

UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE
PRODUÇÃO

Método Sistemático para Seleção de Empresas em Composição de
Portfólio e sua Otimização com Incorporação de Indicadores
Financeiros

Pedro José Papandréa

Itajubá
Abril 2018

UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE
PRODUÇÃO

Pedro José Papandréa

Método Sistemático para Seleção de Empresas em Composição de
Portfólio e sua Otimização com Incorporação de Indicadores
Financeiros

Tese submetida ao Programa de Pós-Graduação em Engenharia de Produção como parte dos requisitos para a obtenção do título de Doutor em Ciências em Engenharia de Produção.

Orientador: Prof. Dr. Anderson Paulo de Paiva

Coorientador: Prof. Dr. Rafael Coradi Leme

Itajubá

Abril 2018

UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE
PRODUÇÃO

Pedro José Papandréa

Método Sistemático para Seleção de Empresas em Composição de
Portfólio e sua Otimização com Incorporação de Indicadores
Financeiros

Tese submetida para banca examinadora em
7 de Março de 2018, como parte de requisito
ao título de Doutor em Ciências em Enge-
nharia de Produção.

Banca Examinadora:

Prof. Dr. Anderson Paulo de Paiva (Orien-
tador)

Prof. Dr. Rafael Coradi Leme (Co-
orientador)

Prof. Dr. José Henrique de Freitas Gomes

Prof. Dr. Luiz Gustavo Dias Lopes

Prof. Dr. Paulo Henrique Campos

Prof. Dr. Ronã Rinston Amaury Mendes

Itajubá
Abril 2018

Especialmente para minha mãe Rita Lopes Papandréa.

AGRADECIMENTOS

Agradeço primeiramente a Deus por ter me indicado o caminho correto a seguir e por ter me aproximado das pessoas que, sem as quais, eu não conseguiria ter prosseguido. Agradeço em especial aos professores, amigos e orientadores Dr. Anderson Paulo de Paiva, Dr. Rafael Coradi Leme, Dr. Xueping Li. Estes que sempre acreditaram no meu trabalho e valorizaram todo meu esforço e dedicação. À minha esposa, Andrezza pela paciência na minha ausência e nos meus momentos de *stress* e ansiedade. À minha mãe Rita por sempre me apoiar incondicionalmente, aos meus irmãos que estão sempre comigo e a minha irmã Rosana, meu primo Antônio Henrique, meu tio Antônio que nos deixaram precocemente, mas que continuam a olhar por mim. À Universidade do Tennessee, Knoxville, por toda a estrutura que me disponibilizaram. Aos amigos que lá deixei. Aos estimados membros da banca, Dr. Ronã Rinston Amaury Mendes, Dr. Paulo Henrique Campos, Dr. Luiz Gustavo Dias Lopes, Dr. José Henrique de Freitas Gomes e ao Dr. Tarcisio Gonçalves de Brito, membro na qualificação. Em especial também ao amigo João Éderson Corrêa cuja ajuda foi crucial. A todos os professores do programa de pós graduação que me auxiliaram na busca de conhecimentos, à secretaria do IEPG. À todos os amigos e colegas do GEPE. E finalmente ao MEC - Fundação coordenação de aperfeiçoamento de pessoal de nível superior (CAPES) que financiou todas as minhas pesquisas em especial pelo projeto 12875-13-9. A palavra Obrigado não é capaz de expressar de forma verdadeira o sentimento de gratidão, amizade e irmandade que tenho pelos meus queridos professores e amigos do Gepe Unifei. Muito Obrigado!

“Nossa maior fraqueza está em desistir. O caminho mais certo de vencer é tentar mais uma vez.”
(Thomas Edison)

RESUMO

Para preencher a lacuna que existe na escolha de empresas para compor portfólios e inclusão de variáveis externas em sua otimização, e sendo também motivado pela crescente visibilidade das empresas verdes no mercado de ações, propõe-se a criação, desenvolvimento e análise de um novo método para escolha de empresas, geração de indicadores e inclusão destes indicadores como variáveis de processo em arranjos de mistura com otimização lexicográfica. A aplicação e validação do método se dá em empresas chamadas de Verdes (*Green Companies*), que são aquelas que possuem ações relevantes em prol do meio-ambiente, como por exemplo, sustentabilidade na produção, uso consciente da água e recursos naturais, sustentabilidade energética. As análises dos portfólios financeiros dessas empresas verdes é feita usando-se experimentos com arranjo de misturas, análise de componentes principais e o conceito de *clusters*. Aliado a isso, os indicadores financeiros, extraídos dos balanços patrimoniais oficiais das empresas candidatas foram analisados e formam as variáveis de processo dos arranjos de mistura. Esta é também uma contribuição importante, já que os modelos tradicionais de portfólio não consideram esses indicadores e não os integra no modelo. As empresas candidatas foram escolhidas a partir da classificação das empresas mais verdes da revista *Newsweek* de 2014. Estabelecemos um portfólio de *clusters* usando o conceito de análise fatorial e análise de componentes principais (PCA). A análise de componentes principais foi aplicada para selecionar e re-classificar as empresas candidatas a partir do *ranking* original da *Newsweek*. O agrupamento dessas empresas foi feito pelo método de Ward com base em indicadores ou taxas de risco financeiro. O PCA foi novamente usado para o cálculo do indicador geral multivariado (MGI) o qual é base para o agrupamento das empresas similares por *clusters*. Essas taxas de risco financeiros foram calculados com base nos relatórios financeiros anuais (Relatório Anual de acordo com a Seção 13 ou 15(d) do Ato regulamentador de 1934 (*Securities Exchange Act, 1934*) para determinar a pontuação dos componentes principais relacionados aos indicadores financeiros. A mineração dos dados foi utilizada para extrair os valores necessários para calcular os índices financeiros de cada variável ou empresa. Os resultados computacionais têm proporcionado novas perspectivas usando a combinação de análise de componentes principais, experimentos com arranjos de mistura, o método de *cluster*, método de otimização lexicográfico e a base dos relatórios financeiros anuais. Tudo isso aplicado no campo financeiro com empresas verdes como alvos da análise. Esta proposta é otimizada para comparação dos resultados com a carteira de modelo global. Este último, tradicional, não considera indicadores financeiros nem variáveis de processo. Ele trata de um portfólio geral contendo todas as empresas que foram selecionadas na primeira etapa do trabalho e otimização com mistura pelo modelo de Markowitz.

Palavras-chaves: Seleção de empresas; Portfólio; Análise de Componentes Principais; Experimentos de Mistura; Indicadores Financeiros; Otimização Lexicográfica.

ABSTRACT

In order to fill the gap that exists in choosing companies to compose portfolios and inclusion of external variables in their optimization and also being motivated by the growing visibility of green companies in the stock market, it is proposed to create, develop and analyze a new method for choice of companies, generation of indicators and inclusion of these indicators as process variables in mix arrangements with lexicographic optimization. The application and validation of the method of Green Companies, which are those that play an important role in the environment, such as sustainability in production, conscious use of water and natural resources, energy sustainability. The data analyzes are organized according to the principal components analysis and the concept of clusters. Allied to this, the financial indicators, extracted from the balance sheets offices of companies candidates were treated and forming the variables of process of mixing arrangements. This is also an important contribution, since programming rights are not considered indicators and are not integrated into the model. Candidates were selected to apply for the greenest companies in Newsweek in 2014. Establish a portfolio of clusters using the concept of factor analysis and component analysis (PCA). A component analysis was used to select and re-rank as candidate companies from the original Newsweek ranking. The grouping of companies was created by the Ward method based on indicators or financial risk rates. The PCA was again used for the general multivariate index (MGI) which is the basis for grouping similar companies by clusters. Annual risk averages were calculated based on the 1934 annual report of *The Securities Exchange Act, 1934* to determine the evolution of financial indicators. The data mining was used to extract the values for the financial results of each variable, the results of the generation of new options using a combination of results analysis. Lexicographic optimization method and the annual financial database has been applied also. All those methods was applied in the financial field with green companies as objects of the analysis. This proposal is optimized for the comparison of results with a global model portfolio. The latter, traditional, is not considered financial vector nor process variable. It deals with a general portfolio of all the companies that were selected in the first stage of the work and optimization with the mixture model using the Markowitz theory.

Key-words: Companies Selection; Portfolio; Principal Component Analysis; Mixture Design of Experiment; Financial Ratios; Lexicographic Optimization.

LISTA DE ILUSTRAÇÕES

Figura 2.1 –Mapa das ações de recuperação ambiental dos E.U.A.	10
Figura 2.2 –Modelo de motivação ecológica empresarial. Fonte: (LYON; SHIMSHACK, 2012).	10
Figura 2.3 –A barra de preços das ações	11
Figura 2.4 –Risco em função do retorno	12
Figura 3.1 –Observador passivo investigando um estado pré-existente	15
Figura 3.2 –Experimentador muda o estado e investiga os resultados obtidos	15
Figura 3.3 –Entradas e saídas do processo. Fonte: Montgomery <i>et al.</i> (1997)	18
Figura 3.4 –Descrição esquemática de experimento cruzado	21
Figura 3.5 –Arranjo de misturas e superfície de resposta para misturas	23
Figura 3.6 –Simplex-Centróide	24
Figura 3.7 –Simplex-Lattice 4,n para polinômio de grau 3	24
Figura 3.8 –Arranjo <i>simplex-lattice</i>	25
Figura 3.9 –Esquema experimental para um arranjo de misturas do tipo <i>Extreme-Vertice</i>	28
Figura 3.10 –Arranjo cruzado de misturas MoM	33
Figura 3.11 – <i>Simplex-centroid</i> em cada ponto do experimento Fatorial Completo 2^3	33
Figura 3.12 –Experimento fatorial 2^3 em cada ponto do <i>Simplex-centroid</i>	34
Figura 3.13 –Maximizar y	45
Figura 3.14 –Minimizar y	46
Figura 3.15 –Função multiobjetivo	46
Figura 4.1 –Retorno em função do risco	49
Figura 4.2 –Efeito da correlação no risco	52
Figura 4.3 –Fronteira Eficiente e Pontos Ótimos	52
Figura 5.1 –Macrofluxo da metodologia proposta	54
Figura 5.2 –Cotações e variação mensal AAPL	55
Figura 5.3 –Capa da revista <i>Newsweek</i> 2014	59
Figura 5.4 –Autocorrelação da empresa AAPL	60
Figura 5.5 –Fluxo do PCA aplicado a <i>Newsweek</i>	65
Figura 5.6 –Logotipo da classificação da <i>Newsweek</i>	66
Figura 5.7 –Agrupamentos gerados pelo PCA incluindo reputação	68
Figura 5.8 –Agrupamento com gráfico de pontos incluindo reputação	68
Figura 5.9 –Agrupamento dos fatores principais sem reputação <i>Newsweek</i>	68
Figura 5.10 –Fluxo para baixa e preparação dos dados	70

Figura 5.11 –Cotações históricas	71
Figura 5.12 –Fluxo para baixa e análise dos balanços patrimoniais	71
Figura 5.13 –Página título do formulário 10K, balanço patrimonial Apple	72
Figura 5.14 –Exemplo de dado retirado para cálculo do indicador financeiro	72
Figura 5.15 –Detalhe de dado retirado para cálculo de indicador financeiro	73
Figura 5.16 –Gráfico multivariado dos indicadores financeiros	75
Figura 5.17 –Fluxo do risco, clusters e variáveis de processo	75
Figura 5.18 –Dendograma, Divisão considerada dos Clusters pelos Indicadores cal- culados	76
Figura 5.19 –Dendograma, Divisão dos Clusters por PCA <i>Newsweek</i>	76
Figura 5.20 –Dendograma, Divisão dos Cluster por Indicadores	76
Figura 5.21 –Agrupamento dos fatores principais dos Indicadores	77
Figura 5.22 –Dendograma, similaridades entre indicadores	78
Figura 5.23 –Cluster em cada PCA	78
 Figura 6.1 –Rampa de otimização para o Cluster 1(-1)	 87
Figura 6.2 –Rampa de otimização para o Cluster 1(+1)	87
Figura 6.3 –Rampa de otimização para o Cluster 2(+1)	87
Figura 6.4 –Rampa de otimização para o Cluster 3(-1)	87
Figura 6.5 –Rampa de otimização para o Cluster 3(+1)	88
Figura 6.6 –Solução de otimização para o Cluster 1(-1) 2D	89
Figura 6.7 –Solução de otimização para o Cluster 1(-1) 3D	89
Figura 6.8 –Solução de otimização para o Cluster 1(+1) 2D	89
Figura 6.9 –Solução de otimização para o Cluster 1(+1) 3D	90
Figura 6.10 –Solução de otimização para o Cluster 2(+1)	91
Figura 6.11 –Solução de otimização para o Cluster 3(-1) 2D	91
Figura 6.12 –Solução de otimização para o Cluster 3(-1) 3D	92
Figura 6.13 –Solução de otimização para o Cluster 3(+1) 2D	92
Figura 6.14 –Solução de otimização para o Cluster 3(+1) 3D	93
Figura 6.15 –Arranjo de misturas com MGI como variável de processo	94
Figura 6.16 –Rampa de otimização do arranjo combinado	96
Figura 6.17 –Solução de otimização para os arranjos combinados 2D	96
Figura 6.18 –Solução de otimização para os arranjos combinados 3D	97
Figura 6.19 –Resultado rampa de otimização do arranjo combinado	97
Figura 6.20 –Arranjo <i>simplex</i> MoM com variáveis de processo (-1) (+1)	98
Figura 6.21 –Rampa de otimização CM-DOE	103
Figura 6.22 –Solução de otimização para o arranjo CM-DOE 2D	103
Figura 6.23 –Solução de otimização para o arranjo CM-DOE 3D	104
Figura 6.24 –Resultado rampa de otimização do arranjo CM-DOE	106
Figura 6.25 –Rampa de otimização Markowitz	108

Figura 6.26 – Solução de otimização para o modelo Markowitz 2D	109
Figura 6.27 – Solução de otimização para o modelo Markowitz 3D	109
Figura 6.28 – Resultado rampa de otimização do modelo Markowitz	111
Figura 6.29 – Comparação entre métodos	117
Figura 6.30 – Comparação entre métodos	117
Figura A.1 – Análise de RP do Cluster 1(-1) Contorno	132
Figura A.2 – Análise de RP do Cluster 1(-1) 3D	132
Figura A.3 – Análise de RP do Cluster 1(-1) Normalidade	132
Figura A.4 – Análise de RP do Cluster 1(-1) Cox	133
Figura A.5 – Análise de SP do Cluster 1(-1) Contorno	134
Figura A.6 – Análise de SP do Cluster 1(-1) 3D	134
Figura A.7 – Análise de SP do Cluster 1(-1) Normalidade	135
Figura A.8 – Análise de SP do Cluster 1(-1)Cox	135
Figura A.9 – Análise de RP do Cluster 1(+1) Contorno	136
Figura A.10 – Análise de RP do Cluster 1(+1) 3D	136
Figura A.11 – Análise de RP do Cluster 1(+1) Normalidade	136
Figura A.12 – Análise de RP do Cluster 1(+1) Cox	137
Figura A.13 – Análise de SP do Cluster 1(+1) Contorno	138
Figura A.14 – Análise de SP do Cluster 1(+1) 3D	138
Figura A.15 – Análise de SP do Cluster 1(+1) Normalidade	138
Figura A.16 – Análise de SP do Cluster 1(+1) Cox	139
Figura A.17 – Análise de RP do Cluster 2(+1) Contorno	139
Figura A.18 – Análise de RP do Cluster 2(+1) Normalidade	140
Figura A.19 – Análises de SP Cluster2(+1) Contorno	140
Figura A.20 – Análises de SP Cluster2(+1) Normalidade	141
Figura A.21 – Análise de RP Cluster 3(-1) Contorno	142
Figura A.22 – Análise de RP Cluster 3(-1) 3D	142
Figura A.23 – Análise de RP Cluster 3(-1) Normalidade	143
Figura A.24 – Análise de RP Cluster 3(-1) Cox	143
Figura A.25 – Análise de SP Cluster 3(-1) Contorno	144
Figura A.26 – Análise de SP Cluster 3(-1) 3D	144
Figura A.27 – Análise de SP Cluster 3(-1) Normalizado	145
Figura A.28 – Análise de SP Cluster 3(-1) Cox	145
Figura A.29 – Análise de RP Cluster 3(+1) Contorno	146
Figura A.30 – Análise de RP Cluster 3(+1) 3D	146
Figura A.31 – Análise de RP Cluster 3(+1) Normalizado	146
Figura A.32 – Análise de RP Cluster 3(+1) Cox	147
Figura A.33 – Análise de SP Cluster 3(+1) Contorno	147
Figura A.34 – Análise de SP Cluster 3(+1) 3D	148

Figura A.35 Análise de SP Cluster 3(+1) Normalizado	148
Figura A.36 Análise de SP Cluster 3(+1) Cox	148
Figura B.1 –Análise de RP do arranjo CMO Contorno	150
Figura B.2 –Análise de RP do arranjo CMO 3D	150
Figura B.3 –Análise de RP do arranjo CMO Normalidade	150
Figura B.4 –Análise de RP do arranjo CMO Cox	151
Figura B.5 –Análise de SP do arranjo combinado Contorno	151
Figura B.6 –Análise de SP do arranjo combinado 3D	152
Figura B.7 –Análise de SP do arranjo combinado Normalidade	152
Figura B.8 –Análise de SP do arranjo combinado Cox	152
Figura C.1 –Análise de RP CM-DOE Contorno MGI(-1)	153
Figura C.2 –Análises de RP CM-DOE 3D MGI(-1)	153
Figura C.3 –Análises de RP CM-DOE Cox MGI(-1)	155
Figura C.4 –Análises de RP CM-DOE Contorno MGI(+1)	155
Figura C.5 –Análises de RP CM-DOE 3D MGI(+1)	155
Figura C.6 –Análises de RP CM-DOE Cox MGI(+1)	156
Figura C.7 –Análises de RP CM-DOE Normalizado	156
Figura C.8 –Análise de SP CM-DOE Contorno MGI(-1)	156
Figura C.9 –Análises de SP CM-DOE 3D MGI(-1)	158
Figura C.10 Análises de SP CM-DOE Cox MGI(-1)	158
Figura C.11 Análises de SP CM-DOE Contorno MGI(+1)	158
Figura C.12 Análises de SP CM-DOE 3D MGI(+1)	159
Figura C.13 Análises de SP CM-DOE Cox MGI(+1)	159
Figura C.14 Análises de SP CM-DOE Normalizado	159
Figura D.1 –Análise de RP Contorno do modelo de mistura Markowitz	160
Figura D.2 –Análise de RP 3D do modelo de mistura Markowitz	161
Figura D.3 –Análise de RP Normalizado do modelo de mistura Markowitz	161
Figura D.4 –Análise de RP Cox do modelo de mistura Markowitz	161
Figura D.5 –Análise de SP Contorno do modelo de mistura Markowitz	162
Figura D.6 –Análise de SP 3D do modelo de mistura Markowitz	163
Figura D.7 –Análise de SP Normalizado do modelo de mistura Markowitz	163
Figura D.8 –Análise de SP Cox do modelo de mistura Markowitz	163

LISTA DE TABELAS

Tabela 3.1 –Características básicas dos pacientes	20
Tabela 3.2 –Esquema experimental para arranjo de misturas do tipo <i>Extreme-Vertice</i>	29
Tabela 3.3 –Proporções MDE 2 componentes, resolução de 0,2	31
Tabela 3.4 –Comparação simples CCD x MoM	34
Tabela 4.1 –Exemplo de dois Ativos	51
Tabela 4.2 –Efeito da correlação no risco	51
Tabela 5.1 –Histórico de 24 meses	64
Tabela 5.2 –Quarenta e quatro empresas melhores classificadas Original	67
Tabela 5.3 –Eigen analyse da matriz de correlação PCA	68
Tabela 5.4 –Quarenta e quatro empresas melhores classificadas pelo PCA	69
Tabela 5.5 –Análise de débitos em milhões de dólares por empresa	73
Tabela 5.6 –Análise dos indicadores de taxas de juros	74
Tabela 5.7 –Divisão dos clusters e valores de PCA	78
Tabela 5.8 –Agrupamento das empresas por <i>cluster</i> e variáveis de processo (-1) e (+1)	79
Tabela 6.1 –Experimento <i>Cluster</i> 1 (-1)	82
Tabela 6.2 –Experimento <i>Cluster</i> 1 (+1)	82
Tabela 6.3 –Experimento <i>Cluster</i> 2 (+1)	83
Tabela 6.4 –Experimento <i>Cluster</i> 3 (-1)	83
Tabela 6.5 –Experimento <i>Cluster</i> 3 (+1)	84
Tabela 6.6 –Arranjo para otimização dos <i>clusters</i> combinados com variáveis de processo	95
Tabela 6.7 –Empresas seleccionadas para o CM-DOE	99
Tabela 6.8 –Arranjo de misturas de misturas MoM cruzado CM-DOE 1 de 2	100
Tabela 6.9 –Arranjo de misturas de misturas MoM cruzado CM-DOE 2 de 2	101
Tabela 6.10 –Arranjo de misturas para portfólio de modelo Markowitz	107
Tabela 6.11 –Comparação dos métodos (Tabela 1 de 3)	114
Tabela 6.12 –Comparação dos métodos (Tabela 2 de 3)	115
Tabela 6.13 –Comparação dos métodos (Tabela 3 de 3)	116

LISTA DE QUADROS

Quadro 2.1 – Indicadores	14
Quadro 3.1 – Sequência pré-definida do tratamento	20
Quadro 3.2 – Arranjos de misturas e modelos polinomiais associados	25
Quadro 3.3 – ANOVA Exemplo	35
Quadro 5.1 – Logotipo e códigos das empresas selecionadas	70
Quadro 5.2 – Divisão das empresas por <i>cluster</i>	75
Quadro 6.1 – Funções de otimização para os <i>clusters</i> individuais	84
Quadro 6.2 – Solução de otimização para o Cluster 1(-1)	88
Quadro 6.3 – Solução de otimização para o Cluster 1(+1)	90
Quadro 6.4 – Solução de otimização para o Cluster 2(+1)	90
Quadro 6.5 – Solução de otimização para o Cluster 3(-1)	91
Quadro 6.6 – Solução de otimização para o Cluster 3(+1)	92
Quadro 6.7 – Resultados por cluster	93
Quadro 6.8 – Funções de otimização para os <i>clusters</i> combinados com variáveis de processo	95
Quadro 6.9 – Resultados Combinado	97
Quadro 6.10–Resumo das empresas selecionadas para o CM-DOE	99
Quadro 6.11–Função de otimização CM-DOE MGI(-1)	102
Quadro 6.12–Função de otimização CM-DOE MGI(+1)	102
Quadro 6.13–Função de otimização CM-DOE RP	103
Quadro 6.14–Função de otimização CM-DOE SP	104
Quadro 6.15–Resultados CM-DOE	105
Quadro 6.16–Função de otimização modelo Markowitz	106
Quadro 6.17–Resultados modelo Markowitz (quadro 1 de 2)	109
Quadro 6.18–Resultados modelo Markowitz (quadro 2 de 2)	110
Quadro A.1–ANOVA de RP Cluster 1 (-1)	131
Quadro A.2–ANOVA de SP Cluster 1(-1)	133
Quadro A.3–ANOVA de RP Cluster 1(+1)	134
Quadro A.4–ANOVA de SP Cluster 1(+1)	137
Quadro A.5–ANOVA de RP Cluster 2(+1)	139
Quadro A.6–ANOVA de SP Cluster 2(+1)	140
Quadro A.7–ANOVA de RP Cluster 3(-1)	141
Quadro A.8–ANOVA de SP Cluster 3(-1)	143
Quadro A.9–ANOVA de RP Cluster 3(+1)	145
Quadro A.10–ANOVA de SP Cluster 3(+1)	147
Quadro B.1–ANOVA de RP do arranjo combinado	149

Quadro B.2–ANOVA de SP do arranjo combinado	151
Quadro C.1–ANOVA de RP CM-DOE	154
Quadro C.2–ANOVA de SP CM-DOE	157
Quadro D.1–ANOVA de RP modelo Markowitz	160
Quadro D.2–ANOVA de SP modelo Markowitz	162

LISTA DE ABREVIATURAS E SIGLAS

PCA	Análise de componentes principais ou <i>Principal component analysis</i>
ANOVA	Análise de variância
CM-DOE	Arranjo cruzado de mistura ou <i>Crossed mixture design of experiment</i>
CCD	Arranjo Composto Central ou <i>Box-Wilson Central Composite Design</i>
CAPM	<i>Capital asset pricing model</i>
Corr	Correlação
PC	Componente principal ou <i>Principal component</i>
EUA	Estados Unidos da América
MGI	Indicador geral multivariado ou <i>Multivariate general indicator</i>
MVP	Média variância do portfólio
Mix	Mistura
MoM	Misturas de misturas ou <i>Mixture of mixture</i>
NASDAQ	<i>National association of securities dealers automatic quotation</i>
DOE	Projeto de experimentos ou <i>Design of experiments</i>
MDE	Projeto de experimentos de mistura ou <i>Mixture design of experiments</i>
RP	Retorno do portfólio
SP	Risco do portfólio
SSE	Soma dos erros quadráticos ou <i>Sum of squared errors</i>
S.A.	Sociedade anônima
Lag	Tamanho do passo
Var	Variância

LISTA DE SÍMBOLOS

δ	Beta no modelo de mistura cruzado
γ	<i>Cluster</i>
β	Constante na função do experimento
r_{xy}	Coefficiente de correlação
\neq	Diferente, não igual a
S	Desvio padrão amostral
σ	Desvio padrão populacional
$f(x)$	Função de x
η	Função polinomial
m	Grau do polinômio da mixtura
ν	Grau de resolução da mistura
\ln	Logaritmo natural ou neperiano
\bar{x}	Média amostral
μ	Média populacional
\hat{Y}	Modelo de mistura generalizado
q	Número de componentes do experimento
N	Número de ensaios diferentes
v	Número de réplicas
w	Peso ou percentual
Π	Produtório
Σ	Somatório
E	Valor esperado para a função
Ψ	Valor máximo da mistura
Var	Variância

SUMÁRIO

Agradecimentos	v
Resumo	vii
Abstract	viii
Lista de Ilustrações	viii
Lista de Tabelas	xii
Lista de Quadros	xiii
Lista de Abreviaturas e Siglas	xvi
Lista de Símbolos	xvii
1 Introdução	1
1.1 Considerações Iniciais	1
1.2 Justificativa e Relevância	3
1.3 Objetivos	4
1.3.1 Objetivo Geral	4
1.3.2 Objetivos Específicos	4
1.4 Delimitação do Trabalho	5
1.5 Caráter Inédito	6
1.6 Metodologia	6
1.7 Organização do Texto	7
2 Empresas Verdes e Mercado Financeiro	8
2.1 Considerações Iniciais	8
2.2 Empresas verdes	8
2.2.1 Bolsa de valores	10
2.2.1.1 Ações	12
2.2.2 Indicadores financeiros	13
2.3 Considerações Finais Sobre o Capítulo	14
3 Métodos de Experimentação e Agrupamento	15
3.1 Considerações Iniciais	15
3.2 Delineamento de Experimentos	15
3.2.1 Processo de delineamento do experimento	18

3.2.2	Projeto com um fator categórico	19
3.2.3	Determinação do Erro Experimental	21
3.2.4	Projeto de experimentos de Mistura	22
3.2.5	Arranjo de mistura cruzado	32
3.2.6	Análise dos Experimentos	35
3.3	Programação Lexicográfica	38
3.4	Análise de Agrupamentos (<i>Clusters</i>)	39
3.4.0.1	Medidas de similaridade	40
3.5	Análise de componentes principais	43
3.6	Otimização de processos	44
3.6.1	Método <i>desirability</i>	45
3.7	Considerações Finais Sobre o Capítulo	47
4	Portfólio e Técnicas de Seleção	48
4.1	Considerações Iniciais	48
4.2	Seleção de portfólio	48
4.2.1	Influência da correlação no portfólio	51
4.2.2	Análise de portfólio baseado em MDE	52
4.3	Considerações Finais Sobre o Capítulo	53
5	Metodologia Experimental e Procedimentos Descritivos	54
5.1	Considerações Iniciais	54
5.2	Metodologia	54
5.3	Método indicadores financeiros incorporados ao arranjo de mistura cruzado	55
5.4	Método de simulação	58
5.5	Procedimento descritivo da coleta dos dados	59
5.6	Procedimento descritivo do PCA	60
5.7	Procedimento descritivo dos arranjos experimentais	61
5.8	Procedimento descritivo das análises	62
5.9	Procedimento descritivo das otimizações	62
5.10	Descrição dos dados	63
5.11	Fases do Macrofluxo	65
5.11.1	Classificação da <i>Newsweek</i> com as 500 empresas mais verdes	65
5.11.2	PCA aplicado nos indicadores originais criando uma nova classificação pelo MGI	66
5.11.3	Análise dos dados Nasdaq das 20 empresas melhores classificadas pelo MGI	69
5.11.4	Análise dos Balanços Patrimoniais anuais	71
5.11.5	Cálculo dos indicadores financeiros dos Balanços Patrimoniais	71

5.11.6	Replicação dos valores de fechamentos mensais das empresas selecionadas por modelo de autocorrelação	73
5.11.7	Cálculo dos riscos e retornos com réplicas	74
5.11.8	Geração de <i>Clusters</i> baseados nos indicadores calculados para formação de misturas menores	74
5.11.9	PCA aplicado nos indicadores financeiros para definição das variáveis (-1) e (+1)	77
5.12	Limitações do Modelo	77
5.13	Superando as Limitações	79
5.14	Considerações Finais Sobre o Capítulo	79
6	Seleção e Otimização de Portfólios com Indicadores Financeiros	80
6.1	Considerações Iniciais	80
6.1.1	Modelagem de cada <i>cluster</i> considerando variáveis de processo . . .	81
6.1.2	Metodologia das demonstrações gráficas nas otimizações	85
6.1.3	Otimização dos <i>Clusters</i> individuais	87
6.1.4	Modelagem combinada de todos os <i>clusters</i> considerando método lexicográfico e variáveis de processo (CMO)	94
6.1.5	Otimização dos arranjos combinados	96
6.1.6	Modelagem por <i>cluster</i> e variáveis de processo juntos utilizando CM-DOE	98
6.1.7	Otimização do arranjo CM-DOE	101
6.1.8	Modelagem de todas as empresas juntas utilizando o modelo Markowitz	106
6.1.9	Otimização utilizando o modelo de Markowitz com otimização por MDE	108
6.2	Comparação dos métodos	112
6.3	Considerações Finais Sobre o Capítulo	113
7	Conclusão	118
	Conclusão	118
7.0.1	Sugestões para trabalhos futuros	120
	Referências	121
	Apêndices	130
	APÊNDICE A ANOVA e Gráficos com Análise de cada <i>Clusters</i> com Variáveis de Processo	131

APÊNDICE B	ANOVA e Gráficos com Análise combinada de todos os <i>clusters</i> considerando método lexicográfico e variáveis de processo (CMO)	149
APÊNDICE C	ANOVA e Gráficos com Análises por <i>cluster</i> e variáveis de processo juntos utilizando CM-DOE	153
APÊNDICE D	ANOVA e Gráficos com Análise utilizando o modelo Markowitz	160
Anexos		164
ANEXO A	Optimization methodology with principal component analysis applied in portfolio: Simulated results in Green Companies	165
ANEXO B	ISERC 2015 - Crossed mixture design of experiments applied on multi-period portfolio	179
ANEXO C	RISCO: MEDIDAS DE RISCO, AVERSÃO E SEU GERENCIAMENTO	181

1 INTRODUÇÃO

1.1 CONSIDERAÇÕES INICIAIS

O estudo de portfólios é uma área amplamente pesquisada e publicada. Neste contexto há sempre um ponto comum, em todos os trabalhos há uma escolha prévia das empresas que irão compor os portfólios. Esta é a primeira lacuna a ser trabalhada: desenvolver uma forma de selecionar empresas para compor portfólios que seja embasada em indicadores calculados. Estes indicadores podem ser de diversos tipos como indicadores de sustentabilidade ou indicadores financeiros por exemplo. Na proposta desta tese utiliza-se um indicador chamado de *Verde* (Green Companies). Este indicador é ranqueado e publicado anualmente pela revista Newsweek ([CAPITAL, 2014](#)) e tem o objetivo de divulgar as empresas que trabalham de forma mais sustentável, ou seja, com mínimos impactos ao meio ambiente; ou que tenham ainda ações de melhoria para os impactos que suas atividades possam causar ao meio ambiente. A escolha deste indicador inicial teve como base a publicação de [Lyon e Shimshack \(2012\)](#) que concluíram, em seu estudo, que as cem empresas melhor classificadas na edição da Newsweek do ano de 2012 no quesito *Verde* tiveram uma anomalia positiva em seus rendimentos, uma valorização dos seus ativos de 0,6 a 1,0% superior ao das demais quatrocentas empresas. Este foi o motivador para o uso das *Green Companies* como base de empresas para o estudo.

Outra lacuna que identificamos diz respeito ao uso de variáveis externas como parte integrante das análises. No portfólio tradicional trabalha-se sobre os resultados dos fechamentos periódicos dos ativos, isto quer dizer que, quando há um aumento no valor da ação de um período para outro, houve ganho. Quando há uma redução do valor da ação de um período em relação a outro, houve perda. [Markowitz \(1952\)](#) foi o ganhador do prêmio Nobel de Economia com seu trabalho sobre portfólio. Nele se fez uma formulação na qual deseja-se, com um risco especificado, ter um maior retorno, ganho, possível em um conjunto de empresas ao invés de uma empresa, um ativo apenas. A inserção de variáveis externas nesse modelo é uma grande contribuição porque tornará as decisões de investimento mais robustas e embasadas. Considerando também um conjunto de empresas, portfólio.

Para que o conjunto de indicadores possa ter também um embasamento, escolheu-se o trabalho de [Barua e Saha \(2015\)](#), no qual listou-se um conjunto de indicadores importantes e relevantes para se ter medidas sobre a capacidade financeira da empresa. A montagem desses indicadores dá-se pelo cálculo de valores retirados da base de dados dos balanços patrimoniais das empresas escolhidas para a formação dos portfólios. Os balanços patrimoniais anuais são uma obrigatoriedade para todas as empresas que possuem capital

aberto, ou seja, todas as empresas que possuem ações à venda em bolsa de valores.

Voltando ao entendimento, percebe-se que a fase inicial para a tomada de decisão em investimentos deve ser a escolha de qual investimento fazer, em qual ou quais empresas. Esta etapa tão importante é negligenciada em análise de portfólios. Por esta razão viu-se a necessidade da criação de um método que considera, inicialmente, a escolha das empresas que formarão o(s) portfólio(s). A teoria de portfólio é uma área das finanças amplamente estudada com uma grande diversidade de métodos, começando com [Markowitz \(1952\)](#) e continuando com muitos outros podendo citar-se [Markowitz \(1959\)](#), [Cass e Stiglitz \(1970\)](#), [Amihud e Barnea \(1974\)](#), [Mayers \(1974\)](#), [Ngun *et al.* \(2014\)](#), [Costa e Paiva \(2002\)](#), [Bertland e Asgharian \(2008\)](#), [Chiarawongse *et al.* \(2012\)](#). Nestes casos, as análises e otimizações são feitas com base em portfólios pré-formados, ou seja, com os investimentos, ativos, empresas previamente selecionadas. Este estudo tem início antes desta formação, indicando inclusive quais ativos de quais empresas considerar para as análises.

O alvo da pesquisa são ativos financeiros de empresas que possuem capital aberto em bolsa de valores, destas foram selecionadas aquelas consideradas “empresas verdes”. É fato, que o risco de uma crise energética está transformando a maneira de como os países e as empresas fazem negócios. De acordo com [Rodrigues e Halmeman \(2012\)](#) desde o ano de 2001 ocorreram diversos blecautes conhecidos como “apagões” por falta de energia elétrica no Brasil. Mostrou ainda que de 2011 a 2020 a demanda energética total do país deverá crescer em mais de 60%, prevendo que em 2020, aproximadamente 66% do consumo total deverá ser incorporado pelos setores industrial e de transporte. A utilização de fontes renováveis de energia, combinadas com novas maneiras de usar os recursos naturais existentes, deverá ser, em breve, um requisito para todos os países e, conseqüentemente, para as empresas e corporações. Estas exigências são também uma motivação para os clientes escolherem comprar produtos das empresas sustentáveis, que se preocupam e agem em prol de melhorias para o meio ambiente, ao invés de adquirir daquelas não sustentáveis. Isto pode levar os investidores do mercado de Ações a dar mais atenção aos chamados “Ativos Verdes”, aqueles relacionados a empresas consideradas verdes por operarem de forma sustentável, do ponto de vista ecológico.

Esse nível de comprometimento das empresas com a sustentabilidade pode ser medido por índices. Anualmente são divulgados alguns índices de sustentabilidade e ações praticadas em prol da melhoria do meio ambiente das maiores corporações mundiais. Estes índices dão origem a uma classificação publicada pela revista *Newsweek* ([CAPITAL, 2014](#)) com um ranqueamento chamado de ranque verde. Tal classificação demonstrou impacto positivo no valor das ações das empresas com maiores pontuações.

Para segregar um número interessante de empresas e formar nossa própria classificação, utilizamos da metodologia de análise de componentes principais (PCA) ([HO-](#)

TELLING, 1933) aplicado sobre os indicadores originais da revista supracitada. Esta nova classificação delimitou as empresas a serem analisadas mais profundamente. Este aprofundamento contemplou inclusive a análise dos balanços patrimoniais para desenvolvimento de indicadores de performance financeira.

A base de dados para pesquisa é em parte provida dos fechamentos mensais dos valores das ações destas *Green Companies* bem como dos seus balanços patrimoniais anuais, no caso, os balanços referentes ao ano de 2013 e que foram divulgados em 2014.

Os modelos tradicionais de portfólio que utilizam os métodos de Markowitz (1952) tem como base de cálculo os riscos e retornos para cada investimento. Neste estudo, além do risco e retorno, contemplamos os indicadores de sustentabilidade, indicadores financeiros de performance, e método de agrupamento (*clusters* (JR, 1963)); com otimização multiobjetivo. Todos com o auxílio das estratégias de mistura de misturas com variáveis de processo e otimização lexicográfica. Tais variáveis de processo são definidas pelos PCAs.

O modelo de programação por metas com priorização (Lexicographic goal programming) LGP ordena os objetivos de acordo com sua importância, fazendo uma priorização baseada em decisores. No caso desta tese, esses decisores são uma otimização. Cada estágio de uma mistura é otimizado e o ponto ótimo encontrado será o ponto inicial da próxima otimização até que se tenha um valor de ótimo global. O ótimo global é, pois, a melhor combinação dos ótimos locais.

O delineamento de experimentos (DOE) e, em alguns casos, delineamento de experimentos de mistura (MDE) é bastante usado na otimização de processos e produtos (SCHEFFÉ, 1958; SCHEFFÉ, 1963; CHAN, 1988; DUINEVELD *et al.*, 1993; ERIKSSON *et al.*, 1998; VIEIRA; BELLO, 2006; ZHANG; WONG, 2013). MDE é normalmente utilizado em campos de estudo da química e da indústria alimentícia. Oliveira *et al.* (2011), utilizou MDE para otimizar portfólios no mercado de energia elétrica pela primeira vez. Também Mendes *et al.* (2016) utilizou MDE e superfície de resposta como forma de modelagem para otimização de portfólio. Análises comparativas entre métodos também são demonstradas com resultados muito satisfatórios através de simulação de investimento de \$ 100.000,00. Nesta tese desenvolveu-se um método sistemático agrupando diversas técnicas para auxiliar os investidores a tomarem as melhores decisões partindo da escolha das empresas, sobre aplicação de capital neste novo seguimento de mercado chamado de empresas verdes.

1.2 JUSTIFICATIVA E RELEVÂNCIA

Propõe-se o desenvolvimento de um método sistemático para composição de portfólios e análise com otimização lexicográfica baseado no delineamento de experimentos de misturas com variáveis de processo. As variáveis de processo são identificadas pelo

PCA aplicado em indicadores financeiros. Estes indicadores são calculados a partir das informações contidas nos balanços patrimoniais das empresas escolhidas. Esta etapa de cálculo também é uma contribuição do trabalho. O cálculo é baseado em informações dos balanços patrimoniais. Cada número relevante ao cálculo é identificado e extraído dos balanços de forma manual e então agregado ao cálculo final dos indicadores. A escolha das empresas começa no *ranking* divulgado pela revista *Newsweek* do ano de 2014.

Percebe-se uma carência neste processo de análise e otimização de portfólios. Esta diz respeito ao método de escolha desses investimentos. Como escolher, neste caso, as empresas e ativos participantes, que irão compor os portfólios. Isto é um desafio até então não transposto. É relevante, visto que a análise de portfólios é amplamente estudada por diversos autores como por exemplo [Markowitz \(1952\)](#), [Markowitz \(1959\)](#), [Amihud e Barnea \(1974\)](#), [Mayers \(1974\)](#), [Li e Ng \(2000\)](#), [Garlappi *et al.* \(2007\)](#) e ainda [Babaei *et al.* \(2015\)](#), [Liu *et al.* \(2015\)](#) entre outros. As análises destes trabalhos são feitas a partir de portfólios já formados. Portfólios que já estão compostos por certas empresas, investimentos, ativos pré definidos, que são então otimizados e/ou apenas analisados.

A utilização de diversificação funcional com o uso de *clusters*, ou conjuntos, têm tido foco em pesquisas no campo da ecologia a vários anos como citado por [Podani e Schmera \(2006\)](#) e será utilizado aqui para aglutinar, formar *clusters* com empresas e indicadores financeiros. Agrupar esse método com delineamento de experimentos de misturas e teoria de portfólios com aplicação em ativos de empresas verdes considerando também indicadores de performance financeira com otimização lexicográfica será inédito.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

O objetivo deste trabalho é desenvolver um método sistemático que possibilitará que o investidor inicie todo seu processo de aplicação a partir da escolha de quais empresas deve compor seu portfólio. Além disso o cálculo de indicadores financeiros baseados em informações específicas retiradas uma a uma dos balanços patrimoniais de cada empresa escolhida. Inclusão desses indicadores em forma de PCA nos arranjos de mistura como variável de processo para que os arranjos sejam otimizados de forma lexicográfica. Validação da metodologia e comparação dos resultados.

1.3.2 Objetivos Específicos

Os objetivos específicos deste trabalho incluem:

- Selecionar, com base em PCA aplicado a indicadores, ativos de empresas verdes para compor portfólios;

- classificar empresas verdes com o uso de MGI dos PCAs de seus indicadores de sustentabilidade publicados na revista *Newsweek* gerando uma nova classificação;
- calcular os indicadores financeiros com base em dados específicos extraídos do balanço patrimonial de cada empresa;
- agrupar empresas em clusters com uso do método de Ward ([Jr \(1963\)](#)) pelos indicadores financeiros;
- analisar arranjos de experimentos de mistura com variáveis de processo definidas por valores de PCA;
- otimizar os portfólios pela função *desirability* descrita em [Myers Raymond H. \(2009\)](#);
- comprovar a utilidade e usabilidade do método, comparando os resultados com outros modelos de otimização de portfólios, principalmente como o modelo que é referência, Markowitz ([Markowitz \(1959\)](#)).

1.4 DELIMITAÇÃO DO TRABALHO

Esta tese limita-se às seguintes condições de contorno:

- (a) Desenvolvimento da metodologia de seleção de empresas e análise de portfólios com otimização por programação lexicográfica com arranjos de mistura e variáveis de processo geradas por indicadores financeiros;
- (b) uso da classificação publicada pela revista *Newsweek* para identificar as empresas norte-americanas mais ecologicamente sustentáveis chamadas empresas verdes;
- (c) aplicação de PCA para agrupamento de indicadores financeiros;
- (d) utilização de variáveis de processo convertidas por indicadores financeiros oriundos de informações coletadas nos balanços patrimoniais;
- (e) as empresas consideradas para o estudo são todas norte-americanas;
- (f) a publicação da *Newsweek* utilizada foi a do ano base 2014;
- (g) os balanços patrimoniais analisados foram do ano base 2014;
- (h) as cotações históricas foram compostas por 120 meses decrescentes a 07 de Julho de 2015;
- (i) empresas que sofreram fusão ou tinham histórico de cotações menor do que sessenta meses foram descartadas;

- (j) o tipo de transação Venda a Descoberto não foi considerado;
- (k) o método de otimização com método lexicográfico e otimizado pela função *desirability* com o uso do *software* Design Expert®7.0 versão de teste.

1.5 CARÁTER INÉDITO

O presente trabalho pretende inovar em relação ao estado atual da pesquisa sobre o tema nos seguintes aspectos:

- (A) seleção, aglomeração e formação de portfólios com as empresas verdes selecionadas por PCA com MGI;
- (B) geração de indicadores baseados em informações contidas nos balanços patrimoniais das empresas selecionadas;
- (C) redução da dimensionalidade das variáveis para formação de *clusters* e dos portfólios;
- (D) incorporação dos indicadores financeiros nos modelos de mistura como variáveis de processo;
- (E) uso do método lexicográfico para a otimização de portfólios;
- (F) comparação com o modelo de portfólio de Markowitz ([Markowitz \(1959\)](#));
- (G) programação de código fonte para gerar matrizes de misturas sem limites de quantidades de componentes utilizando o *software* Matlab®2008b versão de teste.

1.6 METODOLOGIA

A análise da literatura revela que os estudos sobre seleção e otimização de portfólios seguem uma sequência de atividades que não consideram um método para escolha das empresas que irão compor esses portfólios ([Markowitz \(1952\)](#), [Markowitz \(1959\)](#), [Amihud e Barnea \(1974\)](#), [Mayers \(1974\)](#), [Li e Ng \(2000\)](#), [Garlappi *et al.* \(2007\)](#) e ainda [Babaei *et al.* \(2015\)](#), [Liu *et al.* \(2015\)](#)). Estes concentram-se em definir os pesos a investir e na melhor relação risco-retorno da carteira.

O método de pesquisa experimental com simulação será o mais adequado. Tendo-se em vista que é possível, pelo uso de ferramentas computacionais, calcular e replicar retornos e riscos das ações das empresas. A metodologia do projeto de experimentos (DOE) é utilizada aqui, com ênfase nas técnicas de experimentos de misturas com variáveis de processo. Os fatores do DOE (as variáveis independentes) foram, na fase de coleta de dados, as cotações históricas dos valores das ações das empresas selecionadas, em que o planejamento experimental teve por meta agrupar as cotações históricas com indicadores

financeiros como variáveis de processo. Em uma segunda etapa, a fase de análise de dados, os indicadores financeiros transformados em variáveis de processo (-1) e (+1) foram incluídos. Isso foi feito com o objetivo de verificar se estes influenciam no portfólio e obter a configuração ótima de cada portfólio obtendo-se o maior retorno a partir da definição de um risco máximo. Para cada etapa, os resultados obtidos foram analisados e interpretados. A validação do método proposto foi realizada por meio de simulação. Esta deu origem a conjuntos de retornos e riscos de investimentos. Os valores percentuais ótimos obtidos pela aplicação do método foram utilizados para determinar a melhor configuração de investimento. Desta forma, pode-se classificar o trabalho realizado, quanto à sua natureza, como pesquisa aplicada. Em relação ao seu objetivo, o trabalho se classifica como sendo exploratório com pesquisa documental (BERTRAND; FRANSOO, 2002), com abordagem quantitativa (CRESWELL *et al.*, 2011). O método utilizado foi o experimental (BRYMAN *et al.*, 2003). Simulações foram executadas para obtenção de resultados necessários para comparação.

1.7 ORGANIZAÇÃO DO TEXTO

Esta tese está dividida em uma estrutura de 7 capítulos nomeados a partir do segundo e descritos como:

Capítulo 2: Empresas verdes (*Green companies*) e a Bolsa de Valores.

O que são empresas verdes e qual sua visibilidade nas novas tendências e preocupações ecológicas.

Capítulo 3: DOE, PCA, programação lexicográfica e *clusters*.

Descreve o projetos de experimentos e as variações do projeto de experimento de misturas. Além de conceituar PCA, o método de programação lexicográfica e a metodologia de *clusters*.

Capítulo 4: Teoria de portfólio.

Este capítulo tratará da teoria de portfólio e alguns de seus campos de estudo, relacionando-a com os experimentos de mistura.

Capítulo 5: Metodologia Experimental e Procedimentos Descritivos.

Descreve a metodologia e procedimentos de forma sistemática. Apresenta as propostas e modelos de funções da tese para atingir melhores resultados.

Capítulo 6: Seleção e Otimização de Portfólios com Indicadores Financeiros.

Descreve a aplicação do método sistemático com análises e otimizações, além de comparação numérica entre métodos e simulação de aplicação.

Capítulo 7: Conclusão.

Apresenta a conclusão da Tese e sugestões de trabalhos futuros.

2 EMPRESAS VERDES E MERCADO FINANCEIRO

2.1 CONSIDERAÇÕES INICIAIS

Neste capítulo, parte da fundamentação teórica relativa às definições e terminologias importantes de empresas verdes, bem como os indicadores de gestão sustentável e os indicadores que definem o quanto a empresa pode ser considerada verde são descritos. Os índices de sustentabilidade utilizados pela classificação da revista *Newsweek* são identificados. Iniciando-se com uma visão geral de empresas verdes, regulamentação governamental e passando por fatores motivadores dessas empresas como recompensas e incentivos financeiros do Estado e o crescimento da importância dada ao tema.

Os principais atributos e terminologias usadas nos mercados financeiros são descritos através das definições de Bolsas de Valores, Ações e indicadores financeiros, assim como as principais normativas governamentais para as empresas S.A. utilizando para melhor entendimento a regulamentação brasileira descrita na “LEI No 6.404, DE 15 DE DEZEMBRO DE 1976 ”([FEDERAL, 1976](#)).

2.2 EMPRESAS VERDES

A motivação pela escolha do objeto de estudo ter como foco as empresas verdes (*Green Companies*) se deu pelo fato de [Lyon e Shimshack \(2012\)](#) terem identificado em sua pesquisa que as empresas verdes tendem a retornar um resultado mais positivo nas cotações de suas ações. Ser considerada uma empresa verde é um grande desafio. [Saha e Darnton \(2005\)](#) listaram alguns aspectos principais para uma empresa ser reconhecida como verde, são eles: interesse ecológico, conservação da fauna/flora, responsabilidade corporativa social, interesse humanitário, negociações justas, água limpa, bem estar animal, qualidade, sustentabilidade. Para garantir que hajam recursos naturais para o futuro, estes aspectos ambientais devem ser cada vez mais considerados.

O reconhecimento social que empresas verdes recebem é tão importante quanto o reconhecimento financeiro. A médio prazo, empresas não sustentáveis e aquelas poluentes podem deixar de existir, tanto por força governamental, quanto por forças sociais. Esta visibilidade pode ser comparada ao que ocorreu no mercado da década de oitenta com referência a empresas de inovação tecnológica e de produtos citadas por ([COOPER, 1990](#)) na qual descreve que quanto mais inovativas as empresas, mais terão valorização de seus Ativos a longo prazo. Sua fonte de dados para esta afirmação foi a publicação da revista *Fortune and Erdos & Morgan, Inc.*.

O *Newsweek Green Companies Ranking* ([CAPITAL, 2014](#)) é um processo de posi-

cionamento composto pelos indicadores: produtividade energética, produtividade do carbono, produtividade da água, produtividade de desperdício e reputação no mercado. Estes indicadores tem um lastro de 0,00% a 100% no qual maiores percentuais representam melhores (mais verdes) pontuações. A revista é publicada anualmente nos Estados Unidos da América e a cada ano repercute mais nas preferências dos consumidores e usuários de serviços.

Oitenta e oito por cento das cem maiores empresas norte-americanas divulgam seus números sobre gás e energia utilizados com efeito estufa, enquanto sessenta e cinco por cento divulga o uso da água. Em torno da totalidade da classificação, se a empresa não divulga seus números de gases de efeito estufa, energia, uso da água e geração de lixo, é considerada *fora do jogo*. As duas principais razões pelas quais a divulgação destes números é tão importante são: a pressão dos grupos investidores que investem \$87 trilhões de dólares em Ações e o crescente conjunto de leis e regulamentações que exigem a divulgação destes indicadores ambientais. Estas métricas divulgadas, tais como emissão de gases de efeito estufa, agora podem ser incorporadas às estratégias dos investidores. Há ainda dados que mostram que investimentos em empresas verdes são mais inteligentes. A (KNIGHTS, 2016) reportou que um investimento de \$100 igualmente distribuído entre empresas norte-americanas que têm melhores indicadores de emissão de gases com efeito estufa, tiveram um retorno de \$220 nos últimos cinco anos contra \$160 se investidos em outras quinhentas empresas. Investimentos nessas empresas verdes também levaram a 93% menos emissões de gases de efeito estufa considerando iguais investimentos em outras empresas.

Mais de um terço das quinhentas empresas norte-americanas e 71% das empresas globais da classificação da *Newsweek* têm seus números de meio ambiente auditados ou revisados por empresas de consultoria em meio ambiente. Isto inspira mais confiabilidade nos números e aumenta as chances de investidores e consumidores utilizarem estas informações quando considerarem a compra de produtos ou investir em Ações.

O departamento de energias dos Estados Unidos (*U.S. Department of Energy*) motiva, com ajudas monetárias e premiações, empresas sustentáveis e que ajudam na recuperação do meio ambiente. A Figura 2.1 é um exemplo da importância do tema verde “...o Departamento de energias investe mais de US\$31 bilhões para auxiliar a grande gama de projetos de energia limpa em todo país” (CLEARINGHOUSE, 2015). Empresas tanto que projetam quanto as que querem utilizar energia limpa e possuem projetos sustentáveis são suportadas pelo governo. Esta é mais uma evidência do crescimento, em importância e visibilidade, das empresas verdes.

Bansal e Roth (2000) conduziu uma pesquisa em 53 empresas demonstrando as principais motivações para estas serem verdes, Figura 2.2. Lyon e Shimshack (2012) estudaram os impactos de mídia gerados por programas de sustentabilidade das empresas

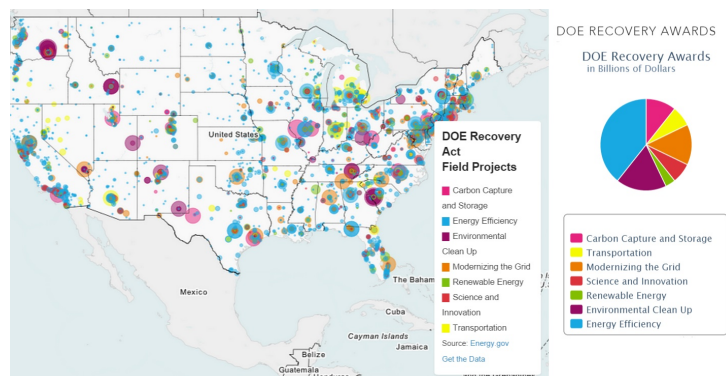


Figura 2.1 – Mapa das ações de recuperação ambiental dos E.U.A.

usando o posicionamento da *Newsweek* 2009 das 500 maiores empresas dos Estados Unidos. Eles confirmaram que as 100 melhores (mais verdes) empresas deste posicionamento experimentaram uma anomalia nos retornos das suas ações, sendo que estas tiveram retornos de 0.6% a 1% maior do que o retorno observado nas demais 400 empresas. Esta é outra evidência de reação positiva do mercado perante a preocupação das empresas com o meio ambiente.

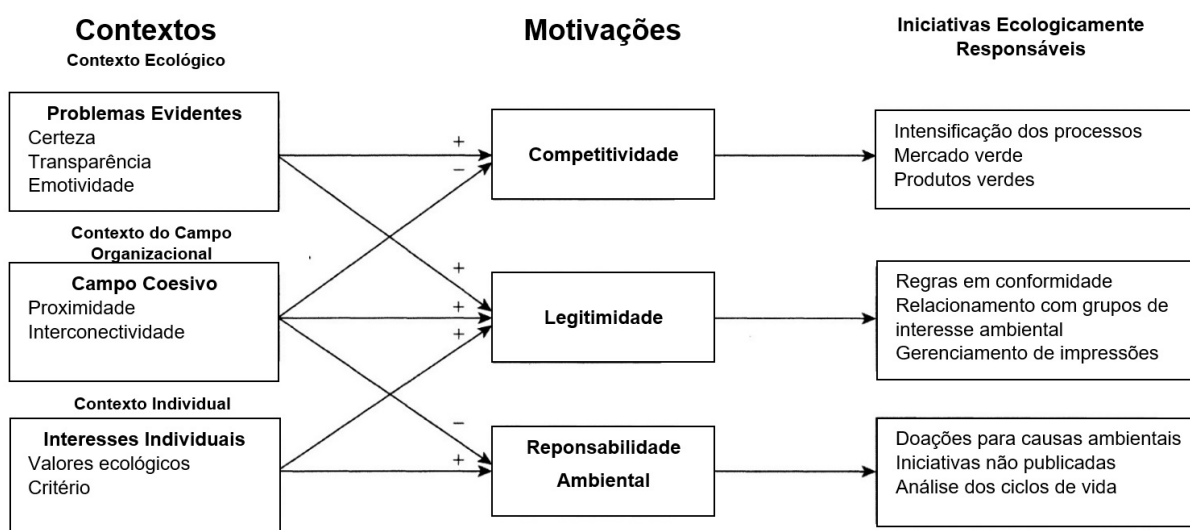


Figura 2.2 – Modelo de motivação ecológica empresarial. Fonte: (LYON; SHIMSHACK, 2012).

2.2.1 Bolsa de valores

Bolsas de valores são associações civis e sem fins lucrativos que funcionam como se fossem clubes privados. Ela é a sede das corretoras de títulos e valores mobiliários. Essas corretoras compram um título da instituição que lhes permite, como associada, fazer as negociações. A administração é feita por um conselho de corretores, eleito entre os membros associados, que indicam o presidente da entidade, para um mandato de quatro anos (SERIFSOY, 2006).

As bolsas de valores de todo o mundo são geralmente classificadas como leilão ou mercados revendedores, sendo este um local de encontro central em que os lances de compradores e vendedores convergem. São grandes mercados nos quais são negociados lotes de Ações que podem ou não ter corretores (especialistas). Em um mercado de Ações, o cliente coloca uma ordem com um corretor e o corretor é responsável por pesquisar o melhor preço entre as empresas que fazem parte da bolsa de valores (MCCONNELL *et al.*, 1996).

Uma barra de preços mostrada na Figura 2.3, simbolizada por uma barra vertical, é o registro pictográfico (ou traçado) de um dia de atividade do preço (um pregão) de um Ativo financeiro (podem ser índices, Ações ou mercadorias agropecuárias/financeiras), na qual cada preço é um consenso momentâneo de valor de todos os participantes do mercado, expresso em movimento. Cada barra de preço fornece alguns pedaços de informação sobre o equilíbrio de forças entre compradores e vendedores (TÍER, 2005). Na barra vertical, através de um traço (tique) horizontal à sua esquerda está representado o nível de preço do primeiro negócio do dia, a abertura. O último negócio do dia, o fechamento, é representado por um tique horizontal à sua direita. As extremidades superior e inferior representam respectivamente à máxima e à mínima atingidas neste dia. Ele classifica também o risco, como uma função do retorno, expresso na Figura 2.4.

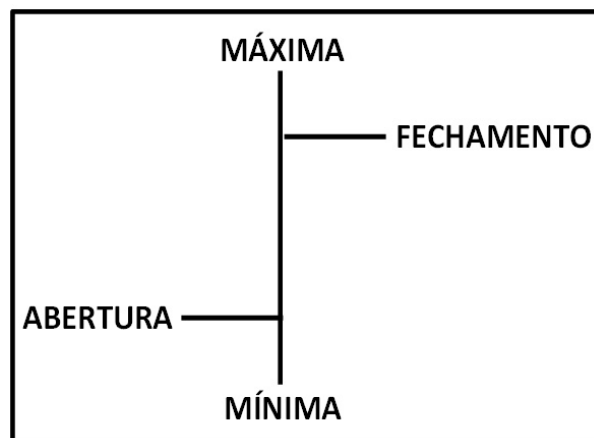


Figura 2.3 – A barra de preços das ações

Segundo Baier *et al.* (2004), a Bolsa de Valores facilita a negociação de Ações contribuindo para o crescimento do comércio, sendo que todas as transações são feitas de forma impessoal e possibilitam a compra e venda de ativos das empresas para os consumidores ou investidores. Ela é formada para diminuir o custo de negociação dos direitos de propriedade das empresas.

Garcia e Giambiagi (2010) citaram que cabem às bolsas orientar e fiscalizar os serviços prestados por seus membros (e, se for o caso, receber reclamações contra as corretoras), facilitar a divulgação constante de informações sobre as empresas de capital aberto (que têm Ações) e os negócios realizados nos pregões diários. Além de central de

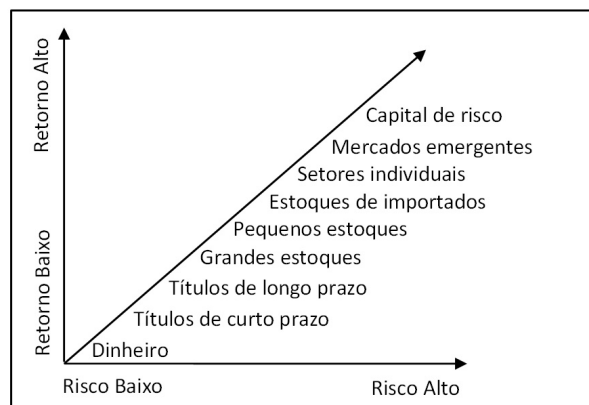


Figura 2.4 – Risco em função do retorno

negócios, as bolsas também controlam a efetivação das operações realizadas e garantem negócios a médio e longo prazo. Há três formas de negociação: à vista, a termo e por meio do mercado de opções. As operações nesses mercados podem ser feitas por meio de pregões eletrônicos, no qual os negócios são realizados diretamente por meio de computador.

Além do efeito de novas listagens no preço de Ações, a covariância entre o retorno de várias Ações tem despertado o interesse dos investidores, por causa de seu papel de destaque como a medida adequada do risco de uma Ação no modelo de precificação de Ações (CAPM - *Capital asset pricing model*). As motivações para os estudos das covariâncias são duas: primeiro, a covariância tem sido estudada como uma questão de curiosidade científica. Segundo, tem sido analisada como uma possível explicação para o aumento no preço das Ações, que acompanham os novos anúncios. O argumento é que, se o CAPM está correto, a covariância mede o risco de uma empresa que determina o seu custo de capital. Se a lista reduz o risco, então a consequente redução, retorno exigido da Ação poderia explicar o aumento de preços de Ações no momento da listagem ([MCCONNELL et al., 1996](#)).

2.2.1.1 Ações

Ação é a menor parte do capital de uma empresa. São títulos de renda variável, emitidos por empresas de capital aberto ou Sociedades Anônimas (S.A.). As ações podem ser escriturais (existirem apenas nos registros da empresa) ou representadas por caute-las e certificados. O investidor em Ações é um coproprietário da sociedade anônima da qual é acionista, participando de seus resultados. As Ações são convertidas em dinheiro, a qualquer tempo, por meio de negociação em bolsas de valores ou no mercado de balcão ([Garcia e Giambiagi \(2010\)](#)). O capítulo III da Lei ([FEDERAL, 1976](#)) trata das definições principais e variações de Ações, sendo destacadas que as Ações podem ser divididas basicamente em duas espécies, Ordinárias e Preferenciais. As ordinárias são aquelas que dão direito de voto em separado para o preenchimento de determinados cargos de órgãos

administrativos aos seus detentores. As preferenciais dão aos seus detentores a prioridade na distribuição de dividendo, lucros a serem distribuídos aos acionistas, fixo ou mínimo além de prioridade no reembolso do capital, com prêmio ou sem ele.

2.2.2 Indicadores financeiros

Indicadores financeiros são uma importante maneira quantitativa de medir a performance de uma empresa. Empresas que falham no mercado exibem significantes diferenças em seus indicadores financeiros quando comparadas com empresas de sucesso. Essas comparações podem ser feitas utilizando cinco indicadores chave descritos em [Altman \(1968\)](#).

O risco de uma empresa, risco do negócio pode se dividido em: de fontes internas e de fontes externas. As fontes internas são em geral, reflexos de variabilidade nos lucro sobre as operações ou também sobre o fluxo de caixa. As fontes externas são aquelas cujas causas são oriundas do mercado, que produz variabilidade nos preços, sejam de matérias primas, sejam dos produtos comercializados pela própria empresa ([GABRIEL; BAKER, 1980](#)). Quanto maior o risco do negócio, maiores as chances do investidor perder dinheiro com esses Ativos, essas Ações. Diferentemente do risco calculado pelos históricos. Neste caso, a o risco é tanto de perder quanto chance de ganhar.

As empresas norte-americanas que possuem ativos em bolsa de valores são regulamentadas pelo Ato de Segurança de Trocas ([SECURITIES, 1934](#)). Essas empresas são obrigadas a publicar anualmente seus indicadores financeiros e os balanços patrimoniais completos. No Brasil, com base no Art. 176 da Lei ([FEDERAL, 1976](#)) “Ao fim de cada exercício social, a diretoria fará elaborar, com base na escrituração mercantil da companhia, as seguintes demonstrações financeiras, que deverão exprimir com clareza a situação do patrimônio da companhia e as mutações ocorridas no exercício:

- I - balanço patrimonial;
- II - demonstração dos lucros ou prejuízos acumulados;
- III - demonstração do resultado do exercício; e
- IV – demonstração dos fluxos de caixa; e
- V – se companhia aberta, demonstração do valor adicionado.”

A maioria destes indicadores têm seu foco nos balancetes e faturamentos. Por outro lado, o fluxo de caixa pode oferecer uma melhor compreensão na análise dos indicadores quando consideram-se períodos contínuos ([BARUA; SAHA, 2015](#)). Os indicadores considerados nesta tese estão listados no Quadro 2.1.

Quadro 2.1 – Indicadores

Análises do uso de débitos		
1	débito sobre capital	total de débitos / capital total
2	débito sobre capital próprio	total de débitos / capital próprio total
Análise dos indicadores de taxas de juros		
3	juros ganhos sobre os impostos	lucros antes de juros e impostos / despesas de juros
4	cobertura de taxa fixa	lucros antes de taxas e impostos / taxas fixas
5	juros ganhos em dinheiro	fluxo de caixa operacional ajustado / despesas de juros
6	taxa fixa de cobertura	fluxo operacional ajustado / taxas fixas
7	despesas de capital	fluxo de caixa sobre as operações / capital
8	caixa sobre débito operacional	fluxo de caixa operacional / total de débitos

Fonte: Adaptado de [Barua e Saha \(2015\)](#)

2.3 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO

Conforme pode ser constatado nesta parte da revisão da literatura, as empresas verdes estão ocupando posição de destaque na preferência de investidores. O investimento em ações é uma forma de aplicação que tem um controle rigoroso nas demonstrações financeiras das empresas S.A. que operam neste mercado. O cálculo de indicadores baseados nessas demonstrações financeiras pode ser de grande valia na tomada de decisão dos investidores sobre qual empresas alocar seus fundos.

3 MÉTODOS DE EXPERIMENTAÇÃO E AGRUPAMENTO

3.1 CONSIDERAÇÕES INICIAIS

Neste capítulo, apresenta-se a fundamentação teórica sobre o processo experimentação baseado em projeto de experimentos, além de processo de experimentos de mistura e suas aplicações. Métodos de agrupamento também são referenciados com ênfase em análise de componentes principais. Além disso também é tratada a programação lexicográfica e sua utilidade para esta tese. Este capítulo é importante para a compreensão de quais métodos serão utilizados e como sua junção diferencia este trabalho de outros.

3.2 DELINEAMENTO DE EXPERIMENTOS

Experimentos são os mais eficientes métodos para o descobrimento. Eles são baseados em interferências no meio para saber como o objeto de estudo reage e, encontrando assim, sua forma de funcionamento. Alternativas para isso são a observação passiva e a experimentação. Na observação passiva, o observador fica afastado assistindo e esperando que algo interessante aconteça (Figura 3.1). A observação passiva obtém grandes montantes de dados, mas são difíceis para isolar a informação que é relevante para o problema particular da pesquisa. A experimentação, por outro lado, provê dados adaptados ao problema observado. Isto significa que experimentos gastam menos esforços e obtêm menos informações consideradas irrelevantes (ANDERSSON, 2012). A grande diferença é que: o observador passivo investiga um estado *pré-existente* (Figura 3.1) enquanto que o experimentador *muda o estado* e investiga o resultado das mudanças (Figura 3.2).



Figura 3.1 – Observador passivo investigando um estado pré-existente

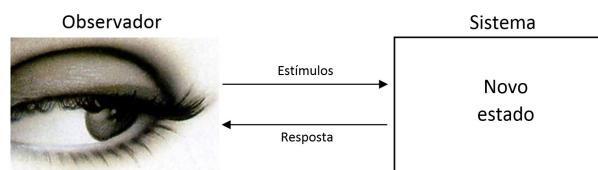


Figura 3.2 – Experimentador muda o estado e investiga os resultados obtidos

O método estatístico padrão ampliado para o estudo de projetos de experimentos é referenciado a mais de sessenta anos com Fisher (1951), Villars *et al.* (1951). O Projeto

de Experimentos (DOE) é uma poderosa ferramenta para o entendimento de sistemas e processos. Na prática, esse conhecimento oferece um caminho imediato para a melhoria (GOOS; JONES, 2011). O DOE tem sido aplicado em diferentes disciplinas. De fato, é uma visão científica de processos e um caminho para o estudo sobre como os sistemas ou processos funcionam. É uma importante ferramenta para a melhoria de desempenho em processos de manufatura, se estendendo ao desenvolvimento de novos processos. As aplicações das técnicas de DOE em desenvolvimento de processos podem resultar em: (MONTGOMERY, 2008)

- Melhoria do rendimento dos processos;
- Redução de variabilidade e aproximação dos resultados reais aos valores nominais ou requisitos alvo do processo;
- Redução do tempo de desenvolvimento;
- Redução dos custos externos.

Quando um modelo descreve um sujeito suficiente e precisamente, a experiência sobre o assunto geralmente é substituída por um experimento sobre o modelo. O modelo geralmente difere do assunto em suas dimensões e, às vezes, em sua natureza. A experiência pode também ser feita com um modelo matemático. Devido a um rápido desenvolvimento da tecnologia em informática, física, modelos são mais frequentemente substituídos por formulações matemáticas.

Um experimento tem um lugar central na ciência, devido à complexidade com que a ciência lida com os problemas. A questão de eficiência de utilização de um experimento é, portanto, imposta. Para aumentar a eficiência da pesquisa, é necessário introduzir algo completamente novo na pesquisa experimental clássica.

DOE é uma abordagem planejada para determinar as causas e relações de efeito. Wassef (1952) fez uso do DOE para melhorar aspectos ligados a qualidade de fotogrametria, que é a técnica de extrair fotografias métricas, utilizadas para cartografia, agrimensuras, entre outras aplicações. Ele testou os tipos de filmes, lentes, relacionando-os com níveis de temperatura em um período de 12 meses sendo que a saída do processo é qualitativa, dividindo a qualidade da foto em boa, ruim, normal, usual, etc. Sellier (1980) usou DOE no campo da genética, aplicando em estudos com porcos. Ele estimou os parâmetros genéticos dando especial atenção para a necessidade do reconhecimento do tamanho real das amostras em comparações com as raças desses animais e considerando a derivação na variância amostral das estimativas da mudança genética. Um exemplo de suas questões é: “A raça Pietrain é melhor do que a raça Large White em relação a eficiência no crescimento e desenvolvimento?”. Anderson e Thomas (1980) concentraram-se na utilização da quantidade mínima de experimentos para comprovar resultados significativamente positivos. Garcá-Ochoa *et al.* (1992) usaram projeto fatorial no campo da química. Com vinte experimentos, sendo quatro fatores com dois níveis e um ponto central, ele obteve a

comprovação estatística que seus fatores de estudo e suas interações afetam mais na sua resposta do que os erros dos experimentos.

Projetos de experimentos estão sendo desenvolvidos ao longo do tempo e cada vez mais sua complexidade é incrementada, tendo novos modelos surgindo e problemas, antes insolúveis sendo solucionados. Pázman e Müller (2001) incrementaram o projeto com restrições do tipo correlacionadas sem possibilidade de replicação com fatores de ruído. Pierlot *et al.* (2008) fizeram uma revisão do uso de delineamento de experimentos em aspersão térmica, relatando trabalhos focados nessa linha de estudo. Abbaszadeh *et al.* (2011) utilizaram para otimização do torque de motores indutivos, enquanto que Beránek e Kolařík (2014) implementaram para otimização da filtragem de partículas por decantação centrífuga de água. Há aplicações inclusive para otimização de modelos matemáticos. Chen e Li (2014) utilizaram delineamento de experimentos para otimizar interpolação baseada em metamodelos.

Um tipo de inovação pode ser aplicar métodos estatísticos ou matemáticos para desenvolver o projeto de experimentos. Por este meio, o que se segue é essencial: redução ou minimização do número total de ensaios; análise simultânea de variação de todos os fatores que formam as atividades do objeto experimentador; escolha de uma estratégia clara que permita soluções confiáveis para serem obtidas após cada sequência de experimentos (LAZIĆ, 2004).

É importante a definição das variáveis dos experimentos. Elas são classificadas como, x sendo aquelas de entrada e controláveis - podem ser o tipo de matéria prima, a corrente elétrica na solda, entre outros - ou fatores que possivelmente exercem influência sobre y que é o resultado ou saída do processo em estudo, pode ser o produto fabricado, os requisitos da qualidade, etc. Há ainda as variáveis que não são controláveis, também chamadas de ruído. Estas podem ser medidas e acompanhadas, porém seus níveis não podem ser alterados a contento, a chuva, o humor dos operadores são exemplos desses fatores identificados como z . A Figura 3.3 mostra o conceito de entradas e saídas.

Os principais objetivos dos experimentos podem ser descritos como: (MONTGOMERY *et al.*, 1997)

- Determinar quais variáveis mais influenciam na resposta y ;
- Determinar os valores de x de forma que y se aproxime mais do valor de interesse;
- Determinar os valores de x para que y tenha uma variação pequena;
- Determinar os valores de x para que os efeitos das variáveis incontroláveis z_1, z_2, \dots, z_q sejam minimizados.

Outros importantes conceitos chave citados por Goos e Jones (2011) podem ser listados como:

- Bons projetos de experimentos direcionam o experimentador para uma estimação precisa de uma ou mais não conhecidas variáveis de interesse, por exemplo, qual a quantidade, ou

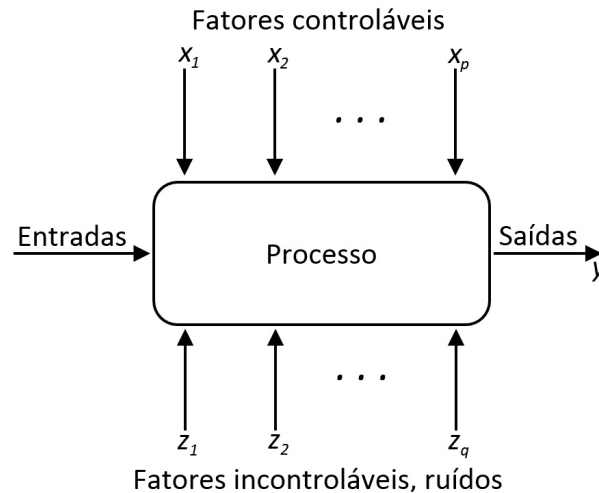


Figura 3.3 – Entradas e saídas do processo. Fonte: [Montgomery et al. \(1997\)](#)

parâmetro causa diferença na média de dois tipos de tratamento? Um parâmetro estimado é mais preciso quando tem uma variância menor;

- Projetos balanceados podem, às vezes, ser ótimos, mas este não é um caso genérico;
- Se dois projetos têm características diferentes, eles genericamente necessitam de projetos diferentes;
- A melhor maneira para alocar testes em um novo experimento é tratando as combinações com as maiores variâncias previstas. Isso pode parecer óbvio, mas é um princípio importante;
- A melhor alocação dos recursos do experimento pode depender do custo relativo de cada corrida, o custo de diferentes combinações deve ser considerado.

3.2.1 Processo de delineamento do experimento

Muitos experimentos científicos podem ser divididos em três estágios: planejamento do experimento; implementação do experimento e; análise e interpretação dos dados coletados. Quando o estágio de planejamento envolve DOE, a implementação implica na utilização de esquemas aleatorizados com possíveis blocagens e quando o estágio de interpretação utiliza análise estatística dos dados, esses estágios compilam o processo de delineamento do experimento ([SMITH, 2005](#)).

