



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE QUÍMICA

ANTÔNIO RAFAEL DE SOUZA ALVES BÔSSO

Desenvolvimento do *Software* PlanEx de Planejamento de
Experimentos *Online* e sua Aplicação Didática na Pós-Graduação

Goiânia – GO, 2012

ANTÔNIO RAFAEL DE SOUZA ALVES BÔSSO

Desenvolvimento do *Software* PlanEx de Planejamento de Experimentos *Online* e sua Aplicação Didática na Pós-Graduação

Tese apresentada ao Instituto de Química da Universidade Federal de Goiás para obtenção do título de Doutor em Química

Área de Concentração:
Química do Cerrado e do Pantanal

Orientador: Prof. Dr. Anselmo Elcana de Oliveira

Co-Orientador: Prof. Dr. Márlon Herbert Flora Barbosa Soares

Goiânia – GO, 2012

Dados Internacionais de Catalogação na Publicação (CIP)

GPT/BC/UFG

B916d Bôsson, Antônio Rafael de Souza Alves.
Desenvolvimento do *Software* PlanEx de planejamento de experimentos *online* e sua aplicação didática na pós-Graduação [manuscrito] / Antônio Rafael de Souza Alves Bôsson. – 2012.
172 f.: figs, tabs.

Orientador: Prof. Dr. Anselmo Elcana de Oliveira; Co-Orientador: Prof. Dr. Márlon Herbert Flora Barbosa Soares .
Tese (Doutorado) – Universidade Federal de Goiás, Instituto de Química, 2012.
Bibliografia.

1. *Software* PlanEx – Planejamento fatorial. 2. PlanEx – *Software* livre. 3. Metodologia de Superfície de Resposta – *Software* livre. I. Título.

CDU: 004.43

Dedico essa tese:

À minha esposa, Eni Pereira Vieira Bôso, com amor, admiração e gratidão por sua compreensão, carinho, presença e incansável apoio ao longo do período de elaboração deste trabalho.

Aos meus Pais, Nivaldo Alves Bôso e Dulcinéa de Souza Neves Bôso pelo apoio incondicional aos estudos, pelo amor e carinho, e, sobretudo pelo que representam para mim.

Aos meus parentes que não fazem parte desse mundo e sempre almejaram para que um dia eu realizasse meu sonho de chegar a um nível mais elevado dos estudos, especialmente meu Pai (*in memorian*), que foi peça fundamental de incentivo para eu fazer a seleção do doutorado e não desistir dos sonhos, meus Avós (*in memorian*) e Titio Adalberto (*in memorian*), que representa para mim fonte de incentivo de profissão.

AGRADECIMENTOS

À Deus pela fonte da vida, sabedoria, humildade e perseverança para concluir este trabalho.

Ao Professor Dr. Anselmo Elcana de Oliveira pela oportunidade de trabalho, pela dedicação e paciência na orientação, pois tive que vencer muitas barreiras para concluir este trabalho, as quais sempre teve participação ativa em ajudar-me.

Ao Professor Dr. Márlon Herbert Flora Barbosa Soares pelas orientações durante a segunda parte do trabalho, validação do *software* como ferramenta de ensino e pesquisa, e pelo apoio apresentado para concluir um futuro artigo do presente trabalho.

Ao meu Pai pelo incentivo moral, financeiro e pelo amor cedido a mim em todos os anos de minha vida, para que eu pudesse tornar uma pessoa melhor com os estudos.

A minha querida Mamãe pelo apoio financeiro, moral, carinho, amor e pelas noites de preocupações que ela ficou sem dormir devido às viagens realizadas na fase de curso de disciplinas.

A minha esposa pelos dias que ela ficou sozinha e com medo pela minha ausência, e pelo apoio incondicional a conclusão deste trabalho.

Às minhas irmãs, Lamara, Eliara, Margareth e Renata pelo carinho e estímulo que me proporcionaram. A todos de minha família que de algum modo influenciaram a conclusão do trabalho.

Ao prezado e estimado colega Ailton Antônio de Araújo que foi peça integrante na fase inicial da implementação dos algoritmos do presente trabalho.

Ao profissional Cleiton da Silva Mendonça que foi condição necessária e suficiente para a realização e conclusão deste trabalho, e pelo trabalho desenvolvido tornamos bons amigos. Eu serei eternamente grato pela paciência, dedicação e carinho prestados ao longo de quatro anos de desenvolvimento do *software*.

Às pessoas que acreditaram em meus objetivos e que diretamente ou indiretamente contribuíram para o desenvolvimento deste trabalho.

Por fim, agradeço a todos os amigos de trabalho da UNITINS, da Universidade Federal de Goiás e de infância que de alguma forma contribuíram para meus estudos: Luiz Carlos, Paulo Vitoriano, Joelson, Cleiton, Moisés, Mário Visintainer, Arlenes, Vailton, Vabson, Carlos Henrique, Sara, Aline Terra, Mariana Godinho, Nilson Filho, River Guimarães e Heloisa Helena.

RESUMO

Desenvolvimento do *Software* PlanEx de Planejamento de Experimentos *Online* e sua Aplicação Didática na Pós-Graduação

Palavras Chaves: Planejamento Fatorial, Metodologia de Superfície de Resposta e *Software* Livre

Esta Tese apresenta o *software* PlanEx que contém as ferramentas do planejamento de experimentos fatorial escritas em linguagem de programação *Java*. O objetivo do trabalho foi desenvolver o *software* PlanEx e validá-lo como uma ferramenta de ensino e pesquisa simples, eficiente, interativa e gratuita na *internet*, sem a necessidade do usuário realizar *download*. A linguagem *Java* foi escolhida por oferecer uma grande quantidade de efeitos gráficos e de bibliotecas livres que contêm cálculos da estatística e da álgebra linear. O *software* PlanEx disponibiliza as ferramentas do planejamento de experimentos fatorial completo 2^k e do fracionário 2^{k-1} com até seis variáveis e com, ou sem, replicatas de todos os experimentos. PlanEx foi desenvolvido para que estudantes e pesquisadores interessados no planejamento de experimentos possam utilizá-lo para realizar uma triagem das variáveis; para verificar quais variáveis provocam efeitos significativos na resposta experimental de um fenômeno em estudo, através de replicatas e/ou de gráficos de probabilidade normal; e para avaliar se o modelo é tendencioso através de gráficos de resíduos em função dos valores estimados. Para otimização ou expansão do planejamento de experimentos fatorial, o *software* PlanEx disponibiliza as ferramentas da Metodologia de Superfície de Resposta (RSM) com até quatro variáveis em três níveis, sendo um ponto central e com, ou sem, replicatas. Através da RSM o usuário pode verificar qual modelo, linear ou quadrático, prediz melhor as respostas experimentais através dos gráficos de superfície de resposta 3D e da Tabela de Análise de Variância (ANOVA). Através de três aplicações realizadas no segundo semestre dos anos de 2009, 2010 e 2011 nas aulas da disciplina de Quimometria II ministrada pelo Professor Dr. Anselmo Elcana de Oliveira no programa de pós-graduação de mestrado em Química da UFG e Doutorado Multiinstitucional em Química UFU/UFG/UFMT, o *software* PlanEx apresentou ser uma ferramenta rápida, interativa, gratuita e simples de ensino e pesquisa em nível de pós-graduação.

ABSTRACT

PlanEx a Web/Online Software for Design of Experiments: Development, and Evaluation in Graduate Chemistry Course

Keywords: Design of experiments, response surface methodology and free *software*

This work describes the design of experiments and response surface methodology (RSM) via factorial designs programmed in PlanEx, a new web/online software implemented using the Java language. The goal was to develop and validate PlanEx as a simple, efficient, and an interactive tool for undergraduate teaching and research. PlanEx was also designed as a free software, and the Java language was chosen since it is a community-developed open source software. PlanEx provides both full factorial design of experiments (2^k) and fractional (2^{k-1}), with up to six variables, and with, or without, replicates. PlanEx was also developed to perform variable screening through replicates, and normal probability plots and also to evaluate the model bias by the residual graphs. PlanEx is based on RSM with up to four variables at three levels, and central points with, or without, replicates. Through RSM the user can check linear or quadratic models to predict experimental responses through 3D surface plots, and analysis of variance (ANOVA). PlanEx was also evaluated in the second term of 2009, 2010, and 2011 in graduate courses of planning and design of experiments. The results have shown that PlanEx is fast, and interactive, and it is a simple design of experiment tool for both teaching and research.

LISTA DE FIGURAS

Figura 2.1 – Gráf. da Dispersão das variáveis x e y da Tab. 2.4 do Ex. 6	38
Figura 2.2 – Gráf. Res. Aleat. em função dos val. Est. da Tab. 2.5	40
Figura 2.3 – Rep. Geo. de um Plan. de Exp. Fat. 2^2	54
Figura 2.4 - Rep. Geo. de um Plan. de Exp. Fat. 2^2 da Tab.2.10	55
Figura 2.5 - Rep. Geo. de um Plan. de Exp. Fat. 2^3	62
Figura 2.6 – Gráf. Res. Aleat. em função dos val. Est. da Tab 2.16	64
Figura 2.7 – Exp. Sel. um Plan. Fat. 2^{3-1} usando como gerador $I = x_1x_2x_3$	70
Figura 2.8 – Plan. de Exp. Fat. 2^2 com ponto central	77
Figura 2.9 – Dis. de Prob. da Variável Aleat. x do Ex. 14	96
Figura 2.10 – Dist. de Prob. Contínua do Ex. 14	98
Figura 2.11 – Dist. de Prob. Padrão do Ex. 14 - $z \approx N(0,1)$	99
Figura 2.12 – Dist. Nor. de Prob. Padrão Cumul. do Ex. 14 - Cur. Monot.	100
Figura 2.13 - Dist. Nor. de Prob. Padrão Cumulativa do Ex. 14 - Reta	100
Figura 2.14 – Gráf. de Prob. Nor. do Ex. 12 do Plan. 2^4 obtido com o <i>soft. PlanEx</i>	102
Figura 2.15 - Gráf. de Prob. Nor. do Ex. 12 do Plan. 2^{4-1} obtido com o <i>soft. PlanEx</i>	103
Figura 5.1 - Fluxograma do Plan. de Exp. Fat. Completo, 2^k	116
Figura 5.2 - <i>Layout</i> do <i>soft. PlanEx</i> para o Plan. 2^3 do Ex. 11	117
Figura 5.3 - <i>Layout</i> do <i>soft. PlanEx</i> para o Plan. 2^3 do Ex. 11	118
Figura 5.4 - <i>Layout</i> do <i>soft. PlanEx</i> para o Plan. 2^3 do Ex. 11	119
Figura 5.5 – Gráf. de Resíduos do Ex. 11 obtido com <i>soft. PlanEx</i>	120
Figura 5.6 – Gráf. de Prob. Nor. do Ex. 11 obtido com <i>soft. PlanEx</i>	121
Figura 5.7 – Fluxograma do Plan. de Exp. Fat. Fracionário, 2^{k-1}	122
Figura 5.8 - <i>Layout</i> do <i>soft. PlanEx</i> para o Plan. 2^{4-1} da Ilustração 1	123
Figura 5.9 – Gráf. de Prob. Nor. da Aplicação 1 obtido com <i>soft. PlanEx</i>	124
Figura 5.10 – Fluxograma do RSM , modelo $2^k + r$ ou $2^k + 2k + r$	125
Figura 5.11 – <i>Layout</i> da RSM do <i>soft. PlanEx</i> da Ilustração 2	127
Figura 5.12 – <i>Layout</i> da RSM do <i>soft. PlanEx</i> da Ilustração 2	128
Figura 5.13 - <i>Layout</i> da RSM do <i>soft. PlanEx</i> da Ilustração 2	129
Figura 5.14 – Tabela de ANOVA da RSM do <i>soft. PlanEx</i> da Ilustração 2	130
Figura 5.15 – Gráf. de Resíduos obtido com o <i>soft. PlanEx</i> da Ilustração 2	130
Figura 5.16 – Gráf. de Sup. em 3D obtido com o <i>soft. PlanEx</i> da Ilustração 2	131
Figura 5.17 - <i>Layout</i> da RSM do <i>soft. PlanEx</i> usando os dados da Ilustração 3	132
Figura 5.18 - <i>Layout</i> da RSM do <i>soft. PlanEx</i> usando os dados da Ilustração 3	133
Figura 5.19 - <i>Layout</i> das Variáveis do <i>soft. PlanEx</i> da Ilustração 3	133
Figura 5.20 – Gráf. de Resíduos obtido com o <i>soft. PlanEx</i> da Ilustração 3	134
Figura 5.21 - <i>Layout</i> da ANOVA do <i>soft. PlanEx</i> da Ilustração 3	135
Figura 5.22 – Gráf. de Sup. obtido com o <i>soft. PlanEx</i> da Ilustração 3	135

Figura 5.23 - <i>Layout</i> da RSM do <i>soft</i> . PlanEx da Ilustração 4	136
Figura 5.24 - <i>Layout</i> da RSM do <i>soft</i> . PlanEx da Ilustração 4	137
Figura 5.25 - <i>Layout</i> das Variáveis do <i>soft</i> . PlanEx da Ilustração 4	138
Figura 5.26 – Gráf. de Resíduos obtido com o <i>soft</i> . PlanEx da Ilustração 4	138
Figura 5.27 - <i>Layout</i> da ANOVA do <i>soft</i> . PlanEx da Ilustração 4	139
Figura 5.28 – Gráf. de Sup. obtido com o <i>soft</i> . PlanEx da Ilustração 4	139
Figura 5.29 - Análise Quant. Questão 1 – Apêndice D	145
Figura 5.30 - Análise Quant. da Questão 2 – Apêndice D	145
Figura 5.31 - Análise Quant. da Questão 3 – Apêndice D	147
Figura 5.32 - Análise Quant. da Questão 4 – Apêndice D	149
Figura 5.33 - Análise Quant. da Questão 5 – Apêndice D	150

LISTA DE TABELAS

Tabela 2.1 - Planejamento de Experimentos do Exemplo 4	33
Tabela 2.2 – Desvio em relação à média aritmética do Exemplo 5	35
Tabela 2.3 - Medidas de Dispersão: Variância e Desvio Padrão do Exemplo 5	36
Tabela 2.4 - Correlação Linear das variáveis do Exemplo 6	37
Tabela 2.5 - Valores estimados e resíduos aleatórios da regressão usando MQO do Exemplo 6	40
Tabela 2.6 - Planejamento de Experimentos da análise de regressão multivariada do Exemplo 8	42
Tabela 2.7 - Matriz de Planejamento de Experimentos Fatorial 2^2 do Exemplo 9	47
Tabela 2.8 - Matriz de Contrastes do Planejamento de Experimentos Fatorial 2^2 do Exemplo 9	47
Tabela 2.9 - Matriz de Planejamento de Experimentos Fatorial 2^2 do Exemplo 10	52
Tabela 2.10 - Matriz de Contrastes de Planejamento de Experimentos Fatorial 2^2 do Exemplo 10	52
Tabela 2.11 - Estimativa da Variância Conjunta do Planejamento 2^2 do Exemplo 10	56
Tabela 2.12 - Efeitos calculados para o Planejamento de Experimentos Fatorial 2^2 da Tabela 2.9	57
Tabela 2.13 - Planejamento de Experimentos Fatorial 2^3 do Exemplo 11	59
Tabela 2.14 - Estimativa da Variância Conjunta do Planejamento 2^3 do Exemplo 11	61
Tabela 2.15 - Efeitos do Planejamento 2^3 do Exemplo 11 - Tabela 2.13	62
Tabela 2.16 - Valores Estimados e Resíduos Aleatórios do Exemplo 11 – Tabela 2.13	63
Tabela 2.17 - Planejamento de Experimentos Fatorial 2^4	65
Tabela 2.18 - Planejamento de Experimentos Fatorial 2^3	67
Tabela 2.19 - Planejamento 2^{3-1} obtido de um Planejamento 2^3 – Tabela 2.18	68
Tabela 2.20 - Relações entre os contrastes da meia fração 2^{3-1} e os efeitos do fatorial 2^3	70
Tabela 2.21 - Planejamento 2^{4-1} obtido de um Planejamento 2^4	70
Tabela 2.22 - Relações entre os contrastes da meia fração 2^{4-1} e os efeitos do fatorial 2^4	71
Tabela 2.23 - Planejamento de Experimentos Fatorial Completo 2^4 do Exemplo 12	72
Tabela 2.24 - Planejamento de Experimentos Fatorial Fracionário 2^{4-1} do Exemplo 12	75
Tabela 2.25 - Relações entre os contrastes da meia fração 2^{4-1} e os efeitos do fatorial completo 2^4	76
Tabela 2.26 - Planejamento Composto Central do Exemplo 13 com 11 experimentos	84
Tabela 2.27 - Tabela de ANOVA para ajuste de um modelo linear	89
Tabela 2.28 - Tabela ANOVA com a possibilidade de julgar o ajuste do modelo	90

Tabela 2.29 - Tabela da Soma Quadrática do Ex. 13 da $\hat{y} = 39,67 - 1,49x_1 + 8,14x_2 - 2,21x_1^2$	92
Tabela 2.30 - Tabela o grau de liberdade do Ex. 13 da $\hat{y} = 39,67 - 1,49x_1 + 8,14x_2 - 2,21x_1^2$	93
Tabela 2.31 - Tabela de ANOVA do Ex. 13 da $\hat{y} = 39,67 - 1,49x_1 + 8,14x_2 - 2,21x_1^2$	94
Tabela 2.32 - Distribuição de Probabilidade no lançamento de quatro moedas – Ex. 14	95
Tabela 2.33 - Efeitos do Planejamento 2^4 do Ex. 12 para encontrar o gráfico de probabilidade normal	101
Tabela – Probabilidade da Distribuição Normal – Apêndice C	171

LISTA DE SÍMBOLOS

$\mathbf{y}_{n \times 1}$: matriz coluna ou vetor coluna de todos os experimentos

\bar{x} : média aritmética dos valores de x_i

K: quantidade de variáveis

d_i : desvio em relação à média aritmética

σ^2 : variância populacional

s^2 : variância amostral

σ : desvio padrão populacional

s: desvio padrão amostral

γ : coeficiente de Pearson que mede o grau da correlação linear entre duas variáveis

y: função de regressão populacional (FRP)

\hat{y} : função de regressão populacional (FRA)

β : parâmetros populacionais da regressão

$\hat{\beta}$: parâmetros amostrais da regressão

u: erro aleatório populacional da regressão

\hat{u} : erro aleatório amostral da regressão

\bar{y} : média aritmética dos valores de y_i

$\hat{\mathbf{y}}$: matriz coluna estimada das respostas experimentais estimadas \hat{y}_i

\mathbf{y} : matriz coluna populacional das respostas experimentais populacionais y_i

$\hat{\beta}$: matriz coluna ou vetor coluna dos parâmetros da regressão β

\mathbf{X} : matriz de todas as variáveis e interações entre elas

\mathbf{X}^t : matriz transposta de todas as variáveis e interações entre elas

$(\mathbf{X}^t\mathbf{X})^{-1}$: matriz inversa da multiplicação da matriz \mathbf{X}^t pela matriz \mathbf{X}

$x_{\text{Nível}}$: codificação da variável no nível desejado

$v_{\text{Nível}}$: o valor experimental da grandeza física realizada no nível i

\bar{v} : média de todos os valores experimentais da grandeza física estudada

$v_{\text{Nível_Menor}}$: menor valor de todos os valores experimentais da grandeza codificada

n : quantidade mínima de experimentos realizados

$Ef_{\text{médio}}$: efeito médio ou média dos experimentos y_i que se denota por \bar{y}

$y_{i(+)}$: respostas experimentais realizadas no nível maior

$y_{i(-)}$: respostas experimentais realizadas no nível menor

X : efeito principal da variável x

X_1X_2 : efeito de interação entre as variáveis x_1 e x_2

\mathbf{m} : matriz coluna resultado da multiplicação de matrizes \mathbf{X}^t por \mathbf{y}

\mathbf{ef} : matriz coluna que contém os efeitos das variáveis e suas interações

X_F : efeito principal da variável no planejamento de experimentos fatorial de meia fração

I – geratriz do planejamento de experimentos fatorial fracionário

2^k – Planejamento de Experimentos Fatorial Completo em dois níveis e com k variáveis

2^{k-1} – Planejamento de Experimentos Fatorial de meia fração em dois níveis

E_C : estimativa da curvatura de uma superfície experimental

α : distância da variável codificada até a origem do sistema de coordenadas

(1): significa o(s) experimento(s) realizado(s) no nível menor de todas as variáveis

x_1 : significa o(s) experimento(s) realizado(s) no nível maior de x_1 e no nível menor das demais

x_2 : significa o(s) experimento(s) realizado(s) no nível maior de x_2 e no nível menor das demais;

x_1x_2 : significa o(s) experimento(s) realizado(s) no nível maior das variáveis x_1 e x_2 ;

o_j : significa o(s) experimento(s) realizado(s) no nível zero ou ponto central das variáveis, com $j = 1, 2, \dots, r$ replicatas

x_{1e-} : significa o(s) experimento(s) realizado(s) no nível menor da variável x_1 do planejamento estrela

x_{1e+} : significa o(s) experimento(s) realizado(s) no nível maior da variável x_1 do planejamento estrela

SQ: soma quadrática da Análise de Variância (ANOVA)

MQ: média quadrática da Análise de Variância (ANOVA)

R^2 : coeficiente de determinação

F: Distribuição de Fisher

p: quantidade de parâmetros da regressão do modelo

m: quantidade de níveis diferentes

z: variável aleatória padronizada da distribuição normal

SUMÁRIO

APRESENTAÇÃO: Justificativa e Guia de Trabalho	16
1. INTRODUÇÃO	17
1.1 A Quimiometria no Brasil	17
1.2 O Ensino de Quimiometria	19
1.3 As Áreas da Quimiometria	20
1.4 Alguns <i>Softwares</i> de Quimiometria	21
1.5 Alguns Trabalhos Usando as Ferramentas da Quimiometria	22
1.6 <i>Internet</i> e Quimiometria	25
1.7 Triagem e Otimização	27
1.8 Tipos de Planejamentos Implementados no <i>software</i> PlanEx	28
2. ESTATÍSTICA EXPERIMENTAL	30
2.1 Conceitos Básicos da Estatística	30
2.2 Correlação Linear e Regressão Multivariada	36
2.3 Planejamento de Experimentos Fatorial	45
2.4 Obtenção de efeitos de um Planejamento de Experimentos Fatorial Completo usando os conceitos de Álgebra Linear	57
2.5 Planejamento de Experimentos Fatorial Fracionário	66
2.6 Planejamento de Experimentos Fatorial com Ponto Central	77
2.7 Metodologia de Superfície de Resposta	80

2.8 Teste de ANOVA para Avaliação do Modelo	87
2.9 Gráficos de Probabilidade Normal	94
3. CONCEITOS BÁSICOS DE LINGUAGEM COMPUTACIONAL	104
4. USANDO JAVA PARA RESOLVER O PROBLEMA	107
4.1 Planejamento de Experimentos Fatorial Completo	107
4.2 Planejamento de Experimentos Fatorial Fracionário	109
4.3 Metodologia de Superfície de Resposta	110
5. RESULTADOS E DISCUSSÃO	115
5.1 Resultado do Planejamento de Experimentos Fatorial Completo 2^k	116
5.2 Resultado do Planejamento de Experimentos Fatorial Fracionário 2^{k-1}	121
5.3 Resultado da Metodologia de Superfície de Resposta	125
5.4 Validação do <i>software</i> PlanEx como Ferramenta de Pesquisa e Ensino	141
6. CONCLUSÕES	153
7. BIBLIOGRAFIA	154
8. APÊNDICE	
8.1 Apêndice A	158
8.2 Apêndice B	164
8.3 Apêndice C	171
8.4 Apêndice D	172

APRESENTAÇÃO: Justificativa e Guia do Trabalho

Os recursos computacionais estão presentes em todos os níveis do ensino de química, com base nisso vários *softwares* livres foram desenvolvidos para o ensino de química no nível médio para alunos visualizarem conceitos e estruturas em ambientes tridimensionais e interagir com a *internet*. O Programa Nacional de Informática Educativa (PROINFO) e a Rede Interativa Virtual de Educação (RIVED), endossam a recomendação da utilização das tecnologias nos mais diversos níveis e áreas curriculares (CASTRO FILHO *et al*, 2008 apud BENITE *et al*, 2011). Nesse contexto, o desenvolvimento das tecnologias digitais e a proliferação das redes interativas tendem a colocar o ensino de química diante de um caminho sem volta, pois já não somos como antes.

Na pós-graduação de química, especialmente na área de quimiometria, há pouco material de ensino e pesquisa livres que realizam cálculos dos planejamentos fatoriais, e o pouco que há requer um conhecimento de uma linguagem específica para operar o *software*, por exemplo, o *Octave*. Isso acarreta num dispendimento maior de tempo para execução de cálculos e, conseqüentemente, um tempo menor disponível para fazer análise dos resultados.

Com base nisso, torna-se oportuno usar a *internet* e a falta de *softwares* livres que contém as ferramentas de planejamentos de experimentos fatorial para desenvolver um *software* de ensino e pesquisa em nível de pós-graduação, para que usuários interessados na área possam realizar cálculos para gerar gráficos de superfície de resposta e otimizar experimentos sem custo de manuseio, sendo o *software* implementado na linguagem Java e disponibilizado gratuitamente na *internet*. Uma das vantagens é que os usuários não precisam realizar *downloads* para usar o *software* e nem conhecer conceitos que não tem relação com o conteúdo proposto, como o conhecimento de uma linguagem de operação, o que pode dificultar o ensino e pesquisa de quimiometria.

O *software* desenvolvido denominou-se de PlanEx – Planejamentos de Experimentos - e mostrou ser uma ferramenta rápida, amigável e interativa para analisar e otimizar problemas de experimentos em nível de pós-graduação, sendo uma ferramenta de ensino eficiente que facilita os cálculos e oportuniza um tempo maior para os usuários preocupem com a análise dos resultados.

Inicialmente, apresenta-se a introdução do trabalho, abordando os conceitos básicos da quimiometria e a falta de material gratuitos e dificuldades de ensino e pesquisa. Em seguida, comenta-se os cálculos estatísticos da quimiometria de forma detalhada e com exemplos numéricos para que os usuários possam acompanhar o raciocínio das equações implementadas no PlanEx. Após, apresenta-se algumas definições na área de informática que foram usadas para implementas os algoritmos da quimiometria, como apresenta-se na seção 3 e 4. Na seção 5, apresenta-se os resultados e discussão com as figuras dos *layouts* do PlanEx e gráficos gerados a partir dele. E por fim, as conclusões do trabalho.

1. INTRODUÇÃO

Para Box *et al* (1978) o termo ciência numa pesquisa científica é um processo de aprendizagem guiada. Pesquisa científica constitui-se de ferramentas lógicas de investigação para resolver problemas e encontrar respostas de questões relacionadas a fenômenos naturais. Essas ferramentas ou um conjunto de regras básicas dão origem ao método científico. Nele os cientistas tentam a geração de um conhecimento baseado em caminhos definidos por sua história profissional e de vida e por suas escolhas para produzir o conhecimento científico, quer seja conhecimento novo de uma nova hipótese de conhecimentos evoluídos ou de expandidos da área de abrangência de conhecimentos pré-existente.

No método científico destaca-se a Estatística, sendo uma parte ou ferramenta essencial para tomada de decisão em experimentos que envolvem a incerteza. Estatística é uma parte da matemática aplicada, e conhecida como a "matemática da incerteza" e tida por muitos como uma ciência aplicada (GUJARATI, 2006). A estatística, suas inferências e ferramentas, dependem dos conceitos das áreas particulares às quais é aplicada. Por esse motivo, é frequentemente identificada de acordo com as suas áreas de aplicação, assumindo denominações específicas em muitos casos, tais como estatística experimental, **quimiometria**, biometria, bioestatística, econometria e sociometria, por exemplo.

Medição química é a interpretação literal da palavra **quimiometria**. Embora seja uma parte importante da **quimiometria**, a abrangência é maior. A **quimiometria** é uma disciplina química que usa operações básicas da matemática, da estatística, da álgebra linear e de outros métodos de forma lógica para planejar ou selecionar procedimentos ótimos para medição e experimentação, e obter a máxima informação relevante pela análise de medidas químicas (BOX *et al*, 1978; BRUNS e FAIGLE, 1985; BRUNS *et al*, 2001, 2006; ORTIZ, 2007).

Percebe-se que a **quimiometria** é constituída pelas ferramentas da estatística com aplicação à química. Um dos pesquisadores que iniciou essa aplicação no mundo foi Bruce R. Kowalski, sendo Roy Edward Bruns no Brasil (BRUNS *et al*, 2006).

1.1. A Quimiometria no Brasil

Um dos fatos que mais acentuam a dificuldade de estudantes iniciantes na utilização da estatística é a repetição de cálculos simples, os quais podem apresentar de primeiro momento como um obstáculo ao estudo e aplicação da estatística. Isso não é diferente nas

áreas aplicadas em que a estatística está, por exemplo, na **quimiometria**. Isso se minimiza pelo uso de ferramentas computacionais.

No final da década de 70, os microcomputadores ainda não tinham chegado ao Brasil e os cálculos muito extensos de **quimiometria** eram realizados em computadores de grande porte, que funcionavam em centros de computação (BRUNS *et al*, 2006). Na época, a Unicamp tinha um dos computadores mais avançados das universidades brasileiras, um *PDP-10*, que era multiusuário e tinha nada menos que *1 MB* de memória *RAM*, a ser compartilhada por todos os usuários da universidade (BRUNS *et al*, 2006). Para realizar os cálculos quimiométricos com o *software* ARTHUR (idealizado por Bruce R. Kowalski), foi necessário o uso de estratégias cuidadosas para poupar espaço na memória *RAM*, e o acesso aos resultados de qualquer cálculo realizou-se exclusivamente através de listagens impressas, produzidas por impressoras de linhas de 132 caracteres. Mas havia um ponto positivo: naquele tempo o *software* era gratuito (BRUNS *et al*, 2006).

O estudo de **quimiometria** no Brasil iniciou-se formalmente na primeira metade da década de 70, consolidando-se no momento que o computador foi utilizado como componente de instrumentos de medição no laboratório químico (BRUNS *et al*, 2006). Em novembro de 1980 o Professor Kowalski, mentor do grupo que produziu o *software* ARTHUR, aceitou um convite para ministrar um curso intensivo de **quimiometria** no Instituto de Química da Unicamp. O curso ocorreu no mês de novembro e foi o primeiro no Brasil sobre esse assunto (BRUNS *et al*, 2006). No Brasil a primeira dissertação envolvendo a aplicação de métodos quimiométricos foi defendida por Ieda Spacino Scarminio, em fevereiro de 1981 e, Elias Zagatto a primeira tese usando técnicas quimiométricas (BRUNS *et al*, 2006).

O avanço computacional ocorreu por volta de 1985, com o aparecimento dos microcomputadores com palavras de *16 bits*. Graças a esse advento, a **quimiometria** foi implantada como disciplina acadêmica, já que, além de se oferecer uma base teórica ao aluno, também passava a ser possível colocar os conhecimentos em prática (BRUNS *et al*, 2006).

Em 1984, quatro anos depois do curso pioneiro de Kowalski, Roy Edward Bruns ministrou o segundo curso intensivo de extensão em **quimiometria** no Departamento de Química Fundamental da Universidade Federal de Pernambuco. Nessa ocasião, o departamento estava modificando a matriz curricular de seu curso de graduação e incluiu a disciplina de **Quimiometria** como eletiva (BRUNS *et al*, 2006). O primeiro curso regular de 60 h semestrais em **quimiometria** foi oferecido na pós-graduação do Instituto de Química - Unicamp, atraindo um total de dois alunos, no ano de 1985. Nesse mesmo ano, Roy Edward

Bruns publicou um trabalho de revisão sobre **quimiometria** em Química Nova (BRUNS e FAIGLE, 1985).

1.2. O Ensino de Quimiometria

Ortiz (2007), membro permanente da equipe de Química Analítica da Universidade de Burgos na Espanha, ensina **quimiometria** e relata a importância de ensinar **quimiometria** em aulas teóricas e práticas. No seu trabalho (ORTIZ, 2007) afirmou que o ensino de **quimiometria** em educação superior é desigual, na melhor das hipóteses. Diz que há departamentos ou institutos que ignoram totalmente o tópico, e quando é incluído no currículo químico o tratamento é frequentemente limitado, disperso e algumas vezes desequilibrado. Por exemplo, a metodologia de planejamento de experimentos pode ser sub-representada enquanto a análise de dados é enfatizada.

Ortiz (2007) afirma ainda que, devido à utilidade e ampla aplicação industrial, é importante que o ensino de **quimiometria** permita aos alunos ver o seu poder em resolver problemas, e cita exemplos de atividades que professores de **quimiometria** possam fazer em laboratório.

Márcia Ferreira e Teófilo (2006) realizaram um tutorial de como ensinar, pesquisar e estudar planejamento de experimentos e otimização empregando Metodologias de Superfície de Resposta (RSM), que é uma área da **quimiometria**, através de planilhas eletrônicas, tais como *Excel* (www.microsoft.com) e *BrOffice* (www.openoffice.org.br).

Os trabalhos de Ortiz (2007) e, Marica Ferreira e Teófilo (2006) divulgaram algumas técnicas das áreas de quimiometria e a importância em ensinar e estudar, e relatam também a dificuldade do acesso de pessoas aos *softwares* e a compreensão das operações que ficam mascaradas, visto que muitos *softwares* não apresentam características de amigabilidade e de interatividade. Um *software* amigável deve possuir uma boa usabilidade e um bom *design* gráfico. A usabilidade de um *software* consiste o quão bem os usuários podem usar as funcionalidades do sistema. A capacidade de aprendizado, a eficiência e memorabilidade do *software* são componentes da usabilidade. Uma das características para um bom *design* gráfico de um *software* consiste em possuir simplicidade.

O trabalho de Peralta-Zamora *et al* (2005) apresentou algumas das ferramentas da RSM usando o modelo linear em três casos, sendo os dois primeiros com duas variáveis independentes, e o último com três variáveis independentes. Esse trabalho é um reforço de como as técnicas de estatística de **quimiometria** são importantes, no entanto, o trabalho não

mostra as equações usadas para calcular os efeitos principais e interações de segunda e terceira ordem, e nem a construção de gráfico de superfície de resposta, usado para otimização de fenômenos em estudo, como se apresentará na página 80 da seção da RSM da Estatística Experimental.

Percebe-se que a aplicação da **quimiometria** é ampla, como será discutido em breve, porém há poucos livros, artigos e *softwares* disponíveis para o ensino e pesquisa de **quimiometria**, especialmente na área de planejamento de experimentos.

1.3. As Áreas da Quimiometria

É rotineiro, especialmente em indústrias químicas, aparecerem problemas em que se precisam estudar várias propriedades ao mesmo tempo e estas, por sua vez, são afetadas por um grande número de variáveis experimentais, e por meio da **quimiometria**, é possível extrair do sistema em estudo o máximo de informação útil, fazendo um número mínimo de experimentos (BRUNS *et al*, 2001). A **quimiometria** no Brasil pode ser agrupada em três áreas principais:

- Planejamento e Otimização de Experimentos;
- Reconhecimento de Padrões;
- Calibração Multivariada.

O Planejamento e Otimização de Experimentos é um ramo da **quimiometria** que cresceu rapidamente em vários campos de pesquisa no Brasil, especialmente Química, Engenharia Química, Engenharia de Alimentos e Biotecnologia, tanto em atividades acadêmicas como em indústrias, por exemplo: Oxiteno, Pirelli, Nitroquímica, Unilever, Petrobrás, Petroflex e outros (BOX *et al*, 1978; DIAS *et al*, 2006; FERREIRA, J. *et al*, 2007; FERREIRA, M. *et al*, 2006; MASHRU *et al*, 2008; MULINARI *et al*, 2006; BRUNS *et al*, 2006; ORTIZ, 2007; SHERIDAN *et al*, 2004; SILVA, J. *et al*, 2006).

No curso de administração a palavra planejamento é um dos requisitos necessários para alcançar o sucesso e os objetivos desejados. Na química não é diferente, é preciso planejar os experimentos. Os Tipos de Planejamentos de Experimentos são (BRUNS *et al*, 2001)

- Quadrado Latino;
- Quadrado Greco-Latino;
- Tratamento em Blocos;
- Quadrado Hiper-Greco-Latino;

- Planejamento Fatorial.

O uso de planejamentos de experimentos, em especial os planejamentos de experimentos fatoriais, permite o estudo e a representação de um processo ou sistema utilizando-se um modelo matemático obtido através de técnicas estatísticas (BRUNS, ET AL, 2001), relacionando as variáveis em estudo de maneira sistemática, proporcionando a obtenção das respostas desejadas com um mínimo de tempo, fator imprescindível para o bom andamento da pesquisa (BRUNS *et al*, 2001; OLIVIER *et al*, 2007). De posse do modelo, é possível então a construção de superfícies de resposta que servem para uma análise mais efetiva dos efeitos das variáveis envolvidas no processo e, também, para a otimização do modelo (ALVES *et al*, 1998; DIAS *et al*, 2006; FERREIRA, J. *et al*, 2007; MASHRU *et al*, 2008; MULINARI *et al*, 2006; OLIVIER *et al*, 2007; SHERIDAN *et al*, 2004; SILVA, J. *et al*, 2006).

1.4. Alguns Softwares de Quimiometria

As ferramentas do planejamento de experimentos fatorial exigem após sua execução, cálculos simples, mas muitas vezes tediosos devido à grande quantidade de dados a serem processados, o que pode ser minimizado pelo uso de recursos computacionais. Atualmente há no mercado diversos tipos de *softwares* que realizam cálculos de planejamentos experimentais fatorial, tais como: Statistica (www.statsoft.com), Unscrambler (www.camo.com), Statgraphics (www.statgraphics.com), Design Expert (www.statease.com), Modde (www.umetric.com), Minitab (www.minilab.com), SAS (www.sas.com), EvIEWS (www.eviews.com), *Matlab* (www.mathworks.com) e outros. Um ponto negativo é que a maioria destes *softwares* não é gratuita ou permite uma licença do tipo *DEMO*, ou seja, o usuário tem a permissão para utilizar o *software* durante um intervalo de tempo, geralmente 30 dias, para demonstração das ferramentas e que nem todas podem estar disponíveis. Outro ponto negativo desses *softwares* é que a maioria processa resultados complicados de serem entendidos por um pesquisador inexperiente, e alguns requerem o conhecimento de uma linguagem matemática computacional específica, por exemplo, o *Matlab*.

Entre os *softwares* gratuitos destaca-se o pacote de algoritmos executáveis no sítio (www.iqm.unicamp.br), a partir do link *chemkeys* (BRUNS *et al*, 2001). Apesar da gratuidade e apresentar-se no idioma português, o ambiente de implementação do *software* está ultrapassado, *MS-DOS*, pois consiste no primeiro sistema operacional lançado pela *Microsoft*, e opera por linha de comando e não possui ambiente gráfico. Atualmente não se utiliza no dia-

a-dia das organizações bem como dos usuários, e apresenta-se como um recurso a mais do sistema operacional *Windows* para execução de comandos ou de algum sistema legado que ainda não foi atualizado. Outro *software* gratuito no idioma português são as planilhas eletrônicas do trabalho de Márcia Ferreira e Teófilo (2006), como se comentou na página 19.

Os *softwares* de planejamento de experimentos fatorial disponíveis em português e gratuitos reforçam o estudo de **quimiometria** e aumenta o número de opções de *softwares* (FERREIRA, M., TEÓFILO, 2006; www.iqm.unicamp.br, a partir do link *chemkeys*), pois acredita-se que *softwares* gratuitos tendem a ser uma ferramenta de ensino e pesquisa mais difundidos devido a facilidade de acesso. O trabalho de Márcia Ferreira e Teófilo (2006) apresenta característica amigável, porém há restrições de alguns tipos de gráficos: no planejamento de experimentos fatorial não se percebe a construção do gráfico de resíduos em função dos valores estimados; na RSM o gráfico de superfícies de resposta não se obtém de forma automática, necessita-se digitar a função estimada. O nível de interatividade das planilhas eletrônicas do trabalho de Márcia Ferreira e Teófilo (2006) pode ser considerado baixo, pois o usuário basicamente deve preencher as planilhas com os dados de entrada e implementar a função para gerar o gráfico de superfície de resposta. O restante das operações, as planilhas não requerem nenhum comando de interação, fazendo os cálculos e gráficos automaticamente após a digitação dos dados de entrada. Já a interatividade do *software* disponível em www.iqm.unicamp.br, a partir do link *chemkeys*, é alta, porém a interatividade classifica-se como comando de execução, e não apresenta opções para gráficos de superfície de resposta, e a opção de salvar gráficos não está disponível. Ambos os trabalhos, são obtidos por *downloads*.

1.5. Alguns Trabalhos Usando as Ferramentas de Quimiometria

O emprego de planejamento de experimentos fatorial é amplo e sua aplicação depende da imaginação dos pesquisadores, como se observa mediante a descrição de alguns trabalhos a seguir (ALVES *et al*, 1998; DIAS *et al*, 2006; FERREIRA, J. *et al*, 2007; MASHRU *et al*, 2008; MULINARI *et al*, 2006; OLIVIER *et al*, 2007; SHERIDAN *et al*, 2004; SILVA, J. *et al*, 2006).

O trabalho de Alves *et al* (1998) usou planejamento de experimentos fatorial para reduzir ou remover a concentração de substâncias tóxicas presentes no hidrolisado da hemicelulose de bagaço de cana através da bioconversão xilose em xilitol variando o nível de pH através da combinação de diferentes bases e ácidos em função do tempo de fermentação,

com ou sem adição subsequente de carvão ativo. Realizou-se um planejamento de experimentos fracionário do tipo 2^{5-1} para encontrar os efeitos significativos. Após, um planejamento de experimentos fatorial completo 2^2 com três replicatas dos pontos centrais para analisar efeitos não-lineares na superfície de resposta. Para o desenvolvimento do trabalho utilizou-se os *softwares* comerciais estatísticos STATGRAPHICS versão 6.0 (www.statgraphics.com) e o Statistica versão 5.0 (www.statsoft.com).

O trabalho de Olivier *et al* (2007) empregou um planejamento de experimentos fatorial completo 2^4 acrescido de triplicatas dos pontos centrais para cada tipo de ácido para otimizar o processo de extração de zinco de resíduos galvânicos. Obtiveram-se os cálculos através do *software* estatístico Statistica versão 6.0 (www.statsoft.com).

Segundo Mashru *et al* (2008) a amargura de cloridrato de mefloquina foi reduzida com uma melhor dissolução em meio ácido. Usou-se Eudragit E como polímero e hidróxido de sódio como precipitado. Utilizou-se um planejamento de experimentos fatorial completo com três níveis 3^2 para otimização, em que a concentração da droga e a concentração do polímero foram selecionadas como variáveis independentes, e o motivo da amargura, o tamanho da partícula e a dissolução em vários pH foram selecionados como variáveis dependentes. Realizou-se o trabalho usando o pacote estatístico DOE v 6.0.5 (www.statease.com).

No trabalho de Sheridan *et al* (2004) foram usados dois planejamentos de experimentos fatorial com três níveis, um completo e outro rotativo, para comparar qual planejamento prediz melhor precisão na eficiência de examinar a interação de três tóxicos aquáticos em embriões de sapos. Utilizou-se o programa o S-Plus (S-Plus2000 versão profissional 2, MathSoft Inc, Cantabrigia, MA, E.U.A). Já no trabalho de Joelma Ferreira *et al* (2007) realizaram através de um planejamento de experimentos fatorial completo 2^3 acrescido de três replicatas dos pontos centrais um estudo mais abrangente das variáveis independente, realizando de uma forma mais organizada uma quantidade mínima de experimentos (BRUNS, ET AL, 2001). Obtiveram-se os parâmetros ou coeficientes da regressão através do *software* Statistica, versão 5.0 (www.statsoft.com). Os autores do trabalho de José Silva *et al* (2006) também usaram o programa Statistica, versão 5.0 (www.statsoft.com).

Dias *et al* (2006) investigaram as propriedades eletroquímicas de Azul de metileno imobilizado na celulose TiO_2 e SiO_2/TiO_2 com o auxílio de um planejamento de experimentos fatorial 2^4 . As técnicas quimiométricas usadas no trabalho propiciaram a análise de gráficos de probabilidade acumulada, denominados de gráficos de probabilidade normal, para verificar quais efeitos das variáveis são significativos (BOX *et al*, 1978; BRUNS *et al*, 2001), como se apresentará na página 94 da seção de Gráficos de Probabilidade Normal. Através do

planejamento, encontrou-se que três variáveis das quatro estudadas possuem efeitos principais significativos, e que existem interações secundárias e terciárias que são significativas contendo ou não a variável que não possui efeito principal. Através da representação geométrica de uma figura de um cubo com as variáveis significativas, os autores encontraram a otimização do sistema, como se apresentará na Figura 2.5 (BOX, 1978; BRUNS *et al*, 2001; DIAS *et al*, 2006). Os cálculos foram obtidos usando o *software* FATORIAL (BRUNS *et al*, 2001).

O trabalho de Gouveia *et al* (2002) desenvolveu um estudo para avaliar a cinética de secagem de caju em um secador de leito fixo, utilizando um planejamento de experimentos do tipo planejamento composto central como se apresentará na página 80 da seção da RSM usando o modelo linear e quadrático para avaliar qual modelo prediz melhor as respostas experimentais tendo como variáveis independentes a temperatura de secagem (°C) e a velocidade do ar (m/s).

O trabalho de Joelma Ferreira *et al* (2005) analisou o efeito das fontes de nutrientes sobre a produção de vinagre em escala de bancada através da RSM. Iniciou-se com um planejamento de experimentos fatorial 2^3 com triplicatas do ponto central para avaliar os efeitos das variáveis independentes: concentrações iniciais de etanol, de fontes de fósforo e de nitrogênio. Após esta análise ampliaram-se as faixas das concentrações iniciais das variáveis de fontes de nutrientes (fósforo e nitrogênio) para um planejamento de experimentos fatorial do tipo 2^2 mais configuração estrela denominado de planejamento composto central, para o modelo linear e quadrático, ambos resultando em boa concordância com as respostas experimentais.

O trabalho de Muniz *et al* (2008) estudou a caracterização de Hidrogéis Condutores constituídos por PAAM e PEDOT/PSS por meio da RSM usando o modelo linear com três variáveis independentes com quadruplicatas do ponto central. Realizou-se a estimativa de erro e significância do modelo através da análise de variância (ANOVA), um acrônimo de *Analysis of Variance*.

Os artigos mencionados anteriormente neste trabalho constituem-se uma pequena amostra do conjunto universo de diversidades de aplicações de **quimiometria** que já foram realizadas usando as técnicas do planejamento de experimentos fatorial e RSM.

Na seção 1.6 apresenta-se indícios que a *internet* facilita o aprendizado de química disponibilizando material de estudo e pesquisa, bem como ferramentas para auxílio no desenvolvimento de estudos.

1.6. *Internet e Quimiometria*

No trabalho de Santos et al (2010) cujo título é “*Softwares* educativos livres para o Ensino de Química: Análise e Categorização” os autores discutem a importância dos recursos computacionais de química no ensino médio:

O computador está inserido na sociedade como aparelho essencial para o desenvolvimento de algumas atividades produtivas, participando também do lazer e cada vez mais sendo incluído na educação. Esta inclusão parte do princípio da necessidade da promoção na escola do desenvolvimento de competências e habilidades do cidadão, além de contribuir para a formação de pessoas com senso crítico apurado. Há também a necessidade crescente imposta pelo mercado de trabalho em oferecer oportunidade para pessoas com habilidade para o uso de novas tecnologias, tais como o computador. Deste modo, a escola e seus profissionais devem adequar-se a nova realidade.

Nesse contexto, o desenvolvimento das tecnologias digitais e a proliferação das redes interativas tendem a colocar o ensino de química diante de um caminho sem volta, pois já não somos como antes.

Promoveu-se a partir do acelerado avanço da ciência e tecnologia nas últimas décadas um enorme crescimento da quantidade de informação disponível no estudo de química, e este crescimento, aliado ao surgimento e à popularização da rede mundial de computadores ou *internet*, demandou a construção de ferramentas eficientes de busca destas informações (MORAN, 1997 apud FIRME *et al*, 2008). Após essas tecnologias terem alcançado vários setores da sociedade, a educação é uma das áreas mais afetadas, com a introdução de computadores nas salas de aulas com *softwares* interativos e acesso aos recursos da *internet* (FERREIRA, V., 1998).

No trabalho de Firme *et al* (2008), os autores apresentaram aos estudantes de Química e áreas afins, tanto de graduação quanto de pós-graduação, as fontes de aquisição de informação via *internet* como ferramenta complementar à pesquisa e ao ensino. Além de fornecer uma lista de sítios importantes, também incentivaram a pesquisa bibliográfica utilizando a *internet*, não defendendo a substituição dos meios convencionais e incentivaram o investimento em bibliotecas tradicionais como caminho ao desenvolvimento. No trabalho de Russo (2006 apud FIRME *et al*, 2008) comentou que há uma tendência de aumento da disponibilidade dos artigos de livre acesso, visto que eles são mais imediatamente

reconhecidos e citados que os artigos de acesso restrito (MACCALLUM, 2006 apud FIRME *et al*, 2008). Um estudo da Comunidade Europeia (CE) mostrou que o custo das revistas científicas aumentou 200 a 300% no período entre 1975 e 1995 (http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf). Isso intensificou o debate ao acesso livre à informação científica, “Open-Access Initiative” (RUSSO, 2006 apud FIRME *et al*, 2008). Amparados pelo estudo da CE, mais de 12.000 acadêmicos, incluindo dois vencedores do prêmio Nobel e 163 pesquisadores brasileiros, assinaram uma petição exigindo que a CE disponibilizasse via *internet* os resultados de pesquisas acadêmicas custeadas por verba pública (<http://www.guardian.co.uk/business/2007/jan/30/research.highereducation>). Na medicina, outro exemplo desta manifestação é a revista científica de livre acesso e alto índice de impacto “Public Library of Science” (PLOS), uma biblioteca de livre acesso ao público (www.plos.org/; Butler, 2006 apud FIRME *et al*, 2008).

Com base nos trabalhos descritos anteriormente, percebe-se que as pessoas possuem acesso à *internet*, e que através dela distâncias e barreiras são rompidas. Na *internet* há entretenimento, comunicação, educação, ensino e pesquisa, estando ela no dia a dia, em todos os lugares, na vida das pessoas. Aproveitando o termo pesquisa e praticidade, percebeu-se que seria possível e oportuno conciliar a praticidade da *internet* com a falta de material (estudo e pesquisa) que apresentem as ferramentas do planejamento de experimentos fatorial para desenvolver um *software webonline* e disponibiliza-lo na *internet* as pessoas interessadas, sem fazer *login*, sem fazer *downloads* e sem custo de manuseio, e mostrar que é uma ferramenta de ensino e pesquisa simples, prática, ágil e eficiente em nível de pós-graduação de química, e que aumenta a pequena quantidade de opções de ferramentas de ensino e pesquisa para estudantes na área de planejamento de experimentos fatorial.

