporativo percebia-se, intuitivamente, algumas “falhas” nos cronogramas de projetos. Além disso, os projetos estavam atrasando e isso causava inquietações e um questionamento: será que os atrasos nos projetos não eram causados por “falhas” na confecção dos cronogramas?

Na busca por definir formalmente essas “falhas” nos cronogramas de projetos, surgiu uma pergunta mais intrigante a partir da reflexão de todas essas

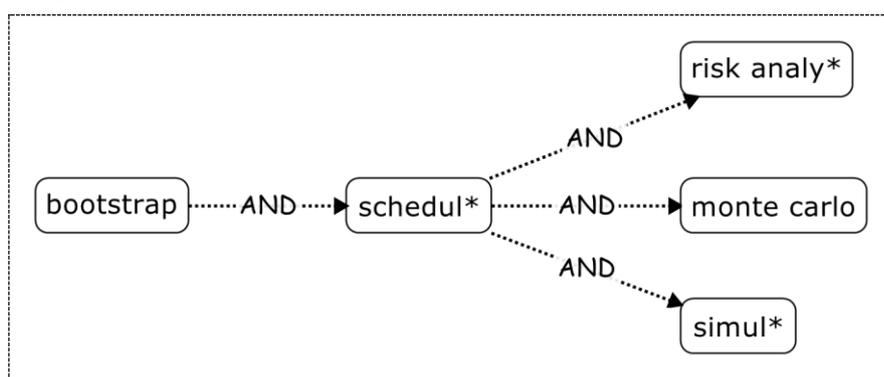
observações: que informações podem ser extraídas do cronograma dos projetos para dar suporte às decisões?

Nas leituras para a definição do problema de pesquisa (item 1.1), foi descoberto que se pode extrair informações do cronograma pela análise de risco do cronograma, o que motivou este pesquisador a explorar técnicas de análise de risco em cronogramas de projeto. Isso destaca a relevância pessoal e empresarial da pesquisa, pois a técnica estudada será aplicada na empresa em que o pesquisador trabalha.

Para verificar a relevância científica, foram feitas buscas em bases de artigos científicos. Como este trabalho trata de análise de riscos em cronogramas de projetos utilizando a distribuição *bootstrap*, as seguintes palavras foram utilizadas na definição da estratégia de busca: *bootstrap*, *schedule* ou *scheduling*, e os termos relacionados a simulação: *risk analysis*, *Monte Carlo* e *simulation*. Os asteriscos nas estratégias significam que o termo a ser buscado pode ser seguido de qualquer combinação de letras. As combinações utilizadas podem ser vistas na Figura 1, resultando em três estratégias:

- a) E1: *bootstrap* AND *schedul\** AND *risk analy\**;
- b) E2: *bootstrap* AND *schedul\** AND *monte carlo*;
- c) E3: *bootstrap* AND *schedul\** AND *simul\**;

Figura 1 – Estratégias de pesquisa utilizadas



Fonte: O autor (2017).

As buscas foram realizadas nas seguintes bases de artigos científicos, teses ou dissertações:

- B01 Ebsco;
- B02 Web Of Science;

- B03 Science Direct;
- B04 Capes Periódicos;
- B05 RI-UFPR;
- B06 BDTD - Capes;
- B07 Oásis BR.

Os resultados das execuções das estratégias de buscas podem ser vistos no Quadro 1.

Quadro 1 - Relação de artigos por base de dados

<b>Estr. \ Base</b>	<b>B01</b>	<b>B02</b>	<b>B03</b>	<b>B04</b>	<b>B05</b>	<b>B06</b>	<b>B07</b>	<b>Total</b>
E01	2	12	4	26	9	1	1	<b>55</b>
E02	1	4	1	36	12	0	0	<b>54</b>
E03	7	25	5	21	9	0	0	<b>67</b>
<b>Total</b>	<b>10</b>	<b>41</b>	<b>10</b>	<b>83</b>	<b>30</b>	<b>1</b>	<b>1</b>	<b>176</b>

Fonte: O autor (2017).

Após análises nos títulos e resumos (quando necessário) dos artigos retornados nas buscas, verificou-se que eles tratam de simulações utilizando a distribuição *bootstrap* e, separadamente, de análise de riscos, mas não da simulação *bootstrap* em cronograma de projetos e da análise de riscos no mesmo artigo.

De acordo com Bytheway (2014), o gerenciamento de projetos é um dos processos que liga a tecnologia da informação com a estratégia empresarial. Já o Programa de Pós-Graduação em Ciência, Gestão e Tecnologia da Informação trata da gestão da informação para a tomada de decisão em negócios. Apenas duas outras dissertações do programa tratam de gerenciamento de projetos, sendo que uma investiga as práticas de apoio ao compartilhamento do conhecimento interprojetos (RORIZ, 2011) e a outra analisa a produção científica sobre o papel do escritório de projetos na gestão do conhecimento (RIEDI, 2015). Como esta dissertação estuda cronograma de projetos, ela traz contribuições ao Programa, pois não se realizou pesquisa que tratasse da análise dos riscos em cronograma de projetos, o que indica a contribuição científica singular desta pesquisa.

E, para finalizar, como esta é uma pesquisa aplicada, que terá como um de seus resultados a criação de um método de análise de riscos em cronogramas de

projetos, que gestores de projetos poderão utilizar sem a necessidade de conhecimentos em estatística, ela realiza uma contribuição social ao socializar conhecimento de especialista em estatística e modelagem.

#### 1.4 ESTRUTURA DA DISSERTAÇÃO

Para cumprir os objetivos propostos para esta pesquisa, esta dissertação está estruturada em cinco capítulos. O primeiro apresenta a introdução ao tema, o problema da pesquisa, o objetivo geral, os objetivos específicos e a justificativa.

No segundo capítulo é apresentada a fundamentação teórica dos conceitos a serem trabalhados e caracteriza-se a gestão da informação em projetos, apresentando o gerenciamento do tempo em projetos. Em seguida, discute-se como se faz a análise dos riscos em cronograma de projetos, identificando-se métricas de qualidade e avalia-se a aplicabilidade da distribuição *bootstrap* na simulação Monte Carlo.

Os métodos utilizados nessa dissertação são descritos e organizados no capítulo três, que também apresenta as características e os motivos da escolha da amostra, como ela foi coletada e apresenta-se as técnicas de estatísticas para analisar os resultados das simulações.

No quarto capítulo, os dados coletados serão apresentados e analisados. Decidiu-se colocar essas duas etapas num mesmo capítulo para facilitar a análise e a visualização dos resultados.

O último capítulo traz as considerações finais, onde são apresentadas as conclusões mais significativas da pesquisa, bem como as principais contribuições. Além disso, as limitações da pesquisa são discutidas e são indicadas possibilidades de desenvolvimento de trabalhos futuros.

## 2 REFERENCIAL TEÓRICO

Este capítulo tem como objetivo definir os conceitos utilizados nesta dissertação. Inicialmente, ele apresenta a definição de gestão da informação empresarial e, na sequência, define conceitos relacionados a gerenciamento do tempo em projetos, simulação e análise de dados e gerenciamento dos riscos em projetos. Por fim, há a explanação sobre a distribuição *bootstrap*.

### 2.1 GESTÃO DA INFORMAÇÃO EMPRESARIAL

A gestão da informação empresarial (GIE) é uma área complexa, tanto que vários modelos foram criados para buscar entender essa complexidade. Em um nível intermediário, a estrutura RAKID aborda a gestão da informação partindo dos resultados que a empresa quer alcançar (MURRAY, 2002), além de existir um modelo para representar o ciclo de vida da informação (CHAFFEY; WHITE, 2012).

O modelo do Corpo de Conhecimento da Gestão da Informação (*Information Management Body of Knowledge* – IMBOK) representa a GIE em seis áreas de conhecimento e quatro processos (BYTHEWAY, 2014). O modelo de Gestão da Informação para Negócios afirma que GIE depende da gestão de três recursos empresariais (CHAFFEY; WHITE, 2012). O Corpo de Conhecimento em Gestão de Dados divide o gerenciamento de dados em dez funções (DAMA, 2010).

Relacionado com todos esses modelos de gestão da informação empresarial, está o gerenciamento de projetos, que é dividido em cinco grupos de processos e dez áreas de conhecimento e, por este motivo, será o primeiro a ser descrito.

#### 2.1.1 Gerenciamento de projetos

Com o objetivo de melhorar sua participação no mercado, tornar sua marca conhecida, aumentar seus lucros, entre outros objetivos, as empresas desenvolvem o planejamento estratégico, que é “uma técnica corporativa que proporciona a análise ambiental de uma empresa, identificando suas oportunidades, ameaças, pontos fortes e fracos para que saia de seu estado atual (missão) e chegue ao estado esperado (visão).” (VALLE et al., 2007, p. 41-42).

Como o planejamento estratégico é utilizado para que a organização evolua, com o objetivo de alcançar a visão determinada por ela, "As ações necessárias para essa transição normalmente são implementadas por meio de projetos." (VALLE et al., 2007, p. 42). Assim, pode-se concluir que projetos podem ser desenvolvidos para ajudar a corporação alcançar seus objetivos definidos no planejamento estratégico.

No entanto, se os projetos não forem bem executados, as empresas não atenderão seus objetivos estratégicos. É nesse contexto que se utiliza o gerenciamento de projetos que, de acordo com Webster Jr & Knutson (2009), "é um ramo do conhecimento que versa sobre o planejamento, monitoramento e controle de empreendimentos exclusivos" (p. 5). Complementando essa definição, o principal objetivo do gerenciamento de projetos é "viabilizar a entrega de projetos individuais que atendam às especificações de prazo, escopo, custo e qualidade acordadas com o cliente." (VALLE et al., 2007, p. 68).

Para melhor entender o que é gerenciamento de projetos é necessário antes ter a definição de projeto. "Projeto é um esforço temporário empreendido para criar um produto, serviço ou resultado único." (PMI, 2013, p. 3). Assim, sendo um "esforço temporário", tem um início e fim determinado; como produz um "resultado único", ele é inédito. Xavier et al. (2009) citam alguns exemplos de projetos:

- lançamento de um novo produto ou serviço;
- construção de uma garagem;
- desenvolvimento de um *software*;
- publicação de um livro;
- planejamento e implementação de uma mudança organizacional;
- construção de uma plataforma de petróleo.

Os exemplos apresentados pelos autores indicam a existência de uma grande diversidade de projetos, que podem variar em duração e custo, indo da construção de uma garagem até a construção de uma plataforma de petróleo. Cada projeto possui algumas características comuns, conforme Webster Jr. & Knutson (2009):

- são empreendimentos exclusivos;
- compõe-se de atividades interdependentes;
- criam entregas com qualidade;
- envolvem múltiplos recursos;
- são diferentes dos produtos que produzem;

- são regidos pelas restrições de prazo, recursos e qualidade.

Esses autores ainda sintetizam essas características ao afirmar que "projetos consistem de atividades que possuem inter-relações, produzem entregas com qualidade aprovada e envolvem múltiplos recursos." (WEBSTER JR.; KNUTSON, 2009, p. 3).

Para o PMI, gerenciamento de projetos "é a aplicação do conhecimento, habilidades, ferramentas e técnicas às atividades do projeto para atender aos seus requisitos." (PMI, 2013, p. 5), e é executado através da integração de 47 processos, agrupados em cinco grupos de processos:

- iniciação: define um novo projeto ou uma nova fase de um projeto existente;
- planejamento: definição do escopo do projeto, refinamento dos objetivos e definição de como alcançar os objetivos;
- execução: execução do trabalho conforme o planejado;
- monitoramento e controle: acompanha, analisa e controla o progresso e desempenho do projeto;
- encerramento: finaliza as atividades com o objetivo de encerrar o projeto ou uma fase.

Além dos grupos de processos, os 47 processos de gerenciamento de projetos são organizados em dez áreas de conhecimento, que: "representam um conjunto completo de conceitos, termos e atividades que compõem um campo profissional [...] ou uma área de especialização." (PMI, 2013, p. 60). A seguir é apresentada a lista das 10 áreas e o objetivo de cada uma delas, conforme descrita em PMI (2013):

- Gerenciamento da integração: "identificar, definir, combinar, unificar e coordenar os vários processos e atividades dentro dos grupos de processos de gerenciamento do projeto." (p. 63);
- Gerenciamento do escopo: "assegurar que o projeto inclui todo o trabalho necessário, e apenas o necessário, para terminar o projeto com sucesso." (p. 105);
- Gerenciamento do tempo: "gerenciar o término pontual do projeto." (p. 141);

- Gerenciamento dos custos: planejar, estimar, financiar, gerenciar e controlar os custos, "de modo que o projeto possa ser terminado dentro do orçamento aprovado." (p. 193);
- Gerenciamento da qualidade: determinar "as políticas de qualidade, os objetivos e as responsabilidades, de modo que o projeto satisfaça às necessidades para as quais foi empreendido." (p. 227);
- Gerenciamento dos recursos humanos: "inclui os processos que organizam, gerenciam e guiam a equipe do projeto." (p. 255);
- Gerenciamento das comunicações: "assegurar que as informações do projeto sejam planejadas [...] e distribuídas [...] de maneira oportuna e apropriada." (p. 287);
- Gerenciamento dos riscos: "aumentar a probabilidade e o impacto dos eventos positivos e reduzir a probabilidade e o impacto dos eventos negativos no projeto." (p. 309);
- Gerenciamento das aquisições: "comprar ou adquirir produtos, serviços ou resultados externos à equipe do projeto." (p. 355);
- Gerenciamento das partes interessadas: "identificar todas as pessoas, grupos ou organizações que podem impactar ou serem impactados pelo projeto, analisar as expectativas [...] e seu impacto no projeto, e desenvolver estratégias de gerenciamento apropriadas para o engajamento eficaz" (p. 391).

Pelo exposto, pode-se observar que os processos do gerenciamento de projeto são divididos em grupos de processos e também em área de conhecimento, sendo que cada uma trata de um determinado tema no projeto. Neste caso, o campo profissional é gerenciamento de projetos, mas, na sequência, será a gestão da informação.

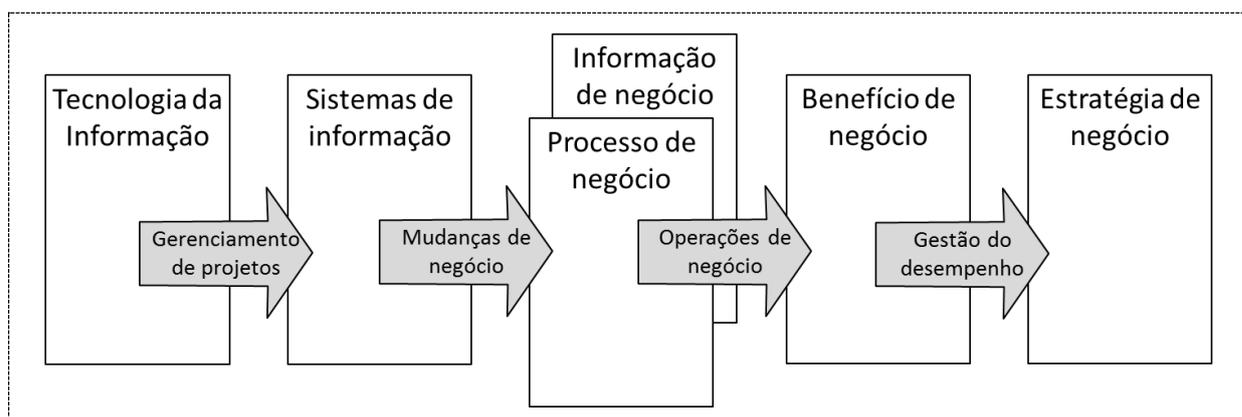
As considerações apresentadas sobre gerenciamento de projetos estão relacionadas com o próximo tópico, em que será abordado o corpo de conhecimento em gestão da informação.

### 2.1.2 Corpo de conhecimento em gestão da informação

De acordo com Bytheway (2014), na visão do Corpo de Conhecimento da Gestão da Informação (IMBOK), a gestão da informação engloba qualquer assunto relacionado à informação em negócio e é dividida em seis áreas de conhecimento e

quatro processos que fazem a transição entre as áreas de conhecimento. O modelo do IMBOK mostra a ligação entre a estratégia empresarial e a tecnologia da informação (Figura 2) e neste tópico esses componentes serão abordados a partir da visão de Bytheway (2014).

Figura 2 – Modelo do Corpo de Conhecimento em Gestão da Informação



Fonte: Adaptada de BYTHEWAY (2014).

Como mostrado na Figura 2, as áreas de conhecimento são tecnologia da informação, sistemas de informação, informação de negócio, processos de negócio, benefícios de negócio e estratégia de negócio e os processos são o gerenciamento de projetos, as mudanças de negócio, as operações de negócio e a gestão do desempenho. O IMBOK é dividido dessa forma porque representa domínios diferentes que devem ser compreendidos pelo corpo gerencial da organização, não apenas por uma pessoa, pois cada um deles tem uma função diferente. Já os processos tratam da migração de ideias de uma área para outra. Para entender o relacionamento entre as áreas de conhecimento e os processos, eles serão descritos a seguir.

A área de tecnologia da informação (TI) refere-se aos componentes tecnológicos, como, por exemplo, *hardware*, *software*, sistemas de comunicações, tudo usado para o desenvolvimento de sistemas de informações. As pessoas que trabalham nessa área devem dominar a tecnologia.

O processo de gerenciamento de projetos é responsável pelo desenvolvimento e entrega de sistemas de informação. Sem ele, a TI não seria capaz de entregar sistemas de informações, pois o gerenciamento de projetos permite que as pessoas se articulem. O foco do gerenciamento de projetos é em tarefas, marcos, entregas e orçamentos. O gerenciamento de projetos é uma

ferramenta que permite a transformação de todo o aparato tecnológico em um sistema de informações.

A próxima área de conhecimento trata dos sistemas de informações (SI). A área de TI representa todo o aparato tecnológico para entregar valor em forma de sistemas de informações. Os sistemas de informações são todos os componentes tecnológicos e humanos para a produção de sistemas e serviços que a empresa necessita, utilizando como ferramenta o gerenciamento de projetos.

Normalmente, implantar um sistema de informação envolve mudanças, então, o processo mudança de negócios trata de planejar as mudanças necessárias para que o sistema possa ser utilizado com o mínimo de resistência possível.

Uma vez que os sistemas estão desenvolvidos, eles fornecem informações para o negócio, de tal forma que os processos possam ser executados. Processo de negócio é um envelope para coordenar as atividades empresariais, com o objetivo de produzir um resultado para os envolvidos nos negócios empresariais. Os processos de negócio geram e se utilizam de um conjunto de dados, e a interpretação desses dados geram as informações do negócio que formam a base para tomada de decisão.

Operações dos negócios são as ações necessárias para que os sistemas sejam operados e os processos de negócios executados, resultando em produtos e serviços para serem entregues aos clientes da empresa.

A boa execução dos processos produz benefícios de negócio, que podem ser em forma de um sistema de informação sendo executado corretamente, a identificação de um novo mercado ou um relacionamento mais próximo com os clientes. Esses benefícios dependem da maturidade do negócio e dos objetivos que a empresa deseja atingir.

A gestão do desempenho é o processo que leva a empresa a atingir um nível de eficiência desejado. Tanto eficiência interna como externa devem ser consideradas. Como resultado, a estratégia é atingida com a gestão de desempenho de suas operações.

Os objetivos a serem alcançados são definidos pela sexta área de conhecimento, a estratégia do negócio. A estratégia trata de mudanças e isso significa saber onde a empresa está; onde poderia estar; e como atingir esses objetivos. Estratégia é uma série de ações coordenadas com objetivo de fortalecer e tornar a empresa sustentável a longo prazo.

Apesar da ordem apresentada neste texto, o modelo do IMBOK deve ser utilizado de forma invertida. Primeiro, deve-se definir a estratégia que a empresa quer adotar para então identificar seus benefícios. Após identificar os benefícios, os processos necessários para atingir esses benefícios devem ser criados e as informações necessárias identificadas. Os sistemas de informações devem prover essas informações e a tecnologia da informação deve dar suporte a todo o processo de desenvolvimento de sistemas e manutenção da infraestrutura necessária. Esta é uma abordagem dirigida pela estratégia, onde a estratégia, em uma ponta, define os recursos tecnológicos necessários para a empresa, em outra.

O IMBOK trata a gestão da informação como um modelo que é composto de seis áreas de conhecimento e quatro processos. No próximo tópico será visto o modelo de gestão da informação para negócios.

### 2.1.3 Gestão da informação para negócios

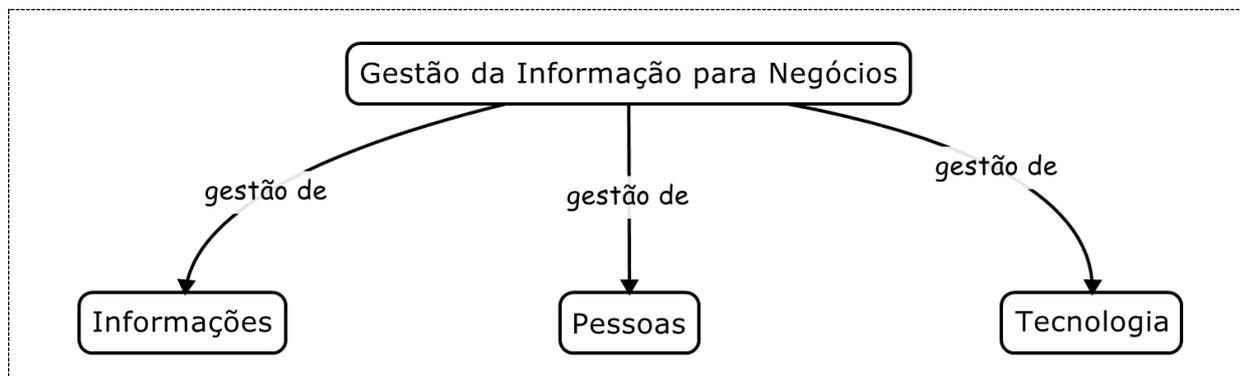
O modelo Gestão da Informação para Negócios (*Business Information Management – BIM*) considera "gestão da informação" e "gestão da informação para negócios" como sendo sinônimos e é definido como:

O processo de gerenciar a informação como um recurso estratégico para melhorar o desempenho organizacional. Esse processo desenvolve estratégias e introduz sistemas e controles para aumentar a qualidade da informação para entregar valor. (CHAFFEY; WHITE, 2012, p. 16)

A partir dessa definição, algumas conclusões podem ser tiradas: a informação é um recurso estratégico e deve ser gerenciada como tal; as informações que interessam às organizações são as que têm valor para a estratégia da empresa; os sistemas de informação são desenvolvidos para aumentar a qualidade da informação.

O modelo BIM prega a gestão eficiente de três tipos de recursos: informações, pessoas e tecnologia, como pode ser visto na Figura 3. Isso significa que os recursos não podem ser pensados isoladamente, deve haver um equilíbrio entre a gestão dos três recursos. Este tópico descreve cada um desses recursos, a partir de Chaffey & White (2012).

Figura 3 – As bases da gestão da informação para negócios



Fonte: Adaptado de CHAFFEY; WHITE (2012).

Os recursos informacionais são compostos de três elementos: dado, informação e conhecimento. Os dados são fatos discretos e objetivos sobre eventos. Por exemplo, um passageiro comprou uma passagem de Barcelona a Paris e o avião decolou dia 24 de abril, às 20:45. Eventos como esse ocorrem milhares de vezes em uma companhia aérea, gerando várias linhas de dados, mas não é possível analisá-los individualmente para tomar uma decisão.

Assim, informações são dados organizados em um contexto relevante para a tomada de decisão, ou seja, os dados são transformados em informação por meio de categorização, cálculos, correções e condensações. Essa transformação ocorre por meio dos sistemas de informações. No exemplo da companhia aérea, 80% dos voos são de Barcelona para Paris, quer dizer, daqueles milhares de dados, chegou-se a um número: 80%, que no contexto específico, indica o percentual de voos de Barcelona para Paris.

Conhecimento é a combinação de dados e informação nos quais são colocadas a opinião, as habilidades e a experiência de especialistas para a tomada de decisão. Assim, o conhecimento depende da experiência das pessoas, que, agregado com dados e informações, podem tomar decisões. Como foi visto no exemplo da empresa aérea, ter 80% dos voos em uma rota pode ser perigoso, pois se acontecer algum desastre nessa rota, a empresa pode perder uma grande fatia de clientes. Os gestores precisam tomar uma decisão: solicitar um estudo para entender porque essa concentração; fazer uma promoção para aumentar os voos em outras rotas; ou não fazer nada, porque está tudo bem. Então, com a mesma

informação, conforme a experiência dos gestores, várias ações podem ser executadas.

O próximo recurso a ser tratado são as pessoas ou os recursos humanos. Esses recursos podem ser a equipe interna ou externa, de outras organizações, como clientes, fornecedores, distribuidores, organizações governamentais e a mídia. Cada um desses grupos de pessoas tem necessidades de informação diferente, que devem ser consideradas na gestão da informação para negócio para a criação de sistemas de informações que atendam a essas necessidades. Essas pessoas são conhecidas como *knowledge workers*, "pessoas que gastam uma significativa quantidade de seu tempo em criar, utilizar e distribuir informação aplicar conhecimento para tomar decisão" (CHAFFEY; WHITE, 2012, p. 32).

O Quadro 2 mostra algumas necessidades de informação dos grupos usuários finais, gerente de departamento CEO, CIO e clientes.

Quadro 2 - Necessidade de informações de alguns grupos de pessoas

Grupo	Necessidade
Usuários finais internos	<ul style="list-style-type: none"> <li>• Qualidade dos dados;</li> <li>• Qualidade dos serviços;</li> <li>• Habilidades em desenvolvimento da TI;</li> <li>• Aplicações adequadas ao trabalho.</li> </ul>
Gestor de departamento ou processo	<ul style="list-style-type: none"> <li>• Selecionar fornecedores;</li> <li>• Selecionar aplicações;</li> <li>• Integrar com outras funções;</li> <li>• Gestão do conhecimento.</li> </ul>
<i>Chief executive officer</i> (CEO)	<ul style="list-style-type: none"> <li>• Contribuição dos sistemas de informação;</li> <li>• Impacto estratégico dos sistemas de informação;</li> <li>• Implicações legais da gestão da informação.</li> </ul>
<i>Chief information officer</i> (CIO)	<ul style="list-style-type: none"> <li>• Estratégia de investimento em sistemas de informação;</li> <li>• Gestão de mudança e minimização de riscos;</li> <li>• Terceirizar ou fazer em casa?</li> <li>• Segurança dos serviços de informações.</li> </ul>
Clientes	<ul style="list-style-type: none"> <li>• Qualidade da informação fornecida pela empresa;</li> <li>• Segurança da própria informação;</li> <li>• Integração de sistemas;</li> <li>• Suporte.</li> </ul>

Fonte: CHAFFEY; WHITE (2012).

Pelo que foi demonstrado no Quadro 2, cada grupo de pessoas possui necessidades diferentes, mas há necessidades que são compartilhadas entre as pessoas de vários grupos, como a qualidade da informação, e há necessidades de um grupo que influencia em outros, como a estratégia de investimento em sistemas de informações.

A seleção de recursos tecnológicos é fundamental para a implementação da estratégia. A seguir, algumas classes de recursos de TI:

- *softwares*: entre esses recursos estão as aplicações e os sistemas, estes com o objetivo de prover computação básica, aquele com o objetivo prover suporte aos processos de negócio;
- infraestrutura: *hardware* e redes de comunicações, responsáveis por armazenar, processar e transmitir *software* e informações para a organização;
- sistemas de informação: são utilizados para coletar, processar, armazenar, utilizar e disseminar informação, ou seja, se utilizam de tecnologia como uma ferramenta para fazer a gestão da informação.

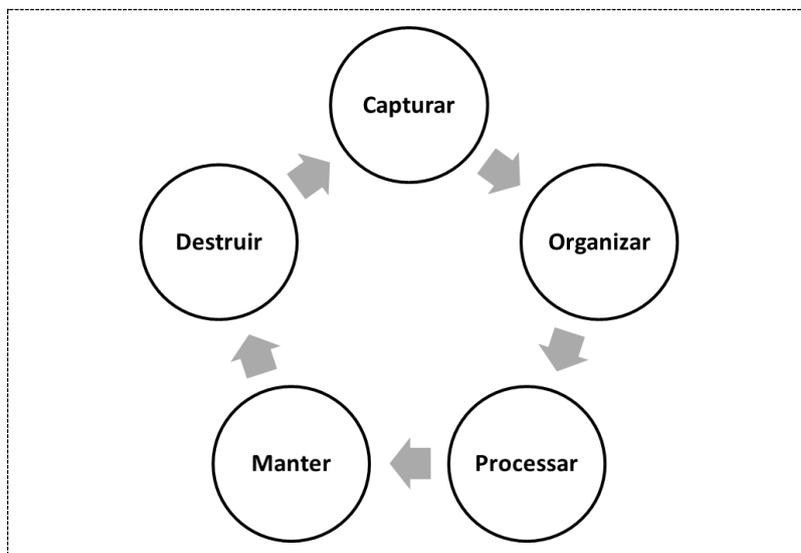
O modelo Gestão da Informação para Negócios se apoia na gestão dos recursos informacionais, humanos e tecnológicos. Assim, para atingir os resultados esperados nas organizações, deve-se fazer a gestão do ciclo de vida da informação, que é um processo dinâmico, envolvendo a coleta da informação dos processos de negócio, a organização dessa informação, o seu processamento e a sua manutenção.

O Ciclo de Vida da Informação são "as sequências de atividades envolvidas na gestão da informação, da criação até a exclusão permanente da informação" (CHAFFEY; WHITE, 2012, p. 166). Como é mostrado na Figura 4, o ciclo completo da informação é capturar, organizar, processar, manter e destruir, podendo algumas dessas atividades se repetirem e outras ocorrerem apenas uma vez.

#### 2.1.4 Corpo de conhecimento em gestão de dados

O Corpo de Conhecimento em Gestão de Dados (*DAMA Guide to the Data Management Body of Knowledge – DMBOK*) ressalta a importância de dados e informações quando afirma que "Dados e informações criadas a partir dos dados estão sendo largamente reconhecidos como um ativo empresarial" (DAMA, 2010, p. 1), mas apesar de considerar ambos importantes, o foco do guia são os dados.

Figura 4 – Ciclo de vida da informação



Fonte: CHAFFEY; WHITE (2012, p. 166).

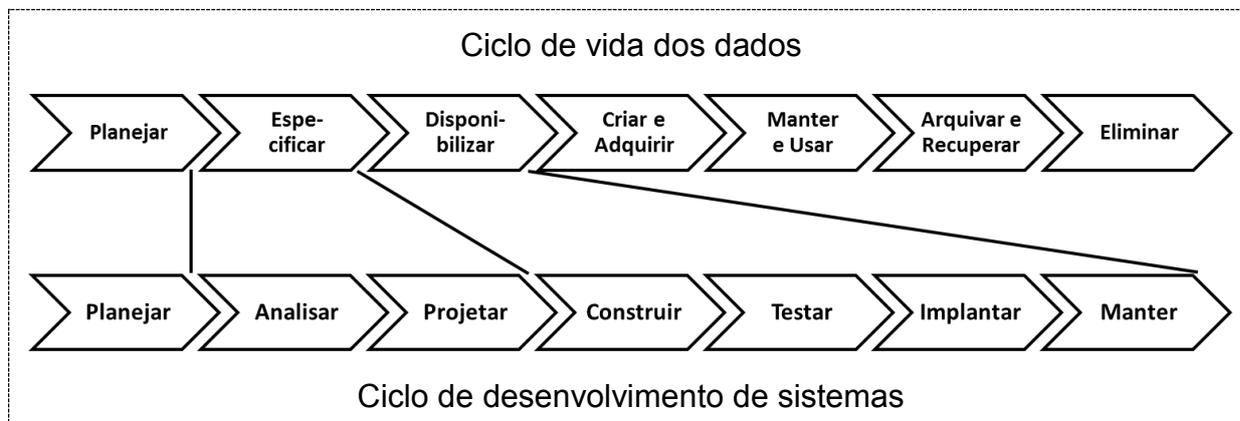
Seguem as definições fornecidas pelo guia:

- dado: representação de fatos como textos, números, gráficos, imagens, sons ou vídeos;
- informação: dado em um contexto, sendo inútil se não houver contexto;
- conhecimento: informação em uma perspectiva, integrada em um ponto de vista baseado no reconhecimento e interpretação de padrões.

Assim como a informação, o dado também tem um ciclo de vida, e além disso, também existe o ciclo de desenvolvimento de sistemas. A partir de uma necessidade corporativa, a criação de um sistema é planejada, assim como o modelo dos dados que ele tratará. Para especificar como o dado será tratado, o sistema deve ser analisado e projetado. O dado deve ser disponibilizado com a construção, teste, implantação e manutenção do sistema. A partir desta etapa, é possível utilizar o sistema para criar ou adquirir o dado, para manter e usar o dado, arquivar e recuperar e eliminar o dado.

Pode-se ver na Figura 5 que o ciclo do dado é mais extenso que o do sistema, mas, ao mesmo tempo, sendo necessário, pode-se fazer uma melhoria no sistema ou mesmo desenvolver um novo.

Figura 5 – Ciclo de vida: dos dados x de desenvolvimento de sistemas



Fonte: DAMA, (2010, p. 4).

O gerenciamento de dados é responsável por planejar, controlar e disponibilizar dados e informações e, para cumprir esses objetivos, ele é dividido em dez funções, sendo que cada função cuida de determinadas características dos dados. Um resumo dessas funções pode ser visto no Quadro 3.

Quadro 3 – Funções do gerenciamento de dados

Função	Objetivo
Governança	<ul style="list-style-type: none"> <li>Planejar, supervisionar e controlar o uso e a gestão de dados.</li> </ul>
Gestão da arquitetura	<ul style="list-style-type: none"> <li>Definir a arquitetura para o gerenciamento de dados.</li> </ul>
Desenvolvimento	<ul style="list-style-type: none"> <li>Analisar, projetar, implementar, testar, disponibilizar e manter os sistemas que cuidam do ciclo de vida dos dados.</li> </ul>
Gestão da operação	<ul style="list-style-type: none"> <li>Prover suporte desde a aquisição dos dados até a sua eliminação.</li> </ul>
Gestão da segurança	<ul style="list-style-type: none"> <li>Garantir a privacidade, confidencialidade e acesso apropriado.</li> </ul>
Gestão da qualidade	<ul style="list-style-type: none"> <li>Definir, monitorar e melhorar a qualidade dos dados.</li> </ul>
Referências e gerenciamento master	<ul style="list-style-type: none"> <li>Gerenciar e replicar dados valiosos.</li> </ul>
Gestão de <i>data warehousing</i> e <i>business intelligence</i>	<ul style="list-style-type: none"> <li>Possibilitar análise de dados e emissões de relatórios.</li> </ul>
Gestão de documento e conteúdo	<ul style="list-style-type: none"> <li>Gerenciar dados externos ao banco de dados.</li> </ul>
Gestão de meta-dados	<ul style="list-style-type: none"> <li>Integrar, controlar e prover meta-dados (dados sobre a definição de dados).</li> </ul>

Fonte: Adaptado de DAMA (2010, p. 7).

A visão do DMBOK é tratar dos dados, seu ciclo de vida e como garantir que esteja disponível. Os usuários não utilizam os dados diretamente, mas através dos

sistemas de informações, que disponibilizam dados e informações. Os projetos são executados para atender o ciclo de desenvolvimento de sistemas de informações.