No curso do planejamento experimental, uma decisão principal é qual condição deve ser variada (fatores) e qual resposta ou respostas serão medidas. Então uma hipótese que qualquer resposta a ser observada e medida está relacionada com os níveis ou valores de um ou mais fatores. Isto pode ser representado analiticamente na Equação (3.1). Na qual: y é a resposta relacionada à função dos níveis dos fatores x_i e que no caso de misturas

será a proporção de cada componente w_i .

$$y = f(x_i) \quad (3.1)$$

A formulação pode ser entendida como uma mistura composta por um ou mais componentes que podem ser proporções, estas são não independentes entre si - se a proporção de um componente é aumentada, então a proporção de um ou mais componentes decai se o total dos pesos ou porções da mistura continua sendo a mesma. As proporções da mistura dos componentes podem ser consideradas como níveis dos fatores, porém a palavra *fatores* é usualmente reservada para variáveis de não mistura que oferecem, embora não necessariamente, independência entre si. Fatores são chamados de variáveis do processo, por exemplo o tempo, temperatura, pressão, velocidade de revestimento, etc.

Quando testes de significância são usados, é importante não confundir os testes de hipóteses estatísticos com os testes de hipóteses científicos. Os testes científicos vão além dos conjuntos de dados e fazem declarações genéricas sobre o Mundo ([ANDERSSON, 2012](#)). Eles podem ser a base para uma explanação teórica, se envolverem mecanismos de qualquer natureza. Testes estatísticos, no entanto, são não exploratórios. Eles podem ser feitos com uma simples declaração sobre relacionamentos matemáticos. Estes sugerem que a média de uma amostra é maior do que outra, mas faz declarações sobre a razão por trás dessa diferença.

3.2.2 Projeto com um fator categórico

Experimentos projetados com base em uma condição específica que investiga os efeitos em uma saída específica (y). O interesse pode ser saber como tipos de tratamento de superfície afetam as propriedades de certo material, como o tempo de cozimento afeta no sabor de um alimento, como um certo medicamento interfere na condição física de um paciente, entre outros. As variáveis de entrada (x_i) são chamadas de fatores categóricos desde que descrevam estados que possam ser rotulados, mas não ordenados ou regulados uns em relação aos outros. Se a variação natural nas medições das variáveis for grande e as amostras forem pequenas, pode haver dificuldade no discernimento do que está realmente afetando a saída e o que é simples ruído do processo. Em tais casos é necessário um projeto que considere os fatores de ruído. Isto é muito comum em estudos com medicamentos, por exemplo, os pacientes são afetados por uma grande variedade de fatores sociais, de meio ambiente e outros indiretos que não estão sob controle do experimento ([ANDERSSON, 2012](#)). Experimentos controlados são usados para contornar o problema citado anteriormente e é útil também para identificar tais fatores.

Para minimizar o risco de que outros fatores de ruído possam influenciar no resultado, as amostras devem ser alocadas aleatoriamente para diferentes grupos de resultados.

Tais experiências são referidas como experimentos controlados e aleatorizados. Os dados são analisados utilizando um teste t de duas amostras ou, se o procedimento for generalizado para mais do que uma resposta, utilizando a análise de variância (ANOVA). É importante a utilização de tamanhos de amostras suficientes para diminuir o risco de erros de Tipo II que é aquele que aparece no caso da análise não rejeitar a hipótese nula, quando esta é falsa. Um tamanho de amostra apropriado é encontrado por análise do poder das amostras.

Um bom exemplo de experimento com uma categoria de fatores é a investigação de um medicamento em pacientes, neste caso os pacientes são separados em dois grupos, um que está tomando o medicamento alvo do estudo e outro que está tomando placebo. Pode-se daí investigar os efeitos do medicamento real nos pacientes e em contra-partida investigar o efeito de variáveis do tipo ruído nos pacientes que estão tomando placebo, já que os pacientes que tomam placebo não são afetados por componentes químicos do medicamento alvo do experimento. Mainardi *et al.* (2014) estudaram e avaliaram a eficácia e tolerabilidade do uso de duas doses de trometamol dexketoprofeno (DKP) em comparação com o placebo para o tratamento da enxaqueca e chamou de *ataque* cada rodada do experimento. A enxaqueca é uma doença incapacitante que pode afetar significativamente a qualidade de vida de uma pessoa. É interessante explorar um pouco esse trabalho utilizando algumas tabelas. O Quadro 3.1 mostra a sequência aleatorizada para os experimentos e a Tabela 3.1 é a demonstração de como algumas características foram escolhidas e controladas para o experimento.

Quadro 3.1 – Sequência pré-definida do tratamento

	Ataque 1	Ataque 2	Ataque 3
Grupo 1	DKP 25 mg	DKP 50 mg	Placebo
Grupo 2	Placebo	DKP 25 mg	DKP 50 mg
Grupo 3	DKP 50 mg	Placebo	DKP 25 mg

Tabela 3.1 – Características básicas dos pacientes

	Placebo (N = 75)	DKP 25 mg (N = 74)	DKP 50 mg (N = 74)
Características demográficas			
Mulheres	46	45	45
Homens	29	29	29
Idade (anos)	40,5 ± 10,9	40,5 ± 11,0	40,5 ± 11,0
Peso (kg)	70,4 ± 13,7	67,8 ± 13,6	67,8 ± 13,6
Altura (cm)	168,6 ± 9,0	168,8 ± 9,0	168,8 ± 9,0
Ataques por mês	3,7 ± 1,3	3,7 ± 1,3	3,7 ± 1,3
Duração dos ataques (h)	15,4 ± 13,5	15,8 ± 14,0	15,8 ± 14,0

A mudança no tipo de tratamento de cada grupo em cada ataque comprova que o experimento é também do tipo cruzado. A Figura 3.4 exemplifica o caso (ANDERSSON, 2012).

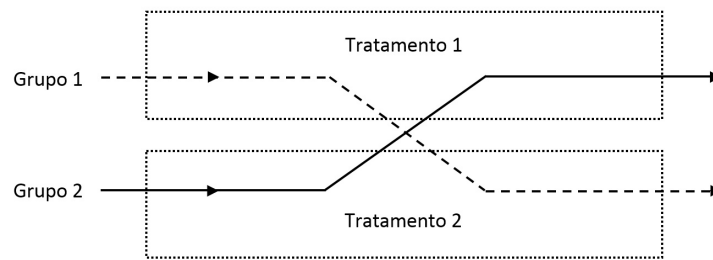


Figura 3.4 – Descrição esquemática de experimento cruzado

3.2.3 Determinação do Erro Experimental

Um pesquisador que pretende aplicar o projeto de experimentos deve ter uma ideia clara sobre os métodos de processamento dos resultados experimentais, no caso real e antes de colocar em funcionamento, de modo a facilitar a pesquisa que define o objetivo e na elaboração das conclusões. A teoria matemática de experimentos diferencia vários tipos de erros no processamento de resultados experimentais, cada um dos quais, é característico de uma fase definitiva da análise dos dados.

Cada experiência é constituída por certo número de pontos experimentais do projeto, cada projeto do ponto experimental com um ou mais ensaios replicados ou não, e um ponto único de uma ou mais medidas replicadas (determinações). De acordo com esta experiência podemos distinguir os erros (variância da reprodutibilidade). O teste da variância dos erros dos ensaios replicados e dos erros de medição será o erro de determinação. O erro de medição é local para um experimento, e pode vir, ou de um instrumento, ou de uma diferença amostral (erro de amostragem) ou ambos. Ele pode ser reduzido fazendo-se repetidas medições e calculando-se as médias destas medições. Tais médias dependem da sua magnitude em relação ao erro experimental.

O desvio padrão do erro de medição pode ser estimado a partir de medições repetidas dentro do mesmo ensaio. A variância da medição é o quadrado do desvio-padrão do erro das medições. O reconhecimento do ponto de julgamento do projeto e o erro experimental são necessários para uma verificação da significância dos coeficientes de regressão e a falta de ajuste de um modelo de regressão. O erro experimental de tentativa e erro também pertence ao grupo de erros aleatórios de modo que ao se estimarem os seus valores usamos a mesma abordagem para a medição aleatória dos erros.

A fim de determinar um erro de medição deve-se levar em conta o número de medições replicadas (u), o número de tentativas repetidas (n), e em um erro experimental, número de ensaios diferentes (N). A replicação de um ensaio não deve ser confundida com a replicação de medições. Ao determinar um erro experimental, estima-se o desvio padrão dos ensaios replicados. Isso pode ser estimado pelo cálculo do desvio padrão dos vários ensaios cujo fator de controle, configurações, são os mesmos. O ideal seria criar e executar o mesmo processo várias vezes. As pequenas diferenças em cada configuração são um

componente importante do erro experimental.

O erro de replicação é composto de duas partes: o erro do experimento e o erro de medição. A variância da réplica é o quadrado do desvio padrão replicado. No caso de erro experimental estimamos a variância de reprodutibilidade. Antes do cálculo do erro experimental é necessário verificar a homogeneidade da variância em diferentes ensaios (LAZIC, 2004).

3.2.4 Projeto de experimentos de Mistura

Projetos de experimentos de mistura fazem parte de uma classe especial de experimentos do tipo superfície de resposta (RSM) nos quais o produto sob investigação é composto por proporções definidas de outros elementos. Estes experimentos são muito úteis porque, em inúmeras aplicações industriais, existem atividades ou processos que envolvem algum tipo de formulação, composto ou mistura. O sistema triangular de coordenadas mostrado na Figura 3.5 permite que os relacionamentos entre os componentes de uma mistura, em um exemplo de três elementos, sejam visualizados. MDE é um tipo especial de experimento de superfície de resposta, em que os fatores são proporções dos componentes de uma mistura (CORNELL, 2002; MYER; MONTGOMERY, 2002). Ao avaliar a resposta experimental para a combinação de vários fatores, pode-se estimar a superfície de resposta do processo em análise. O projeto é escolhido de tal forma que os pontos se espalham ao longo do espaço experimental viável. No MDE, em particular, este espaço é conhecido como espaço *simplex*. Além disso, o *Centroid simplex* e *simplex-lattice* são amplamente utilizados na análise de experimentos MDE. A fim de se obter um modelo polinomial de, pelo menos, grau m de componentes q , deve-se ter $m + 1$ valores igualmente espaçados, para cada componente da mistura. A Figura 3.8 mostra um *simplex-lattice* q, m combinado, em que $q = 3$ e $m = 2$. Em uma mistura, os componentes se restringem mutuamente, uma vez que a sua soma deve ser igual à unidade. Este sistema, portanto, permite que os valores mínimos e máximos dos três elementos da mistura x_1, x_2 e x_3 sejam mostrados. Os componentes são representados em função de sua proporção em relação ao total, e esse total é igual a 100%. Cada vértice do triângulo representa o que se denomina “Mistura Pura” (MONTGOMERY, 2008).

Os pontos médios localizados nos lados do triângulo representam uma mistura na qual dois componentes estão presentes, cada qual na proporção de 50%. Quando os lados forem seccionados em três partes, a mistura terá 2/3(66.7%) de um dos componentes e 1/3(33.3%) do outro. No centróide, a mistura contém exatamente 1/3 de cada um dos três elementos. Há três arranjos específicos de mistura: o arranjo *simplex-centróide* (experimento no qual os componentes sempre aparecerem em iguais proporções), o arranjo *simplex-lattice* e o arranjo de vértices restritos ou *extreme-vertice*. O nome *lattice* é usado como referência à grade de pontos de uma matriz.

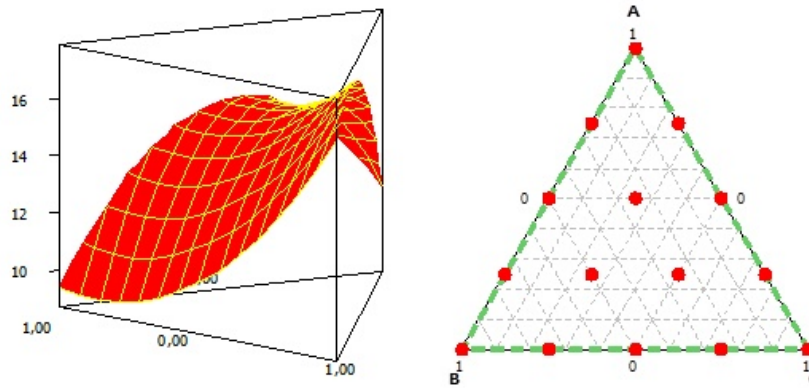


Figura 3.5 – Arranjo de misturas e superfície de resposta para misturas

O *simplex-centróide* é um arranjo que inclui observações para as misturas nas quais os componentes sempre aparecem em iguais proporções. Este arranjo compreende: 2^{q-1} pontos, correspondentes às q permutações de $(1, 0, \dots, 0)$ ou à mistura de q componentes puros; permutações de $(0.5, 0.5, 0, \dots, 0)$ ou todas as misturas binárias; permutações de $(1/3, 1/3, 1/3, 0, \dots, 0)$ ou todas as misturas ternárias, e por fim, o centróide $(1/q, 1/q, \dots, 1/q)$.

A principal característica do arranjo *simplex-lattice* é que os pontos são distribuídos uniformemente por toda a região compreendida pelo *simplex*. Um *simplex-lattice* para q componentes está associado a um polinômio de grau m , e pode ser denotado como um arranjo *simplex-lattice* q, m . As proporções para cada um dos q componentes assumem os possíveis $(m + 1)$ valores igualmente espaçados entre 0 e 1. A Figura 3.6 apresenta um exemplo de um arranjo *simplex-lattice* para 3, n componentes. Nota-se que este tipo de arranjo é restrito por fronteiras, isto é, todos os pontos do arranjo estão nas faces da região *simplex*. A Figura 3.7 mostra o *simplex-lattice* 4, n . Para se pesquisar o comportamento da mistura no interior da região *simplex*, o arranjo pode ser incrementado com pontos interiores em seu interior.

O número de experimentos requeridos por arranjo do tipo *simplex-lattice* q, m que possui q componentes e que deve ser construído para permitir o ajuste de um polinômio canônico do m -ésimo grau e é dado pela Equação (3.2).

$$N = \frac{(q + m - 1)!}{m!(q - 1)!} \quad (3.2)$$

Cada arranjo apresentado tem uma determinada aplicação e permite a estimação de um modelo polinomial diferente. Na Figura 3.2 é apresentado uma relação entre os principais arranjos e os respectivos modelos que são possíveis de serem construídos com pontos experimentais projetados em cada um deles.

Scheffé (1958) considera experimentos com misturas (MDE) aqueles cujas pro-

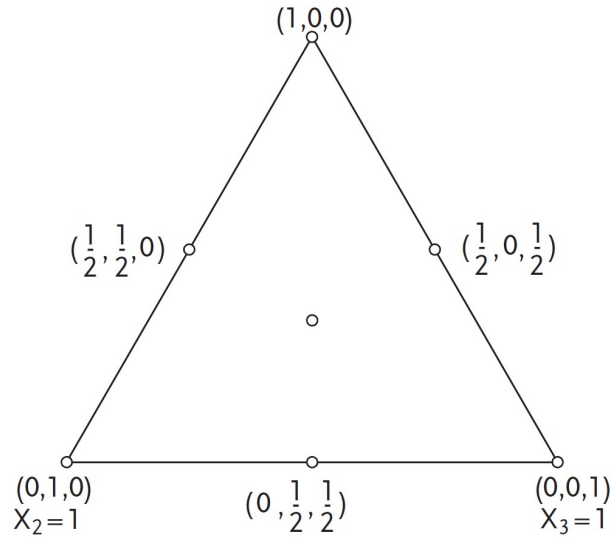


Figura 3.6 – Simplex-Centróide

Fonte: Lazić (2004)

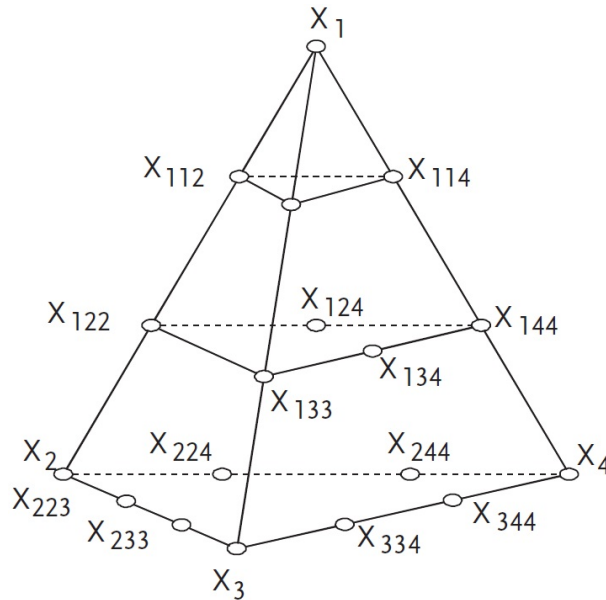


Figura 3.7 – Simplex-Lattice 4,n para polinômio de grau 3

Fonte: Lazić (2004)

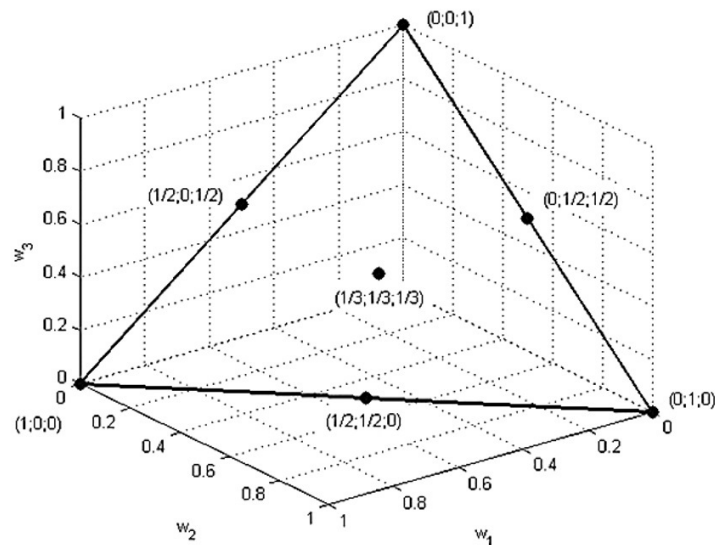
priedades estudadas são dependentes das proporções dos componentes presentes na sua composição, mas não necessariamente do montante da mistura. Em uma mistura de q componentes ($q \geq 3$) considere x_i uma proporção. A proporção pode ser denotada por volume, pesos, percentuais, etc. dos i s componentes na mistura. Diz-se mistura pelo fato da região dimensional ficar dentro do intervalo entre 0 e 1, na qual 0 significa ausência total de certo componente, enquanto 1 significa que a composição é de somente um componente. Quando se estuda as propriedades de um q -componente da mistura dependente apenas de seu percentual, o espaço fatorial é regular, $(q - 1)$ simples e o relacionamento da mistura pode ser escrito como na Equação (3.3), (LAZIĆ, 2004). Na qual X_i é a concentração

Quadro 3.2 – Arranjos de misturas e modelos polinomiais associados

TIPO DE SIMPLEX		MODELO DE AJUSTE POSSÍVEL	
		SIMPLES	INCREMENTADO
CENTRÓIDE		Cúbico Especial	Ajuste parcial em Cúbico Completo
LATTICE	GRAU 1	Linear	Quadrático Parcial
	GRAU 2	Quadrático	Cúbico Especial Parcial
	GRAU 3	Cúbico Completo	Cúbico Completo

percentual do componente no produto e q é o número de componentes na mistura.

$$\sum_{i=1}^q X_i = 1; 0 \leq X_i \leq 1 \quad (3.3)$$

Figura 3.8 – Arranjo *simplex-lattice*

O espaço representado na Figura 3.8 pelo experimento de mistura para os componentes do sistema de coordenadas descreve o *simplex*. Os vértices da região convexa representam a mistura pura; os pontos dentro da região são misturas em que nenhum dos componentes está ausente. O centroide é a mistura com proporções iguais de cada componente $(1/3, 1/3, 1/3)$.

Uma vez que os experimentos são realizados, o MDE permite estabelecer a relação entre as variáveis de resposta e a proporção relativa dos componentes em termos de uma equação matemática. Ele fornece a identificação da influência da proporção de cada elemento e as interações entre fatores sobre a variável de resposta. Geralmente, a relação funcional entre a variável de resposta e as proporções dos elementos n é definida por

um polinômio de grau m , que pode ser estimado com a ajuda de uma regressão linear, quadrática, cúbica ou dependendo dos objetivos do praticante. As propriedades da mistura podem ser expressas por um polinômio com grau definido, de forma que as variáveis independentes X_1, X_2, \dots, X_q sendo X_i os percentuais de cada i -ésimo componente na mistura. O polinômio de grau n com q variáveis terá $\binom{n}{q} + n$ coeficientes visto na Equação (3.4).

$$\hat{Y} = b_0 + \sum_{1 \leq i \leq q} b_i X_i + \sum_{1 \leq i \leq j \leq q} b_{ij} X_i X_j + \sum_{1 \leq i \leq j \leq k \leq q} b_{ijk} X_i X_j X_k + \sum b_{i_1, i_2, \dots, i_n} X_{i_1} X_{i_2} \dots X_{i_n} \quad (3.4)$$

A partir da redução do polinômios da Equação (3.4) é sujeita a normalização da condição imposta pela Equação (3.3) através da soma das variáveis independentes foi sugerido por Scheffé (1958). Lazić (2004) demonstrou como proceder a redução do polinômio de segundo grau devisado em termos de sistema. Seu polinômio generalizado tem a forma descrita na Equação (3.5). Considerando que em (3.5)

$$X_1 + X_2 + X_3 = 1 \quad (3.5)$$

Tem-se a Equação (3.6)

$$b_0 X_1 + b_0 X_2 + b_0 X_3 = b_0 \quad (3.6)$$

Multiplicando a Equação (3.5) por X_1 , X_2 e X_3 sucessivamente teremos a Equação (3.7)

$$X_1^2 = X_1 - X_1 X_2 - X_1 X_3 \quad X_2^2 = X_2 - X_1 X_2 - X_2 X_3 \quad X_3^2 = X_3 - X_1 X_3 - X_2 X_3 \quad (3.7)$$

Substituindo as Equações (3.6) e Equação (3.7) na Equação (3.5) obtém-se a Equação (3.8) necessária para a transformação e em forma de notação na Equação (3.9).

$$\begin{aligned} \hat{Y} = & (b_0 + b_1 + b_{11})X_1 + (b_0 + b_2 + b_{22})X_2 + (b_0 + b_3 + b_{33})X_3 \\ & + (b_{12} - b_{11} - b_{22})X_1 X_2 + (b_{13} - b_{11} - b_{33})X_1 X_3 \\ & + (b_{23} - b_{22} - b_{33})X_2 X_3 \end{aligned} \quad (3.8)$$

$$\beta_i = b_0 + b_i + b_{ii}; \quad \beta_{ij} = b_{ij} - b_{ii} - b_{jj} \quad (3.9)$$

Finalmente é possível a redução do polinômio de segunda ordem em três variáveis (Equação (3.10)). Em casos gerais, para um sistema de q -componentes, diferentes graus de regressão podem ser modelados.

$$\hat{Y} = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 \quad (3.10)$$

Esses coeficientes podem ser calculados com a ajuda do algoritmo de mínimos quadrados ordinário ou abordagem de máxima verossimilhança. A escolha do mais adequado

dependerá do tipo de ajuste para o modelo de regressão que o experimentador desejar, ou aquele que tiver o maior R^2 . O arranjo denominado “incrementado” é aquele que recebe, adicionalmente, pontos axiais interiores e centróides. Além disso, qualquer um dos pontos de um arranjo para mistura podem ser replicados para que se tenha uma estimativa do ajuste ou do erro puro. Os modelos de mistura são ligeiramente diferentes dos modelos gerados pelos demais arranjos DOE, tais como os fatoriais completos ou superfície de resposta.

O espaço fatorial é uma dimensão regular do *simplex* ($q - 1$) sendo uma reta para $q = 2$, um triângulo para $q = 3$, um tetraedro para $q = 4$. Diferentemente dos arranjos fatoriais e dos arranjos de superfície de resposta, neste não há termo independente isolado no modelo polynomial canônico. A constante β_0 será adicionada em todos os termos lineares neste caso. Para um modelo de segunda ordem, tem-se a Equação (3.11) na qual considerando-se um modelo quadrático sem restrições na Equação (3.12) e substituindo a expressão relativa a x_i^2 na Equação (3.13). O modelo quadrático pode ser escrito pela Equação (3.14).

$$\sum_{i=1}^q = 1 \Rightarrow x_i = \left(1 - \sum_{\substack{j=1 \\ j \neq i}}^q x_j \right) \Rightarrow x_i^2 = x_i \left(1 - \sum_{\substack{j=1 \\ j \neq i}}^q x_j \right) \quad (3.11)$$

$$E[f(x)] = \beta_0 = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^q \beta_{ii} x_i^2 + \sum_{i < j}^q \beta_{ij} x_i x_j \quad (3.12)$$

$$E[f(x)] = \beta_0 \left(\sum_{i=1}^q x_i \right) + \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^q \beta_{ii} \left[x_i \left(1 - \sum_{\substack{j=1 \\ j \neq i}}^q x_j \right) \right] + \sum_{i < j}^q \beta_{ij} x_i x_j \quad (3.13a)$$

ou

$$E[f(x)] = (\beta_0 + \beta_i) \left(\sum_{i=1}^q x_i \right) + \sum_{i=1}^q \beta_{ii} \left[\left(x_i - x_i \sum_{\substack{j=1 \\ j \neq i}}^q x_j \right) \right] + \sum_{i < j}^q \beta_{ij} x_i x_j \quad (3.13b)$$

simplificado em

$$E[f(x)] = (\beta_0 + \beta_i + \beta_{ii}) \left(\sum_{i=1}^q x_i \right) - \sum_{i=1}^q \beta_{ii} x_i \sum_{\substack{j=1 \\ j \neq i}}^q x_j + \sum_{i < j}^q \beta_{ij} x_i x_j \quad (3.13c)$$

$$E[f(x)] = \sum_{i=1}^q \beta_i^* x_i + \sum_{i < j}^q \beta_{ij}^* x_i x_j \quad \text{com} \quad \begin{cases} \beta_i^* = \beta_0 + \beta_i + \beta_{ii} \\ \beta_{ij}^* = \beta_{ij} - \beta_{ii} - \beta_{jj} \end{cases} \quad (3.14)$$

O coeficiente da parte linear do modelo quadrático é uma resultante composta por constante, efeitos lineares e efeitos quadráticos. O coeficiente resultante da parte não linear (que se assemelha a uma interação), contém os efeitos sinérgicos ou antagônicos de dois componentes descontando-se os efeitos quadráticos de cada componente envolvido na pseudo-interação. Na Equação (3.15) está apresentado um resumo dos principais modelos possíveis de serem criados com arranjos *simplex-lattice* ou centróide.

$$E(y) = \sum_{i=1}^q \beta_i x_i \quad (3.15a)$$

$$E(y) = \sum \beta_i x_i + \sum_{i < j} x_i x_j \quad (3.15b)$$

$$E(y) = \sum \beta_i x_i + \sum_{i < j} x_i x_j + \sum_{i < j < k} x_i x_j x_k \quad (3.15c)$$

$$E(y) = \sum \beta_i x_i + \sum_{i < j} x_i x_j + \sum_{i < j} \delta_{ij} x_i x_j (x_i - x_j) + \sum_{i < j < k} x_i x_j x_k \quad (3.15d)$$

Para exemplificar a estimação desses coeficientes, considere-se um modelo quadrático completo para 3 componentes conforme Figura 3.9 tendo os dados preenchidos na Tabela 3.2. Este modelo pode ser escrito pela Equação (3.16).

$$y(x) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 \quad (3.16)$$

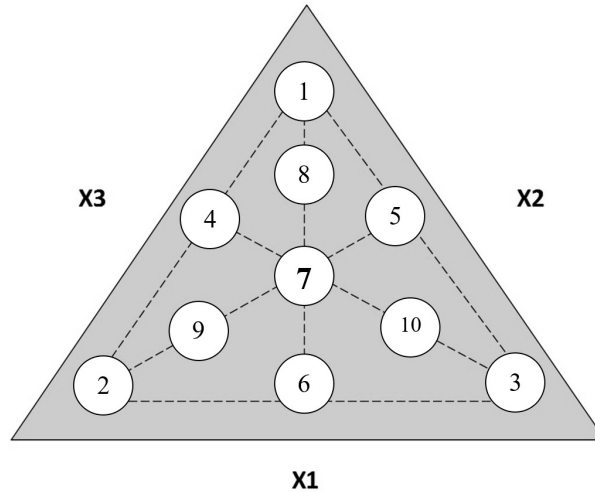


Figura 3.9 – Esquema experimental para um arranjo de misturas do tipo *Extreme-Vertex*

Tabela 3.2 – Esquema experimental para arranjo de misturas do tipo *Extreme-Vertice*

Experimento	Pseudocomponentes			Proporções			Resposta
	X1	X2	X3	X1	X2	X3	Y
1	0	0	1	0,1000	0,1000	0,8000	35,0
2	1	0	0	0,8000	0,1000	0,1000	18,9
3	0	1	0	0,1000	0,8000	0,1000	15,2
4	0	1/2	1/2	0,1000	0,4500	0,4500	31,2
5	1/2	0	1/2	0,4500	0,1000	0,4500	18,9
6	1/2	1/2	0	0,4500	0,4500	0,1000	16,1
7	1/3	1/3	1/3	0,3333	0,3333	0,3333	19,3
8	1/6	1/6	2/3	0,2167	0,2167	0,5667	30,1
9	2/3	1/6	1/6	0,5667	0,2167	0,2167	18,2
10	1/6	2/3	1/6	0,2167	0,5667	0,2167	17,7

Substituindo cada um dos 7 pontos do arranjo *simplex* centróide, tem-se a Equação (3.17).

$$\left\{ \begin{array}{ll} P/x_1 = 1,0 & x_2 = 0,0 & x_3 = 0,0 \Rightarrow y_1 = \beta_1 \\ P/x_1 = 0,0 & x_2 = 1,0 & x_3 = 0,0 \Rightarrow y_2 = \beta_2 \\ P/x_1 = 0,0 & x_2 = 0,0 & x_3 = 1,0 \Rightarrow y_3 = \beta_3 \\ P/x_1 = 0,5 & x_2 = 0,5 & x_3 = 0,0 \Rightarrow y_{12} = 0,50\beta_1 + 0,50\beta_2 + 0,25\beta_{12} \\ P/x_1 = 0,5 & x_2 = 0,0 & x_3 = 0,5 \Rightarrow y_{13} = 0,50\beta_1 + 0,50\beta_3 + 0,25\beta_{13} \\ P/x_1 = 0,0 & x_2 = 0,5 & x_3 = 0,5 \Rightarrow y_{23} = 0,50\beta_2 + 0,50\beta_3 + 0,25\beta_{23} \end{array} \right. \quad (3.17a)$$

Sendo então:

$$\left\{ \begin{array}{l} y_{12} = 0,50\beta_1 + 0,50\beta_2 + 0,25\beta_{12} = \frac{2(\beta_1 + \beta_2) + \beta_{12}}{4} \\ y_{13} = 0,50\beta_1 + 0,50\beta_3 + 0,25\beta_{13} = \frac{2(\beta_1 + \beta_3) + \beta_{13}}{4} \\ y_{23} = 0,50\beta_2 + 0,50\beta_3 + 0,25\beta_{23} = \frac{2(\beta_2 + \beta_3) + \beta_{23}}{4} \end{array} \right. \quad (3.17b)$$

Genericamente escrito em:

$$4y_{ij} = 2(\beta_i + \beta_j) + \beta_{ij} \Rightarrow \beta_{ij} = 4y_{ij} - 2(\beta_i + \beta_j) = 4y_{ij} - 2(y_i + y_j) \quad (3.17c)$$

Caso o arranjo contenha r_i, r_j, r_{ij} réplicas, respectivamente, então:

$$\beta_{ij} = 4\bar{y}_{ij} - 2(\bar{y}_i + \bar{y}_j) \quad (3.17d)$$

Deste modo é possível estimar todos os coeficientes do modelo de misturas. A variância dos coeficientes dos modelos obtidos por arranjos de misturas é escrito na Equação (3.18) sendo analogamente para os termos não lineares escrita na Equação (3.19).

$$Var(\beta_i) = Var\left(\frac{y_i}{r_i}\right) = \left(\frac{1}{r_i^2}\right) Var(y_i) = \left(\frac{1}{r_i^2}\right) Var(\sigma_{y_i}^2) \quad (3.18)$$

$$Var(\beta_{ij}) = Var[4\bar{y}_{ij} - 2(\bar{y}_i + \bar{y}_j)] = 16Var(\bar{y}_{ij}) + 4Var(\bar{y}_i + \bar{y}_j)$$

$$Var(\beta_{ij}) = 16Var\left(\frac{y_{ij}}{r_{ij}}\right) + 4Var\left(\frac{y_i}{r_i} + \frac{y_j}{r_j}\right) \quad (3.19)$$

$$Var(\beta_{ij}) = \frac{16}{r_{ij}}\sigma^2 + 4\sigma^2\left(\frac{1}{r_i} + \frac{1}{r_j}\right)$$

O modelo *simplex-lattice* é referido como $\{q, m\}$ e consiste em pontos cujas coordenadas são combinações dos componentes, as proporções que cada componente assume é $m + 1$ com valores igualmente espaçados de 0 a 1: (SCHEFFÉ, 1958; CORNELL, 2002). E $\{q, m\}$ consiste na combinação de todas as possíveis proporções da mistura. Para entender melhor, veja o exemplo: Suponha que uma xícara de café com leite contenha uma quantidade de café em mililitros, uma quantidade de leite também em mililitros e uma quantidade de açúcar em miligramas. Neste caso $q = 3$ componentes. Pode-se assumir que cada componente tem a proporção $x_i = 0, \frac{1}{2}, 1$ para $i = 1, 2, 3$. Definindo $m = 2$ na Equação (3.20), o modelo que representa a superfície de resposta sobre o triângulo será de segunda ordem. O *simplex-lattice* $\{3, 3\}$ terá seis pontos nas fronteiras do triângulo conforme a Equação (3.21).

$$x_i = 0, \frac{1}{m}, \frac{2}{m}, \dots, 1 \quad (3.20)$$

$$(x_1, x_2, x_3) = (1, 0, 0), (0, 1, 0), (0, 0, 1), \left(\frac{1}{2}, \frac{1}{2}, 0\right), \left(\frac{1}{2}, 0, \frac{1}{2}\right), \left(0, \frac{1}{2}, \frac{1}{2}\right) \quad (3.21)$$

A função polinomial no *simplex* de grau n em q variáveis x_1, x_2, \dots, x_q sujeito a Equação (3.3) é chamado de $\{q, n\}$ polinomial e terá a forma apresentada na Equação (3.22). Na qual β é uma constante e o número de coeficientes da Equação (3.22) é $\binom{n+q}{n}$. Porque a função da Equação (3.22) significa que será apenas restrita na Equação (3.3). Os coeficientes β não são únicos. Ainda de acordo com (SCHEFFÉ, 1958) pode-se, e.g., usar a substituição da Equação (3.23) na Equação (3.22) e a mesma função aparecerá como uma função polinomial em $q - 1$ variáveis x_1, x_2, \dots, x_{q-1} . Isto envolve apenas os coeficientes $\binom{n+q-1}{n}$. O número de pontos será o mesmo no *lattice* $\{q, n\}$.

$$\eta = \beta_0 + \sum_{1 \leq i \leq q} \beta_i x_i + \sum_{1 \leq i \leq j \leq q} \beta_{ij} x_i x_j + \sum_{1 \leq i \leq j \leq k \leq q} \beta_{ijk} x_i x_j x_k + \sum_{1 \leq i_1 \leq i_2 \leq \dots \leq i_n \leq q} \beta_{i_1 i_2 \dots i_n} x_{i_1} x_{i_2} \dots x_{i_n}, \quad (3.22)$$

$$x_q = 1 - \sum_{i=1}^{q-1} x_i \quad (3.23)$$

A generalização do modelo polinomial de grau n em q variáveis pode ser escrita como na Equação (3.24). Na qual a notação $\gamma_{ij}x_i x_j(x_i - x_j)$ torna a forma $\beta_{ii}x_i^2 x_j + \beta_{ijj}x_i x_j^2$ com $\beta_{ii} + \beta_{ijj} = 0$ mais simétrica.

$$\eta = \sum_{1 \leq i \leq q} \beta_i x_i \quad (3.24a)$$

se $n = 1$,

$$\eta = \sum_{1 \leq i \leq q} \beta_i x_i + \sum_{1 \leq i < j \leq q} \beta_{ij} x_i x_j \quad (3.24b)$$

se $n = 2$, e se $n = 3$

$$\eta = \sum_{1 \leq i \leq q} \beta_i x_i + \sum_{1 \leq i < j \leq q} \beta_{ij} x_i x_j + \sum_{1 \leq i < j \leq q} \gamma_{ij} x_i x_j (x_i - x_j) + \sum_{1 \leq i < j < k \leq q} \beta_{ijk} x_i x_j x_k \quad (3.24c)$$

Restrições podem ser também aplicadas em cada componente que se faça necessário. O ponto central continuará recíproco ao número de componentes, sendo $w_i = 1/n$. Há um grande número de aplicações em MDE. Algumas podem ser listadas começando por (SCHEFFÉ, 1958) e continuando em (SCHEFFÉ, 1963) que usou o centróide como ponto central. (BRANDVIK; DALING, 1998) promoveu otimização em desenvolvimento de produtos. (MÅGE; NÆS, 2005) usou MDE com variáveis de processo, ou arranjo fracionado.

Há também diversos trabalhos nos quais os autores utilizaram o conceito original de MDE, i.e., utilizaram MDE para análise e otimização das proporções dos ingredientes em produtos (TEHRANI *et al.*, 2012; LONNI *et al.*, 2012; NGUN *et al.*, 2014) com e sem restrições.

O MDE suporta restrições globais e individuais. Considerando uma mistura com resolução V , sendo η qualquer número natural inteiro, Ψ o valor máximo da mistura e n o número de ativos em um grupo ou portfólios em uma mistura global. Quando V diminui, o número de corridas no experimento aumenta. A Tabela 3.3 é a demonstração das proporções w_1, w_2 na qual $V = 0, 2$ e $n = 2$.

Tabela 3.3 – Proporções MDE 2 componentes, resolução de 0,2

Proporções	
w_1	w_2
0.000	1.000
0.200	0.800
0.400	0.600
0.600	0.400
0.800	0.200
1.000	0.000

3.2.5 Arranjo de mistura cruzado

O arranjo de mistura cruzado (CM-DOE) foi introduzido a partir da combinação de duas misturas diferentes para gerar uma região de interesse como uma terceira mistura (CORNELL, 1971), e depois disso como um experimento de mistura de dois estágios (CORNELL; RAMSEY, 1997). O CM-DOE também é conhecido como mistura de mistura (MoM) (PIEPEL, 1999; BORGES *et al.*, 2007; ZIO *et al.*, 2007; KANG *et al.*, 2011). Este é um tipo de experimento usado para análise e otimização de produtos, principalmente quando existem muitas restrições nas proporções da mistura (KANG *et al.*, 2011).

Cornell e Ramsey (1997) identificaram uma maneira de separar o problema de MoM em duas classes de componentes que chamaram de maiores e menores componentes. Os maiores são aqueles de estágios mais altos na mistura enquanto que os menores são estágios mais baixo ou subgrupos. Exemplificando tem-se a Figura 3.10 com o triângulo maior ao centro como maior componente e os demais que o circundam como subgrupos, menores componentes. Também é possível observar que não há limitações de quantos níveis ou misturas menores podem ser analisadas.

Quando pensamos em variáveis de processo incluídas em arranjos de misturas, é possível visualizar algumas das equações mostradas em Cornell (1971), que define uma função linear dos componentes da mistura com variáveis de processo como w_i ($1 \leq i \leq k - q$). O w_i corresponde a um x_i em uma categoria particular contendo n_l componentes definidos detalhadamente na Equação (3.25).

$$w_i = \frac{v_{S_l} + i + 1 \left(\sum_{j=1}^i h_{S_l}^2 + j \right) - \left[\sum_{j=1}^i v_{S_l} + j h_{S_l} + j \right] h_{S_l} + i + 1}{\left[\left(\sum_{j=1}^i h_{S_l}^2 + j \right) \left(\sum_{j=1}^{i+1} h_{S_l}^2 + j \right) \right]^{0.5}} \quad (3.25)$$

$(i = 1, \dots, n_l - 1)$

Há um modelo de DOE que combina o arranjo de misturas com variáveis de processo. Cornell (2002) descreveu este tipo de modelo como um arranjo do tipo *simplex-lattices* e arranjos fatoriais. No seu caso, havia um experimento de misturas com q componentes acrescidos de n variáveis de processo, na qual esta última poderia ser qualquer n inteiro positivo. A Figura 3.11 foi feita considerando um arranjo de mistura simples composto por três componentes x_n (os triângulos) e três fatores, variáveis de processo como um arranjo de componente central (CCD) Z_n (o cubo). Enquanto que a Figura 3.12 considera arranjos fatoriais dentro de um arranjo de misturas.

A característica ótima de um produto pode ser encontrada a partir de um modelo que combine simultaneamente variáveis de processo e proporções dos componentes no MDE. Como demonstrado na Equação (3.26) (MONTGOMERY, 2008). Os termos que envolvem apenas variáveis de processo não são incluídos no modelo. Ambos, MoM e

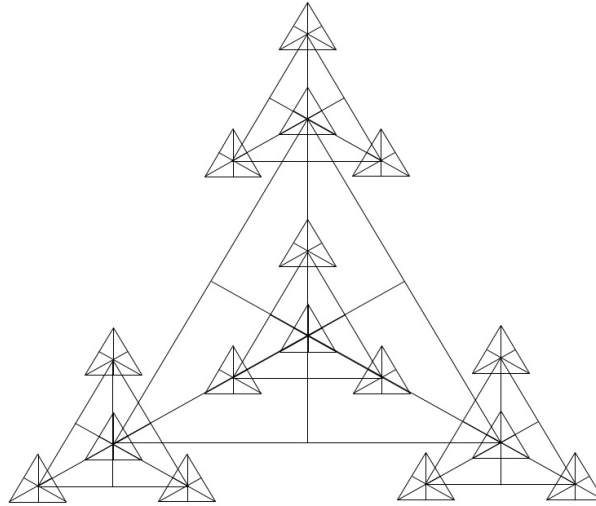
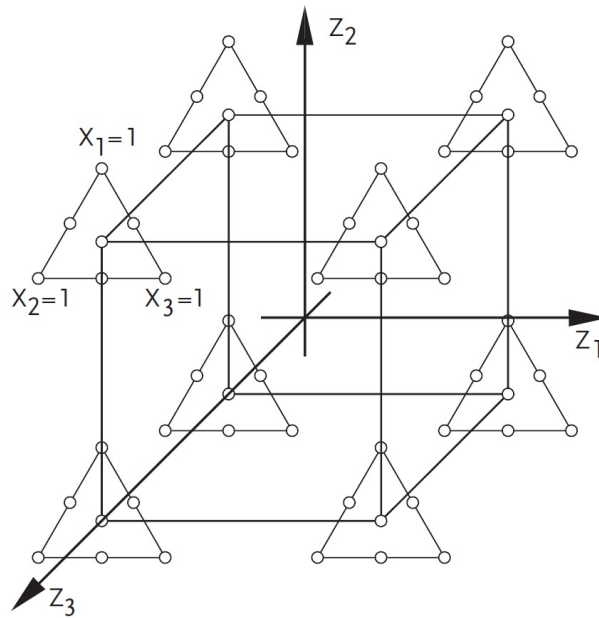


Figura 3.10 – Arranjo cruzado de misturas MoM

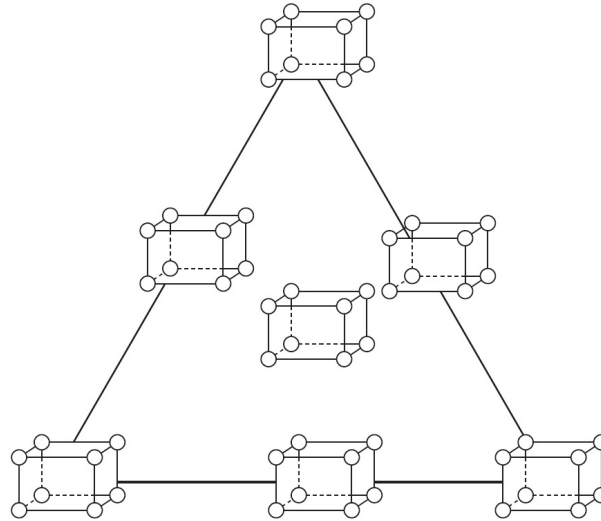
Figura 3.11 – *Simplex-centroid* em cada ponto do experimento Fatorial Completo 2^3 .

Fonte: Lazić (2004)

mistura com variáveis de processo serão usados nas análises desta tese.

$$\begin{aligned}
 E[f(x)] = & \sum_{i=1}^q \beta_i x_i + \sum_{i \leq j}^q \sum_{j=1}^q \beta_{ij} x_i x_j \\
 & + \sum_{k=1}^K \left[\sum_{i=1}^q \alpha_{ik} x_i + \sum_{i \leq j}^q \sum_{j=1}^q \alpha_{ijk} x_i x_j \right] \cdot z_k \\
 & + \left[\sum_{i=1}^q \delta_{ijk} x_i + \sum_{i \leq j}^q \sum_{j=1}^q \delta_{ijk} x_i x_j \right] \cdot z_i z_j
 \end{aligned} \tag{3.26}$$

O arranjo de misturas cruzado é um arranjo do tipo CCD tratado como uma

Figura 3.12 – Experimento fatorial 2^3 em cada ponto do *Simplex-centroid*.Fonte: [Lazić \(2004\)](#)

mistura, ou seja, cada componente pode variar de 0 a 1. Então uma mistura de 2×2 seria como um CCD de 4 componentes, a diferença é que a soma de cada linha da mistura seria 2 e a análise seria compilando cada valor entre 0 e 1. No CCD (0) e (1) ou (-1) e $(+1)$ são códigos para o menor e o maior valor de uma configuração do processo. Usando o CCD como mistura o resultado varia de 0 até 1 para cada componente, respeitando o valor da soma, ou restrição máxima, no caso 2×2 , a soma é dois. então: $2 \times 1 : \Sigma = 1$; $2 \times 2 : \Sigma = 2$; $2 \times 3 := \Sigma 3$; $n \times m = \Sigma m$. M é o número de portfólios, n é a quantidade de empresas em cada portfólio.

Por exemplo, comparando um arranjo para 7 empresas usando arranjo normal de mistura teremos 15 linhas para cada corrida. Em um arranjo MoM 4×2 simplificado teremos 20 linhas com um ponto central. E usando MoM 4×2 completo teremos 150 linhas com um ponto central. Os resultados de um experimento simulado são mostrados na Tabela 3.4 apenas como exemplificação e comparação inicial.

Tabela 3.4 – Comparação simples CCD x MoM

a	b	c	d	e	f	g	retorno	risco	resultado	arranjo utilizado
0,000	0,000	0,470	0,530	0,000	0,000	0,000	0,023	0,005	0,888	Mistura normal
0,000	0,000	0,000	0,806	0,194	1,000	0,000	0,036	0,007	0,980	MoM simplificado 22 linhas similar CCD
0,000	0,000	0,000	0,850	0,261	0,889	0,000	0,036	0,006	0,984	MoM completo 150 linhas

Como a soma em mais de um portfólio será maior do que 1. O resultado final deve ser dividido por m para que se tenha a proporção exata em relação a 100% do valor que será investido. A fronteira eficiente deixará de existir porque agora há um grupo de média variância. Se usar o indicador de performance da ação haverá apenas um objetivo para a formulação sendo possível acrescentar outros, por exemplo correlação.

Quando as ações são de empresas diferentes e fortemente correlacionadas, a otimi-

zação é como se fosse apenas uma ação de uma, porém há uma divisão entre elas dentro do portfólio.

3.2.6 Análise dos Experimentos

Uma importante método para as análises dos experimentos é baseada no conceito de ANOVA. Ela faz testes de hipótese de que as médias de duas ou mais populações são iguais, avaliam ainda o grau de importância de um ou mais fatores do modelo, comparando as médias de variáveis de resposta (y) nos diferentes níveis de fator de entrada (x) (NEWBOLD *et al.*, 2012).

Pela utilização do software Design Expert, que disponibiliza uma série de informações sobre os dados que gera, tem-se quadros de dados para as análises. Exemplificando com o Quadro 3.3.

Quadro 3.3 – ANOVA Exemplo

Resposta 2 SP					
ANOVA para Modelo de Mistura Quadrático (Distorcido)					
Componente da Mistura Codificado em Pseudo Componente					
Tabela Análise de Variância [Soma parcial dos quadrados]					
Modelos	Soma dos Quadrados	df	Média dos Quadrados	F Valor	p-value Prob > F
Testados	0,007327661	7	0,001046809	299,9555467	< 0,0001*
Mistura Linear	0,006483328	3	0,002161109	619,2503953	< 0,0001
AB	2,02118E-05	1	2,02118E-05	5,791552324	0,0369
AC	5,06571E-06	1	5,06571E-06	1,451541964	0,256
AD	7,26622E-05	1	7,26622E-05	20,82084157	0,001
BC	4,46414E-05	1	4,46414E-05	12,79167815	0,005
BD	0,000000	0			
CD	0,000000	0			
Residual	3,48988E-05	10	3,48988E-06		
Falta de Ajuste	5,38427E-12	1	5,38427E-12	1,38854E-06	0,9991**
Erro Puro	3,48988E-05	9	3,87764E-06		
Corr Total	0,00736256	17			
(*) significativa; (**) não significativa					

No quadro são mostrados, na parte superior, o nome da resposta, seu número e o nome dado quando o modelo foi criado. A próxima linha fornece uma breve descrição do modelo sendo ajustado, seguido pelo tipo de soma dos quadrados usados para os cálculos. Dentro do quadro tem-se nas linhas:

- Bloco: A linha do bloco mostra quanta variação na resposta é atribuída aos blocos. A variação do bloco é removida da análise. Os blocos não são testados (sem valor F ou p) porque são considerados um fator não replicável e difícil de alterar. Os blocos são assumidos como não interagindo com os fatores. Se não houver blocos no modelo, essa linha não estará presente. O que é o caso deste exemplo.

- **Modelo:** A linha do modelo mostra quanta variação na resposta é explicada pelo modelo junto com o teste geral do modelo para significância.
- **Termos:** O modelo é separado em termos individuais e testado de forma independente.
- **Residual:** a linha residual mostra quanta variação na resposta ainda está sem explicação.
- **Falta de Ajuste:** é a quantidade informações que as previsões do modelo perdem sobre as observações.
- **Erro Puro:** é a quantidade da diferença entre as execuções replicadas.
- **Corr Total:** Esta linha mostra a quantidade de variação em torno da média das observações. O modelo explica parte disso, o resíduo explica o resto.

Nas colunas:

- **Fonte:** um nome significativo para as linhas.
- **Soma dos Quadrados:** Soma das diferenças quadradas entre a média geral e a quantidade de variação explicada por essa fonte de linhas.
- **df:** Graus de Liberdade é o número de parâmetros estimados usados para calcular a soma dos quadrados da fonte.
- **Média dos Quadrados:** A soma dos quadrados divididos pelos graus de liberdade. Também chamado de variância.
- **Valor F:** teste para comparar o quadrado médio da fonte com o quadrado médio residual.
- **Prob > F:** (p-value) é a probabilidade de ter o valor F observado se a hipótese nula for verdadeira (não há efeitos de fator). Valores pequenos de probabilidade exigem rejeição da hipótese nula. A probabilidade é igual a integral sob a curva da distribuição F que está além do valor F observado.

Se o valor Prob > F for muito pequeno (menor que 0,05 por padrão), a fonte testou significância. Os termos do modelo são significativos e provavelmente têm um efeito real na resposta. Uma falta significativa deste ajuste indicaria que o modelo não se ajusta aos dados dentro da variação entre as replicadas observadas.

A modelagem estatística para estes dados podem ser adaptadas de [Newbold et al. \(2012\)](#), que pela explicação utilizada pelo software em uso, tem-se que o quadro possui

uma linha por termo estimado no modelo. O número de colunas é dependente do tipo de análise.

Fator são variáveis experimentais selecionadas para inclusão no modelo preditivo. A estimativa do coeficiente de regressão representa a mudança esperada na resposta y por unidade de mudança no fator de entrada x quando todos os fatores restantes são mantidos constantes. Em desenhos de dois níveis ortogonais, é igual a metade do efeito fatorial. Estimativa do Coeficiente para Desenhos Fatoriais Gerais são os coeficientes para fatores categóricos de vários níveis não são tão simples de interpretar. β_1 é a diferença da média do nível 1 da média geral. β_2 é a diferença da média do nível 2 da média geral. $\beta_{12}(k-1)$ é a diferença de nível $(k-1)$ da média geral. A soma negativa dos coeficientes será a diferença do nível k da média geral. Para interpretação destes coeficientes deve-se usar os gráficos do modelo.

df: Graus de Liberdade é igual a um para testar coeficientes. O Erro Padrão é o desvio padrão associado ao coeficiente de estimação. 95% CI Alto e Baixo: Se este intervalo se estender por 0 (um limite é positivo e o outro negativo), então o coeficiente de 0 pode ser verdadeiro, indicando que o termo não é significativo, pode ser descartado das análises.

VIF é o fator de inflação de variância. Ele mede o quanto a variância ao redor do coeficiente estimado é inflada pela falta de ortogonalidade no projeto do experimento. Se o fator for ortogonal a todos os outros fatores no modelo, o VIF é um deles. Valores maiores do que 10 indicam que os fatores estão muito correlacionados entre si (eles não são independentes). Os VIFs são uma estatística menos importante quando se trabalha com projetos de mistura e projetos de superfície de resposta restrita. No caso desta tese o uso é arranjo de misturas, projeto de misturas.

O modelo preditivo é listado em termos reais e codificados. (Para experimentos de mistura, as equações de predição são dadas em valores pseudo, reais e reais dos componentes). A equação codificada (ou pseudo) é útil para identificar a significância relativa dos fatores comparando os coeficientes do fator. Todas as equações fornecem previsões idênticas quando a hierarquia é aplicada. Essas equações, usadas para previsão, não têm efeitos de bloqueio. O bloqueio é uma restrição à aleatorização do experimento, usado para reduzir o erro. Blocos são usados apenas para ajustar a observação para estes experimentos, não para fazer previsões.

Apenas para modelos de mistura linear: A tabela de coeficientes é aumentada para modelos de mistura linear para incluir estatísticas sobre os efeitos lineares ajustados. Como os coeficientes lineares não podem ser comparados a zero, o efeito linear do componente i é medido pela diferença do coeficiente i dos outros coeficientes $(q-1)$. O teste t é aplicável à diferença nas estimativas dos coeficientes de mistura. Quando o espaço de experimento não é um *simplex*, a fórmula para calcular os efeitos do componente é ajustada para as

diferenças nos intervalos. O vetor gradiente será a inclinação estimada através da superfície de resposta linear projetada através de uma mistura de referência na direção de Cox e Piepel (JR; BEVERLY, 1997). O efeito total de um componente é o gradiente vezes o intervalo em que o componente variou. Esses efeitos são plotados como um gráfico de contorno.

Prob > t: Esta é a probabilidade de obter este valor t se as duas médias não forem realmente diferentes. Um valor inferior a 0,05 indica que existe uma diferença estatisticamente significativa entre as médias.

3.3 PROGRAMAÇÃO LEXICOGRÁFICA

Conforme descrito por Bartolomeu (2016) o método de programação lexicográfico é considerado um método de otimização intuitivo para funções multicritério. Considera-se funções objetivo multiníveis sendo maximizados ou minimizados problemas de otimização monocritério e, adicionando gradualmente, restrições baseadas nos valores obtidos anteriormente otimizados para as funções objetivos adjacentes. Ele demonstrou genericamente para um problema bicritério (podendo ser de maximização ou minimização) como visto na Equação 3.27. Nesta, f_1 e f_2 representam dois critérios em análise, no caso, ambos de minimização. X é o conjunto de soluções admissíveis para o problema em questão. O método lexicográfico irá resolver o problema com visão monocritério (Equação 3.28) e finalmente agregando esta na Equação 3.29. Sendo f_1^* o valor ótimo obtido para o problema anterior. Este valor ótimo é agora uma restrição, neste caso de igualdade, no valor da função objetivo f_1 . A função da Equação 3.29 é finalmente a função objetivo que irá otimizar o seguinte problema monocritério. Este método poderá ser aplicado para mais do que duas funções.

$$\begin{aligned} \min \quad & f_1(x) \\ \min \quad & f_2(x) \\ \text{s.a : } & x \in X \end{aligned} \tag{3.27}$$

$$\begin{aligned} \min \quad & f_1(x) \\ \text{s.a : } & x \in X \end{aligned} \tag{3.28}$$

$$\begin{aligned} \min \quad & f_2(x) \\ \text{s.a : } & x \in X \\ & f_1(x) = f_1^* \end{aligned} \tag{3.29}$$

No caso desta tese, cada função monocritério é uma otimização da função objetivo de um arranjo de misturas.

3.4 ANÁLISE DE AGRUPAMENTOS (*CLUSTERS*)

A análise de agrupamentos é um processo de particionamento de uma população considerada, dividindo-a em subgrupos compostos de acordo com as semelhanças entre grupos de variáveis (DONI, 2004). Estes grupos são determinados para obter-se homogeneidade dentro dos grupos e ao mesmo tempo heterogeneidade entre eles. Isto torna as variáveis dentro do grupo muito similares e distancia as variáveis similares das não similares ou diferentes. As semelhanças destes termos pode ser considerada como a distância entre pares de termos. Aqueles que possuem menores distâncias são os mais semelhantes entre si. Para isto pode-se utilizar o cálculo de distância Euclidiana. A análise de agrupamentos tem como foco estudar um conjunto de relações interdependentes. Esta não tem intenção de distinção entre variáveis dependentes e independentes, isto é, variáveis do tipo causa e efeito, o que é fator determinante no estudo da regressão (BUSSAB *et al.*, 1990). Esta técnica tem como finalidade, reunir os objetos (indivíduos, elementos) verificados nos grupos em que exista homogeneidade dentro do grupo e heterogeneidade entre os grupos, com o objetivo de propor classificações. Esta análise numérica multivariada, objetiva a proposição de classificações estruturais, ou de reconhecimento da existência de grupos com objetivo mais específico de dividir o conjunto de observações em um número de grupos homogêneos, segundo algum critério de homogeneidade, critério de agrupamento.

Há varias técnicas para auxiliar na formação e divisão dos *clusters*. Zaiane *et al.* (2002) listaram alguns típicos requisitos para a escolha de uma boa técnica de cluster conforme abaixo.

- Escalabilidade: o método deve ser aplicável em grandes volumes de dados e sua performance cairá linearmente conforme o tamanho dos dados aumente;
- Versatilidade: os objetos dos *clusters* podem ser de tipos diferentes, numéricos, do tipo verdadeiro ou falso, ou dados categóricos;
- Habilidade para identificar *clusters* com tamanhos diferentes: isto é importante para *clusters* com dados cujo espaço não seja de tamanho esférico;
- Mínimos parâmetros de entrada: o método deve permitir uma entrada mínima de parâmetros que contenham um domínio conhecido para que haja uma correta formação do *cluster*. A maioria dos métodos possui alguns parâmetros chave que por vezes não podem ser aplicados no mundo real;
- Robustez considerando fatores de ruído: isto é importante porque existe o fator ruído em qualquer problema prático. Um bom algoritmo deve ser capaz de funcionar mesmo quando houver a presença de fatores de ruído intenso;
- Não afetável pela ordem de entrada dos dados: o método de *cluster* deve ter resultado consistentes independente da ordem com que os dados são apresentados;
- Escalabilidade a grandes dimensões: habilidade de funcionar com grandes dimensões de dados que na realidade ofereçam também multidimensionalidade.

Historicamente não há um algoritmo que possa atender sozinho a todos estes requisitos. O mais importante é escolher aquele que melhor atenda a necessidade que os dados apresentam.

3.4.0.1 Medidas de similaridade

A análise de *clusters* requer uma medida de similaridade, normalmente expressa como função distância ou métrica de distância. Podani e Schmera (2006) sugeriu uma formulação para o cálculo de similaridade na qual n é o número de traços (variáveis). Os pesos $w_{ijk} = 0$ se a comparação entre os grupos j e k não pertencerem a variável i , há dados faltantes; caso contrario $w_{ijk} = 1$ na Equação (3.30). A medida de não similaridade entre os grupos j e k para a variável i será S_{ijk} . Esta definição dependerá do tipo de escala utilizado. Para variáveis nominais na Equação (3.31), em que X_{ij} é a pontuação dos dados brutos para a variável i no grupo j .

$$d_{jk} = \frac{\sum_{i=1}^n w_{ijk} S_{ijk}}{\sum_{i=1}^n w_{ijk}} \quad (3.30)$$

$$S_{ijk} = 1, \text{ se } X_{ij} \neq X_{ik} \quad (3.31a)$$

$$S_{ijk} = 0, \text{ se } X_{ij} = X_{ik} \quad (3.31b)$$

Para variáveis do tipo ordinais, todas as pontuações X_{ij} serão substituídas por sua classificação r_{ij} . A não similaridade entre dois grupos para a variável i é obtida utilizando a Equação (3.32). Finalmente para variáveis mensuráveis ou com intervalos escalares será calculado na Equação (3.33).

$$S_{ijk} = |r_{ij} - r_{ik}| / [\max r_i - \min r_i] \quad (3.32)$$

$$S_{ijk} = |X_{ij} - X_{ik}| / [\max x_i - \min x_i] \quad (3.33)$$

A escala das medidas, segundo Manly (2004) influencia na estimativa da distância Euclidiana. Para mitigar este problema, pode-se proceder na padronização escalar das variáveis. As distâncias calculadas podem ser encontradas com a utilização de diversos métodos. Entre eles destacamos: (DONI, 2004)

o Distância Euclidiana é a distância geométrica em um espaço multidimensional, então a distância euclidiana entre dois elementos X_i, Y_j é definida por (3.34). No caso da distância euclidiana quadrática, a raiz quadrada na Equação (3.34) é retirada sendo então definida

na Equação (3.35). Esta pode ser considerada uma medida de dissimilaridade, ao contrário do coeficiente de correlação que é uma medida de similaridade dos termos ou objetos.

$$d_{xy} = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (3.34)$$

$$d_{xy} = \sum_{i=1}^n (X_i - Y_i)^2 \quad (3.35)$$

Sendo a distância Euclidiana média calculada pela média das distâncias euclidianas. A média aritmética dará a distância entre aglomerados considerando um mesmo peso a eles. Esta média pode ser calculada pela Equação (3.36).

$$d = \sum_{i=1}^n \left\{ \frac{(X_i - Y_i)^2}{X_i} \right\} \quad (3.36)$$

o Distância de Mahalanobis mede a similaridade com relação a um conjunto de características correlacionadas. A distância considera o grau de dependência entre as variáveis amostrais. Quando existe repetição de dados, a distância de Mahalanobis (D^2) é a mais indicada. Apesar de D^2 ser o valor do quadrado da distância de Mahalanobis, este é chamado apenas de distância de Mahalanobis (MAESSCHALCK *et al.*, 2000).

o Distância de Manhattan na Equação (3.37) na qual o efeito de uma grande diferença entre dimensões é minimizado pela ausência do termo quadrático.

$$d_{xy} = \sum_{i=1}^n |X_i - Y_i| \quad (3.37)$$

o Distância Chebychev usada para o quando há necessidade de forçar a diferença entre elementos quando há apenas uma dimensão diferente na Equação (3.38).

$$d_{xy} = \max(X_{in} - Y_{in}) \quad (3.38)$$

o Método Hierárquico consiste em agrupamentos com divisões sucessivas dos elementos, nos quais estes são agregados ou desagregados de certo grupo. Geralmente são representados por um dendograma, diagrama de árvore. O dendograma é um diagrama bidimensional composto por ramos e raízes. Os ramos são os elementos e as raízes são os agrupamentos. O número de agrupamentos dado pelas raízes será especificado de acordo com os objetivos das análises. Este corte é então subjetivo e a cargo do pesquisador.

O método hierárquico de Ward (JR, 1963) é usado para que cada membro do grupo possua uma similaridade máxima para uma característica especificada. É sugerido

em estudos nos quais uma solução ótima para um certo número de grupos não pode ser praticada. Ele fez a inferência dados n conjuntos, seu método permite uma redução para $n - 1$ mutuamente exclusivo considerando a união de todos os possíveis $n(n - 1)/2$ pares e selecionando aquelas uniões cujas funções de relação são máximas de acordo com o critério escolhido pelo investigador. Repetindo esse procedimento, apenas um grupo restará sendo que uma estrutura hierárquica e uma estimativa das perdas associadas em cada estágio nos agrupamentos pode ser obtida e analisada. O exemplo de Ward será repetido aqui para auxiliar no entendimento. Seja um conjunto composto por 10 elementos, $\{2, 6, 5, 6, 2, 2, 2, 0, 0, 0\}$. Uma prática comum de análise é utilizar a média como medida representativa para o conjunto. A perda de informação resultante da análise dos 10 componentes como um grupo único utilizando essa média de 2,5 pode ser indicada por um valor numérico de reflexão, a soma dos erros quadrados (SSE). A função da soma dos erros quadrados é dada por uma função de relacionamento na Equação (3.39). Utilizando os dados do conjunto de 10 elementos, teremos a Equação(3.40).

$$SSE = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \quad (3.39)$$

$$SSE_{umgrupo} = \sum_{i=1}^{10} x_i^2 - \frac{1}{10} \left(\sum_{i=1}^{10} x_i \right)^2 = 113 - 62,5 = 50,5 \quad (3.40)$$

Similarmente se os 10 elementos forem classificados de acordo com seus valores em quatro grupos, $\{0, 0, 0\}$, $\{2, 2, 2, 2\}$, $\{5\}$, $\{6, 6\}$, estes grupos podem ter seus valores de erros calculados pela soma dos erros quadrados dos quatro grupos, $SSE_{quatrogrupos} = SSE_{grupo1} + SSE_{grupo2} + SSE_{grupo3} + SSE_{grupo4}$. O objetivo da função de relacionamento da perda de informação deve ser seu valor mínimo, 0,0. No caso do exemplo, considerando os dados como um grupo único, a perda de informação dada pela soma do erro quadrático foi de 50,5 contra a perda de 0,0 com quatro conjuntos, ou grupos, ou clusters.

o Métodos Aglomerativos, cada elemento é inicialmente um grupo que se liga a outros elementos por similaridade a cada passo dado. No último passo haverá apenas um grupo contendo todos os elementos iniciais.

o Coeficiente de correlação de Pearson mede a relação entre as variáveis. A correlação r_{xy} , segundo Newbold *et al.* (2012) é a descrição do relacionamento entre duas variáveis aleatórias, veja a Equação (3.41) que considera o cálculo pela covariância e que pode ainda ser simplificada na Equação (3.42). Esta relação pode ser medida em uma escala de coeficiente que vai de $-1,0$, passando por $0,0$ até $+1,0$. O coeficiente de correlação igual a $-1,0$ significa que a variável x tem um relacionamento linear perfeitamente negativo com a variável y . O coeficiente de correlação igual a $0,0$ é a indicação de que as duas variáveis não tem relação entre si e são completamente independentes uma da outra. Finalmente, o coeficiente de correlação igual a $+1,0$ indica que o relacionamento entre x

e y é linear perfeitamente positivo. Entre os valores extremos e o zero, há o indicativo de relacionamento.

$$r_{xy} = \frac{1}{n} \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\left(\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \right) \cdot \left(\sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}} \right)} \quad (3.41)$$

$$r_{XY} = \frac{n \sum XY - \sum X \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2] \cdot [n \sum Y^2 - (\sum Y)^2]}} \quad (3.42)$$

3.5 ANÁLISE DE COMPONENTES PRINCIPAIS

A análise de componentes principais (PCA - principal component analysis) é provavelmente a melhor e mais antiga técnica para análise de problemas multivariados (JOLIFFE, 2002). Hotelling (1933) introduziu o método de PCA como uma técnica matemática da análise multivariada usada também para a análise estatística multivariada. Isto é utilizado para inferências sobre a estrutura de variância-covariância de um conjunto de dados. Este conceito usa combinações lineares das variáveis originais. Existem dois objetivos principais do PCA: um é a redução na dimensionalidade e o segundo é prover uma interpretação dos dados. A relação entre os componentes é revelada e isto oferece um passo intermediário para uma visão mais ampla desse relacionamento (JOHNSON; WILCHERN, 2002; RENCHER, 2003). O PCA identifica medidas responsáveis pelas maiores variações entre o quadro de elementos-variáveis numéricas mantendo suas características de informação.

Conforme a demonstração feita por Paiva (2006), no caso de uma não reconhecimento dos parâmetros da população dos dados, adota-se a matriz de variância-covariância amostral mostrado na Equação 3.43. Os componentes principais amostrais em termos de \mathbf{S} e \mathbf{R} serão tal como descrito em 3.44 e 3.45.

$$S = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{12} & S_{22} & \cdots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{1p} & S_{12} & \cdots & S_{pp} \end{bmatrix} = \begin{bmatrix} \frac{\sum_{j=1}^n (x_{1j} - \bar{x}_1)^2}{n} & \cdots & \frac{\sum_{j=1}^n (x_{1j} - \bar{x}_1)(x_{pj} - \bar{x}_p)}{n} \\ \vdots & \vdots & \ddots \\ \frac{\sum_{j=1}^n (x_{1j} - \bar{x}_1)(x_{pj} - \bar{x}_p)}{n} & \cdots & \frac{\sum_{j=1}^n (x_{pj} - \bar{x}_p)^2}{n} \end{bmatrix} \quad (3.43)$$

$$\sum_{i=1}^p S_{ii} = \hat{\lambda}_1 + \hat{\lambda}_2 + \cdots + \hat{\lambda}_p \quad (3.44)$$

$$r_{(\hat{y}_i, x_k)} = \frac{Cov(x_k, \hat{y}_i)}{\sqrt{Var(\hat{y}_i)Var(x_k)}} = \frac{\hat{e}_{ki}\sqrt{\hat{\lambda}_i}}{\sqrt{S_{kk}}}, \quad i, k = 1, 2, \dots, p \quad (3.45)$$

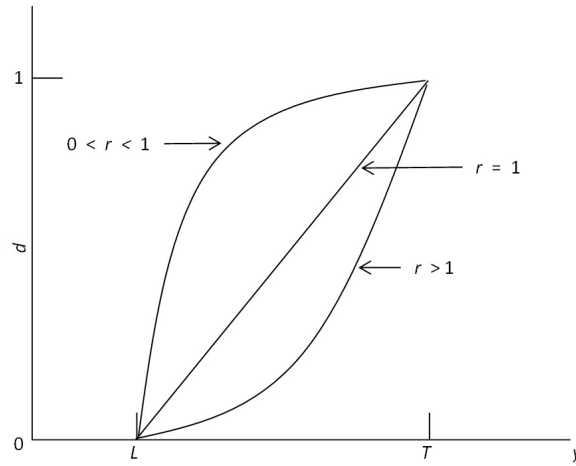
Bouroche e Saporta (1982) considera o PCA como um método de análise fatorial que constrói variáveis sintéticas oriundas da combinação linear das variáveis originais, fazendo assim uma redução de variáveis em fatores para as análises. Estes fatores são as variáveis sintéticas. Para que esta redução seja possível, as variáveis não podem ser independentes, devem ter coeficiente de correlação diferente de zero.

Karhunen e Joutsensalo (1995) aplicaram PCA em otimização de redes neurais. Swiniarski e Skowron (2003) aplicaram na projeção de características e redução dessas características com um algoritmo de reconhecimento de padrões em um conjunto grosseiro de dados. Eles também aplicaram para uma seleção de características que identificassem subclasses dos dados originais tornando possível a formação de padrões desses dados. O reconhecimento de vogais utilizando análise de discriminantes lineares e PCA foi aplicado por Wang e Paliwal (2003). Lin e Zhang (2005) comparou PCA com análise de componentes dependentes na separação de apenas uma fonte de recurso. Flutuações evento por evento em colisão núcleo-núcleo em energia ultra-relativista foi também analisado com a aplicação de PCA por Bhalerao *et al.* (2015). Estes são apenas alguns exemplos que ilustram brevemente a diversidade de problemas que podem ser analisados aplicando-se PCA. Basicamente, qualquer situação na qual se necessite de simplificação dos dados sem que haja a perda das características da informação original, pode ser beneficiada com o uso do PCA. Finalmente, Paiva (2006) demonstra sobre a entidade estatística chamada de escore de componentes principais (PC_k) que é uma representação das características das variáveis originais. Ele escreve sua função na Equação 3.46.

$$PC_k = Z^T e = \begin{bmatrix} \left(\frac{x_{11} - \bar{x}_1}{\sqrt{s_{11}}} \right) & \left(\frac{x_{21} - \bar{x}_2}{\sqrt{s_{22}}} \right) & \dots & \left(\frac{x_{p1} - \bar{x}_p}{\sqrt{s_{pp}}} \right) \\ \left(\frac{x_{12} - \bar{x}_1}{\sqrt{s_{11}}} \right) & \left(\frac{x_{22} - \bar{x}_2}{\sqrt{s_{22}}} \right) & \dots & \left(\frac{x_{p2} - \bar{x}_p}{\sqrt{s_{pp}}} \right) \\ \vdots & \vdots & \ddots & \vdots \\ \left(\frac{x_{1n} - \bar{x}_1}{\sqrt{s_{11}}} \right) & \left(\frac{x_{2n} - \bar{x}_2}{\sqrt{s_{22}}} \right) & \dots & \left(\frac{x_{pn} - \bar{x}_p}{\sqrt{s_{pp}}} \right) \end{bmatrix} \cdot \begin{bmatrix} e_{11} & e_{12} & e_{1p} \\ e_{21} & e_{22} & e_{2p} \\ \vdots & \vdots & \ddots \\ e_{1p} & e_{2p} & e_{pp} \end{bmatrix} \quad (3.46)$$

3.6 OTIMIZAÇÃO DE PROCESSOS

Alguns métodos de otimização, diferentes dos tradicionais métodos de programação matemática estão sendo concebidos. Estes são conhecidos como modernos ou não tradicionais métodos de otimização. A maioria destes é baseado em certas características e comportamentos similares a biológicos, moleculares, enxames de insetos e sistemas neurobiológicos (RAO; RAO, 2009).

Figura 3.13 – Maximizar y

3.6.1 Método *desirability*

Há muitas técnicas numéricas que podem ser usadas para resolver problemas de otimização. Algumas dessas técnicas referem-se a métodos de programação não linear.

Montgomery (2008) explica que uma forma generalizada de otimização multiobjetivo é utilizando a função *desirability* que irá converter as respostas y_i em uma função de otimização d_i , estas estarão dentro do intervalo na Equação (3.47). Esta equação pode ser explicada pela variação padronizado pelo escore Z em cada variável, aproximando as mais similares, formando assim os *PCs*.

$$0 \leq d_i \leq 1 \quad (3.47)$$

No qual a resposta quanto y_i é o alvo ou ponto ótimo, então $d_i = 1$ e se a resposta está fora da região de solução ou região de aceitação, $d_i = 0$. As variáveis do modelo escolhido para maximizar a função *desirability* serão dadas na Equação (3.48).

$$D = (d_1 * d_2 * \dots * d_m)^{1/m} \quad (3.48)$$

A seguir estão demonstradas as estruturas de otimização individual da função *desirability* é linear. Escolhendo $r > 1$ será dado ênfase nos pontos mais próximos do valor alvo e escolhendo $0 < r < 1$ então este será menos importante. A Equação (3.49) e Equação (3.13) demonstram quando o objetivo (alvo) é maximizar y no qual o peso $r = 1$. Sendo T o objetivo (alvo), U o valor máximo e L o valor mínimo da função.

$$d = \begin{cases} 0 & y < L \\ \left(\frac{y-L}{T-L}\right)^r & L \leq y \leq T \\ 1 & y > T \end{cases} \quad (3.49)$$

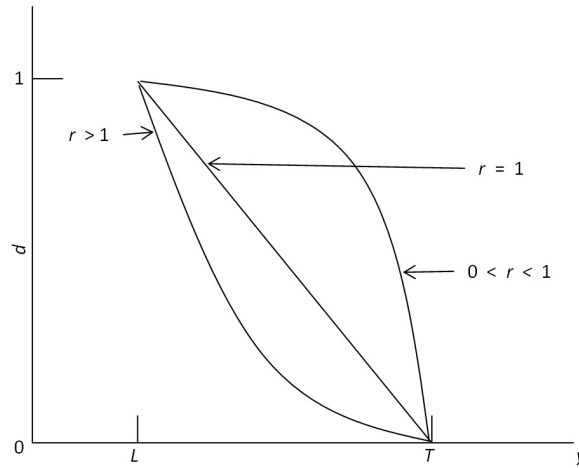
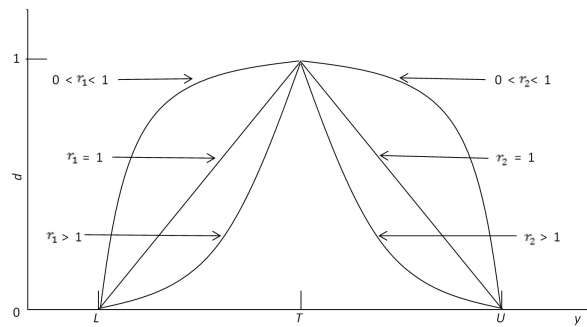
Figura 3.14 – Minimizar y 

Figura 3.15 – Função multiobjetivo

A Equação (3.50) e Equação (3.14) demonstram quando o objetivo (alvo) é minimizar y no qual o peso $r = 1$.

$$d = \begin{cases} 1 & y < T \\ \left(\frac{U-y}{U-T}\right)^r & T \leq y \leq U \\ 0 & y > U \end{cases} \quad (3.50)$$

A Equação (3.51) e a Equação (3.15) demonstram uma função multiobjetivo. Se refere a um alvo que está situado entre um valor mínimo L e um valor máximo U observando y no qual o peso $r = 1$.

$$d = \begin{cases} 0 & y < L \\ \left(\frac{y-L}{T-L}\right)^{r_1} & L \leq y \leq T \\ \left(\frac{U-y}{U-T}\right)^{r_2} & T \leq y \leq U \\ 0 & y > U \end{cases} \quad (3.51)$$

Há programas computacionais que resolvem de maneira rápida e eficiente problemas dessa natureza, este é um dos motivos do método *desirability* ser tão difundido como um ótimo algoritmo de otimização multiobjetivo.

