

**UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA
MARIA ELENA SANTOS TAQUEDA**

**A ESTATÍSTICA COMO CATALISADOR DO CONHECIMENTO NO
ESTUDO E DIAGNÓSTICO DE PROCESSOS**

**São Paulo
2009**

**CONSULTA
FT-2713**

OK

MARIA ELENA SANTOS TAQUEDA

**A ESTATÍSTICA COMO CATALISADOR DO
CONHECIMENTO NO ESTUDO E DIAGNÓSTICO DE
PROCESSOS**

Texto sistematizado apresentado à Escola Politécnica da Universidade de São Paulo no concurso visando a obtenção do título de Livre-Docente, junto ao Departamento de Engenharia Química, na área “Separações Térmicas e Mecânicas”.

Defendido em 17/6/09.

São Paulo

2009

MARIA ELENA SANTOS TAQUEDA

**A ESTATÍSTICA COMO CATALISADOR DO
CONHECIMENTO NO ESTUDO E DIAGNÓSTICO DE
PROCESSOS**

Texto sistematizado apresentado à Escola Politécnica da Universidade de São Paulo no concurso visando a obtenção do título de Livre-Docente, junto ao Departamento de Engenharia Química, na área “Separações Térmicas e Mecânicas”.

São Paulo

2009

À meu esposo Roberto, a meus filhos Ana
Cristina, João Carlos, Luiz Augusto, Dina Lúcia e
Carolina a meu neto Arthur, e a minha irmã
Eliana (*In Memoriam*) com amor e gratidão pelo
apoio constante ao meu trabalho.

AGRADECIMENTOS

À Deus, por estar sempre presente em minha vida ajudando, iluminando e me capacitando. Tenho muito a agradecer.

À meus pais Francisca e João (*In Memoriam*) que não mediram esforços na dedicação à educação dos filhos, com amor e disciplina, agradeço.

À meu esposo, que soube cumprir o juramento do altar: “na alegria e na tristeza, na saúde e na doença”, meu companheiro de quase 40 anos que tem sabido acompanhar minha vida acadêmica com amor e dedicação.

À meus filhos agradeço e peço desculpas de coração por muitas vezes não os ter acompanhado como devia, passo a passo, o desenvolvimento de cada um. Mas juro, filhos! Eu os amo!

À minhas irmãs, Eleide, Stela, Maria Lúcia e Eliana (*In Memoriam*), que são minhas maiores amigas e confidentes, estiveram sempre presentes em minha vida como um todo, agradeço com muito amor.

Ao meu colega Wilson Miguel Salvagnini, pela grande parceria e cumplicidade no LSTM com nossos alunos no desenvolvimento dos trabalhos de graduação e pós-graduação. Não esqueça: estou sempre aprendendo com você. Por seu apoio e amizade, meu muito obrigada.

Ao meu colega José Luis Pires Camacho, com quem tenho uma relação de mãe para filho, agradeço por todo o apoio e também pelas sugestões em minha vida acadêmica.

Aos professores Maurício Torloni pela acolhida como orientador de doutorado e Priscila Aya Shimizu (*In Memoriam*) por acreditar que eu tinha algo a contribuir como docente da EPUSP.

Aos meus alunos de pós-graduação, como aos meus filhos, peço desculpas por muitas vezes não lhes ter dado a assistência devida, principalmente neste momento em que tenho ficado mais ausente. Obrigada por compreenderem este meu momento e pela ajuda concreta no desenvolvimento de seus trabalhos. Isto me permitiu a realização deste.

À minha amiga Profa. Idalina Vieira Aoki, com quem tenho trabalhado em vários projetos, por confiar no potencial dos métodos estatísticos e me proporcionar a oportunidade de aplicá-los em vários trabalhos científicos, além das sugestões dadas na confecção deste texto meu sincero agradecimento.

Aos amigos Isabel e Hercílio, com quem tenho produzido alguns trabalhos, empregando planejamentos experimentais, obrigada pela confiança.

Aos meus colegas de outros departamentos e outras unidades da USP e de outras universidades que me ajudaram com suas valiosas parcerias a construir a minha produção científica, meus sinceros agradecimentos.

À Elisete, que começou aqui bem jovem, pelo exemplo que é de profissionalismo e dedicação meu muito obrigada.

À Sônia, por seu apoio com as disciplinas, e especialmente agora, na elaboração deste texto, tenho muito a agradecer.

À Denise pelo apoio na organização de documentos, pelo apoio no Laboratório com nossos(as) alunos(as), o meu muito obrigada.

Aos demais colegas do LSTM, Tah, Pacheco, Paiva e Magnani (*In Memoriam*), obrigada por todas as contribuições à minha carreira.

Ao Prof. Terron, com quem iniciei trabalhando na pesquisa em 1988, no LABIEQ, obrigada.

Aos funcionários, Graça, Alexandre e Terezinha, agradeço pelo apoio às minhas necessidades profissionais, obrigada.

Ao Engenheiro Caio, pela sua disponibilidade em ajudar sempre no que é necessário, obrigada.

Às funcionárias da biblioteca Maria Lúcia e Fátima, por todo o apoio a mim e aos meus alunos em nossas necessidades e dificuldades no resgate de material bibliográfico, obrigada.

À Neida, ao Fernando e ao Sérgio por estarem sempre disponíveis em todos os momentos, o meu muito obrigada.

À Escola Politécnica da Universidade de São Paulo especialmente ao Departamento de Engenharia Química que possibilitou o meu desenvolvimento profissional agradeço por tudo que representa em minha vida profissional.

Aos órgãos oficiais de fomento à pesquisa, FAPESP, CNPq, CAPES e FINEP, agradeço pelo auxílio financeiro, sem o qual não teria chegado até aqui.

"Feliz aquele que transfere o que sabe e aprende o que ensina"
Cora Coralina

RESUMO

Este texto apresenta a experiência de ensino e de uso de métodos estatísticos aplicados ao desenvolvimento e análise de experimentos. Evidencia a importância do emprego por meio da apresentação de algumas técnicas como a análise de variância, (ANOVA), para efeitos fixos (um fator, dois fatores, fatorial com 2 fatores, dentre outros) e ANOVA para efeitos aleatórios (os componentes da variância), seguido da aplicação de um caso prático intitulado, um ciclone lavador adaptado de um ciclone tipo americano. A proposta deste caso foi analisar o desempenho de um ciclone lavador de gases industriais modificado pela instalação de bicos nebulizadores em seu corpo cilíndrico para umidificar a poeira e coletar o material particulado, mantendo a portabilidade de um ciclone tipo americano original, porém apresentando eficiência de coleta superior. Os projetos fatoriais completos em dois níveis foram apresentados como uma excelente opção de medida de efeitos de fatores, usado na base do método seqüencial. Dois casos de estudo realizados na indústria aplicando estes planejamentos foram apresentados, a saber: melhoria do processo de produção de filmes de polipropileno usando um projeto fatorial 2^4 e aplicando EVOP (Evolutionary Operation) ao processo de fabricação de acrilamida. Estes casos visavam a melhoria dos processos industriais em questão. Em ambos os casos houve um aumento de produtividade e melhoria da qualidade dos produtos. Em seguida vieram os projetos fracionados em dois níveis como uma alternativa econômica para seleção de variáveis quando se suspeita de que mais de 4 fatores afetam o fenômeno. Neste capítulo abordaram-se dois exemplos do bom uso destes planejamentos, encorajando os pesquisadores a ganhar o conhecimento necessário usando tempo e custo da experimentação mínimos. Lembrando que a escolha do projeto e a interpretação dos resultados são essenciais ao estudo. A metodologia da superfície de resposta e noções de regressão linear múltipla, como técnicas de otimização de uma resposta de interesse. O estudo de caso aqui apresentado abordou a destilação azeotrópica como uma alternativa na rota de purificação da glicerina bruta, oriunda do biodiesel, visando menor consumo de energia. A rota como um todo está sendo otimizado no LSTM e está em processo de patenteamento. Finalmente está colocada a opinião da autora, baseada também em opiniões de outros autores, quanto aos métodos estatísticos no ensino de engenharia e o seu uso na pesquisa científica e tecnológica.

Palavras-chave: Análise de variância, planejamento experimental, planejamentos fracionados em 2 níveis, metodologia da superfície de resposta

ABSTRACT

The text presents the experience of teaching and using the statistical methods applied to development and analysis of experiments. It makes clear the importance of its employ through the presentation of some techniques such as the analysis of variance, (ANOVA), of fixed effects model (one-way, two-way, two-way (factorial) design, among others) and ANOVA of the random effects (the components of variance), following the use for an actual case entitled the cyclone washer adapted from an American-type cyclone separator. The proposal of this case was to analyze the performance of a cyclone washer of the industrial gases modified by the installation of spray nozzles at its cylindrical body in order to humidifying the dust, collecting the particulate material, keeping the portability of an original American type cyclone, presenting, however, a superior collecting efficiency. The factorial designs at two-levels were presented as an excellent option to measure effects of one or more variables on a response used at the basis of the sequential method. Two study cases performed at the industry using these designs were presented: improvement of a polypropylene film production process using a 2^4 factorial design and using EVOP (Evolutionary Operation) to the acrylamide production process. Theses cases aimed the improvement of the involved industrial processes. In both cases there were a productivity increase and a quality improvement of the products. In sequence came the fractional factorial design at two levels as an economical alternative in factor screening experiments when there is the suspicion that more than 4 factors affect the phenomenon. In this chapter two examples of positive use of these designs were broached, encouraging the researchers to get the required knowledge, spending minimal experimentation time and cost. It is important to reminding that the design's choice and the results' interpretation are essential for the comprehension of study. The response surface methodology and notions of multiple linear regression are discussed as optimization techniques on a response of interest. The case study here presented broached the azeotrope distillation as an alternative route for the purification of the raw glycerin, deriving from biodiesel, aiming smaller energy consumption. The route as a whole is under optimization at LSTM and the patent process progresses. Finally the view of the author is placed, also based in the view of other authors, in relation to statistical methods at the engineering teaching and its use at the scientific and technological research.

Key-words: analysis of variance, experimental design, 2-level fractional factorial designs at two levels, response surface methodology

LISTA DE ILUSTRAÇÕES

Figura 1.	Diagrama de pontos dos dados originais para os métodos de ensino.....	41
Figura 2.	Box plot dos dados originais para os métodos de ensino	42
Figura 3.	Resíduos em função dos valores ajustados para o exemplo da experiência de ensino.....	48
Figura 4.	Resíduos em função da ordem padrão dos ensaios para o exemplo da experiência de ensino	48
Figura 5.	Diagrama de pontos dos resíduos para o exemplo da experiência de ensino	49
Figura 6.	Geometria da análise de variância (ANOVA One-way)	50
Figura 7.	Geometria da ANOVA do projeto com blocos randomizados	52
Figura 8	Esquema de montagem de um experimento hierárquico $k \times 2 \times 4$, para estimar os componentes da variância.....	59
Figura 9.	Distribuição do tamanho das partículas de quartzo micronizado ($95\% < 20 \mu\text{m}$), usado como poeira neste estudo.....	64
Figura 10.	Efeito da interação posição da chaminé×aquecedores ativos na diferença de temperatura entre a entrada (após aquecimento) e a saída do gás. $F_{8,32} = 9,82$ e $p = 0,00000$. Barras verticais denotam os intervalos de confiança	70
Figura 11.	Efeito da interação número de bicos abertos×aquecedores ativos na diferença de temperatura entre a entrada (após aquecimento) e a saída do gás. $F_{16,32} = 9,82$ e $p = 0,00000$. Barras verticais denotam os intervalos de confiança	71
Figura 12.	Efeito da posição da chaminé, dos aquecedores ativos e do número de bicos abertos na diferença de temperatura entre a entrada (após aquecimento) e a saída do gás. $F_{2,32} = 9,71$; $p = 0,00051$; $F_{4,32} = 2150$; $p = 0,00000$; $F_{4,32} = 52,34$; $p = 0,00000$. Barras verticais denotam os intervalos de confiança	71
Figura 13.	Diagramas de testes de adequação dos resultados ao modelo proposto. (a) Resíduos em escala de probabilidade normal e (b) valores preditos em função dos valores observados.....	72
Figura 14.	Efeito da concentração de particulado na eficiência de coleta do ciclone lavador	75

Figura 15. Representação geométrica do projeto fatorial 2^3 , o exemplo dos poluentes.....	86
Figura 16. Representação gráfica dos efeitos em escala de probabilidade normal, para o exemplo dos poluentes.....	94
Figura 17. A interação de dois fatores para o exemplo do poluente, vista com dois métodos gráficos diferentes	95
Figura 18. Efeitos da resposta Opacidade (y_1) traçados em (a) Pareto, (b) gráfico em escala de probabilidade Normal	106
Figura 19. Efeitos da resposta TRDM (y_2) traçados em (c) Pareto, (d) gráfico em escala de probabilidade Normal	107
Figura 20. Efeitos da resposta TRDT (y_3) traçados em (e) Pareto, (f) gráfico de probabilidade Normal	107
Figura 21. Efeitos da resposta ALDM (y_4) traçados em (g) Pareto, (h) gráfico de probabilidade Normal	108
Figura 22. Efeitos da resposta ALDT (y_5) traçados em (i) Pareto, (j) gráfico em escala de probabilidade Normal	108
Figura 23. Efeitos da resposta EDDM (y_6) traçados em (l) Pareto, (m) gráfico em escala de probabilidade Normal	109
Figura 24: Efeitos da resposta EDDT (y_7) traçados em (n) Pareto, (o) gráfico em escala de probabilidade Normal	109
Figura 25. Linhas de nível das superfícies de rendimento, a) escala reduzida, e b) escala de produção. Fonte: (BOX e DRAPER, 1969).....	123
Figura 26. Informações obtidas de uma planta petroquímica de um programa EVOP, com duas variáveis ao final das fases I, II e III com objetivo de redução de custo por tonelada (fonte: BOX; HUNTER, W, HUNTER, J. 1978).....	127
Figura 27. Esquema EVOP para fase 1 da otimização da produção de acrilamida a partir de acrilonitrila.....	131
Figura 28. Ficha utilizada para coleta de dados de condições operacionais na produção de acrilamida.....	134
Figura 29. Folha de marcha com os dados das corridas do ciclo 13 da fase 1	135

Figura 30. Os efeitos representados em diagramas para o DOE 2^2 , do estudo. No exemplo: (a) efeito da pressão; (b) efeito da relação volumétrica: água/acrilonitrila; (c) interação dos efeitos A e B	138
Figura 31. Resultado extraído da folha de marcha após o ciclo 13 da fase 1.....	140
Figura 32. Esquemas para as alterações para fase seguinte de um plano EVOP, mostrando que os pontos devem ser contíguos.....	141
Figura 33. Seção transversal de uma superfície de resposta. Mudança na média indica curvatura. (Fonte: JURAN, 1993).....	142
Figura 34. Resultado do EVOP da otimização de produção da acrilamida após 6 ciclos da fase 1, com o lote de catalisador K561. Os resultados são em tempo de reação expressos em minutos.....	146
Figura 35. Resultado do EVOP da otimização de produção da acrilamida após 9 ciclos da fase 1, com o lote de catalisador 576. Os resultados são em tempo de reação expressos em minutos.....	148
Figura 36. Resultado do EVOP da otimização de produção da acrilamida após 15 ciclos da fase 1. Os resultados são em tempo de reação expressos em minutos.....	150
Figura 37. Representação esquemática do EVOP para processo de fabricação de acrilamida. As condições 1 a 5 são para fase 1 e as condições de 1' a 5' são para fase 2.....	153
Figura 38. Resultado final do programa EVOP para otimização de produção de acrilamida. A variável objetiva é o tempo de reação em minutos.....	155
Figura 39. Arranjo de um projeto fatorial 2^3 em dois blocos de tamanho 4. Fonte: (BOX, HUNTER W., HUNTER, J. 2005).....	174
Figura 40. Representação gráfica dos efeitos gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta rendimento de fitocianina do projeto fatorial fracionado 2^{4-1}	187
Figura 41. Representação gráfica dos efeitos em gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta L^* do projeto fatorial fracionado 2^{4-1}	187
Figura 42. Representação gráfica dos efeitos em gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta do projeto fatorial fracionado 2^{4-1} .	188

Figura 43. Representação gráfica dos efeitos em gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta hab do projeto fatorial fracionado 2^{4-1}	188
Figura 44. Representação gráfica dos efeitos gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta rendimento de fitocianina do projeto fatorial fracionado 2^4	189
Figura 45. Representação gráfica dos efeitos em gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta L^* do projeto fatorial fracionado 2^4	189
Figura 46. Representação gráfica dos efeitos em gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta do projeto fatorial fracionado 2^4	190
Figura 47. Representação gráfica dos efeitos em gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta hab do projeto fatorial fracionado 2^4	190
Figura 48. Esquema da aparelhagem utilizada na destilação azeotrópica da mistura glicerina-água montado no LSTM montado no LSTM.....	213
Figura 49. Representação gráfica (a) dos resíduos em função dos valores preditos e (b) valores preditos em função dos valores observados para verificação da adequação do modelo para da destilação azeotrópica do projeto de face centrada	222
Figura 50 - Linhas de contorno da concentração final da glicerina predita pelo modelo da equação 53 obtido com os dados do projeto de face centrada	224
Figura 51. Ilustração da interação entre a aplicação de um método estatístico e o desenvolvimento ou a melhoria de um processo ou atividade. (fonte: BOX et al 1978).....	230
Figura 52. Esquema mostrando como a Estatística Descritiva, a Amostragem e o Cálculo de Probabilidades são importantes na análise por meio da Estatística Indutiva.....	231
Figura 53 Diagrama de representação dos diversos tipos de variável.....	233
Figura 54 Exemplo de histograma, traçado para uma variável genérica.....	235
Figura 55. Exemplo de figuras com aspectos diferentes quanto à simetria.....	240
Figura 56. Exemplo de distribuições com diferentes curtoses.....	241

Figura 57. Distribuição normal com indicação das áreas das caudas com $x > \eta + 1\sigma$ e $x > \eta + 2\sigma$	244
Figura 58. Distribuição normal padrão nas duas formas mais comuns de apresentação, função densidade probabilidade (PDF) e a função distribuição probabilidade acumulada (CDF).....	245
Figura 59. Ilustração da distribuição lognormal base e. O exemplo é a distribuição ajustada para triglicérides séricos em cálculo biliar.....	246
Figura 60. Evolução da aproximação da distribuição to da distribuição normal padrão em função do número de graus de liberdade, (a) com 1 g. l; (b) com 8 g.l. e (c) com ∞ g. l. ou distribuição normal padrão z.	248
Figura 61. (a) distribuição χ^2 com 5 g. l. e (b) distribuição χ^2 com 15 g. l.....	250
Figura 62. Distribuição F_{ν_1, ν_2} , no caso $\nu_1 = 10$ e $\nu_2 = 10$	251
Figura 63. Esquema ilustrativo apresentando as principais partes do Box-and-Wisker plots ou simplesmente box plot.....	262
Figura 64. Ilustração da matriz correlação com algumas variáveis tratadas neste trabalho .	263

LISTA DE TABELAS

Tabela 1 - Desempenho de para 45 alunos (divididos em 5 grupos de 9) tirados aleatoriamente de uma instituição de Ensino Fundamental aos quais foram aplicados 5 métodos de ensino de aritmética	40
Tabela 2 - Decomposição Aritmética a partir da grande média $\bar{y} = 20$	43
Tabela 3 - Tabela de análise de Variância (ANOVA) para o exemplo da experiência de ensino	45
Tabela 4 - Decomposição algébrica das somas de quadrados da ANOVA One-way -.....	45
Tabela 5 - Matriz de ensaios para o exemplo da penicilina (ANOVA com dois fatores e blocos randomizados).....	51
Tabela 6 - Dados obtidos para análise do planejamento fatorial com dois fatores no exemplo da bateria.....	54
Tabela 7 - Esquema geral da ANOVA para a separação dos componentes da variância associadas com bateladas, amostras e testes, (Box et al. 1978).....	59
Tabela 8 - Resultados das medições da resistência, em daN/cm ² , feitas para o exemplo dos 10 fardos de “Filter Tow”	61
Tabela 9 - Análise de variância para o teste da resistência do exemplo dos 10 fardos de “Filter Tow”.....	61
Tabela 10 - Os componentes da variância para exemplo dos 10 fardos de “Filter Tow”	62
Tabela 11 - As várias condições de operação, assumindo todos os níveis para o planejamento fatorial com três fatores, usado para obter os resultados da etapa um. Os valores da diferença de temperatura, oC, entre o gás de entrada (após o aquecimento) e o gás de saída, estão no corpo da tabela	68
Tabela 12 - Análise de variância do projeto fatorial three-way, mostrando a significância dos efeitos principais, das interações de dois fatores e o erro baseado no residual.....	69
Tabela 13 - Análise de variância para o modelo proposto.....	69
Tabela 14 - Resultados obtidos para o teste de confirmação das condições impostas e validação do modelo. Todos os resultados estão no corpo da tabela	73

Tabela 15 - Eficiência de coleta para partículas com diâmetro médio 7,48 μm , quando o ciclone lavador operou com 4 aquecedores ligados, 4 bicos abertos, chaminé na posição 3, vazão de ar 0,358 m^3/s e $\Delta P=1,47$ kPa.....	76
Tabela 16 - Matriz de ensaios em variáveis originais e codificadas para o exemplo dos poluentes.....	81
Tabela 17 - Notação mais conveniente e usual para a matriz de ensaios em variáveis originais e codificadas.....	82
Tabela 18 - Modelo de geração da matriz, para a estimativa dos efeitos pelo método rápido	88
Tabela 19 - Os efeitos do exemplo dos poluentes. Projeto 2^3 sem replicação.....	93
Tabela 20 - Efeitos para o exemplo dos poluentes ignorando o efeito da temperatura, tratando-se o projeto 2^3 como duas réplicas genuínas 2^2 , levando em conta apenas a marca química e a agitação.....	96
Tabela 21 - Intervalo de variação para as variáveis de processo, originais e codificadas.....	101
Tabela 22 - Matriz de ensaios com os resultados das sete respostas para o estudo do comportamento do processo de fabricação do BOPP.....	104
Tabela 23 - Efeitos estimados para as sete respostas (propriedades mecânicas do BOPP)do projeto fatorial completo 2^4	105
Tabela 24 - Efeitos considerados na análise do DOE 2^4 , para o processo de obtenção do filme transparente de polipropileno.....	110
Tabela 25 - Modelos simplificados obtidos a partir dos efeitos considerados de maior influência, para as sete respostas estudadas.....	111
Tabela 26 - Uma comparação entre valores preditos pelos modelos e valores observados para o projeto fatorial 2^4 , aplicado na fabricação de filmes de polipropileno transparente.....	117
Tabela 27 - Um teste de validação dos modelos para algumas bobinas produzidas após a melhoria aplicada na fabricação de filmes de polipropileno transparente.....	118
Tabela 28 - Condições operacionais para fase 1 do estudo EVOP, para otimização da reação de obtenção de acrilamida.....	133