Para validação do *software* como ferramenta de ensino e pesquisa houve a colaboração e orientação do Prof. Dr. Márlon Herbert Flora Barbosa Soares que foi mediador dos conhecimentos necessários para introduzir na pós-graduação uma ferramenta gratuita de ensino e pesquisa a fim de visualizar cálculos e gráficos com interação entre os estudos adquiridos durante o curso de **quimiometria** e o *software*, pois o intuito não é apenas ser um *software* gratuito, e sim uma ferramenta de auxílio entre professores e alunos proporcionando interatividade, praticidade, além de promover ou requerer conhecimentos necessários para sua utilização. Realizaram-se as aplicações do *software* nas aulas da disciplina de Quimiometria II, Planejamento e Otimização de Experimentos, ministrada pelo Professor Dr. Anselmo Elcana de Oliveira no programa de pós-graduação de mestrado em Química da UFG e

Doutorado Multiinstitucional em Química UFU/UFMG/UFMT no segundo semestre dos anos de 2009, 2010 e 2011.

O *software* que contém as ferramentas do planejamento de experimentos fatorial do presente trabalho denominou-se de *Software PlanEx* de Planejamento de Experimentos Fatorial, e está disponibilizado no sítio www.quimica.ufg.br na página do professor Anselmo Elcana de Oliveira gratuitamente.

Para desenvolver a implementação do *software* PlanEx, escolheu-se a linguagem *Java*, pois *Java* oferece um grande campo de possibilidade de efeitos gráficos e uma grande quantidade de bibliotecas livres de cálculos estatísticos e de operações de matrizes. Além disso, o *software* será executado no *browser* do próprio cliente, sem a necessidade de realizar *downloads*. Uma das vantagens disto é que as pessoas interessadas no manuseio do *software* terão a disposição cálculos de planejamento de experimentos fatorial e gráficos de otimização de superfície de resposta num mesmo programa com interatividade, amigabilidade, agilidade e poderão utilizá-lo em qualquer lugar que tenha *internet*.

1.7. Triagem e Otimização

Dependendo do objetivo proposto num problema com planejamento de experimentos fatorial, necessita-se distinguir quais variáveis são realmente significativas num fenômeno em estudo, ou seja, quais desempenham alteração na resposta experimental ou na superfície de resposta. Para esse objetivo, o presente trabalho apresentou-se as ferramentas do planejamento de experimentos fatorial completo e fracionário para que o usuário realize uma triagem das variáveis. A triagem consiste em obter cálculos uteis para analisar quais são significativas, através da estimativa do erro obtida de todas as replicatas dos experimentos e da opção de gerar gráficos de probabilidade normal.

Para a etapa de otimização, o presente trabalho disponibilizou as ferramentas da RSM usando o modelo linear (planejamentos de experimentos fatorial com replicatas do ponto central) e o modelo quadrático (planejamento de experimentos fatorial acrescido do planejamento estrela com replicatas do ponto central, denominado de planejamento composto central) para que o usuário analise quais são os efeitos significativos através dos erros a partir das replicatas do ponto central, avalie o grau de curvatura da superfície de experimentos, e realize a otimização do fenômeno em estudo através do caminho de máxima inclinação e de gráficos de superfície de resposta. E através da ANOVA disponível na plataforma, o usuário tem a possibilidade de avaliar os modelos lineares e quadráticos da RSM.

No *software* PlanEx os coeficientes ou parâmetros da regressão dos modelos foram implementados usando o método dos mínimos quadrados ordinários (MQO), e a avaliação dos efeitos realizou-se de quatro maneiras distintas e independentes:

- empregando replicatas de todos os experimentos do planejamento de experimentos fatorial;
- incluindo replicatas de experimentos de ponto central da RSM;
- observando os maiores valores dos efeitos obtidos tanto no planejamento de experimentos fatorial como na RSM;
- construindo gráficos de probabilidade normal para os planejamentos de experimentos fatorial.

1.8. Tipos de Planejamentos Implementados na Plataforma PlanEx

O primeiro passo numa pesquisa é definir, com o máximo de clareza, o objetivo que se pretende alcançar com os experimentos, para a escolha de o planejamento ser a mais adequada (BRUNS *et al*, 2001). Se um pesquisador deseja descobrir qual a influência que uma variável provoca nas respostas experimentais e/ou o valor ou tipo da variável que resultará numa resposta máxima, rendimento, e/ou até quando ele poderá alterar as variáveis sem alterar o rendimento ou a qualidade do produto final, recomenda-se para realização do sistema o uso do planejamento de experimentos fatorial (BOX *et al*, 1978; BRUNS *et al*, 2001).

No presente estudo foi desenvolvido o *software* PlanEx que contém as ferramentas do planejamento de experimentos fatorial de dois níveis, com, ou sem, replicatas com até seis variáveis, e as ferramentas da RSM com até quatro variáveis usando o modelo linear e o modelo quadrático, como se comentou na seção 1.7. A quantidade de replicatas do planejamento de experimentos fatorial foi limitada em quatro, e a replicatas do ponto central na RSM foi limitada em cinco.

Na seção da Estatística Experimental apresenta-se as equações de todas as ferramentas da estatística, da álgebra linear com exemplos numéricos fictícios e de trabalhos obtidos na literatura usadas na implementação do *software* PlanEx. Na seção de Conceitos Básicos de Linguagem Computacional apresenta-se algumas definições importantes que foram usadas ao longo do trabalho. Já na seção Usando *Java* para Resolver o Problema apresenta-se de forma resumida como as ferramentas estatísticas foram implementadas na linguagem *Java* para desenvolver o *software* PlanEx. Na seção dos Resultados e Discussões apresenta-se ilustrações empregando o *software* PlanEx de vários trabalhos realizados por diferentes

autores e de problemas fictícios, e comenta-se a interpretação dos resultados numéricos. E na seção Conclusões faz-se uma abordagem contemplando as ferramentas estatísticas disponíveis no *software* com os resultados obtidos relacionando-os com os comentários dos usuários nas aplicações realizadas na pós-graduação para validar o *software* PlanEx como ferramenta de ensino e pesquisa.

2. ESTATÍSTICA EXPERIMENTAL

Apresenta-se inicialmente nesta parte do trabalho um referencial teórico da estatística experimental abordando os conceitos básicos da estatística, tais como população, amostra e a metodologia de como selecionar elementos nos diferentes tipos de amostragem, e analisar se há ligação ou relação linear entre duas variáveis através da correlação linear. Em sequência, apresentam-se as ferramentas da regressão linear e os pontos principais dos modelos estimados pela regressão através do MQO, abordando de modo detalhado e com exemplos numéricos os cálculos necessários para compreender as teorias de uma regressão linear.

Após a regressão, apresentam-se as teorias e ferramentas do planejamento de experimentos fatorial de dois níveis, destacando as particularidades do planejamento de experimentos fatorial completo e do fracionário, tanto nos cálculos para obter os efeitos e fazer estimativas dos erros através de replicatas como nos possíveis gráficos. Posteriormente, comenta-se o benefício de utilizar replicatas do ponto central na seção de planejamentos de experimentos fatorial com ponto central (nível zero).

Na fase de otimizar experimentos apresentam-se as ferramentas da RSM, tanto para obtenção dos efeitos como nos possíveis gráficos da RSM, e na fase de verificar a significância dos coeficientes da regressão na RSM destaca-se as equações da ANOVA com exemplos numéricos de problemas fictícios e de diferentes autores obtidos na literatura.

Por fim, apresentam-se a metodologia dos gráficos de probabilidade normal para analisar quais efeitos são significativos num planejamento de experimentos fatorial tanto do tipo completo como do fracionário.

2.1 *Conceitos Básicos da Estatística*

Como se comentou na introdução, a estatística é uma ferramenta essencial e necessária do método científico, e para entendê-la necessita-se de alguns conceitos básicos, tais como: população, amostra, variável, tipos de amostragem, medidas de tendência central, medidas de dispersão, correlação e regressão. A seguir apresentam-se tais conceitos com aplicações.

Na estatística, conjunto universo ou população é o todo que se pretende estudar, e uma parte desse todo se denomina de amostra. A finalidade de utilizar amostra é que em alguns casos só se tem acesso a ela, por exemplo, se o objetivo é saber a qualidade da água do rio São Francisco e como não se tem acesso a toda a água do rio, o uso de amostra é

imprescindível. Também, justifica-se o uso de amostra pelo aspecto financeiro e temporal (BRUNS *et al*, 2001; SILVA, E. , 1999; SPIEGEL, 2004).

Na estatística variável ou fator é o possível resultado de um fenômeno em estudo numa pesquisa científica, classifica-se em duas categorias (GUJARATI, 2006; BRUNS *et al*, 2001; SILVA, E., 1999; SPIEGEL, 2004):

- variável quantitativa: podendo ser discreta ou contínua;
- variável qualitativa.

Uma variável é classificada como quantitativa quando é representada por um número. Se o número for proveniente de uma medição provavelmente possuirá incertezas na precisão, assim costuma-se representá-lo por um intervalo numérico, e a variável classifica-se como sendo uma variável quantitativa contínua (SILVA, E., 1999; SPIEGEL, 2004).

Exemplo 1. Ao realizar experimentos num laboratório, se o conjunto universo for os experimentos realizados no mês de março do ano corrente, e se o objetivo é saber a quantidade de gramas do solvente X usada em todos os experimentos, tem-se um exemplo de variável quantitativa contínua, pois ao medir a quantidade de gramas necessita-se de um aparelho (balança), que fornecerá um resultado duvidoso, podendo ser o último dígito do número 3570 gramas, por exemplo. Por esse motivo não se utiliza um valor fixo para representar tal grandeza, e sim um intervalo numérico. Nesse caso, o valor poderia estar, por exemplo, no intervalo de 3200 gramas a 4200 gramas, (quanto menor a amplitude do intervalo melhor é a precisão), sendo o intervalo fechado ou aberto, pois a probabilidade de uma variável contínua num ponto específico é zero, como se apresentará na seção de gráficos de probabilidade normal na página 94 (SILVA, E., 1999; SPIEGEL, 2001).

Quando a variável quantitativa for proveniente de uma contagem ou enumeração finita a variável se denomina de variável quantitativa discreta e representa-se por um número específico (SILVA, E., 1999; SPIEGEL, 2004).

Exemplo 2. Ao realizar experimentos num laboratório, se o conjunto universo for os experimentos realizados no mês de março do ano corrente, e se o objetivo é saber a quantidade de experimentos realizados no laboratório no mês de março, tem-se uma variável quantitativa discreta, pois o valor é enumerável e finito, e representa-se por um número específico, por exemplo, 32 experimentos.

Quando a variável for definida por uma qualidade ou característica se denomina de variável qualitativa (SILVA, E., 1999; SPIEGEL, 2004).

Exemplo 3. Ao realizar experimentos num laboratório, se o conjunto universo for os experimentos realizados no mês de março do ano corrente, e se o objetivo é saber qual o tipo de ácido (forte, moderado ou fraco) usado nos experimentos no mês de março, tem uma variável qualitativa, pois se trata de uma qualidade ou classificação.

Percebe-se nos exemplos de 1 a 3 que o conjunto universo é o mesmo, mas o objetivo de cada exemplo difere, podendo alterar a classificação da variável em estudo.

Dependendo do tipo de estudo, necessita-se de específicos procedimentos para escolher os elementos da amostragem do fenômeno em estudo. E um dos requisitos básicos da escolha, é garantir a casualidade dos eventos, ou seja, garantir que os eventos sejam independentes. Os tipos de amostragem mais utilizados são:

- **Amostragem Casual ou Aleatória Simples** – utiliza-se quando a probabilidade de cada elemento da população é equivalente, e semelha ao sorteio lotérico, em que o procedimento consiste em enumerar os elementos da população e escolher ao acaso alguns desses elementos;
- **Amostragem Proporcional Estratificada** – utiliza-se quando existe uma divisão natural da população em grupos com número de elementos diversos, por exemplo, escolher 8 substâncias de 40, sendo 30 substâncias ácidas e 10 básicas.
- **Amostragem Sistemática** – utiliza-se quando a população já se encontra em ordem, por exemplo, residências numa rua - também há a possibilidade do pesquisador impor a sua metodologia, por exemplo, num laticínio o pesquisador pode escolher a cada hora 40 litros de leite para fazer uma análise da qualidade;
- **Amostragem por conveniência** – utiliza-se quando o pesquisador decide escolher a amostragem que já conhece.

No presente trabalho, o tipo de amostragem dependerá da escolha do usuário e do objetivo do estudo, porém em todos os tipos deve garantir que a escolha dos experimentos seja aleatória (estocástica).

Exemplo 4. Um pesquisador deseja analisar a influência da temperatura (20°C, 30°C, 40°C) e da concentração de um ácido (40 mg/cm³; 60 mg/cm³) na resposta experimental de uma variável que depende dessas duas variáveis independentes. Para a realização de tal tarefa necessita-se no mínimo de seis experimentos (3 níveis de temperaturas x 2 níveis de concentrações). A realização desses seis experimentos realiza-se de forma aleatória, como mostra a coluna 2 da Tabela 2.1.

Tabela 2.1- Planejamento de Experimentos do Exemplo 4

Experimento	Ordem de Realização do Experimento	Temperatura (°C)	Concentração (mg/cm ³)	Resposta Experimental
1	2	20	40	y ₁
2	3	20	60	y ₂
3	5	30	40	y ₃
4	6	30	60	y ₄
5	1	40	40	y ₅
6	4	40	60	y ₆

Percebe-se na Tabela 2.1 que há seis experimentos como é apresentado pela coluna 1. As colunas 3 e 4 representam as seis combinações possíveis das duas variáveis (a menor quantidade possível nesse caso de experimentos) para obter a resposta experimental. Ao longo deste trabalho, a resposta experimental representa-se pelo símbolo y_i , em que $i = 1, 2, \dots, n$ experimentos, e a matriz coluna ou vetor coluna de todos os experimentos se denota por $\mathbf{y}_{n \times 1}$.

Nota. Chama-se uma matriz linha ou uma matriz coluna de **vetor** na matemática, e se representa um **vetor** por letras **minúsculas em negritos**, e as matrizes que não são vetores representam-se por letras **maiúsculas em negritos**.

Na coluna 2 da Tabela 2.1 percebe-se que o experimento 2 foi realizado primeiro, seguido do experimento 3, do 5, do 6, do 1 e do 4. Para realizar o sorteio, devem-se enumerar os seis experimentos de 1 a 6 e colocar a enumeração numa urna, por exemplo, e escolher um experimento de seis (1/6). Em seguida, volta-se o bilhete sorteado para a urna para garantir que a probabilidade do próximo experimento, nesse caso, o experimento 3 tenha a mesma probabilidade do primeiro. Se sortear um mesmo número, repete-se o processo até obter os seis experimentos distintos com a mesma probabilidade, dessa forma garante-se que a escolha é casual e independente.

Na estatística trabalha-se frequentemente com uma grande quantidade de dados, no entanto, na fase de discussão e representação dos dados necessita-se das medidas de tendência central para resumir os dados, dentre as quais se destaca a média aritmética (BRUNS *et al*, 2001; SILVA, E., 1999; SPIEGEL, 2004).

A média aritmética \bar{x} de uma série de dados é dada pela razão do somatório de todas as n observações pela quantidade n de observações:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}. \quad (1)$$

Se houver repetições f_i de observações x_i a Eq. 1 torna-se:

$$\bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i}. \quad (2)$$

Exemplo 5. Um químico encontrou que o pH médio, $\overline{\text{pH}}$, da água de uma piscina ao longo de uma semana (7,1; 7,1; 7,4; 7,4; 7,3; 7,3; 7,3) é $\overline{\text{pH}} = 7,27$.

No Exemplo 5 obteve-se o pH médio, $\overline{\text{pH}} = 7,27$, através das Eq. 1 ou 2:

$$\overline{\text{pH}} = \frac{7,1+7,1+7,4+7,4+7,3+7,3+7,3}{7} = 7,27;$$

$$\overline{\text{pH}} = \frac{2 \cdot 7,1 + 2 \cdot 7,4 + 3 \cdot 7,3}{7} = 7,27.$$

A medida de tendência central mais utilizada é a média aritmética, porém a média só tem significado quando a distribuição dos dados é semelhante à forma de um sino, denominada de distribuição Normal ou distribuição de Gauss, como se apresentará na seção de gráficos de probabilidade normal na página 94. No presente trabalho, usou-se como medida de tendência central a média aritmética.

Para um químico saber qual é a variação de um determinado valor do pH do Exemplo 5 em relação a média, usa-se o desvio em relação a média, dado por:

$$d_i = x_i - \bar{x}. \quad (3)$$

Tabela 2.2 - Desvio em relação à média aritmética do Exemplo 5

	pH	pH - \bar{pH}
	7,1	-0,1714
	7,1	-0,1714
	7,4	0,1286
	7,4	0,1286
	7,3	0,0286
	7,3	0,0286
	7,3	0,0286
Soma	50,9	0,0000
Média	7,271429	

Percebe-se na Tabela 2.2 que a soma de todos os desvios em relação à média é nulo, que se denota por:

$$\sum_{i=1}^n d_i = \sum_{i=1}^n x_i - \bar{x} = 0. \quad (4)$$

Se houver repetição f_i de observações x_i a Eq. 4 torna-se:

$$\sum_{i=1}^n f_i d_i = \sum_{i=1}^n f_i \cdot (x_i - \bar{x}) = 0. \quad (5)$$

Na Tabela 2.2 têm-se sete observações ($n = 7$), porém através de apenas seis ($n-1 = 6$) é possível saber o valor que está faltando para o somatório da coluna 2 ser nulo. Isto se denomina de grau de liberdade (gl) e diz que o grau de liberdade da soma dos desvios em relação à média é $n - 1$.

Para um químico saber qual a variação dos dados (pH) em média, usa-se as medidas de dispersão mais utilizadas: a variância e o desvio padrão. A variância populacional σ^2 define-se pela razão dos somatórios dos desvios em relação à média ao quadrado pela quantidade n de observações:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}. \quad (6)$$

Se houver repetição f_i das observações x_i a Eq. 6 torna-se:

$$\sigma^2 = \frac{\sum_{i=1}^n f_i \cdot (x_i - \bar{x})^2}{n}. \quad (7)$$

As Eq. 6 e 7 são específicas da variância populacional σ^2 , a qual possui n grau de liberdade. No entanto, essas equações quando se referem à amostra possuem-se $n - 1$ grau de

liberdade, como se comentou nas Eq. 4 e 5. Representa-se a variância populacional pelo símbolo σ^2 , e a variância amostral pelo símbolo s^2 .

A variância é uma boa medida de dispersão em média, porém não se utiliza frequentemente, pois a unidade de medida da variância torna-se elevada ao quadrado da unidade de medida original dos dados, isso dificulta a análise dos dados. Para resolver esse problema de análise ou comparação, usa-se o desvio padrão populacional σ ou amostral s , que se define pela raiz quadrada da variância. A Tabela 2.3 apresenta a variância e o desvio padrão do Exemplo 5.

Tabela 2.3– Medidas de Dispersão: Variância e Desvio Padrão do Exemplo 5

	pH	pH - $\bar{\text{pH}}$	$(\text{pH} - \bar{\text{pH}})^2$
	7,1	-0,1714	0,0294
	7,1	-0,1714	0,0294
	7,4	0,1286	0,0165
	7,4	0,1286	0,0165
	7,3	0,0286	0,0008
	7,3	0,0286	0,0008
	7,3	0,0286	0,0008
Soma	50,9	0,0000	0,0943
Média	7,2714		
Variância σ^2	0,0135	Variância s^2	0,0157
Desvio Padrão σ	0,1161	Desvio Padrão s	0,1254

2.2 Correlação Linear e Regressão Multivariada

Em pesquisa científica depara-se com a análise de várias variáveis sobre um fenômeno em estudo. Para uma análise significativa, investiga se a variação de uma variável x afeta os dados de uma variável y , e de que forma. A análise realiza-se pelas medidas de correlação linear e pelas teorias da análise de regressão, como se apresentará a seguir.

2.2.1 Correlação Linear

O Coeficiente de Pearson γ é uma grandeza quantitativa que mede a intensidade da correlação (relação ou ligação) de duas variáveis linearmente (reta) nas n observações, dado por:

$$\gamma = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{\sqrt{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] \cdot \left[n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right]}}; \quad (8)$$

$$\gamma = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}.$$

O coeficiente de Pearson é um número que está dentre $-1 \leq \gamma \leq 1$ ou $-100\% \leq \gamma \leq 100\%$, e se $\gamma = 0$ afirma-se que não há correlação linear, porém pode existir outro tipo de correlação que não seja linear, por exemplo, correlação quadrática ou trigonométrica. Para realizar inferências estatísticas a intensidade da correlação deve estar no intervalo $0,6 \leq |\gamma| \leq 1$ para as previsões terem pesos significativos (SILVA, E., 1999; SPIEGEL, 2004).

Exemplo 6. Um químico encontrou que a intensidade da correlação linear entre a quantidade de fermento x_i utilizado na produção de pães y_i usando a Eq. 8 nos dados da Tabela 2.4 é 98,27%. Isso significa que ao usar 100% dos valores da quantidade de fermento x_i para prever ou explicar a quantidade de pães y_i acerta-se em média 98,27%, e vice-versa. Nesse exemplo considera-se que os outros ingredientes na produção de pães são proporcionais.

Tabela 2.4 – Correlação Linear das variáveis do Exemplo 6

	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
	3,5	1	3,5	12,25	1
	4,5	1,3	5,85	20,25	1,69
	5,5	1,5	8,25	30,25	2,25
	2,5	0,6	1,5	6,25	0,36
	3	0,9	2,7	9	0,81
Soma	19	5,3	21,8	78	6,11
Correlação	γ	0,9827			
Legenda	x: Fermento (1 tablete = 15 gramas)				
	y: Pães (1000 gramas)				

A partir de um gráfico de dispersão obtido da Tabela 2.4, percebe-se que os dados aproximam-se de uma reta, Figura 2.1, o motivo que justifica a nomenclatura: correlação linear.

Após saber a intensidade da correlação linear, é possível obter uma função, por exemplo, o modelo linear com dois parâmetros $y = \beta_1 x + \beta_0 + u$. A obtenção dessa função é um processo denominado de regressão, que descreve a forma ou maneira que as variáveis se relacionam, enquanto a correlação linear mede a intensidade ou grau dessa relação linear - reta (GUJARATI, 2006). Na regressão a variável y é aleatória (estocástica) e denomina-se de variável dependente, e a variável x é fixa ou não estocástica e denomina-se de variável independente. Já na correlação, as variáveis x e y são aleatórias entre si (GUJARATI, 2006).

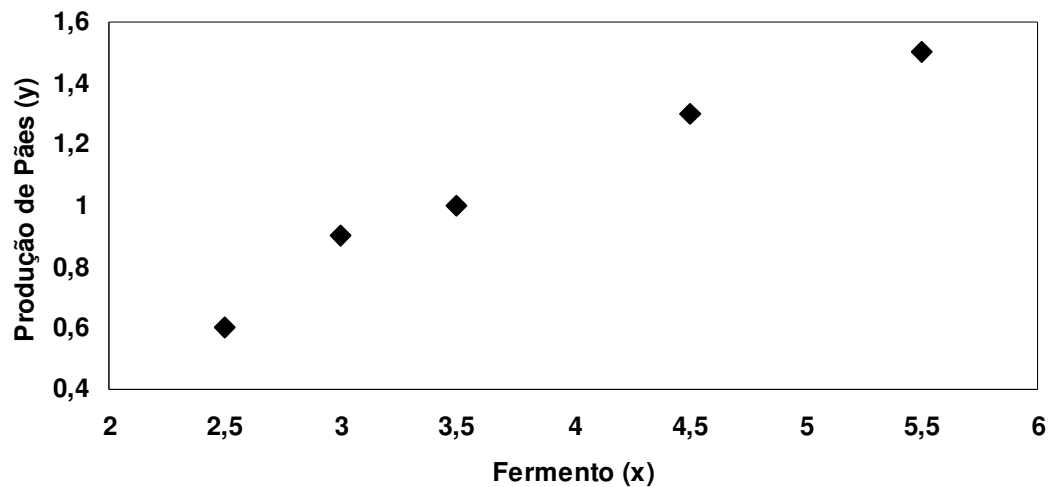


Figura 2.1 Gráfico da Dispersão das variáveis x e y da Tabela 2.4 do Exemplo 6

Na próxima seção apresenta-se inicialmente o modelo de regressão mais simples $y = \beta_1 x + \beta_0 + u$ denominado de regressão bivariada ou regressão linear simples, em seguida apresenta-se o modelo de regressão linear multivariada (mais de uma variável independente) que é uma extensão do modelo de regressão linear simples.

2.2.2 Regressão Linear Simples

Para um químico realizar previsões usando os dados da Tabela 2.4, obrigatoriamente o modelo linear mais simples apresenta-se na forma da Eq. 9 e denomina-se de função de regressão populacional (FRP):

$$y = \beta_0 + \beta_1 x + u. \quad (9)$$

Onde:

- y é uma função de x , que se denota por $y(x)$;

- β_0 e β_1 são os coeficientes populacionais da regressão;
- u é o erro aleatório populacional ou resíduo aleatório populacional.

Rotineiramente trabalha-se com amostragem, isso implica que a Eq. 9, FRP, não é observada diretamente. Com base nisso, obtém-se uma função estimada através da amostragem, a qual se denomina de função de regressão amostral (FRA),

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x. \quad (10)$$

Onde:

- \hat{y} é a FRA que depende de x , o símbolo $\hat{}$ significa estimado e lê-se y chapéu;
- $\hat{\beta}_0$ e $\hat{\beta}_1$ são os coeficientes amostrais da regressão.

Na Eq. 10, se adicionar o erro aleatório amostral \hat{u} , obtém-se a FRP a partir de parâmetros amostrais,

$$y = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{u}. \quad (11)$$

Onde:

- y é a FRP que depende de x ;
- $\hat{\beta}_0$ e $\hat{\beta}_1$ são os coeficientes amostrais da regressão;
- \hat{u} é o erro aleatório amostral ou resíduo aleatório amostral.

Com base no modelo $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ e nos dados do Exemplo 6, a quantidade de pães depende da quantidade de fermento, e não poderia ser diferente. Na correlação, as variáveis x e y são aleatórias, na regressão não (GUJARATI, 2006).

Substituindo a Eq. 10 na Eq. 11 encontra-se o resíduo aleatório amostral em função da FRP e da FRA,

$$\hat{u} = y - \hat{y}. \quad (12)$$

Para obter os coeficientes ou parâmetros (Betas) da regressão, usa-se o MQO, que é um dos métodos de análise de regressão mais poderosos e difundidos, e é atribuído a Carl Friedrich Gauss (GUJARATI, 2006). Os coeficientes ou parâmetros da regressão estimados pelo MQO são obtidos usando a Eq. 13, cuja demonstração está no **Apêndice A**:

$$\hat{\beta}_1 = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}; \quad (13)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

Exemplo 7. Ao considerar o modelo $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ e aplicar a Eq. 13 nos dados da Tabela 2.5 o químico obtém a FRA dada por $\hat{y} = -0,0276 + 0,2862x$. A Tabela 2.5 apresenta os valores estimados \hat{y}_i , os valores populacionais y_i e o resíduo aleatório amostral \hat{u}_i .

Tabela 2.5 – Valores estimados e resíduos aleatórios da regressão usando MQO do Exemplo 6

	x_i	y_i	\hat{y}_i	$\hat{u} = y_i - \hat{y}_i$
	3,5	1	0,9741	0,0259
	4,5	1,3	1,2603	0,0397
	5,5	1,5	1,5466	-0,0466
	2,5	0,6	0,6879	-0,0879
	3	0,9	0,8310	0,0690
Soma	19,0000	5,3000	5,3000	0,0000

Usa-se frequentemente o gráfico de resíduos em função dos valores estimados para analisar se o modelo da regressão considerado apresenta indícios de aspectos tendenciosos. Um modelo bom de regressão apresenta-se graficamente resíduos aleatórios, ou seja, o gráfico de resíduos em função dos valores estimados deve ser estocástico (GUJARATI, 2006; BRUNS *et al*, 2001). Na Figura 2.2 apresenta-se o gráfico de resíduos aleatório em função dos valores estimados da Tabela 2.5.

Percebe-se no gráfico da Figura 2.2 que os resíduos são aleatórios, indicando que o modelo considerado é bom. Caso contrário, o modelo seria ruim e/ou a quantidade de variáveis explicativas (independentes) é insuficiente, podendo o pesquisador julgar em alterar o tipo do modelo ou expandir o modelo com o acréscimo de variáveis.

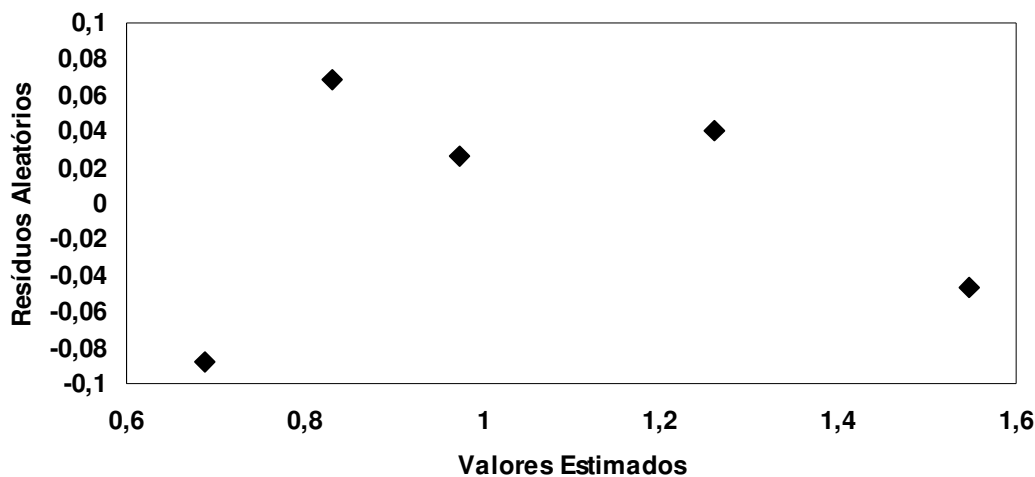


Figura 2.2 – Gráfico de resíduos aleatório em função dos valores estimados da Tabela 2.5

2.2.3 Regressão Linear Multivariada

Uma FRP para duas variáveis independentes é dada por:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u. \quad (14)$$

No dia a dia trabalha-se com amostragem, isso implica que a Eq. 14, FRP, não é obtida através dos parâmetros populacionais, e sim através dos parâmetros amostrais. A FRA para duas variáveis é dada por:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2. \quad (15)$$

Na Eq. 15, se adicionar o erro aleatório amostral \hat{u} , obtém-se a FRP a partir de parâmetros amostrais,

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{u}. \quad (16)$$

Na **econometria** o modelo linear estimado para duas variáveis apresenta-se na forma da Eq. 15, e denomina-se de modelo linear para duas variáveis aditivo. Porém, na **quimiometria** o modelo linear para duas variáveis apresenta-se na forma da Eq. 17 que é o resultado do acréscimo da interação entre as variáveis com a Eq. 15, pois na química pretende-se investigar se as variáveis possuem sinergismo entre si:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_1 x_2. \quad (17)$$

Onde $\hat{\beta}_3$ é o coeficiente da interação entre as variáveis x_1 e x_2 , e se $\hat{\beta}_3$ não for significativo o modelo é aditivo e representado pela Eq. 15.

Em notação matricial a Eq. 17 é dada pela Eq. 18, onde a simbologia usada está de acordo com a **Nota** na página 33,

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}. \quad (18)$$

Onde:

- $\hat{\mathbf{y}}$ é a matriz coluna (vetor) de todas as i -ésima observações;
- \mathbf{X} é a matriz de todas as variáveis e interações entre elas;
- $\hat{\boldsymbol{\beta}}$ é a matriz coluna (vetor) que apresenta os coeficientes da regressão da FRA.

A dimensão da matriz coluna $\hat{\mathbf{y}}$ dependerá da quantidade de n experimentos realizados. A dimensão da matriz \mathbf{X} dependerá do modelo considerado, se o modelo for igual a Eq. 15 a matriz \mathbf{X} possuirá n linhas e 3 colunas (3 coeficientes), e se for igual a Eq. 17 a matriz \mathbf{X} possuirá n linhas e 4 colunas (4 coeficientes), e assim por diante. A matriz coluna $\hat{\boldsymbol{\beta}}$ dependerá do modelo utilizado, se o modelo for igual a Eq. 15 a dimensão será 3 linhas

(3coeficientes) x 1 coluna, e se for igual a Eq. 17 a dimensão será 4 linhas (4 coeficientes) x 1 coluna. Mais detalhes no **Apêndice B** e na seção 4.

Os parâmetros de uma análise de regressão multivariada são obtidos pelo MQO dados pela Eq. 19, como se apresenta no **Apêndice B**:

$$\hat{\beta} = (\mathbf{X}^t\mathbf{X})^{-1} \mathbf{X}^t\mathbf{y}. \quad (19)$$

Onde a simbologia usada está de acordo com a **Nota** na página 33 e:

- $\hat{\beta}$ é a matriz coluna (vetor) dos coeficientes da FRA;
- \mathbf{X} é a matriz das variáveis independentes e das interações entre elas;
- \mathbf{X}^t é a matriz transposta da matriz \mathbf{X} ;
- $(\mathbf{X}^t\mathbf{X})^{-1}$ é a matriz inversa da multiplicação da matriz transposta de \mathbf{X} pela matriz \mathbf{X} ;
- \mathbf{y} é a matriz coluna (coluna) que apresenta as respostas experimentais populacionais.

Exemplo 8. No trabalho de Peralta-Zamora *et al* (2005) os autores realizaram uma regressão multivariada para aperfeiçoar o percentual de degradação de fenóis totais em 30 minutos de tratamento, usando como variáveis independentes o pH e a concentração de peróxido de hidrogênio, como apresenta-se na Tabela 2.6. O modelo estimado obtido mediante o uso da Eq. 15 apresenta-se na forma $\hat{y} = 73,1429 - 6,5000x_1 - 3,0000x_2$, e através da Eq. 17 $\hat{y} = 73,1429 - 6,5000x_1 - 3,0000x_2 - 15,0000x_1x_2$.

Tabela 2.6 – Planejamento de Experimentos da análise de regressão multivariada do Exemplo 8

H ₂ O ₂ (mg/L)	pH	Percentual de Degradação de fenóis em 30 minutos de tratamento	x ₁ *	x ₂ **	X ₁ X ₂ ***
100	4	62	-1	-1	1
300	4	79	1	-1	-1
100	8	86	-1	1	-1
300	8	43	1	1	1
200	6	80	0	0	0
200	6	80	0	0	0
200	6	82	0	0	0

* é a codificação da concentração de peróxido de hidrogênio (água oxigenada); ** é a codificação do pH e *** representa a interação entre as variáveis obtida através da multiplicação algébrica dos sinais das colunas 4 e 5.

A maior parte das variáveis independentes da estatística experimental na química é controlável pelo pesquisador, logo é possível planejar os experimentos. Observa-se na Tabela 2.6 que as variáveis concentração de peróxido de hidrogênio (H₂O₂) e pH foram selecionadas em dois níveis diferentes (100 mg/L e 300 mg/L para a concentração de H₂O₂ e, 4 e 8 para o

pH) acrescentando ao planejamento experimental triplicatas do ponto central, sendo 200 mg/L e 6 os pontos centrais das variáveis concentração de H₂O₂ e pH, respectivamente.

No planejamento de experimentos de dois níveis e de k variáveis, que se denota por 2^k realiza-se rotineiramente uma codificação das variáveis, atribuindo as variáveis quantitativas com nível maior o valor de +1 e ao nível menor o valor de -1. Quando a variável for qualitativa, por exemplo, tipo de catalisador, a escolha não influenciará na análise dos dados e fica arbitrária ao usuário, pois não existem valores menores ou maiores. Com este raciocínio, é possível construir as colunas de sinais para todas as interações e, assim, elaborar a matriz de coeficientes de contraste, Tabela 2.6 (BOX *et al*, 1978; BRUNS *et al*, 2001).

Ao se trabalhar com planejamentos de experimentos com mais de dois níveis usa-se a Eq. 20 para realizar as codificações das variáveis:

$$x_{\text{Nível}} = \frac{v_{\text{Nível}} - \bar{v}}{\bar{v} - v_{\text{Nível_Menor}}} \quad (20)$$

Onde:

- $x_{\text{Nível}}$ é a codificação do nível desejado;
- $v_{\text{Nível}}$ é o valor experimental da grandeza física realizada no nível i;
- \bar{v} é a média de todos os valores experimentais da grandeza física de uma única variável codificada por vez;
- $v_{\text{Nível_Menor}}$ é o menor valor de todos os valores experimentais da grandeza codificada por vez.

Ao aplicar a Eq. 20 nas colunas 1 e 2 da Tabela 2.6 obtém-se, respectivamente, as colunas 4 e 5 da Tabela. Obteve-se a coluna 6 multiplicando as colunas 4 (x_1) e 5 (x_2) da Tabela para investigar se as variáveis possuem interação significativa usando o modelo da Eq. 17. Despreza-se a coluna 6 ao considerar o modelo da Eq. 15.

Mais detalhe da codificação da Tabela 2.6 apresenta-se a seguir:

- A codificação da variável x_1 - aconteceu em três níveis (100, 200 e 300), sendo o nível menor $v_{\text{Nível_Menor}} = 100$ e a média deles $\bar{v} = 200$:

$$x_{\text{Nível}} = \frac{v_{\text{Nível}} - 200}{200 - 100} \Rightarrow x_{\text{Nível}} = \frac{v_{\text{Nível}} - 200}{100} \left\{ \begin{array}{l} x_{\text{Nível_Menor}} = \frac{100 - 200}{100} = -1 \\ x_{\text{Nível_Central}} = \frac{200 - 200}{100} = 0 ; \\ x_{\text{Nível_Maior}} = \frac{300 - 200}{100} = 1 \end{array} \right.$$

- o A codificação da variável x_2 - aconteceu em três níveis (4, 6 e 8), sendo o nível menor

$v_{\text{Nível_Menor}} = 4$ e a média deles $\bar{v} = 6$:

$$x_{\text{Nível}} = \frac{v_{\text{Nível}} - 6}{6 - 4} \Rightarrow x_{\text{Nível}} = \frac{v_{\text{Nível}} - 6}{2} \begin{cases} x_{\text{Nível_Menor}} = \frac{4 - 6}{2} = -1 \\ x_{\text{Nível_Central}} = \frac{6 - 6}{2} = 0 \\ x_{\text{Nível_Maior}} = \frac{8 - 6}{2} = 1 \end{cases} .$$

Os cálculos para obter as funções estimadas do Exemplo 8 apresentam-se a seguir.

Eq. 17 – Modelo Completo: $\hat{y} = 73,1429 - 6,5000x_1 - 3,0000x_2 - 15,0000x_1x_2$.

A matriz \mathbf{X} denomina-se de matriz de contrastes do planejamento de experimentos. Neste caso possui sete linhas e quatro colunas, sendo a coluna 1 acrescentada para calcular a média dos experimentos \bar{y} ou o coeficiente independente do modelo, a coluna 2 refere-se à codificação da variável x_1 , a 3 a x_2 , e a 4 a x_1x_2 .

$$\mathbf{X}_{7 \times 4} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} ; \mathbf{X}'\mathbf{y} = \begin{bmatrix} 512 \\ -26 \\ -12 \\ -60 \end{bmatrix} ; \hat{\boldsymbol{\beta}} = \begin{bmatrix} 73,1429 \\ -6,5000 \\ -3,0000 \\ -15,0000 \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 0,1429 & 0,0000 & 0,0000 & 0,0000 \\ 0,0000 & 0,2500 & 0,0000 & 0,0000 \\ 0,0000 & 0,0000 & 0,2500 & 0,0000 \\ 0,0000 & 0,0000 & 0,0000 & 0,2500 \end{bmatrix} \begin{bmatrix} 512 \\ -26 \\ -12 \\ -60 \end{bmatrix} = \begin{bmatrix} 73,1429 \\ -6,5000 \\ -3,0000 \\ -15,0000 \end{bmatrix} .$$

Eq. 15 – Modelo aditivo: $\hat{y} = 73,1429 - 6,5000x_1 - 3,0000x_2$.

A matriz \mathbf{X} do modelo aditivo possui sete linhas e três colunas, sendo a coluna 1 acrescentada para calcular a média dos experimentos \bar{y} ou o coeficiente independente do modelo, a coluna 2 refere-se à codificação da variável x_1 e a 3 a x_2 .

$$\mathbf{X}_{7 \times 3} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$(\mathbf{X}^t \mathbf{X})^{-1} = \begin{bmatrix} 0,1429 & 0,0000 & 0,0000 \\ 0,0000 & 0,2500 & 0,0000 \\ 0,0000 & 0,0000 & 0,2500 \end{bmatrix}; \mathbf{X}^t \mathbf{Y} = \begin{bmatrix} 512 \\ -26 \\ -12 \end{bmatrix}; \hat{\boldsymbol{\beta}} = \begin{bmatrix} 73,1429 \\ -6,5000 \\ -3,0000 \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}$$

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 0,1429 & 0,0000 & 0,0000 \\ 0,0000 & 0,2500 & 0,0000 \\ 0,0000 & 0,0000 & 0,2500 \end{bmatrix} \begin{bmatrix} 512 \\ -26 \\ -12 \end{bmatrix} = \begin{bmatrix} 73,1429 \\ -6,5000 \\ -3,0000 \end{bmatrix}.$$

2.3 Planejamento de Experimentos Fatorial

Na introdução comentou-se que o ato de planejar está intimamente relacionado com o sucesso e com os objetivos de um fenômeno em estudo em todas as ciências. Também comentou-se anteriormente que existem variáveis que o pesquisador possui o controle sobre elas. Com base nisso, o pesquisador pode planejar os experimentos definindo os níveis das variáveis em estudo e abordando todas as possíveis combinações das variáveis independentes controláveis. Dessa forma, ele pode analisar quais variáveis (fatores) e/ou quais interações das variáveis têm efeitos significativos na resposta experimental através do uso de replicatas e/ou do uso de gráficos de probabilidade normal, como se comentará ao longo da seção da Estatística Experimental. Este planejamento de experimentos denomina-se de planejamento de experimentos fatorial.

Um dos passos iniciais na **quimiometria** na área de planejamento de experimentos fatorial é definir a quantidade de níveis que cada variável será planejada para realizar a quantidade mínima de experimentos, ou seja, os valores que as variáveis assumem ou categoria das variáveis, no caso da variável ser qualitativa. No Exemplo 4, na página 33, a quantidade mínima de experimentos realizados foi seis (3 níveis de temperaturas x 2 níveis de concentrações = 6 experimentos no mínimo). Com o mesmo raciocínio, um químico pode investigar a influência que as variáveis temperatura (40°C, 60°C, 80°C e 100°C) e tipos de

catalisadores (A, B e C) provocam nas respostas experimentais, para isso necessita manipular ou variar de nível, e observar o efeito que essa variação produz sobre a resposta (BRUNS *et al*, 2001). Para execução de determinada tarefa realiza-se todos os experimentos nos diferentes tipos de níveis, ou seja, no mínimo 12 experimentos ($12 = 4 \times 3$, 4 níveis de temperatura e 3 níveis de catalisadores). A quantidade mínima de experimentos é dada por (BOX *et al*, 1978; BRUNS *et al*, 2001): $n_1 \times n_2 \times \dots \times n_k$, onde n_1 é a quantidade de níveis da variável x_1 , n_2 é a quantidade de níveis da variável x_2 , ..., e n_k da variável x_k .

Os planejamentos de experimentos fatoriais de dois níveis são úteis em investigações preliminares e simples de executar, quando pretende saber se determinadas variáveis tem ou não influência sobre a resposta experimental, e que podem ser ampliados a fim de conhecer melhor a relação entre a resposta e as variáveis importantes (BOX *et al*, 1978; BRUNS *et al*, 2001). Neste trabalho inicia-se com planejamentos de dois níveis e estende-se aos planejamentos fatoriais com pontos centrais apresentados na seção da RSM.

2.3.1 Planejamento de Experimentos Fatorial de Dois Níveis

Considerando planejamentos de experimentos de dois níveis com k variáveis, a quantidade mínima de experimentos para serem realizados obtém-se pelo produto de cada nível das k variáveis, que se denota pela Eq. 21:

$$n = \underbrace{2 \cdot 2 \cdot 2 \cdot 2}_{k} = 2^k. \quad (21)$$

Onde n é a quantidade mínima de experimentos realizados.

Um planejamento com a quantidade definida pela Eq. 21 denomina-se Planejamento de Experimentos Fatorial Completo 2^k (BOX *et al*, 1978).

Um planejamento de experimentos fatorial é uma regressão com variáveis codificadas como se apresentou no Exemplo 8, página 42, sendo as variáveis de entrada do sistema (variáveis independentes) controláveis pelo pesquisador e as variáveis de saída do sistema (variável dependente) obtida ao realizar os experimentos pelo pesquisador.

Percebe-se no Exemplo 8, página 42, que através da regressão é possível obter uma função estimada para um fenômeno em estudo envolvendo planejamento de experimentos fatorial. No entanto, antes de estudar essa relação com mais detalhes, apresenta-se a partir de agora um modo de obter o efeito principal e os efeitos de interações de variáveis, para analisar se tais variáveis afetam a resposta experimental através das Eq. 22 e 23.

Exemplo 9. Um químico deseja analisar o efeito que a temperatura (60°C e 80°C) e/ou o tipo de catalisador (A e B) provoca no rendimento (em gramas) de uma resposta experimental através do planejamento de experimentos fatorial 2^k . Para isso ele realiza no mínimo quatro experimentos (2^2 , dois níveis de duas variáveis), como se apresenta na Tabela 2.7.

Tabela 2.7 - Matriz de Planejamento de Experimentos Fatorial 2^2 do Exemplo 9

Experimento	Temperatura	Catalisador	Rendimento (g)
1	60 °C	A	55
2	80 °C	A	88
3	60 °C	B	51
4	80 °C	B	60

Percebe-se que através do planejamento de experimentos fatorial 2^k aplicado na Tabela 2.7 os experimentos foram realizados de forma organizada abordando todo o domínio experimental. A dimensão dos cálculos da Tabela 2.7 é pequena, permitindo fazer algumas análises sem realizar cálculos estatísticos. Percebe-se que há um aumento de 33 g quando se varia a temperatura de 60°C para 80°C (88 g – 55 g = 33 g, média do rendimento dos experimentos 1 e 2); no entanto, nos experimentos 3 e 4 a variação da temperatura de 60°C para 80°C (60 g – 51 g = 9 g, média do rendimento dos experimentos 3 e 4) provoca um aumento no rendimento de apenas 9 g; com base nisso, infere-se que o rendimento dos experimentos não depende apenas da temperatura, também depende do tipo de catalisador utilizado, o que indica que as variáveis possuem sinergismo entre si, modelo da Eq. 17. Para realizar uma investigação com mais detalhes, realiza-se uma codificação de variáveis.

A codificação das variáveis da Tabela 2.8 realizou-se através da Eq. 20 aplicando-a na Tabela 2.7, da mesma maneira como se desenvolveu na Tabela 2.6 do Exemplo 8, página 42, visto que na Tabela 2.8 acrescentou-se a coluna 1 para cálculo da média dos experimentos \bar{y} , como comentou-se na matriz de contraste X do Exemplo 8.

Tabela 2.8 - Matriz de Contrastes do Planejamento de Experimentos Fatorial 2^2 do Exemplo 9

Média	x_1^*	x_2^{**}	$x_1x_2^{***}$	Símbolo dos Experimentos	Rendimento (g)
1	-1	-1	1	(1)	55
1	1	-1	-1	x_1	88
1	-1	1	-1	x_2	51
1	1	1	1	x_1x_2	60

* é a codificação da temperatura; ** é a codificação do catalisador; e *** é a codificação da interação da temperatura com o catalisador obtida pela multiplicação de x_1 e x_2 .

Com base na Tabela 2.8 é possível obter a média de todos os experimentos e os efeitos que cada variável sozinha (efeito principal) ou junta (interação – sinergismo entre as variáveis) provoca na resposta experimental. Na coluna 5 da Tabela estão os códigos dos experimentos, em que:

- (1) significa o(s) experimento(s) realizado(s) no nível menor das variáveis;
- x_1 significa o(s) experimento(s) realizado(s) no nível maior da variável x_1 e menor das demais;
- x_2 significa o(s) experimento(s) realizado(s) no nível maior da variável x_2 e menor das demais;
- x_1x_2 significa o(s) experimento(s) realizado(s) no nível maior das variáveis x_1 e x_2 .

A média dos experimentos obtém-se usando a Eq. 22:

$$Ef_{\text{médio}} = \frac{\sum_{i=1}^n y_i}{n}. \quad (22)$$

Onde:

- $Ef_{\text{médio}}$ é o efeito médio ou média dos experimentos y_i que se denota por \bar{y} ;
- y_i são as respostas experimentais (variáveis de entrada ou variáveis dependentes);
- n é a quantidade total de experimentos realizados, sem levar em consideração se houve replicatas.

O efeito principal de cada variável e o efeito de interação entre as variáveis são obtidos pela Eq. 23,

$$Ef = \frac{\sum_{i=1}^{n/2} y_{i(+)} - \sum_{i=1}^{n/2} y_{i(-)}}{n/2}. \quad (23)$$

Onde:

- Ef é o efeito obtido da variável analisada por vez;
- $y_{i(+)}$ são as respostas experimentais realizadas no nível maior;
- $y_{i(-)}$ são as respostas experimentais realizadas no nível menor;
- n é a quantidade total de experimentos realizados, sem levar em consideração se houve replicatas.