Rao e Rao (2009) explicam outros métodos também comumente usados para otimização de processos. Em anos recentes, alguns métodos de otimização, diferentes dos tradicionais métodos de programação matemática estão sendo concebidos. Estes são conhecidos como modernos ou não tradicionais métodos de otimização. A maioria destes é baseado em certas características e comportamentos similares a biológicos, moleculares, enxames de insetos e sistemas neurobiológicos. Estes métodos podem ser divididos em seis grupos.

3.7 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO

Neste capítulo, discutiram-se diferentes técnicas para o processo experimentação baseado em projeto de experimentos, além de processo de experimentos de mistura e suas aplicações. Foram referenciados métodos de agrupamento com ênfase em análise de componentes principais. Foram apresentadas etapas necessárias para cada fase de experimentação e aglutinação e o uso da programação lexicográfica. Também referenciando-se o método de otimização *desirability* e como o software Design Expert entrega as informações para as análises.

4 PORTFÓLIO E TÉCNICAS DE SELEÇÃO

4.1 CONSIDERAÇÕES INICIAIS

Neste capítulo, apresenta-se a fundamentação teórica sobre seleção de portfólio. Há também a demonstração de como o portfólio é influenciado pela correlação, além de referenciar a análise de portfólio com o uso de MDE.

O intuito principal deste capítulo está no entendimento de como funciona o processo de seleção de portfólio, sua fundamentação na história e algumas técnicas utilizadas.

4.2 SELEÇÃO DE PORTFÓLIO

Harry Markowitz introduziu o modelo de seleção de portfólios ([MARKOWITZ, 1952](#)). A seleção ou otimização de portfólio é uma técnica econométrica usada no auxílio a tomada de decisão pelos investidores. Estes devem decidir quanto de sua riqueza aplicar em cada investimento. Os métodos de seleção de portfólio, porém, podem ser aplicados em diferentes tipos de investimentos ([AMIHUD; BARNEA, 1974](#); [MAYERS, 1974](#); [TRIPPI, 1989](#); [BETTER; GLOVER, 2006](#); [DELARUE *et al.*, 2011](#)).

O conceito de portfólio pode ser comparado a uma cesta cheia de ovos. Se a cesta cair, todos os ovos serão perdidos. Mas dividindo a mesma quantidade de ovos em cestas diferentes, o risco de que todas caiam e que todos os ovos se percam é reduzido. Isto é conhecido como diversificação.

O risco pode ser determinado pela possibilidade de mudança nos valores entre duas datas ([ARTZNER *et al.*, 1999](#)). Isto pode ser considerado, na maioria dos casos, como a variância dos retornos ou pela volatilidade quando considerado o desvio padrão, em um período conhecido. O investidor sempre deseja ter mais retorno do que menos retorno e menos risco do que mais risco. O objetivo da seleção de portfólio é balancear estes objetivos inversos, retorno e risco. [Tíer \(2005\)](#) utilizou a Figura 4.1 para demonstrar a relação destes dois objetivos.

O risco de um investimento, utilizando a teoria de portfólio, pode ser menor do que o risco ao se optar por apenas um investimento, como provado por ([MARKOWITZ, 1952](#)). Se um investidor diversifica entre dois portfólios, a variância ou risco da composição dos dois portfólios será menor do que a variância de um ou outro portfólio separadamente. Se e somente se, ambos os portfólios forem perfeitamente correlacionados, a variância não será reduzida. Nos demais casos, a variância reduzirá.

A decisão mais importante que um investidor pode fazer sobre seu investimento

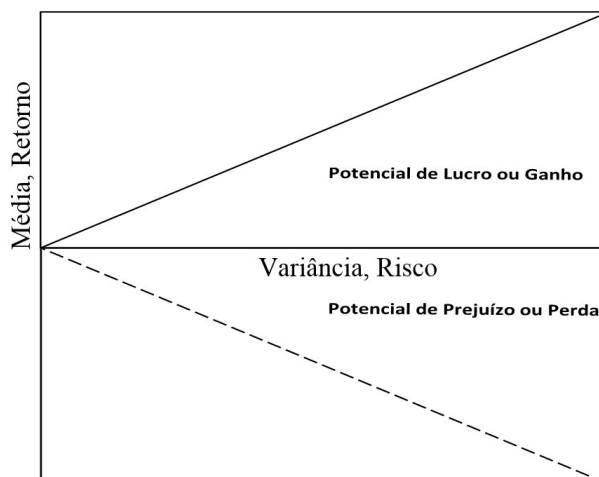


Figura 4.1 – Retorno em função do risco

está no nível de risco que ele está disposto a suportar. O desejo, no investimento, é obter o maior retorno com o menor risco mesmo sabendo que retornos maiores estão sempre associados a riscos maiores (BRENTANI, 2004). O objetivo da gestão de portfólio é, portanto, reunir diversos títulos e outros Ativos em carteiras que atendam às necessidades dos investidores. É trabalho do gestor da carteira gerir o vasto leque de informações disponíveis e transformar esta carteira em um investimento bem sucedido. A gestão de investimentos inclui a formação de um portfólio ótimo, ou seja, a determinação da melhor relação risco-retorno.

“O processo de seleção de uma carteira pode ser dividido em duas etapas. A primeira etapa começa com a observação e experiência e termina com crenças sobre os desempenhos futuros de títulos disponíveis. A segunda etapa começa com as crenças relevantes sobre os desempenhos futuros e termina com a escolha da carteira”(MARKOWITZ, 1952).

A abordagem proposta por Markowitz foi pioneira na construção de modelos que buscam o portfólio ótimo de ativos com o uso da manipulação do binômio risco-retorno. Markowitz modelou o risco como sendo a variância dos retornos e com o uso da programação quadrática conseguiu determinar o quanto deve ser investido em cada um dos ativos candidatos, de modo a se atingir um determinado retorno a um mínimo risco. Esta primeira abordagem motivou muitos trabalhos e é ainda amplamente citada em publicações.

Markowitz (1959) mostra ainda a relação entre o retorno esperado, concebido como a média dos retornos, e a variância do retorno, concebido como o risco do Ativo. É analisado que, se houver um ativo com uma variância (risco) muito alta em relação aos outros ou uma média (retorno) muito baixa em relação aos outros ativos do portfólio, é possível que a formulação descarte este Ativo. Importante também: se dois ou mais ativos

tiverem o mesmo comportamento média-variância, a escolha de qualquer um deles ou a combinação deles resultará em peso igual na composição do portfólio.

Cass e Stiglitz (1970) descrevem que primeiro o investidor deve decidir qual proporção dos ativos de risco deve comprar. Então decide como dividir seu investimento total entre os ativos. Este tipo de separação é um caso especial mais generalizado de alocação ótima do investimento em um portfólio, ou seja, em um dado mercado em que existem ativos diferentes, todas as oportunidades relevantes para a decisão do investidor poder ser fornecidas por um conjunto de fundos de investimentos, um conjunto que pode ser definido por combinações lineares das proporções de cada Ativo no portfólio.

A administração da carteira pode ser ativa ou passiva. As carteiras mais produtivas são geralmente aquelas que tratam o investimento como Ativo. Ao contrário de títulos de investimentos, o Ativo pode ser melhorado de inúmeras maneiras (referido na indústria como tendo “potencial de crescimento”). Portanto, o desempenho da carteira depende da elaboração e execução de um plano de ação ideal de longo prazo para cada propriedade. O desempenho do investimento é marcadamente sensível a fatores ambientais externos. Além das condições gerais de negócios, o sucesso do investimento é sensível às taxas de juros, taxas de inflação (por exemplo, o custo de cláusulas de escalção, aluguel em locações), as políticas fiscais e fatores locais (TRIPPI, 1989).

As explicações matemáticas de como o portfólio funciona serão feitas utilizando o trabalho de Markowitz (1952). Supondo que há n ativos com retornos individuais $r_i, (i = 1, 2, \dots, n)$, o retorno do portfólio será dado por r_p . Como o portfólio é composto por diversos ativos, considere-se que w_i será o peso de cada um deles no montante do investimento. O retorno total pode ser então sumarizado em (4.1) e o risco em termos de variância dado na Equação (4.2) na qual Σ é a matriz de covariância contendo em sua diagonal principal as variâncias de cada ativo. Com estas equações é possível calcular a fronteira eficiente de Pareto. Cada ponto sobre a fronteira é considerado um portfólio ótimo contendo um retorno com seu risco relacionado.

$$\begin{aligned} r_p &= w_i \cdot x_i \\ r_p &= \mathbf{w}^T \cdot \mathbf{r} \end{aligned} \tag{4.1}$$

$$\sigma_p^2 = \mathbf{w}^T \Sigma \mathbf{w} \tag{4.2}$$

Obedecendo-se as restrições da Equação (4.3) (MARKOWITZ, 1959).

$$E = \sum_{i=1}^n X_i \mu_i \tag{4.3a}$$

$$\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n X_i X_j \sigma_{ij} \tag{4.3b}$$

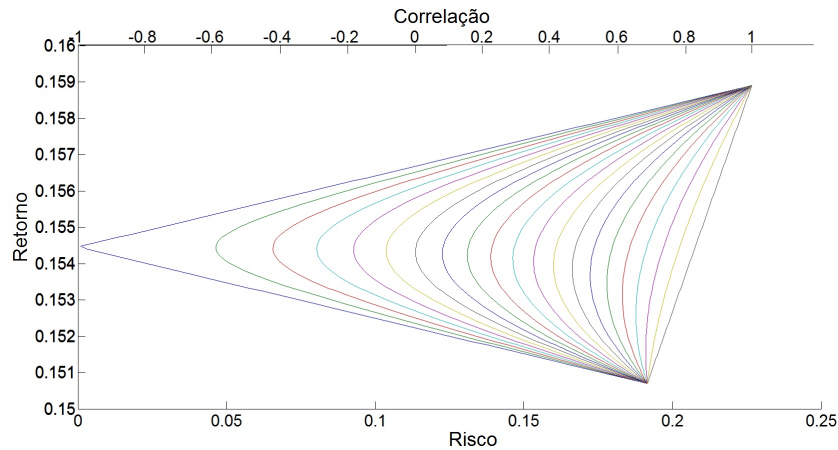


Figura 4.2 – Efeito da correlação no risco

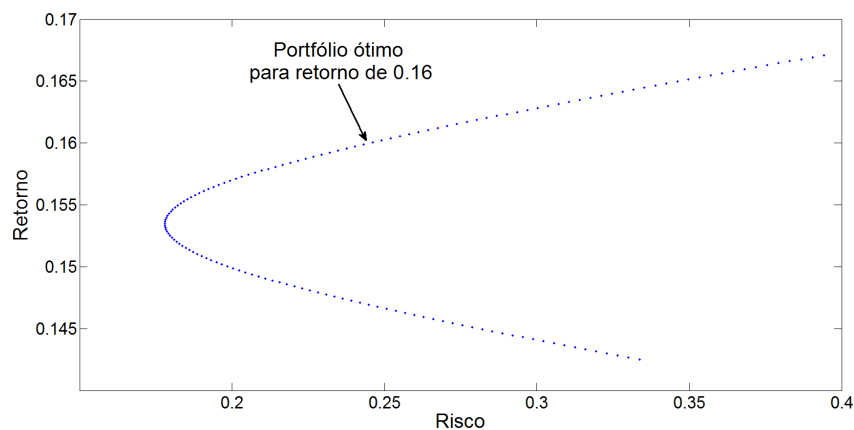


Figura 4.3 – Fronteira Eficiente e Pontos Ótimos

A proposta de utilização do método de *clusters* irá mitigar este problema causado pela correlação. Em termos gerais, a correlação entre ativos, quando fortemente positiva, pode fazer com que investimentos em empresas diferentes se comportem como sendo em apenas uma. Isto acontece porque neste caso as variações nos valores, dos ativos positivamente correlacionados, serão no mesmo sentido, subirão ou cairão juntas. No caso de serem fortemente correlacionados negativamente, o ganho ou perda de um ativo pode ser anulado por outro já que quando um preço da ação subir em um deles, cairá em relativa igualdade, no outro. Este é um efeito indesejado para o investidor e o *cluster* irá minimizá-lo porque empresas com situação similar dos seus indicadores financeiros serão colocadas em um mesmo grupo, *cluster*.

4.2.2 Análise de portfólio baseado em MDE

Suponha agora que os pesos w_i ou quantidades do modelo de média-variância de portfólio (MVP) são considerados proporções de uma mistura com as restrições de convexidade, por exemplo, $\sum w_i = 1$ e $w_i \geq 0$. Este é exatamente o caso de um projeto de experimentos de misturas conhecido como MDE.

Quando se compara o MDE para a abordagem MVP tradicional, é possível perceber que as equações de média-variância podem ser escritas como uma superfície de resposta da mistura, em que os montantes de investimento de capital em ativos n são definidos pelo tipo de experimento de mistura (*simplex-lattice* ou *simplex-centroid*). Além disso, a restrição orçamentária é análoga à restrição de convexidade do MDE. Uma vez que o experimento de misturas é definido, pode-se estimar a superfície de resposta para a média, variância e qualquer outra resposta de interesse, de modo que se pode comparar a influência de cada ativo para o desempenho do portfólio (OLIVEIRA *et al.*, 2011). No entanto, o modelo tradicional MDE não considera, por exemplo, venda a descoberto na análise do portfólio, de forma que uma generalização deve ser considerada. Oliveira *et al.* (2011) foram pioneiros na abordagem de experimentos de misturas para otimização de portfólios com foco em investimentos no setor energético e utilizado também por Mendes *et al.* (2016).

Um portfólio é considerado eficiente (ou Pareto ótimo) se todos os outros portfólios tiverem pelos menos um valor maior para ao menos uma coordenada do portfólio, caso contrário terá os mesmos valores para todas as coordenadas (BUCKLEY *et al.*, 2008).

4.3 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO

Neste capítulo, discutiram-se detalhes sobre seleção de portfólios e a relação de risco-retorno. Foram detalhadas as influências da correlação e relação com melhores ou piores resultados no portfólio. A interligação de portfólio e experimentos de mistura também foi citada. Esta é uma das características principais que possibilitaram o uso de metodologias, voltadas principalmente para engenharia química, em aplicações diretas a otimização de portfólios financeiros. É importante perceber também que portfólios podem ser compostos por outros objetos de estudo além do campo financeiro. O a programação da produção, quando voltada para uma variedade de produtos, por exemplo, pode se utilizar desta metodologia para escolher, através da função custo versus preço, os níveis de produção ótimos de cada produto.

5 METODOLOGIA EXPERIMENTAL E PROCEDIMENTOS DESCRITIVOS

5.1 CONSIDERAÇÕES INICIAIS

Nos capítulos anteriores, foram discutidos os conceitos fundamentais e necessários para a criação de um método de seleção de empresas e otimização considerando a incorporação de indicadores financeiros aos experimentos de mistura. As empresas consideradas como verdes e classificadas pela revista *Newsweek* foram utilizadas como alvo do estudo por causa da crescente visibilidade dessas empresas no mercado mundial e também pelas crescentes exigências governamentais voltadas a sustentabilidade e preservação do meio-ambiente.

A fim de avaliar a utilização de PCA e incorporação de indicadores financeiros para esta finalidade, foi necessário recorrer a um projeto experimental e coletar dados em forma de *data mining* com base histórica de dez anos. Este estudo pode trazer benefícios aos investidores por agrupar mais variáveis aos modelos tradicionais de portfólios.

5.2 METODOLOGIA

Esta pesquisa foi conduzida na forma de coleta de dados no campo exploratório. O método de pesquisa experimental com simulação é aplicado à metodologia proposta procedendo-se também a análise dos resultados. A Figura 5.1 é o macrofluxo da metodologia. Cada processo completo será dividido em partes e explicado individualmente.

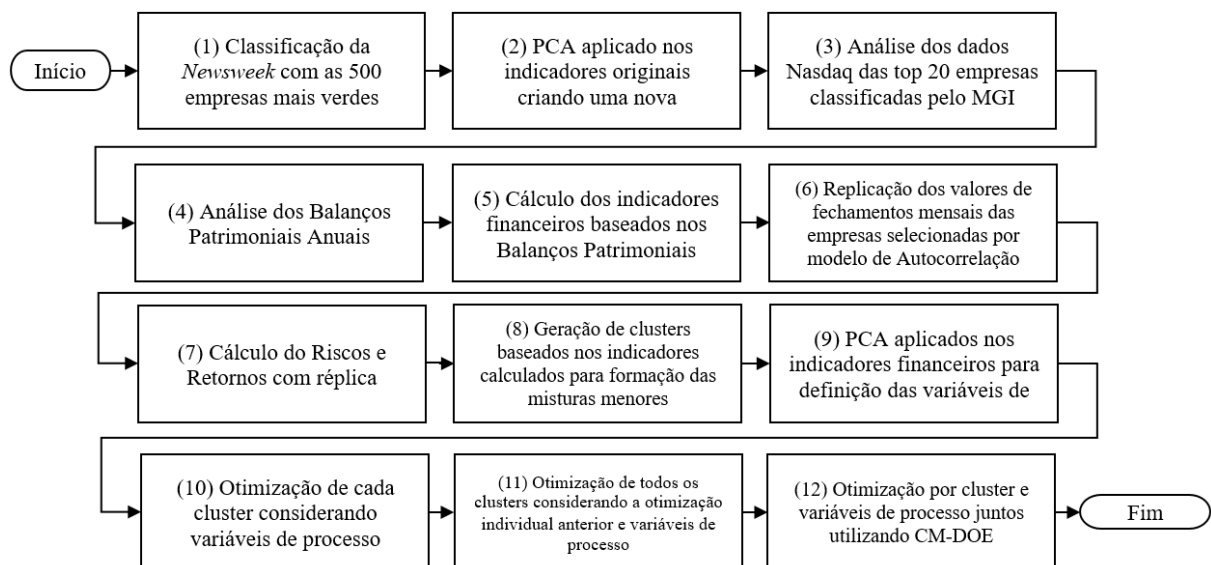


Figura 5.1 – Macrofluxo da metodologia proposta

5.3 MÉTODO INDICADORES FINANCEIROS INCORPORADOS AO AR-RANJO DE MISTURA CRUZADO

Para iniciar a modelagem identificou-se a necessidade primária de calcular os retornos das ações de cada empresa considerada. Diariamente as Bolsas de Valores inicia suas atividades em um horário determinado e encerra também em horário determinado. A diferença entre o valor da ação na abertura e o valor desta no fechamento do dia dá ao investidor o ganho ou perda do dia. Os valores considerados nesta tese foram os valores de fechamento mensais. Para identificar a variação percentual, ou seja, o retorno da ação em um dado fechamento de um mês qualquer para o mês anterior, aplicou-se a Equação (5.1), na qual Ln é logaritmo neperiano do fechamento atual e foi calculado sobre o fechamento do mês, ou período anterior. A mesma análise pode ser feita por dia, ano ou qualquer período que se deseje analisar. O logaritmo irá padronizar as escalas de valores para os fechamentos de todas as empresas, facilitando o entendimento do comportamento da ação no tempo e padronizando também a escala numérica. Com isso pode-se comparar o comportamento de ações que tenham preços diferentes.

A Figura 5.2 é a demonstração do comportamento da ação da empresa Apple Inc. (AAPL) mensalmente em dez anos. O eixo primário do gráfico é a cotação dada pelo fechamento mensal da ação. Parece haver uma ascensão no valor da ação, mas não é possível identificar a variação mês a mês. Olhando agora para o eixo secundário, identificamos tal variação com a utilização da Equação (5.1). Quando Rt_i for negativo, quer dizer que houve prejuízo no período calculado, quando Rt_i for positivo, houve lucro e finalmente, se Rt_i for zero, não houve variação no período.

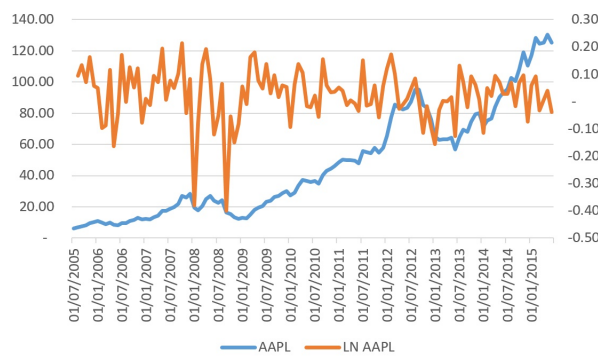


Figura 5.2 – Cotações e variação mensal AAPL

$$Rt_i = Ln \left(\frac{Pt}{Pt_{-1}} \right) \quad (5.1)$$

Com os valores de Rt_i calculados para cada empresa, podemos calcular os retornos das Ações e calcular também o risco de cada uma delas. Os retornos e riscos individuais são dados pela média de Rt na Equação (5.2) e o risco pela variância de Rt na Equação

(5.3). Para que o valor do risco tenha uma medida que faça sentido, no quesito numérico, utilizou-se ainda a medida de volatilidade da ação que é dada pela Equação 5.4.

$$\bar{R}t_i = \frac{Rt_1 + Rt_2 + Rt_3 + \dots + Rt_n}{n} \therefore \bar{R}t_i = \sum \frac{Rt_i}{n} \quad (5.2)$$

$$St_i^2 = \frac{(Rt_i - \bar{R}t_i)^2}{n - 1} \quad (5.3)$$

$$St_i = \sqrt{\frac{(Rt_i - \bar{R}t_i)^2}{n - 1}} \quad (5.4)$$

Os arranjos de mistura foram gerados conforme um algoritmo. Este algoritmo leva em consideração o tamanho do passo, ou resolução v , valor máximo da soma do portfólio pm , quantidade de ativos q , número de portfólios n , sendo que pm deverá ser um número inteiro divisível por v . O algoritmo é recursivo funcionando com a geração de todas as possíveis matrizes.

Considerando um caso de apenas três empresas com um portfólio único e resolução 0,5, tem-se a configuração $v = 0,5$, $pm = 1$ e $n = 2$ na Equação (5.5). Para o caso de quatro empresas composto por um MoM de dois portfólios e resolução igual a 0,5, a configuração será $v = 0,5$, $pm = 2$ e $n = 4$ na Equação (5.6).

$$\begin{bmatrix} w_1 & w_2 & w_3 \\ 0,0 & 0,0 & 1,0 \\ 0,0 & 0,5 & 0,5 \\ 0,5 & 0,5 & 0,0 \\ 0,5 & 0,0 & 0,5 \\ 0,0 & 1,0 & 0,0 \\ 1,0 & 0,0 & 0,0 \end{bmatrix} \sum w_i = 1 \quad (5.5)$$

$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ 1,0 & 1,0 & 0,0 & 0,0 \\ 1,0 & 0,0 & 1,0 & 0,0 \\ 1,0 & 0,0 & 0,0 & 1,0 \\ 0,0 & 1,0 & 1,0 & 0,0 \\ 0,0 & 0,0 & 1,0 & 1,0 \\ 0,5 & 0,5 & 0,5 & 0,5 \\ 0,5 & 0,5 & 1,0 & 0,0 \\ 0,5 & 0,5 & 0,0 & 1,0 \\ 0,5 & 1,0 & 0,5 & 0,0 \\ 0,5 & 0,0 & 0,5 & 1,0 \end{bmatrix} \sum w_i = 2 \quad (5.6)$$

Computacionalmente o algoritmo gera todas as possíveis matrizes com todos os números que atendam ao valor de v . Com as matrizes geradas, apenas as linhas que atendem a restrição $\sum w_i = pm$ são salvas e estas compõem a matriz da mistura. O número de combinações possíveis para 12 empresas seria $(n-1)! = (12-1)! = 39.916.800$. Porém para cada empresa pode-se escolher o valor percentual mínimo e máximo a aplicar, tornando esse número ainda maior. Para 24 empresas 282.429.536.481 (282 trilhões) de combinações possíveis. Esse problema foi resolvido pelo algoritmo gerador das matrizes da mistura que integra as contribuições do trabalho.

Seja as restrições para o valor resolução da mistura: $v = \gamma_i \therefore \Pi \gamma_1 * v = Pm$ e $\Pi \gamma_2 * v = n$.

γ deve ser um número inteiro \in Naturais

q = quantidade de empresas

n = quantidade de portfólios

v = resolução da mistura, quanto menor, maior o número de experimentos

Pm = percentual máximo a investir em apenas uma empresa.

O Pm tem uma outra função importantíssima, pode-se forçar a diversificação reduzindo seu valor para menor do que um. Desta forma a empresa não poderá compor sozinha um portfólio. Seja uma mistura formada por outras duas misturas. A matriz pode ser escrita como exemplificada na Equação (5.7). E em sua forma genérica na Equação (5.8). Cada termo da matriz γ é composto por uma mistura w . A quantidade de misturas dentro da matriz não é limitada. Esta formulação ainda torna possível criar restrições em qualquer nível da mistura e dimensionar restrições até o nível de variável w_{ij} . Isto quebra uma barreira deixando as restrições livres à necessidade do pesquisador, ou investidor.

$$\left[\begin{array}{cc} \gamma_1 & \gamma_2 \\ \begin{bmatrix} 0,0 & 1,0 \\ 1,0 & 0,0 \\ 0,0 & 1,0 \\ 1,0 & 0,0 \end{bmatrix} & \begin{bmatrix} 0,0 & 1,0 \\ 1,0 & 0,0 \\ 0,0 & 1,0 \\ 1,0 & 0,0 \end{bmatrix} \end{array} \right] \sum \gamma_i = 1 \quad (5.7)$$

$$\left[\begin{array}{cc} \gamma_1 & \gamma_2 \\ \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} & \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \end{array} \right] \sum w_i = 1; \sum \gamma_i = 1 \quad (5.8)$$

O investidor pode por exemplo optar impedir que seu investimento em um determinado portfólio seja concentrado em apenas uma ação, ou ainda, que haja concentração do investimento em um portfólio apenas. O conhecimento do investidor é peça fundamental nesse processo. E o modelo dá liberdade para utilizar essas restrições na Equação

(5.9) de \geq ou \leq , tanto em Ações individualmente quanto em portfólios, no caso do uso de MoM. O modelo pode ainda conter as variáveis de processo, que no nosso caso é definida pelo PCA. Na Equação (5.10) é exemplificado um caso com MoM (γ_i) e duas variáveis de processo (x_i), simplificado pela Equação (5.11).

$$\left[\begin{array}{c} \gamma_1 \\ \begin{bmatrix} 0,3 & 0,7 \\ 1,0 & 0,0 \\ 0,0 & 1,0 \\ 1,0 & 0,0 \end{bmatrix} \\ \gamma_2 \\ \begin{bmatrix} 0,1 & 0,9 \\ 1,0 & 0,0 \\ 0,4 & 0,6 \\ 1,0 & 0,0 \end{bmatrix} \end{array} \right] \sum \gamma_i = 1 \quad (5.9)$$

$$\left[\begin{array}{cccc} \gamma_1 & \gamma_2 & x_1 & x_2 \\ 1,0 & 0,0 & -1 & -1 \\ 0,0 & 1,0 & -1 & -1 \\ 0,5 & 0,5 & -1 & -1 \\ 1,0 & 0,0 & -1 & 1 \\ 0,0 & 1,0 & -1 & 1 \\ 0,5 & 0,5 & -1 & 1 \\ 1,0 & 0,0 & 1 & -1 \\ 0,0 & 1,0 & 1 & -1 \\ 0,5 & 0,5 & 1 & -1 \\ 1,0 & 0,0 & 1 & 1 \\ 0,0 & 1,0 & 1 & 1 \\ 0,5 & 0,5 & 1 & 1 \end{array} \right] \sum \gamma_i = 1 \quad (5.10)$$

$$\left[\begin{array}{c} \gamma_1 \\ \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \\ \gamma_2 \\ \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \\ x_1 & x_2 \\ -1 & -1 \\ -1 & 1 \\ 1 & -1 \\ -1 & -1 \end{array} \right] \sum w_i = 1; \sum \gamma_i = 1 \quad (5.11)$$

5.4 MÉTODO DE SIMULAÇÃO

A parte prática desta pesquisa foi realizada com o auxílio de simulação a partir de dados reais, históricos de fechamento das Ações. A simulação foi aplicada principalmente para replicar os valores destes fechamentos. Por causa dos dados terem a característica de

serem determinísticos, não seria possível calcular os erros entre as medições sem o uso das réplicas. A estratégia adotada foi desenvolvida combinando a Análise de Componentes Principais (PCA) para selecionar empresas para o estudo, indicadores de performance de empresas verdes, análise dos balanços patrimoniais das empresas selecionadas, criação de indicadores financeiros, medição de riscos e retornos, Delineamento de Experimentos (DOE), Experimentos de Misturas de Misturas (MoM), Experimentos Cruzados de Misturas (CM-DOE) e finalmente otimização multiobjetivo com a função *desirability*.

Este capítulo descreve as condições com que as análises e simulações foram feitas, tipos de algoritmos, técnicas de replicação bem como o desenvolvimento dos experimentos simulados conforme descrito pelo macrofluxo 5.1. Essas metodologias serão aplicadas nos estudos numéricos.

5.5 PROCEDIMENTO DESCRITIVO DA COLETA DOS DADOS

A seleção das empresas tem como objeto inicial de estudo as chamadas empresas verdes. A base com os nomes destas empresas vem da publicação da revista *Newsweek* edição de 2014 (Figura 5.3). Esta publicação é anual e lista as quinhentas melhores empresas no quesito sustentabilidade. A decisão em se utilizar os retornos destas empresas foi motivada pela publicação de [Lyon e Shimshack \(2012\)](#). Eles identificaram que as empresas verdes tendem a ter uma valorização de seus ativos após a publicação da revista. Considerou-se então os quesitos para seleção das empresas verdes como sendo o primeiro indicador a ser usado.

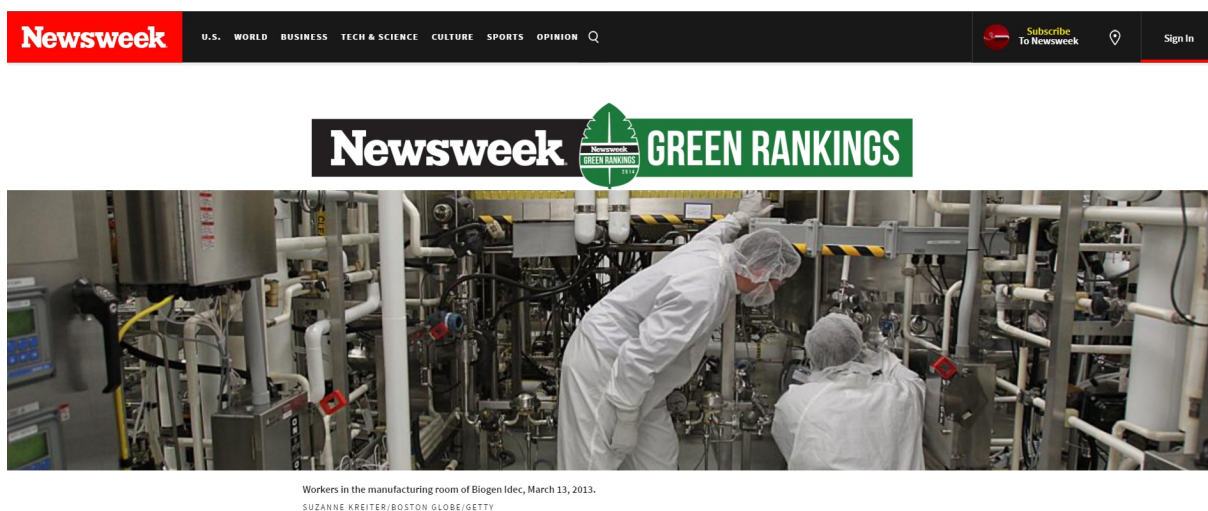


Figura 5.3 – Capa da revista *Newsweek* 2014

A *Newsweek* é responsável por divulgar as mais importantes métricas sobre meio-ambiente envolvendo as maiores empresas Norte-americanas e mundiais. Esta foi a primeira referência para iniciar a escolha das empresas que irão compor os portfólios.

Para aplicar o PCA, todos os indicadores foram considerados. São eles: produtividade energética, produtividade do carbono, produtividade da água, produtividade de desperdício e reputação no mercado. O PCA foi performado em três componentes, PC1, PC2 e PC3. Isto se deu porque este número foi suficiente para um grau de explicação das variáveis de 91,4%, conforme será visto nas fases do macrofluxo, etapa do PCA.

O processo de replicação dos dados dos fechamentos é feito sobre os históricos de mensais de cada ação. Os meses são então replicados por uso do modelo de Autocorrelação com passo de tamanho um para séries temporais. Isto possibilitou replicar os dados mensais mantendo as informações de tempo. O tamanho do passo utilizado então foi padronizado em $lag=1$ por este ter-se demonstrado aplicável para todas as empresas analisadas. A Figura 5.4 exemplifica o uso deste recurso na empresa Apple (AAPL).

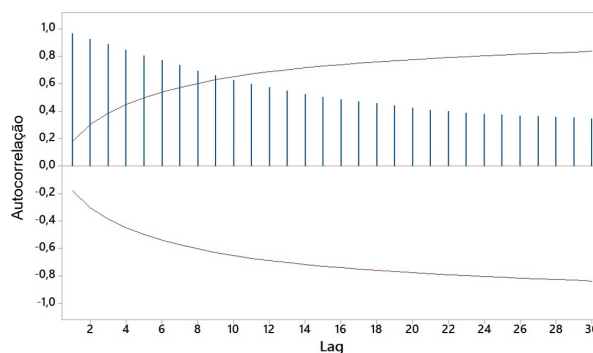


Figura 5.4 – Autocorrelação da empresa AAPL

5.6 PROCEDIMENTO DESCRITIVO DO PCA

O PCA permite que muitas variáveis sejam analisadas por meio de componentes principais. Permite ainda identificar discrepâncias entre elas. O foco no caso deste estudo foi agrupar os indicadores da revista *Newsweek* para formar uma nova classificação, uma classificação que considere matematicamente as interações entre os indicadores de empresas verdes.

O MGI, do inglês Multivariate general indicator, em português chamado de indicador geral multivariado, também foi utilizado para transformar os valores dos PCAs em indicadores de performance para cada empresa. A primeira aplicação foi na classificação original da *Newsweek*. Pela nova classificação do MGI foi possível eliminar um indicador, que para o investidor pode não ser tão importante, no caso, a reputação. Este indicador foi descartado pelo PCA porque é discrepante do conjunto analisado, não compôs por isso os componentes principais. De posse desta nova classificação pode-se passar a segunda aplicação do PCA.

A segunda aplicação do PCA é sobre a base de dados financeiros de cada empresa. Anualmente as empresas de capital aberto, conhecidas como sociedade anônima (S.A.),

que possuem ações em Bolsa de Valores, devem disponibilizar seus balanços patrimoniais anuais. Os balanços foram analisados para que as informações financeiras relevantes e necessárias para preencher e calcular os indicadores descritos na Tabela 2.1 fossem extraídas, preparados e finalmente calculados. Depois de calculados, todos foram submetidos ao PCA. Esta nova aplicação do PCA gerou o indicador MGI financeiro que será a base para as variáveis de processo consideradas nos modelos de arranjo de mistura cruzado, os maiores valores do MGI, em relação a média deles, serão as variáveis de processo classificadas como (+1) e os menores valores, em relação também às médias, serão aquelas classificadas como (-1).

5.7 PROCEDIMENTO DESCRITIVO DOS ARRANJOS EXPERIMENTAIS

Para otimização dos portfólios de ativos financeiros foram usados mais de um tipo de arranjo experimental. Em todos os casos, as empresas que são objeto de estudo foram pré-selecionadas pela classificação do PCA pelo MGI aplicado nas empresas verdes. Isto quer dizer que não há aleatoriedade na composição de nenhum tipo de portfólio de agora em diante nesta tese.

O primeiro experimento é por arranjo de misturas com *clusters*. Neste foram incluídas as empresas em grupos distintos para que ao final pudesse ser feita a otimização pelo método lexicográfico. As otimizações feitas nos *clusters* são as restrições para a otimização destes quando forem agrupados em um experimento global. Para isso, obteve-se cinco arranjos individuais. O primeiro para o *cluster* 1 com variável de processo relativa MGI (+1). O segundo para o *cluster* 1 com variável de processo relativa MGI (+1). O terceiro para o *cluster* 2 com variável de processo relativa MGI (+1). O quarto para o *cluster* 3 com variável de processo relativa MGI (+1). O quinto para o *cluster* 3 com variável de processo relativa MGI (-1). O *cluster* 2 com MGI (-1) não aparece porque seu grupo foi composto de apenas uma empresa. Mesmo assim esta empresa fará parte da otimização agrupada, chamada aqui de combinada.

Outro arranjo deve ser feito para o modelo combinado (CMO), que consiste em um arranjo global que agrupa todos os *clusters*. Este combina os pontos ótimos de cada *cluster* em um novo arranjo considerando as variáveis de processo (+1) e (-1), consiste então em um arranjo de misturas com variáveis de processo. O CMO é o primeiro método proposto que será testado e cuja eficiência deve ser comprovada nesta tese.

O próximo é um arranjo especial do tipo mistura de mistura com variável de processo (CM-DOE), neste os *clusters* são misturados entre si, sem uma otimização lexicográfica. É feita uma otimização direta considerando cada *cluster* como um vértice da mistura maior e suas empresas são como pontos nas misturas menores. Os PCAs irão formar as variáveis de processo (-1) e (+1). Este também é um método proposto na tese

e deverá ser comprovado assim como o CMO.

Finalmente, para efeito comparativo, um arranjo global no modelo Markowitz. Considera-se todas as empresas selecionadas pelo PCA sem separações ou agrupamentos e sem variáveis de processo. Tudo feito em uma mistura única a ser otimizada. Ainda para comparação utiliza-se, na simulação, um modelo que não possui arranjo ou otimização. Este modelo consiste em simplesmente dividir todo o investimento em partes iguais para as empresas participantes, no caso 100%/20 empresas. Lembrando que mesmo este modelo igualitário teve uma seleção prévia das empresas via PCA pelo MGI nos indicadores de empresas verdes.

5.8 PROCEDIMENTO DESCRITIVO DAS ANÁLISES

As análises são feitas em cada método separadamente e explicadas utilizando-se sumários das somas dos quadrados, análise de variância (ANOVA). Os gráficos dos resíduos normalizados, de avaliação dos erros padronizados para o arranjo, transformação Box-Cox e gráfico dos resultados de retorno (RP) e risco (SP) desses modelos estão agrupados nos Apêndices 7.0.1.

Para se manter as características e descritivos o mais corretos possível, optou-se por manter os gráficos com inscrições na língua original de sua geração, o inglês. As explicações pertinentes serão feitas em português.

5.9 PROCEDIMENTO DESCRITIVO DAS OTIMIZAÇÕES

Todos os arranjos apresentam diversas soluções. Estas soluções formam uma região ótima. Nesta região, somente um ponto é selecionado, aquele que obtiver o maior valor para o *desirability*. No caso do CMO, cada solução passa para a próxima otimização, seguindo o método lexicográfico. Após todas as soluções terem se agrupado, é então calculada a solução global para o CMO. Esta deve ser a melhor solução, pois é calculada a partir de soluções ótimas.

Todas as soluções são apresentadas em forma de tabela e aquela selecionada tem seu detalhamento feito com análise numérica e gráficos 2D e 3D da função *desirability*. A maior parte dos gráficos, quadros e tabelas deste processo serão alocados nos 7.0.1. Finalmente o ponto ótimo é mostrado por meio de tabela.

Todas as otimizações tem o mesmo objetivo alvo: Risco de 0,05 e maximização do Retorno. Porém se em algum caso este objetivo de risco não for possível, o mais próximo deste será considerado.

5.10 DESCRIÇÃO DOS DADOS

Os dados coletados, a partir da escolha das empresas classificadas como melhores *Green Companies* pelo PCA através do MGI aplicado na classificação da revista *Newsweek*, fazem parte de um grupo grande de empresas que possuem capital aberto e ações negociadas na Bolsa de Valores da Associação Nacional de Corretores de Títulos de Cotações Automáticas, NASDAQ (*National Association of Securities Dealers Automatic Quotation*) (NASDAQ, 2015). A NASDAQ iniciou suas operações em fevereiro de 1971. A principal contribuição desta foi prover uma melhor comunicação e reduzir os tempos de respostas para transações entre investidores e empresas (SANGER; MCCONNELL, 1986).

Sanger e McConnell (1986) enumeraram uma sequência de procedimentos para que uma empresa pudesse listar suas ações na NASDAQ: primeiro a empresa deve atender a alguns requisitos padrões como um faturamento mínimo anual; deve ter um valor mínimo de capital para disponibilizar para a venda; e um número mínimo de sócios que suportam o capital. Após isto, um documento deve ser preenchido na Divisão de Listagem de Ações. A empresa deve ainda registrar todas as Ações que irá disponibilizar para o mercado na Comissão de Seguros e Ações (SEC) na seção doze do *Securities* (1934). Este registro irá garantir, por um órgão externo, que a empresa preenche os requisitos necessários para negociar Ações na Bolsa de Valores. Este conjunto de dados tem como base os valores de compra e venda das ações de todas as empresas selecionadas. Pelo fato do mercado de ações ser muito dinâmico foi necessário estabelecer um ponto de corte para a extração dos dados que são base para os experimentos, todas as simulações, análises e cálculos. A data de corte foi sete de agosto de dois mil e quinze (07/07/2015) com início em sete de agosto de dois mil e nove (07/07/2009), totalizando cento e vinte fechamentos distribuídos em setenta e dois meses. As simulações para comprovar o desempenho de cada método descrito nesta tese datará, no entanto, até trinta e um de Julho de dois mil e dezessete. A Tabela 5.1 contém os vinte e quatro últimos fechamentos, em dólar, considerando as datas especificadas anteriormente e as empresas escolhidas pelo critério PCA com classificação por MGI.

Os históricos de valores de fechamentos mensais foram transformados em retornos e riscos mensais utilizando a Equação (5.1), formando assim a base para o cálculo dos retornos e riscos globais de cada ação para cada empresa. O retorno de cada ação é calculado pela média (Equação (5.2)) e o risco é calculado pela volatilidade (Equação (5.4)) de cada valor transformado. Esta operação retornará apenas um valor de média e um valor de volatilidade, ou seja, tem-se o retorno e o risco de cada ação por empresa selecionada. Para que o erro quadrático possa ser calculado foram feitas réplicas. Com a replicação dos experimentos é possível se calcular, através da ANOVA, as diferenças entre resultados de cada experimento, isto é medido pelo erro quadrático. As fases do macro-fluxo da Figura 5.1 são detalhadas nas próximas seções.

Tabela 5.1 – Histórico de 24 meses

ADBE	ADI	AAPL	BHI	BLL	BIIB	CLX	CCE	CMA	EMC	IHS	MAS	MET	TAP	MOS	NKE	QCOM	SIAL	S	WYN
79.87	64.19	125.43	61.70	71.43	403.94	104.02	43.44	51.32	26.39	128.63	23.81	55.99	69.81	46.85	108.02	62.63	139.35	4.56	81.91
79.09	67.96	130.28	64.46	70.99	396.99	107.66	44.23	48.95	26.34	123.40	24.17	52.26	73.38	45.85	101.67	69.68	139.30	4.65	84.91
76.06	61.84	125.15	68.46	73.41	373.93	106.10	44.41	47.41	26.91	125.47	23.65	51.29	73.51	44.00	98.84	68.00	138.92	5.13	85.40
73.94	63.00	124.43	63.58	70.64	422.24	110.39	44.20	45.13	25.56	113.76	23.84	50.55	74.45	46.06	100.33	69.34	138.25	4.74	90.47
79.10	58.54	128.46	62.51	71.71	409.59	108.64	46.20	45.78	28.94	117.53	23.38	50.83	75.89	53.26	97.12	72.51	138.06	5.12	91.48
70.13	52.11	117.16	57.99	63.33	389.16	106.71	42.10	41.50	25.93	115.13	22.18	46.50	75.93	48.69	92.25	62.46	137.52	4.30	83.79
72.70	55.52	110.38	56.07	68.17	339.45	104.21	44.22	46.84	29.74	113.88	22.50	54.09	74.52	45.65	96.15	74.33	137.27	4.15	85.76
73.68	54.64	118.93	57.00	67.07	307.69	101.62	43.94	46.61	30.35	122.46	21.60	55.61	77.35	45.77	99.29	72.90	136.60	5.12	83.36
70.12	49.62	108.00	52.96	64.43	321.08	99.50	43.35	47.74	28.73	131.03	19.70	54.24	74.38	44.31	92.97	78.51	135.91	5.93	77.67
69.19	49.49	100.75	65.06	63.27	330.81	96.04	44.36	49.86	29.26	125.19	21.35	53.72	74.44	44.41	89.20	74.77	136.01	6.34	81.26
71.90	51.12	102.50	69.14	64.10	343.04	88.60	47.78	50.34	29.53	142.47	20.95	54.74	73.95	47.76	78.55	76.10	104.00	5.61	80.94
69.25	49.63	95.60	68.77	61.26	334.39	86.87	45.45	50.26	29.30	131.37	18.57	52.60	67.53	46.11	77.13	73.72	100.42	7.35	75.55
72.36	54.07	92.93	74.45	62.68	315.31	91.40	47.78	50.16	26.34	135.67	19.82	55.56	74.16	49.45	77.55	79.20	101.48	8.53	75.72
64.54	52.38	90.43	70.52	60.36	319.37	89.62	45.64	47.97	26.56	125.91	19.02	50.93	65.73	49.99	76.91	80.45	98.53	9.55	73.93
61.69	51.29	84.30	69.90	56.19	287.12	90.70	45.44	48.24	25.80	120.63	17.94	52.35	59.97	50.04	72.95	78.71	96.21	8.50	71.34
65.74	53.14	76.68	65.02	54.81	305.87	88.01	47.76	51.80	27.41	121.50	19.83	52.80	58.86	50.00	73.86	78.86	93.38	9.19	73.23
68.63	50.82	75.18	63.28	55.56	340.68	87.28	47.08	48.18	26.37	119.88	20.85	50.67	56.83	48.86	78.30	75.29	94.41	8.74	72.88
59.19	48.27	71.51	56.64	51.19	312.64	88.27	43.29	45.80	24.24	113.41	18.89	49.05	52.64	44.66	72.85	74.22	92.97	8.27	70.94
59.88	50.93	80.15	55.26	51.66	279.57	92.76	44.13	47.54	25.15	119.70	20.33	53.92	56.15	47.27	78.64	74.25	94.01	10.75	73.69
56.78	48.22	79.44	56.96	49.98	290.97	93.17	41.94	45.35	23.85	114.43	20.02	52.19	52.67	47.90	79.14	73.58	86.24	8.39	71.71
54.22	49.30	74.67	58.09	48.89	244.19	90.19	41.73	43.30	24.07	109.05	18.86	47.31	54.00	45.85	75.76	69.49	86.43	6.73	66.40
51.94	47.05	68.11	49.10	44.88	240.76	81.72	40.21	39.31	25.56	114.18	19.00	46.95	50.13	43.02	72.64	67.32	85.30	6.22	60.97
45.75	46.28	69.60	46.49	44.42	213.02	82.70	37.40	40.84	25.78	107.15	16.89	46.19	48.79	41.65	62.82	66.28	82.47	6.71	59.36
47.28	49.38	64.65	47.43	44.79	218.13	85.94	37.54	42.54	26.15	109.78	18.32	48.42	50.06	41.09	62.92	64.56	83.56	5.96	62.30

5.11 FASES DO MACROFLUXO

5.11.1 Classificação da *Newsweek* com as 500 empresas mais verdes

“A revista *Newsweek* teve sua data de fundação em 17 de fevereiro de 1933. Ela é uma revista de primeira linha, composta também por um site, que leva jornalismo de alta qualidade para leitores de todo o mundo há mais de 80 anos. A *Newsweek* fornece as últimas notícias, análises aprofundadas e ideias sobre questões internacionais, tecnologia, negócios, cultura e política. Além de sua presença *on-line* e móvel, a *Newsweek* publica edições impressas semanais em inglês nos Estados Unidos, Europa / Oriente Médio / África e Ásia, além de edições em japonês, coreano, polonês, sérvio e espanhol. A *Newsweek* é de propriedade da *Newsweek Media Group*.”(CAPITAL, 2014).

A Figura 5.5 simplifica parte do Macrofluxo. Esta é a primeira mineração de dados.

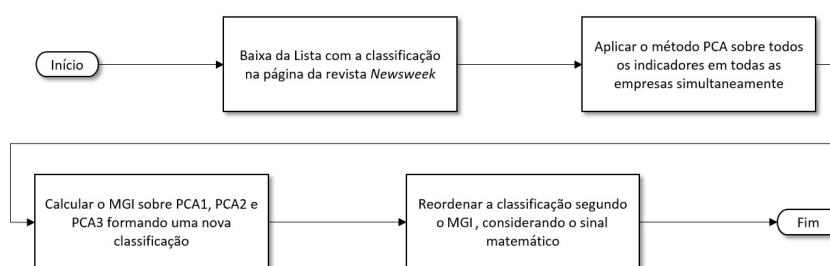


Figura 5.5 – Fluxo do PCA aplicado a *Newsweek*

Mineração de dados ou *data mining* é um termo usado para busca de dados em plataformas de internet. Consiste em baixa de dados, geralmente em grandes quantidades e a partir destes dados fazer uma seleção, extração, processamento, para poder transformá-los em informações relevantes e consistentes para os objetos de pesquisa ou análises. Empresas utilizam a mineração inclusive para distinguir gostos de consumidores a partir das suas buscas por produtos na internet.

A classificação da revista *Newsweek* (CAPITAL, 2014) Figura 5.6 é uma publicação anual que contempla quinhentas empresas norte-americanas dos setores de bens de consumo, energia, finanças, indústria, materiais, produtos de consumo, saúde, serviços de telecomunicação, tecnologia da informação e utilidades. Esta classificação é feita com base em indicadores de produtividade energética, produtividade do carbono, produtividade da água, produtividade de desperdício e reputação no mercado. Estes indicadores tem um lastro de 0,00% a 100% no qual maiores percentuais representam melhores (mais verdes) pontuações.

Para exemplificar, as quarenta e quatro empresas com sede nos Estados Unidos da América, melhores classificadas são listadas na Tabela 5.2. Esta tabela contém a classifi-



Figura 5.6 – Logotipo da classificação da *Newsweek*

cação original, antes do PCA. Note que há uma lacuna de seis empresas que não aparecem listadas na classificação de zero a cinquenta melhores. Estas foram retiradas por não possuírem histórico de transações em quantidades suficientes para os cálculos propostos ou que passaram por fusões com outras empresas no período selecionado. A tabela completa pode ser encontrada em [Papandrea e Paiva \(2016c\)](#).

5.11.2 PCA aplicado nos indicadores originais criando uma nova classificação pelo MGI

O PCA considera todos os indicadores originais e é dividido em PC1, PC2 e PC3. Para criar a nova classificação, o MGI (Multivariate general indicator) é calculado pela Equação (5.12). Os valores do MGI são então classificados em ordem crescente considerando-se o sinal positivo. A nova classificação é detalhada nas Tabela 5.4.

$$MGI = \sum W_{PC_{ij}} * Autovalor_{PC_i} \quad (5.12)$$

A Figura 5.7 possibilita a visão gráfica de três grupos distintos. Um composto apenas pela reputação, outro composto pela produtividade energética e produtividade do carbono e finalmente um composto por produtividade da água e produtividade do desperdício.

De posse dessa classificação, o PCA foi aplicado para que os indicadores pudessem ser agrupados (Figura 5.7). Com o PCA foi possível reduzir os indicadores de cinco para quatro (Figura 5.8) sendo estes quatro agrupados em apenas dois *clusters*. Finalmente a Tabela 5.3 tem os valores calculados nos PCAs e a Figura 5.9 mostra o agrupamento das empresas sem considerar a reputação.

Foram escolhidos três componentes principais porque o acumulado do percentual de explicação da análise de componentes principais foi de 91,4% para 3PCs. Esse montante percentual é considerado suficiente para as análises e é uma redução muito significativa na dimensionalidade do modelo. Esta é a principal característica e objetivo da utilização de componentes principais nas análises.

Destas quarenta e quatro, foram selecionadas as vinte melhores classificadas para as aplicações numéricas do método. Os seus logotipos, códigos e nomes estão listados na

Tabela 5.2 – Quarenta e quatro empresas melhores classificadas Original

Classific.	Empresa	Sector de Atuação	Subdivisão	Produtividade Energética	Produtividade do Carbono	Produtividade da Água	Produtividade de Desperdício	Reputação	Pontuação Verde Newsweek
1	Allergan	Saúde	Farmacêutico, Biotecnologia	72.20%	85.20%	61.10%	82.00%	100.00%	85.10%
2	Adobe Systems	Tec. Informação	Programas e Serviços	82.70%	87.10%	99.20%	91.90%	51.40%	84.40%
3	Ball Corporation	Materiais	Materiais	70.00%	75.20%	76.30%	83.20%	88.80%	83.50%
4	Ecolab	Materiais	Materiais	73.20%	80.10%	84.30%	59.60%	90.10%	82.60%
5	Sigma-Aldrich	Materiais	Materiais	78.30%	77.90%	88.30%	87.00%	100.00%	79.70%
6	McCormick	Produtos de Consumo	Comida, Bebida e Tabaco	85.80%	77.30%	79.40%	41.40%	100.00%	77.60%
7	Bigen Idec	Saúde	Farmacêutico, Biotecnologia	69.20%	82.70%	84.50%	97.00%	53.40%	75.70%
8	Rockwell Automation	Indústria	Bens de Capital	70.00%	47.90%	89.40%	83.80%	50.00%	73.70%
9	Cardinal Health	Saúde	Saúde, Equipamentos e Serviços	75.50%	70.10%	81.90%	64.60%	61.00%	71.00%
10	Agilent Technologies	Saúde	Farmacêutico, Biotecnologia	79.20%	71.50%	54.20%	56.70%	83.50%	70.90%
11	Baker Hughes	Energia	Energia	75.40%	74.90%	60.80%	14.10%	85.30%	70.80%
12	Harley-Davidson	Bens de Consumo	Automotivos Componentes	83.50%	55.90%	90.70%	68.30%	100.00%	69.80%
13	NIKE	Bens de Consumo	Consumo Duráveis Roupas	88.90%	61.60%	92.40%	79.90%	5.50%	69.50%
14	Juniper Networks	Tec. Informação	Tec. Equipamentos	35.90%	28.60%	65.40%	87.30%	84.40%	69.50%
16	Consolidated Edison	Utilidades	Utilidades	46.80%	52.90%	55.10%	77.40%	64.90%	67.80%
17	Applied Materials	Tec. Informação	Semicond.Equip.Semicondutores	60.90%	64.70%	71.90%	31.90%	89.30%	67.30%
18	Las Vegas Sands	Bens de Consumo	Serviços ao Consumidor	35.90%	60.60%	69.20%	49.00%	72.20%	66.60%
19	EMC	Tec. Informação	Tec. Equipamentos	44.90%	57.10%	81.60%	14.70%	83.50%	66.40%
20	IHS	Indústria	Serviços Prof.Comerciais	68.50%	78.40%	77.70%	83.60%	100.00%	66.20%
21	Mosaic	Materiais	Materiais	84.80%	50.20%	33.20%	18.30%	88.50%	65.70%
22	Coca-Cola Enterp.	Prod. Consumo	Comida, Bebida e Tabaco	59.00%	36.20%	59.60%	91.30%	4.40%	65.00%
23	Masco	Indústria	Bens de Capital	63.30%	84.90%	80.10%	35.40%	85.30%	64.40%
24	Altera	Tec. Informação	Semicond.Equip.Semicond.	75.60%	71.50%	69.10%	73.40%	100.00%	63.40%
25	QUALCOMM	Tec. Informação	Tec. Equipamentos	64.90%	74.10%	91.30%	85.80%	28.80%	63.20%
26	Autodesk	Tec. Informação	Programas e Serviços	63.10%	94.00%	30.40%	61.90%	100.00%	62.40%
27	NVIDIA	Tec. Informação	Semicond.Equip.Semicond.	80.80%	73.50%	93.90%	47.50%	64.90%	62.30%
28	MetLife	Finanças	Seguros	59.20%	92.10%	95.90%	81.50%	14.50%	62.20%
30	Praxair	Materiais	Materiais	27.30%	20.90%	54.70%	87.20%	67.60%	62.00%
31	Wyndham Worldwide	Bens de Consumo	Serviços ao Consumidor	38.80%	61.70%	35.90%	20.20%	90.60%	61.60%
32	Sprint	Serviços de Telecom.	Serviços em Telecom.	40.20%	35.10%	62.50%	49.10%	67.50%	61.50%
33	CA Incorporated	Tec. Informação	Programas e Serviços	34.30%	49.20%	57.80%	34.10%	74.40%	61.20%
34	United Technologies	Indústria	Bens de Capital	71.60%	38.30%	73.60%	42.80%	31.00%	60.10%
35	Johnson & Johnson	Saúde	Farmacêutico, Biotecnologia	71.70%	69.40%	73.60%	45.80%	5.00%	60.10%
37	General Mills	Produtos de Consumo	Comida, Bebida e Tabaco	54.50%	51.10%	68.10%	64.00%	21.40%	59.90%
38	Hewlett-Packard	Tec. Informação	Tec. Equipamentos	52.60%	67.30%	55.40%	83.30%	4.10%	59.60%
39	Delta Air Lines	Indústria	Transporte	12.10%	29.50%	88.80%	77.20%	41.60%	59.40%
40	Invesco	Finanças	Finanças Diversificadas	69.20%	48.80%	68.60%	38.60%	76.20%	59.00%
41	Molson Coors Brewing	Produtos de Consumo	Comida, Bebida e Tabaco	42.50%	60.40%	34.70%	9.50%	84.40%	58.90%
43	Clorox	Produtos de Consumo	Casa Produtos Pessoais	32.30%	32.80%	33.60%	49.40%	80.10%	58.20%
44	Activis	Saúde	Farmacêutico, Biotecnologia	71.20%	52.00%	66.80%	85.30%	84.40%	58.20%
46	Hershey	Produtos de Consumo	Comida, Bebida e Tabaco	85.50%	60.80%	67.50%	66.90%	26.60%	57.40%
47	Cisco Systems	Tec. Informação	Tec. Equipamentos	50.00%	60.50%	59.90%	82.30%	20.80%	57.10%
48	Apple	Tec. Informação	Tec. Equipamentos	91.80%	96.30%	95.80%	94.20%	1.00%	56.90%
50	Analog Devices	Tec. Informação	Semicond.Equipamentos Semicond.	55.30%	47.90%	59.40%	29.60%	86.40%	56.10%

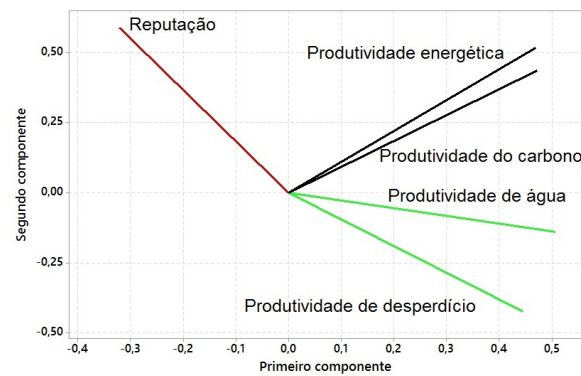


Figura 5.7 – Agrupamentos gerados pelo PCA incluindo reputação

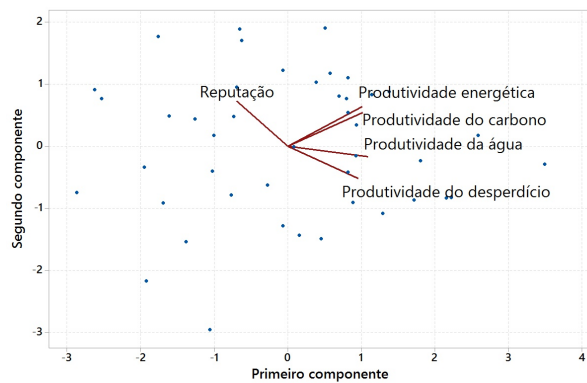


Figura 5.8 – Agrupamento com gráfico de pontos incluindo reputação

Tabela 5.3 – Eigen análise da matriz de correlação PCA

Autovalor	2,037	0,998	0,621	0,343
Proporção	0,509	0,250	0,155	0,086
Acumulado	0,509	0,759	0,914	1,000
Variável	PC1	PC2	PC3	
Produtividade energética	0,544	-0,463	-0,234	
Produtividade do carbono	0,532	-0,415	0,482	
Produtividade da água	0,506	0,387	-0,669	
Produtividade de desperdício	0,406	0,681	0,515	

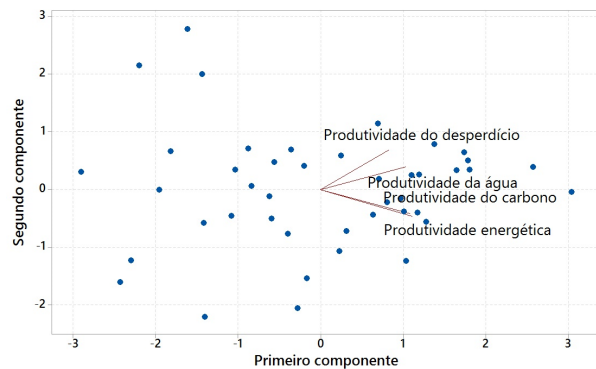


Figura 5.9 – Agrupamento dos fatores principais sem reputação *Newsweek*

Tabela 5.4 – Quarenta e quatro empresas melhores classificadas pelo PCA

Nome da Empresa	Pontuação PCA	Classificação MGI	Classificação Original	Produtividade Energética	Produtividade do Carbono	Produtividade da Água	Produtividade de Desperdício	Reputação	Verde Newsweek Pontuação
Apple	6.30	1	43	91.80%	96.30%	95.80%	94.20%	1.00%	56.90%
Adobe Systems	5.60	2	2	82.70%	87.10%	99.20%	91.90%	51.40%	84.40%
Biogen Idec	4.58	3	7	69.20%	82.70%	84.50%	97.00%	53.40%	75.70%
MetLife	4.31	4	27	59.20%	92.10%	95.90%	81.50%	14.50%	62.20%
Sigma-Aldrich	4.06	5	5	78.30%	77.90%	88.30%	87.00%	100.00%	79.70%
QUALCOMM	3.58	6	24	64.90%	74.10%	91.30%	85.80%	28.80%	63.20%
NIKE	3.21	7	13	88.90%	61.60%	92.40%	79.90%	5.50%	69.50%
IHS	3.00	8	19	68.50%	78.40%	77.70%	83.60%	100.00%	66.20%
Ball Corporation	2.77	9	3	70.00%	75.20%	76.30%	83.20%	88.80%	83.50%
Coca-Cola Enterprises	2.61	10	21	59.00%	84.20%	59.60%	91.30%	4.40%	65.00%
Allergan	2.43	11	1	72.20%	85.20%	61.10%	82.00%	100.00%	85.10%
Rockwell Automation	2.12	12	8	70.00%	47.90%	89.40%	83.80%	50.00%	73.70%
Ecolab	1.85	13	4	73.20%	80.10%	84.30%	59.60%	90.10%	82.60%
Harley-Davidson	1.85	14	12	83.50%	55.90%	90.70%	68.30%	100.00%	69.80%
Altera	1.63	15	23	75.60%	71.50%	69.10%	73.40%	100.00%	63.40%
Cardinal Health	1.62	16	9	75.50%	70.10%	81.90%	64.60%	61.00%	71.00%
NVIDIA	1.36	17	26	80.80%	73.50%	93.90%	47.50%	64.90%	62.30%
Activis	1.23	18	40	71.20%	52.00%	66.80%	85.30%	84.40%	58.20%
Hershey	0.78	19	41	85.50%	60.80%	67.50%	66.90%	26.60%	57.40%
Hewlett-Packard	0.76	20	35	52.60%	67.30%	55.40%	83.30%	4.10%	59.60%
Cisco Systems	0.51	21	42	50.00%	60.50%	59.90%	82.30%	20.80%	57.10%
McCormick	0.46	22	6	85.80%	77.30%	79.40%	41.40%	100.00%	77.60%
Agilent Technologies	-0.29	23	10	79.20%	71.50%	54.20%	56.70%	83.50%	70.90%
Johnson & Johnson	-0.32	24	33	71.70%	69.40%	73.60%	45.80%	5.00%	60.10%
Autodesk	-0.48	25	25	63.10%	94.00%	30.40%	61.90%	100.00%	62.40%
Consolidated Edison	-0.58	26	15	46.80%	52.90%	55.10%	77.40%	64.90%	67.80%
General Mills	-0.72	27	34	54.50%	51.10%	68.10%	64.00%	21.40%	59.90%
Juniper Networks	-0.85	28	14	35.90%	28.60%	65.40%	87.30%	84.40%	69.50%
Delta Air Lines	-0.89	29	36	12.10%	29.50%	88.80%	77.20%	41.60%	59.40%
Las Vegas Sands	-1.74	30	17	35.90%	60.60%	69.20%	49.00%	72.20%	66.60%
Applied Materials	-1.95	31	16	60.90%	64.70%	71.90%	31.90%	89.30%	67.30%
Praxair	-2.07	32	28	27.30%	20.90%	54.70%	87.20%	67.60%	62.00%
United Technologies	-2.15	33	32	71.60%	38.30%	73.60%	42.80%	31.00%	60.10%
Invesco	-2.25	34	37	69.20%	48.80%	68.60%	38.60%	76.20%	59.00%
Masco	-2.62	35	22	63.30%	36.90%	80.10%	35.40%	85.30%	64.40%
Baker Hughes	-2.92	36	11	75.40%	74.90%	60.80%	14.10%	85.30%	70.80%
Sprint	-3.31	37	30	40.20%	35.10%	62.50%	49.10%	67.50%	61.50%
EMC	-3.47	38	18	44.90%	57.10%	81.60%	14.70%	83.50%	66.40%
Analog Devices	-3.81	39	44	55.30%	47.90%	59.40%	29.60%	86.40%	56.10%
CA Incorporated	-4.07	40	31	34.30%	49.20%	57.80%	34.10%	74.40%	61.20%
Mosaic	-5.15	41	20	84.80%	50.20%	33.20%	18.30%	88.50%	65.70%
Clorox	-5.19	42	39	32.30%	32.80%	35.90%	49.40%	80.10%	58.20%
Wyndham Worldwide	-5.50	43	29	38.80%	61.70%	35.90%	20.20%	90.60%	61.60%
Molson Coors Brewing	-6.30	44	38	42.50%	60.40%	34.70%	9.50%	84.40%	58.90%

Tabela 5.1. O número vinte foi escolhido por ser um número suficiente para aplicação de todas as fases da proposta.

5.11.3 Análise dos dados Nasdaq das 20 empresas melhores classificadas pelo MGI

A Figura 5.10 simplifica parte do Macrofluxo. Esta é a segunda mineração de dados. Todas as cotações das vinte empresas melhores pontuadas pelo MGI referentes aos últimos

Quadro 5.1 – Logotipo e códigos das empresas selecionadas





















Logotipo	Código	Nome
	AAPL	Apple
	ADBE	Adobe Systems
	ADI	Analog Devices
	BHI	Baker Hughes
	BIIB	Biogen
	BLL	Ball Corporation
	CCE	Coca-cola Co.
	CLX	Clorox
	CMA	Comerica Inc.
	EMC	EMC Corporation
	IHS	IHS Incorporation
	MAS	Masco
	MET	Met Life
	MOS	Mosaic
	NKE	Nike
	QCOM	Qualcomm
	S	Sprint
	SIAL	Sigma
	TAP	Molson Coors
	WYN	Wyndham Worldwide



Figura 5.10 – Fluxo para baixa e preparação dos dados

dez anos foram baixadas diretamente da base de dados da Bolsa de Valores Nasdaq-EUA. Cada empresa teve seus valores analisados e os fechamentos mensais segregados para serem utilizados nas análises. Estes valores são a base de dados para os cálculos de retornos e riscos de cada empresa. As tabelas completas podem ser encontradas em [Papandrea e Paiva \(2016b\)](#).

Para efeito ilustrativo, a Figura 5.11 mostra o comportamento histórico das Ações das empresas selecionadas, do mais recente para o mais antigo valor da ação de cada uma

das empresas selecionadas pelo PCA.

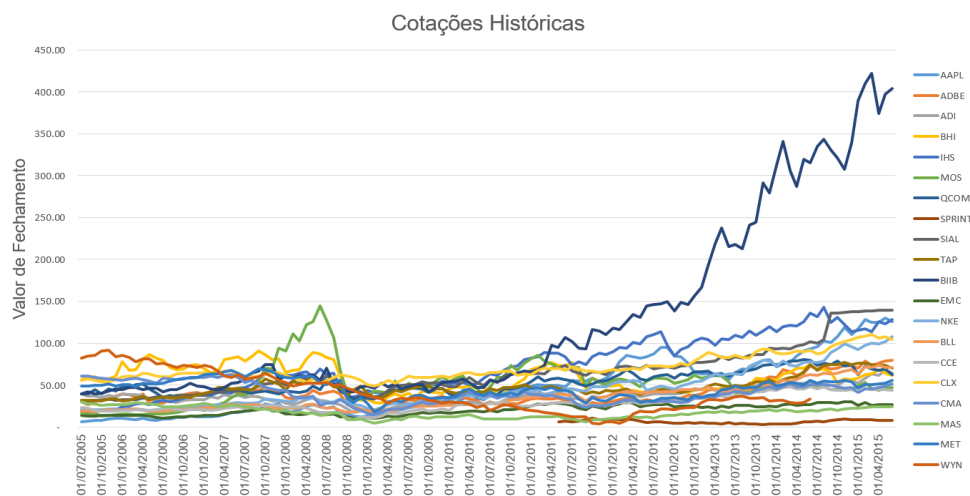


Figura 5.11 – Cotações históricas

5.11.4 Análise dos Balanços Patrimoniais anuais

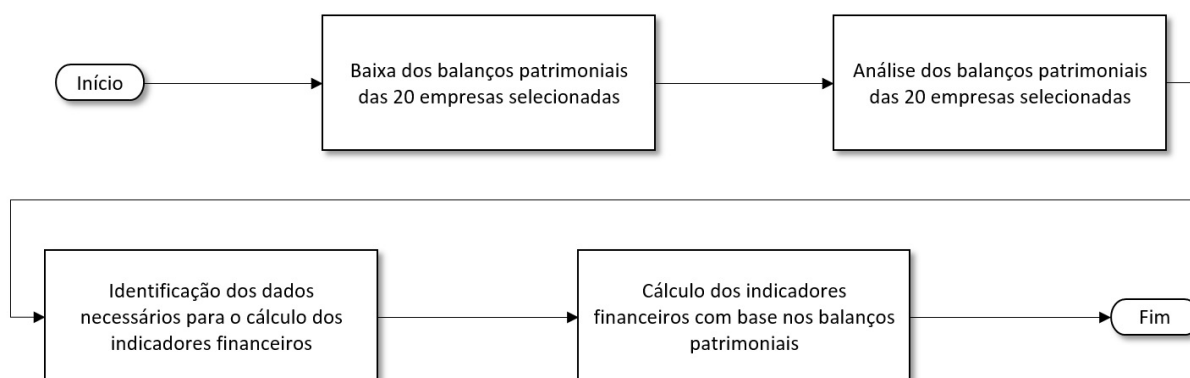


Figura 5.12 – Fluxo para baixa e análise dos balanços patrimoniais

A Figura 5.12 simplifica parte do Macrofluxo. Esta é a terceira mineração de dados. Para as mesmas vinte empresas, os Balanços Patrimoniais anuais foram baixados individualmente. A divulgação desses balanços é obrigatória para todas as empresas que possuem capital aberto. Esses dados podem ser encontrados nos *websites* de cada empresa ou na própria base de dados da Bolsa de Valores Nasdaq-EUA. As informações contidas nesses balanços são necessárias para o cálculo dos indicadores financeiros de cada empresa. Esses indicadores compõem as variáveis de processo dos arranjos de mistura cruzados.

5.11.5 Cálculo dos indicadores financeiros dos Balanços Patrimoniais

Os dados necessários para o cálculo de cada indicador descrito na Tabela 2.1 foram localizados e extraídos dos balanços patrimoniais e então calculados um a um. Os balanços obrigatório *10K* tem uma formatação padrão, porém é livre às empresas a forma

de divulgação das informações, contanto que aquelas obrigatórias estejam presentes. Essas informações podem assumir diversas formas. A mineração de dados foi importante para localização de cada uma das variáveis em cada um dos balanços patrimoniais de cada uma das empresas. Essa mineração também é uma contribuição deste trabalho. As variáveis de cálculo e os indicadores são mostrados nas Tabelas 5.5 e 5.6. As planilhas completas estão também disponíveis em (PAPANDREA; PAIVA, 2016d).

Indicadores financeiros são uma medida de risco das empresas. Este é um dos motivos da obrigatoriedade na divulgação dos balanços patrimoniais das empresas listadas nas bolsas de valores. O investidor deve se manter informado sobre a “saúde financeira” das empresas das quais possui ou pretende possuir Ações, se tornar investidor. Infelizmente os modelos tradicionais de gestão de portfólio não contemplam esses indicadores em seus cálculos. A inovatividade do uso desses indicadores como variáveis de processo aglomeradas por PCA é uma das contribuições importantes deste trabalho. A Figura 5.13 mostra, como exemplo a página título do modelo padrão de divulgação do balanço patrimonial da empresa Apple Corporation. A Figura 5.14 tem marcado alguns dos dados extraídos para o cálculo dos indicadores financeiros, estes dados são ainda, detalhados na Figura 5.15.

APPLE INC.
(Exact name of registrant as specified in its charter)

California
(State or other jurisdiction of incorporation or organization)
1 Infinite Loop
Cupertino, California
(Address of principal executive offices)

94-2404110
(I.R.S. Employer Identification No.)
95014
(Zip Code)

Registerant's telephone number, including area code: (408) 996-1010

Common Stock, \$0.00001 par value per share
(Title of class)

Securities registered pursuant to Section 12(b) of the Act
Securities registered pursuant to Section 12(g) of the Act: None

The NASDAQ Stock Market LLC
(Name of exchange on which registered)

Indicate by check mark if the registrant is a well-known seasoned issuer, as defined in Rule 405 of the Securities Act: Yes ☐ No ☒

Indicate by check mark if the registrant is not required to file reports pursuant to Section 13 or Section 15(d) of the Act: Yes ☐ No ☒

Indicate by check mark whether the registrant (1) has filed all reports required to be filed by Section 13 or 15(d) of the Securities Exchange Act of 1934 during the preceding 12 months (or for such shorter period that the registrant was required to file such reports), and (2) has been subject to such filing requirements for the past 90 days: Yes ☒ No ☐

Indicate by check mark whether the registrant has submitted electronically and posted on its corporate Web site, if any, every Interactive Data File required to be submitted and posted pursuant to Rule 405 of Regulation S-T (§322.405 of this chapter) during the preceding 12 months (or for such shorter period that the registrant was required to submit and post such files): Yes ☒ No ☐

Indicate by check mark if disclosure of delinquent filers pursuant to Item 405 of Regulation S-K (§229.405 of this chapter) is not contained herein, and will not be contained, to the best of the registrant's knowledge, in definitive proxy or information statements incorporated by reference in Part III of this Form 10-K or any amendment to this Form 10-K: ☒

Indicate by check mark whether the registrant is a large accelerated filer, an accelerated filer, a non-accelerated filer, or a smaller reporting company. See the definitions of "large accelerated filer," "accelerated filer" and "smaller reporting company" in Rule 12b-2 of the Exchange Act.