Tabela 29 - Resultados obtidos após 13 ciclos da fase 1 do programa EVOP para a melhoria da fabricação de acrilamida.....	139
Tabela 30 - Dados operacionais obtidos no estudo EVOP - fase 1, da reação de obtenção de acrilamida com o lote de catalisador K561.....	145
Tabela 31 - Dados operacionais obtidos no método EVOP, fase 1, da reação de acrilonitrila e água para produção de acrilamida, com catalisador lote L576.....	147
Tabela 32 - Dados operacionais obtidos no estudo EVOP - obtenção de acrilamida com catalisador lote A502.....	149
Tabela 33 - Tabela de acompanhamento com as condições operacionais para fase 2 do programa EVOP do processo de fabricação da acrilamida.....	154
Tabela 34 - Resultados operacionais obtidos no estudo EVOP - fase 2, da reação de obtenção de acrilamida.....	156
Tabela 35 - Matriz de sinais do projeto fracionado 2^{4-1} , definindo a variável E=ABCD, e relação entre as colunas e o padrão representado pelos contrastes, ℓ	163
Tabela 36 - Matriz de sinais do projeto fracionado 2^{4-1} , definindo a variável E = ABCD, e relação entre as colunas e o padrão representado pelos contrastes ℓ'	165
Tabela 37 - Relação entre colunas e padrão de confundimentos para a fração 2_{III}^{5-2}	169
Tabela 38 - Matriz de sinais para o projeto fracionado 2_{III}^{5-2}	170
Tabela 39 - A matriz de sinais do projeto fracionado 2_{III}^{7-4} , cujas identidades são: I=ABD, I=ACE, I=BCF. e I=ABCG.....	171
Tabela 40 – Exemplo de geração de blocos para DOE 2^k e DOE 2^{k-p} , bloco=ABC.....	173
Tabela 41 - Projeto fracionado específico que não confunde efeitos principais com as interações AB e AE. Estas últimas não se confundem entre si.....	178
Tabela 42 - Projeto fracionado específico que não confunde efeitos principais com as interações AB e AE. Elas também não se confundem entre si.	179
Tabela 43 - Tabela 1 do artigo original, apresentado níveis e códigos dos fatores e as 4 respostas para extração de antocianina da pele de jaboticaba. Fonte: (Montes et al. 2005).....	182

Tabela 44 - Matriz de ensaios com variáveis originais e codificadas e resultados de 4 respostas do projeto 2^4 , retirados dos dados obtidos por Montes e colaboradores (2005) para extração de antocianina da pele de jaboticaba.....	185
Tabela 45 - Matriz de ensaios com variáveis originais e codificadas e resultados de 4 respostas do projeto fracionado 2^{4-1} , retirados dos dados obtidos por Montes e colaboradores (2005) para extração de antocianina da pele de jaboticaba.....	185
Tabela 46 - Efeitos estimados para os dados de Montes e colaboradores (2005); para os projetos 2^4 , a fração 2^{4-1} e o projeto misto $2^1 \times 3^2$	186
Tabela 47 - Comparação dos efeitos usando o projeto misto $2^1 \times 3^2$ e o projeto fracionado 2^{4-1}	191
Tabela 48 - Comparação dos efeitos usando o projeto misto $2^1 \times 3^2$ e o projeto completo 2^4	192
Tabela 49 - Coeficientes obtidos por meio da regressão dos resultados de um planejamento genérico.....	208
Tabela 50 - ANOVA da regressão quando não há repetições genuínas ou réplicas do projeto.....	209
Tabela 51 - ANOVA da regressão quando há repetições genuínas ou réplicas do projeto ...	209
Tabela 52 - Apresentação das variáveis e seus níveis originais e codificados, para o estudo da destilação azeotrópica da solução de glicerina.....	214
Tabela 53 - Matriz do planejamento fatorial com 6 pontos centrais indicando os resultados obtidos para concentração final de glicerina.....	216
Tabela 54 - Análise de variância mostrando a significância dos efeitos agrupados para o planejamento fatorial 24 com 6 pontos centrais.....	217
Tabela 55 - Matriz do planejamento fatorial com 6 pontos centrais e 8 pontos axiais indicando os resultados obtidos para concentração final de glicerina.....	218
Tabela 56 - Efeitos principais; efeitos quadráticos e interações de dois fatores para o projeto composto de face centrada, com 30 experimentos.....	220
Tabela 57 - Efeitos e coeficientes para o projeto de face centrada, ignorando o efeito da temperatura na destilação azeotrópica da glicerina.....	221

Tabela 58 - ANOVA da regressão com a significância sendo estimada com base no erro puro,
para o projeto de face centrada ignorando a variável temperatura..... 222

Tabela 59 - Resultados da aplicação de um teste de hipóteses, com as respectivas
probabilidades 255

NOMEMCLATURA

A = solvente (codificada)

A = temperatura (codificada)

A = pressão (codificada)

a = tipos de fertilizantes (codificada)

A, B, C, D, E = notação usada nos programas estatísticos

ALDM = alongamento na direção longitudinal

ALDT = alongamento na direção transversal

ANOVA = “analysis of variance” - análise de variância

B = ácido (codificada)

B = batelada

B = relação volumétrica água/acrilonitrila (codificada)

B = concentração (codificada)

b = departamentos(codificada)

b_0, b_1, b_{ii}, b_{ij} = coeficientes de regressão não conhecidos

BOPP = polipropileno biorientado

C = % solvente (codificada)

c = fornecedores(codificada)

C = pH(codificada)

C^* = chroma (atributo de cor)

C_{ent} = concentração de material particulado na entrada do lavador de gases (mg/m^3)

C_{sai} = concentração de material particulado na saída do lavador de gases(mg/m^3)

CEP = controle estatístico de processos

Cov = covariância

D = agitação(codificada)

d = diferença entre os resultados dos dois experimentos repetidos

d = máquinas(codificada)

D = pH (codificada)

D_c = diâmetro do ciclone (mm)

DM = grau ou razão de estiro na direção da máquina longitudinal

DOE = “design of experiments” - planejamento de experimentos

DT = grau ou razão de estiro na direção transversal

E = catalisador(codificada)

e = processos(codificada)

EDDM = estabilidade dimensional na direção longitudinal

EDDT = estabilidade dimensional na direção transversal

EVOP = “evolutionary operation” - operação evolutiva

f = métodos de análise química

F = razão entre variâncias (variância do efeito/ variância do erro)

FTU = unidade nefelométrica de turbidez

g = operadores

GL, *g* = graus de liberdade

G_{li} = concentração de glicerina (% em massa)

$G_{li_{máx}}$ = concentração máxima de glicerina (% em massa)

$G_{li_{médio}}$ = concentração média de glicerina (% em massa)

Gli_{\min} = concentração mínima de glicerina (% em massa)

h_{as} = hue (atributo de cor)

\mathcal{H} = umidade absoluta (g/kg)

h = níveis de temperatura

\mathcal{H}_r = umidade relativa (%)

K = fator multiplicador para estimar desvio padrão, é função de número de ciclos e número de observações por ciclo (box, 1969, pag222, tabela v)

k = número de variáveis independentes sob estudo (projeto fracionado)

L = $2/\sqrt{n}$

l = padrão de confundimentos “confounding”

L^* = lightness (atributo de cor)

LSTM = laboratório de separações térmicas e mecânicas

M = $1,79/\sqrt{n}$

MDO = máquina de orientação longitudinal

MFI = índice de fluidez do polímero

ME = “margin of error”

MSE = mean squared error

MSR = metodologia de superfície de resposta

m_T = média de desvios quadráticos da matriz T

m_R = média de desvios quadráticos da matriz R

n = número de ciclos

n = número de elementos de uma amostra

n_o = número de repetições

NTU = unidade nefelométrica de turbidez

OPAC = opacidade

OPP = o mesmo que bopp

p = a probabilidade que é tomada como risco de rejeitarmos a hipótese nula se ela fosse verdadeira

p = número de identidades geradoras (projeto fracionado)

P = probabilidade estatística

P = pressão no ciclone (kPa)

PE = polietileno

PP = polipropileno

PVC = poli (cloreto de vinila)

PVdC = poli (cloreto de vinilideno)

PSE = "pseudo Standard error"

R = coeficiente de determinação

Radj. = coeficiente de determinação ajustado para os graus de liberdade do erro

RSM = Response Surface Methodology

s = erro padrão ou desvio padrão estimado

S = número de amostras

s² = variância global estimada

S_D = soma dos elementos da matriz D elevados ao quadrado

S_E = soma de desvios quadráticos devido ao resíduo

S_L = soma de desvios quadráticos devido a falta de ajuste ao modelo ("lack of fit")

S_M = soma de desvios quadráticos devido a todos os coeficientes (modelo)

- S_{PE} = soma de desvios quadráticos devido ao erro puro (repetições ou réplicas genuínas)
- S_R = soma dos elementos da matriz R elevados ao quadrado
- S_R = soma de desvios quadráticos devido à regressão
- S_T = soma dos elementos da matriz T elevados ao quadrado
- S_T = soma dos desvios quadráticos totais
- t = “t” de student
- T = número de análises
- t = temperatura (°C)
- T = valor do desvio padrão normalizado da distribuição de student (t)
- T = variável em unidades originais
- \bar{T} = media entre os valores de nível baixo e alto
- t_{bu} = temperatura do bulbo seco (°C)
- t_w = temperatura do bulbo úmido (°C)
- TDO = temperatura de pré-aquecimento (°C)
- Teste t = teste de hipóteses aplicando a distribuição t (student)
- T_i = temperatura inicial de alimentação da glicerina (°C)
- $T_{máx}$ = temperatura máxima de alimentação da glicerina (°C)
- T_{max} = valor do nível alto da variável original
- $T_{média}$ = temperatura média de alimentação da glicerina (°C)
- $T_{mín}$ = temperatura mínima de alimentação da glicerina (°C)
- T_{min} = valor do nível baixo da variável original
- TRDM = tensão de ruptura na direção longitudinal

TRDT = tensão de ruptura na direção transversal

VAR = variância

V_g = vazão de glicerina alimentada (g/s)

$V_{g \text{ máximo}}$ = vazão de glicerina máxima (g/s)

$V_{g \text{ médio}}$ = vazão de glicerina média (g/s)

$V_{g \text{ mínimo}}$ = vazão de glicerina mínima (g/s)

V_t = vazão de tolueno alimentado (g/s)

$V_{t \text{ máximo}}$ = vazão de tolueno máximo (g/s)

$V_{t \text{ médio}}$ = vazão de tolueno médio (g/s)

$V_{t \text{ mínimo}}$ = vazão de tolueno mínimo (g/s)

x = variável codificada

x_1 = marca química(codificada)

x_1 = vazão de alimentação da glicerina (codificada)

x_1 = grau de estiro na direção longitudinal (codificada)

x_2 = temperatura (codificada)

x_2 = vazão de alimentação de vapor de tolueno (codificada)

x_2 = grau de estiro na direção transversal (codificada)

x_3 = agitação (codificada)

x_3 = concentração inicial de glicerina (codificada)

x_3 = temperatura de pré-aquecimento (codificada)

x_4 = temperatura de entrada da glicerina (codificada)

x_4 = temperatura de estiro (codificada)

X_i, x_i = variáveis de entrada (independentes)

y = diferença de temperatura entre o gás de entrada e o gás de saída

y = valor observado

Y_i, y_i = variáveis resposta (dependentes)

\bar{y} = grande média

\hat{y} = resposta predita pelo modelo

Letras gregas

η = eficiência (%)

η = grande média

ν = número de graus de liberdade

ε_{ij} = erro aleatório

ε = erro puro

τ_i = i-ésimo efeito do tratamento

τ_i = efeito de chaminé

σ^2 = Variância

σ = Desvio padrão

γ_k = temperatura

β_j = vazão de água

Δ = diferença entre uma variável de entrada e outra de saída

ρ = coeficiente de correlação linear

μm = micron (micra)

Sub-índices

t, i, j = número de níveis das colunas, linhas e réplicas

$1, 2, \dots, n$ = número de componentes

A, B, C, D, E = notação usada nos programas estatísticos

SUMÁRIO

Capítulo 1	33
Introdução e justificativa.....	33
Objetivos	35
Objetivo geral	35
Objetivos específicos.....	36
Capítulo 2.....	37
A análise da variância.....	37
2.1. Análise de variância um fator (“One-way”).....	38
2.1.1 O exemplo da Experiência de Ensino.....	39
2.1.2 Verificação gráfica do modelo dos efeitos fixos proposto.....	47
2.1.3 Interpretação geométrica da ANOVA	49
2.2 Análise de variância com dois fatores (“Two-way”)	51
2.2.1 Blocos randomizados.....	51
2.2.2 Projeto fatorial com dois fatores (“Two-way”).....	53
2.3 Outros planejamentos com ANOVA.....	55
2.4 Considerações sobre a análise de variância.....	56
2.5 Os componentes da variância.....	57
2.5.1 O exemplo dos 10 fardos de “Filter Tow”	60
Capítulo 3.....	63
Estudo de caso aplicando um projeto fatorial “three-way”.....	63
Um ciclone lavador adaptado de um ciclone tipo americano	63
3.1 O problema	63
3.2 A proposta de solução	64
3.3 Planejamento e obtenção dos resultados dos testes da primeira etapa.....	65
3.4 Resultados dos ensaios do teste da eficiência de coleta	74

3.5	Perspectivas do ciclone lavador	77
3.6	Questões adicionais	77
Capítulo 4		78
Projeto fatorial em dois níveis		78
4.1	Aplicando um DOE 2^k	80
4.1.1	O exemplo de redução de poluentes de uma corrente de resíduos	80
4.1.1.1	O projeto completo com 3 fatores	81
4.1.1.2	A estimativa dos efeitos	83
4.1.2	Outras maneiras de estimar os efeitos das variáveis	86
4.1.3	Vantagens do projeto fatorial sobre o método “uma variável por vez”	88
4.1.4	Ensaio com réplica genuína	89
4.1.4.1	Estimativa da variância do erro e do erro padrão dos efeitos para ensaios replicados	90
4.1.4.2	A interpretação dos efeitos	91
4.1.4.3	Uma estimativa grosseira do erro	91
4.1.4.4	Uso da escala de probabilidade Normal	92
4.1.5	Considerações finais sobre os projetos fatoriais completos em dois níveis	97
Capítulo 5		98
Dois casos aplicando o DOE 2^k em processos industriais		98
5.1	Melhoria do processo de produção de filmes de polipropileno usando um projeto fatorial 2^4	98
5.1.1	O problema	99
5.1.2	A proposta de solução	99
5.1.3	A matriz do projeto fatorial 2^4 e os resultados dos ensaios	102
5.1.4	A estimativa e a interpretação dos efeitos	105
5.1.5	Questões adicionais	119

5.2	Aplicando EVOP (<i>Evolutionary Operation</i>) na melhoria do processo de fabricação de acrilamida	119
5.2.1	O problema.....	119
5.2.2	A proposta de solução.....	120
5.2.3	Uma breve apresentação do programa EVOP.....	120
5.2.3.1	Uma comparação do EVOP com operação estática	124
5.2.3.2	O planejamento EVOP.....	125
5.2.3.3	O esquema EVOP para 2 variáveis.....	126
5.2.4	O projeto experimental para o caso da acrilamida	129
5.2.4.1	A escolha das variáveis.....	129
5.2.4.2	O planejamento experimental.....	130
5.2.4.3	A folha de marcha e o seu manuseio	135
5.2.4.4	Alterações.....	141
5.2.4.5	Mudança na média.....	141
5.2.5	Resultados da fase 1	142
5.2.5.1	Resultados e discussão com o lote de catalisador K561	143
5.2.5.2	Resultados e discussão da fase 1 com catalisador L576.....	146
5.2.5.3	Resultados e discussão da fase 1 com catalisador A502.....	148
5.2.5.4	Discussão dos resultados ao final de 15 ciclos da fase 1 com os lotes: K561+L576+A502 do catalisador.....	150
5.2.6	Resultados e discussão da fase 2.....	153
5.2.7	Considerações finais sobre o programa EVOP	158
	Capítulo 6.....	159
	Projetos fatoriais fracionados em dois níveis, DOE 2^{k-p}	159
6.1	A Construção	160
6.1.1	A fração $\frac{1}{2}$	161

6.1.2	A escolha das 16 observações	162
6.1.3	O complemento da fração $\frac{1}{2}$	164
6.1.4	Resolução de um projeto fatorial fracionado	166
6.1.5	A relação completa de definição de um projeto	168
6.1.6	A estrutura de um projeto fracionado saturado	170
6.2	A estimativa dos efeitos	172
6.2.1	A interpretação dos efeitos	172
6.3	O efeito de blocos nos DOE 2^k e DOE 2^{k-p}	172
6.4	Como usar a fração mais adequada à necessidade	174
6.4.1	Um exemplo de projeto que atende a uma necessidade específica	176
6.4.2	Uma solução como uma alternativa mais econômica de um projeto fatorial misto	180
6.4.2.1	A interpretação dos efeitos da Tabela 46	187
6.5	Comentários sobre o DOE 2^{k-p}	192
Capítulo 7		193
Metodologia da superfície de resposta, noções da regressão linear múltipla e estudo de caso		193
7.1	Introdução	193
7.2	Alguns tipos de delineamentos usados na RSM	196
7.2.1	O gradiente ascendente	198
7.2.2	O delineamento DOE 2^k , DOE 2^{k-p} e ambos com ponto central	199
7.2.3	projetos e modelos de segunda ordem (projeto composto)	200
7.3	Regressão linear múltipla	201
7.3.1	A regressão linear simples e múltipla	202
7.3.2	A estimativa dos coeficientes do modelo e sua interpretação	204
7.3.2.1	Obtenção dos parâmetros da regressão por meio das equações normais ...	204

7.3.2.2	A interpretação dos parâmetros da regressão	206
7.4	Estudo de caso: Influência das variáveis de processo na desidratação da glicerina por destilação azeotrópica	211
7.4.1	O problema.....	211
7.4.2	A proposta de solução do problema	212
7.4.3	A seleção das variáveis e o projeto experimental.....	212
7.4.3.1	Planejamento fatorial composto	217
7.4.4	A análise e a interpretação do planejamento de 2ª ordem	219
7.4.5	Perspectivas da destilação azeotrópica para a desidratação da glicerina.....	225
Capítulo 8	226
Considerações finais	226
Apêndice A	229
Conhecimentos Básicos de Estatística	229
A.1	Generalidades.....	229
A.2	Amostragem.....	232
A.3	Estatística descritiva	232
A.3.1	Principais elementos da estatística descritiva	234
A.3.1.1	Descrição gráfica	234
A.3.1.2	Medidas de posição.....	236
A.3.1.3	Medidas de dispersão.....	238
A.3.1.4	Medidas de assimetria.....	240
A.3.1.4	Medidas de achatamento ou curtose.....	241
A.4	Distribuições	241
A.4.1	Distribuição Normal	241
A.4.1.1	Características importantes da distribuição normal.....	243
A.4.1.2	Distribuição normal Padrão.....	244

A.4.1.3	Distribuição Lognormal	246
A.4.2	Outras Distribuições	246
A.4.2.1	Distribuição t.....	246
A.4.2.1.1	Graus de liberdade de uma estatística	249
A.4.3	Distribuição χ^2	249
A.4.4	Distribuição F de Snedecor.....	250
A.5	Estatística Indutiva.....	251
A.5.1	Intervalo de confiança para a média da população quando σ é desconhecido.....	252
A.5.2	Intervalo de confiança para a variância e o desvio padrão da população	253
A.5.3	Testes de hipóteses	253
A.5.4	Teste de normalidade de Anderson Darling.....	255
A.5.5	Teste t de Student	256
A.5.5.1	Teste t para variâncias iguais.....	257
A.5.5.2	Teste t para variâncias desiguais	258
A.5.6	Teste F.....	259
A.6	Visualização de dados	260
A.6.1	Digrama de pontos.....	260
A.6.1	Box and Whisker plots.....	260
A.6.2	Matriz de correlações.....	262
Capítulo 9	264
Referências Bibliográficas	264
ANEXO A	269
ANEXO B	276

Capítulo 1

Introdução e justificativa

Segundo Tomasello (1999), renomado psicólogo especializado em cognição e aprendizado sociais, o diferencial da cultura humana (com relação aos traços culturais que vêm sendo evidenciados em outros animais) é que ela é essencialmente cumulativa. Ou seja, uma nova descoberta (ou conhecimento adquirido) é somada a uma anterior em uma corrente que não volta para o ponto inicial. Conhecida como "efeito catraca", essa teoria está na base para a compreensão do acelerado desenvolvimento humano e a capacidade que temos em transformar o mundo ao nosso redor. A potencialização significativa do processo de aquisição e aprendizagem, por sua vez, pode ser em grande parte creditado ao momento em que o conhecimento científico pôde lançar mão dos métodos estatísticos.

O potencial de impacto dos métodos estatísticos na investigação racional da natureza é inegável. Seja a ciência pura ou a aplicada, a natural ou a social, qualquer campo sistemático de estudo atualmente tem sua credibilidade e viabilidade cada vez mais associados a aplicação de um planejamento estatístico de experimentos, ou seja, um delineamento experimental.

Montgomery (2004), enfatiza que engenheiros resolvem problemas de interesse da sociedade pela aplicação eficiente de princípios científicos. Neste contexto, o uso desses princípios é fundamental na formulação dos problemas. O autor ainda menciona as 8 etapas que compõem o método, a saber:

1. Desenvolver uma descrição clara e concisa do problema;
2. Identificar, no mínimo tentar, os fatores importantes que afetam esse problema ou que possam desempenhar um papel em sua solução;
3. Propor um modelo para o problema, usando conhecimento científico ou de engenharia do fenômeno em estudo. Estabelecer quaisquer limitações ou suposições ao modelo;
4. Conduzir experimentos apropriados e coletar dados para testar e validar o modelo-tentativa ou conclusões feitas nas etapas 2 e 3;
5. Refinar o modelo com base nos dados observados;
6. Manipular o modelo de modo a ajudar o desenvolvimento da solução do problema;

7. Conduzir um experimento para confirmar que a solução proposta para o problema é efetiva e eficiente;
8. Tirar conclusões ou fazer recomendações baseadas na solução do problema.

Vê-se, assim, que a experimentação é uma parte essencial do método científico. Vale lembrar, no entanto, que existem três categorias principais do estado de conhecimento sobre um fenômeno que se quer estudar. Em um dos extremos, existem aqueles já bem conhecidos e estudados, cujos modelos que descrevem seu mecanismo físico são conhecidos e chamados de modelos mecanísticos. Em uma situação intermediária estão os fenômenos que, embora seus mecanismos físicos sejam conhecidos, os modelos que os descrevem são muito complexos para permitir uma solução analítica sem muitas simplificações. E, por fim, existe o extremo em que não se conhece ainda nada sobre o mecanismo físico que descreve o fenômeno. Neste caso, para fenômenos complexos ou desconhecidos, os modelos aproximados por métodos estatísticos são denominados de modelos empíricos. Um bom modelo empírico pode ser manipulado por um técnico ou um cientista da mesma forma que um modelo mecanístico.