Aplicando a Eq. 22 no Exemplo 9, Tabela 2.8, especificamente nas colunas 1 e 6, encontra-se que o valor médio dos experimentos, (\bar{y}) , é 63,5 g, ou seja, a média dos quatro experimentos é 63,5 g:

$$\begin{aligned} \text{Ef}_{\text{médio}} &= \frac{55+88+51+60}{4} \\ \text{Ef}_{\text{médio}} &= 63,5 \text{ g} \\ \bar{y} &= 63,5 \text{ g} \end{aligned}$$

Usando a Eq. 23 nas colunas 2 e 6 na Tabela 2.8, encontra-se que o X_1 (efeito principal da variável x_1 - Temperatura) aumenta em média 21 g o rendimento dos experimentos, quando se passa do nível menor da temperatura para o nível maior:

$$\begin{aligned} \text{Ef} &= \frac{\sum_{i=1}^{n/2} y_{i(+)} - \sum_{i=1}^{n/2} y_{i(-)}}{n/2} \\ \text{Ef}_{(x_1)} &= \frac{(88+60)-(51+55)}{2} \\ \text{Ef}_{(x_1)} &= 21 \text{ g} \\ X_1 &= 21 \text{ g.} \end{aligned}$$

Usando a Eq. 23 nas colunas 3 e 6 na Tabela 2.8, encontra-se que o X_2 (efeito principal da variável x_2 - Catalisador) diminui em média 16 g o rendimento dos experimentos, quando se passa do nível menor do catalisador para o nível maior:

$$\begin{aligned} \text{Ef} &= \frac{\sum_{i=1}^{n/2} y_{i(+)} - \sum_{i=1}^{n/2} y_{i(-)}}{n/2} \\ \text{Ef}_{(x_2)} &= \frac{(51+60)-(55+88)}{2} \\ \text{Ef}_{(x_2)} &= -16 \text{ g} \\ X_2 &= -16 \text{ g.} \end{aligned}$$

Usando a Eq. 23 nas colunas 4 e 6 na Tabela 2.8, encontra-se que o X_1X_2 (efeito de interação da variável x_1 com a variável x_2 - Temperatura com o Catalisador) diminui em média 12 g o rendimento dos experimentos, quando se passa do nível menor para o nível maior ao mesmo tempo nas duas variáveis:

$$\text{Ef} = \frac{\sum_{i=1}^{n/2} y_{i(+)} - \sum_{i=1}^{n/2} y_{i(-)}}{n/2}$$

$$\text{Ef}_{(x_1 x_2)} = \frac{(60+55) - (88+51)}{2} = -12 \text{ g}$$

$$X_1 X_2 = -12 \text{ g.}$$

O efeito médio dos experimentos, $\bar{y} = 63,5 \text{ g}$, é o coeficiente independente $\hat{\beta}_0$ do modelo linear $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$, $\hat{\beta}_0 = \bar{y} = 63,5 \text{ g}$. O efeito $X_1 = 21 \text{ g}$ representa a temperatura necessária (variável x_1) que deve fornecer para que a resposta experimental passe do nível menor para o nível maior, ou seja, **duas** unidades de medidas da codificação da variável. No modelo $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$ o coeficiente $\hat{\beta}_1$ significa o valor necessário para que a variável varie **uma** unidade de medida de codificação da variável, logo através de uma igualdade algébrica simples percebe-se que o coeficiente da variável x_1 é igual à metade do efeito obtido da variável x_1 , que se denota por $\hat{\beta}_1 = \frac{X_1}{2} = 10,5 \text{ g}$. O mesmo acontece para os outros coeficientes da regressão, exceto para $\hat{\beta}_0$. O modelo estimado com os coeficientes brutos (sem a unidade de medida) para o Exemplo 9 é

$$\hat{y} = 63,5 + \frac{21}{2} x_1 - \frac{16}{2} x_2 - \frac{12}{2} x_1 x_2 \text{ ou } \hat{y} = 63,5 + 10,5 x_1 - 8 x_2 - 6 x_1 x_2.$$

Nota. O símbolo x_i (minúsculo) representa a codificação da variável i , enquanto X_i (maiúsculo) representa o efeito da variável. No exemplo 9 a unidade da grandeza escalar considerada como variável dependente, rendimento, é a grama, no entanto o raciocínio das Eq. 22 e 23 aplica-se a qualquer tipo de grandeza escalar.

Através de experimentos realizados com repetições pode estimar o erro experimental de uma resposta individual e analisar se há erro experimental que possam ser atribuídos à ação das variáveis (FERREIRA, J. *et al*, 2007; BRUNS *et al*, 2001), e ainda analisar quais coeficientes ou parâmetros da regressão são significativos.

As estimativas dos erros para os planejamentos de experimentos fatoriais foram implementados de quatro maneiras independentes, como se comentou na seção 1.7 da Introdução, página 27:

- empregando replicatas de todos os experimentos do planejamento de experimentos fatorial;
- incluindo replicatas de experimentos de ponto central da RSM;
- observando os maiores valores dos efeitos obtidos tanto no planejamento de experimentos fatorial como na RSM;
- construindo gráficos de probabilidade normal para os planejamentos de experimentos fatorial.

Para obter a estimativa do erro experimental usando r replicatas de todos os experimentos utiliza-se a Eq. 7 substituindo o grau de liberdade populacional pelo amostral, com base nisso a Eq. 7 torna-se:

$$s_i^2 = \frac{\sum_{i=1}^r (y_i - \bar{y}_i)^2}{(r-1)}. \quad (24)$$

Se forem realizados r duplicatas em n experimentos diferentes, a quantidade total de experimentos é dada pelo produto: $n \cdot r$. Realizando r replicatas nos n experimentos diferentes obtém-se n variâncias. Com isto é possível obter uma média ponderada das variâncias que se denomina de variância conjunta dos efeitos, em que os pesos da média ponderada é o grau de liberdade amostral da quantidade de replicatas em cada nível de experimentos:

$$s_{\text{Conjunta}}^2 = \frac{\sum_{i=1}^n [(r_i - 1) \cdot s_i^2]}{\sum_{i=1}^n (r_i - 1)}. \quad (25)$$

No início da seção da Estatística Experimental apresentou-se que a variância é uma medida de dispersão pouco utilizada devido à dificuldade de realizar análise, pois a unidade de medida de dispersão está elevada ao quadrado da unidade medida original dos experimentos, porém tal dificuldade resolve-se pelo uso do desvio padrão, que se define pela raiz quadrada da variância. Através da variância conjunta torna possível obter uma estimativa do erro do efeito médio dos experimentos definido pela Eq. 26 e do erro do efeito de cada variável e de suas interações definido pela Eq. 27:

$$s_{\text{Média}}^2 = \frac{s_{\text{Conjunta}}^2}{n \cdot r} \Rightarrow s_{\text{Média}} = \sqrt{s_{\text{Média}}^2}; \quad (26)$$

$$s_{\text{Efeitos}}^2 = \frac{s_{\text{Conjunta}}^2}{\frac{n}{2}} \Rightarrow s_{\text{Efeitos}} = \sqrt{s_{\text{Efeitos}}^2}. \quad (27)$$

Exemplo 10. Um químico realizou replicatas de todos os experimentos, duplicatas ($r = 2$), para obter além dos efeitos principais e dos efeitos de interações, uma estimativa do erro experimental do rendimento de uma resposta experimental afeta pela temperatura e pelo tipo de catalisador, como se apresenta na Tabela 2.9.

Tabela 2.9 - Matriz de Planejamento de Experimentos Fatorial 2^2 do Exemplo 10

Experimento	Temperatura	Catalisador	Rendimento (g)		Média (g)
1	60 °C	A	55	59	57
2	80 °C	A	88	84	86
3	60 °C	B	51	59	55
4	80 °C	B	60	64	62

Para um planejamento de experimentos fatorial 2^k usa-se frequentemente uma **representação geométrica** para melhor visualizar e analisar o efeito obtido ao variar uma ou mais variáveis ao mesmo tempo num fenômeno em estudo. Para um planejamento de experimentos fatorial 2^2 a representação geométrica consiste num quadrado disposto num plano cartesiano, como se apresenta na Figura 2.3. E para um planejamento de experimentos fatorial 2^3 a representação geométrica consiste num cubo disposto num espaço cartesiano, como se apresenta na Figura 2.5.

Para calcular os efeitos quando tem replicatas, considera-se a média de cada experimento i , ou seja, a coluna 6 da Tabela 2.9. Usando a codificação das variáveis através da Eq. 20, a matriz de planejamento da Tabela 2.9 pode ser escrita da forma da Tabela 2.10 – Matriz de Contrastes: colunas 1 a 4.

Tabela 2.10 - Matriz de Contrastes de Planejamento de Experimentos Fatorial 2^2 do Exemplo 10

Média	x_1^*	x_2^{**}	$x_1x_2^{***}$	Símbolo dos Experimentos	Média dos Rendimentos (g)
1	-1	-1	1	(1)	57
1	1	-1	-1	x_1	86
1	-1	1	-1	x_2	55
1	1	1	1	x_1x_2	62

* é a codificação da temperatura; ** é a codificação do catalisador; e *** é a codificação da interação da temperatura com o catalisador obtida pela multiplicação de x_1 e x_2 .

Operando com a Eq. 22 nas colunas 1 e 6 da Tabela 2.10, percebe-se que a média dos experimentos é 65 g:

$$Ef_{\text{médio}} = \frac{57 + 86 + 55 + 62}{4}$$

$$Ef_{\text{médio}} = 65 \text{ g}$$

$$\bar{y} = 65 \text{ g.}$$

Operando com a Eq. 23 nas colunas 2 e 6, 3 e 6, e 4 e 6, encontra-se os resultados a seguir:

$$Ef_{(x_1)} = \frac{(62+86)-(57+55)}{2} = 18 \Rightarrow X_1 = 18 \text{ g};$$

$$Ef_{(x_2)} = \frac{(62+55)-(86+57)}{2} = -13 \Rightarrow X_2 = -13 \text{ g};$$

$$Ef_{(x_1x_2)} = \frac{(62+57)-(86+55)}{2} = -11 \Rightarrow X_1X_2 = -11 \text{ g}.$$

O valor encontrado, $X_1 = 18 \text{ g}$, significa que o rendimento aumenta em média 18 g quando se passa do nível menor para o nível maior (60°C para 80°C). O valor encontrado, $X_2 = -13 \text{ g}$, significa que o rendimento diminui em média 13 g quando se passa do nível menor para o nível maior (A para B). Se não houvesse interação, o efeito da temperatura seria o mesmo para qualquer tipo de catalisador.

Através de uma leitura rápida e cálculos simples na Tabela 2.10 percebe-se que ao variar a temperatura e catalisador do experimento 1 para o experimento 4 há um aumento de 5 g (62 g -57 g); já a variação do experimento 2 para o experimento 3 há uma redução de 31 g (55 g – 86 g). O efeito de interação X_1X_2 encontrado representa que em média diminui 11 g o rendimento dos experimentos, quando varia ao mesmo tempo a passagem do nível menor para o nível maior, em ambas as variáveis. Para uma melhor visualização, usa-se a representação de um quadrado como se comentou anteriormente, e que se apresenta a seguir nas Figuras 2.3 e 2.4 (BOX *et al*, 1978; BRUNS *et al*, 2001).

Na Figura 2.3 percebe-se que:

- y_1 é a resposta experimental realizada no nível menor das variáveis x_1 e x_2 que se denota por (1);
- y_2 é a resposta experimental realizada no nível maior da variável x_1 e menor da variável x_2 que se denota por x_1 ;
- y_3 é a resposta experimental realizada no nível maior da variável x_2 e menor da variável x_1 que se denota por x_2 ;
- y_4 é a resposta experimental realizada no nível maior das variáveis x_1 e x_2 que se denota por x_1x_2 ;
- todos os símbolos experimentais estão coluna 5 da Tabela 2.10.

Substituindo os valores das respostas experimentais da Tabela 2.10, na Figura 2.3, obtém-se a Figura 2.4. Fazendo uma análise na Figura 2.4 percebe-se que no nível menor da temperatura a variação do nível maior para o menor do catalisador há uma redução de 2 g, e do nível maior da temperatura a variação do nível maior para o menor do catalisador há uma redução de 24 g. Já no nível menor do catalisador, quando se varia do nível menor da temperatura para o nível maior da temperatura há um aumento de 29 g, e no nível maior do catalisador, quando se varia do nível menor para o nível maior da temperatura há um aumento de 7 g. Percebe-se que a forma de representação do quadrado, Figuras 2.3 e 2.4, apresenta ao usuário a informação com mais clareza para investigações preliminares.

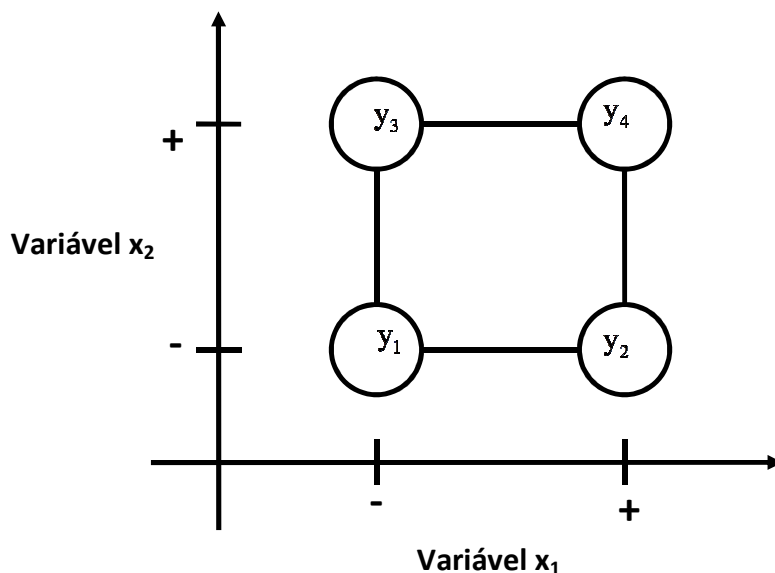


Figura 2.3 - Representação Geométrica de um Planejamento de Experimentos Fatorial 2^2

Como foi mencionado anteriormente, o fato de ter a disposição replicatas é possível obter o erro dos efeitos. Na Tabela 2.11 apresentam-se as respostas experimentais nas colunas 2 e 3, a média das respostas experimentais na coluna 4, na coluna 5 apresenta a variância amostral s_i^2 de cada nível experimental i usando a Eq. 24, na coluna 6 o grau de liberdade amostral ($gl = r_i - 1$) de cada nível experimental i e, na coluna 7 o produto do respectivo grau de liberdade pela sua respectiva variância amostral obtendo a variância conjunta

$$s_{\text{Conjunta}}^2 = \frac{56}{4} = 14 \text{ g}^2 \text{ usando a Eq. 25.}$$

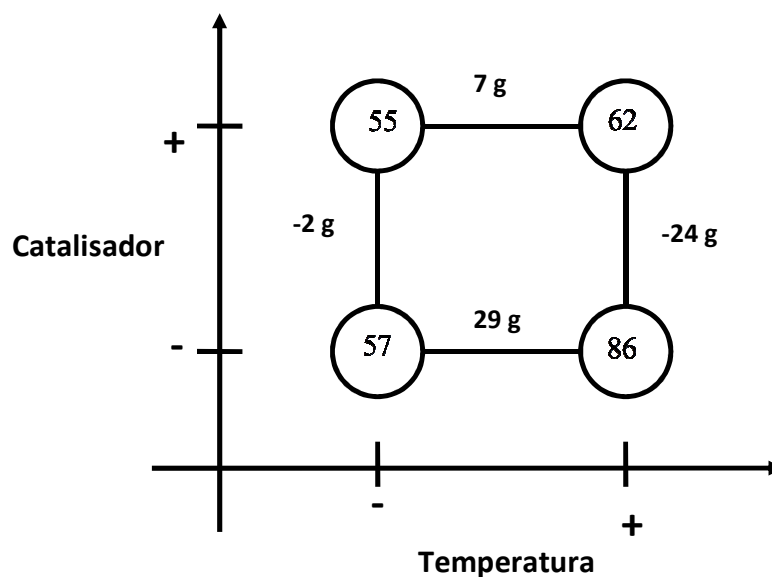


Figura 2.4 - Representação Geométrica do Planejamento de Experimentos Fatorial 2^2 da Tabela 2.10

Como os experimentos foram realizados em duplicatas ($r = 2$) no Exemplo 10 o grau de liberdade em cada nível experimental i será $r - 1 = 1$. Mais detalhes da Tabela 2.11 a seguir:

- As replicatas: $r_1 = r_2 = r_3 = r_4 = 2 \rightarrow$ duplicatas ;
- No experimento $i = 1$, tem-se média $\bar{y}_1 = \frac{55+59}{2} = 57$ e variância amostral

$$s_1^2 = \frac{(55-57)^2 + (59-57)^2}{2-1} = 8;$$
- No experimento $i = 2$, tem-se média $\bar{y}_2 = \frac{88+84}{2} = 86$ e variância amostral

$$s_2^2 = \frac{(88-86)^2 + (84-86)^2}{2-1} = 8;$$
- No experimento $i = 3$, tem-se média $\bar{y}_3 = \frac{51+59}{2} = 55$ e variância amostral

$$s_3^2 = \frac{(51-55)^2 + (59-55)^2}{2-1} = 32;$$
- No experimento $i = 4$, tem-se média $\bar{y}_4 = \frac{60+64}{2} = 62$ e variância amostral

$$s_4^2 = \frac{(60-62)^2 + (64-62)^2}{2-1} = 8;$$

$$\circ S_{\text{Conjunta}}^2 = \frac{(r_1 - 1)s_1^2 + (r_2 - 1)s_2^2 + (r_3 - 1)s_3^2 + (r_4 - 1)s_4^2}{(r_1 - 1) + (r_2 - 1) + (r_3 - 1) + (r_4 - 1)} = \frac{1 \cdot 8 + 1 \cdot 8 + 1 \cdot 32 + 1 \cdot 8}{1 + 1 + 1 + 1} = 14.$$

Tabela 2.11 – Estimativa da Variância Conjunta do Planejamento de Experimentos Fatorial 2^2 do Exemplo 10

Experimento i	Rendimento (g)	Média (g)	s_i^2	gl	$(r_i - 1) \cdot s_i^2$
1	55	59	8	1	8
2	88	84	8	1	8
3	51	59	32	1	32
4	60	64	8	1	8
				4	56

Na Tabela 2.12 estão listados a média dos experimentos, o efeito principal da temperatura e do catalisador, o efeito de interação, e os desvios padrões de cada efeito. O cálculo da variância e do desvio padrão de cada efeito apresenta-se a seguir obtidos pelas Eq. 26 e 27:

$$\circ S_{\text{Média}}^2 = \frac{S_{\text{Conjunta}}^2}{n \cdot r} = \frac{14}{4 \cdot 2} = 1,75 \Rightarrow s_{\text{Média}} = \sqrt{1,75} = \pm 1,32;$$

$$\circ S_{\text{Efeitos}}^2 = \frac{S_{\text{Conjunta}}^2}{\frac{n}{2}} = \frac{14}{\frac{4}{2}} = 7 \Rightarrow s_{\text{Efeitos}} = \sqrt{7} = \pm 2,646.$$

Percebe-se na Tabela 2.12 que os efeitos são significativos, com base nisto o modelo linear estimado significativo para o planejamento de experimentos fatorial completo do Exemplo 10 apresenta-se a seguir, onde os valores entre parênteses são os respectivos desvios padrão dos coeficientes do modelo, os quais são o dobro do desvio padrão dos efeitos obtidos pela Eq. 27, exceto a média dos experimentos:

$$\hat{y} = \underbrace{65}_{(\pm 1,322)} + \underbrace{\frac{18}{2}x_1}_{\frac{(\pm 2,646)}{2}} - \underbrace{\frac{13}{2}x_2}_{\frac{(\pm 2,646)}{2}} - \underbrace{\frac{11}{2}x_1x_2}_{\frac{(\pm 2,646)}{2}} \quad \text{ou} \quad \hat{y} = \underbrace{65}_{(\pm 1,322)} + \underbrace{9x_1}_{(\pm 1,323)} - \underbrace{6,5x_2}_{(\pm 1,323)} - \underbrace{5,5x_1x_2}_{(\pm 1,323)}.$$

Nota. No Exemplo 9, página 47, comentou-se que o efeito de cada variável e das interações entre elas é o dobro do coeficiente de cada variável numa regressão, exceto a média dos experimentos. Isso se explica pelo fato do efeito de uma variável ser o valor necessário para que o experimento passe do nível menor (-1) para o nível maior (+1), por exemplo, ou seja, duas unidades de medida, enquanto o coeficiente significa a codificação em valor unitário para que o experimento passe do nível menor (-1) para o nível zero (0), por exemplo. Com base nisso, o erro de cada efeito é o **dobro** do erro de cada coeficiente das variáveis numa regressão, como se apresentou na função estimada significativa do Exemplo 10.

Tabela 2.12 - Efeitos calculados para o Planejamento de Experimentos Fatorial 2^2 da Tabela 2.9

Média Global	65	$\pm 1,322$
Efeitos Principais		
X_1	18	$\pm 2,646$
X_2	-13	$\pm 2,646$
Efeitos de Interação de Segunda Ordem		
X_1X_2	-11	$\pm 2,646$

Consideram-se os coeficientes da regressão como significativos quando os valores são maiores que os respectivos erros (desvio padrão), portanto, todos os coeficientes do modelo linear estimado do Exemplo 10 são significativos, e quando isto ocorre os valores estimados pelo modelo linear completo são iguais aos verdadeiros valores experimentais. Esta conclusão realizou-se pelo uso de replicatas de todos os experimentos. Se os experimentos não forem realizados com replicatas, é possível analisar quais efeitos são significativos (geralmente os maiores em módulo) usando o gráfico de probabilidade normal, como se apresentará na seção desses gráficos na página 94.

2.4 Obtenção de efeitos de um Planejamento de Experimentos Fatorial Completo usando os conceitos da Álgebra Linear

Percebem-se no Exemplo 8, página 42, que os parâmetros de uma regressão de um planejamento de experimentos fatorial são obtidos através da regressão e também observa-se que a obtenção dos efeitos das variáveis em estudo é oriunda dos cálculos realizados pelas Eq. 24 e 25. Nesta seção do trabalho objetiva-se obter os efeitos das variáveis em estudo utilizando a notação de álgebra linear pelas Eq. 28.1 e 28.2 em que os símbolos estão de acordo com a **Nota** na página 33:

$$\mathbf{m} = \mathbf{X}^t \mathbf{y}. \quad (28.1)$$

Onde:

- \mathbf{X}^t é a matriz transposta da matriz de contraste \mathbf{X} ;
- \mathbf{m} é uma matriz coluna (vetor) resultado da multiplicação de matrizes;
- \mathbf{y} é a matriz coluna (vetor) das respostas experimentais realizadas.

Para encontrar a matriz dos efeitos (vetor) de um planejamento de experimentos fatorial 2^k (dois níveis e k fatores), divide-se a linha 1 da matriz coluna \mathbf{m} por 2^k e as demais linhas por 2^{k-1} , que se denota por:

$$\mathbf{ef} = \begin{bmatrix} \frac{m_{11}}{2^k} \\ \frac{m_{21}}{2^{k-1}} \\ \vdots \\ \frac{m_{\beta 1}}{2^{k-1}} \end{bmatrix}. \quad (28.2)$$

Onde β é a quantidade de parâmetros ou coeficientes da regressão.

Aplicando as Eq. 28.1 e 28.2 no Exemplo 10, Tabela 2.10, obtém-se:

$$\begin{array}{c} \bar{y} \quad T \quad C \quad TC \\ \mathbf{X} = \begin{bmatrix} +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 \end{bmatrix} \Rightarrow \mathbf{X}^t = \begin{bmatrix} +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \Rightarrow \mathbf{y} = \begin{bmatrix} 57 \\ 86 \\ 55 \\ 62 \end{bmatrix} \\ \mathbf{m} = \mathbf{X}^t \mathbf{y} \\ \mathbf{m} = \begin{bmatrix} +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 \end{bmatrix} \cdot \begin{bmatrix} 57 \\ 86 \\ 55 \\ 62 \end{bmatrix} \Rightarrow \mathbf{m} = \begin{bmatrix} 260 \\ 36 \\ -26 \\ -22 \end{bmatrix} \Rightarrow \mathbf{ef} = \begin{bmatrix} 260/4 \\ 36/2 \\ -26/2 \\ -22/2 \end{bmatrix} \Rightarrow \mathbf{ef} = \begin{bmatrix} 65 \\ 18 \\ -13 \\ -11 \end{bmatrix}. \end{array}$$

A linha 1 da matriz coluna (vetor) \mathbf{ef} representa a média dos experimentos, a linha 2 representa o efeito principal da variável x_1 (temperatura), a linha 3 o efeito principal da variável x_2 (catalisador), e a linha 4 o efeito de interação de segunda ordem das variáveis x_1 e x_2 (temperatura e catalisador), o que concorda com os resultados comentados anteriormente.

Exemplo 11. Um químico realizou um planejamento de experimentos fatorial completo 2^3 para investigar o efeito que a temperatura (60°C, 80°C), o catalisador (Tipo A, Tipo B) e a concentração (1,0 M; 2,0 M) e suas interações provocam no rendimento em gramas (g) na resposta experimental de um fenômeno em estudo, como se apresenta na Tabela 2.13.

Através de um planejamento de experimentos fatorial, $2^k = 2^3 = 8$, os experimentos realizaram-se de uma forma mais organizada varrendo todas as possíveis combinações das variáveis em estudo obtendo assim a menor quantidade possível de experimentos realizados (BRUNS, ET AL, 2001).

Tabela 2.13 - Planejamento de Experimentos Fatorial 2^3 do Exemplo 11

Experimentos	x_1^*	x_2^{**}	$x_1x_2^{***}$	Símbolos dos Experimentos	Rendimento (g)	
1	-1	-1	-1	(1)	42	44
2	1	-1	-1	x_1	66	80
3	-1	1	-1	x_2	64	68
4	1	1	-1	x_3	88	84
5	-1	-1	1	x_1x_2	40	44
6	1	-1	1	x_1x_3	66	70
7	-1	1	1	x_2x_3	56	60
8	1	1	1	$x_1x_2x_3$	60	56

* é a codificação da temperatura; ** é a codificação do catalisador; e *** é a codificação da interação da temperatura com o catalisador obtida pela multiplicação de x_1 e x_2 .

A matriz de contrastes X para o planejamento de experimentos fatorial completo com dois níveis e três variáveis, Exemplo 11, é uma matriz quadrada de ordem oito, como se observa a seguir:

$$\bar{y} \quad x_1 \quad x_2 \quad x_3 \quad x_1x_2 \quad x_1x_3 \quad x_2x_3 \quad x_1x_2x_3$$

$$\mathbf{X} = \begin{bmatrix} +1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 \\ +1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \end{bmatrix}$$

Onde:

- coluna 1: é a coluna para calcular a média dos experimentos;
- coluna 2: representa a codificação dos níveis da variável x_1 - temperatura;
- coluna 3: representa a codificação dos níveis da variável x_2 - catalisador;
- coluna 4: representa a codificação dos níveis da variável x_3 - concentração;
- coluna 5: representa a interação das variáveis x_1 e x_2 - temperatura e catalisador;
- coluna 6: representa a interação das variáveis x_1 e x_3 - temperatura e concentração;
- coluna 7: representa a interação das variáveis x_2 e x_3 - catalisador e concentração;
- coluna 8: representa a interação das variáveis x_1 e x_2 e x_3 - temperatura, catalisador e concentração;

Os efeitos do Exemplo 11, Tabela 2.13, obteve-se pelas Eq. 28.1 e 28.2 da mesma maneira que os efeitos da Tabela 2.10:

$$\mathbf{m} = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 \\ -1 & -1 & +1 & +1 & -1 & -1 & +1 & +1 \\ -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ -1 & +1 & +1 & -1 & +1 & -1 & -1 & +1 \end{bmatrix} \cdot \begin{bmatrix} 43 \\ 73 \\ 66 \\ 86 \\ 42 \\ 68 \\ 58 \\ 58 \end{bmatrix} \Rightarrow \mathbf{m} = \begin{bmatrix} 494 \\ 76 \\ 42 \\ -42 \\ -36 \\ -24 \\ -30 \\ -16 \end{bmatrix} \Rightarrow \mathbf{ef} = \begin{bmatrix} 494/8 \\ 76/4 \\ 42/4 \\ -42/4 \\ -36/4 \\ -24/4 \\ -30/4 \\ -16/4 \end{bmatrix} \Rightarrow \mathbf{ef} = \begin{bmatrix} 61,75 \\ 19 \\ 10,5 \\ -10,5 \\ -9 \\ -6 \\ -7,5 \\ -4 \end{bmatrix}.$$

A linha 1 do vetor \mathbf{ef} representa a média dos experimentos \bar{y} , a linha 2 representa o X_1 (o efeito da variável x_1 - temperatura), a linha 3 o X_2 (efeito da variável x_2 - catalisador), a linha 4 o X_3 (efeito da variável x_3 - concentração), a linha 5 o X_1X_2 (efeito de interação de segunda ordem da variável x_1 e x_2), a linha 6 o X_1X_3 (efeito de interação de segunda ordem da variável x_1 e x_3), a linha 7 o X_2X_3 (efeito de interação de segunda ordem da variável x_2 e x_3), e a linha 8 o $X_1X_2X_3$ (efeito de interação de terceira ordem das três variáveis).

Na Tabela 2.14 apresentam-se as respostas experimentais nas colunas 2 e 3, a média das respostas experimentais na coluna 4, na coluna 5 apresenta a variância amostral s_i^2 de cada nível experimental i usando a Eq. 24, na coluna 6 o grau de liberdade amostral ($gl = r_i - 1$) de cada nível experimental i e, na coluna 7 o produto do respectivo grau de liberdade pela sua respectiva variância amostral obtendo a variância conjunta $s_{\text{Conjunta}}^2 = \frac{148}{8} = 18,5 \text{ g}^2$ usando a Eq. 25.

Como os experimentos foram realizados em duplicatas ($r = 2$) no Exemplo 11 o grau de liberdade em cada nível experimental i será $r - 1 = 1$. Mais detalhes da Tabela 2.14 a seguir:

- As replicatas: $r_1 = r_2 = r_3 = \dots = r_8 = 2 \rightarrow$ duplicatas ;
- No experimento $i = 1$, tem-se média $\bar{y}_1 = \frac{42 + 44}{2} = 43$ e variância amostral

$$s_1^2 = \frac{(42 - 43)^2 + (44 - 43)^2}{2 - 1} = 2;$$

- No experimento $i = 2$, tem-se média $\bar{y}_2 = \frac{66+80}{2} = 73$ e variância amostral

$$s_2^2 = \frac{(66-73)^2 + (80-73)^2}{2-1} = 98;$$

- Nos demais experimentos o processo é análogo;
- $s_{\text{Conjunta}}^2 = \frac{(r_1-1)s_1^2 + (r_2-1)s_2^2 + \dots + (r_8-1)s_8^2}{(r_1-1) + (r_2-1) + \dots + (r_8-1)} = \frac{1 \cdot 2 + 1 \cdot 98 + 1 \cdot 8 + \dots + 1 \cdot 8}{1+1+1+\dots+1} = 18,5.$

A discriminação dos efeitos principais, dos efeitos de segunda e terceira ordem, dos erros e da média dos experimentos apresenta-se na Tabela 2.15, onde o cálculo da variância e do desvio padrão de cada coeficiente apresenta-se a seguir obtidos pelas Eq. 26 e 27:

- $s_{\text{Média}}^2 = \frac{s_{\text{Conjunta}}^2}{n \cdot r} = \frac{18,5}{8 \cdot 2} = 1,1563 \Rightarrow s_{\text{Média}} = \sqrt{1,1563} = \pm 1,0753;$

- $s_{\text{Efeitos}}^2 = \frac{s_{\text{Conjunta}}^2}{\frac{n}{2}} = \frac{18,5}{\frac{8}{2}} = 4,6250 \Rightarrow s_{\text{Efeitos}} = \sqrt{4,6250} = \pm 2,1506.$

Tabela 2.14 – Estimativa da Variância Conjunta do Planejamento de Experimentos Fatorial 2^3 do Exemplo 11

Experimento i	Rendimento (g)		Média (g)	s_i^2	gl	$(r_i - 1) \cdot s_i^2$
1	42	44	43	2	1	2
2	66	80	73	98	1	98
3	64	68	66	8	1	8
4	88	84	86	8	1	8
5	40	44	42	8	1	8
6	66	70	68	8	1	8
7	56	60	58	8	1	8
8	60	56	58	8	1	8
					8	148

A partir da Tabela 2.15 conclui-se que todos os efeitos são significativos, pois são maiores que os respectivos desvios padrão, obtidos pelas Eq. 26 e 27. Com base nisto e no comentário do Exemplo 10, o modelo estimado significativo do Exemplo 11 apresenta-se a seguir, onde os valores entre parênteses são os respectivos desvio padrão dos coeficientes da regressão, sendo a metade dos erros dos respectivos efeitos da Tabela 2.15:

$$\hat{y} = \underbrace{61,75}_{(\pm 1,0753)} + \frac{19}{\left(\frac{\pm 2,1506}{2}\right)} \cdot x_1 + \frac{10,5}{\left(\frac{\pm 2,1506}{2}\right)} \cdot x_2 - \frac{10,5}{\left(\frac{\pm 2,1506}{2}\right)} \cdot x_3 - \frac{9}{\left(\frac{\pm 2,1506}{2}\right)} \cdot x_1 x_2 - \frac{6}{\left(\frac{\pm 2,1506}{2}\right)} \cdot x_1 x_3 - \frac{7,5}{\left(\frac{\pm 2,1506}{2}\right)} \cdot x_2 x_3 - \frac{4}{\left(\frac{\pm 2,1506}{2}\right)} \cdot x_1 x_2 x_3$$

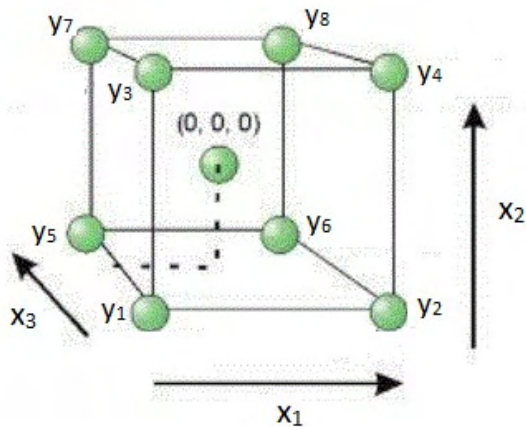
$$\hat{y} = \underbrace{61,75}_{(\pm 1,0753)} + \underbrace{9,5}_{(\pm 1,0753)} \cdot x_1 + \underbrace{5,25}_{(\pm 1,0753)} \cdot x_2 - \underbrace{5,25}_{(\pm 1,0753)} \cdot x_3 - \underbrace{4,5}_{(\pm 1,0753)} \cdot x_1 x_2 - \underbrace{3}_{(\pm 1,0753)} \cdot x_1 x_3 - \underbrace{3,75}_{(\pm 1,0753)} \cdot x_2 x_3 - \underbrace{2}_{(\pm 1,0753)} \cdot x_1 x_2 x_3$$

Tabela 2.15. Efeitos do Planejamento de Experimentos Fatorial 2^3 do Exemplo 11 - Tabela 2.13

Média Global	61,75	$\pm 1,0753$
Efeitos Principais		
X_1	19	$\pm 2,1506$
X_2	10,5	$\pm 2,1506$
X_3	-10,5	$\pm 2,1506$
Efeitos de Interação de Segunda Ordem		
$X_1 X_2$	-9	$\pm 2,1506$
$X_1 X_3$	-6	$\pm 2,1506$
$X_2 X_3$	-7,5	$\pm 2,1506$
Efeitos de Interação de Terceira Ordem		
$X_1 X_2 X_3$	-4	$\pm 2,1506$

A Figura 2.5 consiste na representação geométrica, gráfico do cubo, de um planejamento de experimentos fatorial 2^3 , como se comentou anteriormente (BOX *et al*, 1978; FERREIRA, M. e TEÓFILO, 2006; BRUNS *et al*, 2001).

Figura 2.5 – Representação Geométrica de um Planejamento de Experimentos Fatorial 2^3



Na Figura 2.5 observa-se que:

- y_1 é a resposta experimental realizada no nível menor das variáveis que se denota por (1);

- y_2 é a resposta experimental realizada no nível maior da variável x_1 e menor das demais, que se denota por x_1 ;
- y_3 é a resposta experimental realizada no nível maior da variável y_3 e menor das demais, que se denota por x_2 ;
- y_4 é a resposta experimental realizada no nível maior das variáveis x_1 e x_2 e no nível menor das demais, que se denota por x_1x_2 ;
- y_5 é a resposta experimental realizada no nível maior da variável x_3 , e no nível menor das demais, que se denota por x_3 ;
- y_6 é a resposta experimental realizada no nível maior das variáveis x_1 e x_3 , e no nível menor das demais, que se denota por x_1x_3 ;
- y_7 é a resposta experimental realizada no nível maior das variáveis x_2 e x_3 , e no nível menor das demais, que se denota por x_2x_3 ;
- y_8 é a resposta experimental realizada no nível maior das variáveis x_1 , x_2 e x_3 , e no nível menor das demais, que se denota por $x_1x_2x_3$;
- todos os símbolos experimentais estão coluna 5 da Tabela 2.13.

Obteve-se a Figura 2.6 através dos dados da Tabela 2.16, onde os valores estimados foram calculados pela função estimada significativa:

$$\hat{y} = 61,75 + 9,5x_1 + 5,25x_2 - 5,25x_3 - 4,5x_1x_2 - 3x_1x_3 - 3,75x_2x_3 - 2x_1x_2x_3.$$

Como todos os coeficientes foram significativos no modelo linear completo, os valores estimados e os valores experimentais apresentam-se idênticos na Tabela 2.16, sendo nulos os resíduos na coluna 9 da tabela 2.16.

Tabela 2.16 – Valores Estimados e Resíduos Aleatórios do Exemplo 11 – Tabela 2.13

Experimentos	Rendimento (g)	Média (g)	x_1	x_2	x_3	\hat{y}	\hat{u}	
1	42	44	43	-1	-1	-1	43	0
2	66	80	73	1	-1	-1	73	0
3	64	68	66	-1	1	-1	66	0
4	88	84	86	1	1	-1	86	0
5	40	44	42	-1	-1	1	42	0
6	66	70	68	1	-1	1	68	0
7	56	60	58	-1	1	1	58	0
8	60	56	58	1	1	1	58	0

No presente trabalho, programou-se todos os cálculos do planejamento de experimentos fatorial completo baseados nas teorias apresentadas, sendo os planejamentos constituídos de dois níveis e com até 6 variáveis, 2^2 a 2^6 , com ou sem replicatas de todos os experimentos, limitando a quantidade de replicatas em até quatro (quadruplicatas), com a possibilidade de gerar gráficos de resíduos em função dos valores estimados e gráficos de probabilidade normal, como se apresentará na seção desses gráficos.

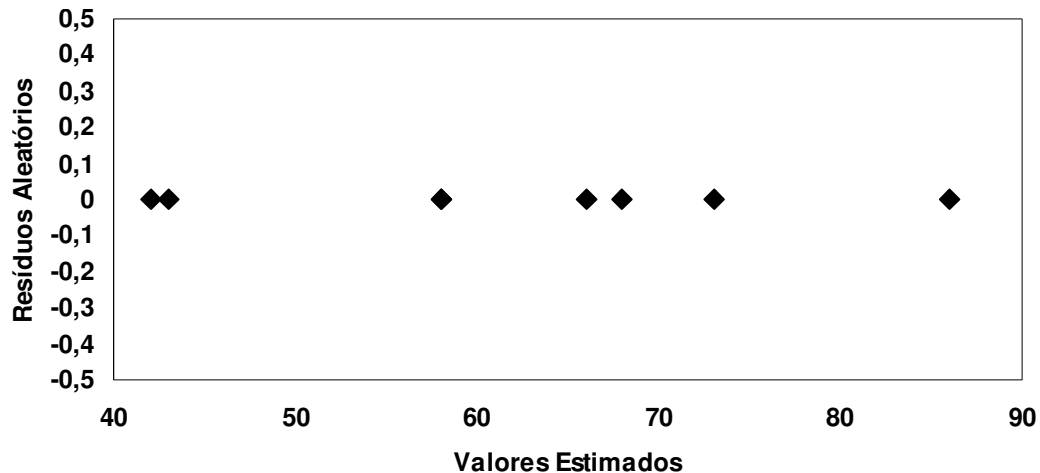


Figura 2.6 - Gráfico de resíduos aleatório versus valores estimados da Tabela 2.16.

Na Tabela 2.17, apresentam-se os símbolos dos experimentos para um planejamento de experimentos fatorial completo do tipo 2^4 , onde há no mínimo 16 experimentos realizados de forma organizada para varrer todas as possíveis combinações das variáveis em estudo com um mínimo de tempo (FERREIRA, M. e TEÓFILO, 2006; BRUNS *et al*, 2001; OLIVIER *et al*, 2007). Nas colunas 2 a 5 da Tabela 2.17 representam a codificação das variáveis x_1 , x_2 , x_3 e x_4 obtidas através da Eq. 20. A coluna 6 possui os símbolos dos experimentos correlacionadas com a simbologia da coluna 7. Mais detalhes:

- y_1 é a resposta experimental realizada no nível menor das variáveis que se denota por (1);
- y_2 é a resposta experimental realizada no nível maior da variável x_1 e menor das demais, que se denota por x_1 ;
- y_2 é a resposta experimental realizada no nível maior da variável x_2 e menor das demais, que se denota por x_2 ;

- y_4 é a resposta experimental realizada no nível maior das variáveis x_1 e x_2 , e no nível menor das demais, que se denota por x_1x_2 ;
- y_5 é a resposta experimental realizada no nível maior da variável x_3 , e no nível menor das demais, que se denota por x_3 ;
- y_6 é a resposta experimental realizada no nível maior das variáveis x_1 e x_3 , e no nível menor das demais, que se denota por x_1x_3 ;
- y_7 é a resposta experimental realizada no nível maior das variáveis x_2 e x_3 , e no nível menor das demais, que se denota por x_2x_3 ;
- y_8 é a resposta experimental realizada no nível maior das variáveis x_1 , x_2 e x_3 , e no nível menor das demais, que se denota por $x_1x_2x_3$;
- y_9 é a resposta experimental realizada no nível maior da variável x_4 e menor das demais, que se denota por x_4 ;
- y_{10} é a resposta experimental realizada no nível maior das variáveis x_1 e x_4 , e no nível menor das demais, que se denota por x_1x_4 , e assim por diante;
- todos os símbolos experimentais estão coluna 6 da Tabela 2.17.

Tabela 2.17 – Planejamento de Experimentos Fatorial 2^4

Experimentos	x_1	x_2	x_3	x_4	Símbolos dos Experimentos	Respostas Experimentais
1	-1	-1	-1	-1	(1)	y_1
2	1	-1	-1	-1	x_1	y_2
3	-1	1	-1	-1	x_2	y_3
4	1	1	-1	-1	x_1x_2	y_4
5	-1	-1	1	-1	x_3	y_5
6	1	-1	1	-1	x_1x_2	y_6
7	-1	1	1	-1	x_2x_3	y_7
8	1	1	1	-1	$x_1x_2x_3$	y_8
9	-1	-1	-1	1	x_4	y_9
10	1	-1	-1	1	x_1x_4	y_{10}
11	-1	1	-1	1	x_2x_4	y_{11}
12	1	1	-1	1	$x_1x_2x_4$	y_{12}
13	-1	-1	1	1	x_3x_4	y_{13}
14	1	-1	1	1	$x_1x_3x_4$	y_{14}
15	-1	1	1	1	$x_2x_3x_4$	y_{15}
16	1	1	1	1	$x_1x_2x_3x_4$	y_{16}

Para os outros tipos de planejamentos de experimentos fatorial completo, 2^5 e 2^6 , os símbolos dos experimentos são realizados obtendo os valores codificados dos níveis das variáveis em estudo, como se apresentou nas Tabelas 2.8, 2.10, 2.13 e 2.17.

Qualquer problema a ser resolvido por um planejamento fatorial completo necessita no mínimo 2^k experimentos para sua execução, sendo k a quantidade de variáveis. Com quatro variáveis a quantidade mínima é 16, e aumentando mais uma variável, a quantidade de experimentos dobra. Portanto, a principal desvantagem é o grande número de experimentos que são realizados a cada variável adicionada ao estudo (BRUNS *et al*, 2001).

Quando a quantidade de variáveis for relativamente grande, por exemplo, maior que quatro, são possíveis que algumas interações não tenham influência significativa sobre a resposta, neste caso, um planejamento de experimentos fatorial completo seria inviável. O recomendado seria a realização de uma triagem com uma quantidade menor de experimentos utilizando um planejamento de experimentos fatorial fracionário (BOX *et al*, 1978, BRUNS *et al*, ET AL, 2001).

Se efeitos de altas ordens não são significativos à realização de experimentos para estimar tais efeitos de interação pode ser irrelevante, dessa maneira com um número menor de experimentos através do planejamento de experimentos fatorial fracionário é possível obter informações daqueles efeitos mais importantes e retirar, na maioria das vezes, as mesmas conclusões caso fosse realizado um planejamento de experimentos fatorial completo (BOX *et al*, 1978; FERREIRA, M. e TEÓFILO, 2006; BRUNS *et al*, 2001).

2.5 Planejamento de Experimentos Fatorial Fracionário

A quantidade de experimentos de um planejamento de experimentos fatorial completo de dois níveis e com k variáveis se dada por 2^k , enquanto um planejamento de experimentos fatorial fracionário de meia fração ($1/2$) obtém-se pela Eq. 29,

$$n_{\text{Fracionário}} = \frac{1}{2} \cdot 2^k = 2^{k-1}. \quad (29)$$

Onde $n_{\text{Fracionário}}$ é a quantidade mínima de experimentos realizados para um planejamento de experimentos fatorial de meia fração.

Na literatura há diferentes tipos de planejamentos de experimentos fatorial fracionários, por exemplo, as frações $1/2$, $1/4$, $1/8$, $1/16 \dots 1/2^a$ de um planejamento 2^{k-a} , em que k é o número de variáveis e a é o tamanho da fração (BOX *et al*, 1978; FERREIRA, M. e TEÓFILO, 2006; BRUNS *et al*, 2001). No presente trabalho consideraram-se apenas os

planejamentos de experimentos fatoriais fracionários de meia fração, os quais reduzem pela metade a quantidade de experimentos.

2.5.1 Construção de um Planejamento de Experimentos Fatorial de Meia Fração

Obtém-se um planejamento de experimentos fatorial de meia fração através da codificação de um planejamento de experimentos fatorial completo. A Tabela 2.18 apresenta-se um planejamento de experimentos fatorial completo 2^3 , enquanto a Tabela 2.19 trata-se de um planejamento de experimentos fatorial de meia fração, sendo um subconjunto da Tabela 2.18.

Para obter a Tabela 2.19, constrói-se a Tabela 2.18, ou seja, um planejamento de experimentos fatorial completo para as variáveis x_1 , x_2 e x_3 . Após tal tarefa, observa-se nos experimentos 5, 2, 3 e 8, respectivamente, que o produto das variáveis x_1 e x_2 é igual a variável x_3 , ou seja, considera-se a identidade $x_3 = x_1x_2$. A primeira consequência é que os efeitos X_3 e X_1X_2 se tornam idênticos, já que são determinados pelos mesmos sinais.

Tabela 2.18. Planejamento de Experimentos Fatorial 2^3

Experimentos	Símbolos dos Experimentos	Variáveis			Interações			
		x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
1	(1)	-1	-1	-1	1	1	1	-1
2	x_1	1	-1	-1	-1	-1	1	1
3	x_2	-1	1	-1	-1	1	-1	1
4	x_3	1	1	-1	1	-1	-1	-1
5	x_1x_2	-1	-1	1	1	-1	-1	1
6	x_1x_3	1	-1	1	-1	1	-1	-1
7	x_2x_3	-1	1	1	-1	-1	1	-1
8	$x_1x_2x_3$	1	1	1	1	1	1	1

Na Tabela 2.18 percebe-se que:

- x_1 é a codificação da variável 1;
- x_2 é a codificação da variável 2;
- x_3 é a codificação da variável 3,
- x_1x_2 é a codificação das variáveis 1 e 2 obtidas pela multiplicação de x_1 e x_2 ,
- e assim por diante.

Tabela 2.19. Planejamento de Experimentos Fatorial Fracionário 2^{3-1} obtido de um Planejamento de Experimentos Fatorial Completo 2^3 – Tabela 2.18

Experimentos	Símbolos dos Experimentos	Variáveis		
		x_1	x_2	$x_3 = x_1x_2$
5	x_1x_2	-1	-1	1
2	x_1	1	-1	-1
3	x_2	-1	1	-1
8	$x_1x_2x_3$	1	1	1

Nota. Os efeitos de um planejamento de experimentos fatorial completo denotam-se neste trabalho pela letra X maiúscula, enquanto que a média dos experimentos pelo símbolo \bar{y} . No planejamento de experimentos fatorial de meia fração haverá um confundimento de efeitos oriundos do planejamento de experimentos fatorial completo. Para distinguir a simbologia dos efeitos do planejamento completo e do fracionário, acrescenta-se neste trabalho um índice inferior com a letra F, que se denota por X_F , para o planejamento fracionário.

Para a obtenção dos diversos contrastes do planejamento de experimentos fatorial fracionário empregou-se duas propriedades da multiplicação das colunas dos sinais. A primeira, Eq. 30, é que a multiplicação de uma coluna por ela mesma, isto é, o quadrado de cada elemento produz uma coluna contendo apenas sinais positivos. Essa coluna, por sua vez, quando multiplicada por outra a deixa inalterada, ou seja, a coluna contendo apenas os sinais positivos será o elemento identidade da álgebra do planejamento de experimentos fatorial fracionário, e por isso usa-se a letra I para representa-lo, como é de costume em álgebra (BOX *et al*, 1978; BRUNS *et al*, 2001). Assim, por exemplo, escreve-se,

$$x_1x_1 = x_2x_2 = x_3x_3 = x_4x_4 = \dots = I. \quad (30)$$

A segunda propriedade, Eq. 31, apenas reconhece que a multiplicação das colunas é comutativa e associativa (BOX *et al*, 1978; BRUNS *et al*, 2001). Por exemplo:

$$x_1x_2x_3 = (x_1)(x_2x_3) = (x_2x_3)(x_1) = (x_1x_2)(x_3) = (x_3x_2x_1) = \dots \quad (31)$$

Com as duas propriedades anunciadas, Eq. 30 e 31, obtiveram-se as relações entre os vários contrastes nas Tabelas 2.20 e 2.22.

A meia fração obtida na Tabela 2.19 foi escrita partir da equação $x_3 = x_1x_2$, que frequentemente apresenta-se na forma equivalente $I = x_1x_2x_3$, obtida multiplicando os dois membros da equação por x_3 . Nessa forma, onde o elemento identidade aparece isolado, a expressão é conhecida como **geratriz** ou **relação geradora** do planejamento de experimentos fatorial fracionário.

Quando o objetivo é saber o equivalente do contraste, manipulam-se as variáveis para aparecer sozinho num dos lados da equação. Por exemplo, se a intenção é saber qual o contraste equivalente a x_2 na Tabela 2.19, multiplica-se os dois membros da equação utilizada como geratriz do planejamento de experimentos fatorial fracionário, $x_3 = x_1x_2$, por x_1 :

$$(x_1)(x_3) = (x_1)(x_1x_2) = (x_1x_1)(x_2) = I(x_2) = x_2.$$

Do lado esquerdo obteve-se o produto x_1x_3 , e daí conclui-se que o efeito da variável 2 é igual ao efeito de interação das variáveis 1 e 3, que se denota por $X_2 = X_1X_3$. Na terminologia estatística, diz que a utilização da meia fração confunde o efeito principal X_2 com a interação de 2ª ordem X_1X_3 , e que o valor do efeito calculado pelo contraste x_2 ou x_1x_3 , é na verdade uma estimativa da soma dos dois efeitos do planejamento de experimentos fatorial completo e que se denota por,

$$X_{F_1} = X_2 + X_1X_3. \quad (32)$$

Na Eq. 32 o efeito da variável 1 do planejamento de experimentos fatorial fracionário X_{F_1} da *primeira meia fração* é a soma do efeito da variável 2, X_2 , com o efeito de interação das variáveis 1 e 3, X_1X_3 , do planejamento de experimentos fatorial completo, ou seja, há um confundimento dos efeitos, porém isso não deve ser obstáculo para impedir o uso do planejamento de experimentos fatorial fracionário, como se comentou anteriormente.

Os experimentos restantes da Tabela 2.18, aqueles que não estão **sublinhados**, não foram selecionados nos cálculos do planejamento de experimentos fatorial fracionário, e constituem a outra metade, isto é, a *outra meia fração* do planejamento de experimentos fatorial completo. Percebe-se facilmente que os sinais da variável x_3 nesses experimentos são o contrário dos sinais do produto das variáveis x_1 e x_2 , logo essa outra fração é gerada pela relação $x_3 = -x_1x_2$, ou equivalente, que sua **geratriz** é $\mathbf{I} = -x_1x_2x_3$. Com um argumento semelhante, mostra-se que os contrastes agora são estimativas da diferença entre dois efeitos do planejamento de experimento fatorial completo (BOX *et al*, 1978; BRUNS *et al*, 2001),

$$X_{F_1}^* \Rightarrow X_2 - X_1X_3. \quad (33)$$

As relações entre os novos contrastes, identificados pelo asterisco, e os efeitos do planejamento são as mesmas da coluna 2 da Tabela 2.20, só com os sinais de negativos (BOX *et al*, 1978; BRUNS *et al*, 2001). Também se pode usar a *segunda meia fração*, que é chamada de *fração complementar da primeira*, para estimar os efeitos do fatorial fracionário. Os resultados finais levariam as mesmas conclusões (BOX *et al*, 1978; BRUNS *et al*, 2001).