#### 2.1.5 Relacionamento entre gestão de projetos e gestão da informação

Este tópico mostrará que os projetos podem ser vistos sob dois aspectos: primeiro, que eles são utilizados para o desenvolvimento de sistemas de gestão da informação; segundo, que também deve ser feita a gestão da informação em projetos.

Até este momento, neste capítulo, foi apresentada a visão da gestão da informação do IMBOK, do BIM e do DMBOK. Também foi apontado que, conforme o contexto, dados são informações ou informações são dados, sendo tratados aqui como informações. Todos os modelos analisados citaram que, para executar a gestão da informação, são necessários sistemas de informações e estes são criados por meio de projetos. Segue abaixo um resumo de como cada modelo aborda os projetos.

O IMBOK trata de projetos na produção de sistemas de informação, ou seja, a área de tecnologia da informação, por meio de projetos, produz os sistemas de informações. Além disso, para o desenvolvimento da estratégia de negócio são necessários outros tipos de projetos, não só de sistemas de informação. A estratégia de negócio, então, também depende de projetos. O gerenciamento de projetos é uma das ferramentas utilizadas na implementação da estratégia, além de auxiliar na criação de sistemas de informação, portanto, essa ferramenta é utilizada em diversos níveis do IMBOK.

O BIM trata da gestão de recursos informacionais, recursos humanos e recursos tecnológicos. Para este modelo, a área de tecnologia da informação também cria os sistemas de informações por meio de projetos; além disso, o desenvolvimento empresarial deve ser feito por meio do planejamento estratégico, que é desenvolvido por projetos. Para o BIM, os projetos também são fundamentais para a crescimento empresarial.

Para o DMBOK, os projetos também são utilizados para o desenvolvimento de sistemas de informações, para que estes possam tratar o ciclo de vida da informação (para o DMBOK, o ciclo de vida do dado).

A partir das visões dos três modelos, pode-se concluir que os projetos são importantes para a gestão da informação, pois é por meio deles que os sistemas de informações são criados para tratar do ciclo de vida da informação.

Por outro lado, os projetos também geram informações e, para que se possa utilizá-las em benefício do projeto, é necessário fazer a gestão dessas informações. Como exemplos pode-se citar o conjunto de requisitos de um sistema de informação e também a lista de atividades de um cronograma e suas durações, com as quais pode-se calcular a duração do projeto

Essas ponderações mostram, então, que os projetos são utilizados para a criação de sistemas de informações e, ao mesmo tempo, eles geram informações para serem gerenciadas, ou seja, os projetos são essenciais para a gestão da informação, ao mesmo tempo que a gestão da informação em projetos também é essencial para o sucesso do projeto.

## 2.2 GERENCIAMENTO DO TEMPO EM PROJETOS

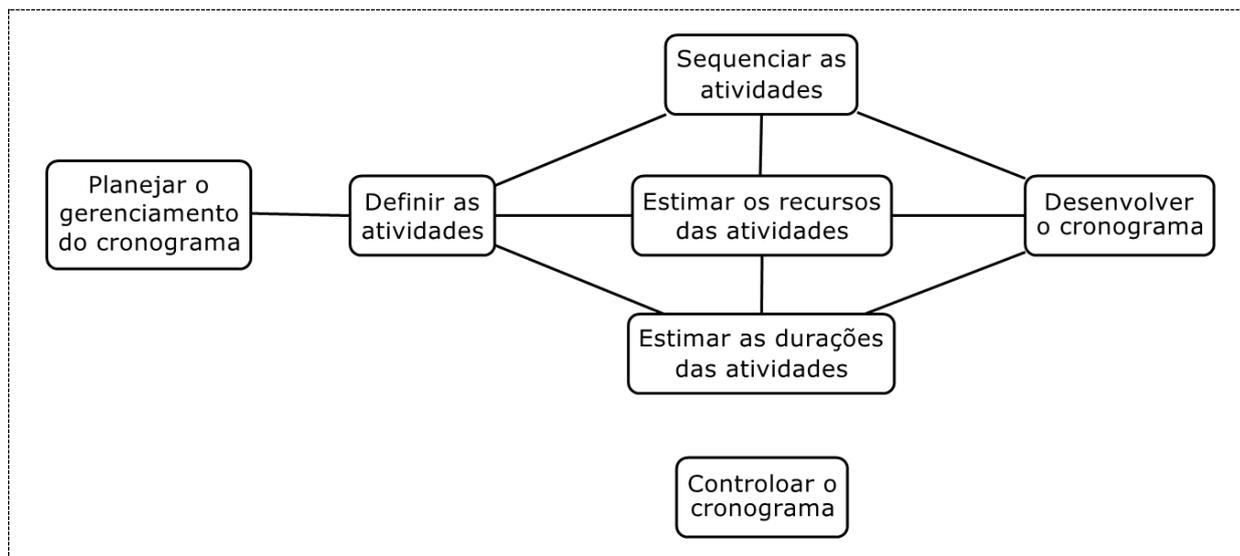
O tempo é um elemento fundamental no gerenciamento de projetos e isso pode ser confirmado a partir das definições de projetos de diversos autores. Para Valle et al. (2007, p. 28), "Um projeto é formado por um esforço, não-permanente, ou seja, temporário, para a criação de um produto ou serviço"; já para Heldman (2011), "Projetos são temporário por natureza; possuem datas de início e fim bem definidas; produzem um único produto, serviço ou resultado." (p. 3). Mulcahy & Diethelm (2011) definem projeto como "um empreendimento temporário, com um início e um fim. Cria um produto, serviço ou resultado exclusivo." (p. 19).

Apesar dessas definições serem baseados na definição do PMI (2013), "Projeto é um esforço temporário empreendido para criar um produto, serviço ou resultado único. A natureza temporária dos projetos indica que eles têm um início e um término definidos." (p. 3), nenhum dos autores citados nega que o tempo seja importante. Eles também concordam, de uma forma ou outra, que o projeto deve ter uma data de início e uma data de fim. Para chegar-se a essas datas, usam-se os processos da área de conhecimento gerenciamento do tempo em projetos.

O gerenciamento do tempo do projeto engloba um conjunto de processos para garantir a finalização do projeto no prazo estimado. Os processos realizados no gerenciamento do tempo do projeto estão no grupo de processos planejamento e no grupo monitoramento e controle. No planejamento, os processos são planejar o

gerenciamento do cronograma, definir as atividades, sequenciar as atividades, estimar os recursos das atividades, estimar as durações das atividades e desenvolver o cronograma. No grupo de processos monitoramento e controle, apenas o processo controlar o cronograma. O relacionamento entre os processos é mostrado na Figura 6.

Figura 6 – Processos do gerenciamento do tempo do projeto



Fonte: PMI (2013, p. 141).

Como todos os outros processos, os processos do gerenciamento do tempo não são sequenciais. Existem várias iterações entre os eles e, apesar de a Figura 6 mostrar o processo desenvolver o cronograma por último, ao desenvolvê-lo, pode-se encontrar alguma questão que exija alteração nos processos anteriores. Então, no planejamento do tempo do projeto, executam-se várias vezes os processos.

A partir de agora, cada um dos processos do gerenciamento do tempo do projeto será definido.

### 2.2.1 Planejamento do gerenciamento do cronograma e definição das atividades

O projeto é planejado e executado por diversas pessoas, com diversas habilidades e, muitas vezes, sem terem conhecido umas às outras. Para que todas essas pessoas tenham uma direção de como elaborar e atualizar o cronograma do projeto deve haver uma orientação. O processo planejar o gerenciamento do cronograma é responsável por "estabelecer as políticas, os procedimentos e a documentação para o planejamento, desenvolvimento, gerenciamento e execução e

controle do cronograma do projeto." (PMI, 2013, p. 145). O resultado desse processo é o plano de gerenciamento do cronograma com várias diretrizes sobre as ações feitas no cronograma do projeto.

Uma vez que o plano de gerenciamento do cronograma está pronto, deve-se identificar as atividades que compõem o cronograma. Essa tarefa é feita no processo definir as atividades, que é responsável pela "identificação e documentação das ações específicas a serem realizadas para produzir as entregas do projeto." (PMI, 2013, p. 149). As entregas do projeto serão produzidas por meio das atividades executadas ao longo do projeto, portanto, todas as atividades necessárias devem ser identificadas. O resultado desse processo é uma lista de atividades, com seus atributos como identificação, título e descrição e uma lista de marcos.

Entre os atributos das atividades está a sua duração, que será definida no processo estimar a duração das atividades. Em alguns casos, a duração da atividade é zero, e, dessa forma, elas são chamadas de marcos. "Um marco é um ponto ou evento significativo no projeto" (PMI, 2013, p. 153), que serve para indicar seu início e seu fim, ou o início e o fim de uma fase intermediária do projeto.

### 2.2.2 Sequenciamento das atividades

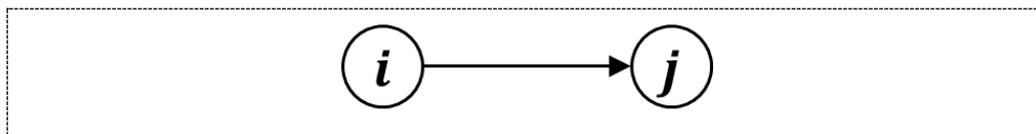
A partir do momento que há uma lista de atividades para executar o projeto, é necessário sequenciá-las em uma ordem específica com o objetivo de atender às necessidades do projeto. "Sequenciar as atividades é o processo de identificação e documentação dos relacionamentos entre as atividades do projeto." (PMI, 2013, p. 153).

O conjunto de todas as atividades sequenciadas forma o diagrama de rede do projeto e, para criá-lo, utiliza-se o método do diagrama de precedência (*Precedence Diagram Method – PDM*, em português, MDP). O MDP foi criado por John Fondahl (Stanford University), entre 1962 e 1964; e, nesse método, as atividades são representadas por nós e as relações por setas (BARCAUI et al., 2006).

Os projetos podem ser representados por um grafo  $G(N, A)$ , sendo  $N$  um conjunto de nodos, representando as atividades e  $A$  um conjunto de arcos, representando as ligações entre as atividades. Para identificar as atividades, elas

são numeradas de  $i = 1 \dots n$ , sendo que  $n$  indica a quantidade de atividades do cronograma ( $n = |N|$ ). O primeiro e o último nodo são marcos para indicar o início e o fim do projeto e possuem duração zero. Graficamente, o nodo pode ser representado por um círculo ou retângulo e o arco por uma seta ou flecha (Figura 7).

Figura 7 – Representação do relacionamento entre as atividades  $i$  e  $j$



Fonte: VANHOUCKE (2013, p. 17).

Por exemplo, entre as atividades  $i = 'Levantar a parede'$  e  $j = 'Pintar a parede'$ , existe um relacionamento lógico, ou seja, as atividades não podem ser executadas em qualquer ordem. A atividade  $i$  precisa ser executada antes da atividade  $j$ .

O conjunto de atividades predecessoras de uma atividade  $i$  será representado por  $P_i$  e o conjunto das sucessoras da atividade  $i$  será representado por  $S_i$ . O relacionamento entre as atividades  $i$  e  $j$  será sempre de término-início, indicando que, logo após o término da atividade  $i$ , será iniciada a execução da atividade  $j$ .

Para montar a rede do projeto, devem-se seguir três passos:

- a) desenhar um nodo para cada atividade da rede;
- b) desenhar um arco para cada relação de precedência direta entre duas atividades;
- c) adicionar um marco de início e um de final da rede para forçar que a rede comece e termine com um única atividade.

Para demonstrar a montagem da rede, serão utilizados dois projetos fictícios, esses dois projetos possuem a mesma quantidade de atividades e elas possuem a mesma duração, em cada um dos projetos. O que diferencia um projeto do outro é a estrutura de relacionamento entre as atividades.

O primeiro projeto, chamado de P1, possui oito atividades, sendo que a 1 e a 8 são marcos para indicar o início e o fim do projeto. A Tabela 1 mostra as atividades e suas sucessoras.

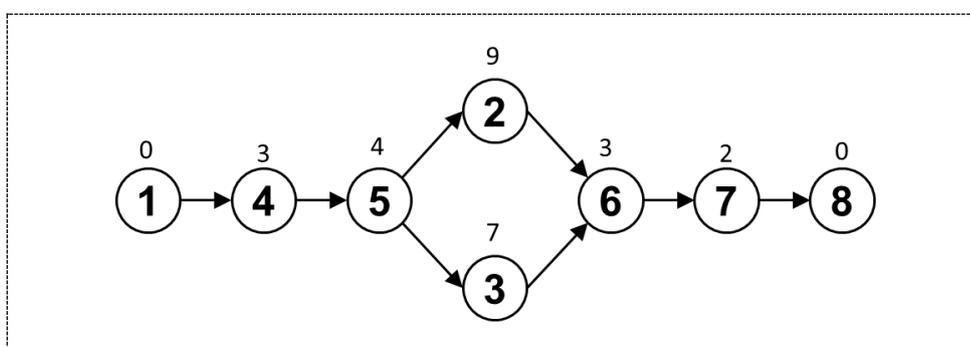
Tabela 1 – Lista de atividades com sucessoras imediatas para o projeto P1

Atividade	1	2	3	4	5	6	7	8
Sucessoras	4	6	6	5	2, 3	7	8	-

Fonte: O autor (2017).

Com base na Tabela 1, montou-se o diagrama de redes do projeto P1, que pode ser visto na Figura 8. O círculo representa as atividades e o número dentro dele, o número da atividade. O número acima do círculo é a duração em dias da atividade.

Figura 8 – Rede para o projeto exemplo P1, baseado na Tabela 1



Fonte: O autor (2017).

As definições para o projeto P2 estão na Tabela 2. A maioria das atividades possui um sucessor, as exceções são as atividades 1 e 3, que possuem três e duas, respectivamente.

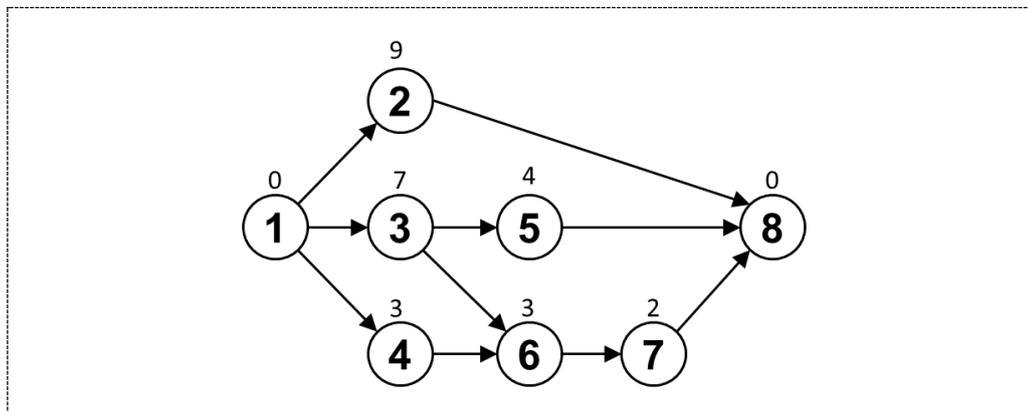
Tabela 2 - Lista de atividades com sucessoras imediatas para o projeto P2

Atividade	1	2	3	4	5	6	7	8
Sucessoras	2, 3, 4	8	5, 6	6	8	7	8	-

Fonte: O autor (2017).

A partir da Tabela 2 foi montado o diagrama de rede do projeto P2 (Figura 9). Apesar de possuírem a mesma quantidade de atividades, as estruturas dos projetos P1 e P2 são diferentes.

Figura 9 – Rede para o projeto exemplo P2, baseado na Tabela 2



Fonte: O autor (2017).

O resultado desta etapa do planejamento é o diagrama de rede do projeto, com todos os relacionamentos entre as atividades identificados. Os próximos passos são estimar os recursos e as durações das atividades e desenvolver o cronograma.

### 2.2.3 Estimação dos recursos e das durações das atividades

A partir do diagrama de rede do projeto, é preciso estimar os recursos necessários para que se possa executar as atividades. O PMI afirma que "Estimar os recursos das atividades é o processo de estimativa dos tipos e quantidades de material, pessoas, equipamentos ou suprimentos que serão necessários para realizar cada atividade." (PMI, 2013, p. 160), portanto, deve-se identificar o que é necessário para a execução das atividades e em qual quantidade.

Para se estimar os recursos necessários à execução das atividades, pode-se fazer uso da opinião especializada, analisar alternativas ou buscar dados publicados, cujo resultado será uma lista de requisitos de recursos das atividades.

Outra estimativa que deve ser feita é da duração de cada atividade. "Estimar as durações das atividades é o processo de estimativa do número de período de trabalho que serão necessários para terminar atividades específicas com os recursos estimados." (PMI, 2013, p. 165). O principal resultado dessa etapa é a quantidade tempo necessário para execução de cada atividade, que influencia diretamente no tempo do projeto.

Como exemplo, as estimativas e as atividades sucessoras para cada uma das atividades do projeto P1 são mostradas na Tabela 3.

Tabela 3 – Lista de atividades com durações para o projeto P1

<b>Atividade</b>	1	2	3	4	5	6	7	8
<b>Sucessoras</b>	4	6	6	5	2, 3	7	8	-
<b>Duração (dias)</b>	0	9	7	3	4	3	2	0

Fonte: O autor (2017).

Da mesma forma, seguem as estimativas para o projeto P2, mostradas na Tabela 4.

Tabela 4 – Lista de atividades com durações para o projeto P2

<b>Atividade</b>	1	2	3	4	5	6	7	8
<b>Sucessoras</b>	2, 3, 4	8	5, 6	6	8	7	8	-
<b>Duração (dias)</b>	0	9	7	3	4	3	2	0

Fonte: O autor (2017).

Existem diversas técnicas para estimar a duração do projeto, tais como: opinião especializada, estimativa análoga, estimativa paramétrica, estimativa três pontos, entre outras. A partir da lista de atividades sequenciadas com suas durações estimadas, pode-se desenvolver o cronograma.

#### 2.2.4 Desenvolvimento do cronograma

Um cronograma deve conter a data de início e término de cada uma das atividades, bem como a duração total do projeto. Como as atividades estão sequenciadas, cada uma delas influencia na data de início da próxima atividade. "Desenvolver o cronograma é o processo de análise de sequências das atividades, suas durações, recursos necessários e restrições do cronograma, visando criar o modelo do cronograma do projeto." (PMI, 2013, p. 172).

Para calcular a duração do projeto, é necessário calcular a duração de cada caminho da rede, e, para que o projeto seja considerado finalizado, todas as atividades devem ser completadas. Um caminho é formado por um conjunto de atividades interconectadas, do início do projeto até o fim (VANHOUCKE, 2013), ou seja, desde a primeira atividade até a última. A duração de cada caminho é a soma da duração das atividades pertencentes àquele caminho. A Tabela 5 lista todos os caminhos do projeto exemplo P1 e a Tabela 6, do P2.

Tabela 5 – Caminhos para o projeto P1

	<b>Caminho</b>	<b>Duração</b>
(1)	1 → 4 → 5 → 2 → 6 → 7 → 8	21
(2)	1 → 4 → 5 → 3 → 6 → 7 → 8	19

Fonte: O autor (2017).

O caminho com a maior duração na rede é o caminho crítico e as atividades pertencentes a esse caminho são chamadas de atividades críticas. A duração do caminho crítico determina a duração do projeto. No projeto P1, Tabela 5, o caminho crítico é o (1) e, conseqüentemente, a duração desse projeto são 21 dias. Para o projeto P2, Tabela 6, as atividades que pertencem ao caminho crítico, o (3), indicam uma duração de 12 dias para o projeto.

Tabela 6 – Caminhos para o projeto P2

	<b>Caminho</b>	<b>Duração</b>
(1)	1 → 2 → 8	9
(2)	1 → 3 → 5 → 8	11
(3)	1 → 3 → 6 → 7 → 8	12
(4)	1 → 4 → 6 → 7 → 8	8

Fonte: O autor (2017).

O que pode ser percebido, a partir do caminho crítico, é que não é a quantidade de atividades pertencentes ao caminho que determina a duração do projeto, mas a soma da duração de cada uma das atividades pertencentes ao caminho. Como exemplo, os caminhos (3) e (4) do projeto P2 possuem a mesma quantidade de atividades, mas o (3) é o caminho crítico, com duração de 12 dias, e o (4) possui duração de 8 dias, a menor duração de todos os caminhos desse projeto.

As atividades pertencentes ao caminho crítico demandam maior atenção, pois um atraso em qualquer uma delas atrasa todo o projeto (VANHOUCKE, 2013). No entanto, de acordo com Mulcahy & Diethel (2011), deve-se ter cuidado também no caminho quase crítico, que é o caminho com a duração quase igual ao caminho crítico. No projeto P1, a duração do caminho (2) difere em pouco mais de 10% do caminho crítico, portanto, um atraso com mais de 2 dias torna esse caminho o crítico. No projeto P2, a diferença entre o caminho crítico (3) e o quase crítico (2) é

menos de 10%, o que pode tornar o (2) em crítico se um eventual atraso ocorrer em alguma de suas atividades.

Em projetos reais, é impraticável gerar todos os caminhos da rede devido ao alto consumo de recursos computacionais, então, é necessário seguir um procedimento de três passos (conforme MULCAHY; DIETHELM, 2011; PMI, 2013; VANHOUCKE, 2013):

- a) calcular a data de início mais cedo da atividade  $IMC_i$ ;
- b) calcular a data de início mais tarde da atividade  $IMT_i$ ;
- c) calcular a folga total de cada atividade  $FTT_i$ .

Esses são os passos do método do caminho crítico (*Critical Path Method* – CPM) que identificam as atividades críticas do cronograma do projeto e permitem definir o início mais cedo e mais tarde das atividades, conforme será descrito a seguir.

**Data de início mais cedo da atividade  $IMC_i$ .** O cálculo da data de início mais cedo é feito através do caminho de ida do cronograma, ou seja, começa com a primeira atividade e percorre todos os caminhos do cronograma até a última atividade. Seguem as formulações para os cálculos indicados por Vanhoucke (2013, p. 27):

$$IMC_1 = 0$$

$$TMC_i = IMC_i + d_i$$

$$IMC_j = \max(TMC_i \mid i \in P_j)$$

Onde:

- $d_i$ : duração da atividade  $i$ ;
- $IMC_i$ : início mais cedo da atividade  $i$ ;
- $TMC_i$ : término mais cedo da atividade  $i$ .

Para o projeto P1, a data de início mais cedo para as atividades fica da seguinte forma:  $IMC_1 = 0$ ,  $IMC_4 = 0$ ,  $IMC_5 = 3$ ,  $IMC_2 = 7$ ,  $IMC_3 = 7$ ,  $IMC_6 = 16$ ,  $IMC_7 = 19$ ,  $IMC_8 = 21$ . A duração do projeto P1 são 21 dias, sendo a data mais cedo para ele finalizar, conforme já havia sido calculado anteriormente a partir dos caminhos demonstrados.

**Data de término mais tarde da atividade  $TMT_i$ .** O cálculo da data de término mais tarde é feito através do caminho de volta do cronograma, começa com a última atividade e percorre todos os caminhos do cronograma até a primeira. As

formulações para os cálculos estão abaixo representadas, segundo Vanhoucke (2013, p. 27), iniciando pela atividade  $n$ :

$$TMT_n = TMC_n$$

$$IMT_i = TMT_i - d_i$$

$$TMT_i = \min(IMT_j \mid j \in S_i)$$

Onde:

- $d_i$ : duração da atividade  $i$ ;
- $IMT_i$ : início mais tarde da atividade  $i$ ;
- $TMT_i$ : término mais tarde da atividade  $i$ .

Continuando com o exemplo, o projeto P1 com duração de 21 dias, a data de término mais tarde das atividades fica:  $IMT_1 = 0$ ,  $IMT_4 = 0$ ,  $IMT_5 = 3$ ,  $IMT_2 = 7$ ,  $IMT_3 = 9$ ,  $IMT_6 = 16$ ,  $IMT_7 = 19$ ,  $IMT_8 = 21$ .

Tabela 7 – Cálculo da folga para o projeto P1

Atividade	1	2	3	4	5	6	7	8
Duração (dias)	0	9	7	3	4	3	2	0
IMC	0	7	7	0	3	16	19	21
TMC	0	16	14	3	7	19	21	21
IMT	0	7	9	0	3	16	19	21
TMT	0	16	16	3	7	19	21	21
Folga	0	0	2	0	0	0	0	0

Fonte: O autor (2017).

**Folga total  $FTT_i$ .** As atividades que não estão no caminho crítico possuem certa folga para ser executada. Folga total da atividade  $i$  é a quantidade de tempo que ela pode atrasar sem atrasar todo o projeto. O cálculo da folga total é feito da seguinte forma:

$$FTT_i = IMT_i - IMC_i = TMT_i - TMC_i$$

Onde:

- $FTT_i$ : folga total da atividade  $i$ ;
- $IMT_i$ : início mais tarde da atividade  $i$ ;
- $IMC_i$ : início mais cedo da atividade  $i$ ;
- $TMT_i$ : término mais tarde da atividade  $i$ ;
- $TMC_i$ : término mais cedo da atividade  $i$ .

As atividades críticas possuem folga total igual a zero e, se qualquer uma delas atrasar, atrasa todo o projeto, portanto, elas devem ser o foco do gerente de projeto.

Na Tabela 7 e na Tabela 8 são demonstrados os cálculos desses conceitos para os projetos P1 e P2 respectivamente.

Tabela 8 – Cálculo da folga para o projeto P2

<b>Atividade</b>	1	2	3	4	5	6	7	8
<b>Duração (dias)</b>	0	9	7	3	4	3	2	0
<b>IMC</b>	0	0	0	0	7	7	10	12
<b>TMC</b>	0	9	7	3	11	10	12	12
<b>IMT</b>	0	3	0	4	8	7	10	12
<b>TMT</b>	0	12	7	7	12	10	12	12
<b>Folga</b>	0	3	0	4	1	0	0	0

Fonte: O autor (2017).

Para o projeto P1, apenas a atividade  $i = 3$  não é crítica e para o P2, não são críticas as atividades  $i \in \{2, 4, 5\}$ . O projeto P1 é menos maleável que o P2, pois naquele quase todas as atividades são críticas. Já o P2 possui mais atividades que não são críticas, portanto, maior maleabilidade nas atividades.

O cálculo do caminho crítico normalmente é feito por uma ferramenta computadorizada, mas é importante saber como ela funciona porque "os cálculos do caminho crítico ajudam o gerente do projeto a focar nas atividades essenciais, as quais atenções e recursos devem ser dedicados." (VANHOUCKE, 2013, p. 28).

No próximo tópico serão vistos os indicadores topológicos que extraem informações da rede do projeto desconsiderando a duração das atividades.

### 2.2.5 Indicadores topológicos

Os indicadores topológicos fornecem informações sobre a estrutura do cronograma. Será utilizada a nomenclatura "indicador" para se referir aos quatro tipos de indicadores. Essa nomenclatura foi utilizada para se diferenciar dos índices de sensibilidade do cronograma, que serão explicados mais adiante.

Os indicadores topológicos foram utilizados por Vanhoucke (2009) para gerar um conjunto aleatório de 4.100 projetos e serão utilizados para agrupar a amostra dos projetos desta pesquisa. Primeiro, serão demonstradas as formulações

topológicas básicas para, em seguida, explicar os indicadores topológicos. Os projetos exemplos P1 e P2, já definidos, serão utilizados para exemplificar os cálculos dos indicadores, os quais serão comparados para mostrar as diferenças entre a estrutura dos projetos.

A rede de atividades do cronograma, ou simplesmente cronograma, pode ser organizada de diversas formas, com estruturas diferentes. Os indicadores topológicos buscam analisar a estrutura de um cronograma e são denominados sequencial/paralelo, distribuição das atividades, comprimento dos arcos e folga topológica. No Quadro 4 são apresentadas algumas definições básicas.

Quadro 4 – Características topológicas básicas da rede do cronograma

Nome	Descrição
$n$	Quantidade de atividades do cronograma com duração diferente de zero
$N$	Índices das atividades do cronograma com duração diferente de zero, de 1 a $n$
$P_i$	Índices das atividades predecessoras de $i$
$S_i$	Índices das atividades sucessoras de $i$

Fonte: VANHOUCKE (2009, p. 56).

Seguindo com os projetos exemplos, P1 e P2 (Tabela 1 e Tabela 2), as definições básicas ficam da seguinte forma (Quadro 5):

Quadro 5 – Características topológicas básicas dos projetos P1 e P2

Nome	P1	P2
$n$	6	6
$N$	{2, 3, 4, 5, 6, 7}	{2, 3, 4, 5, 6, 7}
$P_i$	$P_2=\{5\}$ , $P_3=\{5\}$ , $P_4=0$ , $P_5=\{4\}$ , $P_6=\{2, 3\}$ , $P_7=\{6\}$	$P_2=P_3=P_4=0$ , $P_5=\{3\}$ , $P_6=\{3, 4\}$ , $P_7=\{6\}$
$S_i$	$S_2=\{6\}$ , $S_3=\{6\}$ , $S_4=\{5\}$ , $S_5=\{2, 3\}$ , $S_6=\{7\}$ , $S_7=0$	$S_2=0$ , $S_3=\{5, 6\}$ , $S_4=\{6\}$ , $S_5=0$ , $S_6=\{7\}$ , $S_7=0$

Fonte: VANHOUCKE (2009, p. 56).

Além dessas definições, também é necessário estabelecer para cada atividade o nível progressivo ( $NP_i$ ) e o nível regressivo ( $NR_i$ ) para calcular os indicadores topológicos.

Cada atividade do cronograma pode ter zero ou mais atividades predecessoras. Se a atividade não tiver predecessora, o seu nível progressivo é um. Se ela tiver uma ou mais predecessoras, o seu nível progressivo é o maior nível progressivo das predecessoras acrescentado de um. O nível progressivo das últimas atividades do cronograma, ou seja, aquelas que não têm sucessoras, determina maior nível progressivo do cronograma (Quadro 6).

Quadro 6 - Formulações para o nível progressivo e nível regressivo

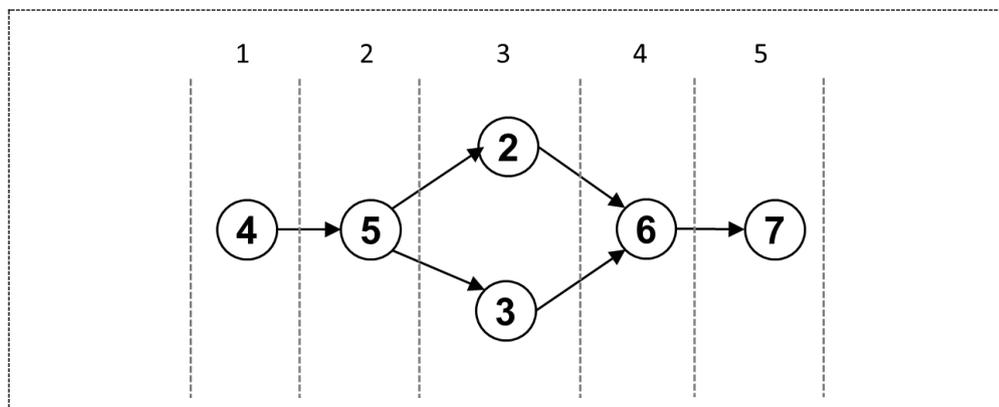
Nível progressivo	Nível regressivo
$NP_i = \begin{cases} 1 & \text{se } P_i = \emptyset \\ \max_{j \in P_i} NP_j + 1 & \text{se } P_i \neq \emptyset \end{cases}$	$NR_i = \begin{cases} m & \text{se } S_i = \emptyset \\ \min_{j \in S_i} NR_j - 1 & \text{se } S_i \neq \emptyset \end{cases}$
$m = \max_{i \in N} NP_i$ O maior nível progressivo do cronograma.	

Fonte: VANHOUCKE (2009, p. 57).

Com relação ao cálculo do nível regressivo, as últimas atividades do cronograma terão seu nível regressivo definido como o maior nível progressivo do cronograma. Se as atividades possuírem sucessoras, o seu nível regressivo será o menor nível regressivo de suas sucessoras, decrescido de uma unidade (Quadro 6).

Todo o cálculo dos níveis progressivo e regressivo é baseado no conjunto de atividades do projeto com duração maior que zero. Como exemplos, a Figura 10 mostra os níveis progressivo e regressivo do projeto P1 e a Figura 11, do P2.

Figura 10 – Níveis progressivo e regressivo do projeto exemplo P1



Fonte: O autor (2017), baseado em VANHOUCKE (2009, p. 59).

Colocando de outra maneira, o nível progressivo conta os níveis de cada atividade a partir do início, e o nível regressivo a partir do fim. Para o projeto P1, na

Figura 10, os níveis progressivo e regressivo são iguais, pois tanto contando do início como do fim o valor é o mesmo (Quadro 7).

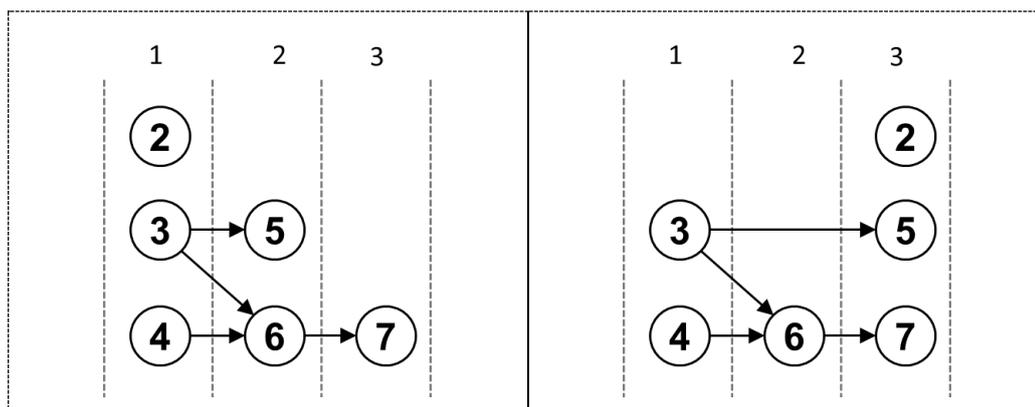
Quadro 7 – Nível progressivo e regressivo para o projeto P1

$i$	2	3	4	5	6	7
$NP_i$	3	3	1	2	4	5
$NR_i$	3	3	1	2	4	5

Fonte: Baseado em VANHOUCKE (2009, p. 60).

Já para o projeto P2, a contagem a partir do início e a partir do fim são diferentes (Figura 11), pois os vários caminhos que existem no projeto não possuem o mesmo tamanho. Então, os seus níveis progressivo e regressivo são diferentes (Quadro 8).

Figura 11 – Níveis progressivo e regressivo do projeto exemplo P2



Fonte: O autor (2017), baseado em VANHOUCKE (2009, p. 59).

Essas diferenças nos níveis progressivo e regressivo impactam no cálculo dos indicadores, principalmente no indicador de folga topológica, como será visto mais adiante.

Quadro 8 – Nível progressivo e regressivo para o projeto P2

$i$	2	3	4	5	6	7
$NP_i$	1	1	1	2	2	3
$NR_i$	3	1	1	3	2	3

Fonte: O autor (2017), baseado em VANHOUCKE (2009, p. 60).