Segundo Schmidt e Launsby (1996), um planejamento experimental consiste de mudanças deliberadas nas entradas (fatores) de um processo/atividade, a fim de se observar as mudanças correspondentes na saída (resposta). Um projeto experimental eficiente não só permite responder às questões propostas na concepção do problema (com erro experimental pequeno) e analisar os dados obtidos em determinado estudo com sensibilidade (indicando com mais robustez o que pode ou não ser concluído a respeito da hipótese proposta), como através dele podem surgir novas idéias. Também é através dos projetos estatísticos de experimentos que se obtêm ganhos em tempo e custo, documentação e um conhecimento muito mais apurado do processo ou da atividade que se investiga. Ou seja, há uma otimização do processo de conhecimento, uma vez que se pode conseguir o máximo de informações com o mínimo de esforço. O tempo e o custo gastos com cada experimento, tanto para desenvolvimento e melhoria de processos em geral quanto para qualquer atividade científica menos aplicada, são variáveis que devem ser bem controladas. Com isso pode-se obter o conhecimento previsto no estudo com mínimos tempo e gasto, e máximas produção e boa qualidade.

Considerando particularmente o conhecimento técnico-científico, seu ganho por meio de métodos estatísticos está necessariamente ligado ao desenvolvimento de novos produtos e sistemas, melhoria dos já existentes e ao planejamento, desenvolvimento e melhoria dos processos de produção. Porque o conhecimento para esses setores aplicados é a chave para a

inovação e o lucro financeiro. Um processo ou uma atividade, no contexto aqui referido, consiste do uso combinado de máquinas, materiais, métodos, pessoas, ambientes e medições, que, em conjunto, desempenham um serviço ou finalizam um produto. Portanto, assim como em tantas outras áreas do conhecimento, em um único tema, questão ou processo estão envolvidos um número de fatores (variáveis) que muitas vezes desenvolvem entre si relações mais complexas. Por essa razão principal, obter dados usando a estratégia “uma variável por vez” torna-se extremamente inadequado e ineficiente na otimização de produtos, projetos e processos. O planejamento experimental, então, é capaz de avaliar não só os efeitos de múltiplas variáveis simultaneamente como também avaliar a interação entre elas.

Como exemplo prático da apresentação aqui realizada estão as maiores corporações do mundo, que têm obtido resultados financeiros nunca antes alcançados. Aponta-se o *Programa 6 σ* como um dos principais responsáveis pelos lucros extraordinários que essas corporações estão experimentando. Entre elas podem-se citar: GE, Motorola, Sony, Nokia, Texas Instruments, Allied Signal, Canon, Hitachi, American Express, Toshiba, Du Pont e Polaroid, entre outras.

Afinal o que é o programa ou estratégia 6 σ ? Segundo Austin Thomas (2006) é um esforço produtivo, com metodologia extremamente disciplinada que usa uma vasta gama de ferramentas (estatísticas e não estatísticas) para melhorar o desempenho operacional das companhias, eliminando defeitos nos projetos de produtos, nos processos de produção, nas transações de negócios e nos serviços relacionados com os processos. O uso da estratégia 6 σ ajuda a atingir o nível de defeitos a quase zero. Por exigência dos clientes, a competitividade das empresas provoca uma busca disciplinada de dados, fatos e análise estatística para analisar, melhorar e controlar processo.

Considerando o exposto, é essencial ponderar que para desenvolver qualquer tipo de trabalho que envolva experimentação é necessário não somente obter, tratar e interpretar dados, mas também conhecer e dispor da potencialidade das ferramentas estatísticas.

Objetivos

Objetivo geral

O objetivo deste texto é mostrar uma conduta básica de uso de algumas técnicas de planejamento estatístico de experimentos que a autora vem usando em suas linhas de pesquisa e na prestação de serviço à comunidade nos últimos 20 anos.

Objetivos específicos

Para atingir o objetivo geral, o texto foi estruturado de modo a apresentar capítulos com a apresentação dos planejamentos experimentais, seguidos por estudos de casos, desenvolvidos por alunos de pós-graduação, aplicando os métodos apresentados. Os casos foram selecionados contemplando dois tipos de aplicação: a aplicação em processos industriais visando buscar as melhores condições operacionais (os dois casos do capítulo 5) e a aplicação em desenvolvimento de equipamento (capítulo 3) e um caso de desenvolvimento de processo (fechando o capítulo 7). O capítulo que recebeu uma ênfase diferente foi o que trata dos planejamentos fatoriais fracionados em 2 níveis (capítulo 6), por ter sido verificado ao longo desses 20 anos, que este tipo de planejamento não é usado com todo o potencial por ele oferecido. Neste capítulo foram apresentados dois exemplos extraídos da literatura cuja discussão teve o intuito de mostrar como uma das formas práticas de uso dos projetos fracionados.. Isto não quer dizer que não se possa usar todos os outros tipos de planejamento, mas, o desperdício de tempo e recurso para experimentação deve ser muito bem avaliado. O capítulo 8, intitulado como “considerações finais” apresenta algumas idéias que revelam o entusiasmo da autora com o ensino e a aplicação desses métodos. O apêndice A é um texto com alguns conceitos fundamentais da estatística básica aplicada para servir de apoio em algumas situações que necessita de conhecimentos anteriores. Como anexo A está uma cópia do artigo discutido no capítulo 6 e finalmente no Anexo B algumas Tabelas estatísticas.

Capítulo 2

A análise da variância

Existem dois marcos principais no desenvolvimento da estatística aplicada. O primeiro deles foi a proposta do químico inglês William Sealy Gosset (1876 – 1937), quando este apresentou a distribuição t com o pseudônimo Student. A distribuição t permitiu fazer a inferência estatística a partir da amostra, substituindo o desvio padrão σ pelo desvio padrão da amostra s . O segundo marco foi a proposta da Análise de Variância (ANOVA do Inglês *Analysis of Variance*) sugerida pelo inglês Ronald Aylmer Fisher (1890 - 1962) que na época revolucionou a pesquisa na agricultura. A ANOVA é um método estatístico que determina como uma variação pode ser atribuída a certos fatores qualitativos. Ela permite separar qualitativamente, por um lado, os efeitos das variáveis de interesse e, por outro, o erro.

Esta abordagem é muito empregada quando se tem uma resposta contínua na comparação, principalmente, de variáveis qualitativas. Assim, podem-se comparar, por exemplo, vários fornecedores de matéria prima (fatores qualitativos) em relação ao prazo de entrega (em dias) do material. Ela é muito empregada também na odontologia, (através da comparação do desempenho de materiais odontológicos), na comparação de máquinas tendo como resposta a qualidade das peças em termos de dimensões, na comparação de operadores de máquinas, na agricultura, pela comparação de fertilizantes, em culturas agrícolas, e assim por diante.

A ANOVA pode ter duas aplicações: ANOVA para efeitos fixos e ANOVA para efeitos aleatórios.

A primeira aplicação é utilizada nos casos em que os níveis dos fatores são especificados. As conclusões obtidas com a análise dos efeitos fixos não podem ser generalizadas. Esta aplicação da ANOVA inclui os projetos um fator (*one-way, single-factor*), dois fatores (*two-way*), projeto fatorial com dois fatores, projeto fatorial geral, *split plot*, entre outros.

A segunda aplicação é utilizada nos casos em que os efeitos são tirados aleatoriamente de uma população. As conclusões tiradas com esta análise podem ser generalizadas. Essa aplicação inclui o estudo dos componentes da variância.

A decomposição da variabilidade de um processo qualquer em efeitos e erros qualquer que seja o número de fontes de variação e níveis de cada fonte tem uma importância muito grande no julgamento das hipóteses adotadas ajudando o pesquisador a discernir sobre qual é a

melhor alternativa. Então, neste capítulo, será mostrado passo a passo no esquema da ANOVA um fator como é feita esta decomposição e também como ela é interpretada. As demais aplicações consistem em generalização da mesma técnica.

2.1. Análise de variância um fator (“One-way”)

O modelo empregado para o estudo com efeitos fixos na análise de variância um fator (*one-way*) é chamado de modelo das médias, e é uma declaração do tipo:

$$y_{ij} = \eta_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, k \\ j = 1, 2, \dots, n \end{cases} \quad (1)$$

Em que, η_i é a i -ésima média do nível do fator e ε_{ij} é o erro aleatório ao qual se incorporam outras fontes de variabilidade no experimento. Essas incluem as medições, os tipos de amostragem, os fatores não controláveis (como variações ambientais), eventualmente o ruído de fundo do processo, dentre outros.

Uma outra alternativa ao modelo da equação é η_i :

$$\eta_i = \eta + \tau_i \begin{cases} i = 1, 2, \dots, k \end{cases} \quad (2)$$

Em que, η é a grande média (média global de todas as observações que compõem o planejamento) e τ_i é o i -ésimo efeito do tratamento.

Assim o modelo das médias pode ser reescrito como:

$$y_{ij} = \eta + \tau_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, k \\ j = 1, 2, \dots, n \end{cases} \quad (3)$$

Alternativamente o modelo da equação 3 é denominado de “modelo dos efeitos”.

Embora ambos os modelos sejam válidos, o modelo dos efeitos é o mais comum na literatura. Também deve ser dito que ambos os modelos são modelos estatísticos lineares e, especificamente para as equações 1 e 3, eles são os modelos da ANOVA um fator ou simples fator (*one-way, single-factor*), dois fatores, projeto fatorial com dois fatores, projeto fatorial geral, *split plot*, entre outros.

O levantamento dos dados de um planejamento dessa natureza deve ser completamente aleatório. Pelo teste de hipóteses, os erros do modelo são considerados como distribuídos normalmente com média zero e variância σ^2 . A variância σ^2 é considerada constante para todos os níveis do fator. Isto implica que as quantidades η , τ_i e σ^2 são mutuamente independentes.

2.1.1 O exemplo da Experiência de Ensino

Em um experimento para comparar 5 diferentes métodos de ensino de aritmética, 45 estudantes foram aleatoriamente divididos em 5 grupos de mesmo tamanho. Os dois primeiros grupos foram instruídos por dois métodos correntes, enquanto que os três restantes foram instruídos cada um por um dos três novos métodos. O resultado do desempenho dos cinco grupos estão apresentados na Tabela 1. As três últimas linhas da tabela já contém as médias dos tratamentos, a grande média e as diferenças.

Tabela 1 - Desempenho de para 45 alunos (divididos em 5 grupos de 9) tirados aleatoriamente de uma instituição de Ensino Fundamental aos quais foram aplicados 5 métodos de ensino de aritmética.

	Métodos de ensino de aritmética				
	A	B	C	D	E
	17	19	28	19	19
	14	20	30	27	13
	23	13	29	26	12
	20	15	24	25	19
	24	13	25	19	15
	21	17	30	23	15
	16	19	27	24	10
	13	21	28	23	17
	14	16	22	21	15
Média de tratamentos	18	17	27	23	15
Grande Média	20	20	20	20	20
Diferença	2	3	-7	-3	5

É importante dizer aqui que, sempre que forem levantados os dados de um planejamento proposto, é conveniente que eles sejam inicialmente visualizados com gráficos do tipo diagrama de pontos, Figura 1 ou *box plot*, Figura 2. Assim, é possível obter algumas características preliminares do comportamento desses métodos antes mesmo de se elaborar a tabela ANOVA.

Ambos os gráficos indicam que os grupos de alunos aos quais foram aplicados os métodos C e D apresentaram maior desempenho médio. Os diagramas de pontos também indicam que os dados se distribuem aleatoriamente em torno das médias dos tratamentos. Em outras palavras os gráficos de visualização já indicam que um dos métodos novos, o Método E, não trouxe benefícios claros em relação aos Métodos A e B correntes.

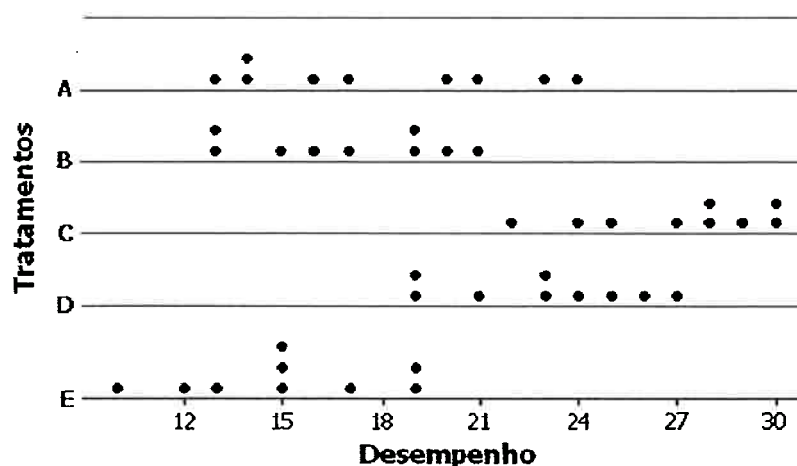


Figura 1. Diagrama de pontos dos dados originais para os métodos de ensino

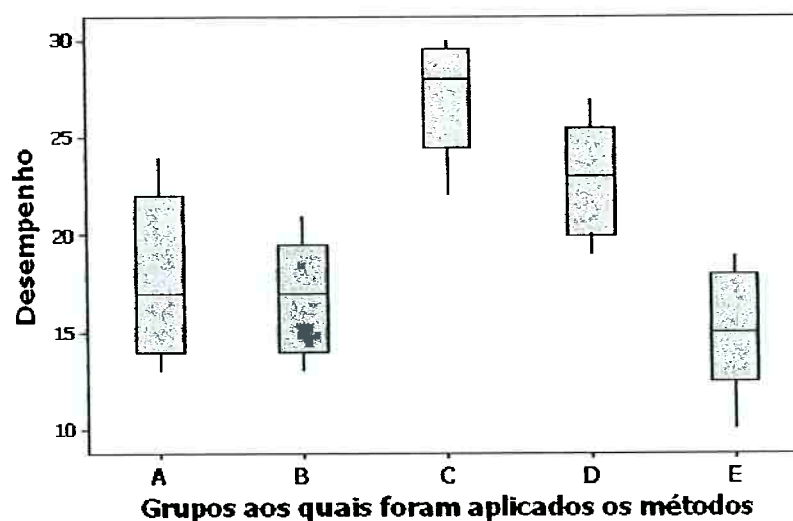


Figura 2. Box plot dos dados originais para os métodos de ensino

Porém, pode-se estar interessado em analisar a questão com maior profundidade, fazendo-se a pergunta: existe evidência que indique diferenças entre os valores das médias associadas com os diferentes métodos de ensino?

A idéia é verificar se as discrepâncias entre as médias dos tratamentos são maiores do que as que poderiam ser esperadas da variação que ocorre dentro dos tratamentos (em cada grupo de 9 alunos). Isto pode ser verificado fazendo-se a decomposição dessa variabilidade usando-se o modelo dos efeitos fixos proposto na equação 3. A amostra é composta das 45 observações dispostas na Tabela 2.

Para entender melhor as quantidades expressas na ANOVA, Box, Hunter, W e Hunter, J. (2005), ilustram a quebra aritmética ou decomposição das somas de quadrados para assim aproximar, pela amostra, o modelo dos efeitos fixos. Esta é uma das formas mais didáticas para o entendimento da ANOVA existente na literatura. A Tabela 2 ilustra a idéia que acabou de ser mencionada.

Tabela 2 - Decomposição aritmética a partir da grande média $\bar{y} = 20$

Observações	Desvios da Grande média	Desvios dos tratamentos	Resíduos
y_{ij}	$y_{ij} - \bar{y}$	$\bar{y}_i - \bar{y}$	$y_{ij} - \bar{y}_i$
17 19 28 19 19	-3 -1 8 -1 -1	-2 -3 7 3 -5	-1 2 1 -4 4
14 20 30 27 13	-6 0 10 7 -7	-2 -3 7 3 -5	-4 3 3 4 -2
23 13 29 26 12	3 -7 9 6 -8	-2 -3 7 3 -5	5 -4 2 3 -3
20 15 24 25 19	0 -5 4 5 -1	-2 -3 7 3 -5	2 -2 -3 2 4
24 13 25 19 15	4 -7 5 -1 -5	-2 -3 7 3 -5	6 -4 -2 -4 0
21 17 30 23 15	1 -3 10 3 -5	-2 -3 7 3 -5	3 0 3 0 0
16 19 27 24 10	-4 -1 7 4 -10	-2 -3 7 3 -5	-2 2 0 1 -5
13 21 28 23 17	-7 1 8 3 -3	-2 -3 7 3 -5	-5 4 1 0 2
14 16 22 21 15	-6 -4 2 1 -5	-2 -3 7 3 -5	-4 -1 -5 -2 0
Y	D=Y-20	T	R
Soma de quadrados	1272	864	408
Graus de Liberdade	44	4	40

Estas representações revelam como os desvios da grande média são decompostos e como cada entrada na ANOVA pode ser estimada.

Trazendo de volta a equação dos efeitos fixos, pode-se verificar que a variabilidade em torno da grande média é representada por $\tau_i + \varepsilon_{ij} = D$, que está indicada na notação da tabela pela matriz D. Já a matriz T dos desvios τ_i e R é a matriz dos ε_{ij} , denominada de “matriz dos resíduos” na análise de variância por conter o erro aleatório e variabilidades não controladas que são adicionados ao erro. Finalmente y_{ij} é a matriz dos valores observados, representada por Y na notação vetorial.

A estimativa pelos dados da amostra está genericamente apresentada na linha acima das matrizes, isto é:

$$D = y_{ij} - \bar{y}; \quad T = \bar{y}_i - \bar{y} \quad \text{e} \quad R = y_{ij} - \bar{y}_i, \quad (4)$$

Assim, chamando S_D a soma dos elementos da matriz D elevados ao quadrado, S_T a soma dos elementos da matriz T elevados ao quadrado, e S_R a soma dos elementos da matriz R elevados ao quadrado, tem-se a entrada da ANOVA denominada de soma de quadrados.

Mais explicitamente:

$$S_D = (-3)^2 + (-1)^2 + (8)^2 + (-1)^2 + \dots + (-5)^2 = 1272 \quad (5)$$

$$S_T = 9 \cdot (-2)^2 + 9 \cdot (-3)^2 + 9 \cdot (-7)^2 + 9 \cdot (3)^2 + 9 \cdot (-5)^2 = 864 \quad (6)$$

$$S_R = (-1)^2 + (2)^2 + (1)^2 + (-4)^2 + \dots + (-0)^2 = 408 \quad (7)$$

Logo $S_D = S_T + S_R$ que para o exemplo $1272 = 864 + 408$. Esta é a propriedade de aditividade explicitada no modelo, que é válida para a soma de quadrados de qualquer tabela similar de números divididos com o esquema da ANOVA um fator.

A Tabela 3 mostra a ANOVA para o conjunto de dados do exemplo que está sendo apresentado.

Tabela 3 - Tabela de análise de Variância (ANOVA) para o exemplo da experiência de ensino

Fonte de variação	Soma de quadrados	Graus de Liberdade	Média de Quadrados	Razão F
Entre tratamentos	$S_T=864$	$v_T = 4$	$m_T = 216,0$	$F_{4,40}=21,18$
Dentro dos tratamentos	$S_R=408$	$v_R = 40$	$m_R = 10,2$	
Total em torno da grande média	$S_D=1272$	$v_D = 44$		

A maioria dos autores apresenta a construção das entradas da tabela ANOVA por meio da decomposição geométrica da variabilidade, como o que é mostrado na Tabela 4.

Tabela 4 - Decomposição algébrica das somas de quadrados da ANOVA One-way -

Fonte de variação	Soma de quadrados	Graus de liberdade	Média de desvios quadráticos
Entre tratamentos	$S_T = \sum_i^k n_j (\bar{y}_i - \bar{y})^2$	$v_T=k-1$	$m_T=S_T/v_T$
Dentro dos tratamentos	$S_R = \left(\sum_i^k \sum_j^{n_j} (y_{ij} - \bar{y}_i)^2 \right)$	$v_R=N-k$	$m_R=S_R/v_R$

Para as duas decomposições (aritmética e geométrica) os resultados são idênticos.

Outra entrada da ANOVA é o número de Graus de Liberdade, v , isto é, o número de elementos dentro de cada matriz da tabela de decomposição que podem ser especificados arbitrariamente. Isto é, o vetor D (matriz com 45 elementos) tem 44 graus de liberdade, porque conhecendo-se 44 elementos desta matriz, o 45º elemento pode ser determinado pela restrição para a qual a soma dos desvios de sua média deve ser igual a zero. O cálculo dos graus de liberdade deve ocorrer também para os elementos de T, no caso com 4 graus de liberdade, lembrando que as linhas da matriz são iguais e os elementos de cada linha também devem somar zero. Neste caso, conhecendo-se 4 elementos o 5º fica determinado pela

restrição. Finalmente os elementos de R são restringidos de duas maneiras: os elementos de cada coluna deve somar zero e a soma total de todos os elementos também devem somar zero.

Assim, o número de graus de liberdade do resíduo é: $45 - 1 - 4 = 40$.

É importante lembrar que para qualquer ANOVA não somente a soma de quadrados é aditiva, como também os graus de liberdade os são.

A terceira entrada da ANOVA é a média de desvios quadráticos (“*mean squares*”) m_T e m_R que são obtidos pela divisão de S_T por v_T e S_R por v_R . Para fechar a ANOVA, considera-se que não existe diferenças entre os tratamentos (métodos de ensino). A média de desvios quadráticos m_T e m_R poderiam ser estimativas independentes da variância do erro σ^2 e sua razão poderia ser uma distribuição F com v_T e v_R graus de liberdade.

Cálculos por computador ou consultando-se tabelas da estatística $F_{4,40}$ mostram que a probabilidade de $F_{4,40} \geq 21,8$ é menor do que 0,001. Logo é compreensível que a hipótese nula (de médias iguais, por exemplo) é insustentável e os métodos de ensino são estatisticamente diferentes.

Logo a tabela de análise de variância responde a questão exposta no início da discussão do problema.

Se os erros são independentemente distribuídos e apresentam distribuições idênticas (a mesma variância) espera-se que:

$$E(m_T) = \sum \tau_i^2 + \sigma^2 \quad (8)$$

$$E(m_R) = \sigma^2 \quad (9)$$

E se não existe diferença entre as médias, isto é, $\tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = 0$ e, portanto,

$\sum \tau_i^2 = 0$, logo m_T e m_R da ANOVA seriam estimativas de σ^2 . Então, m_T e m_R seriam

distribuídos independentemente, e $F = m_T/m_R$ seria a razão de duas estimativas independentes de σ^2 .

Então a estimativa das entradas da tabela ANOVA um fator (modelo apresentado na equação (3)) é realizada com os dados da amostra como:

$$y_{ij} = (y_{ij} - \bar{y}) + (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i - \bar{y}_t + \bar{y}) \quad (10)$$

2.1.2 Verificação gráfica do modelo dos efeitos fixos proposto

A verificação da adequação do modelo proposto pela análise da variância é feito por meio da análise de resíduos. Quando os resíduos se distribuem aleatoriamente não há razão para suspeitar da adequação do modelo e então considera-se o modelo como adequado para representar o conjunto de dados. Se existe alguma anormalidade na distribuição dos resíduos necessita-se uma investigação adicional a esse respeito.