Large accelerated filer ☒ Non-accelerated filer ☐ (Do not check if a smaller reporting company) Accelerated filer ☐ Smaller reporting company ☐

Indicate by check mark whether the registrant is a shell company (as defined in Rule 12b-2 of the Act): Yes ☐ No ☒

The aggregate market value of the voting and non-voting stock held by non-affiliates of the registrant, as of March 28, 2014, the last business day of the registrant's most recently completed second fiscal quarter, was approximately \$462,522,000,000. Solely for purposes of this disclosure, shares of common stock held by executive officers and directors of the registrant as of such date have been excluded because such persons may be deemed to be affiliates. This determination of executive officers and directors as affiliates is not necessarily a conclusive determination for any other purposes.

5,864,840,000 shares of common stock were issued and outstanding as of October 10, 2014.

Figura 5.13 – Página título do formulário 10K, balanço patrimonial Apple

Item 6. Selected Financial Data					
The information set forth below for the five years ended September 27, 2014, is not necessarily indicative of results of future operations, and should be read in conjunction with Part II, Item 7, "Management's Discussion and Analysis of Financial Condition and Results of Operations" and the consolidated financial statements and related notes thereto included in Part II, Item 8 of this Form 10-K to fully understand factors that may affect the comparability of the information presented below (in millions, except number of shares, which are reflected in thousands, and per share amounts)					
	2014	2013	2012	2011	2010
Net sales	\$ 182,795	\$ 170,910	\$ 156,508	\$ 108,249	\$ 65,225
Net income	\$ 39,910	\$ 37,037	\$ 41,733	\$ 25,922	\$ 14,013
Earnings per share:					
Basic	\$ 6.49	\$ 5.72	\$ 6.38	\$ 4.01	\$ 2.20
Diluted	\$ 6.45	\$ 5.68	\$ 6.34	\$ 3.96	\$ 2.16
Cash dividends declared per share	\$ 1.82	\$ 1.64	\$ 0.38	\$ 0	\$ 0
Shares used in computing earnings per share:					
Basic	6,085,672	6,477,320	6,543,726	6,469,806	6,360,224
Diluted	6,122,663	6,521,634	6,617,483	6,556,014	6,472,883
Total cash, cash equivalents and marketable securities	\$ 155,239	\$ 146,761	\$ 121,251	\$ 81,370	\$ 51,011
Total assets	\$ 231,639	\$ 207,000	\$ 176,064	\$ 116,371	\$ 75,183
Commercial paper	\$ 6,308	\$ 0	\$ 0	\$ 0	\$ 0
Long-term debt	\$ 29,987	\$ 16,960	\$ 0	\$ 0	\$ 0
Other long-term obligations (1)	\$ 24,826	\$ 20,208	\$ 16,664	\$ 10,100	\$ 5,631
Total liabilities	\$ 120,292	\$ 83,451	\$ 57,854	\$ 30,156	\$ 27,392
Total shareholders' equity	\$ 111,547	\$ 123,549	\$ 118,210	\$ 76,615	\$ 47,791

(1) Other long-term obligations exclude non-current deferred revenue

Apple Inc. | 2014 Form 10-K | 24

Figura 5.14 – Exemplo de dado retirado para cálculo do indicador financeiro

Os gráficos que demonstram o relacionamento entre os indicadores pode ser visto utilizando-se o gráfico multivariado de pontos na Figura 5.16. Neste gráfico é possível notar relacionamentos fortes entre os indicadores. Estes relacionamentos devem ser absorvidos e utilizados pelo PCA.

2014	2013	2012
\$ 182,795	\$ 170,910	\$ 156,508
\$ 39,510	\$ 37,037	\$ 41,733
\$ 6.49	\$ 5.72	\$ 6.38
\$ 6.45	\$ 5.68	\$ 6.31
\$ 1.82	\$ 1.64	\$ 0.38
6,085,572	6,477,320	6,543,726
6,122,663	6,521,634	6,617,483
\$ 155,239	\$ 146,761	\$ 121,251
\$ 231,839	\$ 207,000	\$ 176,064
\$ 6,308	\$ 0	\$ 0
\$ 28,987	\$ 16,960	\$ 0

Figura 5.15 – Detalhe de dado retirado para cálculo de indicador financeiro

Tabela 5.5 – Análise de débitos em milhões de dólares por empresa

Empresas	Dívida total	Capital total	Débitos sobre capital total	Total de débitos	Capital total próprio	Débitos sobre capital próprio
AAPL	\$ 120.292,00	\$ 231.839,00	51,89%	\$ 120.292,00	\$ 111.547,00	107,84%
ADBE	\$ 4.009,92	\$ 10.785,83	37,18%	\$ 4.009,92	\$ 6.775,91	59,18%
ADI	\$ 2.101,79	\$ 3.811,89	55,14%	\$ 2.101,79	\$ 4.757,90	44,17%
BHI	\$ 4.133,00	\$ 28.827,00	14,34%	\$ 4.161,00	\$ 18.730,00	22,22%
BIIB	\$ 3.502,52	\$ 14.316,56	24,46%	\$ 3.502,52	\$ 10.814,04	32,39%
BLL	\$ 6.331,40	\$ 7.571,00	83,63%	\$ 6.331,40	\$ 1.239,60	510,76%
CCE	\$ 7.112,00	\$ 8.543,00	83,25%	\$ 7.112,00	\$ 1.431,00	497,00%
CLX	\$ 4.104,00	\$ 4.258,00	96,38%	\$ 4.104,00	\$ 154,00	2664,94%
CMA	\$ 61.788,00	\$ 69.190,00	89,30%	\$ 61.788,00	\$ 7.402,00	834,75%
EMC	\$ 22.360,00	\$ 45.885,00	48,73%	\$ 22.360,00	\$ 23.525,00	95,05%
IHS	\$ 3.188,88	\$ 5.348,43	59,62%	\$ 3.188,88	\$ 2.159,55	147,66%
MAS	\$ 6.039,00	\$ 7.167,00	84,26%	\$ 6.039,00	\$ 1.128,00	535,37%
MET	\$ 829.679,00	\$ 902.337,00	91,95%	\$ 829.679,00	\$ 72.560,00	1143,44%
MOS	\$ 7.562,40	\$ 18.283,00	41,36%	\$ 7.562,40	\$ 10.720,60	70,54%
NKE	\$ 7.770,00	\$ 18.594,00	41,79%	\$ 7.770,00	\$ 10.824,00	71,78%
QCOM	\$ 9.408,00	\$ 48.574,00	19,37%	\$ 9.408,00	\$ 39.169,00	24,02%
SIAM	\$ 1.065,00	\$ 2.247,00	47,40%	\$ 1.065,00	\$ 3.130,00	34,03%
SPRINT	\$ 59.377,00	\$ 84.689,00	70,11%	\$ 59.377,00	\$ 25.312,00	234,58%
TAP	\$ 6.110,20	\$ 13.996,30	43,66%	\$ 6.110,20	\$ 7.886,10	77,48%
WYN	\$ 8.422,00	\$ 9.679,00	87,01%	\$ 8.422,00	\$ 1.257,00	670,01%

5.11.6 Replicação dos valores de fechamentos mensais das empresas selecionadas por modelo de autocorrelação

A replicação é necessária para o cálculo dos erros nas simulações. Como os dados são de origem determinística, não haveria possibilidade de medição desses erros sem o uso de alguma técnica de replicação. Os valores de fechamentos mensais das Ações de cada uma das empresas selecionadas foi replicado utilizando-se o modelo ARMA-GARCH com $lag = 1$. O tamanho do passo igual a um foi escolhido porque acolhia todas as empresas selecionadas. Assim este padrão proporcionou a replica de todos os meses e a possibilidade de cálculo dos erros. As replicas foram feitas sobre o valor original dos fechamentos, sem

Tabela 5.6 – Análise dos indicadores de taxas de juros

Empresas	Lucro antes de juros e impostos	Despesas de juros	Juros ganhos sobre impostos	Lucro antes de taxas e impostos	Taxas fixas	Cobertura de taxa fixa	Fluxo de caixa operacional ajustado	Despesas de juros	Juros ganhos em dinheiro
AAPL	\$ 53.483,00	\$ 10.365,00	516%	\$ 53.483,00	\$ 13.973,00	383%	\$ 13.844,00	\$ 10.365,00	134%
ADBE	\$ 361,38	\$ 68,89	525%	\$ 361,38	\$ 92,98	389%	\$ 1.117,40	\$ 68,89	1622%
ADI	\$ 729,35	\$ 34,78	2097%	\$ 729,35	\$ 37,32	1954%	\$ 569,23	\$ 34,78	1636%
BHI	\$ 2.627,00	\$ 232,00	1132%	\$ 2.627,00	\$ 232,00	1132%	\$ 1.740,00	\$ 232,00	750%
BIIB	\$ 3.946,60	\$ 29,50	13378%	\$ 3.946,60	\$ 29,50	13378%	\$ 1.204,92	\$ 29,50	4084%
BLL	\$ 630,00	\$ 183,10	344%	\$ 630,00	\$ 183,10	344%	\$ 1,50	\$ 183,10	1%
CCE	\$ 893,00	\$ 230,00	388%	\$ 893,00	\$ 230,00	388%	\$ 223,00	\$ 230,00	97%
CLX	\$ 861,00	\$ 103,00	836%	\$ 861,00	\$ 103,00	836%	\$ 329,00	\$ 103,00	319%
CMA	\$ 870,00	\$ 1.655,00	53%	\$ 870,00	\$ 1.655,00	53%	\$ 6.071,00	\$ 1.655,00	367%
EMC	\$ 3.762,00	\$ 147,00	2559%	\$ 3.762,00	\$ 147,00	2559%	\$ 6.343,00	\$ 147,00	4315%
IHS	\$ 249,20	\$ 55,38	450%	\$ 249,20	\$ 55,38	450%	\$ 153,16	\$ 55,38	277%
MAS	\$ 575,00	\$ 225,00	256%	\$ 575,00	\$ 225,00	256%	\$ 160,00	\$ 225,00	71%
MET	\$ 5.936,00	\$ 1.151,00	516%	\$ 5.936,00	\$ 1.151,00	516%	\$ 10.808,00	\$ 1.151,00	939%
MOS	\$ 1.217,30	\$ 128,90	944%	\$ 1.217,30	\$ 128,90	944%	\$ 2.400,00	\$ 128,90	1862%
NKE	\$ 3.577,00	\$ 33,00	10839%	\$ 3.577,00	\$ 33,00	10839%	\$ 2.220,00	\$ 33,00	6727%
QCOM	\$ 8.778,00	\$ 1.228,00	715%	\$ 8.778,00	\$ 1.228,00	715%	\$ 7.907,00	\$ 1.228,00	644%
SIAL	\$ 677,00	\$ 8,00	8463%	\$ 677,00	\$ 8,00	8463%	\$ 4,00	\$ 8,00	50%
SPRINT	\$ 95,00	\$ 516,00	18%	\$ 95,00	\$ 516,00	18%	\$ 4.970,00	\$ 516,00	963%
TAP	\$ 586,30	\$ 145,00	404%	\$ 586,30	\$ 145,00	404%	\$ 624,60	\$ 145,00	431%
WYN	\$ 845,00	\$ 113,00	748%	\$ 845,00	\$ 113,00	748%	\$ 183,00	\$ 113,00	162%
AAPL	\$ 13.844,00	\$ 13.973,00	99%	\$ 59.713,00	\$ 11.000,00	543%	\$ 59.713,00	\$ 120.292,00	50%
ADBE	\$ 1.117,40	\$ 92,98	1202%	\$ 1.287,50	\$ 10.785,83	12%	\$ 1.287,50	\$ 4.009,92	32%
ADI	\$ 569,23	\$ 37,32	1525%	\$ 871,60	\$ 178,00	490%	\$ 871,60	\$ 2.101,79	41%
BHI	\$ 1.740,00	\$ 232,00	750%	\$ 1.731,00	\$ 171,00	1012%	\$ 1.731,00	\$ 4.161,00	42%
BIIB	\$ 1.204,92	\$ 29,50	4084%	\$ 2.942,10	\$ 175,00	1681%	\$ 2.942,10	\$ 3.502,52	84%
BLL	\$ 1,50	\$ 183,10	1%	\$ 1.012,50	\$ 390,80	259%	\$ 1.012,50	\$ 6.331,40	16%
CCE	\$ 223,00	\$ 230,00	97%	\$ 982,00	\$ 332,00	296%	\$ 982,00	\$ 7.112,00	14%
CLX	\$ 329,00	\$ 103,00	319%	\$ 558,00	\$ 138,00	404%	\$ 558,00	\$ 4.104,00	14%
CMA	\$ 6.071,00	\$ 1.655,00	367%	\$ 593,00	\$ 179,00	331%	\$ 593,00	\$ 61.788,00	1%
EMC	\$ 6.343,00	\$ 147,00	4315%	\$ 6.523,00	\$ 979,00	666%	\$ 6.523,00	\$ 22.360,00	29%
IHS	\$ 153,16	\$ 55,38	277%	\$ 628,10	\$ 114,45	549%	\$ 628,10	\$ 3.188,88	20%
MAS	\$ 160,00	\$ 225,00	71%	\$ 602,00	\$ 128,00	470%	\$ 602,00	\$ 6.039,00	10%
MET	\$ 10.808,00	\$ 1.151,00	939%	\$ 6.336,00	\$ 906,00	699%	\$ 6.336,00	\$ 829.679,00	1%
MOS	\$ 2.374,60	\$ 128,90	1842%	\$ 1.303,40	\$ 929,10	140%	\$ 1.303,40	\$ 7.562,40	17%
NKE	\$ 2.220,00	\$ 33,00	6727%	\$ 2.693,00	\$ 922,00	292%	\$ 2.693,00	\$ 7.770,00	35%
QCOM	\$ 7.907,00	\$ 1.228,00	644%	\$ 7.534,00	\$ 1.185,00	636%	\$ 7.534,00	\$ 9.408,00	80%
SIAL	\$ 4,00	\$ 8,00	50%	\$ 640,00	\$ 118,00	542%	\$ 640,00	\$ 1.065,00	60%
SPRINT	\$ 4.970,00	\$ 516,00	963%	\$ 522,00	\$ 1.488,00	35%	\$ 522,00	\$ 59.377,00	1%
TAP	\$ 624,60	\$ 145,00	431%	\$ 517,80	\$ 259,50	200%	\$ 517,80	\$ 6.110,20	8%
WYN	\$ 183,00	\$ 113,00	162%	\$ 183,00	\$ 235,00	78%	\$ 183,00	\$ 8.422,00	2%

nenhuma interferência matemática. Somente após isso é que as Equações (5.1), (5.2) e (5.4) foram aplicadas. Todos os fechamentos e réplicas encontram-se em Papandrea e Paiva (2016e).

5.11.7 Cálculo dos riscos e retornos com réplicas

A Figura 5.17 traz parte do Macrofluxo. Os riscos e retornos de cada empresa são calculados pela aplicação das Equações (5.1), (5.2) e (5.4). O mesmo cálculo foi feito tanto para os valores de fechamentos originais quanto para os valores replicados. As cotações replicadas e os valores logarítmicos podem ser acessados em Papandrea e Paiva (2016e).

5.11.8 Geração de *Clusters* baseados nos indicadores calculados para formação de misturas menores

Com os indicadores de cada empresa calculados, pode-se formar os *clusters*. As empresas foram agrupadas pelo método de Ward com distância Euclidiana (Figura 5.18, Tabela 5.2). A Figura 5.19 mostra como seriam os agrupamentos se os dados originais

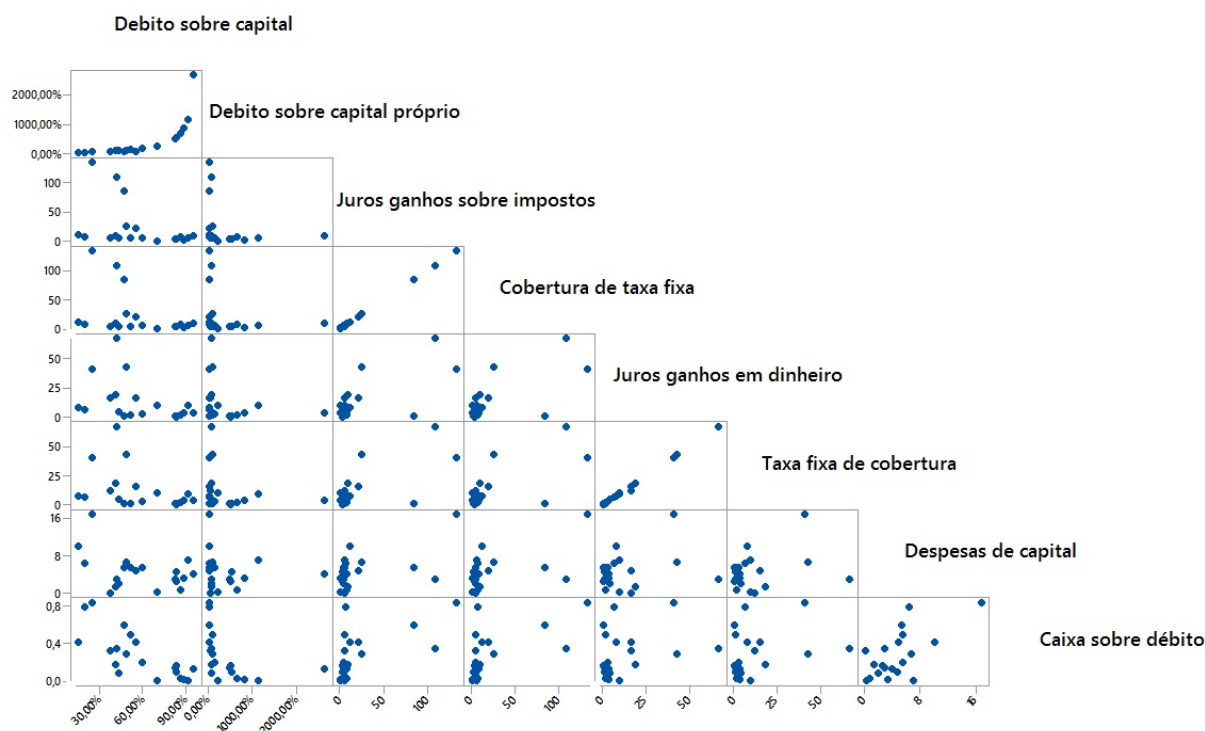


Figura 5.16 – Gráfico multivariado dos indicadores financeiros

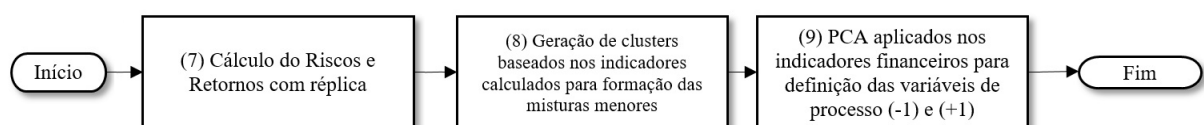


Figura 5.17 – Fluxo do risco, clusters e variáveis de processo

da *Newsweek* fossem utilizados. A Figura 5.21 tem a divisão dos agrupamentos de acordo com a classificação calculada e tem como base os indicadores débito sobre capital, débito sobre capital próprio, juros ganhos sobre os impostos, cobertura de taxa fixa, juros ganhos em dinheiro, taxa fixa de cobertura, despesas de capital e caixa sobre débito operacional da Tabela 2.1.

Quadro 5.2 – Divisão das empresas por *cluster*

Nome da Empresa	Código Nasdaq	Cluster	Nome da Empresa	Código Nasdaq	Cluster
Apple Inc.	AAPL	1	Biogen Idec Inc.	BIIB	2
Adobe Systems Incorporated	ADBE	1	Nike, Inc.	NKE	2
Analog Devices, Inc.	ADI	1	EMC Corporation	EMC	2
Baker Hughes Incorporated	BHI	1	Masco Corporation	MAS	3
IHS Inc.	IHS	1	MetLife, Inc.	MET	3
The Mosaic Company	MOS	1	Ball Corporation	BLL	3
Qualcomm Incorporated	QCOM	1	Coca-Cola Enterprises, Inc.	CCE	3
Sigma-Aldrich Corporation	SIAL	1	The Clorox Company	CLX	3
Sprint Corporation	SPRINT	1	Comerica Incorporated	CMA	3
Molson Coors Brewing Company	TAP	1	Wyndham Worldwide Corporation	WYN	3

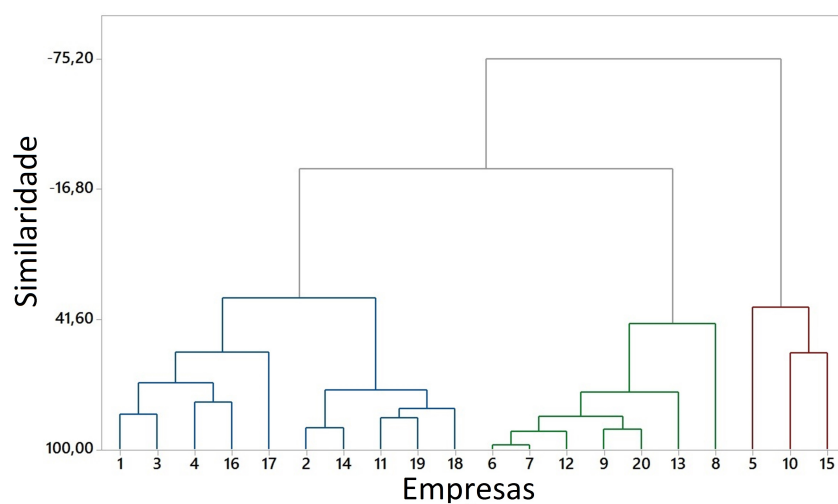


Figura 5.18 – Dendrograma, Divisão considerada dos Clusters pelos Indicadores calculados

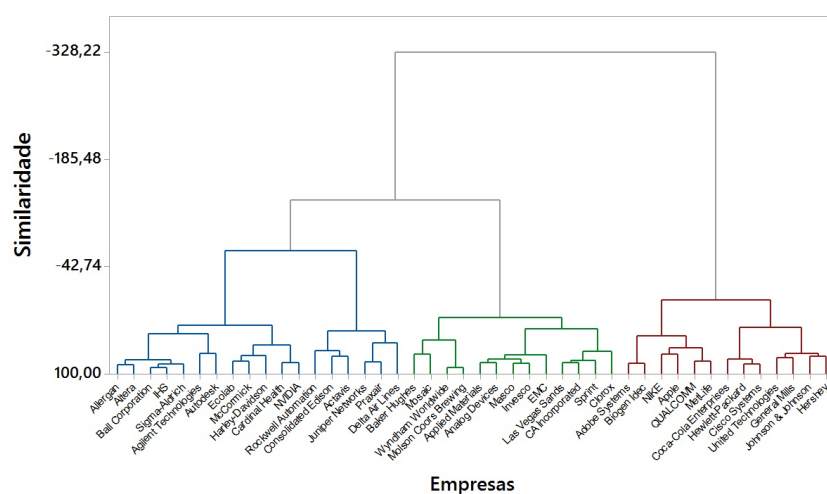


Figura 5.19 – Dendrograma, Divisão dos Clusters por PCA *Newsweek*

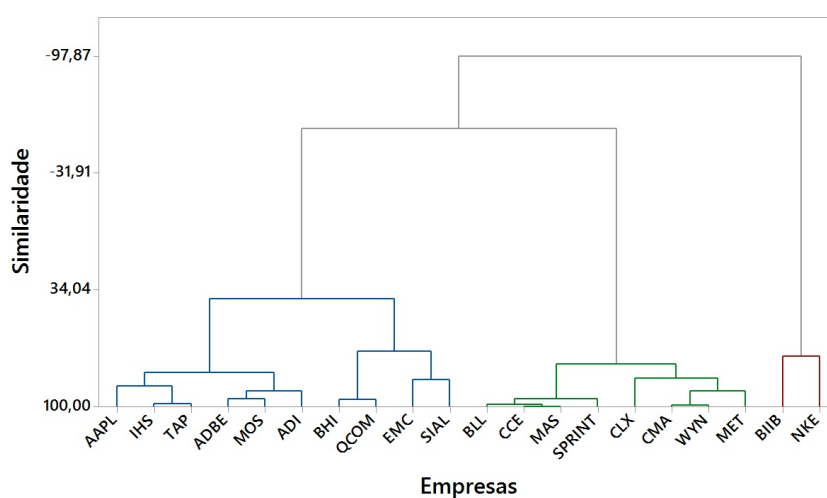


Figura 5.20 – Dendrograma, Divisão dos Cluster por Indicadores

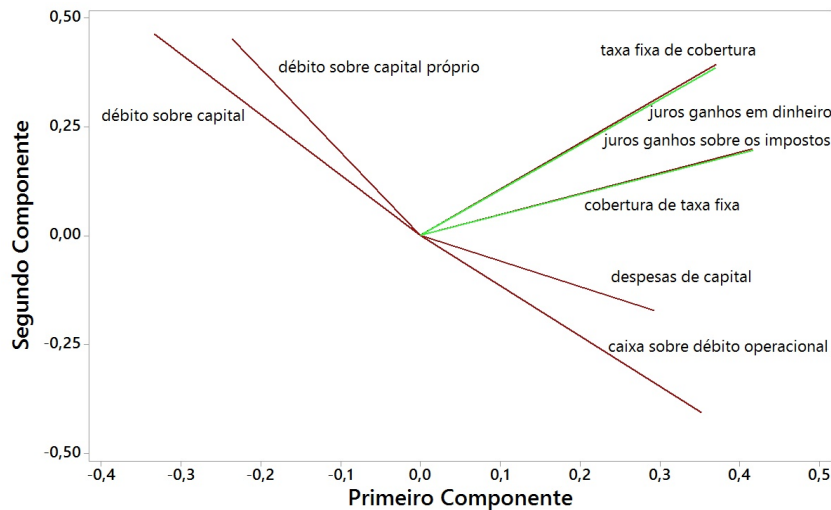


Figura 5.21 – Agrupamento dos fatores principais dos Indicadores

5.11.9 PCA aplicado nos indicadores financeiros para definição das variáveis (-1) e (+1)

Os indicadores financeiros débito sobre capital, débito sobre capital próprio, juros ganhos sobre os impostos, cobertura de taxa fixa, juros ganhos em dinheiro, taxa fixa de cobertura, despesas de capital e caixa sobre débito operacional da Tabela 2.1 calculados e os resultados desses indicadores apresentados nas Tabela 5.5, 5.6 e ??; foram agrupados pelo método de Ward com distância Euclidiana e então relacionados em três *clusters* (Figura 5.22). Formaram-se os PCA1, PCA2 e PCA3. Por similaridade, estes foram agrupados em *clusters* definindo quais empresas integram quais *clusters* (Figura 5.23). É importante saber que cada *cluster* compõe um portfólio e que cada um deles será otimizado. Cada PCA compõe um portfólio com duas variáveis de processo. O valor mais baixo em cada PCA define qual empresa será definida como variável de processo -1 enquanto que o maior valor em cada PCA define qual empresa será a variável de processo +1.

O processo de *cluster* permite agrupar, por similaridade, várias empresas em um mesmo portfólio. Isso permite que mais empresas possam ser analisadas ao mesmo tempo e também proporciona um entendimento de como a composição do portfólio foi feita. Neste caso não é mais uma composição baseada nas experiências pessoais do investidor, mas sim na similaridade dos retornos em um determinado período de tempo. O resultado é apresentado Tabela 5.7 e 5.8.

5.12 LIMITAÇÕES DO MODELO

A limitação do modelo é quanto a geração das matrizes das misturas e é dependente dos recursos computacionais e quantidade de empresas presentes nos portfólios. Se o

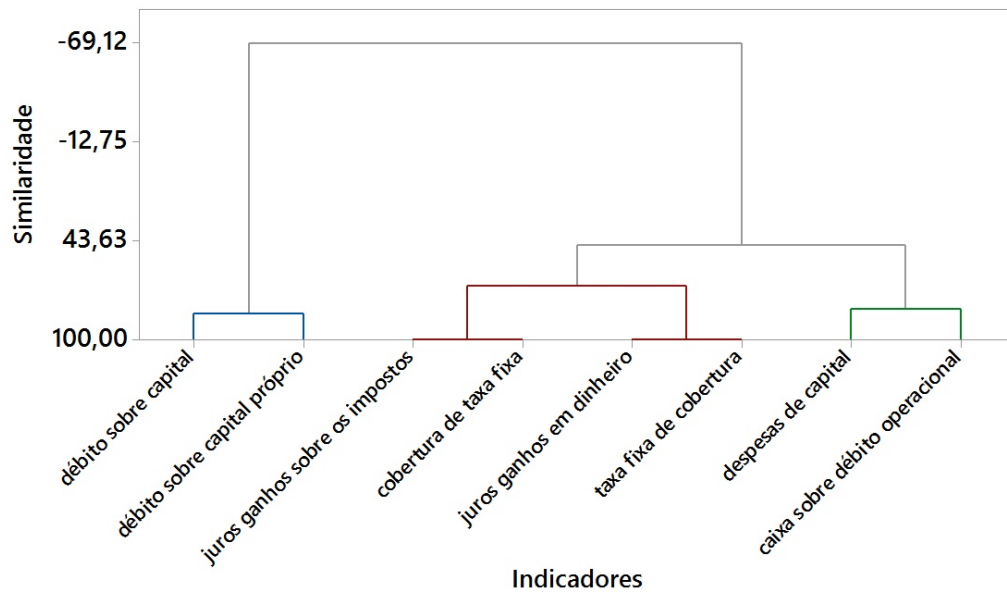


Figura 5.22 – Dendrograma, similaridades entre indicadores

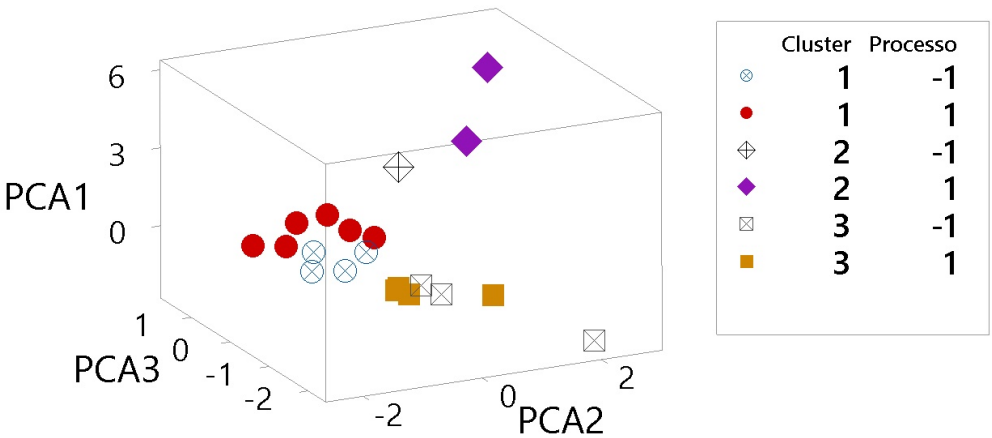


Figura 5.23 – Cluster em cada PCA

Tabela 5.7 – Divisão dos clusters e valores de PCA

Clusters	Cluster 1										Cluster 2			Cluster 3						
Empresas	AAPL	ADBE	ADI	BHI	IHS	MOS	QCOM	SPRINT	SIAL	TAP	BIIB	EMC	NKE	BLL	CCE	CLX	CMA	MAS	MET	WYN
Retorno	0.025	0.008	0.004	0.001	0.017	0.008	0.004	-0.012	0.012	0.007	0.020	0.006	0.014	0.011	0.005	0.005	-0.001	-0.002	0.001	0.008
Risco	0.098	0.094	0.071	0.114	0.078	0.133	0.075	0.147	0.059	0.069	0.089	0.081	0.064	0.059	0.085	0.041	0.084	0.124	0.104	0.161
Retorno Réplica	0.025	0.007	0.005	0.000	0.014	0.010	0.003	-0.018	0.013	0.007	0.020	0.006	0.015	0.012	0.007	0.005	-0.001	-0.002	0.001	0.009
Risco Réplica	0.097	0.095	0.071	0.116	0.079	0.135	0.074	0.147	0.061	0.071	0.092	0.084	0.066	0.059	0.088	0.042	0.086	0.129	0.107	0.167
PCA1	-0.307	-0.210	0.471	0.874	-0.769	-0.089	0.921	1.631	-1.404	-0.906	5.883	1.753	4.436	-1.631	-1.573	-2.293	-1.892	-1.559	-1.424	-1.913
PCA2	-1.401	-0.612	-0.428	-1.876	-0.681	-0.101	-2.324	-0.909	0.364	-0.619	0.052	0.867	2.716	0.076	0.124	2.036	0.754	0.129	1.163	0.670
PCA3	-0.217	1.378	0.294	-0.284	0.016	1.375	-0.131	-1.078	1.085	0.924	-1.993	1.023	1.561	-0.125	-0.116	-2.229	-0.264	-0.381	-0.967	0.130

estudo considerar um grande número de empresas, os recursos de informática devem ser capazes de computar grandes massas de dados para conseguirem gerar as matrizes de misturas.

Tabela 5.8 – Agrupamento das empresas por *cluster* e variáveis de processo (-1) e (+1)

Cluster / Variável de Processo	Empresa	Retorno	Risco	Réplica Retorno	Réplica Risco	MGI Financeiro
Cluster 1 (-1)	AAPL	0,0254	0,0977	0,0247	0,0969	-0,4638
	IHS	0,0165	0,0784	0,0137	0,0788	-0,5653
	SIAL	0,0123	0,0594	0,0130	0,0610	-0,5916
	TAP	0,0067	0,0693	0,0071	0,0706	-0,5085
Cluster 1 (+1)	ADBE	0,0083	0,0944	0,0074	0,0952	-0,0444
	ADI	0,0041	0,0709	0,0047	0,0714	0,2322
	BHI	0,0007	0,1142	0,0001	0,1162	0,1185
	MOS	0,0083	0,1334	0,0102	0,1346	0,1188
	QCOM	0,0039	0,0754	0,0030	0,0739	0,0840
	SPRINT	-0,0116	0,1468	-0,0176	0,1474	0,6227
Cluster 2 (-1)	EMC	0,0055	0,0811	0,0061	0,0836	1,3076
Cluster 2 (+1)	BIIB	0,0196	0,0895	0,0200	0,0921	3,1193
	NKE	0,0138	0,0644	0,0151	0,0656	3,2642
Cluster 3 (-1)	CLX	0,0052	0,0412	0,0050	0,0418	-1,2492
	CMA	-0,0015	0,0835	-0,0011	0,0863	-0,9850
	WYN	0,0084	0,1610	0,0088	0,1670	-0,9588
Cluster 3 (+1)	BLL	0,0111	0,0586	0,0116	0,0594	-0,9408
	CCE	0,0052	0,0855	0,0073	0,0878	-0,8976
	MAS	-0,0020	0,1244	-0,0016	0,1286	-0,9252
	MET	0,0011	0,1037	0,0010	0,1067	-0,7372

5.13 SUPERANDO AS LIMITAÇÕES

Uma opção viável é dividir os portfólios compostos por muitas empresas em portfólios menores e recursivamente ir eliminando do modelos as empresas que forem descartadas pela otimização, ou seja, aquelas em que o $w = 0$. Analogamente ao processo de algoritmos genéticos que leva para as próximas gerações apenas os melhores resultados da geração atual. Utilizar processamento paralelo também pode ser uma opção, dependendo do recurso computacional disponível.

5.14 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO

Neste capítulo, discutiu-se a metodologia utilizada para os experimentos, no caso simulados, e descreveu-se todos os procedimentos para coleta, clusterização e análise dos dados. Também a metodologia utilizada para otimização dos portfólios.

O próximo capítulo se dedica à aplicação desta metodologia analisando cada função e otimizando-as, a fim de consolidar o PCA e o uso de indicadores financeiros como parte integrante na análise e seleção de portfólios.

6 SELEÇÃO E OTIMIZAÇÃO DE PORTFÓLIOS COM INDICADORES FINANCEIROS

6.1 CONSIDERAÇÕES INICIAIS

Nos capítulos anteriores, foram discutidos os conceitos necessários para o entendimento da bibliografia e a criação de um método de seleção de empresas e otimização considerando a incorporação de indicadores financeiros aos experimentos de mistura. Discutiu-se a metodologia utilizada para os experimentos, e descreveu-se todos os procedimentos para coleta, clusterização e análise dos dados. Também a metodologia utilizada para otimização dos portfólios e método lexicográfico foram discutidos.

Este capítulo se reserva à aplicação da metodologia com análises das funções individuais, aplicação do PCA e MGI como variáveis de processo para a otimização dos portfólios. A finalidade é consolidar o PCA e o uso de indicadores financeiros como parte integrante na análise e seleção destes portfólios. A otimização pela metodologia lexicográfica também é uma grande contribuição. Isso torna os portfólios mais robustos e confiáveis para o investidor já que o portfólio combinado será formado apenas por portfólios já otimizados. Poderá haver variação nos tamanhos das letras nas tabelas e quadros para que possam se encaixar melhor nos espaços das páginas. Em cada coluna das tabelas dos experimentos tem-se:

- Padrão: ordem de execução padrão do experimento enumerada crescentemente de um até o número de linhas necessárias para cobrir todos os pontos especificados no arranjo experimental;
- Execução: para experimentos que não utilizam simulação, esta coluna será a ordem de execução real para os experimentos, aleatorizada para evitar qualquer viés durante essas execuções;
- Bloco: A blocagem é uma técnica usada para remover matematicamente a variação causada por alguma alteração identificável durante o curso do experimento. Por exemplo, pode ser necessário usar dois lotes de matérias-primas diferentes para concluir o experimento ou o experimento ou o mesmo pode ocorrer ao longo de vários turnos ou dias. Para cada um desses casos, a alteração pode fazer com que os dados de resposta sejam alterados. A blocagem remove essa mudança e, com efeito, “normaliza” os dados. Para o caso em questão não foi necessário a blocagem;
- Componentes: Cada coluna refere-se a um componente da mistura, tendo sua variação de zero a um no caso de não ter restrição no componente. Nos arranjos desta tese,

os componentes serão as empresas e no arranjo CMO serão os *clusters* otimizados;

- Respostas: são os valores observados em cada linha do experimento, podem também ser calculados quando se trata de portfólios. A resposta 1 é dada pela função do retorno da mistura e calculada pela Equação de RP (4.1). Para a Resposta 2, cálculo do risco, a função é dada pela Equação de SP (4.2).

No caso do arranjo CM-DOE há colunas extras com o nome de Mix-Mix. Estas são relacionadas aos termos de misturas de misturas (MoM) e variam como um componente normal da mistura e também podem ser restritas de acordo com a necessidade do experimento. A otimização das funções geradas, a partir dos experimentos, são colocadas em um subitem ao final das modelagens referentes a cada método. Para que a comparação possa ser feita e entendida mais claramente, os alvos das otimizações foram definidos como: Maximizar o retorno com risco de 0,05. Para os casos em que o risco não puder ser menor ou igual a esse nível, será considerada a minimização também para o risco.

A partir daqui, são feitas as modelagens para cada arranjo e posteriormente sua otimização. Modelagem é uma declaração matemática do relacionamento entre variáveis. É genericamente, uma equação, função que explica quais são as consequências que modificações nos parâmetros de entrada causam nas respostas de saída. Estes modelos são invenções intelectuais que servem para descrever, explicar e prever sobre essas mudanças nas saídas em virtude de alterações nas entradas, parâmetros. Alguns modelos matemáticos são mais acurados do que outros. Se os erros nas previsões forem insignificantes, então esse tipo de modelo será distinguido de outros e chamado de modelo determinístico. Por outro lado se o erro de previsão for significativo, o que é o caso mais usual, esse modelo terá diferentes níveis de explicação. Nestes casos, o erro não pode ser ignorado. O erro de previsão deve ser conhecido e incorporado nos modelos. Quando este é incorporado, o modelo correspondente será denominado de modelo estatístico ou probabilístico. É possível ainda, medir o erro em problemas determinísticos. Uma das formas de se fazer isto é a que foi utilizada nesta tese, a replicação das amostras. Com as réplicas é possível fazer a medição das diferenças entre experimentos, isto trará o resultado como erro de estimação.

6.1.1 Modelagem de cada *cluster* considerando variáveis de processo

A primeira etapa se refere à análise de cada *cluster* que foi gerado como mostrado no Capítulo 5. O projeto de experimento e os valores calculados para o risco e o retorno de cada função são apresentados nas Tabelas 6.1, 6.2, 6.3, 6.4, que estão também disponíveis em Papandrea e Paiva (2016a). Não há uma função nem um experimento referente ao *cluster2* (-1) por este ser formado por apenas uma empresa (veja Tabela 5.8).

A fase *clusters* do macrofluxo foi considerada para otimização individual pela função multiobjetivo de maximização do lucro e, risco de 0,05 utilizando o método *desira-*

Tabela 6.1 – Experimento *Cluster 1* (-1)

Padrão	Execução	Bloco	Componente 1	Componente 2	Componente 3	Componente 4	Resposta 1	Resposta 2
			A:AAPL	B:IHS	C:SIAL	D:TAP	RP	SP
1	1	Bloco 1	1,0000	0,0000	0,0000	0,0000	0,0254	0,0973
2	2	Bloco 1	0,0000	1,0000	0,0000	0,0000	0,0165	0,0781
3	3	Bloco 1	0,0000	0,0000	1,0000	0,0000	0,0123	0,0591
4	4	Bloco 1	0,0000	0,0000	0,0000	1,0000	0,0067	0,0690
5	5	Bloco 1	0,2500	0,2500	0,2500	0,2500	0,0152	0,0518
6	6	Bloco 1	0,6250	0,1250	0,1250	0,1250	0,0203	0,0700
7	7	Bloco 1	0,1250	0,6250	0,1250	0,1250	0,0159	0,0588
8	8	Bloco 1	0,1250	0,1250	0,6250	0,1250	0,0138	0,0502
9	9	Bloco 1	0,1250	0,1250	0,1250	0,6250	0,0110	0,0558
10	10	Bloco 1	1,0000	0,0000	0,0000	0,0000	0,0248	0,0965
11	11	Bloco 1	0,0000	1,0000	0,0000	0,0000	0,0161	0,0784
12	12	Bloco 1	0,0000	0,0000	1,0000	0,0000	0,0127	0,0607
13	13	Bloco 1	0,0000	0,0000	0,0000	1,0000	0,0066	0,0703
14	14	Bloco 1	0,2500	0,2500	0,2500	0,2500	0,0151	0,0523
15	15	Bloco 1	0,6250	0,1250	0,1250	0,1250	0,0199	0,0698
16	16	Bloco 1	0,1250	0,6250	0,1250	0,1250	0,0156	0,0593
17	17	Bloco 1	0,1250	0,1250	0,6250	0,1250	0,0139	0,0515
18	18	Bloco 1	0,1250	0,1250	0,1250	0,6250	0,0108	0,0566

Tabela 6.2 – Experimento *Cluster 1* (+1)

Padrão	Execução	Bloco	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6	Resposta 1	Resposta 2
			A:ADBE	B:ADI	C:BHI	D:MOS	E:QCOM	F:SPRINT	RP	SP
1	1	Bloco 1	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0083	0,0940
2	2	Bloco 1	0,0000	1,0000	0,0000	0,0000	0,0000	0,0000	0,0041	0,0706
3	3	Bloco 1	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	0,0007	0,1137
4	4	Bloco 1	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0083	0,1329
5	5	Bloco 1	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0039	0,0751
6	6	Bloco 1	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	-0,0116	0,1436
7	7	Bloco 1	0,1667	0,1667	0,1667	0,1667	0,1667	0,1667	0,0023	0,0671
8	8	Bloco 1	0,5833	0,0833	0,0833	0,0833	0,0833	0,0833	0,0053	0,0743
9	9	Bloco 1	0,0833	0,5833	0,0833	0,0833	0,0833	0,0833	0,0032	0,0632
10	10	Bloco 1	0,0833	0,0833	0,5833	0,0833	0,0833	0,0833	0,0015	0,0834
11	11	Bloco 1	0,0833	0,0833	0,0833	0,5833	0,0833	0,0833	0,0053	0,0933
12	12	Bloco 1	0,0833	0,0833	0,0833	0,0833	0,5833	0,0833	0,0031	0,0644
13	13	Bloco 1	0,0833	0,0833	0,0833	0,0833	0,0833	0,5833	-0,0047	0,0927
14	14	Bloco 1	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0092	0,0948
15	15	Bloco 1	0,0000	1,0000	0,0000	0,0000	0,0000	0,0000	0,0048	0,0711
16	16	Bloco 1	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	0,0004	0,1157
17	17	Bloco 1	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0090	0,1340
18	18	Bloco 1	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0039	0,0736
19	19	Bloco 1	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	-0,0176	0,1440
20	20	Bloco 1	0,1667	0,1667	0,1667	0,1667	0,1667	0,1667	0,0016	0,0679
21	21	Bloco 1	0,5833	0,0833	0,0833	0,0833	0,0833	0,0833	0,0054	0,0753
22	22	Bloco 1	0,0833	0,5833	0,0833	0,0833	0,0833	0,0833	0,0032	0,0639
23	23	Bloco 1	0,0833	0,0833	0,5833	0,0833	0,0833	0,0833	0,0010	0,0847
24	24	Bloco 1	0,0833	0,0833	0,0833	0,5833	0,0833	0,0833	0,0053	0,0942
25	25	Bloco 1	0,0833	0,0833	0,0833	0,0833	0,5833	0,0833	0,0027	0,0639
26	26	Bloco 1	0,0833	0,0833	0,0833	0,0833	0,0833	0,5833	-0,0080	0,0936

bility. Neste caso, será usado arranjo de mistura considerando os *clusters* e os valores de MGI como sendo as variáveis de processo. O Quadro 6.1 mostra as funções para otimização de cada *cluster* considerando variáveis de processo (-1) e (+1) para risco *SP* e retorno *RP*. Neste, o modelo considerado vai até o nível quadrático por ter uma melhor explicação de acordo com as análises na ANOVA (Apêndice A).

É possível notar que as respostas são influenciadas pelas mudanças nos níveis, proporções, de cada linha dos experimentos. Isto demonstra que os fatores de entrada, as proporções, operam mudanças positivas ou negativas nas respostas Retorno e Risco dos

Tabela 6.3 – Experimento *Cluster 2 (+1)*

Padrão	Execução	Bloco	Componente 1	Componente 2	Resposta 1	Resposta 2
			A:BIIB	B:NKE	RP	SP
1	1	Bloco 1	1,0000	0,0000	0,0196	0,0891
2	2	Bloco 1	0,0000	1,0000	0,0138	0,0641
3	3	Bloco 1	0,5000	0,5000	0,0167	0,0623
4	4	Bloco 1	0,7500	0,2500	0,0181	0,0733
5	5	Bloco 1	0,2500	0,7500	0,0152	0,0588
6	6	Bloco 1	1,0000	0,0000	0,0192	0,0917
7	7	Bloco 1	0,0000	1,0000	0,0144	0,0653
8	8	Bloco 1	0,5000	0,5000	0,0168	0,0639
9	9	Bloco 1	0,7500	0,2500	0,0180	0,0754
10	10	Bloco 1	0,2500	0,7500	0,0156	0,0601

Tabela 6.4 – Experimento *Cluster 3 (-1)*

Padrão	Execução	Bloco	Componente 1	Componente 2	Componente 3	Resposta 1	Resposta 2
			A:CLX	B:CMA	C:WYN	RP	SP
1	1	Bloco 1	1,0000	0,0000	0,0000	0,0052	0,0410
2	2	Bloco 1	0,0000	1,0000	0,0000	-0,0015	0,0832
3	3	Bloco 1	0,0000	0,0000	1,0000	0,0084	0,1603
4	4	Bloco 1	0,3333	0,3333	0,3333	0,0040	0,0782
5	5	Bloco 1	0,6667	0,1667	0,1667	0,0046	0,0533
6	6	Bloco 1	0,1667	0,6667	0,1667	0,0013	0,0760
7	7	Bloco 1	0,1667	0,1667	0,6667	0,0062	0,1172
8	8	Bloco 1	1,0000	0,0000	0,0000	0,0050	0,0416
9	9	Bloco 1	0,0000	1,0000	0,0000	-0,0014	0,0860
10	10	Bloco 1	0,0000	0,0000	1,0000	0,0097	0,1661
11	11	Bloco 1	0,3333	0,3333	0,3333	0,0044	0,0812
12	12	Bloco 1	0,6667	0,1667	0,1667	0,0047	0,0552
13	13	Bloco 1	0,1667	0,6667	0,1667	0,0015	0,0787
14	14	Bloco 1	0,1667	0,1667	0,6667	0,0071	0,1215

portfólios.

Pela análise das ANOVAS (Apêndice A) conclui-se que o modelo estatístico é bom, explicado pelas variáveis, e pode ser usado para previsão e otimização. As funções de otimização tem ajuste maior do que 85% sendo suficientes para uma boa modelagem estatística. As réplicas feitas pela autocorrelação foram suficientes para tornar este modelo determinístico em um modelo normal, provado pela ocorrência e medição dos erros entre réplicas.

Tabela 6.5 – Experimento *Cluster 3 (+1)*

Padrão	Execução	Bloco	Componente 1	Componente 2	Componente 3	Componente 4	Resposta 1	Resposta 2
			A:BLL	B:CCE	C:MAS	D:MET	RP	SP
1	1	Bloco 1	1,0000	0,0000	0,0000	0,0000	0,0111	0,0584
2	2	Bloco 1	0,0000	1,0000	0,0000	0,0000	0,0052	0,0851
3	3	Bloco 1	0,0000	0,0000	1,0000	0,0000	-0,0020	0,1238
4	4	Bloco 1	0,0000	0,0000	0,0000	1,0000	0,0011	0,1032
5	5	Bloco 1	0,2500	0,2500	0,2500	0,2500	0,0038	0,0759
6	6	Bloco 1	0,6250	0,1250	0,1250	0,1250	0,0075	0,0596
7	7	Bloco 1	0,1250	0,6250	0,1250	0,1250	0,0045	0,0761
8	8	Bloco 1	0,1250	0,1250	0,6250	0,1250	0,0009	0,0973
9	9	Bloco 1	0,1250	0,1250	0,1250	0,6250	0,0025	0,0873
10	10	Bloco 1	1,0000	0,0000	0,0000	0,0000	0,0113	0,0591
11	11	Bloco 1	0,0000	1,0000	0,0000	0,0000	0,0056	0,0874
12	12	Bloco 1	0,0000	0,0000	1,0000	0,0000	-0,0012	0,1280
13	13	Bloco 1	0,0000	0,0000	0,0000	1,0000	0,0011	0,1062
14	14	Bloco 1	0,2500	0,2500	0,2500	0,2500	0,0042	0,0786
15	15	Bloco 1	0,6250	0,1250	0,1250	0,1250	0,0078	0,0613
16	16	Bloco 1	0,1250	0,6250	0,1250	0,1250	0,0049	0,0787
17	17	Bloco 1	0,1250	0,1250	0,6250	0,1250	0,0015	0,1007
18	18	Bloco 1	0,1250	0,1250	0,1250	0,6250	0,0027	0,0903

Quadro 6.1 – Funções de otimização para os *clusters* individuais

Cluster 1 (-1)	$SP = 0,096936 * AAPL + 0,078325 * IHS + 0,060004 * SIAL + 0,069745 * TAP$ $- 0,00688 * AAPL * IHS + 0,008469 * AAPL * SIAL - 0,17962 * AAPL$ $* TAP - 0,19589 * IHS * SIAL$
Cluster 1 (-1)	$RP = 0,0251 * AAPL + 0,016307 * IHS + 0,012413 * SIAL + 0,006662 * TAP$
Cluster 1 (+1)	$SP = 0,094518 * ADBE + 0,070957 * ADI + 0,1148 * BHI + 0,133538 * MOS$ $+ 0,074449 * QCOM + 0,143896 * SPRINT + 0,505803 * ADBE * ADI$ $+ 0,471646 * ADBE * BHI - 0,41251 * ADBE * MOS - 0,041351$ $* ADBE * QCOM - 0,55063 * ADBE * SPRINT - 0,89141 * ADI * BHI$
Cluster 1 (+1)	$RP = 0,008754 * ADBE + 0,004469 * ADI + 0,000575 * BHI + 0,008675 * MOS$ $+ 0,00387 * QCOM - 0,0146 * SPRINT$
Cluster 2 (-1)	Somente uma empresa: EMC
Cluster 2 (+1)	$SP = 0,090843 * BIIB + 0,064334 * NKE - 0,05751 * BIIB * NKE$
Cluster 2 (+1)	$RP = 0,019369 * BIIB + 0,014096 * NKE$
Cluster 3 (-1)	$SP = 0,041275 * CLX + 0,084537 * CMA + 0,163165 * WYN - 0,07799 * CLX * CMA$ $- 0,0454 * CLX * WYN - 0,02786 * CMA * WYN$
Cluster 3 (-1)	$RP = 0,00512 * CLX - 0,00143 * CMA + 0,009041 * WYN$
Cluster 3 (+1)	$SP = 0,058726 * BLL + 0,086257 * CCE + 0,125917 * MAS + 0,104706 * MET$ $- 0,05687 * BLL * CCE - 0,02847 * BLL * MAS - 0,10148 * BLL * MET$ $- 0,07954 * CCE * MAS$

6.1.2 Metodologia das demonstrações gráficas nas otimizações

Para elucidar as demonstrações gráficas das otimizações e não tornar a leitura repetitiva, esta subseção explica como são criados e o que os gráficos de otimização significam. Essas informações foram retiradas do próprio *software* utilizado, o Design Expert baseado em Gryze *et al.* (2007).

Os gráficos possuem várias respostas, regiões em que os requisitos atendem simultaneamente às propriedades críticas especificadas e às quais são alvo das otimizações. Ao sobrepor os contornos da resposta crítica em um gráfico de contorno, o melhor ponto pode ser escolhido (o ponto ideal, ótimo).

A otimização gráfica exibe a área de valores de resposta possíveis no espaço fatorial. As regiões que não se enquadram nos critérios de otimização são sombreadas em cinza. Qualquer espaço que não esteja sombreado em cinza satisfaz as metas, ou especificações para cada resposta.

Se o objetivo é minimizar uma resposta, especifica-se o limite superior aceitável. Se estiver maximizando, especifica-se o limite inferior aceitável, nestes casos estarão sendo demarcados os limites das regiões onde estará o ponto ótimo. Se estiver tentando atingir um alvo, especifica-se simultaneamente um limite inferior e superior. As estimativas de intervalo podem ser adicionadas ao gráfico de otimização para considerar a incerteza nas previsões nos pontos. Estas estimativas de intervalo podem ser adicionadas aos gráficos de otimização gráfica para ajudar a entender o impacto da incerteza em atingir as metas do processo. As estimativas de intervalo estão ativadas e mostram o intervalo nos critérios de otimização.

No gráfico, o amarelo brilhante (padrão) mostra onde todo o intervalo de todos os intervalos atende aos critérios especificados. O amarelo ouro escuro corresponde ao local onde a estimativa pontual atende aos requisitos de critérios, mas não a parte correspondente de uma estimativa de intervalo.

O *point prediction* usa os modelos ajustados durante a análise, e as configurações de fatores especificadas na ferramenta de fatores, para calcular as previsões de pontos e as estimativas de intervalo. Os valores previstos são atualizados conforme os níveis são alterados. Intervalos de previsão (PI) são encontrados nos pontos de confirmação. Veja os detalhes matemáticos para o intervalo de confiança $(1 - \alpha) * 100\%$ na Equação 6.1; para o intervalo de previsão na Equação 6.2 e para o intervalo de tolerância na Equação 6.3.

$$\hat{y} \pm t_{(1-\frac{\alpha}{2}, n-p)} \cdot SE_{previsão}$$

na qual,

$$SE_{previsão} = S \cdot \sqrt{x_0(X^T X)^{-1}x_0^T}$$
(6.1)

$$\begin{aligned} & \hat{y} \pm t_{(1-\frac{\alpha}{2}, n-p)} \cdot SE_{previsão} \\ & \text{na qual,} \\ & SE_{previsão} = S \cdot \sqrt{\frac{1}{n} x_0 (X^T X)^{-1} x_0^T} \end{aligned} \quad (6.2)$$

$$\begin{aligned} & \hat{y} \pm TI.Mult \\ & \text{na qual,} \\ & TI.Mult = t_{(1-\alpha, n-p)} \cdot \sqrt{x_0 (X^T X)^{-1} x_0^T} + \phi^{-1}\left(0, 5 + \frac{P}{2}\right) \cdot \sqrt{\frac{n-p}{x_{\alpha, n-p}^2}} \end{aligned} \quad (6.3)$$

Nas equações, a *TI* usa apenas α ao invés de $\frac{\alpha}{2}$ para calcular o intervalo bicaudal. Nas quais:

- \hat{y} = valor previsto em x_0 ,
- S = desvio padrão estimado,
- t = valor crítico t de Student,
- α = taxa de erro aceitável do tipo I (1 - nível de confiança),
- n = número de execuções no experimento,
- N = número de observações na amostra futura,
- p = O número de termos no modelo, incluindo o intercepto,
- P = proporção da população contida no intervalo de tolerância,
- X = matriz do modelo expandido,
- x_0 = vetor de ponto expandido,
- ϕ = função normal inversa para converter a proporção em uma pontuação normal,
- x^2 = valor crítico do qui-quadrado
- $n - p$ demonstra também os graus residuais de liberdade (df) da ANOVA,
- O sobrescrito T indica que a matriz anterior foi transposta,
- O sobrescrito -1 indica que a matriz anterior está invertida.

Os quadros das soluções exibem todas as soluções possíveis, viáveis. Ele mostra o valor percentual a se investir em cada empresa e mostra também o retorno e risco dessa solução. Sempre há mais de uma solução viável, no entanto aquela selecionada será a que obtiver maior valor no *desirability*.

6.1.3 Otimização dos *Clusters* individuais

Os *setups* para as otimizações foram feitos de modo a maximizar os retornos com um valor fixado de 0,05 para o risco. Quando o risco mínimo for maior do que 0,05, então o *setup* é de minimização para o risco. As Figuras 6.1, 6.2, 6.3, 6.4, 6.5 são as rampas de otimização e demonstram os valores mínimos e máximos de cada modelo a ser otimizado pelo método *desirability*. Essas rampas são também uma interessante demonstração gráfica onde estão os valores mínimos encontrados pelas funções e os valores máximos, tanto para retorno (RP) quanto para risco (SP). Estes valores estão alocados em cada *cluster*. Após a otimização, em cada rampa haverá um ponto marcado, que será o ponto ótimo que atende a todas as especificações feitas nos *setups*.

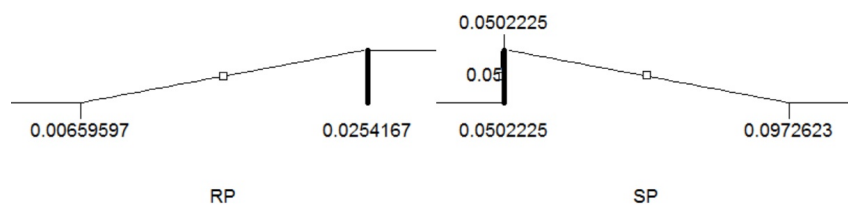


Figura 6.1 – Rampa de otimização para o Cluster 1(-1)

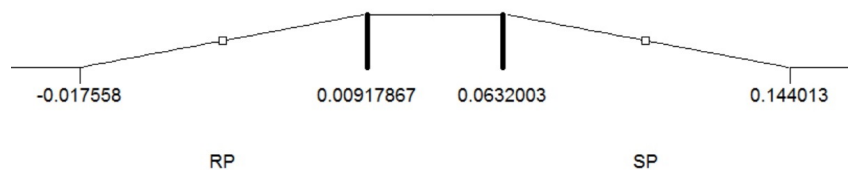


Figura 6.2 – Rampa de otimização para o Cluster 1(+1)

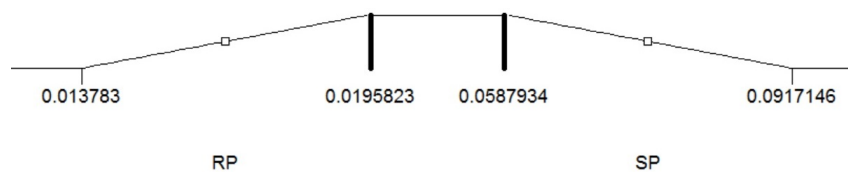


Figura 6.3 – Rampa de otimização para o Cluster 2(+1)

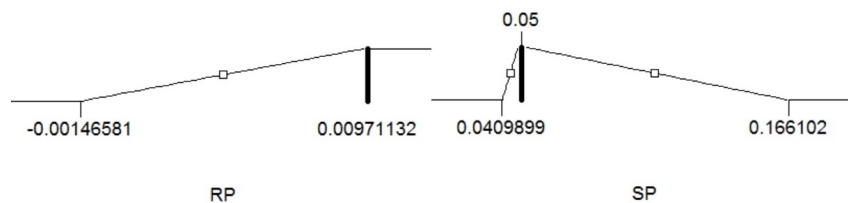


Figura 6.4 – Rampa de otimização para o Cluster 3(-1)

É possível notar que as funções são diferentes nos níveis (-1) e (+1) de cada variável, o que comprova mais uma vez que as variáveis de processo do PCA, os indicadores

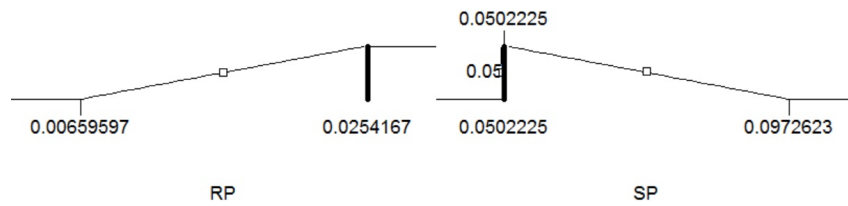


Figura 6.5 – Rampa de otimização para o Cluster 3(+1)

financeiros, têm influência sobre os resultados. Isto ocorre em todas as análises nas quais as variáveis de processo estão presentes. Dentre as soluções apresentadas em cada fase, o investidor pode ainda escolher aquela que mais lhe agrada. Sabendo que todas são soluções viáveis. Para este estudo, escolheu-se a solução com maior valor da função *desirability*. Também pode-se escolher o grau de importância que se deseja dar a cada ativo dos portfólios. Os Quadros 6.2, 6.3, 6.4, 6.5, 6.6 têm os resultados das otimizações por *cluster*, por variável de processo. Note que há diferenças em todos os níveis da otimização. Os gráficos dos modelos estão representados pelas Figuras 6.6, 6.7, 6.8, 6.9 6.10, 6.11, 6.12, 6.13 e 6.14.

Quadro 6.2 – Solução de otimização para o Cluster 1(-1)

		Limite	Limite	Menor	Maior		
Nome	Objetivo	Inferior	Superior	Peso	Peso	Importância	
AAPL	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000	
IHS	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000	
SIAL	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000	
TAP	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000	
RP	maximizar	0,0066	0,0254	1,0000	1,0000	3,0000	
SP	alvo = 0,05	0,0500	0,0973	1,0000	1,0000	3,0000	
Soluções	AAPL	IHS	SIAL	TAP	RP	SP	Desirability
1	0,6910	0,0000	0,0000	0,3090	0,0194	0,0502	0,8257*
2	0,2520	0,4280	0,3200	0,0000	0,0173	0,0502	0,7545
3	0,2420	0,4580	0,3000	0,0000	0,0173	0,0502	0,7541
4	0,2230	0,5030	0,2740	0,0000	0,0172	0,0502	0,7519
5	0,2670	0,3490	0,3840	0,0000	0,0172	0,0502	0,7507
5 Soluções encontradas - (*) selecionado							

O resumo dos resultados das otimizações individuais dos *clusters* por variável de processo podem ser vistos no Quadro 6.7. Este quadro demonstra qual o valor percentual a se investir em cada empresa de cada *cluster*. Se o investidor quiser *forçar* o investimento em todas as empresas, deve-se restringir o modelo levando em consideração a opção de colocar um valor mínimo para cada Ação de cada empresa. Assim o *desirability* irá considerar esta opção. Não foram consideradas restrições nos casos apresentados, mesmo assim é importante frisar que isto é possível nestes modelos conforme descrito no Capítulo 5.

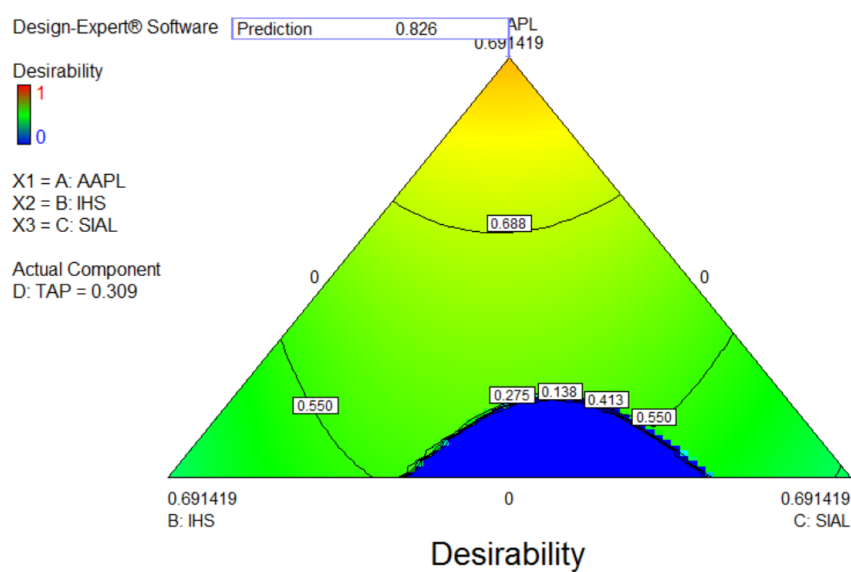


Figura 6.6 – Solução de otimização para o Cluster 1(-1) 2D

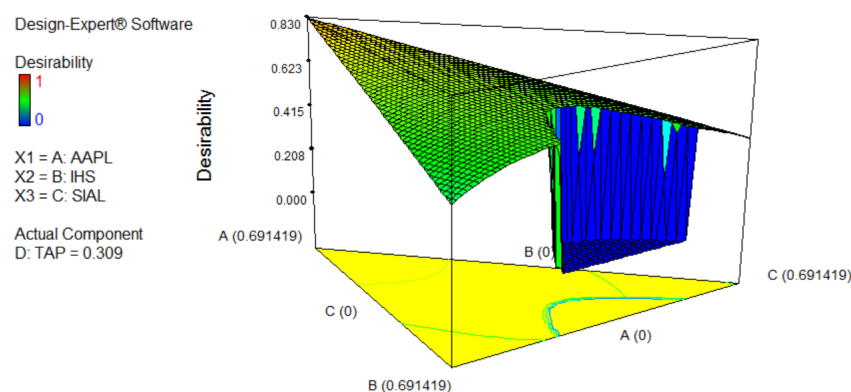


Figura 6.7 – Solução de otimização para o Cluster 1(-1) 3D

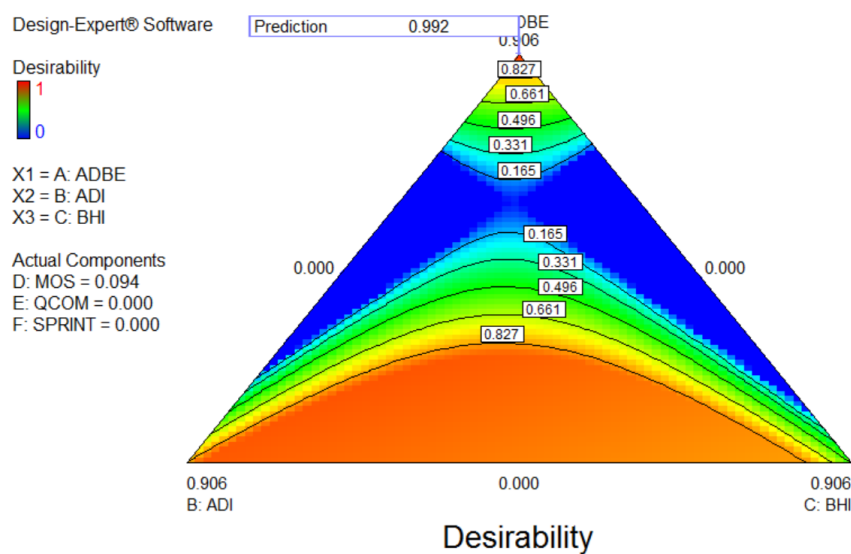


Figura 6.8 – Solução de otimização para o Cluster 1(+1) 2D

Quadro 6.3 – Solução de otimização para o Cluster 1(+1)

Nome	Objetivo	Limite Inferior	Limite Superior	Menor Peso	Maior Peso	Importância				
ADBE	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000				
ADI	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000				
BHI	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000				
MOS	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000				
QCOM	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000				
SPRINT	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000				
RP	maximizar	-0,0176	0,0092	1,0000	1,0000	3,0000				
SP	minimize	0,0632	0,1440	1,0000	1,0000	3,0000				
Soluções	ADBE	ADI	BHI	MOS	QCOM	SPRINT	RP	SP	Desirability	
1	0,9060	0,0000	0,0000	0,0940	0,0000	0,0000	0,0087	0,0632	0,9919*	
2	0,5000	0,0000	0,0000	0,5000	0,0000	0,0000	0,0087	0,0109	0,9913	
3	0,9160	0,0000	0,0000	0,0200	0,0640	0,0000	0,0084	0,0622	0,9861	
4	0,9220	0,0000	0,0000	0,0000	0,0780	0,0000	0,0084	0,0632	0,9848	
5	0,0000	0,3700	0,1350	0,4950	0,0000	0,0000	0,0060	0,0632	0,9391	
6	0,0050	0,3840	0,1280	0,4820	0,0000	0,0000	0,0060	0,0632	0,9391	
7	0,0000	0,4060	0,1180	0,4760	0,0000	0,0000	0,0060	0,0632	0,9389	
8	0,0260	0,3730	0,1360	0,4650	0,0000	0,0000	0,0060	0,0632	0,9389	
9	0,0730	0,3710	0,1450	0,4110	0,0000	0,0000	0,0059	0,0632	0,9376	
10	0,0180	0,4640	0,0990	0,4200	0,0000	0,0000	0,0059	0,0632	0,9372	
11	0,0250	0,2450	0,2310	0,4990	0,0000	0,0000	0,0058	0,0632	0,9342	
12	0,1200	0,2790	0,1670	0,2920	0,1410	0,0000	0,0055	0,0632	0,9282	
13	0,1790	0,2990	0,1600	0,1440	0,2180	0,0000	0,0051	0,0632	0,9204	
14	0,1530	0,6500	0,1240	0,0000	0,0730	0,0000	0,0046	0,0632	0,9104	
14 Soluções encontradas - (*) selecionado										

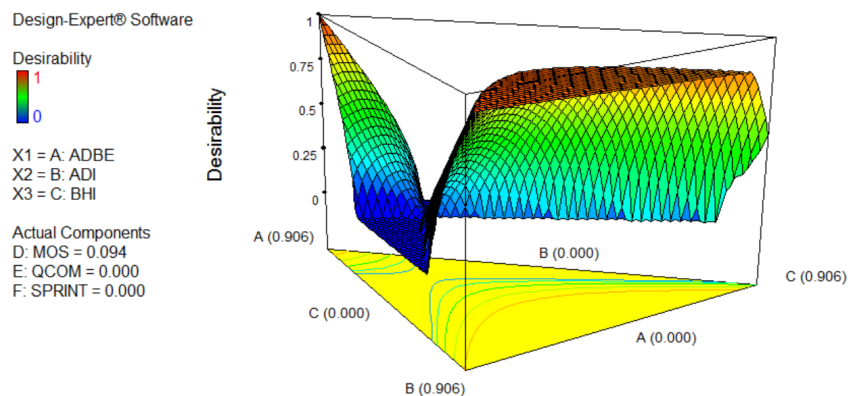


Figura 6.9 – Solução de otimização para o Cluster 1(+1) 3D

Quadro 6.4 – Solução de otimização para o Cluster 2(+1)

Nome	Objetivo	Limite Inferior	Limite Superior	Menor Peso	Maior Peso	Importância
BIIB	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000
NKE	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000
RP	maximizar	0,0138	0,0196	1,0000	1,0000	3,0000
SP	minimize	0,0588	0,0917	1,0000	1,0000	3,0000
Soluções	BIIB	NKE	RP	SP	Desirability	
1	0,602	0,3980	0,0173	0,0665	0,6784*	
1 Solução encontrada - (*) selecionado						

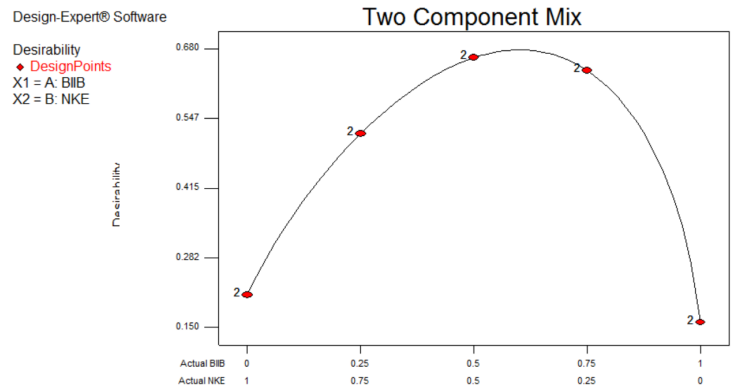


Figura 6.10 – Solução de otimização para o Cluster 2(+1)

Quadro 6.5 – Solução de otimização para o Cluster 3(-1)

Nome	Objetivo	Limite Inferior	Limite Superior	Menor Peso	Maior Peso	Importância
CLX	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000
CMA	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000
WYN	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000
RP	maximizar	-0,0015	0,0097	1,0000	1,0000	3,0000
SP	alvo = 0.05	0,0410	0,1661	1,0000	1,0000	3,0000
Soluções	CLX	CMA	WYN	RP	SP	Desirability
1	0,8930	0,0000	0,1070	0,0055	0,0500	0,7917*
1 Solução encontrada - (*) selecionado						

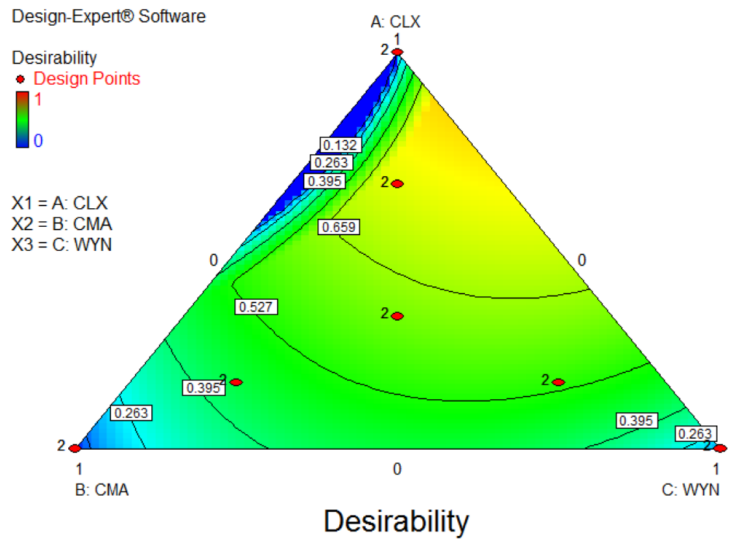


Figura 6.11 – Solução de otimização para o Cluster 3(-1) 2D

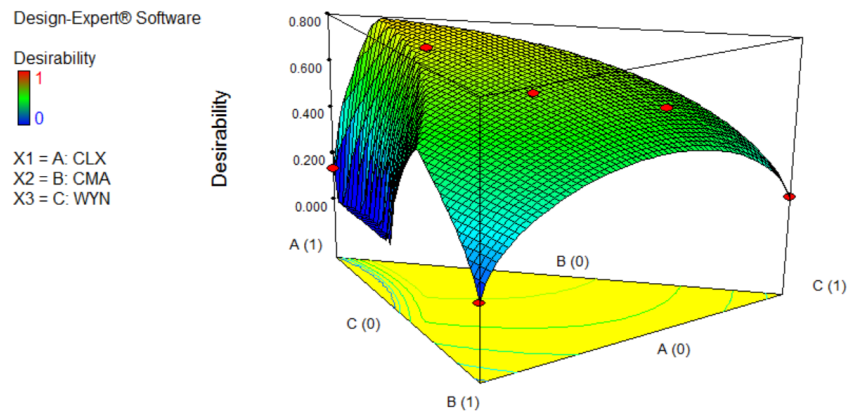


Figura 6.12 – Solução de otimização para o Cluster 3(-1) 3D

Quadro 6.6 – Solução de otimização para o Cluster 3(+1)

		Limite	Limite	Menor	Maior		
Nome	Objetivo	Inferior	Superior	Peso	Peso	Importância	
BLL	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000	
CCE	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000	
MAS	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000	
MET	no intervalo	0,0000	1,0000	1,0000	1,0000	3,0000	
RP	maximizar	-0,0020	0,0113	1,0000	1,0000	3,0000	
SP	minimize	0,0584	0,1280	1,0000	1,0000	3,0000	
Soluções	BLL	CCE	MAS	MET	RP	SP	Desirability
1	1,0000	0,0000	0,0000	0,0000	0,0112	0,0587	0,9937*
1 Solução encontrada - (*) selecionado							

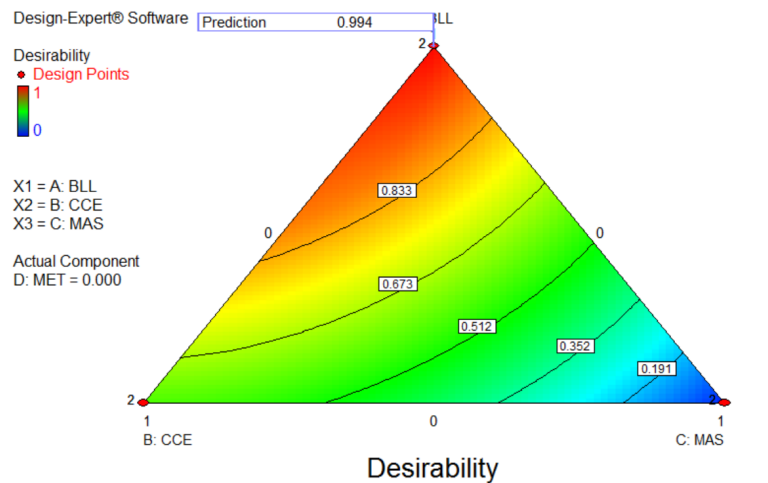


Figura 6.13 – Solução de otimização para o Cluster 3(+1) 2D

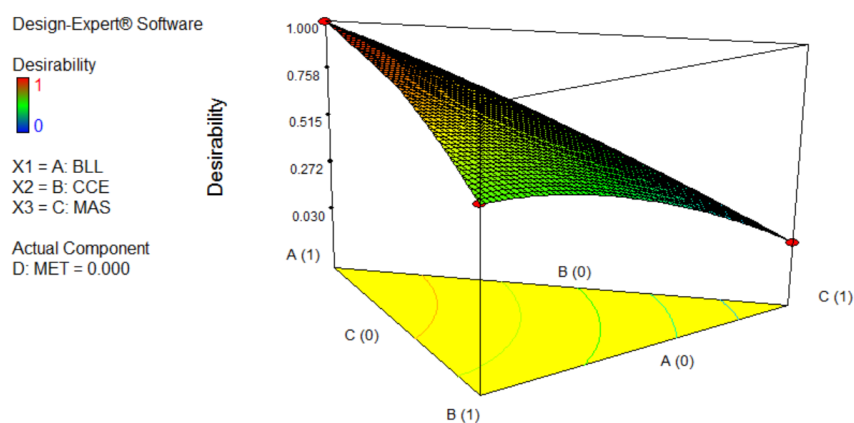


Figura 6.14 – Solução de otimização para o Cluster 3(+1) 3D

Quadro 6.7 – Resultados por cluster

Tipo	Empresas/PCA/Cluster/MGI						Resultados		
Cluster1 (-1)	AAPL	IHS	SIAL	TAP			RP	SP	Desirability
w	0,6910	0,0000	0,0000	0,3090			0,0194	0,0502	0,8257
Cluster1 (+1)	ADBE	ADI	BHI	MOS	QCOM	SPRINT	RP	SP	Desirability
w	0,9060	0,0000	0,0000	0,094	0,0000	0,0000	0,0087	0,0632	0,9919
Cluster2 (-1)	EMC						RP	SP	Desirability
w	1,0000						0,0055	0,0811	-
Cluster2 (+1)	BIIB	NKE					RP	SP	Desirability
w	0,6010	0,3990					0,0173	0,0665	0,6784
Cluster3 (-1)	CLX	CMA	WYN				RP	SP	Desirability
w	0,8930	0,0000	0,1070				0,0055	0,0500	0,7917
Cluster3 (+1)	BLL	CCE	MAS	MET			RP	SP	Desirability
w	0,9930	0,0000	0,0000	0,0070			0,0112	0,0584	0,9938

6.1.4 Modelagem combinada de todos os *clusters* considerando método lexicográfico e variáveis de processo (CMO)

Com os valores ótimos calculados na Subseção 6.1.3 faz-se agora a junção dos *cluster* em um novo arranjo de misturas, considerando como misturas internas os seus percentuais previamente especificados no Quadro 6.7, otimizados. Este novo arranjo é otimizado de forma combinada, também considerando os valores de MGI como variáveis de processo. O método *desirability* foi aplicado para maximizar os retornos com risco igual a 0,05. A programação lexicográfica se dá pela otimização de arranjos que já são os melhores, suas combinações de pesos já foram escolhidas previamente. O que se tem é um novo experimento que só irá ter como resposta a melhor proporção entre os *clusters*. O arranjo é demonstrado na Figura 6.15. Entenda que este resultado trará como melhor a quantidade a se investir no *cluster* 1, 2 ou 3 e dentro deles já estará definido a proporção a se investir em cada empresa. Nesta fase, diferenciador dos portfólios será também a variável de processo, tendo os arranjos novos modificados principalmente por elas. Esta forma demonstrou que realmente há diferença nas funções dos portfólios com a introdução desta variável advinda dos indicadores financeiros. O arranjo completo está discriminado na Tabela 6.6. As funções para otimização são listadas no Quadro 6.8 e está também disponível em Papandrea e Paiva (2016a).

