

PEDRO HENRIQUE ENDO NICOLINI

ANÁLISE DE SIMILARIDADES PARA A CLUSTERIZAÇÃO DE
GRANDES PROJETOS DE ENGENHARIA DO SETOR DE
PETRÓLEO E GÁS

São Paulo

2014

PEDRO HENRIQUE ENDO NICOLINI

ANÁLISE DE SIMILARIDADES PARA A CLUSTERIZAÇÃO DE
GRANDES PROJETOS DE ENGENHARIA DO SETOR DE
PETROLEO E GAS

Trabalho de Formatura apresentado à Escola
Politécnica da Universidade de São Paulo
para obtenção do Diploma de Engenheiro de
Produção

Orientador: Profa. Dra. Marly Monteiro da
Silva

São Paulo

2014

FICHA CATALOGRÁFICA

Nicolini, Pedro Henrique Endo
Análise de similaridades para a clusterização de
grandes
projetos de engenharia do setor de petróleo e gás /
P.H.E. Nicolini. -- São Paulo, 2014.
115 p.

Trabalho de Formatura - Escola Politécnica da
Universidade
de São Paulo. Departamento de Engenharia de
Produção.

1.Administração de projetos 2.Estatística
multivariada I.Uni-
versidade de São Paulo. Escola Politécnica.
Departamento de Engenharia de Produção II.t.

Áqueles que merecem meu respeito e admiração: vós sabeis

Agradecimentos

Agradeço especialmente a todos aqueles que me auxiliaram no momento final, apesar das dificuldades exógenas que apareceram pelo caminho. Grande seria a luta se tais não estivessem ao meu lado.

Mais especialmente aos esforços do prof. Franco Caron pela idealização do projeto. E mais profundamente a profa. Marly pela aceitação e aperfeiçoamento daquilo que fora antes produzido. Sem seu auxílio e conhecimento com certeza o trabalho não teria chegado aos níveis que chegaram.

Ademais, agradeço a todos que no início da jornada acadêmica fizeram parte desse fortuito caminho que foi a Escola Politécnica da Universidade de São Paulo. E mais especialmente os anos glórios de *Politecnico di Milano*, onde a sinergia entre conhecimento teórico e prático ganhou um novo significado.

Por fim, agradeço ao núcleo familiar e de amigos onde discussões de grande nível sempre foram tratadas, não somente em termos acadêmicos, mas em termos de afeto e sinceridade.

A dificuldade real não reside nas novas idéias, mas em conseguir escapar das antigas.
(John Maynard Keynes)

Resumo

Quando o assunto é dinheiro e oportunidades para se ganhar mais, os holofotes acabam virando nessa direção. Grandes investimentos que geram retorno financeiro ou até mesmo vantagem competitiva para uma empresa estão baseados em projetos. Independentemente do tamanho desse projeto, ou da duração do mesmo, o importante é que se tenha sucesso.

O sucesso em um projeto serve como incentivo a empresa de realizar mais. Mas como garantir o sucesso para os projetos subsequentes? É preciso antes de tudo aprender com os projetos passados, tendo estes havido sucesso ou não. Mas aprender não é uma tarefa fácil, é preciso otimizar o aprendizado e tirar proveito de projetos semelhantes ao que se pretende realizar.

Mas semelhança entre projetos também não é uma tarefa trivial. É preciso saber os pontos em comum entre diferentes ocorrências saber balizar bem a forma de agrupar. É justamente esse o ponto focal do trabalho. Estabelecer critérios para melhor agrupar projetos em *clusters*, a fim de proporcionar um melhor entendimento dos mesmos. Um melhor entendimento que por sua vez facilita o aprendizado para futuros projetos, e ajuda na compreensão do que é esperado para os mesmos.

Assim sendo, pode-se afirmar que uma metodologia que agrupe bem os diferentes projetos com características semelhantes é de grande interesse àqueles que desejam que o projeto obtenha sucesso. E por sucesso leia-se retorno e oportunidades.

E justamente para um ambiente de grandes projetos que será focalizada essa metodologia. O cerne do trabalho é analisar como projetos no setor de Petróleo & Gás podem ser agrupados e estudados, a fim de se ter um facilitador de aprendizado para questões que envolvem grandes investimentos, muitas incertezas, e consideráveis retornos.

Palavras-chave: sucesso de projeto, aprendizado, projeto, cluster, Petróleo & Gás.

Abstract

When we talk about money, and opportunities to raise more, a great attention is given. Huge investments that generates financial return (or capital gain) or even competitive advantage for a company are always based on internal or external projects. It doesn't matter the size, duration, cost, or requirements – the key issue is to achieve success.

The project success serves as an incentive for the company to do more. But, how do we guarantee success for subsequent projects? First, it is necessary to obtain knowledge from past projects, being those successful or not. Learning is not an easy task, it is also necessary to improve and optimize this process and have consequently significant gains from similar projects.

Measuring similarities between projects is not a trivial argument, though. It is necessary first to establish a rigorous criteria and common points that should group the projects. The clustering methodology seems to give a proper way to do so, and this is the focal point of this paper. A robust clustering process gives us a better understanding about a range of projects, and a better understanding leverages the future projects assessment.

Given that, we can say that a cluster methodology is of great interest from those who are concerned about project success. By success we can read return and opportunities.

Finally, we will approach this methodology from the view of Oil & Gas industry. Normally they tend to be huge projects in terms of size, cost, return and opportunities. Moreover, they tend to carry a great portion of uncertainty. So it will be our challenge the attempt of give this industry a different approach for their projects.

Key words: project success, knowledge, project, clustering, Oil & Gas.

Lista de Figuras

<i>Figura 1.1 – Estrutura do Trabalho com abordagem Qualitativa e Quantitativa</i>	<i>25</i>
<i>Figura 2.1 – Fluxograma da revisão literária de projetos.....</i>	<i>27</i>
<i>Figura 2.2 – Categorias fundamentais de variáveis de projeto.</i>	<i>29</i>
<i>Figura 2.3 – Os três paradigmas da Gestão de Projetos.....</i>	<i>31</i>
<i>Figura 2.4 – Trade-off de custo de mudança e influência nos resultados de um projeto.....</i>	<i>33</i>
<i>Figura 2.5 – Hierarquia de critérios e fatores críticos de sucesso estandarizados.....</i>	<i>34</i>
<i>Figura 2.6 – Modelo de análise de CFS's segundo a tipologia de projetos.....</i>	<i>36</i>
<i>Figura 2.7 – Linhas gerais de metodologia para obter sucesso em projetos.....</i>	<i>37</i>
<i>Figura 2.8 – Esquema de exposição dos Dez Drivers para sucesso de projetos.....</i>	<i>38</i>
<i>Figura 2.9 – Modelo SERVQUAL para Definição de Projeto.....</i>	<i>41</i>
<i>Figura 2.10 – As Cinco Práticas Imutáveis para sucesso de um projeto.</i>	<i>43</i>
<i>Figura 2.11 – Etapas da definição da Motivação do estudo conduzido.</i>	<i>46</i>
<i>Figura 2.12 – Linha do tempo da evolução do EVM.....</i>	<i>48</i>
<i>Figura 2.13 – Ilustração gráfica do EVMS, com as curvas que o compõe.....</i>	<i>51</i>
<i>Figura 2.14 – SPI e CPI de um projeto mal executado.....</i>	<i>53</i>
<i>Figura 2.15 – SPI e CPI de um projeto bem executado.....</i>	<i>53</i>
<i>Figura 2.16 – Visão multi-projeto versus projeto único.</i>	<i>57</i>
<i>Figura 2.17 – Impactos da implementação GPP em um único projeto.</i>	<i>57</i>
<i>Figura 2.18 – Dimensões positivas para empresa que aplica o GPP.....</i>	<i>58</i>
<i>Figura 2.19 – Ciclo essencial do GPP.....</i>	<i>60</i>
<i>Figura 2.20 – Visão sistêmica de projetos únicos formando clusters de projetos.</i>	<i>61</i>
<i>Figura 3.1 – Exemplo de gráfico de dispersão para duas e três variáveis.....</i>	<i>66</i>
<i>Figura 3.2 – Exemplo de representação de estrela para observações distintas.</i>	<i>67</i>
<i>Figura 3.3 – Gráfico de Dispersão para capturar o comportamento das variáveis.</i>	<i>71</i>
<i>Figura 3.4 – Dendograma representativo de métodos MHC.</i>	<i>73</i>
<i>Figura 3.5 – Análise visual da diferença dos MHC.</i>	<i>74</i>
<i>Figura 3.6 – Espaço com probabilidade alfa de conter uma observação genérica.....</i>	<i>77</i>
<i>Figura 4.1 – Estrutura do Modelo desenvolvido.</i>	<i>85</i>
<i>Figura 5.1 – Gráficos de Dispersão entre duplas de variáveis escolhidas para o modelo.....</i>	<i>97</i>
<i>Figura 5.2 – Gráfico de Dispersão com a introdução da variável binária não considerada no modelo.</i>	<i>98</i>
<i>Figura 5.3 – Resultado do Single Linkage para a base de projetos.....</i>	<i>102</i>
<i>Figura 5.4 – Gráficos de Dispersão para os pares de variáveis correlatas.....</i>	<i>103</i>
<i>Figura 7.1 – Algoritmo de estandarização de variáveis.....</i>	<i>117</i>
<i>Figura 7.2 – Representação de Estrela para todos os projetos planejados.</i>	<i>120</i>

Lista de Tabelas

<i>Tabela 2.1 – Terminologia proposta pelo padrão do EVMS.</i>	<i>51</i>
<i>Tabela 2.2 – O papel do Gerente de Operações e do Gerente de Projetos dentro da empresa.</i>	<i>57</i>
<i>Tabela 3.1 – Dados multivariados dispostos em forma de matriz.</i>	<i>64</i>
<i>Tabela 5.1 – Base de dados dos projetos planejados.</i>	<i>93</i>
<i>Tabela 5.2 – Base de dados dos projetos acabados.</i>	<i>94</i>
<i>Tabela 5.3 – Índice de Pearson para os pares de variáveis.</i>	<i>100</i>
<i>Tabela 5.4 – Aplicação da MANOVA para teste de H3.</i>	<i>104</i>
<i>Tabela 5.5 – Tabela com os valores da Análise de Discriminantes.</i>	<i>106</i>
<i>Tabela 7.1 – MANOVA expressa em formato de tabela expositiva de resultados.</i>	<i>118</i>
<i>Tabela 7.2 – Matriz de similaridades entre projetos planejados.</i>	<i>121</i>
<i>Tabela 7.3 – Matriz de similaridades entre projetos finalizados.</i>	<i>122</i>
<i>Tabela 7.4 – Resultado de iterações do k-médias para projetos planejados.</i>	<i>123</i>
<i>Tabela 7.5 – Resultado de iterações do k-médias para projetos finalizados.</i>	<i>124</i>

SUMÁRIO

1	INTRODUÇÃO.....	23
1.1	MOTIVAÇÃO.....	24
1.2	ESTRUTURA DO TRABALHO.....	24
1.2.1	<i>Revisão da Literatura.....</i>	<i>25</i>
1.2.2	<i>Desenvolvimento, Aplicação e Conclusão do Modelo.....</i>	<i>26</i>
2	REVISÃO DA LITERATURA DE GESTÃO DE PROJETOS.....	27
2.1	ENTENDENDO O PROJETO.....	27
2.1.1	<i>Como definir um Projeto.....</i>	<i>28</i>
2.1.2	<i>Grupos de Variáveis a serem consideradas.....</i>	<i>28</i>
2.1.3	<i>As fases do projeto.....</i>	<i>31</i>
2.1.4	<i>Definindo Sucesso de Projetos.....</i>	<i>33</i>
2.1.5	<i>Os Dez Drivers para Sucesso.....</i>	<i>37</i>
2.1.6	<i>Os Cinco Princípios Imutáveis.....</i>	<i>39</i>
2.1.7	<i>As Cinco Práticas Imutáveis.....</i>	<i>42</i>
2.1.8	<i>Um Resumo do Objeto de Estudo.....</i>	<i>44</i>
2.2	CAPTANDO A MOTIVAÇÃO DO ESTUDO.....	45
2.2.1	<i>A visão de um único projeto.....</i>	<i>47</i>
2.2.2	<i>Gestão de Portfólio de Projetos.....</i>	<i>56</i>
2.2.3	<i>Resumo do Parágrafo.....</i>	<i>61</i>
2.3	RESUMO DO CAPÍTULO.....	62
3	REVISÃO DA LITERATURA MATEMÁTICA.....	63
3.1	FERRAMENTAS ESTATÍSTICAS E REPRESENTAÇÃO DOS DADOS.....	63
3.1.1	<i>Estatística Descritiva Básica.....</i>	<i>63</i>
3.1.2	<i>Técnicas de Representação Gráfica.....</i>	<i>65</i>
3.1.3	<i>Métodos para Cálculo de Distâncias.....</i>	<i>67</i>
3.1.4	<i>Comentários Finais.....</i>	<i>68</i>
3.2	TEORIA DOS CLUSTERS.....	69
3.2.1	<i>O Cluster e a Análise de Clusters.....</i>	<i>69</i>
3.2.2	<i>Exigências do Conjunto de Dados.....</i>	<i>71</i>
3.2.3	<i>Algoritmos de Clusterização.....</i>	<i>72</i>
3.2.4	<i>Análise de Discriminante.....</i>	<i>76</i>
3.3	RESUMO DO CAPÍTULO.....	77
4	DESENVOLVIMENTO DO MODELO.....	78
4.1	O OBJETIVO.....	78
4.2	AS HIPÓTESES.....	79
4.2.1	<i>Hipótese Um.....</i>	<i>80</i>
4.2.2	<i>Hipótese Dois.....</i>	<i>81</i>
4.2.3	<i>Hipótese Três.....</i>	<i>82</i>
4.2.4	<i>Hipótese Quatro.....</i>	<i>83</i>
4.3	DESENVOLVIMENTO DO MODELO.....	84
4.3.1	<i>Tratamento dos Dados.....</i>	<i>85</i>
4.3.2	<i>Similaridades.....</i>	<i>86</i>
4.3.3	<i>Clusterização.....</i>	<i>87</i>
4.3.4	<i>Análise Final.....</i>	<i>88</i>
5	APLICAÇÃO DO MODELO E RESULTADOS OBTIDOS.....	90
5.1	O RECOLHIMENTO DOS DADOS E O CONTEXTO.....	90
5.1.1	<i>Recolhimento do Dados.....</i>	<i>90</i>
5.1.2	<i>Contexto.....</i>	<i>90</i>
5.2	RODANDO O MODELO.....	96

5.2.1	<i>Primeira Etapa: Tratamento dos Dados</i>	96
5.2.2	<i>Segunda Etapa: Similaridades</i>	100
5.2.3	<i>Terceira Etapa: Clusteriação</i>	101
5.2.4	<i>Quarta Etapa: Análise Final</i>	105
5.3	DISCUSSÃO FINAL	107
6	CONCLUSÃO	109
	APÊNDICE A	111
	APÊNDICE B	115
	APÊNDICE C	116
	APÊNDICE D	118
	APÊNDICE E	120
	APPENDIX F	121
	APÊNDICE H	123
8	BIBLIOGRAFIA	125

1 Introdução

Grandes projetos de engenharia não tomam forma sem que haja por traz um estruturado pensamento acerca do assunto. Normalmente, estes envolvem grandes quantidades de recursos, e por consequência a tolerância a erro é zero. Tudo isso é uma questão anacrônica, afinal é um raciocínio que permeia muitas civilizações ao decorrer dos tempos.

Mas antes de aprofundar nesse tópico, seria interessante abordar o conceito de Ciclo PDCA para Gestão da Qualidade: planejar (Plan), executar (Do), verificar (Check) e atuar (Act) [Shewhart, 1931, 1939]. Grandes projetos levam em consideração essa abordagem também, mas com certas ponderações. A fase de planejamento tem uma importância mais visível, afinal é o momento do projeto onde ainda existem *gaps* para erros – tudo ainda está no papel, então o custo de mudança ainda é baixo. A fase de execução também tem grande relevância, afinal aquilo que foi planejado deve ser executado. Aqui, os custos de mudança são consideráveis. A fase de verificação é o pesadelo dos projetistas. Aqui, se aquilo que foi planejado não foi bem executado, o projeto inteiro é um fracasso – os custos de mudança são praticamente os custos totais do projeto. Finalmente, a fase de atuação determina se o resultado do projeto está de acordo com as necessidades sentidas ex-ante. Nada se pode fazer nessa etapa, a não ser aceitar o projeto acabado.

Imagine-se agora um arquiteto com a missão de desenhar um prédio. Gastam-se horas e mais horas desenhando, desde os *sketches* primários até o CAD finalizado. O planejamento do arquiteto sobre o prédio é este, e assim é finalizado. Ele não sabe, tampouco quer saber se aquilo pode ser construído – ele só se importa em ver sua construção de pé.

Mas, pensando como um engenheiro, os detalhes da execução tem que ser medidos. A física das estruturas são de suma importância, nem tudo que o arquiteto desenha pode ser construído. Obviamente, tem-se um conflito de interesses. Assim, surge a importância do planejamento, esse é o momento de mudar o projeto e se adequar a todos *stakeholders*. O resultado final é fruto dessa interação inicial, além de todos os possíveis conflitos que podem surgir durante as outras fases.

Dessa forma, entende-se um projeto como sendo algo planejado, executado, verificado e com um propósito. Além disso, é um esforço que toma em consideração a visão e interesse de vários *stakeholders*. Por fim, o projeto depende da mobilidade de uma grande quantidade de recursos. Em suma, é algo que deve ter-se muito cuidado e atenção.

Fazer errado um projeto, portanto, significa não somente um dispêndio de recursos, mas também uma perda de tempo e de credibilidade do realizador. Mas então como fazer para não errar um projeto?

1.1 Motivação

A motivação do estudo surge unicamente de uma palavra: erro. Projetos são esforços com nenhuma tolerância a erros, afinal, são *outcomes* únicos, e portanto, um erro significa “fazer tudo de novo”.

Um ótimo jeito de acabar com os erros é aprender com eles. Os projetos que já passaram, e que não obtiveram sucesso não devem ser simplesmente jogados no lixo. Um estudo aprofundado acerca das causas do erro podem servir como ótima base para que futuros projetos não sigam a mesma linha.

É justamente na onda do “knowledge and learning” que o dado trabalho se coloca. Um trabalho em conjunto com a *Politecnico di Milano*, onde o professor Franco Caron possui um grupo de estudos sobre projetos no mercado de Petróleo & Gás. Dar-se-á, portanto, continuidade aos trabalhos com a finalidade de analisar a capacidade de projetos semelhantes serem utilizados para “ensinar” projetos novos a obterem sucesso.

A grande discussão nesse estudo será justamente entender como projetos podem conversar, ou seja, quais projetos podem trazer ensinamentos para quais outros projetos? A palavra chave, portanto, é “semelhança”. Projetos semelhantes tendem a ter caminhos e resultados semelhantes, mas antes de validar essa hipótese, o desafio maior é identificar similaridades entre projetos e agrupá-los segundo tais similaridades.

1.2 Estrutura do Trabalho

De maneira resumida, este trabalho será dividido em cinco partes: uma revisão literária sobre projetos, uma revisão sobre ferramentas matemáticas, desenvolvimento de um modelo, a aplicação do mesmo e por fim a conclusão do estudo. A Figura 1.1 a seguir demonstra o andamento proposto para o trabalho. Vale aqui uma ressalva de que durante todo o estudo, transitar-se-á por medidas diversas de análises qualitativas e quantitativas. Assim sendo, propõe-se uma fronteira para essas duas abordagens que ficará clara durante a realização do estudo.

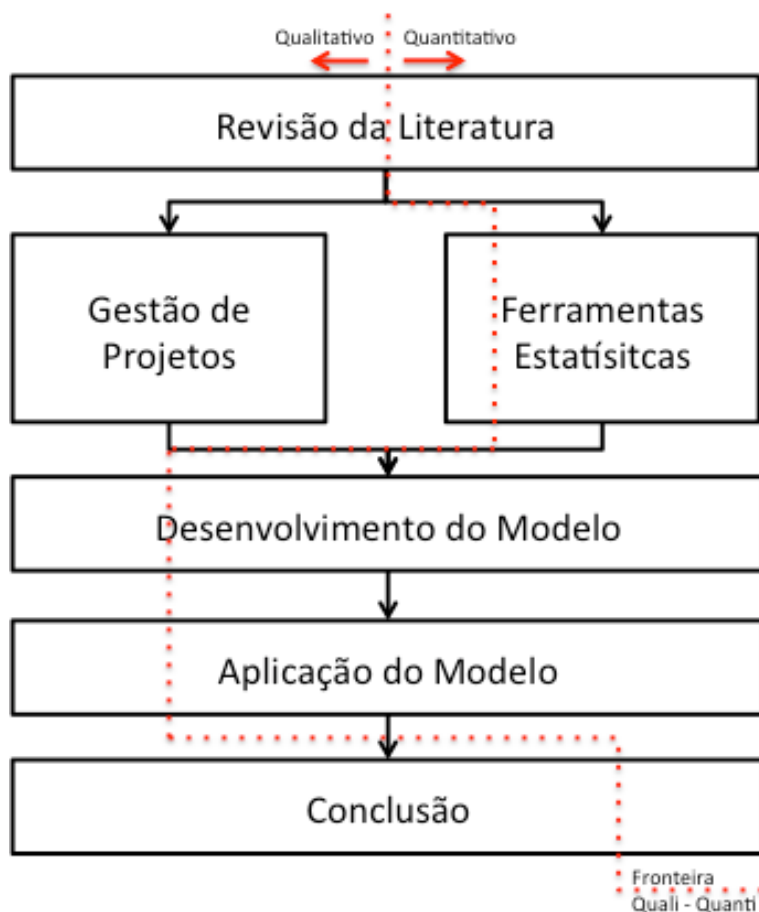


Figura 1.1 – Estrutura do Trabalho com abordagem Qualitativa e Quantitativa

Mas vale agora expor de maneira um pouco mais detalhada como será estruturado o trabalho.

1.2.1 Revisão da Literatura

A revisão da literatura sobre Gestão de Projetos consistirá na pesquisa bibliográfica acerca tópicos relevantes ao estudo. Serão contemplados nessa etapa apenas os conceitos básicos de gestão de projetos, assim como conceitos relacionados ao sucesso de projetos. De maneira geral, serão abordadas cada fase da gestão de projetos, assim como feito pelo Project Management Institute®, e as variáveis envolvidas em cada etapa.

É de suma importância entender bem as variáveis que envolvem um projeto, pois será a partir destas que discutir-se-ão os fatores críticos de sucesso de um projetos. Além disso, são as variáveis de um projeto que determinam como ele se comporta, ou seja, será o ponto focal no momento de classificar dois projetos como semelhantes.

Já na parte do revisão da literatura matemática, será abordado todo ferramental necessário para realização do modelo a ser desenvolvido. Esta etapa é de grande importância porque pretende-se utilizar um ferramental estatístico bastante sofisticado, e portanto, tudo

deve estar muito bem esclarecido. É importante ressaltar também que a abordagem desse capítulo não será somente em aspectos quantitativos mas também qualitativos.

Haja visto que este trabalho é uma continuação de uma série de discussões propostas pelo prof. Franco Caron, os dados fornecidos não são numerosos o suficiente para se confiar exclusivamente numa análise estatística puramente quantitativa. Dessa forma, este trabalho se propõe a dar mais robustez as análises, inserindo um viés qualitativo as mesmas.

1.2.2 Desenvolvimento, Aplicação e Conclusão do Modelo

O desenvolvimento do modelo terá como base tanto a revisão de conceitos de gestão de projetos, como também as ferramentas estatísticas envolvidas. Por isso que citou-se que o trabalho irá transitar em diferentes frentes qualitativas e quantitativas. Isso irá garantir maior robustez do método. Ademais, a abordagem do modelo será no formato linear, ou seja, haverá hipóteses que servem como *milestones* durante o trabalho, respondendo a pergunta: pode-se prosseguir com a análise?

Assim, pode-se partir para a aplicação do modelo, tendo em vista que ainda deve-se considerar a abordagem qualitativa, pois devido o número limitado de projetos e as ferramentas sofisticadas, uma abordagem puramente quantitativa seria extremamente frágil. Lembrando também que o modelo será testado com dados de projetos de Petróleo & Gás, dados que foram fornecidos pelo grupo de estudo do professor Franco Caron.

Finalmente, concluir-se-á o estudo com uma conclusão específica dos resultados obtidos, assim como um indicativo da potencialidade dos mesmos frente a todo o grupo de estudos já realizados nesse âmbito. Ademais, vale lembrar que o objetivo principal aqui é desenvolver o modelo de similaridades de projetos, e ver se este modelo é coerente com a realidade. Todo trabalho subsequente a este é sumariamente bem-vindo.

2 Revisão da Literatura de Gestão de Projetos

Neste capítulo serão apresentados o objetivo e a motivação para o estudo a ser realizado. O principal objetivo deste capítulo é, portanto, dar as bases qualitativas para apresentação do modelo nos capítulos subsequentes. Vale lembrar que não serão abordados com detalhes todos os conceitos de gestão de projetos, tampouco todo o ferramental utilizado no mesmo. Isso porque se tem em mente que é apenas necessário abordar os temas que de fato interagem com o modelo a ser construído.

2.1 Entendendo o Projeto

Primeiramente é necessário entender os macro tópicos referente a gestão de projetos. Isso garante que hajam fronteiras delimitando o sistema a ser estudado. Aqui surge explicitado pela primeira vez o objetivo do trabalho.

O objetivo deste estudo é identificar semelhanças dentro de um range de projetos, coloca-los em grupos – clusters – de acordo com as semelhanças, estudar as características de cada agrupamento e classificar as tipologias de projeto.

Dado o objetivo do trabalho, seria interessante começa-lo abordando o objeto – no caso o projeto em si. Mais especificamente tratar de Gestão de Projetos, já que engloba o projeto em si, e todo o sistema que o permeia.

Para esse propósito, portanto, a abordagem seguida para definir todo o sistema utilizado no estudo pode ser vista na Figura 2.1 a seguir. Mais uma vez, deve-se ressaltar a linearidade do estudo, vendo tópico a tópico em sequência, e movendo somente após entendimento de cada ponto.

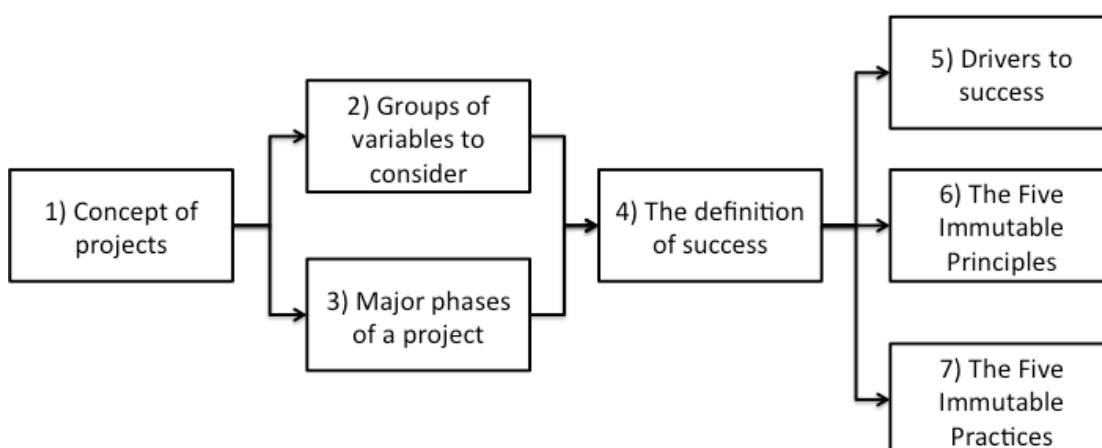


Figura 2.1 – Fluxograma da revisão literária de projetos.

Serão portanto sete passos para revisão da literatura de projetos. Para deixar tudo mais compreensível, o dado capítulo será dividido em sub grupos, cada um contendo um ponto acima mencionado. E por fim um sumário com todas as informações apresentadas.

2.1.1 Como definir um Projeto

Projeto é um processo não-repetitivo, conduzido com a finalidade de produzir um produto ou serviço. Por exemplo, um projeto pode ser algo simples como uma viagem de férias com os amigos, até algo complexo como uma planta industrial. Em outras palavras, um projeto é definido pelo Project Management Institute¹ como sendo:

“ Um projeto é um esforço temporário realizado com a finalidade de criar um produto, serviço ou resultado único.”

(Project Management Institute – Book of Knowledge, PMI, 2013)

Assim sendo, surge de cara duas características latentes e singulares referentes a projetos. A primeira faz menção a escala de tempo: um projeto fornece na sua conclusão um certo resultado dentro de um certo horizonte temporal. A segunda característica é relacionada com o resultado em si: o projeto é uma ocorrência única, e portanto gera um produto ou serviço customizado, ou seja, único também (Caron, 2009).

Tais características são a base das principais marcas registradas de um projeto. A primeira marca que todo projeto tem é o fato de ser “a única chance de fazer a coisa certa”. A segunda marca registrada de todo projeto é que este é governado por um certo conjunto de variáveis que podem ser dependentes ou independentes – em outras palavras, algumas variáveis podem ser controladas, outras simplesmente não (Alleman, 2014).

Daí surge toda complexidade de um projeto. Uma responsabilidade de fazer a coisa certa, com apenas uma chance de realização, tendo em vista que nem todas as variáveis são passíveis de serem controladas.

2.1.2 Grupos de Variáveis a serem consideradas

As variáveis que compõe um projeto foram por muito tempo fundamentalmente divididas em três categorias: variáveis relacionadas a custos, a calendário e medidas de

¹ Project Management Institute: maior associação sem fins lucrativos do mundo voltada para projetos, programas e gestão de portfólios. Além de cuidar de tais assuntos, é responsável também pela normatização dos profissionais da área. Possui mais de 2,9 milhões de profissionais espalhados pelo mundo. Fonte: <http://www.pmi.org/About-Us.aspx>

performance técnica (Technical Performance Measures – TPM). É o chamado velho paradigma da gestão de projetos. A Figura 2.2 a seguir identifica cada categoria, que está relacionada com as seguintes questões:

- Quanto um projeto vai custar?
- Quanto vai demorar para ficar concluído?
- As entregas serão de acordo com os requerimentos?

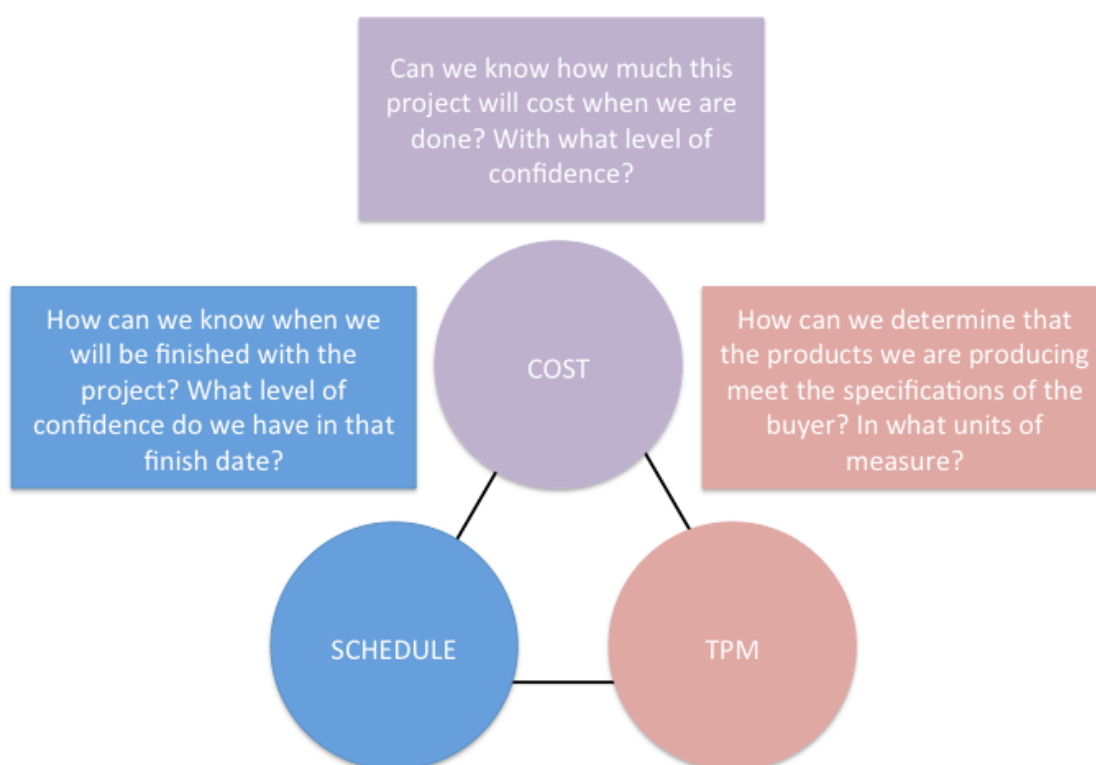


Figura 2.2 – Categorias fundamentais de variáveis de projeto.
(Adaptado de ALLEMAN, 2014.)