Os gráficos de análise de resíduos mais comuns são aqueles que contrapõe resíduos em função dos valores estimados pelo modelo. No modelo estatístico de efeitos fixos, $\hat{y}_{ij} = \eta + \tau_i$, a estimativa pelos dados da amostra seria: $\hat{y}_{ij} = \bar{y} + (\bar{y}_i - \bar{y})$. Logo os resíduos seriam $y_{ij} - \hat{y}_{ij}$; resíduos em função da ordem temporal do ensaio; e, diagrama de pontos do resíduo.

Para o exemplo da experiência de ensino de aritmética, esses gráficos estão nas Figura 3, Figura 4 e Figura 5.

Box, Hunter, W e Hunter, J (2005) ainda ressaltam que as considerações (aditividade, erros independentemente distribuídos, normalidade das variáveis e variância constante) são formidáveis, mas elas não são tão limitantes no caso da ANOVA. Isso ocorre porque a ANOVA é completamente robusta (insensível) a moderadas não normalidade e desigualdade de variâncias. A consideração mais importante, no entanto, é a de independência entre os erros

para os projetos randomizados, por causa da correlação que pode ocorrer entre os dados coletados. Portanto, modelos randomizados implicam em independência estatística.

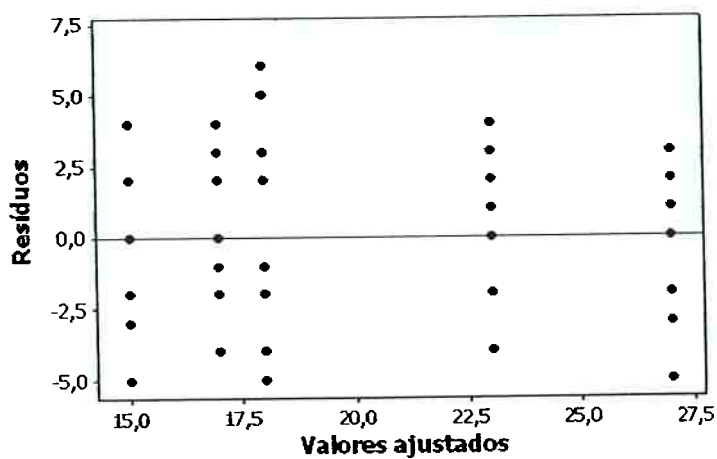


Figura 3. Resíduos em função dos valores ajustados para o exemplo da experiência de ensino

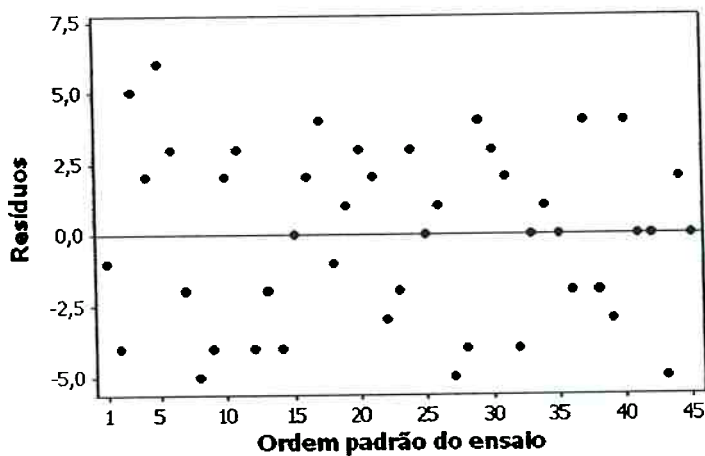


Figura 4. Resíduos em função da ordem padrão dos ensaios para o exemplo da experiência de ensino

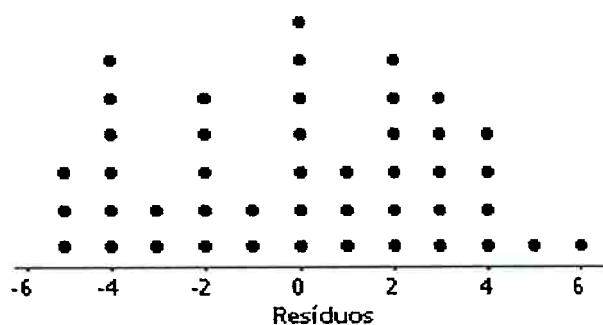


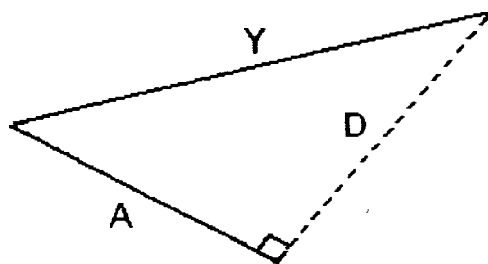
Figura 5. Diagrama de pontos dos resíduos para o exemplo da experiência de ensino

2.1.3 Interpretação geométrica da ANOVA

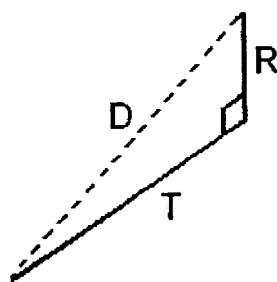
Olhando os dados da Tabela 2 (decomposição aritmética) e os 45 número de cada matriz, vê-se que eles constituem-se dos respectivos vetores Y, D, T e R. Da geometria, se a soma dos produtos dos 45 elementos dois a dois (conhecido como produto interno) é zero os vetores são considerados ortogonais, quaisquer que seja o número de dimensões.

Além disso, para qualquer serie de números de uma tabela desta natureza, por causa da restrição colocada entre seus elementos, os vetores T e R são sempre ortogonais. Como D é a hipotenusa do triângulo retângulo, os vetores D, T e R, pela aditividade da soma de quadrados já demonstrada na decomposição, segue por extensão o teorema de Pitágoras para n dimensões.

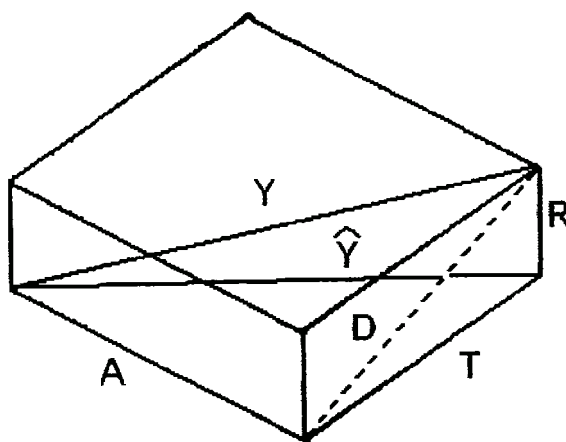
Genericamente falando, a geometria da ANOVA “one-way” pode ser vista no esquema da Figura 6.



(a) representação geométrica da decomposição das observações do vetor Y em termos do vetor A (que corresponde a grande média) e o vetor D (desvios da grande média)



(b) representação geométrica da decomposição do vetor D em termos do vetor T (desvios devido aos tratamentos) e vetor R (resíduos).



(c) representação geométrica da análise de variância mostrando a decomposição ortogonal do vetor Y em termos dos vetores A , T e R .

Figura 6. Geometria da análise de variância (ANOVA One-way)

2.2 Análise de variância com dois fatores (“Two-way”)

Quando se tem duas fontes de variação ou dois fatores (*two-way*), o experimento pode ser planejado de acordo com o tipo de fonte de variação. Portanto, dois modelos de efeitos fixos podem ser testados: blocos randomizados e planejamento fatorial com dois fatores.

2.2.1 Blocos randomizados

Neste caso, uma das fontes de variação é de interesse para estudo e a outra fonte é um fator perturbador, ou bloco (*nuisance factor*), cuja variabilidade deseja-se eliminar. Box, Hunter, W. e Hunter, J. (2005) apresentam um exemplo que contempla uma fonte de variação desejável, no caso, 4 variantes (tratamentos) de um processo de fabricação de penicilina (4 níveis da variável que se deseja investigar); e o bloco, que é a variação de propriedades de uma importante matéria prima (o licor de milho fermentado), que varia de lote para lote. Então, o experimento foi conduzido com 4 tratamentos e 5 misturas do licor (níveis para a variável indesejável).

No planejamento com blocos randomizados, usam-se os níveis de uma das variáveis nas colunas e os níveis da outra nas linhas, um exemplo de estrutura para este planejamento está na Tabela 5.

Tabela 5 - Matriz de ensaios para o exemplo da penicilina (ANOVA com dois fatores e blocos randomizados)

Bloco ↓	Tratamentos			
	A	B	C	D
Mistura 1	89 ⁽¹⁾	88 ⁽³⁾	97 ⁽²⁾	94 ⁽⁴⁾
Mistura 2	84 ⁽⁴⁾	77 ⁽²⁾	92 ⁽³⁾	79 ⁽¹⁾
Mistura 3	81 ⁽²⁾	87 ⁽¹⁾	87 ⁽⁴⁾	85 ⁽³⁾
Mistura 4	87 ⁽¹⁾	92 ⁽³⁾	89 ⁽²⁾	84 ⁽⁴⁾
Mistura 5	79 ⁽³⁾	81 ⁽⁴⁾	80 ⁽¹⁾	88 ⁽²⁾

Na tabela os números entre parêntesis e sobre escrito representam a ordem de realização do ensaio dentro do bloco e os números maiores é a resposta de interesse. Neste caso, o rendimento do processo de produção.

A solução desse problema é idêntica à da análise único fator. Só que, nesse caso, foi introduzida mais uma fonte de variação, os blocos, que estão nas linhas. Então o modelo da análise de variância para este caso é:

$$y_{ij} = \eta + \tau_i + \beta_j + \varepsilon_{ij} \quad \left. \begin{array}{l} i = 1, 2, \dots, k \\ j = 1, 2, \dots, n \end{array} \right\} \quad (11)$$

Em que o termo β_j inserido ao modelo é o termo referente ao efeito dos blocos. A solução deste exemplo pode ser visto em Box, Hunter, W. e Hunter, J. (2005), com as decomposições aritmética e algébrica, a interpretação geométrica, gráficos de resíduos e todas as considerações pertinentes.

Analogamente ao que foi feito para análise de variância um fator, tem-se

$$D = B + T + R \quad (12)$$

A representação geométrica para este caso está apresentada na Figura 7

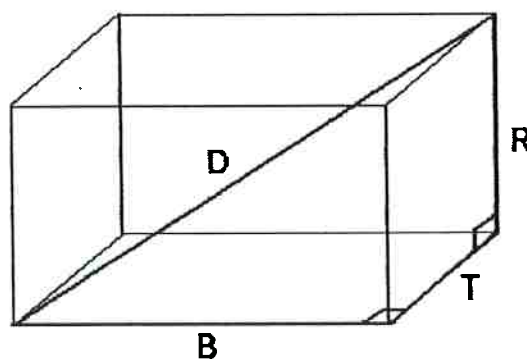


Figura 7. Geometria da ANOVA do projeto com blocos randomizados

Cabe aqui ressaltar que as vantagens deste tipo de projeto são:

1. proporcionar a oportunidade de eliminar as variações de mistura para mistura (isto é os desvios devido as misturas são eliminados do erro);
2. proporcionar uma base indutiva mais ampla do que um experimento feito com um material uniforme (lembrar que os tratamentos são testados com 5 tipos de matéria prima diferentes).

2.2.2 Projeto fatorial com dois fatores (“Two-way”)

Neste caso têm-se duas fontes de variação com igual interesse. É desejável estudar os efeitos de ambas e mais a possibilidade de interação entres essas fontes.

O modelo para este tipo de planejamento não é o mesmo que o do item 2.2.1 (blocos randomizados) porque existe a possibilidade da significância do efeito da interação, causando uma não linearidade ao modelo que é declarado pela equação:

$$y_{ij} = \eta + \tau_t + \beta_i + (\tau\beta)_{it} + \varepsilon_{ij} \begin{cases} t = 1, 2, \dots, k \\ i = 1, 2, \dots, \ell \\ j = 1, 2, \dots, r \end{cases} \quad (13)$$

Sendo, η , τ e β a grande média, o efeito dos níveis do fator que está nas colunas e o fator que está na linha, respectivamente. Estes efeitos são denominados de efeitos principais. Já o termo $(\tau\beta)$ é o efeito de interação entre os dos fatores. Os índices t , i e j são os números de níveis das colunas, linhas e réplicas (a série experimental que se constitui no planejamento), respectivamente. Um arranjo desta natureza é chamado de Projeto Fatorial $k \times \ell \times r$. A razão das r réplicas é possibilitar um maior número de graus de liberdade para a estimativa da interação.

Na Tabela 6 é apresentado um exemplo dessa organização dos dados para o fatorial com dois fatores. Os dados foram obtidos por um engenheiro que estava desenvolvendo um projeto de uma bateria. O interesse era encontrar um material para a confecção das placas que resistisse a variações bruscas de temperatura. Ele dispunha de três tipos de materiais cuja robustez frente à variação de temperatura foi testada em 36 ensaios. A vida útil de cada bateria foi medida em 4 réplicas alocadas aleatoriamente aos três materiais em três níveis de temperatura.

Tabela 6 - Dados obtidos para análise do planejamento fatorial com dois fatores no exemplo da bateria

Tipo de material ↓	Temperatura em °F		
	15	70	125
1	130	34	20
	74	80	82
	155	40	70
	180	75	58
2	150	136	70
	159	106	45
	188	122	25
	126	115	58
3	138	174	96
	168	150	82
	110	120	104
	160	139	60

A Tabela ANOVA para este planejamento pode ser obtida seguindo a linha de decomposição da variabilidade realizada por Box, Hunter, W. e Hunter, J. (1978), na qual o cálculo das

entradas da tabela da análise de variância é feito em duas etapas. Os detalhes dessa decomposição estão nesta referência, no exemplo dos agentes tóxicos.

A adequação do modelo pode ser vista também por meio dos gráficos dos resíduos. Se os resíduos se distribuírem de forma não aleatória (muitas vezes na forma de um funil ou qualquer outro padrão), quando representados graficamente em função dos valores ajustados, os dados não representam o fenômeno investigado. Esses e outros comportamentos podem ser indícios de que é necessária uma transformação matemática na resposta para que haja uma estabilização da variância, isto é, para que a variância se torne constante (resíduos distribuídos aleatoriamente em torno da média). Se a variância é estabilizada pela transformação matemática (log, inversa, raiz quadrada, etc.) é alcançado um modelo linear transformado e diz-se que a interação é do tipo transformável. Tukey (1949) fez a decomposição do resíduo no caso da interação ser significativa e os resíduos indicarem um comportamento não aleatório. Isso permitiu verificar se a transformação em aditividade pode ser conseguida com a transformação matemática. Detalhes sobre estes tipos de transformação podem ser vistos em quase todos os livros sobre planejamento experimental.

2.3 Outros planejamentos com ANOVA

Seguindo o raciocínio vigente, com relação a interpretação da ANOVA, é possível aumentar o número de fontes de variação de acordo com a necessidade. A medida que se aumenta o número de fatores, no entanto, aumenta-se também o número de ensaios, o que torna este tipo de planejamento, com múltiplos fatores, mais oneroso.

Seguindo, então, a linha dos blocos randomizados, Fisher mostrou que é possível eliminar mais do que uma variável de bloco, com um menor número de ensaios. Esses arranjos são denominados quadrados latinos, greco-latinos e hiper-greco-latinos com três, quatro e cinco

fontes variação, respectivamente. No esquema dos quadrados, os modelos testados são sempre aditivos. Para mais informações sobre aplicação destes projetos consultem Box, Hunter, W; Hunter, J; (2005); Montgomery, (2004), Akhnazarova e Kafarov, (1982).

Montgomery (2005) apresenta os planejamentos fatoriais com três fontes de variação, onde é possível estimar até a interação de 3 fatores, e se houver réplicas estima-se o erro experimental puro.

Finalmente, merece ser mencionado, ainda, o “*Split-Plot Design*” (em português conhecido como “planejamento com unidades subdivididas”), que é um projeto fatorial multifator, no qual, por alguma restrição das condições de realização dos experimentos, a ordem dos ensaios não pode ser completamente aleatória. É um projeto de grande interesse na experimentação industrial, porque as comparações são feitas com um erro pequeno, comparado aos planejamentos mencionados até agora.

Segundo Box, Hunter, W. e Hunter, J. (2005) o “*Split-Plot Design*” foi originalmente desenvolvido por R. A. Fisher e F. Yates para experimentos na agricultura, daí o termo “*plot*” que quer dizer lote. Por sugestão de Cox, devido à utilidade desse esquema também fora da agricultura a denominação “*Split-Unit Design*” é mais apropriada.

Exemplos de uso dessa técnica podem ser vistos nos livros de planejamento estatístico de experimentos.

2.4 Considerações sobre a análise de variância

O esquema da análise de variância em geral é útil para verificar a existência ou não de efeitos de fontes de variação, por meio da comparação pelo teste F das médias quadráticas entre tratamentos com as médias quadráticas dentro dos tratamentos. Em outras palavras a ANOVA

testa a hipótese nula (de igualdade das médias) e a hipótese alternativa (pelo menos uma das médias é estatisticamente diferente). Então, empregando os projetos adequados a cada situação, pode-se, por exemplo, verificar se existe diferença ou não entre a tipos de fertilizantes, b departamentos, c fornecedores, d máquinas, e processos, f métodos de análise química, g operadores, h níveis de temperatura, etc., sendo a, b, c, d, e, f, g e h números inteiros e maiores do que 2.

Em geral, quando se deseja comparar apenas duas entidades, pela simplicidade, é conveniente usar o teste t de Student, apresentado no Apêndice-A (noções de estatística básica). O uso do teste t para vários níveis de variação é exaustiva e ineficiente, além de causar uma inflação no erro tipo I. (MONTGOMERY, 2005).

A verificação dos grupos homogêneos de médias é feita com testes de comparações múltiplas, entre eles podem-se citar: testes de Tukey e LSD (*Least Significant Difference*). Cada teste emprega estatísticas próprias o que faz com que uns sejam mais conservativos do que outros. Estes métodos estatísticos ajudam na discussão e na tomada de decisão em relação as questões propostas na descrição do problema.

Finalmente, as saídas gráficas dos resíduos, gráficos de barras das médias com seus intervalos de confiança, Box-plot, etc. se complementam e convergem para um melhor aproveitamento das informações conseguidas por meio de uso dessa metodologia, “A ANOVA para modelos com efeitos fixos”.

2.5 Os componentes da variância

Os componentes da variância podem ser estudados sempre que se deseje identificar quem está causando maior variabilidade em um processo que supostamente está operando sob condições constantes. O procedimento é o mesmo que é feito com o modelo dos efeitos fixos, (usando a

tabela ANOVA), levando-se em conta as definições de cada componente. Para ilustrar o comentário será apresentado o modelo de planejamento experimental para uma situação com três componentes, nos quais a hierarquia dos componentes seria: processo (p. ex. batelada), amostragem do produto (plano de amostragem) e medição (variação analítica, ou do próprio teste).

Quando um dado processo, que supostamente deveria estar operando em condições estabelecidas, apresenta uma grande variabilidade, é importante saber quais são as etapas da fabricação até a medição final da propriedade que qualifica o produto. A seguir verificar o número de etapas desta seqüência e fazer um planejamento hierárquico do tipo apresentado na Figura 8. Após a realização dos experimentos proceder a análise estatística para o modelo dos efeitos aleatórios e interpretação dos resultados. Uma verificação deste tipo traz uma série de vantagens, entre elas:

1. saber onde se direcionar maior esforço para reduzir a variabilidade;
2. esclarecer se é adequado estimar a variância do erro em um experimento comparativo;
3. indicar que é possível fazer um projeto de amostragem a um custo mínimo testando o procedimento para qualquer produto particular.

Um arranjo alternativo para o estudo dos componentes da variância é chamado de “*nested*” ou hierárquico. A Figura 8 mostra um arranjo desta natureza para um exemplo no qual seriam realizadas **B** bateladas, cada batelada foi amostrada duas vezes ($S = 2$) e de cada amostra foram feitas duas análises (testes) ($T = 2$).

Neste caso os componentes da variância seriam:

$$\sigma_y^2 = \sigma_p^2 + \sigma_s^2 + \sigma_T^2$$

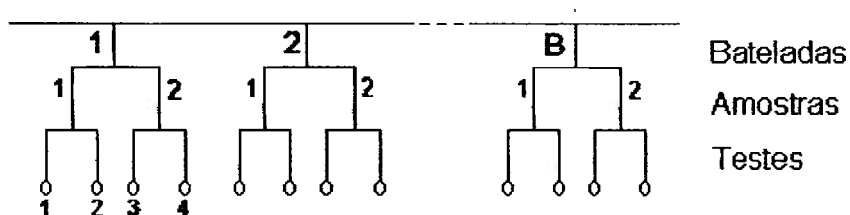


Figura 8 Esquema de montagem de um experimento hierárquico $k \times 2 \times 4$, para estimar os componentes da variância

As quantidades estimadas, usando os resultados da tabela de análise de variância para o projeto esquematizado na Figura 8, seria como ilustrado na Tabela 7 a seguir.

Tabela 7 - Esquema geral da ANOVA para a separação dos componentes da variância associadas com bateladas, amostras e testes, (Box et al. 1978)

Fonte de variação	Soma de quadrados	Graus de liberdade	Média de desvios quadráticos	Valor esperado da média de desvios quadráticos
Grande Média	$BST\bar{y}^2$	1		
Bateladas	$ST \sum_b (\bar{y}_b - \bar{y})^2$	B-1	m_B	$ST\sigma_B^2 + T\sigma_s^2 + \sigma_T^2$
Amostras	$T \sum_b \sum_s (\bar{y}_{bs} - \bar{y}_b)^2$	B(S-1)	m_S	$T\sigma_s^2 + \sigma_T^2$
Testes	$\sum_b \sum_s \sum_t (y_{bst} - \bar{y}_{bs})^2$	BS(T-1)	m_T	σ_T^2
Total	$\sum_b \sum_s \sum_t y_{bst}^2$	BST		

Os valores estimados para os componentes são encontrados por meio da solução do sistema de equações:

$$ST\sigma_B^2 + T\sigma_s^2 + \sigma_T^2 = m_B$$

$$T\sigma_s^2 + \sigma_T^2 = m_S \quad (14)$$

$$\sigma_T^2 = m_T$$

Uma vez conhecidos os componentes, pode-se também estimar a porcentagem de variação de cada componente em relação a variabilidade total. Esta é uma saída do programa MINITAB.

2.5.1 O exemplo dos 10 fardos de “*Filter Tow*”

Para ilustrar um arranjo hierárquico será apresentado um exemplo desenvolvido pelo aluno Roberto Nasser Júnior em sua Dissertação de Mestrado (Impacto das variáveis de produção na qualidade do “*Filter Tow*”), o exemplo dos 10 fardos de “*Filter Tow*”.

Após o tratamento dos dados históricos e construção de modelos empíricos para serem usados na compreensão da influência das variáveis, ainda permanecia uma grande variabilidade não explicada pelo modelo. Então a dúvida era, qual a causa dessa variabilidade? O processo? A amostragem? As medições?