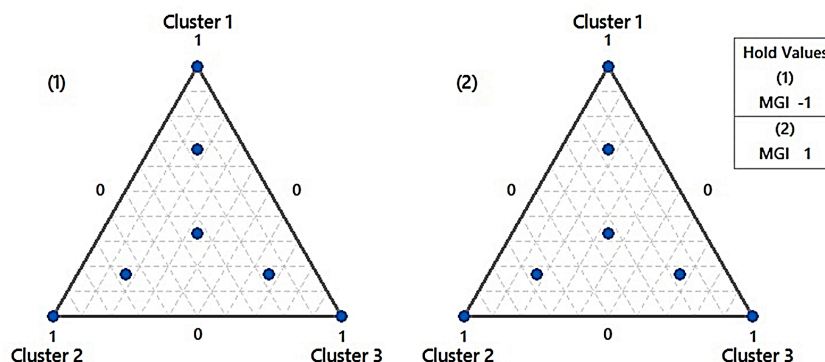


Figura 6.15 – Arranjo de misturas com MGI como variável de processo

É possível notar na Tabela 6.6 que as respostas são influenciadas pelas mudanças nos níveis, proporções, de cada linha dos experimentos. Isto demonstra que os fatores de entrada, as proporções, operam mudanças positivas ou negativas nas respostas Retorno e Risco dos portfólios combinados (CMO) assim como a variável externa, MGI traz alterações nessas respostas.

Pela análise das ANOVAS (Apêndice B) conclui-se que o modelo estatístico é bom, explicado pelas variáveis, e pode ser usado para previsão e otimização. As funções de otimização tem ajuste maior do que 85% sendo suficientes para uma boa modelagem estatística. As variáveis de processo, externas, chamadas aqui de MGI têm também in-

Tabela 6.6 – Arranjo para otimização dos *clusters* combinados com variáveis de processo

Padrão	Execução	Bloco	Componente 1 A:Cluster1	Componente 2 B:Cluster2	Componente 3 C:Cluster3	Fator 4 D:MGI	Resposta 1 RP	Resposta 2 SP
5	1	Bloco 1	0,5000	0,0000	0,5000	+1	0,0097	0,0538
11	2	Bloco 1	0,0000	1,0000	0,0000	+1	0,0173	0,0660
13	3	Bloco 1	0,6667	0,1667	0,1667	-1	0,0149	0,0527
18	4	Bloco 1	0,3333	0,3333	0,3333	-1	0,0101	0,0394
3	5	Bloco 1	0,5000	0,5000	0,0000	+1	0,0128	0,0560
23	6	Bloco 1	0,3333	0,3333	0,3333	+1	0,0122	0,0421
25	7	Bloco 1	0,0000	0,0000	1,0000	-1	0,0052	0,0410
24	8	Bloco 1	0,0000	0,0000	1,0000	+1	0,0111	0,0584
7	9	Bloco 1	0,0000	0,5000	0,5000	+1	0,0142	0,0440
19	10	Bloco 1	0,3333	0,3333	0,3333	-1	0,0101	0,0394
1	11	Bloco 1	0,0000	0,0000	1,0000	+1	0,0111	0,0584
14	12	Bloco 1	0,6667	0,1667	0,1667	+1	0,0109	0,0628
21	13	Bloco 1	0,3333	0,3333	0,3333	+1	0,0126	0,0427
10	14	Bloco 1	0,0000	1,0000	0,0000	-1	0,0061	0,0812
22	15	Bloco 1	0,3333	0,3333	0,3333	+1	0,0126	0,0427
12	16	Bloco 1	0,0000	0,0000	1,0000	-1	0,0050	0,0416
9	17	Bloco 1	1,0000	0,0000	0,0000	+1	0,0092	0,0914
2	18	Bloco 1	0,5000	0,5000	0,0000	-1	0,0126	0,0553
17	19	Bloco 1	0,3333	0,3333	0,3333	-1	0,0101	0,0394
8	20	Bloco 1	1,0000	0,0000	0,0000	-1	0,0192	0,0751
15	21	Bloco 1	0,1667	0,6667	0,1667	-1	0,0081	0,0560
20	22	Bloco 1	0,3333	0,3333	0,3333	-1	0,0101	0,0394
4	23	Bloco 1	0,5000	0,0000	0,5000	-1	0,0121	0,0429
16	24	Bloco 1	0,1667	0,6667	0,1667	+1	0,0149	0,0487
6	25	Bloco 1	0,0000	0,5000	0,5000	-1	0,0056	0,0456

Quadro 6.8 – Funções de otimização para os *clusters* combinados com variáveis de processo

Combinado $\gamma_i^* = \text{Cluster}_i$ ótimo	$SP = 0,084046 * \gamma_1^* + 0,074109 * \gamma_2^* + 0,049858 * \gamma_3^* - 0,0948 * \gamma_1^* * \gamma_2^* - 0,07719 * \gamma_1^* * \gamma_3^* + 0,008263 * \gamma_1^* * MGI - 0,07262 * \gamma_2^* * \gamma_3^* - 0,00789 * \gamma_2^* * MGI + 0,008501 * \gamma_3^* * MGI + 0,001494 * \gamma_1^* * \gamma_2^* * MGI - 0,01001 * \gamma_1^* * \gamma_3^* * MGI - 0,00418 * \gamma_2^* * \gamma_3^* * MGI$
Combinado $\gamma_i^* = \text{Cluster}_i$ ótimo	$RP = 0,014122 * \gamma_1^* + 0,011652 * \gamma_2^* + 0,008099 * \gamma_3^* - 0,00513 * \gamma_1^* * MGI + 0,005585 * \gamma_2^* * MGI + 0,003004 * \gamma_3^* * MGI$
Combinado $\gamma_i^* = \text{Cluster}_i$ ótimo MGI(-1)	$SP = 0,07578 * \gamma_1^* + 0,081994 * \gamma_2^* + 0,041357 * \gamma_3^* - 0,09629 * \gamma_1^* * \gamma_2^* - 0,06718 * \gamma_1^* * \gamma_3^* - 0,06844 * \gamma_2^* * \gamma_3^*$
Combinado $\gamma_i^* = \text{Cluster}_i$ ótimo MGI(-1)	$RP = 0,019249 * \gamma_1^* + 0,006067 * \gamma_2^* + 0,005095 * \gamma_3^*$
Combinado $\gamma_i^* = \text{Cluster}_i$ ótimo MGI(+1)	$SP = 0,092308 * \gamma_1^* + 0,066223 * \gamma_2^* + 0,05836 * \gamma_3^* - 0,0933 * \gamma_1^* * \gamma_2^* - 0,0872 * \gamma_1^* * \gamma_3^* - 0,0768 * \gamma_2^* * \gamma_3^*$
Combinado $\gamma_i^* = \text{Cluster}_i$ ótimo MGI(+1)	$RP = 0,008994 * \gamma_1^* + 0,017236 * \gamma_2^* + 0,011103 * \gamma_3^*$

fluência nos resultados e provavelmente a otimização se demonstrará ser a mais robusta, já que cada *cluster* foi otimizado antes de fazer parte deste arranjo de misturas CMO.

6.1.5 Otimização dos arranjos combinados

Os *setups* para as otimizações foram para maximizar os retornos e um valor fixado de 0,05 para o risco. Quando o risco mínimo for maior do que 0,05, então o setup é de minimização para o risco. A Figura 6.16 mostra a rampa de otimização e demonstra os valores mínimos e máximos de cada modelo a ser otimizado pelo método *desirability*. Igualmente aos *clusters* individuais, aqui pode-se restringir o quanto investir em cada *cluster*. Essas rampas são também uma interessante demonstração gráfica onde estão os valores mínimos encontrados pelas funções e os valores máximos, tanto para retorno (RP) quanto para risco (SP). Estes valores estão alocados em cada *cluster* já otimizado. Após a otimização combinada, em cada rampa haverá um ponto marcado, que será o ponto ótimo que atende a todas as especificações feitas nos *setups*.

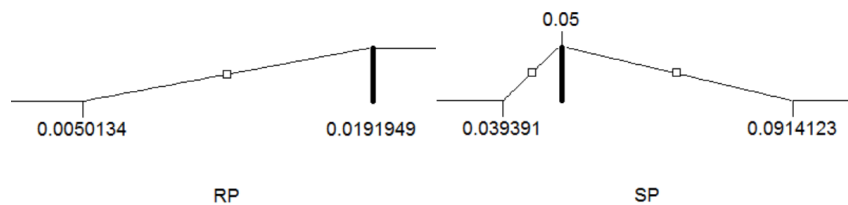


Figura 6.16 – Rampa de otimização do arranjo combinado

O Quadro 6.9 mostra que há dez soluções viáveis para o problema apresentado, das quais a que teve o maior valor para o *desirability* D foi escolhida. Note que o investidor tem dez opções diferentes. Essas diferenças associam-se inclusive à opção de investir ou não em certo *cluster* e mostra a diversificação quanto a variável de processo $(+1)(-1)$. Um outra característica desta proposta é exatamente dar opções ótimas de escolha para o investidor, não apenas apontar uma solução, mas várias.

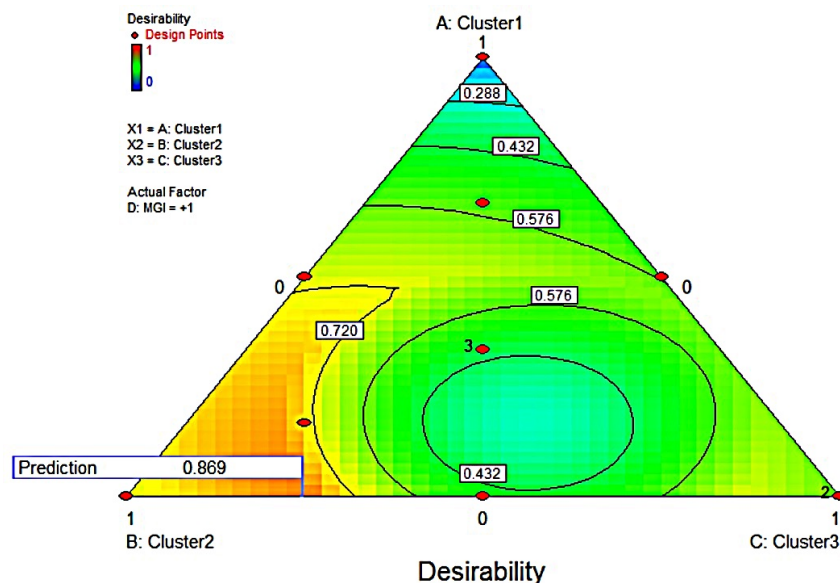


Figura 6.17 – Solução de otimização para os arranjos combinados 2D

Quadro 6.9 – Resultados Combinado

Nome	Objetivo	Limite	Limite	Menor	Maior	Importância	
		Inferior	Superior	Peso	Peso		
Cluster1	no intervalo	0	1	1	1	3	
Cluster2	no intervalo	0	1	1	1	3	
Cluster3	no intervalo	0	1	1	1	3	
MGI	no intervalo	-1	+1	1	1	3	
RP	maximizar	0,0050	0,0192	1	1	3	
SP	no alvo = 0,05	0,0394	0,0914	1	1	3	
Soluções para 2 combinações de categorias e níveis fatoriais							
Número	Cluster1	Cluster2	Cluster3	MGI	RP	SP	Desirability
1	0,0000	0,7530	0,2470	+1	0,0157	0,0500	0,8689*
2	0,6790	0,0380	0,2830	-1	0,0147	0,0502	0,8268
3	0,6860	0,0000	0,3140	-1	0,0148	0,0505	0,8259
4	0,2110	0,7890	0,0000	+1	0,0155	0,0562	0,7929
5	0,7390	0,2610	0,0000	-1	0,0158	0,0588	0,7739
6	0,0000	0,1450	0,8550	+1	0,0120	0,0500	0,7014
7	0,5000	0,5000	0,0000	-1	0,0127	0,0548	0,6902
8	0,2180	0,0000	0,7820	+1	0,0106	0,0509	0,6232
9	0,0000	0,6230	0,3770	-1	0,0057	0,0506	0,2186
10	0,0000	0,0000	1,0000	-1	0,0051	0,0414	0,0328
10 Soluções encontradas - (*) selecionado							

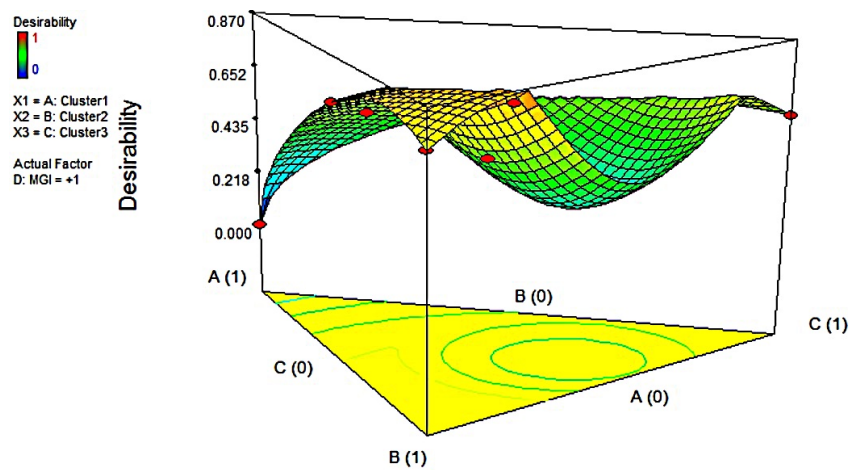


Figura 6.18 – Solução de otimização para os arranjos combinados 3D

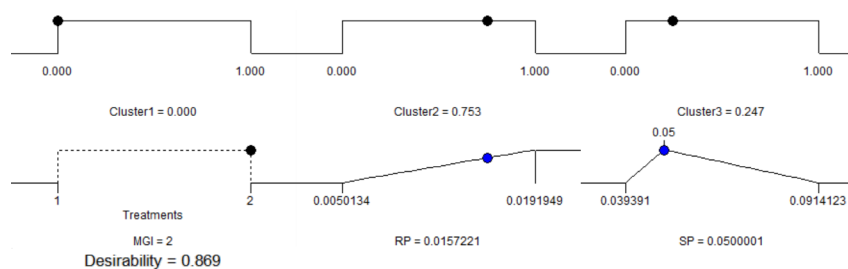


Figura 6.19 – Resultado rampa de otimização do arranjo combinado

6.1.6 Modelagem por *cluster* e variáveis de processo juntos utilizando CM-DOE

Neste caso é aplicado o arranjo de “misturas de misturas”(MoM) com variáveis de processo. O primeiro nível da mistura é composto pelos PCAs calculados sobre a classificação da *Newsweek* e o segundo nível é formado pelos *clusters*; os valores das variáveis de processo são os maiores e menores valores de PCA1, PCA2, PCA3 que foram reduzidos ao MGI. O método de otimização aplicado é o *desirability* sendo os alvos, maximizar o retorno e risco igual a 0,05 ou minimizar o risco quando este valor não for possível. O arranjo dessa mistura com variáveis de processo é demonstrado na Figura 6.20. Os gráficos e figuras das análises encontram-se no Apêndice C.

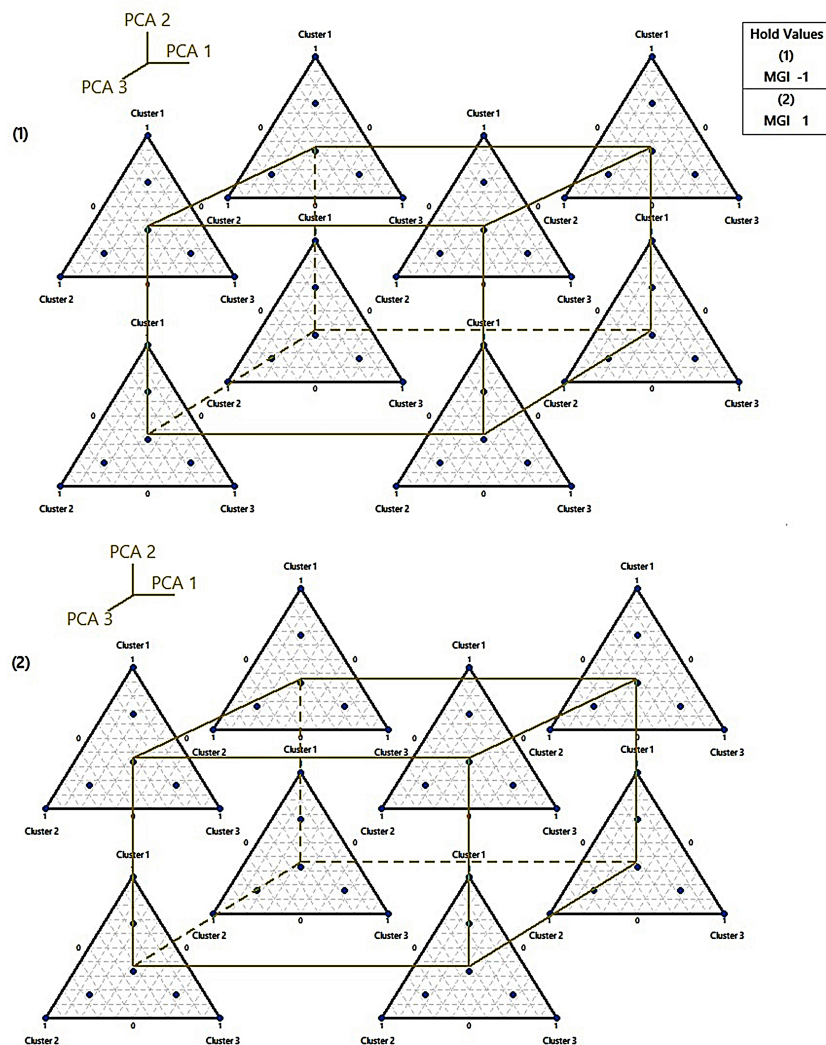


Figura 6.20 – Arranjo *simplex* MoM com variáveis de processo (-1) (+1)

A Tabela 6.7 lista todas as empresas por *cluster* e seus valores na classificação do PCA. Cada portfólio agora é composto pelas empresas com menores PCAs para a variável (-1) e maiores PCAs para a variável (+1) do arranjo. Neste caso, a variável de processo pode alterar quais empresas participarão de qual portfólio. O Quadro 6.10 é o resumo

com apenas as empresas que participarão dessas análises com os respectivos valores para variável de processo (-1) e (+1).

Tabela 6.7 – Empresas selecionadas para o CM-DOE

Cluster / Variável Processo	Empresa	Classificação PCA1	Classificação PCA2	Classificação PCA3
Cluster 1 (-1)	AAPL (1)	-0.31	-1.40	-0.22
	IHS (11)	-0.77	-0.68	0.02
	SIAL (17)	-1.40	0.36	1.08
	TAP (19)	-0.91	-0.62	0.92
Cluster 1 (+1)	ADBE (2)	-0.21	-0.61	1.38
	ADI (3)	0.47	-0.43	0.29
	BHI (4)	0.87	-1.88	-0.28
	MOS (14)	-0.09	-0.10	1.37
	QCOM (16)	0.92	-2.32	-0.13
	SPRINT (18)	1.63	-0.91	-1.08
Cluster 2 (-1)	EMC (10)	1.75	0.87	1.02
Cluster 2 (+1)	BIIB (5)	5.88	0.05	-1.99
	NKE (15)	4.44	2.72	1.56
Cluster 3 (-1)	CLX (8)	-2.29	2.04	-2.23
	CMA (9)	-1.89	0.75	-0.26
	WYN (20)	-1.91	0.67	0.13
Cluster 3 (+1)	BLL (6)	-1.63	0.08	-0.12
	CCE (7)	-1.57	0.12	-0.12
	MAS (12)	-1.56	0.13	-0.38
	MET (13)	-1.42	1.16	-0.97

Quadro 6.10 – Resumo das empresas selecionadas para o CM-DOE

PCA	MGI	Cluster1	Cluster2	Cluster3
PCA1	-1	SIAL	EMC	CLX
	1	SPRINT	BIIB	MET
PCA2	-1	QCOM	BIIB	BLL
	1	SIAL	NKE	CLX
PCA3	-1	SPRINT	BIIB	CCE
	1	ADBE	NKE	WYN

Os arranjos de misturas de misturas (MoM) cruzado considerando as variáveis de processo (CM-DOE) é mostrado nas Tabelas 6.8 e 6.9 e está também disponível em Papandrea e Paiva (2016a). Relembrando que neste caso tanto o PCA quanto o MGI são os indicadores para seleção das empresas que irão compor cada nível de cada portfólio.

Tabela 6.8 – Arranjo de misturas de misturas MoM cruzado CM-DOE 1 de 2

Padrão	Execução	Bloco	Mix 1-1 A:PCA1	Mix 1-2 B:PCA2	Mix 1-3 C:PCA3	Mix 2-1 D:Cluster1	Mix 2-2 E:Cluster2	Mix 2-3 F:Cluster3	Fator 7 G:MGI	Resposta 1 RP	Resposta 2 SP
54	1	Bloco 1	0,5000	0,0000	0,5000	0,0000	0,0000	1,0000	+1	0,0345	0,1792
10	2	Bloco 1	1,0000	0,0000	0,0000	0,5000	0,0000	0,5000	+1	0,0209	0,2188
32	3	Bloco 1	0,0000	0,0000	1,0000	0,5000	0,5000	0,0000	+1	0,0586	0,2100
23	4	Bloco 1	0,0000	1,0000	0,0000	0,0000	0,5000	0,5000	-1	0,0677	0,1396
46	5	Bloco 1	0,5000	0,5000	0,0000	0,5000	0,0000	0,5000	+1	0,0321	0,1295
79	6	Bloco 1	0,0000	0,0000	1,0000	0,0000	0,0000	1,0000	+1	0,0452	0,2358
7	7	Bloco 1	1,0000	0,0000	0,0000	0,5000	0,5000	0,0000	-1	0,0477	0,1220
50	8	Bloco 1	0,5000	0,0000	0,5000	1,0000	0,0000	0,0000	+1	0,0288	0,1748
24	9	Bloco 1	0,0000	1,0000	0,0000	0,0000	0,5000	0,5000	+1	0,0623	0,1105
76	10	Bloco 1	0,6667	0,1667	0,1667	0,1667	0,6667	0,1667	+1	0,0517	0,1480
27	11	Bloco 1	0,0000	0,0000	1,0000	0,0000	1,0000	0,0000	-1	0,0578	0,2146
60	12	Bloco 1	0,5000	0,0000	0,5000	0,0000	0,5000	0,5000	+1	0,0507	0,1545
70	13	Bloco 1	0,0000	0,5000	0,5000	0,5000	0,0000	0,5000	+1	0,0428	0,1280
5	14	Bloco 1	1,0000	0,0000	0,0000	0,0000	0,0000	1,0000	-1	0,0446	0,1211
6	15	Bloco 1	1,0000	0,0000	0,0000	0,0000	0,0000	1,0000	+1	0,0237	0,2391
38	16	Bloco 1	0,5000	0,5000	0,0000	1,0000	0,0000	0,0000	+1	0,0292	0,1454
37	17	Bloco 1	0,5000	0,5000	0,0000	1,0000	0,0000	0,0000	-1	0,0334	0,1158
9	18	Bloco 1	1,0000	0,0000	0,0000	0,5000	0,0000	0,5000	-1	0,0361	0,1194
8	19	Bloco 1	1,0000	0,0000	0,0000	0,5000	0,5000	0,0000	+1	0,0371	0,2133
17	20	Bloco 1	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	-1	0,0561	0,1433
57	21	Bloco 1	0,5000	0,0000	0,5000	0,5000	0,0000	0,5000	-1	0,0311	0,1211
26	22	Bloco 1	0,0000	0,0000	1,0000	1,0000	0,0000	0,0000	+1	0,0395	0,2314
44	23	Bloco 1	0,5000	0,5000	0,0000	0,5000	0,5000	0,0000	+1	0,0483	0,1240
47	24	Bloco 1	0,5000	0,5000	0,0000	0,0000	0,5000	0,5000	-1	0,0619	0,0938
40	25	Bloco 1	0,5000	0,5000	0,0000	0,0000	1,0000	0,0000	+1	0,0673	0,1277
72	26	Bloco 1	0,0000	0,5000	0,5000	0,0000	0,5000	0,5000	+1	0,0618	0,1236
69	27	Bloco 1	0,0000	0,5000	0,5000	0,5000	0,0000	0,5000	-1	0,0369	0,1270
67	28	Bloco 1	0,0000	0,5000	0,5000	0,5000	0,5000	0,0000	-1	0,0485	0,1295
64	29	Bloco 1	0,0000	0,5000	0,5000	0,0000	1,0000	0,0000	+1	0,0781	0,1263
18	30	Bloco 1	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	+1	0,0460	0,1352
41	31	Bloco 1	0,5000	0,5000	0,0000	0,0000	0,0000	1,0000	-1	0,0504	0,0975
56	32	Bloco 1	0,5000	0,0000	0,5000	0,5000	0,5000	0,0000	+1	0,0479	0,1534
19	33	Bloco 1	0,0000	1,0000	0,0000	0,5000	0,5000	0,0000	-1	0,0592	0,1442
53	34	Bloco 1	0,5000	0,0000	0,5000	0,0000	0,0000	1,0000	-1	0,0396	0,1227
29	35	Bloco 1	0,0000	0,0000	1,0000	0,0000	0,0000	1,0000	-1	0,0346	0,2044
52	36	Bloco 1	0,5000	0,0000	0,5000	0,0000	1,0000	0,0000	+1	0,0669	0,1571
16	37	Bloco 1	0,0000	1,0000	0,0000	0,0000	1,0000	0,0000	+1	0,0785	0,1131
68	38	Bloco 1	0,0000	0,5000	0,5000	0,5000	0,5000	0,0000	+1	0,0590	0,1225
28	39	Bloco 1	0,0000	0,0000	1,0000	0,0000	1,0000	0,0000	+1	0,0776	0,2138

As funções para otimização do modelo podem ser divididas pelo MGI (-1) e $(+1)$ nos Quadros 6.11 e 6.12. As funções completas considerando todas as variáveis da mistura e as variáveis de processo são demonstradas nos Quadros 6.13 para os retornos RP e 6.14 para os riscos SP.

As análises de ANOVA e análises gráficas são feitas para os retornos RP e os riscos SP. É possível notar que as respostas são influenciadas pelas mudanças nos níveis, proporções, de cada linha dos experimentos. Isto demonstra que os fatores de entrada, as proporções, operam mudanças positivas ou negativas nas respostas Retorno e Risco dos portfólios.

Pela análise das ANOVAS (Apêndice C) conclui-se que o modelo estatístico é bom, explicado pelas variáveis, e pode ser usado para previsão e otimização. As funções de otimização tem ajuste maior do que 85% sendo suficientes para uma boa modelagem

Tabela 6.9 – Arranjo de misturas de misturas MoM cruzado CM-DOE 2 de 2

Padrão	Execução	Bloco	Mix 1-1 A:PCA1	Mix 1-2 B:PCA2	Mix 1-3 C:PCA3	Mix 2-1 D:Cluster1	Mix 2-2 E:Cluster2	Mix 2-3 F:Cluster3	Fator 7 G:MGI	Resposta 1 RP	Resposta 2 SP
73	40	Bloco 1	0,6667	0,1667	0,1667	0,6667	0,1667	0,1667	-1	0,0374	0,0968
48	41	Bloco 1	0,5000	0,5000	0,0000	0,0000	0,5000	0,5000	+1	0,0511	0,1251
49	42	Bloco 1	0,5000	0,0000	0,5000	1,0000	0,0000	0,0000	-1	0,0227	0,1410
4	43	Bloco 1	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	+1	0,0562	0,2170
20	44	Bloco 1	0,0000	1,0000	0,0000	0,5000	0,5000	0,0000	+1	0,0594	0,1094
63	45	Bloco 1	0,0000	0,5000	0,5000	0,0000	1,0000	0,0000	-1	0,0685	0,1388
74	46	Bloco 1	0,6667	0,1667	0,1667	0,6667	0,1667	0,1667	+1	0,0327	0,1554
51	47	Bloco 1	0,5000	0,0000	0,5000	0,0000	1,0000	0,0000	-1	0,0628	0,1329
61	48	Bloco 1	0,0000	0,5000	0,5000	1,0000	0,0000	0,0000	-1	0,0284	0,1469
65	49	Bloco 1	0,0000	0,5000	0,5000	0,0000	0,0000	1,0000	-1	0,0454	0,1286
43	50	Bloco 1	0,5000	0,5000	0,0000	0,5000	0,5000	0,0000	-1	0,0535	0,0984
25	51	Bloco 1	0,0000	0,0000	1,0000	1,0000	0,0000	0,0000	-1	0,0177	0,2227
77	52	Bloco 1	0,6667	0,1667	0,1667	0,1667	0,1667	0,6667	-1	0,0459	0,0891
75	53	Bloco 1	0,6667	0,1667	0,1667	0,1667	0,6667	0,1667	-1	0,0559	0,0936
58	54	Bloco 1	0,5000	0,0000	0,5000	0,5000	0,0000	0,5000	+1	0,0281	0,1593
34	55	Bloco 1	0,0000	0,0000	1,0000	0,5000	0,0000	0,5000	+1	0,0434	0,2157
39	56	Bloco 1	0,5000	0,5000	0,0000	0,0000	1,0000	0,0000	-1	0,0735	0,1080
78	57	Bloco 1	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	+1	0,0480	0,1354
55	58	Bloco 1	0,5000	0,0000	0,5000	0,5000	0,5000	0,0000	-1	0,0372	0,1241
3	59	Bloco 1	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	-1	0,0682	0,1315
12	60	Bloco 1	1,0000	0,0000	0,0000	0,0000	0,5000	0,5000	+1	0,0346	0,2153
66	61	Bloco 1	0,0000	0,5000	0,5000	0,0000	0,0000	1,0000	+1	0,0486	0,1485
22	62	Bloco 1	0,0000	1,0000	0,0000	0,5000	0,0000	0,5000	+1	0,0422	0,1150
1	63	Bloco 1	1,0000	0,0000	0,0000	1,0000	0,0000	0,0000	-1	0,0228	0,1398
62	64	Bloco 1	0,0000	0,5000	0,5000	1,0000	0,0000	0,0000	+1	0,0370	0,1442
2	65	Bloco 1	1,0000	0,0000	0,0000	1,0000	0,0000	0,0000	+1	0,0070	0,2357
42	66	Bloco 1	0,5000	0,5000	0,0000	0,0000	0,0000	1,0000	+1	0,0333	0,1503
33	67	Bloco 1	0,0000	0,0000	1,0000	0,5000	0,0000	0,5000	-1	0,0177	0,2035
35	68	Bloco 1	0,0000	0,0000	1,0000	0,0000	0,5000	0,5000	-1	0,0404	0,2015
11	69	Bloco 1	1,0000	0,0000	0,0000	0,0000	0,5000	0,5000	-1	0,0570	0,1176
13	70	Bloco 1	0,0000	1,0000	0,0000	1,0000	0,0000	0,0000	-1	0,0333	0,1624
59	71	Bloco 1	0,5000	0,0000	0,5000	0,0000	0,5000	0,5000	-1	0,0487	0,1194
14	72	Bloco 1	0,0000	1,0000	0,0000	1,0000	0,0000	0,0000	+1	0,0364	0,1310
71	73	Bloco 1	0,0000	0,5000	0,5000	0,0000	0,5000	0,5000	-1	0,0540	0,1254
30	74	Bloco 1	0,0000	0,0000	1,0000	0,0000	0,0000	1,0000	+1	0,0492	0,2361
21	75	Bloco 1	0,0000	1,0000	0,0000	0,5000	0,0000	0,5000	-1	0,0448	0,1422
31	76	Bloco 1	0,0000	0,0000	1,0000	0,5000	0,5000	0,0000	-1	0,0289	0,2061
45	77	Bloco 1	0,5000	0,5000	0,0000	0,5000	0,0000	0,5000	-1	0,0396	0,0962
15	78	Bloco 1	0,0000	1,0000	0,0000	0,0000	1,0000	0,0000	-1	0,0788	0,1540
36	79	Bloco 1	0,0000	0,0000	1,0000	0,0000	0,5000	0,5000	+1	0,0652	0,2112

estatística. As réplicas feitas pela autocorrelação foram suficientes para tornar este modelo determinístico em um modelo normal, provado pela ocorrência e medição dos erros entre réplicas.

6.1.7 Otimização do arranjo CM-DOE

Os *setups* para as otimizações foram para maximizar os retornos e um valor fixado de 0,05 para o risco. Quando o risco mínimo for maior do que 0,05, então o *setup* é de minimização para o risco. A Figura 6.21 mostra a rampa de otimização e demonstra os valores mínimos e máximos de cada modelo a ser otimizado pelo método *desirability*. Essas rampas são também uma interessante demonstração gráfica onde estão os valores mínimos encontrados pelas funções e os valores máximos, tanto para retorno (RP) quanto para risco (SP). Estes valores estão alocados em cada ativo das empresas selecionadas,

Quadro 6.11 – Função de otimização CM-DOE MGI(-1)

CM-DOE $\gamma_i = Cluster_i$ MGI (-1)	$ \begin{aligned} SP = & 0.139818 * PCA_1 * \gamma_1 + 0.131522 * PCA_1 * \gamma_2 + 0.121049 * PCA_1 * \gamma_3 + 0.162353 \\ & * PCA_2 * \gamma_1 + 0.154046 * PCA_2 * \gamma_2 + 0.14332 * PCA_2 * \gamma_3 + 0.222739 \\ & * PCA_3 * \gamma_1 + 0.214569 * PCA_3 * \gamma_2 + 0.204411 * PCA_3 * \gamma_3 - 0.14114 \\ & * PCA_1 * PCA_2 * \gamma_1 - 0.13915 * PCA_1 * PCA_2 * \gamma_2 - 0.13887 * PCA_1 \\ & * PCA_2 * \gamma_3 - 0.16094 * PCA_1 * PCA_3 * \gamma_1 - 0.1606 * PCA_1 * PCA_3 * \gamma_2 \\ & - 0.16007 * PCA_1 * PCA_3 * \gamma_3 - 0.05836 * PCA_1 * \gamma_1 * \gamma_2 - 0.04761 \\ & * PCA_1 * \gamma_1 * \gamma_3 - 0.03821 * PCA_1 * \gamma_2 * \gamma_3 - 0.18245 * PCA_2 * PCA_3 \\ & * \gamma_1 - 0.18214 * PCA_2 * PCA_3 * \gamma_2 - 0.18104 * PCA_2 * PCA_3 * \gamma_3 \\ & - 0.05413 * PCA_2 * \gamma_1 * \gamma_2 - 0.04068 * PCA_2 * \gamma_1 * \gamma_3 - 0.03439 \\ & * PCA_2 * \gamma_2 * \gamma_3 - 0.04848 * PCA_3 * \gamma_1 * \gamma_2 - 0.03835 * PCA_3 * \gamma_1 * \gamma_3 \\ & - 0.03033 * PCA_3 * \gamma_2 * \gamma_3 - 0.01941 * PCA_1 * PCA_2 * \gamma_1 * \gamma_2 \\ & - 0.01948 * PCA_1 * PCA_2 * \gamma_1 * \gamma_3 - 0.0259 * PCA_1 * PCA_2 * \gamma_2 * \gamma_3 \\ & - 0.02098 * PCA_1 * PCA_3 * \gamma_1 * \gamma_2 - 0.02098 * PCA_1 * PCA_3 * \gamma_1 * \gamma_3 \\ & - 0.02499 * PCA_1 * PCA_3 * \gamma_2 * \gamma_3 - 0.01524 * PCA_2 * PCA_3 * \gamma_1 * \gamma_2 \\ & - 0.0218 * PCA_2 * PCA_3 * \gamma_1 * \gamma_3 - 0.01039 * PCA_2 * PCA_3 * \gamma_2 * \gamma_3 \end{aligned} $
CM-DOE $\gamma_i = Cluster_i$ MGI (-1)	$ \begin{aligned} RP = & 0.024991 * PCA_1 * \gamma_1 + 0.06792 * PCA_1 * \gamma_2 + 0.044626 * PCA_1 * \gamma_3 + 0.03559 \\ & * PCA_2 * \gamma_1 + 0.078837 * PCA_2 * \gamma_2 + 0.056129 * PCA_2 * \gamma_3 + 0.018769 \\ & * PCA_3 * \gamma_1 + 0.057786 * PCA_3 * \gamma_2 + 0.034634 * PCA_3 * \gamma_3 + 0.002517 \\ & * PCA_1 * \gamma_1 * \gamma_2 + 0.00837 * PCA_1 * \gamma_1 * \gamma_3 + 0.002358 * PCA_1 * \gamma_2 * \gamma_3 \\ & + 0.1384 * PCA_2 * \gamma_1 * \gamma_2 - 0.00005 * PCA_2 * \gamma_1 * \gamma_3 + 0.000307 \\ & * PCA_2 * \gamma_2 * \gamma_3 - 0.03343 * PCA_3 * \gamma_1 * \gamma_2 - 0.02365 * PCA_3 * \gamma_1 * \gamma_3 \\ & - 0.02311 * PCA_3 * \gamma_2 * \gamma_3 \end{aligned} $

Quadro 6.12 – Função de otimização CM-DOE MGI(+1)

CM-DOE $\gamma_i = Cluster_i$ MGI (+1)	$ \begin{aligned} SP = & 0.235752 * PCA_1 * \gamma_1 + 0.217093 * PCA_1 * \gamma_2 + 0.239021 * PCA_1 * \gamma_3 + 0.131006 \\ & * PCA_2 * \gamma_1 + 0.113068 * PCA_2 * \gamma_2 + 0.135289 * PCA_2 * \gamma_3 + 0.231413 \\ & * PCA_3 * \gamma_1 + 0.213733 * PCA_3 * \gamma_2 + 0.235958 * PCA_3 * \gamma_3 - 0.15154 \\ & * PCA_1 * PCA_2 * \gamma_1 - 0.14902 * PCA_1 * PCA_2 * \gamma_2 - 0.14804 * PCA_1 \\ & * PCA_2 * \gamma_3 - 0.23491 * PCA_1 * PCA_3 * \gamma_1 - 0.23291 * PCA_1 * PCA_3 \\ & * \gamma_2 - 0.23405 * PCA_1 * PCA_3 * \gamma_3 - 0.05101 * PCA_1 * \gamma_1 * \gamma_2 \\ & - 0.07334 * PCA_1 * \gamma_1 * \gamma_3 - 0.05032 * PCA_1 * \gamma_2 * \gamma_3 - 0.14808 \\ & * PCA_2 * PCA_3 * \gamma_1 - 0.1485 * PCA_2 * PCA_3 * \gamma_2 - 0.14859 * PCA_2 \\ & * PCA_3 * \gamma_3 - 0.05143 * PCA_2 * \gamma_1 * \gamma_2 - 0.07291 * PCA_2 * \gamma_1 * \gamma_3 \\ & - 0.05534 * PCA_2 * \gamma_2 * \gamma_3 - 0.05091 * PCA_3 * \gamma_1 * \gamma_2 - 0.07237 \\ & * PCA_3 * \gamma_1 * \gamma_3 - 0.05486 * PCA_3 * \gamma_2 * \gamma_3 + 0.013731 * PCA_1 * PCA_2 \\ & * \gamma_1 * \gamma_2 + 0.006716 * PCA_1 * PCA_2 * \gamma_1 * \gamma_3 - 0.00368 * PCA_1 * PCA_2 \\ & * \gamma_2 * \gamma_3 + 0.012695 * PCA_1 * PCA_3 * \gamma_1 * \gamma_2 + 0.01621 * PCA_1 * PCA_3 \\ & * \gamma_1 * \gamma_3 - 0.00042 * PCA_1 * PCA_3 * \gamma_2 * \gamma_3 + 0.004128 * PCA_2 * PCA_3 \\ & * \gamma_1 * \gamma_2 - 0.00041 * PCA_2 * PCA_3 * \gamma_1 * \gamma_3 + 0.002075 * PCA_2 * PCA_3 \\ & * \gamma_2 * \gamma_3 \end{aligned} $
CM-DOE $\gamma_i = Cluster_i$ MGI (+1)	$ \begin{aligned} RP = & 0.0111616 * PCA_1 * \gamma_1 + 0.056164 * PCA_1 * \gamma_2 + 0.0225869 * PCA_1 * \gamma_3 \\ & + 0.0378316 * PCA_2 * \gamma_1 + 0.0784969 * PCA_2 * \gamma_2 + 0.0469619 * PCA_2 \\ & * \gamma_3 + 0.040142 * PCA_3 * \gamma_1 + 0.0776281 * PCA_3 * \gamma_2 + 0.0473679 \\ & * PCA_3 * \gamma_3 + 0.0152689 * PCA_1 * \gamma_1 * \gamma_2 + 0.0130564 * PCA_1 * \gamma_1 * \gamma_3 \\ & - 0.0135031 * PCA_1 * \gamma_2 * \gamma_3 + 0.0052147 * PCA_2 * \gamma_1 * \gamma_2 \\ & + 0.0014793 * PCA_2 * \gamma_1 * \gamma_3 - 0.0011886 * PCA_2 * \gamma_2 * \gamma_3 \\ & - 0.0011437 * PCA_3 * \gamma_1 * \gamma_2 - 0.0064707 * PCA_3 * \gamma_1 * \gamma_3 \\ & + 0.0084601 * PCA_3 * \gamma_2 * \gamma_3 \end{aligned} $

Quadro 6.14 – Função de otimização CM-DOE SP

<p>CM-DOE $\gamma_i = Cluster_i$</p>	$ \begin{aligned} SP = & 0.187785 * PCA_1 * \gamma_1 + 0.174308 * PCA_1 * \gamma_2 + 0.180035 * PCA_1 * \gamma_3 + 0.14668 \\ & * PCA_2 * \gamma_1 + 0.1333557 * PCA_2 * \gamma_2 + 0.139305 * PCA_2 * \gamma_3 \\ & + 0.227076 * PCA_3 * \gamma_1 + 0.214151 * PCA_3 * \gamma_2 + 0.220184 * PCA_3 * \gamma_3 \\ & - 0.14634 * PCA_1 * PCA_2 * \gamma_1 - 0.14408 * PCA_1 * PCA_2 * \gamma_2 - 14345 \\ & * PCA_1 * PCA_2 * \gamma_3 - 0.19792 * PCA_1 * PCA_3 * \gamma_1 - 0.19676 * PCA_1 \\ & * PCA_3 * \gamma_2 - 0.19706 * PCA_1 * PCA_3 * \gamma_3 - 0.05468 * PCA_1 * \gamma_1 * \gamma_2 \\ & - 0.06048 * PCA_1 * \gamma_1 * \gamma_3 + 0.047967 * PCA_1 * \gamma_1 * MGI - 0.04427 \\ & * PCA_1 * \gamma_3 + 0.042785 * PCA_1 * \gamma_2 * MGI + 0.058986 * PCA_1 * \gamma_3 \\ & * MGI - 0.16526 * PCA_2 * PCA_3 * \gamma_1 - 0.16532 * PCA_2 * PCA_3 * \gamma_2 \\ & - 0.16482 * PCA_2 * PCA_3 * \gamma_3 - 0.05278 * PCA_2 * \gamma_1 * \gamma_2 - 0.0568 \\ & * PCA_2 * \gamma_1 * \gamma_3 - 0.01567 * PCA_2 * \gamma_1 * MGI - 0.04487 * PCA_2 * \gamma_2 \\ & * \gamma_3 - 0.02049 * PCA_2 * \gamma_2 * MGI - 0.00402 * PCA_2 * \gamma_3 * MGI \\ & - 0.0497 * PCA_3 * \gamma_1 * \gamma_2 - 0.05536 * PCA_3 * \gamma_1 * \gamma_3 + 0.004337 \\ & * PCA_3 * \gamma_1 * MGI - 0.0426 * PCA_3 * \gamma_2 * \gamma_3 - 0.00042 * PCA_3 * \gamma_2 \\ & * MGI + 0.015774 * PCA_3 * \gamma_3 * MGI - 0.00284 * PCA_1 * PCA_2 * \gamma_1 * \gamma_2 \\ & - 0.00638 * PCA_1 * PCA_2 * \gamma_1 * \gamma_3 - 0.0052 * PCA_1 * PCA_2 * \gamma_1 * MGI \\ & - 0.01479 * PCA_1 * PCA_2 * \gamma_2 * \gamma_3 - 0.00494 * PCA_1 * PCA_2 * \gamma_2 * MGI \\ & - 0.00459 * PCA_1 * PCA_2 * \gamma_3 * MGI - 0.00414 * PCA_1 * PCA_3 * \gamma_1 * \gamma_2 \\ & - 0.0065 * PCA_1 * PCA_3 * \gamma_1 * \gamma_3 - 0.03699 * PCA_1 * PCA_3 * \gamma_1 * MGI \\ & - 0.01271 * PCA_1 * PCA_3 * \gamma_2 * \gamma_3 - 0.03615 * PCA_1 * PCA_3 * \gamma_2 * MGI \\ & - 0.03669 * PCA_1 * PCA_3 * \gamma_3 * MGI + 0.003677 * PCA_1 * \gamma_1 * \gamma_2 * MGI \\ & - 0.01286 * PCA_1 * \gamma_1 * \gamma_3 * MGI - 0.00605 * PCA_1 * \gamma_2 * \gamma_3 * MGI \\ & - 0.00556 * PCA_2 * PCA_3 * \gamma_1 * \gamma_2 - 0.01111 * PCA_2 * PCA_3 * \gamma_1 * \gamma_3 \\ & + 0.017186 * PCA_2 * PCA_3 * \gamma_1 * MGI - 0.00416 * PCA_2 * PCA_3 * \gamma_3 \\ & * \gamma_3 + 0.16821 * PCA_2 * PCA_3 * \gamma_2 * MGI + 0.016226 * PCA_2 * PCA_3 \\ & * \gamma_3 * MGI + 0.001353 * PCA_2 * \gamma_1 * \gamma_2 * MGI - 0.01612 * PCA_2 * \gamma_1 \\ & * \gamma_3 * MGI - 0.01047 * PCA_2 * \gamma_2 * \gamma_3 * MGI - 0.00122 * PCA_3 * \gamma_1 * \gamma_1 \\ & * MGI - 0.01701 * PCA_3 * \gamma_1 * \gamma_3 * MGI - 0.01227 * PCA_3 * \gamma_2 * \gamma_3 \\ & * MGI + 0.016572 * PCA_1 * PCA_2 * \gamma_1 * \gamma_2 * MGI + 0.013098 * PCA_1 \\ & * PCA_2 * \gamma_1 * \gamma_3 * MGI + 0.011106 * PCA_1 * PCA_2 * \gamma_2 * \gamma_3 * MGI \\ & + 0.016836 * PCA_1 * PCA_3 * \gamma_1 * \gamma_2 * MGI + 0.022711 * PCA_1 * PCA_3 \\ & * \gamma_1 * \gamma_3 * MGI + 0.012286 * PCA_1 * PCA_3 * \gamma_2 * \gamma_3 * MGI + 0.009685 \\ & * PCA_2 * PCA_3 * \gamma_1 * \gamma_2 * MGI + 0.010697 * PCA_2 * PCA_3 * \gamma_1 * \gamma_3 \\ & * MGI + 0.006231 * PCA_2 * PCA_3 * \gamma_2 * \gamma_3 * MGI \end{aligned} $
--	---

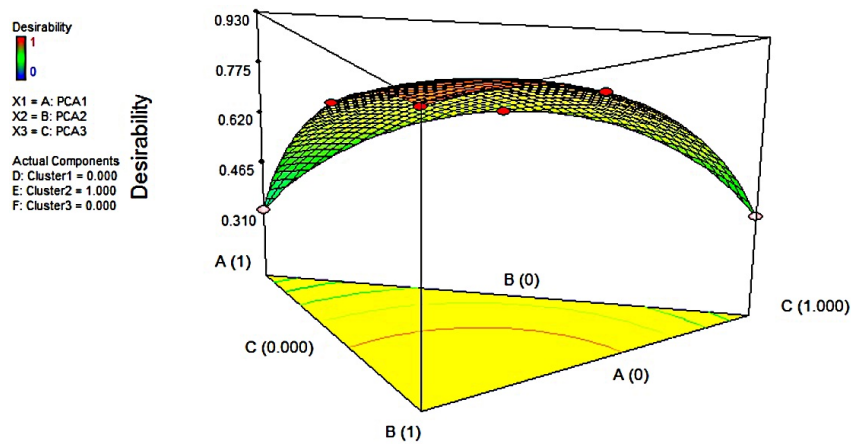


Figura 6.23 – Solução de otimização para o arranjo CM-DOE 3D

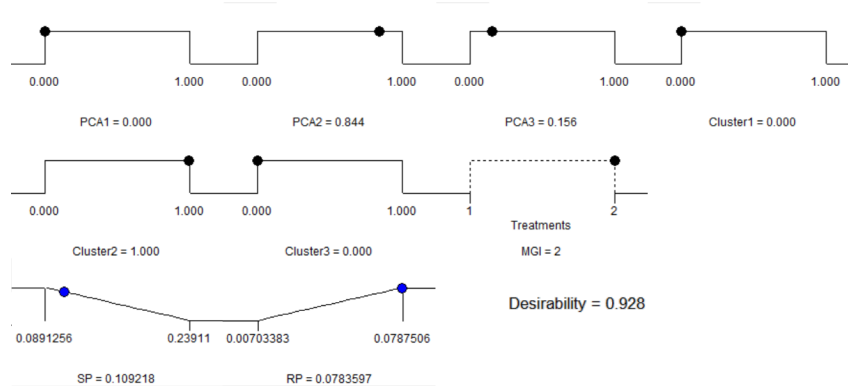


Figura 6.24 – Resultado rampa de otimização do arranjo CM-DOE

6.1.8 Modelagem de todas as empresas juntas utilizando o modelo Markowitz

Esta etapa compreende a análise, simultaneamente, das vinte empresas selecionadas pelo PCA aplicado a classificação da *Newsweek*. Um portfólio comportará todas as empresas, sem considerar variáveis de processo nem *clusters*. Este modelo é usado como comparativo perante os outros. As análises são as mesmas e a otimização segue o mesmo padrão de maximizar os retornos com o risco igual a 0,05 ou minimizar o risco se esse valor não for possível. O arranjo de misturas completo está mostrado na Tabela 6.10 e está também disponível em Papandrea e Paiva (2016a). As funções de risco SP e retorno RP estão na Tabela 6.16. As análises completas podem também ser encontradas em Papandrea Pedro; Paiva (2016). Os gráficos e tabelas das análises completas encontram-se no Apêndice D.

Quadro 6.16 – Função de otimização modelo Markowitz

Convencional	$ \begin{aligned} SP = & 0.096906 * AAPL + 0.09446 * ADBE + 0.070885 * ADI + 0.114727 * BHI \\ & + 0.078295 * IHS + 0.133466 * MOS + 0.074377 * QCOM + 0.143823 \\ & * SPRINT + 0.059973 * SIAL + 0.069714 * TAP + 0.09045 * BIIB \\ & + 0.082051 * EMC + 0.064761 * NKE + 0.058764 * BLL + 0.086295 \\ & * CCE + 0.041334 * CLX + 0.084597 * CMA + 0.125956 * MAS \\ & + 0.104745 * MET + 0.163224 * WYN + 10.88067 * AAPL * ADBE \\ & + 10.83849 * AAPL * ADI - 1.30735 * AAPL * BHI - 1.58431 * AAPL \\ & * IHS - 1.53092 * AAPL * MOS - 1.28834 * AAPL * QCOM - 1.98679 \\ & * AAPL * SPRINT - 1.22483 * AAPL * SIAL - 1.335 * AAPL * TAP \\ & - 1.6369 * AAPL * BIIB - 1.32151 * AAPL * EMC - 1.2895 * AAPL \\ & * NKE - 1.33663 * AAPL * BLL - 1.19171 * AAPL * CCE - 1.39784 \\ & * AAPL * CLX - 1.33886 * AAPL * CMA - 1.19363 * AAPL * MAS \\ & - 1.03278 * AAPL * MET - 1.16717 * AAPL * WYN - 12.0023 * ADBE \\ & * ADI \end{aligned} $
Convencional	$ \begin{aligned} RP = & 0.025035 * AAPL + 0.007888 * ADBE + 0.004399 * ADI + 0.000433 * BHI \\ & + 0.015121 * IHS + 0.009266 * MOS + 0.003434 * QCOM - 0.0146 \\ & * SPRINT + 0.012674 * SIAL + 0.006934 * TAP + 0.019782 * BIIB \\ & + 0.0005801 * EMC + 0.014454 * NKE + 0.011358 * BLL + 0.006218 \\ & * CCE + 0.005103 * CLX - 0.00128 * CMA - 0.00182 * MAS \\ & + 0.001043 * MET + 0.008582 * WYN \end{aligned} $

Tabela 6.10 – Arranjo de misturas para portfólio de modelo Markowitz

[illegible]

É possível notar que as respostas são influenciadas pelas mudanças nos níveis, proporções, de cada linha dos experimentos. Isto demonstra que os fatores de entrada, as proporções, operam mudanças positivas ou negativas nas respostas Retorno e Risco dos portfólios.

Pela análise das ANOVAS (D) conclui-se que o modelo estatístico é bom, explicado pelas variáveis, e pode ser usado para previsão e otimização. As funções de otimização tem ajuste maior do que 85% sendo suficientes para uma boa modelagem estatística. As réplicas feitas pela autocorrelação foram suficientes para tornar este modelo determinístico em um modelo normal, provado pela ocorrência e medição dos erros entre réplicas. Este modelo é o mais simples porque considera todas as empresas em uma mistura única. As empresas participantes desta modelagem foram previamente selecionadas pelo MGI aplicado aos indicadores de empresas verdes, sendo que não houve aleatoriedade nas escolhas, mesmo para este modelo comparativo.

6.1.9 Otimização utilizando o modelo de Markowitz com otimização por MDE

Os *setups* para as otimizações foram especificados para maximizar os retornos contendo um valor fixado de 0,05 para o risco. A Figura 6.25 mostra a rampa de otimização e demonstra os valores mínimos e máximos a serem otimizados pelo método *desirability*. Essas rampas são também uma interessante demonstração gráfica onde estão os valores mínimos encontrados pelas funções e os valores máximos, tanto para retorno (RP) quanto para risco (SP). Estes valores estão alocados em cada ativo das empresas selecionadas. Após a otimização, em cada rampa haverá um ponto marcado, que será o ponto ótimo que atende a todas as especificações feitas nos *setups*.

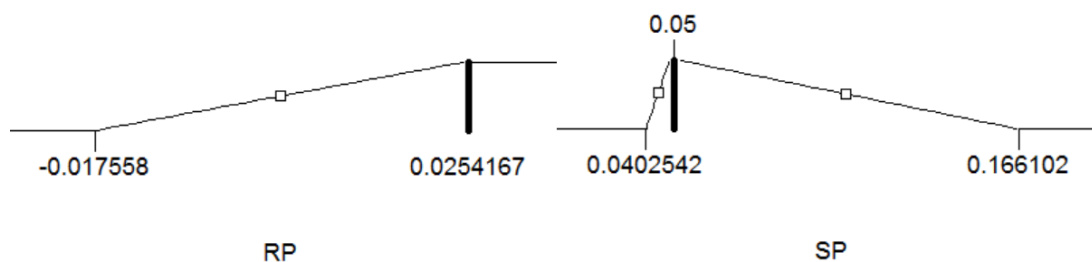


Figura 6.25 – Rampa de otimização Markowitz

Quadro 6.17 – Resultados modelo Markowitz (quadro 1 de 2)

Nome	Objetivo	Limite Inferior	Limite Superior	Menor Peso	Maior Peso	Importância
AAPL	no intervalo	0	1	1	1	3
ADBE	no intervalo	0	1	1	1	3
ADI	no intervalo	0	1	1	1	3
BHI	no intervalo	0	1	1	1	3
IHS	no intervalo	0	1	1	1	3
MOS	no intervalo	0	1	1	1	3
QCOM	no intervalo	0	1	1	1	3
SPRINT	no intervalo	0	1	1	1	3
SIAL	no intervalo	0	1	1	1	3
TAP	no intervalo	0	1	1	1	3
BIIB	no intervalo	0	1	1	1	3
EMC	no intervalo	0	1	1	1	3
NKE	no intervalo	0	1	1	1	3
BLL	no intervalo	0	1	1	1	3
CCE	no intervalo	0	1	1	1	3
CLX	no intervalo	0	1	1	1	3
CMA	no intervalo	0	1	1	1	3
MAS	no intervalo	0	1	1	1	3
MET	no intervalo	0	1	1	1	3
WYN	no intervalo	0	1	1	1	3
RP	maximizar	-0,0176	0,0254	1	1	3
SP	no alvo = 0,05	0,0403	0,1661	1	1	3

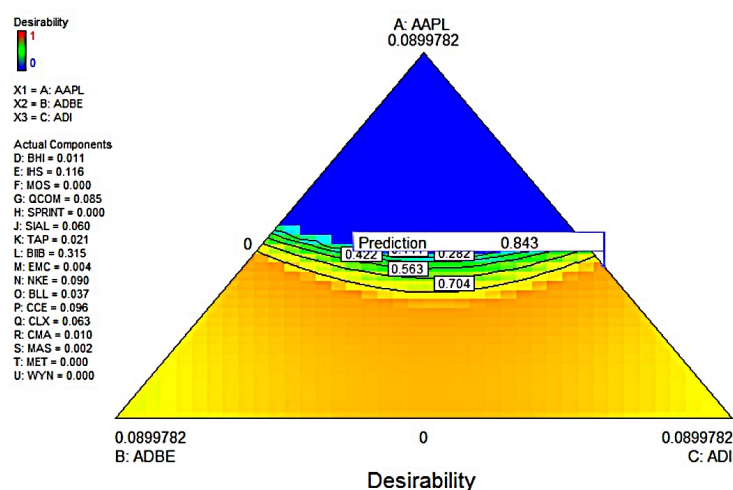


Figura 6.26 – Solução de otimização para o modelo Markowitz 2D

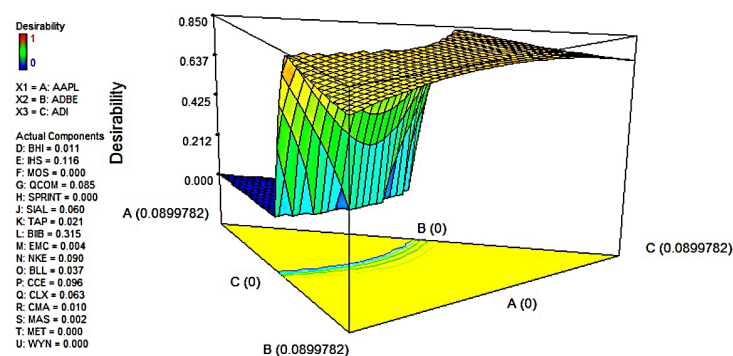


Figura 6.27 – Solução de otimização para o modelo Markowitz 3D

Quadro 6.18 – Resultados modelo Markowitz (quadro 2 de 2)

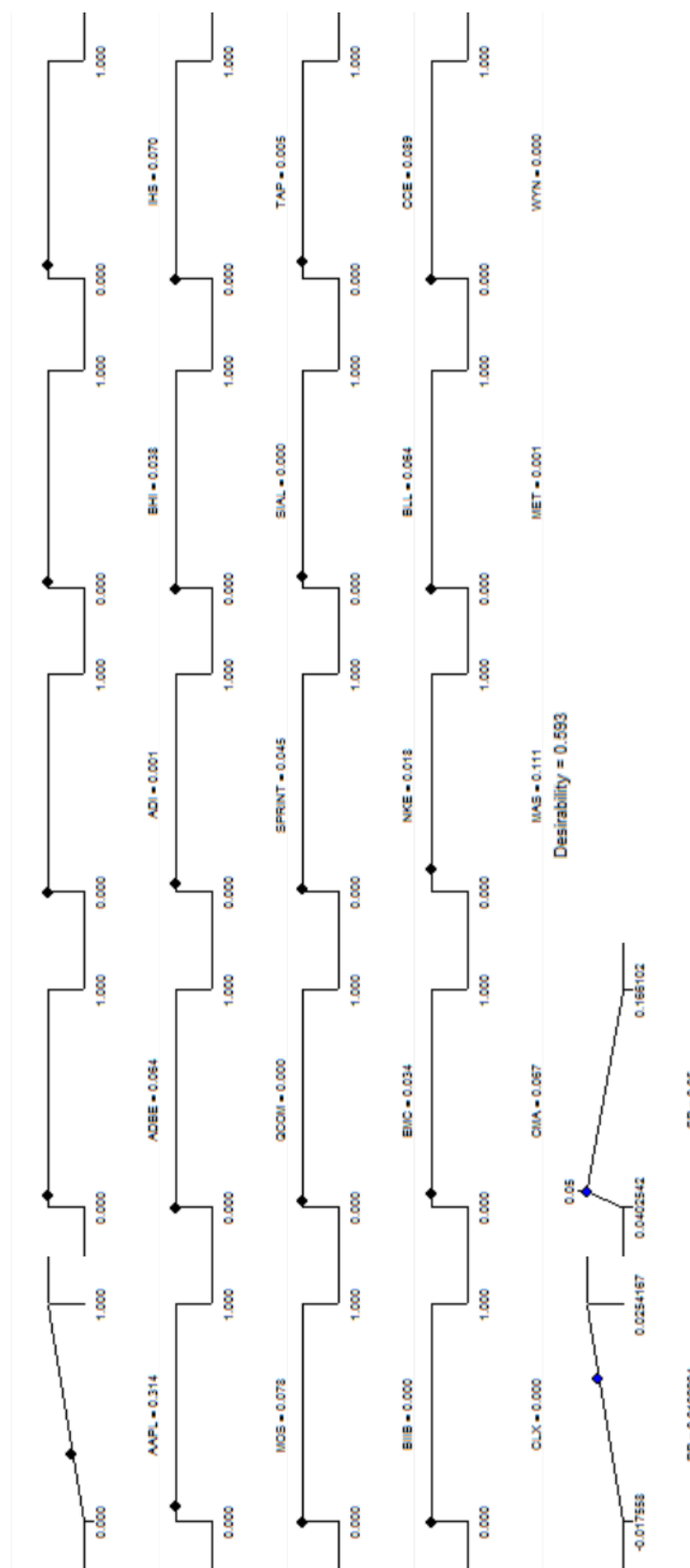


Figura 6.28 – Resultado rampa de otimização do modelo Markowitz

6.2 COMPARAÇÃO DOS MÉTODOS

Para uma comprovação da eficácia ou não dos métodos CMO e CM-DOE sobre o método de Markowitz, esta seção traz uma comparação feita através de uma simulação de investimentos. A comparação é feita com um investimento simulado de R\$100.000,00. Este investimento considera ano inicial de 2007 com venda dos ativos em 2017, totalizando 10 anos de movimentação. Os valores, proporções deste montante que será investido em cada empresa participante dos portfólios advém dos resultados das otimizações feitas anteriormente. A partir destes investimentos, a simulação retorna os valores correspondentes a cada método utilizado. Há um método não mencionado nas otimizações que também será demonstrado, 1/20. Este é um investimento igualitário para todas as empresas, como uma forma extra de comparação. É fundamental observar que para todos os métodos, as empresas selecionadas provieram da seleção feita por PCA com MGI nos indicadores de empresas verdes da *Newsweek*. Essa pré-seleção das empresas já é uma melhoria, pois considera empresas com melhor visibilidade do mercado. Essa visibilidade é tendenciosa para um aumento nos preços das ações. Empresas verdes, além disso, recebem incentivos governamentais pelas suas contribuições no que tange à sustentabilidade, consumo inteligente de recursos energéticos e naturais entre outras formas de redução na degradação do meio ambiente e melhorias no mesmo.

As Tabelas 6.11, 6.12 e 6.12 mostram os resultados dos investimentos simulados descritos anteriormente. Nelas, tem-se a coluna Mês com todos os fechamentos mensais relativos aos investimentos. Para efeito das tabelas, estes fechamentos vão de 28 de Setembro de 2007 a 02 de Agosto de 2017.

A coluna CMO refere-se a uma das propostas desta tese, o arranjo combinado que utiliza modelo lexicográfico para otimização de misturas de misturas com as variáveis financeiras acopladas como variáveis de processo. Nesta houveram divisões das empresas selecionadas em *clusters*, os *clusters* foram otimizados e depois agrupados em um novo experimento de mistura que também foi otimizado. O resultado final foi uma otimização que definiu o percentual a se investir em cada *cluster* e finalmente em cada empresa participante destes. O método CMO foi o melhor em todos os testes e comprovando sua robustez e aplicabilidade.

A coluna CM-DOE refere-se também a uma proposta desta tese, este arranjo considera misturas com variáveis de processo. Neste modelo, as empresas participantes foram selecionadas por MGI aplicado em indicadores financeiros. Este modelo, no entanto teve a pior performance dentre os modelos testados, mesmo assim resultou em ganhos já que as empresas foram pré-selecionadas através da aplicação do PCA com MGI nos indicadores de empresas verdes. Por não ter tido resultados superiores ao CMO, este modelo não aparece nos gráficos comparativos (Figuras 6.29, 6.30).

A coluna Markowitz traz os resultados de um arranjo de misturas sem variável de processo. Neste arranjo estão contidas todas as 20 empresas pré-selecionadas. Estes resultados demonstraram uma excelente escolha, ficando em segundo lugar.

Finalmente a coluna 1/20 traz os resultados de um investimento igualitário para todas as empresas pré-selecionadas. Neste caso houve também um bom retorno no investimento.

A Figura 6.29 é a demonstração gráfica dos investimentos simulados para cada método de comparação. Nota-se que em todos os casos houve uma queda brusca nos valores das ações no período que norteia o ano de 2008. Isto se deu pela crise ocorrida nos Estados Unidos da América em 2008 e que refletiu-se em todo o mundo. A data inicial que aparece no gráfico é primeiro de Setembro de 2007 e a data final considerada foi primeiro de Julho de 2017. É perceptível que os três modelos comparados tem um retorno relativamente similar até Janeiro de 2013. Isto é explicado porque os indicadores financeiros que foram compilados e gerados neste estudo datam da divulgação dos balanços patrimoniais, em 2014. Estes balanços referem-se ao período contábil de 2013. Portanto essa diferença e avanço nos resultados do modelo proposto CMO.

A Figura 6.30 é um detalhamento das diferenças entre investimentos com uma linha de corte em 2013. Note-se que o arranjo denominado CMO, proposta desta tese, teve o melhor desempenho, seguido do método de Markowitz. Está análise gráfica reforça a eficiência do método proposto. Todas as empresas em todas as simulações provieram da primeira fase desta tese que foi a seleção das empresas por PCA com MGI aplicado ao ranqueamento da revista *Newsweek*.

As comparações serviram para, além de comprovar a viabilidade prática da metodologia proposta, entender como os indicadores financeiros impactam nos investimentos em ações, no mercado financeiro. Esses indicadores foram inseridos aqui como variável de processo em arranjos de mistura com metodologia de otimização lexicográfica, o que resultou em um retorno melhor e mais robusto do que o modelo de Markowitz. Esta contribuição foi parte fundamental desta tese.

6.3 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO

Neste capítulo, discutiu-se a maneira como diferentes métodos de seleção de portfólios levam a resultados diferentes quanto a porcentagem de investimento em cada Ativo. Foram apresentadas as deduções matemáticas, as principais estatísticas de teste, as vantagens e as desvantagens de cada um dos métodos, além dos gráficos de cada modelo.

O próximo capítulo se dedica à conclusão do trabalho, consolidando o potencial dos modelos desenvolvidos em comparação ao modelo tradicional.