A partir das definições dos níveis progressivo e regressivo para o cronograma, pode-se derivar os conceitos abaixo.

- largura: a largura  $w_a$  de cada nível progressivo,  $a = 1, \dots, m$ , é definida como a quantidade de atividades de cada nível;
- comprimento do arco: o comprimento  $c$  do arco  $(i, j)$  é igual a diferença entre o nível progressivo do nó final  $j$  e o nó inicial  $i$ , ou seja,  $PL_j - PL_i$ ;
- folga topológica: a folga topológica de uma atividade  $i$  é igual a diferença entre o nível regressivo e o nível progressivo da atividade  $i$ , ou seja,  $NR_i - NP_i$ .

Importante observar que Vanhoucke (2009) utiliza duas vezes o termo folga topológica: uma como mostrada aqui e, a outra, como um indicador topológico, que será definido a seguir. O autor não comenta, mas fazendo uma analogia com o método do caminho crítico, em que as atividades que fazem parte do caminho crítico são as que tem folga total igual a zero, as atividades que fazem parte do caminho topológico são as que têm folga topológica igual a zero, e isso é totalmente independente da sua duração.

Feitas as definições topológicas básicas, é possível conceituar os indicadores topológicos. Cada um dos indicadores representa um aspecto estrutural do cronograma, são eles: indicador serial/paralelo, indicador de distribuição das atividades, indicador de comprimento dos arcos e indicador de folga topológica. O Quadro 9 mostra os termos em português e inglês para comparação e, na sequência, é apresentada a definição de cada um deles.

Quadro 9 – Nome dos indicadores topológicos em português e em inglês

Termo em português	Termo em inglês
Indicador sequencial/paralelo (SP)	<i>Serial or parallel indicator (SP)</i>
Indicador de distribuição das atividades (DA)	<i>Activity distribution (AD)</i>
Indicador de comprimento dos arcos (CA)	<i>Length of arcs (LA)</i>
Indicador de folga topológica (FT)	<i>Topological float (TF)</i>

Fonte: VANHOUCKE (2009).

**Indicador sequencial/paralelo SP:** o indicador  $SP \in [0,1]$  mede o quão perto a rede está de ser sequencial ou paralela. A medida que SP se aproxima de zero, as atividades do cronograma tornam-se paralelas, até que, quando  $SP = 0$ ,

todas as atividades são paralelas. Por outro lado, a medida que SP se aproxima de um, as atividades do cronograma tornam-se sequencias, até que, quando  $SP = 1$ , o cronograma é totalmente sequencial. Segue abaixo a formulação do SP.

$$SP = \begin{cases} 1 & \text{se } n = 1 \\ \frac{m-1}{n-1} & \text{se } n > 1 \end{cases}$$

Onde:

- $m$ : número máximo de níveis progressivos;
- $n$ : quantidade de atividades com duração diferente de zero.

**Indicador de distribuição das atividades DA:** mede a distribuição das atividades nos níveis. Se  $DA = 0$ , todos os níveis contêm a mesma quantidade de atividades. Se  $DA = 1$ , um nível contém o número máximo de atividades e, os outros, apenas uma. O DA é formulado da seguinte forma:

$$DA = \begin{cases} 0 & \text{se } m \in \{1, n\} \\ \frac{\sum_{a=1}^m |w_a - \bar{w}|}{2(m-1)(\bar{w}-1)} & \text{se } m \notin \{1, n\} \end{cases}$$

Onde:

- $n$ : quantidade de atividades com duração diferente de zero;
- $m$ : número máximo de níveis progressivos;
- $w$ : quantidade de atividades em cada nível;
- $\bar{w}$ : quantidade média de atividades em cada nível, calculado da seguinte forma:  $\bar{w} = \frac{n}{m}$ .

**Indicador de comprimento dos arcos CA:** mede a relação de precedência direta da rede. Quando CA for igual a zero, as relações de precedências estão em níveis distantes, podendo a atividade mudar de nível. Se CA for igual a um, existirão várias relações com precedência direta, o que resulta em atividade com sucessoras no próximo nível da rede, e, portanto, não existindo liberdade para uma atividade mudar de nível.

Para poder definir o indicador CA, o parâmetro  $n'_l$  é definido como o número de arcos na rede com comprimento  $l$ , sendo que  $l$  pode variar entre 1 e  $m-1$ . Considerando que  $l = 1$ ,  $n'_1$  indica quantas ligações entre duas atividades de níveis adjacentes existem na rede. De acordo com esse parâmetro, o indicador CA mede quantidade de precedência direta,  $l = 1$ . A formulação para o CA fica:

$$CA = \begin{cases} 1 & \text{se } D = n - w_1 \\ \frac{n'_1 - n + w_1}{D - n + w_1} & \text{se } D > n - w_1 \end{cases}$$

O parâmetro  $D$  é o número máximo de ligações entre duas atividades de dois níveis sequenciais, formulado como  $D = \sum_{a=1}^{m-1} w_a w_{a+1}$ .

Onde:

- $n$ : quantidade de atividades com duração diferente de zero;
- $m$ : número máximo de níveis progressivos;
- $w$ : quantidade de atividades em cada nível.

**Indicador de folga topológica FT:** indica o número de níveis que cada atividade pode se deslocar sem violar o nível máximo da rede. Quando FT é igual a zero, a estrutura da rede é totalmente densa e nenhuma atividade pode deslocar de nível. Uma rede com FT igual a um consiste de uma cadeia de atividades sem folga topológica, com outras atividades completamente desvinculadas, tendo a maior folga possível. Segue abaixo a formulação para o indicador FT.

$$FT = \begin{cases} 0 & \text{se } m \in \{1, n\} \\ \frac{\sum_{i=1}^n NR_i - NP_i}{(m-1)(n-m)} & \text{se } m \notin \{1, n\} \end{cases}$$

Onde:

- $n$ : quantidade de atividades com duração diferente de zero;
- $m$ : número máximo de níveis progressivos;
- $NP_i$ : nível progressivo da atividade  $i$ ;
- $NR_i$ : nível regressivo da atividade  $i$ .

Tanto o indicador comprimento do arco (CA) como o indicador de folga topológica (FT) indicam a liberdade das atividades mudarem de nível. Ambos são calculados com base nas precedências diretas, mas a diferença entre eles é que o FT utiliza todas as relações de precedência no seu cálculo.

Cálculo dos indicadores topológicos para o projeto P1.

- $SP = \frac{5-1}{6-1} = 0,8$
- $DA = \frac{|1-1,2|+|1-1,2|+|2-1,2|+|1-1,2|+|1-1,2|}{2(5-1)(1,2-1)} = \frac{1,6}{1,6} = 1$
- $CA = \frac{6-6+1}{6-6+1} = \frac{1}{1} = 1$

$$\bullet \quad FT = \frac{(3-3)+(3-3)+(1-1)+(2-2)+(4-4)+(5-5)}{(5-1)(6-5)} = \frac{0}{4} = 0$$

Cálculos dos indicadores topológicos para o projeto P2.

$$\bullet \quad SP = \frac{3-1}{6-1} = 0,4$$

$$\bullet \quad DA = \frac{|3-2|+|2-2|+|1-2|}{2(3-1)(2-1)} = \frac{2}{4} = 0,5$$

$$\bullet \quad CA = \frac{4-6+3}{8-6+3} = \frac{1}{5} = 0,2$$

$$\bullet \quad FT = \frac{(3-1)+(1-1)+(1-1)+(3-2)+(2-2)+(3-3)}{(3-1)(6-3)} = \frac{3}{6} = 0,5$$

Onde:

- **SP**: Indicador sequencial/paralelo;
- **DA**: Indicador de distribuição das atividades;
- **CA**: Indicador de comprimento dos arcos;
- **FT**: Indicador de folga total.

Comparando os indicadores de ambos os projetos,  $SP_{P1} > SP_{P2}$ , o que indica que o P1 é mais sequencial do que o P2. Já com relação à distribuição das atividades, o P2 é mais uniforme que o P1, porque  $DA_{P2} < DA_{P1}$ . O CA do projeto P1 indica que as ligações existentes são todas em níveis adjacentes e, por isso, nenhuma atividade pode mudar de nível e isso é confirmado pelo  $FT=0$ . Já para o P2, o CA indica que as ligações existentes são entre atividades com diferença de níveis maior do que um. O FT também mostra isso.

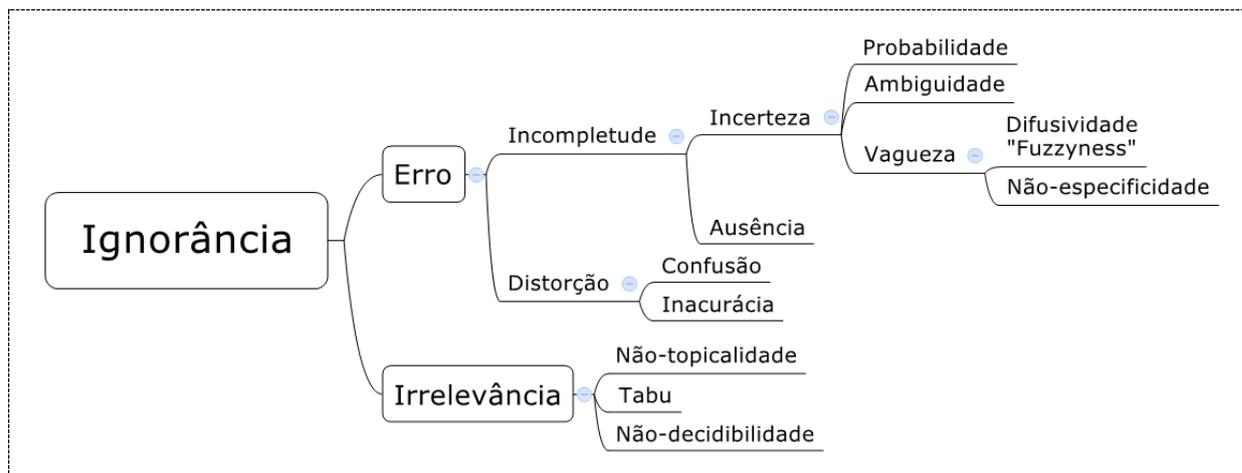
Neste tópico foram vistos os indicadores topológicos, que servem para identificação e avaliação da estrutura do cronograma. Foi visto que todos os indicadores variam entre zero e um e que cada um deles mostra um aspecto da estrutura da rede de atividades.

### 2.3 A RELEVÂNCIA DA TAXONOMIA DA IGNORÂNCIA NA MEDIÇÃO DA INCERTEZA

O objeto de estudo desta pesquisa é a análise de riscos em cronogramas de projetos, feita de diversas formas. Os riscos são analisados porque existe uma dúvida: será que o projeto finalizará no tempo previsto no cronograma? Ou seja, existe incerteza sobre o término do projeto, ignora-se a verdadeira data e ela pode variar. Este tópico explica o relacionamento entre as palavras incerteza, ignorância e variabilidade, contextualizando sua importância para esta pesquisa.

Existem várias definições sobre ignorância e Smithson (1989) discute algumas dessas definições, aponta algumas conclusões e as organiza em uma taxonomia. Esse autor coloca a ignorância como sendo a raiz da taxonomia da Ignorância (Figura 12), e afirma que a incerteza ocupa uma posição especial como uns dos tipos de ignorância mais gerenciável.

Figura 12 – Taxonomia da Ignorância



Fonte: SMITHSON (1989); BRACARENSE COSTA (1999).

A ignorância se desdobra em níveis, sendo que a principal distinção emerge a partir da distinção entre "ignorar" e "ser ignorante". O ato de ignorar é tornar determinada questão irrelevante, ao passo que ser ignorante é um estado cognitivo errôneo e o autor chamou esses dois estados de irrelevância e de erro.

A irrelevância é não dar a devida importância, ou seja, não querer saber, e isso é uma estratégia negativa. Já o erro denota um conhecimento incompleto ou distorcido, sendo uma estratégia positiva e que resulta em inclusão.

Descendo mais um nível, o erro é subdividido em distorção e incompletude. A distorção, em grau, se refere a inacuracidade ou desvio (*bias*); já em tipo, se refere a confusão. Então Smithson (1989) subdivide a distorção em confusão e inacurácia.

A incompletude é mais corrigível do que a distorção, por isso o autor a subdividiu em mais níveis. Incompletude, em tipo, será chamada de ausência, e, em grau, será incerteza. A incerteza é um dos mais gerenciáveis tipos de ignorância, subdividida em probabilidade, ambiguidade e vagueza.

Esta pesquisa está totalmente voltada para o estudo da incerteza, utilizando-se a probabilidade para a verificação das hipóteses de pesquisa. Em outro exemplo de pesquisa, Bracarense Costa (1999) estudou a incerteza dando ênfase na

vagueza, por meio do estudo da Lógica Fuzzy. Como será visto adiante e na literatura, é possível estudar a incerteza utilizando-se de vários métodos.

Outros autores diferenciam a incerteza e a variabilidade (VOSE, 2008) e apresentam outras definições para incerteza (HUBBARD, 2014; BECKER, 2015). Para esclarecer essas questões, neste tópico será comentado sobre esses conceitos e como eles são relacionados com a Taxonomia da Ignorância.

A incerteza e a variabilidade são dois elementos que dificultam a previsão assertiva sobre o futuro. Isso ocorre porque a incerteza é confundida com a variabilidade. Em um contexto em que um avaliador está modelando um sistema, incerteza é "a falta de conhecimento do avaliador sobre os parâmetros que caracterizam o sistema físico que está sendo modelado." (VOSE, 2008, p. 48). Essa definição diz que não se tem conhecimento dos parâmetros de um sistema. Ela pode ser reduzida, fazendo-se novas medições ou consultando especialistas do sistema, ou seja, depende da pessoa que está analisando e, portanto, é subjetiva (VOSE, 2008).

Outra definição sobre o mesmo termo caracteriza incerteza como falta de conhecimento, e diz que incerteza é "a existência de mais de uma possibilidade, sendo que o verdadeiro resultado não é conhecido" (HUBBARD, 2014, p. 84). Nessa definição, o autor aponta que para existir incerteza, deve haver várias possibilidades. Para que se possa medir a incerteza, "atribui-se probabilidades a cada uma dessas possibilidades" (HUBBARD, 2014, p. 84). Esse autor fala também que medir é reduzir a incerteza, baseada em observações, e não necessariamente eliminá-la completamente.

A última definição a ser vista é a de Becker (2015), que afirma categoricamente que "O ponto fundamental é que a **incerteza** é uma propriedade do **nosso conhecimento** a respeito dos eventos, e **não** dos eventos em si, todas as incertezas sendo intrinsecamente do mesmo tipo." (p. 131). Esse autor fala que as incertezas são do mesmo tipo porque não importa se a incerteza é sobre o futuro ou o passado, mas afirma que para se mensurar a incerteza, atribui-se probabilidade a ela.

Já a variabilidade está ligada ao sistema em análise e é o efeito aleatório em função das características do sistema. Em consequência disso, para poder alterá-la, o próprio sistema deve ser alterado (VOSE, 2008). Por exemplo, considerando um dado (um cubo com seus lados numerados de um a seis lados) como um sistema,

ao lançá-lo certa quantidade de vezes, é possível calcular a média e o desvio padrão desses lançamentos. Se repetir o experimento, esses valores serão aproximadamente os mesmos. Existe uma variabilidade no sistema (dado), que, para que possa ser reduzida, é necessário alterar a estrutura do sistema. A variabilidade é importante, pois, para que se possa detectar um fenômeno, é necessário que exista variabilidade (BECKER, 2015).

A incerteza e a variabilidade estão relacionadas com a Taxonomia da Ignorância de Smithson (1989). A Taxonomia da Ignorância cria uma hierarquia para definir o que é ignorância e como ela é subdividida. Ela mostra explicitamente a incerteza, sendo aderente ao que foi discutido. Porém, ela não explica o que é variabilidade, dizendo que a probabilidade é um tipo de incerteza.

Pela definição de Vose (2008), variabilidade são os possíveis valores que podem ocorrer no lançamento de um dado ou uma moeda, ou seja está associada a probabilidade, e essa também é a definição de Hubbard (2014). Já em incerteza, os parâmetros do sistema sendo modelados não são conhecidos, então não se sabe quais são os valores possíveis de ocorrer.

As definições de incerteza e variabilidade de Vose (2008) encaixam no ramo probabilidade da Taxonomia da Ignorância, porque em ambos os casos o autor utiliza um sistema probabilístico, diferenciando entre si o conhecimento ou não dos seus parâmetros.

Todo esse trabalho lida com a incerteza, se encaixando no ramo probabilidade, logo abaixo da incerteza, na Taxonomia da Ignorância de Smithson (1989).

### 2.3.1 Níveis de mensuração

Para que se possa identificar a variabilidade de um sistema, é necessário mensurá-lo. Existem diferentes formas de definir mensuração, por exemplo, "é o processo de aplicar ou atribuir números a objetos ou observações." (SIEGEL; CASTELLAN JR., 2006, p. 51), ou seja, para se medir, é necessário retirar amostras ou fazer uma estimativa.

Se "medir é reduzir a incerteza baseada em amostras" (HUBBARD, 2014, p. 31) e incerteza é a falta de conhecimento de um ou mais parâmetros (VOSE, 2008), ao medir, aumenta-se o conhecimento sobre os parâmetros, sendo importante destacar que a medição não elimina toda a incerteza, apenas aumenta o

conhecimento até um determinado limite em que uma decisão possa ser tomada (HUBBARD, 2014).

A medição não precisa ser necessariamente sobre quantidade (HUBBARD, 2014), pois existem diversos níveis de mensuração, conforme afirmou inicialmente o psicologista Stanley Smith Stevens (HUBBARD, 2014). Além disso, diversos pesquisadores (MOORE et al., 2006; HUBBARD, 2014; BECKER, 2015) identificam esses níveis de mensuração como: nominal, ordinal, intervalar e razão, os quais também são chamados de escalas.

Assim, na escala nominal, "números ou outros símbolos são usados simplesmente para classificar um objeto, pessoa ou características" (SIEGEL; CASTELLAN JR., 2006, p. 43). Neste caso, a única comparação que pode ser feita é se um objeto é igual a outro, e as únicas operações permitidas na escala nominal são a contagem, o cálculo da frequência e o da moda, que é o elemento que mais aparece. Nesta escala não são permitidas operações matemáticas, como adições e, portanto, não se pode calcular a média e o desvio padrão; conseqüentemente, a principal característica desta escala é a categorização de objetos.

A escala ordinal se caracteriza por apresentar uma ordem entre os objetos, seguindo algum critério. Então, além dos objetos pertencerem a uma categoria, essa categoria tem uma relação de ordem. Como exemplos, podem ser citados o sistema de posição em serviços militares e os graus de avaliação atribuídos em um curso. Na escala ordinal, os objetos podem ser comparados entre si, resultando em um objeto igual a outro, um maior ou um menor do que outro.

Além de possuir as mesmas características da escala ordinal, na escala intervalar "as distâncias ou diferenças entre quaisquer dois números na escala têm significado" (SIEGEL; CASTELLAN JR., 2006, p. 47). Nesta escala, o zero e a unidade de medida são atribuídos arbitrariamente e intervalos iguais correspondem diferenças iguais (BECKER, 2015, p. 13) e, além disso, todas as operações matemáticas e estatísticas podem ser efetuadas com os dados coletados.

A escala razão possui um ponto zero verdadeiro como sua origem, além de todas as características da escala intervalar: "as operações da aritmética são permitidas sobre os valores numéricos atribuídos aos objetos, assim como sobre os intervalos entre números" (SIEGEL; CASTELLAN JR., 2006, p. 50). As escalas razão somente são caracterizadas quando possuem as seguintes operações:

- a) equivalência;

- b) maior do que;
- c) razão conhecida entre dois intervalos quaisquer;
- d) razão conhecida entre dois valores quaisquer.

Com a escala razão, podem-se fazer quaisquer operações matemáticas e testes estatísticos, respeitadas as suposições do teste.

Existe uma relação hierárquica entre as escalas (Quadro 10), pois a subsequente adquire as características da anterior, sendo que a escala razão possui todas as características. Essa questão é importante porque, para cada tipo de escala, há um conjunto de determinados testes que podem ser feitos, por exemplo, dados medidos em escalas nominais ou ordinais só podem ser analisados por métodos não paramétricos; os medidos em escala intervalar ou razão podem ser analisados tanto por métodos não paramétricos como por métodos paramétricos, desde que respeitadas as suposições do método.

Quadro 10 – Relação entre níveis de mensurações

Características \ Níveis	Níveis			
	Nominal	Ordinal	Intervalar	Razão
Equivalência	XXX	XXX	XXX	XXX
Maior do que		XXX	XXX	XXX
Razão conhecida de dois intervalos quaisquer			XXX	XXX
Razão conhecida de dois valores quaisquer				XXX

Fonte: O autor (2017), baseado em SIEGEL; CASTELLAN JR. (2006, p. 52).

Este tópico tratou da definição de incerteza e variabilidade e o seu relacionamento com os níveis de mensurações. O próximo tratará como modelar um sistema e identificar as características desses sistemas por meio de simulação.

### 2.3.2 Simulação

Utiliza-se simulação para avaliar diversas possibilidades em um sistema, pois é uma ferramenta de análise de cenários (CHWIF; MEDINA, 2015), e é executada para se buscar respostas acerca do sistema real quando o sistema sob análise é complexo, dinâmico e aleatório. Assim, o modelo de simulação captura as características de um sistema real, simplificando-o, buscando repetir em um

computador suas características para avaliar a variabilidade de seus parâmetros e tomar decisões.

Como a simulação utiliza um modelo matemático para analisar o comportamento de um sistema, neste trabalho, define-se sistema conforme apresentado por Paula Filho (2009):

1. Conjunto de elementos, materiais ou ideias, entre os quais se possa encontrar ou definir alguma relação.
2. Disposição das partes ou dos elementos de um todo, coordenados entre si, e que funcionam como estrutura organizada.
3. Reunião de elementos naturais da mesma espécie, que constituem um conjunto intimamente relacionado. (PAULA FILHO, 2009, p. 4)

Se os sistemas são formados por partes coordenadas entre si, cada parte influencia no funcionamento de outras, o que significa que há uma relação de causa e efeito entre as partes. O cronograma de um projeto pode ser considerado um sistema, pois as atividades estão relacionadas entre si e influenciam-se umas às outras. Então, pode-se concluir que o cronograma de um projeto é um sistema.

Além disso, podem-se classificar as simulações em duas categorias: a computacional, que utiliza o computador para sua execução, e não computacional, que utiliza outros meios para se executar, por exemplo, a simulação de uma aeronave em um túnel de vento. A simulação de interesse neste texto é a computacional.

Há, ainda, diversas formas de se criar um modelo simulação, como a simulação de eventos discretos, a simulação contínua e a simulação Monte Carlo.

A simulação de eventos discretos e a de sistemas dinâmicos utilizam fundamentalmente da passagem do tempo, pois consideram a mudança do sistema ao longo do tempo. A simulação de eventos discretos modela a evolução de um sistema no tempo, assim, para cada elemento do sistema é definida a sua evolução e como interage com os outros, então, a passagem do tempo é simulada e o estado de cada elemento é observado (VOSE, 2008). Já a simulação contínua considera o tempo continuamente (CHWIF; MEDINA, 2015) e o comportamento do sistema ao longo do tempo.

A simulação Monte Carlo utiliza números aleatórios para simular um sistema físico ou matemático, sendo que o tempo não é uma variável explícita, pois "representa um sistema em determinado instante do tempo" (CHWIF; MEDINA, 2015, p. 185) e, justamente por isso, é conhecida também como simulação estática.

A simulação Monte Carlo é a mais adequada para esta pesquisa porque o resultado que interessa é a duração final do projeto.

A simulação Monte Carlo "envolve amostra aleatória de cada distribuição de probabilidade dentro do modelo para produzir centenas ou mesmo milhares de cenários" (VOSE, 2008, p. 45). Então, a partir de distribuições de probabilidade, ela se utiliza do computador para gerar vários cenários. Para cada cenário é calculado o resultado de uma fórmula para determinado parâmetro, como, por exemplo, a duração do projeto, e, ao final, existe uma distribuição dos valores do parâmetro em análise.

A estrutura central de uma análise de riscos é a variabilidade do sistema (VOSE, 2008). Como um projeto é um sistema, é possível utilizar-se da simulação Monte Carlo na análise de riscos de cronogramas. As vantagens e desvantagens da utilização da simulação Monte Carlo, na análise quantitativa dos riscos de projetos podem ser vistas no Quadro 11.

Quadro 11 – Vantagens e desvantagens da utilização da simulação Monte Carlo

Vantagens	Desvantagens
<ul style="list-style-type: none"> <li>• Usada principalmente para análise de cronograma e de custo em projetos, para dar suporte às decisões estratégicas;</li> <li>• Permite que todos os riscos variem simultaneamente;</li> <li>• Calcula as estimativas quantitativas dos riscos gerais do projeto;</li> <li>• Provê respostas para perguntas do tipo (1) Qual a probabilidade de sucesso da linha de base do plano? (2) Quanto de contingência no tempo e no custo é necessário para atingir determinado nível de confiança? (3) Quais são as atividades mais importantes em determinar o risco geral do projeto?</li> </ul>	<ul style="list-style-type: none"> <li>• Cronogramas não são simples e frequentemente não podem ser usados em simulação, sem um envolvimento significativo dos especialistas em cronogramas;</li> <li>• A qualidade da entrada dos dados depende fortemente do julgamento dos especialistas e do esforço e expertise do analista de risco;</li> <li>• A simulação é geralmente rejeitada pela gestão como não sendo necessária ou muito sofisticada comparada com outras ferramentas tradicionais de gerenciamento;</li> <li>• A simulação Monte Carlo requer <i>software</i> especializado que deve ser adquirido e aprendido, causando barreira em seu uso;</li> <li>• Para produzir resultados realísticos, os dados de entrada devem considerar as ameaças e oportunidades.</li> </ul>

Fonte: PMI (2009, p. 93)

Quando se tem a estimativa de duração das atividades de um projeto e a duração do próprio projeto, não existe variabilidade, pois essa é uma estimativa pontual. Quando se executa uma simulação utilizando-se uma distribuição de probabilidade, têm-se diversas durações para as atividades e para o projeto.

Existem variabilidades nos dados e, neste caso, é possível detectar várias informações do projeto, como a probabilidade de ser finalizado na duração estimada, calcular a duração necessária para se ter 80% de probabilidade de cumprir o prazo e identificar as atividades que mais influenciam na duração do projeto, tal como foi mostrado no Quadro 11.

Se, como foi definido anteriormente, medir é reduzir a incerteza, é preciso haver uma linha de base para comparação, ou seja, responder à pergunta "Quanto você sabe agora?" (HUBBARD, 2014, p. 93). Essa questão é respondida pelas estimativas das durações das atividades e o cálculo da duração do projeto. Essa duração significa o quanto é conhecido inicialmente no projeto. Medir esse valor é simular o cronograma para ver a probabilidade de se finalizar o cronograma na duração estimada, e, dessa forma, reduzir a incerteza acerca do cronograma.

Para Chwif & Medina (2015), o resultado de todo esse processo é um modelo probabilístico utilizado para representar a aleatoriedade do sistema. Assim, ao se construir um modelo, deve-se criar uma abstração do sistema real, ou seja, deve-se "identificar o que realmente é importante no sistema e trazer para o modelo." (p. 43). O modelo abstrato imaginado pelos especialistas deve ser colocado no papel, utilizando-se técnicas de representação de modelos. O modelo de simulação pode ser especificado por um documento que contenha os seguintes tópicos:

- a) objetivos: indicam quais são os objetivos que a simulação quer alcançar;
- b) conteúdo: indica o escopo do modelo e o nível de detalhamento;
- c) entradas: indicam os parâmetros necessários para a simulação;
- d) saídas: indicam as medidas de desempenho que estão sendo avaliadas, de acordo com os objetivos;
- e) hipóteses e simplificações do modelo: hipóteses indicam uma simplificação do sistema e todas devem ser listadas;
- f) definições de iterações: indicam quantos cenários serão executados na simulação, sendo que cada cenário é um conjunto fixo de parâmetros escolhido aleatoriamente.

Uma vez criado o modelo conceitual, é necessário implementá-lo utilizando-se o computador. A implementação computacional pode ser feita de três formas: utilizando uma linguagem de simulação, um simulador ou uma linguagem de programação (CHWIF; MEDINA, 2015).

Nesta dissertação, a implementação computacional foi feita utilizando a linguagem de programação Java. Foi também criado um simulador de cronogramas que opera tendo como entrada um cronograma de projetos com as atividades, suas durações e as relações de dependências. A saída do simulador é um conjunto de estatísticas acerca do cronograma do projeto e suas atividades.

O modelo de uma simulação depende da definição da aleatoriedade do sistema, que são definidas por meio da distribuição de probabilidade dos seus componentes. Algumas distribuições de probabilidades são descritas no próximo tópico.

### 2.3.3 Distribuições de probabilidades

Foi visto anteriormente que medir é atribuir probabilidades a diversos resultados possíveis (HUBBARD, 2014). Esses resultados possíveis poderiam ser a duração média dos projetos ou a categoria em que os projetos estão em determinado momento (adiantado, no prazo ou atrasado). "A distribuição de probabilidade de uma variável aleatória  $X$  nos informa quais valores a variável  $X$  pode assumir e como atribuir probabilidades a esses valores." (MOORE et al., 2006, p. 217), ou seja, existe uma forma de distribuir as probabilidades aos diversos resultados possíveis.

A distribuição de probabilidade busca uma aproximação da distribuição amostral, que é "aquela distribuição que teríamos se tomássemos todas as possíveis amostras de mesmo tamanho da mesma população, extraíndo cada uma aleatoriamente." (SIEGEL; CASTELLAN JR., 2006, p. 32). A diferença é que a distribuição de probabilidade é baseada em fórmulas e a amostral é uma distribuição teórica.

Por exemplo, se a variável de interesse na população fosse normalmente distribuída e fossem extraídas todas as possíveis amostras dessa população, a distribuição amostral também seria normal e possuiria a mesma média. A distribuição de probabilidade normal é uma formulação matemática para expressar o comportamento de uma população e é caracterizada pela sua média e seu desvio padrão. Portanto, uma distribuição de probabilidade é uma formulação matemática sobre o comportamento de uma população ou amostra.

A probabilidade é um valor entre zero e um que pode ser distribuída de diversas formas: todas as probabilidades podem ser iguais; as probabilidades podem ser maiores no centro, e à medida que se afastam do centro, diminuem simetricamente; ou podem ser maiores para determinado valor e diminuir à medida que se afastam desse valor, não simetricamente. Essas são as características de algumas distribuições de probabilidade, tais como a uniforme, a normal, a triangular/PERT e a *bootstrap*. Essas distribuições serão descritas nesse tópico, exceto a *bootstrap*, pois será descrita em um capítulo à parte.

Assim, a análise de riscos depende totalmente da escolha das distribuições de probabilidade, as quais representam as incertezas e variabilidades do sistema, e a característica mais básica das distribuições é se elas são discretas ou contínuas (HAUWERMEIREN, 2008). As discretas aceitam um conjunto de valores finitos com suas respectivas probabilidades; as contínuas podem ter qualquer valor dentro de uma faixa pré-estabelecida.

Outra característica das distribuições de probabilidades é ser paramétricas ou não paramétricas. As paramétricas são baseadas em modelos, ou seja, a forma dessa distribuição origina-se a partir de um modelo matemático. As distribuições não paramétricas são empíricas, ou seja, a sua forma é definida a partir de uma amostra.

Na sequência, as descrições das distribuições de probabilidade binária, uniforme, normal e triangular.

A Distribuição Binomial modela o número de sucessos a partir de  $n$  amostras independentes, com probabilidade  $p$  de sucesso. Exemplos de processos (ou sistemas) que são binomiais ou que podem ser reduzidos a binomiais, são, por exemplo:

- a) número de falhas na partida de um carro em  $n$  tentativas;
- b) número de itens falhos em  $n$  de uma linha de produção;
- c) número de máquinas que duraram mais de  $T$  horas de operação sem falhas.

O Quadro 12 mostra um resumo da distribuição binomial.

Quadro 12 – Resumo da distribuição binomial

<b>Parâmetros</b>	$n = 1, 2, 3, \dots$ $p (0 < p < 1)$	$n$ : tamanho da amostra $p$ : probabilidade de sucesso
<b>Notação</b>	$Binomial(n, p)$	
<b>Domínio</b>	Discreto $x = \{0, 1, 2, \dots, n\}$	
<b>Valor esperado</b>	$np$	
<b>Variância</b>	$np(1 - p)$	

Fonte: BECKER (2015).

A Distribuição Uniforme atribui probabilidade igual a todos os valores entre seu mínimo e máximo e, segundo Hauwermeiren (2008), pode ter os seguintes usos:

- estimativa aproximada: utilizada para um modelo em que se tem pouco ou nenhum dado disponível;
- análise de sensibilidade: atribui a distribuição uniforme a um parâmetro e executa a simulação para observar a influência dessa parâmetro no sistema;
- distribuição prévia desconhecida: utilizada quando não se conhece a distribuição prévia em uma inferência Bayesiana.

O Quadro 13 mostra um resumo da Distribuição Uniforme.

Quadro 13 – Resumo da distribuição uniforme

<b>Parâmetros</b>	$min$ $max (min < max)$	$min$ : valor mínimo da distribuição $max$ : valor máximo da distribuição
<b>Notação</b>	$Uniforme(min, max)$	
<b>Domínio</b>	Discreto ou contínuo $min \leq x \leq max$	
<b>Valor esperado</b>	$\frac{min+max}{2}$	
<b>Variância</b>	$\frac{(max-min)^2}{12}$	

Fonte: BECKER (2015).

A Distribuição Normal possui maior frequência na média e, à medida em que se afasta dela, a frequência diminui, aproximando-se da forma de um sino. A estatística clássica foi desenvolvida no conceito da distribuição normal. Seguem

abaixo alguns modelos que possuem a distribuição normal, de acordo com Hauwermeiren (2008):

- a) variáveis que ocorrem naturalmente: a distribuição normal ocorre em várias aplicações devido, em parte, ao teorema central do limite, que é uma boa aproximação de outras distribuições;
- b) distribuição de erros;
- c) aproximação de distribuições desconhecida: a regra básica em estatística é que quanto mais dados se tem disponível, mais a distribuição desconhecida se aproxima da normal.

O Quadro 14 mostra algumas características da Distribuição Normal.

Quadro 14 – Resumo da distribuição normal

<b>Parâmetros</b>	$\mu$ $\sigma (\sigma > 0)$	$\mu$ : média $\sigma$ : desvio padrão
<b>Notação</b>	$Normal(\mu, \sigma)$	
<b>Domínio</b>	Contínuo $-\infty \leq x \leq +\infty$	
<b>Valor esperado</b>	$\mu$	
<b>Variância</b>	$\sigma^2$	

Fonte: BECKER (2015).

A Distribuição Triangular é caracterizada por três parâmetros: o valor mínimo, o valor máximo e o valor mais provável (HAUWERMEIREN, 2008). Ela é utilizada como uma ferramenta de modelagem quando se tem uma ideia do seu valor mínimo e máximo, como na distribuição uniforme, e também quando se tem uma estimativa para o valor mais provável.

Apesar de usar os mesmos parâmetros da distribuição PERT, a distribuição triangular é diferente da PERT, pois a triangular enfatiza as caudas (valores extremos), ao passo que a PERT dá pouca ênfase nas caudas.