O modelo a ser construído e toda a análise a ser feita será baseada em variáveis pertencentes ao velho paradigma, como por exemplo o custo total do projeto, sua duração e adequação do projeto aos requerimentos do cliente.

Entretanto, vale salientar que trata-se de um velho paradigma, cuja denominação “velho” advém do fato de que já não se mensura projetos com base nessas três categorias de variáveis. Isso porque entende-se que tais categorias não são MECE², ou seja, são interdependentes.

² MECE: conceito bastante utilizado, que indica que uma base de informações é Mutuamente Excludente, Coletivamente Exaustiva (MECE). Em outras palavras, um elemento informativo não deve sobrepor outro, e todos os elementos juntos devem ser capazes de explicar completamente um fenômeno sem que falte conteúdo. Fonte: www.caseinterview.com/mece

A primeira evolução, portanto, do velho paradigma foi uma tentativa de trazer informações dentro de um contexto MECE. O segundo paradigma, assim como é chamado, envolve um certo grau de dinamicidade ao projeto, e divide as variáveis nos seguintes grupos: Progresso do projeto, Grau de Modificação e Riscos. Note que Custos, Calendário e Performance são parte integrante do segundo paradigma, mas de maneira espalhada, a fim de que não se sobreponham.

O terceiro paradigma já é um refinamento do pensamento trazido pelo segundo paradigma, e classifica as variáveis nos seguintes grupos: Risco, Valor e Linhas Bases. Esse último paradigma da gestão de projeto considera e engloba os paradigmas anteriores, mas propõe modificações. As variáveis de Valor descrevem os requerimentos dos *stakeholders*, apoiando-se no que estes consideram como valor agregado ao projeto. As Linhas Bases consideram a interação entre custos e tempos de realização do projeto, e demonstram frente as Variáveis de Valor, se o projeto tem capacidades de gerar sucesso. Finalmente, as variáveis de Risco indicam como o Valor e as Linhas Bases podem falhar na interação entre si, e assim o projeto não obter sucesso. Ou seja, não se trata mais apenas de realizar o projeto dentro de certo *budget* e horizonte temporal, segundo alguns requerimentos. Trata-se de trazer valor aos *stakeholders*, considerando as linhas de custo e tempo, e os riscos envolvidos na realização do projeto (Caron, 2009).

A Figura 2.3 a seguir coloca os três paradigmas como tripés na realização de projetos. Lembrando que para os fins desse trabalho, será considerado apenas o primeiro paradigma, pois ele fornece uma visão estática do projeto, e por ser uma abordagem com variáveis interdependentes (custo depende do tempo, e vice-versa), será possível tirar algumas conclusões mais sofisticadas dos resultados. O objetivo principal do estudo é, portanto, fornecer uma linha de comparação baseada na ideia trazida pelos economistas de “estática comparada”. A estática comparada é uma ferramenta que associa duas situações de equilíbrio, ou seja, quando todas as variáveis do modelos já se mostram em estado inercial. Dessa forma é possível traçar conclusões a partir da comparação de dois estados de equilíbrio (Boitani, 1988).

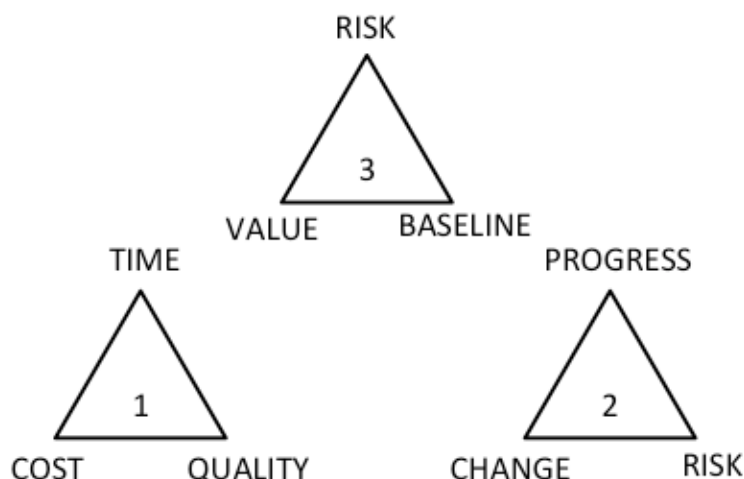


Figura 2.3 – Os três paradigmas da Gestão de Projetos.
(Adapted from CARON, 2009)

Mas vale dizer que na conclusão do estudo, surgirá novamente o terceiro paradigma, pois a continuação proposta deste trabalho envolve uma análise dinâmica da robustez dos resultados aqui obtidos. Em todo o caso, o paradigma primeiro é aquele que deve ser considerado durante a evolução do trabalho.

2.1.3 As fases do projeto

Todo projeto pode ser dividido de maneira padronizada em um determinado número de fases, as quais definirão seu ciclo de vida. Na Introdução foi falado sobre as fases do projeto segundo abordagem do ciclo PDCA. Entretanto, utilizar-se-á agora uma tipologia diferente daquela abordada anteriormente. Segundo PMI, um projeto possui as seguintes fases:

- I. Iniciação
- II. Planejamento e Organização
- III. Execução
- IV. Fechamento

A primeira fase do projeto – Iniciação – está relacionada com a definição do projeto. É nessa fase que serão dialogados os principais objetivos do projeto e a estratégia por traz do mesmo. O produto ou serviço a ser entregue pelo projeto é desenhado em termos de requerimentos dos clientes³.

A segunda fase – Planejamento e Organização – define com alto grau de detalhamento as entregas do projeto (tanto intermediárias como finais). Nessa segunda fase, como já foi falado, o custo de mudanças ainda é relativamente baixo

³ Nesse caso utiliza-se o termo “cliente” que pode tanto servir para designar clientes internos, como clientes externos. Afinal, o projeto pode tanto ser feito para benefício de terceiros, como para benefício da própria empresa que o faz.

A terceira fase – Execução – é o momento em que os recursos do projeto serão ativados, e os *milestones* de conclusão são alcançados. Tanto o *budget* quanto o tempo de realização são determinantes nessa etapa, e as linhas bases tomam grande importância na gestão do projeto.

A última fase – Fechamento – é o momento quando o resultado do projeto é transferido ao cliente, e também o importante momento quando os aprendizados do projeto finalizado são estocados numa base de dados para uso posterior em projetos semelhantes (Caron, 2009).

O trabalho proposto aqui está principalmente relacionado às etapas I, II e IV. A execução toma como referência as linhas bases do projeto, o que já foi dito anteriormente se tratar de um paradigma que não é o escopo do estudo em questão. Lembrando também que se trata de uma análise de estática comparada, portanto, será mais relevante entender projetos com planejamentos semelhantes, e compará-los no momento de seus fechamentos.

Uma questão relevante nesse caso, assim como apontado na introdução, faz referência ao custo de mudanças. As Fases I e II são caracterizadas por custos de mudança muito baixos, ou seja, é possível delinear um outro andamento para o projeto, sem que isso comprometa o *budget* total proposto *ex-ante* (vide Figura 2.4). Isso explica o porque esse estudo foca tanto nas primeiras etapas do projeto, afinal compreendê-las seria tornar o projeto mais susceptível ao sucesso. Em outras palavras, a perfeição alcançada nas duas primeiras etapas minimiza erros na execução, e facilita o Fechamento do projeto. Vale dizer também que grande foco se dará no processo de aprendizado proporcionado pela última etapa do projeto – o Fechamento. É graças a essa etapa que se pode aprender e não cometer erros futuros, além de fornecer insights para as fase de iniciação e planejamento de outros projetos.

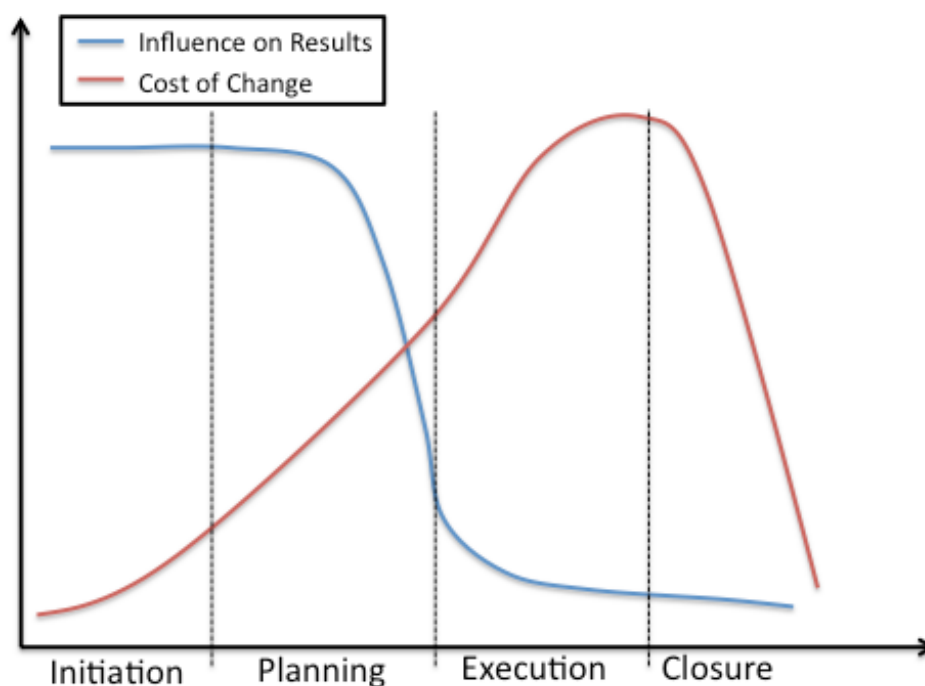


Figura 2.4 – Trade-off de custo de mudança e influência nos resultados de um projeto.
(Adaptado de CARON, 2009)

Assim sendo, um projeto feito e analisado sob a ótica do velho paradigma compromete-se a ter sucesso quando as etapas I, II e IV são bem sucedidas. Como numa linha de montagem, os projetos também podem possuir um aprendizado estruturado, resultando num melhoramento contínuo de sua produção.

2.1.4 Definindo Sucesso de Projetos

Muitos autores indicam que sucesso de um projeto é associado ao velho paradigma, ou seja, é associado a custo, tempo e performance técnica. De maneira simplificada, quando o projeto tem o match entre os valores das variáveis planejadas e os valores do que realmente aconteceu.

Por outro lado, é cada vez mais notado que alguns projetos, mesmo não chegando aos seus objetivos de custo, tempo ou qualidade são considerados sucesso. São casos bastante conhecidos como o Windows da Microsoft, o Macintosh da Apple, o Taurus da Ford. Esses três casos nasceram de projetos que não obtiveram o match de custo ou tempo, mas mesmo assim geraram valor para os *stakeholders* (terceiro paradigma claro aqui) (Shenhar & Dvir, 2007).

Assim sendo, é notório que o primeiro paradigma não preenche a lacuna do sucesso de projetos de maneira satisfatória. Tem-se de considerar também a efetividade de um projeto em alcançar seus objetivos em sentido amplo – ou seja, o valor que ele agrega aos *stakeholders* (Cserhádi & Szabó, 2013).

Na realidade, vê-se então que considerar se um projeto obteve sucesso é uma tarefa bastante complexa. Surge aqui a necessidade de considerar, portanto, os critérios e fatores de sucesso – o que na sua essência é diverso, sendo um critério um regra ou princípio que testa algo, e um fator é um elemento que contribui para um resultado. Todo time de projeto deve levantar os critérios e fatores de sucesso na fase de inicialização, e todo o projeto tem seu próprio *set* de fatores e critérios. Por exemplo, a indústria financeira considera as políticas de estado e de Banco Central como um fator de sucesso em seus projetos (Ika, Diallo, & Thuillier, 2012). No caso de grandes eventos, a satisfação do cliente é um critério que se toma para observar o sucesso de um projeto (Cserháti & Szabó, 2013). Por fim, em projetos urbanos, o grau de satisfação médio dos *stakeholders* é um fator de sucesso, enquanto os conflitos de interesses são os critérios que permeiam o sucesso do projeto (Yu & Kwon, 2011).

Como se pode observar, não é trivial também endereçar os critérios e fatores críticos de sucesso em projetos, devido sua natureza mutável dentro de diferentes tipologias de projetos. Entretanto, existe sim uma standardização dessas classes em alguns macro grupos, dependendo assim somente a ponderação para cada indústria em que o projeto se insere.

Assim sendo, para melhorar a compreensão acerca do assunto, é possível expor os macro tópicos de critérios e fatores críticos de sucesso (Shenhar & Dvir, 2007). A Figura 2.5 mostra como essa relação se dá.

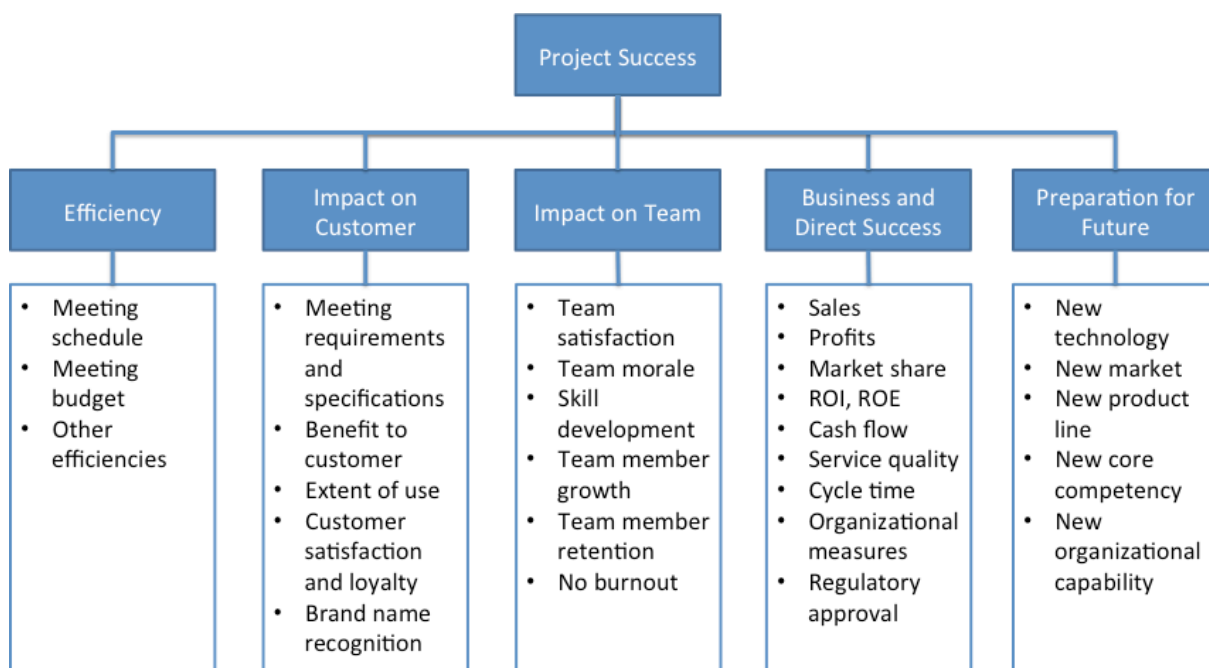


Figura 2.5 – Hierarquia de critérios e fatores críticos de sucesso estandardizados.
(Shenhar & Dvir, 2007)

A Figura 2.5 indica as mais comuns áreas de medida de sucesso de projetos. A tentativa do autor aqui é ser MECE. A figura também elenca os fatores críticos de sucesso (CFS – Critical Factors of Success) dentro de cada macro critério. E vale novamente ressaltar que cada projeto, dentro de cada indústria, irá ponderar aqueles fatores e critérios que melhor se adequam a sua realidade. Um exemplo interessante que surge é o projeto de construção do metrô de Los Angeles (Shenhar & Dvir, 2007). Seria interessante aqui, para melhor explicar o tópico, analisar esse case de maneira resumida. Assim sendo, demonstra-se que os CFS's desse projeto são:

- Eficiência: o projeto alcançou todos os valores planejados de custo, tempo e performance (velho paradigma). Inclusive foi eleito o “Projeto do Ano”, prêmio concebido pelo PMI em 1993;
- Impacto ao Consumidor: esse aspecto foi realmente um problema para o projeto. Sendo L.A. uma cidade basicamente de carros, os consumidores não aderiram a ideia de uso de metro;
- Impacto ao Time: o projeto gerou um grande grau de aprendizado na área, e foi um sucesso para o time de projeto;
- Business e Sucesso Direto: o projeto não obteve a demanda esperada, então o retorno sobre o investimento não justificou todo dispêndio de recursos;
- Preparação para o Futuro: nesse aspecto o projeto foi um sucesso, pois sua realização trouxe a sociedade um alto grau de aprendizagem, e portanto, o melhoramento técnico de novos projetos da mesma indústria.

Vale dizer que o projeto de metrô de Los Angeles foi um projeto sem sucesso. Isso porque, quando analisados os fatores críticos e critérios, o peso dado a cada macro tópico é diferente. Mesmo o projeto tendo sido sucesso em termos de Eficiência, Impacto ao Time e Preparação para o Futuro, esses não são critérios que realmente geraram valor ao projeto. O fato de não ter Impacto ao Consumidor gerou um fracasso, e o metrô não foi considerado um sucesso em âmbito.

Conclusivamente, esse *case* mostra claramente que para obter sucesso, um projeto deve contemplar todos os CFS's. Entretanto, os exemplos mostrados anteriormente dizem o contrário. Surge então a pergunta: De fato, quais são os CFS's que devem estar de acordo para que um projeto obtenha sucesso?

Para responder tal pergunta, seria interessante mudar a perspectiva de análise. Não é apenas o fato de elencar todos os CFS's e verificar sua validade. Seria interessante propor

algum modelo que case a tipologia de projetos com os critérios fortes para seu sucesso (Samset, 1998). A Figura 2.6 mostra a dinâmica desse modelo.

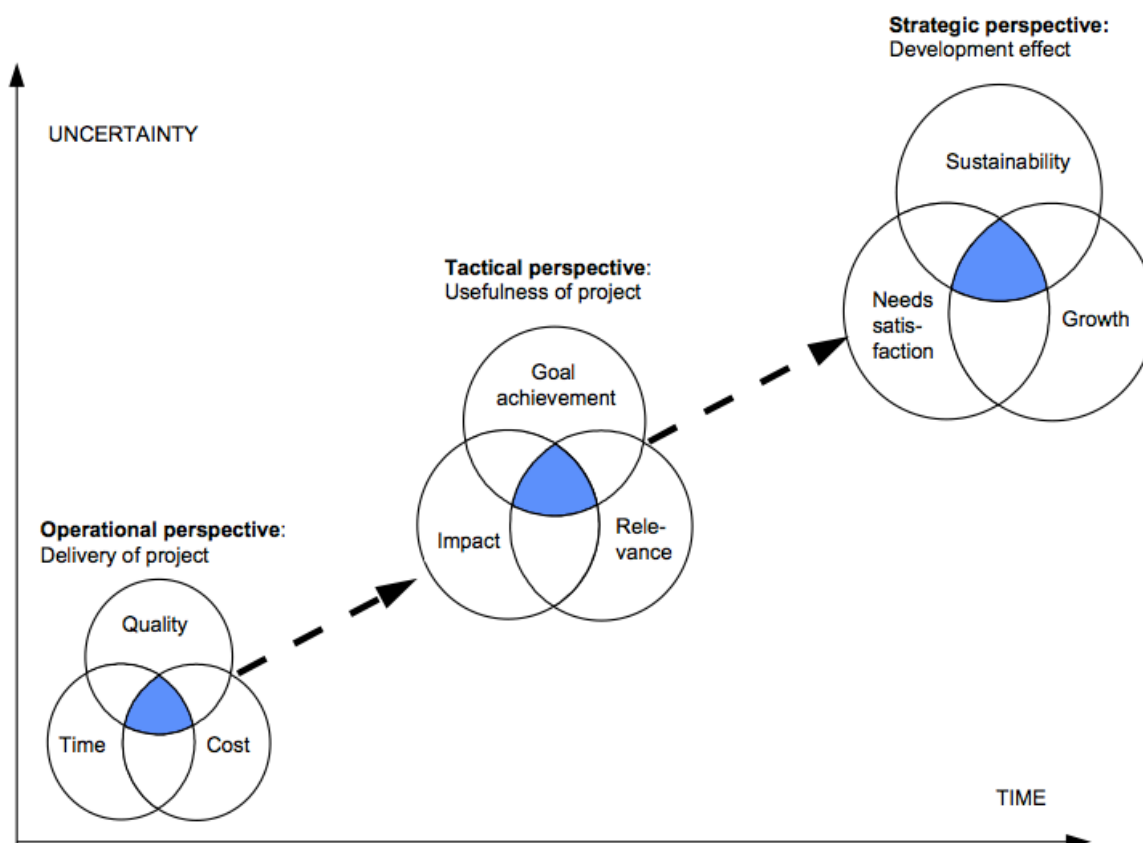


Figura 2.6 – Modelo de análise de CFS's segundo a tipologia de projetos.
(Samset, 1998)

O Modelo exposto acima associa as variáveis de incerteza e tempo sobre os critérios de sucesso de um projeto, e assim, estrutura o raciocínio de quais CFS's que devem ser tomados em consideração para que um projeto seja sucesso. De maneira mais clara, pode-se dizer que a Figura 2.5 e a Figura 2.6 devem ser lidas em conjunto, assim a análise será mais robusta.

Valem então dois comentários para sintetizar tudo expresso até então:

- I. O respeito das restrições impostas pelo velho paradigma indicam que o projeto foi bem sucedido em termos de eficiência, ou seja, foi bem gerido. Mas isso não significa que o sucesso geral foi alcançado.
- II. Considerações sobre incerteza e tempo que permeiam o projeto geram uma dinâmica que expõe os verdadeiros CFS's de um projeto bem sucedido.

Uma vez entendidos o papel dos critérios e fatores para o sucesso de projetos, seria ideal agora propor algum *framework* capaz de auxiliar o alcance do sucesso para o projeto.

Lembrando que no contexto desse trabalho, os projetos são do setor de Petróleo & Gás, e tem-se uma amostra fixa de projetos que serão analisados.

Além disso, o foco desse estudo é a relação entre projetos, e como já dito antes, será utilizado para classificar os projetos o velho paradigma, custo, tempo e performance. Assim sendo, o *framework* que está sendo buscado seria algum que leve em consideração somente esses grupos de variáveis. Vale então a máxima de que nesse caso, o gerente de projeto deve estar preocupado somente com o projeto dentro do budget, tempo e restrições de performance (Shenhar & Dvir, 2007).

Esse caminho é o ideal a ser seguido por dois motivos. Primeiro porque o foco desse estudo é o projeto, sem tomar conta de todo impacto estratégico para a empresa que está desenvolvendo-o. O segundo motivo a ser elencado seria o fato que o grande personagem na análise é o gerente de projeto, pois é ele que se beneficia do aprendizado de projetos passados, para seu uso em projetos futuros. Mais do que isso, é ele que garante a boa execução, e portanto, é ele que deve ter o pleno conhecimento acerca do assunto.

Assim sendo, finalmente pode-se propor um *framework* ideal para alcançar o sucesso de um projeto, tendo em vista o velho paradigma. A metodologia portanto será aquela que, em linhas gerais, está exposta na Figura 2.7 (Alleman, 2014).

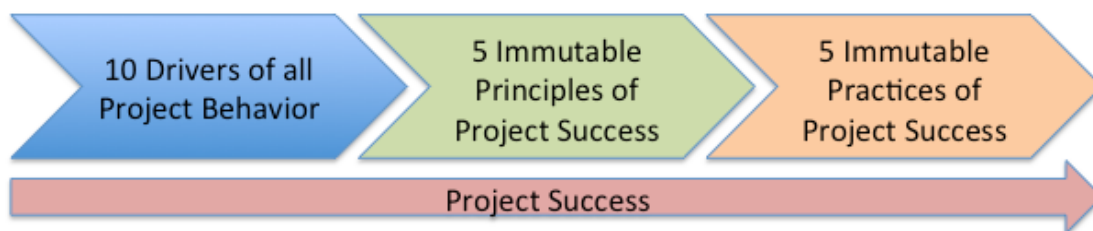


Figura 2.7 – Linhas gerais de metodologia para obter sucesso em projetos.
(Adaptado de ALLEMAN, 2014)

Esse processo proposto acima será agora desmembrado em suas diferentes etapas. Trata-se de uma metodologia não-trivial, mas bastante flexível, abrangendo uma vasta gama de indústrias e tipologias de projetos que podem ser explicados e geridos segundo o velho paradigma.

2.1.5 Os Dez Drivers para Sucesso

A primeira ressalva que se deve ter na busca pelo sucesso de projeto é a consideração dos Dez Drivers que governam o comportamento do velho paradigma. Os Dez Drivers são divididos em três grupos distintos, cada um fazendo menção a uma etapa do ciclo de vida do projeto: Planejando, Executando e Gerindo a Performance.

Ao invés de listar todos os Dez Drivers, será mais proveitoso expô-los na Figura 2.8 a seguir. Depois será feita a discussão com base no que foi visto.

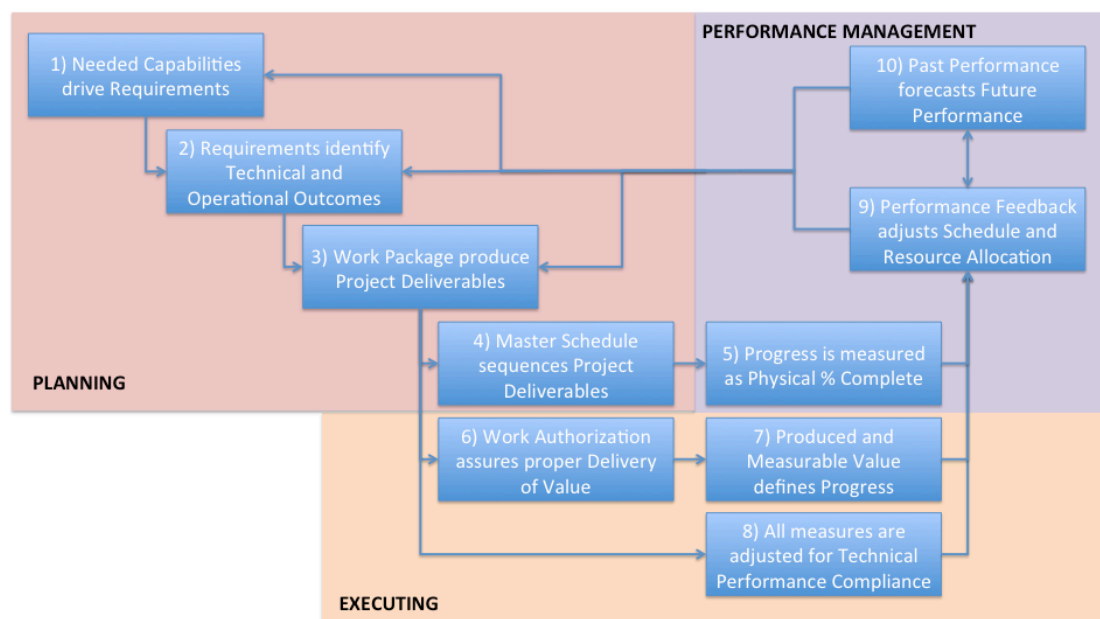


Figura 2.8 – Esquema de exposição dos Dez Drivers para sucesso de projetos.
(Adapted from ALLEMAN, 2014)

Numa primeira análise, pode-se notar que os drivers (9) e (10) são a base para os drivers (1), (2) e (3). Mais do que isso, nota-se que os drivers são cíclicos, ou retroalimentados. Dessa forma, pode-se dizer que Planejar, Executar e Gerir a Performance formam um processo produtivo passível de melhoramento, e portanto, susceptível a aprendizado. Nesse sentido, a primeira hipótese que surge dessa metodologia é que o sucesso é alcançado nas mesmas bases que um processo repetitivo (Alleman, 2014).

Na realidade, o próprio autor comenta que não se trata de um ciclo perfeito, pois se trata de um projeto, com começo, meio e fim. Mas vale uma ressalva de que os projetos passados podem alimentar a base de dados de projetos futuros, e com isso propor um simulacro de processo contínuo.

Essa primeira conclusão é extremamente importante para o estudo proposto – usar-se de outros projetos é de suma importância para atingir o sucesso de futuros projetos, mas fica ainda a questão: tendo um projeto no pipeline, qual pode servir de modelo passado? Surge bem claro aqui a noção de similaridade, algo que será buscado nesse trabalho.

Em resumo, o trabalho deve ser bastante focado nos drivers (9) e (10), pois estes determinam as fronteiras do estudo. É importante também mencionar que o papel dos *stakeholders* é vital para o sucesso dos projetos, mas nesse estudo proposto não será muito aparente, uma vez que já foi dito que o conceito aqui é da estática comparada das situações de equilíbrio, não importando muito as fase de Iniciação.

Finalmente, pode-se dizer que os drivers de (2) a (8) são as consequências do trabalho, uma vez que mensuram a dinâmica do processo de gestão de projetos. O que não significa que não devem ser analisados. O grande ponto em questão aqui é que a metodologia dos Dez Drivers considera o velho paradigma, portanto aborda todos os aspectos da fase de Execução de gestão de projetos, mas como não é escopo do estudo, não será o foco aqui.

2.1.6 Os Cinco Princípios Imutáveis

Os Cinco Princípios Imutáveis são a extensão dos Dez Drivers. Estes são expostos em formas de pergunta, e carregam em si a soma de todos os drivers. Além disso, mais uma vez pode-se ver a correlação com o ciclo de vida do projeto, e acima de tudo a correlação com o velho paradigma.

As cinco indagações que compõe os Cinco Princípios Imutáveis são (Alleman, 2014):

- I. O que o cliente está pedindo⁴?
- II. Como alcançar o pedido do cliente dentro do budget e do tempo pré-estabelecido, e com uma performance coerente?
- III. Como acessar os recursos?
- IV. Como lidar com imprevistos e riscos?
- V. Como mensurar progresso dentro do projeto?

Mais uma vez vale dizer qual o foco desse trabalho, agora no contexto dos Cinco Princípios Imutáveis. A atenção deve principalmente aos questionamentos (I) e (II), e um pequeno comentário sobre o (III). Já o princípio (IV) fala sobre riscos e imprevistos, e dá certa lembrança ao novo paradigma de variáveis. Na realidade é de suma importância pensar nos riscos do projeto, mas no escopo desse trabalho, onde o que conta é a estática comparada, não existe muita abertura para os riscos a imprevistos. Da mesma forma, o princípio (V) carrega com si todo o dinamismo da gestão de projeto, o que também não é escopo desse trabalho, e sim daqueles que serão continuação ativa deste.

Assim sendo, é bom focar apenas naquilo que importa de fato. Começando portanto com uma explicação mais detalhada do princípio (I). Este afirma em sua indagação que o cliente procura algo quando aciona sua necessidade de realizar um projeto. Mas, como identificar o que o cliente de fato precisa? É necessário de fato ter pleno conhecimento daquilo que está sendo esperado do projeto. Esse é com certeza um princípio bastante atuante em prol do sucesso do projeto.

⁴ Independe do fato de ser cliente interno ou externo.

A razão pelo qual o cliente é um ponto focal pode parecer trivial, mas talvez tenha um detalhe que as vezes passa despercebido pelos gerentes de projeto. Saber o que o cliente realmente deseja é a melhor maneira de se estabelecer as fronteiras do projeto, ou seja, categorizar tudo aquilo que é intrínseco à perfeita entrega, e aquilo que é supérfluo e que, portanto, é somente um gasto de recursos que não gera valor ao projeto.

A grande atuação do princípio (I) se dá em duas fases do ciclo de vida do projeto. A primeira, obviamente é a Iniciação, onde serão discutidas todas as características gerais do projeto. Outra fase que fica bastante aparente o papel do cliente é na fase de Planejamento, onde todas as linhas base serão desenhadas, a fim de que sejam implementadas durante a execução do projeto.

Muitos autores caracterizam o primeiro princípio imutável como sendo a *Definição do Projeto*. O momento em que o primeiro sketch é desenvolvido, ou metaforicamente falando, onde é plantada a semente.

Definição de Projeto, assim como é encontrado na literatura, é o processo utilizado para identificar as necessidades dos *stakeholders* e transportar essas necessidades para dentro do projeto. Assim como dito anteriormente, esse processo é bastante aparente no início do ciclo de vida de todo o projeto (Cano & Lidón, 2011).