Na Eq. 33 o efeito da variável 1 do planejamento de experimentos fatorial fracionário X_{F_1} da *segunda meia fração* é a diferença do efeito da variável 2, X_2 , com o efeito de

interação das variáveis 1 e 3, X_1X_3 , do planejamento de experimentos fatorial completo, ou seja, há um confundimento, como se comentou anteriormente.

Na Figura 2.7 percebe-se que os experimentos selecionados para um planejamento de experimentos fatorial de meia fração 2^{3-1} usando a geratriz $I = x_1x_2x_3$ a partir de um planejamento de experimentos fatorial 2^3 distribuem-se no domínio experimental para as três variáveis, com base nisso obtém-se as mesmas conclusões daquelas percebidas a partir de um planejamento de experimentos fatorial completo.

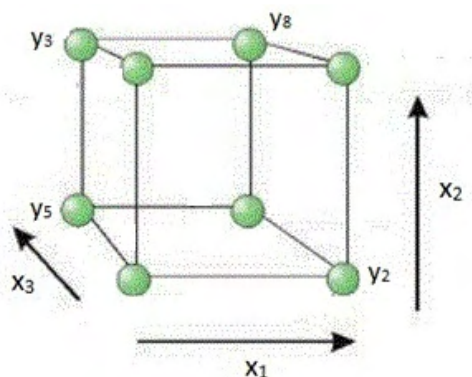


Figura 2.7 – Experimentos Selecionados de um Planejamento Fatorial 2^{3-1} usando como gerador $I = x_1x_2x_3$

A Tabela 2.20 apresenta todas as relações entre os contrastes calculados na meia fração e os efeitos obtidos com o planejamento completo 2^3 da Tabela 2.18, os chamados padrões de confundimento, e a Tabela 2.22 apresentam os padrões de confundimento para um fatorial completo 2^4 e para sua meia fração.

Tabela 2.20 - Relações entre os contrastes da meia fração 2^{3-1} e os efeitos do fatorial completo 2^3

Relações entre as colunas de sinais	Contrastes da meia fração 2^{3-1} em termos de efeitos do fatorial 2^3
$x_1 = x_2x_3$	$X_{F_1} = X_{F_2}X_{F_3} \rightarrow X_1 + X_2X_3$
$x_2 = x_1x_3$	$X_{F_2} = X_{F_1}X_{F_3} \rightarrow X_2 + X_1X_3$
$x_3 = x_1x_2$	$X_{F_3} = X_{F_1}X_{F_2} \rightarrow X_3 + X_1X_2$
$I = x_1x_2x_3$	$\bar{y}_F \rightarrow \bar{y} + 1/2(X_1X_2X_3)$

A Tabela 2.21 apresenta o planejamento de experimentos fatorial fracionário 2^{4-1} obtido a partir de um planejamento de experimentos fatorial completo 2^4 usando como relação geradora $I = x_1x_2x_3x_4$.

Tabela 2.21 - Planejamento Fatorial Fracionário 2^{4-1} obtido de um Planejamento Fatorial Completo 2^4

Experimentos	Símbolos dos Experimentos	x_1	x_2	x_3	x_4	x_1x_2	x_1x_3	x_1x_4
1	(1)	-1	-1	-1	-1	1	1	1
10	x_1x_4	1	-1	-1	1	-1	-1	1
11	x_2x_4	-1	1	-1	1	-1	1	-1
4	x_1x_2	1	1	-1	-1	1	-1	-1
13	x_3x_4	-1	-1	1	1	1	-1	-1
6	x_1x_2	1	-1	1	-1	-1	1	-1
7	x_2x_3	-1	1	1	-1	-1	-1	1
16	$x_1x_2x_3x_4$	1	1	1	1	1	1	1

Percebe-se na Tabela 2.21 que a coluna dos experimentos tem numeração aleatória da mesma forma que a Tabela 2.19. Com a construção de um planejamento de experimentos fatorial completo 2^4 e usando a relação geradora ou geratriz $I = x_1x_2x_3x_4$, percebe-se que o contraste x_4 é igual ao produto dos contrastes das variáveis **1**, **2** e **3**, ou seja, $x_4 = x_1x_2x_3$, com base nisto, os experimentos que possuem o sinal do contraste x_4 igual ao produto das variáveis **1**, **2** e **3** são os experimentos que estão na coluna 1 da Tabela 2.21. Neste caso o efeito X_4 é confundido com o efeito de interação das variáveis 1, 2 e 3, $X_4 = X_1X_2X_3$.

Tabela 2.22. Relações entre os contrastes da meia fração 2^{4-1} e os efeitos do fatorial completo 2^4

Relações entre as colunas de sinais	Contrastes da meia fração 2^{4-1} em termos de efeitos do fatorial 2^4
$x_1 = x_2x_3x_4$	$X_{F_1} = X_{F_2}X_{F_3}X_{F_4} \rightarrow X_1 + X_2X_3X_4$
$x_2 = x_1x_3x_4$	$X_{F_2} = X_{F_1}X_{F_3}X_{F_4} \rightarrow X_2 + X_1X_3X_4$
$x_3 = x_1x_2x_4$	$X_{F_3} = X_{F_1}X_{F_2}X_{F_4} \rightarrow X_3 + X_1X_2X_4$
$x_4 = x_1x_2x_3$	$X_{F_4} = X_{F_1}X_{F_2}X_{F_3} \rightarrow X_4 + X_1X_2X_3$
$x_1x_2 = x_3x_4$	$X_{F_1}X_{F_2} = X_{F_3}X_{F_4} \rightarrow X_1X_2 + X_3X_4$
$x_1x_3 = x_2x_4$	$X_{F_1}X_{F_3} = X_{F_2}X_{F_4} \rightarrow X_1X_3 + X_2X_4$
$x_1x_4 = x_2x_3$	$X_{F_1}X_{F_4} = X_{F_2}X_{F_3} \rightarrow X_1X_4 + X_2X_3$
$I = x_1x_2x_3x_4$	$\bar{y}_F \rightarrow \bar{y} + 1/2(X_1X_2X_3X_4)$

A fração do planejamento 2^{4-1} não mistura os efeitos principais com interações de duas variáveis, mas mistura com interações de três variáveis, que em princípio devem ser menos significativos. Se essas interações forem realmente desprezíveis, os contrastes fornecerão ótimas aproximações dos efeitos principais calculados a partir do fatorial completo (BOX *et al*, 1978; FERREIRA, M. e TEÓFILO, 2006; BRUNS *et al*, 2001). Na prática, espera-se que o efeito do planejamento fracionário da *primeira meia fração* seja igual ao efeito do planejamento fracionário da *segunda meia fração* e igual ao efeito do planejamento completo, que se denota por:

$$X_{F_i} \cong X_{F_i}^* \cong X_i.$$

Ao usar um planejamento de experimentos fatorial fracionário é necessário saber qual é a **resolução** do planejamento para alcançar os objetivos propostos. O planejamento 2^{4-1} é um exemplo de planejamento de experimentos fatorial fracionário de resolução quatro, que não confunde efeitos principais entre si e nem com efeitos de interação entre duas variáveis, mas confunde efeitos principais com efeitos de interação entre três variáveis e os efeitos entre duas variáveis se confundem com outros efeitos, inclusive entre eles. A notação empregada para representar a resolução de um planejamento é um índice em algarismos romanos. Para este fracionário, escreve-se 2_{IV}^{4-1} . Os de resolução cinco os efeitos principais são confundidos com efeitos de interação entre quatro variáveis e os efeitos de interação entre duas variáveis são confundidos com efeitos de interação entre três variáveis (BOX *et al*, 1978; FERREIRA, M. e TEÓFILO, 2006; BRUNS *et al*, 2001).

A resolução de um planejamento de experimentos fatorial fracionário é determinada pelas relações geradoras ou geratrizes. O número de variáveis que compõem o termo mais curto presente nessas relações é, por definição, a resolução do planejamento. Para definir uma meia fração, basta uma relação geradora. Na relação $I = \pm x_1 x_2 x_3 x_4$ contém quatro variáveis, e por isso a resolução do fatorial 2^{4-1} é quatro, 2_{IV}^{4-1} . Na prática, as resoluções mais usadas ficam dentre três e cinco.

Tabela 2.23 – Planejamento de Experimentos Fatorial Completo 2^4 do Exemplo 12

Experimentos	X_1	X_2	X_3	X_4	Símbolos dos Experimentos	Respostas Experimentais	Efeitos Obtidos
<u>1</u>	-1	-1	-1	-1	(1)	102	$\bar{y} = 120,062$
2	1	-1	-1	-1	x_1	110	$X_1 = 0,375$
3	-1	1	-1	-1	x_2	90	$X_2 = -0,875$
4	1	1	-1	-1	$x_1 x_2$	92	$X_3 = 8,625$
5	-1	-1	1	-1	x_3	115	$X_4 = 15,375$
6	1	-1	1	-1	$x_1 x_2$	120	$X_1 X_2 = -13,625$
7	-1	1	1	-1	$x_2 x_3$	140	$X_1 X_3 = -4,125$
8	1	1	1	-1	$x_1 x_2 x_3$	130	$X_1 X_4 = -0,875$
9	-1	-1	-1	1	x_4	112	$X_2 X_3 = -0,375$
10	1	-1	-1	1	$x_1 x_4$	140	$X_2 X_4 = -2,125$
11	-1	1	-1	1	$x_2 x_4$	150	$X_3 X_4 = -19,125$
12	1	1	-1	1	$x_1 x_2 x_4$	130	$X_1 X_2 X_3 = -0,125$
13	-1	-1	1	1	$x_3 x_4$	125	$X_1 X_2 X_4 = -8,375$
14	1	-1	1	1	$x_1 x_3 x_4$	140	$X_1 X_3 X_4 = -0,375$
15	-1	1	1	1	$x_2 x_3 x_4$	125	$X_2 X_3 X_4 = -16,625$
16	1	1	1	1	$x_1 x_2 x_3 x_4$	100	$X_1 X_2 X_3 X_4 = 2,125$

Exemplo 12. Na Tabela 2.23 apresenta-se um planejamento de experimentos fatorial completo 2^4 de um problema fictício elaborado para fins didáticos, em que as variáveis independentes foram codificadas usando a Eq. 20 e a resposta experimental representa o rendimento de um fenômeno qualquer. Os efeitos obtidos do planejamento de experimentos fatorial completo estão na coluna 8 da Tabela 2.23.

Utilizou-se o modelo linear para as quatro variáveis do Exemplo 12 do tipo completo, ou seja, com os coeficientes das interações de segunda, terceira e quarta ordem, além dos coeficientes da média experimental e dos efeitos principais:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_1 x_2 + \hat{\beta}_6 x_1 x_3 + \hat{\beta}_7 x_1 x_4 + \hat{\beta}_8 x_2 x_3 + \hat{\beta}_9 x_2 x_4 + \hat{\beta}_{10} x_3 x_4 + \hat{\beta}_{11} x_1 x_2 x_3 + \hat{\beta}_{12} x_1 x_2 x_4 + \hat{\beta}_{13} x_1 x_3 x_4 + \hat{\beta}_{14} x_2 x_3 x_4 + \hat{\beta}_{15} x_1 x_2 x_3 x_4.$$

Com base nisto, a ordem da dimensão da matriz de contraste do planejamento 2^4 é 16, ou seja, 16 linhas e 16 colunas, e a dimensão da matriz de resposta é 16 linhas e 1 coluna. Ambas as matrizes apresentam-se a seguir:

$$\mathbf{X} = \begin{matrix} \bar{y} & x_1 & x_2 & x_3 & x_4 & A & B & C & D & E & F & G & H & I & J & K \\ \left[\begin{array}{cccccccccccccccc} 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \right] \end{matrix} \mathbf{y} = \begin{matrix} \left[\begin{array}{c} 102 \\ 110 \\ 90 \\ 92 \\ 115 \\ 120 \\ 140 \\ 130 \\ 112 \\ 140 \\ 150 \\ 130 \\ 125 \\ 140 \\ 125 \\ 100 \end{array} \right] \end{matrix}.$$

$$A = x_1 x_2 \quad B = x_1 x_3 \quad C = x_1 x_4 \quad D = x_2 x_3 \quad E = x_2 x_4 \quad F = x_3 x_4 \quad G = x_1 x_2 x_3$$

$$H = x_1 x_2 x_4 \quad I = x_1 x_3 x_4 \quad J = x_2 x_3 x_4 \quad K = x_1 x_2 x_3 x_4$$

Obteve-se a matriz \mathbf{m} pela Eq. 27, $\mathbf{m} = \mathbf{X}'\mathbf{y}$. Os valores obtidos da matriz \mathbf{m} e os efeitos obtidos pela Eq. 28 apresentam-se a seguir e na Tabela 2.23:

$$\mathbf{m} = \begin{bmatrix} 1921 \\ 3 \\ -7 \\ 69 \\ 123 \\ -109 \\ -33 \\ -7 \\ -3 \\ -17 \\ -153 \\ -1 \\ -67 \\ -3 \\ -133 \\ 17 \end{bmatrix}; \mathbf{ef} = \begin{bmatrix} \frac{M_{11}}{2^k} \\ \frac{M_{21}}{2^{k-1}} \\ \vdots \\ \frac{M_{\beta 1}}{2^{k-1}} \end{bmatrix} \quad k = 4 \text{ variáveis} \Rightarrow \mathbf{ef} = \begin{bmatrix} 120,062 \\ 0,375 \\ -0,875 \\ 8,625 \\ 15,375 \\ -13,625 \\ -4,125 \\ -0,875 \\ -0,375 \\ -2,125 \\ -19,125 \\ -0,125 \\ -8,375 \\ -0,375 \\ -16,625 \\ 2,125 \end{bmatrix}$$

Já para o planejamento de experimentos fatorial fracionário 2^{4-1} considerou-se o modelo linear completo para 3 variáveis, ou seja, com os coeficientes das interações de segunda e terceira ordem, além dos coeficientes da média experimental e dos efeitos principais:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_1 x_2 + \hat{\beta}_5 x_1 x_3 + \hat{\beta}_6 x_2 x_3 + \hat{\beta}_7 x_1 x_2 x_3.$$

Com base nisto, a ordem da dimensão da matriz de contraste do planejamento 2^{4-1} é 8, ou seja, 8 linhas e 8 colunas, e a dimensão da matriz de resposta é 8 linhas e 1 coluna. Ambas as matrizes apresentam-se a seguir:

$$\mathbf{X} = \begin{matrix} \bar{y} & x_1 & x_2 & x_3 & x_4 & A & B & C \\ \begin{bmatrix} 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} & \mathbf{y} = \begin{bmatrix} 102 \\ 140 \\ 150 \\ 92 \\ 125 \\ 120 \\ 140 \\ 100 \end{bmatrix} \end{matrix}.$$

$$x_4 = x_1 x_2 x_3 \quad A = x_1 x_2 = x_3 x_4 \quad B = x_1 x_3 = x_2 x_4 \quad C = x_1 x_4 = x_2 x_3$$

Obteve-se a matriz \mathbf{m} pela Eq. 27, $\mathbf{m} = \mathbf{X}'\mathbf{y}$, e a matriz dos efeitos pela Eq. 28, em que os detalhes dos cálculos matriciais apresentam-se a seguir e na Tabela 2.24.

$$\mathbf{m} = \begin{bmatrix} 969 \\ -65 \\ -5 \\ 1 \\ 61 \\ -131 \\ -25 \\ -5 \end{bmatrix}; \text{EF} = \begin{bmatrix} \frac{M_{11}}{2^k} \\ \frac{M_{21}}{2^{k-1}} \\ \vdots \\ \frac{M_{\beta 1}}{2^{k-1}} \end{bmatrix} \quad k = 3 \text{ variáveis} \Rightarrow \mathbf{ef} = \begin{bmatrix} 121,125 \\ -16,25 \\ -1,25 \\ 0,25 \\ 15,25 \\ -32,75 \\ -6,25 \\ -1,25 \end{bmatrix}.$$

Percebe-se na Tabela 2.23 que existem alguns experimentos, a metade – *meia fração positiva* ou *primeira meia fração*, em que a codificação de x_4 é igual ao produto das codificações x_1 , x_2 e x_3 , onde se utilizou a geratriz $I = x_1x_2x_3x_4$. Os experimentos selecionados com os respectivos símbolos experimentais estão na Tabela 2.24, em que os níveis menores foram selecionados primeiros, como se apresentou na Tabela 2.21.

Tabela 2.24 – Planejamento de Experimentos Fatorial Fracionário 2^{4-1} do Exemplo 12

Exp.	Símbolos dos Experimentos	Respostas Experimentais	x_1	x_2	x_3	x_4	x_1x_2	x_1x_3	x_1x_4	Efeitos Obtidos
1	(1)	102	-1	-1	-1	-1	1	1	1	$\bar{y}_F = 121,125$
10	x_1x_4	140	1	-1	-1	1	-1	-1	1	$X_{F_1} = X_{F_2} X_{F_3} X_{F_4} = -16,25$
11	x_2x_4	150	-1	1	-1	1	-1	1	-1	$X_{F_2} = X_{F_1} X_{F_3} X_{F_4} = -1,25$
4	x_1x_2	92	1	1	-1	-1	1	-1	-1	$X_{F_3} = X_{F_1} X_{F_2} X_{F_4} = 0,25$
13	x_3x_4	125	-1	-1	1	1	1	-1	-1	$X_{F_4} = X_{F_1} X_{F_2} X_{F_3} = 15,25$
6	x_1x_2	120	1	-1	1	-1	-1	1	-1	$X_{F_1} X_{F_2} = X_{F_3} X_{F_4} = -32,75$
7	x_2x_3	140	-1	1	1	-1	-1	-1	1	$X_{F_1} X_{F_3} = X_{F_2} X_{F_4} = -6,75$
16	$x_1x_2x_3x_4$	100	1	1	1	1	1	1	1	$X_{F_1} X_{F_4} = X_{F_2} X_{F_3} = -1,25$

Percebe-se na Tabela 2.22 que os efeitos principais do planejamento de experimentos fatorial fracionário 2^{4-1} são misturados com os efeitos terciários ou de terceira ordem do planejamento 2^4 , e que os efeitos do fracionário são uma soma dos efeitos do fatorial completo. Na Tabela 2.25 apresentam-se os efeitos do planejamento de experimentos fatorial fracionário 2^{4-1} que são resultados da soma dos efeitos do completo 2^4 do Exemplo 12, e que através do fatorial fracionário é possível obter às mesmas conclusões daquelas oriundas de um fatorial completo, dependendo do objetivo do fenômeno em estudo e da significância dos efeitos.

Tabela 2.25. Relações entre os contrastes da meia fração 2^{4-1} e os efeitos do fatorial completo 2^4

Contrastes da meia fração 2^{4-1} em termos de efeitos do fatorial 2^4	Efeitos do Fracionário	Efeitos do Completo
$X_{F_1} = X_{F_2} X_{F_3} X_{F_4} = X_1 + X_2 X_3 X_4$	-16,25	$X_1 + X_2 X_3 X_4 = 0,375 + (-16,625) = -16,25$
$X_{F_2} = X_{F_1} X_{F_3} X_{F_4} = X_2 + X_1 X_3 X_4$	-1,25	$X_2 + X_1 X_3 X_4 = -0,875 + (-0,375) = -1,25$
$X_{F_3} = X_{F_1} X_{F_2} X_{F_4} = X_3 + X_1 X_2 X_4$	0,25	$X_3 + X_1 X_2 X_4 = 8,625 + (-8,375) = 0,25$
$X_{F_4} = X_{F_1} X_{F_2} X_{F_3} = X_4 + X_1 X_2 X_3$	15,25	$X_4 + X_1 X_2 X_3 = 15,375 + (-0,125) = 15,25$
$X_{F_1} X_{F_2} = X_{F_3} X_{F_4} = X_1 X_2 + X_4 X_3$	-32,75	$X_1 X_2 + X_4 X_3 = -13,625 + (-19,125) = -32,75$
$X_{F_1} X_{F_3} = X_{F_2} X_{F_4} = X_1 X_3 + X_2 X_4$	-6,25	$X_1 X_3 + X_2 X_4 = X_1 X_3 = -4,125 + (-2,125)$
$X_{F_1} X_{F_4} = X_{F_2} X_{F_3} = X_1 X_4 + X_2 X_3$	-1,25	$X_1 X_4 + X_2 X_3 = 0,875 + (-0,375) = -1,25 -$
$\bar{y}_F = \bar{y} + 1/2(X_1 X_2 X_3 X_4)$	121,125	$\bar{y} + 1/2(X_1 X_2 X_3 X_4) = 120,062 + 1/2(2,125) = 121,125$

As funções estimadas completas para o planejamento 2^4 e 2^{4-1} do Exemplo 12 apresentam-se a seguir, respectivamente:

$$\hat{y} = 120,062 + \frac{0,375}{2}x_1 - \frac{0,875}{2}x_2 + \frac{8,625}{2}x_3 + \frac{15,375}{2}x_4 - \frac{13,625}{2}x_1x_2 - \frac{4,125}{2}x_1x_3 - \frac{0,875}{2}x_1x_4 - \frac{0,375}{2}x_2x_3 - \frac{2,125}{2}x_2x_4 - \frac{19,125}{2}x_3x_4 - \frac{0,125}{2}x_1x_2x_3 - \frac{8,375}{2}x_1x_2x_4 - \frac{0,375}{2}x_1x_3x_4 - \frac{16,625}{2}x_2x_3x_4 + \frac{2,125}{2}x_1x_2x_3x_4;$$

$$\hat{y} = 121,125 - \frac{16,25}{2}x_1 - \frac{1,25}{2}x_2 + \frac{0,25}{2}x_3 + \frac{15,25}{2}x_1x_2 - \frac{32,75}{2}x_1x_3 - \frac{6,25}{2}x_2x_3 - \frac{1,25}{2}x_1x_2x_3.$$

Nas duas funções estimadas os coeficientes da regressão são a metade dos respectivos efeitos obtidos, exceto a média, como se comentou anteriormente. Para distinguir quais efeitos são significativos, geralmente consideram-se os valores maiores, caso não seja realizados experimentos com replicatas. Desse modo, infere-se que a função estimada significativa para os dois planejamentos 2^4 e 2^{4-1} apresentam-se a seguir, respectivamente:

$$\hat{y} = 120,062 + \frac{8,625}{2}x_3 + \frac{15,375}{2}x_4 - \frac{13,625}{2}x_1x_2 - \frac{19,125}{2}x_3x_4 - \frac{8,375}{2}x_1x_2x_4 - \frac{16,625}{2}x_2x_3x_4;$$

$$\hat{y} = 121,125 - \frac{16,25}{2}x_1 + \frac{15,25}{2}x_1x_2 - \frac{32,75}{2}x_1x_3.$$

Na seção de gráficos de probabilidade normal será apresentada uma metodologia de como visualizar a partir dos gráficos quais efeitos são significativos para o Exemplo 12, por exemplo.

Todas as ferramentas do planejamento de experimentos fatorial de meia fração, 2^{k-1} , apresentados anteriormente foram implementados no *software* PlanEx sendo k a quantidade de variáveis no intervalo $3 \leq k \leq 6$. Como se comentou anteriormente, o fato de escolher uma ou outra geratriz não interfere nas conclusões dos resultados, portanto, no *software* PlanEx

implementou todos os planejamentos 2^{k-1} , $3 \leq k \leq 6$, usando como geratriz a *meia fração positiva*. A quantidade r de replicatas de todos os experimentos implementada no *software* PlanEx consiste no intervalo $1 \leq r \leq 4$ e se $r = 1$ não há replicatas.

2.6 Planejamento de Experimentos Fatorial com Ponto Central

Em várias situações envolvendo planejamento de experimentos a realização de replicatas de todos os experimentos (repetições autênticas) pode ser inconveniente por diversas razões, por exemplo, o aspecto temporal e financeiro. Para contornar esta dificuldade e obter uma estimativa dos erros, inclui-se um experimento no centro do planejamento de experimentos fatorial, Figura 2.8, em que emprega-se o valor médio dos níveis de todas as variáveis, que se denomina de experimentos no ponto central ou no nível zero.

Sabe-se que através de um planejamento de experimentos fatorial completo ou fracionário de dois níveis estima-se quais são os efeitos principais e quais são as interações, julgando quais são significativos com a realização de replicatas e/ou com a análise de gráficos de probabilidade normal, como se apresentará na seção desses gráficos. Já com um planejamento de experimentos fatorial com ponto central, ou seja, três níveis, além de saber quais são os efeitos principais e as interações, é possível calcular o grau de curvatura da superfície (BOX *et al*, 1978; FERREIRA, M. e TEÓFILO, 2006; BRUNS *et al*, 2001) e obter uma estimativa de erro dos coeficientes da regressão realizando uma quantidade menor de replicatas, pois através de replicatas do ponto central é possível obter uma estimativa dos erros, Figura 2.8.

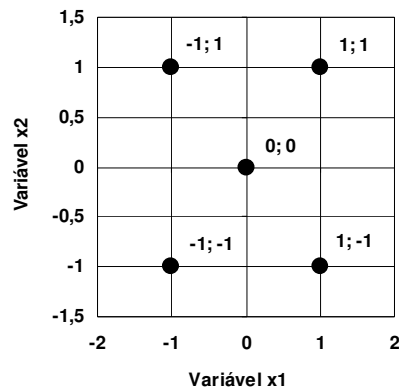


Figura 2.8 – Planejamento de Experimentos Fatorial 2^2 com ponto central

A estimativa do erro dos efeitos de um planejamento de experimentos fatorial com ponto central através das replicatas do ponto central obtém-se pela Eq. 34.1:

$$s_{\text{Efeitos}}^2 = s_{\text{Nível_Central}}^2 \cdot (\mathbf{X}^t \mathbf{X})^{-1} \Rightarrow s_{\text{Efeitos}} = \sqrt{s_{\text{Nível_Central}}^2 \cdot (\mathbf{X}^t \mathbf{X})^{-1}}. \quad (34.1)$$

E usando a Eq. 34.2 obtém-se o grau de curvatura ou estimativa de curvatura de uma superfície de experimentos, permitindo analisar se o modelo linear adequa-se aos dados ou não, podendo substituí-lo por outro modelo, por exemplo, o quadrático:

$$E_C = \bar{y}_{\text{Fatorial}} - \bar{y}_{\text{Central}}. \quad (34.2)$$

Onde:

- E_C é a estimativa da curvatura de uma superfície experimental;
- $\bar{y}_{\text{Fatorial}}$ é a média dos experimentos do planejamento de experimentos fatorial;
- \bar{y}_{Central} é a média das replicatas dos experimentos realizados no ponto central.

E a estimativa do erro da $E_C (\bar{y}_{\text{Fatorial}} - \bar{y}_{\text{Central}})$ obtém-se pela Eq. 34.3:

$$s_{E_C}^2 (\bar{y}_{\text{Fatorial}} - \bar{y}_{\text{Central}}) = s_{E_C}^2 (\bar{y}_{\text{Fatorial}}) + s_{E_C}^2 (\bar{y}_{\text{Central}}) = \left(\frac{\sigma}{\sqrt{n}} \right)^2 + \left(\frac{\sigma}{\sqrt{n}} \right)^2. \quad (34.3)$$

Onde o desvio padrão (erro) dos efeitos σ é o dobro do desvio padrão (erro) dos coeficientes da regressão, como se comentou anteriormente no Exemplo 10 na página 47, exceto a média dos experimentos.

No Exemplo 8, página 42, apresentou-se um planejamento de experimentos fatorial 2^2 acrescido de triplicatas do ponto central do trabalho de Peralta-Zamora *et al* (2005). Com base nos dados do trabalho de Peralta-Zamora *et al* (2005) foi aplicado a Eq. 34.1 para analisar quais coeficientes do modelo linear são significativos, e com as Eq. 34.2 e 34.3 para verificar se o modelo linear adequa aos dados experimentais. Os detalhes dos cálculos apresentam-se a seguir.

Obteve-se $s_{\text{Nível_Central}}^2 = 1, \bar{3}$ como a variância dos experimentos realizados no nível central do Exemplo 8, sendo 7 linhas e 4 colunas a dimensão da matriz de contraste \mathbf{X} para o modelo linear, com base nisto a matriz $(\mathbf{X}^t \mathbf{X})^{-1}$ possui dimensão 4 x 4. Obteve-se a matriz dos erros dos efeitos pela Eq. 34.1:

$$S_{\text{Efeitos}}^2 = S_{\text{Nível_Central}}^2 \cdot (\mathbf{X}'\mathbf{X})^{-1}$$

$$S_{\text{Efeitos}}^2 = 1, \bar{3} \cdot \begin{bmatrix} 0,1429 & 0 & 0 & 0 \\ 0 & 0,25 & 0 & 0 \\ 0 & 0 & 0,25 & 0 \\ 0 & 0 & 0 & 0,25 \end{bmatrix}$$

$$S_{\text{Efeitos}}^2 = \begin{bmatrix} 0,1905 & 0 & 0 & 0 \\ 0 & 0, \bar{3} & 0 & 0 \\ 0 & 0 & 0, \bar{3} & 0 \\ 0 & 0 & 0 & 0, \bar{3} \end{bmatrix}$$

$$S_{\text{Efeitos}} = \sqrt{\begin{bmatrix} 0,1905 & 0 & 0 & 0 \\ 0 & 0, \bar{3} & 0 & 0 \\ 0 & 0 & 0, \bar{3} & 0 \\ 0 & 0 & 0 & 0, \bar{3} \end{bmatrix}}$$

$$S_{\text{Efeitos}} = \begin{bmatrix} 0,4364 & 0 & 0 & 0 \\ 0 & 0,5774 & 0 & 0 \\ 0 & 0 & 0,5774 & 0 \\ 0 & 0 & 0 & 0,5774 \end{bmatrix}$$

Onde o desvio padrão da média é 0,4364, e os demais valores da diagonal principal representam o desvio padrão dos efeitos da variável 1, da variável 2, e o efeito de interação das variáveis 1 e 2, respectivamente, do modelo linear. O modelo estimado com os respectivos desvios padrão dos coeficientes do Exemplo 8 apresenta-se a seguir e indica que todos os coeficientes são significativos:

$$\hat{y} = \underbrace{73,1429}_{(\pm 0,4364)} - \underbrace{6,5000x_1}_{(\pm 0,5774)} - \underbrace{3,0000x_2}_{(\pm 0,5774)} - \underbrace{15,0000x_1x_2}_{(\pm 0,5774)}$$

A média dos experimentos do planejamento de experimentos fatorial, a média das replicatas do ponto central e a estimativa da curvatura do Exemplo 8 apresentam-se a seguir:

$$\bar{y}_{\text{Fatorial}} = 67,50 \quad \bar{y}_{\text{Central}} = 80,67$$

$$E_c = \bar{y}_{\text{Fatorial}} - \bar{y}_{\text{Central}}$$

$$E_c = 67,50 - 80,67$$

$$E_c = -13,17.$$

O desvio padrão da Eq. 34.3 é o dobro do desvio padrão dos coeficientes da regressão, com base nisto o desvio padrão da estimativa conjunta $s_{E_c}(\bar{y}_{\text{Fatorial}} - \bar{y}_{\text{Central}})$ apresenta-se:

$$s_{E_c}^2(\bar{y}_{\text{Fatorial}} - \bar{y}_{\text{Central}}) = s_{E_c}^2(\bar{y}_{\text{Fatorial}}) + s_{E_c}^2(\bar{y}_{\text{Central}})$$

$$s_{E_c}^2(\bar{y}_{\text{Fatorial}} - \bar{y}_{\text{Central}}) = \left(\frac{\sigma}{\sqrt{n}}\right)^2 + \left(\frac{\sigma}{\sqrt{n}}\right)^2$$

$$s_{E_c}^2(\bar{y}_{\text{Fatorial}} - \bar{y}_{\text{Central}}) = \left(\frac{1,1548}{\sqrt{7}}\right)^2 + \left(\frac{1,1548}{\sqrt{7}}\right)^2$$

$$s_{E_c}^2(\bar{y}_{\text{Fatorial}} - \bar{y}_{\text{Central}}) = 0,3810$$

$$s_{E_c}(\bar{y}_{\text{Fatorial}} - \bar{y}_{\text{Central}}) = \sqrt{0,3810} = 0,6173.$$

Com os cálculos obtidos do Exemplo 8 percebe-se que o coeficiente do efeito de interação das variáveis x_1 e x_2 é significativo:

$$\hat{\beta}_3 = -15 \pm 0,5774 \Rightarrow \text{significativo}.$$

Isso indica que o modelo não será um plano, pois as variáveis possuem sinergismo. O valor da estimativa conjunta da superfície corrobora o fato da superfície não ser um plano:

$$E_c = -13,17 \pm 0,6173 \Rightarrow \text{significativo}.$$

Apesar dos coeficientes do modelo linear serem significativos do Exemplo 8, é possível melhorar a estimativa da regressão. O acréscimo de termos quadráticos na regressão para obter um modelo quadrático provavelmente resultará numa melhor representação da superfície experimental, podendo quantificar qual modelo prediz melhor a superfície experimental, como se apresentará na seção da RSM e na seção da ANOVA.

2.7 Metodologia de Superfície de Resposta - RSM

O planejamento de experimentos fatorial é uma técnica estatística útil e importante numa investigação preliminar na realização de experimentos, e pode fornecer bons resultados. Com o objetivo de aperfeiçoar os resultados obtidos por um planejamento de experimentos fatorial ou até mesmo encontrar o ponto ótimo de um problema experimental George. E. P. Box introduziu por volta de 1950 uma técnica de modelagem utilizada em diversos processos industriais com o intuito de otimizá-los baseada nos planejamentos de experimentos fatorial

denominada de metodologia de superfície de resposta, que é frequentemente representada pela sigla RSM, de *Response Surface Methodology* (BOX *et al*, 1978; BRUNS *et al*, 2001).

Modelagem e deslocamento são as duas etapas distintas da RSM, as quais são repetidas tantas vezes quantas forem necessárias, com o objetivo de encontrar um ponto ótimo ou uma região ótima da superfície de experimentos investigada. A etapa da modelagem na química é realizada inicialmente através de modelos simples, em geral lineares ou quadráticos, ajustando-os as respostas experimentais obtidas com planejamentos de experimentos fatorial completos ou fracionários ou até mesmo com planejamentos fatoriais ampliados (BOX, 1978; BRUNS, ET AL, 2001). A etapa de deslocamento realiza-se na direção do caminho de máxima inclinação de um determinado modelo, que é a trajetória na qual a resposta varia de forma mais pronunciada.

No presente trabalho, a RSM foi implementada de dois modos, modelo linear e modelo quadrático, com até 4 variáveis independentes. O modelo linear constitui-se de planejamento de experimentos fatorial completo acrescido de replicatas r ($1 \leq r \leq 5$) do ponto central, enquanto o modelo quadrático constitui-se de planejamento de experimentos fatorial completo acrescido de planejamento estrela e replicatas r ($1 \leq r \leq 5$) do ponto central, que se denomina de planejamento composto central.

2.7.1 Modelo Linear da RSM

Um modelo estimado linear, \hat{y} , com k variáveis é dado na notação de somatório pela Eq. 35:

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i x_i + \sum_{i < j} \sum_j \hat{\beta}_{ij} x_i x_j. \quad (35)$$

Utiliza-se o modelo linear, Eq. 35, em qualquer tipo de planejamento de experimentos fatorial. Se um pesquisador sabe que o modelo está funcionando bem, porém ele gostaria melhorar ainda mais o resultado, uma boa escolha é fazer um planejamento com outros níveis. O primeiro passo do pesquisador, para atacar o problema, é investigar a superfície de experimentos em torno das condições habituais de funcionamento do processo, usando o planejamento de experimentos fatorial com ponto central, mostrado na Figura 2.8, e usar um modelo linear como Eq. 35 (BRUNS *et al*, 2001).

Note que o planejamento contém um ponto central, e por isso varre três níveis de cada variável, e não apenas dois. Isto permite verificar se há ou não falta de ajuste para um modelo

linear, o que seria impossível se tivesse usado apenas dois níveis, como se comentou nas Eq. 34.2 e 34.3 aplicados ao Exemplo 8.

No presente trabalho ao considerar a RSM através de um modelo linear, o pesquisador terá opção de incluir repetições autênticas, replicatas r , do ponto central no limite de 5 repetições, proporcionando a ele a possibilidade de obter estimativas do erro experimental (BOX *et al*, 1978; FERREIRA, M. e TEÓFILO, 2006; BRUNS *et al*, 2001; OLIVIER *et al*, 2007), sendo a quantidade de experimentos dada por $2^k + r$.

2.7.2 Modelo Quadrático da RSM

Ao usar um planejamento fatorial com ponto central para um modelo linear, se o pesquisador perceber que as respostas experimentais estão aumentando e diminuindo a partir de um valor de referência, é hora de pensar em mudar de modelo, e uma boa escolha seria um modelo quadrático, como se apresenta em notação de somatório na Eq. 36,

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i x_i + \sum_{i=1}^k \hat{\beta}_{ii} x_i^2 + \sum_{i<j} \sum_j \hat{\beta}_{ij} x_i x_j. \quad (36)$$

Onde k é a quantidade de variáveis.

No presente trabalho os modelos quadráticos foram implementados usando as teorias dos planejamentos compostos centrais, apresentados por Box e Wilson (1951) como uma evolução dos planejamentos 3^k , que necessitavam de muitos experimentos para um pequeno número de variáveis, mesmo para planejamentos fracionários (BOX *et al*, 1978; FERREIRA, M. e TEÓFILO, 2006; BRUNS *et al*, 2001; OLIVIER *et al*, 2007).

Para construir um planejamento composto central define-se:

- a quantidade de variáveis a serem estudadas (k);
- o planejamento de experimentos fatorial a ser empregado (completo 2^k ou fracionário 2^{k-b});
- e quantos experimentos serão realizados no ponto central.

No presente trabalho a quantidade de experimentos do modelo quadrático da RSM é dada pela $2^k + 2k + r$, onde k é a quantidade de variáveis e r é a quantidade de replicatas do ponto central. O termo 2^k refere-se ao *planejamento fatorial completo* e $2k$ ao *planejamento estrela*, que faz um giro de 45° em relação ao planejamento de experimentos fatorial inicial.

A codificação dos níveis de um planejamento estrela nos pontos axiais é dada por $x_i = \pm \alpha$, com $x_j \neq x_i$, e:

$$\alpha = \sqrt[4]{2^k}.$$

Onde k é a quantidade de variáveis e os pontos axiais são situados nos eixos do sistema de coordenadas com distância $\pm \alpha$ da origem (BOX e WILSON, 1951; BOX *et al*, 1978; BRUNS *et al*, 2001).

A matriz de Contraste para um planejamento composto central para duas variáveis representa-se na forma geral pela Eq. 38:

$$\mathbf{X} = \begin{matrix} & \bar{y} & x_1 & x_2 & x_1^2 & x_2^2 & x_1x_2 \\ \begin{bmatrix} 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -\alpha & 0 & \alpha^2 & 0 & 0 \\ 1 & \alpha & 0 & \alpha^2 & 0 & 0 \\ 1 & 0 & -\alpha & 0 & \alpha^2 & 0 \\ 1 & 0 & \alpha & 0 & \alpha^2 & 0 \end{bmatrix} \end{matrix} \quad (38)$$

$$\mathbf{X} = \begin{matrix} & \bar{y} & x_1 & x_2 & x_3 & x_1^2 & x_2^2 & x_3^2 & x_1x_2 & x_1x_3 & x_2x_3 & x_1x_2x_3 \\ \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -\alpha & 0 & 0 & 0 & \alpha^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & \alpha & 0 & 0 & 0 & \alpha^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -\alpha & 0 & 0 & 0 & \alpha^2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & \alpha & 0 & 0 & 0 & \alpha^2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -\alpha & 0 & 0 & 0 & \alpha^2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & \alpha & 0 & 0 & 0 & \alpha^2 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix} \quad (39)$$

Na Eq. 38 as 4 primeiras linhas se referem ao planejamento fatorial completo 2^2 . A linha do meio se refere ao ponto central ou nível zero e poderá ser estendida até 5 replicatas neste trabalho, e as últimas quatro linhas se referem ao planejamento estrela, em que o valor de α é obtido pela Eq. 37. As colunas 4 e 5 se referem ao termo quadrático do modelo

considerado obtidas pelo quadrado das colunas 2 e 3, respectivamente. E a coluna 6 se refere à codificação do efeito de interação das variáveis x_1 e x_2 .

A matriz de Contraste para um planejamento composto central para três variáveis representa-se na forma geral pela Eq. 39, onde as 8 primeiras linhas se referem ao planejamento fatorial completo 2^3 . A linha 9 se refere ao nível central e poderá ser estendida até 5 replicatas neste trabalho, e as últimas seis linhas se referem ao planejamento estrela, em que o valor de α é obtido pela Eq. 37. As colunas 5, 6 e 7 se referem ao termo quadrático do modelo considerado obtidas pelo quadrado das colunas 2, 3 e 4, respectivamente. E as demais colunas representam a codificação dos efeitos de interação de segunda e terceira ordem das variáveis.

Exemplo 13. Na Introdução deste trabalho comentou-se que o trabalho de Joelma Ferreira *et al* (2005) analisou o efeito das fontes de nutrientes sobre a produção de vinagre em escala de bancada usando inicialmente um planejamento de experimentos fatorial 2^3 com triplicatas do ponto central para avaliar os efeitos das variáveis independentes: concentrações iniciais de etanol, de fontes de fósforo e de nitrogênio. Após esta análise as faixas das concentrações iniciais das variáveis de fontes de nutrientes (fósforo e nitrogênio) foram ampliadas para o planejamento de experimentos fatorial do tipo 2^2 mais configuração estrela, denominado de planejamento composto central, em que os dados do planejamento fatorial apresentam-se na Tabela 2.26.

Tabela 2.26 – Planejamento Composto Central do Exemplo 13 com 11 experimentos

Exp.	\bar{y}	X_1	X_2	X_1^2 **	X_2^2 ***	X_1X_2	Símbolos dos Experimentos	Respostas Experimentais
1	1	-1	-1	1	1	1	(1)	29,33
2	1	1	-1	1	1	-1	x_1	31,96
3	1	-1	1	1	1	-1	x_2	44,33
4	1	1	1	1	1	1	x_1x_2	40,29
5	1	0	0	0	0	0	o_1	38,93
6	1	0	0	0	0	0	o_2	43,35
7	1	0	0	0	0	0	o_3	36,72
8	1	-1,4142	0	2	0	0	x_1e-	40,42
9	1	1,4142	0	2	0	0	x_1e+	32,98
10	1	0	-1,4142	0	2	0	x_2e-	27,30
11	1	0	1,4142	0	2	0	x_2e+	56,87

** x_1x_1 e *** x_2x_2 são os termos quadráticos do modelo quadrático da RSM.

Fonte: (FERREIRA, J. *et al*, 2005).

A Tabela 2.26 representa um planejamento composto central para o modelo quadrático com duas variáveis

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{11} x_1^2 + \hat{\beta}_{22} x_2^2 + \hat{\beta}_{12} x_1 x_2$$

e com triplicatas do ponto central, onde o valor de $\alpha = \pm 1,41$ obteve-se pela Eq. 37. Foi inserido a coluna 2 da Tabela 2.26 para obter a média dos experimentos. A coluna 3 e 4 representam as codificações das variáveis x_1 e x_2 , respectivamente. As colunas 5 e 6 obtiveram-se pelo quadrado dos sinais das colunas 3 e 4, respectivamente, representando o termo quadrático das variáveis x_1 e x_2 , nessa ordem. E a coluna 7 representa a codificação da interação das variáveis x_1 e x_2 e obteve-se pela multiplicação dos sinais das colunas 3 e 4.

Na coluna 8 da Tabela 2.26 apresentam-se os símbolos utilizados neste trabalho de doutorado nos planejamentos da RSM para duas variáveis:

- O símbolo (1) significa o(s) experimento(s) realizado(s) no nível menor de todas as variáveis;
- O símbolo x_1 significa o(s) experimento(s) realizado(s) no nível maior de x_1 e no nível menor das demais;
- O símbolo x_2 significa o(s) experimento(s) realizado(s) no nível maior de x_2 e no nível menor das demais;
- O símbolo $x_1 x_2$ significa o(s) experimento(s) realizado(s) no nível maior das variáveis x_1 e x_2 ;
- O símbolo o_j , com $j = 1, 2, \dots, r$ replicatas, significa o(s) experimento(s) realizado(s) no nível zero ou ponto central das variáveis;
- O símbolo x_{1e-} significa o(s) experimento(s) realizado(s) no nível menor da variável x_1 do planejamento estrela;
- O símbolo x_{1e+} significa o(s) experimento(s) realizado(s) no nível maior da variável x_1 do planejamento estrela;
- O símbolo x_{2e-} significa o(s) experimento(s) realizado(s) no nível menor da variável x_2 do planejamento estrela;
- O símbolo x_{2e+} significa o(s) experimento(s) realizado(s) no nível maior da variável x_2 do planejamento estrela.

Nota. Os sete primeiros símbolos da coluna 8 da Tabela 2.26 também se referem a RSM para o modelo linear com duas variáveis e com triplicatas do ponto central.

A matriz de contraste do Exemplo 13 para o modelo quadrático completo para duas variáveis apresenta-se na Tabela 2.26, sendo constituída das colunas de 2 a 7, com dimensão de ordem 6. A matriz de respostas experimentais, coluna 9 da Tabela 2.26, possui 11 linhas e 1 coluna. A quantidade de experimentos, 11, obteve-se através da equação $2^k + 2k + r$, sendo $k = 2$ variáveis e $r = 3$ replicatas do ponto central. Os parâmetros ou coeficientes da regressão obtiveram-se usando a Eq. 19:

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1,4142 & 0 & 2 & 0 & 0 \\ 1 & 1,4142 & 0 & 2 & 0 & 0 \\ 1 & 0 & -1,4142 & 0 & 2 & 0 \\ 1 & 0 & 1,4142 & 0 & 2 & 0 \end{bmatrix}; \mathbf{y} = \begin{bmatrix} 29,33 \\ 31,96 \\ 44,33 \\ 40,29 \\ 38,93 \\ 43,35 \\ 36,72 \\ 40,42 \\ 32,98 \\ 27,30 \\ 56,87 \end{bmatrix}; \hat{\boldsymbol{\beta}} = \begin{bmatrix} 39,67 \\ -1,49 \\ 8,14 \\ -2,21 \\ 0,48 \\ -1,67 \end{bmatrix}.$$

Através da replicatas do ponto central é possível obter uma estimativa dos erros. A variância amostral das replicatas do ponto central aplicada na Eq. 34.1 encontra-se:

$$s^2_{\text{Nível_Central}} = 11,396$$

$$s^2_{\text{Efeitos}} = s^2_{\text{Nível_Central}} \cdot (\mathbf{X}^t \mathbf{X})^{-1}$$

$$s^2_{\text{Efeitos}} = 11,396 \cdot \begin{bmatrix} 0,3 & 0 & 0 & -0,167 & -0,167 & 0 \\ 0 & 0,125 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,125 & 0 & 0 & 0 \\ -0,167 & 0 & 0 & 0,177 & 0,052 & 0 \\ -0,167 & 0 & 0 & 0,052 & 0,177 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,250 \end{bmatrix}$$

$$s^2_{\text{Efeitos}} = \begin{bmatrix} 3,799 & 0 & 0 & -1,899 & -1,899 & 0 \\ 0 & 1,425 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1,425 & 0 & 0 & 0 \\ -1,899 & 0 & 0 & 2,018 & 0,594 & 0 \\ -1,899 & 0 & 0 & 0,594 & 2,018 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2,849 \end{bmatrix}$$

$$S_{\text{Efeitos}} = \begin{bmatrix} 3,799 & 0 & 0 & -1,899 & -1,899 & 0 \\ 0 & 1,425 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1,425 & 0 & 0 & 0 \\ -1,899 & 0 & 0 & 2,018 & 0,594 & 0 \\ -1,899 & 0 & 0 & 0,594 & 2,018 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2,849 \end{bmatrix}$$

$$S_{\text{Efeitos}} = \begin{bmatrix} 1,949 & 0 & 0 & \bar{X} \in \mathbb{R} & \bar{X} \in \mathbb{R} & 0 \\ 0 & 1,194 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1,194 & 0 & 0 & 0 \\ \bar{X} \in \mathbb{R} & 0 & 0 & 1,421 & 0,770 & 0 \\ \bar{X} \in \mathbb{R} & 0 & 0 & 0,770 & 1,421 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1,688 \end{bmatrix}.$$

Os elementos na diagonal principal da matriz S_{Efeitos} são os desvios padrão ou erro dos coeficientes da regressão, sendo 1,949 o erro da média dos experimentos, e os demais são os erros dos coeficientes das variáveis e das interações entre elas. O modelo quadrático estimado pela MQO é

$$\hat{y} = \underbrace{39,67}_{(\pm 1,949)} - \underbrace{1,49}_{(\pm 1,194)} x_1 + \underbrace{8,14}_{(\pm 1,194)} x_2 - \underbrace{2,21}_{(\pm 1,421)} x_1^2 + \underbrace{0,48}_{(\pm 1,421)} x_2^2 - \underbrace{1,67}_{(\pm 1,688)} x_1 x_2.$$

Onde os valores entre parênteses são os respectivos desvios padrão dos coeficientes da regressão. Pela equação acima, percebe-se que o coeficiente 0,48 do termo quadrático da variável x_2 e o coeficiente -1,67 da interação das variáveis x_1 e x_2 não são significativos, pois são menores que os seus respectivos desvios padrão (erros), logo o modelo quadrático estimado significativo para o Exemplo 13 da Tabela 2.26 é

$$\hat{y} = \underbrace{39,67}_{(\pm 1,949)} - \underbrace{1,49}_{(\pm 1,194)} x_1 + \underbrace{8,14}_{(\pm 1,194)} x_2 - \underbrace{2,21}_{(\pm 1,421)} x_1^2.$$

Com o modelo estimado é possível construir gráficos de superfície de resposta, como se apresentam na seção de Resultados e Discussão deste Trabalho, para obter a otimização do fenômeno em estudo (BOX *et al*, 1978; FERREIRA, M. e TEÓFILO, 2006; BRUNS, *et al*, 2001).

2.8 Teste de ANOVA para Avaliação do Modelo

De acordo com o livro do autor Gujarati (2006) a análise dos resíduos é de grande importância para avaliar a qualidade do ajuste de qualquer modelo. Em princípio os resíduos

devem ser pequenos e aleatórios, caso contrário o modelo que deixa resíduos consideráveis ou tendenciosos é chamado de modelo ruim. No modelo ideal, todas as previsões coincidiram exatamente com os valores experimentais observados, logo não haveria resíduo algum.

Um dos métodos mais confiáveis para analisar o ajuste de um modelo é ANOVA. Por meio da ANOVA é possível saber se o modelo pode ser usado ou não para fazer estimativas para a etapa do deslocamento e para conclusões sobre a região ótima (ALVES *et al*, 1998; BOX *et al*, 1978; FERREIRA, M. e TEÓFILO, 2006; GUJARATI, 2006; BRUNS *et al*, 2001).