O Quadro 15 mostra os resumos das distribuições triangular e PERT. Analisando esse quadro, são percebidas as seguintes diferenças entre as duas distribuições:

- a) a restrição dos parâmetros: na triangular, a moda pode ser igual ao mínimo ou ao máximo e na PERT não pode;

- b) o cálculo do valor esperado: o PERT dá mais ênfase na média, porque atribui peso maior a ela;
- c) o cálculo da variância.

Quadro 15 – Resumo das distribuições triangular e PERT

<b>Parâmetros</b>	$min$ $max$ ( $min < max$ ) $moda$ ( $min \leq moda \leq max$ )	$min$ $max$ ( $min < max$ ) $moda$ ( $min < moda < max$ )
<b>Notação</b>	$Triangular(min, max, moda)$	$PERT(min, max, moda)$
<b>Domínio</b>	Contínuo $min \leq x \leq max$	Contínuo $min \leq x \leq max$
<b>Valor esperado</b>	$\frac{min+max+moda}{3}$	$\frac{min+max+4moda}{6} \equiv \mu$
<b>Variância</b>	$\frac{min^2+max^2+moda^2-min\ max-min\ mode-max\ mode}{18}$	$\frac{(\mu-min)(max-\mu)}{7}$
	$min$ : valor mínimo da distribuição $max$ : valor máximo da distribuição $moda$ : é o valor que mais ocorre	

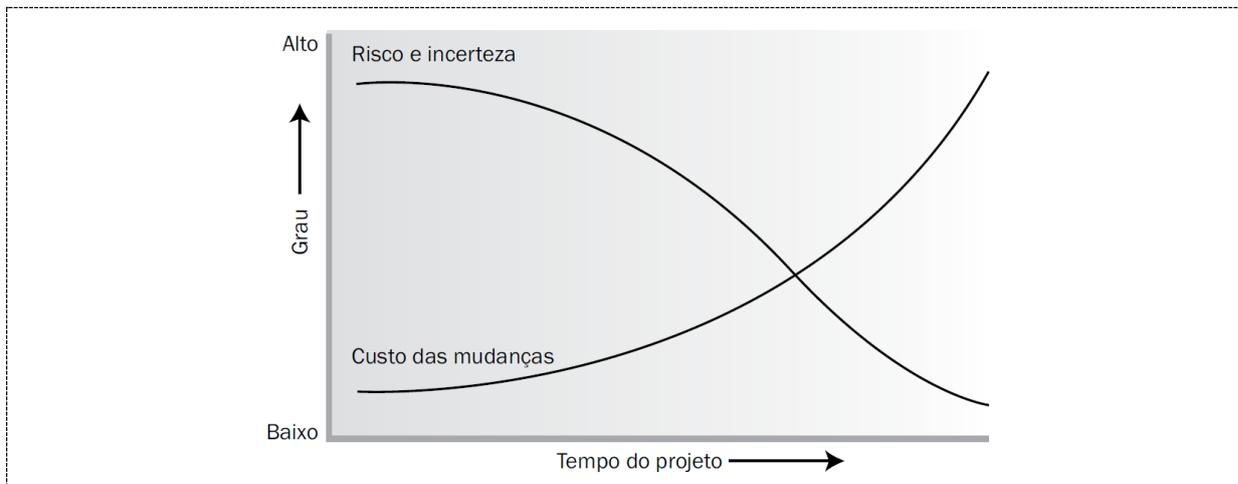
Fonte: BECKER (2015) e HAUWERMEIREN (2008).

Este tópico descreveu as distribuições de probabilidades e destacou que elas são utilizadas para atribuir incertezas e variabilidade às variáveis que fazem parte do modelo de um sistema.

## 2.4 GERENCIAMENTO DOS RISCOS

Existe um relacionamento entre o tempo de projeto e as incertezas que podem ocorrer (Figura 13). À medida que o tempo de execução do projeto avança, os custos de mudanças aumentam e as incertezas diminuem. O início do projeto é o momento de maior incerteza, então, realizar a análise dos riscos do cronograma do projeto, ainda em planejamento, e identificar problemas no cronograma, o custo de sua alteração não é alto, porque o projeto ainda está em fase de planejamento.

Figura 13 – Custo de mudanças x riscos e incerteza



Fonte: PMI (2013, p. 40).

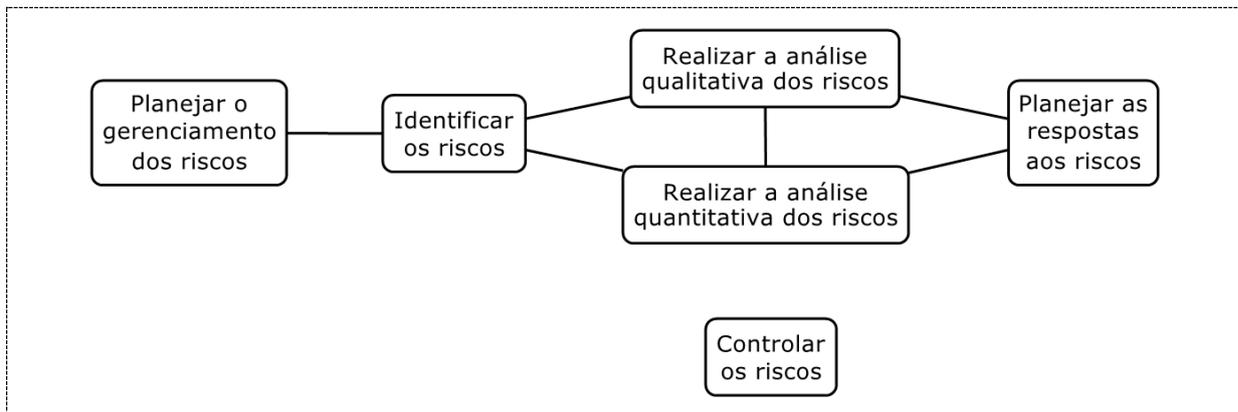
Assim, este tópico trata do gerenciamento dos riscos do projeto. Para PMI (2013), risco pode ser definido como "um evento ou condição incerta que, se ocorrer, provocará um efeito positivo ou negativo em um ou mais objetivos do projeto tais, como escopo, cronograma, custo e qualidade." (p. 310). Conseqüentemente, gerenciamento de risco é a execução dos "processos de identificação, análise, planejamento de respostas e controle de riscos de um projeto" (p. 309).

A explanação a seguir está dividida em duas partes. A primeira explica os processos de gerenciamento dos riscos do projeto de uma maneira geral; a segunda detalha a análise quantitativa de riscos, mostrando como analisar as incertezas do cronograma do projeto porque "tempo, custos e riscos estão intimamente associados" (FINOCCHIO JÚNIOR, 2013, p. 95).

#### 2.4.1 Processos do gerenciamento dos riscos

No gerenciamento dos riscos do projeto existem seis processos (Figura 14), sendo que os cinco primeiros pertencem ao grupo de processos planejamento e o último pertence ao grupo de processos de monitoramento e controle. Cada um dos processos será descrito de forma resumida baseando-se em PMI (2009) e em PMI (2013).

Figura 14 – Processos do gerenciamento dos riscos do projeto



Fonte: PMI (2013, p. 141).

O objetivo geral do processo Planejar o gerenciamento dos riscos é definir como serão conduzidas as atividades de gerenciamento dos riscos do projeto. Possui como objetivos específicos desenvolver a estratégia geral do gerenciamento de riscos; decidir como os processos do gerenciamento dos riscos serão executados; integrar o gerenciamento dos riscos do projeto com as outras atividades de planejamento; estabelecer o patamar de risco aceitável para o projeto; identificar a tolerância a risco das principais partes interessadas; resolver eventuais diferenças e comunicar as conclusões à equipe de projeto; definir os termos para análise qualitativa de probabilidade e impacto sobre o projeto; definir potenciais fontes de riscos para o projeto. O resultado desse processo é o plano de gerenciamento dos riscos do projeto.

Já o processo Identificar os riscos tem como objetivo geral identificar os riscos que podem afetar o projeto, documentar as características dos riscos e identificar potenciais respostas aos riscos. Tem como resultado a lista de riscos com suas potenciais respostas.

Realizar a análise qualitativa dos riscos tem como objetivo geral analisar as características individuais dos riscos para ação posterior e seus objetivos específicos são avaliar a probabilidade e o impacto da ocorrência dos riscos e categorizar os riscos de acordo com a causa raiz e os efeitos comuns. Uma vez que forem acrescentadas novas informações para os riscos, o resultado desse processo serão atualizações na lista dos riscos.

O processo Realizar a análise quantitativa dos riscos tem como o objetivo geral analisar numericamente o efeito global dos riscos identificados nos objetivos

do projeto; os objetivos específicos são avaliar a probabilidade de atingir os objetivos do projeto e estimar as reservas de contingência de tempo e custo.

O objetivo geral de Planejar as respostas aos riscos é desenvolver um conjunto de ações que ampliem as chances de sucesso do projeto, ao mesmo tempo em que se atende às restrições do projeto. Os objetivos específicos são desenvolver um plano para cada risco considerado importante, para aproveitar uma oportunidade ou tratar uma ameaça aos objetivos do projeto; identificar riscos secundários (originados em decorrência do planejamento das respostas) para avaliação e planejamentos das respostas; identificar e documentar riscos residuais; integrar todas as ações de respostas ao plano de gerenciamento de projetos.

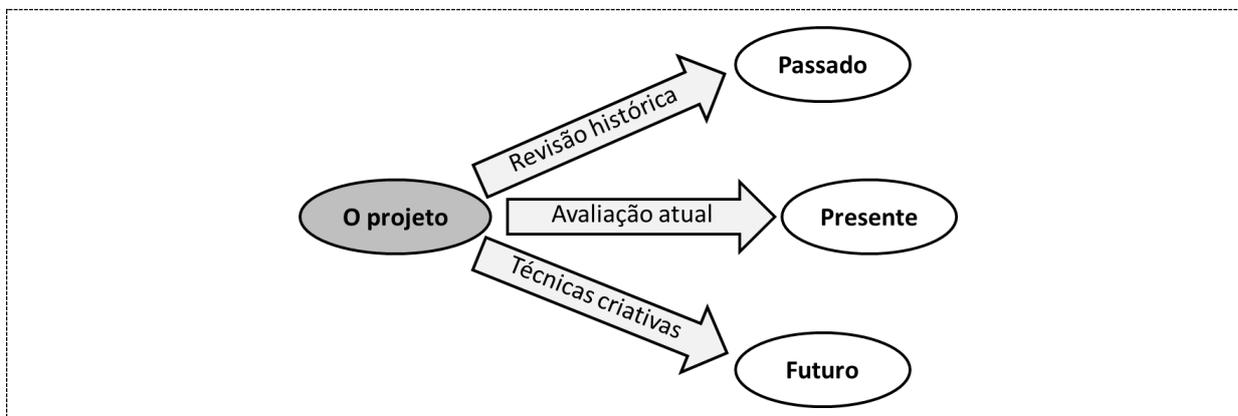
Por fim, Controlar os riscos tem como o objetivo geral monitorar os riscos identificados e os residuais e identificar novos riscos. Os objetivos específicos são garantir que os planos de respostas sejam executados no tempo estipulado; avaliar a eficiência dos planos durante o ciclo do projeto; monitorar os gatilhos dos riscos; executar os planos de respostas.

Neste tópico foi visto, de maneira geral, os processos de gerenciamento dos riscos em projetos. Pode concluir que os processos são planejamento, identificação, análise, respostas e monitoramento (PIARC). No próximo tópico será mostrado como fazer a análise das incertezas do cronograma com o processo de análise quantitativa dos riscos.

#### 2.4.2 Análise quantitativa dos riscos

Existem várias técnicas para fazer o gerenciamento dos riscos do projeto e, de acordo com PMI (2009), elas se agrupam em três categorias: revisão histórica, avaliação atual e técnicas criativas, conforme mostrado na Figura 15. A revisão histórica trata de analisar os eventos já ocorridos em projetos anteriores; a avaliação atual utiliza-se dos dados atuais para identificar incertezas e possíveis falhas no projeto; e as técnicas criativas são utilizadas para identificar cenários que podem ocorrer no futuro.

Figura 15 – Perspectivas da identificação de riscos



Fonte: PMI (2009, p. 27).

Outro ponto de vista defende que a análise dos riscos dos projetos pode ser feita de duas formas: a qualitativa e a quantitativa. As principais diferenças entre as duas são apresentadas no Quadro 16.

Quadro 16 – Diferenças entre a análise qualitativa e quantitativa dos riscos

Análise Qualitativa	Análise Quantitativa
<ul style="list-style-type: none"> <li>• Aborda os riscos de forma individual e descritiva;</li> <li>• Avalia a probabilidade discreta de ocorrência e o impacto nos objetivos do projeto;</li> <li>• Prioriza os riscos individuais para tratamento subsequente.</li> </ul>	<ul style="list-style-type: none"> <li>• Estima a probabilidade dos resultados do projeto baseado no efeito combinado dos riscos;</li> <li>• Usa uma distribuição de probabilidade para caracterizar a probabilidade e impacto dos riscos;</li> <li>• Usa um modelo de cálculo para, por exemplo, o cronograma ou os custos;</li> <li>• Usa um método quantitativo, que requer ferramentas especializadas;</li> <li>• Estima a probabilidade de atingir os alvos e contingências necessárias para desejado nível de conforto;</li> <li>• Identifica os riscos com os maiores efeitos nos riscos globais do projeto.</li> </ul>

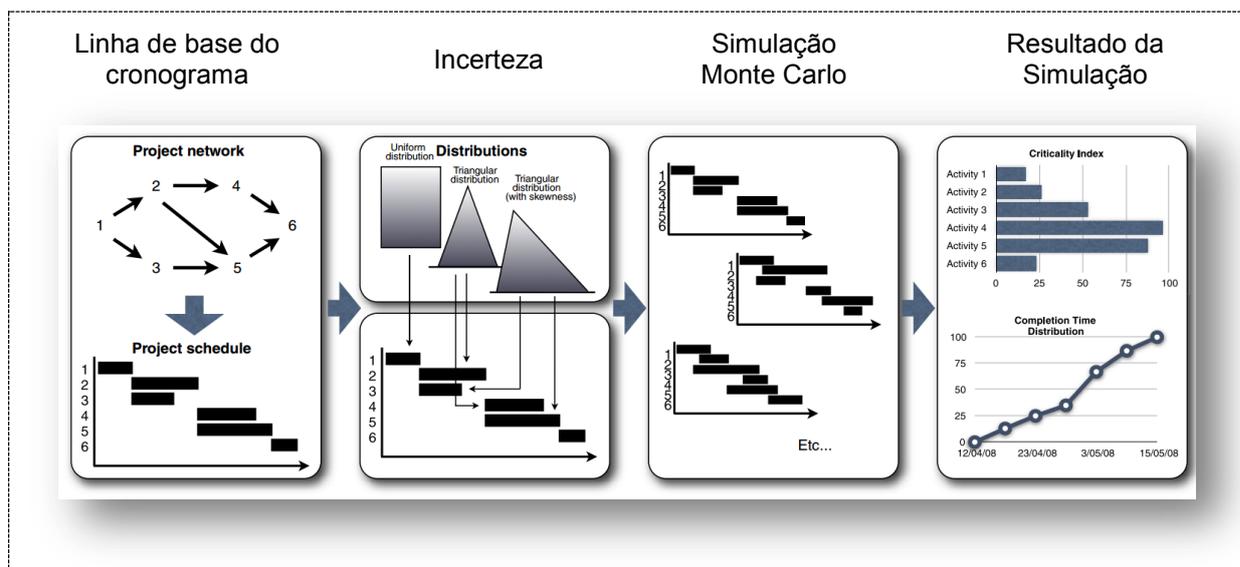
Fonte: PMI (2009, p. 38).

O estudo da análise quantitativa de riscos será feito com base no método Análise de Risco de Cronogramas (*Schedule Risk Analysis*) criado por Vanhoucke (2015) e é dividido em quatro passos (Figura 16):

- a) criar a linha de base do cronograma: planejamento tradicional já feito pelo gerente do projeto;
- b) identificar as incertezas do cronograma: atribuindo uma distribuição de probabilidade para a duração das atividades;

- c) simular a execução do cronograma: utilizando a técnica Monte Carlo;
- d) analisar os resultados da simulação: por meio de métodos estatísticos e métricas de qualidade.

Figura 16 – Passos da análise de riscos dos cronogramas



Fonte: VANHOUCKE (2013, p. 83).

Todos esses passos já foram descritos anteriormente, exceto os índices de sensibilidade. Com os índices de sensibilidade é possível priorizar as atividades por importância, quer elas sejam críticas ou não, para receber uma atenção maior, possibilitando, destinar os melhores recursos às atividades mais importantes. O próximo tópico tratará dos índices de sensibilidade do tempo.

### 2.4.3 Índices de sensibilidade do tempo

O resultado do passo quatro da análise de riscos do cronograma é a análise das simulações efetuadas no passo três (conforme Figura 16 – Passos da análise de riscos dos cronogramas). A saída desse passo são as medidas que definem o grau de sensibilidade e criticidade das atividades do cronograma (VANHOUCKE, 2013). Essas medidas vão além de simplesmente dizer se uma atividade é crítica ou não, elas oferecem um grau de criticidade.

Os índices de sensibilidade do tempo são assim chamados porque indicam as atividades que mais influenciam no tempo de execução do projeto e serão chamados de "índice" para se diferenciar dos indicadores topológicos. Considerando as três formas do índice de crucialidade, eles são em número de seis:

- a) Índice de criticidade (IC): mede a probabilidade de uma atividade estar no caminho crítico;
- b) Índice de significância (IS): mede a importância relativa de uma atividade;
- c) Índice de sensibilidade do cronograma (ISC): mede a importância relativa de uma atividade, levando em consideração o índice de criticidade;
- d) Índice de crucialidade (ICR): mede a correlação entre a duração da atividade e a duração total do projeto de três formas diferentes:
- ICR-r: coeficiente de correlação linear de Pearson;
  - ICR-rho: coeficiente de correlação posto ordem de Spearman
  - ICR-tal: coeficiente de correlação posto ordem de Kendall.

Para fins de comparação, o Quadro 17 mostra os termos traduzidos para o português e seus respectivos originais em inglês.

Quadro 17 – Índices de sensibilidade em português e em inglês

Termo em português	Termo em inglês
Índice de criticidade (IC)	<i>Criticality index (CI)</i>
Índice de significância (IS)	<i>Significance index (SI)</i>
Índice de sensibilidade do cronograma (ISC)	<i>Schedule sensitivity index (SSI)</i>
Índice de crucialidade (ICR)	<i>Cruciatility index (CRI)</i>

Fonte: VANHOUCKE (2009).

### Índice de criticidade (IC)

O índice de criticidade mede a probabilidade de uma atividade pertencer ao caminho crítico. Ele é medido pelo percentual de vezes que a atividade  $i$  pertenceu ao caminho crítico na execução das simulações (VANHOUCKE, 2013). O IC pode ser calculado da seguinte forma:

$$IC = \frac{1}{nrS} \sum_{s=1}^{nrS} Um(ft_i^s = 0)$$

Onde:

- $ft_i^s$ : folga total da atividade  $i$  na simulação  $s$ ;
- $Um(G)$ : é uma operação definida por Vanhoucke (2009), que retorna 1 se a expressão lógica  $G$  for verdadeira e 0, caso contrário, formulada como:

$$Um(G) = \begin{cases} 1, & \text{se } G \text{ é verdadeiro} \\ 0, & \text{se } G \text{ é falso} \end{cases}. \text{ A expressão lógica nesse contexto é se a}$$

folga total da atividade é igual a zero, indicando que ela é crítica;

- $nrS$ : Número total de simulações executadas.

Vanhoucke (2013) aponta como principal falha do índice de criticidade é que ele é apenas baseado na probabilidade de uma atividade ser crítica ou não, sem levar em conta a sua duração.

### Índice de significância (IS)

Com o objetivo de refletir a importância relativa entre as atividades do projeto, o índice de significância de uma atividade  $i$  é formulado da seguinte forma:

$$IS = \frac{1}{nrS} \sum_{s=1}^{nrS} \left( \frac{d_i^s}{d_i^s + ft_i^s} \right) \left( \frac{DTP^s}{E(DTP)} \right)$$

Onde:

- $nrS$ : Número total de simulações executadas;
- $s$ : Índice de cada simulação, variando de 1, ...,  $nrS$ ;
- $d_i^s$ : Duração da  $i$ -ésima atividade na simulação  $s$ ;
- $ft_i^s$ : Folga total da atividade  $i$  na simulação  $s$ ;
- $DTP^s$ : Duração total do projeto na simulação  $s$ ;
- $E(DTP)$ : Valor esperado (média) da duração total do projeto.

O objetivo o IS é expor a significância de cada atividade na duração total do projeto, levando-se em conta não só se ela pertence ou não ao caminho crítico, mas também a folga que ela possui.

### Índice de sensibilidade do cronograma (ISC)

PMI (2013) menciona que um risco deve ser avaliado com a sua probabilidade e o seu impacto. O índice de sensibilidade do cronograma trata dessas duas questões, combinando o índice de criticidade, que leva em consideração a probabilidade da atividade  $i$  ser crítica, com o desvio padrão dessa atividade ( $\sigma_{d_i}$ ) e da duração total do projeto ( $\sigma_{DTP}$ ) (VANHOUCKE, 2013). A sua formulação fica da seguinte forma:

$$ISC = \frac{\sigma_{d_i}}{\sigma_{DTP}} IC$$

Onde:

- $IC$ : Índice de criticidade
- $\sigma_{d_i}$ : Desvio padrão da duração da  $i$ -ésima atividade;
- $\sigma_{DTP}$ : Desvio padrão da duração total do projeto.

### Índice de crucialidade (ICR)

O índice de crucialidade mede a correlação entre a duração da atividade  $i$  e a duração total do projeto (VANHOUCKE, 2013) e, para esse autor, o ICR

[...] reflete a importância de uma atividade de um modo mais intuitivo e mede a porção da incerteza da duração total do projeto que pode ser explicada pela incerteza de uma atividade. (VANHOUCKE, 2013, p. 90)

Esse índice pode ser calculado com as três formas do índice de correlação:

- ICR-r: coeficiente de correlação linear de Pearson:

$$ICR_r = \frac{\sum_{s=1}^{nrS} (d_i^s - \bar{d}_i) (DTP^s - \overline{DTP})}{nrS \sigma_{d_i} \sigma_{DTP}}$$

Onde:

- $d_i^s$ : Duração da  $i$ -ésima atividade na simulação  $s$ ;
- $d_i$ : Duração da  $i$ -ésima atividade;
- $DTP^s$ : Duração total do projeto na simulação  $s$ ;
- $\overline{DTP}$ : Média da duração total do projeto;
- $nrS$ : Número total de simulações executadas;
- $\sigma_{d_i}$ : Desvio padrão da duração da  $i$ -ésima atividade;
- $\sigma_{DTP}$ : Desvio padrão da duração total do projeto.

- ICR-rho: coeficiente de correlação posto ordem de Spearman:

$$ICR_{rho} = 1 - \frac{6}{nrS(nrS^2 - 1)} \sum_{s=1}^{nrS} dif_s^2$$

Onde:

- $nrS$ : Número total de simulações executadas;
- $s$ : Índice de cada simulação, variando de  $1, \dots, nrS$ ;
- $dif_s$ : é a diferença da ordem dos valores da duração da atividade  $i$  com a ordem dos valores da duração total do projeto, na simulação  $s$ .

- ICR-tal: coeficiente de correlação posto ordem de Kendall:

$$ICR_{tal} = \left[ \frac{4}{nrS(nrS - 1)} \sum_{s=1}^{nrS-1} \sum_{t=s+1}^{nrS} Um\{(d_i^t - d_i^s)(DTP^t - DTP^s) > 0\} \right] - 1$$

Onde:

- $nrS$ : Número total de simulações executadas;

- $s, t$ : Índice de cada simulação;
- $d_i^s$ : Duração da  $i$ -ésima atividade na simulação  $s$ ;
- $d_i^t$ : Duração da  $i$ -ésima atividade na simulação  $t$ ;
- $DTP^s$ : Duração total do projeto na simulação  $s$ ;
- $DTP^t$ : Duração total do projeto na simulação  $t$ ;
- $Um(G)$ : é uma operação definida por Vanhoucke (2009), que retorna 1 se a expressão lógica  $G$  for verdadeira e 0 caso contrário, formulada como:  $Um(G) = \begin{cases} 1, & \text{se } G \text{ é verdadeiro} \\ 0, & \text{se } G \text{ é falso} \end{cases}$ . A expressão lógica nesse contexto é se a duração da atividade e a duração total projeto com índices  $t$  e  $s$  possuem o mesmo sinal.

Por medir a correlação linear entre cada uma das atividades e a duração total do projeto (e muitas vezes não existir essa relação), o ICR-r pode não ser adequado. Dessa forma, o índice de crucialidade é medido com os coeficientes não paramétricos, ICR-rho e ICR-tal, que consideram a ordem da cada atividade dentro da simulação (VANHOUCKE, 2014).

Deve-se observar que os índices de sensibilidade são calculados para cada uma das atividades do cronograma e que as atividades com maiores índices serão consideradas as mais sensíveis e as que devem receber uma prioridade maior.

## 2.5 DISTRIBUIÇÃO *BOOTSTRAP*

*Bootstrapping* é um termo que se refere a se colocar em prontidão, em alerta, a partir de seus próprios recursos, possivelmente advinda da expressão idiomática "*to pull yourself up by your bootstraps*" (VOSE, 2008, p. 247) que foi utilizado pela primeira vez por Rudolph Erich Raspe em seu livro *The Singular Travels, Campaigns and Adventures of Baron Munchausen*, em 1786 (CHERNICK; LABUDD, 2011). No contexto do livro, é uma ordem para os soldados apertarem os cordões (*straps*) de suas botas (*boots*) para indicar que estão preparados para a batalha.

Em informática, termo *booting* significa reinicializar o computador, até que ele fique pronto para operar. Outro uso do termo *bootstrap* é na criação de compiladores de uma linguagem de programação. *Bootstrap* é o termo utilizado para indicar que a linguagem já está sendo desenvolvida a partir de um compilador da própria linguagem. Já na área de estatística, o termo *bootstrap* significa retirar

amostras da própria amostra, ou seja, como nos conceitos anteriores, se desenvolver a partir de seus próprios recursos.

Este tópico tratará a distribuição *bootstrap*. Na primeira parte, serão mostrados os conceitos fundamentais da distribuição *bootstrap* e, na sequência, serão vistas algumas aplicações. O próximo tópico mostrará os desvios que podem ocorrer na distribuição *bootstrap* e como tratá-los.

### 2.5.1 Caracterização da distribuição *bootstrap*

A distribuição *bootstrap* foi criada por Bradley Efron, em 1979, e é um método de reamostragem (MOORE et al., 2006). Os métodos de reamostragem permitem quantificar incertezas através do cálculo de erros padrões e intervalos de confiança e, também, realizar testes de significâncias.

O criador da distribuição *bootstrap* a define como sendo "uma técnica baseada em computador para estimar erro padrão, viés, intervalo de confiança e outras medidas estatísticas de precisão." (EFRON, 1992, p. 83), e como ela é baseada em reamostragem da própria amostra, uma distribuição empírica, é um método estatístico não paramétrico (CHERNICK, LABUDDE, 2011). Não paramétrico indica que esse método não está baseado em suposições sobre alguma distribuição pré-existente.

A distribuição *bootstrap* tem se tornado uma técnica útil porque está alinhada às técnicas tradicionais e permite que se avalie a incerteza sobre os parâmetros em estudo (VOSE, 2008).

Para Moore et al. (2006), as vantagens da utilização dos métodos de reamostragem são as seguintes:

- menos suposições: a distribuição da população não necessita ser normal nem a amostra extraída dela ser grande;
- maior precisão: esses métodos permitem obter uma maior precisão do que os tradicionais;
- generalidades: os métodos de permutação se utilizam das mesmas fórmulas estatísticas convencionais;
- função pedagógica: os métodos *bootstrap* são análogos às distribuições amostral, porém são concretos.

Para gerar uma distribuição *bootstrap*, devem ser executados os seguintes passos:

- a) extrair várias reamostras com reposição a partir da amostra original, que serão chamadas de amostra *bootstrap*;
- b) calcular a estatística em análise com base na amostra *bootstrap*, resultando em vários valores da estatística, chamados de distribuição *bootstrap*;
- c) utilizar a distribuição *bootstrap* para verificar a forma, o centro e a dispersão da estatística.

A partir de uma amostra da população de tamanho  $n$ , o *bootstrap* retira amostras com reposição também de tamanho  $n$ . A amostra *bootstrap* tem tamanho  $n^n$ , tornando impossível sua utilização. Para resolver esse problema, Efron (1979) criou a aproximação Monte Carlo da distribuição *bootstrap*. Assim, neste texto, sempre que se falar de *bootstrap*, está se referindo à aproximação Monte Carlo da distribuição *bootstrap*.

O elemento fundamental para que se possa extrair a distribuição *bootstrap* é que já exista uma amostra da população, com uma distribuição qualquer, chamada de distribuição empírica. A distribuição empírica é uma distribuição discreta com cada elemento tendo probabilidade  $\frac{1}{n}$ . A distribuição *bootstrap* utiliza apenas o que se sabe sobre os dados, não faz nenhuma suposição sobre a distribuição da população.

A distribuição *bootstrap* pode ser utilizada para várias estatísticas, desde as mais simples, como média, média aparada, diferença entre médias, até as mais complexas, como o erro padrão e o coeficiente de correlação.

A distribuição *bootstrap* é útil porque possui a forma, o viés e a dispersão da amostra original e, conseqüentemente, da população. A simplicidade é uma de suas características, o que faz com que ela possa ser utilizada para vários tipos de problemas, bastando utilizar a aproximação Monte Carlo.

### 2.5.2 Fontes de variação da distribuição *bootstrap*

Este tópico tratará de duas questões relativas a fontes de variação da distribuição *bootstrap*. Uma é sobre a estimativa da estatística em si; a outra é sobre o cálculo de intervalo de confiança.

A primeira questão é sobre a variação no cálculo da estatística. As extrações aleatórias de amostras da distribuição *bootstrap*, utilizando a aproximação Monte

Carlo, possuem duas fontes de variações que, segundo Chernick & Labudde (2011) são as seguintes:

- a) a amostra original, extraída da população;
- b) as amostras *bootstrap*, feitas a partir da amostra original.

Para verificar o impacto dessas fontes de variações na distribuição *bootstrap*, Moore et al. (2006) fizeram alguns experimentos e chegaram às seguintes conclusões com relação à distribuição *bootstrap*:

- a) é centrada na média amostral ( $\bar{x}$ );
- b) possui a forma e a dispersão da distribuição amostral;
- c) as reamostras acrescentam poucas variações;
- d) se tiver recursos disponíveis, é melhor aumentar o tamanho da amostra da população e utilizar a distribuição *bootstrap* do que retirar várias amostras diretamente da população.

Por exemplo, o cálculo da estatística desejada (duração do projeto) feito com a distribuição *bootstrap* se aproxima do verdadeiro parâmetro na medida em que se aumenta o número de simulações ( $n_{rs}$ ), então a maior parte da variação de uma estatística vem da amostra original, pois pouca variação é acrescentada pela distribuição *bootstrap*.

Outra questão é sobre o Intervalo de confiança *bootstrap*. Para criar um intervalo de confiança *bootstrap* de 95%, por exemplo, basta calcular o percentil 2,5 e 97,5 da distribuição *bootstrap*. Esse intervalo de confiança é denominado intervalo de confiança percentil. Outra forma de calcular o intervalo de confiança é utilizando a distribuição t de *student*, mas, para poder utilizá-la, deve ser feita a suposição de que a distribuição *bootstrap* possui a forma normal, o que pode não ocorrer. Uma forma de verificar se a distribuição é normal, é comparar as duas formas de cálculo de intervalo confiança, a percentil e o *bootstrap*-t. Se os intervalos não forem iguais, a distribuição possui viés, e nenhuma das duas pode ser utilizada.

Se a amostra for enviesada, para o cálculo de intervalo de confiança, deve-se utilizar os métodos *bootstrap* acelerado de correlação de viés (BCa – *bias corrected accelerated*) ou o *bootstrap tilting*. O BCa ajusta os pontos extremos do percentil da distribuição *bootstrap* para corrigir a assimetria. Já o intervalo *tilting* ajusta o "processo de extração aleatória das reamostras." (MOORE et al., 2006, p. 820). Tanto um quanto outro fornecem resultados satisfatórios.

O objetivo de se utilizar a distribuição *bootstrap* é identificar a variabilidade da duração do projeto para que o gerente do projeto analise o cronograma e execute as devidas correções, caso sejam necessárias. Se a variabilidade for corrigida utilizando um dos métodos citados, a informação da variabilidade não será repassada ao gerente do projeto, impossibilitando a tomada de decisão sobre o que corrigir no projeto. Então, nesta pesquisa, será utilizada a distribuição *bootstrap* sem corrigir o viés que ela possa ter.

### 2.5.3 Áreas de aplicação da distribuição *bootstrap*

O objetivo deste tópico é identificar as áreas de aplicação em que já foram desenvolvidos artigos sobre a distribuição *bootstrap*. Para atender o objetivo deste tópico, foi desenvolvida pelo autor uma pesquisa que está disponibilizada no Apêndice A. Este tópico fará um resumo da pesquisa e, em seguida, apresentará as principais conclusões.

O objetivo da pesquisa do Apêndice A foi identificar as áreas de aplicação da distribuição *bootstrap*. Para isso, foi desenvolvida uma pesquisa bibliométrica a partir das bases de artigos científicos *Science Direct* e *Web of Science*, e cada palavra do título, do resumo e as palavras-chave foram analisadas e agrupadas utilizando a técnica TF-IDF (*term frequency-inverse document frequency*), que, de maneira resumida, multiplica a frequência do termo pelo inverso da frequência de documentos em que o termo aparece. Essa técnica é mais apropriada do que buscar os termos mais frequentes porque muitas vezes o termo mais frequente não é o mais relevante.

O texto resultante dessa pesquisa infere as áreas de aplicações da simulação *bootstrap* a partir dos termos mais relevantes e verifica o histórico de aplicações sobre o assunto a ponto de concluir que a simulação *bootstrap* é utilizada em diversas áreas da ciência e, por isso, é uma técnica multidisciplinar. Além disso, observou-se que a produção de artigos aumentou ao longo dos anos, desde a criação da *bootstrap*, apesar de ter uma queda no ano de 2016.

Com base nas conclusões da pesquisa do apêndice, é possível concluir que a simulação *bootstrap* pode ser utilizada também na análise de riscos em cronograma de projetos e que artigos sobre o assunto também são relevantes.

### 3 MATERIAL E MÉTODOS

Esta pesquisa utilizou vários métodos para atingir o seu objetivo, e como metodologia “é o conjunto de métodos ou caminhos utilizados para a condução da pesquisa” (AMADEU et al., 2015, p. 49), esta seção apresenta a metodologia utilizada. A primeira parte efetua a classificação da pesquisa e, logo em seguida, a população e a amostra são definidas. Na sequência, a forma de coleta, tratamento e análise dos dados são explicados.

#### 3.1 CLASSIFICAÇÃO DA PESQUISA

Esta dissertação trata de **pesquisa aplicada**, pois tem a “finalidade de resolver problemas identificados no âmbito das sociedades” (GIL, 2010, p. 26). Ela busca a aquisição de conhecimentos com o objetivo de aplicá-lo em gerenciamento de projetos. Além disso, é uma **pesquisa descritiva**, porque visa “descobrir a existência de associação entre variáveis” (GIL, 2010, p. 28) e descrever características da amostra.