O produto acabado é acondicionado em grandes cubos prensados denominados fardos, dos quais são tiradas as amostras para realização de medidas de propriedades que conferem qualidade ao produto, no caso a medida estudada foi a tensão de ruptura em daN/cm². No projeto hierárquico de verificação dos componentes da variância, a amostragem não foi estudada, porque de cada fardo apenas uma amostra foi retirada e de cada uma delas foram feitas 10 medidas de resistência. Portanto, o exemplo se constitui em 10 fardos representando

a variabilidade devido ao processo e 10 medidas de cada fardo representando o componente devido a variabilidade da medição, com os resultados apresentados na Tabela 8.

Tabela 8 - Resultados das medições da resistência, em daN/cm², feitas para o exemplo dos 10 fardos de “Filter Tow”

Fardo 1	Fardo 2	Fardo 3	Fardo 4	Fardo 5	Fardo 6	Fardo 7	Fardo 8	Fardo 9	Fardo 10
12,3	10,1	11,5	12,8	10,0	12,0	11,9	10,3	9,9	10,1
11,4	12,0	12,6	10,8	11,8	9,3	11,0	11,1	12,9	10,7
8,3	11,6	11,6	11,5	10,4	11,5	12,9	10,2	8,8	12,6
11,3	11,9	10,0	11,3	11,5	12,5	9,8	10,9	9,1	13,8
9,1	12,3	11,0	12,0	10,6	10,4	10,7	9,7	11,0	11,7
11,2	10,9	9,9	9,8	11,1	9,2	11,1	10,4	9,1	10,6
10,9	11,0	10,5	9,7	12,1	10,6	9,0	11,6	7,9	10,5
11,6	10,7	12,2	10,8	11,0	10,1	11,3	10,0	9,0	10,4
12,4	13,9	12,7	11,9	11,9	10,0	11,6	11,1	7,2	11,1
9,7	10,3	11,3	10,9	10,6	10,3	11,1	10,0	10,0	10,1

Neste caso em particular, a falha está na medição e não no processo de fabricação do fardo. Isto remete ao que foi dito anteriormente, todo o esforço para diminuir a variabilidade apresentada deve ser concentrado no método de medição da propriedade. O estudo mostrou que o processo em si está melhor controlado

Tabela 9 - Análise de variância para o teste da resistência do exemplo dos 10 fardos de “Filter Tow”

Fonte de variação	Graus de Liberdade	Soma de quadrados	Média de desvios quadráticos	Razão F	Valor de p
Fardo	9	29,17	3,24	2,636	0,009
Medidas	90	110,67	1,23		
Total	99	139,84			

Tabela 10 - Os componentes da variância para exemplo dos 10 fardos de “*Filter Tow*”

Fonte de variação	Componentes da Variância	% da variabilidade Total	Desvio Padrão
Fardo	0,201	14,06	0,448
Medidas	1,230	85,94	1,109
Total	1,431		1,196

Capítulo 3

Estudo de caso aplicando um projeto fatorial “*three-way*”

Um ciclone lavador adaptado de um ciclone tipo americano

Este caso mostra a aplicação de um projeto fatorial “*three-way*” na verificação do desempenho de um ciclone lavador desenvolvido pelo aluno de Doutorado Ryszard Majewski, para obtenção do título de Doutor em Engenharia pela Escola Politécnica da Universidade de São Paulo em 2006 (Majewski, 2006). O detalhamento do projeto e construção do ciclone intitulado “ciclone lavador” encontra-se na tese ou na publicação (Majewski, R.; Boulos, R. E. e Taqueda, M. E. S., 2008). O enfoque neste estudo será dado ao uso da técnica da análise de variância na forma de um projeto fatorial “*three-way*” com o intuito de verificar a influência dos fatores no desempenho do ciclone lavador.

3.1 O problema

Verificar o comportamento operacional de um ciclone com dimensões características de um ciclone tipo americano modificado (com a instalação de bicos nebulizadores em seu corpo para umidificar e aglomerar a poeira facilitando a separação) para ser empregado na coleta de partículas com dimensão média de 7,48 μm . Um ciclone tipo americano é caracterizado por sua portabilidade. Suas dimensões são caracterizadas por ter altura aproximadamente igual ao diâmetro do corpo. Foi muito usado em minas de carvão, mas sua eficiência de coleta quando ele opera como ciclone seco é inferior a 70% para partículas com diâmetro médio de 20 μm e cerca de 15% para partículas com 7,48 μm , cuja distribuição de tamanhos está apresentado na Figura 9. O desafio deste estudo é modificar o ciclone americano para que ele se torne

mais eficiente para a coleta de partículas menores, de modo que possa ser utilizado na purificação de gases em ambientes fechados.

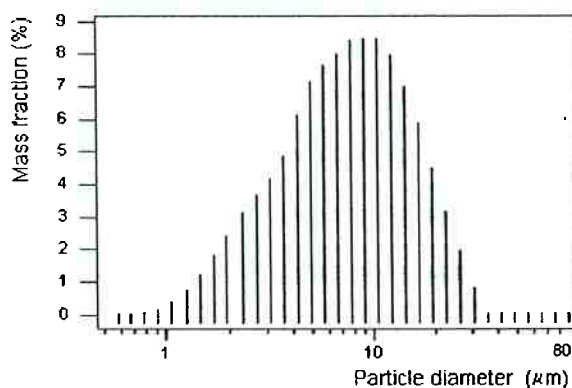


Figura 9. Distribuição do tamanho das partículas de quartzo micronizado (95% < 20 μm), usado como poeira neste estudo

3.2 A proposta de solução

Diferentemente dos ciclones estudados na literatura e encontrados no comércio, o ciclone tipo americano aqui apresentado foi construído de forma modular, fato que possibilita a alteração da altura da parte cilíndrica do ciclone. Na fase de projeto e construção, o ciclone tipo americano lavador foi provido de uma chaminé móvel (duto de saída). Esta característica construtiva possibilitou a variação da altura da chaminé em três posições, a saber: 425 mm, 525 mm e 625 mm. A primeira posição é característica do ciclone americano padrão e as demais são alturas que foram testadas para verificação do efeito da altura da chaminé no desempenho operacional do equipamento. A modificação geral do ciclone também inclui a possibilidade de variação da temperatura do gás a ser purificado, de 20 a 100°C. O indicador de desempenho (variável de resposta da primeira fase de ensaios) foi a diferença de temperaturas de entrada (após o aquecimento) e de saída do gás arrastador. A escolha desta

resposta foi devido ao fato de se avaliar a capacidade de resfriamento do ar durante sua passagem no ciclone lavador, principalmente, quando se deseja purificar gases quentes de combustão de caldeiras e geradores de calor, antes de serem liberados na atmosfera. A verificação do comportamento veio da aplicação do planejamento fatorial citado seguido de um teste de verificação da eficiência de coleta nas melhores condições de operação do equipamento. Portanto, os ensaios foram realizados em duas etapas. O indicador de desempenho (resposta da segunda etapa de ensaios) foi a eficiência de coleta do material particulado, medida experimentalmente. Esta estimativa foi realizada por meio de testes com aerossol (ar mais material particulado).

3.3 Planejamento e obtenção dos resultados dos testes da primeira etapa

Os ensaios foram arranjados de acordo com o planejamento fatorial com três fatores, seguindo o modelo dos efeitos fixos proposto na equação 15. O número total de observações dessa etapa foi $3 \times 5 \times 5 = 75$ realizados em apenas uma réplica. Empregou-se a soma dos quadrados bem como o número de graus de liberdade da interação de três fatores para estimar o erro residual do modelo.

Analisaram-se as seguintes variáveis:

1. Variável de resposta: diferença de temperatura de entrada (após o aquecimento) e saída do portador;
2. Variáveis controláveis: número de aquecedores em funcionamento (0, 1, 2, 3, 4); número dos bicos em funcionamento (0, 1, 2, 3, 4); altura da chaminé (3 posições), 425, 525 e 625 mm;

3. Variáveis monitoráveis: temperaturas de bulbo seco e de bulbo úmido do ar (usado como arrastador), na entrada e na saída do ciclone.

O modelo da análise de variância para os efeitos fixos testado está apresentado na equação 15.

$$y_{ijk} = \eta + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} \quad \left\{ \begin{array}{l} i = 1, 2 \text{ e } 3 \\ j = 0, 1, 2, 3, \text{ e } 4 \\ k = 0, 1, 2, 3, \text{ e } 4 \end{array} \right. \quad (15)$$

Sendo:

y a diferença de temperatura entre o gás de entrada (após aquecimento) e o gás de saída; η a grande média; τ_i é o efeito da chaminé; β_j o efeito da vazão de água; γ_k o efeito da temperatura; $(\tau\beta)_{ij}$, $(\tau\gamma)_{ik}$ e $(\beta\gamma)_{jk}$ os efeitos das interações de dois fatores; $(\tau\beta\gamma)_{ijk}$ o efeito da interação de três fatores, aqui usado para estimar o erro residual.

Os resultados dos 75 experimentos planejados na primeira etapa, foram conduzidos de acordo com o planejamento fatorial com três fatores, e estão apresentados na Tabela 11. As 75 diferenças de temperatura estão no corpo da tabela. A tabela também mostra todo o arranjo dos ensaios em uma ordem padrão, porém a seqüência de realização foi aleatória para cumprir uma das considerações do modelo da equação 15.

A Tabela 11 apresenta a ANOVA com as entradas para todos os efeitos, já a Tabela 12 mostra a ANOVA do modelo, isto é, a hipótese testada é apenas da significância do modelo como um todo em relação ao erro residual.

A significância dos efeitos e do modelo apresentados nas Tabela 12 e Tabela 13 foram julgados a um nível $\alpha = 0,05$. Pela Tabela 12 vê-se que apenas o efeito da interação chaminé**x**bicos não foi significativo ao nível de significância especificado, porque $p > 0,05$. Devido a significância das interações de dois fatores envolvendo os três efeitos principais, estes últimos não podem ser interpretados individualmente. Assim, por exemplo,

considerando-se a chaminé vê-se que o seu efeito depende também do número de aquecedores ativos. Esta interação está representada graficamente na Figura 10. Para os bicos nebulizadores ocorre o mesmo, só que a sua interação é com os aquecedores. A Figura 11 revela este comportamento.

Tabela 11 - As várias condições de operação, assumindo todos os níveis para o planejamento fatorial com três fatores, usado para obter os resultados da etapa um. Os valores da diferença de temperatura, °C, entre o gás de entrada (após o aquecimento) e o gás de saída, estão no corpo da tabela

Chaminé (posição)⇒	1				2				3						
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
Nº de aquecedores ativos ⇒															
Nº de bicos nebulizadores ↓															
0	-0,3	13,0	25,8	36,8	37,7	-2,0	12,8	24,4	37,1	36,7	-2,2	11,7	24,9	37,1	51,4
1	-1,0	15,9	28,9	36,1	45,8	-1,8	12,6	28,0	42,0	48,2	-2,5	13,0	27,5	42,4	57,7
2	-2,7	17,5	32,1	50,2	53,4	2,7	14,2	29,8	45,3	52,9	-1,8	16,0	30,2	46,2	61,0
3	-0,6	18,4	34,4	46,5	57,0	-3,1	15,0	29,9	46,4	55,6	-1,3	16,8	30,5	47,6	62,7
4	0,5	17,8	33,2	50,1	56,8	-1,0	16,4	27,5	48,0	54,8	-0,4	20,5	30,9	46,2	62,7

Tabela 12 - Análise de variância do projeto fatorial three-way, mostrando a significância dos efeitos principais, das interações de dois fatores e o erro baseado no residual

Fonte de variação	G.L.	Soma de Quadrados	Média de desvios quad.	Valor de F	Valor de p
Efeitos principais					
Chaminé	2	63,81	31,91	9,71	0,00051
Aquecedores	4	28269,34	7067,34	2149,97	0,00000
Bicos	4	688,16	172,04	52,34	0,00000
Interações de dois fatores					
Chaminé×Aquecedores	8	258,20	32,28	9,82	0,00000
Chaminé×Bicos	8	22,56	2,82	0,86	0,56095
Aquecedores×Bicos	16	336,22	21,01	6,39	0,00000
Erro	32	105,19	3,29		
Total	74	29743,48			

Tabela 13 - Análise de variância para o modelo proposto

	Soma de Quadrados	G.L.	Média de desvios quad.	Valor de F	Valor de p
Modelo	29638,29	42	705,67	214,67	0,00000
Erro residual	105,19	32	3,29		

$$R^2 = 0,996 \text{ e } R_{aj}^2 = 0,992$$

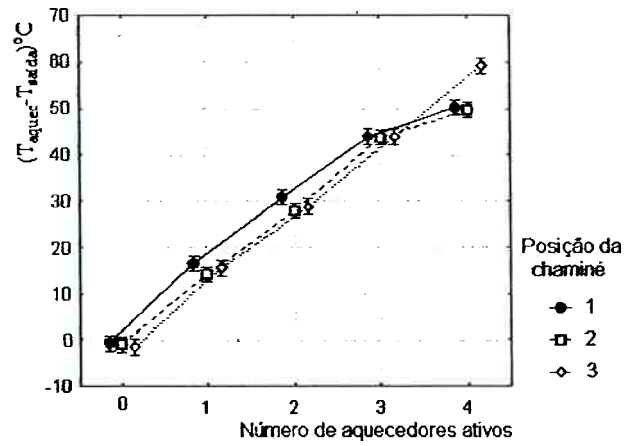


Figura 10. Efeito da interação posição da chaminé \times aquecedores ativos na diferença de temperatura entre a entrada (após aquecimento) e a saída do gás. $F_{8,32} = 9,82$ e $p = 0,00000$. Barras verticais denotam os intervalos de confiança

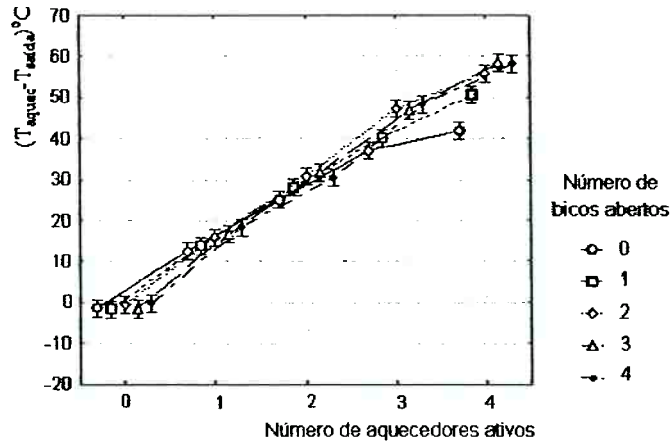


Figura 11. Efeito da interação número de bicos abertos \times aquecedores ativos na diferença de temperatura entre a entrada (após aquecimento) e a saída do gás. $F_{16,32} = 9,82$ e $p = 0,00000$. Barras verticais denotam os intervalos de confiança

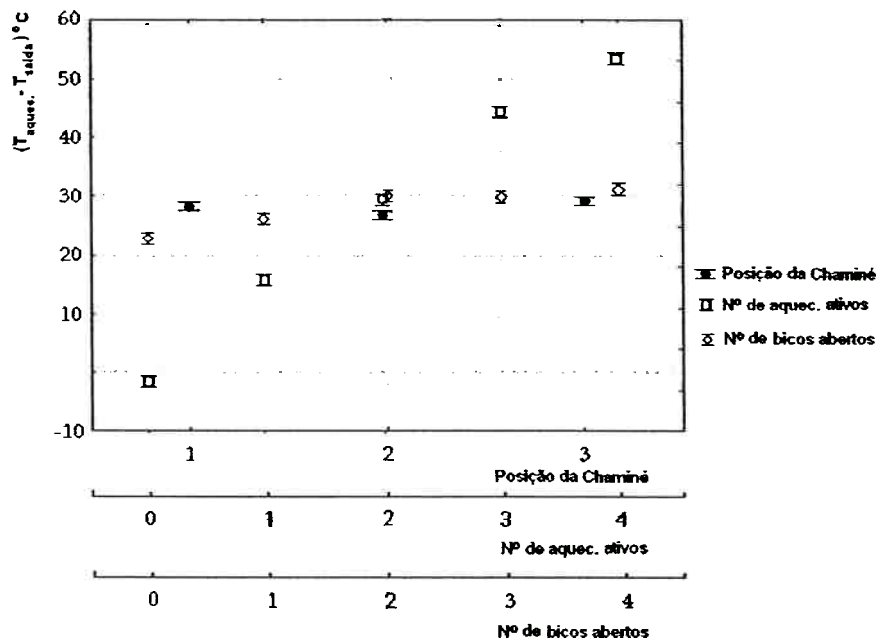
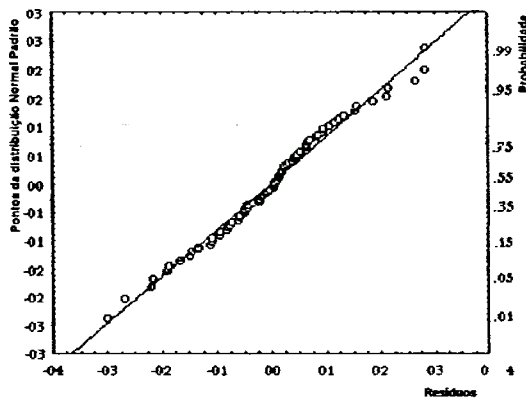


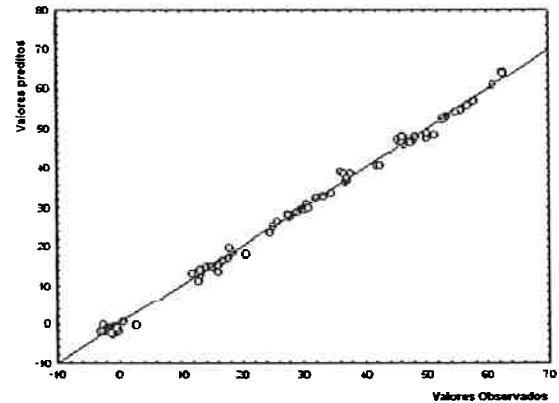
Figura 12. Efeito da posição da chaminé, dos aquecedores ativos e do número de bicos abertos na diferença de temperatura entre a entrada (após aquecimento) e a saída do gás. $F_{2,32} = 9,71$; $p = 0,00051$; $F_{4,32} = 2150$; $p = 0,00000$; $F_{4,32} = 52,34$; $p = 0,00000$. Barras verticais denotam os intervalos de confiança

Este resultado foi corroborado pelas Figura 13(a) e Figura 13(b), respectivamente os resíduos expressos em papel de probabilidade normal (valores se ajustam a uma reta) e os valores preditos em função dos valores observados (valores tendem a uma reta de 45°).

Cinco ensaios de confirmação para as condições de operação mais drástica, na quais seria testado o aerosol, confirmaram a estabilidade do ciclone frente as perturbações impostas. A Tabela 14 apresenta os resultados obtidos.



(a)



(b)

Figura 13. Diagramas de testes de adequação dos resultados ao modelo proposto. (a) Resíduos em escala de probabilidade normal e (b) valores preditos em função dos valores observados

Tabela 14 – Resultados obtidos para o teste de confirmação das condições impostas e validação do modelo. Todos os resultados estão no corpo da tabela

Ens.	ΔP da placa (kPa)	Vazão de gás (m^3/s)	ΔP ciclone (kPa)	Cond. de entrada do Gás			Aquec.			Cond. De saída do gás				$\Delta \mathcal{H}_f$ (g/kg)	Temp. da água $t_{\text{água}}$ ($^{\circ}C$)
				t ($^{\circ}C$)	t_{bu} ($^{\circ}C$)	\mathcal{H}_f (%)	\mathcal{H} (g/kg)	t ($^{\circ}C$)	t ($^{\circ}C$)	t ($^{\circ}C$)	t_{bu} ($^{\circ}C$)	\mathcal{H}_f (%)	\mathcal{H} (g/kg)		
1	1,91	0,358	1,47	14,0	13,6	95,9	10,0	95,0	32,0	25,8	61,8	19,7	9,7	25,5	
2	1,91	0,358	1,47	14,2	13,8	95,9	10,2	95,0	31,5	26,0	65,4	20,2	10,0	25,5	
3	1,91	0,358	1,47	14,0	13,8	97,9	10,3	94,0	32,0	26,2	63,9	20,4	10,1	25,5	
4	1,91	0,358	1,47	14,0	13,8	97,9	10,3	93,0	32,0	26,2	63,9	20,4	10,1	25,5	
5	1,91	0,358	1,47	14,0	13,8	97,9	10,3	96,0	32,0	26,2	63,9	20,4	10,1	25,5	

3.4 Resultados dos ensaios do teste da eficiência de coleta

Os dados obtidos na primeira etapa mostraram que 99% da variabilidade do processo é explicada pelo modelo. Portanto, o modelo proposto representa muito bem os dados do estudo da primeira etapa. Então, escolheu-se um conjunto de condições com bom desempenho da resposta para testar o ciclone com o aerossol (ar+poeira). Nestes testes fez-se o uso simultâneo do sistema de dosagem de material particulado, instalado no duto de entrada do equipamento, e do sistema de medição da concentração residual de material particulado instalado no duto de saída (MAJEWSKY, 2006; MAJEWSKY, BOULOS, e TAQUEDA, 2008).

Para realizar a medição da eficiência de coleta, a câmara cilíndrica do sistema de dosagem do material particulado foi alimentada manualmente com uma massa de aproximadamente 2,8 kg do material particulado (previamente seco em estufa apropriada). Após a introdução do material particulado, a câmara cilíndrica é fechada e ligada à rede de ar comprimido, responsável pela injeção do material.

A determinação do material particulado coletado através da sonda foi definida através da pesagem dos filtros antes e depois da coleta, utilizando-se uma balança digital com 0,1 mg de precisão.

A eficiência de separação foi calculada a partir da relação:

$$\eta = \frac{C_{ent} - C_{sai}}{C_{ent}} \quad (16)$$

Em que:

η é a eficiência (%), $C_{\text{saí}}$ é a concentração de material particulado na saída do lavador de gases (mg/m^3) e C_{ent} é a concentração de material particulado na entrada do lavador de gases (mg/m^3).

Finalmente, após a análise e interpretação do comportamento de operação frente as variáveis testadas, pensou-se em verificar a capacidade de coleta do ciclone em pelo menos um conjunto de condições de operação do ciclone lavador.

Estabeleceu-se uma condição fixa de operação do ciclone, máxima temperatura de aquecimento do gás (4 aquecedores ativos), maior vazão de água (4 bicos abertos) e a chaminé na posição 3. Nestas condições de operação, realizaram os ensaios para o aerossol, estimando assim a eficiência de coleta de ciclone, para a faixa de concentração de poeira até cerca de $500 \text{ mg}/\text{m}^3$.

A Tabela 15 bem como a Figura 14 mostram o comportamento dos testes para a eficiência de coleta.