O desvio de uma resposta individual em relação à média de todas as respostas experimentais observadas, que se denota por $(y_i - \bar{y})$, se escreve como a soma de duas parcelas como apresenta a Eq. 40,

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i). \quad (40)$$

O termo $(\hat{y}_i - \bar{y})$ representa o desvio dos valores estimados pelo modelo em relação à média de todos os experimentos realizados nos níveis i . O outro termo, $(y_i - \hat{y}_i)$, representa o desvio das respostas experimentais em todos os níveis i em relação aos valores estimados pelo modelo. Esse último termo, $(y_i - \hat{y}_i)$, se o modelo for bom o desvio deve ser praticamente nulo, pois o que se espera de um modelo ajustado é que os valores experimentais e os valores estimados sejam praticamente iguais, ambos em relação à média \bar{y} .

Operando com o somatório e com o quadrado em ambos os membros na Eq. 40, percebe-se que o somatório dos produtos $(\hat{y}_i - \bar{y}) \cdot (y_i - \hat{y}_i)$ é nulo, portanto o quadrado da Eq. 40 é dado pela Eq. 41,

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (41)$$

Estas somas de quadrados de desvios são rotineiramente denominadas de somas quadráticas, ou abreviadamente S.Q. Com base nisso, a leitura da Eq. 41 é

$$[\text{S.Q. em torno da média}] = [\text{S.Q. devida à regressão}] + [\text{S.Q. residual}].$$

Usando uma simbologia mais compacta, tem:

$$SQ_{\text{Total}} = SQ_{\text{Regressão}} + SQ_{\text{Resíduos}}.$$

Observando a Eq. 41 percebe-se que uma parte da variação total das observações y_i em torno da média \bar{y} é descrita pela equação de regressão, e o restante fica por conta dos resíduos. Por consequência, quanto maior for a parte descrita pela regressão, melhor será o

ajuste do modelo, o que se quantifica por meio da razão denominada de **coeficiente de determinação**, Eq. 42, que significa a porcentagem de variação explicada,

$$R^2 = \frac{SQ_{\text{Regressão}}}{SQ_{\text{Total}}}. \quad (42)$$

Em que o valor máximo do R^2 é 1, e só ocorrerá se não existir nenhum resíduo e portanto toda a variação em torno da média será explicada pela regressão (BRUNS *et al*, 2001; GUJARATI, 2006). Resumindo, quanto mais perto de 1 estiver o valor de R^2 , melhor será o ajuste do modelo às respostas experimentais.

A Tabela 2.27 apresenta o resumo de uma Tabela de ANOVA, onde p é a quantidade de parâmetros da regressão do modelo, n é a quantidade de experimentos realizados. A média quadrática (M.Q) é a razão da soma quadrática pelo grau de liberdade, ambos respectivamente de cada fonte de variação.

Tabela 2.27. Tabela de ANOVA para ajuste de um modelo linear

Fonte de Variação	S. Q	gl	M. Q
Regressão	$\sum (\hat{y}_i - \bar{y})^2$	$p - 1$	$\sum (\hat{y}_i - \bar{y})^2 / (p - 1)$
Resíduos	$\sum (y_i - \hat{y}_i)^2$	$n - p$	$\sum (y_i - \hat{y}_i)^2 / (n - p)$
Total	$\sum (y_i - \bar{y})^2$	$n - 1$	$\sum (y_i - \bar{y})^2 / (n - 1)$

Com as replicatas do ponto central na RSM é possível fazer uma estimativa do erro aleatório, e com essa estimativa obtém-se uma opinião quantitativa para julgar se o modelo considerado (linear ou quadrático) é uma boa representação das observações, ou se precisa modifica-lo (BOX *et al*, 1978; FERREIRA, J. *et al*, 2007; FERREIRA, M. *et al*, 2006; BRUNS *et al*, 2001).

Na Eq. 40, o desvio residual é uma soma de duas parcelas de desvios, como apresenta-se:

$$(y_{ij} - \hat{y}_i) = (y_{ij} - \bar{y}_i) - (\hat{y}_i - \bar{y}_i). \quad (43)$$

Operando com o somatório e com o quadrado em ambos os membros na Eq. 43, percebe-se que o somatório dos produtos $(y_{ij} - \bar{y}_i) \cdot (\hat{y}_i - \bar{y}_i)$ é nulo, portanto o quadrado da Eq. 43 é dado pela Eq. 44,

$$\sum_i^m \sum_j^{n_i} (y_{ij} - \hat{y}_i)^2 = \sum_i^m \sum_j^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_i^m \sum_j^{n_i} (\hat{y}_i - \bar{y}_i)^2 \quad (44)$$

Na Eq. 42 foram realizados n_i experimentos, em que a quantidade dos n experimentos é dada pela soma dos n_i experimentos, $n = \sum n_i$. As replicatas dos experimentos foram identificadas pela letra j , em que y_{ij} é a j -ésima resposta (replicatas j) obtida para o i -ésimo experimento, e a quantidade de níveis diferentes é dada pelo símbolo m . Desse modo \bar{y}_i é a média de cada nível, lembrando que no nível que não há replicatas a média será igual ao valor experimental, e \hat{y}_i é o valor estimado para cada nível i .

Tabela 2.28 - Tabela ANOVA com a possibilidade de julgar o ajuste do modelo com a soma quadrática do erro puro e da falta de ajuste.

Fonte de Variação	S. Q	Gl	M. Q
Regressão	$\sum_i^m \sum_j^{n_i} (\hat{y}_{ij} - \bar{y})^2$	$p - 1$	$MQ_{Regressão} = \frac{\sum_i^m \sum_j^{n_i} (\hat{y}_{ij} - \bar{y})^2}{p - 1}$
Resíduos	$\sum_i^m \sum_j^{n_i} (y_{ij} - \hat{y}_i)^2$	$n - p$	$MQ_{Resíduos} = \frac{\sum_i^m \sum_j^{n_i} (y_{ij} - \hat{y}_i)^2}{n - p}$
Falta de Ajuste	$\sum_i^m \sum_j^{n_i} (\hat{y}_{ij} - \bar{y}_i)^2$	$m - p$	$MQ_{FaltaAjuste} = \frac{\sum_i^m \sum_j^{n_i} (\hat{y}_{ij} - \bar{y}_i)^2}{m - p}$
Erro Puro	$\sum_i^m \sum_j^{n_i} (y_{ij} - \bar{y}_i)^2$	$n - m$	$MQ_{ErroPuro} = \frac{\sum_i^m \sum_j^{n_i} (y_{ij} - \bar{y}_i)^2}{n - m}$
Total	$\sum_i^m \sum_j^{n_i} (y_{ij} - \bar{y})^2$	$n - 1$	
% de variação explicada:		$R^2 = \frac{SQ_{Regressão}}{SQ_{Total}}$	
% máxima de variação explicável		$\frac{SQ_{Total} - SQ_{ErroPuro}}{SQ_{Total}}$	
$F_{Calculado}$		$F_{p-1, n-p} = \frac{MQ_{Regressão}}{MQ_{Resíduos}}$	
$F_{Calculado}$ pela falta de ajuste		$F_{m-p, n-m} = \frac{MQ_{FaltaAjuste}}{MQ_{ErroPuro}}$	

Com o uso de replicatas a Eq. 41 representa-se na forma da Eq. 45:

$$\sum_i^m \sum_j^{n_i} (y_{ij} - \bar{y})^2 = \sum_i^m \sum_j^{n_i} (\hat{y}_{ij} - \bar{y})^2 + \sum_i^m \sum_j^{n_i} (\hat{y}_{ij} - \hat{y}_i)^2. \quad (45)$$

A Tabela 2.28 apresenta o resumo de uma Tabela ANOVA com a possibilidade de julgar o ajuste do modelo com a soma quadrática do erro puro e da falta de ajuste.

Com os dados da tabela de ANOVA aplicados na distribuição de Fisher é possível fazer um teste para verificar se o modelo de regressão é significativo ou não, de acordo com a Eq. 46:

$$F_{\text{Calculado}} = \frac{MQ_{\text{Regressão}}}{MQ_{\text{Resíduos}}}. \quad (46)$$

O $F_{\text{Calculado}}$ tem grau de liberdade $p-1$ no numerador e $n-p$ grau de liberdade no denominador. Se o $F_{\text{Calculado}} > F_{\text{Tabelado}}$ com $p-1$ grau de liberdade no numerador e $n-p$ grau de liberdade no denominador há evidência estatística para acreditar na existência de uma relação linear ou quadrática das variáveis y e X , e quanto maior o valor de $F_{\text{Calculado}}$ melhor. Com base nas experiências práticas, o valor de $F_{\text{Calculado}}$ deve ser pelo menos dez vezes maior que o valor da distribuição F com o grau de significância escolhido para realizar previsões com o modelo (BOX, 1973 apud FERREIRA, M. e TEÓFILO, 2006; BOX, 1987 apud FERREIRA, M. E TEÓFILO, 2006).

Valores altos de $F_{m-p, n-m} = \frac{MQ_{\text{FaltaAjuste}}}{MQ_{\text{ErroPuro}}}$ significam muita falta de ajuste, e para o

modelo não apresentar falta de ajuste é necessário que $F_{\text{Calculado}}$ pela falta de ajuste seja menor que $F_{m-p, n-m}$ tabelado pela falta de ajuste (BRUNS *et al*, 2001).

No presente trabalho programaram-se os cálculos de ANOVA somente na RSM quando a quantidade de r replicatas estiver no intervalo $2 \leq r \leq 5$. Com base no Exemplo 13, página 84, construiu-se a Tabela de ANOVA, Tabela 2.31, em que os detalhes dos cálculos apresentam-se a seguir e a função estimada com os coeficientes significativos obteve-se na seção de RSM nas páginas 86 e 87,

$$\hat{y} = \underbrace{39,67}_{(\pm 1,949)} - \underbrace{1,49}_{(\pm 1,194)} x_1 + \underbrace{8,14}_{(\pm 1,194)} x_2 - \underbrace{2,21}_{(\pm 1,421)} x_1^2.$$

A média das replicatas, a variância das replicatas e a média de todos os experimentos da Tabela 2.26 apresentam-se, respectivamente:

$$\begin{aligned} \bar{y}_{\text{Replicatas}} &= \frac{38,93 + 43,35 + 36,72}{3} = 39,67 \\ s_{\text{Replicatas}}^2 &= \frac{(38,93 - 39,67)^2 + (43,35 - 39,67)^2 + (36,72 - 39,67)^2}{3 - 1} = 11,40 \\ \bar{y} &= \frac{29,33 + 31,96 + 44,33 + 40,29 + 38,93 + \dots + 56,87}{11} = 38,41. \end{aligned}$$

Nota. Apesar dos cálculos nas fórmulas aparecerem com duas casas decimais, a ANOVA realizou-se com todas as casas decimais.

Tabela 2.29 - Tabela apresentando a Soma Quadrática do Exemplo 13 onde a função estimada significativa apresenta-se na forma $\hat{y} = 39,67 - 1,49x_1 + 8,14x_2 - 2,21x_1^2$

Símbolos	i	y_i	\hat{y}_i	S.Q_Reg	S.Q_Res	S.Q_E.P	S.Q_F.A	S.Q_Total
(1)	1	29,33	30,81	57,72	2,19	0,00	2,19	82,40
x_1	2	31,96	27,83	111,88	17,06	0,00	17,06	41,57
x_2	3	44,33	47,09	75,39	7,62	0,00	7,62	35,08
x_1x_2	4	40,29	44,11	32,52	14,59	0,00	14,59	3,54
o_1	5	38,93	39,67	1,59	0,55	0,54	0,00	0,27
o_2	6	43,35	39,67	1,59	13,54	13,57	0,00	24,43
o_3	7	36,72	39,67	1,59	8,70	8,68	0,00	2,85
x_1e-	8	40,42	37,36	1,10	9,38	0,00	9,38	4,05
x_1e+	9	32,98	33,14	27,71	0,03	0,00	0,03	29,46
x_2e-	10	27,30	28,16	105,04	0,74	0,00	0,74	123,37
x_2e+	11	56,87	51,18	163,18	32,36	0,00	32,36	340,87
Soma		422,48		579,33	106,75	22,79	83,96	687,89
Média		38,41		Regressão	Resíduos	Erro Puro	Falta de Ajuste	TOTAL

Detalhes da Tabela 2.29:

- para obter $\sum_i^m \sum_j^{n_i} (\hat{y}_{ij} - \bar{y})^2$ consideraram-se todos os valores estimados \hat{y}_{ij} e a média de todos os experimentos realizados $\bar{y} = 38,41$;
- para obter $\sum_i^m \sum_j^{n_i} (y_{ij} - \hat{y}_i)^2$ consideraram-se todos os experimentos realizados y_{ij} e todos os i valores estimados \hat{y}_i ;
- para obter $\sum_i^m \sum_j^{n_i} (\hat{y}_{ij} - \bar{y}_i)^2$ consideraram-se todos os valores estimados \hat{y}_{ij} e a média de cada resposta experimental \bar{y}_i , sendo a média das replicatas $\bar{y}_{\text{Replicatas}} = 39,67$ e a média de cada experimental realizado sem replicatas o próprio valor experimental: $\bar{y}_i = y_i$;
- para obter $\sum_i^m \sum_j^{n_i} (y_{ij} - \bar{y}_i)^2$ consideraram-se todos os experimentos realizados y_{ij} e a média de cada resposta experimental \bar{y}_i , sendo a média das replicatas

$\bar{y}_{\text{Replicatas}} = 39,67$ e a média de cada experimental realizado sem replicatas o próprio valor experimental: $\bar{y}_i = y_i$;

- o para obter $\sum_i^m \sum_j^{n_j} (y_{ij} - \bar{y})^2$ consideraram-se todos os experimentos realizados y_{ij} e a média de todos os experimentos realizados \bar{y} .

Dividindo a soma quadrática da Tabela 2.29 pelos respectivos graus de liberdade da Tabela 2.30 obtém-se a tabela de ANOVA, Tabela 2.31.

Tabela 2.30 - Tabela o grau de liberdade do Exemplo 13 onde a função estimada significativa apresenta-se na forma $\hat{y} = 39,67 - 1,49x_1 + 8,14x_2 - 2,21x_1^2$

Descrição da ANOVA			Grau de Liberdade: gl		
Variáveis	k	2	Total	n-1	10
Replicatas do ponto central	j	3	Regressão	p-1	3
Níveis Distintos	m	9	Resíduos	n-p	7
Total de Experimentos	n	11	Erro Puro	n-m	2
Coefficientes do Modelo	p	4	Falta de Ajuste	m-p	5

Com base na Tabela 2.31 e com 95% de confiança o $F_{\text{Tabelado}} = 4,35$ com 3 grau de liberdade no numerador e 7 no denominador é menor que o $F_{\text{Calculado}} = 12,66$, o que indica que o modelo é significativo. E o $F_{\text{Tabelado}} = 19,30$ pela falta de ajuste com 5 grau de liberdade no numerador e 2 no denominador é maior que o $F_{\text{Calculado_F.A}} = 1,47$, o que corrobora o fato de não haver falta de ajuste pelo modelo, e comprova que o modelo $\hat{y} = 39,67 - 1,49x_1 + 8,14x_2 - 2,21x_1^2$ é significativo.

Aplicando a Eq. 42 no Exemplo 8 obtém-se $R^2 = 78,66\%$. Com base nisto, percebe-se que apesar de todos os coeficientes do modelo linear serem significativos ele consegue prever de 100% dos valores experimentais apenas 78,66%. Porém, isso pode ser melhorado com o uso de outro modelo, por exemplo, o quadrático, como se indicou pelo uso da Eq. 34.2 ao obter a estimativa de curvatura, a qual foi significativa, nas páginas 79 a 81.

Tabela 2.31 – Tabela de ANOVA do Exemplo 13 onde a função estimada significativa apresenta-se na forma $\hat{y} = 39,67 - 1,49x_1 + 8,14x_2 - 2,21x_1^2$.

Fonte de Variação	Soma Quadrática	Grau de Liberdade	Média Quadrática
Regressão	$\sum_i^m \sum_j^{n_j} (\hat{y}_{ij} - \bar{y})^2 = 579,33$	$p - 1 = 3$	$MQ_{Regressão} = \frac{\sum_i^m \sum_j^{n_j} (\hat{y}_{ij} - \bar{y})^2}{p - 1} = 193,11$
Resíduos	$\sum_i^m \sum_j^{n_j} (y_{ij} - \hat{y}_i)^2 = 106,75$	$n - p = 7$	$MQ_{Resíduos} = \frac{\sum_i^m \sum_j^{n_j} (y_{ij} - \hat{y}_i)^2}{n - p} = 15,25$
Falta de Ajuste	$\sum_i^m \sum_j^{n_j} (\hat{y}_{ij} - \bar{y}_i)^2 = 83,96$	$m - p = 5$	$MQ_{FaltaAjuste} = \frac{\sum_i^m \sum_j^{n_j} (\hat{y}_{ij} - \bar{y}_i)^2}{m - p} = 16,79$
Erro Puro	$\sum_i^m \sum_j^{n_j} (y_{ij} - \bar{y}_i)^2 = 22,79$	$n - m = 2$	$MQ_{ErroPuro} = \frac{\sum_i^m \sum_j^{n_j} (y_{ij} - \bar{y}_i)^2}{n - m} = 11,40$
Total	$\sum_i^m \sum_j^{n_j} (y_{ij} - \bar{y})^2 = 687,89$	$n - 1 = 10$	
% de variação explicada:		$R^2 = \frac{SQ_{Regressão}}{SQ_{Total}} = 84,22\%$	
% máxima de variação explicável		$\frac{SQ_{Total} - SQ_{ErroPuro}}{SQ_{Total}} = 96,69\%$	
$F_{Calculado}$		$F_{p-1, n-p} = \frac{MQ_{Regressão}}{MQ_{Resíduos}} = 12,66$	
$F_{Calculado}$ pela falta de ajuste		$F_{m-p, n-m} = \frac{MQ_{FaltaAjuste}}{MQ_{ErroPuro}} = 1,47$	

2.9 Gráficos de Probabilidade Normal

Na Estatística variável aleatória x é uma grandeza física que pode assumir qualquer valor dentro do conjunto universo ou populacional de um fenômeno em estudo, em que cada valor específico possui certa probabilidade de ocorrência, determinado por uma distribuição de probabilidade. De posse dessa distribuição é possível realizar previsões e estimativas de erros experimentais.

A distribuição de probabilidade é uma função que representa o comportamento de uma variável aleatória x , podendo ser classificada de acordo com a natureza da variável:

- **Discreta** - a probabilidade da função de distribuição de probabilidade é dada por somatório;

- **Contínua** - a probabilidade da função de distribuição de probabilidade é dada pela integral da função no intervalo considerado.

Exemplo 14. Seja o experimento - *jogar quatro moedas simultaneamente*, e o evento - *obter somente caras*. Com base nisso, afirma-se que a variável aleatória x (quantidade de caras obtidas) possuirá os seguintes valores: 0, 1, 2, 3 e 4 caras, como se apresenta na Tabela 2.32, construída para destacar os n possíveis valores da variável aleatória x com a respectiva probabilidade, onde se obteve a média $\mu = 2$ através de uma média ponderada usando como pesos as respectivas probabilidades ou frequências. Na tabela C significa cara e K significa coroa.

Tabela 2.32 – Distribuição de Probabilidade no lançamento de quatro moedas – Exemplo 14.

	x: quantidade de caras	f(x)	Possibilidades
	0	1	KKKK
	1	4	CKKK
	2	6	CCKK
	3	4	CCCK
	4	1	CCCC
		16	
Média: μ	2		
Desvio Padrão: σ	1		

Com base na Tabela 2.32 obteve-se um gráfico simétrico em torno da média $\mu = 2$ (probabilidade à esquerda e à direita da média são iguais em relação a uma certa quantidade de desvio padrão), Figura 2.9, em que a altura de cada coluna é a probabilidade de cada valor da variável aleatória x , sendo a soma das n probabilidades igual a 1 ou 100%, que se denota pelo somatório, visto que a variável é discreta:

$$\sum_{i=1}^n P_i(x) = f_1(x_1) + f_2(x_2) + \dots + f_n(x_n) = 100\% .$$

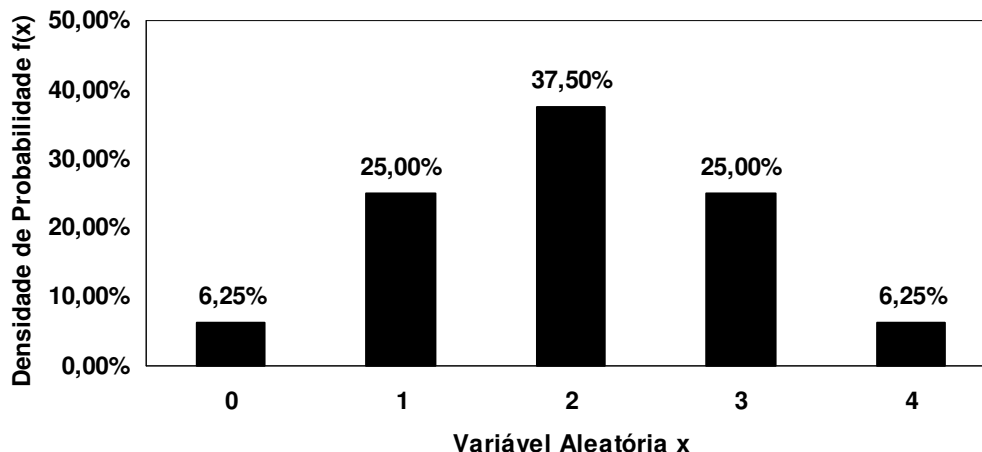


Figura 2.9 – Distribuição de Probabilidade da Variável Aleatória x do Exemplo 14

Apresentou-se no Exemplo 14 uma aplicação da importância de conhecer a distribuição de probabilidade de uma variável aleatória x . A partir de agora, apresenta-se um dos modelos de distribuição mais importantes, denominada de distribuição de Gauss ou distribuição Normal, proposta pelo famoso matemático Karl F. Gauss no início do século XIX para calcular as probabilidades de erros em medições, pois há vários conjuntos de dados estatísticos que podem ser bem representados pela distribuição Normal (BRUNS, ET AL, 2001; GUJARATI, 2006). É tão grande a quantidade de conjuntos de dados representados pelo modelo da distribuição Normal, que ela passou a ser conhecida como o comportamento natural de qualquer tipo de erro experimental, daí o adjetivo *Normal*.

Rotulava dizer se a distribuição dos erros não seguisse uma gaussiana, a culpa era jogada na coleta dos dados, porém, depois de algum tempo, ficou claro que existem muitas situações experimentais em que a distribuição Normal de fato não é válida, mas ela permanece sendo um dos modelos fundamentais da estatística (GUJARATI, 2006; BRUNS *et al*, 2001; SILVA, E., 1999; SPIEGEL, 2004).

A distribuição Normal é uma distribuição contínua da variável aleatória x , a qual pode assumir, pelo menos em principio, qualquer valor real no intervalo $(-\infty, +\infty)$, e define-se pela densidade de probabilidade $f(x)$, que é uma função matemática contendo dois parâmetros: a média populacional μ e variância populacional σ^2 , Eq. 47,

$$f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \quad (47)$$

Onde o produto $f(x)dx$ é, por definição, a probabilidade de ocorrência de um valor da variável aleatória x no intervalo de largura dx em torno do ponto x . Isso significa que, ao

extrair aleatoriamente da população um valor de x , as chances de que esse valor esteja no intervalo de largura infinitesimal que vai de x a $x + dx$ são dadas por $f(x)dx$ (BRUNS *et al*, 2001).

Na prática e no sentido físico consideram-se intervalos finitos, e para calcular as probabilidades correspondentes integra-se a densidade de probabilidade entre os limites apropriados, pois a variável aleatória é contínua, significando que as probabilidades passam a ser calculadas por integrais, e não mais por somatórios, como se apresentou no Exemplo 14. O uso de integral torna nula a probabilidade de se observar exatamente um valor em específico, já que isso corresponderia a fazer dx igual a zero, tornando a integral nula. Com base nisto, para uma distribuição contínua, portanto, não faz diferença se o intervalo considerado é aberto ou fechado, como apresentou no Exemplo 1, página 31, portanto a probabilidade de $a \leq x \leq b$ é igual a probabilidade de $a < x < b$:

$$P(a \leq x \leq b) = P(a < x < b) = \int_a^b f(x) dx = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \quad (48)$$

Considerando a variável aleatória x como contínua e aplicando a Eq. 47 no Exemplo 14, e fazendo o intervalo de x variar infinitesimal $x + dx$ construiu-se o gráfico da Figura 2.10, que é simétrico em torno da média $\mu = 2$. Uma boa noção de como obter o gráfico da figura 2.10 a partir do gráfico da Figura 2.9 é traçar linhas no ponto médio da largura das alturas dos retângulos da Figura 2.9, que é um gráfico simétrico.

A notação $x \approx N(\mu, \sigma^2)$ indica que a variável aleatória x se distribui normalmente, com média populacional μ e variância populacional σ^2 , em que o símbolo \approx pode ser interpretado como “distribui de acordo com”. Quando a média for zero ($\mu = 0$) e a variância igual a um ($\sigma^2 = 1$), diz que a distribuição é **Normal Padronizada** ou **Padrão**, que se denota por $x \approx N(0,1)$, e a Eq. 47 torna na Eq. 49,

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}. \quad (49)$$

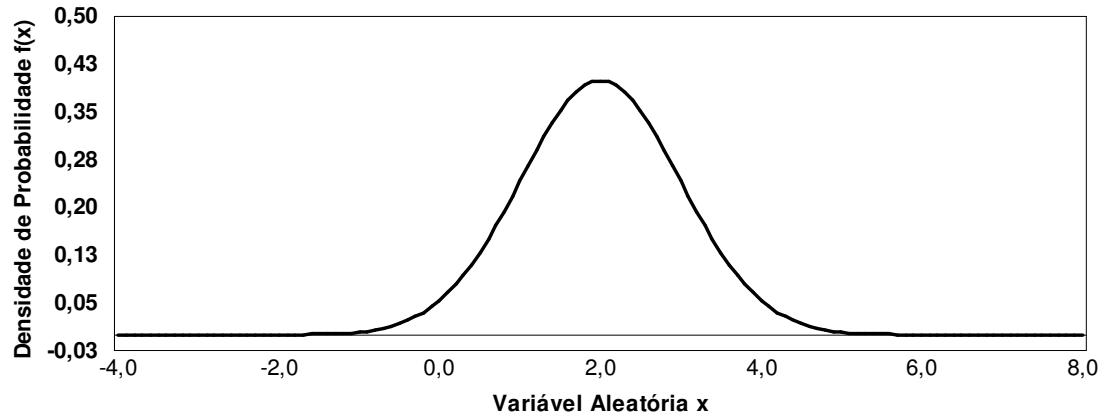


Figura 2.10 – Distribuição de Probabilidade Contínua do Exemplo 14

Observa-se na Eq. 48 que para obter a probabilidade de uma variável contínua com intervalos finitos, calculam-se tantas integrais quantas forem os intervalos. Porém, há uma tabela pronta, **Apêndice C**, que serve de consulta para vários intervalos de uma variável $z \approx N(0,1)$. Apesar da Tabela - **Apêndice C** corresponder à distribuição padrão, com média zero ($\mu=0$) e variância um ($\sigma^2=1 \Rightarrow \sigma=1$), esses valores podem ser usados para fazer inferências a respeito de qualquer distribuição normal. Para consultar a Tabela - **Apêndice C** necessita-se introduzir o conceito de padronização, que por definição, padronizar uma variável aleatória x de média μ e variância σ^2 é construir a partir dela uma nova variável aleatória z , cujos valores são obtidos subtraindo-se de cada valor de x a média populacional e dividindo-se o resultado pelo desvio padrão populacional, que se denota por:

$$z = \frac{x - \mu}{\sigma}. \quad (50)$$

Isolando x na Eq. 50 encontra-se: $x = \mu + z\sigma \Rightarrow dx = \sigma dz$. Substituindo nas Eq. 47 e 49 obtém-se a Eq. 51:

$$f(z) dz = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} \sigma dz. \quad (51)$$

Onde se escreveu $f(z)dz$ do lado esquerdo, porque a função agora depende de z . A Eq. 51 ficou idêntica à Eq. 49. A padronização simplesmente alterou a escala e deslocou a origem do eixo da variável aleatória, transformando a variável original x , que se distribuía de acordo com $x \approx N(\mu, \sigma^2)$, numa nova variável z que segue a distribuição padrão, $z \approx N(0,1)$. Como essa transformação não depende dos valores numéricos de μ e σ sempre se pode usar a

distribuição Normal Padrão para discutir o comportamento de uma distribuição Normal qualquer.

Construiu-se a Figura 2.11 empregando a Eq. 51 no Exemplo 14, onde a média populacional ($\mu=2$) foi convertida em zero ($\mu=0$) pela padronização e com desvio padrão um ($\sigma=1$).

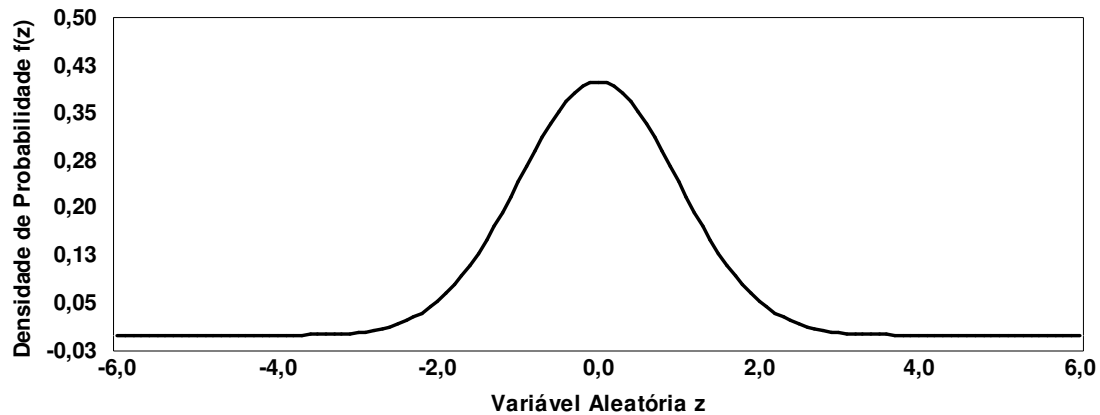


Figura 2.11 – Distribuição de Probabilidade Padrão do Exemplo 14 - $z \approx N(0,1)$

Percebe-se na Figura 2.11 que a curva da distribuição Normal é perfeitamente simétrica em torno do ponto central, que é a média $\mu = 0$. O valor da densidade é máximo sobre a média, e diminui rapidamente quando se afastam dela, em ambas os sentidos. A três desvios padrão de distância da média, a densidade de probabilidade praticamente reduz-se a zero. Com base na Figura 2.11, a maior parte da área sob uma gaussiana está contida no intervalo definido por um desvio padrão em torno da média $\mu \pm \sigma$, e praticamente toda ela está situada entre $\mu \pm 3\sigma$. Para obter os valores numéricos correspondentes a esses fatos, integra-se, entre os limites apropriados, a expressão de $f(x)$ dada pela Eq. 47:

$$P(\mu - \sigma < x < \mu + \sigma) = \int_{\mu - \sigma}^{\mu + \sigma} f(x) dx = 0,6826 = 68,26\%;$$

$$P(\mu - 3\sigma < x < \mu + 3\sigma) = \int_{\mu - 3\sigma}^{\mu + 3\sigma} f(x) dx = 0,9973 = 99,73\%.$$

A Figura 2.12 é um gráfico de probabilidade cumulativa, curva monotonicamente crescente em forma de S que vai (de forma assintótica) de zero à esquerda para 1 à direita, que pode-se transformar numa reta, por meio de uma modificação na escala do eixo das ordenadas, Figura 2.13. Essa modificação consiste em expandir a escala simetricamente em

torno do ponto representando 0,5 (50% de probabilidade acumulativa), de modo a esticar as duas pontas do S, Figura 2.13. De posse do gráfico da Figura 2.13 é possível prever quais efeitos são significativos num planejamento de experimentos fatorial sem a realização de replicatas.

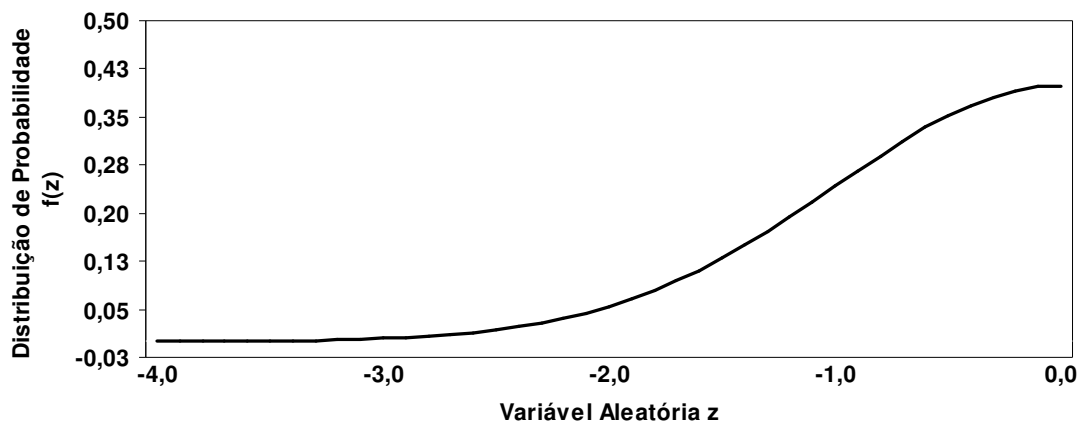


Figura 2.12 – Distribuição Normal de Probabilidade Padrão Cumulativa do Exemplo 14 – Curva Monotonicamente

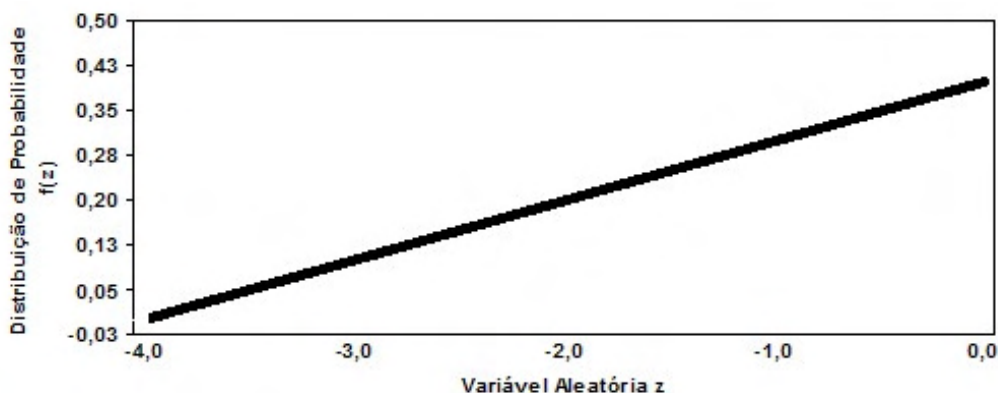


Figura 2.13 – Distribuição Normal de Probabilidade Padrão Cumulativa do Exemplo 14 - Reta

Para construir um gráfico igual da Figura 2.13 são realizados alguns procedimentos, os quais se destacam a seguir:

- **Primeiro passo** - consideram-se os efeitos obtidos do Exemplo 12 (Tabela 2.23) em ordem crescente, excluindo a média dos experimentos, inserindo cada efeito num intervalo de probabilidade da variável z padronizada, como se apresenta na Tabela 2.33, onde o primeiro elemento x_1 será o menor de todos os efeitos do Exemplo 12 e que está no intervalo de 0 a 0,067, o segundo elemento x_2 está no intervalo de 0,067 a 0,133, e assim por diante, de modo que o décimo quinto elemento x_{15} esteja no intervalo de 0,933 a 1,000;

A amplitude de cada intervalo da Tabela 2.33 obteve-se dividindo a probabilidade total 1 ou 100% pela quantidade de efeitos, neste caso 15, $1/15 = 0,067$, e o intervalo da primeira classe (0; 0,067) obteve-se considerando o limite inferior o valor zero e o limite superior a soma do limite inferior com a amplitude, e os demais intervalos considerou o limite inferior igual ao limite superior do intervalo anterior, e o limite superior igual a soma do limite inferior da classe atual com a amplitude, que se denota:

$$\begin{aligned} \text{Intervalo 1} &= \left(\underbrace{0}_{\text{Limite Inferior 1}} ; \underbrace{0 + \text{Amplitude}}_{\text{Limite Superior 1}} \right) \\ \text{Intervalo 2} &= \left(\underbrace{0 + \text{Amplitude}}_{\text{Limite Inferior 2}} ; \underbrace{\text{Limite Inferior 2} + \text{Amplitude}}_{\text{Limite Superior 2}} \right) \\ &\quad \vdots \\ \text{Intervalo 15} &= \left(\underbrace{0 + 14 \cdot \text{Amplitude}}_{\text{Limite Inferior 15}} ; \underbrace{\text{Limite Inferior 15} + \text{Amplitude}}_{\text{Limite Superior 15}} \right). \end{aligned}$$

Tabela 2.33 – Efeitos do Planejamento 2⁴ do Exemplo 12 para encontrar o gráfico de probabilidade normal

Símbolo do Normal Plot	Efeitos	Intervalo de Probabilidade Cumulativa		Ponto Central	Variável z	
1	-19,125	0,000	-	0,067	0,033	-1,834
2	-16,625	0,067	-	0,133	0,100	-1,282
3	-13,625	0,133	-	0,200	0,167	-0,967
4	-8,375	0,200	-	0,267	0,233	-0,728
5	-4,125	0,267	-	0,333	0,300	-0,524
6	-2,125	0,333	-	0,400	0,367	-0,341
7	-0,875	0,400	-	0,467	0,433	-0,168
8	-0,875	0,467	-	0,533	0,500	0,000
9	-0,375	0,533	-	0,600	0,567	0,168
10	-0,375	0,600	-	0,667	0,633	0,341
11	-0,125	0,667	-	0,733	0,700	0,524
12	0,375	0,733	-	0,800	0,767	0,728
13	2,125	0,800	-	0,867	0,833	0,967
14	8,625	0,867	-	0,933	0,900	1,282
15	15,375	0,933	-	1,000	0,967	1,834

O símbolo – no intervalo significa que não existem valores coincidentes com os valores extremos

- **Segundo Passo** – obter o ponto central ou médio de cada intervalo como se apresenta na Tabela 2.33;
- **Terceiro Passo** – obter através da Tabela - **Apêndice C** o valor da variável z que fornece a probabilidade dada pelo ponto central, Tabela 2.33;

- **Quarto Passo** – construir um gráfico usando os valores dos efeitos no eixo das abscissas com os respectivos valores da variável z no eixo das ordenadas, como se apresenta na Tabela 2.33 e na Figura 2.14.

No Exemplo 12, comentou-se que é possível obter a partir dos efeitos uma função estimada para o planejamento 2^4 e 2^{4-1} , as quais foram reproduzidas novamente a seguir para uma melhor acompanhamento dos resultados:

$$\hat{y} = 120,062 + \frac{0,375}{2}x_1 - \frac{0,875}{2}x_2 + \frac{8,625}{2}x_3 + \frac{15,375}{2}x_4 - \frac{13,625}{2}x_1x_2 - \frac{4,125}{2}x_1x_3 - \frac{0,875}{2}x_1x_4 - \frac{0,375}{2}x_2x_3 - \frac{2,125}{2}x_2x_4 - \frac{19,125}{2}x_3x_4 - \frac{0,125}{2}x_1x_2x_3 - \frac{8,375}{2}x_1x_2x_4 - \frac{0,375}{2}x_1x_3x_4 - \frac{16,625}{2}x_2x_3x_4 + \frac{2,125}{2}x_1x_2x_3x_4;$$

$$\hat{y} = 121,125 - \frac{16,25}{2}x_1 - \frac{1,25}{2}x_2 + \frac{0,25}{2}x_3 + \frac{15,25}{2}x_1x_2 - \frac{32,75}{2}x_1x_3 - \frac{6,25}{2}x_2x_3 - \frac{1,25}{2}x_1x_2x_3.$$

Também se comentou que geralmente os efeitos significativos e os coeficientes significativos são os maiores valores e baseados nisto inferiu-se no Exemplo 12 que as funções estimadas significativas para os planejamentos 2^4 e 2^{4-1} são, respectivamente:

$$\hat{y} = 120,062 + \frac{8,625}{2}x_3 + \frac{15,375}{2}x_4 - \frac{13,625}{2}x_1x_2 - \frac{19,125}{2}x_3x_4 - \frac{8,375}{2}x_1x_2x_4 - \frac{16,625}{2}x_2x_3x_4;$$

$$\hat{y} = 121,125 - \frac{16,25}{2}x_1 + \frac{15,25}{2}x_1x_2 - \frac{32,75}{2}x_1x_3.$$

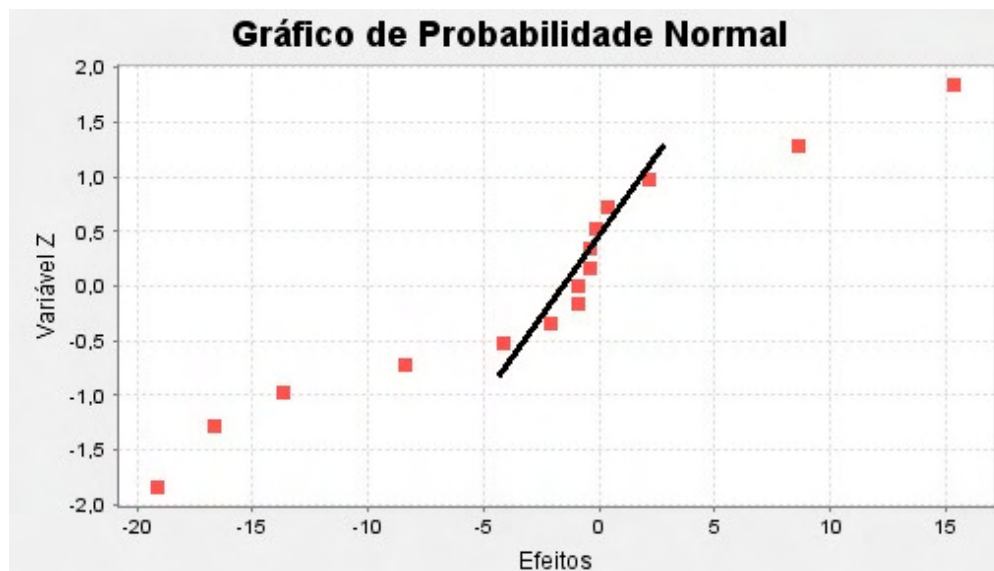


Figura 2.14 – Gráfico de Probabilidade Normal do Exemplo 12 do Planejamento 2^4 obtido com o *software* PlanEx

Agora, baseados no gráfico de distribuição de probabilidade normal é possível prever com mais segurança quais valores são significativos. Nas Figuras 2.14 e 2.15 apresentam-se o

gráfico de probabilidade normal para o Exemplo 12 do planejamento 2^4 e 2^{4-1} , respectivamente, pelos quais se percebem que os efeitos que estão distantes do ajuste da reta são significativos. Os efeitos centrais ajustam muito bem a uma reta que cruza a probabilidade acumulada em 50%, praticamente ao redor do ponto zero do eixo das abscissas. Faz sentido, portanto considerar que esses pontos sejam oriundos de uma população normal com média zero, ou seja, eles não apresentam efeitos físicos. O mesmo não se pode dizer dos efeitos que não se ajustam a curva (reta) e que dificilmente são da mesma distribuição normal dos outros efeitos, logo tais efeitos são ditos significativos, tanto mais quanto estiverem mais afastados da reta, tanto para esquerda ou para direita.

Com base no gráfico das Figuras 2.14 e 2.15 obtém-se às mesmas inferências que se realizaram na seção de planejamento de experimentos fatorial fracionário do Exemplo 12, ou seja, que as funções estimadas significativas são:

$$\hat{y} = 120,062 + \frac{8,625}{2}x_3 + \frac{15,375}{2}x_4 - \frac{13,625}{2}x_1x_2 - \frac{19,125}{2}x_3x_4 - \frac{8,375}{2}x_1x_2x_4 - \frac{16,625}{2}x_2x_3x_4;$$

$$\hat{y} = 121,125 - \frac{16,25}{2}x_1 + \frac{15,25}{2}x_1x_2 - \frac{32,75}{2}x_1x_3.$$

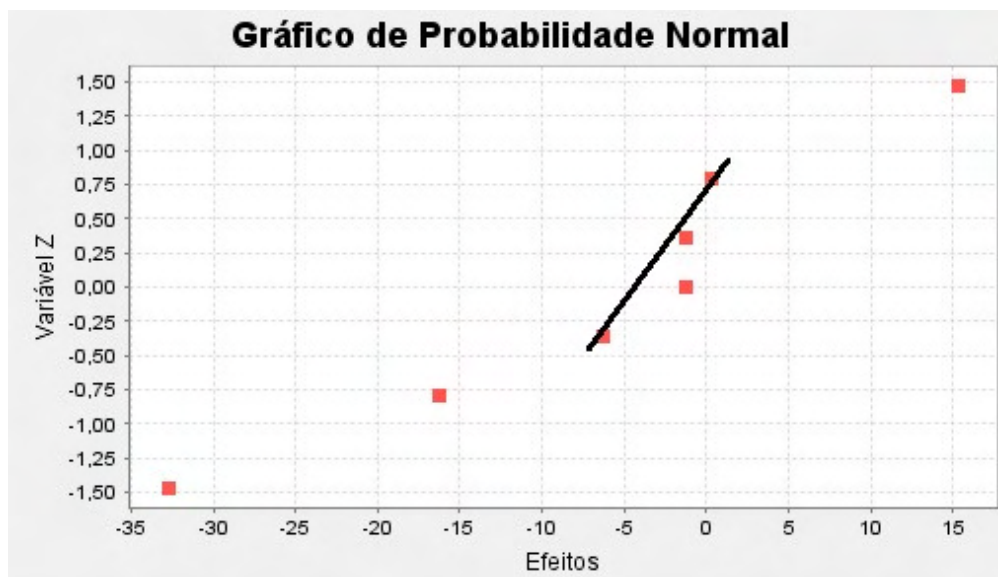


Figura 2.15 – Gráfico de Probabilidade Normal do Exemplo 12 do Planejamento 2^{4-1} obtido com o *software* PlanEx

Os gráficos das Figuras 2.14 e 2.15 foram geradas pelo *software* PlanEx como se comentará na seção de Resultados e Discussão.

3. CONCEITOS BÁSICOS DE LINGUAGEM COMPUTACIONAL

Na Introdução dessa Tese foi comentado que a linguagem de programação escolhida para implementar os algoritmos do planejamento de experimentos fatorial foi *Java*, pois oferece um grande campo de bibliotecas livre para implementação de cálculos estatísticos e a geração de gráficos. Baseado nisso, apresenta-se nesta parte do trabalho as ferramentas e conceitos de programação da linguagem *Java* utilizados ao longo da implementação dos algoritmos do planejamento de experimentos fatorial, tais como: Linguagem de Programação, *Application Programming Interface (API)*, *Integrated Development Environment (IDE)*, Servidor e Cliente.

Na área da informática define-se Servidor como uma máquina computacional que oferece serviços computacionais através de rede local ou pela *internet* a outras máquinas, denominadas de Clientes (DEITEL, H. e DEITEL, P., 2003). Com base nisso, percebe-se que Cliente define-se como uma máquina computacional usada para obter serviços computacionais num servidor e que não compartilha os seus recursos com outra máquina. Já linguagem de Programação constitui-se de um processo padrão de comunicação de instruções para uma máquina computacional. Esse processo compõe-se de regras sintáticas e semânticas utilizadas pelo programador para definir normas e informações traduzidas para o código de máquina por um compilador (DEITEL, H. e DEITEL, P., 2003; SILBERSCHATZ *et al*, 2004).

Java é uma linguagem de programação orientada a objetos desenvolvida na década de 90 por uma equipe de programadores da empresa *Sun Microsystems* incorporada à empresa *Oracle* no ano de 2010 (www.oracle.com/us/sun/index.htm). A diferença considerável da linguagem *Java* para as outras é essas são compiladas para um código nativo enquanto *Java* é compilada para um *bytecode*, sendo executado por sua máquina virtual *Java (JVM)*, um acrônimo a *Java Virtual Machine*.

JVM é uma plataforma que executa *softwares* desenvolvidos em *Java*, que atua convertendo os *bytecode* em códigos executáveis de máquina. Ou seja, interpreta o código *Java* pré-compilado para linguagem de máquina. Por tal peculiaridade os *softwares* desenvolvidos em *Java* funcionam em qualquer plataforma de *hardware* e *software* desde que tenha uma máquina virtual tornando assim independentes de plataforma (DEITEL, H. e DEITEL, P., 2003). O *bytecode* é um código *Java* compilado para um estado intermediário entre a linguagem de alto nível e a linguagem de baixo nível que é interpretado por uma JVM.

Isso faz com que a linguagem de programação *Java* seja considerada híbrida, ou seja, ela é compilada e depois é interpretada por uma JVM para que atinja o seu objetivo final.

Dentre as várias extensões da linguagem *Java*, as mais conhecidas e difundidas são:

- *Java SE (Java Standard Edition)* - é uma plataforma que contém todos os recursos utilizados para desenvolvimento e execução de sistemas *Java*, tais como *JVM*, compilador *Java* e *APIs* necessárias;
- *Java EE (Java Enterprise Edition)* - é utilizada para o desenvolvimento de aplicações em ambiente corporativos e possui todos os recursos do *Java SE* e fornece também os recursos para criação de aplicação distribuídas, baseada em multi-camadas e componentes que podem ser executados em servidores de aplicação, como os servidores *Web*;
- *Java ME (Java Micro Edition)* - é a extensão utilizada para desenvolvimento de sistemas embarcados, para dispositivos compactos, tais como celulares, PDAs, televisão, bem como qualquer dispositivo com pequeno poder de processamento.

O *Java Web Start* é um *software* desenvolvido pela *Oracle* que já vem incorporado ao *Java* desde a sua versão 5, que permite executar *softwares* desenvolvidos em *Java* para ambientes *desktops* que são disponibilizados na *Web*. Assim o *software Web Start* permite ativar aplicativos facilmente com apenas um clique e, oferece garantia de que estará executando sempre a última versão dos aplicativos, pois possui o mecanismo de verificação de versão, além de eliminar procedimentos complexos de instalação ou atualização (www.java.com/pt_BR/download/faq/java_webstart.xml).

Para o desenvolvimento do *software PlanEx* utilizou-se o *Java SE*, que contém todos os componentes padrões para desenvolvimento de aplicativos em *Java*. Na parte que se refere aos cálculos estatísticos usou-se API *Apache Commons Math* (<http://commons.apache.org>), que é uma API matemática para cálculos probabilísticos e estatísticos desenvolvida pela Fundação Apache. Na parte dos gráficos de resíduos aleatórios em função dos valores estimados e dos gráficos de probabilidade normal usou-se a API *JFreeChart*, versão 1.0.10, que é uma biblioteca gráfica totalmente em *Java*, sendo um *software* livre (<http://www.jfree.org/jfreechart/>). E para os gráficos de superfície de resposta 3D utilizou-se a API *jzy3d*, versão 0.8.4, que é uma biblioteca *Java* para gráficos científicos em 3D (www.jzy3d.org).