Como parte dos passos para se atingir o objetivo geral, verificou-se quais fatores contribuem para que haja a correlação entre os cálculos efetuados com as duas distribuições, então, também é uma **pesquisa explicativa** porque tem “como propósito identificar fatores que determinam ou contribuem para a ocorrência de fenômenos” (GIL, 2010, p. 28).

Com relação à natureza dos dados, esta é uma **pesquisa quantitativa**, pois “utiliza a coleta de dados para testar hipóteses, baseando-se na medição numérica e na análise estatística para estabelecer padrões e comprovar teorias” (HERNÁNDEZ SAMPIERI; FERNÁNDEZ COLLADO; BAPTISTA LUCIO, 2013, p. 30).

#### 3.2 POPULAÇÃO

A população deste estudo foi baseada em um conjunto de 4.100 projetos gerados aleatoriamente por Vanhoucke (2009), de acordo com a estrutura da rede. Os projetos foram divididos pelo autor em quatro grupos, baseados nos indicadores topológicos, conforme mostrado no Quadro 18.

Quadro 18 – Definição dos grupos de projetos

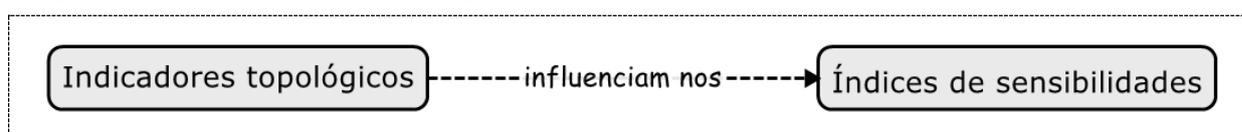
Grupo	SP	DA	CA	FT
1 Sequencial/paralelo (SP) 900 projetos	0,1 0,2 0,3 0,4 0,5 0,6 0,7 0,8 0,9	Aleatório	Aleatório	Aleatório
2 Distribuição das atividades (DA) 800 projetos	0,2 0,5	0,2 0,4 0,6 0,8	Aleatório	Aleatório
3 Comprimento dos arcos (CA) 1.200 projetos	0,2 0,5 0,8	Aleatório	0,2 0,4 0,6 0,8	Aleatório
4 Folga topológica (FT) 1.200 projetos	0,2 0,5 0,8	Aleatório	Aleatório	0,2 0,4 0,6 0,8

Fonte: VANHOUCKE (2009, p. 62-63).

O grupo 1 é formado com o objetivo de controlar o indicador sequencial-paralelo dos projetos. Os valores definidos para esse indicador variaram de 0,1 até 0,9; os outros indicadores são aleatórios. O objetivo do grupo 2 é controlar o indicador de distribuição das atividades e, para isso, foram definidos quatro valores para ele e dois valores para o SP; os outros indicadores foram aleatórios. O grupo 3 tem o objetivo de controlar o indicador de comprimento dos arcos e, subordinado a este, o SP; os indicadores distribuição das atividades e de folga topológica são aleatórios. E, finalmente, o grupo 4 controla o indicador folga topológica, fixando seus valores bem como os valores do SP; os indicadores DA e CA são gerados aleatoriamente. A regra geral que Vanhoucke (2009) utilizou para gerar os projetos foi controlar o valor de um ou dois indicadores topológicos e variar os outros aleatoriamente.

Como, de acordo com esse autor, a estrutura topológica da rede do cronograma influencia diretamente nos valores dos índices de sensibilidades, esta pesquisa avaliou diversos valores para os indicadores topológicos.

Figura 17 – Indicadores topológicos e índices de sensibilidades



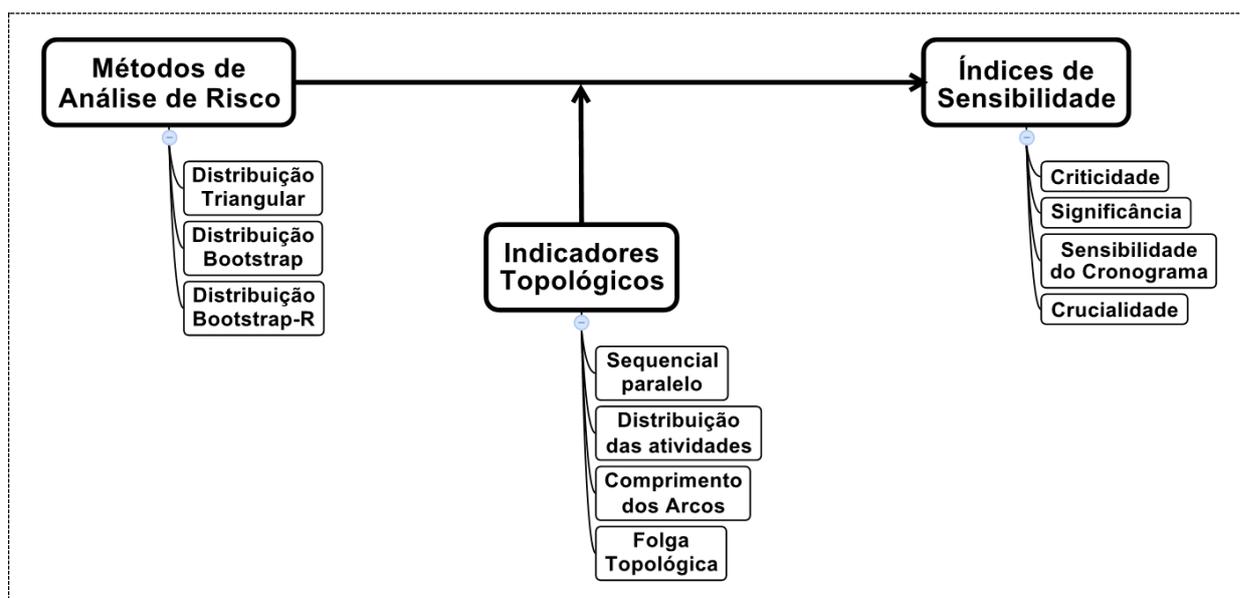
Fonte: O autor (2017) baseado em VANHOUCKE (2009).

Os projetos foram coletados de Vanhoucke (2009), disponibilizados no *site* <[http://www.or-as.be/sites/default/files/files/DataSetEvms\\_Patterson.zip](http://www.or-as.be/sites/default/files/files/DataSetEvms_Patterson.zip)>. Vanhoucke

(2009) informa que os projetos foram gerados aleatoriamente de acordo com a estrutura topológica da rede. Os 4.100 projetos estão disponíveis em arquivos texto, formato Patterson, uma estrutura pré-definida, conforme é explicado no *site* <[http://www.p2engine.com/p2reader/patterson\\_format](http://www.p2engine.com/p2reader/patterson_format)>, e, para facilitar a referência, foram numerados de 1 a 4.100.

A variável independente desta pesquisa é o método de análise de risco, efetuado de três formas diferentes: simulação do cronograma do projeto utilizando a distribuição triangular, a *bootstrap* padrão e a *bootstrap-rate*. As variáveis dependentes são os índices de sensibilidade (Figura 18). Os indicadores topológicos são variáveis de covariância, ou seja, influenciam nos índices de sensibilidade, mas não são objeto de estudo da pesquisa.

Figura 18 – Variáveis do estudo



Fonte: O autor (2017) baseado em VANHOUCKE (2009).

Para fins de padronização das durações dos projetos, a duração das atividades foi normalizada para que o caminho crítico dure 1.000 horas, totalizando 125 dias. Depois, esses projetos serão simulados, utilizando as distribuições deste estudo: triangular, *bootstrap* padrão e *bootstrap-rate*.

### 3.3 GRUPOS DE MÉTODOS

A execução de pesquisa científica pressupõe a adoção de um método de pesquisa e no caso desta pesquisa não é diferente. Como esta pesquisa é sobre o

estudo de um método, ela utiliza vários métodos para atingir seu objetivo, mais precisamente, trabalha com três grupos de métodos.

O primeiro grupo de métodos são os de análise de risco. Esses métodos utilizam a simulação Monte Carlo com a distribuição triangular (TR), *bootstrap* padrão (B-1) e *bootstrap-rate* (B-R). A distribuição utilizada passa a ser a variável independente desta pesquisa, pois é ela que será alterada para observar o resultado dos índices de sensibilidade.

O segundo grupo engloba os métodos de coleta de dados e cálculo das variáveis de análise. Para fazer a coleta dos dados, foram desenvolvidos algoritmos utilizando a linguagem de programação Java. Através do Java, foi programado o cálculo da rede do projeto, a execução do projeto e as simulações, bem como o cálculo dos indicadores topológicos e índices de sensibilidade.

E, finalmente, no último grupo, estão os métodos de validação e de comparação dos resultados. A partir de todo o conjunto de dados gerados, utilizou-se o ambiente de desenvolvimento estatístico **R** para efetuar a análise dos dados. Foram realizados testes estatísticos para verificar se os dados estavam de acordo com o esperado, que incluíram testes de igualdade de valor, de distribuição e de correlação.

### 3.4 ANÁLISE DE DADOS

Este tópico trata da análise de dados e o próximo tratará dos testes estatísticos. Esses dois temas são tratados em dois contextos diferentes. A análise de dados está no contexto de um gerente de projeto que, após ter finalizado o seu cronograma, deseja verificar se as estimativas feitas são adequadas. Então, ele faz a simulação do cronograma e analisa os resultados, executando um teste de hipótese para verificar se a duração estimada do cronograma está coerente com a simulação. Os testes de hipóteses estão no contexto deste trabalho, pois foram utilizados para comprovar a hipótese da pesquisa. Eles foram executados para verificar se as análises dos resultados dos cronogramas utilizando a distribuição triangular são semelhantes com as análises feitas com a distribuição *bootstrap*.

A análise dos dados é a etapa seguinte após a simulação e utiliza testes de hipóteses. Conforme orientam Chwif & Medina (2015), os objetivos da análise dos dados são verificar se existe alguma distribuição que explique o fenômeno e se os parâmetros definidos previamente se adequam aos resultados da simulação,

portanto, serão utilizadas ferramentas da estatística descritiva para explorar os dados e também da estatística inferencial para chegar-se a algumas conclusões.

Segundo esses autores, para se executar a análise dos resultados, deve-se "escolher corretamente os parâmetros que utilizaremos como medidas de desempenho do sistema." (p. 104). Como o interesse desta pesquisa está na duração do projeto e na criticidade das atividades, as medidas de desempenho estão atreladas a essas duas variáveis, que, neste contexto, podem ser resumidas aos índices de sensibilidade do cronograma. Um conjunto de medidas são as probabilidades, por exemplo, de o projeto finalizar no prazo estimado. Considerando que o resultado principal das simulações é a duração média do projeto, será que essa duração pode ser considerada igual à duração estimada do projeto sem a simulação? Para responder a essa questão será feito um teste de hipótese.

Na montagem do cronograma de um projeto, o seu principal resultado é a duração total do projeto. Segundo Chwif & Medina (2015), para que se tenha confiança de que essa duração está coerente, cria-se um "intervalo de valores que contém a média da população, com certa probabilidade" (p. 105), que, neste caso, a média da população é a duração total do projeto. Se a probabilidade for alta, maior a confiança estatística de que a duração total do projeto está no intervalo calculado; no entanto, a precisão depende ao tamanho do intervalo, que deve ser o menor possível, "pois de nada adianta uma alta confiança estatística se o tamanho do intervalo é tão grande que não podemos concluir nada sobre a média da população." (p. 105).

Então, de acordo com esses autores, para confirmar a confiança e a precisão dos parâmetros, é preciso executar projetos de experimentos, pois auxiliam na identificação de quais parâmetros têm o maior efeito no experimento, ou seja, "é um método que permite estimar como as variáveis de entrada afetam a(s) resposta(s) de um experimento." (p. 139).

As simulações que foram executadas nesta dissertação são do tipo sistemas terminais porque os sistemas bases, os projetos, possuem um início e um fim determinado (PMI, 2013) e as simulações terminais possuem exatamente essa característica (CHWIF; MEDINA, 2015). Os resultados das simulações serão analisados por meio de testes estatísticos.

### 3.5 TESTES ESTATÍSTICOS

Os testes estatísticos são realizados para verificar se uma amostra pode ser considerada igual a um determinado valor da população, ou a uma distribuição teórica. Então, para realizá-los, é necessário extrair uma amostra da população de determinado tamanho ( $n$ ).

A escolha de um teste estatístico depende da amostra possuir determinadas características e, de uma maneira geral, existem duas formas de realizá-los: os paramétricos e os não paramétricos, cada um requerendo certo nível de suposições (SIEGEL; CASTELLAN JR., 2006).

Para poder realizar um teste estatístico paramétrico, por exemplo, o teste  $t$ , que compara se as médias de dois grupos são iguais, os autores Siegel e Castellan Jr. (2006) apontam quatro suposições que devem ser verdadeiras para as observações:

- a) serem independentes;
- b) serem extraídas de populações com distribuição normal;
- c) possuírem a mesma variância;
- d) terem sido medidas pelo menos em escala intervalar.

Segundo esses autores, se essas suposições não forem satisfeitas, não é possível utilizar um teste paramétrico; no entanto, o teorema central do limite afirma que, apesar da população não ser normal, na medida em que aumenta o tamanho da amostra, a distribuição da média torna-se normal (MOORE et al., 2006; BRUNI, 2007). Então a afirmação (b) não é uma condição que precisa ser verdadeira, e os testes paramétricos podem ser utilizados, pois, apesar da distribuição da população não ser normal, basta aumentar o tamanho da amostra para que ela se torne normal.

Por questões de padronização na execução dos testes, foi seguido o seguinte procedimento geral, baseado em Siegel e Castellan Jr. (2006) e em Malhotra (2012):

- 1) Estabelecer a hipótese nula  $H_0$  e sua alternativa  $H_1$ .
- 2) Escolher um teste estatístico para testar  $H_0$ .
- 3) Especificar o tamanho da amostra  $n$ .
- 4) Coletar os dados.
- 5) Calcular a estatística do teste selecionado, supondo que  $H_0$  é verdadeira

e estimar a probabilidade valor-p de obter um valor mais extremo do que a estatística calculada.

- 6) Expressar a decisão estatística em termos do problema de pesquisa, de acordo com a análise do valor-p.

Seguem agora as descrições dos testes utilizados nesta dissertação, começando pela explicação do teorema central do limite.

### 3.5.1 Teorema central do limite

O teorema central do limite (TCL) impacta na escolha do teste, se paramétrico ou não paramétrico; no entanto, "Função Paramétrica Populacional é uma característica numérica qualquer da população" (BOLFARINE; BUSSAB, 2005, p. 38), ou seja, é um cálculo efetuado com os dados da população. Também é conhecido como parâmetro populacional e é representado genericamente como  $\theta(D)$ , onde  $D$  representa um conjunto de dados da população. A média ( $\mu$ ), o total ( $\tau$ ) e a variância ( $\sigma^2$ ) são alguns exemplos de parâmetros populacionais.

Estatística é um cálculo efetuado sobre uma amostra. São exemplos de estatísticas a média ( $\bar{x}$ ), o total ( $T$ ) e a variância ( $s^2$ ), calculados a partir de uma amostra. Elas são estimativas dos respectivos parâmetros populacionais sendo que, para cada amostra, possuem um valor diferente.

A lei dos grandes números diz que "À medida que o número de observações aumenta, a média  $\bar{x}$  dos valores observados aproxima-se cada vez mais da média  $\mu$  da população." (MOORE et al, 2006, p. 239). Assim, se forem extraídas várias amostras da população e se for calculada sua média e desvio padrão, a média da distribuição amostral das médias é  $\mu$  e o seu desvio padrão é  $\sigma/\sqrt{n}$ ; então, o TCL diz que, à medida em que se aumenta o tamanho da amostra, a distribuição amostral das médias se aproxima da distribuição normal com média  $\mu$  e desvio padrão é  $\sigma/\sqrt{n}$ .

Dizendo de outra forma, se existirem várias populações assimétricas, como a distribuição de renda de várias cidades, na medida em que se aumenta a quantidade de cidades, a distribuição média das rendas tende para a normalidade.

Como consequência, o TCL torna possível utilizar testes paramétricos para comprovar as hipóteses desta pesquisa, independente da distribuição ser normal ou

não. Por exemplo, para grandes amostras, com  $n \geq 40$ , o teste z pode ser utilizado mesmo que a distribuição seja assimétrica, ou, então,

para valores grandes do tamanho da amostra,  $n$  maior ou igual a 30, a distribuição das médias amostrais se comporta como uma distribuição normal, com média igual à média populacional e desvio padrão igual desvio padrão da variável original dividido pela raiz do tamanho da amostra. (BRUNI, 2007, p. 177)

A variável z tem distribuição normal padronizada com média zero e variância um. Por meio das tabelas de curva normal padronizada ou de um programa estatístico, é possível calcular o valor-p, que é a probabilidade da média amostral ser igual à média sob teste.

Já para um teste estatístico não paramétrico, as condições são mais gerais e não existe nenhuma condição sobre a forma da distribuição populacional, por isso esses testes são chamados de "livres de distribuição" (SIEGEL; CASTELLAN JR., 2006, p. 19). Tan, Steinbach, Kumar, (2009) também reforçam que nesta abordagem não é necessário ter conhecimentos a priori sobre a distribuição de probabilidade.

Vose (2008) também explica a diferença entre as distribuições paramétrica e não paramétrica. A paramétrica é baseada em um modelo matemático, por exemplo, um modelo exponencial. Já a não paramétrica, ou distribuição empírica, é definida pela sua forma e, então, busca-se criar um modelo matemático a partir de seus dados empíricos; o exemplo dado por esse autor é a distribuição triangular.

### 3.5.2 Teste de Kolmogorov-Smirnov

Para verificar se duas variáveis aleatórias possuem a mesma distribuição de probabilidade, utiliza-se um teste de aderência. Nesta dissertação, será utilizado o teste de Kolmogorov-Smirnov (K-S) porque ele é mais poderoso que o teste alternativo (QuiQuadrado).

O teste de Kolmogorov-Smirnov admite que a distribuição da variável subjacente que está sendo testada é contínua, como especificado pela distribuição de frequências acumuladas. Assim, o teste é apropriado para testar a aderência para variáveis que são medidas pelo menos em uma escala ordinal. (SIEGEL; CASTELLAN JR., 2006, p. 71)

O teste K-S será utilizado para verificar se a distribuição da duração total do projeto resultante da simulação triangular e da simulação *bootstrap* são iguais.

### 3.5.3 Testes de correlação

Para verificar o relacionamento entre duas variáveis, é calculado o coeficiente de associação entre elas, e também é feito um teste para verificar se esse coeficiente é significativo. Neste texto serão descritas três formas de cálculo e teste do coeficiente de associação, sendo um paramétrico e dois não paramétricos. Assim, serão descritos o coeficiente de correlação linear de Pearson  $r$  (paramétrico), o coeficiente de correlação posto ordem de Spearman  $\rho$  (não paramétrico) e o Coeficiente  $\tau$  de correlação posto ordem de Kendall (não paramétrico).

Torna-se relevante descrever esses coeficientes porque Vanhoucke (2009) os utiliza para calcular os índices de crucialidade do cronograma do projeto. No entanto, é válido ressaltar que uma coisa é o coeficiente e outra é o teste de hipótese, que diz se o coeficiente é significativo ou não. Vanhoucke (2009) somente calcula os coeficientes de correlação, nas três formas, mas não realiza um teste de hipótese para validá-los.

O coeficiente de correlação linear de Pearson ( $r$ ) é utilizado para avaliar a relação linear entre duas variáveis (BRUNI, 2013; BECKER, 2015). Para Bruni (2013), há algumas hipóteses que devem ser consideradas no cálculo desse coeficiente:

- a) para cada valor de  $x$ , haverá possíveis valores para  $y$ ;
- b) a variável  $y$  é aleatória;
- c) para cada valor de  $x$ , há uma distribuição condicional de  $y$  que é normal;
- d) os desvios padrões de todas as distribuições condicionais são iguais.

Este coeficiente não é livre de distribuição, pois assume-se que as medidas subjacentes sejam normais, além disso, as variáveis relacionadas devem ser medidas no nível intervalar ou razão. Por essas questões, este é um teste paramétrico e, portanto, para poder utilizá-lo é necessário atender aos pressupostos.

O coeficiente de correlação posto ordem de Spearman ( $\rho$ ) mede a correlação entre duas variáveis medidas no mínimo na escala ordinal (SIEGEL; CASTELLAN JR., 2006). Então, segundo Becker (2015), se as duas variáveis forem correlacionadas, o coeficiente de correlação estará próximo a um ( $\rho \approx 1$ ) e se forem inversamente correlacionadas, próximo a menos um ( $\rho \approx -1$ ).

Basicamente atribui-se postos às variáveis  $x$  e  $y$  e verifica-se se existe correlação entre os postos, calculando a diferença ao quadrado entre os postos (SIEGEL; CASTELLAN JR., 2006).

O coeficiente de correlação posto ordem de Kendall ( $\tau$ ) mede a correlação entre duas variáveis e foi criado como uma alternativa ao coeficiente de Spearman. Ele é calculado por meio da diferença entre concordâncias e discordâncias nas ordenações das variáveis, normalizando-se pelo total de comparações possíveis (BECKER, 2015). Siegel & Castellan Jr. (2006) afirmam que o nível de medição mínimo para o cálculo do coeficiente é ordinal para ambas as variáveis.

Os coeficientes de correlação de Spearman e de Kendall não possuem a mesma escala base, portanto seus valores não são comparáveis. Outra diferença que existe é na interpretação das medidas. O coeficiente de Spearman tem a mesma interpretação do coeficiente de correlação de Pearson, ou seja, mede o grau de associação entre as duas variáveis. Já o coeficiente de Kendall mede a diferença entre a probabilidade de  $X$  e  $Y$  estarem na mesma ordem; ambos, no entanto, possuem a mesma habilidade em rejeitar a  $H_0$  (SIEGEL; CASTELLAN JR., 2006).

#### 3.5.4 Protocolos de análise

Para demonstrar que a análise de risco pode ser feita com a distribuição *bootstrap*, será feita a análise de correlação entre os resultados fornecidos pela distribuição *bootstrap* e pelos fornecidos pela distribuição triangular. Se os resultados forem correlacionados, as distribuições fornecem resultados compatíveis e, então, a *bootstrap* pode ser utilizada na análise de riscos.

Cada projeto da amostra possui 30 atividades, que servirão de base para o cálculo dos índices de sensibilidade utilizando a distribuição triangular e as duas modalidades da distribuição *bootstrap* por meio da simulação Monte Carlo. Então, a simulação feita com a distribuição triangular será comparada com a *bootstrap* padrão e a *bootstrap-rate*. Cada simulação resulta em 5.000 durações totais do projeto, atendendo aos requisitos do TCL que é ter uma amostra grande ( $n > 30$ ). Os testes para validar a hipótese da pesquisa estão descritos nos protocolos abaixo.

O protocolo P1 verifica a igualdade entre duração total do projeto (DTP) calculada pelo método do caminho crítico com as simulações efetuadas pelas três distribuições.

- **P1: Teste de igualdade da média:** as durações totais do projeto simuladas ( $DTP^s$ ) podem ser consideradas iguais à duração total do projeto (DTP)?

1)  $H_0: DTP^s = DTP$

$H_1: DTP^s \neq DTP$

2) Teste estatístico: teste z

3) Tamanho da amostra:  $nrS = 5.000$

4) Coleta dos dados: simulação

5) Calcular a estatística e estimar o *valor-p*

6) Decisão: quanto menor o valor-p, menos chances existem da DTP simulada ser igual à DTP do cronograma.

O protocolo P2 testa a correlação entre os valores-p do protocolo P1 da distribuição TR com a B-1 e, também, da TR com a B-R.

- **P2: Teste de correlação dos valores-p entre duas simulações:** existe correlação entre os valores-p da distribuição triangular e da *bootstrap*?

1)  $H_0$ : não existe correlação

$H_1$ : existe correlação

2) Teste estatístico: teste de correlação linear de Pearson

3) Tamanho da amostra:  $nrS = 5.000$

4) Coleta dos dados: simulação

5) Calcular a estatística e estimar o *valor-p*

6) Decisão: quanto menor o valor-p, mais chances existem de as duas distribuições serem correlacionadas.

O protocolo P3 verifica a igualdade entre as distribuições da duração total do projeto (DTP) simulada utilizando a distribuição triangular e com as simuladas com a distribuição *bootstrap*.

- **P3: Teste de igualdade entre duas distribuições:** as durações totais do projeto simuladas ( $DTP^s$ ) com a distribuição TR x B-1 e TR x B-R são as mesmas?

1)  $H_0: TR.DTP^s = BOOT.DTP^s$

$H_1: TR.DTP^s \neq BOOT.DTP^s$

2) Teste estatístico: Kolmogorov-Smirnov

3) Tamanho da amostra:  $nrS = 5.000$

- 4) Coleta dos dados: simulação
- 5) Calcular a estatística e estimar o *valor-p*
- 6) Decisão: quanto menor *valor-p*, menos chances existem de as duas distribuições serem iguais.

O protocolo P4 testa a correlação entre os seis índices de sensibilidade calculados com a distribuição TR e com os calculados com a B-1 e, também, entre os calculados com a TR e a B-R.

- **P4: Teste de correlação entre os índices de sensibilidade**

- 1)  $H_0$ : não existe a correlação  
 $H_1$ : existe a correlação
- 2) Teste estatístico: teste de correlação de Pearson
- 3) Tamanho da amostra:  $nrS = 5.000$
- 4) Coleta dos dados: simulação
- 5) Calcular a estatística e estimar o *valor-p*
- 6) Decisão: quanto menor for o *valor-p*, maior é o indício da existência de correlação entre os índices de sensibilidade.

Todos os testes estatísticos definidos nos protocolos de análises foram realizados no ambiente estatístico **R** (R CORE TEAM, 2016).

### 3.6 ANÁLISE DE RISCOS COM A DISTRIBUIÇÃO *BOOTSTRAP-RATE*

Foi visto que a simulação *bootstrap* trabalha com reamostragem da própria amostra, utilizando amostra com reposição. Em uma amostra com reposição, o valor esperado dela é a média dos elementos da amostra (BOLFARINE; BUSSAB, 2005). Como num projeto existem vários caminhos com quantidade de atividades possivelmente diferentes, ao utilizar a amostra com reposição, as atividades do caminho topológico (folga topológica igual a zero) sempre serão as mais sensíveis do cronograma. Por causa dessa característica, as atividades sensíveis já seriam conhecidas e não seria necessário simular.

Para resolver essa questão, foi criada a distribuição *bootstrap-rate* (B-R), derivada da *bootstrap* padrão (B-1). O cálculo da B-R também é feito partir da duração das atividades do cronograma e tem os seguintes passos:

- Preparação:
  - 1) calcular a média das durações das atividades;

- 2) dividir cada duração pela média;
- Sorteio das durações:
- 3) sortear uma taxa com reposição;
- 4) multiplicar a duração da atividade  $d_i$  pela taxa;
- 5) repetir os passos 3 e 4 para todas as atividades;
- 6) calcular a duração total do projeto da simulação  $s$ ;
- 7) repetir os passos 3 a 6  $nrS=5.000$  vezes.

A distribuição *bootstrap-rate* possui as seguintes propriedades:

- a) a duração sorteada da atividade  $d_i$  é proporcional à sua duração original, pois ela é multiplicada pela taxa;
- b) a variação da duração das atividades espelha a variação das durações de todas as atividades, pois ela está embutida na taxa e esta é calculada a partir da duração de todas as atividades.

Para fins de comparação com a distribuição triangular, os cronogramas dos projetos serão simulados utilizando as duas formas da distribuição *bootstrap*.

### 3.7 DEFINIÇÃO DO INSTRUMENTO DE COLETA DE DADOS

Conforme descrito anteriormente, os projetos utilizados nesta pesquisa estão em formato Patterson e, para extrair os seus dados, foi desenvolvido um algoritmo para ler o arquivo, interpretar o seu conteúdo, conforme definição em [http://www.p2engine.com/p2reader/patterson\\_format](http://www.p2engine.com/p2reader/patterson_format), e disponibilizar os resultados. Da mesma forma, após a leitura de um projeto, a sua execução foi simulada 5.000 vezes através de algoritmos desenvolvidos. Portanto, todos os dados desta pesquisa foram coletados a partir de algoritmos desenvolvidos pelo autor na linguagem de programação Java. O Quadro 19 mostra a especificação do modelo de simulação para os cronogramas de projetos.

No formato Patterson, a duração das atividades está em dias e, nesta dissertação, elas serão convertidas para horas, utilizando uma taxa de conversão de 8h/dia. Somente horas inteiras serão aceitas. Se em um sorteio aleatório ocorrer um número menor do que um, ele será arredondado para uma hora.

O conjunto dos dados dos projetos está dividido em dois grupos. O primeiro grupo de dados são aqueles coletados a partir dos projetos, por exemplo, a sua duração e os indicadores topológicos. Os dados do segundo grupo foram coletados após a execução da simulação; são específicos do projeto e da distribuição utilizada para a simulação.

Quadro 19 – Especificação do Modelo de Simulação de Cronogramas de Projetos

<b>Objetivos</b>	<ul style="list-style-type: none"> <li>• Verificar se o cronograma do projeto tem alguma chance de ser finalizado no prazo calculado;</li> <li>• Identificar as atividades que mais influenciam o cronograma.</li> </ul>
<b>Conteúdo</b>	<ul style="list-style-type: none"> <li>• A simulação levará em conta a duração das atividades dos projetos e suas relações de dependências. Não será considerada a natureza da atividade.</li> </ul>
<b>Entradas</b>	<ul style="list-style-type: none"> <li>• Cronograma do projeto com lista de atividades, suas durações e as dependências entre elas.</li> </ul>
<b>Saídas</b>	<ul style="list-style-type: none"> <li>• A distribuição de probabilidade da duração total do projeto;</li> <li>• O resultado de um teste de hipótese, considerando como <math>H_0</math>: a duração do projeto é igual a duração calculada e <math>H_1</math>: a duração do projeto é diferente da calculada;</li> <li>• A relação de atividades críticas;</li> <li>• A(s) atividade(s) que causa(m) o maior impacto no cronograma.</li> </ul>
<b>Premissas</b>	<ul style="list-style-type: none"> <li>• Uma atividade só pode iniciar após todas as suas predecessoras finalizarem;</li> <li>• A atividade iniciará exatamente depois de suas predecessoras finalizarem;</li> <li>• Será considerado que as durações das atividades informadas pelo gerente do projeto estão corretas.</li> </ul>
<b>Definições</b>	<ul style="list-style-type: none"> <li>• Serão executadas 5.000 iterações.</li> <li>• Cada iteração gerará aleatoriamente a duração para as atividades e, então, será calculada a duração do projeto, serão identificadas as atividades críticas e calculados os índices de sensibilidade do tempo.</li> </ul>

Fonte: O autor (2017), baseado em CHWIF; MEDINA (2015).

Os projetos são compostos de 32 atividades, sendo duas com duração zero, que são o marco inicial e o marco final. As análises foram feitas para as atividades com duração diferente de zero, então, a simulação gerou valores de duração para cada atividade, o que permitiu calcular os índices de sensibilidade de tempo para cada uma das 30 atividades do cronograma.

Na sequência, foram aplicados testes de hipótese para verificar se existia correlação no cálculo dos índices feito com a distribuição triangular e os feitos com a distribuição *bootstrap*.

### 3.7.1 Simulação utilizando a distribuição triangular

Conforme explicado anteriormente, a distribuição triangular foi escolhida para comparar com a *bootstrap* a fim de manter a mesma distribuição utilizada por Vanhoucke (2009) e ter uma base compatível de comparação. Essa distribuição também é recomendada pelo Instituto de Gerenciamento de Projeto (PMI, 2013).

A distribuição triangular é caracterizada por três parâmetros: o valor mínimo, o mais provável e o máximo. Esses valores foram definidos com base em Vanhoucke (2009), conforme mostrado no Quadro 20.

Quadro 20 – Valores utilizados na distribuição triangular

Parâmetro	Valor
Mínimo	duração X 0,1
Mais provável	duração
Máximo	duração X 1,9

Fonte: O autor (2017), baseado em VANHOUCKE (2009).

Sendo que duração representa a duração mais provável da atividade e a duração mínima e a máxima são calculas com base neste valor.

### 3.7.2 Simulação utilizando a distribuição *bootstrap*

A utilização da distribuição *bootstrap* foi feita de duas formas: na primeira, chamada de B-1, a atribuição da duração de cada atividade foi feita através de sorteio aleatório com reposição, a partir da amostra da duração de todas as atividades; na segunda, B-R, a duração cada atividade foi dividida pela média e gerado um índice de variação. No cálculo da duração de uma atividade, foi sorteado um índice (com reposição) e multiplicado pela duração da atividade. Isso foi feito para que a duração simulada da atividade dependa da sua própria duração e da variabilidade da duração das atividades em torno da média.

O Quadro 21 mostra um resumo da forma que a duração das atividades foi atribuída utilizando a distribuição *bootstrap* B-1 e B-R.

Quadro 21 – Sorteio das durações das atividades

<i>Bootstrap</i>	Duração de uma atividade
<b>B-1</b>	Sorteio aleatório com reposição a partir da duração de todas as atividades.
<b>B-R</b>	1º: Dividir a duração de cada atividade pela média das durações; 2º: Sortear um índice aleatoriamente com reposição; 3º: Multiplicar a duração da atividade pelo índice sorteado.

Fonte: O autor (2017).

### 3.7.3 Coleta e tratamento dos dados

Esta pesquisa foi executada, de modo abrangente, com as etapas de desenvolvimento dos algoritmos, coleta e organização dos dados e análise. O protocolo seguido para execução dessas etapas está descrito no Quadro 22.

Quadro 22 – Etapas do protocolo da pesquisa

<b>Etapa</b>	<b>Objetivo</b>	<b>Materiais</b>
<b>1</b> Desenvolvimento dos algoritmos	Desenvolver algoritmos para leitura, simulação e coleta dos dados dos projetos.	Fundamentação teórica e linguagem Java.
<b>2</b> Teste da coleta de dados	Calcular quanto tempo os algoritmos levam para serem executados, para montar uma estratégia de coleta dos dados.	Linguagem Java.
<b>3</b> Correções e otimização dos algoritmos	Corrigir e otimizar os algoritmos para que eles levem menos tempo para serem executados.	Fundamentação teórica e linguagem Java.
<b>4</b> Definição da amostra	Definir como os projetos serão alocados na amostra e sortear os projetos participantes.	Fundamentação teórica e linguagem Java.
<b>5</b> Agendamento da coleta de dados	Dividir a coleta dos dados da amostra dos projetos em grupos e agendar cada grupo.	Linguagem Java.
<b>6</b> Coleta de dados	Coletar os dados da amostra de projetos.	Linguagem Java e ambiente estatístico <b>R</b> .
<b>7</b> Análise dos dados	Desenvolver algoritmos para aplicar testes de hipóteses aos dados.	Ambiente estatístico <b>R</b> .

Fonte: O autor (2017), baseado em GUIMARÃES (2014).

Os algoritmos desenvolvidos calculam o caminho crítico do projeto e sua duração, bem como os indicadores topológicos. Também foram desenvolvidos algoritmos para simular a execução do projeto e calcular os índices de sensibilidade.