Existem algumas metodologias bem estruturadas para se realizar a correta *Definição de Projeto*, e assim, limitar um erro que pode ser carregado por todo ciclo de vida do projeto, e minar seu sucesso. Um ótimo modelo de *Definição de Projeto* que pode ser citado é o chamado SERVQUAL. Esse modelo é bastante enxuto, e trata de todos os *gaps* que podem interferir na correta visualização daquilo que o cliente realmente deseja. O modelo trata das interfaces entre expectativas do cliente, percepções do produtor do serviço, e entrega do serviço propriamente dita (Westbrook & Peterson, 1998). A Figura 2.9 abaixo demonstra todas as relações e gaps encontrados segundo as interfaces mencionadas acima.

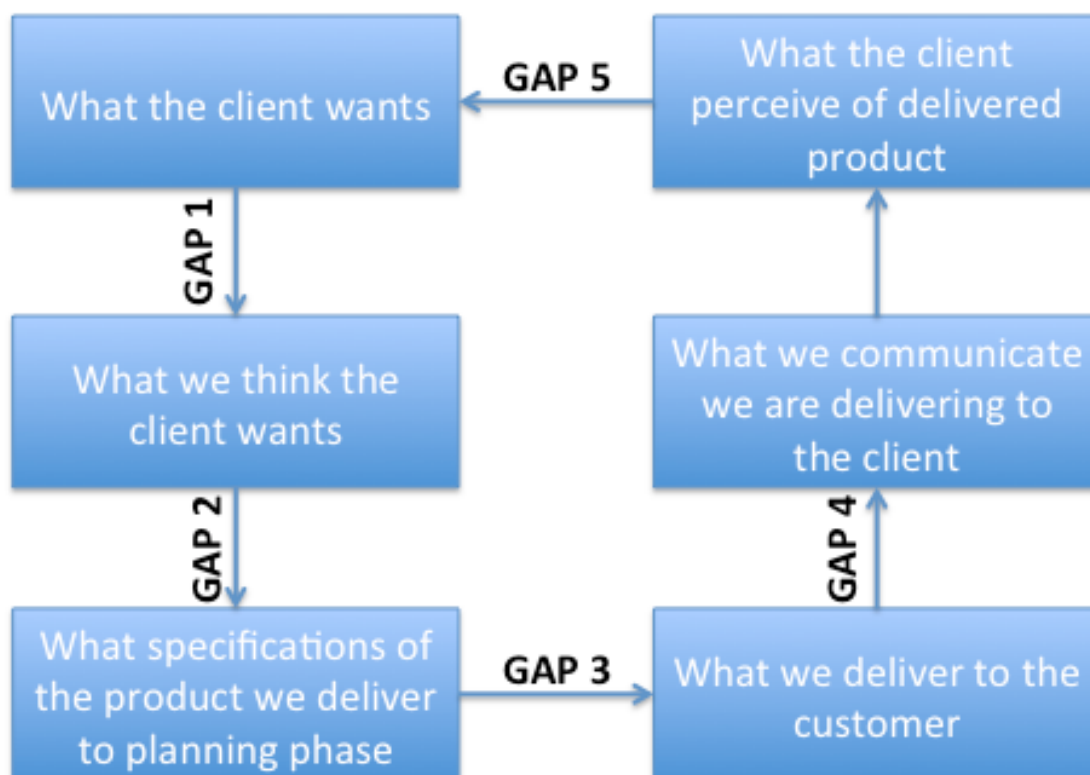


Figura 2.9 – Modelo SERVQUAL para Definição de Projeto.
(Desenvolvido pelo autor)

É possível notar que o primeiro gap é o ponto chave de sucesso para qualquer projeto. Se não houver a percepção correta entre o que o cliente procura, e o que o gerente de projeto executa, o fracasso se torna eminente. Isso reforça mais uma vez a necessidade de foco no princípio imutável número um. Os outros gaps são auxiliares e delineiam como deve ser o andamento do processo de *Definição de Projeto*.

Falando agora do Princípio Imutável 2, pode-se imediatamente notar que este é relacionado diretamente com o paradigma utilizado. O princípio é bastante assertivo quanto ao modo que as variáveis devem ser planejadas *ex-ante*, e como estas devem se comportar ao longo do ciclo de vida do projeto.

Este ponto é também bastante crucial no desenrolar desse estudo como um todo. O modelo a ser proposto é totalmente baseado nas variáveis de custo, tempo e performance. O modelo, assim como o projeto, terá parte do seu sucesso definido pelo correto andamento dessa tríade.

Finalmente, o Princípio Imutável 3 é relacionado com os recursos utilizados durante o projeto. Haja visto que o objetivo do trabalho não afere importância a acessibilidade de recursos, nem a dinâmica de sua utilização, esse princípio acaba perdendo importância. O que vale dizer, entretanto é que os recursos planejados devem ser os mesmos utilizados. Essa comparação *ex-ante/ex-post* será bastante tratada no modelo a ser desenvolvido, pois validará

os resultados obtidos. Assim sendo, deve-se haver um espaço para apontamento desse princípio como sendo um dos principais no sucesso de projetos.

Agora, uma vez enunciados os Dez Drivers, relacionados diretamente com os Cinco Princípios Imutáveis, é possível identificar as práticas que levarão ao sucesso de um projeto.

2.1.7 As Cinco Práticas Imutáveis

Os Dez Drivers e as Cinco Práticas Imutáveis forneceram informações bastante consistentes sobre quais variáveis e quais pontos serão focalizados no modelo a ser desenvolvido. Entretanto, é necessário haver também alguns comentários que tratem dos aspectos qualitativos do modelo. Para haver robustez nos resultados, será necessário entender alguns aspectos de realização dos projetos, e verificar se estes foram feitos de maneira padronizada em relação as melhores práticas para o sucesso. Quando a discussão sobre similaridades de projetos começar a ganhar corpo, será necessário levar como hipótese que todos os projetos foram desenvolvidos segundo as Cinco Práticas Imutáveis.

Caso essa hipótese não seja realística, todo o modelo irá colapsar, afinal como se pode medir similaridades entre variáveis se estas foram modificadas segundo diferentes práticas? Felizmente, haja visto que trata-se de um conjunto de dados de projetos do setor de Petróleo & Gás, as práticas são bastante padronizadas por órgãos externos, logo, essa hipótese é bastante consistente.

Quanto as Cinco Práticas Imutáveis, pode-se dizer que estas são também bastantes correlatas com o ciclo de vida do projeto. Na verdade, elas caminham, segundo o projeto vai se desenvolvendo, desde a Iniciação, até o Fechamento.

Interessante, antes de entrar em detalhes sobre cada prática, entender a palavra “imutável”. Tanto as práticas quanto os princípios mencionados anteriormente ganham esse rótulo. O autor dessa metodologia categoriza dessa forma devido a característica de amplitude de sua metodologia. É considerado aqui que qualquer que seja o tipo de indústria que o projeto se insira, a adoção desse método guiará tal projeto ao sucesso (Alleman, 2014). Entretanto, o autor não explicita isso, mas pode-se entender que isso somente ocorre em casos de projetos governados pelo paradigma custo, tempo e performance (o que é o caso do estudo a ser desenvolvido).

Assim sendo, para sumarizar todas as Cinco Práticas Imutáveis, expõe-se a Figura 2.10 a seguir. Mais do que isso, são demonstradas também as boas práticas para atingir essa metodologia, assim como os benefícios do seu uso.

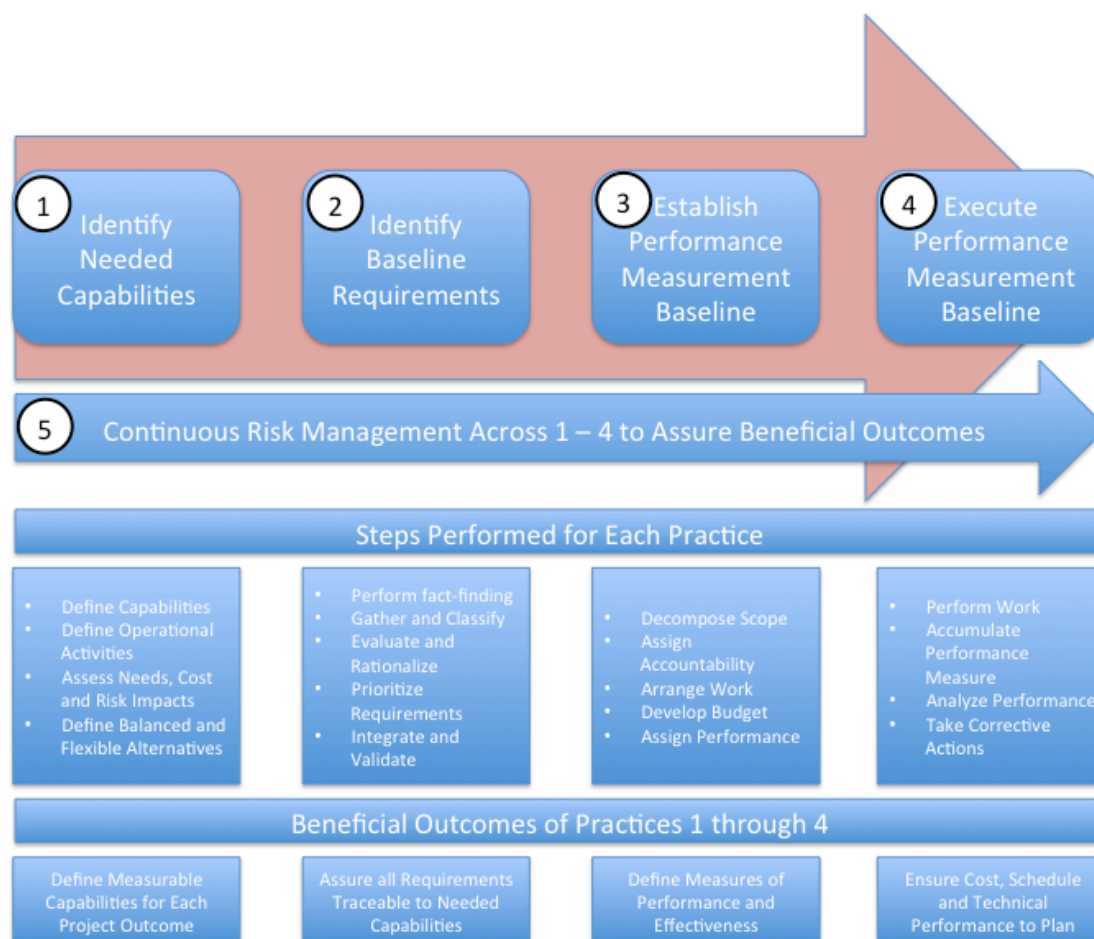


Figura 2.10 – As Cinco Práticas Imutáveis para sucesso de um projeto. (Alleman, 2014)

Numa breve explicação sobre cada prática, pode-se notar que a primeira delas caminha junto com a etapa de Iniciação do projeto. As práticas (2) e (3) já se referem mais a etapa de Planejamento, e a (4) e (5) são puramente relacionadas com a etapa de execução. Haja visto que o escopo desse trabalho, já mencionado anteriormente, é bastante focado nas etapas de Iniciação, Planejamento e Fechamento, deixando um pouco de lado a Execução propriamente dita, pode-se entender que as práticas (4) e (5) não serão de grande importância.

Mais do que isso, é interessante mencionar uma discordância encontrada nessa metodologia, com relação a algo que é considerado vital para sucesso de projetos, mas numa visão de melhoramento contínuo. Se trata da prática de Documentação e Aprendizado, algo bastante forte durante a etapa de Fechamento de um projeto. O autor sequer menciona essa prática, pois não é vista no alcance do sucesso do projeto em si, mas alavanca bastante o sucesso de futuros projetos. Fica então essa crítica a metodologia utilizada, ficando claro que no trabalho proposto o foco também será na prática de Documentação e Aprendizado.

Por fim, vale mencionar que toda a metodologia estudada é guia para o desenvolvimento do modelo, tanto na suas questões quantitativas quanto qualitativas. Mais do que isso, o aprofundamento nos tópicos abordados foi de acordo com a necessidade futura para desenvolvimento do método e robustez das conclusões e resultados a serem obtidos.

2.1.8 Um Resumo do Objeto de Estudo

Nesse parágrafo foi colocado o objetivo do estudo, justamente o ponto central que guia e delimita todo trabalho. Em seguida, toda a base teórica e literária para suportar o objetivo e fornecer *insights* qualitativos foi exposta de maneira linear, sempre tendo em vista as fronteiras do estudo.

O parágrafo foi dividido em blocos para melhor visualização da revisão literária. O primeiro bloco serviu para conceituar o termo projeto, expondo além de sua definição proposta pelo PMI, também as fases do seu ciclo de vida de maneira estandardizada. Além disso, foram colocados os três paradigmas de variáveis que governam o projeto. Foco maior no primeiro e terceiro paradigma. O primeiro porque será aquele adotado no desenvolvimento do modelo, dado suas características. O terceiro por ser aquele vigente nas boas práticas de gestão de projetos, ainda que seja muito complexo para a adoção no estudo proposto.

O próximo bloco definiu o conceito de sucesso de um projeto. O grande uso desse bloco é no fato de mostrar que o sucesso de um projeto é algo complexo e subjetivo. Envolve todos os *stakeholders*, e pode variar conforme a tipologia de projeto a ser desenvolvido.

Finalmente, o terceiro bloco foi voltado para revisão da literatura acerca de uma metodologia estruturada para se alcançar o sucesso de um projeto. O método apresentado tem característica bastante linear, e é de fácil entendimento. Começa mostrando quais são os principais drivers que governam um projeto. Em seguida quais são os princípios imutáveis nos quais esses drivers estão inseridos. E por fim, quais as práticas imutáveis a serem seguidas, tendo em vista os princípios, para que o projeto seja bem-sucedido.

Vale lembrar também que, embora esse método seja bastante robusto, será necessário adaptá-lo ao modelo a ser desenvolvido. Isso porque Alleman (2014) não sugere a importância do processo de Documentação e Aprendizado no Fechamento do projeto, nem direciona esforços a utilizar velhas experiências em novos projetos. Esse é justamente um dos pontos a serem levantados nesse estudo, e portanto deve se ter bastante atenção nesse aspecto. Assim sendo, a metodologia de Alleman (2014) se mostra bastante útil se e somente se for adaptada e reforçada pela visão da Documentação e Aprendizagem.

Para o próximo parágrafo será abordado o assunto motivacional para o estudo em questão. Tanto as ferramentas de gestão de projetos, como de gestão de portfólio podem ser bastante auxiliadas por uma melhor Iniciação e Planejamento do projeto. Por isso, a abordagem agora terá justamente o nome de motivação, pois o modelo a ser desenvolvido é *input* necessário para as boas práticas de gestão de projetos durante a sua execução.

2.2 Captando a Motivação do Estudo

No parágrafo anterior foram discutidos os conceitos de projeto e sucesso de projeto. Pode-se dizer, portanto, que o objeto de estudo é o projeto em si, e todas as variáveis que o permeiam.

É necessário entender bem o que está sendo manuseado antes de propor modelos e soluções de melhoramento. Mais do que isso, o entendimento correto do objeto de estudo já fornece os *gaps* teóricos onde o estudioso pode se aplicar para desenvolver melhorias.

Assim sendo, é hora de introduzir a análise da motivação do estudo a ser desenvolvido. Em outras palavras, uma vez que o projeto para obter sucesso necessita de uma correta Iniciação, um correto Planejamento, uma correta execução e um correto Fechamento, qual destes aspectos o modelo a ser desenvolvido pretende atuar e melhorar? É justamente esse melhoramento proposto que pode ser chamado de motivação do trabalho desenvolvido.

A discussão aqui se inicia com o claro direcionamento: uma vez que foi tratado que o modelo será baseado no velho paradigma – custo, tempo e performance – pode-se induzir que a etapa de Execução é onde estes aspectos são mais latentes. A essência aqui é, portanto, atuar junto dos gerentes de projeto e dos Escritórios de Gerência de Projeto (Project Management Office – PMO⁵). Nesse trabalho conjunto, observar como estes utilizam suas ferramentas de gestão de projetos, e como utilizam o aprendizado adquirido em ocorrências similares as ocorrências de agora para melhorar continuamente a execução dos seus projetos.

Esse parágrafo será dividido em dois momentos diferentes: primeiro a análise de ferramentas de execução e aprendizado em um único projeto, e depois ferramentas de gestão de uma gama de projetos. Esses dois momentos são tratados pois, em sinergia, atuam para o melhoramento contínuo de projetos, ainda que estes sejam uma ocorrência única.

⁵ Escritório de Gerência de Projetos: departamento de uma empresa responsável pelos serviços internos e externos quanto ao desenvolvimento de projetos. São usualmente um departamento com vida limitada a um certo período de tempo. Suas funções são de gestão de dados, coordenação, *governance* e relatório, e atividades administrativas para auxiliar a correta execução de um projeto (Project Management Institute, 2014).

Assim sendo, na primeira etapa desse parágrafo serão vistas as principais ferramentas de gestão de projetos – como as linhas base e EVMS – e em sequência será visto o conceito de Knowledge Management – que consiste na ferramenta de aprendizado de projetos passados para serem utilizados em projetos futuros semelhantes.

A segunda parte desse parágrafo consiste, portanto na visão de múltiplos projetos. O grande foco aqui será no conceito de Gestão de Portfólio de Projetos, que consiste na gestão de múltiplos projetos em prol de um retorno esperado pela empresa. O sucesso de um projeto aqui já não garante o sucesso do portfólio inteiro, ou seja, agora a visão sistêmica ganha força, e os conceitos abordados anteriormente já terão de ser ampliados.

É importante agora analisar sob uma perspectiva macro do parágrafo. A primeira etapa mostra dentro do ambiente de motivação do estudo, aquele momento “onde esse estudo se posicionará” e “como ele será alavancado”, a partir da definição das boas práticas de gestão de projetos. Já a segunda etapa consiste mais numa visão de “por que deve-se ter em mente as boas práticas e por que esse estudo deve ser realizado”. Ainda, para se ter maior clareza da motivação do estudo, é sempre bom ter em mente o objetivo desse estudo, que de maneira resumida irá analisar a semelhança entre diferentes projetos e agrupá-los em *clusters*. Parece finalmente que a coisa faz sentido. Tanto em termos de aprendizado, como gestão de portfólios a semelhança de projetos parece ter um papel de destaque. Em todo o caso, a Figura 2.11 a seguir mostra como será o andamento desse parágrafo.

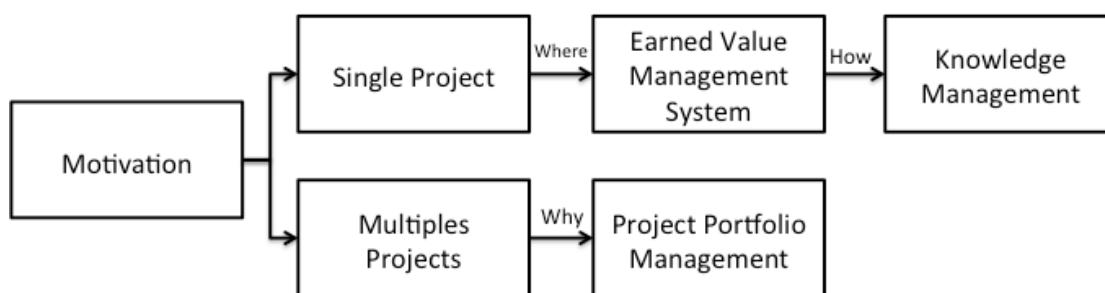


Figura 2.11 – Etapas da definição da Motivação do estudo conduzido.

Como pode ser visto, serão abordados dois caminhos distintos, já explicitados anteriormente. Mas vale a menção aqui de que sempre deve-se ter em mente três coisas nesse parágrafo: o objetivo do estudo, o objeto central do estudo e a motivação para o estudo. Além disso, para facilitar também a compreensão, é sempre bom ter em vista os momentos a serem tratados: onde aplicar, como se posicionar e porque fazer.

2.2.1 A visão de um único projeto

No parágrafo anterior foram considerados os três paradigmas de variáveis que explicam um projeto. Além disso, foi dito que se utilizaria nesse trabalho o velho paradigma – custo, tempo e performance – assim como ilustrado na Figura 2.2.

Assim sendo, o sucesso de um projeto, após percorrer toda literatura sobre o assunto, acabou tratando – sob a ótica do velho paradigma – que o sucesso de um projeto está intimamente ligado com o planejamento e estimativa dos custos, tempos e performance de execução. Mais que isso, para se obter sucesso, aquilo que foi planejado e estimado deve coincidir com aquilo que está sendo executado.

Assim sendo, é boa hora de verificar como fazer a execução de um projeto de maneira que as variáveis não fujam de controle, e não passem dos limites previamente planejados. Para isso, as boas práticas de gestão de projetos apontam duas ferramentas de execução de projetos: o Método do Caminho Crítico (Critical Path Method – CPM) e o Sistema de Gestão do Valor Agregado (Earned Value Management System – EVMS). A primeira ferramenta é relacionada com a variável de calendário, segundo execução temporal dos pacotes de serviços – *work packages* – mas não leva em consideração os custos e performance do projeto. Para esse sentido, o EVMS se mostra mais adequado, pois este integra o escopo (ou performance requerida), custo e tempo dentro do mesmo *framework*. Pode-se dizer que no momento atual da literatura de Gestão de Projetos, o EVMS é o estado de arte. A grande questão que o EVMS propõe é o confronto das variáveis planejadas e estimadas com aquilo que está sendo executado (Acebes, Pajares, Galán, & López-Paredes, 2013). Para melhor entender essa metodologia, é necessário entender seu histórico de desenvolvimento e os critérios adotados para sua execução. Isso será tratado no item a seguir.

2.2.1.1 *Earned Value Management System*

As primeiras ocorrências do EVM podem ser datadas de muito tempo atrás, algo em torno do final dos anos 1800's, aparecendo como uma metodologia básica para controle dos aspectos financeiros de projeto – como por custos e custo no tempo. Em 1967, o Governo dos Estados Unidos introduziu o EVM como forma de entender os aspectos financeiros dos programas militares, e também como forma de gestão de tais programas (Kwak & Anbari, 2011).

Desde então o EVM foi continuamente melhorado até que foram criadas as práticas standardizadas – consequentemente o uso da ferramenta ganhou corpo no mundo da Gestão

de Projetos. Finalmente, em 2005 o PMI lançou o EVM como boas práticas de Gestão de Projetos. A Figura 2.12 a seguir mostra a linha do tempo para a evolução do EVM.

Year	Event
1967	Cost/Schedule Control System Criteria [C/SCSC] introduced by U.S. Department of Defense [DOD].
1972	First C/SCSC Joint Implementation Guide issued to ensure consistency among military departments.
1991	DOD Instruction 5000.2—Defense Acquisition Management Policies and Procedures issued reaffirming use of EVM.
1996	DDDR 5000.2-R—Mandatory Procedures for Major Defense Acquisition Programs and Major Automated Information System Acquisition Programs issued. Draft industry guidelines accepted by Under Secretary of Defense and C/SCSC revised from 35 to 32 criteria.
1998	American National Standards Institute/Electronic Industries Alliance published industry guidelines for EVM Systems [EVMS; ANSI/EIA-748-98].
1999	Under Secretary of Defense adopts ANSI/EIA-748-98 for DOD acquisition.
2000	Simplified EVM Terminology published by Project Management Institute.
2005	<i>Practice Standard for Earned Value Management</i> published by the Project Management Institute [revised; second edition published in 2011].
<small>Note. Details of important milestones in the progress of EVM implementation can be found in several sources, such as http://www.acq.osd.mil/pm/historical/Timeline/EV%20Timeline.htm.</small>	

Figura 2.12 – Linha do tempo da evolução do EVM.
(Kwak & Anbari, 2011)

Dentro desse período que vai do final do século XIX até 2005, o EVM sofreu uma série de mudanças e padronizações. Mas a maior delas pode ser datada dos anos de 1998. Ocorreu quando a American National Standards Institute (ANSI) juntamente com a Electronic Industries Alliance (EIA) publicaram as linhas mestre para o EVM, que se tornava naquela ocasião o EVM System. Foram desenvolvidos 32 critérios (veja o Apêndice A) como forma de padronizar e alargar a utilização do EVMS. Haja visto que são muitos critérios, para melhor compreensão seria bom dividi-los em categorias (Kwak & Anbari, 2011):

- *Organização*: Atividades que definem o escopo do projeto e direcionam responsabilidades para o trabalho a ser realizado;
- *Planejamento e Budget*: Atividades de planejamento, calendarização, precificação e autorização do trabalho a ser realizado;
- *Contabilidade*:: Atividades de acumulação de custos de trabalho e material necessários para completar o trabalho;
- *Análise*: Atividades de comparação entre o orçamentado, o performado, e os custos reais. Além disso, análise da variância desses parâmetros e estimativas do custo final do projeto;
- *Revisão e Manutenção dos Dados*: Atividades para incorporação de mudanças internas e externas no calendário, orçamento e trabalho autorizado.

Note que essas categorias, que determinam o funcionamento do EVMS tem forte correlação com o ciclo de vida de um projeto. As fases de Planejamento e execução são

bastante presentes nessa metodologia. Por tal motivo, o EVMS ganhou forte uso no meio da Gestão de Projetos, uma vez que propõe uma revisão contínua dos dados planejados, sem que simplesmente estes permaneçam imutáveis no início da execução do projeto.

Para os propósitos desse estudo, o foco principal é nas categorias de *Organização e Planejamento e Budget*, as quais implicam a correlação entre o velho paradigma de custo, tempo e performance com a robusta metodologia do EVMS. Mais do que isso, é necessário tomar nota do escopo do projeto, a fim de que as variáveis adotadas sejam corretamente estimadas para o bom funcionamento do sistema proposto.

Para a correta implementação do EVMS é necessário o conhecimento de alguns fatores chave, os quais podem ser agrupados em três categorias: fatores de aceitação do EVMS, fatores de uso do EVMS e fatores de performance do EVMS (Kim, Wells Jr., & Duffey, 2003).

Em termos de aceitação do EVMS, é algo bastante subjetivo, mas também de vital importância para as boas práticas do modelo. Quanto ao termo aceitação, o significado aqui é justamente o fato de que o top management deve ter conhecimento pleno das capacidades de uso do EVMS e literalmente aceitar sua utilização dentro do projeto. A aceitação por parte dos postos mais altos de gestão traz grande força para a metodologia, pois esta será capaz de se difundir por todo o time de projeto. Mas como fazer para que todos os envolvidos tomem consciência da força do EVMS? Muito disso é baseado na documentação de performance passada. O correto processo de documentação se torna evidente aqui, pois só assim que os futuros projetos podem ver quais os benefícios do uso do EVMS e fazer disso uma prática indispensável. Mas é necessário também salientar que o gerente do projeto deve ter maturidade o suficiente para confrontar aquilo que foi documentado com suas próprias expectativas sobre seu projeto, e com isso, personalizar seu EVMS para atender suas necessidades (Kim, Wells Jr., & Duffey, 2003).

Em termos dos fatores de uso, é possível citar alguns detalhes que o gerente de projeto deve se ater (Kim, Wells Jr., & Duffey, 2003):

- Longa experiência no uso do EVMS contribui para todas as etapas do ciclo de vida do projeto;
- O correto uso parte de todos os membros do time de projeto;
- O uso do CPM é complementar e auxilia bastante o EVMS;
- O bom funcionamento depende largamente das relações e comunicação interna do time de projeto.

Finalmente, os fatores de performance são aqueles que determinam os *outputs* da fase de execução do projeto. Uma vez finalizada essa etapa, utilizando-se do EVMS, a documentação será analisada e a performance do modelo será discutida. Esse fator retroalimenta todos os outros fatores do EVMS.

Num paralelo com as etapas do ciclo de vida do projeto, é possível expandir um pouco o que foi falado anteriormente. O EVMS não trata somente das etapas de planejamento e execução, mas é muito ligada também ao fechamento do projeto. Note que muito foi falado durante os fatores que a correta documentação alimenta todo funcionamento do EVMS. E justamente a documentação é correlata com a fase de Fechamento de um projeto.

Seria interessante agora ressaltar portanto o fator motivacional do EVMS para o estudo realizado. Numa visão de um único projeto, pode-se notar que o EVMS é uma forte ferramenta para o sucesso do projeto. Além disso, engloba muitas fases do ciclo de vida de um projeto. Fica bem claro o porquê do EVMS ser o “onde” para motivação do estudo desenvolvido.

Uma vez percebido isso, seria interessante ter uma visão mais completa do seu funcionamento, a fim de estabelecer aqui alguns insights positivos para a elaboração do modelo de clusters a ser realizado. Em outras palavras, haja visto que é um fator motivacional para o estudo, ter em vista sua operação, ainda que de forma resumida, auxilia na elucubração de algumas hipóteses e fronteiras do estudo.

Basicamente, o EVMS controla e monitora os custos e o tempo durante a execução do projeto. Mais do que isso, ele se retroalimenta com os dados que vão surgindo durante o avançamento físico do processo. Tudo isso já foi mencionado anteriormente. A forma com a qual o EVMS opera, entretanto, é o grande diferencial. Existem dentro do modelo basicamente três curvas: curva do que foi planejado, curva do que está sendo executado e curva do valor agregado. O grande diferencial aqui é justamente na curva de valor agregado, que consiste no cruzamento do que foi planejado com o que realmente foi feito. Em outras palavras, a qualquer instante de tempo, tem-se uma medida do que foi feito, em termos de tempo. Naquele tempo, quanto estava planejado para ter sido realizado? É justamente essa a curva do valor agregado. Gráficamente, a Figura 2.13 a seguir mostra as relações propostas entre as curvas de valor agregado, planejado e executado.

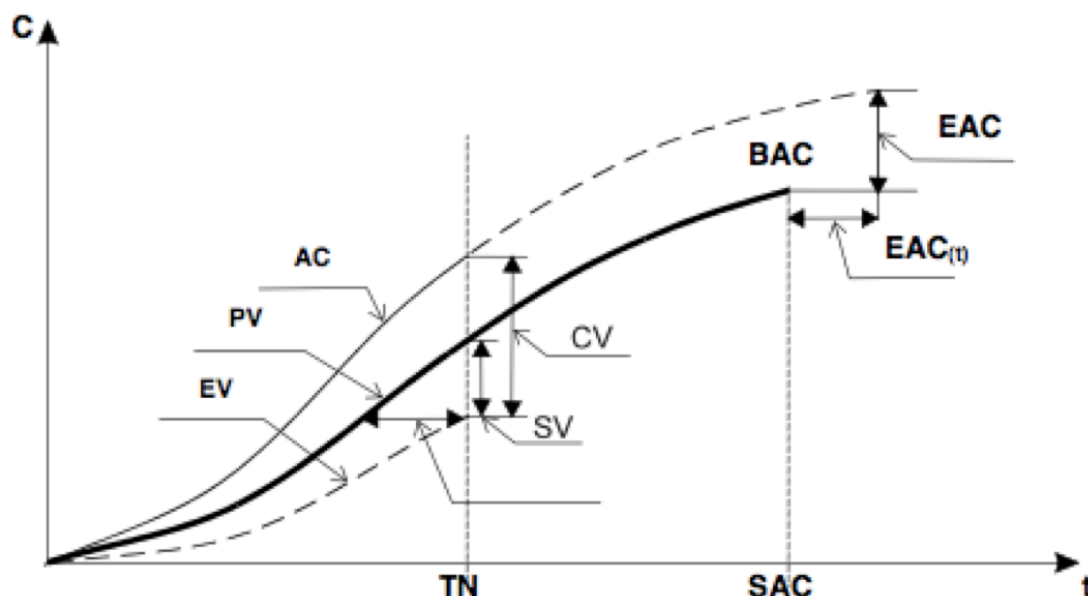


Figura 2.13 – Ilustração gráfica do EVMS, com as curvas que o compõe.
(Merli, 2010)

Para melhor entender a Figura 2.13 seria interessante expor toda a terminologia envolvida no processo do EMVS. A Tabela 2.1 a seguir expõe todas as nomenclaturas utilizadas. Note entretanto que as siglas estão em inglês, pois a standardização do método propõe que assim seja feito.

Tabela 2.1 – Terminologia proposta pelo padrão do EVMS.

PV	Valor Planejado
EV	Valor Agregado
AC	Custo Efetivo
CV	Variação do Custo ($CV = EV - AC$)
SV	Variação do Tempo ($SV = EV - PV$)
CPI	Índice de Performance de Custos ($CPI = EV/AC$)
SPI	Índice de Performance de Tempo ($SPI = EV/PV$)
BAC	Budget Final
SAC	Tempo Total
EAC	Estimativa Total

O Valor Planejado (PV) também é conhecido como *Budget Cost Work Schedule* (BCWS). Uma interessante característica do PV é que ele coincide com o BAC no final do projeto. Além disso, é uma das curvas nas quais o foco é maior pois corresponde a tudo o que foi planejado para o projeto, em termos de custos e tempo, segundo os requerimentos de performance. E a curva do PV em teoria deveria ser exatamente igual a curva de Custo

Efetivo (AC). Mas obviamente isso é bastante improvável de acontecer, dados as incertezas que circundam todo projeto. Lembrando a discussão prévia sobre sucesso de projeto, ao considerar-se apenas o velho paradigma, quanto mais próximo está a curva de PV com a curva de AC, maior a probabilidade de sucesso do projeto.