Para a implementação do algoritmo do planejamento de experimentos fatorial em *Java* utilizou a IDE *NetBeans*. A IDE *NetBeans* é livre e constitui-se de vários recursos para desenvolvimento em *Java* com destaque para criação de interface gráfica.

Utilizou-se o *software Java Web Start* para disponibilizar o *software PlanEx* na *internet*. O mecanismo do *Java Web Star* consiste em localizar a aplicação desenvolvida, PlanEx, e disponibilizar na *internet* através de um servidor de aplicação. Desse modo, o usuário pode ter acesso à aplicação atualizada de qualquer lugar, desde que tenha acesso à *internet*.

4. USANDO JAVA PARA RESOLVER O PROBLEMA

Nessa parte do trabalho serão apresentadas as equações estatísticas que foram implementadas em linguagem *Java*, abordando de forma resumida a implementação. Para uma melhor apresentação, dividiu-se essa seção em três categorias, como se comenta a seguir.

4.1 Planejamento de Experimentos Fatorial Completo

Os comentários dessa seção estão relacionados com os *layouts* semelhantes aos da Figura 5.2 a 5.6 na seção de Resultados e Discussão a partir da página 115.

Para o planejamento de experimentos fatorial completo iniciou-se definindo a quantidade de variáveis, $2 \leq k \leq 6$, na caixa de seleção do tipo *ComboBox*. Em seguida, programou-se numa caixa de seleção do tipo *CheckBox* se há ou não replicatas r , $1 \leq r \leq 4$. Se $r = 1$ não há replicatas e por consequência habilitam-se a aba de variáveis e uma aba de respostas experimentais. Caso contrário, se r estiver no intervalo $2 \leq r \leq 4$, há replicatas e habilita-se a caixa de seleção do tipo *ComboBox* com a quantidade correspondente a r , e por consequência a aba de variáveis e as abas r de respostas experimentais são habilitadas.

Habilitou-se a quantidade de linhas na tabela dentro da aba das variáveis como sendo a quantidade definida por k variáveis selecionadas no *ComboBox*. E a quantidade de linhas dentro de cada aba das respostas experimentais pela quantidade 2^k , sendo 2 a quantidade de níveis experimentais.

Para obter os efeitos do planejamento de experimentos fatorial completo 2^k programaram-se várias matrizes de contrastes que são selecionadas de acordo com a quantidade k de variáveis. Definiu-se, inicialmente, a dimensão da matriz de contraste, $\mathbf{X}_{n \times m}$, como n sendo as possíveis combinações 2^k das variáveis k em dois níveis (linha) e m sendo a quantidade 2^k de coeficientes do modelo linear **completo** (coluna). Desse modo, definiu-se 2^k como sendo a **ordem** da dimensão da matriz \mathbf{X} . Por exemplo, se $k = 3$ no *ComboBox* o sistema gera automaticamente a matriz \mathbf{X} quadrada de ordem 2^3 , como apresenta a seguir:

$$\mathbf{X} = \begin{bmatrix} +1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 \\ +1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \end{bmatrix}.$$

Em relação aos cálculos dos efeitos, programaram-se as Eq. 28.1 e 28.2 como se apresentou nas páginas 57 e 58, para realizar os cálculos necessários:

$$\mathbf{m} = \mathbf{X}^t \mathbf{y}; \quad (28.1)$$

$$\mathbf{ef} = \begin{bmatrix} \frac{m_{11}}{2^k} \\ \frac{m_{21}}{2^{k-1}} \\ \vdots \\ \frac{m_{\beta 1}}{2^{k-1}} \end{bmatrix}. \quad (28.2)$$

Os efeitos foram implementados numa tabela, de modo que eles possam ser incluídos e/ou excluídos na função estimada através de uma caixa de seleção do tipo *CheckBox*. A função estimada após realizar a seleção no *CheckBox* implementou-se usando a notação de matriz pela Eq. 18, como se apresentou na página 41:

$$\hat{\mathbf{y}} = \mathbf{X} \hat{\boldsymbol{\beta}}. \quad (18)$$

A matriz $\hat{\boldsymbol{\beta}}$ constituiu-se pelos coeficientes da regressão (metade dos efeitos, exceto a média) escolhidos na caixa de seleção do tipo *CheckBox*, de modo que os efeitos não escolhidos são considerados com valor nulo, zero.

Em relação à estimativa do erro, programou-se de modo que $r = 1$ não há replicatas e apresenta-se o valor do desvio padrão em branco. Caso contrário, $2 \leq r \leq 4$, calcula-se o desvio padrão. Iniciou-se a implementação do erro através do cálculo da variância de cada nível experimental i usando a API *Commons-math-1.2* do projeto *Apache Commons Math Bundle* (<http://commons.apache.org/>), Eq. 24. Em seguida, programaram-se as Eq. 25 a 27, respectivamente, para obter os erros da média, dos efeitos principais e de suas interações:

$$s_i^2 = \frac{\sum_{i=1}^r (y_i - \bar{y}_i)^2}{r-1}; \quad (24)$$

$$s_{\text{Conjunta}}^2 = \frac{\sum_{i=1}^n [(r_i - 1) \cdot s_i^2]}{\sum_{i=1}^n r_i - 1}; \quad (25)$$

$$s_{\text{Média}}^2 = \frac{s_{\text{Conjunta}}^2}{n \cdot r} \Rightarrow s_{\text{Média}} = \sqrt{s_{\text{Média}}^2}; \quad (26)$$

$$s_{\text{Efeitos}}^2 = \frac{s_{\text{Conjunta}}^2}{\frac{n}{2}} \Rightarrow s_{\text{Efeitos}} = \sqrt{s_{\text{Efeitos}}^2}. \quad (27)$$

Sendo n a quantidade de experimentos realizados em cada aba definidos por $n = 2^k$.

Em relação ao gráfico de resíduos em função dos valores estimados programaram-se no eixo da abscissa os valores estimados, $\hat{\mathbf{y}}$, definidos pela Eq. 18 e, no eixo das ordenadas os resíduos aleatórios, $\hat{\mathbf{u}}$, obtidos pela Eq. 12 escrita na forma de vetor (matriz coluna):

$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}. \quad (12)$$

Onde $\hat{\mathbf{y}}$ é a matriz coluna (vetor) obtida pela Eq. 18 e \mathbf{y} é a matriz coluna das médias das respostas experimentais realizadas nos diferentes níveis i , sendo $i = 1, 2, 3, \dots, n$, e digitadas nas r aba(s) da(s) respostas experimentais, lembrando que se $r = 1$ a média de cada nível é o próprio valor observado no nível i . Para gerar o gráfico de resíduos aleatórios em função dos valores estimados usou-se a API *jfreechart-1.0.10-swt.jar* (<http://www.jfree.org/jfreechart/>).

Para gerar o gráfico de probabilidade normal programaram-se no eixo da abscissa os valores dos efeitos (exceto a média) em ordem crescente e no eixo das ordenadas o valor da variável z relacionada com a respectiva probabilidade, e usou-se a mesma API para gerar os gráficos de resíduos aleatórios em função dos valores estimados.

4.2 Planejamento de Experimentos Fatorial Fracionário

Os comentários dessa seção estão relacionados com o fluxograma apresentado na Figura 5.7 na seção de Resultados e Discussão a partir da página 121.

Para o planejamento de experimentos fatorial fracionário iniciou-se definindo a quantidade de variáveis, $3 \leq k \leq 6$, na caixa de seleção do tipo *ComboBox*. Em seguida, programou-se numa caixa de seleção do tipo *CheckBox* se há ou não replicatas r , $1 \leq r \leq 4$. Se

$r = 1$ não há replicatas e por consequência habilitam-se a aba de variáveis e uma aba de respostas experimentais. Caso contrário, se r estiver no intervalo $2 \leq r \leq 4$, há replicatas e habilita-se o *ComboBox* com a quantidade correspondente a r , e por consequência a aba de variáveis e as abas r de respostas experimentais são habilitadas.

Habilitou-se a quantidade de linhas na tabela dentro da aba das variáveis como sendo a quantidade definida por k variáveis selecionadas no *ComboBox*. E a quantidade de linhas dentro de cada aba das respostas experimentais pela quantidade 2^{k-1} , sendo 2 a quantidade de níveis experimentais.

Para obter os efeitos do planejamento de experimentos fatorial fracionário 2^{k-1} programaram-se várias matrizes de contrastes que são selecionadas de acordo com a quantidade k de variáveis. Definiu-se, inicialmente, a dimensão da matriz de contraste, $\mathbf{X}_{p \times q}$, como p sendo as possíveis combinações 2^{k-1} das variáveis k em dois níveis (linha) e q sendo a quantidade 2^{k-1} de coeficientes do modelo linear **completo** (coluna). Desse modo, definiu-se 2^{k-1} como sendo a **ordem** da dimensão da matriz \mathbf{X} . Por exemplo, se $k = 4$ no *ComboBox* o sistema gera automaticamente a matriz \mathbf{X} quadrada de ordem 2^3 , como apresenta a seguir:

$$\mathbf{X} = \begin{array}{c} \bar{y} \quad x_1 \quad x_2 \quad x_3 \quad x_4 \quad A \quad B \quad C \\ \left[\begin{array}{ccccccc} 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array} \right] \end{array} .$$

$$x_4 = x_1x_2x_3 \quad A = x_1x_2 = x_3x_4 \quad B = x_1x_3 = x_2x_4 \quad C = x_1x_4 = x_2x_3$$

Para obter os efeitos, a função estimada e os gráficos do planejamento de experimentos fatorial fracionário 2^{k-1} o processo de implementação é similar ao descrito para o planejamento de experimentos fatorial completo 2^k .

4.3 Metodologia de Superfície de Resposta – RSM

Os comentários dessa seção estão relacionados com os *layouts* semelhantes aos da Figura 5.10 a 5.28 na seção de Resultados e Discussão a partir da página 125.

Para a RSM iniciou-se definindo o tipo de modelo (linear ou quadrático) na caixa de seleção do tipo *ComboBox*. Em seguida, definiu-se a quantidade de variáveis k , $2 \leq k \leq 4$, na caixa de seleção do tipo *ComboBox*. Após, programou-se numa caixa de seleção do tipo *ComboBox* a quantidade de replicatas r do ponto central, $1 \leq r \leq 5$. Se $r = 1$ não há replicatas e não se calcula o desvio padrão dos coeficientes da regressão e nem a tabela de ANOVA. Caso contrário, $2 \leq r \leq 5$, há replicatas e calcula-se a o desvio padrão dos coeficientes da regressão e a tabela de ANOVA. Após, habilita-se a aba de variáveis e a aba de respostas experimentais.

Habilitou-se a quantidade de linhas na tabela dentro da aba das variáveis como sendo a quantidade definida por k variáveis selecionadas no *ComboBox*. E a quantidade de linhas dentro da aba das respostas experimentais implementou-se em função do tipo de modelo escolhido no primeiro *ComboBox*:

- Modelo linear a quantidade de linhas é dada por $2^k + r$;
- Modelo quadrático a quantidade de linhas é dada por $2^k + 2k + r$.

Para obter os efeitos da RSM programaram-se várias matrizes de contrastes que são selecionadas de acordo com:

- o tipo de modelo;
- a quantidade k de variáveis;
- a quantidade r de replicatas do ponto central.

Se o modelo for linear, definiu-se, inicialmente, a dimensão da matriz de contraste, $\mathbf{X}_{t \times m}$, como t sendo a quantidade de linhas dadas por $2^k + r$ e m sendo a quantidade 2^k de coeficientes do modelo linear **completo** (coluna). Por exemplo, se $k = 3$ e $r = 3$ nos *ComboBox*, o sistema gera automaticamente a matriz \mathbf{X} de ordem 11×8 , como apresenta a seguir:

$$\mathbf{X} = \begin{bmatrix} +1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 \\ +1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Se o modelo for quadrático, definiu-se, inicialmente, a dimensão da matriz de contraste, $\mathbf{X}_{u \times v}$, como u sendo a quantidade de linhas dadas por $2^k + 2k + r$ e v sendo a quantidade $2^k + k$ de coeficientes do modelo quadrático **completo** (coluna). Por exemplo, se $k = 2$ e $r = 3$ nos *ComboBox*, o sistema gera automaticamente a matriz \mathbf{X} de ordem 11×6 , como apresenta a seguir:

$$\mathbf{X} = \begin{bmatrix} +1 & -1 & -1 & +1 & +1 & +1 \\ +1 & -1 & +1 & +1 & +1 & -1 \\ +1 & -1 & +1 & +1 & +1 & -1 \\ +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & 0 & 0 & 0 & 0 & 0 \\ +1 & 0 & 0 & 0 & 0 & 0 \\ +1 & 0 & 0 & 0 & 0 & 0 \\ +1 & -\alpha & 0 & \alpha^2 & 0 & 0 \\ +1 & \alpha & 0 & \alpha^2 & 0 & 0 \\ +1 & 0 & -\alpha & 0 & \alpha^2 & 0 \\ +1 & 0 & \alpha & 0 & \alpha^2 & 0 \end{bmatrix}.$$

Onde α é calculado pela Eq. 37, $\alpha = \sqrt[4]{2^k}$.

Em relação aos cálculos dos coeficientes do modelo da regressão programou-se a Eq. 19 para realizar os cálculos necessários:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}. \quad (19)$$

Os efeitos foram implementados numa tabela, lembrando que os coeficientes obtidos pela Eq. 19 são a metade dos efeitos, exceto a média. Programaram os efeitos numa tabela de modo que eles possam ser incluídos e/ou excluídos na função estimada através de uma caixa de

seleção do tipo *CheckBox*. A função estimada após a seleção do *CheckBox* é escrita em notação de matriz pela Eq. 18:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}. \quad (18)$$

A matriz $\hat{\boldsymbol{\beta}}$ constituiu-se pelos coeficientes da regressão (metade dos efeitos, exceto a média) escolhidos na caixa de seleção do tipo *CheckBox*, de modo que os efeitos não escolhidos são considerados com valor nulo, zero.

Em relação à estimativa do erro, programou-se de modo que $r = 1$ não há replicatas e apresenta-se o valor do desvio padrão dos coeficientes em branco e o botão da *Tabela de ANOVA* desativa-se. Caso contrário, $2 \leq r \leq 5$, calcula-se o desvio padrão e após usar a Eq. 18 o botão da *Tabela ANOVA* ativa-se. Iniciou-se a implementação do erro através do cálculo da variância nos níveis centrais das r replicatas do ponto central usando a API *Commons-math-1.2* do projeto *Apache Commons Math Bundle* (<http://commons.apache.org/>), Eq. 24.

$$s_i^2 = \frac{\sum_{i=1}^r (y_i - \bar{y}_i)^2}{(r-1)}. \quad (24)$$

Em seguida, programou-se a Eq. 34.1 para obter os erros da média, dos efeitos principais e de suas interações, que estão na diagonal principal da matriz:

$$s_{\text{Efeitos}}^2 = s_{\text{Nível_Central}}^2 \cdot (\mathbf{X}^t \mathbf{X})^{-1} \Rightarrow s_{\text{Efeitos}} = \sqrt{s_{\text{Nível_Central}}^2 \cdot (\mathbf{X}^t \mathbf{X})^{-1}}; \quad (34.1)$$

Em relação ao gráfico de resíduos em função dos valores estimados programaram-se no eixo da abscissa os valores estimados, $\hat{\mathbf{y}}$, pela Eq. 18 e, no eixo das ordenadas os resíduos aleatórios, $\hat{\mathbf{u}}$, obtidos pela Eq. 12 escrita na forma de vetor (matriz coluna):

$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}. \quad (12)$$

Onde $\hat{\mathbf{y}}$ é a matriz coluna (vetor) obtida pela Eq. 18 e \mathbf{y} é a matriz coluna das médias das respostas experimentais realizadas nos diferentes níveis i , sendo $i = 1, 2, 3, \dots, n$, e digitadas nas r aba(s) da(s) respostas experimentais, lembrando que se $r = 1$ a média de cada nível é o próprio valor observado no nível i . Para gerar o gráfico de resíduos aleatórios em função dos valores aleatórios usou-se a API *jfreechart-1.0.10-swt.jar* (<http://www.jfree.org/jfreechart/>).

Em relação ao gráfico de superfície de resposta 3D programou-se de forma que o usuário deve escolher no máximo duas variáveis e/ou sua interação na caixa de seleção do tipo *CheckBox* dos efeitos para gerar o gráfico. No eixo z atribuiu-se aos valores da matriz da função estimada, $\hat{\mathbf{y}}$, obtida pela Eq. 18, e nos demais eixos programaram as variáveis (no

máximo duas) escolhidas no *CheckBox* dos Efeitos. Para gerar o gráfico de superfície de resposta 3D usou a API *org.jzy3d-0.8.4.jar* do projeto *Uma Biblioteca Java de Gráficos Científicos 3D*.

Para gerar a Tabela de ANOVA programaram as equações apresentadas na Tabela 2.28 e comentadas na seção 2.8 - Teste de ANOVA para Avaliação do Modelo.

5. RESULTADOS E DISCUSSÃO

Apresentam-se nessa seção problemas fictícios e de trabalhos obtidos na literatura de planejamentos de experimentos fatorial usando o *software* PlanEx como ferramenta de pesquisa em nível de pós-graduação. No fim da seção comenta-se a validação do *software* PlanEx como ferramenta de ensino no nível de pós-graduação, realizada em três aplicações na disciplina Quimiometria II, Planejamento e Otimização de Experimentos, ministrada pelo Professor Dr. Anselmo Elcana de Oliveira no programa de pós-graduação de mestrado em Química da UFG e Doutorado Multiinstitucional em Química UFU/UFG/UFMT nos segundos semestres dos anos de 2009, 2010 e 2011.

Para uma melhor apresentação e discussão dos resultados, dividiu-se a discussão em quatro categorias:

- Resultados do Planejamento de Experimentos Fatorial Completo;
- Resultados do Planejamento de Experimentos Fatorial Fracionário;
- Resultados da Metodologia de Superfície de Resposta (RSM);
- Validação do *software* PlanEx como ferramenta de pesquisa e ensino.

Para o planejamento de experimentos fatorial completo apresenta-se a aplicação do *software* PlanEx nos dados do Exemplo 11 da seção da Estatística Experimental para obter: os cálculos dos efeitos, a função estimada com os resíduos, o gráfico de resíduos em função dos valores estimados, e o gráfico de probabilidade normal. Já para o planejamento de experimentos fatorial fracionário, aplicou-se o *software* PlanEx nos dados do trabalho de Calazans (2007), Ilustração 1.

Para a RSM aplicou-se o *software* PlanEx nos trabalhos de Joelma Ferreira *et al* (2005), Mansilla *et al* (2002) e Muniz *et al* (2008), respectivamente, para obter: os cálculos dos efeitos, a função estimada com os resíduos, o gráfico de resíduos em função dos valores estimados, o gráfico de superfície de resposta e a Tabela de ANOVA.

Para a validação do *software* PlanEx realizaram-se três aplicações em turmas diferentes na Universidade Federal de Goiás com o objetivo dos alunos de pós-graduação usarem o *software* PlanEx para conhecer as ferramentas e opinar sobre as características observadas.

5.1. Resultado do Planejamento de Experimentos Fatorial Completo 2^k

Na Figura 5.1 apresenta-se o fluxograma do *software* PlanEx para as ferramentas de um planejamento de experimentos fatorial completo, 2^k , sendo k a quantidade de variáveis definida no intervalo $2 \leq k \leq 6$, visto que a maior parte dos estudos realizados está compreendida neste intervalo.

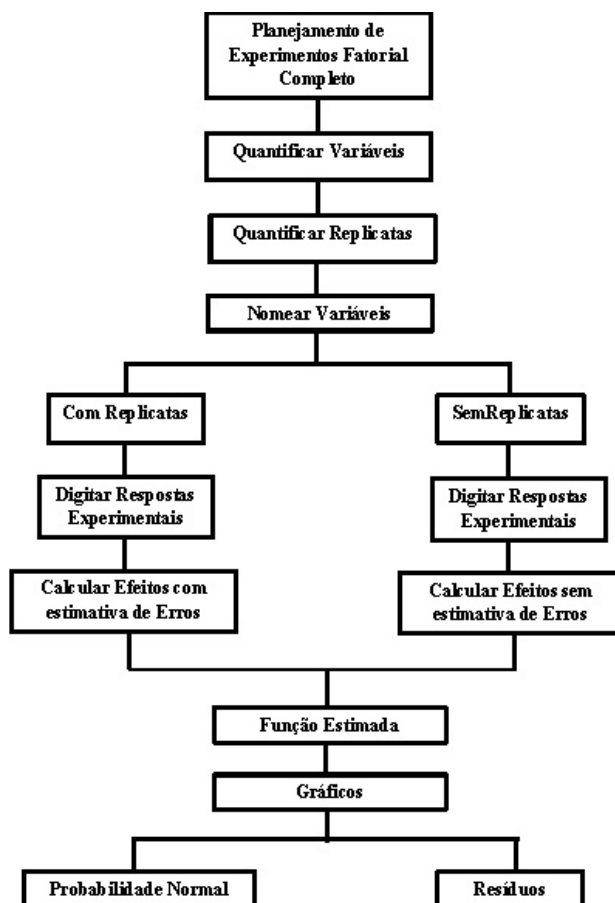


Figura 5.1– Fluxograma do Planejamento de Experimentos Fatorial Completo, 2^k

Percebe-se na Figura 5.1 que após escolher o planejamento de experimentos fatorial completo, o usuário para obter os efeitos e os gráficos do planejamento precisa realizar algumas etapas na seguinte ordem:

- quantificar a quantidade de variáveis k , $2 \leq k \leq 6$;
- quantificar a quantidade de replicatas r , $1 \leq r \leq 4$, se $r = 1$ não há replicatas;
- nomear as variáveis;
- digitar as respostas experimentais de acordo com o símbolos dos experimentos;

- calcular os efeitos através do botão *Calcular*;
- analisar quais efeitos são significativos para gerar a função estimada com os resíduos;
- gerar gráficos de probabilidade normal e resíduos em função dos valores estimados.

A Figura 5.2 refere-se ao *layout* da tela inicial do *software* PlanEx para o planejamento de experimentos fatorial completo 2^3 com os dados do Exemplo 11 da seção da Estatística Experimental, página 58. Percebe-se que a quantidade de variáveis selecionadas foi três e a quantidade de replicatas de todos os experimentos realizados foi duas (duplicatas), e os nomes das variáveis x_1 , x_2 e x_3 são respectivamente, temperatura, tipo de catalisador e concentração. O *software* foi implementado para o planejamento de experimentos fatorial 2^k , desse modo fica subtendido que todas as variáveis são estudadas em dois níveis, um menor (-1) e outro maior (+1), não havendo a necessidade de especificar quais são os valores desses níveis e nem de realizar codificações das variáveis usando a Eq. 20.

As duas *abas* (respostas 1 e respostas 2) na Figura 5.2 referem-se as respostas experimentais que o usuário realizou e deve digitar de acordo a nomenclatura dos símbolos, como apresenta-se na Figura 5.3.

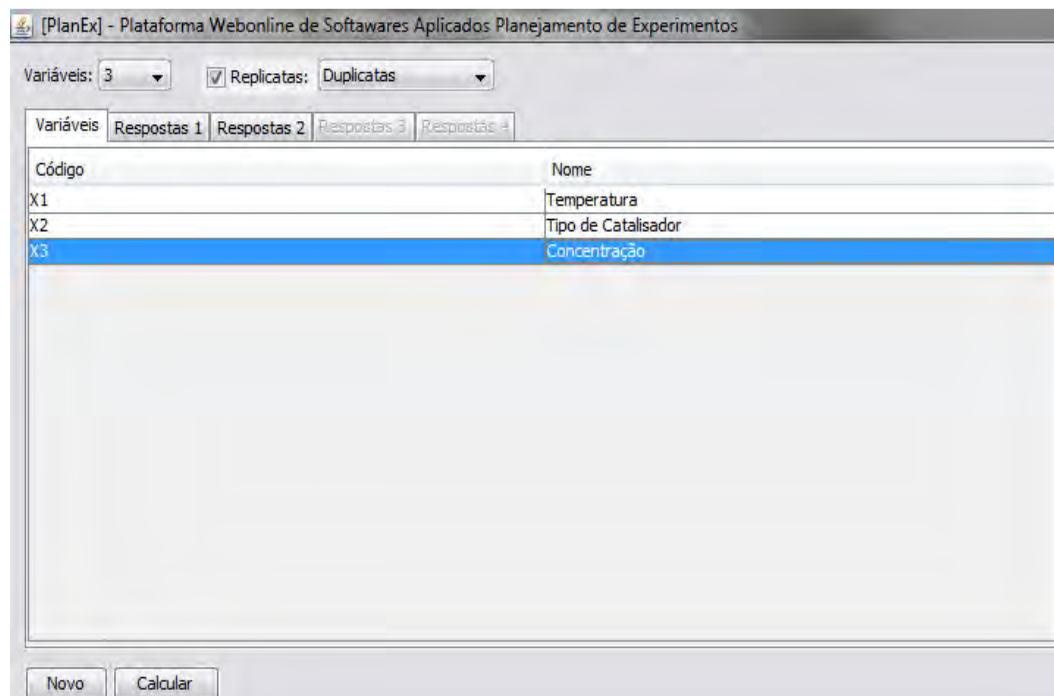


Figura 5.2 – *Layout* da Tela Inicial do *software* PlanEx para o Planejamento de Experimentos Fatorial Completo 2^3 usando os dados do Exemplo 11 da seção do Estatística Experimental

Apresentam-se na Figura 5.3 os símbolos experimentais (teoria) comentados na Tabela 2.13 relacionados com as respostas experimentais (prática). Na Figura 5.3 aparecem-se nas

abas de respostas experimentais, *Respostas 1* e *Respostas 2*, os experimentos do Exemplo 11 da seção da Estatística Experimental, página 58, que foram digitadas pelo usuário: 42 (44), 66 (80), 64 (68), 88 (84), 40 (44), 66 (70), 56 (60), 60 (56), em que o valor entre parênteses são as respostas experimentais em duplicatas, *Respostas 2*. O usuário ao digitar as respostas deve-se lembrar dos símbolos dos experimentos (teoria) apresentados na Tabela 2.13, tais como: o símbolo (1) significa o(s) experimento(s) realizado(s) no nível menor de todas as variáveis, x_1 significa o(s) experimento(s) realizado(s) no nível maior da variável x_1 e menor das demais, x_2 significa o(s) experimento(s) realizado(s) no nível maior da variável x_2 e menor das demais, x_1x_2 significa o(s) experimento(s) realizado(s) no nível maior das duas variáveis, e assim por diante. Após realizar os passos da Figura 5.2 e Figura 5.3, o usuário para obter os efeitos deve interagir com o *software* PlanEx através de um click no botão *Calcular*.

The screenshot shows the PlanEx software interface. At the top, it says "[PlanEx] - Plataforma Webonline de Softwares Aplicados Planejamento de Experimentos". Below that, there are controls for "Variáveis: 3" and "Replicatas: Duplicatas". There are tabs for "Variáveis", "Respostas 1", "Respostas 2", "Respostas 3", and "Respostas 4". The "Respostas 2" tab is active, showing a table with the following data:

Experimentos	Valor
(1)	42
x_1	66
x_2	64
x_1x_2	88
x_3	40
x_1x_3	66
x_2x_3	56
$x_1x_2x_3$	60

At the bottom of the interface, there are two buttons: "Novo" and "Calcular".

Figura 5.3 - Layout da Tela Inicial do *software* PlanEx para o Planejamento de Experimentos Fatorial Completo 2^3 usando os dados do Exemplo 11 da seção do Estatística Experimental

A Figura 5.4 apresenta os efeitos com os respectivos erros (desvio padrão). Através do *software* Matlab v R2011b (7.13.0.564) obtém-se os mesmos efeitos da Figura 5.4, porém o *software* requer o conhecimento de uma linguagem computacional específica de representação e operações de matrizes e operações numéricas, o que pode tornar o trabalho repetitivo e cansativo, pois realiza-se todos os procedimentos comentados no Exemplo 11 a

partir da página 58. Além disso, o *Matlab* v R2011b (7.13.0.564) pode ser inconveniente também em relação a tempo e custo. Já no *software* PlanEx, Figura 5.4, os efeitos foram obtidos sem o usuário digitar a matriz de contraste **X**, e também sem a necessidade de conhecer comandos específicos de operações de matrizes e de operações numéricas de linguagem computacional. Isso, facilita os cálculos e evita que o usuário realize procedimentos repetitivos que são desnecessários ao uso do *software* PlanEx como ferramenta de pesquisa.

Tela de Resultados dos Efeitos das Variáveis

Opções de Cálculo

Símbolos dos Efeitos	Efeitos	Desvio Padrão dos Efeitos	Opção de Cálculo
Média	61,75	1,075	<input checked="" type="checkbox"/>
X1	19	2,151	<input checked="" type="checkbox"/>
X2	10,5	2,151	<input checked="" type="checkbox"/>
X3	-10,5	2,151	<input checked="" type="checkbox"/>
X1X2	-9	2,151	<input checked="" type="checkbox"/>
X1X3	-6	2,151	<input checked="" type="checkbox"/>
X2X3	-7,5	2,151	<input checked="" type="checkbox"/>
X1X2X3	-4	2,151	<input checked="" type="checkbox"/>

$\hat{y} = (61,750) + (9,500X1) + (5,250X2) + (-5,250X3) + (-4,500X1X2) + (-3,000X1X3) + (-3,750X2X3) + (-2,000X1X2X3)$

Respostas Experimentais	Valores Estimados	Resíduos
43	43	0
73	73	0
66	66	0
86	86	0
42	42	0
68	68	0
58	58	0
58	58	0

Calculador Opções de Gráficos Seleccione o Gráfico Gráficos

Figura 5.4 – Layout da Tela de Resultados dos Efeitos do *software* PlanEx para o Planejamento de Experimentos Fatorial 2^3 do Exemplo 11 da seção do Estatística Experimental

Percebe-se na Figura 5.4 que na coluna 1 estão os símbolos dos efeitos, em que X_1 (maiúsculo) é o efeito da variável x_1 (minúsculo) (temperatura) sobre os experimentos, ou seja, para o(s) experimento(s) variar do nível menor para o nível maior mantendo as outras variáveis fixas, a temperatura (variável x_1) em média deve aumentar 19°C , e assim por diante. E na coluna 2 estão a média experimental e os respectivos efeitos das variáveis e suas interações, como apresentou-se na Tabela 2.15. Pelos dados da coluna 3 o usuário pode distinguir quais efeitos são significativos, se os experimentos forem realizados com replicatas. Na seção da Estatística Experimental comentou-se que os efeitos significativos do planejamento de experimentos fatorial são aqueles que o valor é maior que o respectivo erro (desvio padrão) quando houver replicatas. Com base nisto, o usuário concluirá que todos os

efeitos das variáveis são significativos, com isso pode gerar a função estimada e os resíduos escolhendo no botão *Opção de Cálculo* os efeitos que são convenientes e clicando no botão *Calcular* no *layout* da tela de resultados.

Na Figura 5.4, através da escolha da função estimada o usuário pode gerar valores estimados e os resíduos aleatórios no botão *Calcular* sem a necessidade de realizar operações com matrizes e numéricas, ou seja, o usuário obtém os resultados desejados de forma rápida, prática e eficiente. Após obter os resíduos o usuário pode gerar gráficos. Primeiramente escolhe-se o tipo de gráfico, sendo gráfico de probabilidade normal e gráfico de resíduos em função dos valores estimados. Seguindo clica-se no botão *Gráficos* para gerar os gráficos, Figuras 5.5 e 5.6.

O *software* de planejamento de experimentos fatorial pode gerar gráficos de resíduos em função dos valores estimados para analisar se o modelo da regressão apresenta indícios tendenciosos, ou seja, para verificar se o modelo é bom. Percebe-se que a Figura 5.5 gerada pelo *software* de planejamento de experimentos fatorial da plataforma PlanEx concorda perfeitamente com a Figura 2.6 da seção da Estatística Experimental gerada pelo *Excel*.



Figura 5.5 – Gráfico de Resíduos em função dos Valores Estimados do Exemplo 11 da seção da Estatística Experimental obtido com *software* PlanEx

Na Figura 5.6 apresenta-se o gráfico de probabilidade normal do *software* PlanEx útil para analisar quais efeitos são significativos, quando não há a presença de replicatas de experimentos ou mesmo para corroborar com a análise. Percebe-se no Exemplo 11, Figura 5.4, que todos os efeitos são significativos através dos erros obtidos dos efeitos pelas

replicatas, o que se confirma no gráfico de probabilidade normal, Figura 5.6, pois os efeitos significativos são aqueles que estão distantes da reta no valor zero no eixo das abscissas, ou seja, todos.

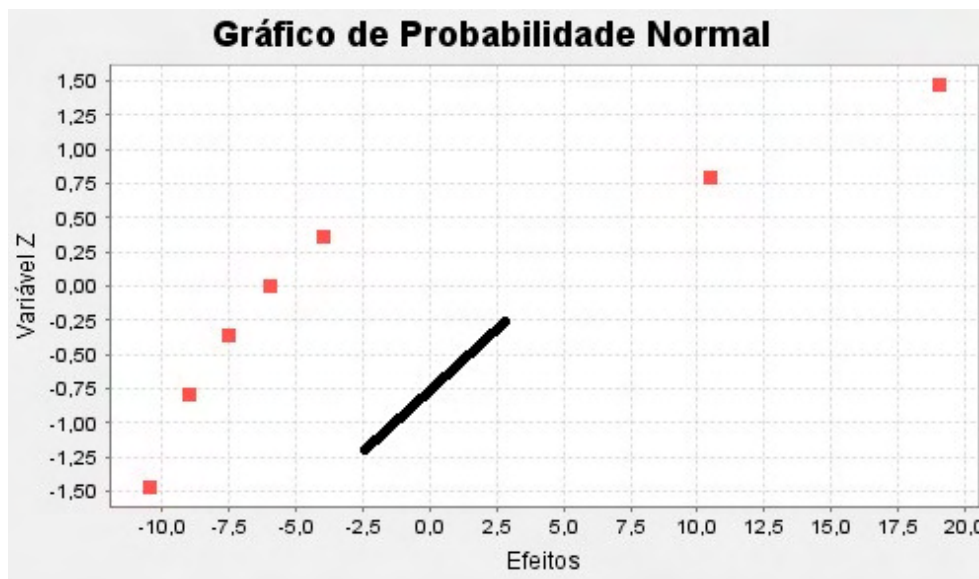


Figura 5.6 – Gráfico de Probabilidade Normal do Exemplo 11 da seção da Estatística Experimental obtido com *software* PlanEx

5.2. Resultado do Planejamento de Experimentos Fatorial Fracionário 2^{k-1}

Na Figura 5.7 apresenta-se o fluxograma do *software* de planejamento de experimentos fatorial de meia fração, 2^{k-1} , sendo k a quantidade de variáveis definida no intervalo $3 \leq k \leq 6$, visto que a maior parte dos estudos realizados está compreendida neste intervalo.

Percebe-se na Figura 5.7 que após escolher o planejamento de experimentos fatorial fracionário, o usuário para obter os efeitos e os gráficos do planejamento precisa realizar algumas etapas na seguinte ordem:

- quantificar a quantidade de variáveis k , $3 \leq k \leq 6$;
- quantificar a quantidade de replicatas r , $1 \leq r \leq 4$, se $r = 1$ não há replicatas;
- nomear as variáveis;
- digitar as respostas experimentais de acordo com o símbolos dos experimentos;
- calcular os efeitos através do botão *Calcular*;
- analisar quais efeitos são significativos para gerar a função estimada com os resíduos;
- gerar gráficos de probabilidade normal e resíduos versus valores estimados.

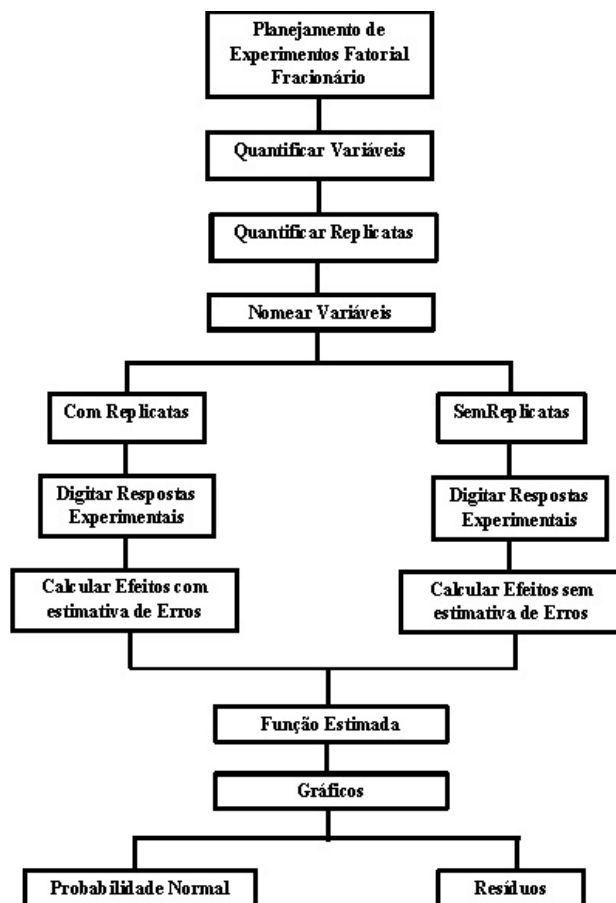


Figura 5.7 – Fluxograma do Planejamento de Experimentos Fatorial Fracionário, 2^{k-1}

O *layout* da Tela Inicial do *software* PlanEx para o planejamento de experimentos fatorial fracionário parece-se ao do completo, Figura 5.2, exceto a quantidade de experimentos e os símbolos das respostas experimentais, como se apresentou no Exemplo 12 na seção da Estatística Experimental, a partir da página 73. Percebem-se na Figura 5.7 e Figura 5.1 que as ferramentas do planejamento fracionário comparadas com as do completo são as mesmas, podendo chegar às vezes às mesmas conclusões quando efeitos de altas ordens não são significativos.

Ilustração 1. No trabalho de Calazans *et al* (2007) usou-se planejamento de experimentos fatorial fracionário 2^{4-1} com ponto central na produção de *Zymomonas mobilis* para quantificar a concentração de etanol, levana e biomassa na fermentação inicial e final, por exemplo. As variáveis independentes consideradas foram concentração inicial em g/L de sacarose x_1 (150, 200, 250), Temperatura em °C x_2 (20, 25, 30), concentração inicial de

extrato de fermento em g/L x_3 (2,0; 3,5, 5,0) e agitação na produção em rpm x_4 (50, 75, 100). Ao todo foram realizados 27 experimentos, sendo 24 relacionados às triplicatas do planejamento de experimentos fatorial fracionário 2^{4-1} para investigar se há erro experimental, e 3 experimentos relacionados a triplicatas do ponto central para investigar se há falta de ajuste do modelo linear, como se comentou na seção de RSM da Estatística Experimental. No artigo disponibilizaram-se tabelas contendo apenas as médias dos experimentos (média da concentração de etanol, por exemplo) e os efeitos obtidos. Os cálculos realizaram-se através dos *softwares Matlab* versão 4.2 c.1 e *Statistica* versão 5.1.

Após digitar as respostas experimentais da média da concentração de Levana no *layout* do *software PlanEx* da Tela Inicial do planejamento de experimentos fatorial fracionário o usuário para obter os efeitos da Ilustração 1 necessita dar o comando no botão *Calcular* para gerar os cálculos da regressão e obtenção dos efeitos, como apresenta-se na Figura 5.8.

Na coluna 1 da Figura 5.8 apresenta-se os símbolos dos efeitos obtidos para o planejamento de experimentos fatorial fracionário. Os símbolos X_1 e X_2 , por exemplo, significam o efeito obtido para o planejamento de experimentos fatorial fracionário da variável 1 e da variável 2, respectivamente, que se denotaram na seção de planejamento fatorial fracionário da Estatística Experimental por X_{F_1} e X_{F_2} , respectivamente, e o símbolo X_1X_2 significa o efeito de interação obtido para o planejamento de experimentos fatorial fracionário das variável 1 e 2, que se denotou na seção de planejamento fatorial fracionário da Estatística Experimental por $X_{F_1}X_{F_2}$, e assim por diante.

Símbolos dos Efeitos	Efeitos	Desvio Padrão dos Efeitos	Opção de Cálculo
Média	5,641		<input type="checkbox"/>
X1	3,127		<input type="checkbox"/>
X2	-6,177		<input type="checkbox"/>
X3	-0,658		<input type="checkbox"/>
X1X2	3,208		<input type="checkbox"/>
X1X3	-1,232		<input type="checkbox"/>
X2X3	-2,402		<input type="checkbox"/>
X1X2X3	1,252		<input type="checkbox"/>

Figura 5.8 – *Layout* da Tela de Resultados dos Efeitos do *software PlanEx* para o Planejamento de Experimentos Fatorial 2^{4-1} da Ilustração 1

Na coluna 2 da Figura 5.8, desvio padrão dos efeitos, apresenta-se sem descrição de valores, pois os autores do trabalho da Aplicação 1 não disponibilizaram as respostas

experimentais individuais (somente as médias para cada nível experimental), o que resulta na impossibilidade de obtenção dos erros. No entanto, através das médias experimentais no *software* PlanEx obteve-se praticamente aos mesmos efeitos obtidos do trabalho usando *software* Matlab v R2011b (7.13.0.564), sem considerar o nível zero nos cálculos.

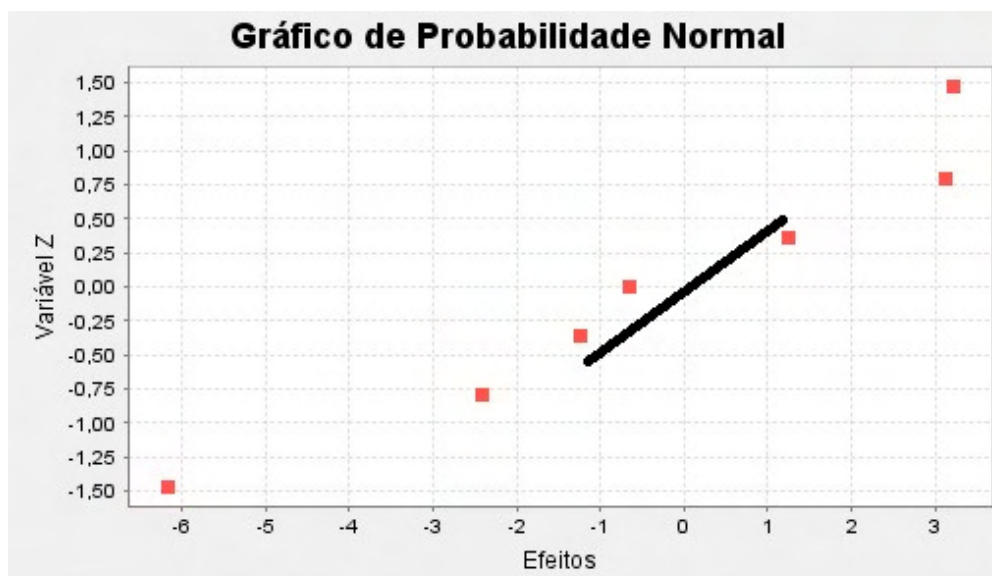


Figura 5.9 – Gráfico de Probabilidade Normal da Aplicação 1 (CALAZANS *et al*, 2007) obtido com *software* PlanEx

Na Figura 5.9, gráfico de probabilidade normal obtido pelo *software* PlanEx, percebe-se que os efeitos significativos são aqueles distantes da reta no valor zero no eixo das abscissas. Com base nisto e nos efeitos obtidos, Figura 5.8, ambos gerados pelo *software* PlanEx chegam-se nas mesmas conclusões dos autores do trabalho da Ilustração 1 usando os *softwares* Matlab versão 4.2 c.1 e Statistica versão 5.1. Portanto, infere-se que as ferramentas do PlanEx são eficientes para estudar um problema e que fornece ao pesquisador os mesmos resultados daqueles obtidos com um *software* comercial, por exemplo, o *Matlab*.

5.3. Metodologia de Superfície de Resposta - RSM

Na Figura 5.10 apresenta-se o fluxograma do *software* PlanEx da RSM, sendo a RSM dividida em dois modelos: linear e quadrático.

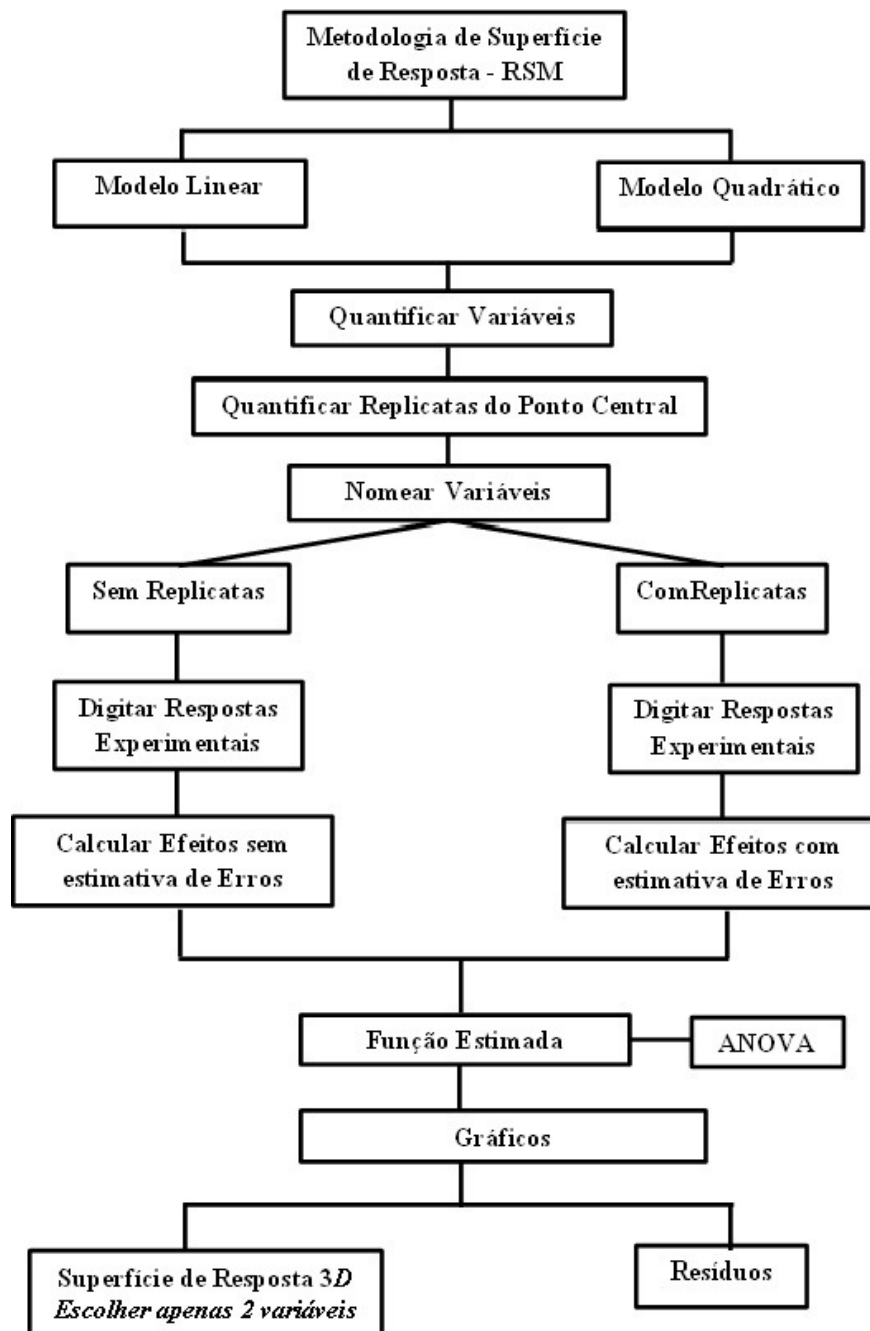


Figura 5.10 – Fluxograma do RSM , modelo $2^k + r$ ou $2^k + 2k + r$

Os modelos da RSM possuem as características:

- **Modelo Linear** – quantidade de experimentos $2^k + r$, sendo k a quantidade de variáveis definida no intervalo $2 \leq k \leq 4$ e r a quantidade de replicas do ponto central ou nível zero definida no intervalo $1 \leq r \leq 5$ e se $r = 1$ não há replicas e não realiza ANOVA;

- **Modelo Quadrático** – quantidade de experimentos $2^k + 2k + r$, sendo k a quantidade de variáveis definida no intervalo $2 \leq k \leq 4$ e r a quantidade de replicatas do ponto central ou nível zero definida no intervalo $1 \leq r \leq 5$ e se $r = 1$ não há replicatas e não realiza ANOVA, sendo o termo 2^k referente ao planejamento fatorial e $2k$ ao planejamento estrela.

Percebe-se na Figura 5.10 que após escolher a RSM, o usuário para obter os efeitos, os gráficos e a tabela de ANOVA da RSM precisa realizar algumas etapas na seguinte ordem:

- escolher o tipo de modelo, linear ou quadrático;
- quantificar a quantidade de variáveis k , $2 \leq k \leq 4$;
- quantificar a quantidade de replicatas r do ponto central, $1 \leq r \leq 5$, se $r = 1$ não há replicatas;
- nomear as variáveis;
- digitar as respostas experimentais de acordo com o símbolos dos experimentos;
- calcular os efeitos através do botão *Calcular*;
- analisar quais efeitos são significativos para gerar a função estimada com os resíduos;
- gerar a tabela de ANOVA, caso haja replicatas;
- escolher o efeito principal de uma variável ou de duas e/ou o efeito de interação entre elas para gerar gráfico de superfície de resposta em 3D, caso contrário o botão de *Gráficos* não será ativado;
- gerar gráficos de resíduos em função dos valores estimados.

Ilustração 2. No trabalho de Joelma Ferreira *et al* (2005) investigou-se o efeito das fontes de nutrientes sobre a produção de vinagre em escala de bancada, usando-se a RSM. Iniciou-se com o planejamento de experimentos fatorial completo 2^3 para investigar quais efeitos são significativos das variáveis concentrações iniciais de etanol, de fontes de nitrogênio (N) e de fósforo (P). Após esta análise, realizou-se uma ampliação das faixas das concentrações iniciais das variáveis de fontes de nutrientes (N e P) para otimizar os experimentos através da RSM. Realizaram-se os cálculos e gráficos de quimiometria através do *software* Statistica 5.0.

Com base no *software Matlab* v R2011b (7.13.0.564) e na Ilustração 2, seria necessário conhecer a linguagem computacional específica do *software* sobre operações e representações de matrizes e, operações numéricas, para obter os efeitos, e comandos específicos para gerar os gráficos. Além de requerer um tempo maior na digitação de matriz e

de dificuldade oriundas dos conhecimentos necessários a execução do *software*, necessita-se de uma licença comercial para o manuseio do *software*, que geralmente não é acessível à maioria dos usuários devido a dimensão financeira.

No *layout* da Tela Inicial da RSM do *software* PlanEx, primeiramente o usuário escolhe o modelo da RSM (linear ou quadrático), em seguida quantifica a quantidade de variáveis e replicatas do ponto central. Após, digita os símbolos das variáveis que são necessários para gerar a rotulação do gráfico 3D da superfície de resposta. Por fim, o usuário deve digitar as respostas experimentais observando os símbolos dos experimentos como se comentou na seção da RSM da Estatística Experimental.