Após o desenvolvimento dos algoritmos, as etapas 2 e 3 foram repetidas até que não houvesse mais erros e o tempo de execução fosse otimizado. Passou-se, então, à fase de pré-testes.

## 3.8 PRÉ-TESTE DO INSTRUMENTO DE COLETA DE DADOS

O pré-teste dos instrumentos de coleta de dados foi feito em duas fases. A fase um era a de pré-qualificação, na qual a decisão final foi coletar uma amostra de 410 projetos, num tempo de 12 dias.

Na fase dois, pós-qualificação, com os resultados da fase um e a utilização de outros métodos para calcular a rede do projeto, chegou-se à conclusão que

poderiam ser utilizados todos os 4.100 projetos da amostra original de Vanhoucke (2009), com um tempo total de execução da simulação de uma hora e quinze minutos.

### 3.8.1 Fase um

Na fase um, o cálculo da rede foi feito utilizando a linguagem de programação **R**, seguindo a definição do cálculo do caminho crítico. Neste formato, descobrem-se todos os caminhos e calcula-se a duração de cada caminho somando as durações das atividades pertencentes ao caminho. O caminho com a maior duração é o crítico.

Foram executadas algumas simulações, com variações nos parâmetros das simulações. Os dados da execução de testes são mostrados nas duas primeiras linhas do Quadro 23.

Quadro 23 – Coleta dos dados de testes

Linha	Dados	Núm. Projs	Núm. Distrs.	Núm. Simul.	Tempo (h)	Tempo
1	Teste	4.100	1	50	8,635	0,36 dia
2	Teste	4.100	7	15	19,597	0,82 dia
3	Projeção	4.100	7	10.000	13.064,33	1,5 ano
4	Projeção	410	2	5.000	186,633	7,8 dias
5	Projeção	410	3	5.000	279,95	11,7 dias
6	Projeção	34	3	5.000	23,215	0,97 dia
7	Projeção	36	3	5.000	24,581	1,02 dia
<b>8</b>	<b>Real</b>	<b>4.100</b>	<b>3</b>	<b>5.000</b>	<b>1,25</b>	<b>0,05 dia</b>

Fonte: O autor (2017).

A linha um do Quadro 23 mostra o primeiro teste efetuado. Para esse teste, além de terem sido considerados os 4.100 projetos, foi também considerada uma distribuição para a simulação Monte Carlo, e foram executadas 50 simulações para cada projeto. O tempo total foi de 8,635 horas.

Como a ideia inicial deste projeto era comparar a distribuição *bootstrap* com outras seis distribuições, totalizando sete distribuições para cada projeto, esse é o cenário da linha dois: os 4.100 projetos foram testados com sete distribuições, sendo que cada distribuição foi simulada 15 vezes, totalizando 19,597 horas.

A partir dos dados desse segundo teste, foram feitas projeções para os outros cenários. Avaliando os 4.100 projetos, com sete distribuições, executando

10.000 simulações, o tempo total de execução somaria um ano e meio, o que foi considerado inviável para essa pesquisa.

Em seguida, foi decidido tirar uma amostra de 410 projetos, e, para cada projeto, avaliar duas distribuições, a triangular (TR) e as duas modalidades de *bootstrap* (B-1 e B-R), cada uma delas em metade dos projetos. Considerando 5.000 simulações, o tempo total chegou a 7,8 dias. Finalmente, chegou-se ao cenário final desta pesquisa, na fase um: aplicar as três distribuições nas simulações (a triangular e as duas modalidades da *bootstrap*), chegando a um total de 280 horas (11,7 dias).

### 3.8.2 Fase dois

Na fase dois, após a qualificação da dissertação, buscaram-se novas formas para tornar o cálculo da rede projeto mais rápido e ser possível o aumento da amostra.

Segundo Goodrich, Tamassia e Goldwasser (2014), o diagrama de rede do projeto é um grafo, composto de arestas e vértices. Os vértices são o conjunto de atividades e as arestas são as ligações entre as atividades. Para o cálculo do caminho crítico, em última instância, utiliza-se do caminhamento em profundidade em um grafo, e, à medida que se percorre os vértices e arestas, analisa-se o caminho do projeto e calcula-se o início e término das atividades. Esse processo gera várias combinações, através da utilização de várias recorrências, e ele é executado repetidamente, a cada iteração da simulação, aumentando o tempo de cálculo da rede do projeto.

Para resolver esse problema, foi utilizada a ordenação topológica, conforme orientam esses autores. Essa técnica ordena todas as ligações entre as atividades de tal forma que as atividades respeitam todas as suas precedências. O resultado da ordenação topológica é uma lista das ligações (arestas) entre as atividades ordenadas.

A ordenação topológica é feita uma vez para cada projeto, antes da simulação. Na simulação dos projetos, que ocorrem 5.000 vezes, a lista das ligações é percorrida de maneira sequencial para calcular a rede do projeto.

Essa mudança permitiu simular os 4.100 projetos, com as três distribuições, 5.000 vezes, totalizando 61,5 milhões de execuções de projetos em menos de duas horas. Por utilizar essa estratégia foi possível incluir na amostra final todos os 4.100 projetos.

## 4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Conforme apresentado no capítulo anterior, nesta pesquisa foram analisados 4.100 projetos fictícios; então, por economia de espaço, este tópico começa com a descrição de apenas um projeto individual, considerando a visão de um gerente de projetos. Na sequência, o conjunto dos projetos será analisado e descrito como um todo, considerando os resultados das simulações feitas com a distribuição triangular, *bootstrap* padrão e *bootstrap-rate*.

### 4.1 APRESENTAÇÃO DE UM PROJETO

Do cronograma do projeto é calculada a duração total do projeto  $DTP$ . Assim, a partir dos cenários da simulação, foi possível calcular as seguintes estatísticas:

- A probabilidade da duração total do projeto cumprir o prazo, que é a quantidade de vezes em que  $DTP_s \leq DTP$ ;
- Duração total do projeto média  $\overline{DTP}$ ;
- Intervalo de confiança da duração total do projeto média  $\overline{DTP}$ ;
- O percentil 80 da duração total do projeto dos cenários  $DTP_{p80}$ ;
- A variação da  $DTP$  para se chegar à  $DTP_{p80}$ .

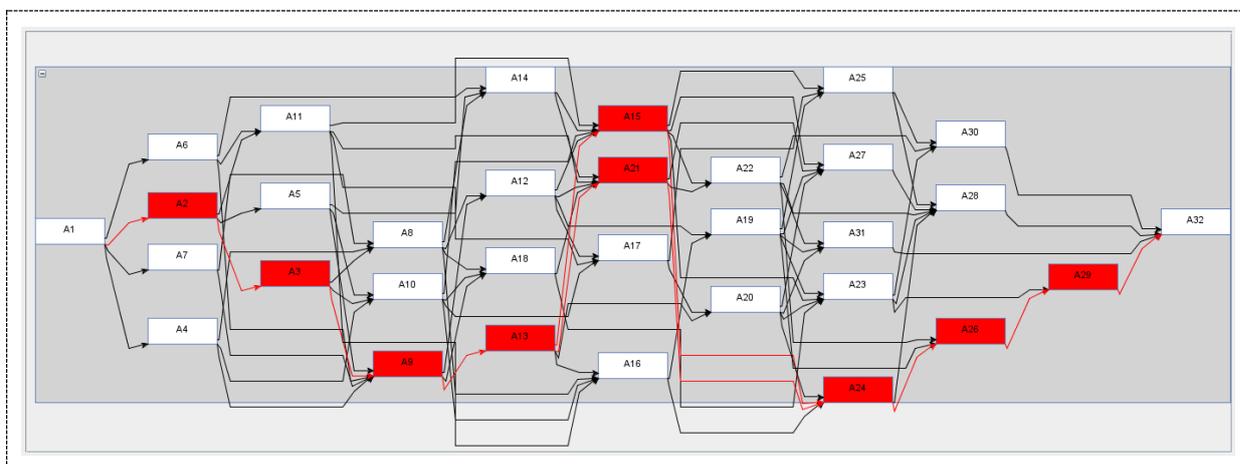
Essa análise foi elaborada a partir da visão de um gerente de projeto, como se ele tivesse feito um relatório a partir dos dados e dos diagramas do projeto. O projeto a ser estudado é o P0210, com nome de arquivo P0210.prj. A seguir são apresentadas as principais características desse projeto:

- Data de início: 02/01/2017.
- Data de término: 23/06/2017.
- Duração: 125 dias úteis, ou seja, 1.000 horas, desde o início até o fim.
- Esforço necessário: 2.504 homem/hora.
- Quantidade de atividades críticas: 9.
- Atividades críticas: A2, A3, A9, A13, A15, A21, A24, A26 e A29.

O diagrama de rede do projeto (Figura 19) mostra que existem atividades de execução em paralelo e sequencial e que essa estrutura influencia a distribuição das atividades no projeto. As atividades em vermelho (as mais escuras) são as críticas;

as em branco, as não críticas. A estrutura da rede pode ser analisada pelos indicadores topológicos:

Figura 19 – Diagrama de rede do projeto P0210



Fonte: O autor (2017).

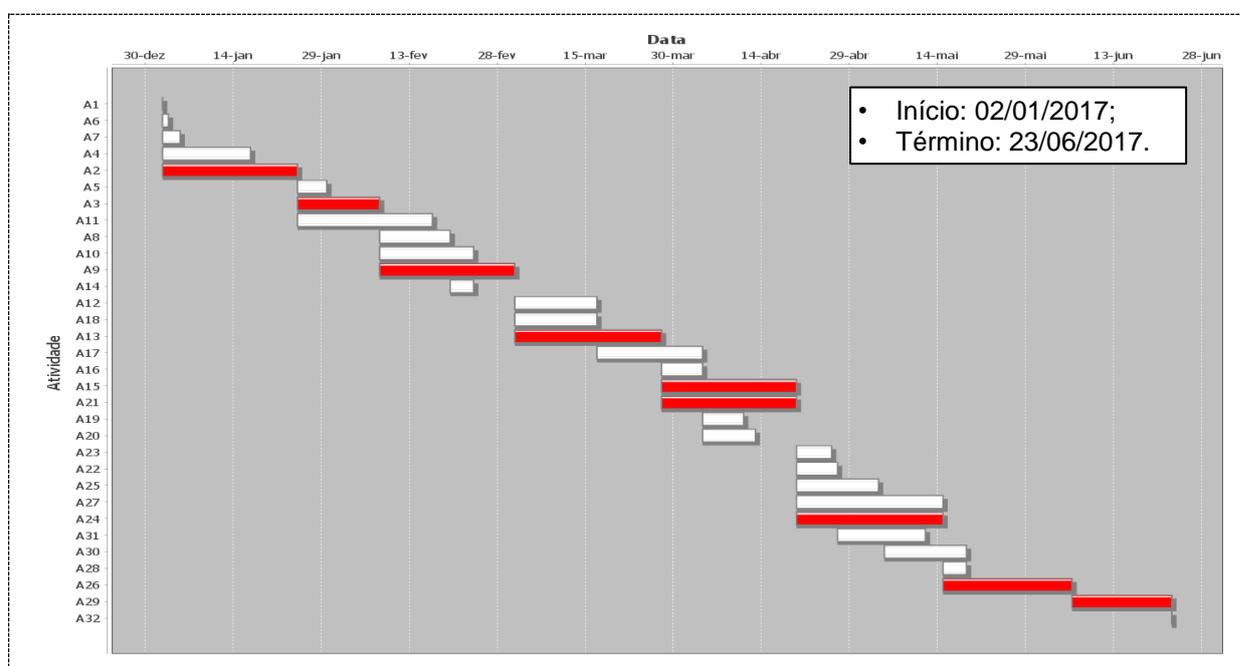
- Indicador sequencial/paralelo,  $SP=0,276$ : indica que a rede está mais próxima de ser totalmente paralela do que sequencial. Isso é confirmado, também, pelo baixo percentual de 30% de atividades críticas;
- Indicador de distribuição das atividades,  $DA=0,196$ : as atividades desse projeto se distribuem de maneira aproximadamente uniforme, conforme observado na Figura 19 e no valor do indicador;
- Indicador de comprimento dos arcos,  $CA=0,294$ : um arco representa a ligação entre uma atividade e a próxima, e, quando essas atividades não são de níveis adjacentes, o comprimento do arco é maior que um. Como exemplo, as ligações  $(A6) \rightarrow (A14)$ ,  $(A11) \rightarrow (A15)$ . O valor desse indicador próximo a zero indica que as ligações que mais predominam são entre atividades de níveis diferente;
- Indicador de folga topológica,  $FT=0,125$ : por esse indicador, a rede é considerada densa, pois as atividades não podem mudar de nível sem que se aumente o nível máximo, pois apesar de existirem ligações em níveis distantes, também existem ligações entre níveis adjacentes.

Resumindo, a partir dos indicadores topológicos, é possível dizer que esse projeto executa várias atividades paralelas e que seu caminho crítico é curto em relação ao total de atividades, as quais são regularmente distribuídas ao longo da

rede. Existem várias ligações entre atividades de níveis distantes, mas elas não podem se deslocar por causa da folga topológica ser baixa.

O diagrama de rede permite analisar a estrutura do projeto, já o gráfico de Gantt permite identificar a distribuição das atividades ao longo do tempo (Figura 20). Nele é possível observar as atividades críticas marcadas em vermelho (as mais escuras); e, se qualquer uma delas atrasar, o projeto atrasa como um todo. O gráfico de Gantt respeita totalmente as relações de precedências indicadas pelo diagrama de rede.

Figura 20 – Gráfico de Gantt do projeto P0210



Fonte: O autor (2017).

As barras paralelas no gráfico de Gantt indicam que várias atividades serão executadas em paralelo e, neste caso, serão necessárias várias equipes, chegando até a ter cinco equipes em paralelo.

Todas as atividades do projeto são importantes, mas muitas vezes não existem recursos para colocar a melhor equipe em todas elas, então é preciso escolher quais delas receberão os melhores recursos. Uma forma natural é atribuir os melhores recursos às atividades críticas, mas essa pode não ser a melhor decisão.

O esforço total desse projeto é de 2.504 horas, sendo que a duração determinada pelo caminho crítico é 1.000 horas, ou seja, existem 1.504 horas do

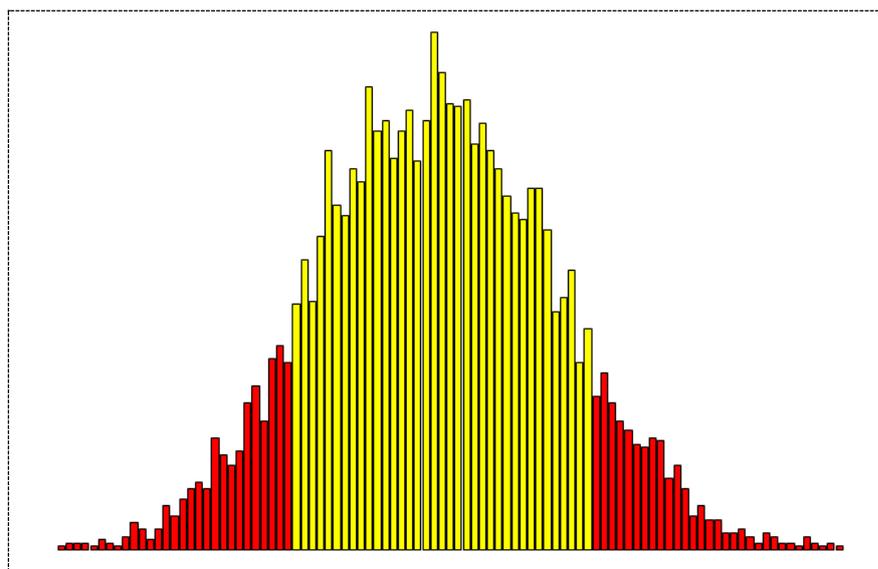
projeto que não pertencem ao caminho crítico, então, 60% do esforço total não são críticos. Focar os melhores recursos no caminho crítico significa que em 60% do total do projeto não serão alocados os melhores recursos, então é necessário priorizar as atividades do projeto não só considerando o caminho crítico. E é nesses casos que utilizar a simulação Monte Carlo pode ajudar.

Com a simulação Monte Carlo é possível ter uma distribuição da duração do projeto e, então, calcular vários índices para identificar quais as atividades mais influentes na duração do projeto. Além disso, também é possível prover estimativas da duração do projeto em termos de intervalo.

As principais informações no planejamento de um projeto são fornecidas a seguir, calculadas através da simulação Monte Carlo e utilizando a distribuição triangular:

- Data de início: 02/01/2017;
- Data de término média: 03/07/2017;
- Duração média: 131 dias úteis, ou 1.047 horas, desde o início até o fim;
- Data de término no décimo percentil: 07/06/2017;
- Data de término no nonagésimo percentil: 28/07/2017;
- Provável data de término: entre 07/06/2017 e 28/07/2017.

Figura 21 – Distribuição da duração total do projeto P0210



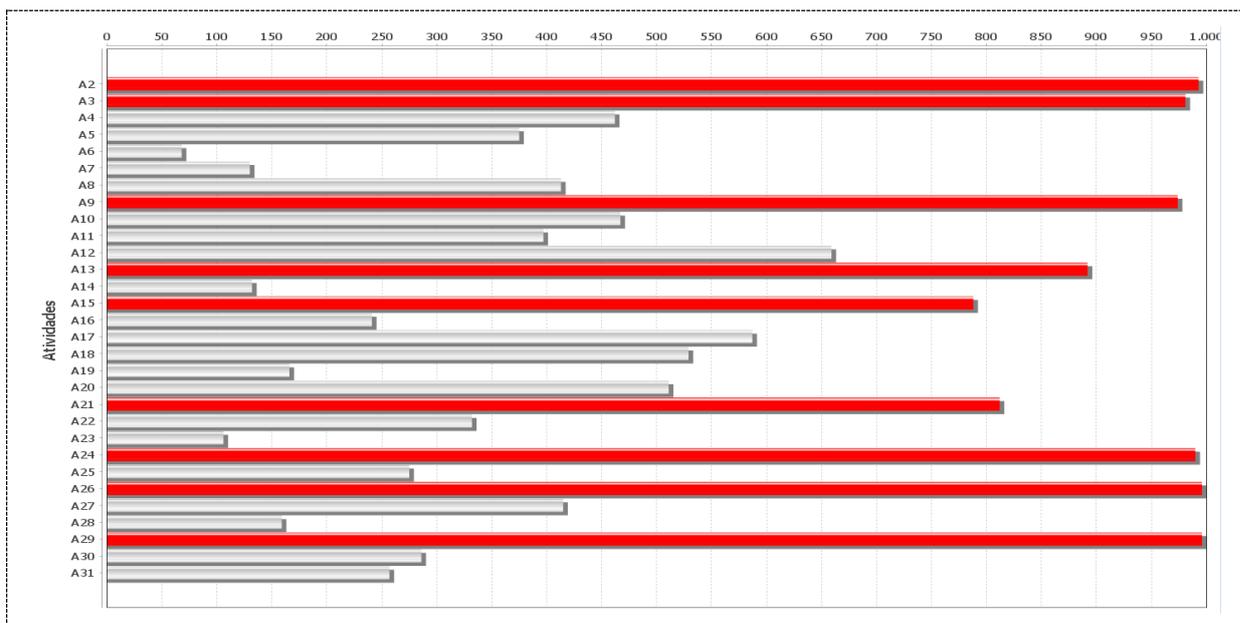
Fonte: O autor (2017).

A Figura 21 mostra a distribuição de probabilidade da duração total do projeto para o projeto P0210. Na faixa central amarela (cor mais clara) é possível

observar os intervalados de datas dos percentis 10 e 90, que correspondem às datas 07/06/2017 e 28/07/2017. O percentil 10 indica que o projeto tem uma probabilidade de 10% de ser finalizando até a data 07/06/2017; da mesma forma, o percentil 90 indica que o projeto tem 90% de chance de ser finalizado até 28/07/2017. Dito de outra forma, o projeto tem probabilidade de 80% de finalizar entre essas duas datas.

Considerando a duração total do projeto que foi estimada em 125 dias, planejada para finalizar em 23/06/2017, o projeto tem uma probabilidade de 36% de cumprir este prazo. É melhor que a estimativa seja um intervalo, mas, se não for possível, pode-se definir estimativa com maior probabilidade de cumpri-la. Neste caso foi escolhido o percentil 90, mas outro valor poderia ser escolhido, dependendo dos riscos que se quer aceitar.

Figura 22 – Índice de significância das atividades do projeto P0210



Fonte: O autor (2017).

Com a simulação, é possível calcular os índices de sensibilidade do projeto para cada atividade do cronograma, colocando-as em ordem de importância. Como exemplo, a Figura 22 indica o índice de significância para as atividades do projeto escolhido, por ordem do número da atividade, indicando em vermelho (cor escura) as atividades críticas. O Quadro 24 mostra a lista das atividades por ordem de importância, utilizando como critério o índice de significância.

Quadro 24 – Índices de significância de ordenados por importância

Ordem	Ativ.	Valor	Ordem	Ativ.	Valor
<b>1</b>	<b>a26</b>	<b>996</b>	16	a27	415
<b>2</b>	<b>a29</b>	<b>996</b>	17	a8	413
<b>3</b>	<b>a2</b>	<b>993</b>	18	a11	397
<b>4</b>	<b>a24</b>	<b>990</b>	19	a5	375
<b>5</b>	<b>a3</b>	<b>981</b>	20	a22	332
<b>6</b>	<b>a9</b>	<b>974</b>	21	a30	286
<b>7</b>	<b>a13</b>	<b>892</b>	22	a25	275
<b>8</b>	<b>a21</b>	<b>812</b>	23	a31	257
<b>9</b>	<b>a15</b>	<b>788</b>	24	a16	241
10	a12	659	25	a19	166
11	a17	587	26	a28	159
12	a18	529	27	a14	132
13	a20	511	28	a7	130
14	a10	467	29	a23	106
15	a4	462	30	a6	68

Fonte: O autor (2017).

As atividades críticas do projeto são as primeiras nove atividades mais importantes mostradas no Quadro 24; e isso ocorre também para os outros índices de sensibilidade. Com os índices é possível priorizar as atividades por importância e não apenas informar se a atividade é crítica ou não. Por exemplo, as atividades 12, 17, 18 e 20 também são importantes além das críticas.

Para fins de comparação, a Figura 23 mostra os seis índices de sensibilidade, com as atividades críticas destacadas em vermelho (cor escura). É possível observar nessa figura que os índices priorizam as mesmas atividades, apesar dos valores dos índices serem diferentes. Entre outras palavras, os índices não apresentam os mesmos valores, mas são correlacionados.

Figura 23 – Gráfico dos índices de sensibilidade para as atividades



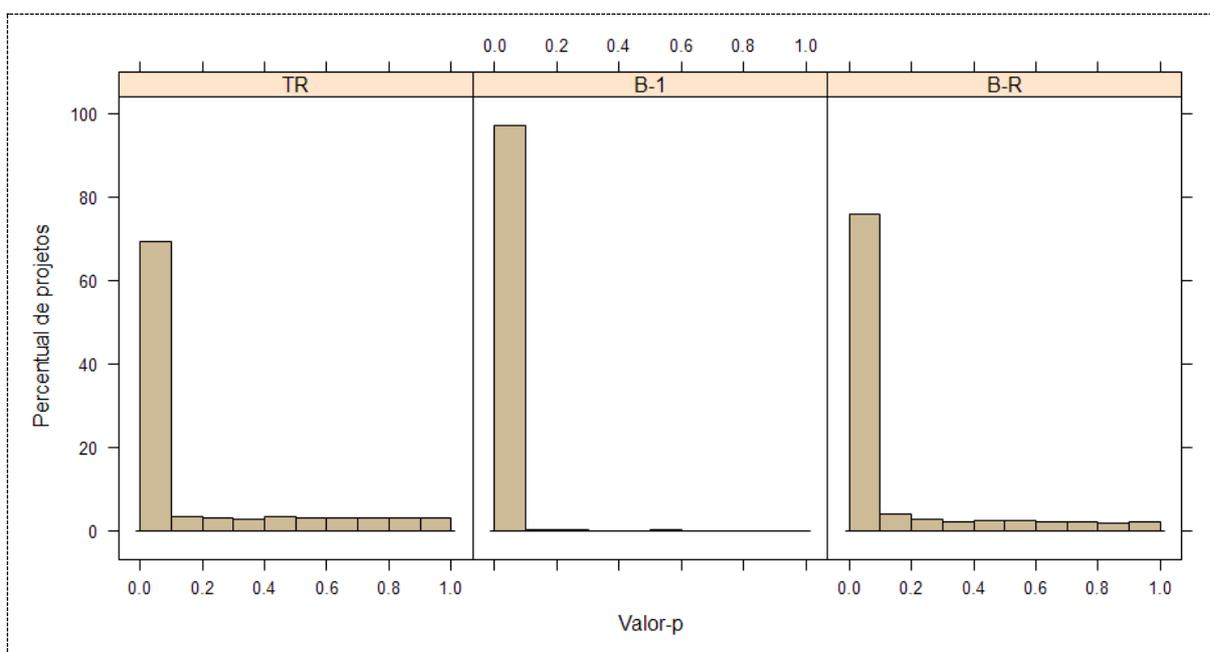
Fonte: O autor (2017).

Após a análise dos dados de um projeto, foram analisados os dados para a totalidade dos projetos da amostra. Como o objetivo desta pesquisa é demonstrar que a distribuição *bootstrap* pode ser utilizada na análise dos riscos em cronograma de projetos, os próximos tópicos comparam os resultados obtidos com a simulação Monte Carlo utilizando a distribuição triangular (TR), a *bootstrap* padrão (B-1) e a *bootstrap-rate* (B-R); além disso, esses resultados são comparados entre si considerando os protocolos definidos.

## 4.2 IGUALDADE DA DURAÇÃO PLANEJADA

Para verificar se a duração total do projeto planejada a partir da linha de base do projeto pode ser considerada igual à duração total do projeto simulada, ou seja, se  $DTP \cong DTP^S$ , foi utilizado o teste z, conforme definido no protocolo P1. A hipótese nula desse teste é que a simulação tem média igual a DTP e, a alternativa, indica que elas são diferentes. O protocolo P1 foi aplicado nos 4.100 projetos, considerando as três distribuições de probabilidade: triangular (TR), *bootstrap* padrão (B-1) e *bootstrap-rate* (B-R). O resultado é um conjunto de 4.100 valores-p para cada distribuição de probabilidade, resumidos na Figura 24.

Figura 24 – Proporção de igualdade entre  $DTP \cong DTP^S$



Fonte: O autor (2017).

Para analisar o gráfico, foi considerado um limite hipotético de 0,1, o que indica que projetos com o valor-p maior que esse limite possuem média simulada igual à da linha de base. Com esse limite, as distribuições triangular (TR), *bootstrap* padrão (B-1) e a *bootstrap-rate* (B-R) apresentam 31%, 3% e 24% de projetos com as durações iguais. Isso destaca que nas três distribuições a quantidade de diferenças das durações é maior do que a da igualdade.

Outro destaque é a semelhança entre as distribuições TR e B-R e a diferença da distribuição TR e B-1, que apresenta quase 100% de diferença. Isso ocorre porque, na simulação com a B-1, o caminho crítico é aquele com o maior

número de atividades, independente da duração, e é o caminho crítico que indica a duração total do projeto. Ou seja, a distribuição *bootstrap* B-1 faz com que todas as atividades tenham a mesma variabilidade, independente da duração original. O grupo que estiver no caminho com a maior quantidade de atividades será o crítico, já para a B-R o que tem mais peso é o caminho crítico, conforme calculado pelo método do caminho crítico.

Essas proposições são confirmadas pelo protocolo P2, que identifica a correlação entre os valores-p do protocolo P1, considerando as distribuições, duas a duas: TR x B-1 e TR x B-R.

Quadro 25 - Análise de correlação entre as distribuições

TR x B-1	TR x B-R
Valor-p: 0.5488	Valor-p: < 2.2e-16
Intervalo de 95% de confiança: -0.02125155 -- 0.03996535	Intervalo de 95% de confiança: 0.5352927 -- 0.5775481
Índice de correlação: 0.009365676 Coeficiente de determinação R <sup>2</sup> : 0.00	Índice de correlação: 0.5567805 Coeficiente de determinação R <sup>2</sup> : 0.31

Fonte: O autor (2017).

O Quadro 25 mostra o resultado da análise de correlação entre o valor-p da distribuição triangular (TR) com o da *bootstrap* padrão (B-1). As distribuições TR x B-1 não são correlacionadas, pois quanto maior o valor-p, menores são as chances de existir a correlação, e, nesse caso, o valor-p é igual a 0,55, muito acima de um nível de significância de 0,05 ou 0,01. O seu índice de correlação é estatisticamente igual zero, pois um intervalo de confiança de 95% de confiança inclui o valor zero.

Já, em relação às distribuições triangular (TR) e *bootstrap-rate* (B-R), há grandes indícios da existência de correlação, pois seu valor-p é muito menor que 0,001, quase igual a zero. O índice correlação é igual a 0,58, sendo que seu coeficiente de determinação é de 31%. Isso significa que a igualdade entre a DTP da linha de base e a simulação da distribuição B-R explica 31% dos mesmos resultados em relação à distribuição triangular (TR).

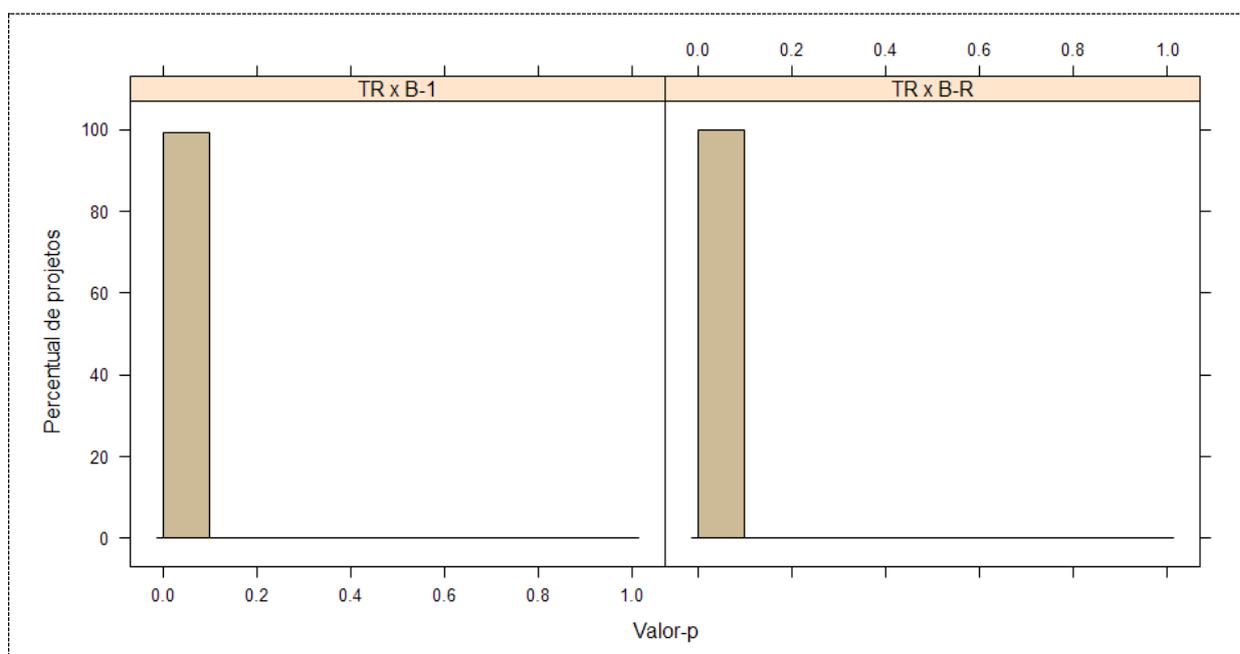
Então, apesar das estimativas da DTP serem bem diferentes nas três distribuições, os dados desta pesquisa demonstram que a distribuição *bootstrap-rate* (B-R) apresentou resultado correlacionado com a distribuição triangular (TR).

### 4.3 IGUALDADE DAS DISTRIBUIÇÕES DA DURAÇÃO TOTAL DO PROJETO

Foi verificado também se as distribuições da duração total do projeto podem ser consideradas as mesmas para os resultados da simulação Monte Carlo. As comparações foram feitas com as distribuições TR x B-1 e TR x B-R e foi utilizado o teste de Kolmogorov-Smirnov, conforme definido no protocolo P2.

No gráfico da Figura 25, quanto maior o valor-p, maiores as chances de as distribuições serem consideradas iguais. Como a maioria dos projetos ficou com valor-p bem abaixo de 0,1, por exemplo, significa que as distribuições de probabilidades das durações dos projetos são diferentes.

Figura 25 – Percentual de igualdade das distribuições da DTP<sup>s</sup>



Fonte: O autor (2017).

Os testes de igualdades entre as distribuições demonstram que tanto na verificação entre TR x B-1, como na verificação entre TR x B-R, as distribuições não podem ser consideradas as mesmas, ou seja, 100% dos projetos apresentaram valores-p abaixo de 0,1.

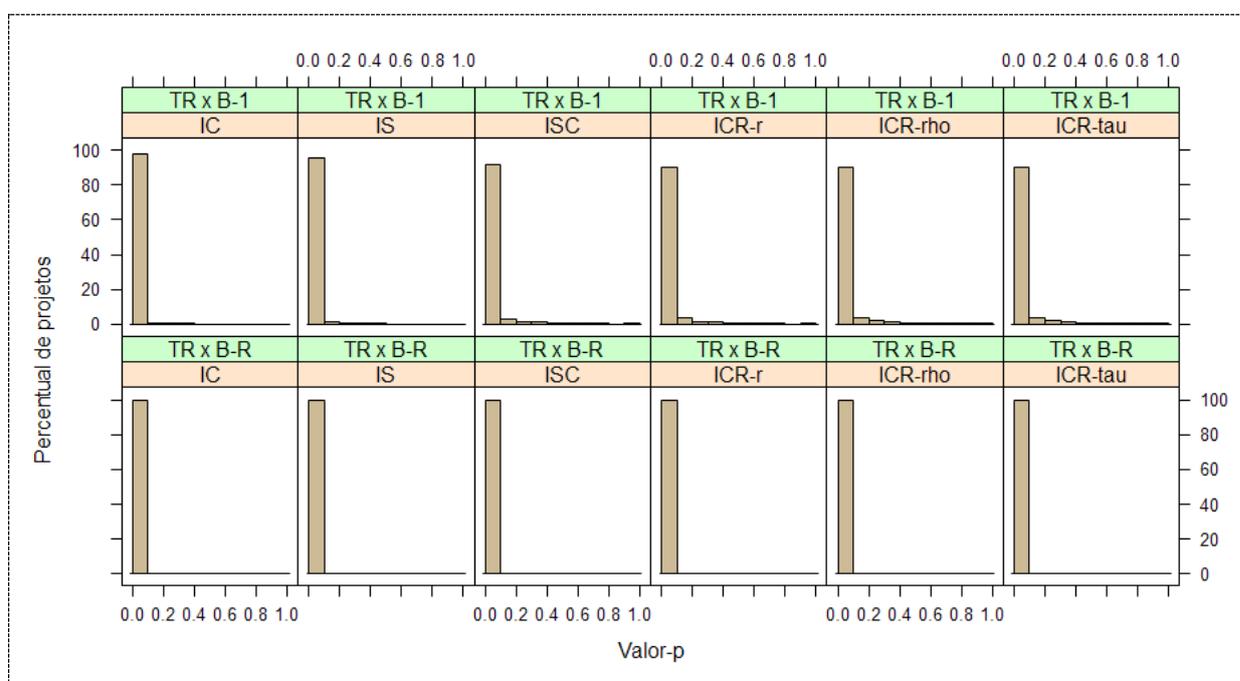
Os dois grupos de testes efetuados até este momento levaram em consideração a duração total do projeto (DTP): um verificou se a DTP é igual a DTP simulada; o outro, se a distribuição da DTP da triangular e da *bootstrap* podem ser considerados iguais. Tão relevante quanto esses resultados é verificar a influência da duração das atividades na duração total do projeto. Para isso, foram calculados

os índices de sensibilidade para a simulação Monte Carlo feita com as distribuições TR, B-1 e B-R, para cada um dos 4.100 projetos da amostra.