Tabela 6.11 – Comparação dos métodos (Tabela 1 de 3)

Mês	CMO	CM-DOE	Markowitz	1/20
2-Aug-17	R\$ 585.073,67	R\$ 285.918,85	R\$ 511.251,55	R\$ 400.700,68
31-Jul-17	R\$ 587.778,41	R\$ 281.861,58	R\$ 509.344,44	R\$ 393.843,20
30-Jun-17	R\$ 567.109,38	R\$ 281.622,91	R\$ 495.091,88	R\$ 386.208,26
31-May-17	R\$ 529.610,13	R\$ 252.935,56	R\$ 484.938,92	R\$ 391.159,46
28-Apr-17	R\$ 556.910,48	R\$ 264.486,87	R\$ 490.797,22	R\$ 384.446,66
31-Mar-17	R\$ 558.218,72	R\$ 266.014,32	R\$ 490.674,17	R\$ 382.493,28
28-Feb-17	R\$ 577.281,11	R\$ 272.840,10	R\$ 494.151,92	R\$ 377.446,50
31-Jan-17	R\$ 539.532,40	R\$ 252.505,97	R\$ 469.613,17	R\$ 356.178,32
30-Dec-16	R\$ 542.630,56	R\$ 242.625,30	R\$ 466.448,93	R\$ 348.181,48
30-Nov-16	R\$ 552.851,71	R\$ 238.997,61	R\$ 468.520,09	R\$ 342.952,63
31-Oct-16	R\$ 539.299,65	R\$ 239.522,67	R\$ 462.719,64	R\$ 338.846,33
30-Sep-16	R\$ 581.416,13	R\$ 251.312,65	R\$ 485.523,11	R\$ 347.016,11
31-Aug-16	R\$ 578.803,94	R\$ 275.131,26	R\$ 478.738,56	R\$ 340.297,21
29-Jul-16	R\$ 553.293,46	R\$ 264.916,47	R\$ 462.588,33	R\$ 332.550,61
30-Jun-16	R\$ 502.012,82	R\$ 263.484,49	R\$ 431.700,01	R\$ 313.728,57
31-May-16	R\$ 553.739,64	R\$ 263.579,95	R\$ 458.540,85	R\$ 323.536,17
29-Apr-16	R\$ 542.581,08	R\$ 281.336,52	R\$ 450.446,23	R\$ 318.493,68
31-Mar-16	R\$ 530.325,29	R\$ 293.412,89	R\$ 450.400,81	R\$ 329.487,09
29-Feb-16	R\$ 526.241,05	R\$ 293.985,68	R\$ 434.220,97	R\$ 309.570,50
29-Jan-16	R\$ 541.959,98	R\$ 295.990,45	R\$ 441.401,30	R\$ 310.509,23
31-Dec-15	R\$ 582.392,12	R\$ 298.329,36	R\$ 469.450,93	R\$ 329.812,46
30-Nov-15	R\$ 564.462,90	R\$ 315.704,06	R\$ 466.669,18	R\$ 341.988,94
30-Oct-15	R\$ 566.906,09	R\$ 312.720,76	R\$ 468.978,37	R\$ 344.801,96
30-Sep-15	R\$ 558.445,67	R\$ 293.484,49	R\$ 455.376,15	R\$ 326.137,95
31-Aug-15	R\$ 562.464,96	R\$ 274.988,07	R\$ 459.981,07	R\$ 331.533,50
31-Jul-15	R\$ 581.981,42	R\$ 257.804,30	R\$ 476.192,05	R\$ 344.002,12
30-jun-15	R\$ 584.250,26	R\$ 242.649,16	R\$ 481.454,86	R\$ 348.666,55
29-mai-15	R\$ 575.766,71	R\$ 235.894,99	R\$ 479.294,51	R\$ 352.635,87
30-abr-15	R\$ 548.480,32	R\$ 239.451,07	R\$ 463.496,76	R\$ 344.393,32
31-mar-15	R\$ 602.512,28	R\$ 231.789,98	R\$ 483.116,76	R\$ 346.880,30
27-fev-15	R\$ 579.034,58	R\$ 220.167,06	R\$ 479.847,14	R\$ 352.728,11
30-jan-15	R\$ 561.412,13	R\$ 229.474,94	R\$ 459.491,40	R\$ 330.965,92
31-Dec-14	R\$ 505.722,12	R\$ 236.968,97	R\$ 433.348,29	R\$ 323.096,28
28-Nov-14	R\$ 462.907,40	R\$ 221.885,44	R\$ 421.793,32	R\$ 327.906,69
31-Oct-14	R\$ 474.872,53	R\$ 212.887,83	R\$ 424.430,32	R\$ 318.935,95
30-Sep-14	R\$ 478.977,75	R\$ 187.470,17	R\$ 421.737,49	R\$ 313.396,46
29-Aug-14	R\$ 490.214,48	R\$ 184.081,15	R\$ 430.021,49	R\$ 316.997,67
31-Jul-14	R\$ 481.462,55	R\$ 185.083,53	R\$ 414.636,78	R\$ 303.864,30
30-Jun-14	R\$ 457.537,77	R\$ 183.556,09	R\$ 406.553,99	R\$ 305.784,06
30-May-14	R\$ 456.663,27	R\$ 174.105,01	R\$ 397.288,41	R\$ 296.589,96

Tabela 6.12 – Comparação dos métodos (Tabela 2 de 3)

Mês	CMO	CM-DOE	Markowitz	1/20
30-Apr-14	R\$ 419.282,46	R\$ 280.558,66	R\$ 372.180,14	R\$ 176.276,85
31-Mar-14	R\$ 444.549,27	R\$ 275.751,96	R\$ 381.846,66	R\$ 186.873,51
28-Feb-14	R\$ 477.904,59	R\$ 271.121,24	R\$ 394.577,50	R\$ 173.866,35
31-Jan-14	R\$ 450.057,18	R\$ 255.967,38	R\$ 371.227,85	R\$ 187.684,96
31-Dec-13	R\$ 411.258,43	R\$ 273.854,49	R\$ 362.013,19	R\$ 188.878,28
29-Nov-13	R\$ 421.252,75	R\$ 264.170,57	R\$ 360.251,37	R\$ 180.811,46
31-Oct-13	R\$ 362.524,21	R\$ 252.741,81	R\$ 327.320,34	R\$ 173.365,16
30-Sep-13	R\$ 351.233,37	R\$ 240.884,41	R\$ 320.726,62	R\$ 149.928,40
30-Aug-13	R\$ 319.589,84	R\$ 236.343,48	R\$ 301.001,43	R\$ 150.167,06
31-Jul-13	R\$ 323.928,01	R\$ 233.827,54	R\$ 302.382,90	R\$ 151.980,91
28-Jun-13	R\$ 320.142,35	R\$ 222.127,03	R\$ 294.711,81	R\$ 147.159,90
31-May-13	R\$ 347.822,67	R\$ 234.978,67	R\$ 314.707,20	R\$ 151.789,98
30-Apr-13	R\$ 325.371,29	R\$ 229.932,28	R\$ 299.326,18	R\$ 140.835,32
28-Mar-13	R\$ 289.749,61	R\$ 233.800,45	R\$ 286.424,76	R\$ 129.976,13
28-Feb-13	R\$ 259.249,91	R\$ 228.855,23	R\$ 270.283,40	R\$ 128.997,61
31-Jan-13	R\$ 245.830,71	R\$ 229.405,34	R\$ 265.070,83	R\$ 123.150,36
31-Dec-12	R\$ 232.541,42	R\$ 231.853,60	R\$ 257.539,98	R\$ 116.324,58
30-Nov-12	R\$ 232.286,13	R\$ 235.835,22	R\$ 257.203,19	R\$ 109.045,35
31-Oct-12	R\$ 220.694,40	R\$ 232.126,03	R\$ 247.003,56	R\$ 113.257,76
28-Sep-12	R\$ 234.146,26	R\$ 251.854,67	R\$ 267.032,93	R\$ 116.181,38
31-Aug-12	R\$ 229.295,36	R\$ 255.451,25	R\$ 270.981,59	R\$ 111.396,18
31-Jul-12	R\$ 226.081,54	R\$ 244.569,94	R\$ 263.458,96	R\$ 104.749,40
29-Jun-12	R\$ 231.027,65	R\$ 238.166,03	R\$ 261.086,09	R\$ 129.093,08
31-May-12	R\$ 217.802,32	R\$ 227.988,40	R\$ 246.215,60	R\$ 133.496,42
30-Apr-12	R\$ 221.075,37	R\$ 237.814,78	R\$ 255.729,85	R\$ 129.403,34
30-Mar-12	R\$ 209.857,80	R\$ 240.105,98	R\$ 250.962,31	R\$ 128.782,82
29-Feb-12	R\$ 196.946,13	R\$ 229.846,30	R\$ 239.956,30	R\$ 124.093,08
31-Jan-12	R\$ 193.613,15	R\$ 210.651,03	R\$ 227.244,52	R\$ 115.000,00
30-Dec-11	R\$ 184.043,21	R\$ 194.001,47	R\$ 211.868,94	R\$ 114.773,27
30-Nov-11	R\$ 189.400,17	R\$ 192.349,86	R\$ 215.143,28	R\$ 114.976,13
31-Oct-11	R\$ 184.856,37	R\$ 198.153,60	R\$ 215.389,80	R\$ 102.040,57
30-Sep-11	R\$ 161.672,90	R\$ 179.420,49	R\$ 192.178,15	R\$ 103.400,95
31-Aug-11	R\$ 166.018,18	R\$ 195.906,79	R\$ 204.864,53	R\$ 107.577,57
29-Jul-11	R\$ 174.598,78	R\$ 202.471,75	R\$ 211.078,24	R\$ 107.374,70
30-Jun-11	R\$ 179.120,82	R\$ 199.862,99	R\$ 217.376,88	R\$ 100.775,66
31-May-11	R\$ 162.895,33	R\$ 206.836,97	R\$ 214.130,53	R\$ 98.233,89
29-Apr-11	R\$ 162.951,07	R\$ 208.757,74	R\$ 215.540,43	R\$ 90.334,13
31-Mar-11	R\$ 140.033,39	R\$ 205.714,83	R\$ 202.092,94	R\$ 106.241,05
28-Feb-11	R\$ 131.508,75	R\$ 207.819,51	R\$ 199.207,81	R\$ 98.424,82
31-Jan-11	R\$ 128.187,07	R\$ 199.438,19	R\$ 192.166,05	R\$ 101.933,17

Tabela 6.13 – Comparação dos métodos (Tabela 3 de 3)

Mês	CMO	CM-DOE	Markowitz	1/20
31-Dec-10	R\$ 129.555,60	R\$ 193.102,87	R\$ 188.724,59	R\$ 102.780,43
30-Nov-10	R\$ 123.813,15	R\$ 180.398,84	R\$ 176.531,76	R\$ 97.183,77
29-Oct-10	R\$ 120.122,87	R\$ 179.872,85	R\$ 174.712,35	R\$ 95.632,46
30-Sep-10	R\$ 107.995,11	R\$ 170.502,17	R\$ 162.134,09	R\$ 83.532,22
31-Aug-10	R\$ 107.323,10	R\$ 155.478,65	R\$ 151.949,77	R\$ 87.875,89
30-Jul-10	R\$ 105.802,74	R\$ 158.185,04	R\$ 151.920,30	R\$ 80.608,59
30-Jun-10	R\$ 96.652,52	R\$ 145.732,75	R\$ 138.245,10	R\$ 86.372,32
28-May-10	R\$ 99.184,37	R\$ 150.344,18	R\$ 140.916,27	R\$ 90.584,73
30-Apr-10	R\$ 105.142,87	R\$ 158.791,18	R\$ 147.999,53	R\$ 87.708,83
31-Mar-10	R\$ 107.955,44	R\$ 156.909,07	R\$ 150.154,24	R\$ 80.668,26
26-Feb-10	R\$ 102.759,07	R\$ 146.523,75	R\$ 141.405,15	R\$ 76.073,99
29-Jan-10	R\$ 102.418,33	R\$ 139.875,53	R\$ 137.913,07	R\$ 78.842,48
31-Dec-09	R\$ 100.978,92	R\$ 149.591,19	R\$ 144.960,27	R\$ 77.434,37
30-Nov-09	R\$ 92.421,39	R\$ 143.169,60	R\$ 136.616,57	R\$ 74.200,48
30-Oct-09	R\$ 87.742,02	R\$ 136.969,45	R\$ 131.106,36	R\$ 77.207,64
30-Sep-09	R\$ 93.843,03	R\$ 139.213,67	R\$ 135.970,84	R\$ 66.097,85
31-Aug-09	R\$ 93.889,20	R\$ 133.119,29	R\$ 132.125,98	R\$ 67.589,50
31-Jul-09	R\$ 88.046,92	R\$ 133.336,89	R\$ 129.765,64	R\$ 61.789,98
30-Jun-09	R\$ 85.426,04	R\$ 121.264,10	R\$ 121.360,85	R\$ 68.078,76
29-May-09	R\$ 90.765,20	R\$ 122.633,18	R\$ 124.455,25	R\$ 62.613,37
30-Apr-09	R\$ 86.620,14	R\$ 111.091,08	R\$ 113.897,52	R\$ 55.954,65
31-Mar-09	R\$ 88.368,33	R\$ 99.609,25	R\$ 109.071,41	R\$ 49.558,47
27-Feb-09	R\$ 81.708,09	R\$ 93.482,01	R\$ 102.204,58	R\$ 53.997,61
30-Jan-09	R\$ 87.862,27	R\$ 95.881,20	R\$ 106.510,18	R\$ 60.859,19
31-Dec-08	R\$ 85.854,67	R\$ 97.552,43	R\$ 104.809,85	R\$ 63.544,15
28-Nov-08	R\$ 80.548,75	R\$ 95.399,99	R\$ 99.676,60	R\$ 68.770,88
31-Oct-08	R\$ 85.873,91	R\$ 103.672,18	R\$ 106.237,97	R\$ 79.832,94
30-Sep-08	R\$ 94.691,92	R\$ 134.881,09	R\$ 132.199,30	R\$ 72.326,97
29-Aug-08	R\$ 131.992,23	R\$ 169.598,56	R\$ 175.271,56	R\$ 144.653,94
31-Jul-08	R\$ 151.419,11	R\$ 176.315,92	R\$ 189.919,70	R\$ 140.047,73
30-Jun-08	R\$ 138.159,47	R\$ 182.739,23	R\$ 189.336,64	R\$ 142.267,30
30-May-08	R\$ 156.607,71	R\$ 184.651,61	R\$ 194.933,30	R\$ 163.174,22
30-Apr-08	R\$ 152.775,97	R\$ 177.866,05	R\$ 190.987,14	R\$ 159.427,21
31-Mar-08	R\$ 149.719,20	R\$ 163.805,40	R\$ 179.766,94	R\$ 162.291,17
29-Feb-08	R\$ 139.094,33	R\$ 162.966,28	R\$ 174.769,49	R\$ 143.675,42
31-Jan-08	R\$ 144.001,66	R\$ 156.316,69	R\$ 170.186,16	R\$ 146.372,32
31-Dec-07	R\$ 140.926,11	R\$ 174.515,89	R\$ 177.754,93	R\$ 153.317,42
30-Nov-07	R\$ 162.570,43	R\$ 167.892,02	R\$ 184.990,11	R\$ 156.682,58
31-Oct-07	R\$ 165.540,07	R\$ 173.684,50	R\$ 189.207,79	R\$ 158.138,42
28-Sep-07	R\$ 153.447,29	R\$ 155.777,18	R\$ 170.658,96	R\$ 140.000,00

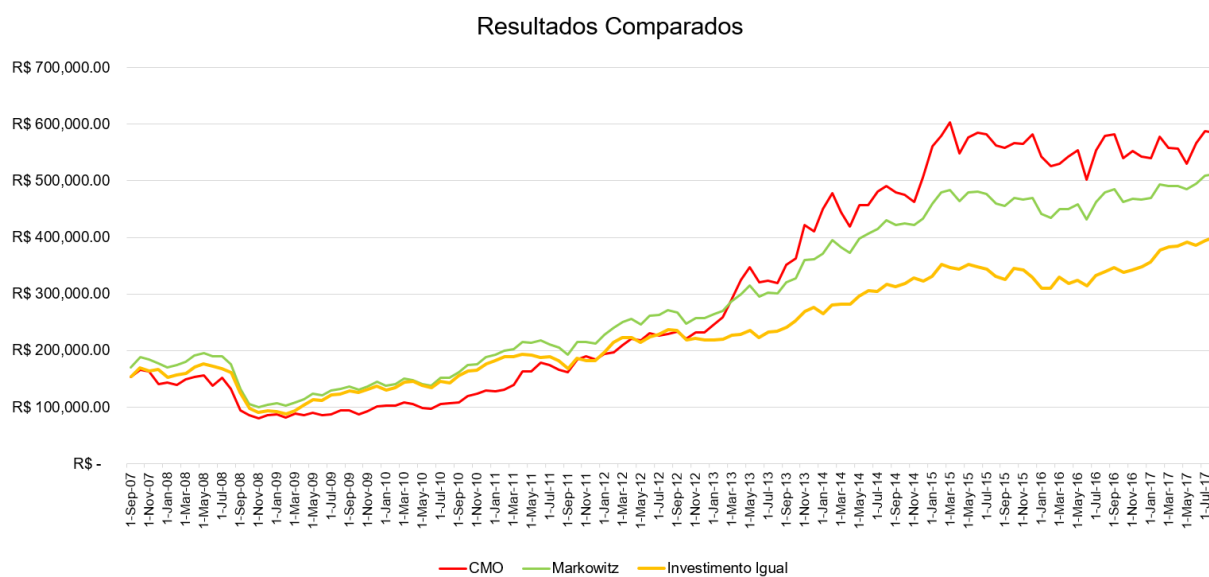


Figura 6.29 – Comparação entre métodos

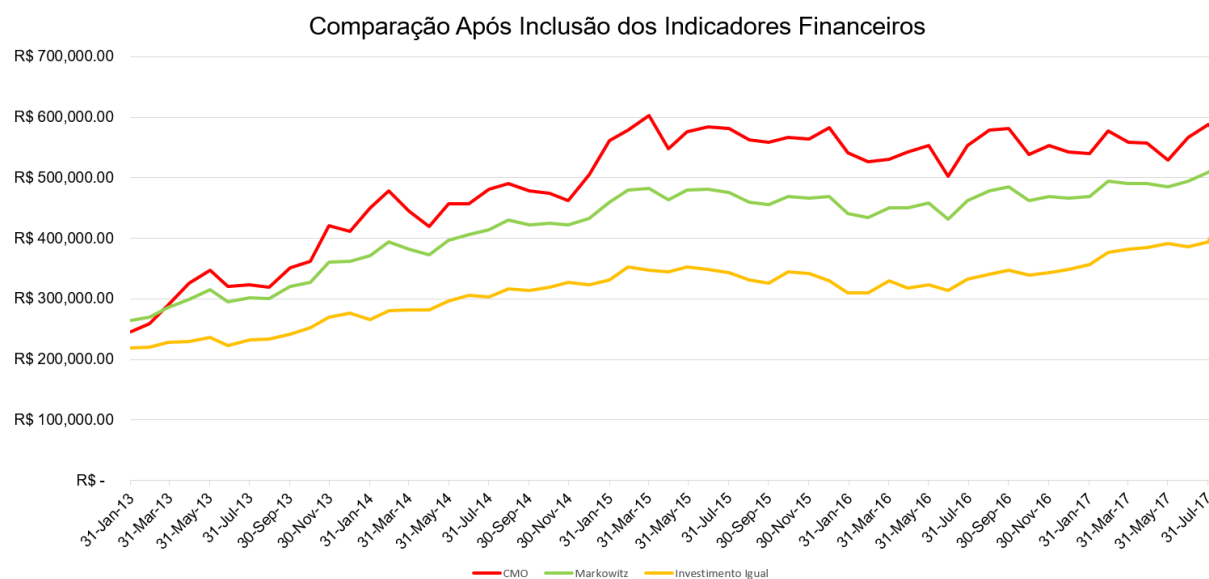


Figura 6.30 – Comparação entre métodos

7 CONCLUSÃO

Este trabalho teve como objetivo o desenvolvimento de um método sistemático para composição de portfólios capaz de tornar a escolha do investidor mais robusta, minimizando seus riscos e maximizando seus retornos. Essa metodologia conta com a integração de delineamento de experimentos com variáveis de processo, além de PCA e *clusters* com otimização por método lexicográfico.

Inicialmente foi necessário escolher as empresas participantes das análises e simulações. Para isto, foi aplicada a análise de componentes principais (PCA) em indicadores de sustentabilidade chamados de indicadores de empresas verdes (*Green Companies*). Os indicadores considerados vêm da publicação anual da revista *Newsweek*. O PCA então, foi indexado pelo MGI e a partir deste foi feito um ranqueamento dessas empresas verdes. Deste ranqueamento, as vinte empresas com maior MGI foram selecionadas para as análises e simulações. Esta técnica traz um embasamento para o investidor escolher as empresas que irão compor seu portfólio de investimento e esta forma foi uma inovação proposta neste trabalho.

A etapa seguinte foi obter os valores dos fechamentos mensais das ações das empresas selecionadas. Estes valores estão disponíveis na plataforma eletrônica da bolsa de valores, no caso estudado aqui, a NASDAQ. Os fechamentos sofrem flutuações tendo altas e baixas por conta de valorização ou desvalorização das empresas perante o mercado. Além desses fechamentos, os balanços patrimoniais também foram considerados. A divulgação anual dos balanços é uma exigência para as empresas de capital aberto. Com os fechamentos armazenados, foi possível calcular a média dos retornos, e a variação destes retornos, os riscos. Foi necessário também utilizar um método de replicação dos riscos e retornos, utilizou-se a autocorrelação com $lag = 1$. Assim obteve-se mais de uma série de fechamentos para o uso nos experimentos. Com os balanços patrimoniais disponíveis, fez-se o cálculo dos indicadores financeiros. Cada indicador necessitou de certos dados e informações que foram extraídas manualmente uma a uma desses balanços. Estes indicadores não estão disponíveis em nenhuma plataforma, foram portanto criados para os fins dos estudos aqui apresentados.

Baseado nos indicadores gerados, foram feitas aglutinações das empresas que demonstraram-se similares financeiramente, esta similaridade foi definida pelo método de Ward. Então foram criados os *clusters*. Estes formaram os fatores dos arranjos de misturas juntamente com as variáveis de processo, os PCAs dos indicadores financeiros. Cada *cluster* foi otimizado via função *desirability*, sendo esta a primeira otimização lexicográfica. Após este estágio, um novo arranjo de mistura foi definido. Os valores otimizados dos arranjos anteriores são agora as restrições para o arranjo combinado. Neste, os fatores não serão

empresas, mas sim, os *clusters* otimizados.

Para o próximo experimento, os PCAs foram usados como selecionadores extras. Por seus níveis mais altos e mais baixos fez-se a escolha das empresas e formação do modelo chamado de CM-DOE, no qual foram feitos experimentos considerando empresas diferentes como variáveis de processo, dependendo dos valores a elas definidas dos PCAs.

O modelo de Markowitz, por ser referência na modelagem de portfólios foi escolhido para as comparações nos resultados. Os resultados foram providos por uma simulação de investimento a partir dos valores ótimos encontrados para cada modelo. Esses valores ótimos definem, de forma prática, o percentual do montante de investimento a se alocar em cada empresa participante.

A partir das considerações anteriores, três métodos diferentes foram analisados e tiveram suas funções otimizadas. Destes três, dois métodos são propostas deste trabalho: *clusters* com variáveis de processo otimizados lexicograficamente e *clusters* utilizando arranjos de mistura combinados com variáveis de processo. Através da análise de cada modelo foi possível distinguir que as variáveis de processo atuam sobre os resultados dos portfólios, tornando-os diferentes entre si assim como os modelos são diferentes. Este é um resultado importante pois torna válido o uso de indicadores financeiros incorporados nas análises dos portfólios. Estes indicadores modificam os resultados quando são considerados.

Vale ainda ressaltar que este trabalho apresenta uma proposta de método sistemático para seleção de empresas, composição de portfólio e incorporação de indicadores financeiros advindos dos balanços patrimoniais, diferente da maioria dos trabalhos encontrados na literatura, que têm como foco a análise e otimização de portfólios já formados previamente. Os resultados apresentados comprovam que o método se sobressai aos métodos de mistura que não consideram indicadores financeiros e nem a aglomeração por PCA e este método pode dar mais segurança ao investidor, destacando outra importante contribuição.

Finalmente, o método CMO foi o que obteve melhor resultado nas simulações e foi proposto neste trabalho. Para todos os métodos, as empresas escolhidas são advindas também de proposta deste trabalho, a escolha e ranqueamento pelo uso do PCA com MGI aplicado em indicadores de empresas verdes. A escolha das empresas tem agora uma metodologia sistemática e matemática, isso passa a escolha das empresas de subjetiva para objetiva. Os outros métodos testados também obtiveram bons resultados, visto que, a escolha das empresas não foi aleatória, isto tornou mais robustos também, os outros métodos testados.

7.0.1 Sugestões para trabalhos futuros

Algumas investigações podem ser realizadas em trabalhos futuros a fim de consolidar ainda mais a utilização dos *clusters* e indicadores financeiros para análises de investimentos:

- Avaliar como o fechamento do balanço patrimonial de anos diferentes pode impactar nos PCAs;
- Avaliar a influência de outros indicadores no resultado das otimizações;
- Testar o método em outros tipos de investimentos com outros tipos de indicadores externos ao processo;
- Avaliar a aplicação de outras métricas de análise de risco e retorno;
- Avaliar o uso do método Média-Gini para métricas de risco e retorno.

É possível também a aplicação das metodologias desenvolvidas em outros campos de estudo. A programação da produção pode fazer uso quando agrupa seus produtos com restrições de capacidade e usando a função custo versus lucratividade como variável de processo, por exemplo.

REFERÊNCIAS

- ABBASZADEH, K.; ALAM, F. R.; SAIED, S. Cogging torque optimization in surface-mounted permanent-magnet motors by using design of experiment. *Energy Conversion and Management*, Elsevier, v. 52, n. 10, p. 3075–3082, 2011. Citado na página 17.
- ALTMAN, E. I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, Wiley Online Library, v. 23, n. 4, p. 589–609, 1968. Citado na página 13.
- AMIHUD, Y.; BARNEA, A. Portfolio selection for managerial control. *Omega*, Elsevier, v. 2, n. 6, p. 775–783, 1974. Citado 4 vezes nas páginas 2, 4, 6 e 48.
- ANDERSON, D. A.; THOMAS, A. M. Weakly resolvable iv. 3 search designs for the pn factorial experiment. *Journal of Statistical Planning and Inference*, Elsevier, v. 4, n. 3, p. 299–312, 1980. Citado na página 16.
- ANDERSSON, O. *Experiment!: Planning, Implementing and Interpreting*. [S.l.]: John Wiley & Sons, 2012. Citado 3 vezes nas páginas 15, 19 e 20.
- ARTZNER, P.; DELBAEN, F.; EBER, J.-M.; HEATH, D. Coherent measures of risk. *Mathematical finance*, Wiley Online Library, v. 9, n. 3, p. 203–228, 1999. Citado na página 48.
- BABAEI, S.; SEPEHRI, M. M.; BABAEI, E. Multi-objective portfolio optimization considering the dependence structure of asset returns. *European Journal of Operational Research*, Elsevier, v. 244, n. 2, p. 525–539, 2015. Citado 2 vezes nas páginas 4 e 6.
- BAIER, S. L.; JR, G. P. D.; TAMURA, R. Does opening a stock exchange increase economic growth? *Journal of International Money and Finance*, Elsevier, v. 23, n. 3, p. 311–331, 2004. Citado na página 11.
- BANSAL, P.; ROTH, K. Why companies go green: A model of ecological responsiveness. *Academy of management journal*, Academy of Management, v. 43, n. 4, p. 717–736, 2000. Citado na página 9.
- BARTOLOMEU, D. C. S. Otimização do planeamento das inspeções de aeronaves na tap manutenção e engenharia. 2016. Citado na página 38.
- BARUA, S.; SAHA, A. K. Traditional ratios vs. cash flow based ratios: Which one is better performance indicator? *Advances in Economics and Business*, v. 3, n. 6, p. 232–251, 2015. Citado 3 vezes nas páginas 1, 13 e 14.
- BERÁNEK, L.; KOLAŘÍK, K. Surface integrity analysis of duplex steel by design of experiment approach. *Procedia Engineering*, Elsevier, v. 69, p. 630–637, 2014. Citado na página 17.
- BERTLAND, D.; ASGHARIAN, H. The perfect portfolio. 2008. Citado na página 2.

- BERTRAND, J. W. M.; FRANSOO, J. C. Operations management research methodologies using quantitative modeling. *International Journal of Operations & Production Management*, MCB UP Ltd, v. 22, n. 2, p. 241–264, 2002. Citado na página 7.
- BETTER, M.; GLOVER, F. Selecting project portfolios by optimizing simulations. *The Engineering Economist*, Taylor & Francis, v. 51, n. 2, p. 81–97, 2006. Citado na página 48.
- BHALERAO, R. S.; OLLITRAULT, J.-Y.; PAL, S.; TEANEY, D. Principal component analysis of event-by-event fluctuations. *Physical review letters*, APS, v. 114, n. 15, p. 152301, 2015. Citado na página 44.
- BORGES, C. N.; BRUNS, R. E.; ALMEIDA, A. A.; SCARMINIO, I. S. Mixture–mixture design for the fingerprint optimization of chromatographic mobile phases and extraction solutions for camellia sinensis. *Analytica chimica acta*, Elsevier, v. 595, n. 1, p. 28–37, 2007. Citado na página 32.
- BOUROCHE, J.-M.; SAPORTA, G. Análise de dados. *Rio de Janeiro. Zaher Editores*, 1982. Citado na página 44.
- BRANDVIK, P. J.; DALING, P. S. Optimisation of oil spill dispersant composition by mixture design and response surface methods. *Chemometrics and intelligent laboratory systems*, Elsevier, v. 42, n. 1, p. 63–72, 1998. Citado na página 31.
- BRENTANI, C. Portfolio management in practice (essential capital markets). Butterworth-Heinemann, 2004. Citado na página 49.
- BRYMAN, A. *et al. Research methods and organization studies*. [S.l.]: Routledge, 2003. Citado na página 7.
- BUCKLEY, I.; SAUNDERS, D.; SECO, L. Portfolio optimization when asset returns have the gaussian mixture distribution. *European Journal of Operational Research*, Elsevier, v. 185, n. 3, p. 1434–1461, 2008. Citado na página 53.
- BUSSAB, W. d. O.; MIAZAKI, É.; ANDRADE, D. Introdução à análise de grupamentos: In: Anais do ix simpósio brasileiro de probabilidade e estatística. *São Paulo: IME-USP*, 1990. Citado na página 39.
- CAPITAL, C. K. *America's Greenest Companies 2014*. 2014. Disponível em: <<http://www.newsweek.com/green/americas-greenest-companies-2014>>. Citado 4 vezes nas páginas 1, 2, 8 e 65.
- CASS, D.; STIGLITZ, J. E. The structure of investor preferences and asset returns, and separability in portfolio allocation: A contribution to the pure theory of mutual funds. *Journal of Economic Theory*, Elsevier, v. 2, n. 2, p. 122–160, 1970. Citado 2 vezes nas páginas 2 e 50.
- CHAN, L.-Y. Optimal design for a linear log contrast model for experiments with mixtures. *Journal of statistical planning and inference*, Elsevier, v. 20, n. 1, p. 105–113, 1988. Citado na página 3.

- CHEN, E. J.; LI, M. Design of experiments for interpolation-based metamodels. *Simulation Modelling Practice and Theory*, Elsevier, v. 44, p. 14–25, 2014. Citado na página 17.
- CHIARAWONGSE, A.; KIATSUPAIBUL, S.; TIRAPAT, S.; ROY, B. V. Portfolio selection with qualitative input. *Journal of Banking & Finance*, Elsevier, v. 36, n. 2, p. 489–496, 2012. Citado na página 2.
- CLEARINGHOUSE, O. R. *RECOVERY ACT*. 2015. Disponível em: <<http://www.energy.gov/recovery-act>>. Citado na página 9.
- COOPER, R. G. Stage-gate systems: a new tool for managing new products. *Business horizons*, Elsevier, v. 33, n. 3, p. 44–54, 1990. Citado na página 8.
- CORNELL, J. *Experiments with mixtures*. 28. [S.l.]: New York: John Wiley & Sons, Inc, 2002. Citado 3 vezes nas páginas 22, 30 e 32.
- CORNELL, J. A. Process variables in the mixture problem for categorized components. *Journal of the American Statistical Association*, Taylor & Francis, v. 66, n. 333, p. 42–48, 1971. Citado na página 32.
- CORNELL, J. A.; RAMSEY, P. J. Modeling the component linear and nonlinear blending properties in a two-stage mixture experiment. *Nonlinear Analysis: Theory, Methods & Applications*, Elsevier, v. 30, n. 7, p. 4041–4050, 1997. Citado na página 32.
- COSTA, O.; PAIVA, A. Robust portfolio selection using linear-matrix inequalities. *Journal of Economic Dynamics and Control*, Elsevier, v. 26, n. 6, p. 889–909, 2002. Citado na página 2.
- CRESWELL, J. W.; KLASSEN, A. C.; CLARK, V. L. P.; SMITH, K. C. Best practices for mixed methods research in the health sciences. *Bethesda (Maryland): National Institutes of Health*, p. 2094–2103, 2011. Citado na página 7.
- DELARUE, E.; JONGHE, C. D.; BELMANS, R.; D’HAESELEER, W. Applying portfolio theory to the electricity sector: Energy versus power. *Energy Economics*, Elsevier, v. 33, n. 1, p. 12–23, 2011. Citado na página 48.
- DONI, M. V. Análise de cluster: métodos hierárquicos e de particionamento. *São Paulo: Universidade Presbiteriana Mackenzie, São Paulo*, 2004. Citado 2 vezes nas páginas 39 e 40.
- DUINEVELD, C.; SMILDE, A.; DOORNBOS, D. Designs for mixture and process variables applied in tablet formulations. *Analytica chimica acta*, Elsevier, v. 277, n. 2, p. 455–465, 1993. Citado na página 3.
- ERIKSSON, L.; JOHANSSON, E.; WIKSTRÖM, C. Mixture design—design generation, pls analysis, and model usage. *Chemometrics and Intelligent Laboratory Systems*, Elsevier, v. 43, n. 1, p. 1–24, 1998. Citado na página 3.
- FEDERAL, B. G. *LEI No 6.404, DE 15 DE DEZEMBRO DE 1976*. 1976. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/L6404compilada.htm>. Citado 3 vezes nas páginas 8, 12 e 13.

FISHER, R. The design of experiments 6th ed. *Olivier and Boyd, London, UK*, 1951. Citado na página 15.

GABRIEL, S. C.; BAKER, C. B. Concepts of business and financial risk. *American Journal of Agricultural Economics*, Oxford University Press, v. 62, n. 3, p. 560–564, 1980. Citado na página 13.

GARCÍA-OCCHOA, F.; SANTOS, V.; FRITSCH, A. Nutritional study of xanthomonas campestris in xanthan gum production by factorial design of experiments. *Enzyme and microbial technology*, Elsevier, v. 14, n. 12, p. 991–996, 1992. Citado na página 16.

GARCIA, M.; GIAMBIAGI, F. *Risco e Regulação: por que o Brasil enfrentou bem a crise financeira e como ela afetou a economia mundial*. [S.l.]: Rio de Janeiro: Elsevier, 2010. Citado 2 vezes nas páginas 11 e 12.

GARLAPPI, L.; UPPAL, R.; WANG, T. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies*, Soc Financial Studies, v. 20, n. 1, p. 41–81, 2007. Citado 2 vezes nas páginas 4 e 6.

GOOS, P.; JONES, B. *Optimal design of experiments: a case study approach*. [S.l.]: John Wiley & Sons, 2011. Citado 2 vezes nas páginas 16 e 17.

GRYZE, S. D.; LANGHANS, I.; VANDEBROEK, M. Using the correct intervals for prediction: A tutorial on tolerance intervals for ordinary least-squares regression. *Chemometrics and intelligent laboratory systems*, Elsevier, v. 87, n. 2, p. 147–154, 2007. Citado na página 85.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, Warwick & York, v. 24, n. 6, p. 417, 1933. Citado 2 vezes nas páginas 3 e 43.

JOHNSON, R.; WICHERN, D. *Applied Multivariate Analysis, 5/e*. [S.l.]: Prentice Hall, 2002. Citado na página 43.

JOLLIFFE, I. *Principal component analysis*. [S.l.]: Wiley Online Library, 2002. Citado na página 43.

JR, J. H. W. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, Taylor & Francis, v. 58, n. 301, p. 236–244, 1963. Citado 3 vezes nas páginas 3, 5 e 41.

JR, W. F. S.; BEVERLY, T. A. Generating linear and quadratic cox mixture models. *Journal of quality technology*, American Society for Quality, v. 29, n. 2, p. 211, 1997. Citado na página 38.

KANG, L.; JOSEPH, V. R.; BRENNEMAN, W. A. Design and modeling strategies for mixture-of-mixtures experiments. *Technometrics*, Taylor & Francis, v. 53, n. 2, p. 125–136, 2011. Citado na página 32.

KARHUNEN, J.; JOUTSENSALO, J. Generalizations of principal component analysis, optimization problems, and neural networks. *Neural Networks*, Elsevier, v. 8, n. 4, p. 549–562, 1995. Citado na página 44.

KNIGHTS, C. Corporate knights. *Corporate Knights*, n. 1, 2016. Citado na página 9.

- LAZIĆ, Ž. R. *Design and Analysis of Experiments: Section 2.3*. [S.l.]: Wiley Online Library, 2004. Citado 6 vezes nas páginas 17, 22, 24, 26, 33 e 34.
- LI, D.; NG, W.-L. Optimal dynamic portfolio selection: Multiperiod mean-variance formulation. *Mathematical Finance*, Wiley Online Library, v. 10, n. 3, p. 387–406, 2000. Citado 3 vezes nas páginas 4, 6 e 51.
- LIN, J.; ZHANG, A. Fault feature separation using wavelet-ica filter. *NDT & E International*, Elsevier, v. 38, n. 6, p. 421–427, 2005. Citado na página 44.
- LIU, W.; ZHOU, Z.; LIU, D.; XIAO, H. Estimation of portfolio efficiency via dea. *Omega*, Elsevier, v. 52, p. 107–118, 2015. Citado 2 vezes nas páginas 4 e 6.
- LONNI, A. A. S. G.; LONGHINI, R.; LOPES, G. C.; MELLO, J. C. P. de; SCARMINIO, I. S. Statistical mixture design selective extraction of compounds with antioxidant activity and total polyphenol content from trichilia catigua. *Analytica chimica acta*, Elsevier, v. 719, p. 57–60, 2012. Citado na página 31.
- LYON, T. P.; SHIMSHACK, J. P. Environmental disclosure: evidence from newsweek's green companies rankings. *Business & Society*, SAGE Publications, p. 0007650312439701, 2012. Citado 6 vezes nas páginas viii, 1, 8, 9, 10 e 59.
- MAESSCHALCK, R. D.; JOUAN-RIMBAUD, D.; MASSART, D. L. The mahalanobis distance. *Chemometrics and intelligent laboratory systems*, Elsevier, v. 50, n. 1, p. 1–18, 2000. Citado na página 41.
- MÅGE, I.; NÆS, T. Split-plot design for mixture experiments with process variables: A comparison of design strategies. *Chemometrics and intelligent laboratory systems*, Elsevier, v. 78, n. 1, p. 81–95, 2005. Citado na página 31.
- MAINARDI, F.; MAGGIONI, F.; PEZZOLA, D.; ZAVA, D.; ZANCHIN, G. Dexketoprofen trometamol in the acute treatment of migraine attack: a phase ii, randomized, double-blind, crossover, placebo-controlled, dose optimization study. *The Journal of Pain*, Elsevier, 2014. Citado na página 20.
- MANLY, B. F. *Multivariate statistical methods: a primer*. [S.l.]: CRC Press, 2004. Citado na página 40.
- MARKOWITZ, H. Portfolio selection*. *The journal of finance*, Wiley Online Library, v. 7, n. 1, p. 77–91, 1952. Citado 9 vezes nas páginas 1, 2, 3, 4, 6, 48, 49, 50 e 51.
- MARKOWITZ, H. Portfolio selection: efficient diversification of investments. *New York*, 1959. Citado 6 vezes nas páginas 2, 4, 5, 6, 49 e 50.
- MAYERS, D. Portfolio theory, job choice and the equilibrium structure of expected wages. *Journal of Financial Economics*, Elsevier, v. 1, n. 1, p. 23–42, 1974. Citado 4 vezes nas páginas 2, 4, 6 e 48.
- MCCONNELL, J. J.; DYBEVIK, H. J.; HAUSHALTER, D.; LIE, E. A survey of evidence on domestic and international stock exchange listings with implications for markets and managers. *Pacific-Basin Finance Journal*, Elsevier, v. 4, n. 4, p. 347–376, 1996. Citado 2 vezes nas páginas 11 e 12.

- MENDES, R.; PAIVA, A.; PERUCHI, R. S.; BALESTRASSI, P. P.; LEME, R. C.; SILVA, M. Multiobjective portfolio optimization of arma-garch time series based on experimental designs. *Computers & Operations Research*, Elsevier, v. 66, p. 434–444, 2016. Citado 2 vezes nas páginas 3 e 53.
- MONTGOMERY, D. C. *Design and analysis of experiments*. [S.l.]: John Wiley & Sons, 2008. Citado 4 vezes nas páginas 16, 22, 32 e 45.
- MONTGOMERY, D. C.; MONTGOMERY, D. C.; MONTGOMERY, D. C. *Design and analysis of experiments*. [S.l.]: Wiley New York, 1997. Citado 3 vezes nas páginas viii, 17 e 18.
- MYER, R.; MONTGOMERY, D. C. Response surface methodology: process and product optimization using designed experiment. *John Wiley and Sons, New York*, p. 343–350, 2002. Citado na página 22.
- MYERS RAYMOND H., M. D. C. A.-C. C. M. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. [S.l.]: John Wiley-New York, 2009. Citado na página 5.
- NASDAQ, I. *National Association of Securities Dealers Automatic Quotation*. 2015. Disponível em: <<http://www.nasdaq.com>>. Citado na página 63.
- NEWBOLD, P.; CARLSON, W.; THORNE, B. *Statistics for Business and Economics*. [S.l.]: Pearson Higher Ed, 2012. Citado 3 vezes nas páginas 35, 36 e 42.
- NGUN, B. K.; MOHAMAD, H.; KATSUMATA, K.-i.; OKADA, K.; AHMAD, Z. A. Using design of mixture experiments to optimize triaxial ceramic tile compositions incorporating cambodian clays. *Applied Clay Science*, Elsevier, v. 87, p. 97–107, 2014. Citado 2 vezes nas páginas 2 e 31.
- OLIVEIRA, F. A. d.; PAIVA, A. P. D.; LIMA, J. W. M.; BALESTRASSI, P. P.; MENDES, R. R. A. Portfolio optimization using mixture design of experiments: Scheduling trades within electricity markets. *Energy Economics*, Elsevier, v. 33, n. 1, p. 24–32, 2011. Citado 3 vezes nas páginas 3, 51 e 53.
- PAIVA, A. Metodologia de superfície de resposta e análise de componentes principais em otimização de processos de manufatura com múltiplas respostas correlacionadas. *Metodologia de superfície de resposta e análise de componentes principais em otimização de processos de manufatura com múltiplas respostas correlacionadas*, 2006. Citado 2 vezes nas páginas 43 e 44.
- PAPANDREA, P.; PAIVA, A. Design of experiment sheets of green companies. *Mendeley Data*, Mendeley, v2, n. DOI: 10.17632/9kvzx7k5yh.2, p. <http://dx.doi.org/10.17632/9kvzx7k5yh.2>, 2016. Citado 4 vezes nas páginas 81, 94, 99 e 106.
- PAPANDREA, P.; PAIVA, A. Monthly quotation of twenty green companies. *Mendeley Data*, Mendeley, v2, n. DOI: 10.17632/ccpzrpx4xk.2, p. <http://dx.doi.org/10.17632/ccpzrpx4xk.2>, 2016. Citado na página 70.
- PAPANDREA, P.; PAIVA, A. The pca applied into the newsweek green companies ranking of 2014. *Mendeley Data*, Mendeley, v2, n. DOI: 10.17632/cpvr7gdy9r.2, p. <http://dx.doi.org/10.17632/cpvr7gdy9r.2>, 2016. Citado na página 66.

- PAPANDREA, P.; PAIVA, A. Pca applied on calculated financial indicators and quotations analysis. *Mendeley Data*, Mendeley, v2, n. DOI: 10.17632/s2kny3549v.2, p. <http://dx.doi.org/10.17632/s2kny3549v.2>, 2016. Citado na página 72.
- PAPANDREA, P.; PAIVA, A. Replication of the monthly quotation. *Mendeley Data*, Mendeley, v2, n. DOI: 10.17632/vjsmh4b7zt.2, p. <http://dx.doi.org/10.17632/vjsmh4b7zt.2>, 2016. Citado na página 74.
- PAPANDREA PEDRO; PAIVA, A. L. R. Analysis of doe applied on green companies portfolios. *Mendeley Data*, Mendeley, v1, n. DOI: 10.17632/pc4m3hf8sh.1, p. <http://dx.doi.org/10.17632/pc4m3hf8sh.1>, 2016. Citado 5 vezes nas páginas 106, 153, 155, 160 e 162.
- PÁZMAN, A.; MÜLLER, W. G. Optimal design of experiments subject to correlated errors. *Statistics & probability letters*, Elsevier, v. 52, n. 1, p. 29–34, 2001. Citado na página 17.
- PIEPEL, G. F. Modeling methods for mixture-of-mixtures experiments applied to a tablet formulation problem. *Pharmaceutical development and technology*, Informa UK Ltd UK, v. 4, n. 4, p. 593–606, 1999. Citado na página 32.
- PIERLOT, C.; PAWLOWSKI, L.; BIGAN, M.; CHAGNON, P. Design of experiments in thermal spraying: A review. *Surface and Coatings technology*, Elsevier, v. 202, n. 18, p. 4483–4490, 2008. Citado na página 17.
- PODANI, J.; SCHMERA, D. On dendrogram-based measures of functional diversity. *Oikos*, Wiley Online Library, v. 115, n. 1, p. 179–185, 2006. Citado 2 vezes nas páginas 4 e 40.
- RAO, S. S.; RAO, S. *Engineering optimization: theory and practice*. [S.l.]: John Wiley & Sons, 2009. Citado 2 vezes nas páginas 44 e 47.
- RENCHE, A. C. *Methods of multivariate analysis*. [S.l.]: John Wiley & Sons, 2003. Citado na página 43.
- RODRIGUES, S. A.; HALMEMAN, J. R. Matriz energética brasileira: uma reflexão sobre a situação atual e possíveis riscos de “apagões”. *Bioenergia em Revista: Diálogos (ISSN: 2236-9171)*, v. 2, n. 1, p. 11–26, 2012. Citado na página 2.
- SAHA, M.; DARNTON, G. Green companies or green conpanies: Are companies really green, or are they pretending to be? *Business and Society Review*, Wiley Online Library, v. 110, n. 2, p. 117–157, 2005. Citado na página 8.
- SANGER, G. C.; MCCONNELL, J. J. Stock exchange listings, firm value, and security market efficiency: The impact of nasdaq. *Journal of Financial and Quantitative Analysis*, Cambridge Univ Press, v. 21, n. 01, p. 1–25, 1986. Citado na página 63.
- SCHEFFÉ, H. Experiments with mixtures. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 344–360, 1958. Citado 5 vezes nas páginas 3, 23, 26, 30 e 31.
- SCHEFFÉ, H. The simplex-centroid design for experiments with mixtures. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 235–263, 1963. Citado 2 vezes nas páginas 3 e 31.

- SECURITIES, U. Securities exchange act of 1934. *Retrieved May*, v. 23, p. 2012, 1934. Citado 2 vezes nas páginas 13 e 63.
- SELLIER, P. Design and analysis of quantitative genetic experiments: A review with particular reference to pigs. *Livestock Production Science*, Elsevier, v. 7, n. 6, p. 539–554, 1980. Citado na página 16.
- SERIFSOY, B. *Essays on stock exchange efficiency, business models, and governance*. Tese (Doutorado) — Ph. D. dissertation, Johann Wolfgang Goeth-University, Frankfurt/Main, 2006. Citado na página 10.
- SHALIT, H.; YITZHAKI, S. Mean-gini, portfolio theory, and the pricing of risky assets. *The Journal of Finance*, Wiley Online Library, v. 39, n. 5, p. 1449–1468, 1984. Citado na página 51.
- SMITH, W. F. *Experimental design for formulation*. [S.l.]: Siam, 2005. Citado na página 18.
- SWINIARSKI, R. W.; SKOWRON, A. Rough set methods in feature selection and recognition. *Pattern recognition letters*, Elsevier, v. 24, n. 6, p. 833–849, 2003. Citado na página 44.
- TEHRANI, M. M.; TARZJAN, M. S. *et al.* Optimization of low-cholesterol-low-fat mayonnaise formulation: Effect of using soy milk and some stabilizer by a mixture design approach. *Food Hydrocolloids*, v. 28, 2012. Citado na página 31.
- TÍER, M. *Os Segredos de George Soros & Warren Buffett*. [S.l.]: Rio de Janeiro: Elsevier, 2005. Citado 2 vezes nas páginas 11 e 48.
- TRIPPI, R. R. A decision support system for real estate investment portfolio management. *Information & Management*, Elsevier, v. 16, n. 1, p. 47–54, 1989. Citado 2 vezes nas páginas 48 e 50.
- VIEIRA, A. F. d. C.; BELLO, L. H. A. D. Experimentos com mistura para otimização de processos: uma aplicação com respostas não normais. *Pesquisa Operacional*, SciELO Brasil, v. 26, n. 3, p. 605–623, 2006. Citado na página 3.
- VILLARS, D. S. *et al.* Statistical design and analysis of experiments for development research. *Statistical design and analysis of experiments for development research.*, Wm. C. Brown Company, Dubuque, Iowa., 1951. Citado na página 15.
- WANG, X.; PALIWAL, K. K. Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. *Pattern recognition*, Elsevier, v. 36, n. 10, p. 2429–2439, 2003. Citado na página 44.
- WASSEF, A. Some aspects of the design, analysis and presentation of photogrammetric experiments. *Photogrammetria*, Elsevier, v. 8, p. 146–167, 1952. Citado na página 16.
- YU, M.; WANG, S.-Y.; LAI, K. K.; CHAO, X. Multiperiod portfolio selection on a minimax rule. *Dynamics of continuous discrete and impulsive systems series B*, WATAM PRESS, v. 12, n. 4, p. 565, 2005. Citado na página 51.

ZAIANE, O. R.; FOSS, A.; LEE, C.-H.; WANG, W. On data clustering analysis: Scalability, constraints, and validation. In: *Advances in knowledge discovery and data mining*. [S.l.]: Springer, 2002. p. 28–39. Citado na página 39.

ZHANG, C.; WONG, W. K. Optimal designs for mixture models with amount constraints. *Statistics & Probability Letters*, Elsevier, v. 83, n. 1, p. 196–202, 2013. Citado na página 3.

ZHANG, W.-G.; LIU, Y.-J.; XU, W.-J. A possibilistic mean-semivariance-entropy model for multi-period portfolio selection with transaction costs. *European Journal of Operational Research*, Elsevier, v. 222, n. 2, p. 341–349, 2012. Citado na página 51.

ZIO, M. D.; GUARNERA, U.; ROCCHI, R. A mixture of mixture models for a classification problem: The unity measure error. *Computational statistics & data analysis*, Elsevier, v. 51, n. 5, p. 2573–2585, 2007. Citado na página 32.

Apêndices

APÊNDICE A – ANOVA E GRÁFICOS COM ANÁLISE DE CADA CLUSTERS COM VARIÁVEIS DE PROCESSO

Quadro A.1 – ANOVA de RP Cluster 1 (-1)

Resposta 1 RP					
ANOVA para Modelo de Mistura Linear					
Componente da Mistura Codificado em Pseudo Componente					
Tabela Análise de Variância [Soma parcial dos quadrados]					
	Soma dos		Média dos	F	p-value
Modelos	Quadrados	df	Quadrados	Valor	Prob > F
Testados	0,000449504	3	0,000149835	4345,704802	< 0,0001*
Mistura Linear	0,000449504	3	0,000149835	4345,704802	< 0,0001
Residual	4,82703E-07	14	3,44788E-08		
Falta de Ajuste	0,000000	5	0,000000	0,000000	1,0000**
Erro Puro	4,82703E-07	9	5,36336E-08		
Corr Total	0,000449986	17			
(*) significante; (**) não significante					

O valor F de 4345,70 do Quadro A.1 implica que o modelo é significativo. Existe apenas a chance de 0,01% que o F-Valor desta magnitude poderia ocorrer devido ao ruído. Os valores de “Prob> F ”inferiores a 0,0500 indicam que os termos do modelo são significativos. Neste caso os componentes da mistura Linear são termos significativos do modelo. Valores superiores a 0,1000 indicam os termos do modelo não são significativos. Se há muitos termos não significativos no modelo (não contando os necessários para suportar a hierarquia), a redução do modelo pode melhorar as funções. A falta de ajuste do Valor-F (“F-value ”) de 0,00 implica que a falta de ajuste não é significativa em relação ao erro puro. Há uma chance de 100,00% que uma falta de ajuste F-value tão grande poderia ocorrer devido ao ruído. A falta não significativa de ajuste significa que o modelo é bom. As Figuras A.1, A.2, A.3 e A.4 demonstram as análises de forma gráfica.

O valor F de 841,19 do Quadro A.2 implica que o modelo é significativo. Existe uma chance de apenas 0,01% que um F-Valor sendo grande poderia ocorrer devido ao ruído. Os valores de “Prob> F ”inferior a 0,0500 indica que os termos do modelo são significativos. Neste caso componentes Lineares da mistura AD, BC são termos significativos para o modelo. A falta de ajuste F-value de 4,68 implica há uma chance de 5,87% que uma falta de ajuste por F-value grande poderia ocorrer devido ao ruído, falta de ajuste é ruim. Esta probabilidade relativamente baixa (< 10%) é preocupante, porém não excludente, mas o modelo se demonstra aplicável para otimização multiobjetivo. As Figuras A.5, A.6, A.7 e A.8 tem as análises gráficas.

O valor F de 155,79 do Quadro A.3 implica que o modelo é significativo. Existe

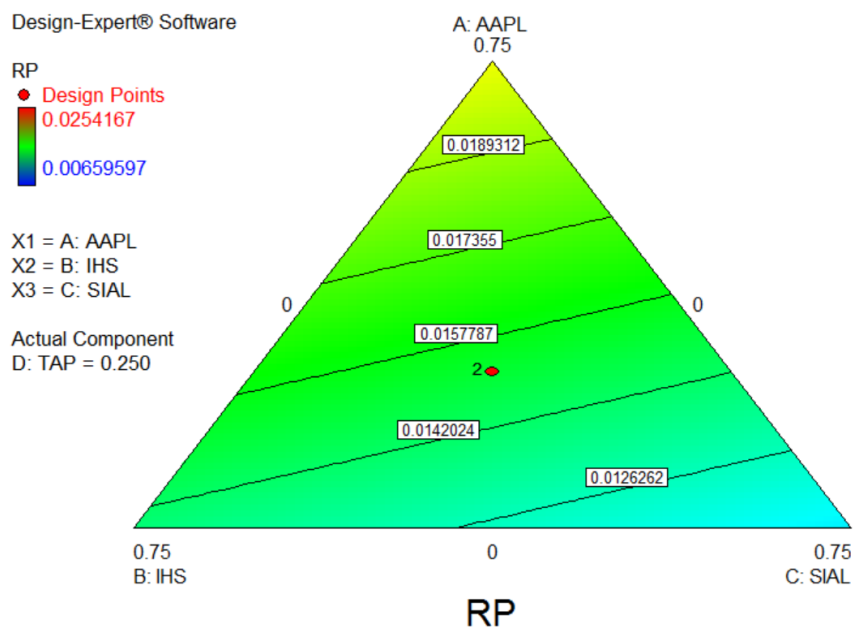


Figura A.1 – Análise de RP do Cluster 1(-1) Contorno

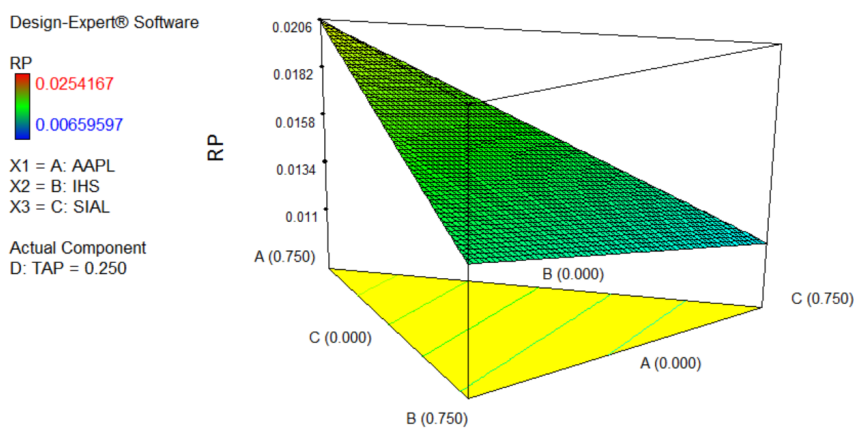


Figura A.2 – Análise de RP do Cluster 1(-1) 3D

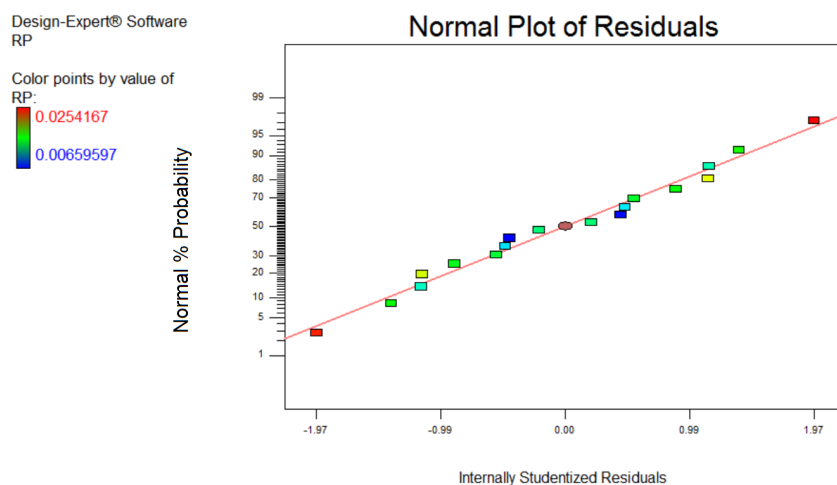


Figura A.3 – Análise de RP do Cluster 1(-1) Normalidade

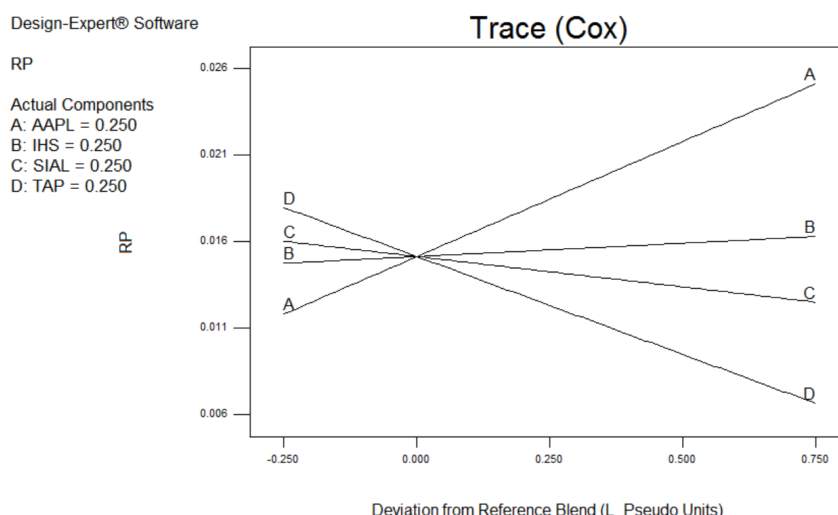


Figura A.4 – Análise de RP do Cluster 1(-1) Cox

Quadro A.2 – ANOVA de SP Cluster 1(-1)

Resposta 2 SP					
ANOVA para Modelo de Mistura Quadrático (Distorcido)					
Componente da Mistura Codificado em Pseudo Componente					
Tabela Análise de Variância [Soma parcial dos quadrados]					
Modelos Testados	Soma dos Quadrados	df	Média dos Quadrados	F Valor	p-value Prob > F
Mistura Linear	0,003471	7	0,000496	841,1878	< 0,0001*
AB	2,96E-07	1	2,96E-07	0,501524	0,495
AC	4,48E-07	1	4,48E-07	0,760364	0,4037
AD	0,000228	1	0,000228	386,1135	< 0,0001
BC	0,000271	1	0,000271	459,2641	< 0,0001
BD	0,000000	0			
CD	0,000000	0			
Residual	5,9E-06	10	5,9E-07		
Falta de Ajuste	2,02E-06	1	2,02E-06	4,682636	0,0587**
Erro Puro	3,88E-06	9	4,31E-07		
Corr Total	0,003477	17			
(*) significante; (**) não significante					

apenas a chance de 0,01% que o F-Valor desta magnitude poderia ocorrer devido ao ruído. Os valores de “Prob> F ”inferiores a 0,0500 indicam que os termos do modelo são significativos. Neste caso os componentes da mistura Linear são termos significativos do modelo. Valores superiores a 0,1000 indicam os termos do modelo não são significativos. Se há muitos termos não significativos no modelo (não contando os necessários para suportar a hierarquia), a redução do modelo pode melhorar as funções. A falta de ajuste do Valor-F (“F-value ”) de 0,00 implica que a falta de ajuste não é significativa em relação ao erro puro. Há uma chance de 100,00% que uma falta de ajuste F-value tão grande poderia ocorrer devido ao ruído. A falta não significativa de ajuste significa que o modelo é bom. As Figuras A.9, A.10, A.11 e A.12 demonstram as análises de forma gráfica.

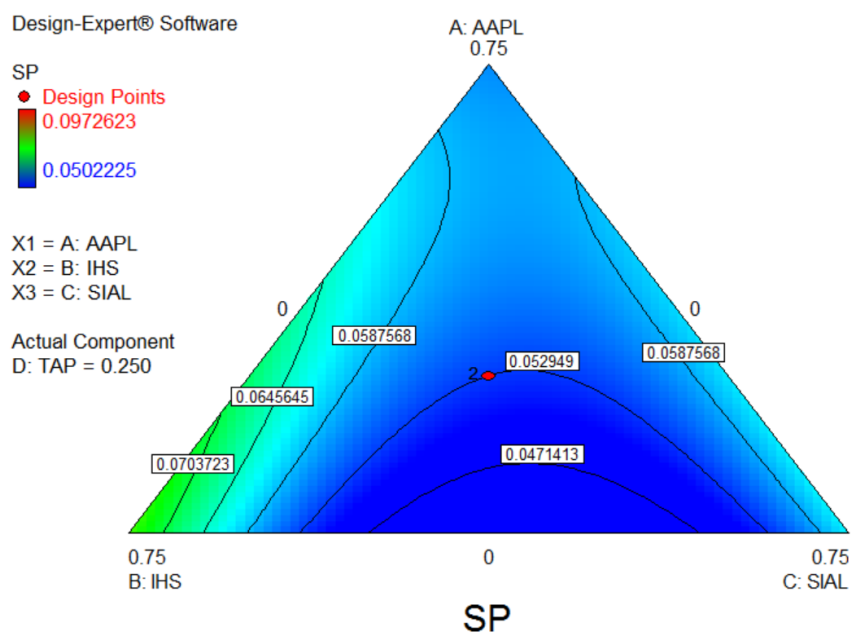


Figura A.5 – Análise de SP do Cluster 1(-1) Contorno

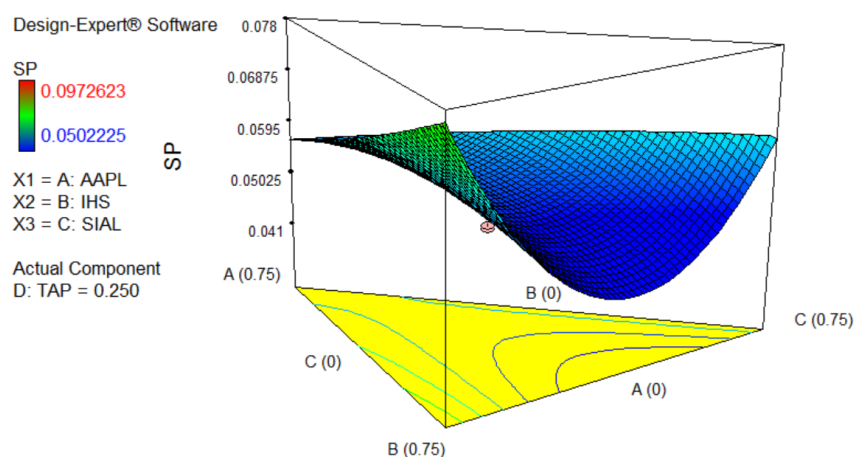


Figura A.6 – Análise de SP do Cluster 1(-1) 3D

Quadro A.3 – ANOVA de RP Cluster 1(+1)

Resposta 1 RP					
ANOVA para Modelo de Mistura Linear					
Componente da Mistura Codificado em Pseudo Componente					
Tabela Análise de Variância [Soma parcial dos quadrados]					
	Soma dos	Média dos	F	p-value	
Modelos	Quadrados	df	Quadrados	Valor	Prob > F
Testados	0,00094333	5	0,000188666	155,7915229	< 0,0001*
Mistura Linear	0,00094333	5	0,000188666	155,7915229	< 0,0001
Residual	2,42203E-05	20	1,21102E-06		
Falta de Ajuste	0,000000	7	0,000000	0,000000	1**
Erro Puro	2,42203E-05	13	1,8631E-06		
Corr Total	0,00096755	25			
(*) significante; (**) não significante					

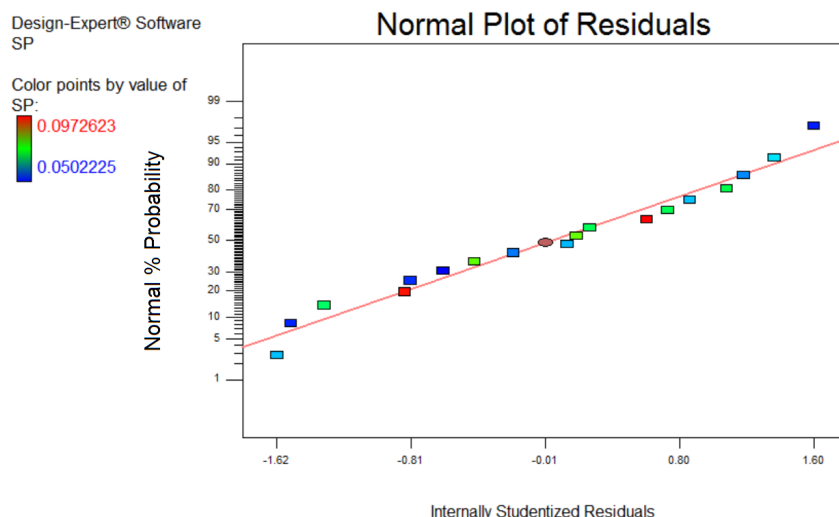


Figura A.7 – Análise de SP do Cluster 1(-1) Normalidade

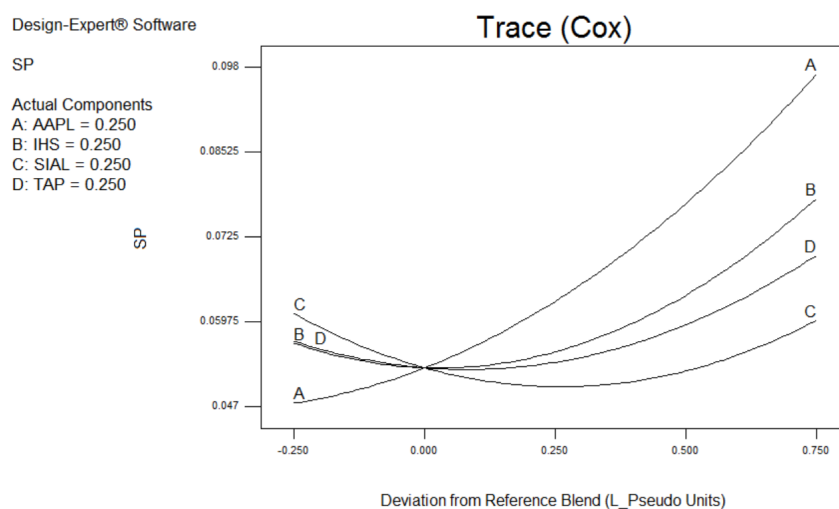


Figura A.8 – Análise de SP do Cluster 1(-1)Cox

O valor F de 1173,75 do Quadro A.4 implica que o modelo é significativo. Neste caso componentes da mistura Linear, AB, AC, AD, AE, AF, BC são termos significativos do modelo. Valores superiores a 0,1000 indicam os termos do modelo não são significativas. A falta de ajuste F-value de 18,65 implica a falta de ajuste é significativo. Há apenas uma chance de 0,08% que uma esta poderia ocorrer devido ao ruído. A.13, A.14, A.15 e A.16 demonstram graficamente as análises.

O valor F de 747,73 do Quadro A.5 implica que o modelo é significativo. Existe apenas a chance de 0,01% que o F-Valor desta magnitude poderia ocorrer devido ao ruído. Os valores de “Prob> F ”inferiores a 0,0500 indicam que os termos do modelo são significativos. Neste caso os componentes da mistura Linear são termos significativos do modelo. Valores superiores a 0,1000 indicam os termos do modelo não são significativos. Se há muitos termos não significativos no modelo (não contando os necessários para suportar a hierarquia), a redução do modelo pode melhorar as funções. A falta de ajuste do Valor-F

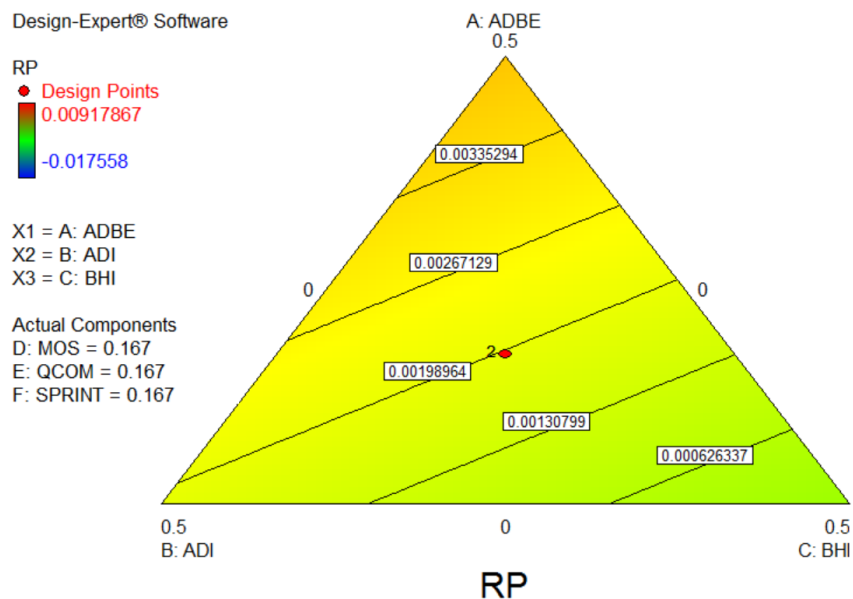


Figura A.9 – Análise de RP do Cluster 1(+1) Contorno

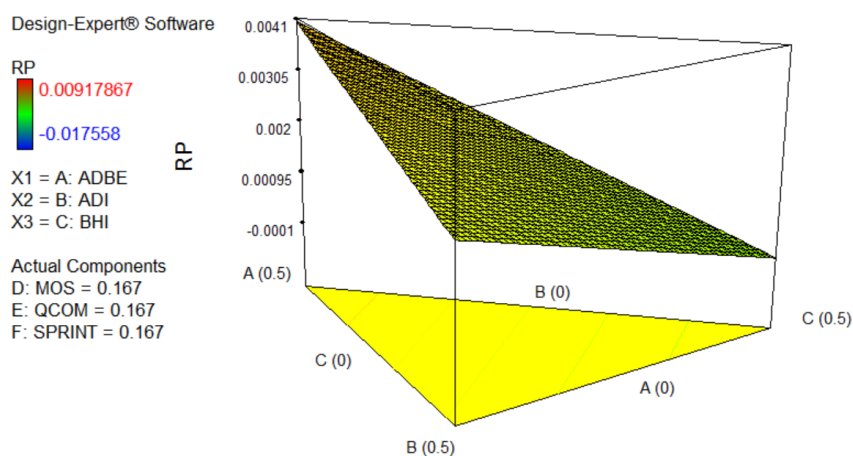


Figura A.10 – Análise de RP do Cluster 1(+1) 3D

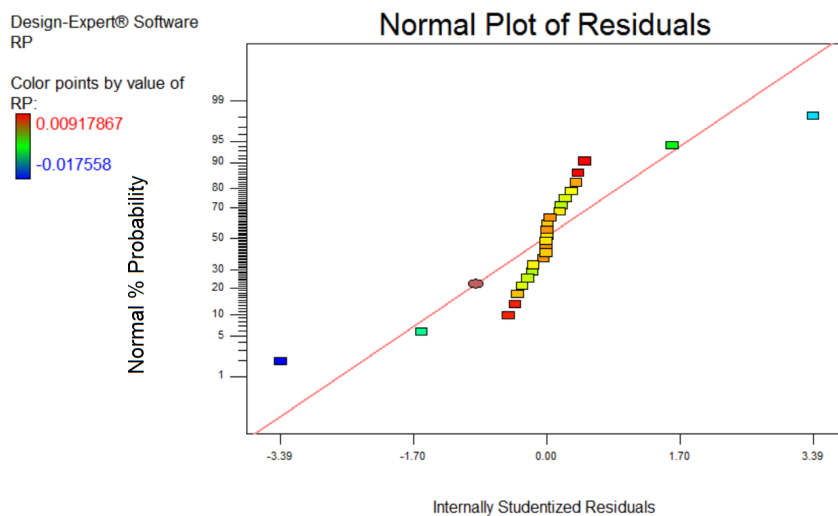


Figura A.11 – Análise de RP do Cluster 1(+1) Normalidade

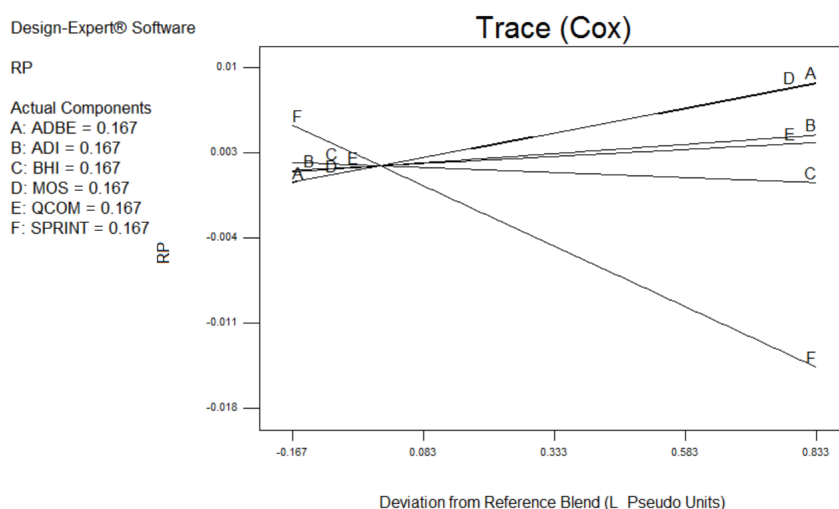


Figura A.12 – Análise de RP do Cluster 1(+1) Cox

Quadro A.4 – ANOVA de SP Cluster 1(+1)

Resposta 2 SP					
ANOVA para Modelo de Mistura Quadrático (Distorcido)					
Componente da Mistura Codificado em Pseudo Componente					
Tabela Análise de Variância [Soma parcial dos quadrados]					
Modelos	Soma dos Quadrados	df	Média dos Quadrados	F Valor	p-value Prob > F
Testados	0,016353637	11	0,001486694	1173,754144	< 0,0001*
Mistura Linear	0,011038608	5	0,002207722	1743,009616	< 0,0001
AB	0,000496621	1	0,000496621	392,0855058	< 0,0001
AC	0,000431812	1	0,000431812	340,9184042	< 0,0001
AD	0,000507714	1	0,000507714	400,8435278	< 0,0001
AE	0,000510193	1	0,000510193	402,8003207	< 0,0001
AF	0,000904649	1	0,000904649	714,225497	< 0,0001
BC	0,001803397	1	0,001803397	1423,792687	< 0,0001
Residual	1,77326E-05	14	1,26661E-06		
Falta de Ajuste	1,0448E-05	1	1,0448E-05	18,64518318	0,0008*
Erro Puro	7,28464E-06	13	5,60357E-07		
Corr Total	0,016371369	25			
(*) significante; (**) não significante					

("F-value ") de 0,00 implica que a falta de ajuste não é significativa em relação ao erro puro. Há uma chance de 100,00% que uma falta de ajuste F-value tão grande poderia ocorrer devido ao ruído. A falta não significativa de ajuste significa que o modelo é bom. As Figuras A.17 e A.18 demonstram as análises de forma gráfica.