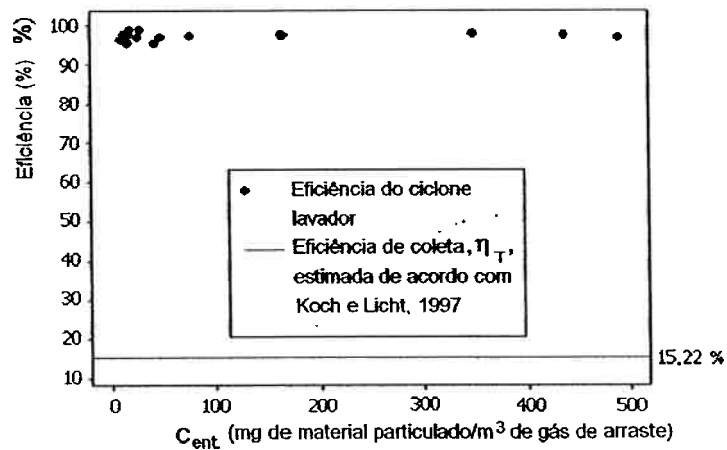


Figura 14. Efeito da concentração de particulado na eficiência de coleta do ciclone lavador

Tabela 15 - Eficiência de coleta para partículas com diâmetro médio 7,48 μm , quando o ciclone lavador operou com 4 aquecedores ligados, 4 bicos abertos, chaminé na posição 3, vazão de ar 0,358 m^3/s e $\Delta P=1,47$ kPa

Ens.	C_{ent} (mg/m^3)	$C_{\text{saída}}$ (mg/m^3)	Eficiência		Cond. de ent. do gás				Aquec.				Cond. de saída do gás				$\Delta \mathcal{H}$ (g/kg)	Temp. da água, $t_{\text{água}}$ ($^{\circ}\text{C}$)
			η_c (%)	η_e (%)	t ($^{\circ}\text{C}$)	T_{bu} ($^{\circ}\text{C}$)	\mathcal{H}_f (%)	\mathcal{H} (g/kg)	t ($^{\circ}\text{C}$)	t ($^{\circ}\text{C}$)	tw ($^{\circ}\text{C}$)	\mathcal{H}_f (%)	\mathcal{H} (g/kg)	t ($^{\circ}\text{C}$)	tw ($^{\circ}\text{C}$)	\mathcal{H}_f (%)		
1	47,37	1,45	96,94	96,94	30,8	22,8	51,1	15,1	94,8	39,0	31,5	59,5	28,3	13,2	31,5			
2	76,49	2,07	97,29	97,29	32,2	22,8	45,6	14,6	98,0	39,0	31,4	59,0	28,0	13,4	31,5			
3	164,98	4,00	97,58	97,58	31,8	21,6	41,4	12,9	98,0	39,0	31,0	57,2	27,0	14,1	31,0			
4	348,68	8,00	97,71	97,71	29,6	21,8	51,3	14,1	97,2	38,0	30,4	58,4	26,1	12,0	30,0			
5	14,01	0,55	96,07	96,07	19,2	15,8	71,2	10,4	82,0	32,0	25,8	61,8	19,7	9,3	26,0			
6	9,13	0,34	96,28	96,28	23,6	20,6	76,9	14,9	89,0	36,0	29,6	63,1	25,3	10,4	30,0			
7	16,21	0,34	97,90	97,90	21,8	20,8	91,7	15,9	84,0	34,0	28,4	66,2	23,7	7,8	27,0			
8	25,34	0,76	97,00	97,00	22,0	20,6	88,5	15,5	82,0	31,0	25,4	64,5	19,4	3,9	24,0			
9	17,48	0,21	98,80	98,80	22,8	20,2	79,5	14,6	88,0	35,0	29,4	66,7	25,3	10,7	29,0			
10	27,38	0,34	98,76	98,76	20,7	16,9	69,2	11,15	84,0	32,5	26,9	65,4	21,45	10,3	27,0			
11	42,02	2,00	95,24	95,24	18,0	15,4	76,9	10,5	78,0	28,0	24,4	74,9	18,9	8,4	21,0			
12	10,31	0,34	96,70	96,70	20,2	16,4	68,8	10,7	82,0	32,5	26,4	62,6	20,5	9,8	26,0			
13	489,37	16,00	96,73	96,73	30,2	19,9	39,3	11,1	95,3	35,1	26,5	51,7	19,6	8,5	25,8			
14	435,99	11,52	97,36	97,36	30,6	19,8	37,4	10,8	96,5	36,6	28,3	54,2	22,4	11,5	28,0			
15	11,10	0,28	97,48	97,48	22,0	17,4	64,3	11,18	86,6	33,5	27,1	61,7	21,4	10,3	27,0			
16	14,74	0,69	95,32	95,32	20,8	16,4	64,8	10,48	84,0	32,9	26,6	61,6	20,7	10,2	26,2			

3.5 Perspectivas do ciclone lavador

A análise do ciclone transformado para este estudo mostrou as informações necessárias para a adequação de seus elementos ao consumo mínimo de energia elétrica, o que o torna mais rentável. Além disso, apesar de compacto, esse equipamento pode exercer várias funções, tais como: umidificação, resfriamento, retenção de partículas, tratamento bioquímico da água e resfriamento abaixo da temperatura ambiente (ar condicionado).

O referido equipamento pode ser montado sobre uma carreta, o que confere uma característica de mobilidade, tornando-o muito útil para deslocamentos (dentro ou fora do ambiente fabril).

3.6 Questões adicionais

Apesar do amplo estudo relatado neste caso, algumas questões centrais permaneceram:

1. A eficiência deste ciclone poderia ser melhorada para partículas $<1 \mu\text{m}$?
2. Caso fosse melhorada a sua eficiência para coletar partículas ainda menores, serviria para esterilizar ambientes hospitalares?
3. Quais as variáveis que poderiam ser adicionadas em experiências futuras?

Capítulo 4

Projeto fatorial em dois níveis

Como visto no capítulo anterior, a análise de variância é muito útil na comparação de fatores qualitativos como, por exemplo, tipos de catalisadores, departamentos, fornecedores, dentre outros. Embora possam ser incluídos fatores quantitativos, como foi observado no capítulo 2, a ANOVA (para o modelo de efeitos fixos) apenas indica a significância ou não do efeito dos fatores. As comparações entre as médias, duas a duas, poderá ser adicionada por outro método de comparação múltiplas, como o teste de Tukey.

Os projetos fatoriais gerais, em contrapartida, são capazes de medir efeitos, ou seja eles mensuram o quanto as variações impostas aos fatores são desviadas da média. Isto pode ser feito com variáveis tanto qualitativas quanto quantitativas. Se as variáveis em estudo são quantitativas eles permitem ainda a construção de modelos empíricos.

O que é, então, um projeto fatorial geral? É um arranjo experimental onde podem ser feitas todas as combinações dos níveis dos fatores. A série experimental que se constitui no planejamento é chamada de réplica (“*replication*”) do experimento. Supondo-se que se queira estudar: a catalisadores, b níveis de temperatura, c níveis de agitação e d níveis da vazão de alimentação em um reator, a série experimental seria constituída de $a \times b \times c \times d$ experimentos por réplica. A questão mais importante neste caso seria: quanto vale cada um desses níveis? Um experimentador mais cético poderia pensar em $a=3$, $b=5$, $c=4$ e $d=3$. Então, neste caso, a série experimental seria constituída de 180 ensaios. Dependendo do custo e do tempo gastos por ensaio, a pesquisa seria inviável. No entanto se: $a=b=c=d=2$, ter-se-ia $2^4=16$ ensaios, isto é, uma série experimental cerca de 11 vezes menor do que a série anterior. Então quando se tem k variáveis (fatores) todos estudados em dois níveis, o esquema é chamado de planejamento fatorial em dois níveis muito conhecido como DOE 2^k (“*design of experiments*”).

ou *experimental design*”). Quando é realizado a série 2^k de experimentos ela é chamada de projeto completo (“*full factorial design*”).

Segundo Box, Hunter, W e Hunter, J. (2005), esses projetos são de especial importância porque:

1. requerem relativamente poucos ensaios por fator;
2. a interpretação das observações produzidas pelo projeto pode ser feita com bom senso, usando aritmética elementar e gráficos de visualização de dados (saída comum nos dos *softwares* estatísticos);
3. eles indicam uma promissora direção para ensaios adicionais quando os fatores são quantitativos, embora sejam incapazes de explorar uma larga faixa do espaço fator;
4. podem ser convenientemente aumentados quando se torna necessária uma investigação minuciosa do espaço fator, em um processo chamado “*seqüential assembly*”;
5. são base para os projetos fatoriais fracionados em dois níveis, que serão discutidos no capítulo 6. Os fracionados podem ser uma parte cuidadosamente escolhida de um fatorial completo. Eles são particularmente empregados em investigações em que se suspeitam de muitos fatores que podem causar variabilidade. Neste caso, eles servem para selecionar as variáveis mais importantes e o processo é conhecido como “*factor-screening*”. Ainda podem ser usados na construção de blocos de modo a apresentar soluções requintadas de problemas complexos, em uma estratégia seqüencial;
6. os projetos fatoriais completos, e os projetos fatoriais fracionados correspondentes, se ajustam naturalmente à estratégia seqüencial, uma característica essencial do método científico, mencionado no capítulo 1.

4.1 Aplicando um DOE 2^k

De posse das variáveis independentes que potencialmente podem afetar um determinado processo, a construção do projeto fatorial em dois níveis consiste em estabelecer os limites de variação de cada variável. Estes são os níveis que devem ser codificados em nível baixo (–) e nível alto (+), para as variáveis independentes x_i . Já a variável independente (resposta) em geral é identificada por y . Para variáveis quantitativas é fácil enxergar os limites superior e inferior dos níveis. Já para variáveis qualitativas é necessário que o experimentador defina como serão estabelecidos os níveis, baixo e alto. Como por definição, o projeto fatorial completo em 2 níveis é construído com todas as combinações dos níveis, então a série experimental é constituída de 2^k experimentos, em que k é o número de fatores.

Tanto o método de construção do projeto quanto os métodos de estimativa dos efeitos serão ilustrados por um exemplo simples tirado da literatura.

4.1.1 O exemplo de redução de poluentes de uma corrente de resíduos

Com o intuito de reduzir a quantidade de poluentes em uma corrente de resíduos de uma pequena fábrica de moldes de plásticos, deve-se tratar a mesma antes de ser descartada. A lei declara que diariamente o limite de poluentes não deve ultrapassar 4,5 kg. Um experimento com 8 ensaios foi conduzido com o objetivo de estabelecer um procedimento do tratamento da corrente antes de ser descartada, atendendo a legislação vigente.

Foram testados as marcas químicas A e B, dois níveis de temperatura, e dois níveis de agitação.

4.1.1.1 O projeto completo com 3 fatores

A matriz de ensaios na ordem padrão de construção do projeto, em variáveis originais e codificadas para o exemplo dos poluentes está apresentada na Tabela 16.

Tabela 16 - Matriz de ensaios em variáveis originais e codificadas para o exemplo dos poluentes

	Marca química		Temperatura (°C)		Agitação		
	-	+	-	+	-	+	
	A	B	22	38	não	sim	
Ordem padrão dos ensaios	Unidade codificada das variáveis						Poluente (kg/dia)
	x_1		x_2		x_3		
1	-		-		-		2,27
2	+		-		-		13,62
3	-		+		-		2,72
4	+		+		-		14,98
5	-		-		+		1,82
6	+		-		+		1,36
7	-		+		+		2,27
8	+		+		+		1,82
Unidades originais das variáveis							
Ordem padrão dos ensaios	Marca Química M		Temperatura (°C) T		Agitação S		Poluente (kg/dia)
1	A		22		não		2,27
2	B		22		não		13,62
3	A		38		não		2,72
4	B		38		não		14,98
5	A		22		sim		1,82
6	B		22		sim		1,36
7	A		38		sim		2,27
8	B		38		sim		1,82

Observação: os ensaios foram realizados em ordem aleatória para garantir o uso dos métodos estatísticos.

Pode-se observar que existe um padrão de geração da tabela. Assim, a primeira coluna é formada por uma seqüência de sinais - (negativo) e sinais + (positivo) alternados; a segunda

coluna de seqüências de grupos de dois – e grupos de dois +; a terceira coluna de seqüências de grupos de quatro – e grupos de quatro +; e assim por diante, numa progressão geométrica. O projeto pode ser construído com esse ou outro padrão para assegurar a individualidade de cada ensaio, mas é importante reforçar que a seqüência de realização dos ensaios deve ser aleatória. O resultado obtido para cada ensaio está na última coluna como número de kg/dia de poluente descarregado. A Tabela 17 mostra como são construídos os projetos fatoriais completos em dois níveis para qualquer número de fatores.

Tabela 17 - Notação mais conveniente e usual para a matriz de ensaios em variáveis originais e codificadas

Ensaio	Variáveis independentes			
	x_1	x_2	x_3	x_4
1	–	–	–	–
2	+	–	–	–
3	–	+	–	–
$4=2^2$	+	+	–	–
5	–	–	+	–
6	+	–	+	–
7	–	+	+	–
$8=2^3$	+	+	+	–
9	–	–	–	+
10	+	–	–	+
11	–	+	–	+
12	+	+	–	+
13	–	–	+	+
14	+	–	+	+
15	–	+	+	+
$16=2^4$	+	+	+	+

4.1.1.2 A estimativa dos efeitos

Antes da realização de qualquer cálculo é interessante observar a tabela com os dados na seqüência padrão porque eles revelam coisas muito interessantes. Por exemplo, observando-se a variação dos níveis de poluentes do ensaio 2 para o 1, nota-se que eles foram realizados nos níveis baixos das outras duas variáveis (temperatura e agitação); do 4 para o 3, foram realizados no nível alto de temperatura e no nível baixo de concentração; do 6 para o 5 foram realizados no nível baixo de temperatura e no nível alto de concentração; e, finalmente, do 8 para o 7, nos níveis altos das duas variáveis. Em suma, a variação dos poluentes foram estudadas sob todas as combinações de níveis das outras. Isto pode ser observado com a temperatura, analisando-se os ensaios: 3 e 1; 4 e 2; 7 e 5; e 8 e 6 (aqui há um salto de 3 para 1, por exemplo), porque é assim que a variável temperatura altera os seus níveis. E finalmente, seguindo o mesmo raciocínio, para agitação, o salto é de 5 para 1; de 6 para 2; de 7 para 3; e, de 8 para 4. Assim, nota-se que as três variáveis foram observadas de maneira análoga. Quando se faz esse mesmo raciocínio e, adicionalmente, calculam-se as diferenças nas respostas para estes experimentos, vê-se claramente como a variação ocorre em cada situação. Daí é possível, sem nenhum conhecimento adicional, perceber o que ocorre com estas variações. A partir desse raciocínio, o efeito principal (individual) de cada variável foi definido como a média das quatro diferenças encontradas, duas a duas, como foi mostrado anteriormente. Para ilustrar melhor, será feito este raciocínio em forma de tabela, para a variável 3 (agitação).

Medidas individuais dos efeitos para a variação da agitação de não (-) para sim (+)	Condições sob as quais as comparações foram feitas	
	Marca química M	Temperatura T
$y_5 - y_1 = 1,82 - 2,27 = -0,45$	A	22
$y_6 - y_2 = 1,36 - 13,62 = -12,26$	B	22
$y_7 - y_3 = 2,27 - 2,72 = -0,450$	A	38
$y_8 - y_4 = 1,82 - 14,98 = -13,16$	B	38
Efeito principal da agitação, $x_3 = -6,58$		

Assim, usando o mesmo raciocínio, encontra-se que: efeito principal do marca química, $x_1 = 5,68$; efeito principal da temperatura, $x_2 = 0,68$.

Observando a matriz de ensaios, vê-se que a descarga de poluente é menor quando a corrente é agitada. Assim, a variável marca química (x_1) e agitação não se comportam aditivamente. Só se pode interpretar o efeito da marca química explicitando se a corrente está ou não sendo agitada. Nesta situação, diz-se que há uma interação entre marca química e agitação, que em variáveis codificadas é a interação $x_1 \times x_3$. A estimativa da interação é feita pela diferença entre o efeito médio da marca química com a corrente agitada e o efeito médio da marca química com a corrente sem agitação. Por convenção, o efeito da interação é metade dessa diferença.

Assim, para a interação entre a marca química e a agitação ($x_1 \times x_3$), tem-se:

Agitação	Efeito médio da marca química
(+) sim	- 0,455
(-) não	- 11,802
Diferença	-12, 26
a interação $x_1 \times x_3 = \frac{-12,6}{2}$	- 6,13

Seguindo-se o mesmo raciocínio, encontram-se as outras duas interações de dois fatores, isto é:

$$x_1 \cdot x_2 = 0,23 \quad (17)$$

$$x_2 \cdot x_3 = -0,23 \quad (18)$$

Para a estimativa da interação de três fatores o procedimento é o descrito a seguir:

Duas interações $x_1 \times x_2$ são estimadas, uma para corrente com agitação e a outra para a corrente sem agitação.

Interação $x_1 \times x_2$, com agitação, x_3 , no nível **sim (+)**:

$$\frac{(y_8 - y_7) - (y_6 - y_5)}{2} = \frac{(1,82 - 2,27) - (1,36 - 1,82)}{2} = \frac{-0,45 - (-0,46)}{2} = 0,005 \quad (19)$$

Interação $x_1 \times x_2$, com agitação, x_3 , no nível **não (-)**

$$\frac{(y_4 - y_3) - (y_2 - y_1)}{2} = \frac{(14,98 - 2,72) - (13,62 - 2,27)}{2} = \frac{12,26 - 11,35}{2} = -0,455 \quad (20)$$

A diferença mede a coerência da interação da marca química com a temperatura, para os dois níveis de agitação. Metade desta diferença é por definição a interação de três fatores.

$$\text{Interação } x_1 \times x_2 \times x_3 = \frac{0,005 - 0,455}{2} = -0,23 \quad (21)$$

Pode se verificar que para o cálculo de quaisquer efeitos, usam-se os resultados dos 8 experimentos. Portanto, cada estimativa é feita com precisão de uma diferença repetida quatro vezes.

2. para a temperatura, x_2 (codificado), plano (-) delimitados por (y_1, y_2, y_5 e y_6) e plano (+) delimitados por (y_3, y_4, y_7 e y_8);
3. para a agitação, x_3 (codificado), plano (-) delimitados por (y_1, y_2, y_3 e y_4) e plano (+) delimitados por (y_5, y_6, y_7 e y_8).

Assim, os efeitos principais também podem ser estimados pela diferença entre as médias do plano (+) e do plano (-).

Genericamente falando:

O efeito principal = $\bar{y}_+ - \bar{y}_-$, que desdobrando para cada variável, tem-se:

$$\text{efeito da marca química, } x_1 = \frac{13,62 + 14,98 + 1,36 + 1,82}{4} - \frac{2,27 + 2,72 + 1,82 + 2,27}{4} = 5,68$$

$$\text{efeito da temperatura, } x_2 = \frac{2,72 + 14,98 + 2,27 + 1,82}{4} - \frac{2,27 + 13,62 + 1,82 + 1,36}{4} = 0,68$$

$$\text{efeito da marca química, } x_3 = \frac{1,82 + 1,36 + 2,27 + 1,82}{4} - \frac{2,27 + 13,62 + 2,72 + 14,98}{4} = -6,58$$

Nota-se que qualquer dos dois métodos de cálculo dos efeitos, até agora apresentados, levam aos mesmos valores. As interações também podem ser calculadas pelas diferenças entre os planos. Isto pode ser visto em detalhes em Box, Hunter, W. e Hunter, J. (2005).

Ainda existem dois métodos de cálculo para os efeitos, a saber: o método rápido e o algoritmo de Yates. Este último é muito útil, na sua forma reversa, para estimar resíduos, quando não se tem recurso computacional disponível.

Para encerrar os métodos de cálculo, será apresentado o método rápido:

A partir do planejamento original, as colunas das interações podem ser geradas, na matriz, multiplicando-se os sinais das colunas apropriadas (anotando-se o sinal resultante na coluna onde foi especificada a interação, como mostra a Tabela 18).

Tabela 18 - Modelo de geração da matriz, para a estimativa dos efeitos pelo método rápido

Ensaio #	Designação dos efeitos							Resposta (saída)
	x_1	x_2	x_3	$x_1 \times x_2$	$x_1 \times x_3$	$x_2 \times x_3$	$x_1 \times x_2 \times x_3$	
1	-	-	-	+	+	+	-	y_1
2	+	-	-	-	-	+	+	y_2
3	-	+	-	-	+	-	+	y_3
4	+	+	-	+	-	-	-	y_4
5	-	-	+	+	-	-	+	y_5
6	+	-	+	-	+	-	-	y_6
7	-	+	+	-	-	+	-	y_7
8	+	+	+	+	+	+	+	y_8

O cálculo de cada efeito é realizado antepondo-se a coluna de sinais, correspondente ao efeito desejado, à coluna de respostas. Em seguida procede-se a operação dos números resultantes dividindo-se o valor obtido pela metade do número de ensaios que compõem o projeto. No caso do exemplo aqui explorado, 4.

Os valores obtidos por qualquer dos 4 métodos é o mesmo.

Vale aqui mencionar que a grande média é estimada pela soma de todos os resultados dividida pelo número total de ensaios, no caso do exemplo, 8.

4.1.3 Vantagens do projeto fatorial sobre o método “uma variável por vez”

O método de experimentação “uma variável por vez” é um método clássico de experimentação no qual um fator é variado, enquanto os demais são mantidos constantes. A desvantagem deste método é que ele proporciona somente a estimativa do efeito de um único

fator variando em condições pré-estabelecidas em que os demais fatores ficam e fixos. Para que tal estimativa tivesse uma relevância mais geral seria necessário considerar que o efeito seria o mesmo em todas as outras condições dos demais fatores, isto é, os fatores afetariam a resposta aditivamente (em uma dependência linear). Porém, se o fator só atua linearmente, o projeto fatorial faz o mesmo trabalho com maior precisão. E, se os fatores não atuam aditivamente (a dependência é não linear) o fatorial, diferentemente do método “uma variável por vez”, pode detectar e estimar a interação que é a medida da não aditividade. Lembrando que, todas as observações são usadas na estimativa dos efeitos de um projeto fatorial em dois níveis, e que os efeitos são estimados como diferenças entre médias, então, para que o efeito estimado pelo método “uma variável por vez” tivesse a mesma precisão do projeto fatorial, 2^3 , seriam necessários 8 experimentos para cada efeito. Isto resultaria em 24 ensaios para a estimativa de 3 três efeitos. De uma maneira geral, para k fatores, seriam necessários $k \times 2^k$ ensaios, no método “uma variável por vez” para se ter a mesma precisão de um projeto fatorial 2^k .

4.1.4 Ensaios com réplica genuína

O experimento dos poluentes poderia ter sido conduzido com 16 ensaios ao invés de 8, por exemplo, tomando-se cada série como uma réplica genuína. Como “réplica genuína” entende-se que a variação de ensaio para ensaio, realizados nas mesmas condições experimentais, é o reflexo da variabilidade total ensaio para ensaio. Isto requer considerações cuidadosas:

1. realizar os 16 ensaios de uma forma totalmente aleatória, para realmente assegurar uma réplica genuína. Isto requer a realização de todas as etapas do ensaio novamente, o que muitas vezes é difícil;

2. a replicação de somente uma análise química de uma amostra de um único ensaio proporcionaria apenas a estimativa da variância analítica. Esta é apenas uma parte da variância experimental ensaio a ensaio (Lembrar dos componentes da variância, vista no capítulo 2, item 2.5);

De maneira semelhantemente, a replicação de amostras tiradas do mesmo ensaio, poderia proporcionar a estimativa da variância da amostragem adicionada a variância analítica.