Para entender melhor as discrepâncias entre as curvas de PV e AC que surge, portanto a curva de Valor Agregado (EV). Qualquer deslize na curva AC determina uma mudança na EV, estas duas são bastante correlatas, segundo o modelo. Quanto surge a palavra desliza, isso significa qualquer movimento da variância de custos (CV) e variância de tempo (SV). Consequentemente, os índices CPI e SPI também são modificados. Estes dois índices que são o *core* do EVMS. Juntos, correspondem a fonte de dados de controle e monitoramento do progresso do projeto. Traduzem as variações de custo e tempo, e podem ser facilmente comparados com outros projetos. Qualquer mudança nesses índices indica também que há uma mudança no BAC em relação ao EAC – o ideal é que eles fossem iguais.

Muitos estudos são realizados para cálculo correto do EAC baseado nos índices de CPI e SPI. Existem em grande maioria abordagens estatísticas que tratam os eventos probabilisticamente [(Merli, 2010) e (Caron, Ruggeri, & Merli, 2012)]. Os estudos feitos nesse campo tratam o EVMS de maneira mais dinâmica, havendo assim a possibilidade de calcular o EAC com um intervalo de valores que é mutável, e portanto fornece melhor visão do progresso no exato instante.

Para estimas a duração de um projeto não é tão simples quanto determinar o EAC. Para isso, foi desenvolvida a evolução do EVMS, que além de tratar do Valor Agregado, trata agora também do Tempo Agregado (Lipke, 2003). Nessa abordagem, o autor não utiliza mais o SPI para calcular a estimativa de tempo. O que o autor descobriu foi que o SPI traz uma relação entre EV e PV, então existe a convergência para o valor 1, mesmo que o projeto tenha sido feito fora do prazo planejado. Para sanar esse problema, o autor propôs um novo meio de calcular o SPI, utilizando agora a projeção da curva PV sobre o tempo. Assim sendo, o calculo do SPI se dá como:

$$SPI = \frac{Te}{T}$$

Onde, *Te* é o Tempo Agregado, obtido a partir da projeção mencionada dada pela curva EV no tempo dado.

Aprofundando um pouco mais o estudo, basta agora fazer a ponte entre a velha tratativa do EVMS com os dados de Tempo Agregado. Juntos, os dois fornecem uma robusta ferramenta de controle e monitoramento de projetos durante a sua execução (Lipke, 2006).

Agora, uma vez entendido as ferramentas, é interessante fazer uma chamada da motivação do estudo novamente. Uma vez que o EVMS se mostra uma forte ferramenta de controle e monitoramento durante a execução do projeto, seria interessante auxiliar o planejamento e estimativas das linhas bases de PV. É notório que não basta o projeto ser bem executado, ele deve haver *inputs* de planejamento que sejam capazes de serem seguidos. A seguir encontram-se dois interessantes exemplos de projetos. O primeiro (Figura 2.14) mal planejado e mal executado, com os valores de SPI e CPI tendendo a valores bastante baixos, a medida que o projeto está sendo executado. O segundo exemplo (Figura 2.15) já mostra um projeto sob controle, onde os valores de SPI e CPI giram em torno de um.

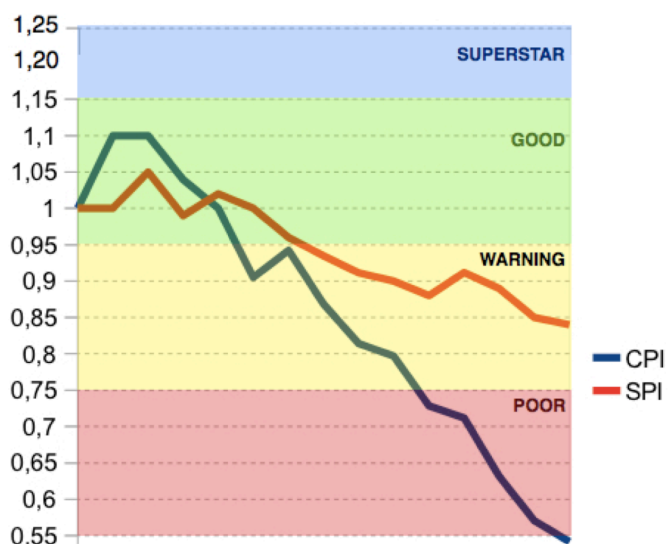


Figura 2.14 – SPI e CPI de um projeto mal executado (Merli, 2010)

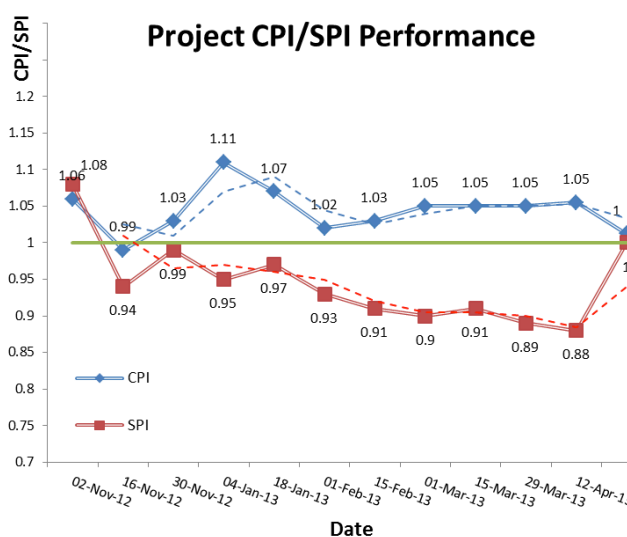


Figura 2.15 – SPI e CPI de um projeto bem executado. Source: www.projectcontrolacademy.com

O modelo a ser desenvolvido tem como motivação intrínseca auxiliar na estabilização dos índices CPI e SPI, uma vez que estes já traduzem em consistência como o projeto foi planejado e como está sendo executado. Lembrando que o PV pode e deve conter dados de projetos já realizados, projetos semelhantes. Assim sendo, tratar de similaridades de projetos nada mais é como uma ferramenta de auxílio na hora de estimar o andamento de futuros projetos, e consequentemente, convergir os valores de CPI e SPI para um intervalo mais confortável.

2.2.1.2 Knowledge Management

Haja visto todas as nuances do EVMS, algo surge com bastante clareza. Para que este funcione com maior significância, é necessário que os dados imputados no começo do processo sejam os mais corretos possíveis. Mas para que isso ocorra, é necessário acima de

tudo, que se tenha um bom planejamento. Mas, então, o que determina um bom planejamento?

Essa pergunta é crucial para sucesso de todo o projeto. Saber como planejar é ponto-chave dentro de um projeto bem-sucedido. E o planejamento é muito baseado em duas variáveis qualitativas do gerente de projeto. A primeira é seu conhecimento acerca do processo de planejar. E a segunda é seu conhecimento sobre o assunto tratado em questão.

Pode-se afirmar, portanto, que o conhecimento de processo é algo inato ao gerente de projeto. Mas e quanto o conhecimento mais “mercadológico” acerca de um determinado tipo de projeto? Essa segunda variável é muito determinada em função do aprendizado que o gerente tem com projetos passados que sejam semelhantes. Surge aqui um conceito bastante robusto que é o de Knowledge Management.

O conceito mais geral de Knowledge Management se refere a identificação de conhecimentos coletivos dentro de uma organização com o propósito de alavancar a competitividade. Mais do que isso, a correta gestão do Knowledge Management System (KMS) é de suma importância, pois é no KMS onde todas as informações e conhecimento estão estruturados e disponíveis para a organização (Alavi & Leidner, 2001).

A massiva quantidade de dados dentro do KMS vem de um processo chamado Criação de Conhecimento, definido por:

“Organizational knowledge creation involves developing new content or replacing existing content within the organization’s tacit and explicit knowledge. Through social and collaborative processes as well as an individual’s cognitive process (e.g. reflection), knowledge is created, share, amplified, enlarged and justified in organizational settings”

(Alavi & Leidner, 2001).

Aplicado ao estudo proposto, o processo de Criação de Conhecimento se trata do “aprender com velhos projetos”. A documentação gerada no Fechamento do projeto deve ser introduzida no KMS, onde servirá de auxílio para o gerente de projeto e toda sua equipe quando estes se depararem com algum projeto futuro semelhante.

A importância do KMS não é somente em dados de projetos semelhantes, mas também em todo um conhecimento de como utilizar as boas práticas para execução de um projeto. Um bom uso é demonstrado pelo *PMBOK Guide*[®], e a constante renovação dos conhecimentos ali descritos que são utilizados por muitos estudos acerca do tema de gestão de

projetos (Chou & Yang, 2012). Assim sendo, de maneira geral, pode-se inferir ao KMS as seguintes atribuições:

- I. Dados para Inicialização e Planejamento do projeto;
- II. Acesso fácil às técnicas de gestão de projetos com exemplos de aplicação;
- III. Desenvolvimento de *skills* para o correto uso de modelos, técnicas e ferramentas de gestão;
- IV. Melhoramento contínuo dos dados que serão guardados e que estão guardados no KMS.

Como conclusão, pode-se dizer que todas as fases do ciclo de vida do projeto alimentam e são alimentados pelo KMS. Mas como o foco desse estudo é algo em torno do planejamento, vale dizer mais algumas palavras sobre a relação do KMS e esta fase.

A fase de planejamento terá como poder extrair de projetos passados possíveis comportamentos que se traduzirão na formulação das linhas base do projeto. Essas linhas base serão utilizadas na execução, e alimentará o KMS com os dados reais do que está acontecendo. Dessa forma, o sistema se alimenta com dados mais verídicos e se adequa para os futuros projetos.

No Fechamento do projeto, novamente o KMS será alimentado, mas dessa vez com os dados do projetos finalizado, e conseqüentemente pronto para ser utilizado no próximo projeto a ser desenvolvido. Vale dizer que a gestão do KMS é bastante complexa, uma vez que existem várias tipologias de projetos, e portanto, os dados não podem se cruzar (Ahern, Leavy, & Byrne, 2014).

Haja visto então as relações com o projeto, seria interessante também mostrar como o KMS se relaciona com a empresa em geral. Mais do que a Criação de Conhecimento, surge aqui a necessidade de Alinhamento de Conhecimento. Esse conceito diz que para que o KMS tenha plena difusão dentro da empresa é necessário que o conhecimento contido neste esteja alinhado com o conhecimento intrínseco à organização. O Alinhamento do Conhecimento é um conceito necessário pois é este que determina os limites e fronteiras em termos de tecnologias, skills e recursos da empresa (Reich, Gemino, & Sauer, 2013).

Mais do que isso, quando todos esses conceitos estão bastante difundidos, tanto o EVMS, quanto o KMS junto ao Alinhamento do Conhecimento, a empresa pode sair da zona de conforto de análise de um único projeto e passar a pensar de maneira mais global. Surge aqui o conceito de Gestão de Portfólio de Projetos, algo que analisaremos em seguida e que compõe a etapa do “por que” realizar o estudo proposto.

2.2.2 Gestão de Portfólio de Projetos

Até então a motivação do estudo foi discutida sob a ótica do projeto único. O movimento se iniciou com a apresentação do EVMS para controle e monitoramento do projeto, e em seguida foi mostrada a importância do KMS para a correta formulação do planejamento do projeto para correta execução do mesmo. Em resumo, foi demonstrado onde e como o modelo a ser desenvolvido pode servir como ferramenta de auxílio para o sucesso do projeto.

Finalmente agora pode-se expandir os horizontes e mostrar por quê o modelo deve ser desenvolvido, no sentido de agregar valor não somente ao projeto, mas numa visão sistêmica, como agregar valor para toda empresa.

A primeira pergunta que surge nesse contexto é: a empresa desenvolve projetos em paralelo? E a partir disso, uma série de indagações devem surgir: Como os projetos estão conectados entre si? Existe algum tipo de priorização? Por quê? Para esclarecer essas perguntas, deve ser formalizado agora o conceito de Gestão de Portfólio de Projetos (GPP).

“Project Portfolio Management is a set of business practices that brings the world of projects into tight integration with other business operations. It brings projects into harmony with the strategies, resources, and executive oversight of the enterprise and provides structure and processes for project portfolio governance.”

(Levine, 2005)

A partir da definição dada, é possível dizer que numa abordagem multi-projeto o responsável pela gerência não é mais o gerente de projeto. Existem agora inúmeros players que estão em contato com o projeto e atuam diretamente na sua realização. De maneira bem geral, a Figura 2.16 a seguir mostra essa visão multi-projeto.



Figura 2.16 – Visão multi-projeto versus projeto único.
(Project Management Institute, 2014)

A visão de um único projeto mostra a momento de agora de uma empresa, mas não carrega com si a visão estratégica por traz da realização daquele projeto, muito menos as sinergias entre este projeto e outros que a companhia pode estar rodando em paralelo. A Figura 2.17 a seguir mostra as relações entre projetos, como o GPP atua melhorando a visão de um único projeto.

	Highly effective at portfolio management	Minimally effective at portfolio management	% Increase
AVERAGE PERCENTAGE OF PROJECTS:			
Completed on time	68%	50%	36%
Completed on budget	64%	54%	19%
Met original goals and business intent	77%	65%	18%
Met/Exceeded forecasted ROI	62%	48%	29%

Figura 2.17 – Impactos da implementação GPP em um único projeto.
(Project Management Institute, 2014)

Mas sabe-se que o projeto em si está na alçada do Gerente de Projeto, mas e quanto ao GPP? A visão do GPP é um pouco mais complexa, e acaba envolvendo vários *stakeholders*, como já dito anteriormente. As figuras centrais no caso podem ser o Gerente de Projeto – para cada projeto em paralelo – e o Gerente de Operações. A Tabela 2.2 a seguir mostra o papel de cada um dentro do GPP.

Tabela 2.2 – O papel do Gerente de Operações e do Gerente de Projetos dentro da empresa.

Gerente de Operações	Gerente de Projetos
Strategies	Schedule/time
Objectives, goals	Project Cost
Business Performance	Project Performance

Stockholder Satisfaction	Stakeholder Satisfaction
Project Selection and Mix	Scope/change Control
Resource Availability	Resource Utilization
Cash flow, Income	Cash Usage

Source: (Levine, 2005)

Assim, o GPP surge como figura de união entre as duas partes, cuidando não somente da visão fechada do projeto, mas também a visão integrada da empresa. Mais do que isso, a intensidade de conexão dos dois papéis joga um papel decisivo no sucesso do portfólio. Para que o portfólio tenha um sucesso e gere vantagem competitiva para a empresa é de suma importância que o Gerente de Operações e o Gerente de Projetos estejam em perfeito alinhamento. Obviamente, o sucesso do portfólio traduz-se em sucesso da estratégia da empresa (e maiores bônus para os gerentes) (Beringer, Jonas, & Gemünden, 2012).

O ponto deve ser até estendido. Para que haja sucesso do portfólio, e consequentemente da empresa, todos os stakeholders devem se envolver ao máximo e procurar um alto grau de sinergia entre si. Assim, conclusivamente, pode-se observar alguns motivos pelos quais a empresa deve realizar o GPP, com várias dimensões de sucesso, assim como é mostrado na Figura 2.18.

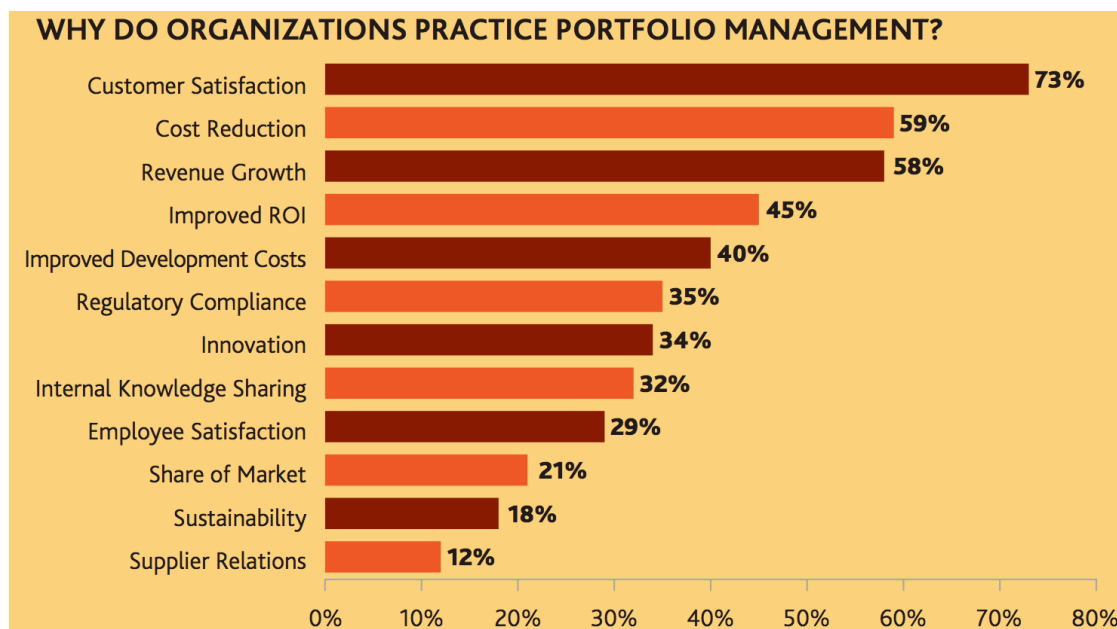


Figura 2.18 – Dimensões positivas para empresa que aplica o GPP.
(Project Management Institute, 2014)

A Figura 2.18 acima mostra os objetivos de qualquer organização e como o GPP pode melhorar seus resultados. Para propósito desse trabalho, obviamente, o foco seria em todas as dimensões que envolvem o velho paradigma – como Redução de Custos e Melhoramentos no Custos de Desenvolvimento .

Vale dizer que os dados da Figura 2.17 e Figura 2.18 foram extraídos do relatório anual do PMI chamado *Pulse of Profession*, e se trata de uma pesquisa realizada com diversas empresas e contribuintes.

Voltando os olhos agora para o modelo a ser desenvolvido, é possível focalizar o GPP com olhares voltados para as etapas de Planejamento e Fechamento do único projeto. Mas quando isso é expandido para uma visão multi-projeto, talvez os ganhos possam ser alavancados. Quando foi falado anteriormente que uma Redução de Custos pode ser um alvo do estudo, a figura em questão muda, pois não é simplesmente a redução de custos do projeto, mas da empresa como um todo.

Assim sendo, quando fala-se do GPP, ou da Redução de Custo da empresa causada pelo GPP é necessário ver como isso é possível, e como a visão do projeto único contribui para essa assertiva. A primeira coisa então é tentar entender como funciona o GPP, ou quais são suas etapas constituintes. Mas vale aqui um ressalva, pois o GPP não pode ser constituído por etapas, uma vez que ele não tem começo, meio e fim. Então chamar-se-á de ciclo essencial.

O ciclo essencial do portfólio é composto por cinco ações:

- I. Identificação das necessidades e oportunidades da empresa;
- II. Seleção da melhor combinação de projetos;
- III. Planejamento e Execução desses projetos;
- IV. Visualização dos *outcomes* dos projetos;
- V. Realização dos lucros e benefícios.

Note que as ações (I) e (II) são essencialmente ações que envolvem o Gerente de Operações como peça-chave. As ações (III) e (IV) já são de escopo maior do Gerente de Projetos. A ação (V) é um resultado esperado por ambos, e muito importante pois mede a saúde do portfólio, e mede a competência da equipe em gerar valor para a empresa (Levine, 2005). A Figura 2.19 a seguir mostra o início do ciclo essencial, lembrando que o foco principal do estudo está na etapa de Gestão de Projetos assinalada na figura.

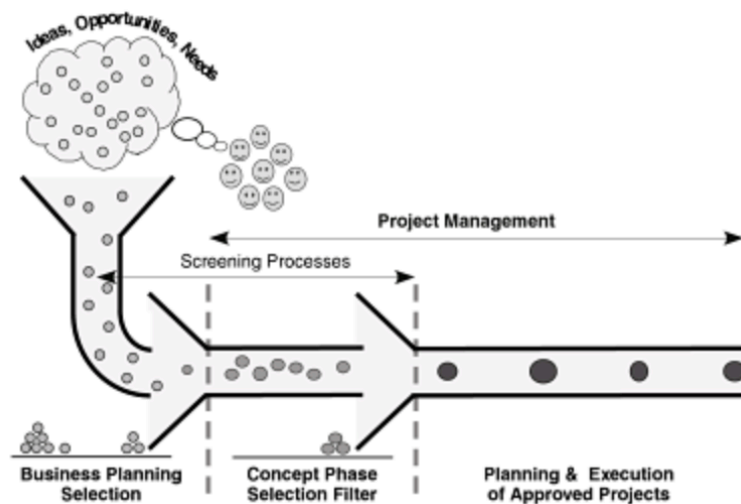


Figura 2.19 – Ciclo essencial do GPP.
(Wideman, 2004)

É possível notar que uma vasta gama de projetos podem ser desenvolvidos, mas somente alguns entram no *pipeline*. Por que existe tal segregação? Nem todos os projetos são potenciais retornos para a empresa, ou mesmo que fossem, não possuem a sinergia necessária com os outros projetos para serem passíveis de entrar no *pipeline*.

O processo de seleção dos projetos que de fato merecem ser tocados a diante toma como *input* também toda base de informações do KMS. Assim sendo, é interessante ter uma visão única das variáveis e dos projetos contidos no KMS, e estabelecer por semelhança aqueles que deram retorno e que portanto merecem ter algo semelhante no *pipeline*. Note que é justamente isso a proposta do modelo a ser desenvolvido.

É possível ainda quebrar a ação de seleção do pipeline em algumas fases:

- Preparação da proposta do projeto;
- Valoração do projeto em termos de retornos e benefícios;
- Valoração dos riscos do projeto;
- Alinhamento do projeto-candidato com a estratégia da empresa;
- Determinação dos recursos a serem alocados;
- Ranqueamento dos projetos de acordo com os seguintes critérios:
 - Execução da estratégia e linha táticas;
 - Manutenção do inventário de recursos;
 - Estabelecimento do budget;
 - Decisão do tamanho ótimo do portfólio;
 - Limites aceitáveis de risco;
- Seleção dos projetos.

Para a correta seleção dos projetos é necessário portanto um robusto método que leve em consideração todas as etapas acima descritas. Mais uma vez, a relação com o estudo desenvolvido se mostra eminente. Pode-se associar um novo projeto com um velho projeto, segundo suas semelhanças, e isso já é meio caminho andado para a correta seleção do portfólio.

2.2.3 Resumo do Parágrafo

Através desse parágrafo foi possível observar a motivação do estudo desenvolvido. Foram discutidas as visões de um único projeto, e a importância de um bom instrumento de previsão e estimativas para auxílio das ferramentas como o EVMS.

Mais do que isso, foi visto também a importância da documentação dos projetos acabados para alimentar o KMS, e assim, fornecer dados fáceis para o desenvolvimento de novos projetos.

Finalmente, a visão de um único projeto foi alargada e surgiu então o conceito de Gestão de Portfólio de Projetos. A importância aqui se deu para a visão de múltiplos projetos e como a sinergia entre estes pode gerar resultados concretos para a empresa. Além disso, a discussão mostrou bastante interessante quando os aspectos do KMS se mostraram novamente presentes na seleção dos projetos para o *pipeline*.

De maneira resumida, a Figura 2.20 a seguir mostra como foi o andamento de todo o parágrafo, e onde o estudo a ser desenvolvido deve focar.

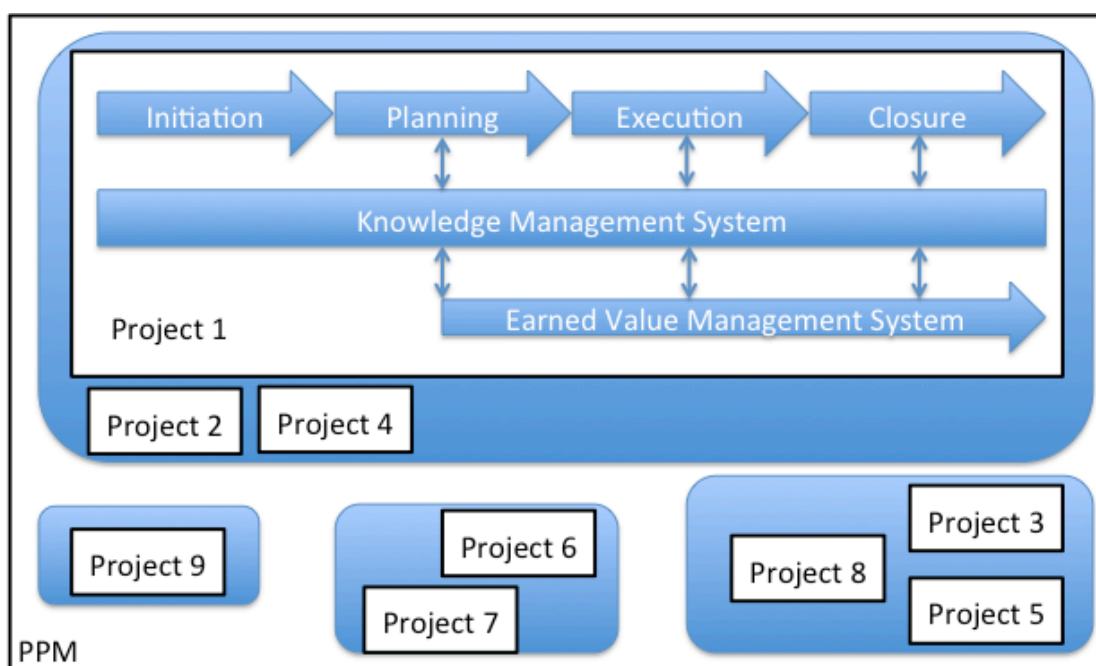


Figura 2.20 – Visão sistêmica de projetos únicos formando clusters de projetos.

Apenas note, antes de passar para o próximo capítulo como os projetos podem se agrupar segundo suas similaridades. E mais do que isso, como o portfólio pode ficar ainda mais robusto, quando analisado em grupos de projetos.

2.3 Resumo do Capítulo

Esse capítulo teve um foco bastante intenso na literatura, com o objetivo de mostrar para o leitor quais as bases teóricas que serão utilizadas no desenvolvimento do modelo de clusters por semelhanças entre projetos.

A abordagem seguiu a lógica de expor primeiro o objeto do estudo, que no caso são os projetos propriamente falando. Assim que nomeou-se o objeto, foi a vez de conceitualizá-lo, expor suas variáveis e fases, definir fatores e critérios para seu sucesso e fornecer uma metodologia para obtenção do sucesso.

O parágrafo subsequente focalizou na motivação do estudo. Uma vez definido o objeto, seria interessante mostrar um pouco como este se encaixa num contexto real, e quais as práticas adotadas para que ele funcione bem. Nesse sentido foi abordado o método de EVMS, juntamente com a importante ferramenta do KMS. Juntos, esses dois termos tendem a trazer grandes benefícios para o projeto, quando analisado de maneira singular.

Entretanto, é necessário também mostrar o lado da sinergia entre projetos, uma vez que toda empresa possui não somente um projeto no pipeline, mas diversos, que devem estar totalmente alinhados com a estratégia da empresa e prontos para, juntos, trazerem valor para a organização.

Note portanto que essa revisão bibliográfica nada mais é que um instrumento de análise qualitativa para o modelo a ser desenvolvido. A noção de tudo que foi abordado será utilizado para que se tenha maior robustez no método apresentado, uma vez que as ferramentas estatísticas que serão apresentadas são extremamente complexas, e portanto necessitam de uma boa base qualitativa. Somado a isso, já será antecipado que devido ao número reduzido da amostra de projetos, mais uma vez uma boa base teórica pode ajudar a suprir essa falta de dados quantitativos.

3 Revisão da Literatura Matemática

Nesse capítulo será tratado todo o ferramental matemático para construção do modelo. O principal propósito aqui é adquirir conhecimento necessário para formulação correta dos aspectos quantitativos do estudo. É necessário também sublinhar o fato que o propósito não é ser uma revisão bibliográfica puramente matemática. As ferramentas apresentadas serão aprofundadas somente com a necessidade apresentada pelo modelo.

3.1 Ferramentas Estatísticas e Representação dos Dados

Nesse parágrafo será exposto todo background estatístico que é base do modelo. Serão apresentados os seguintes pontos, na seguinte ordem: estatística descritiva básica, técnicas gráficas de representação de dados, métodos para cálculo de distâncias entre pontos e comentários finais.

Um breve por que quanto aos assuntos abordados se mostra interessante aqui. No caso da estatística básica seu uso será para manipulação dos dados. No caso das técnicas de representação, é interessante ver uma abordagem mais visual dos dados, e junto com a revisão bibliográfica do capítulo anterior, fornecer uma boa base qualitativa para o modelo.

Já os métodos de distância são essenciais para verificar quão próximos estão os dados, e isso significa quão similar são os projetos entre si. E finalmente, serão feitos alguns comentários finais com o objetivo de deixar tudo bem claro e pronto para ser utilizado pelo modelo.

3.1.1 Estatística Descritiva Básica

Nesse item serão discutidos três pontos: como apresentar os dados em forma de matrizes, como calcular a média da amostra, e como calcular a matriz de covariâncias da amostra.

Antes de tudo é necessário, porém, discutir um pouco sobre os dados. Uma vez que falou-se em média e covariância da amostra, é de se pensar: e quais inferências pode-se fazer sobre a população? Primeiramente, é necessário dizer que como os dados foram obtidos numa pesquisa setorial (setor de Petróleo & Gás), não tem-se como pegar todos os dados de população de projetos dentro desse setor. O que vale dizer é que os cálculos apresentados serão aproximados para uma população, mesmo que não esteja se falando de uma.

A maneira mais simples de apresentar um montante de dados dentro da estatística multivariada é por meio de matrizes. A Tabela 3.1 a seguir mostra como os dados podem ficar dispostos.

x_{jk} = measurement of the k^{th} variable on the j^{th} project

Tabela 3.1 – Dados multivariados dispostos em forma de matriz.

X =

	Variable 1	Variable 2	...	Variable k	...	Variable p
Project 1	x_{11}	x_{12}	...	x_{1k}	...	x_{1p}
Project 2	x_{21}	x_{22}	...	x_{2k}	...	x_{2p}
...
Project j	x_{j1}	x_{j2}	...	x_{jk}	...	x_{jp}
...
Project n	x_{n1}	x_{n2}	...	x_{nk}	...	x_{np}

Considerando essa representação, é possível calcular a média das amostras. Na estatística básica, a média é calculada segundo uma variável apenas, assim será também na estatística multivariada, mas como são muitas variáveis, as médias formarão também uma matriz de dados.

$$\bar{x}_k = \frac{1}{n} \sum_{j=1}^n x_{jk} \quad k = 1, 2, \dots, p$$

Todo valor k-ésimo valor representa um subgrupo de valores para uma dada variável, no caso de haverem p-variáveis, então serão calculados p-médias. Assim sendo, o conjunto de todas as médias para todas as variáveis pode ser representado da seguinte forma:

$$\bar{\mathbf{X}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}$$

Finalmente, em termos de variâncias e covariâncias, segue o mesmo princípio. Mas antes seria bom conceituar esses dois termos. A variância é um valor que mostra, dentro de uma dada amostra ou população, como os dados estão dispersos. Mais do que isso, a variância mostra os desvios de cada observação para a média (Costa Neto, 2002).

A covariância, por sua vez, expressa a associação linear entre duas variáveis. É um conceito muito importante para o propósito desse estudo, pois mostra como as variáveis interagem entre si. Essa interação pode entrar no cálculo das distâncias estatísticas (Johnson & Wichern, 2007). A matematização da variância e covariância é dado por:

$$s_k^2 = s_{kk} = \frac{1}{n} \sum_{j=1}^n (x_{jk} - \bar{x}_k)^2 \quad k = 1, 2, \dots, p$$

$$s_{ik} = \frac{1}{n} \sum_{j=1}^n (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k) \quad i \neq k, \quad i, k = 1, 2, \dots, p$$

Novamente, em termos de representação visual, tanto a variância como a covariância são expressas em formato matricial.

$$\mathbf{S}_n = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1p} \\ s_{21} & s_{22} & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & s_{pp} \end{bmatrix}$$

Para tornar os cálculos mais simples, e computadorizáveis, é possível expressar tanto a média como a variância e covariância como:

$$\bar{\mathbf{x}} = \frac{1}{n} \mathbf{X}' \mathbf{1} \quad (1)$$

$$\mathbf{S}_n = \frac{1}{n} \mathbf{X}' \left(\mathbf{I} - \frac{1}{n} \mathbf{1} \mathbf{1}' \right) \mathbf{X} \quad (2)$$

O cálculo para chegar a esses resultados pode ser visto no Apêndice B.\

3.1.2 Técnicas de Representação Gráfica

Representações gráficas são importantes, mas são geralmente negligenciadas pelos estudiosos (Johnson & Wichern, 2007). Por conta disso, o trabalho em questão não deixará passar a oportunidade de expressar os dados e análises em formas gráficas, pois muitos *insights* preciosos podem surgir antes mesmo de complexos modelos matemáticos. Basicamente, serão tratados dois tipos de representação gráfica aqui nessa revisão: os gráficos de dispersão e a representação de estrela. Ambos serão muito úteis no desenvolver do modelo.