O valor F de 362,73 do Quadro A.6 implica que o modelo é significativo. Neste caso, os componentes lineares da mistura, AB são termos significativos do modelo. A falta de ajuste do F-value de 1,00 implica que a falta de ajuste não é significativa em relação ao erro puro. Há uma chance de 43,03% que uma falta de ajuste poderia ocorrer devido ao ruído. Falta não significativa de ajuste é bom. As Figuras A.19 e A.20 demonstram as

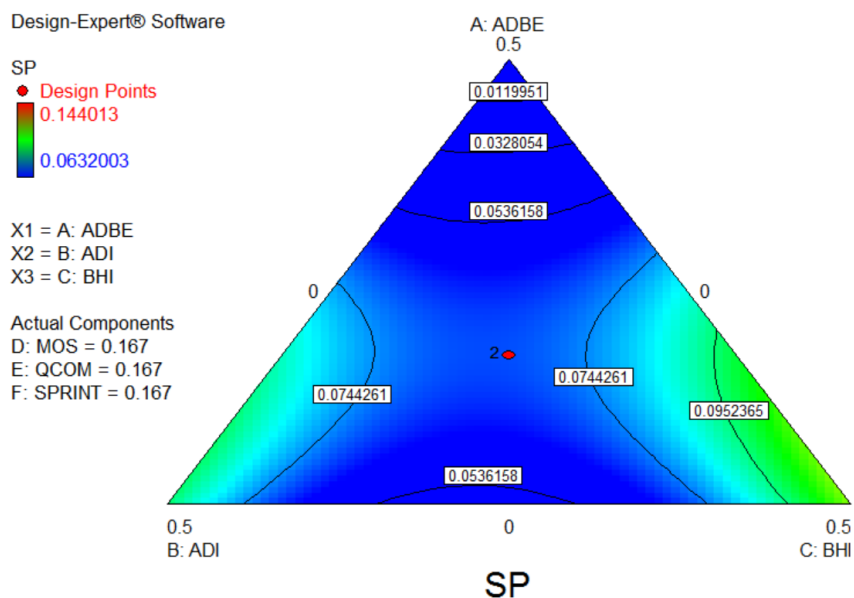


Figura A.13 – Análise de SP do Cluster 1(+1) Contorno

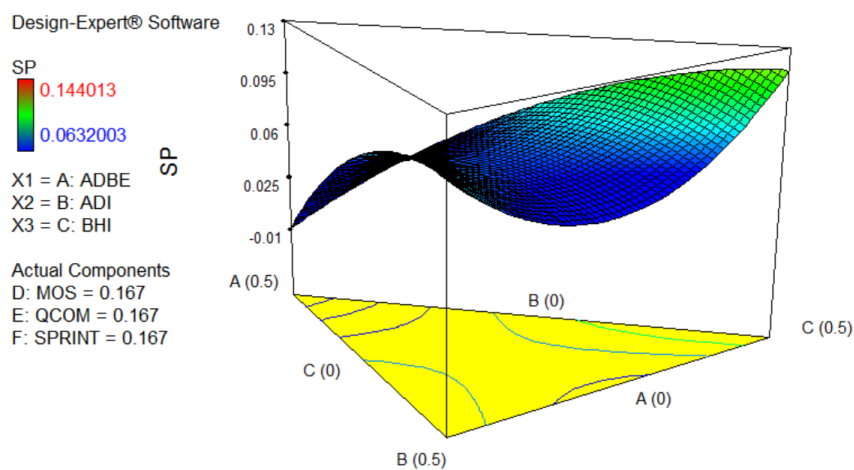


Figura A.14 – Análise de SP do Cluster 1(+1) 3D

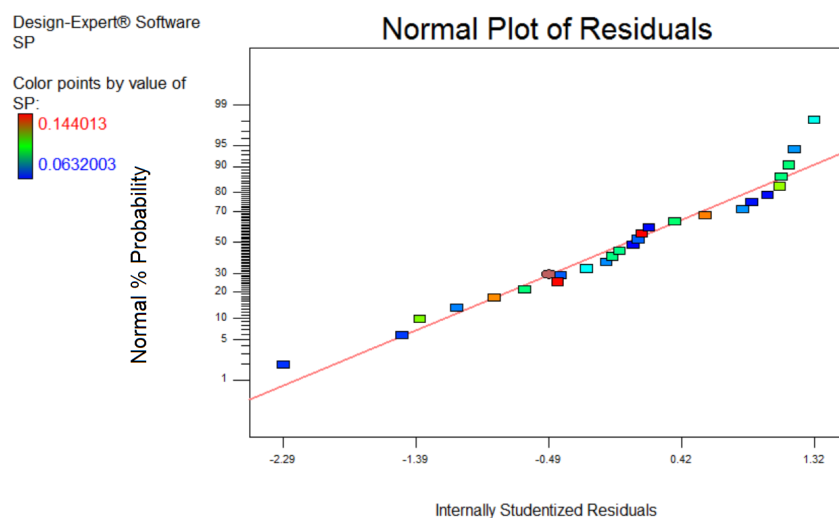


Figura A.15 – Análise de SP do Cluster 1(+1) Normalidade

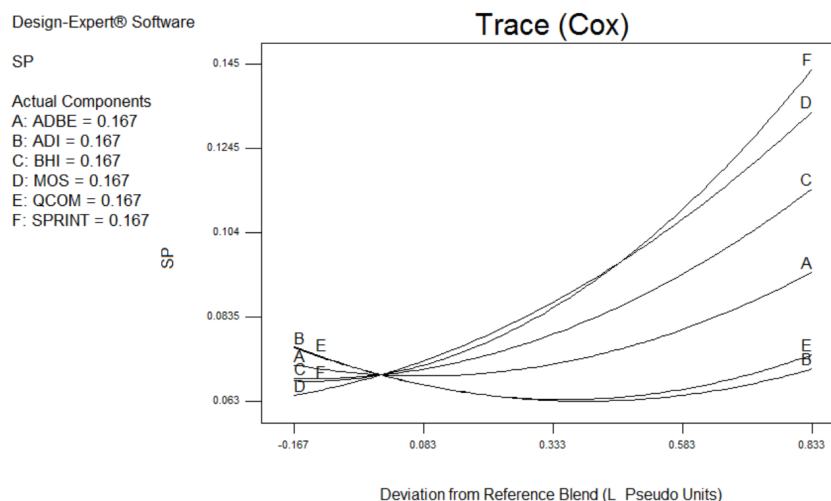


Figura A.16 – Análise de SP do Cluster 1(+1) Cox

Quadro A.5 – ANOVA de RP Cluster 2(+1)

Resposta 1 RP					
ANOVA para Modelo de Mistura Linear					
Componente da Mistura Codificado em Pseudo Componente					
Tabela Análise de Variância [Soma parcial dos quadrados]					
Modelos	Soma dos Quadrados	df	Média dos Quadrados	F Valor	p-value Prob > F
Testados	3,47477E-05	1	3,47477E-05	747,7319279	< 0,0001*
Mistura Linear	3,47477E-05	1	3,47477E-05	747,7319279	< 0,0001
Residual	3,71766E-07	8	4,64708E-08		
Falta de Ajuste	0,000000	3	0,000000	0,000000	1,000000**
Erro Puro	3,71766E-07	5	7,43533E-08		
Corr Total	3,51195E-05	9			
(*) significativa; (**) não significativa					

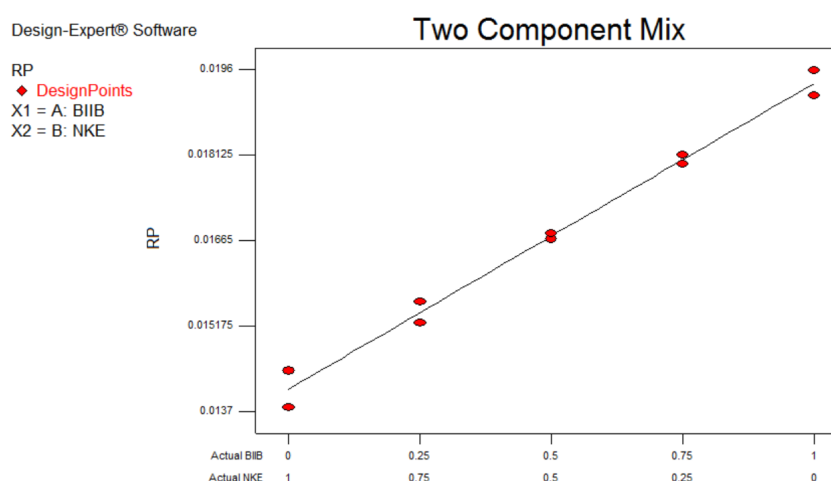


Figura A.17 – Análise de RP do Cluster 2(+1) Contorno

análises.

O valor F de 543,05 do Quadro A.7 implica que o modelo é significativo. Existe

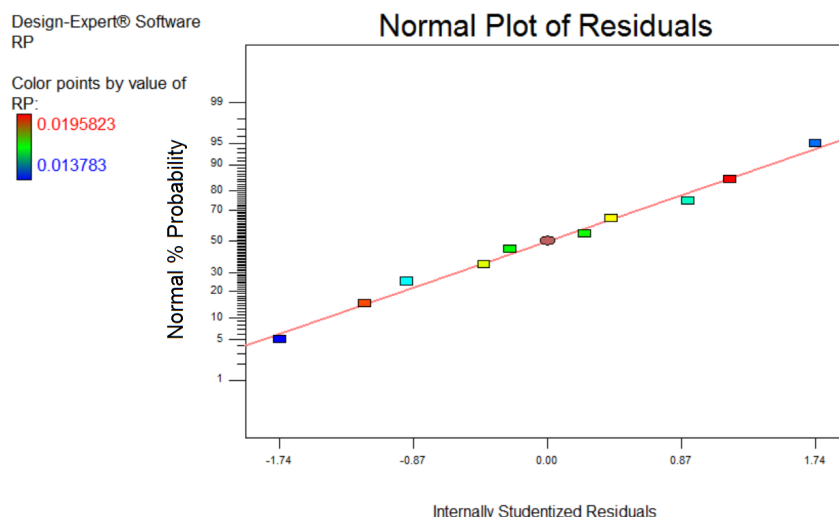


Figura A.18 – Análise de RP do Cluster 2(+1) Normalidade

Quadro A.6 – ANOVA de SP Cluster 2(+1)

Resposta 2 SP					
ANOVA para Modelo de Mistura Quadrático					
Componente da Mistura Codificado em Pseudo Componente					
Tabela Análise de Variância [Soma parcial dos quadrados]					
	Soma dos	Média dos	F	p-value	
Modelos	Quadrados	df	Quadrados	Valor	Prob > F
Testados	0,001240149	2	0,000620074	362,7315834	< 0,0001*
Mistura Linear	0,000878445	1	0,000878445	513,8735434	< 0,0001
AB	0,000361704	1	0,000361704	211,5896233	< 0,0001
Residual	1,19662E-05	7	1,70946E-06		
Falta de Ajuste	3,42573E-06	2	1,71287E-06	1,002793813	0,4303**
Erro Puro	8,54047E-06	5	1,70809E-06		
Corr Total	0,001252115	9			
(*) significante; (**) não significante					

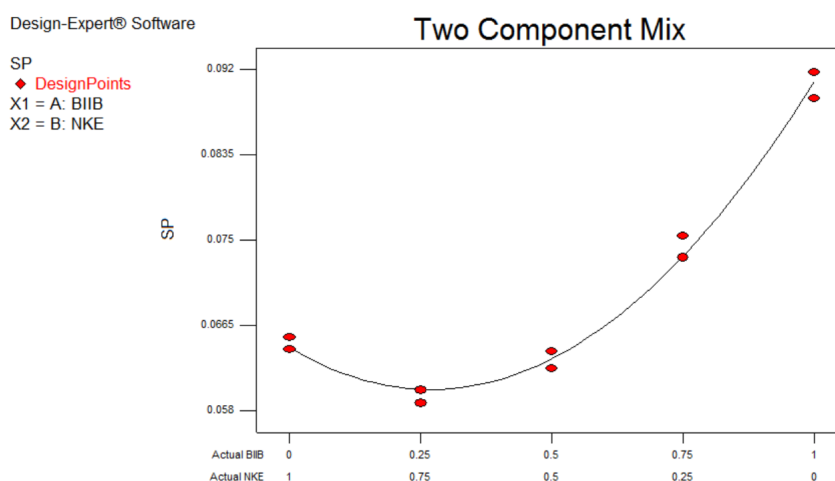


Figura A.19 – Análises de SP Cluster2(+1) Contorno

apenas a chance de 0,01% que o F-Valor desta magnitude poderia ocorrer devido ao

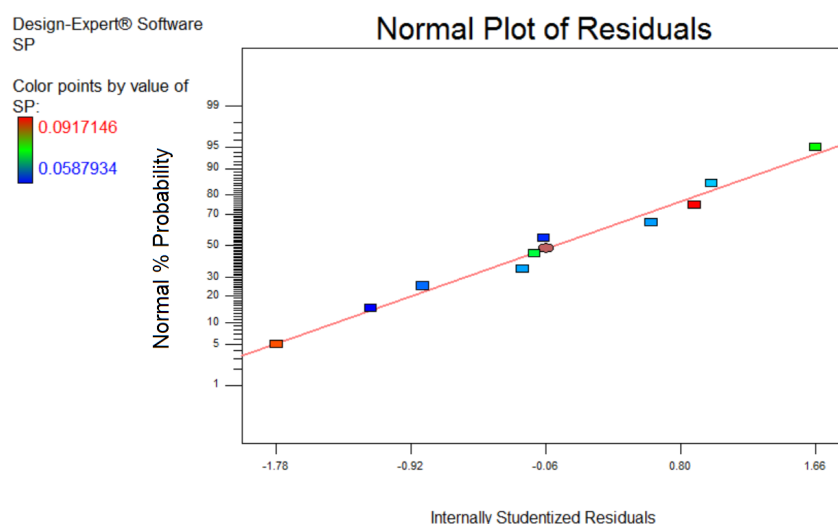


Figura A.20 – Análises de SP Cluster2(+1) Normalidade

ruído. Os valores de “Prob> F ”inferiores a 0,0500 indicam que os termos do modelo são significativos. Neste caso os componentes da mistura Linear são termos significativos do modelo. Valores superiores a 0,1000 indicam os termos do modelo não são significativos. Se há muitos termos não significativos no modelo (não contando os necessários para suportar a hierarquia), a redução do modelo pode melhorar as funções. A falta de ajuste do Valor-F (“F-value ”) de 0,00 implica que a falta de ajuste não é significativa em relação ao erro puro. Há uma chance de 100,00% que uma falta de ajuste F-value tão grande poderia ocorrer devido ao ruído. A falta não significativa de ajuste significa que o modelo é bom. As Figuras A.21, A.22, A.23 e A.24 demonstram as análises de forma gráfica.

Quadro A.7 – ANOVA de RP Cluster 3(-1)

Resposta 1 RP					
ANOVA para Modelo de Mistura Linear					
Componente da Mistura Codificado em Pseudo Componente					
Tabela Análise de Variância [Soma parcial dos quadrados]					
Modelos	Soma dos Quadrados	df	Média dos Quadrados	F Valor	p-value Prob > F
Testados	0,000139907	2	6,99533E-05	543,0546442	< 0,0001*
Mistura Linear	0,000139907	2	6,99533E-05	543,0546442	< 0,0001
Residual	1,41696E-06	11	1,28815E-07		
Falta de Ajuste	0,000000	4	0,000000	0,000000	1,000000**
Erro Puro	1,41696E-06	7	2,02423E-07		
Corr Total	0,000141324	13			
(*) significante; (**) não significante					

O valor F de 799,27 do Quadro A.8 implica que o modelo é significativo. A falta de ajuste F-value de 0,02 implica que a falta de ajuste não é significativa em relação ao erro puro erro. Há uma chance de 89,21% que uma falta de ajuste poderia ocorrer devido ao ruído. Falta não significativa de ajuste é bom. As Figuras A.25, A.26, A.27 e A.28 têm a demonstração gráfica das análises.

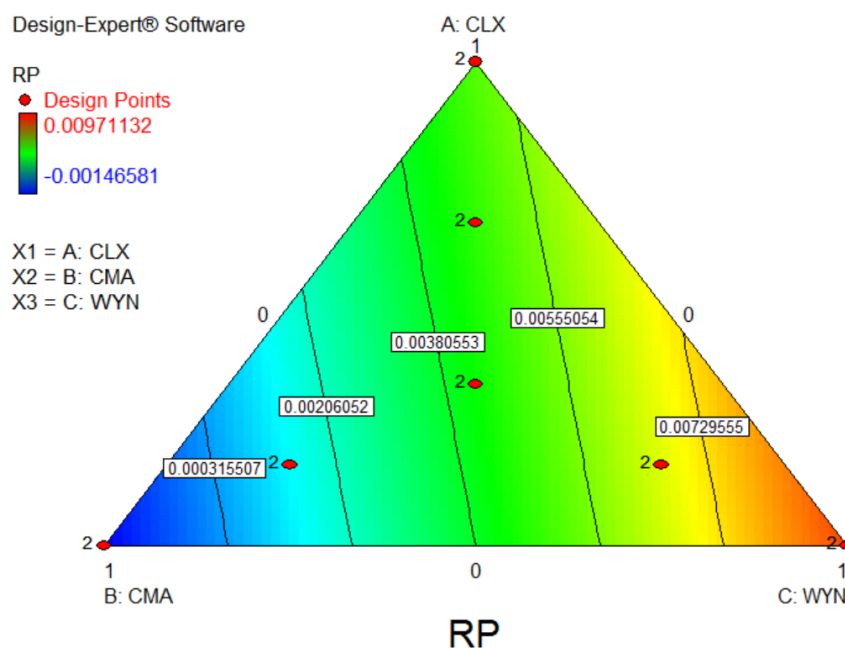


Figura A.21 – Análise de RP Cluster 3(-1) Contorno

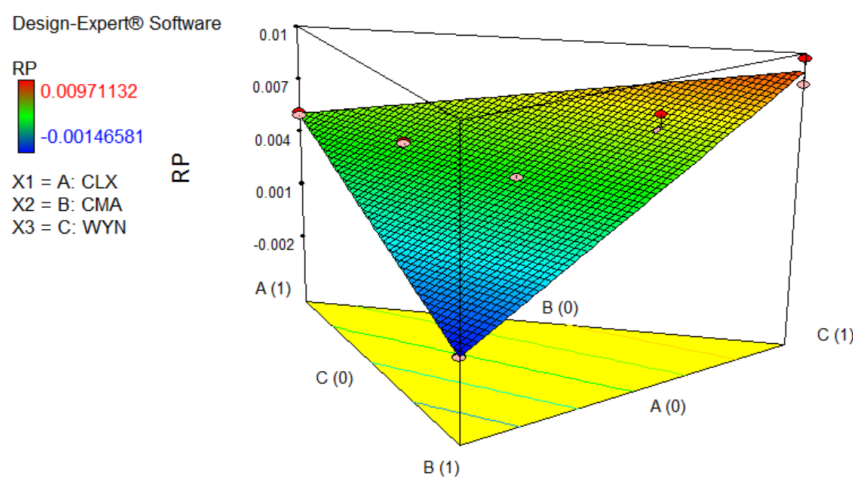


Figura A.22 – Análise de RP Cluster 3(-1) 3D

O valor F de 1237,81 do Quadro A.9 implica que o modelo é significativo. Existe apenas a chance de 0,01% que o F-Valor desta magnitude poderia ocorrer devido ao ruído. Os valores de “Prob> F ”inferiores a 0,0500 indicam que os termos do modelo são significativos. Neste caso os componentes da mistura Linear são termos significativos do modelo. Valores superiores a 0,1000 indicam os termos do modelo não são significativos. Se há muitos termos não significativos no modelo (não contando os necessários para suportar a hierarquia), a redução do modelo pode melhorar as funções. A falta de ajuste do Valor-F (“F-value ”) de 0,00 implica que a falta de ajuste não é significativa em relação ao erro puro. Há uma chance de 100,00% que uma falta de ajuste F-value tão grande poderia ocorrer devido ao ruído. A falta não significativa de ajuste significa que o modelo é bom. As Figuras A.29, A.30, A.31 e A.32 demonstram as análises de forma gráfica.

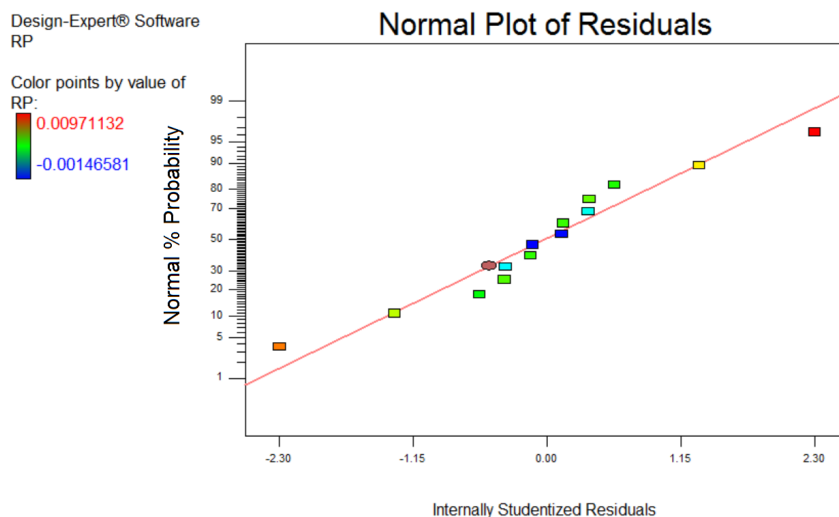


Figura A.23 – Análise de RP Cluster 3(-1) Normalidade

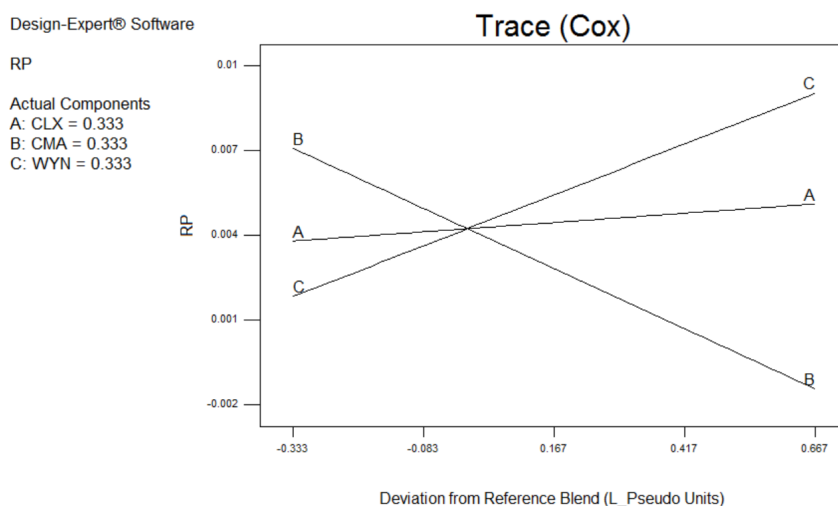


Figura A.24 – Análise de RP Cluster 3(-1) Cox

Quadro A.8 – ANOVA de SP Cluster 3(-1)

Resposta 2 SP					
ANOVA para Modelo de Mistura Quadrático					
*** Componente da Mistura Codificado em Pseudo Componente. ***					
Tabela Análise de Variância [Soma parcial dos quadrados]					
Modelos	Soma dos Quadrados	df	Média dos Quadrados	F Valor	p-value Prob > F
Testados	0,020300058	5	0,004060012	799,2707877	< 0,0001 significativa
Mistura Linear	0,019619975	2	0,009809987	1931,234944	< 0,0001
AB	9,52969E-05	1	9,52969E-05	18,76054521	0,0025
AC	3,23024E-05	1	3,23024E-05	6,359175893	0,0357
BC	1,21622E-05	1	1,21622E-05	2,394300751	0,1604
Residual	4,06372E-05	8	5,07964E-06		
Falta de Ajuste	1,14597E-07	1	1,14597E-07	0,019795787	0,8921 não significativa
Erro Puro	4,05226E-05	7	5,78894E-06		
Corr Total	0,020340695	13			

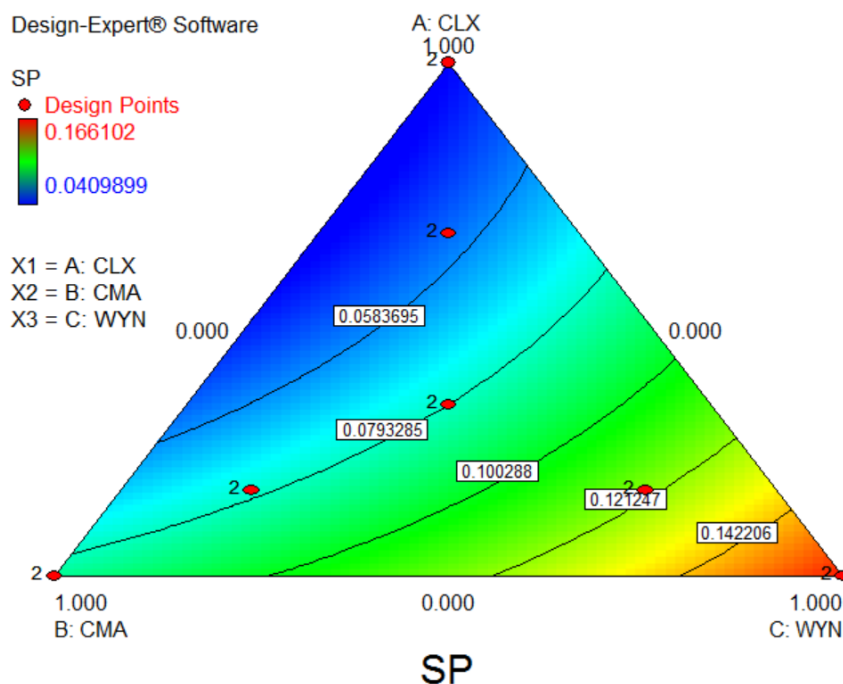


Figura A.25 – Análise de SP Cluster 3(-1) Contorno

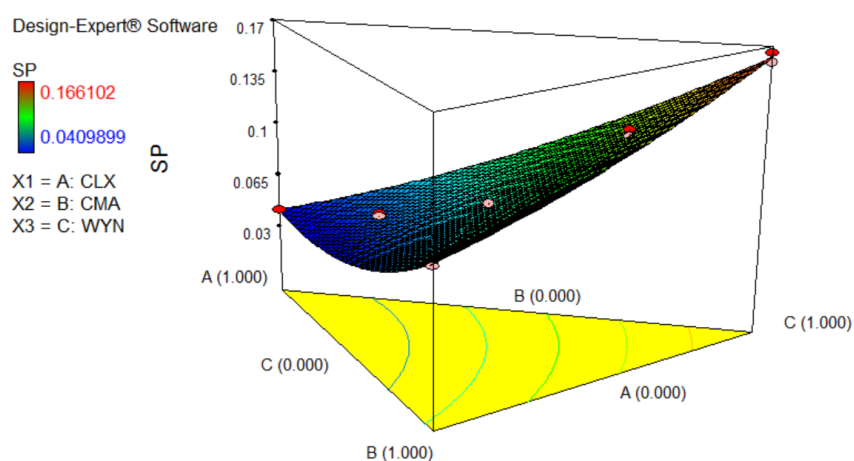


Figura A.26 – Análise de SP Cluster 3(-1) 3D

O valor F de 299,96 do Quadro A.10 implica o modelo é significativo. A falta de ajuste do F-value de 0,00 implica a falta de ajuste não é significativa em relação ao erro puro. Há uma chance de 99,91% que uma falta de ajuste poderia ocorrer devido ao ruído. Falta não significativa de ajuste é bom. As Figuras A.33, A.34, A.35 e A.36 têm as demonstrações gráficas das análises.

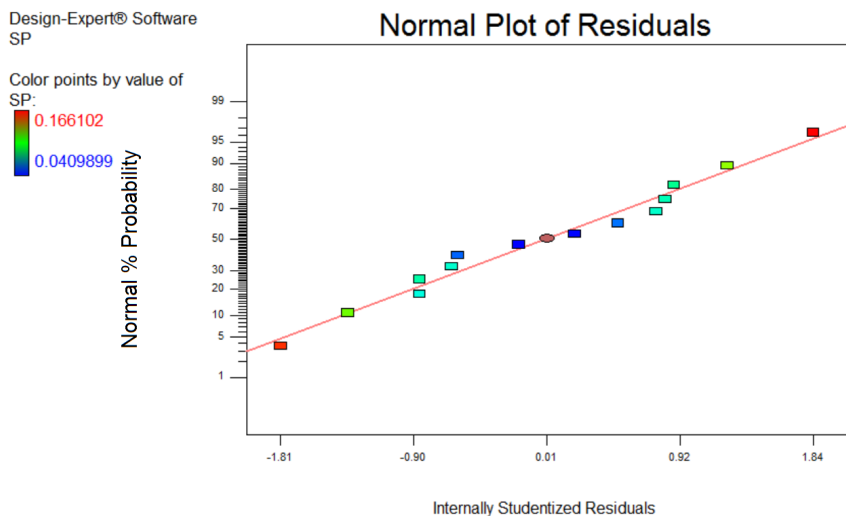


Figura A.27 – Análise de SP Cluster 3(-1) Normalizado

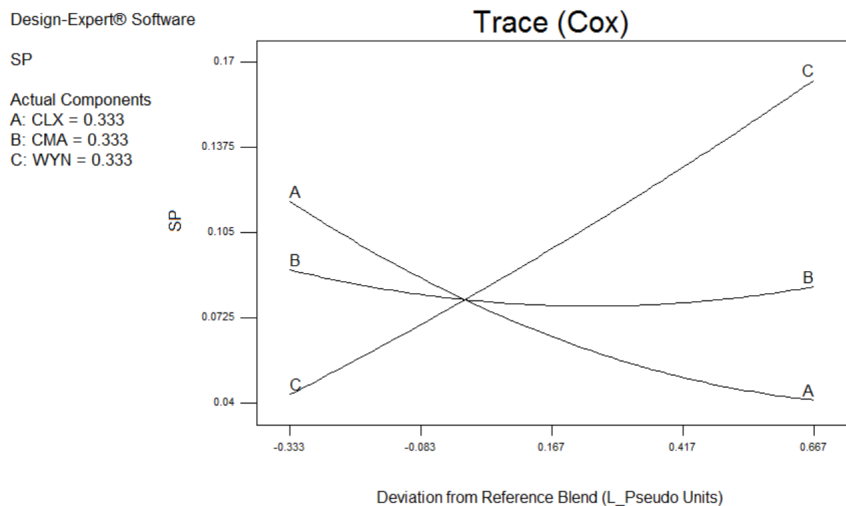


Figura A.28 – Análise de SP Cluster 3(-1) Cox

Quadro A.9 – ANOVA de RP Cluster 3(+1)

Resposta 1 RP					
ANOVA para Modelo de Mistura Linear					
Componente da Mistura Codificado em Pseudo Componente					
Tabela Análise de Variância [Soma parcial dos quadrados]					
	Soma dos	Média dos	F	p-value	
Modelos	Quadrados	df	Quadrados	Valor	Prob > F
Testados	0,000235051	3	7,83502E-05	1237,813483	< 0,0001*
Mistura Linear	0,000235051	3	7,83502E-05	1237,813483	< 0,0001
Residual	8,86162E-07	14	6,32973E-08		
Falta de Ajuste	0,000000	5	0,000000	0,000000	1,000000**
Erro Puro	8,86162E-07	9	9,84625E-08		
Corr Total	0,000235937	17			
(*) significante; (**) não significante					

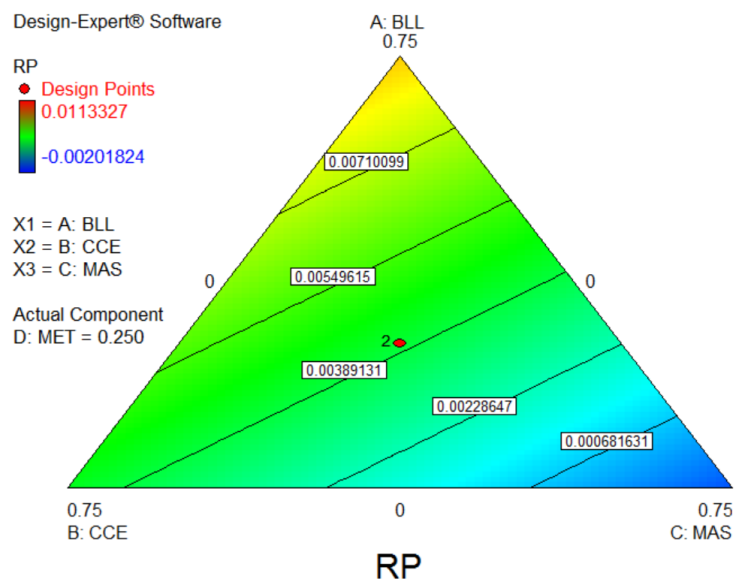


Figura A.29 – Análise de RP Cluster 3(+1) Contorno

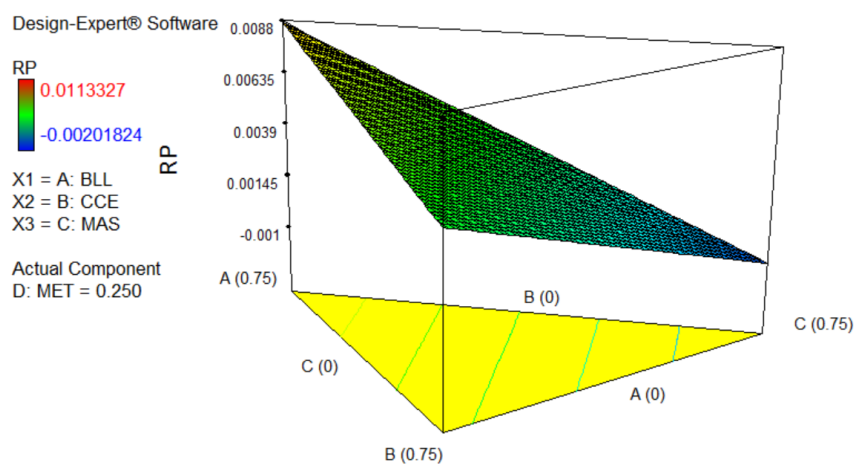


Figura A.30 – Análise de RP Cluster 3(+1) 3D

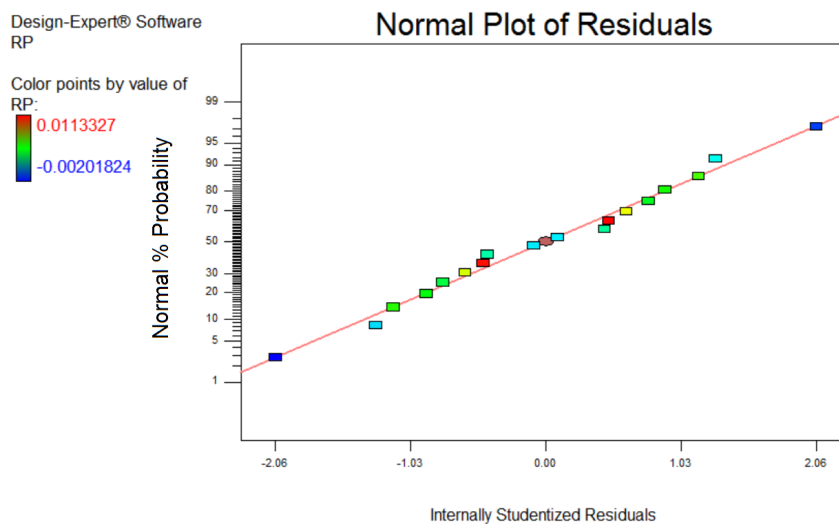


Figura A.31 – Análise de RP Cluster 3(+1) Normalizado

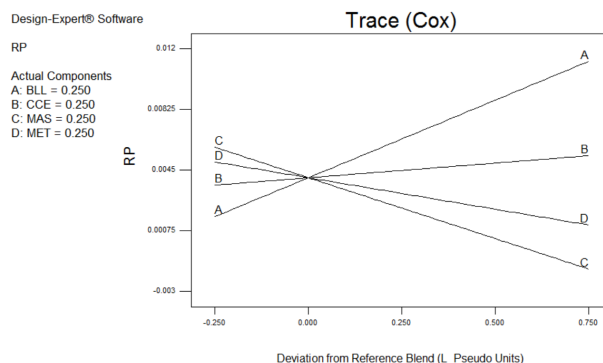


Figura A.32 – Análise de RP Cluster 3(+1) Cox

Quadro A.10 – ANOVA de SP Cluster 3(+1)

Resposta 2 SP					
ANOVA para Modelo de Mistura Quadrático (Distorcido)					
Componente da Mistura Codificado em Pseudo Componente					
Tabela Análise de Variância [Soma parcial dos quadrados]					
Modelos	Soma dos Quadrados	df	Média dos Quadrados	F Valor	p-value Prob > F
Testados	0,007327661	7	0,001046809	299,9555467	< 0,0001*
Mistura Linear	0,006483328	3	0,002161109	619,2503953	< 0,0001
AB	2,02118E-05	1	2,02118E-05	5,791552324	0,0369
AC	5,06571E-06	1	5,06571E-06	1,451541964	0,256
AD	7,26622E-05	1	7,26622E-05	20,82084157	0,001
BC	4,46414E-05	1	4,46414E-05	12,79167815	0,005
BD	0,000000	0			
CD	0,000000	0			
Residual	3,48988E-05	10	3,48988E-06		
Falta de Ajuste	5,38427E-12	1	5,38427E-12	1,38854E-06	0,9991**
Erro Puro	3,48988E-05	9	3,87764E-06		
Corr Total	0,00736256	17			
(*) significante; (**) não significante					

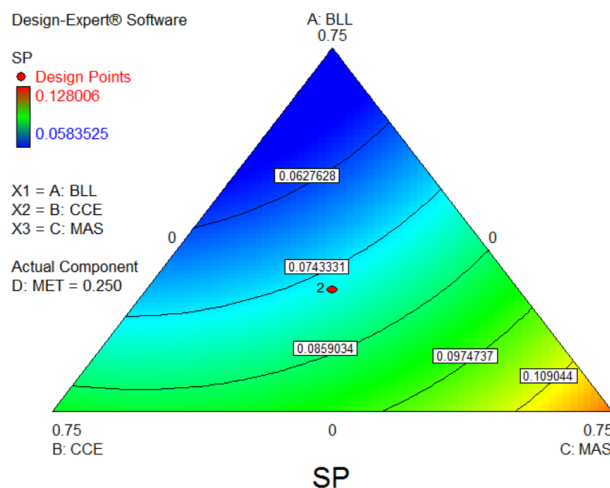


Figura A.33 – Análise de SP Cluster 3(+1) Contorno

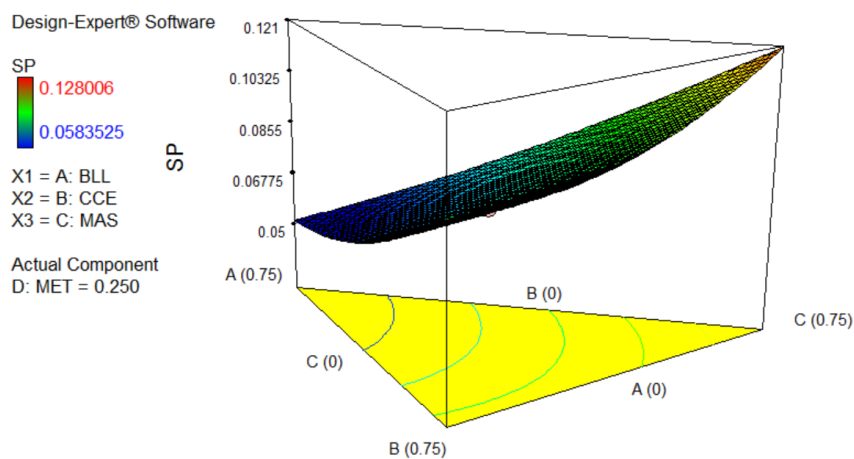


Figura A.34 – Análise de SP Cluster 3(+1) 3D

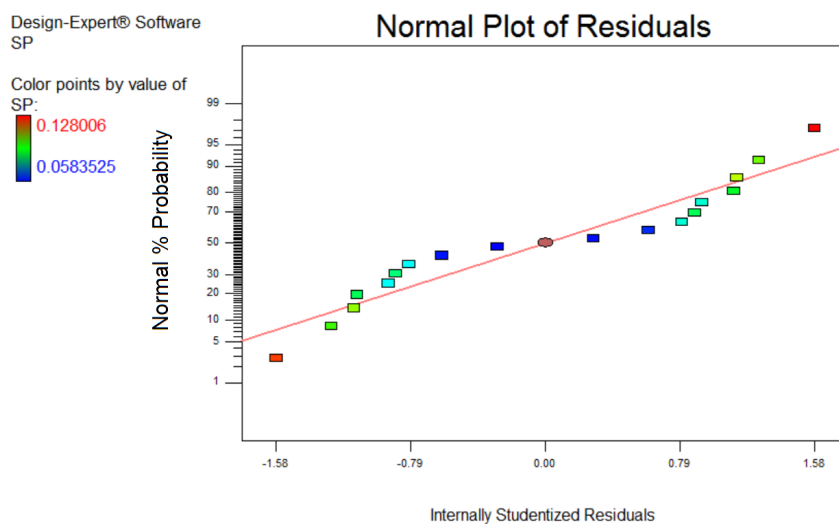


Figura A.35 – Análise de SP Cluster 3(+1) Normalizado

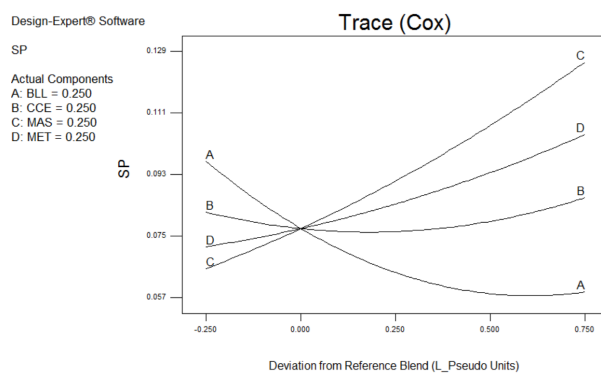


Figura A.36 – Análise de SP Cluster 3(+1) Cox

APÊNDICE B – ANOVA E GRÁFICOS COM ANÁLISE COMBINADA DE TODOS OS *CLUSTERS* CONSIDERANDO MÉTODO LEXICOGRÁFICO E VARIÁVEIS DE PROCESSO (CMO)

O valor F de 2737,06 da Tabela B.1 implica que o modelo é significativo. Neste caso os componentes lineares da mistura, AD, BD, CD são termos significativos do modelo. A falta de ajuste do F-value de 1,83 implica que a falta de ajuste não é significativa em relação ao erro puro. Há uma chance de 21,59% que uma falta de ajuste poderia ocorrer devido ao ruído. Falta não significativa de ajuste é bom. As Figuras B.1, B.2, B.3, B.4 são a demonstração gráfica das análises.

Quadro B.1 – ANOVA de RP do arranjo combinado

Resposta 1 RP					
ANOVA para Combinação Linear x Modelo de Efeitos Principais Componente da Mistura Codificado em Pseudo Componente Tabela Análise de Variância [Soma parcial dos quadrados]					
Modelos	Soma dos Quadrados	df	Média dos Quadrados	F Valor	p-value Prob > F
Testados	0,000299719	5	5,99438E-05	2737,064286	< 0,0001*
Mistura Linear	6,95673E-05	2	3,47836E-05	1588,239442	< 0,0001
AD	0,000102827	1	0,000102827	4695,122609	< 0,0001
BD	0,000121987	1	0,000121987	5569,992442	< 0,0001
CD	4,63502E-05	1	4,63502E-05	2116,374675	< 0,0001
Residual	4,16114E-07	19	2,19007E-08		
Falta de Ajuste	3,15489E-07	12	2,62907E-08	1,828909245	0,2159**
Erro Puro	1,00626E-07	7	1,43751E-08		
Corr Total	0,000300135	24			
(*) significante; (**) não significante					

O valor F de 181,35 da Tabela B.2 implica que o modelo é significativo. Neste caso os componentes lineares da mistura, AB, AC, AD, BC, BD, CD, ACD são termos significativos do modelo. A falta de ajuste do F-value de 74,28 implica a falta de ajuste é significativo. Há apenas uma está 0,01% de chance de que uma falta de ajuste poderia ocorrer devido ao ruído. Significativa falta de ajuste é ruim o modelo deve adaptar-se a esse ajuste, mas o modelo se demonstra aplicável para otimização multiobjetivo. As Figuras B.5, B.6, B.7 e B.8 demonstram as análises de forma gráfica.

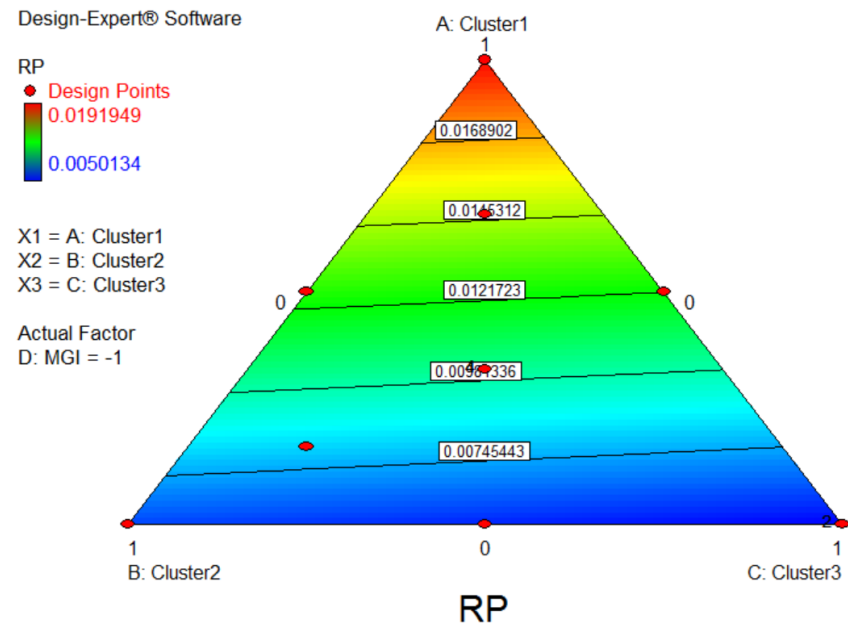


Figura B.1 – Análise de RP do arranjo CMO Contorno

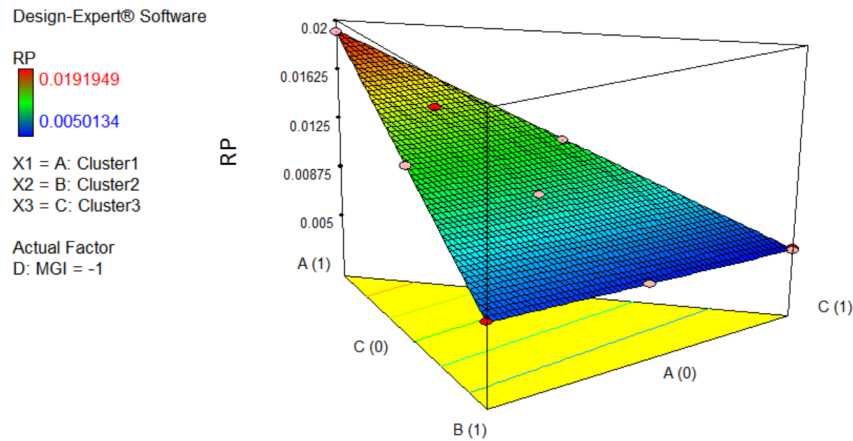


Figura B.2 – Análise de RP do arranjo CMO 3D

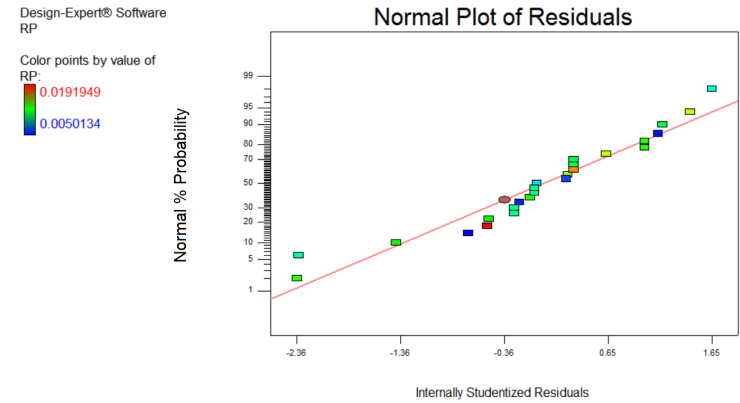


Figura B.3 – Análise de RP do arranjo CMO Normalidade

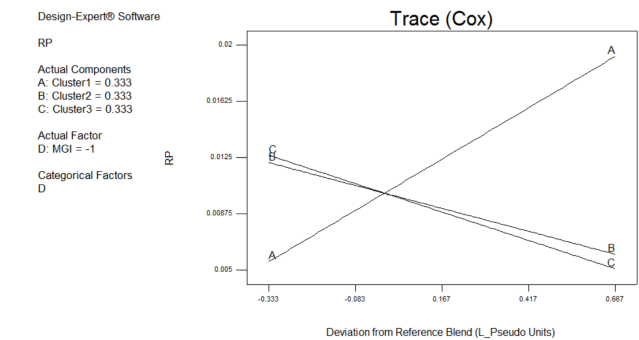


Figura B.4 – Análise de RP do arranjo CMO Cox

Quadro B.2 – ANOVA de SP do arranjo combinado

Resposta 2 SP					
ANOVA para Quadrático Combinado x Modelo de Efeitos Principais					
Componente da Mistura Codificado em Pseudo Componente					
Tabela Análise de Variância [Soma parcial dos quadrados]					
Modelos	Soma dos		Média dos	F	p-value
Testados	Quadrados	df	Quadrados	Valor	Prob > F
Mistura Linear	0,001058919	2	0,000529459	226,3173107	< 0,0001
AB	0,00098623	1	0,00098623	421,5637125	< 0,0001
AC	0,000704273	1	0,000704273	301,0412605	< 0,0001
AD	0,000147961	1	0,000147961	63,24592952	< 0,0001
BC	0,00062334	1	0,00062334	266,4463747	< 0,0001
BD	0,000134765	1	0,000134765	57,60509871	< 0,0001
CD	0,000291818	1	0,000291818	124,7375177	< 0,0001
ABD	2,44926E-07	1	2,44926E-07	0,104693525	0,7514
ACD	1,18335E-05	1	1,18335E-05	5,058231534	0,0425
BCD	2,06857E-06	1	2,06857E-06	0,884209131	0,3642
Residual	3,04129E-05	13	2,33946E-06		
Falta de Ajuste	2,99427E-05	6	4,99044E-06	74,28179638	< 0,0001*
Erro Puro	4,70278E-07	7	6,71826E-08		
Corr Total	0,004697381	24			

(*) significante; (**) não significante

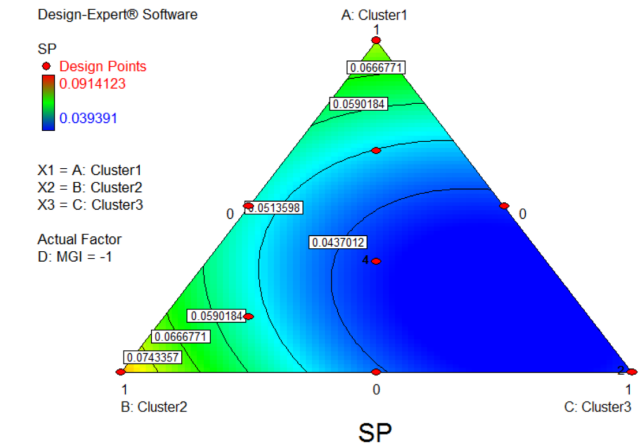


Figura B.5 – Análise de SP do arranjo combinado Contorno

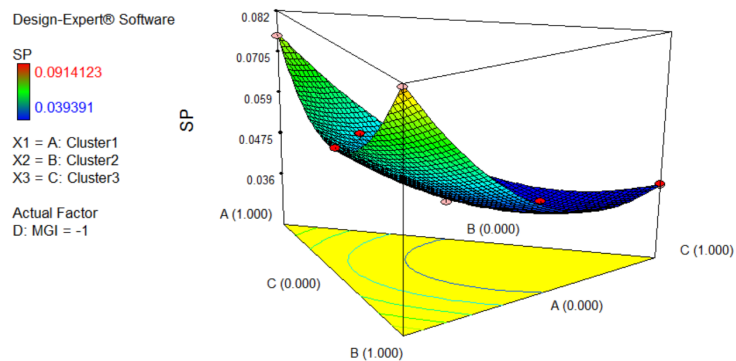


Figura B.6 – Análise de SP do arranjo combinado 3D

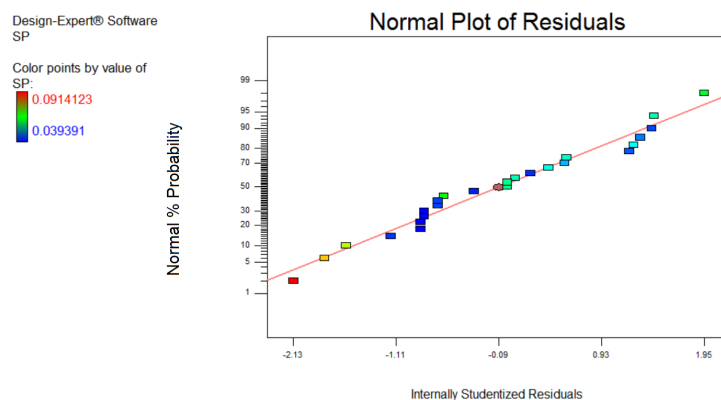


Figura B.7 – Análise de SP do arranjo combinado Normalidade

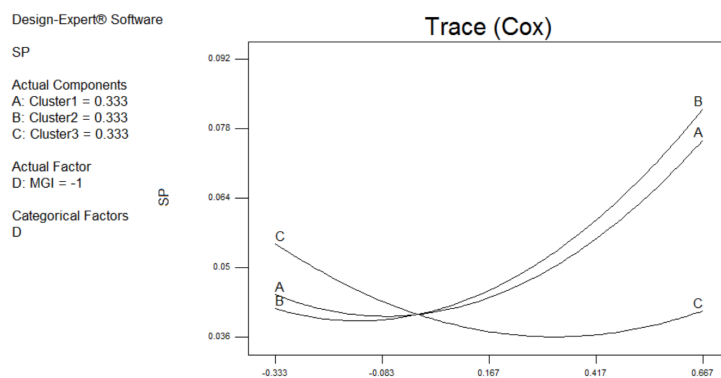


Figura B.8 – Análise de SP do arranjo combinado Cox

APÊNDICE C – ANOVA E GRÁFICOS COM ANÁLISES POR *CLUSTER* E VARIÁVEIS DE PROCESSO JUNTOS UTILIZANDO CM-DOE

O valor F de 138,52 do Quadro C.1 implica que o modelo é significativo. Existe apenas a chance de 0,01% que o F-Valor poderia ocorrer devido ao ruído. Os valores de “Prob> F” inferior a 0,0500 indica que os termos do modelo são significativos. Neste caso os componentes lineares da mistura, ADG, AEG, AFG, BFG, CDE, CDF, CDG, CEG, CFG, CDEG, CEFG são termos significativos do modelo. Valores superiores a 0,1000 indicam os termos do modelo que não são significativos. A falta de ajuste F-value de 0,80 indica que a falta de ajuste não é significativa em relação erro ao puro. Há uma chance de 70,41% que a falta de ajuste do F-value poderia ocorrer devido ao ruído. Falta não significativa de ajuste é bom. A análise gráfica pode ser vista nas Figuras C.1, C.2, C.7, C.3, C.4, C.5, C.6, C.7. As análises completas podem também ser encontradas em Papandrea Pedro; Paiva (2016).

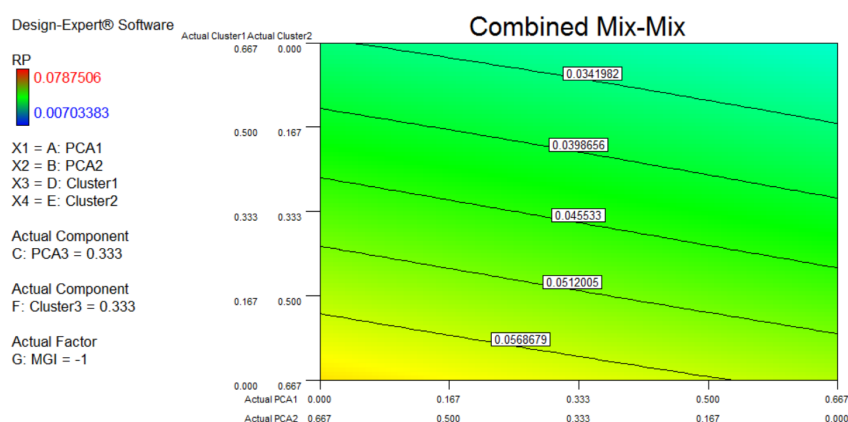


Figura C.1 – Análise de RP CM-DOE Contorno MGI(-1)

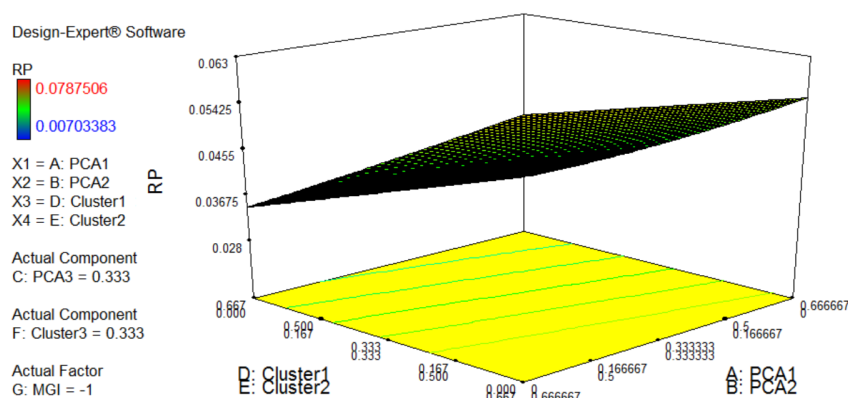


Figura C.2 – Análises de RP CM-DOE 3D MGI(-1)

Quadro C.1 – ANOVA de RP CM-DOE

Resposta 1 RP						
ANOVA para CM-DOE x Quadrático x Modelo de Efeitos Principais						
Componente da Mistura Codificado em Pseudo Componente						
Análise dos Componentes Maiores da Mistura						
Tabela Análise de Variância [Soma parcial dos quadrados]						
		Misturas		Componentes		Total
		Mistura 1		A B C		1
		Mistura 2		D E F		1
Modelos	Soma dos	df	Média dos	F	p-value	
Testados	Quadrados		Quadrados	Valor	Prob > F	
Linear x Mistura Linear	0,019102575	35	0,000545788	138,522101	< 0.0001*	
ADE	0,015907753	8	0,001988469	504,677621	< 0.0001	
ADF	9,8751E-06	1	9,8751E-06	2,506322	0.1207	
ADG	1,42446E-05	1	1,42446E-05	3,615302	0.0640	
AEF	0,000138965	1	0,000138965	35,269528	< 0.0001	
AEG	3,85433E-06	1	3,85433E-06	0,978237	0.3282	
AFG	0,000100727	1	0,000100727	25,564716	< 0.0001	
BDE	0,000354396	1	0,000354396	89,946531	< 0.0001	
BDF	1,08085E-05	1	1,08085E-05	2,743227	0.1049	
BDG	6,2332E-08	1	6,2332E-08	0,015820	0.9005	
BEF	3,59048E-06	1	3,59048E-06	0,911271	0.3451	
BEG	2,39432E-08	1	2,39432E-08	0,006077	0.9382	
BFG	8,28142E-08	1	8,28142E-08	0,021018	0.8854	
CDE	7,57328E-05	1	7,57328E-05	19,221153	< 0.0001	
CDF	3,55767E-05	1	3,55767E-05	9,029431	0.0044	
CDG	2,79827E-05	1	2,79827E-05	7,102072	0.0108	
CEF	0,000326323	1	0,000326323	82,821582	< 0.0001	
CEG	6,6149E-06	1	6,6149E-06	1,678876	0.2020	
CFG	0,000281264	1	0,000281264	71,385304	< 0.0001	
ADEG	0,000146141	1	0,000146141	37,090963	< 0.0001	
ADFG	5,07627E-06	1	5,07627E-06	1,288368	0.2626	
AEFG	6,81616E-07	1	6,81616E-07	0,172995	0.6795	
BDEG	7,80606E-06	1	7,80606E-06	1,981194	0.1665	
BDFG	2,21461E-06	1	2,21461E-06	0,562073	0.4575	
BEFG	7,28155E-08	1	7,28155E-08	0,018481	0.8925	
CDEG	6,90023E-08	1	6,90023E-08	0,017513	0.8953	
CDFG	3,10243E-05	1	3,10243E-05	7,874041	0.0075	
CEFG	9,10463E-06	1	9,10463E-06	2,310775	0.1358	
Residual	3,07257E-05	1	3,07257E-05	7,798248	0.0078	
Falta de Ajuste	0,000169423	43	3,94008E-06			
Erro Puro	0,000159652	41	3,89396E-06	0,797030	0.7041**	
Corr Total	9,77116E-06	2	4,88558E-06			
	0,019271999	78				
(*) significante; (**) não significante						

O valor F de 171,97 do Quadro C.2 implica que o modelo é significativo. Neste caso os componentes lineares da mistura, ABD, ABE, ABF, ACD, ACE, ACF, ADE, ADF, ADG, AEF são significativos para o modelo. A falta de ajuste do F-value de 612,54

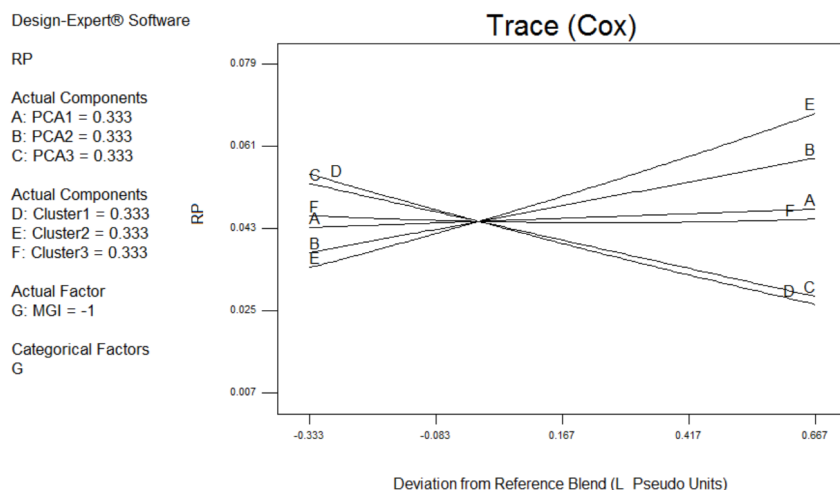


Figura C.3 – Análises de RP CM-DOE Cox MGI(-1)

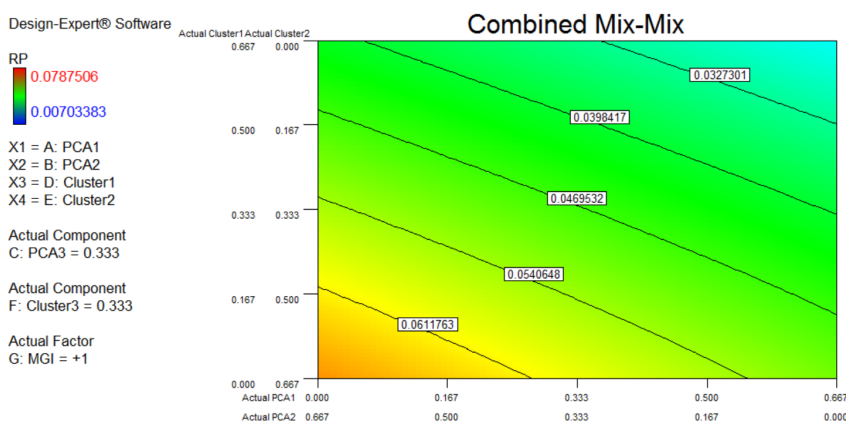


Figura C.4 – Análises de RP CM-DOE Contorno MGI(+1)

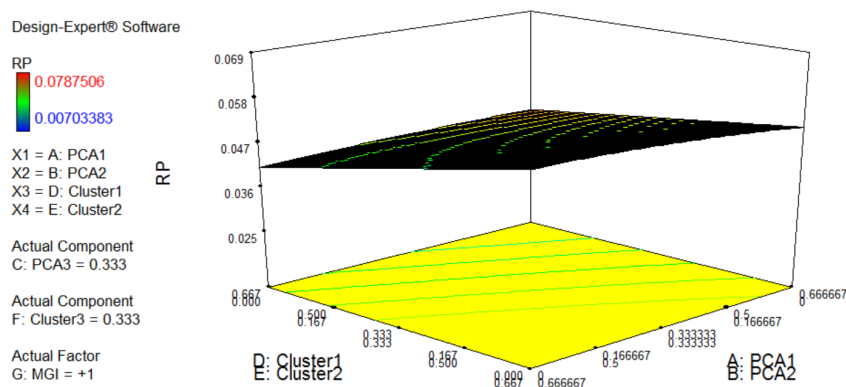


Figura C.5 – Análises de RP CM-DOE 3D MGI(+1)

implica a falta de ajuste é significativa. Há apenas uma chance de 0,16% de que a falta de ajuste poderia ocorrer devido ao ruído. Significativa falta de ajuste é ruim, mesmo assim o modelo é significativo para a otimização. As análises gráficas podem ser vistas nas Figuras C.8, C.9, C.10, C.12, C.13 C.14. As análises completas podem também ser encontradas em Papandrea Pedro; Paiva (2016).

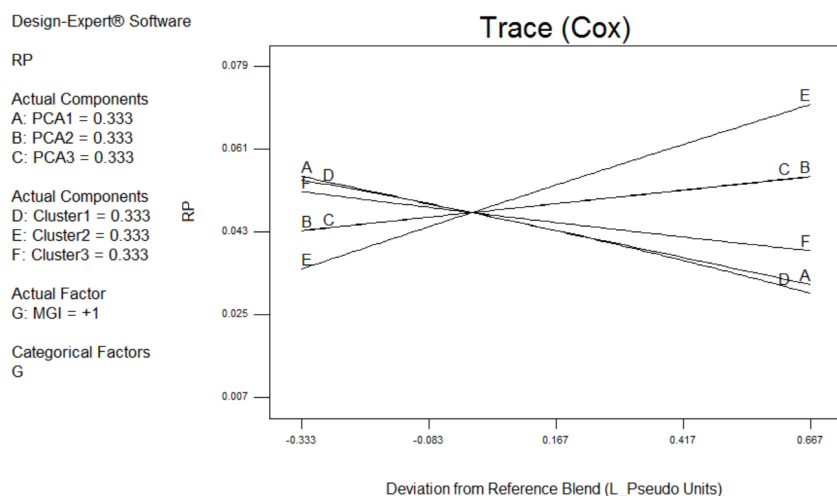


Figura C.6 – Análises de RP CM-DOE Cox MGI(+1)

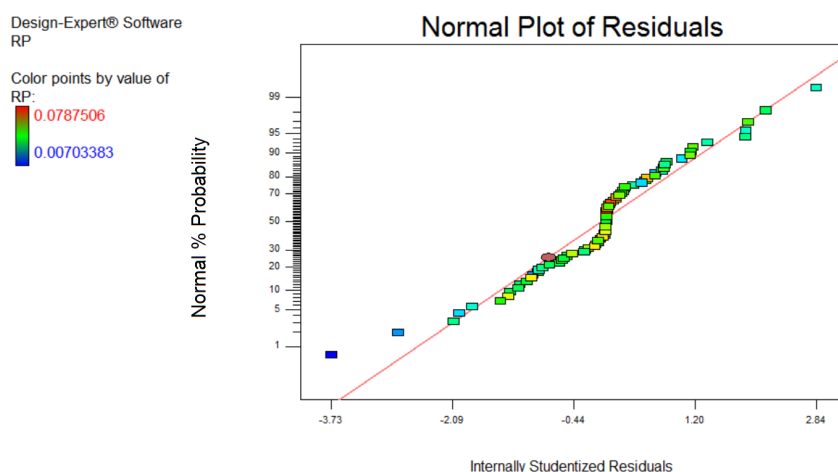


Figura C.7 – Análises de RP CM-DOE Normalizado

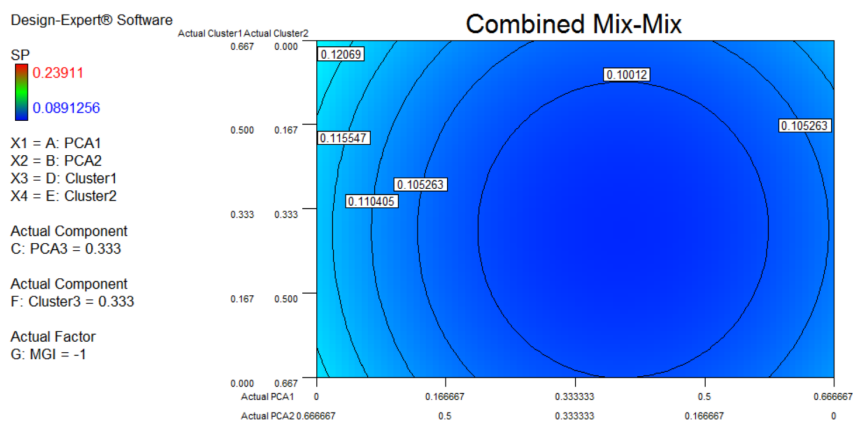


Figura C.8 – Análise de SP CM-DOE Contorno MGI(-1)

Quadro C.2 – ANOVA de SP CM-DOE

Resposta 2 SP					
ANOVA para CM-DOE x Quadrático x Modelo de Efeitos Principais					
Componente da Mistura Codificado em Pseudo Componente					
Análise dos Componentes Maiores da Mistura					
Tabela de Análise de Variância [Soma dos quadrados parcial - Tipo III]					
			Misturas	Componentes	Total
			Mistura 1	A B C	1
			Mistura 2	D E F	1
Modelos	Soma dos	Média dos	F	p-value	
Testados	Quadrados	df	Quadrados	Valor	Prob > F
Linear x Mistura Linear	0,138122927	71	0,001945393	171,96819	< 0.0001*
ABD	0,049671002	8	0,006208875	548,84995	< 0.0001
ABE	0,001795601	1	0,001795601	158,72688	< 0.0001
ABF	0,001740667	1	0,001740667	153,8709	< 0.0001
ABF	0,001797267	1	0,001797267	158,87421	< 0.0001
ACD	0,003284635	1	0,003284635	290,35396	< 0.0001
ACE	0,003246042	1	0,003246042	286,9425	< 0.0001
ACF	0,003391504	1	0,003391504	299,80097	< 0.0001
ADE	0,0002511	1	0,0002511	22,196671	0.0022
ADF	0,000306773	1	0,000306773	27,117977	0.0012
ADG	0,004615511	1	0,004615511	408,00029	< 0.0001
AEF	0,000164352	1	0,000164352	14,528362	0.0066
AEG	0,00367222	1	0,00367222	324,61562	< 0.0001
AFG	0,00697337	1	0,00697337	616,42949	< 0.0001
BCD	0,002280502	1	0,002280502	201,59101	< 0.0001
BCE	0,002282051	1	0,002282051	201,72794	< 0.0001
BCF	0,002472947	1	0,002472947	218,60273	< 0.0001
BDE	0,000232602	1	0,000232602	20,561461	0.0027
BDF	0,000280946	1	0,000280946	24,834943	0.0016
BDG	0,000491689	1	0,000491689	43,464138	0.0003
BEF	0,000175342	1	0,000175342	15,49986	0.0056
BEG	0,000840207	1	0,000840207	74,272327	< 0.0001
BFG	4,30221E-05	1	4,30221E-05	3,8030538	0.0921
CDE	0,000206195	1	0,000206195	18,22712	0.0037
CDF	0,000266911	1	0,000266911	23,594268	0.0018
CDG	3,76415E-05	1	3,76415E-05	3,3274234	0.1109
CEF	0,00015802	1	0,00015802	13,968614	0.0073
CEG	3,49803E-07	1	3,49803E-07	0,0309218	0.8654
CFG	0,000663873	1	0,000663873	58,68479	0.0001
ABDE	2,84812E-08	1	2,84812E-08	0,0025177	0.9614
ABDF	1,44286E-07	1	1,44286E-07	0,0127545	0.9133
ABDG	2,26424E-06	1	2,26424E-06	0,2001531	0.6681
ABEF	7,75096E-07	1	7,75096E-07	0,0685167	0.8010
ABEG	2,04227E-06	1	2,04227E-06	0,1805316	0.6837
ABFG	1,83706E-06	1	1,83706E-06	0,162392	0.6990
ACDE	6,04672E-08	1	6,04672E-08	0,0053452	0.9438
ACDF	1,49738E-07	1	1,49738E-07	0,0132365	0.9116
ACDG	0,000114704	1	0,000114704	10,139537	0.0154
ACEF	5,72262E-07	1	5,72262E-07	0,0505866	0.8285
ACEG	0,000109594	1	0,000109594	9,6878278	0.0170
ACFG	0,000119493	1	0,000119493	10,562884	0.0141
ADEG	1,13531E-06	1	1,13531E-06	0,1003591	0.7606
Residual	7,91876E-05	7	1,13125E-05		
Falta de Ajuste	7,9136E-05	5	1,58272E-05	612,53658	0.0016**
Erro Puro	5,16775E-08	2	2,58388E-08		
Corr Total	0,138202114	78			
(*) significante; (**) não significante					

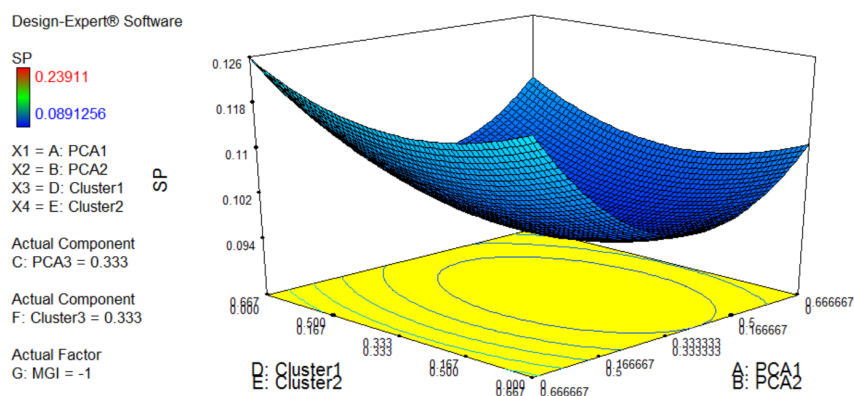


Figura C.9 – Análises de SP CM-DOE 3D MGI(-1)

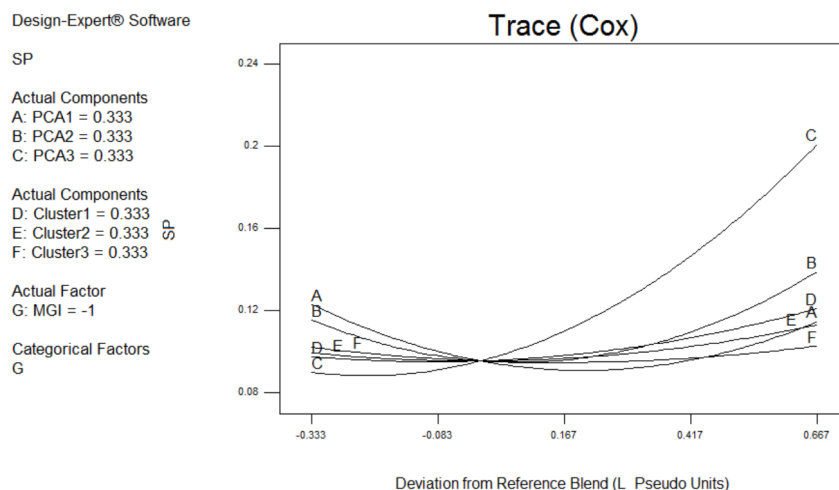


Figura C.10 – Análises de SP CM-DOE Cox MGI(-1)

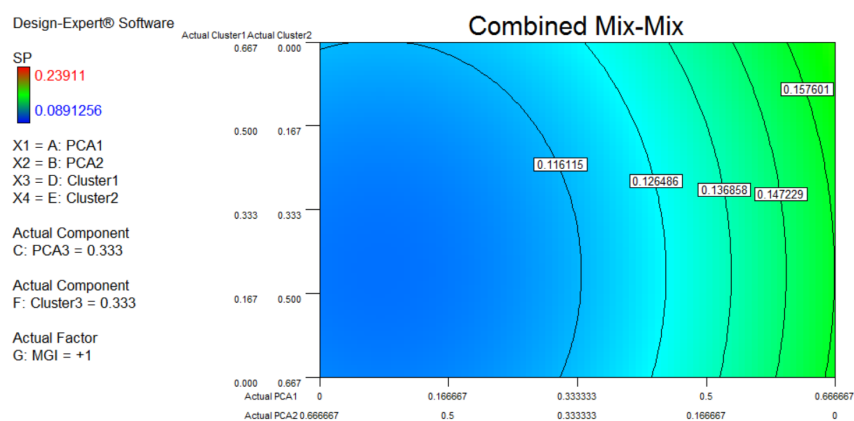


Figura C.11 – Análises de SP CM-DOE Contorno MGI(+1)

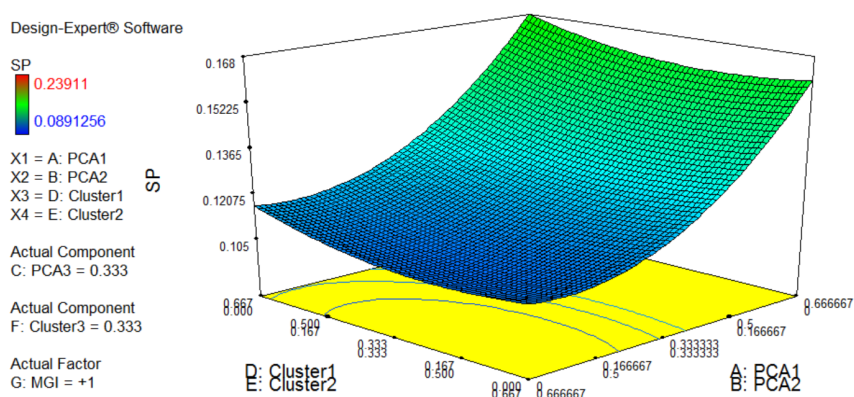


Figura C.12 – Análises de SP CM-DOE 3D MGI(+1)

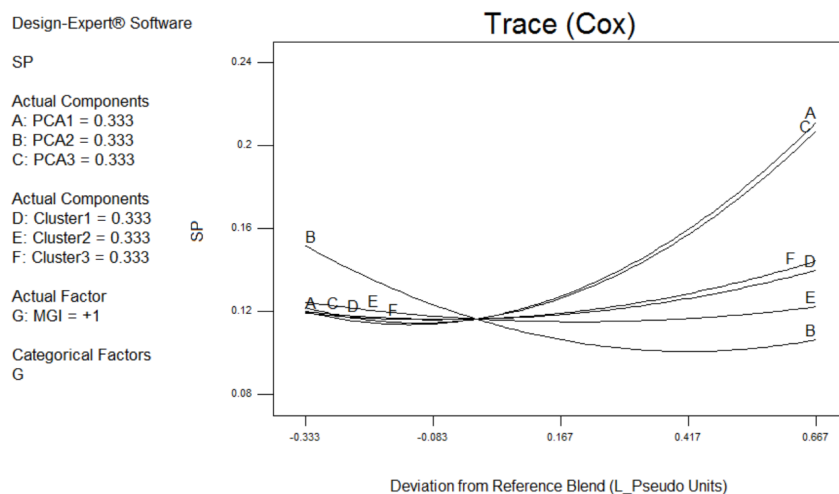


Figura C.13 – Análises de SP CM-DOE Cox MGI(+1)

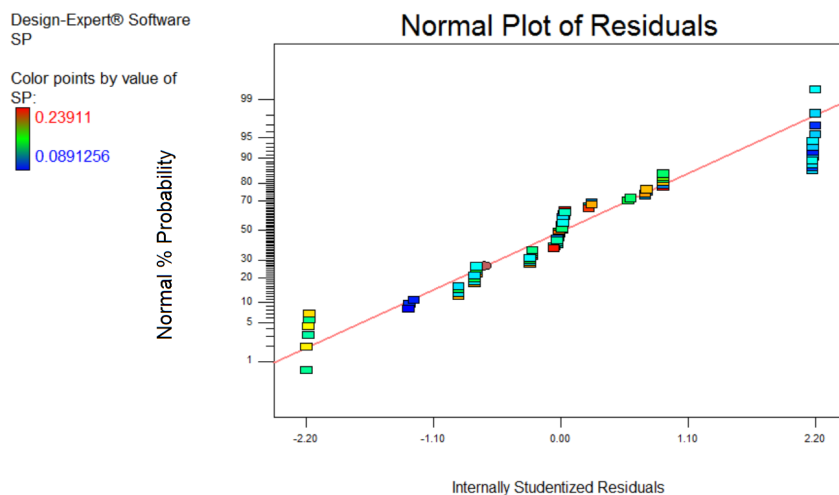


Figura C.14 – Análises de SP CM-DOE Normalizado

APÊNDICE D – ANOVA E GRÁFICOS COM ANÁLISE UTILIZANDO O MODELO MARKOWITZ

O valor F de 315,30 do Quadro D.1 implica que o modelo é significativo. Neste caso os componentes lineares da mistura termos significativos. A falta de ajuste do F-value de 0,00 implica que a falta de ajuste não é significativa em relação ao erro puro. Falta não significativa de ajuste é bom. As Figuras D.1, D.2, D.3 e D.4 mostram as análises gráficas. As análises completas podem também ser encontradas em Papandrea Pedro; Paiva (2016).