Deve-se, portanto, tomar muito cuidado ao falar de réplicas genuínas.

4.1.4.1 Estimativa da variância do erro e do erro padrão dos efeitos para ensaios replicados

Uma estimativa da variância com 1 grau de liberdade pode ser dado por $s^2 = d^2/2$, sendo d a diferença entre os resultados dos dois experimentos repetidos genuinamente. Se um projeto 2^3 é replicado genuinamente, gera um par de respostas para cada ensaio, isto é, 8 diferenças d . É importante lembrar, ainda, que o divisor é 2 porque não foi consumido um grau de liberdade para a estimativa da média. A médias dessas 8 variâncias estimadas para cada par, com 8 graus de liberdade, produz a estimativa da variância do erro, s^2

Como a estimativa de cada efeito é uma diferença entre duas médias de 8 observações (16 ensaios no total, ver item 3.2.4), a variância de um efeito é dado por:

$$V(\text{efeito}) = \left(\frac{1}{8} + \frac{1}{8} \right) s^2 \quad (22)$$

A raiz quadrada dessa variância é o erro padrão do efeito.

De uma maneira geral, a variância global estimada para o experimento a partir de g fatores, seria:

$$s^2 = \frac{\nu_1 s_1^2 + \nu_2 s_2^2 + \dots + \nu_g s_g^2}{\nu_1 + \nu_2 + \dots + \nu_g} \quad (23)$$

Em que $\nu = \nu_1 + \nu_2 + \dots + \nu_g$ é o número de graus de liberdade com os quais foi feita a estimativa. No caso de g duplicatas a equação 23 reduz-se a:

$$s^2 = \frac{\sum d^2}{2g}, \text{ com } g \text{ graus de liberdade.} \quad (24)$$

4.1.4.2 A interpretação dos efeitos

A interpretação dos resultados é feita usando uma regra grosseira na qual, efeitos duas ou três vezes maiores do que o erro padrão são considerados significantes. Cada razão *efeito/SE(efeito)*, se distribui como uma distribuição t com ν graus de liberdade. Isto vem da consideração que n observações repetidas são independentes e identicamente distribuídas, consideração freqüentemente abreviada para IID. Se adicionalmente considera-se que esta distribuição é aproximadamente normal, a notação torna-se NIID.

4.1.4.3 Uma estimativa grosseira do erro

Em uma investigação muitas vezes é preferível realizar 16 ensaios com 4 variáveis do que suprimir uma variável da qual não se conhece seu efeito, para realização de duas réplicas. A vantagem de se realizar um projeto 2^4 é que as quatro variáveis que, em princípio, podem interferir no processo têm seus efeitos avaliados. Um projeto com 4 variáveis, estima 15 contrastes além da grande média. São eles: quatro efeitos principais, seis interações de dois fatores, quatro interações de três fatores e uma interação de quatro fatores.

Usando-se a regra da hierarquia dos efeitos, isto é, geralmente os efeitos principais são maiores do que as interações de dois fatores, que são maiores do que as de três fatores e assim

por diante, pode-se negligenciar efeitos de interações de ordem maiores do que dois ou três e assim usar estes valores como resíduos. Embora possa haver exceções essa regra é fundamental para o uso dos planejamentos fatoriais em dois níveis.

Após a estimativa dos efeitos, constata-se, por exemplo para o 2^4 , que as interações de terceira ordem em diante não apresentam valores expressivos, atribuindo-se esses valores apenas como ruído. Usam-se então os valores dessas interações como um conjunto de referência para os outros efeitos. Assim, pode-se estimar o erro padrão dos efeitos, neste caso, pela equação:

$$[SE(\text{efeito})]^2 = \frac{(\text{efeito123})^2 + (\text{efeito124})^2 + (\text{efeito134})^2 + (\text{efeito234})^2 + (\text{efeito1234})^2}{5} \quad (25)$$

Esse procedimento pode ser estendido a outras situações, e a interpretação é feita da mesma maneira que o caso do item 4.1.4.2. O t usado como referência, neste caso, tem 5 graus de liberdade.

4.1.4.4 Uso da escala de probabilidade Normal

Considere o exemplo dos poluentes, em que nenhuma replicação foi realizada. Os efeitos estimados estão na Tabela 19.

Tabela 19 - Os efeitos do exemplo dos poluentes. Projeto 2^3 sem replicação

Média	5,11
Efeitos principais	
Marca química, x_1	5,68
Temperatura, x_2	0,68
Agitação x_3	-6,58
Interações de dois fatores	
$x_1 \times x_2$	0,23
$x_1 \times x_3$	-6,13
$x_2 \times x_3$	-0,23
Interação de três fatores	
$x_1 \times x_2 \times x_3$	-0,23

Observa-se na Tabela que para 8 observações (resultados de 8 ensaios) foram estimadas 8 quantidades: a média, três efeitos principais, três interações de dois fatores e uma interação de três fatores. Portanto, foram consumidos todos os graus de liberdade para estas estimativas. Com isso, não sobraram graus de liberdade para a estimativa do erro padrão dos efeitos. Então como saber quais são os efeitos significativos? Na falta de graus de liberdade para estimativa do erro, ocorrem dois problemas na avaliação dos efeitos:

1. podem ocorrer interações expressivas de alta ordem, (uma exceção a regra da hierarquia dos efeitos); e,
2. acima de qual valor do efeito ele deve ser levado em conta na seleção. Existem efeitos maiores e menores, porém o maior é excepcionalmente maior?

A solução para estas duas situações foi resolvida quando Daniel (1959) propôs um método no qual os efeitos são plotados em gráfico em um papel de probabilidade normal. Este método, hoje, está disponível como saída de quase todos os *softwares* estatísticos.

A Figura 16, ilustra um gráfico desta natureza para o exemplo dos poluentes.

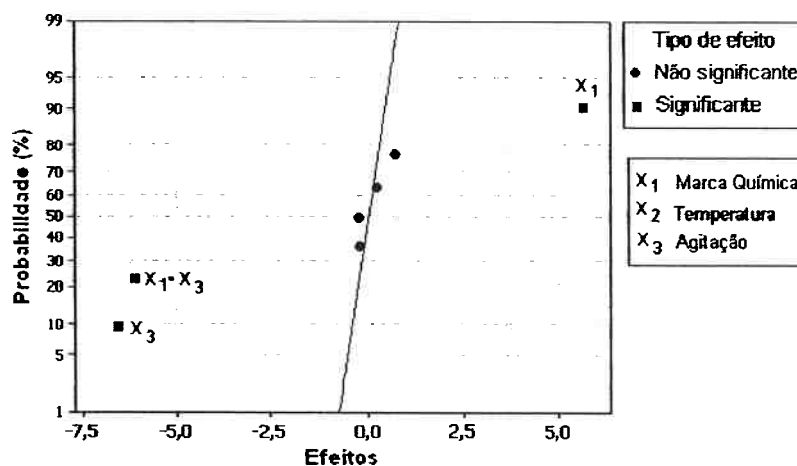


Figura 16. Representação gráfica dos efeitos em escala de probabilidade normal, para o exemplo dos poluentes

Para o exemplo em questão, supondo que os 8 dados ocorreram apenas como resultado da variação aleatória em torno da média e que as mudanças nos níveis dos fatores não tragam efeitos consideráveis na resposta, os sete contrastes entre pares de médias que teriam uma distribuição grosseiramente normal, centrada em zero. Os valores destes contrastes representados graficamente em escala de probabilidade normal teriam a aparência de uma reta, como mostra a Figura 16. Nota-se pela figura que 4 das 7 estimativas se ajustam bem à linha reta, conhecida como linha do erro, porque se espera que essas variações sejam causadas apenas pelo ruído. As outras três estimativas não se ajustam bem à linha de erro, revelando que essas ocorrências não são explicadas por chance. Então, pode-se concluir que os efeitos da marca química, da agitação e da interação entre elas são significativos e estão assinalados em vermelho.

Um comentário adicional a esta interpretação é que a temperatura não interfere no processo de redução do poluente. Por outro lado, a marca química e a agitação apresentam efeitos

principais e de interação significantes, não podendo ser interpretados separadamente. A importância da interação pode ser melhor visualizada com as Figura 17 (a) e Figura 17 (b).

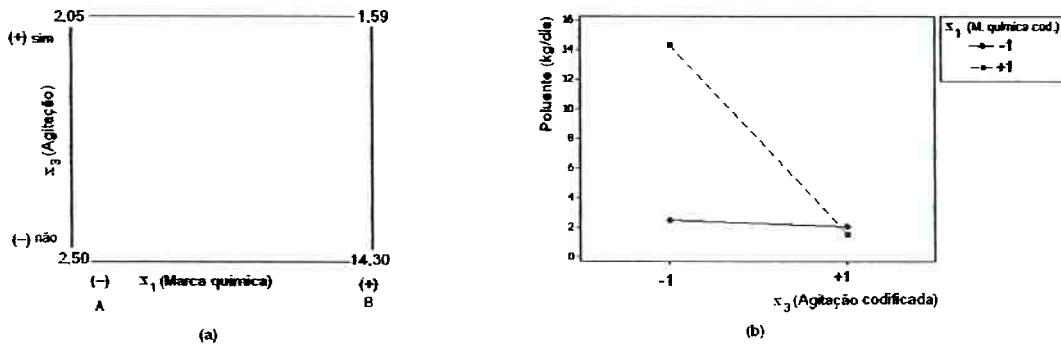


Figura 17 A interação de dois fatores para o exemplo do poluente, vista com dois métodos gráficos diferentes

Em ambas as figuras os valores representados graficamente são os mesmos só que estão em diagramas diferentes, isto é: em (a) a interação está apresentada usando os valores médios do poluente liberado nos vértices do quadrado delimitado pelos pontos fatoriais enquanto que em (b) o poluente liberado está em função da agitação tendo como parâmetro a marca química. Se não houvesse interação, as diferenças entre os dois valores médios dos níveis altos de agitação seriam próximas das diferenças entre os dois valores médios dos níveis baixos, isto é $1,59 - 2,05 \neq 14,30 - 2,50$. Neste caso, a interação é evidente porque há uma diferença sensível da quantidade de poluente liberada na corrente residual para cada nível de agitação. Sem agitação, a quantidade de poluente de marca química A na corrente é muito maior do que a da marca química B. Já, com a corrente agitada, não há diferença importante para as quantidades médias.

Os valores indicados nos vértices do quadrado são as médias dos valores das quantidades de poluente diária para as variações de níveis, indicadas na Figura 17 (a), para os níveis baixo e

alto de temperatura. No caso, o interesse prático é que a marca química B libera uma quantidade menor de poluente quer a corrente esteja ou não agitada.

A Figura 17 (b) diz o mesmo em um outro diagrama. Aqui as quantidades de poluentes que aparecem nos vértices do quadrado da Figura 17 (a) é a ordenada da Figura 17 (b). A abscissa é a agitação, e usa-se como parâmetro a marca química. Os quatro pontos são representados graficamente e, para melhor visualização da interação, traçam-se duas retas unindo os pontos para cada nível da variável usada como parâmetro. Se não houver interação, ou esta for muito fraca, as retas são aproximadamente paralelas. O contrário ocorre quando há intersecção entre as retas. Este é o caso do exemplo dos poluentes: ambas as figuras remetem à mesma conclusão. Cada experimentador usa a figura que tem maior significado para ele.

Convém salientar que, como a temperatura não foi significante neste exemplo, e os experimentos foram realizados de uma forma totalmente aleatória, os 8 experimentos poderiam ser empregados como duas réplicas de um projeto fatorial 2^2 . No caso, caso o t usado como referência tem 4 graus de liberdade. Como ilustração, é apresentada a Tabela 20, com estes resultados e com todos os valores estimados, como se tivesse realizado duas réplicas genuínas.

Tabela 20 - Efeitos para o exemplo dos poluentes ignorando o efeito da temperatura, tratando-se o projeto 2^3 como duas réplicas genuínas 2^2 , levando em conta apenas a marca química e a agitação

Efeito médio	Estimativa \pm o erro padrão
Efeitos principais	
x_1 (m. química)	5,675 \pm 0,197
x_3 (agitação)	-6,583 \pm 0,197
Interação de dois fatores	
$x_1 \times x_3$ (m. química \times Agitação)	-6,129 \pm 0,197

A conclusão é a mesma tirada com o uso da escala de probabilidades.

Finalmente, é importante mencionar que existem outros métodos para a avaliação de efeitos de projetos fatoriais completos que não foram replicados. Entre eles vale a pena mencionar o método de Lenth, proposto em 1989 que Montgomery (2005) o recomenda como um suplemento aos gráficos de probabilidade Normal.

Bisgaard, (1998-1999) propôs uma nova técnica gráfica chamada de carta de inferência condicional, que também auxilia a interpretação dos gráficos de probabilidade normal.

Maiores detalhes sobre estas técnicas podem ser vistas em Box, Hunter, W; Hunter, J. (2005) e Montgomery (2005).

4.1.5 Considerações finais sobre os projetos fatoriais completos em dois níveis

Como o número de ensaios cresce geometricamente em função do número de variáveis, recomenda-se o uso de projetos completos com até 4 variáveis (SCHMIDT E LAUNSBY, 1996). Isto não tira o brilho destes projetos por todas as razões já apresentadas neste texto. Atualmente eles são tão usados na melhoria de processos industriais, quanto em desenvolvimento de produtos, trazendo conclusões importantes para os estudos por eles direcionados. Eles podem ser usados na construção de blocos, quando uma variável que não é de interesse na pesquisa ou na melhoria pode interferir na análise estatística. No capítulo 5 serão apresentados dois casos de melhoria de processos industriais, usando planejamento completo em dois níveis.

Capítulo 5

Dois casos aplicando o DOE 2^k em processos industriais

No capítulo 4 foi mencionada uma série de vantagens do uso do planejamento fatorial em dois níveis (DOE 2^k). Aqui serão apresentadas duas aplicações de interesse prático na indústria. Eles também são largamente empregados no desenvolvimento de produtos e de processos. O primeiro caso é o uso de um projeto fatorial 2⁴ usado na melhoria do processo de produção de filmes de polipropileno biorientado. Este estudo é parte do trabalho desenvolvido pelo aluno Wanderley Altomani Júnior para obtenção do título de Mestre em Engenharia pela Escola Politécnica da USP (Altomi Jr, 1996).

O segundo caso mostra uma aplicação peculiar dos projetos fatoriais em dois níveis, descrita por Box e Draper, (1998) que chega a se constituir como uma filosofia de trabalho, por envolver toda equipe de produção, na condução do programa. Esta aplicação é conhecida como EVOP (*evolutionary operation*) aplicada na melhoria do processo de produção de acrilamida. Será apresentado parte do trabalho desenvolvido por Nam Ting Kwok para obtenção do título de Mestre em Engenharia pela Escola Politécnica da USP.

Para melhor compreensão dessa aplicação específica será feita uma introdução sobre os procedimentos de uso da técnica.

5.1 Melhoria do processo de produção de filmes de polipropileno usando um projeto fatorial 2⁴

5.1.1 O problema

A Koppol Filme S A era uma empresa situada no Pólo petroquímico de Mauá, cidade que pertence à grande São Paulo. O processo de produção de filmes dessa empresa vinha atravessando momentos delicados em decorrência de paradas inesperadas, que provocavam grande reciclo de material processado. Wanderley, engenheiro de processo da empresa, estava interessado em empregar em seu programa de Mestrado uma metodologia estatística que permitisse conhecer melhor o processo e minimizasse as paradas inesperadas. O estudo foi realizado usando o planejamento fatorial em dois níveis com quatro fatores (variáveis independentes) e sete variáveis de resposta (propriedades mecânicas que conferem ao filme a qualidade desejada).

5.1.2 A proposta de solução

Empregar um planejamento fatorial em dois níveis foi uma proposta viável para a solução desse problema, porque ele é capaz de estimar os efeitos principais e as interações, fazendo com que viesse à tona o mistério sobre as paradas inesperadas da planta. Como mencionado anteriormente, a planta onde o trabalho foi realizado produz filmes de polipropileno biaxialmente orientados (BOPP), coextrudados (ou multicamada), pelo processo stenter, descrito em Altomani Jr, (1996). É uma planta de tecnologia Brückner, alemã, com capacidade para produção de 9000 toneladas por ano de filmes.

Para o uso do planejamento sugerido, inicialmente foram levantadas as variáveis chaves do processo que poderiam potencialmente ser causadoras dos problemas operacionais da fábrica, ou seja, as variáveis independentes. Os níveis de variação foram estabelecidos próximos dos valores das condições operacionais usuais, porque os testes foram realizados com o processo

em andamento. A seleção das variáveis e de seus níveis de variação foi baseada na experiência do corpo técnico da empresa e na enorme experiência do fabricante do equipamento, que além de acumular conhecimentos por cerca de 30 anos de trabalho e desenvolvimentos, até o momento do estudo, também é quem desenvolveu a tecnologia deste tipo de processo para produção de BOPP.

As variáveis de resposta escolhidas foram as que mais interferem nas propriedades óticas e mecânicas do filme, pois são estas as principais características que conferem qualidade ao BOPP em relação a outros tipos de filme.

As variáveis de entrada (fatores ou variáveis independentes) escolhidas, inicialmente, foram:

1. grau ou razão de estiro na direção da máquina (DM);
2. grau ou razão de estiro na direção transversal (DT);
3. temperatura de pré-aquecimento do TDO;
4. temperatura de estiro do TDO;
5. temperatura do “*chill-roll*”;
6. temperatura do “*water bath*”.

Experimentos preliminares mostraram que as duas últimas variáveis temperaturas do “*chill roll*” e do “*water bath*” não exerciam influências relevantes nas propriedades óticas e mecânicas do filme dentro das faixas estudadas (faixa limite que o equipamento pode fornecer). Em vista disto, elas não foram incluídas como variáveis no estudo.

Os intervalos de variação para cada variável ficaram de acordo com a Tabela 21.

Tabela 21 - Intervalo de variação para as variáveis de processo, originais e codificadas

	Variável original		Variável codificada	
grau de estiro na DM (*)	5,0	5,3	-	+
grau de estiro na DT (*)	8,47	8,65	-	+
temperatura de pré aquecimento do TDO (°C)	170	174	-	+
temperatura de estiro do TDO (°C)	156	160	-	+

(*) o grau de estiro é uma grandeza adimensional e representa quantas vezes o filme foi estirado em relação as suas dimensões originais.

Com o processo de orientação das macromoléculas do polipropileno as propriedades do filme são alteradas da seguinte forma:

1. as propriedades mecânicas são melhoradas;
2. o alongamento é diminuído;
3. a opacidade é diminuída;
4. o brilho é aumentado;
5. a permeabilidade ao gás e ao vapor é diminuída.

Como as variáveis de processo causam efeitos nas propriedades citadas anteriormente, a qualidade do filme produzido é medida a partir de restrições impostas a estas características.

Neste trabalho, foram estudadas sete respostas para avaliar as propriedades óticas e mecânicas do filme:

1. opacidade (OPAC), y_1 ;
2. tensão de ruptura na direção longitudinal (TRDM), y_2 ;
3. tensão de ruptura na direção transversal (TRDT), y_3 ;
4. alongamento na direção longitudinal (ALDM), y_4 ;

5. alongamento na direção transversal (ALDT), y_5 ;
6. estabilidade dimensional na direção longitudinal (EDDM), y_6 ;
7. estabilidade dimensional na direção transversal (EDDT), y_7 ;

Detalhes sobre as definições, os métodos de medição e respectivas normas técnicas estão em Altomani, Jr. (1996).

5.1.3 A matriz do projeto fatorial 2^4 e os resultados dos ensaios

O projeto fatorial completo em dois níveis, 2^4 , com as variáveis originais, codificadas e os respectivos resultados estão na Tabela 22. O planejamento foi realizado para a produção do filme transparente, envolvendo as 4 variáveis de entrada: x_1 - grau de estiro na direção longitudinal, x_2 - grau de estiro na direção transversal, x_3 - temperatura de pré-aquecimento e x_4 - temperatura de estiro.

Para a execução dos ensaios, que foram realizados com a planta em operação normal, foram seguidos os procedimentos de operação usuais, tomando-se os seguintes cuidados:

1. cada experimento, ou seja, cada ponto de operação, corresponde a uma bobina jumbo. Fazendo uma comparação com a idéia original do EVOP, cada bobina jumbo representa um ponto fatorial;
2. no início de produção de uma bobina jumbo, estabelecer as condições de processo da tabela;
3. apenas considerar como iniciado o experimento quando as condições alteradas já estiverem estabilizadas;
4. não efetuar qualquer alteração adicional de condições durante o experimento;

5. todas as demais condições de processo que não fazem parte do estudo devem se manter inalteradas durante a realização de todo o projeto experimental;
6. a bobina referente a cada experimento deve ser identificada com o número do ensaio;
7. ocorrendo problemas durante um ensaio que deixem dúvidas quanto a sua validade, este deve ser abortado e reiniciado.

Tabela 22 - Matriz de ensaios com os resultados das sete respostas para o estudo do comportamento do processo de fabricação do BOPP

Ensaio.	Variáveis codificadas				Opacidade (%)		Tensão de ruptura (N/(mm) ²)			Alongamento (%)			Estabilidade (%)	
	x ₁	x ₂	x ₃	x ₄	OPAC	(TRDM) y ₂	(TRDT) y ₃	(ALDM) y ₄	(ALDT) y ₅	(EEDM) y ₆	(EEDT) y ₇			
1	-	-	-	-	1,1	148	283	177	49	2,2	1,2			
2	+	-	-	-	1,2	164	282	170	42	2,0	0,8			
3	-	+	-	-	1,3	129	285	151	51	3,0	1,5			
4	+	+	-	-	1,1	146	280	158	58	2,3	1,0			
5	-	-	+	-	1,0	146	274	190	51	2,2	1,0			
6	+	-	+	-	1,2	147	253	165	40	2,7	2,9			
7	-	+	+	-	1,1	145	259	188	39	2,0	0,8			
8	+	+	+	-	1,1	155	275	166	54	2,2	1,0			
9	-	-	-	+	1,3	136	264	180	69	2,0	0,7			
10	+	-	-	+	1,3	143	277	163	55	2,2	1,0			
11	-	+	-	+	1,2	134	270	187	57	3,0	1,5			
12	+	+	-	+	1,2	146	277	167	59	2,5	0,2			
13	-	-	+	+	1,0	135	251	189	53	1,8	0,2			
14	+	-	+	+	1,0	132	246	153	48	2,8	1,7			
15	-	+	+	+	1,1	135	258	196	56	2,0	1,0			
16	+	+	+	+	1,2	152	258	172	51	2,0	0,3			

5.1.4 A estimativa e a interpretação dos efeitos

A estimativa dos efeitos principais e efeitos interativos das variáveis independentes causados nas variáveis de resposta foi feita com o programa Minitab versão 15 seguindo a metodologia descrita no capítulo 4. Na Tabela 23 estão os efeitos para as sete respostas estudadas.