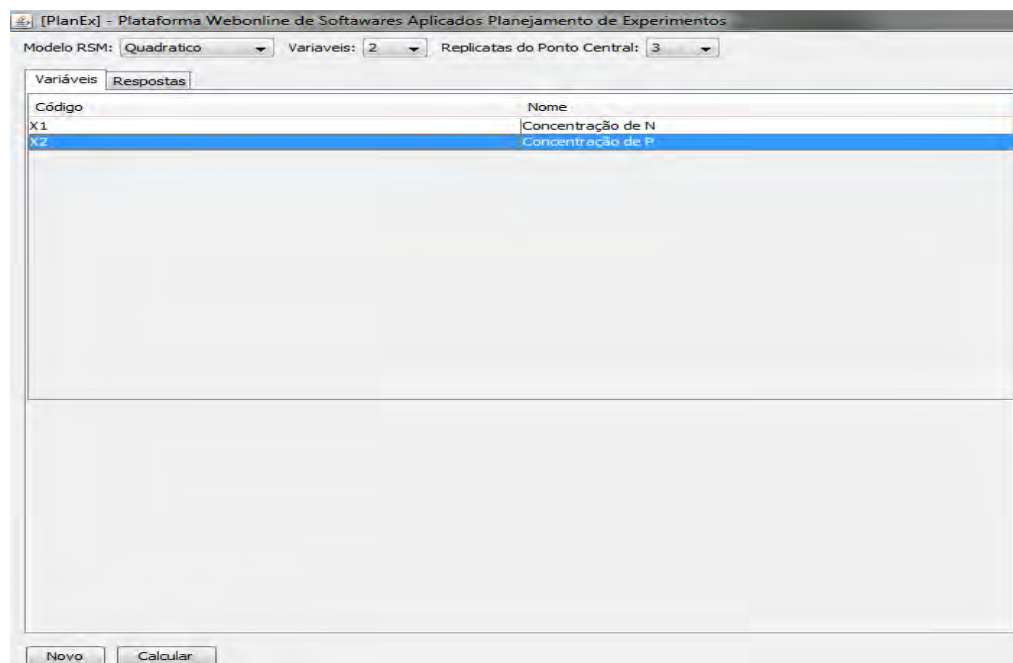


Figura 5.11 – *Layout* da Tela Inicial da RSM do *software* PlanEx usando os dados da Ilustração 2 (FERREIRA, J. et al, 2005)

Na Figura 5.11 apresenta-se o *layout* da Tela Inicial da RSM do *software* PlanEx usando os dados da Ilustração 2. Pelo enunciado da Ilustração 2 e com os dados do artigo de Ferreira (2005), observou-se que para a otimização dos experimentos escolheu-se o modelo quadrático da RSM, usando a concentração de N e P como variáveis independentes e triplicatas do ponto central para investigar se há erro experimental. O *software* PlanEx foi implementado para a RSM do modelo quadrático, planejamento composto central, para duas variáveis de modo que as codificações das variáveis pela Eq. 20 resultem na matriz de contraste semelhante a Eq. 38, exceto que a quantidade de replicatas do nível zero é três na

Ilustração 2, ou seja, três linhas do nível zero. O valor de α é obtido pela Eq. 37. Desse modo não se necessita especificar quais os valores destes níveis. Essa praticidade é uma das vantagens do PlanEx, e essa característica tende a atrair usuários com o objetivo de usar o *software* PlanEx como ferramenta de pesquisa, pois o tempo de manuseio é menor do que de outros *softwares*.

Na Figura 5.12 apresentam-se as respostas experimentais dos dados da Ilustração 2. Observou-se no artigo da Ilustração 2 que a ordem dos experimentos tabelados difere-se da ordem dos experimentos implementados no *software* PlanEx da RSM, porém o usuário deve atentar para o significado de cada símbolo dos experimentos e relacioná-los corretamente, como se comentou anteriormente no Exemplo 13.

Experimentos	Valor
(1)	29,33
x1	31,96
x2	44,33
x1x2	40,29
o1	38,93
o2	43,35
o3	36,72
x1e-	40,42
x1e+	32,98
x2e-	27,3
x2e+	55,87

Figura 5.12 – Layout da Tela Inicial da RSM do *software* PlanEx usando os dados da Aplicação 2 (FERREIRA, J. et al, 2005)

Na coluna 1 da Figura 5.13, *layout* da Tela de Resultados dos Efeitos da RSM do *software* PlanEx, apresenta os símbolos dos efeitos do modelo quadrático, onde X_1^2 e X_2^2 significam os símbolos dos efeitos da variável 1 ao quadrado e da variável 2 ao quadrado, que se denotam por X_1^2 e X_2^2 , respectivamente. O restante é similar às ferramentas do planejamento de experimentos fatorial completo e fracionário, exceto que não há gráfico de probabilidade normal, e há gráfico de superfície de resposta e tabela de ANOVA.

Na coluna 2 da Figura 5.13 apresentam-se os efeitos relacionados com os seus respectivos símbolos da coluna 1. Na coluna 3 apresenta o erro ou desvio padrão dos coeficientes da regressão e não o desvio padrão dos efeitos como se comentou no planejamento de experimentos fatorial completo e fracionário.

Na coluna 4 da Figura 5.13 o usuário pode escolher quais efeitos são significativos para obter a função estimada da regressão gerando os valores estimados e os resíduos no botão *Calcular*. Em seguida, o usuário pode gerar a tabela de ANOVA no botão *Tabela de ANOVA*, caso a quantidade de replicatas esteja no intervalo $2 \leq r \leq 5$, Figura 5.14.

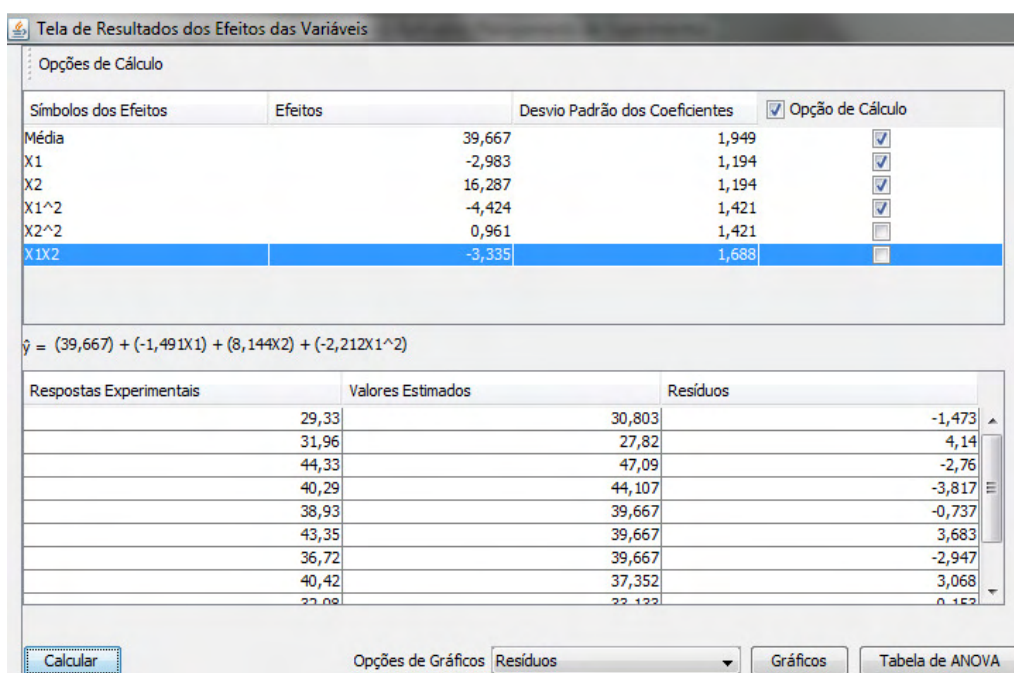


Figura 5.13 – Layout da Tela de Resultados dos Efeitos da RSM do *software* PlanEx usando os dados da Aplicação 2 (FERREIRA, J. *et al*, 2005)

Com base na Figura 5.14, o usuário pode analisar a significância do modelo e decidir em permanecer ou alterar a quantidade de coeficientes ou o tipo do modelo da regressão.

Tabela ANOVA			
Fonte de Variação	Soma Quadrática	Grau de Liberdade	Média Quadrática
Regressão	579,926	3	193,309
Resíduos	106,784	7	15,255
Falta de Ajuste	83,991	5	16,798
Erro Puro	22,792	2	11,396
Total	686,71	10	68,671

R²: 84,45%

% máxima de variação explicável: 96,681%

F_Calculado: 12,672

F_Calculado_Falta_Ajuste: 1,474

Figura 5.14 – Tabela de ANOVA da RSM do *software* PlanEx usando os dados da Ilustração 2 (FERREIRA, J. *et al*, 2005)

Na Figura 5.15, gráfico de resíduos em função dos valores estimados, o usuário pode investigar se o modelo é tendencioso como se comentou na seção da Estatística Experimental, o que pode corroborar com os dados da Figura 5.14.

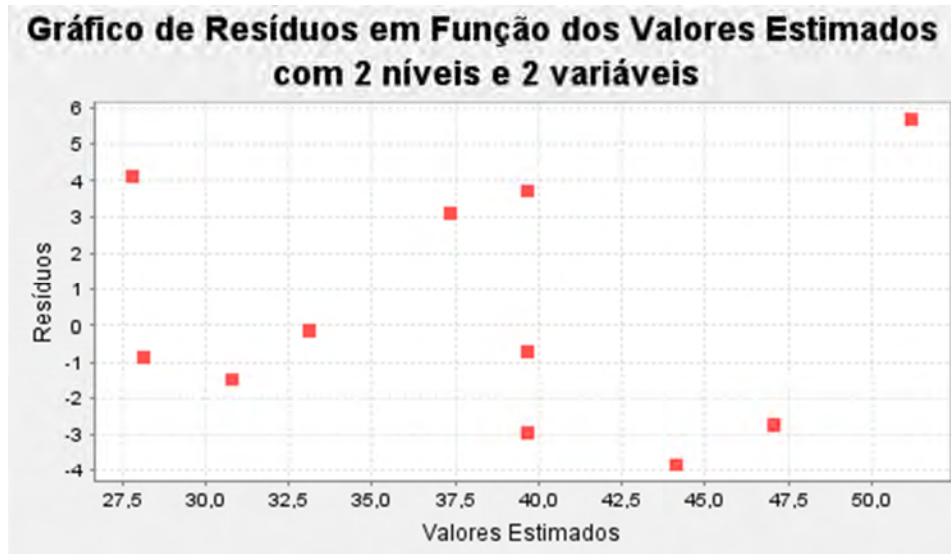


Figura 5.15 – Gráfico de Resíduos em função dos Valores Estimados obtidos com o *software* PlanEx usando os dados da Ilustração 2 (FERREIRA, J. *et al*, 2005)

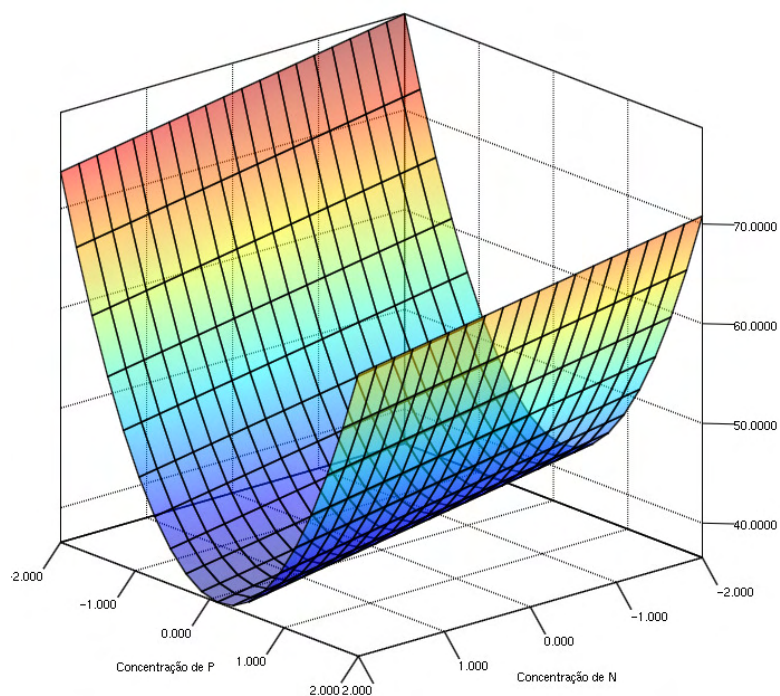


Figura 5.16 – Gráfico de Superfície de Resposta em 3D da RSM obtido com o *software* PlanEx usando os dados da Aplicação 2 (FERREIRA, J. *et al*, 2005)

O gráfico de 3D de superfície de resposta, Figura 5.16, é útil para visualizar com maior eficiência a superfície experimental a fim de encontrar o ponto ótimo do sistema em estudo.

O usuário para gerar o gráfico, Figura 5.16, deve escolher no máximo duas variáveis independentes e/ou a interação entre elas na coluna 4 da Figura 5.13 no botão *Opção de Cálculo*, para selecionar o gráfico de superfície na aba *Opções de Gráficos*, e por fim gera-lo no botão *Gráficos*. Caso contrário, o botão de *Gráficos* não será ativado.

Percebe-se na Figura 5.16 que a rotulação dos eixos das variáveis (Concentração de N e P) equivale à nomeação das variáveis na Figura 5.11, como se comentou anteriormente nesta seção.

No trabalho de Joelma Ferreira *et al* (2005), Ilustração 2, realizou-se os cálculos com o *software* Statistica 5.0, e percebe-se que os cálculos e gráficos realizados são equivalentes aos obtidos com o *software* PlanEx da RSM, sendo uma ferramenta eficiente, rápida e gratuita.

Ilustração 3. No trabalho de Mansilla *et al* (2002) realizaram-se experimentos envolvendo descoloração de Orange II utilizando a metodologia de planejamento fatorial para a simulação

das três variáveis críticas que afetam a descoloração de corantes: concentração de Orange II, pH e concentração de TiO₂.

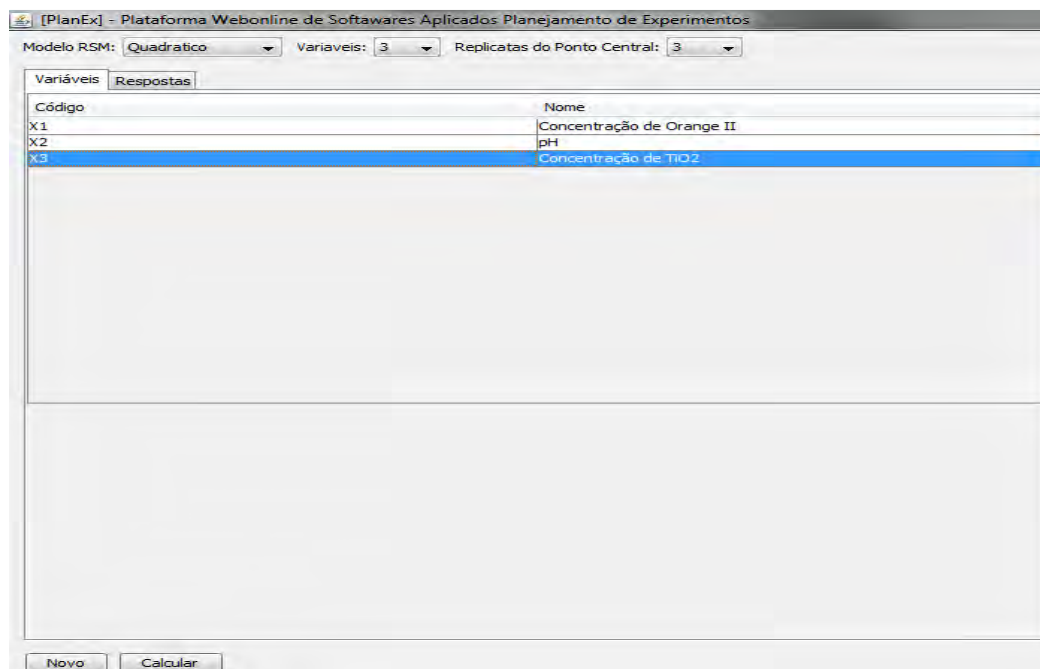


Figura 5.17– Layout da Tela Inicial da RSM do *software* PlanEx usando os dados da Ilustração 3 (MANSILLA *et al*, 2002)

Na Figura 5.17 apresenta o *layout* da Tela Inicial da RSM do *software* PlanEx usando os dados da Ilustração 3. Observa-se que o modelo da RSM selecionado foi o quadrático com três variáveis independentes: concentração de Orange II, pH e concentração de TiO₂ e com triplicatas do ponto central para obter uma estimativa dos erros dos coeficientes da regressão. A digitação dos nomes das variáveis é obrigatória para a rotulação do gráfico de superfície de resposta.

O *software* PlanEx foi implementado para a RSM do modelo quadrático, planejamento composto central, para três variáveis de modo que as codificações das variáveis pela Eq. 20 resultem na matriz de contraste semelhante a Eq. 39, exceto que a quantidade de replicatas do nível zero é três na Ilustração 3, ou seja, três linhas do nível zero. O valor de α é obtido pela Eq. 37. Desse modo não se necessita especificar quais os valores destes níveis. O mesmo acontece para o caso de quatro variáveis, ou seja, não há a necessidade de especificar os níveis, desde que a codificação pela Eq. 20 e pela Eq. 37 resulte na metodologia do planejamento composto central e na quantidade de experimentos definida por $2^k + 2k + r$.

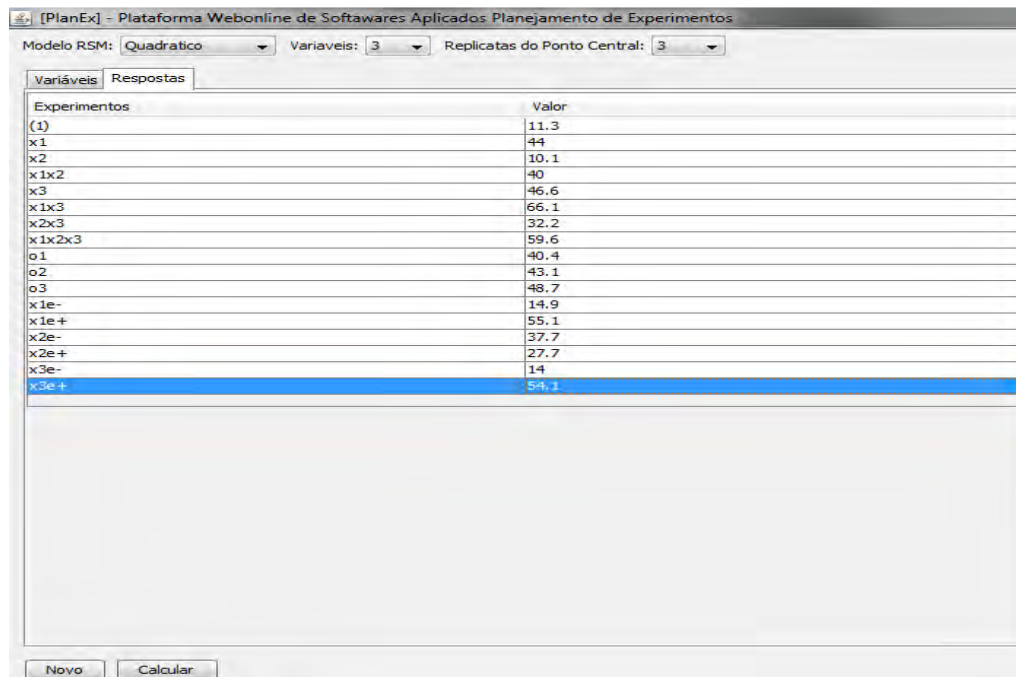


Figura 5.18 – Layout da Tela Inicial da RSM do *software* PlanEx usando os dados da Ilustração 3 (MANSILLA *et al*, 2002)

Na Figura 5.18 apresenta-se as respostas experimentais da Ilustração 3 digitadas de acordo com o símbolos dos experimentos como se comentou na Tabela 2.26 do Exemplo 13, exceto que na Figura 5.15 há alguns símbolos a mais, porém o significado é similar ao comentado no Exemplo 13.

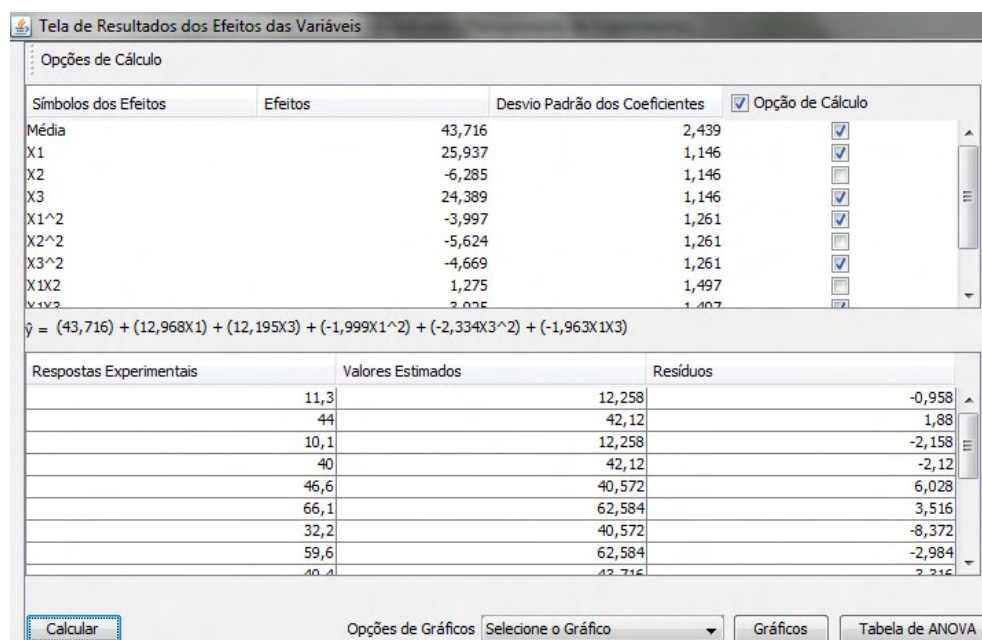


Figura 5.19 – Layout da Tela de Resultados dos Efeitos das Variáveis do *software* PlanEx usando os dados da Aplicação 3 (MANSILLA *et al*, 2002)

Na Figura 5.19 apresenta-se o *layout* da Tela de Resultados do *software* PlanEx dos dados da Ilustração 3. O comentário da Figura 5.19 é semelhante ao comentário ao da Figura 5.13, exceto que na Figura 5.19 o usuário deve escolher no máximo duas variáveis independentes e/ou sua interação para habilitar o botão de *Gráficos* para gerar o gráfico de superfície de resposta, Figura 5.22, como se comentou anteriormente.

Na Figura 5.20, gráfico de resíduos em função dos valores estimados, o usuário pode investigar se o modelo é tendencioso como se comentou na seção da Estatística Experimental, o que pode corroborar com os dados da Figura 5.19.



Figura 5.20 – Gráfico de Resíduos em função dos Valores Estimados obtido com o *software* PlanEx usando os dados da Ilustração 3 (MANSILLA *et al*, 2002)

Com base na Figura 5.21, o usuário pode analisar a significância do modelo e decidir em permanecer ou alterar a quantidade de coeficientes ou o tipo do modelo da regressão.

Fonte de Variação	Soma Quadrática	Grau de Liberdade	Média Quadrática
Regressão	4,540,576	5	908,115
Resíduos	514,033	11	46,73
Falta de Ajuste	478,186	9	53,132
Erro Puro	35,847	2	17,923
Total	5.054,609	16	315,913

R²: 89,83%

% máxima de variação explicável: 99,291%

F_Calculado: 19,433

F_Calculado_Falta_Ajuste: 2,964

Figura 5.21 – Layout da Tabela de ANOVA da RSM do *software* PlanEx usando os dados da Ilustração 3 (MANSILLA *et al*, 2002)

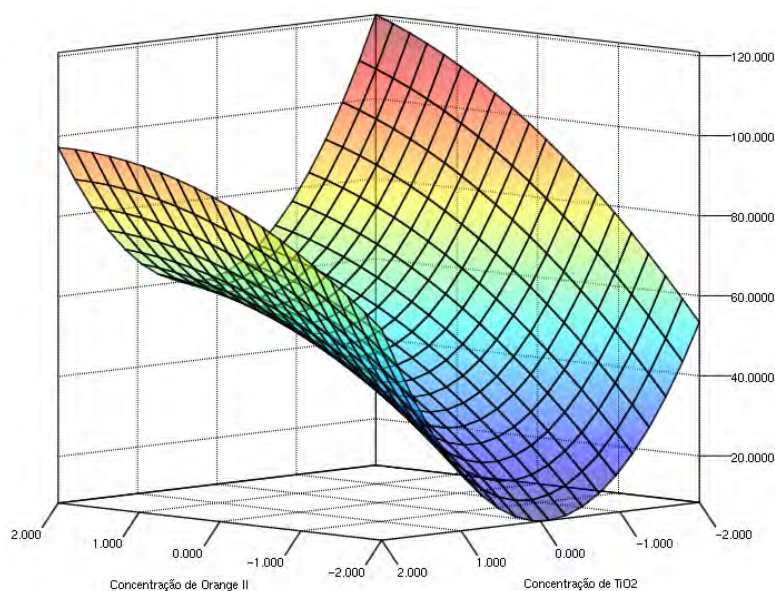


Figura 5.22 – Gráfico de Superfície de Resposta da RSM obtido com o *software* PlanEx usando os dados da Ilustração 3 (MANSILLA *et al*, 2002)

Após escolher os efeitos significativos ou de sua escolha, o usuário deve obter os valores estimados com os resíduos aleatórios através do botão *Calcular* no *layout* da Tela de Resultados do *software* PlanEx para gerar os gráficos e a Tabela de ANOVA.

A Figura 5.22 foi obtida pela função estimada, \hat{y} , no *layout* da Figura 5.19 abaixo do primeiro quadrado, a qual deve possuir no máximo duas variáveis independentes para o gráfico possui 3 dimensões, 3D, (1 dimensão da variável dependente e 2 dimensões das variáveis independentes).

Ilustração 4. No trabalho de Muniz *et al* (2008) os autores sintetizaram hidrogéis condutores do tipo semi-IPN constituídos por poliacrilamida (PAAm) e polímero intrinsecamente condutor poli(3,4-etileno dioxitiofeno)/poli(estireno-sulfonato) (PEDOT/PSS) e investigaram as propriedades hidrofílicas, mecânicas, condutoras e ópticas por meio de planejamento fatorial completo 2^3 com ponto central. Para isso, utilizaram-se três fatores: concentrações de acrilamida (AAm) e N,N'-metileno-bis-acrilamida (MBAAm), e quantidade de PEDOT/PSS, em três níveis, sendo um deles o ponto central. O planejamento foi construído para avaliar os efeitos desses fatores (efeitos principais e efeitos de interação entre os fatores) nas seguintes propriedades dos hidrogéis estudados (respostas): grau de intumescimento, tensão máxima de compressão, condutância iônica e intensidade relativa de luz transmitida através dos hidrogéis estudados.

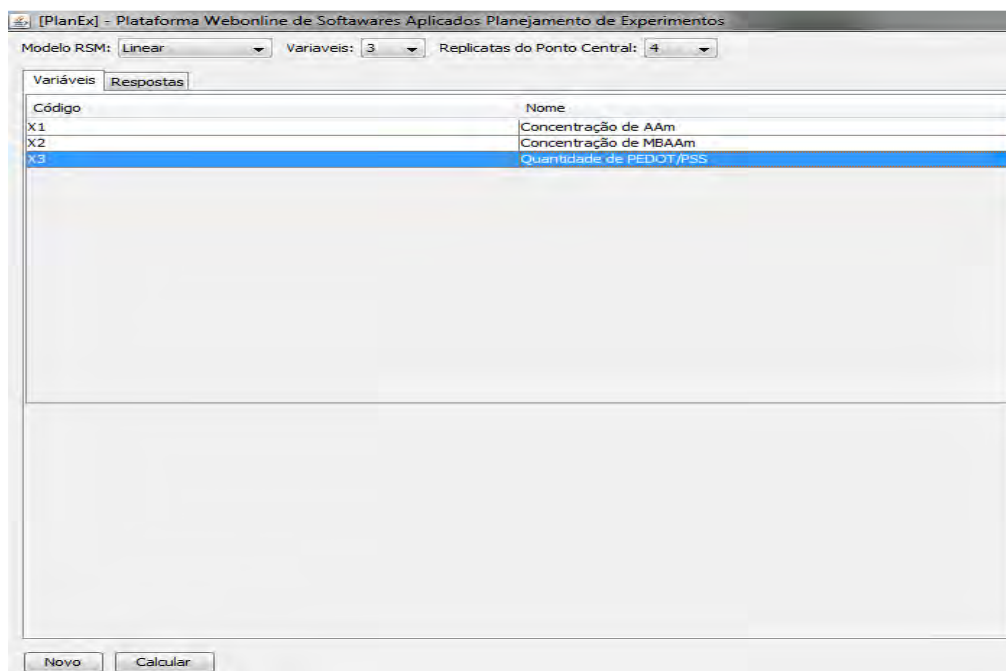


Figura 5.23 – Layout da Tela Inicial da RSM do software PlanEx usando os dados da Ilustração 4 (Muniz *et al*, 2008)

A Figura 5.23 apresenta o *layout* da Tela Inicial do software PlanEx da RSM onde escolheu-se o modelo linear com três variáveis independentes (concentrações de AAm, concentração de MBAAm e quantidade de PEDOT/PSS) e com triplicatas do ponto central.

O software PlanEx da RSM para o modelo linear foi implementado de forma que se subentende-se que as variáveis independentes serão analisadas em três níveis, logo não há a

necessidade de especificar o valor do módulo de cada variável, desde que a codificação delas pela Eq. 20 da seção da Estatística experimental resulte num nível: menor (-1), central ou zero (0) e maior (+1), e a quantidade de experimentos seja definida por $2^k + r$, sendo $2 \leq k \leq 4$ e $1 \leq r \leq 5$.

Experimentos	Valor
(1)	9,93
x1	6,81
x2	5,86
x1x2	4,13
x3	10,12
x1x3	7,27
x2x3	5,97
x1x2x3	4,22
o1	5,48
o2	5,55
o3	5,51
o4	5,55

Figura 5.24 – Layout da Tela Inicial da RSM do software PlanEx usando os dados da Ilustração 4 (MUNIZ *et al.*, 2008)

Na Figura 5.24 apresentam-se as respostas experimentais (grau de intumescimento g/g da Ilustração 4) digitadas pelo usuário de acordo com a simbologia dos sete primeiros experimentos da Tabela 2.26 do Exemplo 13.

Na Figura 5.25 apresenta-se o layout da Tela de Resultados da RSM do software PlanEx em que o comentário é similar aos comentários realizados das Figuras 5.19 e 5.13.

Na Figura 5.26 apresenta-se o gráfico de resíduos em função dos valores estimados gerado pelo software PlanEx obtido pela função estimada, \hat{y} , abaixo do primeiro quadro da figura, onde o usuário escolheu os efeitos significativos ou de sua escolha.

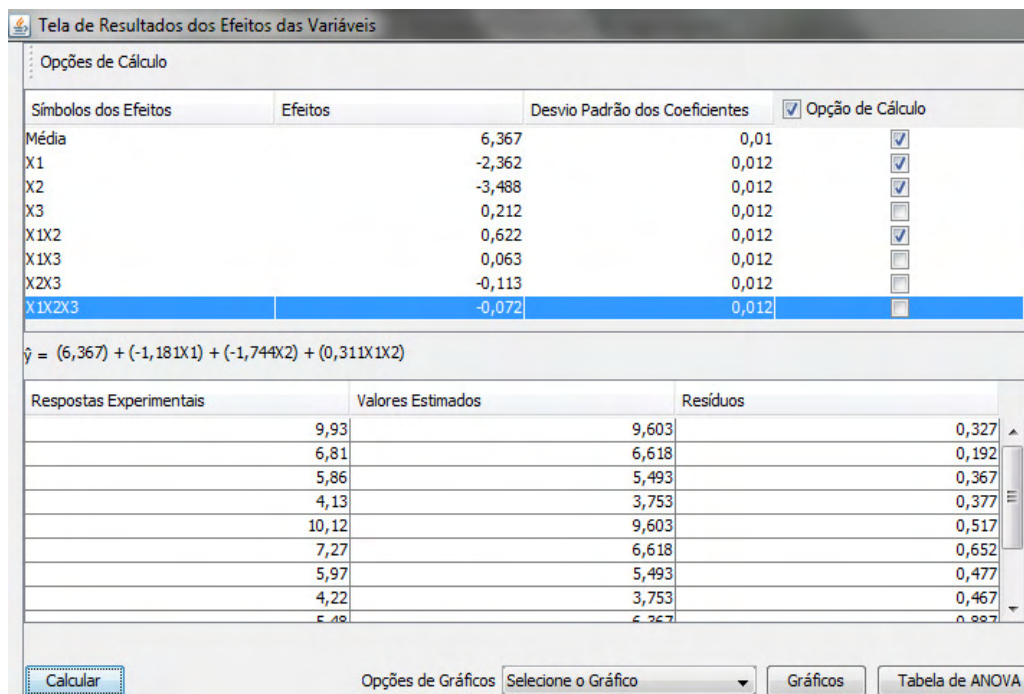


Figura 5.25 – Layout da Tela de Resultados dos Efeitos das Variáveis do *software* PlanEx usando os dados da Ilustração 4 (MUNIZ *et al*, 2008)



Figura 5.26 – Gráfico de Resíduos em função dos Valores Estimados obtido com o *software* PlanEx usando os dados da Ilustração 4 (MUNIZ *et al*, 2008)

Na Figura 5.27 apresenta a Tabela de ANOVA obtida após o usuário estimar a função na Figura 5.25. Na Figura 5.28 apresenta-se o gráfico de superfície de resposta obtida pela

função estimada na Figura 5.25, visto que a função deve possuir no máximo duas variáveis independentes e/ou sua interação, como se comentou anteriormente.

Fonte de Variação	Soma Quadrática	Grau de Liberdade	Média Quadrática
Regressão	36,263	3	12,088
Resíduos	4,413	8	0,552
Falta de Ajuste	4,41	5	0,882
Erro Puro	0,003	3	0,001
Total	40,676	11	3,698

R²: 89,151%
 % máxima de variação explicável: 99,991%
 F_Calculado: 21,912
 F_Calculado_Falta_Ajuste: 761,379

Figura 5.27 – Layout da Tabela de ANOVA da RSM do *software* PlanEx usando os dados da Ilustração 4 (MUNIZ *et al*, 2008)

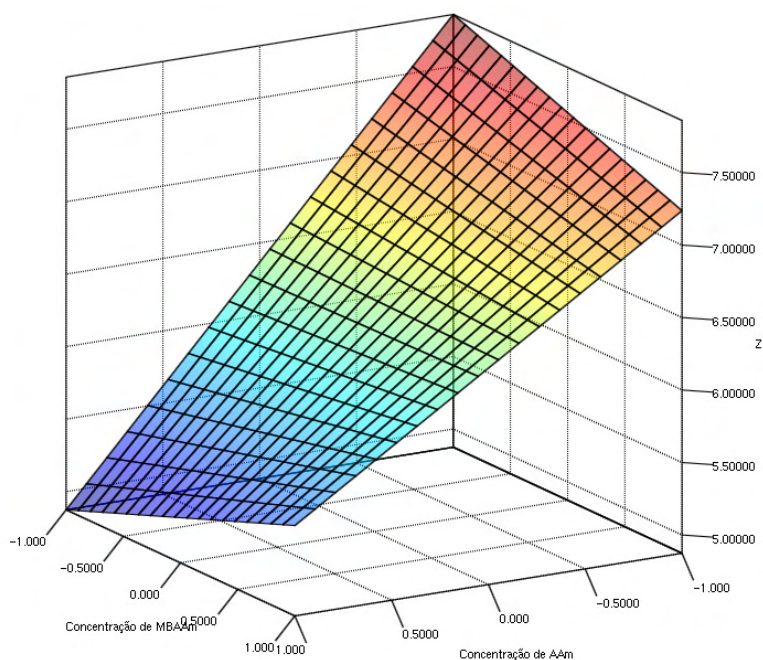


Figura 5.28 – Gráfico de Superfície de Resposta da RSM obtido com o *software* PlanEx usando os dados da Ilustração 4 (MUNIZ *et al*, 2008)

Percebem-se através dos comentários e das figuras desta seção que a fato do *software* PlanEx não requerer o conhecimento de uma linguagem computacional específica como o *Matlab* tende a incentivar o acesso e o uso dele (*software* PlanEx) como ferramenta de ensino

e pesquisa, o que é corroborado pelo fato do PlanEx não requer a digitação de matrizes e nem requer as operações entre elas e nem o processo de codificação de variáveis para obter a matriz de contraste através da Eq. 20. Pois o fato de um *software* requerer o conhecimento de uma linguagem específica isso acarreta um despendimento de um tempo maior de estudo em conteúdos que não são específicos para resolver o problema, portanto, o PlanEx pode permitir ao usuário um tempo maior para realizar análises do problema, facilitando o ensino e a pesquisa.

5.4 Validação do *software* PlanEx Como Ferramenta de Ensino e Pesquisa

A etapa de validação do *software* PlanEx realizou-se em três datas em distintas turmas da disciplina de Quimiometria II, Planejamento e Otimização de Experimentos, ministrada pelo Professor Dr. Anselmo Elcana de Oliveira no segundo semestre dos anos de 2009, 2010 e 2011. A quantidade de alunos em cada semestre foi 7, 16 e 13 alunos, respectivamente.

A implementação do *software* PlanEx realizou-se em três etapas, sendo planejamento de experimentos fatorial completo 2^k , planejamento de experimentos fatorial fracionário 2^{k-1} e metodologia de superfície de resposta – RSM, respectivamente. Na medida em que se desenvolveram as implementações do *software* PlanEx realizaram aplicações, ou seja, disponibilizou-se o *software* PlanEx aos alunos da disciplina de Quimiometria II, Planejamento e Otimização de Experimentos, ministrada pelo Prof. Dr. Anselmo Elcana de Oliveira no programa de pós-graduação de Mestrado em Química da UFG e de Doutorado Multiinstitucional UFU/UFG/UFMT para eles conhecerem e usarem as ferramentas disponíveis em cada época.

5.4.1 Primeira Aplicação do *Software* PlanEx aos usuários de Quimiometria II

Na primeira aplicação com os sete alunos, segundo semestre de 2009, o *software* PlanEx possuía apenas as ferramentas do planejamento de experimentos fatorial completo 2^k . Naquele momento, os usuários obtiveram:

- efeitos de um planejamento de experimentos 2^k com ou sem replicatas,
- a função estimada **completa** do modelo linear considerado;
- gráficos de resíduos aleatórios em função dos valores estimados
- gráfico de probabilidade normal.

Como o *software* PlanEx estava em fase inicial, o objetivo da primeira aplicação, segundo semestre de 2009, era disponibilizá-lo para que os usuários pudessem dar as opiniões sobre as ferramentas, apontando pontos positivos e pontos negativos os quais seriam analisados para melhorar a implementação do *software*. Nessa época, as opiniões foram enviadas por e-mail, pois o objetivo era adquirir indícios para melhoria do *software* PlanEx.

Com base em vários e-mails e nas conversas observadas pelo vídeo gravado durante a primeira aplicação percebeu-se que as características mais atrativas do *software* PlanEx eram o *layout* simples e a rapidez em obter os efeitos, como se pode observar nos trechos transcritos abaixo de alguns alunos:

“Gostei muito do PlanEx, pois a maneira de obter os efeitos é rápido e simples. Bem prático!”

“É impressionante o tempo economizado ao resolver as atividades de Planejamento Fatorial Completo do semestre da disciplina de Quimiometria II no PlanEx. É rápido! Queria ter conhecido o PlanEx antes de ter realizado a primeira avaliação.”

“Apesar do PlanEx não ser chamativo, ele funciona bem e rápido. impressionante o tempo economizado ao resolver as atividades de Planejamento Fatorial Completo do semestre da disciplina de Quimiometria II no PlanEx. É rápido! Queria ter conhecido o PlanEx antes de ter realizado a primeira avaliação.”

A agilidade do PlanEx deve-se pela ausência de matrizes e de codificações de variáveis. Um exemplo: nessa aplicação os alunos resolveram todas as atividades de planejamento de experimentos fatorial completo práticas do semestre disponibilizadas na página do Professor Dr. Anselmo Elcana de Oliveira em menos de uma hora, sendo que no *software* *Octave* despenhou-se um tempo maior.

Na primeira aplicação obteve-se algumas sugestões de melhoria e que foram implementadas, tais como:

- gerar a função estimada de acordo com a escolha dos efeitos significados ou efeitos que o usuário queria escolher, como se apresentou no botão *Opção de Cálculo* na Figura 5.;
- apresentar o desvio padrão quando há replicatas.

5.4.2 Segunda Aplicação do Software PlanEx aos usuários de Quimiometria II

Na segunda aplicação com os 16 alunos no segundo semestre de 2010, o *software* possuía as ferramentas do planejamento de experimentos fatorial completo 2^k e as ferramentas do planejamento de experimentos fatorial fracionário 2^{k-1} e melhorias sugeridas pelos alunos na primeira aplicação. Durante a segunda aplicação surgiram algumas dúvidas, oriundas do *software* não possuir na época um manual ou tutorial de ajuda. Essas dúvidas foram registradas no vídeo gravado, e a seguir são transcritas as duas principais.

1. “Qual é a geratriz do planejamento de experimentos fatorial fracionário?”;
2. “É possível realizar qual tipo de fração do planejamento fatorial fracionário?”.

Em relação à pergunta 1, o fato de usar uma ou outra geratriz (*meia fração positiva* ou *meia fração complementar*) não interfeririam nos resultados, pois uma análise adequada levaria às mesmas conclusões. Com base nisso, programou-se somente a *meia fração positiva*.

Em relação à pergunta 2, observou-se que a maior parte dos estudos realizados com planejamento de experimentos fatorial fracionário usa-se a meia fração que reduz pela metade a quantidade de experimentos realizados. Com base nisso, programou-se somente planejamento de experimentos fatorial de meia fração pelo fato de ser o mais utilizado e fornecer resultados obtidos através de uma análise mais simples.

Nessa aplicação o objetivo era disponibilizar o *software* PlanEx para os usuários utilizarem e realizar após o manuseio uma comparação em relação aos outros *softwares* que eles conheciam, como o *Ocatave*, e utilizavam em sala de aula. Também, objetivou-se analisar a evolução do *software* PlanEx como ferramenta de pesquisa e ensino através de sugestões e críticas enviadas por e-mail de forma espontânea pelos alunos.

Pelo vídeo gravado percebeu-se que os alunos estavam concentrados no *software* PlanEx e que a pequena dispersão em sala se deve ao fato da discussão entre os alunos em relação a economia de tempo, facilidade e simplicidade em resolver problemas de planejamento de experimentos fatorial usando o PlanEx. A maior parte dos alunos realizaram comentários de elogios sobre o PlanEx e sobre o tema do trabalho de doutorado, como se pode observar pelos trechos transcritos de alguns alunos:

“Nossa! O trabalho é fantástico. Vocês estão de parabéns.”

“O software é simples, mas funciona muito bem. O jeito de obter os resultados e gráficos sem a necessidade de digitar e saber operar com matrizes facilitou o nosso estudo. Teremos mais tempo para realizar a análise do problema, pois é isso que conta no fim das contas.”

5.4.3 Terceira Aplicação do Software PlanEx aos usuários de Quimiometria II

Na terceira aplicação com os 13 alunos, segundo semestre de 2011, o *software* PlanEx estava completo, ou seja, possuía além das ferramentas comentadas nas seções anteriores (5.4.1 e 5.4.2) as ferramentas da RSM. Os usuários nessa aplicação obtiveram e/ou visualizaram:

- efeitos com os respectivos erros de um planejamento fatorial usando o modelo linear com replicatas ou o modelo quadrático do planejamento composto central;
- a função estimada após a escolha dos efeitos;
- os resíduos aleatório para gerar a tabela de ANOVA;
- gráfico de resíduos em função dos valores estimados;
- gráfico de superfície de resposta em 3D de modo que o gráfico possa girar automaticamente para uma melhor visualização da otimização dos experimentos.

Nessa aplicação objetivou-se validar o *software* PlanEx como ferramenta de pesquisa e ensino, pois se acreditava que continha as ferramentas necessárias e suficientes para resolver um problema de planejamento de experimentos fatorial de dois e três níveis, sendo um nível zero. Para a validação, disponibilizou o *software* PlanEx aos alunos para eles conhecerem as ferramentas do planejamento de experimentos fatorial completo 2^k , do planejamento de experimentos fatorial fracionário 2^{k-1} e, da RSM.

Após o manuseio, pediu-se aos usuários para responder um questionário com 8 perguntas, **Apêndice D**, para obter uma medida quantitativa para validar ou não o *software* PlanEx como ferramenta de pesquisa e ensino. Apresenta-se a seguir os comentários realizados com base nas medidas quantitativas e as justificativas dos comentários através de uma análise qualitativa.

Na **Questão 1**, perguntou se as ferramentas do planejamento fatorial completo 2^k do *software* PlanEx são suficientes para analisar um problema.

Comentou-se que o planejamento fatorial completo com dois níveis é útil para realizar uma investigação preliminar de um problema em estudo e, dependendo da necessidade, pode

ser ampliado. As ferramentas do planejamento fatorial completo são: os efeitos, a função estimada, o gráfico dos resíduos, o gráfico de probabilidade normal e a representação geométrica, sendo um quadrado com um planejamento com duas variáveis e um cubo com três variáveis. Para um planejamento com mais de três variáveis não é possível fazer a representação geométrica.

No *software* PlanEx programaram-se praticamente todas as ferramentas do planejamento de experimentos fatorial completo 2^k , exceto a representação geométrica. No entanto, a falta dessa ferramenta não compromete a análise de um problema. Dos 13 alunos entrevistados, apenas 7 responderam as perguntas, visto que os alunos estavam resolvendo os problemas em dupla. Dos sete alunos, 100% opinaram favorável de que as ferramentas são suficientes para analisar um problema, e desses alunos 42,86% justificaram a resposta fornecida dizendo que o *software* PlanEx é prático e que abarca os cálculos e gráficos de interesse, e que a obtenção dos efeitos sem conhecimento prévio de uma linguagem de programação ou codificação facilita o trabalho do usuário e que isso torna interessante o manuseio da plataforma, como se apresenta na Figura 5.29 e nos trechos transcritos dos alunos na terceira aplicação:

“São muito práticas e abarca os cálculos e gráficos de interesse”

“ Foi interessante, pelo fato do cálculo ser direto, não havendo a necessidade de conhecer códigos e/ou programação.”

“ Sim. Porém, poderia ter opções de manipulação dos gráficos como:

- Rótulo dos pontos dos gráficos;
- Modificar as opções dos eixos (fontes, limites dos eixos, escolha de variáveis). Poderia ter essas opções na própria janela do gráficos.”

Na **Questão 2**, perguntou se as ferramentas do planejamento fatorial fracionário 2^{k-1} do *software* PlanEx são suficientes para analisar um problema.

As ferramentas do planejamento de experimentos fatorial fracionário são as mesmas daquelas do planejamento de experimentos fatorial completo. Todos os usuários disseram que as ferramentas são eficazes para analisar um problema, e desses 42,86% justificaram a resposta dada. Alguns disseram que o *software* PlanEx é de fácil acesso e demanda pouco tempo do analista ou usuário no tratamento de dados, e outros disseram que a agilidade é maior em relação aos outros *softwares* gratuitos usados com a mesma finalidade, como se apresenta na Figura 5.30.

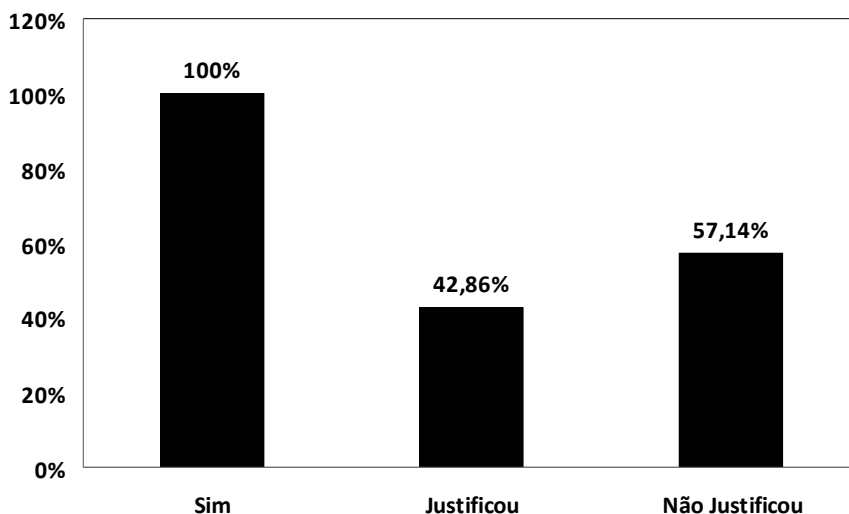


Figura 5.29 – Análise Quantitativa da Questão 1 – Apêndice D

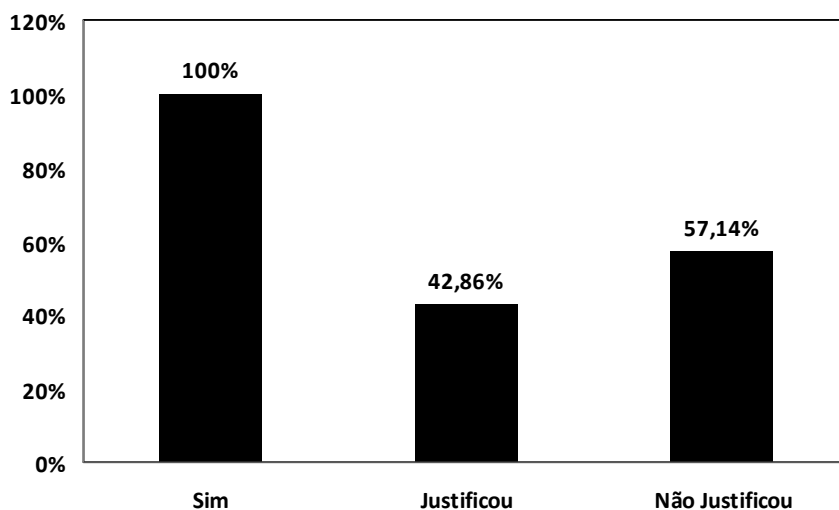


Figura 5.30 – Análise Quantitativa da Questão 2 – Apêndice D

Na **Questão 3**, perguntou se as ferramentas da RSM do *software* PlanEx são suficientes para analisar um problema.

Comentou-se que a RSM com três níveis neste trabalho constitui-se por um planejamento de experimentos fatorial completo acrescido de ponto central com ou sem replicatas do ponto central, permitindo a investigação de indícios de curvatura na superfície de resposta e de avaliação do modelo. As ferramentas da RSM são:

- os efeitos do planejamento;

- a função estimada através da escolha dos efeitos significativos ou da escolha do usuário;
- o gráfico dos resíduos aleatórios em função dos valores estimados;
- o gráfico de superfície de resposta;
- a tabela de ANOVA.