3.1.2.1 Gráficos de Dispersão

A primeira técnica de representação é o chamado gráfico de dispersão GD. Basicamente, o GD expõe uma série de dados multivariados num diagrama de duas ou três dimensões. É chamado de dispersão pois não há a priori qualquer tipo de correlação entre as variáveis, e os dados ficam simplesmente “jogados” no plano cartesiano. A Figura 3.1 a seguir mostra um exemplo de duas e três variáveis.

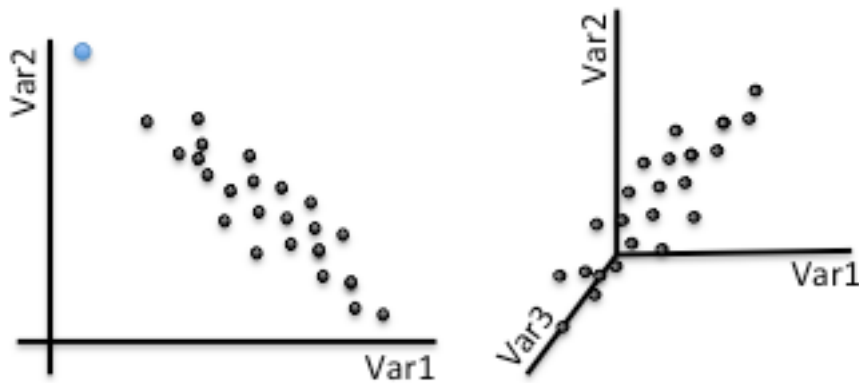


Figura 3.1 – Exemplo de gráfico de dispersão para duas e três variáveis.

Para o propósito desse estudo, a GP auxilia e muito na visualização de relações entre as variáveis. Por exemplo, na Figura 3.1, tanto para o caso bi-dimensional quanto para o caso tri-dimensional, pode-se notar a presença de correlação entre as variáveis.

Mais do que a percepção visual, o diagrama de dispersão consegue dar *insights* sobre a possível presença de *outliers* dentro do espaço amostral que corresponde a uma determinada população. Além disso, já é possível verificar a presença de clusters “óbvios”, ou agrupamentos dentro da mesma amostra.

3.1.2.2 Representação de Estrela

A segunda técnica de representação a ser utilizada é a representação de estrela (RE). Esse tipo de técnica gráfica se insere no contexto das chamadas representações pictóricas. O poder dessas representações é bastante grande pois induz o cérebro do ser humano de maneira diferente. Os olhos reagem a esse tipo de representação mais sensivelmente, e portanto é possível captar as nuances da amostra de maneira mais fácil (Johnson & Wichern, 2007).

O problema da representação de estrela é que esta é restrita a um conjunto de observações cujos valores das variáveis que o representam são positivos. Além disso, é necessário um conjunto de variáveis $p \geq 3$.

Em termos de construção, a representação de estrela é bastante simples. Cada eixo representa uma variável e o centro representa o zero absoluto. Se alguma variável possui valor maior que zero, a estrela apontará esse fato.

Também é importante salientar que a representação de estrela denota apenas uma observação. Diferentemente do caso do gráfico de dispersão, a representação de estrela não mostra o relacionamento entre variáveis, mas sim o comportamento destas para cada

observação. A Figura 3.2 abaixo mostra um exemplo de representação de estrela para quatro observações distintas.

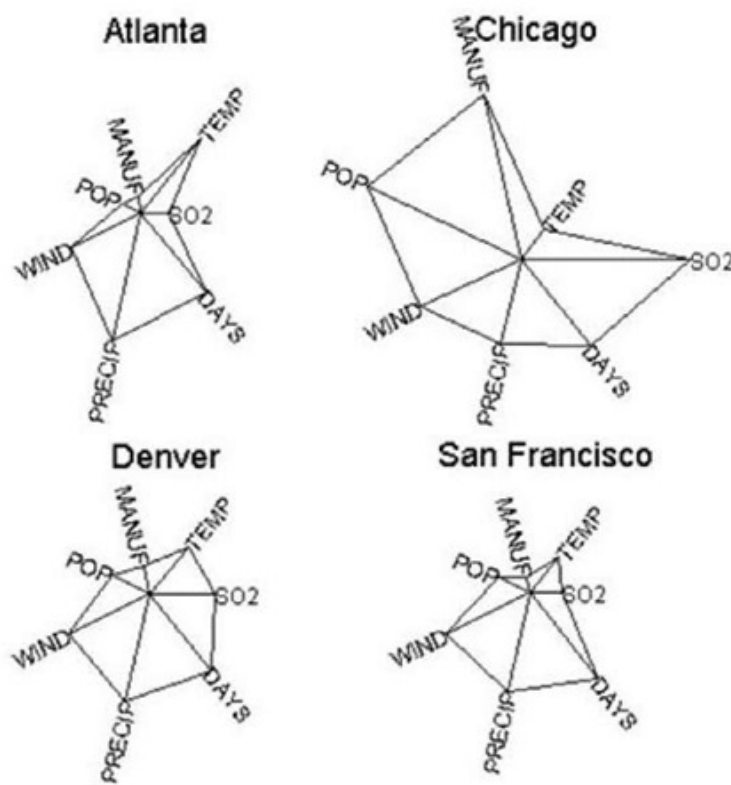


Figura 3.2 – Exemplo de representação de estrela para observações distintas.
(Source: <http://what-when-how.com/statistics/skewness-to-systematic-review-statistics>)

3.1.3 Métodos para Cálculo de Distâncias

Os métodos para cálculo de distâncias são o ponto central na teoria dos clusters. Para se produzir uma boa estrutura de agrupamento é necessário se ter em mente quão “próximas” estão as observações. De maneira análoga, quanto mais próximos os dados, maior a semelhança entre eles.

Talvez um fator que gere certos problemas nesses métodos é paradoxalmente a subjetividade que os cálculos possuem. É importante, antes de escolher o melhor método, entender da onde vêm as variáveis, quais suas escalas de medida e porque foram escolhidas tais variáveis (Johnson & Wichern, 2007). Primeiramente e mais fácil é saber qual escala de medidas utilizarem. Para esse propósito, favor verificar o Apêndice C.

Serão aqui expostos três métodos de cálculo de distâncias: a Distância Euclidiana, a Distância Estatística e a Distância de Minkowski. Cada uma possui suas sutilezas e sua subjetividade.

A distância Euclidiana é a mais conhecida e mais comum para as práticas de clusterização de dados. Também conhecida como a distância da linha reta, porque toma como parâmetro a menor distância entre dois pontos. A formulação matemática pode ser tida como:

$$d_E(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} = \sqrt{(\mathbf{x} - \mathbf{y})'(\mathbf{x} - \mathbf{y})} \quad (3)$$

Já a distância Estatística pega como base a Distância Euclidiana, mas ajusta os dados segundo a covariância entre as variáveis. A formulação matemática se dá como:

$$d_S(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})' \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y})} \quad (4)$$

Finalmente, a Distância de Minkowski incorpora da distância Euclidiana a metodologia de cálculo, mas enfraquecendo distâncias pequenas e grandes. Ou seja, quanto maior a distância, menor o valor de Minkowski, e ao contrário, quanto menor a distância, maior o valor de Minkowski, em comparação a distância Euclidiana. A formulação matemática é dada como:

$$d_M(\mathbf{x}, \mathbf{y}) = \left[\sum_{i=1}^p |x_i - y_i|^m \right]^{\frac{1}{m}} \quad (5)$$

Como já mencionado anteriormente, a escolha do método mais adequado baseia-se em questões qualitativas. Assim sendo, antes de determinar qual método a ser utilizado, é necessário examinar a tabela de dados.

Finalmente, é possível ainda entender a relação entre distância e similaridade de observações. Isso pode ser esclarecido apenas por uma formulação matemática simples dada a seguir:

$$s_{ik} = \frac{1}{1 + d_{ik}} \quad (6)$$

Pode ser visto da relação acima que a similaridade é dada pelo inverso da distância entre observações. Mais do que isso, haja visto que as distâncias são sempre valores positivos, é possível inferir que a medida de similaridade é algo dentro do intervalo $0 < s_{ik} \leq 1$.

3.1.4 Comentários Finais

Nesse parágrafo foram apresentadas as ferramentas estatísticas básicas, que dão suporte a uma análise quantitativa do modelo. Em seguida as representações gráficas tomaram lugar na revisão literária, pois são de suma importância para a análise qualitativa do modelo.

Por fim, foi mostrado o coração do modelo de clusters, que são os métodos de cálculo de distâncias entre observações.

Mais do isso valem aqui alguns comentários. O primeiro diz respeito a dependência da variáveis dentro do modelo. Se as variáveis são independentes, a matriz de covariância se

torna uma matriz diagonal. Por essa razão, as distâncias Euclidianas e de Minkowski são a melhor escolha.

No caso das observações serem independentes, essa hipótese é fundamental para o funcionamento do modelo, pois só assim a matriz de covariâncias será diferente de zero. Caso isso ocorra, é interessante utilizar a Distância Estatística para cálculo de similaridades.

Por fim, é necessário que o número de observações seja maior que o número de variáveis. Isso garante a mesma relação que citada anteriormente, que seria a inversibilidade da matriz de covariâncias.

3.2 Teoria dos Clusters

No parágrafo anterior foi discutido o tratamento dos dados de maneira geral, tanto com estatística básica como métodos de representação gráfica e métodos para cálculo de distâncias.

Já nesse parágrafo, um importante passo para concretização do modelo proposto será apresentado. Trata-se da Teoria dos Clusters. Serão apresentados os conceitos e algoritmos de clusterização de dados, tanto os hierárquicos como os não hierárquicos. Mas antes de tudo, será exposto o tratamento dos dados necessário antes da aplicação de qualquer algoritmo.

E por fim, será exposto aqui o conceito de Análise de Discriminantes, um importante conceito que será o fechamento do modelo a ser desenvolvido.

3.2.1 O Cluster e a Análise de Clusters

“Even though there is an increasing interest in the use of clustering methods in pattern recognition, image processing and information retrieval, clustering has a rich history in other disciplines such as biology, psychiatry, psychology, archaeology, geology, geography, and marketing”

(Jain, Murty, & Flynn, Data Clustering: a Review, 1999).

Um bom jeito de introduzir a análise de clusters é mostrando suas aplicações crescentes, e como essa teoria revoluciona amplas áreas do conhecimento. Importante salientar que o cluster é baseado numa questão de reconhecimento de padrões. Assim sendo, a missão central de qualquer algoritmo de clusterização é encontrar padrões que sejam ditos “semelhantes”.

Assim sendo, é possível agora conceituar o Cluster e a análise de Clusters. O que realmente está por trás dessa palavra? Ou, pura e simplesmente, o que significa dizer “cluster”? Algumas definições podem ser explicitadas.

“[Cluster]... is the organization of a collection of patterns (usually represented as a vector of measurements, or a point in a multidimensional space) into clusters based on similarity.”

(Jain, Murty, & Flynn, 1999)

“Cluster analysis seeks to reduce the dimensionality of data by grouping objects into a small number of groups, or clusters, whose members are more similar to each other than they are to objects in other clusters.”

(Manton, Lowrimore, Yashin, & Kovtun, 2005)

“La cluster analysis è una tecnica di analisi multivariata attraverso la quale è possibile raggruppare le unità statistiche, in modo da minimizzare la ‘lontananza logica’ interna a ciascun gruppo e di massimizzare quella tra i gruppi.”⁶

Tais definições tornam possível a compreensão de alguns detalhes vitais sobre a Teoria dos Clusters. E tais detalhes são guia para uso dessa teoria dentro do modelo proposto. Em resumo, é importante guardar as seguintes nuances sobre clusters:

- a. Cluster é um grupo de objetos com comportamentos similares, expressidos por um dado conjunto de variáveis;
- b. Análise de Clusters é um procedimento de agrupamento de observações dentro de diferentes clusters com o propósito de estudar diferentes comportamentos de grupo.

Por fim, qualquer procedimento de clusterização segue uma metodologia básica, baseada em alguns passos (Jain & Dubes, Algorithms for Clustering Data, 1988).

- I. Representação dos dados;
- II. Definição do que é similaridade;
- III. Aplicação do algoritmo correto;

⁶ Fonte: http://host.uniroma3.it/facolta/economia/db/materiali/insegnamenti/185_903.pdf

IV. Análise dos grupos formados;

V. Estudo dos clusters obtidos;

Note que tanto a representação dos dados, como a definição das similaridades já foram discutidas no parágrafo anterior. Faltam agora mostrar os algoritmos de clusterização, e a Análise de Discriminantes. Mais do que isso, antes de tudo é necessário abordar algumas exigências do conjunto de dados para que o modelo rode de maneira confiável.

3.2.2 Exigências do Conjunto de Dados

Antes de discutir algoritmos e análises de clusters, é necessário lembrar e aplicar os princípios abordados no capítulo anterior quanto a representação gráfica e métodos para cálculo de distâncias. A importância desses dois procedimentos está intimamente ligada à correta escolha do algoritmo de clusterização. Mais do que isso, uma estrutura mais robusta de clusters deve ter ex-ante um tratamento de dados. A consequência do tratamento dos dados é a escolha correta do melhor algoritmo, além de resultados mais confiáveis.

Assim sendo, o primeiro ponto a ser tratado seria uma análise qualitativa a partir da utilização de representações gráficas. Por exemplo, a partir de um conjunto de dados, contendo n -observações num espaço p -dimensional, um gráfico de dispersão mostra o comportamento das variáveis, segundo a Figura 3.3.

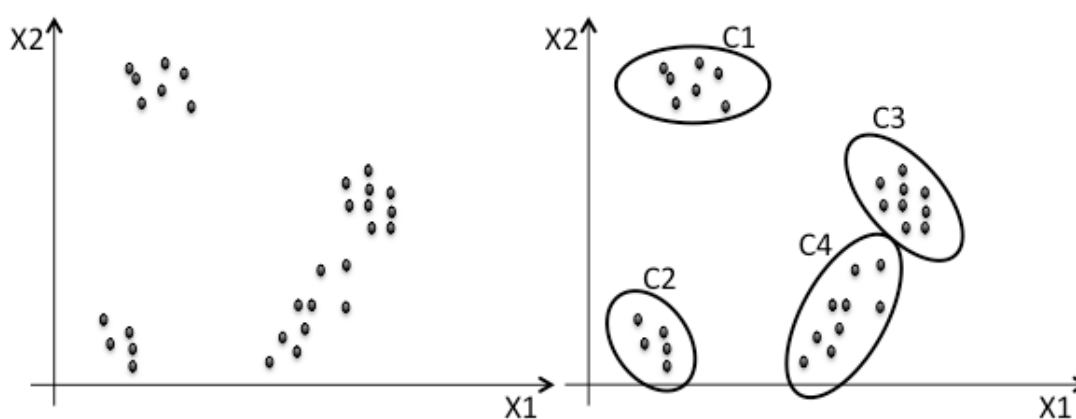


Figura 3.3 – Gráfico de Dispersão para capturar o comportamento das variáveis.
(Desenvolvido pelo autor)

Note-se que a análise no exemplo já captura um padrão de clusters e já dá alguns indícios de quatro padrões populacionais dentro da amostra. Essa primeira ideia é bastante útil na hora de escolher o melhor algoritmo de partição dos dados.

O segundo ponto que merece atenção é quanto ao método de cálculo de distância. No caso do exemplo, pode-se verificar que a correlação entre as variáveis é zero. Os dados estão bastante dispersos no plano cartesiano, logo a escolha pela Distância Estatística não seria interessante.

Para verificar essa conclusão numericamente é necessário calcular a matriz de correlação, onde cada elemento é dado por:

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}}$$

Em termos computacionais, pode-se derivar a matriz de correlação como:

$\mathbf{R} = \mathbf{D}^{-1/2} \mathbf{S} \mathbf{D}^{1/2}$ (7)

No caso da Equação (7), o parâmetro D representa a matriz de desvio padrão da amostra, definida pelos desvios padrão na diagonal principal, e todos os outros valores iguais a zero.

Mas qual a importância da matriz de correlação? A principal informação que esta fornece é quanto à independência das variáveis. Se os valores de correlação são muito próximos de zero, isso significa que uma variável não explica o andamento da outra.

Voltando ao exemplo da Figura 3.3, pode-se notar que as variáveis não estão correlacionadas, e, portanto a Distância Estatística não é a melhor escolha. Mas e quanto a distância Euclidiana e Minkowski, qual a melhor a ser adotada?

Para finalizar a escolha de cálculo de distâncias é necessário analisar o contexto. É necessário verificar como os dados foram obtidos, e qual o escopo do estudo. O principal motivo pelo qual se faz essa análise é porque a Distância de Minkowski tem a habilidade de enfraquecer distâncias grandes e pequenas, fazendo estas girarem em torno de um centroide. Quando as variâncias são grandes, é interessante aplicar a distância de Minkowski. Isso se mostra normalmente com dados difíceis de serem capturados, pois os erros de medida podem ampliar a variância das observações.

É interessante sublinhar também que a distância Euclidiana é a mais utilizada justamente por ser um método intermediário, que inclui características dos dois outros métodos.

3.2.3 Algoritmos de Clusterização

Todo algoritmo de clusterização se baseia na mesma função objetivo. Esse objetivo seria minimizar as distâncias, ou seu dual que é maximizar similaridades. Matematicamente falando, o padrão a ser adotado é:

$$\min f = \sum Distances$$

Assim sendo, são discutidos nesse parágrafo duas classes de algoritmos: os hierárquicos e os não-hierárquicos. Cada classe possui seu pró e contra, além de pontos fortes dependendo do contexto e do resultado esperado.

3.2.3.1 Métodos Hierárquicos de Clusterização (MHC)

“Técnicas Hierárquicas de Clusterização ocorrem com uma série sucessiva de fusões ou divisões.” (Johnson & Wichern, 2007). Além de da função objetivo, todo cluster hierárquico é motivado pela lei descrita acima. Mais do que isso, não existe uma determinação prévia de quantos clusters serão obtidos.

Existem dois algoritmos principais dentro dos MHC que traduzem a principal característica desse grupo de métodos: o single linkage e o complete linkage. Tecnicamente, os resultados obtidos são os mesmos, mas as premissas adotadas são diferentes. Ambos são chamados métodos aglomerativos, pois funcionam a partir de fusões de observações. Todo algoritmo MHC segue o mesmo padrão de passos a serem seguidos (Johnson & Wichern, 2007):

1. O *input* inicial é de N clusters, cada um contendo uma única observação, gerando uma matriz $N \times N$ de distâncias $D = \{d_{ik}\}$.
2. Pesquisa inicial dentro da matriz de distâncias pelo par de observações mais próximo.
3. Agrupar as duas observações, formando um novo cluster.
4. Atualizar a matriz de distâncias.
5. Repetir os passos 2, 3 e 4 até obter um total de um cluster somente.

Para compreender melhor o procedimento acima descrito, é interessante mostrar de maneira gráfica. Assim sendo, a Figura 3.4 apresenta o procedimento de clusterização. Tal representação gráfica recebe o nome de dendograma, e é bastante utilizada pelos MHC.

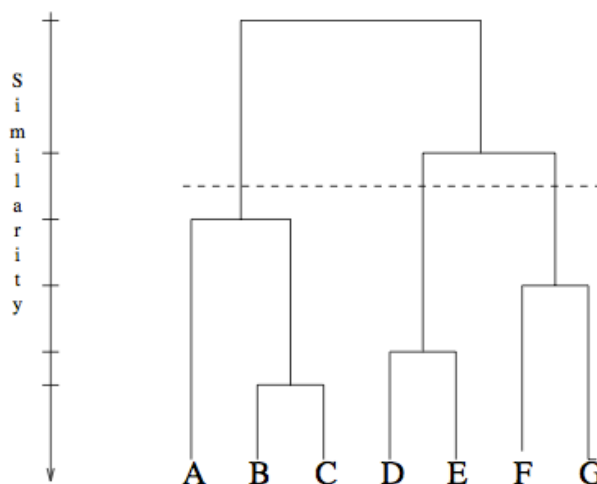


Figura 3.4 – Dendograma representativo de métodos MHC.

(Jain, Murty, & Flynn, 1999)

É possível notar que o número de clusters pode variar de N , onde esse é o número total de observações, até um cluster somente. Esse é o grande ganho dos MHC, pois dependendo do grau de similaridade que se deseja obter, o número de clusters resultantes pode variar. Ou seja, a variável exógena deixa de ser o número de clusters requeridos e passa a ser o nível de proximidade entre observações.

3.2.3.1.1 Algoritmo Single Linkage

O mecanismo de Single Linkage segue os passos básicos já descritos, a diferença, como já dito, está nas premissas adotadas. Nesse caso, a premissa principal está no fato de que o recálculo da matriz de distâncias utiliza a mínima distância entre as duas observações agrupadas e uma terceira observação. Matematicamente falando:

$$d_{(UV)W} = \min \{d_{UW}, d_{VW}\}$$

3.2.3.1.2 Algoritmo Complete Linkage

O mecanismo de Complete Linkage também segue os passos tratados acima, e novamente difere na questão da premissa adotada para o quarto passo. Nesse caso, a matriz de distâncias é recalculada com base na máxima distância entre as observações e o cluster formado na iteração. Matematicamente falando:

$$d_{(UV)W} = \max \{d_{UW}, d_{VW}\}$$

3.2.3.1.3 Diferenças entre Algoritmos MHC

A diferença entre os dois algoritmos é bastante visual. A Figura 3.5 abaixo mostra como se dão as diferentes premissas abordadas pelos dois tipos de algoritmo MHC.

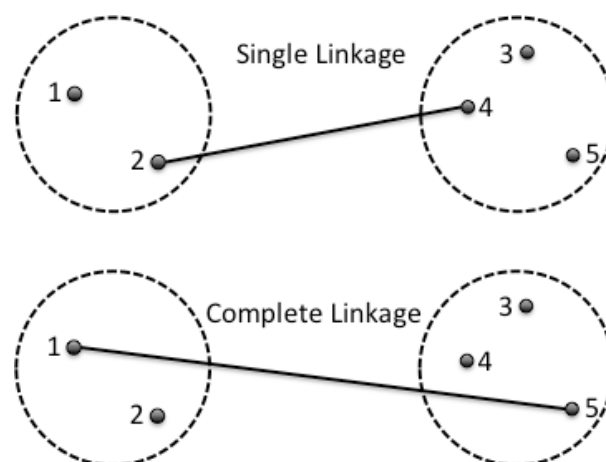


Figura 3.5 – Análise visual da diferença dos MHC.
(Johnson & Wichern, 2007)

Noe que no Single Linkage a distância entre os dois clusters é dada por d_{24} (distância mínima), e no caso do Complete Linkage, a distância escolhida é expressa por d_{15} (distância máxima).

Além disso, é interessante dizer que no caso do modelo desenvolvido nesse estudo o ponto de força desses dois algoritmos está na escolha intermediária dos clusters, segundo a escolha de um “limite de distância tolerável”. Ou seja, o importante aqui é o resultado intermediário.

3.2.3.2 Métodos de Clusterização Não Hierárquicos

Os métodos de clusterização não hierárquicos (MCNH), também conhecidos como métodos particionais, tem como característica básica a adoção prévia de um determinado número de clusters K . Bem diferente do caso anterior, onde o *input* inicial era o nível de distância tolerável. Uma outra característica desses tipos de algoritmos é que estes devem ser rodados algumas vezes para fornecer um resultado consistente.

Para efeito desse estudo, o foco será apenas em um tipo de MCNH, o Algoritmo das K-médias.

3.2.3.2.1 Algoritmo das K-médias

Este é com certeza o algoritmo mais popular para clusterização de dados, devido sua simplicidade e robustez nos resultados. Algumas premissas e hipóteses devem ser expressadas antes de conhecer a fundo o passo-a-passo do algoritmo (Hartigan, 1975).

- Existe um número pré determinado de clusters K ;
- Seja $L(i)$ L -ésimo cluster que contem a i -ésima observação;
- Seja $N(L)$ o número de observações que o L -ésimo cluster contem;
- Seja $B(L, j)$ a média da j -ésima variável que explicam cada cluster L ;
- Seja $D(i, L)$ a distância entre o i -ésimo caso e o L -ésimo cluster.

O objetivo do primário do algoritmo é minimizar as distâncias, a partir de um objetivo secundário que é minimizar o erro de partição. Matematicamente falando:

$$\min error = \sum_{i=1}^n D[i, L(i)]^2$$

Finalmente, os passos para o algoritmo são:

1. Relacionar cada observação a um cluster inicial;
2. Computar todas as medias dos clusters inicialmente formados e o erro inicial;
3. Para o primeiro caso, computar para cada cluster L :

$$ie = \frac{N(L)D(1,L)^2}{N(L) + 1} - \frac{N[L(1)]D[1,L(1)]^2}{N[L(1)] - 1}$$

Que indica o aumento do erro ao transferir o primeiro caso do cluster L(1) para um cluster L. Se ie é negativo, é interessante fazer a mudança.

4. Ajustar as médias dos clusters e o erro;
5. Repetir os passos 3 e 4 para todos os casos;
6. Se não houver nenhum movimento, parar. Se não, voltar ao passo 3.

O principal nesse algoritmo é justamente a partição dos dados. A questão que surge então é: como alocar inicialmente uma observação em um cluster? Para sanar tal indagação a proposta é combinar os dois métodos de clusterização vistos.

3.2.4 Análise de Discriminante

Depois de estabelecer um método de formação de clusters a partir de um determinado algoritmo, é interessante encontrar um jeito de determinar se uma nova observação pertence a um determinado cluster. Seria como uma análise de afiliação de um determinado projeto a uma determinada tipologia. Essa análise se chama Análise de Discriminante.

Para análises de observações e dois grupos possíveis, é bem simples alocar as observações, caso ambos os grupos sejam representações de populações normalmente distribuídas. Entretanto, para mais grupos, a análise se torna um pouco mais complexa (Johnson & Wichern, 2007).

Mas por questão de simplificação do estudo, e relaxando a hipótese de normalidade da população, é possível dizer que uma observação multivariada pertence a um determinado grupo se, e somente se:

$$(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \leq \chi_p^2(\alpha) \quad (8)$$

Onde $\boldsymbol{\Sigma}^{-1}$ é a inversa da matriz de covariâncias, e $\chi_p^2(\alpha)$ é a distribuição chi-quadrado, com p graus de liberdade (Johnson & Wichern, 2007). Fazendo isso, é possível dizer que com certa probabilidade α a nova observação pertence ao cluster analisado. A Figura 3.6 mostra essa relação.

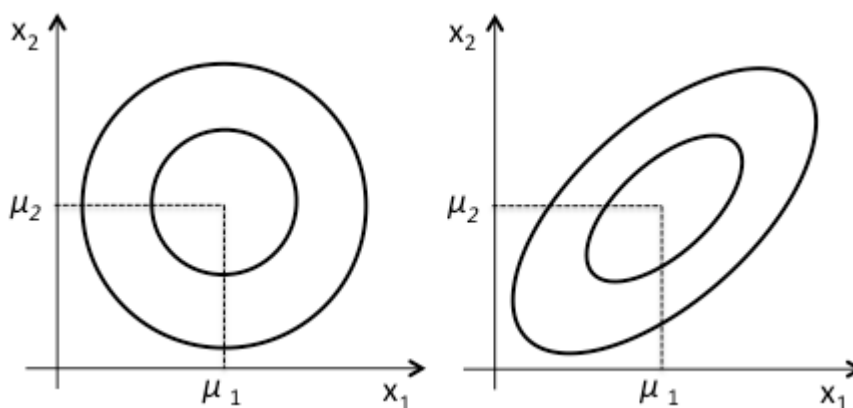


Figura 3.6 – Espaço com probabilidade alfa de conter uma observação genérica.
(Johnson & Wichern, 2007)

Somente note que em caso de Σ^{-1} ser uma matriz diagonal, ou seja, as variáveis sem correlação, o espaço de maior probabilidade toma a forma de uma circunferência. Também note que o círculo com raio maior tem maior probabilidade associada, e vice-versa.

3.3 Resumo do Capítulo

Neste capítulo foram discutidos todos os aspectos matemáticos que farão parte do modelo. Isso tudo a fim de garantir uma boa base quantitativa para o estudo.

Mas, para ir um pouco além, foram abordados alguns aspectos que também auxiliam na boa percepção qualitativa do modelo. Logo, esse capítulo conseguiu explorar bem os dois lados que se complementam, e portanto, tornam o estudo muito mais consistente e robusto.

4 Desenvolvimento do Modelo

Nesse capítulo finalmente é desenvolvido o modelo matemático do estudo proposto, uma vez que no capítulo anterior foram expostas todas as ferramentas necessárias, não somente para desenvolver um modelo puramente quantitativo. A força desse estudo está justamente no mix ótimo de quantitativo e qualitativo que ele aborda.

Para correta elaboração do modelo, é necessário antes entender o objetivo do trabalho em geral, e elaborar hipóteses que guiam e formam as fronteiras do estudo. E são justamente esses dois tópicos que serão abordados agora.

4.1 O Objetivo

Uma vez que toda teoria estão consolidada e foi vista com base numa vasta revisão bibliográfica, é interessante agora dedicar um capítulo para considerações do autor sobre o estudo. Um início necessário é entendendo qual a proposta do estudo, descrita pelo seu objetivo.

Foi comentado no primeiro capítulo que o objetivo do estudo é: identificar semelhanças dentro de um conjunto de dados, agrupar esses dados segundo tais semelhanças e estudar os agrupamentos a fim de prever o comportamento de dados futuros semelhantes aos dados coletados.

Relembrando também o parágrafo anterior, onde foi tratado a teoria dos clusters, pode-se dizer que o estudo proposto nada mais é que uma aplicação prática de um processo de clusterização. Uma vez que existe um espaço amostral contendo um determinado número de projetos, o objetivo aqui é agrega-los em grupos de projetos parecidos, segundo algumas variáveis que os determinam. Assim sendo, é possível entender – ou ao menos esse é o intuito – como algumas tipologias de projetos se comportam na média.

Através desse ponto de vista, é possível personificar um pouco o objetivo do trabalho: identificar semelhanças entre uma série de projetos, agrupá-los segundo tais semelhanças a fim de entender seus comportamentos médios e traçar alguns perfis de tipologia de projetos, com o intuito de prever o comportamento base de futuros projetos a serem desenvolvidos.

Mas toda essa explicação só leva em conta os aspectos quantitativos, então é necessário retomar aquilo que foi apresentado no primeiro capítulo, que no caso determinava a grande motivação do estudo realizado. Entender projetos, segundo seus pares semelhantes significa melhorar o processo de Planejamento de futuros projetos. Significa também alimentar o KMS de maneira confiável e de fácil entendimento. E principalmente significa auxiliar a execução de um projeto, já que se conhece ex-ante seu comportamento médio, e

consequentemente gerar sucesso do projeto. Sucesso de um projeto, que expandido, se torna sucesso da empresa como um todo.

Assim sendo, vale somente uma ressalva. Para que o objetivo seja alcançado, mais do que entender o procedimento matemático, é necessário entender as escolhas subjetivas a serem realizadas, como por exemplo, a escolha das variáveis e escolha dos algoritmos.

E finalmente, vale dizer que a análise de impacto do modelo numa empresa não é escopo desse trabalho, mas apenas uma consequência motivacional para a realização do mesmo. Essa confusão não pode ser presente, e nem será, uma vez que as fronteiras do estudo serão bem traçadas e seguidas a partir de agora.

Assim sendo, é hora de formalizar o objetivo do trabalho, e assim, propor a linha de chegada do mesmo, sem que se faça mais, ou que se faça menos.

O objetivo deste estudo é (I) identificar semelhanças dentro de um range de projetos, (II) coloca-los em grupos – clusters – de acordo com as semelhanças, (III) estudar as características de cada agrupamento e (IV) classificar as tipologias de projeto.

4.2 As Hipóteses

Uma vez esclarecido o ponto final do trabalho, no sentido de determinar o limite do estudo, é interessante traçar um caminho ótimo para chegar a tal limite. Justamente esse é o papel das hipóteses a serem determinadas. Mais do que traçar um caminho, o papel de uma hipótese é garantir que o estudo não fuja do escopo e fique sempre dentro de uma fronteira ótima.