#### 4.4 CORRELAÇÃO DOS ÍNDICES DE SENSIBILIDADE

Após calcular os índices de sensibilidade, foi feita a análise de correlação entre os resultados, comparando TR x B-1 e TR x B-R, conforme protocolo P4, para verificar se existe correlação entre os índices de sensibilidade do tempo calculados entre as distribuições desta pesquisa.

Figura 26 – Percentual de projetos em relação ao valor-p da correlação entre os índices de sensibilidade

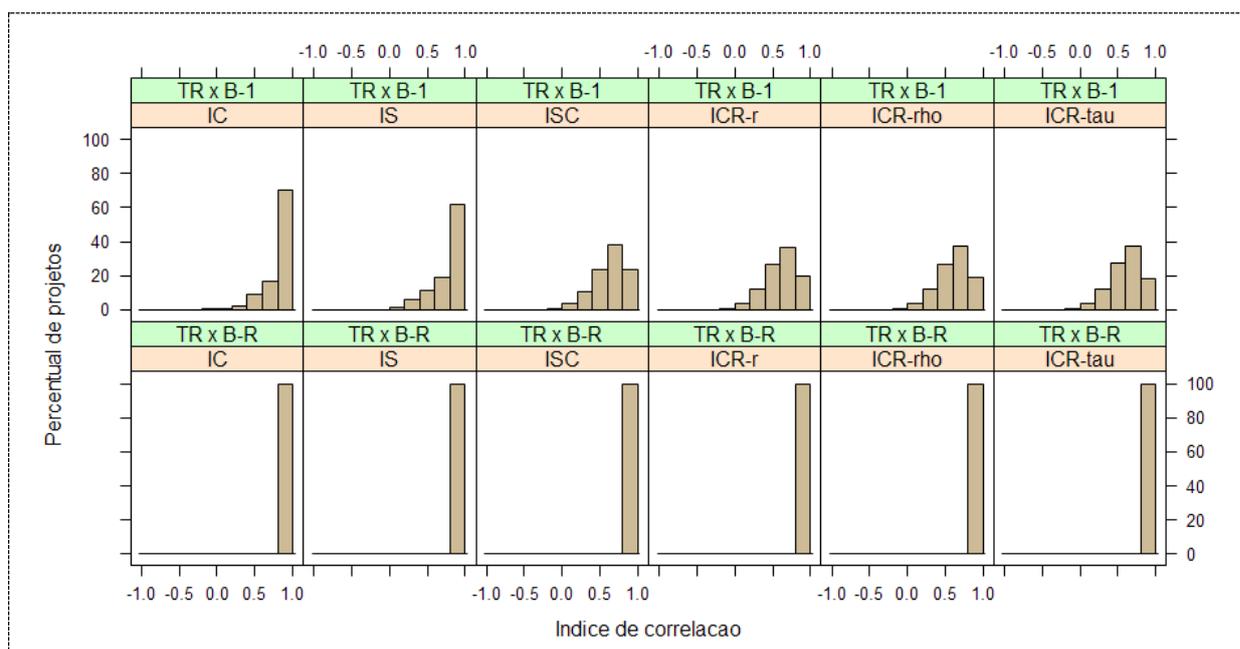


Fonte: O autor (2017).

A Figura 26 apresenta o gráfico do percentual de projetos em relação ao valor-p da correlação entre os índices de sensibilidade. A hipótese nula do teste é de não existência de correlação; então, quanto menor o valor-p, mais forte é a evidência de existência de correlação. A correlação da TR com a B-1 varia entre 90% e 97% dos projetos que apresentam valores-p menores do que 0,1; a correlação da TR com a B-R apresentou 100% dos projetos estão com valores-p menores do que 0,1. Isso é um forte da existência de correlação entre os índices de sensibilidade do cronograma.

Foi visto que existe correlação entre os índices de sensibilidade, agora será verificado qual é o valor do índice de correlação. A Figura 27 apresenta gráficos que indicam a proporção de projetos que possuem determinado índice de correlação. Os gráficos mostram que a correlação entre as distribuições B-R e TR é perfeita, pois 100% dos projetos tiveram um índice de correlação próximo a um, e, quanto mais o índice for próximo a um, mais correlacionados as variáveis são.

Figura 27 – Índice de correlação para os índices de sensibilidade do cronograma



Fonte: O autor (2017).

O índice de correlação pode variar de -1 a 1 e quando ele é negativo, existe uma correlação inversa, ou seja, uma variável aumenta e a outra diminui. Quando o índice é um, a correlação é perfeita e os dois valores variam sempre no mesmo sentido. A correlação dos índices de sensibilidade do cronograma da distribuição *bootstrap* padrão e da triangular não tiveram o mesmo desempenho da B-R, pois a proporção de projetos com índice próximo a um foi bem mais baixo. No gráfico é difícil de observar, mas a B-1 apresentou para alguns projetos correlação negativa.

Esses dados mostram que até existe a correlação entre os índices de sensibilidade do cronograma, mas para a distribuição B-1 os resultados não são satisfatórios, pois apresentaram correlação negativa. O mesmo não ocorre para a distribuição B-R, que demonstrou ser totalmente correlacionada em todos os índices com a distribuição triangular (TR).

## 4.5 DISCUSSÃO

A discussão dos resultados foi dividida em duas partes. A primeira parte faz um paralelo entre os resultados desta pesquisa com as pesquisas efetuadas pelos principais autores da referência, a segunda discute as diferenças dos resultados obtidos com as distribuições *bootstrap* padrão e a *rate*.

### 4.5.1 Comparação com outros trabalhos

Assim como nesta pesquisa, outros pesquisadores também simularam a execução de projeto por meio da simulação.

Tavares (1999) e Tavares, Ferreira e Coelho (1999) utilizaram simulação para identificar os riscos de atraso em projetos em termos da morfologia da rede do projeto. A palavra "morfologia" tem o mesmo sentido de "topologia", usada nesta pesquisa. Os autores analisaram a correlação da duração do projeto com os indicadores topológicos, considerando as distribuições de probabilidade normal, lognormal e exponencial invertida. Eles criaram um modelo para estimar o risco de atraso em projetos a partir da topologia da rede.

Demeulemeester, Vanhoucke e Herroelen (2003) criaram um gerador aleatório de rede de projetos, denominado RanGen, comparando-o com outros geradores aleatórios em termos de desempenho de processamento e de parametrização. Os parâmetros de entrada do RanGen são os indicadores topológicos e o resultado é a geração de dezenas e até milhares de redes de projetos para fins de pesquisa.

Em outra pesquisa, Vanhoucke et al. (2008) avaliaram, novamente, os geradores de rede aleatórios e criaram o RanGen2. Esses autores compararam quatro geradores, analisando a correlação entre os indicadores topológicos e a quantidade de redes diferentes que foram geradas. Ao todo foram geradas mais de 26 milhões de redes de projetos e os autores concluíram que o RanGen2 foi o melhor de todos. Novamente, os parâmetros utilizados foram os indicadores topológicos.

Vanhoucke (2015) utilizou o método Análise de Risco de Cronograma (*Schedule Risk Analysis – SRA*), que é base desta dissertação, para analisar a sensibilidade do tempo e do custo das atividades do projeto. Ele considerou os índices de sensibilidade do cronograma e foram utilizados os indicadores

topológicos para gerar as redes do projeto. Esse autor também publicou outros artigos nos quais analisa diversas características do projeto, como custo, duração, valor agregado, utilizando a simulação gerada a partir dos indicadores topológicos e levando em consideração os índices de sensibilidade do cronograma.

Os resultados desta dissertação são baseados nas técnicas desenvolvidas pelos autores citados, pois analisa a correlação dos índices de sensibilidade entre duas distribuições de probabilidade, a triangular e a *bootstrap*, resultando nas comparações TR x B-1 e TR x B-R. Apesar desses autores terem realizado diversas pesquisas, não compararam os resultados das simulação Monte Carlo com duas distribuições e nenhum deles citou a distribuição *bootstrap*.

#### 4.5.2 Comparação entre as distribuições

Na apresentação dos resultados, percebeu-se que existe diferença no desempenho das distribuições B-1 e B-R quando comparadas com a distribuição TR. Aqui neste sub-tópico será explicado o motivo dessa diferença.

A distribuição *bootstrap* foi criada a partir de reamostras com reposição da amostra original. As características de uma amostra com reposição já foi estudada (BOLFARINE; BUSSAB, 2005) e a soma dos elementos da amostra é a média multiplicada pela quantidade de elementos. No entanto, em um cronograma existem vários caminhos com quantidade diferentes de atividades e, ao utilizar a distribuição *bootstrap* padrão (B-1), o caminho com a maior quantidade de atividades determinará a duração do projeto. Isso é diferente do resultado de uma distribuição triangular (TR), pois o sorteio das durações é proporcional à duração da atividade e as atividades críticas serão as que pertencem ao caminho com a maior duração (e não o caminho com a maior quantidade de atividades), por isso a distribuição B-1 não apresenta resultados semelhantes à distribuição triangular (TR).

A distribuição *bootstrap-rate* (B-R) é baseada na média das durações das atividades e na duração da atividade que está sendo sorteada, ou seja, também é proporcional à duração da atividade. Por isso ela é correlacionada com a distribuição triangular (TR) e apresenta resultados semelhantes.

Outra questão que os gráficos e tabelas apresentados ilustram é que as distribuições da duração total do projeto não terão valores iguais à duração calculada pelo método do caminho crítico, e muito menos as distribuições da DTP serão iguais entre as distribuições TR x B-1 e TR x B-R. No entanto, ressalta-se que,

apesar disso, a correlação entre os índices de sensibilidade do cronograma é perfeita para a distribuição *bootstrap-rate* (B-R). Já para a *bootstrap* padrão (B-1) a correlação não é perfeita (apesar de os testes de correlação apontarem a existência de alguma correlação), pois o índice de correlação não é próximo a um para alguns projetos, e chega até a ter índices com correlação negativa.

## 5 CONSIDERAÇÕES FINAIS

Este capítulo final apresenta uma síntese da pesquisa. Para melhor organização, ele foi dividido em quatro partes. Na primeira são feitas as considerações finais propriamente ditas. A segunda parte relata as limitações encontradas nesta pesquisa. As sugestões para trabalhos futuros foram colocadas na terceira parte e, finalmente, as principais contribuições desta pesquisa estão descritas na última parte.

Esta pesquisa teve o objetivo de demonstrar que a distribuição *bootstrap* pode ser utilizada na análise dos riscos em cronograma de projetos, o qual foi alcançado duplamente.

Primeiro, considerando que a distribuição *bootstrap* padrão (B-1) teve desempenho muito variado, pois em alguns projetos teve correlação próximo a um e, em outros, próximo a 0,6 e até correlação negativa; Com esses resultados não é possível afirmar que a distribuição *bootstrap* padrão pode ser utilizada para a análise de riscos em cronograma de projetos.

Segundo, como a distribuição *bootstrap-rate* (B-R) teve um desempenho consistente, com índice de correlação acima de 0,98, seu desempenho pode ser considerado igual ao da triangular (TR). O cálculo dos índices de sensibilidade do cronograma feito com a B-R é totalmente correlacionado com os feitos com a distribuição triangular (TR), que é a distribuição recomendada pelo PMI (2013) e utilizada em pesquisas de vários autores, como Vose (2008), Hulett (2009) e Vanhoucke (2013).

Assim, esta pesquisa apontou que a distribuições de durações da B-R não é igual à TR, mas é correlacionada, significando que ela indica as mesmas atividades mais sensíveis do cronograma que a TR aponta. **Portanto, a distribuição *bootstrap-rate* (B-R) pode ser utilizada na análise de riscos de cronograma de projetos.**

O benefício da *bootstrap-rate* em relação à distribuição triangular é tornar mais rápida a utilização da simulação por não necessitar da definição de distribuições de probabilidade nem de seus parâmetros, além de não exigir conhecimentos nesses assuntos, ao mesmo tempo em que mantém as mesmas características da distribuição triangular. O método aqui desenvolvido utiliza como entrada o diagrama de rede do projeto com a duração estimada das atividades, sem

necessidade de fazer suposições sobre a distribuição de probabilidade das atividades, muito menos estimar seus parâmetros, e mantendo a variabilidade intrínseca das atividades e a duração aleatória proporcional à duração da atividade.

Deve-se ressaltar que a distribuição *bootstrap* não substitui o método do caminho crítico (CPM), muito pelo contrário, depende dele. Além disso, os cálculos dos índices de sensibilidade dependem do cálculo da rede do projeto, o qual é feito com o CPM. Outra questão relevante é que a simulação Monte Carlo é um método computacional, ou seja, deve ser feito com auxílio do computador e uma linguagem de programação. Independente se é feita com a distribuição triangular ou com a *bootstrap-rate*, é necessário o auxílio de um programa de computador.

## 5.1 LIMITAÇÕES

Esta pesquisa foi baseada em uma amostra de 4.100 cronogramas fictícios sorteados aleatoriamente por Vanhoucke (2009), abrangendo diversas estruturas de cronogramas, dessa forma não foram utilizados cronogramas reais. Uma limitação desta pesquisa é a não utilização de cronogramas reais.

Além disso, a pesquisa focou apenas na fase de planejamento do gerenciamento de projetos. Poderia focar também no grupo de processo de execução dos projetos, alimentando os cronogramas e simulando novamente para verificar se as questões apontadas pela simulação no planejamento estão ocorrendo na fase de execução.

## 5.2 SUGESTÕES PARA TRABALHOS FUTUROS

Dois trabalhos futuros surgem diretamente das limitações da pesquisa: um é realizar a pesquisa com cronogramas de projetos reais e o outro é uma pesquisa para abranger todo o ciclo de vida do projeto, principalmente planejamento, execução e monitoramento e controle, realizando simulações periodicamente para comparar seus resultados.

Assim como proposto por Vanhoucke (2009), esta dissertação abordou seis índices de sensibilidade diferentes, o que pode levar a novos questionamentos: algum desses índices é melhor que os outros? Não é redundante calcular os seis índices? Os índices apontam na mesma direção ou são conflitantes? Então, outra sugestão de trabalho seria comparar os índices de sensibilidade entre si para ver se

existe algum que seja o “melhor” e responder às questões apontadas com o objetivo de reduzir o cálculo a apenas dois índices a fim de tornar mais ágil sua utilização.

Nesta pesquisa foram vistos também os indicadores topológicos, mas não foi verificada a influência desses indicadores nos índices de sensibilidade, então, outro trabalho futuro poderia analisar os índices de sensibilidade em relação aos índices de estrutura topológica da rede.

Vanhoucke (2009) calcula o índice de crucialidade, considerando as três formas de índice de correlação, mas ele não faz um teste de hipótese para ver se existe a correlação entre a atividade e o projeto. Além disso, para o cálculo dos índices, ele compara a duração das atividades uma a uma, com a duração do projeto. Duas propostas relacionadas a essas observações podem ser abordadas em um futuro estudo: a importância do teste de hipótese, para verificar se existe correlação mesmo, e a execução de uma análise multivariada, considerando, como variáveis independentes, a duração das atividades, e, como dependente, a duração total do projeto.

### 5.3 CONTRIBUIÇÕES DA PESQUISA

Na sessão 1.3 Justificativa, foram apontadas questões relevantes para esta pesquisa as quais estão ligadas diretamente às contribuições da pesquisa.

Contribuição pessoal: são de interesse do autor desta pesquisa a análise de risco, a simulação e a análise estatística, portanto, este trabalho trouxe uma contribuição pessoal, uma vez que foram desenvolvidos estudos nessas áreas e foi possível a aquisição de conhecimentos em outras áreas, como mineração de dados, apesar de não terem sido utilizados diretamente na pesquisa.

Contribuição empresarial: a pesquisa é uma atividade muito importante, mas quando ligada ao ambiente empresarial, essa importância é ampliada. O desenvolvimento desta pesquisa também permitiu a troca de experiências dentro do ambiente empresarial, principalmente em discussões sobre a aplicação da técnica desenvolvida, sendo que a mesma será aplicada nos projetos da empresa na qual o pesquisador trabalha, além de ser possível sua aplicação em outras empresas.

Contribuição científica: apesar de existirem milhares de artigos que estudam a aplicação da distribuição *bootstrap*, em diversas áreas de pesquisa, esta pesquisa é um dos poucos trabalhos que trata da simulação *bootstrap* na análise de riscos de cronograma, o que permite oferecer uma contribuição científica base, que servirá

para desenvolvimento de novas pesquisas, tanto na área de gerenciamento de projetos como em outras áreas que envolvam simulação.

Contribuição ao Programa: o gerenciamento de projetos é um dos processos fundamentais para a gestão da informação, tema central do Programa de Pós-Graduação em Ciência, Gestão e Tecnologia da Informação, mas, apesar disso, apenas duas dissertações foram elaboradas no programa que aborda o tema gerenciamento de projetos. Uma investiga as práticas de apoio ao compartilhamento do conhecimento interprojetos e sua importância para os resultados do projeto (RORIZ, 2011); a outra analisa a produção científica sobre o papel do escritório de projetos na gestão do conhecimento (RIEDI, 2015). Assim, esta dissertação apresenta mais uma contribuição para o programa, abrindo possibilidades para o desenvolvimento de outros estudos relacionados ao tema.

Contribuição à comunidade: a simulação Monte Carlo é uma técnica difundida na área de gerenciamento de projetos; ao prover alternativas em que o seu uso seja menos custoso e exija menos conhecimento técnico em estatística, será possível que ela seja ainda mais utilizada. Esta pesquisa traz contribuições para a comunidade de gerenciamento de projetos, diretamente, e também para a comunidade em geral, pois pode minimizar os atrasos e o estouro de orçamentos em obras públicas e particulares.

## REFERÊNCIAS

- AMADEU, Maria Simone Utida dos Santos *et al.* **Manual de normalização de documentos científicos de acordo com as normas da ABNT.** Curitiba: Ed. UFPR, 2015.
- BARCAUI, André B. **Gerenciamento do tempo em projetos.** 2. ed., Rio de Janeiro: Ed. FGV, 2006.
- BECKER, João Luiz. **Estatística básica: transformando dados em informação.** Porto Alegre: Bookman, 2015.
- BOLFARINE, Heleno; BUSSAB, Wilton O. **Elementos de amostragem.** São Paulo: Blucher, 2005.
- BRACARENSE COSTA, Paulo Afonso. **Um enfoque segundo a lógica difusa para a meta-análise.** 155 f. Tese (Doutorado em Engenharia de Produção) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 1999.
- BRUNI, Adriano Leal. **Estatística aplicada à gestão empresarial.** 4 ed. São Paulo: Atlas, 2013.
- BYTHEWAY, Andy. **Investing in information: the Information Management Body of Knowledge (IMBOK).** New York: Spring, 2014.
- CHAFFEY, Dave; WHITE, Gareth. **Business information management: improving performance using information systems.** 2 ed. Harlow: Pearson, 2012.
- CHERNICK, Michael R.; LABUDDE, Robert A. **An introduction to bootstrap methods with applications to R.** Hoboken: Wiley, 2011.
- CHWIF, Leonardo; MEDINA, Afonso C. **Modelagem e simulação de eventos discretos: teorias & aplicações.** 4 ed. Rio de Janeiro: Elsevier, 2015.
- DAMA INTERNATIONAL (DAMA). **The DAMA guide to the data management body of knowledge: DAMA-DMBOK Guide.** Bradley Beach: Technics, 2010.
- DEMEULEMEESTER, Erik; VANHOUCKE, Mario; HERROELEN Willy. RanGen: a random network generator for activity-on-the-node networks. **Journal of Scheduling.** v. 6, p. 17-38, 2003.
- EFRON, Bradley. Bootstrap methods: another look at the Jackknife. **The Annals of Statistics.** California, v. 7, n. 1, p. 1-26, 1979.
- \_\_\_\_\_. Jackknife-after-bootstrap standard errors and influence functions. **Journal of the Royal Statistical Society.** v. 54, n. 1, p. 83-127, 1992.
- FINOCCHIO JÚNIOR, José. **Project model canvas.** Rio de Janeiro: Elsevier, 2013.
- GIL, Antonio Carlos. **Como elaborar projetos de pesquisas.** 5 ed. São Paulo: Atlas, 2010.
- GOODRICH, Michael T.; TAMASSIA, Roberto; GOLDWASSER, Michael H.. **Data Structures and Algorithms in Java™.** 6. ed., Hoboken: Wiley, 2014.

GUIMARÃES, André José Ribeiro. **Gestão do conhecimento, criatividade, inovatividade e desempenho inovador em empresas de publicidade em Curitiba-PR**. 230 f. Dissertação (Mestrado em Ciência, Gestão e Tecnologia da Informação) – Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, Curitiba, 2014.

HAUWERMEIREN, Michael van. A compendium of distributions. In: VOSE, David. **Risk analysis**. 3 ed. West Sussex: John Wiley, 2008. 585-713.

HELDMAN, Kim. **PMP Project Management Exam: study guide**. 6 ed. Indianapolis: John Wiley, 2011.

HERNÁNDEZ SAMPIERI, Roberto; FERNÁNDEZ COLLADO, Carlos; BAPTISTA LUCIO, María del Pilar. **Metodologia de pesquisa**. 5 ed. São Paulo: Penso, 2013.

HUBBARD, Douglas W. **How to measure anything: finding the value of “intangibles” in business**. 3 ed. New Jersey: Wiley, 2014.

HULETT, David. **Practical schedule risk analysis**. Burlington: Gower, 2009.

MALHOTRA, Naresh K. **Pesquisa de marketing: uma orientação aplicada**. 6 ed. Porto Alegre: Bookman, 2012.

MOORE, David S. *et al.* **A prática da estatística empresarial: como usar dados para tomar decisões**. Rio de Janeiro: LTC, 2006.

MULCAHY, Rita; DIETHELM, Laurie. **Preparatório para o exame de PMP: aprendizado rápido para passar no exame de PMP do PMI – na primeira tentativa**. [s. l.]: RMC, 2011.

MURRAY, Peter. Knowledge management as a sustained competitive advantage. **Ivey Business Journal: improving the practice of management**. Março/Abril, 2002. Disponível em: <<http://iveybusinessjournal.com/publication/knowledge-management-as-a-sustained-competitive-advantage/>>. Acesso em: 2 mar. 2016.

PAULA FILHO, Wilson de Pádua. **Engenharia de software: fundamentos, métodos e padrões**. 3 ed. Rio de Janeiro: LTC, 2009.

PROJECT MANAGEMENT INSTITUTE (PMI). **Practice standard for project risk management**. Newton Square: PMI, 2009.

\_\_\_\_\_. **Um guia do conjunto de conhecimentos em gerenciamento de projetos (PMBOK)**. 5 ed. Newton Square: Project Management Institute, 2013.

R CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2016. Disponível em: <<http://www.R-project.org/>>. Acesso em: 1 dez. 2016.

RORIZ, Maysa Silva. **Uma investigação sobre as práticas de apoio ao compartilhamento do conhecimento interprojetos em uma empresa orientada a projetos**. 149 f. Dissertação (Mestrado em Ciência, Gestão e Tecnologia da Informação) – Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, Curitiba, 2011.

SIEGEL, Sidney; CASTELLAN JR, Norman John. **Estatística não paramétrica para ciências do comportamento**. 2 ed. Porto Alegre: Artmed, 2006.

SMITHSON, Michael. **Ignorance and uncertainty: emerging paradigms**. New York: Springer-Verlag, 1989.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao data mining – mineração de dados**. Rio de Janeiro: Ciência Moderna, 2009.

TAVARES, Luís Valadares. **Advanced models for project management**. New York: Springer, 1999.

TAVARES, Luís Valadares; FERREIRA, José Antunes; COELHO, José. The risk of delay of a project in terms of the morphology of its network. **European Journal of Operational Research**. v. 119, p. 510-537, 1999.

VALLE, André Bittencourt do *et al.* **Fundamentos do gerenciamento de projetos**. Rio de Janeiro: FGV, 2007.

VANHOUCKE, Mario. **Measuring time: improving project management using earned value management**. London: Springer, 2009.

\_\_\_\_\_. **Project management with dynamic scheduling: baseline scheduling, risk analysis and project control**. 2 ed. London: Springer, 2013.

\_\_\_\_\_. **Integrated Project management and control: first comes the theory, then the practice**. New York: Spring, 2014.

\_\_\_\_\_. On the use of schedule risk analysis for project management. **The Journal of Modern Project Management**. Curitiba, v. 2, n. 3, p. 108-117, jan./apr., 2015.

VANHOUCKE, Mario *et al.* An evaluation of the adequacy of project network generators with systematically sampled networks. **European Journal of Operational Research**. n. 187, p. 511-524, 2008.

VOSE, David. **Risk analysis**. 3 ed. West Sussex: John Wiley, 2008.

WEBSTER JR., Francis M.; KNUTSON, Joan. O que é gerenciamento de projetos? Conceitos de gerenciamento de projetos e metodologias. In: DINSMORE, Paul C.; CABANIS-BREWIN, Jeannette. **AMA manual de gerenciamento de projetos**. Rio de Janeiro: Brasport, 2009. 1-10.

XAVIER, Carlos Magno da Silva *et al.* **Metodologia de gerenciamento de projetos Methodware: abordagem prática de como iniciar, planejar, executar, controlar e fechar projetos**. 2 ed. Rio de Janeiro: Brasport, 2009.

## **APÊNDICES**

## APÊNDICE A - ÁREAS DE APLICAÇÃO DA DISTRIBUIÇÃO *BOOTSTRAP*

O objetivo deste texto é identificar áreas de aplicação da distribuição *bootstrap*. Para isso foi feita uma análise bibliométrica de artigos publicados sobre esse assunto e foi aplicada a técnica de mineração de textos para identificar as palavras mais significantes e os periódicos em que esses artigos foram publicados.

### 1 Introdução

**Objetivo: Identificar áreas de aplicação da distribuição *bootstrap*.**

Este tópico identifica algumas áreas em que a distribuição *bootstrap* é aplicada, bem como as técnicas utilizadas em conjunto com ela.

Quadro 26 – Perguntas a serem respondidas

Pergunta	Por quê?
Considerando que existem artigos publicados sobre a simulação <i>bootstrap</i> , essas publicações estão aumentando, estabilizadas ou diminuindo no período pesquisado?	<ul style="list-style-type: none"> <li>Identificar a evolução anual da publicação de artigos sobre a simulação <i>bootstrap</i> ao longo dos anos.</li> </ul>
Quais as palavras relevantes do conjunto de artigos?	<ul style="list-style-type: none"> <li>Analisando as palavras-chave é possível inferir os temas em que a simulação <i>bootstrap</i> foi utilizada, e, com essas palavras-chave, identificam-se os periódicos em que os artigos foram publicados.</li> </ul>
Quais foram os periódicos que mais publicaram sobre o tema em questão?	<ul style="list-style-type: none"> <li>Os periódicos são especializados em determinados assuntos. Identificando os periódicos, é possível inferir os temas em que a simulação <i>bootstrap</i> foi utilizada.</li> </ul>

Fonte: O autor (2017).

### 2 Referencial teórico

Este referencial explicará o princípio do *plug-in* e a técnica de identificação de documentos relevantes baseada na frequência do termo e na quantidade de documentos em que ele aparece.

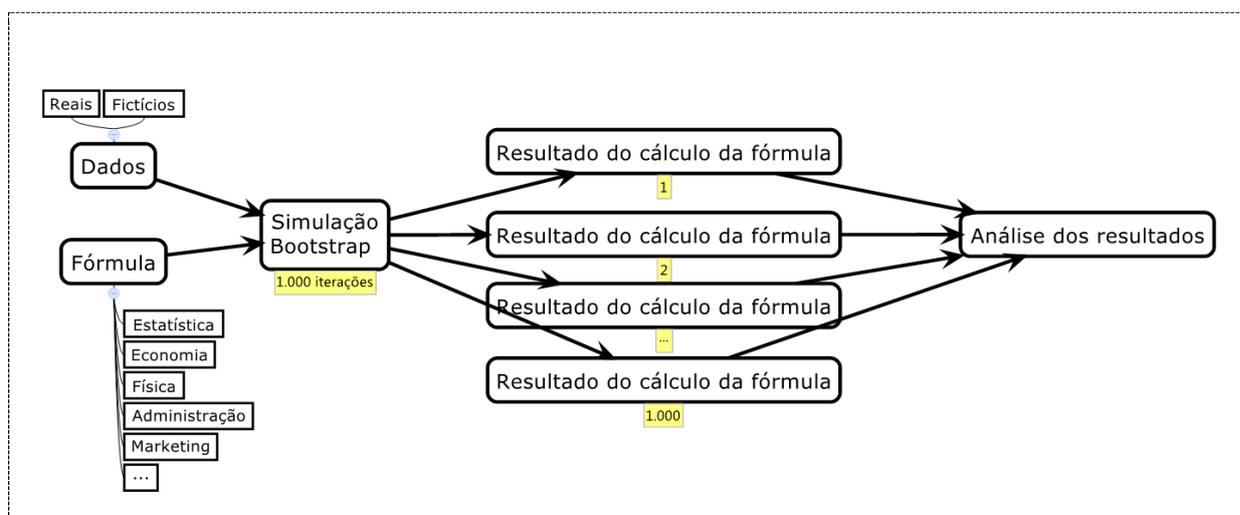
#### 2.1 Princípio *plug-in*

Quando uma estimativa é usada no lugar de um parâmetro, diz-se que se aplicou o princípio *plug-in*. O princípio *plug-in* é a utilização da uma estatística

calculada a partir da amostra no lugar de um parâmetro populacional (CHERNICK; LABUDDE, 2011). Como exemplo, pode-se citar o cálculo de intervalo de confiança quando o desvio padrão populacional não é conhecido. Estima-se o seu valor a partir dos dados da amostra.

A simulação *bootstrap* utiliza esse mesmo princípio, pois com ela são estimados vários parâmetros populacionais a partir de reamostras da amostra.

Figura 28 – Esquema geral da simulação *bootstrap*



Fonte: O autor (2017).

A simulação *bootstrap* está associada a técnicas estatísticas e de outras áreas, generalizada como uma fórmula (Figura 28). Nos artigos analisados, as palavras que mais aparecem associadas à simulação *bootstrap* estão relacionadas à estatística, assim como os periódicos em que foram publicados são fundamentalmente estatísticos. Já é esperado que as maiores associações ocorram com técnicas estatísticas.

## 2.2 Frequência do termo

Este tópico demonstra como selecionar as palavras mais significativas de um conjunto de artigos. A seleção não pode ser feita pela frequência das palavras porque as palavras mais frequentes são as palavras que fizeram parte de uma estratégia de busca e também são aquelas palavras mais comuns da língua, neste caso, o inglês. Reconhecer palavras mais significativas implica que as palavras têm um peso diferente. Para identificar os termos (ou palavras) mais significativos, será utilizada a técnica TF-IDF, explicada passo a passo a seguir.

A primeira abordagem é atribuir o peso de cada termo como sendo a frequência desse termo no artigo, ou seja, a frequência do termo (*term frequency* - TF) (conforme MANNING; RAGHAVAN; SCHÜTZE, 2009). Então, considerando a amostra de artigos desta pesquisa, os cinco termos mais significativos são: *the, of, de, yen, and*. Por esses termos aparecerem frequentemente em todos os documentos, eles não têm um alto poder de discriminação.

Essas palavras são chamadas de *stop words*, ou seja, palavras que não contribuem para a identificação de um documento (INGERSOLL; MORTON; FARRIS, 2013). Nesta pesquisa será utilizada a expressão “palavras comuns” para se referenciar às *stop words*.

A segunda abordagem é considerar a frequência de documentos em que o termo aparece (*document frequency* – DF). Nesta abordagem são contados em quantos documentos o termo aparece na coleção de documentos (MANNING; RAGHAVAN; SCHÜTZE, 2009).

Para a coleção de documentos desta pesquisa, os termos que mais aparecem em documentos diferentes são: *bootstrap, simul, the, of, and*. Os termos utilizados como estratégia de busca foram *simula\** e *bootstrap*, então é natural que esses termos sejam os que apareçam em todos os documentos junto com algumas palavras comuns já conhecidas. Nenhuma dessas palavras pode ser utilizada como palavras significantes dessa coleção de documentos, pois elas aparecem em quase todos os documentos.

A abordagem utilizada para definir o peso de cada termo é o inverso do DF (*inverse document frequency* – IDF), conforme a seguinte fórmula (onde N é a quantidade total de documentos):

$$IDF = \log_{10} \frac{N}{DF}$$

Ou seja, é o logaritmo na base de 10 da quantidade de documentos dividido pela frequência de documentos em que o termo aparece. Quanto mais próximo o DF chega de N, mais próximo a divisão se aproxima de um e o cálculo do logaritmo se aproxima de zero. Significando que termos que aparecem em todos os documentos terão peso zero (MANNING; RAGHAVAN; SCHÜTZE, 2009).

A fórmula final para o cálculo do peso de cada termo nos documentos é:

$$TF-IDF = TF \times IDF$$

Ou seja, o TF-IDF associa um peso ao termo de um documento com o inverso da frequência em que o termo aparece no documento, que, segundo Manning, Raghavan e Schütze (2009), apresenta as seguintes características:

- maior, quando o termo ocorre várias vezes em um pequeno número de documentos;
- menor, quando o termo ocorre poucas vezes em um documento ou ocorre em muitos documentos;
- o mais baixo, quando o termo ocorre em (quase) todos os documentos.

Deve-se observar que o TF-IDF é calculado para cada termo em cada documento, ou seja, o mesmo termo terá TF-IDF diferentes porque, apesar do IDF ser uma propriedade do termo, o TF é a frequência do termo em um documento, ou seja, uma propriedade do termo em um documento.

Os cinco termos com maiores TF-IDF são: *yen, de, hl, propofol, dyslexia*. Esses termos têm o TF-IDF mais alto porque ocorrem entre um e seis documentos, e nesses documentos ocorrem muitas vezes, entre 34 e 55 vezes. O termo *de* ocorre em 55 documentos, por isso ele apareceu entre os mais significativos.

O objetivo desse tópico foi mostrar como se calcula o TF-IDF e que o índice tem alto poder discriminatório. Com ele é possível ordenar os termos de um conjunto de documento considerando a sua frequência no documento e a frequência de documentos em que ele aparece. Na sessão método, será explicado como o TF-IDF foi utilizado nesta pesquisa para que ele não traga termos que aparecem muitas vezes em poucos documentos.

### 3 Método

Para identificar as áreas de aplicações da simulação *bootstrap*, esta pesquisa partiu do pressuposto que os periódicos são especializados em determinado assunto e, então, se existirem artigos publicados em um periódico de medicina, por exemplo, a simulação *bootstrap* é aplicada em medicina. A base para identificar os periódicos foram as palavras mais significativas do conjunto de artigos identificadas através da técnica *term frequency inverse document frequency* (TF-IDF).