Quadro D.1 – ANOVA de RP modelo Markowitz

Resposta 1 RP					
ANOVA para Modelo de Mistura Linear					
Componente da Mistura Codificado em Pseudo Componente					
Tabela de Análise de Variância [Soma parcial dos quadrados - Tipo III]					
	Soma dos		Média dos	F	p-value
Modelos	Quadrados	df	Quadrados	Valor	Prob > F
Testados	0,003466221	19	0,000182433	315,3025842	< 0,0001*
Mistura Linear	0,003466221	19	0,000182433	315,3025842	< 0,0001
Residual	3,58729E-05	62	5,78596E-07		
Falta de Ajuste	0,000000000	21	0,000000000	0,000000000	1,0000**
Erro Puro	3,58729E-05	41	8,74949E-07		
Corr Total	0,003502094	81			

(*) signifiante; (**) não signifiante

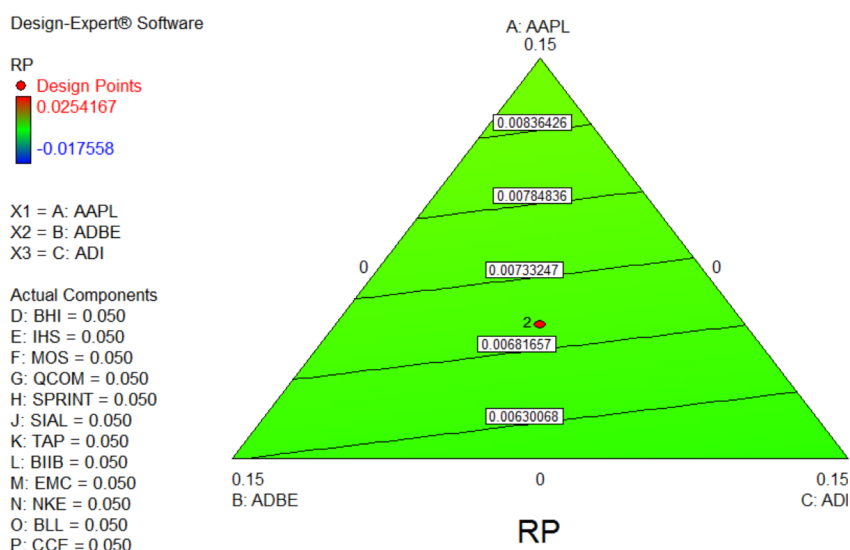


Figura D.1 – Análise de RP Contorno do modelo de mistura Markowitz

O valor F de 619,02 do Quadro D.2 implica que o modelo é significativo. Neste caso, os componentes lineares da mistura AB, AC, AD, AE, AF, AG, AH, AJ, AK, AL, AM, AN, AO, AP, AQ, AR, AS, AT, AU, BC são termos significativos do modelo. A falta

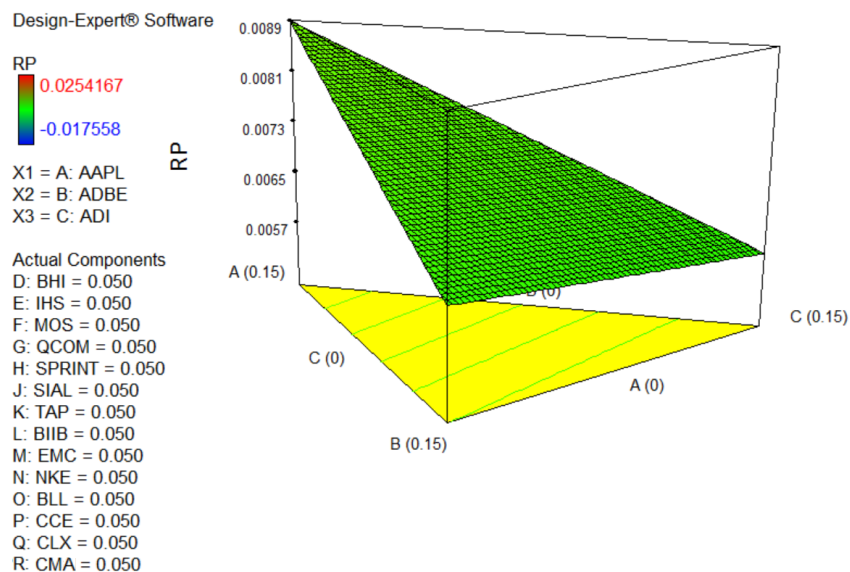


Figura D.2 – Análise de RP 3D do modelo de mistura Markowitz

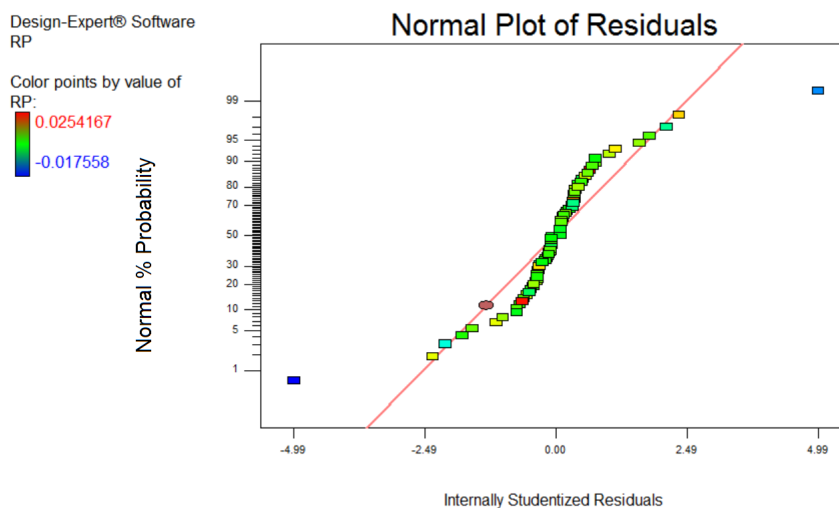


Figura D.3 – Análise de RP Normalizado do modelo de mistura Markowitz

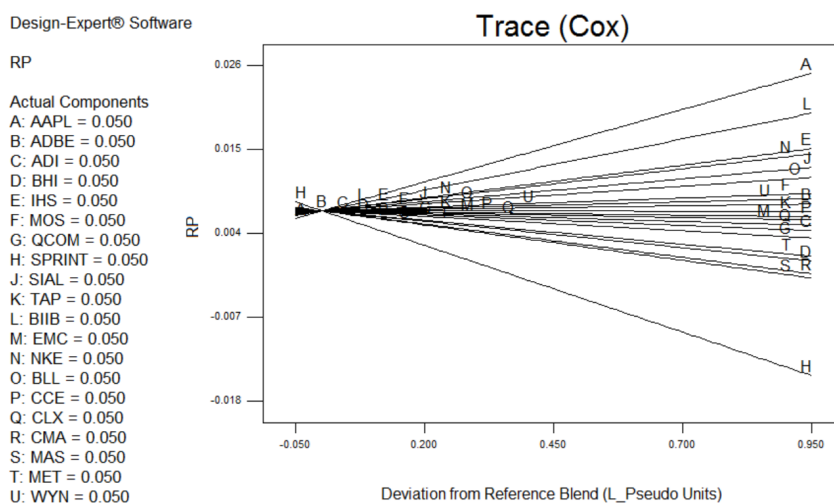


Figura D.4 – Análise de RP Cox do modelo de mistura Markowitz

de ajuste do F-value de 5,23 implica que a falta de ajuste é significativo. Há apenas 2,74% de chance de que a falta de ajuste poderia ocorrer devido ao ruído. Significativa falta de ajuste é ruim, mas o modelo se demonstra aplicável para otimização multiobjetivo. As Figuras D.5, D.6, D.7 e D.8 mostram as análises gráficas. As análises completas podem também ser encontradas em Papandrea Pedro; Paiva (2016).

Quadro D.2 – ANOVA de SP modelo Markowitz

Resposta 2 SP					
ANOVA para Modelo de Mistura Quadrático (Distorcido)					
Componente da Mistura Codificado em Pseudo Componente					
Tabela de Análise de Variância [Soma parcial dos quadrados - Tipo III]					
Modelos	Soma dos Quadrados	df	Média dos Quadrados	F Valor	p-value Prob > F
Testados	0,059331659	39	0,001521325	619,01810	< 0,0001*
Mistura Linear	0,045493562	19	0,002394398	974,26657	< 0,0001
AB	0,007528871	1	0,007528871	3063,45371	< 0,0001
AC	0,007470606	1	0,007470606	3039,74600	< 0,0001
AD	0,000434549	1	0,000434549	176,81552	< 0,0001
AE	0,000638168	1	0,000638168	259,66691	< 0,0001
AF	0,000595881	1	0,000595881	242,46047	< 0,0001
AG	0,000421999	1	0,000421999	171,70909	< 0,0001
AH	0,001003597	1	0,001003597	408,35776	< 0,0001
AJ	0,000381423	1	0,000381423	155,19863	< 0,0001
AK	0,000466801	1	0,000466801	189,93849	< 0,0001
AL	0,000681233	1	0,000681233	277,18989	< 0,0001
AM	0,000444011	1	0,000444011	180,66535	< 0,0001
AN	0,000422761	1	0,000422761	172,01883	< 0,0001
AO	0,000454232	1	0,000454232	184,82440	< 0,0001
AP	0,000361074	1	0,000361074	146,91883	< 0,0001
AQ	0,000496786	1	0,000496786	202,13945	< 0,0001
AR	0,000455749	1	0,000455749	185,44158	< 0,0001
AS	0,000362236	1	0,000362236	147,39190	< 0,0001
AT	0,000271186	1	0,000271186	110,34394	< 0,0001
AU	0,000346355	1	0,000346355	140,93003	< 0,0001
BC	0,00987427	1	0,00987427	4017,78267	< 0,0001
Residual	0,000103221	42	2,45764E-06		
Falta de Ajuste	1,16868E-05	1	1,16868E-05	5,23473	0,0274*
Erro Puro	9,15342E-05	41	2,23254E-06		
Corr Total	0,05943488	81			

(*) significante; (**) não significante

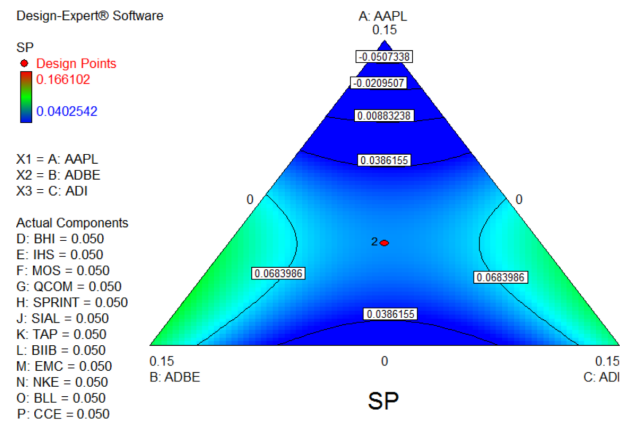


Figura D.5 – Análise de SP Contorno do modelo de mistura Markowitz

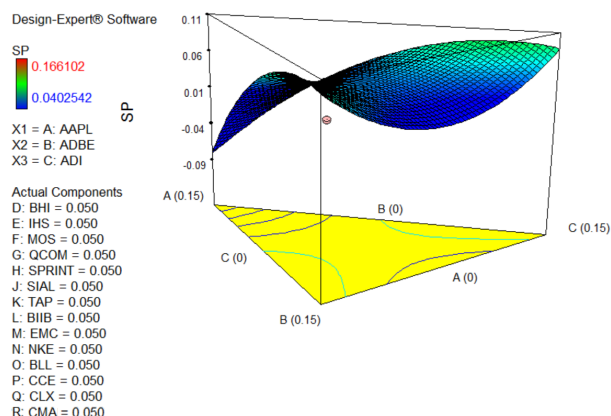


Figura D.6 – Análise de SP 3D do modelo de mistura Markowitz

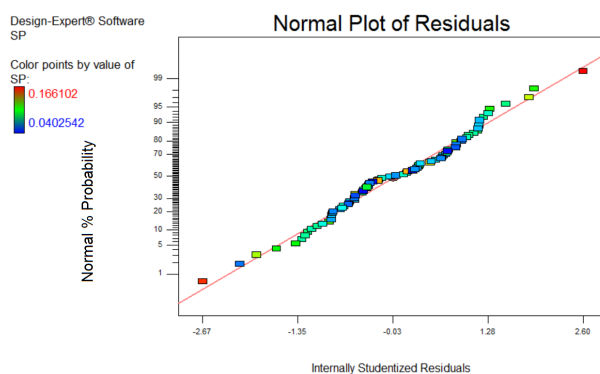


Figura D.7 – Análise de SP Normalizado do modelo de mistura Markowitz

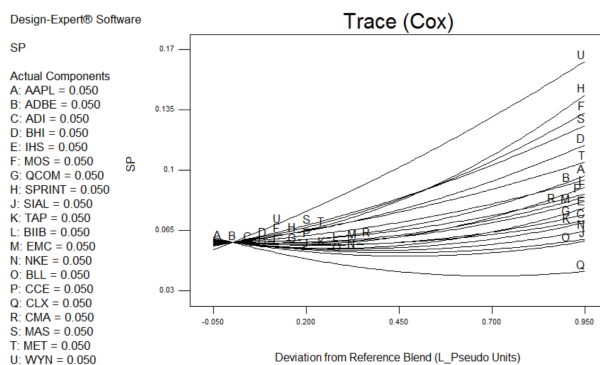


Figura D.8 – Análise de SP Cox do modelo de mistura Markowitz

Anexos

ANEXO A – OPTIMIZATION METHODOLOGY WITH PRINCIPAL COMPONENT ANALYSIS APPLIED IN PORTFOLIO: SIMULATED RESULTS IN GREEN COMPANIES

Manuscript Details

Manuscript number	RED_2018_63
Title	Optimization methodology with principal component analysis applied in portfolio: Simulated results in Green Companies
Article type	Full Length Article

Abstract

This research objectives to develop mathematical models for risk-return that consider simultaneously the linear and non-linear effects of the invested proportions and features of green companies based on their balance sheets/annual reports indicators. The mixture design of experiments is combined with Ward clustered PCA process variables for selecting the most promising companies through generated financial indicators (General Multivariate Indicators method) from the green companies. The green companies are the object of study for this empirical research, they were gathered from Newsweek magazine rank. The proposed Clustered Multilevel Optimization method has showed to be more ro-bust and efficient than the all other tested methods in this research. That means greater security and less risk to the investor.

Keywords principal components analysis; financial indicators; DOE; green companies; portfolio optimization; financial ratios; balance sheets

Taxonomy Statistical Method, Statistical Simulation, Econometric Methodology, Portfolio Optimization, Investment Decisions, Financial Market Efficiency

Corresponding Author Pedro Papandrea

Order of Authors Pedro Papandrea, Anderson Paulo de Paiva, Rafael Leme, Xueping Li, Joao Correa, Joao Turrioni, Vinicius Paes

Submission Files Included in this PDF

File Name [File Type]

CoverLetter.doc [Cover Letter]

PortfolioOptimizationPCA.docx [Manuscript File]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Optimization methodology with principal component analysis applied in portfolio: Simulated results in Green Companies

Pedro José Papandréa^{a*}, Anderson Paulo de Paiva^a, Rafael Coradi Leme^a, Xueping Li^b, João Ederson Corrêa^a, João Batista Turrioni^a, Vinicius de Carvalho Paes^a

^aInstitute of Industrial Engineering and Management, Federal University of Itajuba, AVE. BPS, 1303, Itajuba, Minas Gerais 37500-188, Brazil.

^bDepartment of Industrial and Systems Engineering, University of Tennessee, Knoxville, 520 John D. Tickle Building, 851 Neyland Drive, Knoxville, Tennessee 37996-2315, USA

E-mail addresses: pedro.papandrea@unifei.edu.br, andersonppaiva@unifei.edu.br, xueping.li@utk.edu, leme@unifei.edu.br, edercorrea@unifei.edu.br, joabatu@gmail.com, vinicius.paes@unifei.edu.br

***Corresponding author:** E-mail address: pedro.papandrea@unifei.edu.br. Tel.: +55 35 3425 0104. Cellphone: +55 35 991 329 391. 1303, BPS Avenue, Itajubá, Minas Gerais 37500-188, Brazil.

Abstract

This research objectives to develop mathematical models for risk-return that consider simultaneously the linear and non-linear effects of the invested proportions and features of green companies based on their balance sheets/annual reports indicators. The mixture design of experiments is combined with Ward clustered PCA process variables for selecting the most promising companies through generated financial indicators (General Multivariate Indicators method) from the green companies. The green companies are the object of study for this empirical research, they were gathered from Newsweek magazine rank. The proposed Clustered Multilevel Optimization method has showed to be more robust and efficient than the all other tested methods in this research. That means greater security and less risk to the investor.

Keywords: principal components analysis, financial indicators, DOE, green companies, portfolio optimization, financial ratios, balance sheets.

1. INTRODUCTION

The consideration of financial indicators acquired from balance sheets of several green companies is an extension studied in this article, since it is possible to encounter several works on portfolio analysis in literature. The study of financial indicators is a very broad topic with an open gap: the company selection method and the use of other data sources besides the stock market history in order to optimize the risk-return function. Investors can benefit from this methodology when selecting the companies that compose their portfolios, adding a numerical analysis to their know-how. The investigation is performed incorporating the financial indicators as a process variable in the mixture design of experiments generated through Principal Component Analysis and Clusters. The green companies classified on Newsweek Magazine were selected as investigation objects. Firstly, an original classification was recreated using the general multivariate index. In the second phase, the balance sheet indicators were calculated and added to the model for cluster division. A risk-return analysis was made for comparison with a portfolio selection method using a mixture design of experiments (de Oliveira et al., 2011) that considers all the selected companies and is based on the portfolio selection proposed by (Markowitz, 1952). The conclusion show better stability and

profitability of the portfolios on a 5-year period and also verify that the proposed methodology is more consistent for company selection.

The balance sheets are a variable that influence portfolio composition and should be taken into consideration. Experimental arrangements were used to analyze portfolio formation using clusters. The investment simulation was used to benchmark the 5 portfolio methods analyzed:

1. CMO (Clustered Multilevel Optimization) – proposed method;
2. CM-DOE (Crossed Mixture Design of Experiment) – proposed by (Piepel, 1999);
3. Markowitz method (Markowitz, 1952);
4. EC-CMO (Excluded Companies from Clustered Multilevel Optimization) – since the CMO classifies the best companies for the investment, this method basically uses the excluded ones to compare results;
5. Equal Investment – this method basically sets the same investment in all companies.

The balance sheets are an important source of financial information and open capital companies are obliged by law to publicly release this data. One of the contributions of this study is to include indicators extracted from balance sheets as a process variable based on Design of Experiments (DOE) in portfolio selection models.

Two new methods for grouping companies considering risk-return were introduced in this article: placement on *Newsweek* magazine green company rank and financial indicators. A combination of PCA and Crossed Mixture Design of Experiment (CM-DOE) (Piepel, 1999) in order to generate optimal portfolios including green companies are also proposed as an opportunity to incorporate new tools into the conventional models of portfolio selection. Several financial risk ratios were used as basis for PCA, where companies were gathered into three clusters. The indicators were screen using the concept of data mining, and the balance sheet of each company was individually analyzed so that all the relevant information was extracted to generate the indicators.

2. BACKGROUND

2.1 Green companies.

Being green is a considerable task. (Saha and Darnton, 2005) listed some prime aspects for any company to be considered green: ecological concerns, conservation of the fauna and flora, corporative social responsibility, humanitarian concerns, fair trade, clean water, animal welfare, quality, and sustainability. *Newsweek's* magazine's Greenest Companies ratings is a ranking composed of energy productivity, carbon productivity, water productivity, waste productivity, and reputation. Those indicators have a range from 0.00% to 100.00% with a higher percentage representing better/greener score. The U.S. Department of Energy gives monetary motivation and awards to companies that help to recover the environment. The Energy Department invested more than US\$31 billion to support a wide range of clean energy projects across the nation according to the Clearinghouse of 2015.

In (Bansal, Pratima; Roth, 2000), a survey conducted among 53 companies showed several motivations for companies to be green; (Lyon and Shimshack, 2012) studied the impacts of media generated by companies sustainability programs. Using *Newsweek's* ratings of the 500 largest U.S. firms they found that the top 100 firms of the ranking have experienced abnormal returns of 0.6% to 1.0% higher than the returns of the bottom 400 on the ranking.

2.2 The financial ratios.

Financial ratios are an important quantitative way to measure company performance. Failing enterprises exhibit ratio measurements significantly different than those that are thriving, and these can all be described by five key ratios (Altman, 1968). The business risk is the risk that a company fails and it is divided into internal and external sources. The internal source is the general reflection of the variability of the net operating income or net cash flow. The external source is caused by the market which produces price variability for both outputs and inputs (Gabriel and Baker, 1980). In the United States of America, companies that have assets on the exchange market are regulated by the Securities Exchange Act (1934). Companies need to publish all their financial reports in a complete balance sheet. Most of those ratios are focused on the balance sheet and income statements. Cash flows can offer better insights from the ratio analysis when the period is considered continuously (Barua and Saha, 2015). The considered ratios for this study are listed in Table 1, adapted from (Barua and Saha, 2015). These indicators demonstrate the company's financial health.

Tab. 1: Considered ratios

Analysis of debt usage	
1	$\text{debt to total capital} = \frac{\text{total debt}}{\text{total capital}}$
2	$\text{debt to equity} = \frac{\text{total debt}}{\text{total equity}}$
Analysis of Interest Coverage Ratio	
3	$\text{times interest earned tax} = \frac{\text{earnings before interest and tax}}{\text{interest expense}}$
4	$\text{fixed charge coverage} = \frac{\text{earnings before charges and taxes}}{\text{fixed charges}}$
5	$\text{times interest earned cash} = \frac{\text{adjusted operating cash flow}}{\text{interest expense}}$
6	$\text{fixed charge coverage ratio} = \frac{\text{adjusted operating flow}}{\text{fixed charges}}$
7	$\text{capital expenditure ratio} = \frac{\text{cash flow from operations}}{\text{capital expenditures}}$
8	$\text{cash from operations to debt} = \frac{\text{cash flow from operations}}{\text{total debt}}$

2.3 Principal Component Analysis (PCA).

PCA is probably the best and most aged technique used to analyze multivariate problems (Jolliffe, 2002). (Hotelling, 1933) introduced the PCA method as a technique of the multivariate statistics. It is used to make inferences about the variance-covariance structure in a data set. This concept uses linear combinations of the original variables. There are two main objectives of PCA: Shrink the dimensionality of the dataset and provide interpretation of the data. The relationship between the components is revealed and is often an intermediary step within a broader data group (Johnson, 2002; Puntanen, 2013).

2.4 The mixture and crossed mixture design of experiments.

The experiment is the most efficient method of discovery. It is one of several major research strategies (Montgomery, 2012) and can provide a mechanism of action for a particular phenomenon. An alternative to experimentation is passive observation, where a scientist will observe a process in anticipation of an interesting change. This method can often lead to a large amount of data, where identification of relevant information may be difficult. Experimentation, on other hand, can require less effort and generate less non-relevant data (Andersson, 2012). The main difference is that a

passive observer just monitors a pre-existing process while the experimenter changes the state and investigates the results of these changes.

Mixture designs are experimental designs where the factors of the design are components or ingredients (Draper and Pukelsheim, 1998) of a mixture (Gozálvez-Zafrilla et al., 2013), i.e., different proportions of any component can change the overall result. Experiments with mixtures are those in which the property studied depends of the proportion of the components, but not on the amount of the mixture (Scheffé, 1963). The proportions are more important than the final quantity of the object of study.

Considering each proportion of the mixture as a weight w_i of q components yields the equation (1), which is related to the portfolio constraint of (Markowitz, 1952).

$$\sum_{i=1}^q w_i = 1.0 \quad (1a)$$

$$w_i \geq 0, \quad i = 1, 2, \dots, q \quad (1b)$$

Some of MDE applications can be listed starting with (Misturas, 2008), and (Scheffé, 1963) using the centroid or central point. (Brandvik and Daling, 1998) provided an optimization in product development. (Måge and Næs, 2005) used MDE with process variables or fractional design. There also several works using the original concept of MDE, i.e., using MDE to analyze and optimize the proportions of ingredients of products (Lonni et al., 2012; Ngun et al., 2014; Nikzade et al., 2012) with and without constraints.

MDE supports individual and global constraints. Considering the resolution of the mixture as V , with η being any integer natural number, Ψ the maximum value of the mixture and n the number of assets of the clusters or portfolios in the global mixture shown at the equation (2). When V decreases, the number of runs in the experiment increases.

$$V = \eta_i \therefore \Pi_{\eta_1} * V = \Psi \text{ and } \Pi_{\eta_2} * V = n \quad (2)$$

The crossed mixture design (CM-DOE) was introduced by combining two different mixtures to create a region of interest as a third mixture (Cornell, 1971), and after that as a two-stage mixture experiment (Cornell and Ramsey, 1997). It is also known as mixture of mixture design (MoM) (Borges et al., 2007; Di Zio et al., 2007; Kang et al., 2011; Piepel, 1999). There is one model of DOE that combines mixture with process variables. (Cornell, 1971) described that kind of model as a design of simplex-lattices and factorial arrangements. This case involved a mixture experiment with q components plus n process variables, where n can be any positive integer.

The optimal characteristic of a product can be found from one model that combines both process variables and component's proportions in the MDE. (Montgomery, 2012) has described the process component as new variables z considered in this case as the MGI (Multivariate General Index) variable. The terms that involve only the process variables are not included in the model. Both mixture of mixture and mixture with process variable methods were used to perform the analysis.

2.5 Portfolio selection.

Harry Markowitz introduced the portfolio selection model (Markowitz, 1952). The portfolio selection is an econometric tool designed to help the investor to decide how much of his wealth should be applied towards each investment. In the case approached by this study, the investments are assets of

the stock exchange. However, this can also be applied in different types of investments. (Trippi, 1989) used it for real estate investment; (Better and Glover, 2006) used portfolio in selection of projects; (Delarue et al., 2011) applied it to the electric sector. (Byers et al., 2015) also discussed the application of portfolio theory in a broader sense, including intellectual capital and resources of enterprises.

The risk of the investment using portfolio theory can be higher by only selecting one investment as proven by (Markowitz, 1952), who states that if an investor diversifies between two portfolios, the risk of the compound investment will be smaller than the risk of either single investment. There are many options to solve the portfolio optimization problem, e.g., isomean line (Markowitz, 1952), mean-Gini (Shalit and Yitzhaki, 1984), analytical method (Li and Ng, 2000), min-max rule (Yu et al., 2005), desirability function (de Oliveira et al., 2011), hybrid intelligent algorithm (Zhang et al., 2012) and so on. The choice of the optimization method depends on the analyst, but in all the cases, the goal is the same: a maximum return with the lowest possible risk. The objective in portfolio selection is to make tradeoff between return and risk.

3. METHOD

The Newsweek's magazine has a rank formed by 500 companies as Greenest Companies. The rank is composed of five indicators: energy productivity, carbon productivity, water productivity, waste productivity, and reputation, which are updated annually. The steps of the method are numerated to highlight the process.

1. The principal component analysis (PCA) is applied to the original indicators creating a new classification by the multivariate general index (MGI): the PCA was performed considering all the original indicators with three components to compute and to evaluate the correlation matrix. The MGI is then calculated using equation (4). A new ranking came up assuming the highest values for MGI and their corresponding companies (Papandrea et al., 2016). This phase changes the original order of the ranking, considering just the indicators that most impacts the PCA scores. The companies are now selected according to the MGI based on their sustainability indicators.

$$MGI = \sum W_{PCij} * Eigenvalue_{PCi} \quad (4)$$

2. Obtain the historical data of stock prices from the NASDAQ databank. The considered data comprehends the past 5 years of the selected companies. After downloading the data, each company had their monthly closing prices segregated for the risk calculations and the individual returns (Papandrea and Paiva, 2016a).
3. Download of Balance Sheets of the Annual Report: The same 20 companies have had their financial balance sheets downloaded from their official websites or from NASDAQ's webpage. These sheets had to be analyzed individually in order to extract the information needed to calculate the financial ratios based on balance sheets: the necessary data for the calculation of each of the indicators, described in Tab. 1, were recorded and calculated using those sheets (Papandrea Pedro and Anderson, 2016).
4. Risk (σ) and return (μ_i) calculation with replication: the risk and return of each company was calculated using the average and standard deviation of the standardized values from the equation (5). The same calculation was made to the original values and to the replicated values (Papandrea Pedro and Anderson, 2016). The monthly closing values of each asset of each company were replicated using the ARMA-GARCH model, with $lag = 1$ (Papandrea and Paiva, 2016b).

$$r_i = \frac{\log(P_i)}{\log(P_{i-1})} \quad (5)$$

5. Generation of clusters based on indicators calculated to create smaller mixes: the ratios debt to total capital, debt to equity, times interest earned, fixed-charge coverage, times interest earned - cash basis, fixed charge coverage ratio - cash basis, capital expenditure ratio and CFO to debt were grouped by Ward linkage method (Ward, 1963), Euclidean distance measure and related in three clusters (Papandrea and Paiva, 2016c).
6. PCA applied to financial indicators in order to define the process variables (-1) and (+1): the process from the PCA of item 1 was now done using the variables of the financial ratios from item 3, the MGI was also calculated in the same way.

The steps mentioned above describe the method for creation and selection of process control variables using balance sheet indexes. The consideration of these indexes is made possible on the portfolio indicator analysis through the usage of DOE and mixture process variables.

4. FINDINGS

PCA was applied on the Newsweek's Greenest Companies ratings; this is shown in Fig. 1 and detailed in Tab. 2. That ranking is based upon the following criteria in percent scale: energy productivity, carbon productivity, water productivity, waste productivity, reputation. From here, three groups were formed. The three scores of the PCs calculated and grouped by the MGI (20) lead to a new ranking. The (Papandrea et al., 2016) shows the Newsweek ranking versus the MGI ranking. Forty four companies were considered because they possess all the necessary data for analysis and for not having undergone mergers or closure in considered periods. The top twenty companies were extracted which are detailed in (Papandrea and Paiva, 2016c) with their NASDAQ codes. Fig. 1 shows also the principal components of: 1-Reputation, 2-Energy productivity, 3-Carbon productivity, 4-Water productivity, 5-Waste productivity. The five indicators were gathered in three groups. Of those groups, the twenty highest PCA scores were added to our selection, considering the numerical signal. The second stage was to research and calculate the financial ratios shown in Tab. 1.

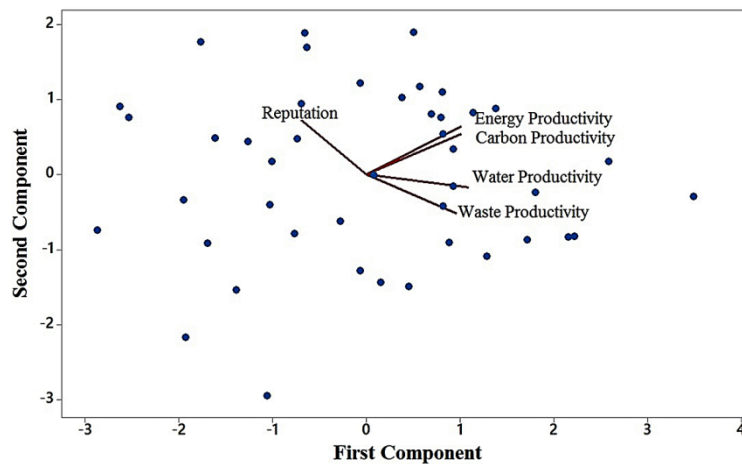


Fig. 1: PCA on *Newsweek* ranking

Tab. 2 Eigen analysis of the *Newsweek* Ranking

	Eigen analysis			
Eigenvalue	2.037	0.998	0.621	0.343
Proportion	0.509	0.250	0.155	0.086

Cumulative	0.509	0.759	0.914	1.000
Variable	PC1	PC2	PC3	
Energy Productivity	0.544	-0.463	-0.234	
Carbon Productivity	0.532	-0.415	0.482	
Water Productivity	0.506	0.387	-0.669	
Waste Productivity	0.406	0.681	0.515	
Reputation	-0.321	0.588	-0.572	

The reputation is out of the considerations because it resembled with no indicator, showing a negative PC value.

The analysis of the annual financial report of each company has been considered to calculate the financial ratios described in Tab. 1. After that, a PCA was applied to gather those ratios in four groups (Fig. 7: A-Debt to capital, B-Debt to equity, C-Fixed-charge covered ratio, D-Times interest earned-cash, E-Fixed-charge coverage, F-Times interest earned-tax, G-Capital expenditure ratio, H-Cash from operations to debt). Three main component scores were identified for each company (PCA1, PCA2, and PCA3). Each PCA score also has a related weight. The weightings for PCA1, PCA2, and PCA3 are $W_{PCA_1} = 0.575$, $W_{PCA_2} = 0.184$, and $W_{PCA_3} = 0.137$. A multivariate global index (MGI) was then calculated using these proportions. The Ward method was applied on the scores to divide the companies into. The MGI was used to classify each cluster in variable (-1), (+1) by the algorithm: if the company's MGI in the cluster is less than the average of the MGI of its cluster, the process variable is -1; otherwise, it is +1. MGI works as process variables in the crossed mixture. The (Papandrea Pedro and Anderson, 2016) have the financial ratios calculated.

The higher the indicators of debt (in millions) the lower will be the MGI while indicators of the interest coverage ratio will increase the MGI. This indicates that the more a company owes, the more opportunity for your MGI be increased. Fig. 2 shows a cluster grouping process by the variable MGI.

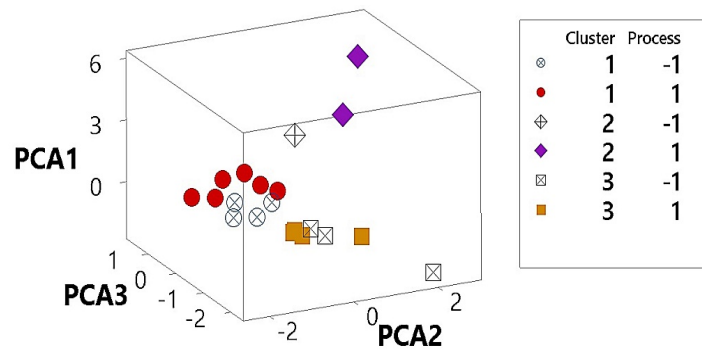


Fig. 2: Clusters and MGI process variables

The return and the risk of the company's asset are considered the average and the standard deviation (volatilities) of the monetary quote value of each closed month. Their values were approximated based on historical spanning 120 months as described in Section 3. The (Papandrea and Paiva, 2016c) shows the allocation of enterprises by cluster and within each cluster divisions. The first cluster indicate the companies that make up the process variables in each cluster by PCA; the second shows the companies that make up the process variables by cluster in the MGI. These divisions are the composition of portfolios. Instead of a random company selection performed by the investor, the cluster process selects which companies will be inside each group and the PCA defines which companies will compose the portfolio. The (Papandrea and Paiva, 2016c) also shows the division of the companies that make up each cluster divided into process variables, it includes the companies that

will compose the crossed mixture portfolios with its structure formed by portfolios of clusters, portfolios of companies with process variables. The CMO arrangement does not consider all companies, only those selected by the MGI values. The PCA values that are higher or lower than the average of MGI indicate the companies that will be selected by the process variables (-1) e (+1). The average values for the clusters are: *Cluster 1* = -0.10, *Cluster 2* = 2.56, and *Cluster 3* = -0.96. All types of design are described in (Papandrea and Paiva, 2016d).

The optimizations were performed using the desirability method. Firstly, each cluster has been optimized completely, considering all the companies that compose it, divided only by MGI, and considering it as process variables. Note that the cluster 2 with (-1) process variable was not optimized because it is composed of only one company. This is considered for the next optimization. Afterwards, the clusters comprise a new optimized portfolio, and in this, the returns and risks of each cluster previously optimized will be the input values of the next optimization. Next, the optimization of the crossed mixture model with process variables is made considering only a few companies in each cluster, depending on their MGI values. Finally a global optimization using the Markowitz model of portfolio, which consists of all companies in a same set. This optimization does not consider process variables, but the values of return and risk.

The Tab. 3 shows the result of each optimization. The lower is the level of individual clusters the highest will be the compound by the mixture of mixture.

Tab. 3 Results of optimization

Type	Companies/PCA/Cluster/MGI							Results		
Cluster1 -1	AAPL	IHS	SIAL	TAP				RP	SP	Desirability
γ_i^*	0.6910	0.0000	0.0000	0.3090				0.0194	0.0502	0.8257
Cluster1 +1	ADBE	ADI	BHI	MOS	QCOM	SPRINT		RP	SP	Desirability
γ_i^*	0.9060	0	0	0.094	0	0		0.0087	0.0632	0.9919
Cluster2 -1	EMC							RP	SP	Desirability
γ_i^*	1.0000							0.0055	0.0811	-
Cluster2 +1	BIIB	NKE						RP	SP	Desirability
γ_i^*	0.6010	0.3990						0.0173	0.0665	0.6784
Cluster3 -1	CLX	CMA	WYN					RP	SP	Desirability
γ_i^*	0.8930	0.0000	0.1070					0.0055	0.0500	0.7917
Cluster3 +1	BLL	CCE	MAS	MET				RP	SP	Desirability
γ_i^*	0.9930	0.0000	0.0000	0.0070				0.0112	0.0584	0.9938
CMO	Cluster1	Cluster2	Cluster3	MGI				RP	SP	Desirability
w_i^*	0.0000	0.7530	0.2470	+1				0.0157	0.0500	0.8690
CM-DOE	PCA1	PCA2	PCA3	Cluster1	Cluster2	Cluster3	MGI	RP	SP	Desirability
w_i^*	0.0000	0.8440	0.1560	0.0000	1.0000	0.0000	+1	0.0784	0.1092	0.9281
Markowitz	AAPL	ADBE	ADI	BHI	IHS	MOS	QCOM	RP	SP	Desirability
	0.0260	0.0430	0.0310	0.0000	0.0860	0.0450	0.0490			
	SPRINT	SIAL	TAP	BIIB	EMC	NKE	BLL			
w_i^*	0.0000	0.0720	0.0490	0.2090	0.0730	0.0970	0.1680	0.0123	0.0508	0.8305
	CCE	CLX	CMA	MAS	MET	WYN				
	0.0320	0.0000	0.0000	0.0130	0.0080	0.0000				

The desirability optimization graphic of the CMO method with clusters and financial ratios is shown in Fig. 4.

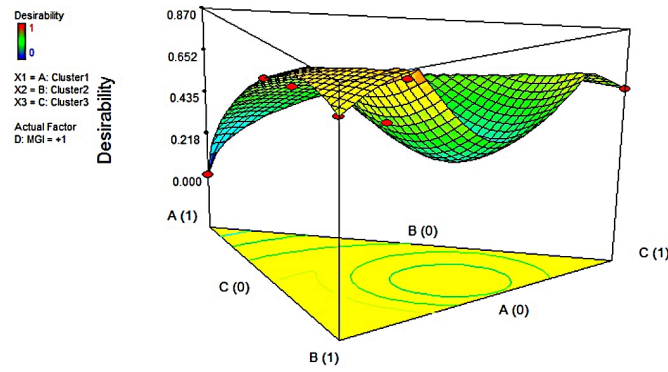


Fig. 4: 3D Desirability optimization of the CMO method with clusters and financial ratios

5. CONCLUSION

In this research, the proposed CMO method is exemplified and validated using real world data of Green Companies. Computational results show that it can help investors to analyze market exchange scenarios more robustly. The diversification seems better when the calculated ratios is considered from the annual financial reports.

The results shown in Table 4 demonstrate that there are considerable differences between the returns and risks of each model. Observing, the proposed method mitigates the risk of negative returns or losses, although the Markowitz model allows it. In the series where the Markowitz model is being used, there are negative values. Observing the CMO and CM-DOE series, in any case of the simulations, there are non-positive returns, meaning that in any case using the proposals the risk of capital loss exists, but only with reduced gain. This pioneering approach makes the proposed models more robust and viable. In the CMO model, the best portfolio will be clustered in a new one, and this new cluster will be optimized. In the other hand, on the CM-DOE model, there is the simultaneous optimization of mixtures of portfolios and companies considering their financial indicators. The graph shown in Fig. 5 illustrates the simulation of a US\$ 100,000.00 investment comparing the different methods. Observing the data provided by the simulation, it can be concluded that the best investment is the one which considers the CMO model with MGI process variables. The Markowitz method applied (de Oliveira et al., 2011) shows to be the second best method, followed by the CM-DOE. The portfolios of companies that were excluded by proposed method and an equal distributed capital portfolio (Equal Investment) is also shown.

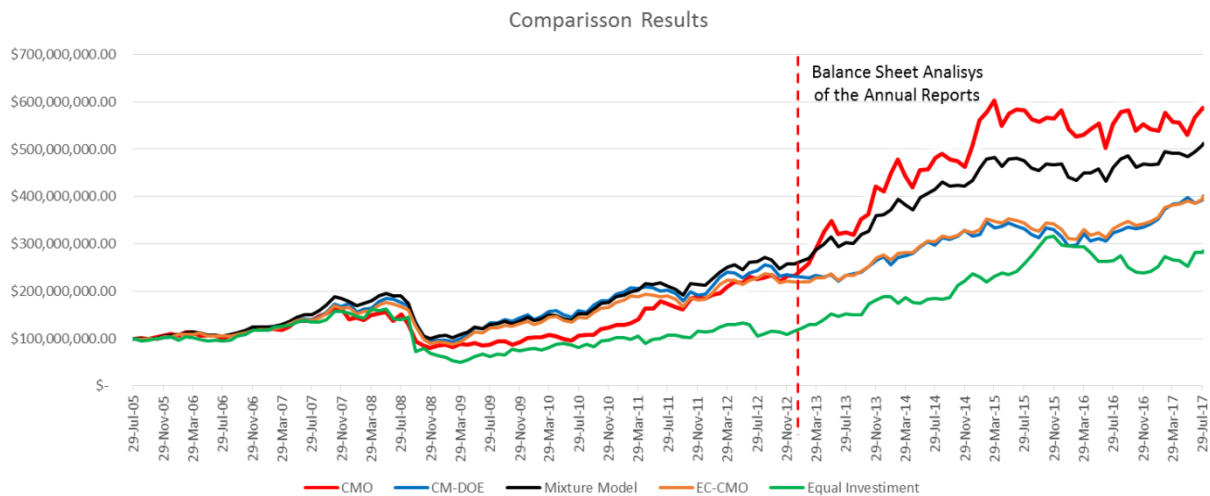


Fig. 5: Simulation result of US\$ 100,000.00 investment.

This is the first attempt in the literature to combine principal components analysis, mixture of mixture and crossed mixtures design of experiments into a single analysis tool that is capable of providing robust results with a minimal computational cost. For future work, the inclusion of several different years of the annual reports and comparing return-risk is proposed. Another proposal would be to use constraints at all levels, to increase the diversification. All the datasets are available at Mendeley Data according to the references.

Acknowledgment

This work was supported by the Brazilian agencies (CNPq, CAPES, project 12875-13-9).

References

- Altman, E.I., 1968. The Prediction of Corporate Bankruptcy: A Discriminant Analysis. *J. Finance* 23, 193–194. doi:10.1111/j.1540-6261.1968.tb00843.x/pdf
- Andersson, Ö., 2012. Experiment!: Planning, Implementing and Interpreting, Experiment!: Planning, Implementing and Interpreting. doi:10.1002/9781118311059
- Bansal, Pratima; Roth, K., 2000. Why Companies Go Green : Responsiveness. *Acad. Manag.* 43, 717–736. doi:10.2307/1556363
- Barua, S., Saha, A.K., 2015. Traditional Ratios vs . Cash Flow based Ratios : Which One is Better Performance Indicator ? *Adv. Econ. Bus.* 3, 232–251. doi:10.13189/aeb.2015.030605
- Better, M., Glover, F., 2006. Selecting Project Portfolios by Optimizing Simulations. *Eng. Econ.* 51, 81–97. doi:10.1080/00137910600695593
- Borges, C.N., Bruns, R.E., Almeida, A. a., Scarminio, I.S., 2007. Mixture-mixture design for the fingerprint optimization of chromatographic mobile phases and extraction solutions for *Camellia sinensis*. *Anal. Chim. Acta.* doi:10.1016/j.aca.2007.02.067
- Brandvik, P.J., Daling, P.S., 1998. Optimisation of oil spill dispersant composition by mixture design and response surface methods. *Chemom. Intell. Lab. Syst.* 42, 63–72. doi:10.1016/S0169-7439(98)00009-4
- Byers, S.S., Groth, J.C., Sakao, T., 2015. Using portfolio theory to improve resource efficiency of invested capital. *J. Clean. Prod.* 98, 156–165. doi:10.1016/j.jclepro.2013.11.014
- Cornell, J.A., Ramsey, P.J., 1997. Modeling the Component Linear and Nonlinear Blending Properties in a Two-Stage Mixture Experiment. *Nonlinear Anal. Theory, Methods Appl.* 30, 4041–4050.
- Cornell, J. a., 1971. Process Variables in the Mixture Problem for Categorized Components. *J. Am. Stat. Assoc.* 66, 42. doi:10.2307/2284844
- de Oliveira, F.A., de Paiva, A.P., Lima, J.W.M., Balestrassi, P.P., Mendes, R.R.A., 2011. Portfolio optimization using Mixture Design of Experiments: Scheduling trades within electricity markets. *Energy Econ.* 33, 24–32. doi:10.1016/j.eneco.2010.09.008
- Delarue, E., De Jonghe, C., Belmans, R., D’haeseleer, W., 2011. Applying portfolio theory to the electricity sector: Energy versus power. *Energy Econ.* 33, 12–23. doi:10.1016/j.eneco.2010.05.003

- Di Zio, M., Guarnera, U., Rocci, R., 2007. A mixture of mixture models for a classification problem: The unity measure error. *Comput. Stat. Data Anal.* 51, 2573–2585. doi:10.1016/j.csda.2006.01.001
- Draper, N.R., Pukelsheim, F., 1998. Mixture models based on homogeneous polynomials. *J. Stat. Plan. Inference* 71, 303–311. doi:10.1016/S0378-3758(98)00012-3
- Gabriel, S.C., Baker, C.B., 1980. Concepts of Business and Financial Risk. *Am. J. Agric. Econ.* 62, 560–564. doi:10.2307/1240215
- Gozálvez-Zafrilla, J.M., Santafé-Moros, a., García-Díaz, J.C., 2013. Crossed mixture-process design approach to model nanofiltration rejection for non-dilute multi-ionic solutions in a given range of solution compositions. *Desalination* 315, 61–69. doi:10.1016/j.desal.2012.08.009
- Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24, 417–441. doi:10.1037/h0071325
- Johnson, M., 2002. Waveform based clustering and classification of AE transients in composite laminates using principal component analysis. *NDT E Int.* 35, 367–376. doi:10.1016/S0963-8695(02)00004-X
- Jolliffe, I.T., 2002. Principal Component Analysis, Second Edition, *Encyclopedia of Statistics in Behavioral Science*. doi:10.2307/1270093
- Kang, L., Joseph, V.R., Brenneman, W. a, 2011. Design and Analysis of Mixture-of-Mixture Experiments. *Technometrics* 53, 125–136.
- Li, D., Ng, W.-L., 2000. Optimal Dynamic Portfolio Selection: Multiperiod Mean-Variance Formulation. *Math. Finance* 10, 387–406. doi:10.1111/1467-9965.00100
- Lonni, A.A.S.G., Longhini, R., Lopes, G.C., De Mello, J.C.P., Scarminio, I.S., 2012. Statistical mixture design selective extraction of compounds with antioxidant activity and total polyphenol content from *Trichilia catigua*. *Anal. Chim. Acta* 719, 57–60. doi:10.1016/j.aca.2011.12.053
- Lyon, T., Shimshack, J., 2012. Environmental Disclosure: Evidence From Newsweek's Green Companies Rankings, *Business & Society*. doi:10.1177/0007650312439701
- Måge, I., Næs, T., 2005. Split-plot design for mixture experiments with process variables: A comparison of design strategies. *Chemom. Intell. Lab. Syst.* 78, 81–95. doi:10.1016/j.chemolab.2004.12.010
- Markowitz, H., 1952. Portfolio Selection*. *J. Finance* 7, 77–91. doi:10.1111/j.1540-6261.1952.tb01525.x
- Montgomery, D.C., 2012. *Design and Analysis of Experiments*, eight. ed. John Wiley & Sons, Inc., New York, United States of America.
- Ngun, B.K., Mohamad, H., Katsumata, K.I., Okada, K., Ahmad, Z.A., 2014. Using design of mixture experiments to optimize triaxial ceramic tile compositions incorporating Cambodian clays. *Appl. Clay Sci.* 87, 97–107. doi:10.1016/j.clay.2013.11.037
- Nikzade, V., Tehrani, M.M., Saadatmand-Tarzan, M., 2012. Optimization of low-cholesterol-low-fat mayonnaise formulation: Effect of using soy milk and some stabilizer by a mixture design approach. *Food Hydrocoll.* 28, 344–352. doi:10.1016/j.foodhyd.2011.12.023
- Papandrea, P., Paiva, A., 2016a. Monthly quotation of twenty green companies. *Mendeley Data*. doi:10.17632/ccpzrpz4xk.2

- Papandrea, P., Paiva, A., 2016b. Replication of the monthly quotation. doi:10.17632/vjsmh4b7zt.2
- Papandrea, P., Paiva, A., 2016c. Functions and optimization of portfolio of Green companies. doi:10.17632/ggyzww74b2.2
- Papandrea, P., Paiva, A., 2016d. Design of experiment sheets of Green Companies. doi:10.17632/9kvzx7k5yh.2
- Papandrea, P., Paiva, A., Leme, R., 2016. The PCA applied into the Newsweek Green Companies Ranking of 2014. Mendeley Data 1. doi:10.17632/CPVR7GDY9R.1
- Papandrea Pedro, Anderson, P., 2016. PCA applied on calculated financial indicators and quotations analysis. doi:10.17632/s2kny3549v.2
- Piepel, G.F., 1999. Modeling methods for mixture-of-mixtures experiments applied to a tablet formulation problem. *Pharm. Dev. Technol.* 4, 593–606. doi:10.1081/PDT-100101398
- Puntanen, S., 2013. Methods of Multivariate Analysis, Third Edition by Alvin C. Rencher, William F. Christensen. *Int. Stat. Rev.* 81, 328–329. doi:10.1111/insr.12020_20
- Saha, M., Darnton, G., 2005. Green Companies or Green Con-panies: Are Companies Really Green, or Are They Pretending to Be? *Bus. Soc. Rev.* 110, 117–157. doi:10.1111/j.0045-3609.2005.00007.x
- Scheffé, H., 1963. The Simplex-Centroid Design for Experiments with Mixtures. *J. R. Stat. Soc.* 25, 235–263.
- Shalit, H., Yitzhaki, S., 1984. Mean-Gini, Portfolio Theory, and the Pricing of Risky Assets. *J. Finance* 39, 1449–1468. doi:10.2307/2327737
- Trippi, R.R., 1989. A decision support system for real estate investment portfolio management. *Inf. Manag.* 16, 47–54. doi:10.1016/0378-7206(89)90026-8
- Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* doi:10.1080/01621459.1963.10500845
- Yu, M., Wang, S., Lai, K.K., Chao, X., 2005. Multiperiod Portfolio Selection on a Minimax Rule 12, 565–587.
- Zhang, W.G., Liu, Y.J., Xu, W.J., 2012. A possibilistic mean-semivariance-entropy model for multi-period portfolio selection with transaction costs. *Eur. J. Oper. Res.* 222, 341–349. doi:10.1016/j.ejor.2012.04.023

ANEXO B – ISERC 2015 - CROSSED MIXTURE DESIGN OF EXPERIMENTS APPLIED ON MULTI-PERIOD PORTFOLIO

IE ANNUAL CONFERENCE & EXPO | 2015



[Author Home \(/iie2015/abstract/author_home.cfm\)](/iie2015/abstract/author_home.cfm)

[Update Profile \(/iie2015/abstract/update_profile.cfm\)](/iie2015/abstract/update_profile.cfm)

[Logout \(/iie2015/logout.cfm\)](/iie2015/logout.cfm)

[Help \(https://www.xcdsystem.com/help/index.cfm\)](https://www.xcdsystem.com/help/index.cfm)

ISERC 2015 Abstract Submission

You are now editing Abstract: I661

[Return to Author Home](#)

Abstract Status

Abstract ID **I661**

Final Status **Abstract Accepted**

Thank you for your submission. We look forward to receiving your paper.

ISERC 2015 Abstract Acceptance

Dear [Firstname],

Abstract ID: [ID]

Abstract Title: [AbTitle]

Dear Colleague,

Congratulations, the abstract that you submitted for the ISERC 2015 Conference has been accepted! We look forward to a high-quality program of research presentations this year.

In this e-mail, you will find: (1) deadlines, (2) important notes, and (3) full paper submission guidelines. Although paper submission is option, we strongly urge you to submit a paper to the conference.

Deadlines - Please note that these dates will be strictly enforced.

2015 Deadlines

Submission of papers (optional but encouraged)	January 30, 2015
Notification of acceptance/declines of papers	February 27, 2015
Final paper due	March 13, 2015
Presenting author registration due	March 22, 2015

Please note the following points:

1. To make the ISERC 2015 Conference a high-quality event, the presenting author must register for the conference by March 22, 2015. Those abstracts that do not have the presenting author registered for the conference by March 22, 2015 will be removed from the conference program and proceedings by the web management system. If the presenting author cannot register by March 22, 2015, a co-author may substitute as the presenting author to present the paper instead and register for the conference. In this case, you must notify the ISERC program co-chairs via email about the change in the presenting author. An individual can be the presenting author for at most two submissions in ISERC 2015.

2. Although paper submission is optional, you are strongly encouraged to submit a full paper. All papers will be double-blind refereed, and if accepted, will appear in the conference proceedings, which are indexed by ProQuest. The deadline for full paper submissions is January 30, 2015, and this deadline will be strictly enforced.

3. Please understand that a full paper submission also implies that you are committing to act as a reviewer for other submitted papers and that you must meet all review deadlines. You will also be required to complete and return the permission to publish form which will be available at <https://www.xcdsystem.com/iie2015/abstract/index.cfm?ID=A1NlbS3> (<https://www.xcdsystem.com/iie2015/abstract/index.cfm?ID=A1NlbS3>).

4. Papers will be reviewed by two reviewers in a double-blind process and will be rated as either 1) Accept as is, 2) Accept with minor revisions, or 3) Reject. We have time for only one review cycle, so any paper that requires more than minor revisions (based on either technical content or writing style and grammar) will be rejected. In the event the paper is rejected, the authors may still present their work at the conference.

All paper submissions and reviews are being handled electronically through the following web site <https://www.xcdsystem.com/iie2015/index.cfm>

ANEXO C – RISCO: MEDIDAS DE RISCO, AVERSÃO E SEU GERENCIAMENTO

RISCO: MEDIDAS DE RISCO, AVERSÃO E SEU GERENCIAMENTO

Pedro José Papandréa

pedro.papandrea@gmail.com

Rafael Coradi Leme

leme@unifei.edu.br

[Catarine Conceição Moura Tenório](#)

catarinemoura_28@hotmail.com

Para referência: Papandrea P.J., Leme R.C., Tenório C.C.M., 2013. MEDIDAS DE RISCO, AVERSÃO E SEU GERENCIAMENTO.

Resumo:

Para um investidor, o risco é a variável que pode ser mais importante na escolha de um investimento, seja ele financeiro, físico ou em qualquer outro campo econômico. A fundamentação teórica é a base para que haja o conhecimento das medidas de risco, conceito de aversão ao risco e utilidade e o gerenciamento, além de possibilitar o conhecimento de técnicas para redução do risco pela aplicação de portfólios. O objetivo é expor ao investidor as variáveis que afetarão as suas decisões e os cálculos que pode ser feitos para a mensuração destas variáveis.

Palavras-Chave: Risco, Medição, Aversão, Gerenciamento.

1. INTRODUÇÃO

As medidas de risco são indicadores que contribuem para a tomada de decisão, principalmente quanto a investimentos. Para essa avaliação o horizonte de tempo é importante. Em Lien et al. (2007) as decisões de gestão têm um longo horizonte de tempo com risco significativo. Normalmente, o horizonte de tempo para esse tipo de investimento, por exemplo, florestal, é de 50-120 anos, envolvendo uma considerável incerteza sobre uma eventual produção e preços. Por outro lado há a aversão ao risco do investidor. O efeito da aversão ao risco pode ser analisado como uma decisão tática, ou

seja, investir agora ou esperar, ou a nível mais estratégico de uma decisão. Lambert (1972) diz que a probabilidade de sucesso de um evento arriscado utilizados em experimentos sobre comparação entre riscos assumidos pelos individualmente e em grupos terá um impacto sobre as conclusões relativas ao risco da transação.

As métricas de risco podem então serem expressas em funções probabilísticas dadas pelos eventos relativo sendo $1/n$ com base em uma melhoria na homogeneidade da variância:

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} \quad (1.1)$$

Na realidade, o problema subjacente é muito mais importante e envolve determinar quais valores relativos, tais valores se atribuem para as diferentes probabilidades sugeridas para eles.

2. ORIGEM DAS MEDIDAS DE RISCO

Na pré-época Markowitz, risco financeiro foi considerado como um fator de correção de retorno esperado e risco ajustado dos retornos. Foram definidos numa base para esta finalidade. Estas medidas primitivas tinham a vantagem de permitir uma imediata ordem preferencial de todos os investimentos.

Markowitz (1952) propôs medir o risco associado ao retorno de cada investimento, com uso de um desvio da média da distribuição, retorno a variância, e no caso de uma combinação (portfólio) de ativos, a fim de avaliar o nível de risco pela covariância entre todos os pares de investimentos, já que a covariância é medida 2 a 2:

$$Cov[X, Y] = E[XY] - E[X]E[Y] \quad (2.1)$$

Na qual X e Y são retornos aleatórios.

Temos que lembrar que o modelo de Markowitz é relacionado com as funções de utilidade, que permitem uma ordenação nos investimentos com as suas combinações. No caso de não normal, embora as distribuições sejam simétricas, as funções de utilidade devem ser quadráticas. Na prática, essa limitação restringe o uso deste modelo de carteiras caracterizadas por conjuntos de regressão da distribuição normal, por exemplo,

para o caso em que o retorno de todos os ativos, bem como a sua estrutura de dependência seja normal (Szegö G., 2005).

Em teoria econômica, denomina-se utilidade, a propriedade que os produtos tangíveis e serviços têm de satisfazer as necessidades e desejos humanos. Os objetos que têm utilidade são considerados bens, do ponto-de-vista econômico. A caracterização dos bens como econômicos, requer também que os mesmos sejam escassos, isso é: estejam disponíveis em quantidades limitadas.

Recentemente, a classe de variáveis aleatórias cuja correlação é linear pode ser usada como uma medida de dependência que foi totalmente identificada. Essa é a classe de caracterizados por distribuições elípticas e tem a propriedade de que as suas superfícies são de equidensidades elipsoides. Assim, o modelo de Markowitz é adequado apenas para o caso de distribuições elípticas, como normal ou distribuição t com variâncias finitas. Note-se que distribuições simétricas não são necessariamente elípticas. O coeficiente de correlação linear, se utilizado no caso de distribuições não elípticas, pode conduzir a resultados incorretos (Szegö G., 2005).

Szegö G. (2005) finalmente, destaca historicamente: Nos anos 60 o conceito de β (volatilidade) foi introduzido. Este desenvolvimento foi motivado pela falta de recursos computacionais. A complexidade da abordagem de média-variância foi considerada muito alta. Depois de quase 40 anos e o progresso gigantesco dos computadores, este não é mais o caso. A segunda motivação para a introdução do β baseado em métodos de portfólio foram os dados insuficientes para calcular a matriz de variância-covariância: como de fato, o número de dados deve ser de pelo menos duas vezes o número de ativos, hoje em dia as técnicas de bootstrapping permitem contornar este problema e os β s são quase abandonados na carteira em favor da função completa dos modelos de variância-covariância.

O nome bootstrap é uma referência à história do Barão Von Münchhausee, o qual se livrou de um pântano por seus próprios esforços, fazendo alusão pelo método ser uma reamos-tragem gerada a partir dos próprios dados de uma amostra. Usando técnicas clássicas, não é fácil obter erros padrão ou intervalos de confiança para os parâmetros de uma série. Usando o *bootstrap*, no entanto, eles são facilmente encontrados. São criadas novas amostras, a partir de um conjunto de dados originais, por reamostragem

dos dados com reposição de valores. Ao realizar este procedimento de reamostragem muitas vezes, uma boa estimativa pode ser obtida da distribuição das estatísticas de interesse. Tais distribuições podem ser vistas como uma aproximação às verdadeiras distribuições dos estimadores e, portanto, estatísticas de interesse, (WEHRENS, 2000).

É importante entender as diferenças entre o risco sistemático e o risco não sistemático. O risco não sistemático é o risco do portfólio em si, e será considerado neste trabalho, enquanto que o risco não sistemático, não diversificável ou risco de mercado é aquele relacionado às flutuações do sistema econômico como um todo e não é abordado na teoria de portfólio.

3. MEDIÇÃO DO RISCO

Uma medida de risco designa um número real de uma variável aleatória. Costa e Araújo (2008) utilizaram o trabalho de Li e NG (2000) para melhor responder aos movimentos drásticos do mercado, como resultado de situações de mudanças devidas as descontinuidades de fatores externos. Formulando um modelo no qual o portfólio pode ser reavaliado de período em período e seus parâmetros possam ser modificados a contento.

Podemos definir o risco em termos de mudanças de valores entre duas datas, defendemos que, por causa do risco estar relacionado à variabilidade do valor futuro de uma posição, devido às mudanças do mercado, ou, mais geralmente a eventos incertos, é melhor considerar apenas os valores futuros.

Considere um conjunto com vetor de realização $V \subseteq \mathbb{R}^S$, no qual S indica o número de estados da natureza. Esse estado ocorre com probabilidade $p_s > 0$ e $\sum_{s=1}^S p_s = 1$. O vetor $X \in V$ representa o lucro possível da carteira (a empresa de, a companhia de seguros, banco, etc.) e perdas das realizações em um futuro comum escolhido no horizonte de tempo, digamos, $t = 1$. A quantidade X_s é o retorno da carteira no estado de natureza s . Valores negativos de X_s correspondem às perdas. A desigualdade $Y \geq X$ significa que $Y_s \geq X_s$ para todos $s = 1, \dots, S$. Definimos $R_+ = [0, \infty)$, $R_{++} = (0, \infty)$, $R_- = (-\infty, 0]$, respectivamente. A variável

aleatória discreta gerada pelo $p \in R_{++}^S$ e $X \in V$ é denotada por \hat{X} , isto é $P(\hat{X} = X_s) = p_s$, para todo $s = 1, \dots, S$.

A medida de risco é então uma função $\rho: V \rightarrow \mathbb{R}$, medindo o risco de uma carteira usando a perspectiva do presente com $(t = 0)$. Isto é a quantidade mínima de dinheiro que o agente regulador tem que adicionar ao seu portfolio e investir em um instrumento de referência hoje, de tal forma que garanta que o risco envolvido na carteira seja aceitável para o regulador. Assumimos que o instrumento de referência tem uma recompensa em cada estado da natureza em $t = 1$, assim, a sua realização é um vetor $\mathbf{1} = (1, \dots, 1) \in V$. O instrumento de referência é sem risco no “sentido clássico”, não tendo nenhuma incerteza em seus retornos. É mais natural que se pense nisso como uma obrigação recompensa igual à zero. O preço do instrumento de referência, os fatores de desconto são denotados por $\delta \in \mathbb{R}_+$. Essas medidas podem ser ajustadas para um domínio generalizado de V (CSÓKA, 2007).

4. AVERSÃO AO RISCO

A crença comum entre os economistas é que os seres humanos são, ou deveriam ser avessos ao risco. A introspecção revela que isto é muito plausível em muitas situações, apesar de comportamento de procura de risco que muitas vezes tem sido observado em estudos experimentais da escolha humana.

Aversão ao risco é uma conclusão necessária se as pessoas enfrentam utilidade marginal decrescente, um pressuposto para o qual há um forte apelo intuitivo. Isso aumentou o apelo de aversão ao risco como a suposição padrão de comportamento humano (von NEUMANN, 1947). Trata-se da preempção do investidor de arriscar-se em um determinado negócio, sendo que a atitude de risco é afetada pela aversão à perda possibilidade de ganhos e distorções na decisão, bem como a curvatura utilidade para ambos os ganhos e perdas (DAVIES E SATCHELL, 2007).

5. GERENCIAMENTO DO RISCO

A gestão de riscos é de fundamental importância, considerando o enorme risco financeiro que a economia está exposta. Os riscos de muitos agentes econômicos são regulados por diversas instituições. Por exemplo, se um comerciante financeiro quer vender opções, que dão direitos ao comprador de comprar ou vender a um determinado preço durante um horizonte de tempo especificado (ou em um determinado momento), ele tem de cumprir os requisitos de margem, ou seja, ele tem que depositar algum dinheiro ou algum outro sem risco e instrumento líquido. Empresa de mediação de uma bolsa de valores, que é responsável pelas promessas a todas as partes das transações de forma segura, sendo concluído, exige depósitos de margem. Uma medida de risco pode ser usada para determinar o requisito de margem. Quanto maior risco da carteira do comerciante, maior a margem exigida deve ser (Csóka et al., 2007).

Travers (2004) classifica risco em função do retorno conforme **Figura 5.1**:

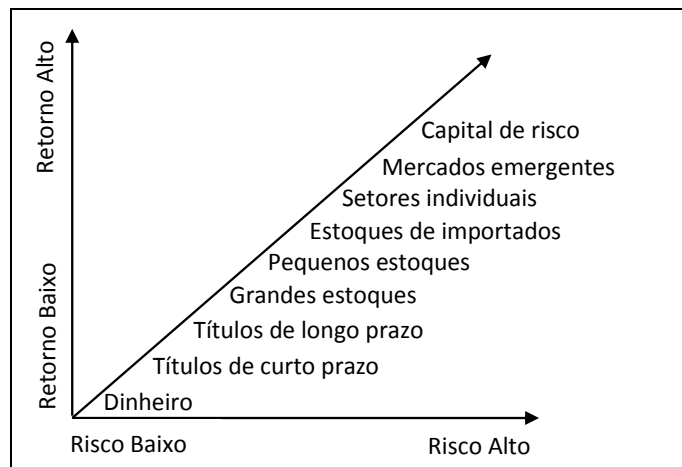


Figura 5.1 – Risco em Função do Retorno

Fonte: TRAVERS, 2004.

A Teoria de Portfólio permite gerenciar o risco em um grupo de ativos para determinar uma combinação que ofereça o menor risco e o maior retorno esperado. Esse grupo é chamado de portfólio ótimo. O portfólio de ativos é uma combinação de n ativos de interesse, cada um tendo retorno r_i , ($i = 1, 2, \dots, n$), o retorno do portfólio,

doravante denotado por r_p , é a média ponderada do retorno do ativo componente, com as proporções de investimentos como w_i sendo os pesos. Organizando r_i e w_i em vetores colunas $r^T = [r_1 \ r_2 \ \dots \ r_n]$ e $w^T = [w_1 \ w_2 \ \dots \ w_n]$, a equação, r_p do portfólio de retorno pode ser definida como:

$$r_p = w^T \cdot r \quad (5.1)$$

A média da distribuição de probabilidade do retorno, ou o rendimento esperado, é uma indicação da rentabilidade esperada. A variância da distribuição indica a amplitude dos resultados possíveis em torno da média, isto é, quanto maior for a variância, o mais incerto é o resultado. Assim, a variância da distribuição é uma indicação intuitiva do risco envolvido.

Se a taxa de retorno evolui de maneira pelo menos fracamente estacionária ao longo do tempo, cada retorno do ativo pode ser representado por um processo estocástico com retorno esperado $\mu_i = E[r_i]$ e a variância $\sigma_i^2 = E[r_i^2]$. Além disso, considerando a covariância entre ativos e $\sigma_{ij} = E[r_i r_j]$, como o retorno esperado do portfólio e sua variância podem ser escritos como:

$$\mu_p = w^T \cdot \mu \quad (5.2)$$

$$\sigma_p^2 = w^T \Sigma w \quad (5.3)$$

no qual $\mu^T = [\mu_1 \ \mu_2 \ \dots \ \mu_n]$, e Σ é a matriz de covariância, com diagonal contendo σ_i^2 , e fora da diagonal com σ_{ij} . Note-se que o σ_{ij} da covariância mede quantos dos retornos de dois ativos se movem em relação uns aos outros. Este é o conhecido modelo abordado por Markowitz: Média-Variância do Portfólio (MVP), estabelecendo a melhor estratégia para minimizar o risco e maximizar o retorno. Ao fazer isso, atinge-se a fronteira eficiente, na qual para um dado nível de variância, não existe outro portfólio com maior retorno esperado. Da mesma forma, para um dado nível de retorno esperado, não existe qualquer outro portfólio com menor variância. A **Figura 5.2** mostra a fronteira eficiente.

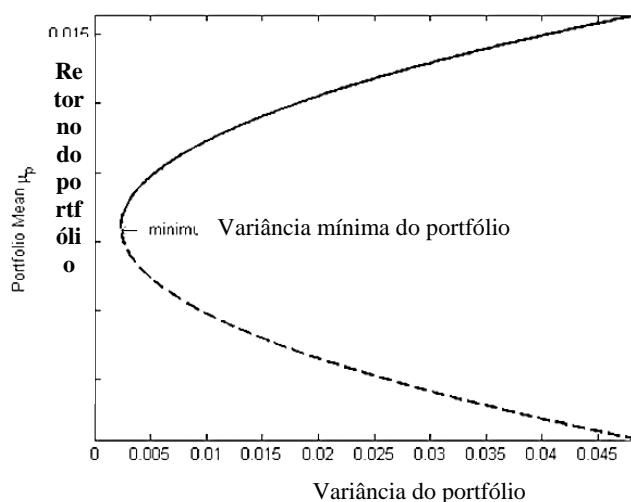


Figura 5.2 - Fronteira eficiente média-variância

6. CONCLUSÃO

Esta revisão possibilitou o entendimento de questões fundamentais sobre risco como variável de decisão para investimentos diversos. A gestão de riscos deve ser um processo contínuo - a análise de risco por si só não é suficiente. As técnicas de portfólio são uma ótima opção para redução do risco envolvido no investimento e é também um bom mecanismo de gerenciamento, o qual possibilita que o investidor mais ousado consiga ter um indicador de seus possíveis ganhos e idem para aquele mais conservador. O risco é por si só um agente de decisão, mas foi evidenciado que há fundamentação para ajuste do valor desse agente no âmbito do conceito de utilidade. É importante que haja aprofundamento dos conceitos e variáveis de risco pelo investidor.

7. REFERENCIAS BIBLIOGRÁFICAS

- Artzner P., Delbaen F., Eber J., Haeth D., 1999. Coherent measures of risk. *Mathematical Finance* 9 (3), 203–228.
- Csóka P., Herings P.JJ., Kóczy L.Á., 2007. Coherent measures of risk from a general equilibrium perspective / *Journal of Banking & Finance* 31 (2007) 2517–2534.
- Davies G.B., Satchell E.S., 2007. The behavioural components of risk aversion / *Journal of Mathematical Psychology* 51 (2007) 1–13.

- Eisenhauer J.G., 2006. Risk aversion and prudence in the large / *Research in Economics* 60 (2006) 179–187.
- Kreiss J.P., Paparoditis E. 2011. Bootstrap methods for dependent data: A review / *Journal of the Korean Statistical Society* 40 (2011) 357–378.
- Lambert R., 1972. Risky Shift in Relation to Choice of Metric / *JOURNAL OF EXPERIMENTAL SOCIAL PSYCHOLOGY* 8, 315-318 (1972).
- Li D., NG W.L., 2000. Optimal dynamic portfolio selection: multiperiod mean-variance formulation. *Mathematical Finance*, Vol. 10, No. 3, 387–406.
- Lien G., Størdal S., Hardaker J.B., Asheim L.J., 2006. Risk aversion and optimal forest replanting: A stochastic efficiency study / *European Journal of operational Research* 181 (2007) 1584–1592.
- Markowitz H., 1952. Portfolio selection / *Journal of Finance* 7, 77–91.
- Markowitz H., 1959. *Portfolio Selection: Efficient Diversification of Investments* / Wiley, New York.
- Prem K.P., Ng D. Pasman H.J., Sawyer M., Guo Y., Mannan M.S. Risk measures constituting a risk metrics which enables improved decision making: Value-at-Risk / *Journal of Loss Prevention in the Process Industries* 23 (2010) 211–219.
- Stoica G., 2006. Relevant coherent measures of risk / *Journal of Mathematical Economics* 42 (2006) 794–806.
- Szegö G., 2005. Measures of risk / *European Journal of Operational Research* 163 (2005) 5–19.
- Travers F. J., 2004. *Investment Manager Analysis*. Wiley Finance, Investment Manager Analysis.
- Von Neumann, J., & Morgenstern, O. (1947). *The theory of games and economic behaviour*. Princeton: Princeton University Press.
- Wehrens R., et al., 2000. *The bootstrap: a tutorial*. Chemometrics and Intelligent Laboratory Systems 54, 35–52.
- Williams T., 1995. A classified bibliography of recent research relating to project risk management / *European Journal of Operational Research* 85 (1995) 18-38.