Em resumo, as hipótese guiam a discussão e o desenvolvimento do modelo. São validadas a medida do avanço do estudo, e são dependentes, ou seja, uma vez que uma hipótese não é verificada, o modelo pode mudar de direção. No caso desse trabalho, as seguintes hipóteses são traçadas:

- I. As variáveis do modelo possuem correlação positiva;
- II. Os clusters formados com os valores de projetos na data de planejamento são os mesmos que formados com valores de projetos finalizados;
- III. Os clusters obtidos são estatisticamente diversos uns dos outros;
- IV. Todo projeto futuro deve pertencer a um, e somente um cluster.

4.2.1 Hipótese Um

A primeira hipótese, assim como todas subsequentes tem a mesma estrutura. Primeiro passo descrevê-la e explicá-la, segundo mostrar sua importância dentro do estudo e por fim demonstrar como será testada a hipótese. Assim sendo, segue a descrição da primeira hipótese:

H1: Existe correlação estatística entre as variáveis que determinam os projetos. Tanto em termos de projetos planejados, como projetos acabados.

A hipótese H1 determina que, dentro de um conjunto de dados, determinados por um conjunto de variáveis uma certa característica estatística pode ser verificada: ao variar o valor de uma variável, dentro de uma certa observação, varia também o valor de outra, segundo alguma proporção. Essa proporção é determinada pela correlação estatística entre variáveis, assunto abordado no capítulo anterior.

A importância de H1 para o estudo é bastante simples. Uma vez que existem diversos métodos para se calcular a distância entre variáveis, saber que estas estão correlacionadas já é um forte indício para se escolher a Distância Estatística. Mais do que isso, numa representação gráfica, saber que as variáveis estão correlacionadas facilita a identificação de *outliers*. E por fim e mais importante, verificar tal propriedade deixa bem claro que a escolha de variáveis foi da melhor forma MECE. Obviamente sempre existirão correlações entre variáveis, sobretudo pelo fato que o foco do estudo é tratar com o velho paradigma – custo, tempo e performance. Entretanto, uma pequena correlação facilita a interpretação dos resultados finais, pois cada explicação se torna mais abrangente.

No caso do teste de H1, é algo bastante simples. O primeiro passo para se determinar se as variáveis estão correlacionadas ou não é entender bem o conjunto de dados presente. O cálculo da matriz de médias e covariâncias é imprescindível nesse caso. Havendo esses dois valores é possível utilizar-se da Equação (7) para calcular a matriz de correlações.

O segundo passo é baseado na análise da matriz encontrada. Todos os valores devem estar no intervalo $-1 \leq r \leq 1$. Se o valor é zero, ou muito próximo de zero, significa que simplesmente não existe correlação significativa entre as variáveis. Inclui-se os valores “próximos” de zero nesse conjunto pois se trata de uma amostra pequena, com alto grau de variação interna, e sobretudo com um nível de significância baixo para altos níveis de detalhe.

Portanto, para valores dentro de um certo limite, é considerado que a correlação é igual a zero.

O limite a ser considerado não é algo baseado em análise subjetiva, e sim num índice chamado Correlação de Produto-momento de Pearson. Tal índice propõe que dada uma população com características próximas a uma normal, a distribuição das correlações entre variáveis é distribuída como uma população t-student. Assim sendo, pode-se comparar o valor de t-student, com certo nível de significância e $n - 2$ graus de liberdade com a seguinte relação:

$$T(r) = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Assim sendo, o limite tolerante para se dizer que duas variáveis são correlatas varia conforme o tamanho da amostra, e grau de significância da análise. Mais do que isso, é interessante sublinhar que o valor exato da correlação não é importante, somente importa saber a variável binária – verdadeiro ou falso – referente a correlação de dois parâmetros.

4.2.2 Hipótese Dois

A Hipótese Dois, da mesma forma que a hipótese um deve ser primeiro explicitada, para depois ser analisado sua importância para o estudo e finalmente como ela deve ser testada. Assim sendo, a Hipótese Dois pode ser expressa da seguinte maneira:

H2: Os clusters originados das observações de projetos em estágio de planejamento são os mesmos originados das observações de projetos finalizados.

A hipótese H2 implica no fato de que, quando analisados os projetos em fase de planejamento, e agrupados em clusters segundo semelhança com outros projetos também planejados, devem ter mesma configuração de agrupamento que no caso de projetos finalizados.

O impacto de H2 para os resultados finais é tremendo. Uma vez analisado o conjunto de dados, e aplicados o algoritmo de clusterização correto, é necessário que H2 seja verdadeira pois caso não seja, é impossível estabelecer uma relação entre planejamento e fechamento de projetos. Isso implica dizer que comportamentos de projetos são 100% imprevisíveis dentro de uma certa escolha de variáveis, tomando assim caminhos distintos dentro de um sistema determinado pelo velho paradigma.

Ampliando um pouco essa afirmação, é possível dizer que caso H2 seja falsa, o processo de aprendizado se torna menos útil dentro do ambiente de gestão de projetos, pois de nada adianta entender o funcionamento de um projeto, se o próximo a ser realizado poderá seguir um caminho totalmente diferente. Assim sendo, note quão vital para o estudo é a hipótese H2.

No caso do teste da segunda hipótese, apesar de sua importância elevada, é bem simples. Uma vez que os clusters já estão formados, tanto aqueles de projetos em fase de planejamento, quanto os projetos acabados, basta comparar cada agrupamento resultante nos dois cenários. Caso as estruturas de clusters sejam iguais, é possível afirmar que H2 é verdadeira.

Quanto uma análise de sensibilidade desse teste de hipótese de H2 são necessárias algumas palavras. A significância do teste proposto é tão maior quanto melhor for desenvolvido e implementado o algoritmo de clusterização. Portanto, para que isso ocorra, é proposto um procedimento mais sofisticado, com duas etapas e dois algoritmos diferentes, a fim de estabelecer maior confiabilidade no agrupamento gerado. Mas isso será discutido no próximo capítulo, com a aplicação do modelo.

4.2.3 Hipótese Três

Seguindo o mesmo padrão, a Hipótese Três deve ser apresentada, seguida pela discussão de sua importância para o estudo. Logo após essa discussão uma breve apresentação do modelo de teste de H3, e finalmente pode-se seguir para formulação da última hipótese. Assim sendo, a terceira hipótese diz que:

H3: O grupos resultantes da aplicação dos algoritmos de clusterização são estatisticamente diferentes entre si.

A terceira hipótese implica que todo grupo resultante da clusterização do conjunto de dados terá características distintas a ponto de poder ser considerados estatisticamente diferentes. Ou seja, o comportamento esperado por tais grupos, em termos de transformação das variáveis pode ser considera exclusivo para cada grupo.

O impacto de H3 para o estudo se mostra tão forte quanto a segunda hipótese. É uma hipótese delicada, e que deve ser tratada com bastante atenção pois dela que surge a validação do argumento de que existem diversas tipologias de projetos, cada uma com seu comportamento médio descrito.

No caso de H3 não ser verdadeira, não se pode saber ao certo se H2 é verdadeira também, ou seja, cai a zero a confiabilidade de H2. Isso porque se H3 indica que os clusters não são estatisticamente diferentes, não se pode afirmar que existem de fato clusters que traduzem diferentes tipologias de projetos. Em outras palavras, fica impossível saber até que nível de sobreposição estão os agrupamentos formados, e qual a sensibilidade desse agrupamento.

Em resumo, H3 serve como uma espécie de encruzilhada, e até por isso vem apresentada antes de H2 no modelo. É H3 que determina se o modelo pode seguir em diante. Sendo H3 falsa, o modelo simplesmente deve acabar e uma revisão dos dados ou da escolha de variáveis deve ser proposta.

Para testar H3 a ferramenta encontrada é um pouco mais sofisticada. Nesse caso, o instrumento necessário para verificar a veracidade dessa hipótese se chama One-Way MANOVA. Essa ferramenta serve para comparar uma série de médias multivariadas de diferentes populações – ou no caso do estudo, de subpopulações. O algoritmo utilizado para cálculo da MANOVA pode ser encontrado no Apêndice D. Vale dizer somente que tal ferramenta, ainda que complexa, fornece resultados bastante robustos, mesmo com uma amostra pequena de dados, como é o caso desse estudo.

4.2.4 Hipótese Quatro

Finalmente chega a discussão da última hipótese. E novamente será apresentada da mesma forma que as anteriores, com sua definição, importância e verificação. Assim sendo, segue abaixo a definição da última hipótese que guia o modelo desenvolvido.

H4: Todo projeto futuro deve pertencer a um, e somente um cluster de acordo com a verificação a partir dos seus dados planejados.

A quarta hipótese é uma continuação da terceira, mas haja visto que esta é um divisor de águas dentro do modelo, as duas acabaram sendo separadas. A H4 diz que um projeto, quando estiver na sua etapa de planejamento, deve ser associado a somente um cluster, dependendo sua proximidade com o comportamento médio de tal. Vale dizer que todo planejamento de projeto é algo que deve ser revisto a fim de garantir uma boa base para execução. Assim sendo, numa primeira análise já é possível associar o projeto a um determinado grupo de projetos passados, e assim refinar o planejamento de acordo com os conhecimentos dos outros esforços.

E é justamente esse o impacto causado pela H4. Ela garante que um novo projeto será associado a somente um grupo de projetos, e seu comportamento pode ser previsto e estimado de acordo com observações passadas.

Mais interessante ainda, H4 define aquilo que é a última parte do objetivo proposto pelo estudo, pois uma vez que é validada, define que as novas observações serão classificadas corretamente como uma tipologia de projeto.

Para testar H4 é aplicada a Análise de Discriminantes descrita no capítulo anterior. É um tipo de análise também mais sofisticado, e toma como base algumas hipóteses que serão discutidas na aplicação do modelo, pois tem um alto grau de subjetividade envolvido.

Vale comentar também que, caso H4 seja invalidada, todo algoritmo de clusterização deve ser revisto, assim como no caso de H2. Isso porque novamente será possível observar sobreposição de grupos, e consequentemente incertezas estatísticas quanto a tipologias distintas de projetos. É interessante ressaltar, portanto, que H2 e H4 são bastante interligados, mesmo que a verificação de tais hipóteses sejam completamente distintas.

4.3 Desenvolvimento do Modelo

Finalmente é hora de discutir o modelo. Para melhor entendimento, este foi estruturado segundo um fluxograma de processos. Além de melhor visualização, a abordagem sugerida confere maior qualidade no formulamento do modelo, pois garante uma solidez na estruturação do mesmo.

Mas o ponto chave do modelo não é a sua estruturação. A maior força se dá pelo seu conteúdo e abordagem tanto quantitativa quanto qualitativa. Assim sendo, tudo que foi visto até então não foi meramente decorativo, e de alguma forma irá complementar o estudo proposto.

Será proposta uma divisão em quatro grandes grupos de estudo: Tratamento de Dados, Similaridades, Clusterização e Análise Final. Cada grupo será dividido em uma série de ações que devem ser feitas, e assim o modelo se constrói.

A Figura 7.2 a seguir mostra de maneira bastante razoável como se dá a construção do modelo. A primeira vista não parece um fluxograma, mas a medida que é esmiuçado, fica claro seu caráter de continuidade de tarefas a serem realizadas.

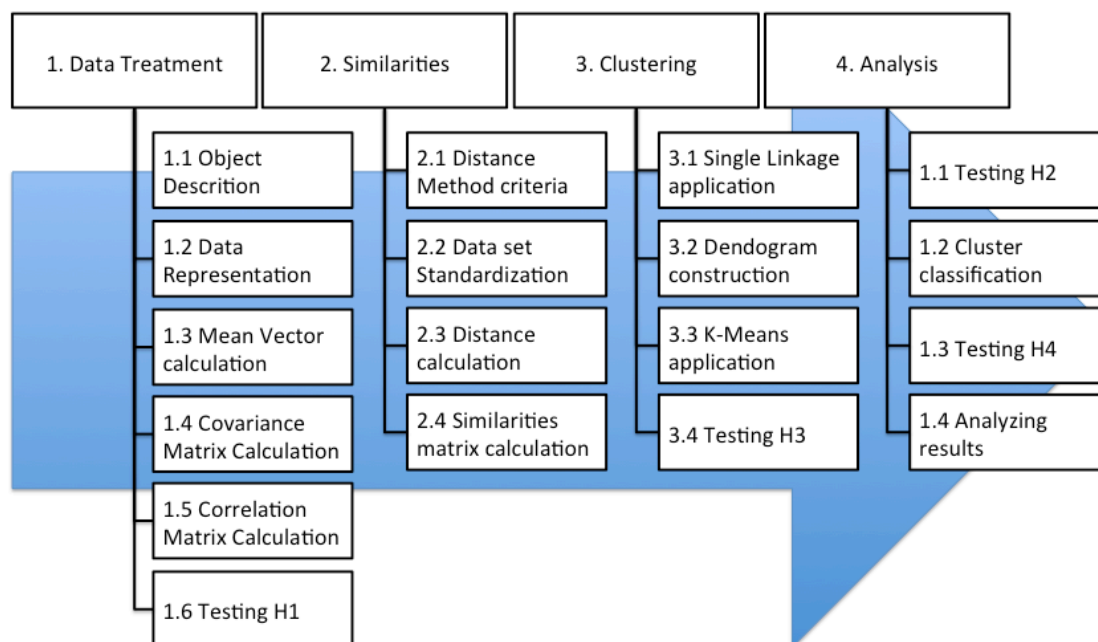


Figura 4.1 – Estrutura do Modelo desenvolvido.

É interessante sublinhar que a divisão de grupos foi feita devido a carga de dificuldade. Ou seja, a idéia é categorizar o modelo dentro de grupos equânimes em termos de dificuldade de realização e entendimento.

4.3.1 Tratamento dos Dados

O primeiro grupo faz literalmente menção ao tratamento dos dados. Uma vez especificadas, as observações serão manuseadas aqui, a fim de proporcionarem subsídios para as outras partes do estudo.

4.3.1.1 Descrição do Objeto

Essa etapa consiste em entender os projetos que compõe o conjunto de dados. Para isso, nada melhor que uma conjunto de perguntas que devem ser respondidas a fim de situar o leitor em questão de sistema analisado.

- Quais são os projetos analisados?
- Quais variáveis melhor explicam o projeto?
- Quantos projetos existem no espaço amostral para representar a população?

4.3.1.2 Representação dos Dados

Considerando então o entendimento dos projetos e do sistema como um todo, é interessante expressar as observações graficamente para entender como estas se comportam, e se representam bem a população de projetos.

Juntamente com a etapa anterior, a representação gráfica dos dados fornece um altíssimo grau de sofisticação para a análise qualitativa dos projetos. É importante entender como as variáveis afetam os projetos, e se já existem alguns padrões associados a subpopulações dentro da população total.

4.3.1.3 Matrizes Notáveis

Estudado tanto o objeto de estudo como as representações gráficas, é possível proceder com a análise quantitativa dos dados. Os principais *outputs* a serem divulgados nessa fase são: vetor de médias, matriz de covariâncias e matriz de correlação. O conjunto desses três dados é chamado matrizes notáveis.

As matrizes notáveis são importantes *inputs* do modelo em si, pois são com elas que as distâncias serão calculadas, e os algoritmos de clusterização serão rodados. Lembrando que sempre haverá revisitação destas, pois um movimento de um projeto de um cluster para outro já muda a perspectiva, tanto da média como da variância.

A matriz de correlação é menos utilizada, mas é fundamental no teste da hipótese H1, que determina por si qual o método utilizado para cálculo das distâncias entre projetos. Dessa forma, ganha esta espaço dentro das matrizes notáveis.

4.3.1.4 Teste de H1

Para fechar o primeiro grupo de atividades do modelo como um todo, é necessário testar H1. Como já foi mencionado, todas as hipóteses funcionam como milestones durante aplicação do modelo. H1 por si também é, mas confere ao modelo o poder de escolha por diferentes caminhos, diferente de outras hipóteses que simplesmente tem o poder de parar o modelo.

Também já foi informado que o teste de H1 é feito com a matriz de correlação. Os dados fornecidos por esta serão confrontados com o índice de de Correlação de Produto-momento de Pearson, segundo cálculo já explicitado.

4.3.2 Similaridades

A segunda fase do modelo é relacionada única e exclusivamente com a análise de similaridades entre projetos. São discutidos nesse ponto a escolha do melhor método de cálculo de distâncias e é feito em seguida o cálculo a partir dessa escolha. Lembrando que é necessário aplicar um método de standardização dos dados antes de calcular as distâncias, algo que é tratado no Apêndice C.

4.3.2.1 A Escolha do Método de Cálculo de Distância adequado

Dentro do cálculo das distâncias, há uma tarefa implícita que é a escolha do melhor método. Para realizar a melhor escolha dentre as três possíveis, é proposto um questionário que tem como *output* a melhor decisão:

- Os dados são bastante esparsos?

Distância de Minkowski tem como propriedade trazer os dados mais para perto de si. Pelo gráfico de dispersão é possível ver quão esparsos são os dados.

- As variáveis são correlatas?

Variáveis correlatas são melhores explicadas pela distância Estatística. A correlação já foi testada por H1.

- Qual o grau de acurácia que se tem em mente?

Um alto grau de acurácia não é preferível utilizar a distância de Minkowski, pois os dados serão agrupados mais facilmente uma vez que as distâncias se encurtam.

- Qual o número de clusters que estão sendo mirados?

Um número pequeno de clusters em relação ao total de observações é melhor talvez com a distância de Minkowski, pois novamente os dados serão agrupados com maior facilidade.

4.3.2.2 Cálculo preliminar das Distâncias

Depois de estabelecer os critérios de escolha do melhor método de cálculo das distâncias, é possível fazer um cálculo preliminar para observar como os dados estão se comportando ex-ante. É importante salientar, entretanto, que os dados devem ser estandardizados, conforme o Apêndice C.

Vale dizer que é aplicado somente um método de cálculo de distâncias, por todo o desenvolvimento do modelo. Mais do que isso, nessa etapa é feito o cálculo preliminar, lembrando que durante a execução do algoritmo de clusterização o método será revisitado, uma vez que os clusters tendem a mudar.

Mais do que isso, segundo a Equação (6) é possível calcular a matriz de similaridades, que permite visualizar com maior grau de detalhes quão semelhantes são as observações entre si, ao invés de ver quão próximas ou distantes são – como é o caso da matriz de distâncias.

4.3.3 Clusterização

Finalmente serão aplicados os algoritmos de clusterização, que fornecem os *outputs* mais importantes do modelo desenvolvido.

Com a finalidade de aumentar a acurácia do modelo, haja visto o número limitado de observações no espaço amostral, é aplicado aqui dois métodos distintos de clusterização. Os dois métodos são de classes distintas de algoritmos, e portanto são complementares. Foram escolhidos para o estudo a aplicação do Método Single Linkage seguido pelo Método das K-Médias.

Assim sendo, primeiro é rodado o Single Linkage, que fornece um mapa estruturado de todos os projetos, e suas distâncias. Com a observação do dendograma é possível estabelecer uma pré-escolha dos clusters, e isso serve como *input* inicial para o algoritmo das k-médias. Dessa forma, pode-se induzir que o algoritmo das k-médias terá um caminho mais curto a percorrer, e com menor erro inicial.

Existem entretanto, alguns critérios que devem ser discutidos antes de toda aplicação dos algoritmos. Pode-se novamente aplicar um questionário com a finalidade de alinhar o modelo de forma certa. Segue o questionário.

- Quantos clusters deseja-se obter?

Pela análise do gráfico de dispersão já é possível se ter a noção de quantas sub-populações compõe a população total.

- Qual o input inicial de distribuição dos projetos dentro dos clusters para inicialização do k-médias?

Basta analisar o dendograma.

Feito tudo isso, é possível finalmente avaliar a hipótese H3. Lembrando que esta fornece a resposta se os clusters são distintos entre si. Dessa forma, se H3 não for validada, o modelo deve ser revisto, e portanto, não pode prosseguir. Para testar H3 é utilizado a MANOVA, cujo funcionamento é descrito no Apêndice D.

4.3.4 Análise Final

A última etapa do modelo consiste na análise. Esta etapa é a mais interessante, pois é dela que sairão os comentários finais, e consequentemente o sucesso ou não do modelo. Sucesso que é medido segundo o ato de alcançar o objetivo proposto anteriormente.

E como toda análise deve começar de maneira robusta, essa etapa começa testando H2, pois é necessário saber se os projetos planejados são os mesmos que os acabados, em termos de distribuição de clusters. Essa hipótese é o guia-mestre de toda etapa de análise final.

4.3.4.1 Consequências de teste de H2

Existem dois caminhos possíveis depois do teste da hipótese H2, que são: H2 é falso e H2 é verdadeiro. Para cada um dos caminhos existem consequências distintas. Segue a análise das consequências.

➤ H2 é falso

- 1) As variáveis escolhidas não são adequadas, ou não correspondem bem ao velho paradigma proposto inicialmente. Pode-se dizer que seu comportamento não é previsível e portanto, não são passíveis de informar possíveis caminhos que os projetos seguem. Voltar a primeira etapa do modelo e mudar as variáveis que explicam os projetos.
- 2) Pode existir algum outlier dentro dos projetos que não foi capturado pelo gráfico de dispersão. Voltar a primeira etapa do modelo e detectar o outlier, removendo-o.

➤ H2 é verdadeiro

- 1) É possível mover com o modelo adiante.

4.3.4.2 Classificando os Clusters e os Futuros Projetos

Assumindo portanto que H2 é verdadeiro, é possível prosseguir com o modelo. Agora é possível classificar os clusters obtidos segundo características que sejam latentes e que sejam facilmente percebíveis segundo análise comportamental das variáveis que os explicam. Dessa forma, todos os clusters obtidos serão rotulados. Vale lembrar que uma vez que foi validado a hipótese que os clusters de projetos planejados são os mesmos dos projetos acabados, estes últimos perdem importância dentro da análise final, e portanto podem ser descartados.

Agora falta somente considerar a análise de Discriminantes e verificar a hipótese H4. Todo novo projeto deve alinhar-se somente a um cluster obtido. Dessa forma, testar-se-á H4 e finalmente poder-se-á analisar a veracidade e robustez do modelo.

Se H4 for verificada, pode-se afirmar que o modelo resulta no objetivo proposto anteriormente. Caso não seja validada H4, a conclusão que se tira é que o novo projeto em questão pode compor um novo cluster distinto, e uma nova análise deve ser feita. Ou, de outra maneira, pode indicar que os clusters não são propriamente MECE, e o modelo deve ser revisto, tendo em vista outras variáveis.

5 Aplicação do Modelo e Resultados Obtidos

Esse capítulo finalmente mostra o modelo em funcionamento, tal como a expressão dos resultados obtidos. A representação aqui abordada segue fielmente o modelo apresentado no capítulo anterior. Entretanto, para se ter uma boa conclusão sobre tudo, é necessário antes abordar um aspecto que tange o passado que não contempla o modelo em si: como os dados foram obtidos.

5.1 O Recolhimento dos Dados e o Contexto

O primeiro que se deve ter é em cima dos dados que são utilizados pelo modelo. Dados que não foram obtidos de maneira correta com certeza não traduzem corretamente as nuances de uma população. Mais do que isso, os dados devem ter um contexto esclarecido, pois todas as análises qualitativas passam por uma adequação do tipo de projetos que está sendo lido.

Em todo o caso, segue a análise sobre o recolhimento dos dados, e em seguida o contexto em que se inserem.

5.1.1 Recolhimento do Dados

Todos os dados para elaboração do modelo foram fornecidos pelo professor Franco Caron, do *Politecnico di Milano*. Os dados fazem parte de uma pesquisa estruturada pelo próprio professor, e alimenta uma série de pesquisas que o mesmo desenvolve junto a um corpo estudantil qualificado. A Tabela 5.1 e a Tabela 5.2 a seguir mostra todos os dados que foram passados para esse estudo.

Vale dizer que não foram feitas mudanças algumas nos dados fornecidos, assim a continuidade do estudo conjunto proposto pelo Prof. Franco Caron pode se manter uniforme, e os resultados condizentes. Mais do que isso, tais dados já foram validados em estudos passados, então será dado aqui que tal conjunto segue um padrão correto de obtenção de dados.

5.1.2 Contexto

De maneira bastante direta, o contexto dos dados é o setor de Petróleo & Gás. Os projetos elencados sugerem grandes projetos de engenharia dentro desse setor da economia. Note que esforços de empresas de Petróleo & Gás trazem sempre grande mobilização de recursos, e portanto devem ser tratados com muita cautela.

Mais do que isso, projetos nessa indústria são altamente complexos, e o ambiente externo é sempre gerido por um alto grau de incerteza. Dessa forma, a escolha das variáveis

que explicam tais projetos deve ser feita de maneira bastante cautelosa, e englobar toda incerteza e complexidade.

Para se entender um pouco melhor o contexto da indústria de Petróleo & Gás é sugerido uma pequena apresentação teórica do assunto.

5.1.2.1 Indústria de Petróleo & Gás

A Indústria de Petróleo & Gás é considerada um dos aspectos chave da economia atual. Mobiliza uma quantidade enorme de recursos, e gera como resultado uma capitalização ainda maior para as empresas do setor. O grande foco dos projetos são plantas de extração de petróleo e gás das reservas espalhadas pelo mundo. E tudo isso com o foco principal de gerar insumos energéticos para o planeta. Por exemplo, mais de 85% da energia consumida pelos EUA em 2008 provinha de combustíveis fósseis.

A indústria como um todo é dividida em dois componentes: *upstream* e *downstream*. A parte *upstream* se refere a exploração direta dos recursos e a produção de petróleo e gás propriamente dita. Já a parte *downstream* se refere a toda distribuição e comercialização dos insumos já prontos⁷.

No caso desse estudo, o foco é nos projetos que compõe o *upstream*. Estes são baseados em dois tipos básicos de atividades: Extração direta de recursos, e Refinarias. Ambos requerem uma quantidade enorme de investimento em bens de capital para que se haja grande produtividade e qualidade no produto. A sensibilidade a erro das grandes indústrias do setor é bastante elevada, ou seja, qualquer que seja o desvio do ótimo pode custar muito aos cofres das empresas.

A complexidade dos projetos advem de duas características interessantes que permeiam qualquer esforço de capital dentro da indústria (Costa Lima & Suslick, 2006):

- Os projetos são em parte irreversíveis: em caso de insucesso do investimento, a empresa não consegue recuperar grande parte do capital investido.
- O futuro incerto afeta o cash flow da empresa: o preço mundial do petróleo, crescimento econômico e taxas de juros afetam em muito os custos e retornos da empresa (Jafarizadeh, 2010).

Assim sendo, é possível observar que a margem para erro é bastante restrita. Um modelo que possa auxiliar o planejamento desses projetos, de modo a ampliar sua capacidade de sucesso é bastante útil nesse contexto. É interessante notar que na indústria de Petróleo & Gás os projetos já tendem a incorporar aprendizados de projetos anteriores, mas nada feito de

⁷ Fonte: www.investopedia.com/features/industryhandbook/oil_services.asp

maneira estruturada, somente um conhecimento que se torna intrínseco ao gerente de projeto, e não acessível a qualquer elemento do time (Amstrong, Galli, Bailey, & Couët, 2004).

Dessa forma, se torna o objetivo não-declarado desse estudo: estruturar um procedimento de aprendizagem para o setor de Petróleo & Gás, com a finalidade de mitigar os riscos inerentes a projetos de grande complexidade e em ambientes de grande incerteza.

5.1.2.2 Variáveis que Descrevem o Modelo

Uma vez explicitado o contexto, é possível então entender a escolha das variáveis que explicam os projeto, e que guiam o modelo como um todo. Antes de tudo, é necessário lembrar que o foi escolhido o velho paradigma como forma de explicação dos projetos. Essa escolha vem sendo explicitada durante todo o desenvolver do estudo.

Como já foi dito, entretanto, o contexto é bastante complexo, e o ambiente bastante incerto. Altos investimentos devem ter essas problemáticas em mente, antes do desembolso de grandes cifras. Por tal motivo, o investidor não põe seu dinheiro em algo que sabe que não trará retornos acima da média – é tudo uma questão de risco/rendimento. Dessa forma, uma das variáveis escolhidas deve ter consigo essa noção do investidor do risco que envolve o projeto. Além disso, existem algumas variáveis base que devem explicar o projeto segundo uma visão do velho paradigma.

Por fim, haja visto que o número de observações é bastante restrito – apenas 16 projetos são analisados – o número de variáveis deve ser também pequeno, a fim de concentrar o mínimo de erro possível nas análises.

Considerando tudo isso, foram escolhidas apenas três variáveis para aplicar o modelo de clusterização proposto:

- Custo;
- Tempo;
- Grau de Alavancagem.

As duas primeiras variáveis são facilmente explicadas, pois constituem o velho paradigma em termos explícitos. O que vale um comentário mais aprofundado aqui é o Grau de Alavancagem. Haja visto toda a discussão acerca as incertezas e riscos para um investidor que coloca seu dinheiro num projeto de Petróleo & Gás, é de se entender que este não irá alavancar seu investimento sem que haja uma certeza maior dos resultados do projeto. Assim sendo, um alto grau de alavancagem determina a crença de que o projeto tem menos riscos envolvidos, e retornar mais certos – logo, o investidor financia o projeto com capital externo, que é certamente mais barato. Da mesma forma, um baixo grau de alavancagem mostra que o

investidor está colocando seu próprio dinheiro a disposição do projeto, logo os riscos são maiores, e os retornos também.

Vale também o comentário que essas três variáveis conseguem ser em bom gênero MECE, pois não há sobreposição de conceitos nas três. A questão da correlação será abordada logo em seguida.

5.1.2.3 A Base de Dados e a Força das Estatísticas

Assim como já foi dito, a base de dados é composta de dezesseis projetos que configuram o espaço amostral, mais dois projetos extras que determinarão a Análise de Discriminantes. Existem dois tipos básicos de projetos: os marítimos e os offshore. Essa distinção não pode ser considerada uma variável para o modelo, ainda que mostre de maneira binária uma distinção entre projetos. Mais do que isso, é possível verificar o lugar de origem dos projetos, ou seja, onde foram executados.

Tanto a variável binária de tipologia quanto a variável de localização indicam talvez uma tendência associada a risco do projeto. Mas, uma vez que o grau de alavancagem já engloba a variável de risco, estas outras duas serão desprezadas. De qualquer forma, é possível observar a Tabela 5.1 e a Tabela 5.2 a seguir com todos os dados, tanto dos projetos planejados quanto aqueles acabados.

Tabela 5.1 – Base de dados dos projetos planejados.

Projeto	Type	Country	CAPEX @FID	Leverage	Duration	CAPEX Equity
1	Subsea	Nigeria	\$434.433.000,00	49,81%	17	\$218.042.000,00
2	Subsea	Norway	\$4.812.281.000,00	92,10%	28	\$380.223.000,00
3	Subsea	Angola	\$1.550.955.000,00	80,00%	30	\$310.191.000,00
4	Subsea	USA	\$587.600.000,00	25,00%	15	\$440.700.000,00
5	Subsea	Norway	\$1.724.064.000,00	70,00%	31	\$517.219.000,00
6	Subsea	Angola	\$1.941.010.000,00	80,00%	29	\$388.202.000,00
7	Subsea	Egypt	\$785.370.000,00	50,00%	18	\$392.685.000,00
8	Subsea	Angola	\$3.749.855.000,00	80,00%	37	\$749.971.000,00
9	Offshore	Italy	\$518.508.000,00	27,36%	26	\$376.621.000,00
10	Offshore	Congo	\$289.896.000,00	0,00%	20	\$289.896.000,00
11	Offshore	Tunisia	\$253.060.000,00	51,00%	15	\$124.000.000,00
12	Offshore	Australia	\$445.870.000,00	0,00%	32	\$445.870.000,00
13	Offshore	Egypt	\$326.340.000,00	50,00%	18	\$163.170.000,00
14	Offshore	Congo	\$227.507.000,00	0,00%	18	\$227.507.000,00
15	Offshore	Egypt	\$202.633.000,00	40,00%	15	\$121.580.000,00
16	Offshore	Egypt	\$190.253.000,00	36,82%	17	\$120.200.000,00
T1	Subsea	Nigeria	\$536.553.000,00	30,00%	17	\$375.587.000,00
T2	Offshore	Tunisia	\$214.980.000,00	51,00%	19	\$105.341.000,00

Tabela 5.2 – Base de dados dos projetos acabados.