Para atender ao objetivo desta pesquisa, ou seja, para identificar as áreas em que a simulação *bootstrap* é aplicada, foram coletados artigos científicos, de qualquer área, que contivessem a expressão "simulação *bootstrap*". Além disso, não

se limitou o período de anos da pesquisa, então, a estratégia de busca foi definida da seguinte forma:

- estratégia: **bootstrap\* AND simula\***;
- campos de busca: **título, resumo / abstract e palavras-chave**;
- Período: **todos os anos**;

Foram definidos como campos de busca o título, o resumo e as palavras-chave para direcionar os artigos para o assunto da pesquisa. Se a busca fosse feita no texto todo, poderia trazer artigos não relevantes, por isso a busca foi limitada aos três campos. A busca em todos os anos disponíveis nas bases foi feita para identificar a evolução anual da quantidade de artigos publicados sobre o assunto.

Após definida a estratégia de busca, deve-se definir a base de artigos onde a busca será feita. Para esta pesquisa, duas bases foram selecionadas: *Web of Science* e *Science Direct*. Essas duas bases foram escolhidas devido à possibilidade de exportar os resultados em arquivos e por abordarem diversas áreas.

Para esta pesquisa foram coletadas as seguintes variáveis: o título do periódico, o tipo, o ano de publicação e as palavras-chave. Como os metadados, cada base é diferente, o Quadro 27 mostra o relacionamento entre as variáveis e os metadados de cada uma das bases.

Quadro 27 - Metadados das variáveis coletadas por periódicos

Variável	Periódico	
	Science Direct (.ris)	Web of Science (.bib)
Título	T1	Title
Periódico	JO	Journal
Ano	PY	Year
Palavras-chave	KW	Keywords Keywords-Plus

Fonte: O autor (2017).

Como apresentado no Quadro 27, ambos os periódicos retornam arquivos do tipo texto com um formato pré-definido. A base de artigos *Science Direct* retorna arquivo do tipo 'ris' e a *Web of Science* do tipo 'bib'. Por serem arquivos do tipo texto, existem vários aplicativos que conseguem ler esses arquivos e é possível até mesmo desenvolver um programa em uma linguagem de programação para analisar os resultados.

Resumidamente, os passos executados do pré-processamento foram:

- 1) definir a estratégia de busca;
- 2) selecionar as bases de artigos científicos;
- 3) efetuar a busca;
- 4) copiar os arquivos com os artigos científicos.

Os arquivos foram processados utilizando-se o ambiente estatístico e linguagem de programação **R** (R Core Team, 2016) e foram seguidos os seguintes passos:

- 1) agrupar os arquivos das bases de periódicos;
- 2) transformar em um estrutura para processar no **R**;
- 3) agrupar em um campo único o título, as palavras-chave e o resumo;
- 4) identificar os radicais de cada palavra;
- 5) calcular o TF-IDF;
- 6) ignorar os radicais que aparecerem em menos de 1% dos artigos;
- 7) ordenar os radicais por ordem decrescente do TF-IDF;
- 8) selecionar os 25 radicais com o maior TF-IDF.

A partir desta etapa já foi possível identificar os termos mais significativos, bem como os periódicos em que esses termos foram publicados, então, conforme a sequência abaixo (pós-processamento):

- 1) identificar os termos relacionados aos radicais;
- 2) identificar os periódicos com maior quantidade de artigos em que os termos aparecem;
- 3) inferir as áreas de aplicação da simulação *bootstrap* a partir dos periódicos em que os artigos foram publicados.

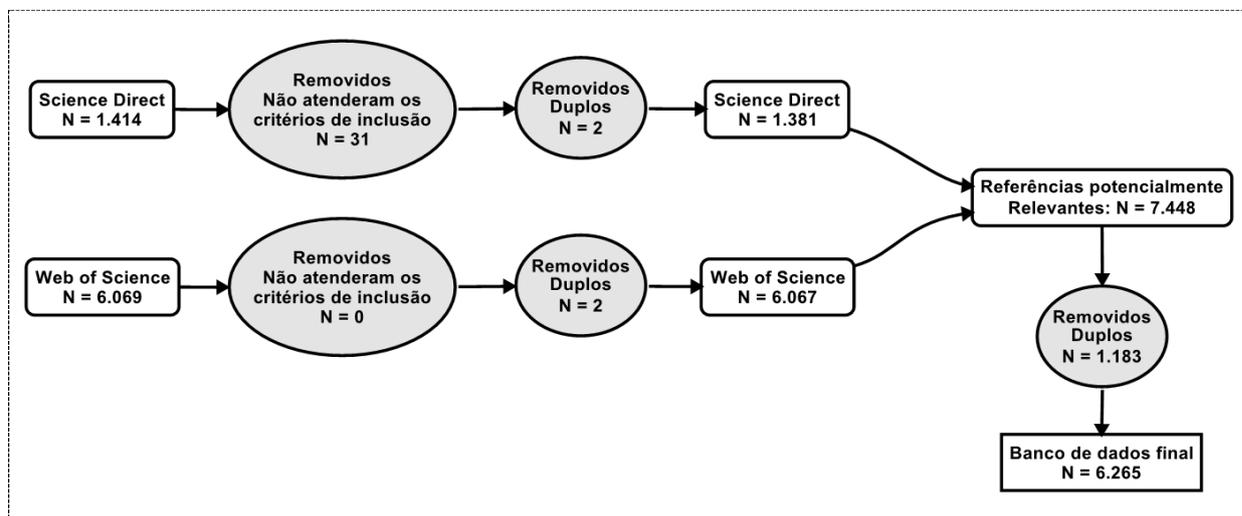
Seguindo esses passos e analisando os resultados obtidos, foi possível identificar as principais áreas de aplicação da simulação *bootstrap*.

#### **4 Apresentação dos resultados**

Passa-se agora à apresentação dos resultados. Os dados foram coletados no dia 11 de setembro de 2016, e a quantidade final de artigos foi 6.265. A Figura 29 mostra como se chegou a esse valor. Alguns artigos coletados não continham as palavras definidas na estratégia de busca e foram eliminados. Além disso, com base no título, foram retirados da amostra os artigos duplicados na própria base e entre as

bases. No caso de duplicações entre as bases, foram mantidos os artigos da base *Web of Science*.

Figura 29 – Resultado da busca nas bases *Science Direct* e *Web of Science*



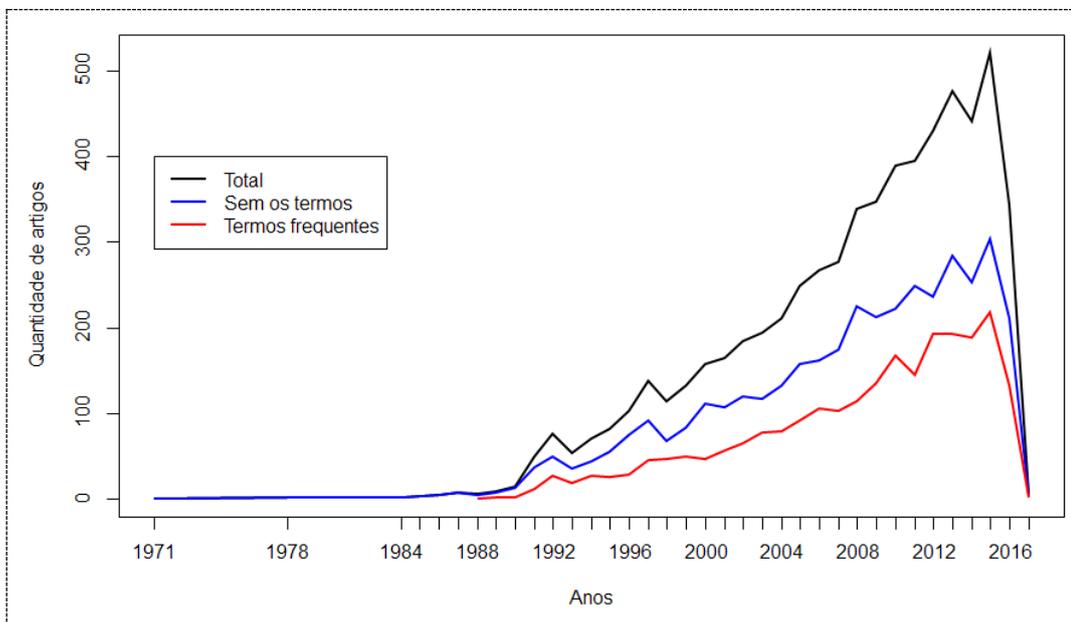
Fonte: O autor (2017).

O gráfico da Figura 30 indica a evolução anual da publicação de artigos relacionados à simulação *bootstrap*. A linha preta é a quantidade total de publicações. A linha vermelha são os artigos que contêm as palavras-chave mais significantes de acordo com o método TF-IDF, e linha a azul são os artigos que não contêm as palavras-chave mais significantes. É possível perceber no gráfico que a publicação sobre simulação *bootstrap* aumenta até o ano de 2015, quando ocorre o pico mais alto. Em 2016, até a data da coleta dos dados, o número de artigos publicados diminuiu consideravelmente.

Essa evolução ao longo do tempo é para as publicações totais, bem como dos artigos que contêm as palavras-chave mais importantes.

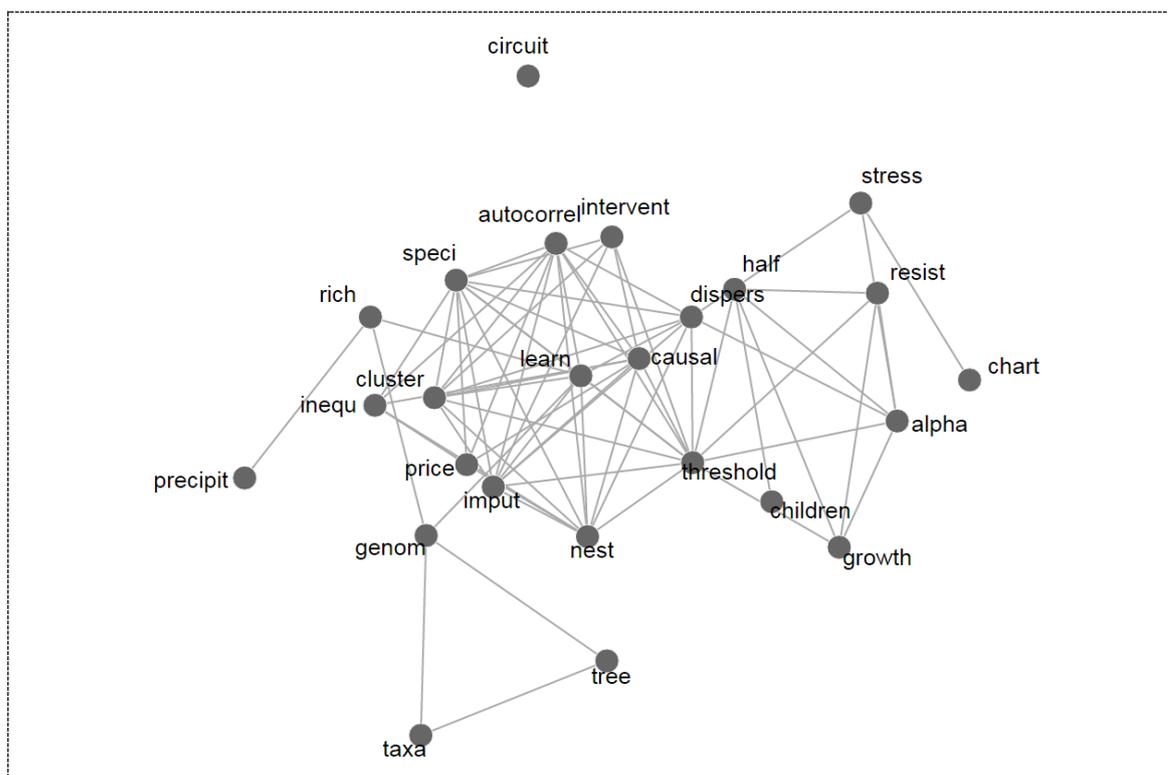
Foram selecionados os radicais dos termos mais relevantes através do filtro TF-IDF e a Figura 31 mostra a rede de relacionamentos entre os radicais e os periódicos em que foram publicadas com mais frequência. Nessa figura, os radicais são representados em círculo; os periódicos nos quais foram publicados artigos são as linhas.

Figura 30 – Quantidade de artigos publicados por ano



Fonte: O autor (2017).

Figura 31 – Relacionamento das palavras-chaves com periódicos em comum



Fonte: O autor (2017).

Percebe-se na figura que o radical *'circuit'* não aparece em publicações relacionado com os outros radicais, já o radical *'chart'* é relacionado ao *'stress'* e o radical *'learn'* é relacionado com vários radicais.

A técnica utilizada para montar a Figura 31 chama-se gráfico de força, pois é uma simulação física da força de relação entre os radicais, baseado nos periódicos publicados. Por causa da força, os elementos mais relacionados ficam no meio e os mais soltos ficam na periferia da figura, caso do radical *'learn'* e do *'chart'* respectivamente.

Cada radical pode representar uma ou mais palavras, com variações para plural, singular ou palavras derivadas. O Quadro 28 mostra a relação de cada um dos radicais com os termos que eles representam. Nesse quadro também é mostrado o periódico em que o radical foi considerado relevante.

Quadro 28 – Relação de termos e seus periódicos

Radical	Termo	Periódico
tree	tree, trees.	<ul style="list-style-type: none"> <li>• Systematic Biology;</li> <li>• Molecular Biology and Evolution.</li> </ul>
chart	chart, charting, charts.	<ul style="list-style-type: none"> <li>• Quality and Reliability Engineering International;</li> <li>• Communications in Statistics-Simulation and Computation.</li> </ul>
cluster	cluster, clustered, clustering, clusterings, clusterization, clusters.	<ul style="list-style-type: none"> <li>• Computational Statistics &amp; Data Analysis;</li> <li>• Statistics in Medicine.</li> </ul>
alpha	alpha, alphas.	<ul style="list-style-type: none"> <li>• Nuclear Fusion;</li> <li>• Communications in Statistics-Theory and Methods;</li> <li>• Physics of Plasmas.</li> </ul>
imput	imputation, imputations, impute, imputed, imputer, imputes, imputing.	<ul style="list-style-type: none"> <li>• Computational Statistics &amp; Data Analysis;</li> <li>• Statistics in Medicine.</li> </ul>
speci	specie, species.	<ul style="list-style-type: none"> <li>• Computational Statistics &amp; Data Analysis;</li> <li>• Statistics in Medicine;</li> <li>• Journal of Econometrics;</li> <li>• Journal of the American Statistical Association.</li> </ul>
genom	genome, genomes, genomic, genomics.	<ul style="list-style-type: none"> <li>• Bioinformatics;</li> <li>• Systematic Biology.</li> </ul>
intervent	intervention, interventional, interventions.	<ul style="list-style-type: none"> <li>• Statistics in Medicine;</li> <li>• Pharmacoeconomics.</li> </ul>
growth	growth, growths.	<ul style="list-style-type: none"> <li>• Physics of Plasmas;</li> <li>• Nuclear Fusion;</li> <li>• Forest Ecology and Management.</li> </ul>
half	half.	<ul style="list-style-type: none"> <li>• Journal of Clinical Pharmacology;</li> <li>• Physics of Plasmas;</li> <li>• Nuclear Fusion.</li> </ul>

Quadro 28 – Relação de termos e seus periódicos

Radical	Termo	Periódico
dispers	dispersal, dispersals, dispersed, disperser, dispersers, dispersing, dispersion, dispersions, dispersive, dispersivities.	<ul style="list-style-type: none"> <li>• Communications in Statistics -Theory and Methods;</li> <li>• Computational Statistics &amp; Data Analysis;</li> <li>• Journal of Statistical Computation and Simulation.</li> </ul>
stress	stress, stressed, stresses, stressing	<ul style="list-style-type: none"> <li>• Communications in Statistics -Theory and Methods;</li> <li>• Communications in Statistics -Simulation and Computation.</li> </ul>
causal	causal, causalities, causality, causally.	<ul style="list-style-type: none"> <li>• Journal of Econometrics;</li> <li>• Statistics in Medicine.</li> </ul>
circuit	circuit, circuits	<ul style="list-style-type: none"> <li>• Electronics Letters;</li> <li>• Analog Integrated Circuits and Signal Processing.</li> </ul>
threshold	threshold, thresholded, thresholding, thresholds.	<ul style="list-style-type: none"> <li>• Nuclear Fusion;</li> <li>• Environmetrics;</li> <li>• Statistics in Medicine;</li> <li>• Computational Statistics &amp; Data Analysis.</li> </ul>
rich	rich, richly, richness.	<ul style="list-style-type: none"> <li>• Plos One;</li> <li>• Water Resources Research;</li> <li>• Bioinformatics;</li> <li>• BMC Bioinformatics.</li> </ul>
autocorrel	autocorrelated, autocorrelation, autocorrelations.	<ul style="list-style-type: none"> <li>• Journal of Econometrics;</li> <li>• Computational Statistics &amp; Data Analysis.</li> </ul>
precipit	precipitable, precipitation, precipitator.	<ul style="list-style-type: none"> <li>• Journal of Hydrology;</li> <li>• Water Resources Research.</li> </ul>
inequ	inequalities, inequality.	<ul style="list-style-type: none"> <li>• Journal of Econometrics;</li> <li>• Econometrica.</li> </ul>
learn	learn, learned, learning, learns.	<ul style="list-style-type: none"> <li>• Computational Statistics &amp; Data Analysis;</li> <li>• Neural Networks;</li> <li>• Bioinformatics.</li> </ul>
taxa	taxa.	<ul style="list-style-type: none"> <li>• Systematic Biology;</li> <li>• Molecular Biology and Evolution.</li> </ul>
price	price, prices, pricing.	<ul style="list-style-type: none"> <li>• Journal of Econometrics;</li> <li>• Computational Economics.</li> </ul>
children	children.	<ul style="list-style-type: none"> <li>• Journal of Clinical Pharmacology;</li> <li>• Antimicrobial Agents and Chemotherapy.</li> </ul>
nest	nest, nested, nesting, nests.	<ul style="list-style-type: none"> <li>• Journal of Econometrics;</li> <li>• Computational Statistics &amp; Data Analysis.</li> </ul>
resist	resist, resistance, resistances, resistant, resisting, resistive, resistivity.	<ul style="list-style-type: none"> <li>• Nuclear Fusion;</li> <li>• Physics of Plasmas.</li> </ul>

Fonte: O autor (2017).

Como o radical '*circuit*' não é relacionado com outros radicais, o periódico em que ele aparece com artigos publicados não aparece para nenhum outro radical.

Já o radical 'learn' aparece em artigos publicados no periódico *Computational Statistics & Data Analysis* e nesse mesmo periódico aparece outros radicais relevantes: 'cluster', 'imput', 'speci', 'disper', 'threshold', 'autocorrel', 'learn' e 'nest'.

Com foi visto anteriormente, a quantidade final de artigos foi de 6.265 unidades e avaliá-los individualmente não seria possível. Como a ideia deste artigo é ter um panorama geral das publicações sobre o tema, apenas alguns artigos foram avaliados individualmente, conforme indicado no Quadro 29.

Quadro 29 – Comentários sobre alguns artigos

Artigo	Comentário
<p><i>Parametric bootstrap tests for continuous and discrete distributions</i></p> <p>SZÜCS, Gábor</p>	<ul style="list-style-type: none"> <li>• <b>Periódico:</b> Metrika</li> <li>• <b>Ano:</b> 2007.</li> <li>• <b>Técnica base:</b> Distribuição binomial negativa</li> <li>• <b>Área de estudo:</b> Distribuição dos números de filhos das famílias da Jordânia.</li> <li>• <b>Resumo:</b> Utiliza a <i>bootstrap</i> paramétrica para identificar parâmetros desconhecidos da distribuição binomial negativa, esta que é considerada a distribuição dos números de filhos das famílias jordanianas.</li> </ul>
<p><i>Robust optimization in simulation: Taguchi and Response Surface Methodology</i></p> <p>DELLINO, Gabriella; KLEIJNEN, Jack P. C.; MELONI, Carlo</p>	<ul style="list-style-type: none"> <li>• <b>Periódico:</b> <i>Int. Journal of Production Economics</i></li> <li>• <b>Ano:</b> 2010.</li> <li>• <b>Técnica base:</b> visão de Taguchi e <i>Response Surface Methodology</i>.</li> <li>• <b>Área de estudo:</b> <i>Economic Order Quantity</i></li> <li>• <b>Resumo:</b> Criação de uma metodologia para a otimização da simulação, utilizando a visão de Taguchi com a técnica <i>Response Surface Methodology</i> (RSM). A metodologia é aplicada utilizando a <i>bootstrap</i> paramétrica no modelo <i>Economic Order Quantity</i> (EOQ).</li> </ul>
<p><i>The simultaneous inference of zooplanktivorous fish and macrophyte density from sub-fossil cladoceran assemblages: a multivariate regression tree approach</i></p> <p>DAVIDSON, Thomas A.; SAYER, Carl D.; PERROW, Martin; BRAMM, Mette; JEPPESSEN, Erik.</p>	<ul style="list-style-type: none"> <li>• <b>Periódico:</b> <i>Freshwater Biology</i></li> <li>• <b>Ano:</b> 2010</li> <li>• <b>Técnica base:</b> Coeficiente de determinação, árvores de regressão</li> <li>• <b>Área de estudo:</b> Biologia, ambiente natural, <i>paleolimnology</i> (estudos de lagos antigos), relacionamento entre espécies e ambiente.</li> <li>• <b>Resumo:</b> Desenvolve a técnica de árvore de regressão multivariada (<i>multivariate regression tree</i> – MRT) para quantificar o relacionamento entre espécies-ambiente de lagos antigos. A simulação <i>bootstrap</i> foi utilizada para comparar o coeficiente de determinação (<math>r^2</math>), bem como outros índices, de diversos modelos.</li> </ul>

## Quadro 29 – Comentários sobre alguns artigos

Artigo	Comentário
<p><i>Bootstrap prediction intervals for Markov processes</i> PAN, Li; POLITIS, Dimitris N.</p>	<ul style="list-style-type: none"> <li>• <b>Periódico:</b> <i>Computational Statistics and Data Analysis</i></li> <li>• <b>Ano:</b> 2016</li> <li>• <b>Técnica base:</b> Séries temporais, estimativa de intervalo, cadeia de Markov.</li> <li>• <b>Área de estudo:</b></li> <li>• <b>Resumo:</b> Compara três modelos de simulação das cadeias de Markov, utilizando a simulação <i>bootstrap</i>, para estimar um intervalo para o próximo valor de uma série temporal.</li> <li>• <b>Observações:</b> Este estudo utiliza dados artificiais, não demonstrando uma área de estudo específica. Sabe-se que o estudo de séries temporais é uma técnica que pode ser aplicada a diversas áreas.</li> </ul>
<p><i>Physical ageing and lifetime prediction of polymer membranes for gas separation processes</i> Nils Müller, Ulrich A. Handge, Volker Abetz</p>	<ul style="list-style-type: none"> <li>• <b>Periódico:</b> <i>Journal of Membrane Science</i></li> <li>• <b>Ano:</b> 2016</li> <li>• <b>Técnica base:</b> Curva de regressão, intervalo de confiança, modelo Struik.</li> <li>• <b>Área de estudo:</b> Tempo de vida de membranas de polímeros.</li> <li>• <b>Resumo:</b> Utiliza o método <i>bootstrap</i> para avaliar a previsão de tempo de vida de membranas de polímeros no modelo Struik.</li> </ul>

Fonte: O autor (2017).

Conforme os dados apresentados, foram avaliados cinco artigos, sendo um de 2007, dois de 2010 e dois de 2016. Os artigos deste quadro mostram que a utilização da simulação *bootstrap* é dividida em pelo menos dois grandes grupos: o primeiro é a área de estudo do artigo, sociologia, economia ou biologia; o segundo é a técnica aplicada no estudo, por exemplo, distribuição binomial negativa, coeficiente de determinação, cadeia de Markov ou modelo Struik. Além disso, pode-se generalizar a utilização da simulação *bootstrap* da seguinte forma: o artigo utilizou a simulação *bootstrap* na área de <área de estudo>, utilizando a técnica <técnica base>.

Neste tópico foi visto a evolução da publicação dos artigos sobre a simulação *bootstrap*, o relacionamento entre os principais radicais com os artigos em que foram publicados e a análise de alguns artigos. O próximo tópico tratará do relacionamento entre todas essas informações e o que isso significa.

## 5 Análise dos resultados

A evolução dos artigos se deu a partir de 1979, que é quando foi criada a técnica *bootstrap*. Houve a publicação de alguns artigos antes desse ano, mas seu número é irrisório em relação ao total de artigos encontrados.

Para utilizar a simulação *bootstrap* são necessários dois grupos de técnicas: as técnicas bases e as da área de estudo. Os artigos de Szücs (2007) e de Dellino; Kleijnen; Meloni (2010) são exemplos dessa questão. Esse é um padrão que se repete nos artigos sobre simulação *bootstrap*: existe uma área de estudo. Nessa área, é possível aplicar alguma técnica numérica (estatística, economia, química, física ou outra área) que pode ser analisada com a simulação *bootstrap*. Pode-se concluir que, assim como as técnicas estatísticas, a simulação *bootstrap* pode ser aplicada a qualquer área de estudo que possuam dados para analisar.

A simulação *bootstrap* também pode ser empregada para analisar uma técnica em si, sem especificar uma área de estudo. Exemplo dessa questão pode ser vista no artigo de Pan & Politis (2016).

Os autores Müller; Handge; Abetz (2016) dão a seguinte definição de *bootstrap*: "*Bootstrapping* é um método estatístico de reamostragem de dados, o qual deriva informações de um único conjunto de dados  $x_1, x_2, \dots, x_n$  através da repetição de análises estatísticas nas várias reamostras desse conjunto." (p. 36). Dito de outra forma, é um método de reamostragem dos dados originais em que se realiza análises estatísticas a cada reamostra. Os autores não especificam as análises estatísticas, significando que a simulação *bootstrap* pode ser utilizada para qualquer análise estatística. Os dados desta pesquisa confirmam essa afirmação.

Após apresentar os 25 termos mais importantes conforme a técnica TF-IDF e utilizá-los para identificar artigos de determinado periódico, a partir do nome dos periódicos foi possível inferir que a simulação *bootstrap* é utilizada nas seguintes áreas, além de estatística: circuito integrado, processamento de sinais, quimioterapia, bioinformática comunicação, computação, econometria, análise de dados, ecologia, farmacologia, hidrologia, biologia molecular, redes neurais, fusão nuclear, medicina e hidrologia.

## 6 Conclusão

A simulação *bootstrap* é uma técnica estatística que pode ser aplicada a diversas áreas de estudos. A publicação de artigos vem evoluindo ao longo do tempo, desde que foi criada em 1979, e ainda existe espaço para pesquisas relacionadas a ela.

As publicações utilizando a simulação *bootstrap* a aplicam em uma determinada área de estudo, como biologia molecular, utilizando determinada fórmula ou técnica, como análise de correlação ou valor do dinheiro no tempo.

As principais conclusões deste artigo é que as publicações sobre a simulação *bootstrap* estão evoluindo ao longo do tempo, que ela tem sido aplicada a várias áreas da ciência e que pode ser aplicada a qualquer técnica que tenha uma fórmula, sendo avaliada a partir de dados empíricos, reais ou fictícios.

## Referências

DAVIDSON, Thomas A. et al. The simultaneous inference of zooplanktivorous fish and macrophyte density from sub-fossil cladoceran assemblages: a multivariate regression tree approach. **Freshwater Biology**. v. 55, p. 546-564, 2010.

DELLINO, Gabriella; KLEIJNEN, Jack P. C.; MELONI, Carlo. Robust optimization in simulation: Taguchi and Response Surface Methodology. **Int. Journal of Production Economics**. v. 1287, p. 52-59, 2010.

INGERSOLL, Grant S.; MORTON, Thomas S.; FARRIS, Andrew L. **Taming text**: how to find, organize, and manipulate it. Shelter Island: Manning, 2013.

MANNING, Christopher D.; RAGHAVAN, Prabhaka; SCHÜTZE, Hinrich. **An introduction to information retrieval**. Cambridge: Cambridge University, 2009. Disponível em: <<http://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf>>. Acesso em: 11 set. 2016.

MÜLLER, Nils; HANDGE, Ulrich A.; ABETZ, Volker. Physical ageing and lifetime prediction of polymer membranes for gas separation processes. **Journal of Membrane Science**. n. 516, p. 33-46, 2016.

PAN, Li; POLITIS, Dimitris N. Bootstrap prediction intervals for Markov processes. **Computational Statistics and Data Analysis**. v. 100, p. 467-494, 2016.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2016. Disponível em: <<http://www.R-project.org/>>. Acesso em: 1 dez. 2016.

SZÜCS, Gábor. Parametric bootstrap tests for continuous and discrete distributions. **Metrika**. v. 67, p. 63-81, 2007.

## APÊNDICE B - ALGORITMOS DESENVOLVIDOS

A metodologia desta pesquisa, de maneira ampla, envolveu coletar a amostra de projetos, simular sua execução e analisar os resultados. Para isso, foram desenvolvidos programas utilizando a linguagem de programação Java e o ambiente estatístico R (R CORE TEAM, 2016).

De uma maneira geral, as simulações foram feitas utilizando o Java e os cálculos estatísticos foram feitos utilizando o **R**. O Quadro 30 relaciona os programas desenvolvidos bem como a linguagem utilizada. Neste quadro, os arquivos que acabam com a extensão **.java** foram desenvolvidos com o Java e os que acabam com a extensão **.R** foram desenvolvidos em **R**.

Quadro 30 – Programas de computador desenvolvidos nesta pesquisa

Programa	Linguagem
Transformar arquivos de projetos de formato Patterson para formato .prj.	PattersonParaPrj.java
Calcular os indicadores topológicos.	CalcIndicTopologicos.java
Coletar a duração planejada das atividades de todos os projetos.	ColetarDuracaoPlanejada.java
Sortear a duração das atividades dos projetos.	SortearDuracoesDasAtividades.java
Calcular a duração e folgas totais dos projetos com as durações sorteadas.	ExecutaProjetoComDadosSimulados.java
Calcular os índices de sensibilidade.	CalcularSensibilidade.java
Efetuar teste de igualdade da duração planejada com as durações simuladas	DtplgualDtps.R
Verificar a igualdade das distribuições da duração total do projeto: TR x B-1, TR x B-R.	MesmasDistrs.R
Verificar se os índices de sensibilidade estão correlacionados: TR x B-1, TR x B-R	CorrelacaoIndiceSensib.R

Fonte: O autor (2017).

Além dos algoritmos citados, foram desenvolvidos algoritmos específicos para o cálculo da rede do projeto, para que se pudesse simular a execução do cronograma, baseados nas definições de Vanhoucke (2013), com algumas adaptações em questão de desempenho.

Para realizar o cálculo da rede do projeto é necessário percorrer os caminhos da rede do início para o fim (calcular a ida) e do fim para o início (calcular a volta). Esse é um processo recursivo com tempo exponencialmente proporcional à quantidade de atividades.

A mudança feita no algoritmo de cálculo do caminho foi efetuar a ordenação topológica (Quadro 31) antes de calcular a ida e a volta. Com a ordenação topológica, as atividades ficam ordenadas de acordo com a relação de precedência entre elas, como consequência, é possível calcular a ida e a volta com tempo linear. O resultado foi que o tempo de simulação passou de 11 dias para menos de duas horas.

#### Quadro 31 – Algoritmo para ordenação topológica

```

Encontrar todos os vértices sem entradas: VSE
Criar uma lista com a ordenação topológica: ordemTopo
Criar uma fila
Colocar na fila os VSE
Enquanto a fila não estiver vazia
  V <- fila.proximoElemento
  Para cada filho de V faça
    ordemTopo.adicionar(ligação V -> filho)
  Se o filho não tiver outros pais E possuir filhos
    Colocar os filhos do filho na fila
  Fim-se
Fim-para
Fim-enquanto
Retorne ordemTopo

```

Fonte: Baseado em GOODRICH; TAMASSIA; GOLDWASSER (2014).

O Quadro 32 mostra o cálculo da ida e o Quadro 33 mostra o cálculo da volta.

### Quadro 32 – Cálculo da ida da rede do projeto

```

A ordemTopológica é uma lista com a ordenação topológica
Configurar como zero o início das atividades sem predecessoras
Para cada relação na ordemTopológica faça
  predecessora = relação.pegueAtividadePredecessora()
  sucessora = relação.pegueAtividadeSucessora()
  proximoInicio = predecessora.pegueTerminoMaisCedo()
  Se não é predecessora.marcoDeControle()
    e não é sucessora.marcoDeControle() Então
    proximoInicio++
Fim-se
proximoInicio += relacao.pegueIntervaloEmHoras()
Se sucessora.pegueInicioMaisCedo() < proximoInicio Então
  sucessora.configureInicioMaisCedo(proximoInicio)
  Se é sucessora.MarcoDeControle() Então
    sucessora.configureTerminoMaisCedo(proximoInicio)
  Senão
    Long dur = sucessora.pegueDuracaoEmHoras()
    sucessora.configureTerminoMaisCedo(proximoInicio + dur - 1)
  Fim-se
Fim-se
Fim-para
duraçãoDoProjeto = máximo(términoMaisCedo)

```

Fonte: Baseado em VANHOUCKE (2013).

É importante observar no cálculo da ida (Quadro 32) que o início das atividades sem predecessoras é configurado como zero, e, como as atividades estão ordenadas de acordo com sua relação de precedência, na análise da relação, todas as atividades predecessoras já estão com o seu início mais cedo configurado, bastando configurar o seu término mais cedo de acordo com a sua duração.

O término do projeto é o maior término mais cedo das atividades que não possuem sucessoras. E, para essas atividades, o término mais tarde será configurado com o valor do término do projeto; então, é calculada a volta.

### Quadro 33 – Cálculo da volta da rede do projeto

```

A ordemTopológica é uma lista com a ordenação topológica
Configurar o término das atividades sem predecessoras com a
duraçãoDoProjeto
Para cada relacao na ordemTopológica, de trás para frente, faça
  Atividade predecessora = relacao.peguePredecessora()
  Atividade sucessora = relacao.pegueSucessora()
  Long proximoTermino = sucessora.pegueInicioMaisTarde()
  Se não é predecessora.marcoDeControle()
    e não é sucessora.marcoDeControle() Então
      proximoTermino--
  Fim-se
  proximoTermino -= relacao.pegueIntervaloEmHoras()
  Se predecessora.pegueTerminoMaisTarde() > proximoTermino Então
    predecessora.configureTerminoMaisTarde(proximoTermino)
    Se é predecessora.marcoDeControle() então
      predecessora.configureInicioMaisTarde(proximoTermino)
  Senão
    duração = predecessora.pegueDuracaoEmHoras()
    predecessora.configureInicioMaisTarde
      (proximoTermino - duração + 1)
  Fim-se
Fim-se
Fim-para

```

Fonte: Baseado em VANHOUCKE (2013).

Nos arquivos Projeto.java, Atividade.java, Relacao.java e Topologia.java estão todos algoritmos desenvolvidos para a execução desta pesquisa. Além desses, há outros arquivos no pacote com.rjr.projeto, responsáveis pela simulação do cronograma e visualização de alguns gráficos.