Dos usuários entrevistados, Figura 5.31, 57,14% disseram que as ferramentas da RSM do *software* são suficientes para analisar um problema, 14,29% disseram que não são suficientes, 28,57% não responderam, e 71,43% justificaram a resposta dada. Os usuários que disseram que as ferramentas da RSM não são suficientes para analisar um problema foram unânimes ao justificar a resposta dada, dizendo que o *software* PlanEx deveria apresentar gráficos em três dimensões da superfície de resposta, 3D, como se pode observar nos trechos transcritos a seguir:

“Não abriu o gráfico de superfície de resposta.”

“ Deve ser acrescentado o rótulo de cores para visualização junto com o gráfico de superfície de resposta.”

Não foi possível testar os gráficos 3D.

Não foi totalmente testado.

O gráfico da superfície não apareceu.

Existem vários estudos que sugerem e comprovam que figuras 3D são úteis para uma melhor compreensão de um fenômeno em estudo. Por exemplo, no Trabalho de Meira *et al* (2011), espectrofluorimetria total 3D e análise de componentes principais (PCA), possibilitaram de uma forma mais adequada diferenciar amostras puras de diesel, biodiesel, óleos vegetais e óleo residual. E ainda foi possível identificar adulterações de diesel, pela adição de óleo residual resultante de fritura ao invés de biodiesel ou em decorrência do excesso de biodiesel em diesel, ultrapassando os valores permitidos por lei. Já no trabalho de Arroio *et al* (2011), os autores afirmam que:

Os avanços na computação têm permitido a construção de *softwares* de alta performance, nomeadamente no que diz respeito ao uso de imagens 3D (tridimensionais) projetadas em duas dimensões que, devido a suas características, permitem uma melhor visualização dos modelos adotados pela comunidade científica para representar os arranjos atômicos e os processos envolvidos nos fenômenos químicos.

No trabalho de Ferk (2003), na área de ensino da química, os autores afirmam que os alunos preferem as representações 3D a outros tipos de representações. Essa preferência deve-se ao fato de uma representação 3D ajudar os alunos a obterem uma melhor percepção do arranjo espacial dos átomos, da conectividade entre átomos e moléculas e das interações permanentes entre as várias entidades sub-microscópicas presentes em qualquer porção de matéria (FERK, 2003 apud ARROIO *et al*, 2011).

Percebe-se que imagens e/ou figuras em 3D, mesmo que sejam plotadas em 2D, permitem uma melhor visualização do fenômeno em estudo. Baseados nisso e na dimensão do mundo real (3D), conclui-se que os usuários, 14,29%, estão corretos ao afirmar que o *software* PlanEx deveria possuir gráficos de superfície de resposta 3D. Isso é um bom indicio, pois no dia da aplicação do *software* PlanEx os computadores não geraram o gráfico de superfície de resposta em três dimensões, 3D, por não possuir o API de *Java* 3D instalado, o que percebeu-se somente após a finalização da aplicação. Porém, os usuários visualizaram o gráfico 3D no computador utilizado como guia de apresentação no data show.

O fato do *software* PlanEx não ter gerado os gráficos de superfície de resposta provocou em 14,29% dos entrevistados o julgamento de insuficiência das ferramentas da plataforma para analisar um problema, porém isso foi um problema momentâneo de atualização de *softwares* de *Java* necessários à utilização. Contudo, acredita-se que uma nova aplicação não seja condição necessária para afirmar que as ferramentas da RSM são suficientes para analisar um problema. Explicou-se ou comentou-se a justificativa da não geração do gráfico 3D para alguns usuários após a aplicação.

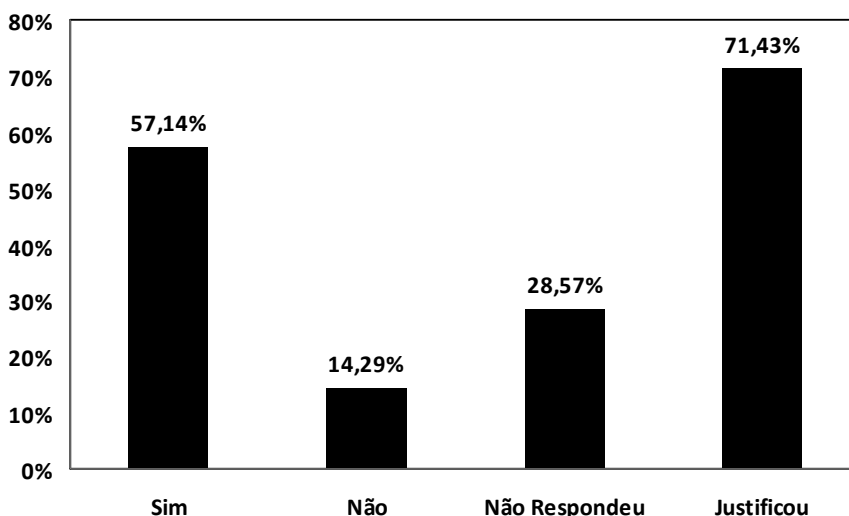


Figura 5.31 – Análise Quantitativa da Questão 3 – Apêndice D

A **Questão 4** afirmou que há *softwares* que apresenta característica de amigabilidade, e com base nisso, a questão objetivou saber se o *software* PlanEx apresenta tal característica.

Dos usuários entrevistados, Figura 5.32, 71,43% disseram que o *software* apresenta uma boa ou simplesmente apresenta a característica de amigabilidade, 28,57% não responderam a pergunta, e 57,14% justificaram a resposta dada.

Em seu livro, Jakob Nielsen (JAKOB e LOGRANGER, 2000) trata da simplicidade em planejar a usabilidade na *web*, onde um *software* amigável deve possuir uma boa usabilidade e um bom *design* gráfico. A capacidade de aprendizado, a eficiência e a memorabilidade do *software* são componentes da usabilidade. No PlanEx percebe-se que é fácil aprender a operar as ferramentas do *software*, as quais apresentam ser eficientes. Como o *layout* é simples, pode inferir que uma vez aprendido a operar o PlanEx não se esquece mais. Uma das características para um bom *design* gráfico de um *software* consiste em possuir simplicidade.

A palavra Amigável aparece em vários trabalhos e em diferentes áreas. Na área farmacêutica o trabalho de Pezza *et al* (2010) apresenta uma metodologia mais amigável e simples de determinação de metoclopramida para não agredir o meio ambiente. No trabalho de Stulzer *et al* (2010), no título do artigo aparece duas palavras (simples e rápido) que são características da palavra amigável, e que os usuários do PlanEx citaram ao referir sobre o *software* na **Questão 4**.

Os usuários que não responderam também não justificaram, e dos que responderam a maioria justificou a resposta com argumentos que comprova que o *software* possui uma boa interface de fácil manuseio e de simples execução, dando a ele a característica de amigabilidade. Durante as três aplicações, 2009 a 2011, o *software* não possuía *menu* de Ajuda. No entanto, após a última aplicação, 2001, implementou-se um *menu* de ajuda, pois alguns usuários sugeriram que se o *software* tivesse um *menu* de ajuda, ele apresentaria mais amigável.

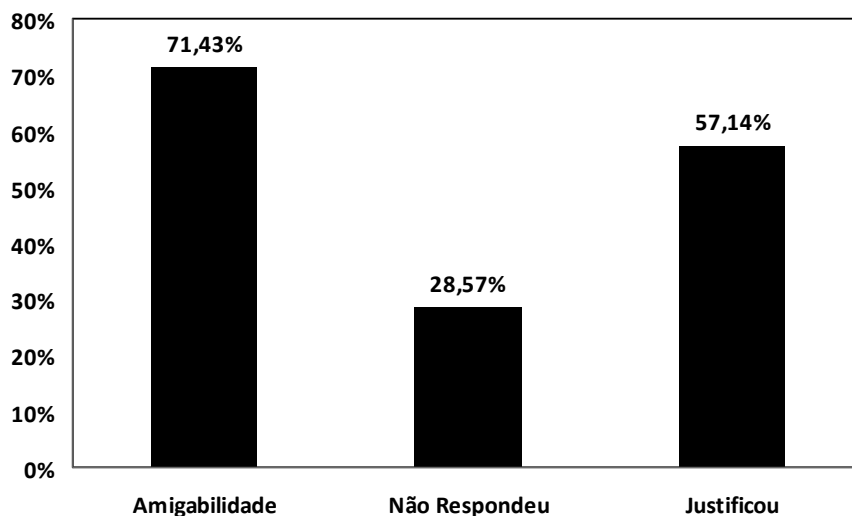


Figura 5.32 – Análise Quantitativa da Questão 4 – Apêndice D

Ao responder a **Questão 5** o usuário teve a possibilidade de comentar sobre a característica de agilidade do *software* PlanEx em relação a de outros *softwares*.

Dos usuários entrevistados, Figura 5.33, 100% disseram que o PlanEx apresenta maior agilidade do que outros *softwares* usados por eles, por exemplo, *Octave* e *Matlab* v R2011b (7.13.0.564), e todos justificaram a resposta dada, como pode observar nos trechos transcritos a seguir:

“Sim. Principalmente em relação ao *Octave* e ao *Matlab* o software em questão é mais rápido.”

“ Comparando com o *Octave*, que é o que nós vínhamos usando, ele apresenta muito maior agilidade.”

“ A plataforma é bem mais ágil que os softwares “*Octave*” e “*Matlab*” pois não é necessário digitar a matriz “X” e o programa efetua a formação de gráficos em uma velocidade considerável.”

“ Sim, e quando ocorre um erro de digitação você pode de maneira rápida fazer uma correção.”

“ Sim, porém falta um tutorial (menu de ajuda).”

“ Comparado com os programas *Octave* e Excel, esta plataforma é mais prática, possibilitando uma enorme rapidez no desenvolvimento.”

“ Sim, pois devido a ausência de ter que calcular percebe-se esta característica.”

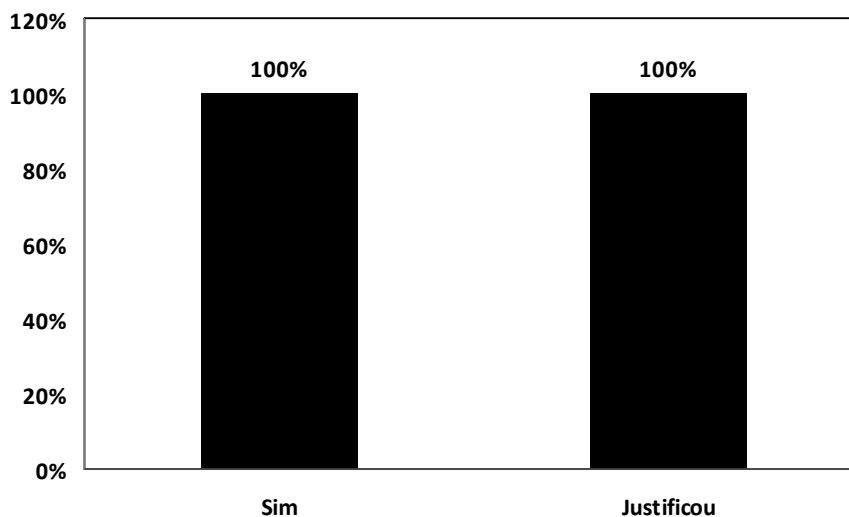


Figura 5.33 – Análise Quantitativa da Questão 5 – Apêndice D

A realização dos cálculos do planejamento de experimentos fatorial completo ou fracionário como da RSM é uma tarefa realizada através da álgebra linear e estatística. Com duas variáveis em dois níveis a dimensão da matriz de contrastes para o modelo linear é de quarta ordem e a dimensão dos cálculos é no mínimo trabalhosa. Quando se aumenta uma variável num planejamento em dois níveis a dimensão da matriz de contraste aumenta-se para oitava ordem, tornando o cálculo cansativo e repetitivo. Em vários *softwares* é necessário realizar uma codificação das variáveis nos níveis em estudo e conhecer uma linguagem própria de cada *software*, tornando o processo de obtenção dos parâmetros da regressão cansativo, repetitivo e com um longo tempo de digitação.

Na experimentação ou produção em indústria, a variável tempo é um fato primordial no que diz respeito a otimização de experimentos, e está presente em vários trabalhos na literatura. Um exemplo, no trabalho de Ruela (2009), os autores apresentaram um método rápido através de cromatografia líquida de alta eficiência para determinação de nimesulida.

Neste trabalho o *software* PlanEx oferece os mesmos resultados e as mesmas ferramentas básicas para análise de um problema através de um planejamento fatorial completo ou fracionário ou de uma RSM com um menor tempo de digitação e execução, como percebeu-se na resposta de todos os entrevistados. O ponto alto do trabalho é oferecer uma ferramenta simples, gratuita e rápida, e que funcione da mesma forma que as ferramentas disponibilizadas pelos *softwares* pagos.

Na **Questão 6** objetivou-se saber qual o grau de interatividade do *software* PlanEx numa escala de 1 (pouco interativo) a 10 (muito interativo) e a justificativa do valor atribuído.

Dos usuários entrevistados, todos atribuíram uma nota, sendo a menor nota 6 e a maior nota 10, e a média de todas as notas foi 8, e 42,86% justificaram a nota atribuída.

Percebe-se que o visual ou *layout* dos *softwares* comerciais são desenvolvidos de forma a atrair os usuários, pois se sabe que um belo visual de uma empresa ou de um *software* é condição necessária para o mercado financeiro. O objetivo do *software* PlanEx é oferecer ferramentas de quimiometria do planejamento de experimentos fatorial gratuitas e que funcionem, e que foram implementadas com uma aparência simples.

A média 8 representa uma interatividade alta e está dentro dos objetivos propostos, pois apesar da plataforma ser de aparência simples conclui-se que ela é interativa. A interatividade inicia-se mesmo na Tela Inicial, quando o usuário deve escolher a quantidade de variáveis e replicatas para poder digitar as respostas experimentais.

Na **Questão 7** o usuário teve a oportunidade de citar a característica mais atrativa do *software* PlanEx de sua opinião, e justificá-la. Dos usuários entrevistados, 28,57% disseram que a característica que lhe mais chamou a atenção foi a agilidade, e o restante, 71,43%, opinaram que a simplicidade de execução foi o que mais lhe chamou a atenção.

Para o presente trabalho, os pontos positivos são oferecer ferramentas para o planejamento de experimentos fatorial que funcione tão bem quanto às ferramentas de outros *softwares*, sendo gratuitas e utilizadas através do acesso da *internet* sem que o usuário necessite fazer o *download* ou ter senha de acesso. Os pontos positivos do *software* PlanEx que a maioria destacou é a simplicidade de execução, pois não requer um conhecimento prévio de uma linguagem de um *software* específico, por exemplo, o *Octave*, o *Scilab* e *Matlab* v R2011b (7.13.0.564). Dos usuários entrevistados, 28,57% disseram que a característica que lhe mais chamou a atenção foi a agilidade, pois a ausência de codificação e de digitação de matrizes e operações de matrizes diminui muito o tempo.

Na última pergunta, **Questão 8**, objetivou saber na opinião do usuário o que menos lhe chamou a atenção no *software* PlanEx fazendo um comentário ao destacar os pontos negativos. Das opiniões negativas sobre o *software*, 14,29% disseram que a opção de gráficos não é suficiente, e que deveria ter opção para gráficos em três dimensões. Porém, como se comentou na Questão 3, no dia da última aplicação, dia que os alunos poderiam usar os gráficos de superfície de resposta em 3D, os gráficos 3D não abriram devido a um erro ou falta da API de *Java* 3D. Acredita-se que essa opinião de 14,29% não seja significativa ou que não existiria se os gráficos fossem gerados.

Outros usuários, também 14,29%, disseram que o *software* PlanEx deveria apresentar o valor do desvio padrão mesmo quando não há replicatas, usando o desvio padrão através de

séries históricas ou através do cálculo dos desvios, porém isso seria praticamente impossível, pois seria necessário arquivar no *software* PlanEx uma quantidade grande de dados de variáveis conhecidas e estudadas ao longo do tempo por vários pesquisadores, tornando o processamento lento do *software*, e frequentemente sem sentido devido ao fato de arquivar dados que não sejam usados.

Outra quantidade, 14,29%, disseram que as tabelas deveriam ser salvas, o que será implementado no futuro. Outra quantidade de 14,29% disse que os planejamentos deveriam ser expandidos para 3 ou 4 níveis, porém na química percebe-se que os problemas são resolvidos usando dois níveis ou três níveis (sendo dois níveis com ponto central), logo não percebe-se a necessidade de tal expansão. E o restante, 14,29% disser que o programa não é chamativo e que não desperta a curiosidade em conhecê-lo, o que é explicado pelo fato do visual simples do *software* e confrontado pelo visual robusto de outros *softwares*.

5. CONCLUSÕES

O *software* PlanEx contém as ferramentas do planejamento de experimentos fatorial em dois níveis com ou sem ponto central desenvolvido no presente trabalho, e apresentou ser uma ferramenta gratuita, rápida, eficiente e amigável de ensino e pesquisa no nível de pós-graduação.

A linguagem de programação *Java* escolhida permite ao usuário do *software* PlanEx a possibilidade de gerar e visualizar gráficos 2D e 3D e a obtenção de efeitos bem como a otimização de planejamentos de experimentos fatorial para modelos lineares e/ou quadráticos através da *internet* e sem fazer *downloads* para o manuseio, pois o *software* rodará no *browser* do computador do próprio usuário e possui uma grande quantidade de biblioteca gráficas livres.

A validação do *software* PlanEx como ferramenta de ensino e pesquisa realizou-se em três etapas. Pelos comentários, críticas e sugestões, percebeu que o *software* PlanEx é eficaz para analisar um problema que envolve os conceitos e particularidades de um planejamento de experimentos fatorial, tanto para a fase de triagem de variáveis como para a fase de otimização. Com base nisso, concluiu-se que o *software* PlanEx é uma ferramenta de pesquisa em nível de pós-graduação, pois através dele obtiveram-se os mesmos resultados em trabalhos realizados com outros *softwares* de uma forma mais rápida e simples.

Através dos conceitos (teoria) trabalhos na disciplina de Quimiometria II, Planejamento e Otimização de Experimentos, ministrada pelo Professor Dr. Anselmo Elcana de Oliveira foi possível com o *software* PlanEx mostrar aos alunos da disciplina uma aplicação (prática) dos conhecimentos adquiridos durante as aulas em problemas fictícios ou reais para **construir o conhecimento** durante o **processo de aprendizagem** da disciplina. Com base nisso, percebeu-se que o *software* PlanEx é uma ferramenta de ensino em nível de pós-graduação, pois a sua utilização requereu dos alunos os conhecimentos adquiridos durante a disciplina, dessa forma o *software* conciliou a teoria com a prática, mostrando ser uma ferramenta de ensino rápida, simples e eficiente.

A evolução do desenvolvimento do *software* realizou-se com a participação dos alunos da disciplina de quimiometria nos anos de 2009, 2010 e 2011. Através das opiniões e críticas obtidas nas três aplicações do *software* pelos alunos, conclui-se que o presente trabalho é um reforço ou uma corroboração para aumentar a divulgação e uso das ferramentas de planejamentos de experimentos fatorial, pois o *layout* simples do *software* tende a atrair jovens pesquisadores.

6. REFERÊNCIAS

ALVES, Lourdes A.; FELIPE, Maria G. A.; SILVA, João B. Almeida E.; SILVA, Silvio S.; PRATA, Arnaldo M. R. **O pré-tratamento da Hemicelulose Hidrosisada do Bagaço de Cana para produção de Xilitol através de *Candida guilliermondii***. *Applied Biochemistry and Biotechnology*, Vol. 70-72, 1998.

ARROIO, Agnaldo; FERREIA, Celeste; REZENDO, Daisy de Brito. **Uso de Modelagem Molecular no Estudo dos Conceitos de Nucleofilicidade e Basicidade**. *Química Nova*, Vol. 34, nº 9, 661-1665, 2011.

BENITE, Anna M. Canavarro; BENITE, Claudio R. Machado; FILHO, Supercil Mendes da Silva. **Cibercultura em Ensino de Química: Elaboração de um Objeto Virtual de Aprendizagem para o Ensino de Modelos Atômicos**. *Química Nova Na Escola*, Vol. 33, Nº 2, MAIO, 2011.

BOX, George E. P.; HUNTER, William G.; HUNTER, J. Stuart. **Statistics for Experimenters**. Estados Unidos: John Wiley e Sons, 1978.

BOX, George E. P.; WILSON, K. B. **On the experimental attainment of optimum conditions** *J. Roy. Statist. Soc., Ser. B*, 1951, 13, 1.

BRUNS, Roy Edward; FAIGLE, José Fernando Gregori. **Quimiometria**. *Química Nova*, Vol. 8, 84-99, 1985.

BRUNS, Roy Edward; NETO, Benício de Barros; SCARMINIO, Ieda Spacino. **Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria**. Campinas: Unicamp, 2001.

BRUNS, Roy Edward; NETO, Benício de Barros; SCARMINIO, Ieda Spacino. **25 Anos de Quimiometria no Brasil**. *Química Nova*, Vol. 29, nº. 6, 1401-1406, 2006.

CALAZANS, Glicia Maria Torres; MELO, I. R.; PIMENTEL, M. F.; LOPES, Carlos Edison. **Application of Fractional Factorial Design to Levan Production by *Zymomonas Mobilis***. *Brazilian Journal of Microbiology*, Vol. 38: 45-5, 2007.

DEITEL, Harvey M.; DEITEL, Paul. J. **Java, como programar**. Tradução de Carlos Arthur Lang Lisboa. Porto Alegre: Elsevier, 2003.

DIAS, Sílvio L. P.; SCOTTI, Rodrigo; LIMA, Eder C.; LIMA; BENVENUTTI, Edilson V.; PIATNICKI, Clarisse M. S.; GUSHIKEM, Yoshitaka; KUBOTA, Lauro T. **Azul De Metileno Imobilizado Na Celulose/TiO₂ e SiO₂/TiO₂ Propriedades Eletroquímicas E Planejamento Fatorial**. *Química Nova*, Vol. 29, nº 2, 208-212, 2006.

DUEWER, David Lee; KOSKINEN, James R.; KOWALSKI, Bruce R. **ARTHUR**, disponível por B. R. Kowalski, Department of Chemistry, BG – 10, University of Washington, Seattle, WA 98195.

FERREIRA, Joelma M.; SWARNAKAR, Ramdayal; SILVA, Flávio L. H. da. **Efeito das fontes de nutrientes sobre a produção de vinagre em escala de bancada, usando-se a metodologia de superfície de resposta.** Revista Brasileira de Engenharia Agrícola e Ambiental, Vol. 9, nº 1, 73-77, 2005.

FERREIRA, Joelma Moraes; SILVA, Flávio Luiz Honorato da; ALSINA, Odelsia Leonor Sanchez; OLIVEIRA, Líbia de Sousa Conrado; CAVALCANTI, Eliane Bezerra; GOMES, Wolia Costa. **Estudo do equilíbrio e cinética da biossorção do Pb²⁺ por Saccharomyces cerevisiae.** Química Nova, Vol. 30, nº 5, 1188–1193, 2007.

FERREIRA, Márcia M. C.; TEÓFILO, Reinaldo F. **Quimiometria II: Planilhas Eletrônicas para Cálculos de Planejamentos Experimentais, Um Tutorial.** Química Nova, Vol. 29, nº 2, 338-350, 2006.

FERREIRA, Vítor F. **As Tecnologias Interativas No Ensino.** Química Nova, Vol. 21 nº 6, 1998.

FIRME, Caio Lima; SANTOS, Anderson Rouge; BARROS, José Celestino. **A Internet Como Fonte De Informação Bibliográfica Em Química.** Química Nova, Vol. 31, nº. 2, 445-451, 2008.

GOUVEIA, Josivanda P. G. de; MOURA, Rodrigo S. F. de; ALMEIDA, Francisco de A. C.; OLIVEIRA, Alexsandra M. de V.; SILVA, Manassés M. da. **Avaliação da cinética de secagem de caju mediante um planejamento experimental.** Revista Brasileira de Engenharia Agrícola e Ambiental, Vol. 6, nº 3, 471-474, 2002.

GUJARATI, Damodar N. **Econometria Básica.** Rio de Janeiro: Elsevier, 2006.

<http://code.google.com/p/jzy3d/>, acessada em janeiro de 2012.

<http://commons.apache.org/>, acessada em janeiro de 2012.

http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf, acessada em Janeiro 2012.

<http://www.chemkeys.com/bra/index.htm>; acessada Julho 2009.

<http://www.guardian.co.uk/business/2007/jan/30/research.highereducation>, acessada em janeiro de 2012.

http://www.java.com/pt_BR/download/faq/java_webstart.xml, acessada em janeiro de 2012.

<http://www.jfree.org/jfreechart/>, acessada em janeiro de 2012.

<http://www.oracle.com/technetwork/java/index.html>, acessada em janeiro de 2012.

<http://www.oracle.com/us/sun/index.htm>, acessada em janeiro de 2012.

<http://www.plos.org/>, acessada em janeiro de 2012.

MANSILLA, Héctor D.; FERNÁNDEZ, J.; KIWI, J.; LIZAMA, C.; FREER, J. **Factorial experimental design of Orange II photocatalytic discolouration.** Journal of Photochemistry and Photobiology A: Chemistry, Vol. 151, 213–219, 2002.

MASHRU, Raja shree C.; SHAH, Punit P.; RANE, Yogesh M.; THAKKAR, Arti. **Planejamento e Otimização de Cloridrato de Mefloquina em micropartículas para mascaramento do sabor amargo.** AAPS PharmSciTech, Vol. 9, nº. 2, Junho 2008.

MEIRA, Marilena; QUINTELLA, Cristina M.; FERRER, Tiago M.; SILVA, Humbervânia R. Gonçalves da; GUIMARÃES, Alexandre K.; SANTOS, Mariana Andrade; NETO, Pedro R. da Costa; PEPE, Iuri M. **Identificação de Adulteração de Biocombustível por Adição de Óleo Residual ao Diesel por Espectrofluorimetria Total 3D e Análise das Componentes Principais.** Quim. Nova, Vol. 34, nº 4, 621-624, 2011.

MULINARI, Daniella Regina; DA SILVA, Gilberto Luís J. P.; DA SILVA, Maria Lúcia Caetano P. **Adsorção de íons dicromato nos compósitos celulose/ZrO₂.nH₂O preparados pelos Métodos da precipitação convencional e em solução homogênea.** Química Nova, Vol. 29, nº. 3, 496-500, 2006.

MUNIZ, Edvani C.; AOUADA, Fauze A.; MOURA, Márcia R. de; GIROTTI, Emerson M.; RUBIRA, Adley F. **Caracterização de Hidrogéis Condutores Constituídos por PAAM e PEDOT/PSS por meio de Planejamento Fatorial.** Polímeros: Ciência e Tecnologia, Vol. 18, nº 2, 126-131, 2008.

NIELSEN, Jakob; LORANGER, Hoa. **Usabilidade na Web: Projetando Websites com Qualidade.** Rio de Janeiro: Elsevier, 2000.

OLIVIER, Samantha; SILVA, Valdinete Lins da; MOTTA, Maurício da; SILVA, José Edson da. **Emprego de planejamento fatorial no desenvolvimento de uma metodologia para Extração de zinco de resíduos galvânicos.** Química Nova, Vol. 30, nº. 7, 1750-1753, 2007.

ORTIZ, Maria Cruz. **Ensino de Quimiometria.** Anal Bioanal Chem, Vol 388, 1557-1560, 2007.

PERALTA-ZAMORA, Patricio; MORAIS, Josmaria Lopes de; NAGATA, Noemi. **Por Que Otimização Multivariada?** Engenharia Sanitária Ambiental. Vol. 10, nº 2, abril -junho, 106-110, 2005.

PEZZA, Leonardo; DORETTO, Keity Margareth; GOTARDO, Mara Andréia; PEZZA, Helena Redigolo. **Desenvolvimento de um Método Reflectométrico Ambientalmente mais Amigável para a determinação de Metoclopramida em Formulações Farmacêuticas.** Química Nova, Vol. 33, nº 2, 461-465, 2010.

RUELA, André Luís Morais; ARAÚJO, Magali Benjamim; PEREIRA, Gislaine Ribeiro. **Desenvolvimento e validação de um método analítico rápido por cromatografia líquida de alta eficiência para determinação de nimesulida em estudos de liberação in vitro.** Química Nova, Vol. 32, nº 1, 165-168, 2009.

SANTOS, Danilo Oliveira; WARTHA, Edson José; FILHO, Juvenal Carolino da Silva. **Softwares educativos livres para o Ensino de Química: Análise e Categorização.** Divisão

de Ensino de Química da Sociedade Brasileira de Química (ED/SBQ) - XV Encontro Nacional de Ensino de Química (XV ENEQ) – Brasília, DF, Brasil – 21 a 24 de julho de 2010.

SHERIDAN, Patrick M.; EDGINTON, Andrea N.; BOERMANS, Herman J.; THOMPSON, Dean G.; HOLT, John D.; STEPHENSON, Gerald R. **A comparação entre dois modelos fatoriais, um fatorial completo 3^3 e um Delineamento Central Composto Rotacional, para uso em experimentos de Resposta Binomial em Toxicologia Aquática.** Arch. Environ. Contam. Toxicol. 46, 216 –223, 2004.

SILBERSCHATZ, Abraham, GALVIN, Peter Baer; GAGNE, Greg. **Sistemas Operacionais: Conceitos e Aplicações.** Rio de Janeiro: Campus, 2004.

SILVA, Elio Medeiros da. **Matemática e estatística aplicada.** São Paulo: Atlas, 1999.

SILVA, José Alexsandro; SILVA, Flávio Luiz Honorato da.; ALVES, Rômulo Romeu da Nóbrega; SANTANA, Davi Pereira de. **Influência das variáveis Nitrogênio, Fósforo, e °Brix na produção de metabólitos secundários contaminantes totais da fermentação alcoólica.** Química Nova, Vol. 29, nº 4, 695-698, 2006.

SPIEGEL, Murray. R. **Probabilidade estatística.** São Paulo: McGraw-Hill, 2004.

STULZER, Hellen Karine; BARBOZA, Fernanda Malaquias; VECCHIA, Débora Dalla; PEREIRA, Airton Vicente Pereira; SILVA, Marcos Antonio Segatto. **Desenvolvimento e validação de um método analítico simples e rápido por espectroscopia UV para quantificação de aciclovir em matrizes hidrofílicas de liberação prolongada.** Química Nova, Vol. 33, nº 3, 747-749, 2010.

APÊNDICE A – Regressão Linear Simples

Esta seção do trabalho apresenta-se de forma detalhada as equações utilizadas para estimar os parâmetros de uma regressão do modelo linear com duas variáveis através do MQO desenvolvido por Carl Friedrich Gauss (GUJARATI, 2006). O modelo com duas variáveis é dado por $y_i = \beta_0 + \beta_1 x_i + u_i$.

O modelo linear mais simples de uma regressão denomina-se modelo binário ou modelo linear com duas variáveis: uma variável dependente (y) e uma variável independente (x). A Eq. A.1 representa o modelo binário populacional, denominado de função de regressão populacional (FRP), em que o índice inferior i indica a resposta y relacionada com a variável x para o nível experimental i , $i = 1, 2, 3, \dots, n$,

$$y_i = \beta_0 + \beta_1 x_i + u_i. \quad (\text{A.1})$$

Onde:

- y_i é uma função de x_i , que se denota por $y_i(x_i)$ do nível experimental i ;
- β_0 e β_1 são os coeficientes populacionais da regressão;
- u_i é o erro aleatório populacional ou resíduo aleatório populacional do nível experimental i .

Um dos objetivos fundamentais da Estatística é trabalhar com amostragem, isso implica que a Eq. A.1 não é observada diretamente. Com base nisso, obtém-se uma função estimada através da amostragem, a qual se denomina de função de regressão amostral (FRA),

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i. \quad (\text{A.2})$$

Onde:

- \hat{y}_i é a FRA que depende de x_i , o símbolo $\hat{}$ significa estimado e lê-se y chapéu;
- $\hat{\beta}_0$ e $\hat{\beta}_1$ são os coeficientes amostrais da regressão.

Ao adicionar erro aleatório amostral, \hat{u}_i , na Eq. A.2 obtém-se a FRP a partir de parâmetros amostrais,

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i. \quad (\text{A.3})$$

Onde:

- y_i é a FRP que depende de x_i ;
- $\hat{\beta}_0$ e $\hat{\beta}_1$ são os coeficientes amostrais da regressão;

- \hat{u}_i é o erro aleatório amostral ou resíduo aleatório amostral.

Substituindo a Eq. A.2 na Eq. A.3 encontra-se o resíduo aleatório amostral em função da FRP e da FRA,

$$\hat{u}_i = y_i - \hat{y}_i. \quad (\text{A.3})$$

Para obter os coeficientes ou parâmetros (Betas) da regressão, usa-se o MQO, que é um dos métodos de análise de regressão mais poderosos e difundidos, e é atribuído a Carl Friedrich Gauss (GUJARATI, 2006).

Para estimar os parâmetros da regressão, opera-se com o quadrado em ambos os lados da Eq. A.3:

$$\hat{u}_i^2 = (y_i - \hat{y}_i)^2. \quad (\text{A.4})$$

Operando com o somatório na Eq. A.4 obtém-se:

$$\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (\text{A.5})$$

Substituindo a Eq. A.2 na Eq. A.5, encontra-se:

$$\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2. \quad (\text{A.6})$$

O MQO escolhe $\hat{\beta}_0$ e $\hat{\beta}_1$ de modo a garantir que $\sum_{i=1}^n \hat{u}_i^2$ seja o menor possível. Para obter $\hat{\beta}_0$ e $\hat{\beta}_1$ deve-se minimizar $\sum_{i=1}^n \hat{u}_i^2$ na Eq. A.6 através dos conceitos do cálculo diferencial. A derivada de primeira ordem de uma função igual à zero resulta no ponto de máximo ou mínimo de uma função. Derivando a Eq. A.6 em função do $\hat{\beta}_0$ obtém-se:

$$\begin{aligned} \frac{\partial \sum_{i=1}^n \hat{u}_i^2}{\partial \hat{\beta}_0} &= 0 \\ \frac{\partial}{\partial \hat{\beta}_0} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 &= 0 \\ \sum_{i=1}^n 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 (-1) &= 0 \end{aligned} \quad (\text{A.7})$$

Derivando a Eq. A.6 em função do $\hat{\beta}_1$, obtém-se:

$$\begin{aligned} \frac{\partial \sum_{i=1}^n \hat{u}_i^2}{\partial \hat{\beta}_1} &= 0 \\ \frac{\partial}{\partial \hat{\beta}_1} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 &= 0 \\ \sum_{i=1}^n 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 (-x_i) &= 0 \end{aligned} \quad (\text{A.8})$$

Isolando $\hat{\beta}_0$ na Eq. A.7, obtém-se:

$$\begin{aligned} \sum_{i=1}^n 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 (-1) &= 0 \\ -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 &= 0 \\ \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\beta}_0 - \sum_{i=1}^n \hat{\beta}_1 x_i &= \frac{0}{-2} \\ \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\beta}_0 - \sum_{i=1}^n \hat{\beta}_1 x_i &= 0 \\ \sum_{i=1}^n \hat{\beta}_0 &= \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\beta}_1 x_i \end{aligned}$$

Percebe-se que $\sum_{i=1}^n \hat{\beta}_0 = n\hat{\beta}_0$ e $\sum_{i=1}^n \hat{\beta}_1 x_i = \hat{\beta}_1 \sum_{i=1}^n x_i$. Substituindo e realizando algumas operações

na $\sum_{i=1}^n \hat{\beta}_0 = \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\beta}_1 x_i$, obtém-se:

$$\begin{aligned} \sum_{i=1}^n \hat{\beta}_0 &= \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\beta}_1 x_i \\ n\hat{\beta}_0 &= \sum_{i=1}^n y_i - \hat{\beta}_1 \sum_{i=1}^n x_i \\ \hat{\beta}_0 &= \frac{\sum_{i=1}^n y_i}{n} - \frac{\hat{\beta}_1 \sum_{i=1}^n x_i}{n} \end{aligned}$$

Acrescentando $\sum_{i=1}^n x_i$ no numerador e denominador do lado direito de $\hat{\beta}_0$, obtém-se:

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i \cdot \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i} - \frac{\hat{\beta}_1 \sum_{i=1}^n x_i \cdot \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i}$$

Usando a propriedade do somatório que $\sum_{i=1}^n x_i = nx_i$ na equação acima, obtém-se:

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i \cdot nx_i}{n \sum_{i=1}^n x_i} - \frac{\hat{\beta}_1 \sum_{i=1}^n x_i \cdot nx_i}{n \sum_{i=1}^n x_i}$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i \cdot x_i}{\sum_{i=1}^n x_i} - \frac{\hat{\beta}_1 \sum_{i=1}^n x_i \cdot x_i}{\sum_{i=1}^n x_i}$$

Usando a propriedade do somatório que $x_i = \frac{\sum_{i=1}^n x_i}{n}$ na equação acima, obtém-se:

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i \cdot \frac{\sum_{i=1}^n x_i}{n}}{\sum_{i=1}^n x_i} - \frac{\hat{\beta}_1 \sum_{i=1}^n x_i \cdot \frac{\sum_{i=1}^n x_i}{n}}{\sum_{i=1}^n x_i}$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i}{n} - \frac{\hat{\beta}_1 \sum_{i=1}^n x_i}{n}$$

Percebe que $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ e $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$. Substituindo, obtém-se:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \quad (\text{A.9})$$

Isolando $\hat{\beta}_0$ na Eq. A.8 obtém-se:

$$\sum_{i=1}^n 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 (-x_i) = 0$$

$$-2 \left(\sum_{i=1}^n y_i x_i - \sum_{i=1}^n \hat{\beta}_0 x_i - \sum_{i=1}^n \hat{\beta}_1 x_i^2 \right) = 0$$

$$\sum_{i=1}^n y_i x_i - \sum_{i=1}^n \hat{\beta}_0 x_i - \sum_{i=1}^n \hat{\beta}_1 x_i^2 = \frac{0}{-2}$$

$$\sum_{i=1}^n y_i x_i - \sum_{i=1}^n \hat{\beta}_0 x_i - \sum_{i=1}^n \hat{\beta}_1 x_i^2 = 0$$

$$\sum_{i=1}^n \hat{\beta}_0 x_i = \sum_{i=1}^n y_i x_i - \sum_{i=1}^n \hat{\beta}_1 x_i^2$$

$$\hat{\beta}_0 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i x_i - \sum_{i=1}^n \hat{\beta}_1 x_i^2$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i} - \frac{\sum_{i=1}^n \hat{\beta}_1 x_i^2}{\sum_{i=1}^n x_i}$$

Com as duas equações do termo $\hat{\beta}_0$ obtidas pelas Eq. A.7 e A.8 chega-se ao sistema a seguir:

$$\left\{ \begin{array}{l} \hat{\beta}_0 = \frac{\sum_{i=1}^n y_i}{n} - \frac{\hat{\beta}_1 \sum_{i=1}^n x_i}{n} \\ \hat{\beta}_0 = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i} - \frac{\sum_{i=1}^n \hat{\beta}_1 x_i^2}{\sum_{i=1}^n x_i} \end{array} \right.$$

Resolvendo o sistema por igualdade, obtém-se:

$$\hat{\beta}_0 = \hat{\beta}_0$$

$$\frac{\sum_{i=1}^n y_i}{n} - \frac{\hat{\beta}_1 \sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i} - \frac{\sum_{i=1}^n \hat{\beta}_1 x_i^2}{\sum_{i=1}^n x_i}$$

$$\frac{\sum_{i=1}^n \hat{\beta}_1 x_i^2}{\sum_{i=1}^n x_i} - \frac{\hat{\beta}_1 \sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i} - \frac{\sum_{i=1}^n y_i}{n}$$

Tirando o mínimo múltiplo comum (m.m.c.) da equação acima, obtém-se:

$$\frac{n \sum_{i=1}^n \hat{\beta}_1 x_i^2}{n \sum_{i=1}^n x_i} - \frac{\hat{\beta}_1 \left(\sum_{i=1}^n x_i \right)^2}{n \sum_{i=1}^n x_i} = \frac{n \sum_{i=1}^n y_i x_i}{n \sum_{i=1}^n x_i} - \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i}$$

$$n \hat{\beta}_1 \sum_{i=1}^n x_i^2 - \hat{\beta}_1 \left(\sum_{i=1}^n x_i \right)^2 = n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i$$

Colocando o $\hat{\beta}_1$ em evidência, obtém-se:

$$\hat{\beta}_1 \left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) = n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i$$

Isolando $\hat{\beta}_1$, obtém-se a Eq. A.10 para calcular o coeficiente angular do modelo binário:

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right)} \quad (\text{A.10})$$

As Eq. A.9 e A.10 são equivalentes a Eq. 13.

APÊNDICE B - Regressão Linear Multivariada

Como se comentou, esta parte do trabalho destina-se a apresentar de forma detalhada a equação para obter os coeficientes de uma regressão linear multivariada através do modelo com k variáveis (Y e X_2, X_3, \dots, X_k) em notação de álgebra matricial. O modelo de k variáveis é uma extensão lógica dos modelos de duas e três variáveis examinados anteriormente. Portanto, este apêndice apresenta poucos conceitos novos, exceto a notação de matrizes. Recomenda-se aos leitores que não tiverem familiarizados com a álgebra matricial, rever tais conceitos através de um bom livro da área.

A vantagem da álgebra matricial sobre a álgebra escalar é que ela oferece um método compacto para lidar com modelos de regressão envolvendo muitas variáveis; uma vez formulado um modelo de k variáveis, a solução se aplica a uma, duas, três ou qualquer número de variáveis (GUJARATI, 2006).

B.1. Modelo de Regressão Linear com k variáveis

Ao generalizar os modelos de regressão linear de duas ou três variáveis apresentados na seção da Estatística Experimental obtém-se o modelo de regressão da população com k variáveis (FRP), envolvendo a variável dependente y e $k-1$ variáveis independentes, x_2, x_3, \dots, x_k da seguinte forma:

$$\text{FRP: } y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_k x_{ik} + u_i \quad i = 1, 2, \dots, n. \quad (\text{B.1})$$

Onde:

- β_1 é o coeficiente independente e se denomina de intercepto da regressão;
- β_2 até β_k são os coeficientes angulares parciais;
- u_i é o erro aleatório ou estocástico da $i = i$ -ésima observação, sendo n o tamanho da população.

Nota. No modelo binário usou-se o símbolo β_0 para representar o intercepto ou coeficiente independente: $y_i = \beta_0 + \beta_1 x_i + u_i$. No modelo multivariado,

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_k x_{ik} + u_i$$

Tabela B.1 - Despesas Familiares de Consumo Semanal Y em função da Renda Familiar Semanal X.

i	X (R\$)	Y (R\$)
1	80	70
2	100	65
3	120	90
4	140	95
5	160	110
6	180	115
7	200	120
8	220	140
9	240	155
10	260	150

Como uma ilustração, aplicou-se a Eq. B.5 nos dados da Tabela B.1 para visualizar com uma maior didática a dimensão das matrizes envolvidas na Eq. B.5 submetida nos dados tabelados, como se apresenta na Eq. B.6.

$$\begin{bmatrix} 70 \\ 65 \\ 90 \\ 95 \\ 110 \\ 115 \\ 120 \\ 140 \\ 155 \\ 150 \end{bmatrix} = \begin{bmatrix} 1 & 80 \\ 1 & 100 \\ 1 & 120 \\ 1 & 140 \\ 1 & 160 \\ 1 & 180 \\ 1 & 200 \\ 1 & 220 \\ 1 & 240 \\ 1 & 260 \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \\ u_9 \\ u_{10} \end{bmatrix}. \quad (\text{B.6})$$

Assim como nos casos de duas e três variáveis, o objetivo de uma regressão é estimar os parâmetros da regressão linear múltipla, Eq. B.1, e obter inferências sobre eles a partir dos dados tabelados. Em notação matricial, isso significa calcular a matriz dos parâmetros da regressão, β , e obter inferências sobre eles.

Para obter os coeficientes da regressão, β , usa-se o método dos mínimos quadrados ordinários (MQO) ou método da máxima verossimilhança (MV). Segundo Gujarati (2006), esses dois métodos produzem resultados estimados idênticos para os coeficientes de regressão, por isso limita-se neste trabalho ao método dos MQO.

B.2 Estimativas dos Coeficientes de uma Regressão Linear Múltipla pelo método dos Mínimos Quadrados Ordinários (MQO)

A função de regressão amostral, FRA, para k variáveis apresenta-se na forma a seguir:

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \dots + \hat{\beta}_k x_{ik} + \hat{u}_i. \quad (\text{B.7})$$

Rotineiramente, escreve-se a Eq. B.7 de modo mais compacto em notação matricial como:

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}. \quad (\text{B.8})$$

E em forma de matriz como:

$$\begin{array}{c} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ n \times 1 \end{array} = \begin{array}{c} \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix} \\ n \times k \end{array} \cdot \begin{array}{c} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{bmatrix} \\ k \times 1 \end{array} + \begin{array}{c} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \end{bmatrix} \\ n \times 1 \end{array} \quad (\text{B.9})$$

Da mesma forma como se apresentou no **Apêndice A**, estimam-se os coeficientes da regressão para de k variáveis através do método dos MQO minimizando o somatório dos resíduos ao quadrado usando o cálculo diferencial:

$$\sum \hat{u}_i^2 = \sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{i2} - \dots - \hat{\beta}_k x_{ik})^2. \quad (\text{B.10})$$

Onde $\sum \hat{u}_i^2$ é a soma dos quadrados dos resíduos ($SQ_{\text{Resíduos}}$), como se apresentou na seção de ANOVA. Em notação matricial, isso equivale a minimizar $\hat{\mathbf{u}}^t \hat{\mathbf{u}}$, pois:

$$\hat{\mathbf{u}}^t \hat{\mathbf{u}} = \begin{bmatrix} \hat{u}_1 & \hat{u}_2 & \cdots & \hat{u}_n \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix} = \hat{u}_1^2 + \hat{u}_2^2 + \dots + \hat{u}_n^2 = \sum \hat{u}_i^2. \quad (\text{B.11})$$

Ao isolar o resíduo aleatório na Eq. B.8, obtém-se:

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}. \quad (\text{B.12})$$

Por conseguinte,

$$\begin{aligned} \hat{\mathbf{u}}^t \hat{\mathbf{u}} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^t (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ \hat{\mathbf{u}}^t \hat{\mathbf{u}} &= \mathbf{y}^t \mathbf{y} - 2\hat{\boldsymbol{\beta}}^t \mathbf{X}^t \mathbf{y} + \hat{\boldsymbol{\beta}}^t \mathbf{X}^t \mathbf{X} \hat{\boldsymbol{\beta}}. \end{aligned} \quad (\text{B.13})$$

Na Eq. B.16 as quantidades conhecidas são: $(\mathbf{X}^t\mathbf{X})$ e $\mathbf{X}^t\mathbf{y}$, ou seja, pretende-se calcular $\hat{\beta}$. Aplicando a álgebra linear, se a inversa de $(\mathbf{X}^t\mathbf{X})$ existe, que se denota por $(\mathbf{X}^t\mathbf{X})^{-1}$, então multiplicando ambos os lados da Eq. B.16 por essa inversa, obtém-se:

$$(\mathbf{X}^t\mathbf{X})^{-1}(\mathbf{X}^t\mathbf{X})\hat{\beta} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}$$

Mas como $(\mathbf{X}^t\mathbf{X})^{-1}(\mathbf{X}^t\mathbf{X}) = \mathbf{I}$, isto é, uma matriz identidade de ordem $k \times k$, obtém-se:

$$\begin{aligned} \mathbf{I}\hat{\beta} &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}; \\ \hat{\beta}_{k \times 1} &= (\mathbf{X}^t\mathbf{X})_{k \times k}^{-1} \mathbf{X}_{k \times n}^t \mathbf{Y}_{n \times 1}. \end{aligned} \quad (\text{B.17})$$

A Eq. B.17 é um resultado fundamental da teoria dos MQO na notação matricial.

Como uma ilustração, aplicou-se a Eq. B.17 nos dados da Tabela B.1 para obter ou estimar os coeficientes da regressão do modelo linear binário:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \quad \mathbf{X}^t\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

$$\mathbf{X}^t\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$\mathbf{X}^t\mathbf{X} = \begin{bmatrix} 10 & 1700 \\ 1700 & 322000 \end{bmatrix} \quad \mathbf{X}^t\mathbf{y} = \begin{bmatrix} 1110 \\ 205500 \end{bmatrix}$$

Usando as regras da inversão de matrizes através dos conceitos de álgebra linear, obtém-se:

$$(\mathbf{X}^t\mathbf{X})^{-1} = \begin{bmatrix} 0,97576 & -0,005152 \\ -0,005152 & 0,0000303 \end{bmatrix}$$

Por conseguinte, tem-se:

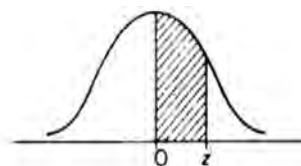
$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 0,97576 & -0,005152 \\ -0,005152 & 0,0000303 \end{bmatrix} \begin{bmatrix} 1110 \\ 205500 \end{bmatrix}$$
$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 24,4545 \\ 0,5079 \end{bmatrix}$$

O modelo estimado é $\hat{y}_i = 24,4545 + 0,5079x_1$.

No caso de considerar um modelo quadrático ou um modelo linear completo, basta acrescentar na matriz \mathbf{X} os termos quadráticos ou o produto de cada variável, como se apresentou na seção da metodologia de superfície de resposta, RSM.

APÊNDICE C – Tabela da Distribuição de Probabilidade

Normal



A tabela a seguir representa a probabilidade de encontrar uma variável x que se distribui normalmente que foi transformada na variável z padronizada com média zero e desvio padrão no intervalo de $0 \leq P(z) \leq z$. A área a esquerda é simétrica a área da direita.

Com base nisso: $0 \leq P(z) \leq z = -z \leq P(z) \leq 0 = 0 < P(z) < z = -z < P(z) < 0$.

z	Segunda casa decimal de Z									
	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986

APÊNDICE D – Questionário da Terceira Aplicação do *Software* PlanEx

Pergunta 1. As ferramentas do planejamento fatorial completo do *software* PlanEx são suficientes para analisar um problema?

Sim (___)

Não (___)

Comente/Justifique.

Pergunta 2. As ferramentas do planejamento fatorial fracionário do *software* PlanEx são suficientes para analisar um problema?

Sim (___)

Não (___)

Comente/Justifique.

Pergunta 3. As ferramentas da metodologia de superfície de resposta do *software* PlanEx são suficientes para analisar um problema?

Sim (___)

Não (___)

Comente/Justifique.

Pergunta 4. Há softwares que apresenta característica de amigabilidade. O que Você acha do *software* PlanEx em relação a essa característica?

Pergunta 5. Você acha que este *software* é mais ágil do que outros softwares? Faça um comentário.

Pergunta 6. Estabeleça um grau de interatividade desse *software*. Coloque em uma escala de 1 (pouco interativo) a 10 (muito interativo). Justifique

Pergunta 7. O que mais lhe chamou a atenção no *software* PlanEx? Faça um comentário abordando os pontos positivos.

Pergunta 8. O que menos lhe chamou a atenção no *software* PlanEx? Faça um comentário abordando os pontos negativos.