Projeto	Type	Country	CAPEX Actual	Leverage	Duration	CAPEX Equity
1	Subsea	Nigeria	\$486.351.000,00	49,81%	27	\$244.100.000,00
2	Subsea	Norway	\$7.732.216.000,00	92,10%	35	\$610.845.000,00
3	Subsea	Angola	\$1.810.417.000,00	80,25%	42	\$357.525.000,00
4	Subsea	USA	\$525.400.000,00	24,78%	22	\$395.200.000,00
5	Subsea	Norway	\$1.623.758.000,00	70,53%	30	\$478.466.000,00
6	Subsea	Angola	\$2.120.870.000,00	80,00%	31	\$424.174.000,00
7	Subsea	Egypt	\$1.009.860.000,00	50,00%	23	\$504.930.000,00
8	Subsea	Angola	\$4.030.929.000,00	80,53%	37	\$785.000.000,00
9	Offshore	Italy	\$598.549.000,00	27,00%	32	\$436.933.000,00
10	Offshore	Congo	\$549.103.000,00	2,94%	32	\$532.950.000,00
11	Offshore	Tunisia	\$345.667.000,00	51,00%	26	\$169.377.000,00
12	Offshore	Australia	\$1.063.100.000,00	0,00%	39	\$1.063.100.000,00
13	Offshore	Egypt	\$550.639.000,00	51,44%	17	\$267.395.000,00
14	Offshore	Congo	\$310.137.000,00	0,00%	17	\$310.137.000,00
15	Offshore	Egypt	\$264.900.000,00	40,02%	20	\$158.900.000,00
16	Offshore	Egypt	\$199.670.000,00	39,45%	19	\$120.910.000,00
T1	Subsea	Nigeria	\$689.500.000,00	28,56%	21	\$492.600.000,00
T2	Offshore	Tunisia	\$476.802.000,00	49,12%	30	\$242.610.000,00

Surge um problema que já foi apontado anteriormente. O número de observações da amostra é bastante reduzido frente a sofisticação estatística que se deseja aplicar no modelo. Dessa forma é preciso um argumento qualitativo bastante sólido para validar toda a análise quantitativa. Assim sendo, segue uma breve explicação de cada passo estatístico a ser seguido, para que sua aplicação possa ocorrer sem preocupações.

A primeira análise a ser feita é em relação aos básicos vetor de médias e matriz de covariância. Uma vez que a amostra é reduzida e o contexto é bastante complexo, a tentativa de explica-lo torna-se bastante frágil se aumentam os graus de detalhamento das observações. Dessa forma, para cálculo de médias e covariâncias um número reduzido de amostra pode ser explicado por um número reduzido de variáveis.

Vale lembrar aqui a escolha das variáveis feitas anteriormente: as três variáveis escolhidas contém uma grande quantidade de informações, sem detalhar nada, mas tampouco excluindo análises necessárias.

Ademais, para se medir a robustez dos cálculos estatísticos básicos é possível extrair a seguinte relação (Johnson & Wichern, 2007):

If $n \leq p$, that is (sample size) \leq (number of variables), then

$$|S| = 0, \text{ for all samples}$$

Tal relação indica que se o número de variáveis é grande, são necessárias mais observações para compor a amostra a fim de se extrair uma matriz de covariâncias com determinante diferente de zero – e consequentemente reversível. Mais do que isso, assumindo a escolha feita de três variáveis, qualquer subamostra escolhida três a três tem garantida sua matriz inversa de covariâncias. Assim sendo, mesmo diminuindo a três o número de observações, ainda sim a matriz de covariância possui inversa, e portanto não é entrave para o desenvolvimento do modelo (Johnson & Wichern, 2007).

O segundo ponto a se tratar é em relação ao método para cálculo das distâncias. Uma vez dado o número de variáveis que explicam o modelo, e que a matriz de covariâncias é reversível, pode-se também induzir-se que existe uma matriz de correlação entre variáveis, e que com grau considerável de significância, é possível dizer se as variáveis são correlatas ou não. Dessa forma, é possível também inferir qual o melhor método de cálculo das distâncias, pois tanto o grau de acurácia a ser escolhido quanto a verificação de correlação entre as variáveis são análises que estão seguras dentro da robustez do método.

Assegurado isso, também é possível notar que o próprio algoritmo de clusterização é robusto também, pois considera somente as distâncias e o número de clusters a serem classificados, como input.

O verdadeiro desafio porém advém do uso da MANOVA e da Análise de Discriminantes. As duas ferramentas utilizam como proxy o fato das populações serem distribuídas normalmente. E a normalidade da população não é algo que pode ser garantido trivialmente. Para relaxar tal hipótese existem duas ponderações:

- I. Todo projeto do setor de Petróleo & Gás é bastante complexo e tolera o mínimo de erro. Mais do que isso, são seguidos padrões que tendem a colocá-los de maneira simétrica em torno de uma média. Pode-se dizer então que dentro da população de projetos existem subpopulações com distribuição normal;
- II. Segundo a ponderação acima, é possível dizer que existem algumas subpopulações normais dentro dos projetos da indústria de Petróleo & Gás. Dessa forma, segundo o Teorema do Limite Central, é possível relaxar a hipótese de normalidade de toda população e considerá-la assim.

Dada toda essa explicação é possível considerar que as ferramentas estatísticas são passíveis de serem utilizadas sem restrições. Toda a análise qualitativa feita é um bom suporte a tal decisão, e os resultados ainda podem ser considerados robustos.

5.2 Rodando o Modelo

Postas todas as considerações prévias, é possível finalmente rodar o modelo desenvolvido no último capítulo. Lembrando que as linhas mestre para o modelo podem ser vistas na Figura 4.1. Serão agora passadas as quatro fases do modelo, e para melhor compreensão expressão dos resultados finais serão colocadas a cada etapa uma Mini Conclusão, com um breve resumo do que foi obtido até então. Posteriormente essas conclusões prévias serão retomadas e contemplarão o resultado completo do modelo.

5.2.1 Primeira Etapa: Tratamento dos Dados

Esta primeira etapa consiste muito mais numa análise qualitativa que quantitativa da base de dados. Dessa forma, seguindo a Figura 4.1 a risca, começa-se a análise em questão pela descrição do objeto de estudo.

5.2.1.1 Descrição do Objeto

Para a primeira ação do modelo, são estruturadas as perguntas que devem definir alguns caminhos a serem seguidos. Todas já foram respondidas anteriormente, mas devem ficar registradas aqui para guiar o modelo, afinal são de suma importância em todos os momentos.

- Quais são os projetos analisados?

Os projetos analisados são grande projetos de engenharia do setor de Petróleo & Gás. Um setor bastante complexo e incerto, logo, com margens de erros bastante pequenas.

- Quais variáveis melhor explicam o projeto?

As variáveis escolhidas para explicar os projetos são: custo, tempo e grau de alavancagem do projeto. Foram escolhidas tendo em base o velho paradigma e também uma análise de risco intrínseca a projetos complexos.

- Quantos projetos existem no espaço amostral para representar a população?

O espaço amostral é bastante enxuto, contendo apenas dezesseis projetos. Entretanto, já foi feita uma análise qualitativa para se medir a robustez do modelo e a eficácia do ferramental estatístico utilizado.

5.2.1.2 Representação do Dados

A representação dos dados da base pode trazer alguns insights sobre o problema como um todo. Na verdade retrata uma visão sistêmica do problema. Assim sendo, para ter essa primeira visualização do todo apresenta-se aqui o gráfico de dispersão por pares de variáveis. Existem dois aspectos sendo investigados aqui: correlação visual entre as variáveis e presença

de alguns clusters naturais, ou seja, facilmente perceptíveis antes da aplicação de qualquer algoritmo.

Mais do que isso, a apresentação dos gráficos de dispersão podem mostrar se a escolha de variáveis é adequada ao modelo a ser desenvolvido. Logo, são analisados aqui, além das variáveis já mencionadas, mostrar-se-á a variável binária de tipologia de projeto. Veja então a Figura 5.1 com as variáveis básicas.

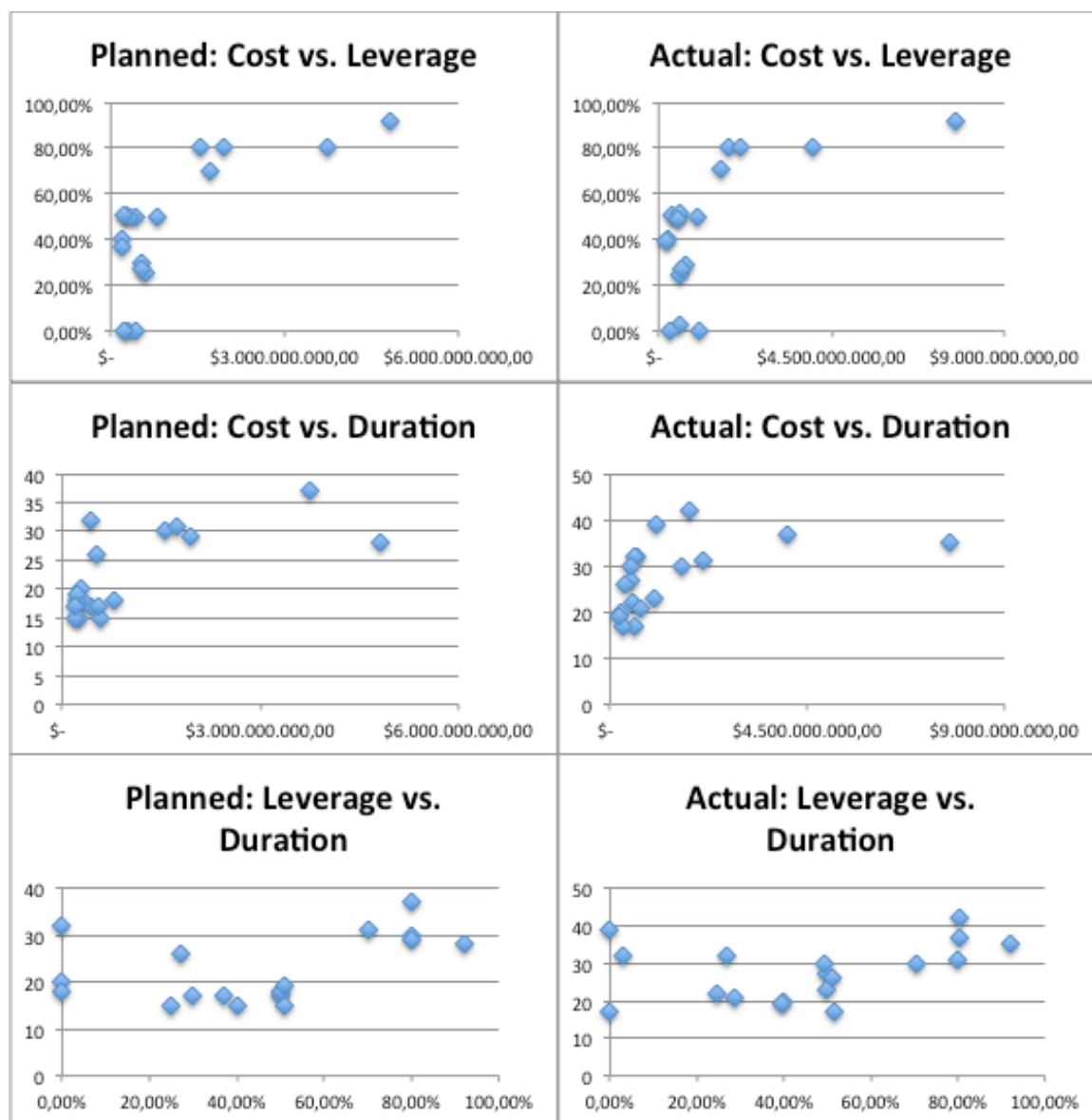


Figura 5.1 – Gráficos de Dispersão entre duplas de variáveis escolhidas para o modelo.

A primeira coisa que vem a mente aqui é a similaridade dos gráficos para variáveis de projetos planejados e acabados. Note como o desenho é bastante parecido, e isso já é um bom indicativo que a Hipótese H2 é verdadeira.

O que emerge em seguida é o senso de correlação entre as duplas de variáveis de Custo & Grau de Alavancagem, e Custo & Tempo. Mais uma vez já é um bom indicativo para que a hipótese H1 seja verdadeira, mesmo que ainda serão testados com maior significância.

O terceiro comentário a ser feito é quanto aos padrões de clusters que surgem. Note que nos gráficos da Figura 5.1 existem algumas áreas bastante densas e outras sem alguma observação. Nesse caso é um bom indicativo de que a hipótese H3 pode ser verdadeira, pois os clusters estão bastante esparsos entre si, logo, são diferentes entre si.

Agora vale a pena considerar a Figura 5.2 onde mostra a relação de uma variável que não foi considerada no modelo com as outras variáveis.

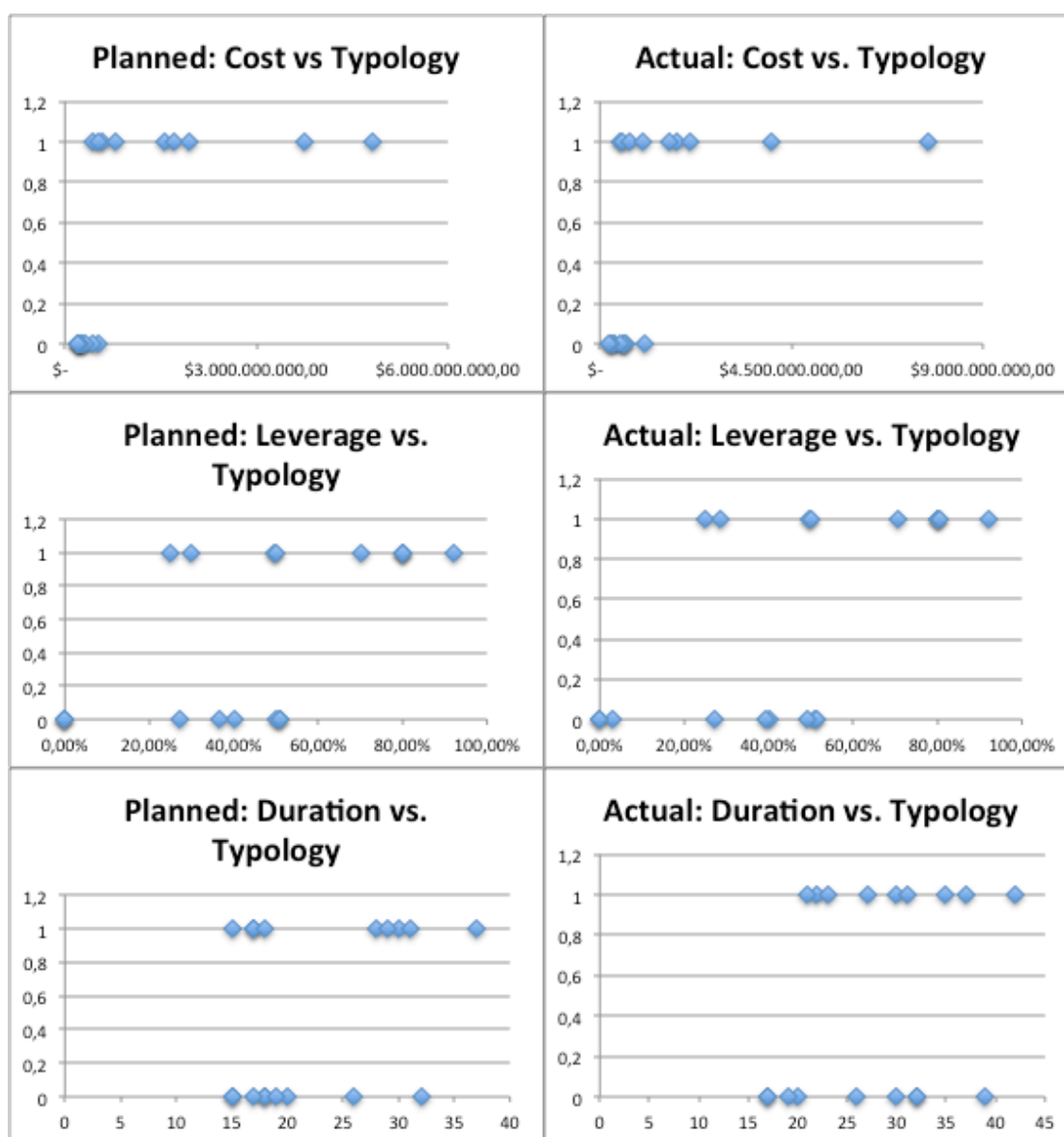


Figura 5.2 – Gráfico de Dispersão com a introdução da variável binária não considerada no modelo.

Nesse caso, pode-se notar que tem termos de tipologia de projeto não há uma clara diferença. No andamento das duas linhas os projetos estão esparsados da mesma maneira,

indicando que de fato a escolha dessa variável não influencia no algoritmo de clusterização. Mais que isso, a introdução de uma nova variável traria erros para o modelo, pois seriam necessárias mais observações para trazer o mesmo resultado de análise.

Assim sendo, haja visto tudo que foi abordado até então é possível escrever aqui a primeira Mini Conclusão, a qual deve ser armazenada para exposição do resultado final ao terminar esse capítulo.

MC1: A tipologia de projetos não é uma variável que deve ser considerada, pois não traz benefícios a análise. Por outro lado, são considerados: custo, tempo e grau de alavancagem. Essa tríade é traz aspectos operacionais e de risco à análise final.

5.2.1.3 Matrizes Notáveis

As matrizes notáveis compreendem: vetor de médias, matriz de covariância e matriz de correlação. Os cálculos são bastante simples, e podem ser realizados sem grandes problemas, assim como exposto no terceiro capítulo. Assim sendo, seguem as matrizes notáveis tanto para projetos planejados como para projetos acabados:

- Projetos Planejados:

$$\bar{x} = \begin{bmatrix} \$1.127.500,00 \\ 46\% \\ 23 \end{bmatrix} \quad S = \begin{bmatrix} 1,8E+18 & 2,8E+08 & 6,3E+09 \\ 2,8E+08 & 8,4E-02 & 8,8E-01 \\ 6,3E+09 & 8,8E-01 & 5,1E+01 \end{bmatrix} \quad R = \begin{bmatrix} 1,00 & 0,73 & 0,67 \\ 0,73 & 1,00 & 0,43 \\ 0,67 & 0,43 & 1,00 \end{bmatrix}$$

- Projetos Acabados:

$$\bar{x} = \begin{bmatrix} \$1.450.000,00 \\ 46\% \\ 28 \end{bmatrix} \quad S = \begin{bmatrix} 3,5E+18 & 3,5E+08 & 7,3E+09 \\ 3,5E+08 & 8,2E-02 & 7,2E-01 \\ 7,3E+09 & 7,2E-01 & 5,9E+01 \end{bmatrix} \quad R = \begin{bmatrix} 1,00 & 0,65 & 0,51 \\ 0,65 & 1,00 & 0,33 \\ 0,51 & 0,33 & 1,00 \end{bmatrix}$$

É interessante notar que os valores das médias e variâncias ainda não foram estandardizados, e por isso encontram-se em unidades de medidas diferentes. Propositamente foi feito dessa forma para conseguir observar em termos reais como as variáveis estão se comportando. Assim sendo, note como as variâncias do custo e do tempo são bastante grandes. Esses valores explicitam as incertezas e complexidade dos projetos de Petróleo & Gas.

5.2.1.4 Teste de H1

Por fim da primeira etapa do modelo é possível testar a hipótese H1, dada a matriz de correlação. O teste escolhido para validar H1 é o chamado Índice de Correlação de Produto-Momento de Pearson. A comparação se dá entre o valor calculado do índice e o valor t-student, com grau de significância α , e $n - 2$ graus de liberdade. Se qualquer valor da matriz

de correlação for maior do que o valor t-student, é possível validar H1 para este caso. Se o valor for menor ou igual ao t-student, considera-se a correlação nula entre tal par de variáveis. Dessa forma, com $\alpha = 5\%$ e $n - 2 = 14$ graus de liberdade, é possível observar que o valor t-student é 1,76. Já o Índice de Pearson é dado pela Tabela 5.3.

Tabela 5.3 – Índice de Pearson para os pares de variáveis.

Correlation Test – T(r)			
Planned Values for the variable			
	V1	V2	V3
V1	0,00	3,18	2,20
V2	3,18	0,00	1,29
V3	2,20	1,29	0,00
Actual Values for the variables			
	V1	V2	V3
V1	0,00	4,05	3,39
V2	4,05	0,00	1,76
V3	3,39	1,76	0,00

Portanto, é possível dizer que com 95% de confiança que existe correlação entre os pares de variável V1 & V2, e V1 & V3. Logo, é possível inferir a segunda Mini Conclusão:

MC2: Com 95% de confiança é possível validar a hipótese H1 e dizer que os pares de variáveis Custo & Grau de Alavancagem, e Custo & Tempo são correlacionados estatisticamente. Essa conclusão já havia sido percebida pelo gráfico de dispersão.

5.2.2 Segunda Etapa: Similaridades

Todos os dados já foram bem tratados, e as similaridades entre alguns projetos já começam a se mostrar presentes. Assim sendo, a essa etapa confere a matematização de algo que foi mostrado visualmente apenas. O que se faz necessário, portanto, é escolher a melhor maneira de calcular as distâncias e similaridades entre projetos.

5.2.2.1 A escolha do melhor método de cálculo de Distâncias

Essa primeira atividade é dada pelo questionário elaborado no capítulo anterior e que fornece como resultado a escolha do melhor método para cálculo de distâncias entre projetos:

- Os dados são bastante esparsos?

Pela análise dos gráficos de dispersão é possível notar que os dados não são muito esparsos, seguem uma regra normal de uma população que foi aproximada para uma normal.

- As variáveis são correlatas?

O teste de H1 indicou que existem dois pares de variáveis correlatas: Custo & Grau de Alavancagem e Custo & Tempo.

- Qual o grau de acurácia que se tem em mente?

Haja visto que existem poucas observações e que estas não são muito esparsas, é possível refinar um pouco o modelo e trazer um grau de acurácia moderado, indicando que a distância de Minkowski talvez não seja a mais indicada.

- Qual o número de clusters que estão sendo mirados?

Pelos gráficos de dispersão já se podem notar algumas tendências de agrupamento, mas não se tem ainda ao certo quantos serão os clusters formados.

5.2.2.2 Cálculo Preliminar das Distâncias e Matriz de Similaridades

Pelo questionário acima proposto já é possível verificar qual o método para cálculo de distâncias que melhor se adequa ao estudo. Primeiro, o tal método deve incorporar as correlações entre as variáveis, o que dá grande parecer a Distância Estatística. Em segundo lugar, haja visto o grau moderado de acurácia que se deseja obter, além do fato dos dados não estarem muito esparsos, isso elimina a possibilidade de utilização da distância de Minkowski. Dessa forma, sem grandes delongas, é escolhido de fato a Distância Estatística para medida de distâncias entre projetos, e posterior cálculo de similaridades.

As distâncias preliminares não são utilizadas pelo algoritmo das k-médias, mas sim pelo Single Linkage. Foram calculadas, mas não traduzem grandes informações ao leitor. Entretanto, com as distâncias foram calculadas as similaridades entre projetos, o que pode ser visto no Apêndice F.

Numa análise preliminar pode-se dizer que os projetos P1, P14 e P16 podem pertencer ao mesmo cluster, assim como P11 e P15. Mas isso é somente uma análise preliminar. Afinal de contas, P1 pode ser muito semelhante a P14, mas que por sua vez é muito discrepante de P16, dessa forma, os três juntos não constituem um cluster. A análise mais profunda fica para o algoritmo de clusterização.

5.2.3 Terceira Etapa: Clusteriação

Uma vez que todas as distâncias já estão computadas e toda a análise de contexto e ambiente já foi executada é possível entrar no processo de clusterização.

Assim como foi dito antes, o modelo se propõe a concatenar dois algoritmos diferentes em prol de uma maior robustez para a formatação final dos grupos. Dessa forma, primeiramente é aplicado o Single Linkage, que fornecerá o input inicial dos grupos para o K-Médias.

5.2.3.1 *Single Linkage e a Representação do Dendograma*

Antes de tudo é necessário tecer um comentário importante para a continuidade do modelo. Haja visto que o algoritmo de Single Linkage é apenas uma ferramenta intermediária para aplicação do k-médias, não é necessário que seja aplicado tanto para os projetos planejados como para os projetos acabados. O input inicial para o k-médias é o mesmo, e como hipótese arbitrária é considerado aqui que a hipótese H2 já é válida. Dessa forma o trabalho se reduz, pois não são necessários esforços para calcular duas vezes o Single Linkage.

Os passos para rodar o algoritmo são aqueles mencionados no terceiro capítulo. Mas o mais importante aqui é o resultado final, ou seja, a representação do dendograma resultante do método. Lembrando que para realização de todos os procedimentos foi utilizada a matriz de distâncias estatísticas calculada anteriormente. Finalmente, a Figura 5.3 revela o resultado do Single Linkage.

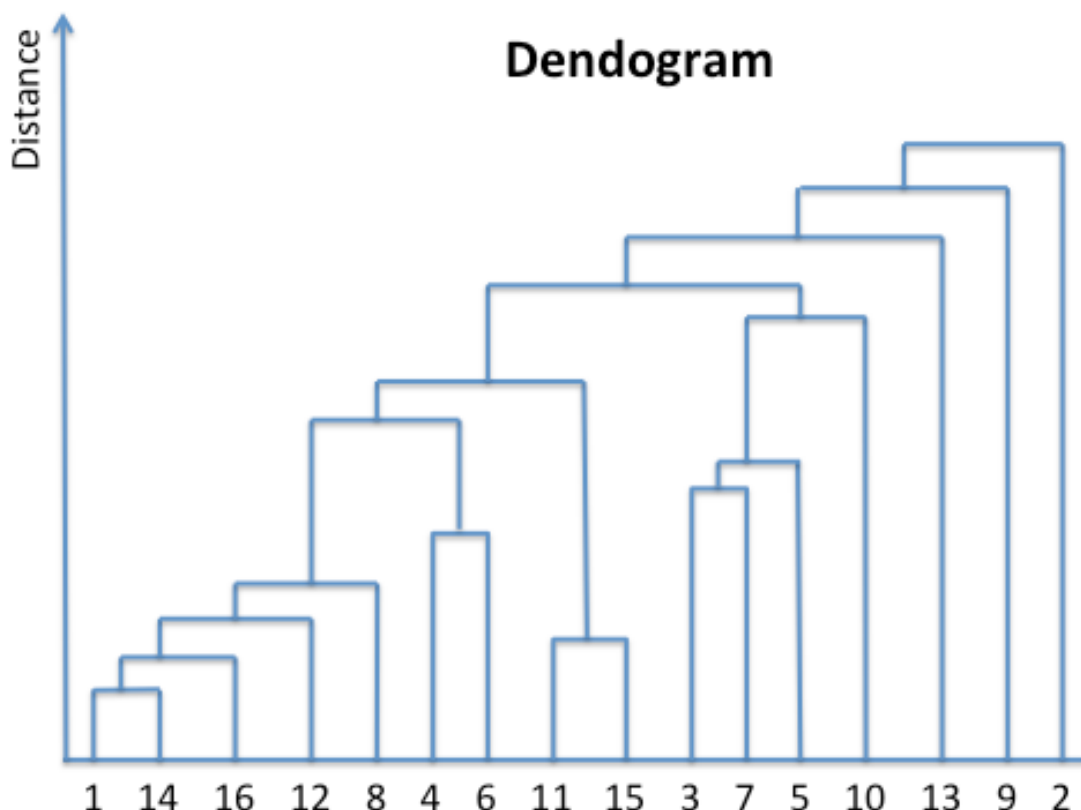


Figura 5.3 – Resultado do Single Linkage para a base de projetos.

É possível notar que de fato o Single Linkage mostrou um resultado que já era esperado: os projetos P1, P14 e P16 muito próximos de si, e os projetos P11 e P15 também. Assim sendo, pode-se já ter uma prévia de que esses dois clusters podem ser inputs iniciais para o algoritmo das k-médias.

5.2.3.2 Algoritmo das K-Médias

Antes de entrar propriamente no existe o questionário exposto no modelo para tratar dos inputs do algoritmo das k-médias.

- Quantos clusters deseja-se obter?

Para tratar desse tópico seria interessante analisar mais de perto os gráficos de dispersão da Figura 5.1. Entretanto estes carregam demasiada informação. Haja visto que já se sabe que o custo é correlacionado com o Grau de Alavancagem e com o Tempo, é possível focalizar somente nessas duas últimas variáveis, pois estas já apresentam uma análise. Assim sendo, a Figura 5.4 mostra o gráfico de dispersão mais direcionado aos propósitos da k-médias.

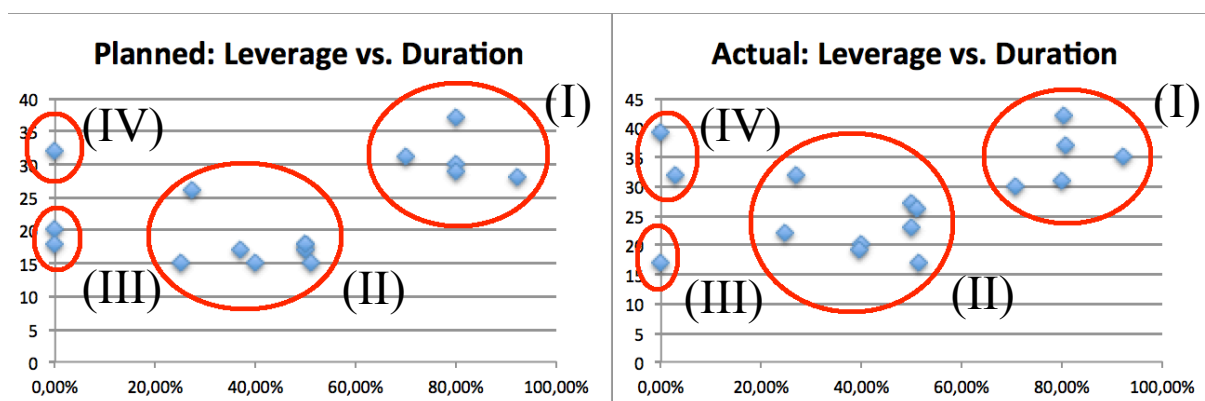


Figura 5.4 – Gráficos de Dispersão para os pares de variáveis correlatas.

É importante ressaltar que essa análise é bastante aberta aos detalhes, então não convém mostrar quais são realmente os projetos em cada agrupamento mostrado acima. O que importa de fato é visualizar que existem quatro padrões observados, e isso que será o *input* do novo algoritmo. Mas já vale também classificar os possíveis tipos de agrupamentos, assim como vistos no padrão da Figura 5.4:

- I. Alta alavancagem e alta duração;
- II. Alta alavancagem e baixa duração;
- III. Baixa alavancagem e baixa duração;
- IV. Baixa alavancagem e alta duração.

Essa primeira análise de tipologia de cluster é bastante importante e precisa ser tratada como uma Mini Conclusão.

MC3: Existem quatro padrões observados dentro da população de projetos, e podem ser considerados os quatro clusters naturais de projetos de Petroleo & Gas. As variáveis que explicam os tipos de projetos são Grau de Alavancagem e Tempo, sendo o Custo uma

correlação de ambas.

- Qual o input inicial de distribuição dos projetos dentro dos clusters para inicialização do k-médias?

Para o segundo input do modelo basta observar o dendograma da Figura 5.3 e verificar quais os clusters iniciais, tendo em vista que são quatro. O arranjo inicial que melhor traduz os quatro clusters requeridos é: (P1, P14, P16, P12, P8, P4, P6), (P11, P15), (P3, P7, P5) e (P10, P13, P9, P2).

Finalmente pode-se aplicar então o Algoritmo das K-Médias, com os inputs iniciais já discutidos. O método foi baseado nos passos apresentados no terceiro capítulo, e auxiliado pelo ferramental desenvolvido pelo autor em código VBA para MS Excel[®]. Os resultados das iterações pode ser visto no Apêndice G. O que importa ser mostrado aqui é o resultado final, ou seja, a configuração que os clusters finalizaram.

- (P1, P7, P11, P13, P15, P16);
- (P2, P3, P5, P6, P8);
- (P4, P9);
- (P10, P12, P14).

5.2.3.3 Teste de H3

A Hipótese H3 deve ser agora testada como forma de garantir o bom resultado do algoritmo das k-médias. Se este foi bem executado, então os clusters podem ser tidos como diferentes entre si. Para teste de H3 é proposto o uso da MANOVA tal qual é apresentado no Apêndice D. O resultado pode ser visto na Tabela 5.4.

Tabela 5.4 – Aplicação da MANOVA para teste de H3.

Source of Variation	Matrix of sum of squares and cross products (SSP)	Degrees of freedom (d.f.)
Treatment	$B = \begin{bmatrix} 0,84 & 0,93 & 0,54 \\ 0,93 & 1,54 & 0,42 \\ 0,54 & 0,42 & 0,42 \end{bmatrix}$	$4 - 1 = 3$
Residual (error)	$W = \begin{bmatrix} 0,37 & 0,09 & 0,03 \\ 0,09 & 0,05 & -0,01 \\ 0,03 & -0,01 & 0,17 \end{bmatrix}$	$16 - 4 = 12$
Total (corrected for the mean)	$B + W = \begin{bmatrix} 1,21 & 1,02 & 0,57 \\ 1,02 & 1,59 & 0,41 \\ 0,57 & 0,41 & 0,59 \end{bmatrix}$	$16 - 1 = 15$

O valor do Parâmetro Lambda de Wilk para a dada tabela é de 0,0069, e consequentemente o Índice de Barlett é 57,24. Mais do que isso, o 5º percentil da distribuição chi-quadrado com $3 \times (4 - 1) = 9$ graus de liberdade é $\chi_9^2(5\%) = 23,59$. Dessa forma pode-se rejeitar a hipótese que as subpopulações possuem mesma média. Em outras palavras, chega-se a conclusão que com 95% de confiança os clusters obtidos pelo algoritmo das k-médias são estatisticamente diferentes entre si.

Assim sendo, é possível formular a quarta Mini Conclusão do estudo em questão:

MC4: Existem dentro dos projetos de Petroleo & Gás quatro clusters naturais que são estatisticamente diferentes entre si.

5.2.4 Quarta Etapa: Análise Final

A última fase do modelo é a análise dos resultados. Essa etapa é bastante importante, não a toa concatena duas hipótese dentro dela, pois é preciso que o modelo esteja bem amarrado nessa etapa. Assim sendo, é necessário já começar validando H2.

5.2.4.1 Teste de H2

O teste de H2 é o mais simples de todos, mas que traz a conclusão mais importante também. A ideia aqui é verificar se os clusters formados por projetos em fase de planejamento são os mesmos de projetos já finalizados. Verificar que H2 é verdadeira é o mesmo que dizer que os projetos tendem a seguir uma linha ou um caminho padrão para cada tipologia observada anteriormente.

Logo, pode-se observar que os clusters formados são os mesmos, tanto em termos planejados quanto em termos acabados, segundo mostra o Apêndice H. É possível fazer essa validação simples graças a validação de H3 que permitiu afirmar que os clusters são diferentes entre si.

5.2.4.2 Classificação Final e Teste de H4

Agora a última etapa antes de poder afirmar que o objetivo do estudo está concretizado. É necessário reafirmar as tipologias de projetos traduzidas pelos clusters encontrados e testar H4 para verificar se novos projetos pertencem somente a um dos clusters estabelecidos. Para tratar das classificações é interessante lembrar o que foi abordado anteriormente. As tipologias encontradas são:

- I. Alta alavancagem e alta duração;
- II. Alta alavancagem e baixa duração;
- III. Baixa alavancagem e baixa duração;

IV. Baixa alavancagem e alta duração.

Pelo Apêndice E é possível verificar que as representações de estrela fornecem bem uma identidade visual de cada um dos clusters. Vale somente dizer antes de tudo que como a hipótese H2 foi validada, os projetos acabados perdem referência dentro da análise final, pois correspondem exatamente aos mesmos clusters que os planejados. Assim sendo, antes de exemplificar o poder da representação de estrela, seria interessante numerar aqui os clusters obtidos:

- C1 = {P2, P3, P5, P6, P8};
- C2 = {P1, P7, P11, P13, P15, P16};
- C3 = {P4, P9};
- C4 = {P10, P12, P14}.

Agora note que pelas representações do Apêndice E, claramente os projetos do cluster C1 são aqueles com alta alavancagem e alta duração, C2 são projetos com alta alavancagem e baixa duração. C3 são projetos com baixa alavancagem e baixa duração, e finalmente C4 são os projetos com baixa alavancagem e alta duração. Assim é conclusivo dizer que de fato foram encontrados padrões de comportamento de projetos dentro do setor de Petróleo & Gás.

Entretanto, antes de finalizar o estudo, é necessário validar a hipótese H4. Para tanto, foram deixadas duas observações em destaque para que fosse feita a análise de discriminantes. São os projetos PT1 e PT2, conforme aparecem na Tabela 5.1 e na Tabela 5.2. Pela representação de estrela, já é possível inferir que PT1 provavelmente cabe dentro do cluster C4 e PT2 provavelmente dentro do cluster C2. Entretanto somente se basear no visual não é a forma correta de fazer. Assim sendo, é aplicada a Análise de Discriminantes, conforme abordado no terceiro capítulo. A Tabela 5.5 apresenta os resultados dos cálculos.

Tabela 5.5 – Tabela com os valores da Análise de Discriminantes.

Clusters	Distance to Clusters' center	
	PT1	PT2
C1	10,22	7,32
C2	0,98	0,50
C3	4,93	4,25
C4	0,44	2,45

Comparando os valores calculados com os valores da distribuição chi-quadrado com três graus de liberdade e 90% de confiança $\chi^2_3(90\%) = 0,58$ pode-se notar que de fato PT1 pertence ao cluster C4 e PT2 pertence ao cluster C2. E mais que isso, pertencem exclusivamente a esses dois agrupamentos. Vale aqui somente um comentário que o valor de significância foi baixo ($\alpha = 90\%$) pois o tamanho da amostra é pequeno, e consequentemente

o tamanho dos clusters acabou ainda mais reduzido. Dessa forma, houve uma perda no nível de significância para a análise de Discriminantes.

O resultado final, entretanto, parece bastante sólido, e portanto, assim se encerra o estudo, com a formulação da quinta e última Mini Conclusão.

MC5: A Análise de Discriminantes mostrou que futuros projetos pertencem a um e somente um cluster dentro as possibilidades elencadas. O nível de confiança para essa análise é de 90%, mas conclusivamente robusto.

5.3 Discussão Final

Para finalizar portanto todas as análises feitas e concluir o estudo é interessante recolocar aqui todas as mini conclusões que foram sendo postas durante a execução do modelo. Essa ferramenta auxilia na visualização do estudo e dos resultados como um todo. Portanto, as cinco mini conclusões foram:

- 1) A tipologia de projetos não é uma variável que deve ser considerada, pois não traz benefícios à análise. Por outro lado, são considerados: custo, tempo e grau de alavancagem. Essa tríade é traz aspectos operacionais e de risco à análise final;
- 2) Com 95% de confiança é possível validar a hipótese H1 e dizer que os pares de variáveis Custo & Grau de Alavancagem, e Custo & Tempo são correlacionados estatisticamente. Essa conclusão já havia sido percebida pelo gráfico de dispersão;
- 3) Existem quatro padrões observados dentro da população de projetos, e podem ser considerados os quatro clusters naturais de projetos de Petroleo & Gas. As variáveis que explicam os tipos de projetos são Grau de Alavancagem e Tempo, sendo o Custo uma correlação de ambas;
- 4) Existem dentro dos projetos de Petroleo & Gás quatro clusters naturais que são estatisticamente diferentes entre si;
- 5) A Análise de Discriminantes mostrou que futuros projetos pertencem a um e somente um cluster dentro as possibilidades elencadas. O nível de confiança para essa análise é de 90%, mas conclusivamente robusto.

Logo, se forem tomadas as mini conclusões e analisadas em paralelo com as hipóteses propostas como gui do modelo é possível ver como foi construído o resultado de maneira linear, e com fronteiras bem definidas, guardando o modelo para a resposta apenas do objetivo proposto.

Mas a conclusão final que deve ser feita é um pouco mais complexa. Dentro de um contexto de projetos de Petróleo & Gás, é possível dizer que poucos grupos de variáveis servem como forma de classificação de tais projetos. E tais variáveis acabam gerando alguns grupos naturais dentro da população. Logo, a escolha das variáveis para o modelo foi totalmente acertada nesse aspecto.

Mas a ressalva fica que as variáveis escolhidas não servem para gerenciar um projeto de tamanha complexidade como são estes. É interessante ver essa dualidade entre classificação e gerenciamento, pois o estudo desenvolvido não tem de fato a característica gerencial, e sim classificatória.

Colocando em linha com o objetivo do estudo, de fato foram percebidas similaridades entre os projetos, e foram formados grupos – clusters – de acordo com as similaridades. Mas em momento algum se tentou entrar no detalhamento da execução do projeto.

Como conclusão final portanto tem-se que existem quatro tipos de projetos dentro daqueles do setor de Petróleo & Gás, e essas classificações são totalmente MECE, segundo ferramental estatístico. Dessa forma é possível utilizar como forma de aprendizado e alimentar o KMS com projetos passados.

A discussão que se move agora é sobre o quanto isso pode afetar no sucesso dos novos projetos, mas essa questão deve ser tratada por estudos subsequentes.

6 Conclusão

Como pode-se concluir no último capítulo, o modelo desenvolvido conseguiu alcançar seu objetivo proposto ainda no começo do estudo. Mais do que isso, conseguiu dar alguns insights valiosos sobre projetos da indústria de Petróleo & Gás.

O objetivo declarado logo no começo do estudo foi: (I) identificar semelhanças dentro de um range de projetos, (II) coloca-los em grupos – clusters – de acordo com as semelhanças, (III) estudar as características de cada agrupamento e (IV) classificar as tipologias de projeto.

É interessante notar, antes de tudo, quão aberto é o estudo, pois em momento algum o objetivo cita que o modelo desenvolvido é único e exclusivo para projetos de Petróleo & Gás. Dessa forma, é possível dizer que o modelo pode ser adaptado para qualquer tipo de indústria.

Mais do que isso, o modelo é flexível também em número de observações, sendo que uma análise qualitativa forte foi considerada, apesar do pequeno espaço amostral que se dispunha para realização dos cálculos. De forma geral, novamente, qualquer indústria cabe dentro do modelo.

E é justamente a flexibilidade do modelo que serve de primeiro comentário final.

O modelo desenvolvido é flexível o bastante para pode ser replicado a diferentes industrias e diferentes tipologias de projetos. O objetivo é amplo e pode ser alcançado em diferentes contextos.

Entretanto, o trabalho não se baseia somente no objetivo, sendo este alcançado ou não. É preciso dizer que o modelo desenvolvido não é uma nova ferramenta para a gestão de projetos. É sim uma ferramenta de auxílio na obtenção de sucesso dentro de um projeto, assim como dito no primeiro capítulo. Um dos resultados obtidos pode ser descrito da seguinte forma: projetos passados podem gerar uma idealização de como um futuro projeto pode se comportar, e isso é um ganho de gestão.

E justamente isso é uma das belezas do modelo: um projeto planejado oferece um comportamento padrão que diz como ele terá acabado. De maneira maquiavélica é possível dizer que os fins justificam os meios. O que significa que o modelo utiliza um comportamento padrão de um determinado projeto para auxiliar no planejamento de um novo projeto. Mais do que isso, pode auxiliar na previsão de comportamento durante a execução desse novo projeto. Em outras palavras é dada a segunda peculiaridade do estudo até então realizado.

O modelo desenvolvido não é uma ferramenta gerencial. É sim uma ferramenta auxiliar para planejamento de novos projetos, segundo uma comparação com projetos passados. Auxilia portanto na obtenção de sucesso para novos projetos.

Dessa forma, é possível dizer que o estudo realizado possui um vasto potencial para auxílio dos gerentes de projetos em todo o ciclo de vida de seus projetos. Mais do que isso, é uma ferramenta poderosa para toda a empresa quando esta desejar configurar o pipeline do seu Portfólio de Projetos.

Entretanto, haja visto que o escopo do trabalho era se ater as fronteiras determinadas pelo objetivo, não houve espaço para discussão de resultados referentes ao sucesso do EVMS tampouco do GPP graças ao modelo desenvolvido. Dessa forma surge aqui uma oportunidade de futuros trabalhos que utilizem o modelo desenvolvido para verificar e validar o sucesso gerencial dos projetos, com auxílio dessa nova ferramenta.

De maneira resumida, foi criada uma ferramenta que associa o início de um projeto com o fim do mesmo, sem que haja alguma noção do que pode ocorrer durante a execução do dado projeto. As implicações para essa ferramenta tem assim um grande potencial de impacto sobre o tema de gestão de projetos. Imagine, por exemplo diminuir drasticamente o intervalo dos SPI e CPI. Assim, resumindo o último comentário sobre o estudo realizado.

Um próximo estudo deve ser conduzido a fim de validar a relação entre o modelo desenvolvido e o sucesso de projetos e portfólio. Mais do que isso, verificar o impacto em termos de previsibilidade que o modelo proporciona na gestão de projetos como um todo, e na gestão de portfólios de projetos.

Dessa forma se encerra o trabalho desenvolvido. Não finalizando entretanto o tema abordado, pois novos estudos devem ser conduzidos após este com o objetivo de complementar um ramo da gestão de projetos que tem ainda muito a melhorar. E mesmo sendo um projeto um esforço temporário e único, foi visto que é possível propor um determinado tipo de melhoramento contínuo dentro desse tema. Assim sendo, o melhoramento contínuo está novamente sendo proposto e clamado para que novas mentes possam dar continuidade a esse vasto campo da ciência.

Apêndice A

Organização

1. Definir os elementos de trabalho autorizados para o programa. O *work breakdown structure* (WBS), adaptado para o controle efetivo da gestão interna, é normalmente utilizado nesse processo.
2. Identificar a estrutura organizacional do programa incluindo a maioria das empresas subcontratadas responsáveis pelo o alcance do trabalho autorizado, além de definir os elementos organizacionais, nos quais o trabalho será planejado e controlado.
3. Proporcionar para a integração do planejamento da empresa o *scheduling*, orçamento, autorização do trabalho bem como os processos de acumulação de custos, e conforme o caso, o programa do WBS e o programa da estrutura organizacional.
4. Identificar a organização da empresa ou a função responsável pelo controle de despesas em geral (custos indiretos)
5. Proporcionar a integração do programa do WBS e o programa da estrutura organizacional de maneira que permita a medição da performance dos custos e do cronograma através de elementos de uma ou ambas estruturas, conforme necessário.

Programação, cronograma e orçamento

6. Programar o trabalho autorizado de maneira que descreva a sequencia de trabalho bem como identificar significantes tarefas interdependentes exigidas para atender os requisitos do programa.
7. Identificar produtos físicos, etapas, performance técnica dos objetivos, ou indicadores que serão utilizados para medir o progresso.
8. Estabelecer e manter a linha base de custos e tempo a nível contabilístico de controle, contra cada programa de performance a ser mensurado. Orçamento para esforços de longo prazo podem ser mantidos em níveis superiores ate um tempo apropriado para a alocação no nível contabilístico de controle. Orçamentos inicialmente estabelecidos como medida de desempenho serão baseados em objetivos internos de gestão ou no custo externo do consumidor incluindo estimativas para trabalho autorizado porem inconcluso. Nos contratos governamentais, se uma linha base for usada para fins de relatórios de medição de desempenho, uma comunicação previsa deve ser proporcionada ao cliente.
9. Estabelecer orçamentos para o trabalho autorizado com identificação de significantes

elementos de custos (mão-de-obra, matéria-prima, etc) utilizados para gestão interna e no controle de empresas subcontratadas.

10. Na medida em que é prático identificar o trabalho autorizado em pacotes de trabalho distintos, estabelecer orçamentos ou seu trabalho em termos de dólar, horas ou em unidade mensuráveis. Aonde todo o controle contabilístico não é subdividido em pacotes de trabalho, identificar esforços de longo prazo em pacotes de planejamento maiores para fins de orçamento e *scheduling*.
11. Prever que a soma de todos os orçamentos de pacotes de trabalho e dos pacotes de orçamento e planejamento dentro de uma conta controle é igual ao orçamento de controle contábil.
12. Identificar e controlar o nível de esforço da atividade através da linha base de orçamentos estabelecidas para esse proposito. Apenas este esforço, o qual é imensurável ou para aqueles com medidas impraticáveis talvez seja classificado como nível de esforço.
13. Estabelecer orçamentos gerais para cada componentes organizacional significativa da companhia para as despesas, as quais tornam-se custos indiretos. Refletir no programa de orçamentos, a um nível apropriado, os montantes que são planejados a serem alocados ao programa de custos indiretos.
14. Identificar reservas gerenciais bem como orçamentos não distribuídos.
15. Prever que a meta de custo do programa é reconciliada com a soma de orçamentos de programas internos e reservas gerenciais.

Considerações Contábeis

16. Registrar custos diretos de uma maneira consistente com um sistema de orçamento formal e controlado por livros gerais contábeis.
17. Quando a *work breakdown structure* é usada, resumir os custos diretos do controle contábil na *work breakdown structure*, sem atribuir um controle contábil para dois ou mais elementos.
18. Resumir os custos diretos do controle contábil em elementos contratuais organizacionais, sem atribuir um controle contábil para dois ou mais elementos organizacionais.
19. Registrar todos os custos indiretos, os quais serão alocado ao contrato.
20. Identificar custos unitários, custos unitários equivalentes ou *lot-costs* quando necessário.
21. Para EVMS, o sistema de contabilidade do material ira proporcionar:

- I.Acumulação de custos e atribuição de custos para controles contábeis de uma maneira consistente com os orçamentos utilizando técnicas de custos aceitáveis e reconhecidas.
- II.Medida de desempenho do custo no momento mais adequado para a categoria ou o material envolvido, mas não antes do tempo de pagamentos progressivos ou efetivo recebimento de material.
- III.Prestação de contas completas de todo o material comprado para o programa incluindo o estoque residual.

Análise e Relatórios de Gestão.

- 22. Pelo menos em uma frequência mensal, gerar as seguintes informações a nível contábil e em outros níveis se necessário com fim de controle gerencial, utilizando dados de custos ou conciliável com o sistema de qualidade:
 - I.Comparação do montante do orçamento planejado com o montante ganho através do trabalho realizado. Essa comparação proporciona uma variação no cronograma.
 - II.Comparação com do montante do orçamento ganho (aplicado se necessário) com os custos diretos do mesmo trabalho.
- 23. Identificar, pelo menos mensalmente, as diferenças significativas entre a performance planejada e a real e também o desempenho de custo, proporcionando razões para as variações em detalhes para o programa gerencial.
- 24. Identificar os custos indiretos orçados e aplicados no nível e frequência necessária pela administração para o controle efetivo, juntamente com as razões para qualquer variação significativa.
- 25. Resumir os dados e as variações adotadas através da organização do programa ou através da *work breakdown structure* para auxiliar em necessidades gerenciais ou qualquer especificação do cliente.
- 26. Implementar ações gerenciais tomadas como resultado de informações de valor agregado.
- 27. Desenvolver estimativas revisadas dos custos na conclusão com base no desempenho até a data, valores autorizados de materiais bem como estimativas de condições futuras. Comparar esta informação com a linha base de medida de desempenho no intuito de identificar variações na conclusão importante para a gestão da empresa e quaisquer requisitos do cliente incluindo declarações de necessidade de financiamento.

Revisão e Manutenção de Dados

- 28. Incorporar as mudanças autorizadas em tempo hábil, registrando os efeitos de tais

mudanças nos orçamentos e cronogramas. Em um direto esforço antes da negociação da modificação, basear tais revisões no montante estimado bem como no orçamento para o programa organizacional

29. Reconciliar atuais orçamentos a orçamentos anteriores em termos de mudanças ao trabalho autorizado bem como um replanejamento interno em detalhes através de um controle gerencial efetivo.
30. Controlar mudanças retroativas ao registros referentes ao trabalho realizado que mudaria previamente diversos relatórios de custos reais, valor agregado ou orçamentos. Ajustes devem ser realizados somente para correções de erros, ajustes de rotina contábil, efeitos do consumidor ou gestão de mudanças direcionadas, ou para aprimorar a integridade da linha base e precisão dos dados de medição de desempenho.
31. Prevenir revisões do programa de budgets exceto por mudanças autorizadas.
32. Mudanças de documentos para a performance de medida da linha base.

Apêndice B

Considere o vetor de media de amostras e o vetor da covariância dados por:

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} \quad \mathbf{S}_n = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{21} & S_{22} & \dots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \dots & S_{pp} \end{bmatrix}$$

Considere o vetor $\mathbf{1}$ ($n \times 1$), sendo que todos os dados incluídos neste vetor são dados pelo valor 1 e \mathbf{I} como a matriz identidade ($n \times n$). Para a Prova #1:

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \sum_{j=1}^n x_{j1} \\ \frac{1}{n} \sum_{j=1}^n x_{j2} \\ \vdots \\ \frac{1}{n} \sum_{j=1}^n x_{jp} \end{bmatrix} = \frac{1}{n} \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{1}{n} \mathbf{X}' \mathbf{1}$$

Para a prova #2, a qual é representada pela equação para obter o vetor covariância, tem-se a seguinte expressão:

$$\mathbf{1} \bar{\mathbf{x}}' = \frac{1}{n} \mathbf{1} \mathbf{1}' \mathbf{X} = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \dots & \bar{x}_p \\ \bar{x}_1 & \bar{x}_2 & \dots & \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_1 & \bar{x}_2 & \dots & \bar{x}_p \end{bmatrix}$$

$$\mathbf{X} - \frac{1}{n} \mathbf{1} \mathbf{1}' \mathbf{X} = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \dots & x_{1p} - \bar{x}_p \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \dots & x_{2p} - \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \dots & x_{np} - \bar{x}_p \end{bmatrix}$$

Desta forma,

$$\begin{aligned} \mathbf{S}_n &= \frac{1}{n} \begin{bmatrix} x_{11} - \bar{x}_1 & x_{21} - \bar{x}_1 & \dots & x_{n1} - \bar{x}_1 \\ x_{12} - \bar{x}_2 & x_{22} - \bar{x}_2 & \dots & x_{n2} - \bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} - \bar{x}_p & x_{2p} - \bar{x}_p & \dots & x_{np} - \bar{x}_p \end{bmatrix} \\ &\quad \times \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \dots & x_{1p} - \bar{x}_p \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \dots & x_{2p} - \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \dots & x_{np} - \bar{x}_p \end{bmatrix} \\ &= \frac{1}{n} \left(\mathbf{X} - \frac{1}{n} \mathbf{1} \mathbf{1}' \mathbf{X} \right)' \left(\mathbf{X} - \frac{1}{n} \mathbf{1} \mathbf{1}' \mathbf{X} \right) = \frac{1}{n} \mathbf{X}' \left(\mathbf{I} - \frac{1}{n} \mathbf{1} \mathbf{1}' \right) \mathbf{X} \end{aligned}$$

Apêndice C

Suponha um conjunto de dados com n observações e p -dimensões. Esse conjunto de dados pode ser expresso por uma matriz de n linhas e p colunas.

$$\mathbf{X} = \begin{array}{c|cccccc} & \text{Variable 1} & \text{Variable 2} & \dots & \text{Variable k} & \dots & \text{Variable p} \\ \hline \text{Observation 1} & x_{11} & x_{12} & \dots & x_{1k} & \dots & x_{1p} \\ \text{Observation 2} & x_{21} & x_{22} & \dots & x_{2k} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \text{Observation j} & x_{j1} & x_{j2} & \dots & x_{jk} & \dots & x_{jp} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \text{Observation n} & x_{n1} & x_{n2} & \dots & x_{nk} & \dots & x_{np} \end{array}$$

Cada vetor representa uma variável, ou seja, cada vetor representa uma coluna e possui n linhas. Para padronizar as colunas de modo que estas sejam comparáveis e passíveis de interação entre si, e consequentemente, resultem em um cálculo de distâncias apropriadas pode-se utilizar dois algoritmos diferentes, ambos com uma abordagem linear.

Algoritmo 1 – Zero Mínimo

1. Procurar pelo valor máximo de cada variável.
2. Implicar o mínimo valor igual a zero em cada uma das variáveis.
3. Para toda variável, dividir cada observação pelo valor máximo.

O resultado será a Distância Euclidiana, representada a seguir:

$$d_{ij} = \sqrt{\frac{(x_{i1} - x_{j1})^2}{x_{1max}^2} + \frac{(x_{i2} - x_{j2})^2}{x_{2max}^2} + \dots + \frac{(x_{ip} - x_{jp})^2}{x_{pmax}^2}}$$

Algoritmo 2 – Max Min Valor

1. Procurar pelo valor máximo e mínimo de cada uma das variáveis.
2. Para cada variável, subtrair de cada observação o valor mínimo.
3. Para toda variável, dividir cada observação pela diferença entre o valor máximo e o valor mínimo.

O resultado será a Distância Euclidiana, representada a seguir:

$$d_{ij} = \sqrt{\frac{(x_{i1} - x_{j1})^2}{(x_{1max} - x_{1min})^2} + \frac{(x_{i2} - x_{j2})^2}{(x_{2max} - x_{2min})^2} + \dots + \frac{(x_{ip} - x_{jp})^2}{(x_{pmax} - x_{pmin})^2}}$$

Diferenças entre o Algoritmo 1 e o Algoritmo 2

Os dois algoritmos, utilizados para padronizar um conjunto de dados segue duas diferentes abordagens mesmo que lineares. A Figura 7.1 ilustra como cada algoritmo transforma as observações de uma variável com o proposito de proporcionar a padronização do conjunto de dados.

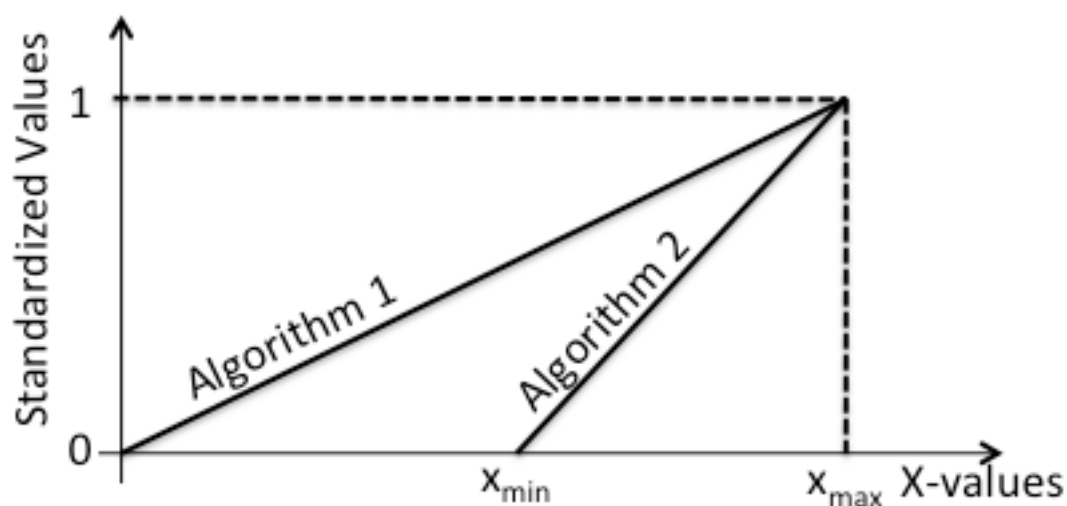


Figura 7.1 – Algoritmo de estandardização de variáveis.

Além disso, pode ser observado que o Algoritmo 2, proporciona um maior peso para grandes distancias entre variáveis. Devido a tal fato, a distancia no caso do algoritmo 2 pode ser maior que um. Esse resultado não ocorre caso seja utilizado o Algoritmo 1.

Finalmente, utilizando o Algoritmo 2, a Distância Minkowski pode resultar redundante, uma vez que este método é utilizado quando busca-se um maior peso para maiores distâncias.

Apêndice D

A Análise Multivariada da Variância (MANOVA) é um método utilizado para comparar as diversas médias multivariadas da população. O principal objetivo da MANOVA é analisar se os vetores da media da população são diferentes ou não, e caso sejam, estabelecer quais componentes os diferenciam de maneira significativa.

Existem algumas premissas relacionadas ao dados aplicados ao MANOVA, os quais são:

- I. As amostras aleatórias de diferentes populações são independentes.
- II. Todas as populações possuem em comum a covariância matriz Σ .
- III. Cada população é multivariada normal.

A terceira premissa pode ser relaxada utilizando o Teorema do Limite Central.

O algoritmo é muito simples e é computado através do MS Excel[®], de acordo com Johnson & Wichern (2007).

A Tabela 7.1 ilustra o resumo do algoritmo MANOVA.

Tabela 7.1 – MANOVA expressa em formato de tabela expositiva de resultados.

Source of Variation	Matrix of sum of squares and cross products (SSP)	Degrees of freedom (d.f.)
Treatment	$\mathbf{B} = \sum_{\ell=1}^g n_{\ell} (\bar{x}_{\ell} - \bar{x})(\bar{x}_{\ell} - \bar{x})'$	$g - 1$
Residual (error)	$\mathbf{W} = \sum_{\ell=1}^g \sum_{j=1}^{n_{\ell}} (\bar{x}_{\ell j} - \bar{x}_{\ell})(\bar{x}_{\ell j} - \bar{x}_{\ell})'$	$\sum_{\ell=1}^g n_{\ell} - g$
Total (corrected for the mean)	$\mathbf{B} + \mathbf{W} = \sum_{\ell=1}^g \sum_{j=1}^{n_{\ell}} (\bar{x}_{\ell j} - \bar{x})(\bar{x}_{\ell j} - \bar{x})'$	$\sum_{\ell=1}^g n_{\ell} - 1$

Source: (Johnson & Wichern, 2007)

O principal objetivo desse algoritmo é testar a hipótese na qual todos os vetores de media são os mesmos:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_g$$

Para testar tal hipótese, será utilizado o parâmetro Wilk's Lambda:

$$\Lambda^* = \frac{|W|}{|B + W|}$$

Com Wilk's Lambda calcula-se o índice de Barlett:

$$-\left(n - 1 - \frac{(p + g)}{2}\right) \ln \Lambda^*$$

Finalmente, o H_0 é rejeitado caso seja observada a seguinte ocorrência:

$$-\left(n - 1 - \frac{(p + g)}{2}\right) \ln \Lambda^* > \chi_{p(g-1)}^2(\alpha)$$

Apêndice E

Da P1 até a P8 é possível observar todos os projetos submarinos, e do P9 até o P16 todos os projetos *offshore*. Além disso, PT1 é um projeto submarino e PT2 é um projeto *offshore*.

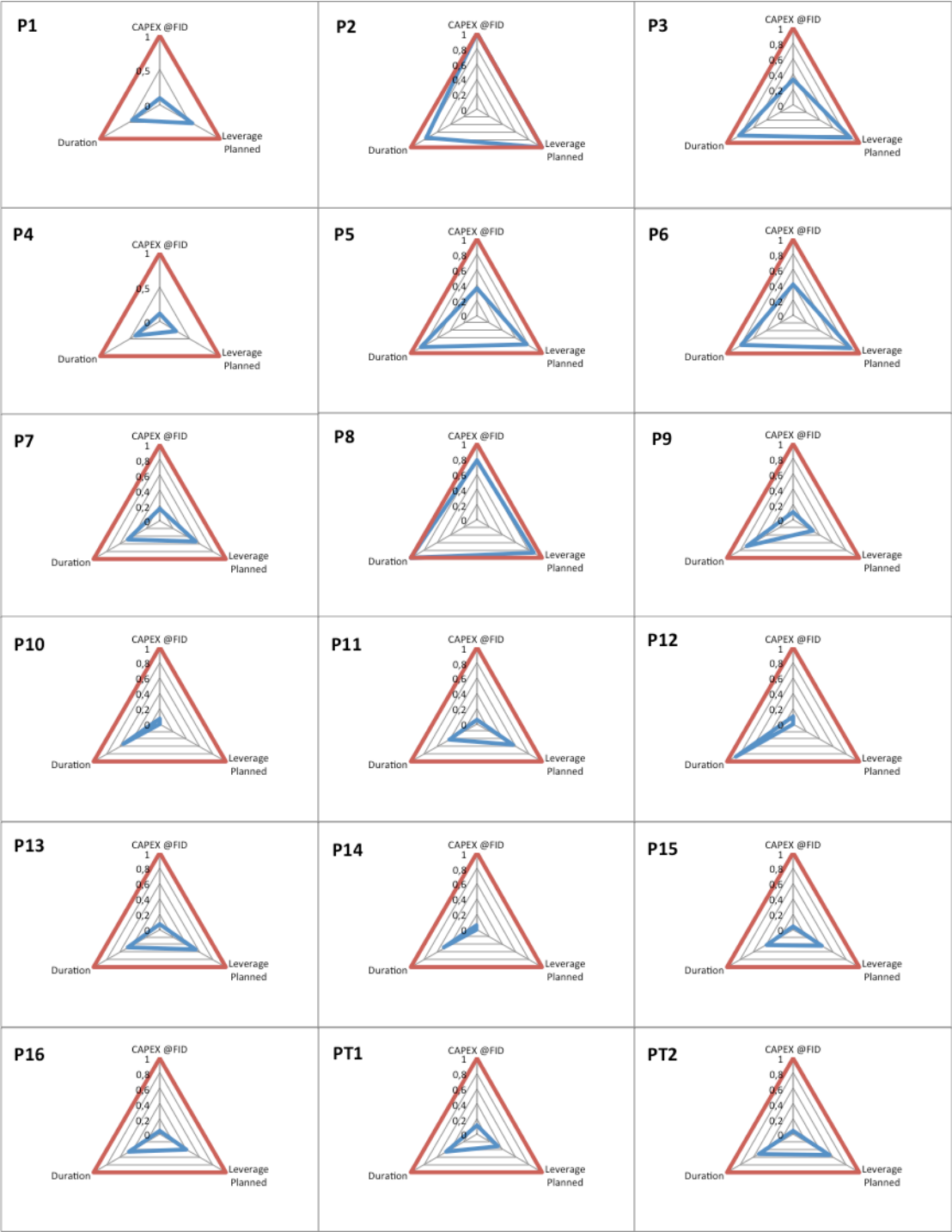


Figura 7.2 – Representação de Estrela para todos os projetos planejados.