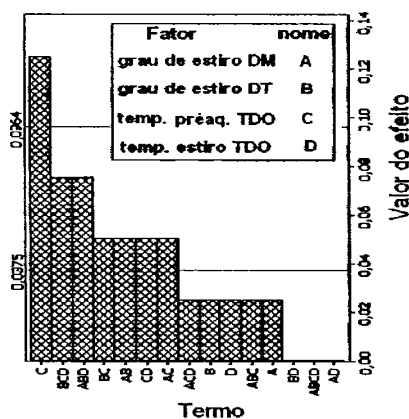
Todas as observações foram consideradas na estimativa dos efeitos, não havendo portanto, graus de liberdade para estimativa do erro.

Tabela 23 - Efeitos estimados para as sete respostas (propriedades mecânicas do BOPP) do projeto fatorial completo 2^4

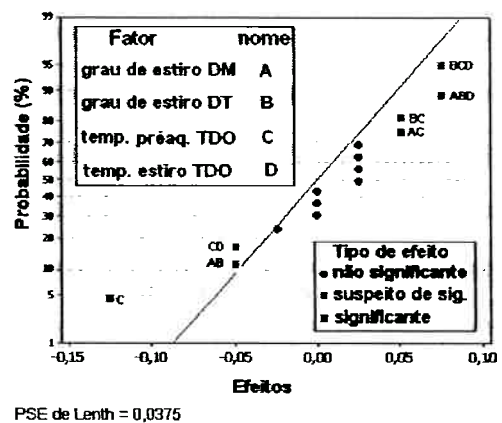
Efeitos estimados ↓	Respostas						
	y_1 OPAC	y_2 TRDM	y_3 TRDT	y_4 ALDM	y_5 ALDT	y_6 EDDM	y_7 EDDT
média	1,15	143,3	268,3	173,3	52,0	2,31	1,05
x_1	0,03	9,6	0,5	-18,0	-2,3	0,06	0,13
x_2	0,03	-1,1	4,0	-0,3	2,3	0,14	-0,28
x_3	-0,13	0,1	-18,0	8,3	-6,0	-0,19	0,13
x_4	0,03	-8,4	-11,3	5,3	8,0	-0,04	-0,45
$x_1 \times x_2$	-0,05	4,4	4,0	3,3	7,0	-0,31	-0,70
$x_1 \times x_3$	0,05	-3,4	-3,0	-8,8	0,8	0,36	0,60
$x_1 \times x_4$	0,00	-1,4	3,3	-6,3	-3,3	0,11	-0,18
$x_2 \times x_3$	0,05	7,9	2,5	6,5	-0,3	-0,46	-0,40
$x_2 \times x_4$	0,00	6,4	2,3	9,5	-2,8	0,04	0,13
$x_3 \times x_4$	-0,05	1,4	-0,8	-5,0	-2,0	-0,09	-0,18
$x_1 \times x_2 \times x_3$	0,03	2,9	6,5	0,5	-0,5	-0,01	-0,28
$x_1 \times x_2 \times x_4$	0,08	1,9	-4,3	-1,0	-3,0	-0,11	-0,25
$x_1 \times x_3 \times x_4$	-0,03	2,1	-3,3	3,0	-0,3	-0,04	-0,15
$x_2 \times x_3 \times x_4$	0,08	-3,1	0,8	-2,8	3,8	-0,01	0,25
$x_1 \times x_2 \times x_3 \times x_4$	0,00	0,9	-3,8	3,3	-3,5	-0,06	0,13

Os níveis das variáveis independentes foram especificados de modo a acompanhar a variações normais do processo porque, como foi mencionado, a fábrica continuou produzindo durante todo o estudo, caso típico do EVOP (*“Evolutionary Operation”*).

Em vista das estreitas faixas de variação das variáveis independentes, seriam necessários muitos ensaios para que o erro padrão fosse minimizado. Devido a dificuldade de realização do trabalho e verificando-se que muitos efeitos eram superiores ao pseudo erro padrão (PSE de Lenth), os efeitos cujos valores estavam entre o PSE e o ME (*Margin Error*), foram considerados suspeitos de significância e então incluídos na análise. O conjunto de gráficos da Figura 18 à Figura 24 ilustram o que está sendo relatado. Todos os efeitos assinalados em verde (suspeitos) e em vermelho altamente significativos estão na Tabela 24 para que se possa fazer uma análise visual de todas as variáveis respostas simultaneamente.

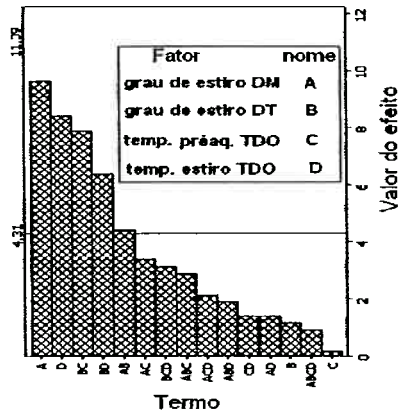


(a)

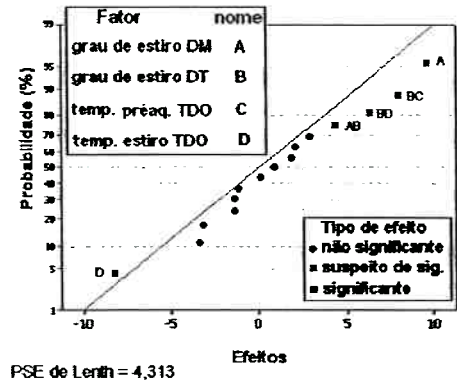


(b)

Figura 18. Efeitos da resposta Opacidade (γ_1) traçados em (a) Pareto, (b) gráfico em escala de probabilidade Normal

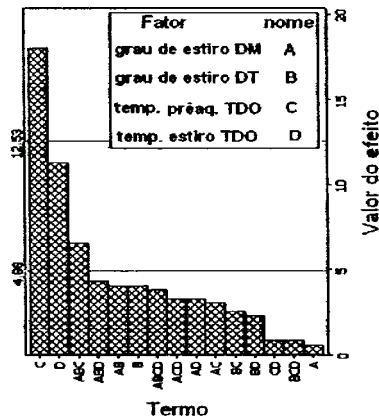


(c)

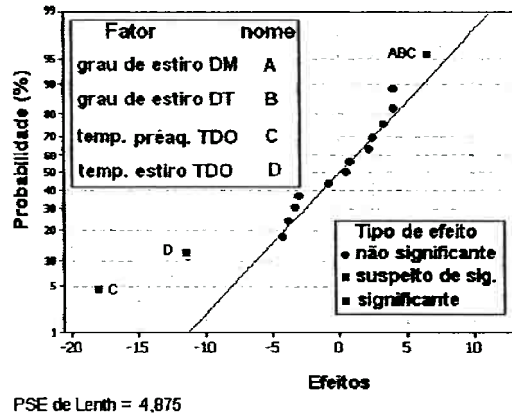


(d)

Figura 19. Efeitos da resposta TRDM (y_2) traçados em (c) Pareto, (d) gráfico em escala de probabilidade Normal

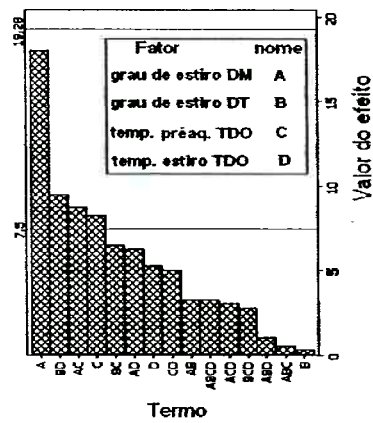


(e)

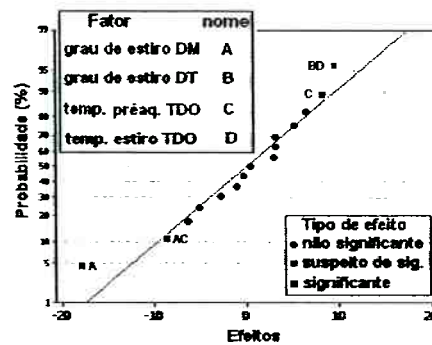


(f)

Figura 20. Efeitos da resposta TRDT (y_3) traçados em (e) Pareto, (f) gráfico de probabilidade Normal

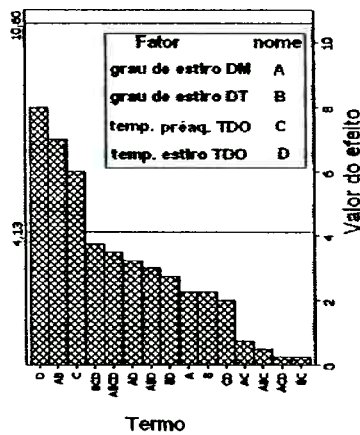


(g)

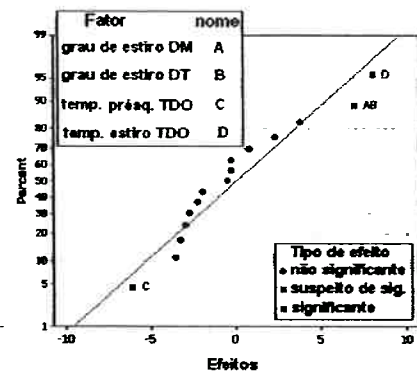


(h)

Figura 21. Efeitos da resposta ALDM (y_4) traçados em (g) Pareto, (h) gráfico de probabilidade Normal

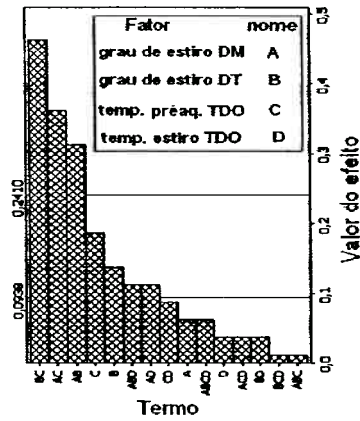


(i)

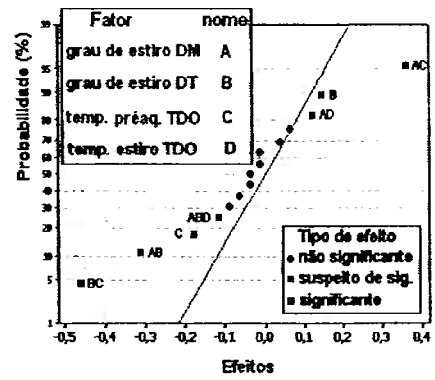


(j)

Figura 22. Efeitos da resposta ALDT (y_5) traçados em (i) Pareto, (j) gráfico em escala de probabilidade Normal

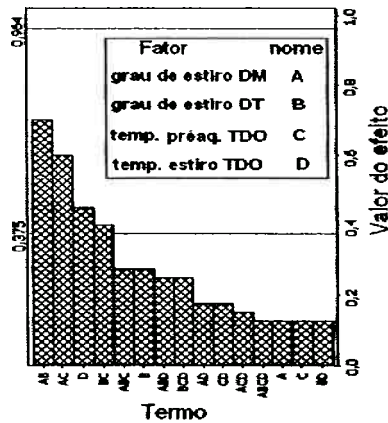


(l)

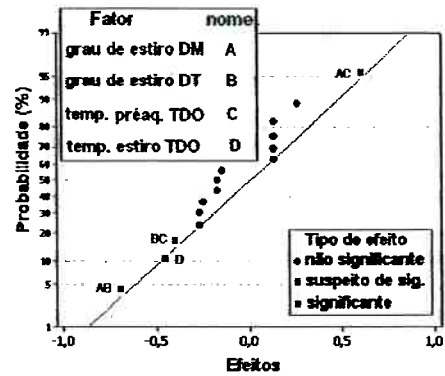


(m)

Figura 23. Efeitos da resposta EDDM (y_6) traçados em (l) Pareto, (m) gráfico em escala de probabilidade Normal



(n)



(o)

Figura 24: Efeitos da resposta EDDT (y_7) traçados em (n) Pareto, (o) gráfico em escala de probabilidade Normal

Tabela 24 - Efeitos considerados na análise do DOE 2⁴, para o processo de obtenção do filme transparente de polipropileno

Efeitos estimados ↓	Respostas						
	y ₁ OPAC	y ₂ TRDM	y ₃ TRDT	y ₄ ALDM	y ₅ ALDT	y ₆ EDDM	y ₇ EDDT
média	1,15	143,3	268,3	173,3	52,0	2,31	1,05
x ₁	0,03	9,6	–	–18,0	–	–	–
x ₂	0,03	–	–	–	–	0,14	–
x ₃	–	–	–18,0	8,3	–6,0	–0,19	–
x ₄	–	–8,4	–11,3	–	8,0	–	–0,45
x ₁ ×x ₂	–0,05	4,4	–	–	7,0	–0,31	–0,70
x ₁ ×x ₃	0,05	–	–	–8,8	–	0,36	0,60
x ₁ ×x ₄	–	–	–	–	–	0,11	–
x ₂ ×x ₃	0,05	7,9	–	–	–	–0,46	–0,40
x ₂ ×x ₄	–	6,4	–	9,5	–	–	–
x ₃ ×x ₄	–0,05	–	–	–	–	–	–
x ₁ ×x ₂ ×x ₃	–	–	6,5	–	–	–	–
x ₁ ×x ₂ ×x ₄	0,08	–	–	–	–	–0,11	–
x ₁ ×x ₃ ×x ₄	–	–	–	–	–	–	–
x ₂ ×x ₃ ×x ₄	0,08	–	–	–	–	–	–
x ₁ ×x ₂ ×x ₃ ×x ₄	–	0,9	–	–	–	–	–

Os projetos fatoriais em dois níveis ainda permitem construção de modelos simplificados indicando as variações da média, com coeficientes iguais a metade do efeito. Estes modelos foram escritos e apresentados na Tabela 25.

Tabela 25 - Modelos simplificados obtidos a partir dos efeitos considerados de maior influência, para as sete respostas estudadas

Resposta estimada ↓	Modelos *
Opacidade	$\hat{y}_1 = 1,150 - 0,025 \cdot x_1 \cdot x_2 + 0,025 \cdot x_1 \cdot x_3 + 0,025 \cdot x_2 \cdot x_3 - 0,025 \cdot x_3 \cdot x_4 + 0,038 \cdot x_1 \cdot x_2 \cdot x_4 + 0,038 \cdot x_2 \cdot x_3 \cdot x_4$
TRDM	$\hat{y}_2 = 143,31 + 4,85 \cdot x_1 - 4,19 \cdot x_4 + 2,19 \cdot x_1 \cdot x_2 + 3,94 \cdot x_2 \cdot x_3 + 3,19 \cdot x_2 \cdot x_4$
TRDT	$\hat{y}_3 = 268,25 - 9,00 \cdot x_3 - 5,63 \cdot x_4 + 3,25 \cdot x_1 \cdot x_2 \cdot x_3$
ALDM**	$\hat{y}_4 = 173,25 - 9,00 \cdot x_1 + 4,13 \cdot x_3 - 4,38 \cdot x_1 \cdot x_3 - 3,13 \cdot x_1 \cdot x_4 + 3,25 \cdot x_2 \cdot x_4 + 4,75 \cdot x_2 \cdot x_4$
ALDT	$\hat{y}_5 = 52,00 - 3,00 \cdot x_3 + 4,00 \cdot x_4 + 3,50 \cdot x_1 \cdot x_2$
EDDM	$\hat{y}_6 = 2,306 + 0,069 \cdot x_2 - 0,094 \cdot x_3 - 0,156 \cdot x_1 \cdot x_2 + 0,181 \cdot x_1 \cdot x_3 + 0,056 \cdot x_1 \cdot x_4 - 0,231 \cdot x_2 \cdot x_3 - 0,056 \cdot x_1 \cdot x_2 \cdot x_4$
EDDT	$\hat{y}_7 = 1,050 - 0,225 \cdot x_4 - 0,350 \cdot x_1 \cdot x_2 + 0,300 \cdot x_1 \cdot x_3 - 0,200 \cdot x_2 \cdot x_3$

*As variáveis x_i , com $i = 1, 2, 3, 4$, são as codificações das variáveis independentes: grau de estiro na DM (x_1), grau de estiro na DT (x_2), temperatura de pré-aquecimento do TDO (x_3) e temperatura de estiro do TDO (x_4).

**O modelo ALDM (Alongamento na direção longitudinal) foi o que apresentou o pseudo erro padrão mais elevado (7,5), por esta razão foram acrescentados mais dois coeficientes ao modelo que estavam na vizinhança dos valores aceitos pelo critério estabelecido.

Para o filme transparente com o qual se fez o estudo, as restrições do produto são: OPAC - opacidade $<2,3\%$; TRDM - tensão de ruptura longitudinal $>125 \text{ N}/(\text{mm})^2$; TRDT - tensão de ruptura transversal $>220 \text{ N}/(\text{mm})^2$; ALDM - alongamento longitudinal $<180\%$; ALDT - alongamento transversal $<70\%$; EDDM - estabilidade dimensional longitudinal $<4\%$ e EDDT - estabilidade dimensional transversal $<2\%$.

Os modelos devem ser avaliados com base nas restrições das propriedades do filme para conferir ao produto a qualidade desejada. Vê-se então que é um pouco difícil a avaliação pela observação direta dos sete modelos propostos. É possível se chegar em várias condições que atendam simultaneamente aos modelos usando uma função objetivo, denominada “*Desirability Function*”, que foi empregada na análise deste trabalho e que não será abordada aqui, mas pode ser vista em Altomani Jr. (1996).

A análise aqui será limitada a verificação das influências das variáveis independentes e seus conflitos com as restrições, além de ser mostrado alguns resultados práticos de verificação dos modelos propostos, lembrando que eles só são válidos para o domínio experimental estudado.

De uma maneira geral, vê-se na Tabela 25 que há uma forte influência de interações em todos os modelos. Esta era a principal causa da dificuldade de ajuste das variáveis de processo (independentes) sem que a qualidade do produto fosse prejudicada. O conhecimento destas interrelações trouxe muitos benefícios ao processo em todos os sentidos, isto é, diminuindo as paradas e, como consequência, o reciclo; aumentando a produtividade; e melhorando a qualidade do produto. O pessoal de operação ficou convencido da complexidade do problema e dos cuidados que devem ser tomados na escolha do conjunto de condições operacionais.

Vê-se também pelas restrições que as tensões de ruptura (longitudinal e transversal) devem ser maximizadas e as demais variáveis devem se ser minimizadas.

Uma análise da influência das variáveis em todos os modelos (sem tecer consideração sobre a individualidade das mesmas porque elas não podem ser analisadas separadamente, devido as interações de dois e três fatores) são concordantes com os estudos apresentados na literatura. Esta discussão será feita a seguir:

Observando o modelo da opacidade, verifica-se que os efeitos mais importantes são interativos, alguns tendendo a aumentar a opacidade e outros tendendo a diminuí-la.

O estiro transversal (x_2) influi apenas junto com outras variáveis, porém de forma mais suave e influenciando tanto de maneira negativa como positiva.

As temperaturas de estiro também apresentam influência predominantemente como interações de três fatores com o grau de estiro. Este fato pode ser explicado: as cadeias moleculares estão mais ou menos emaranhadas em si, de modo desordenado e contorcido, formando uma espécie de novelo que se entrelaça no novelo de moléculas próximas. Se, porém, o produto é submetido a oportunas deformações, isto é, a um estiro, em condições térmicas adequadas, as macromoléculas são, de um certo modo, desembaraçadas, e portanto, estendidas, dispostas paralelamente entre si. Esta orientação provoca uma alteração na cristalinidade do polímero, fazendo com que a opacidade diminua, o que é melhor (MOPLEFAN, 1982).

A tensão de ruptura na DM é influenciada principalmente pelo estiro nesta direção (x_1), positivamente. Apresentam grande influência também a temperatura de estiro x_4 mas de forma negativa.

A orientação de filmes poliolefinicos provoca um rearranjo molecular, de forma que as cadeias do polímero fiquem ordenadas na direção deste estiro. O estiro, e conseqüentemente a orientação, são facilitados pelos aquecimentos aos quais o filme é submetido, até um certo limite a partir do qual começam a prejudicar suas propriedades mecânicas, pois alteram a distribuição de tensões entre as cadeias poliméricas (Bruckner, 1990). As cadeias do polímero

quando sofrem um maior estiro, atingem um grau maior de orientação, suportando uma força maior antes de romper quando solicitado, por outro lado, as variáveis x_3 e x_4 que favorecem o estiramento, quando elevadas, passam a prejudicar, de certa forma, a resistência à tração longitudinal, o que explica o modelo obtido.

A tensão de ruptura na DT é fortemente influenciada, de forma negativa, pelas temperaturas de pré-aquecimento e estiro (x_3 e x_4 respectivamente). O estiro transversal (x_2) contribui de forma predominantemente positiva com esta propriedade, tanto isolado quanto em conjunto, principalmente com o estiro longitudinal (x_1).

O comportamento do modelo para a tensão de ruptura na DT é muito bem explicado pelos comentários feitos anteriormente para o modelo da tensão de ruptura na DM.

Uma das maiores influências sofridas pelo alongamento longitudinal é devido ao estiro longitudinal (x_1), atuando de forma negativa. Na seqüência pode-se perceber a influência das temperaturas de pré-aquecimento e estiro (x_3 e x_4), isoladas e interagindo com os graus de estiro, de forma predominantemente positiva.

Conceitualmente, as propriedades mecânicas, tração e alongamento, têm comportamento inversamente proporcional. Em outras palavras, com o aumento da tensão de ruptura, ocorre diminuição do alongamento do material, sendo este conceito válido para qualquer tipo de material (BRUCKNER, 1989 e 1990).

O modelo do alongamento na DM se comporta de forma contrária ao modelo da tensão de ruptura na DM, o que vai de encontro com a teoria.

As temperaturas de pré-aquecimento e estiro (x_3 e x_4) mostram-se bastante influentes no alongamento longitudinal, de forma negativa e positiva respectivamente. Os estiros (x_1 e x_2) influenciam conjuntamente e de forma ora positiva, ora negativa.

O comportamento do alongamento, conforme comentado anteriormente, deve ser inversamente proporcional ao comportamento da tensão de ruptura. Neste modelo observa-se que a variável x_4 e algumas das interações obedecem esta regra, comportando-se de forma inversa em relação ao modelo da tensão de ruptura na direção transversal.

A estabilidade na DM sofre influência maior das interações dos os estiros (x_1 e x_2) entre si e em conjunto com a temperatura de pré-aquecimento (x_3).

As principais variáveis de processo que influenciam diretamente o mecanismo de encolhimento residual de polímeros, orientados por aquecimento subsequente, são: razão de estiro; temperatura de estiro; e velocidade de estiro (Leroux et al., 1992). Neste trabalho, a velocidade de estiro foi mantida constante.

Quando o filme é estirado em temperaturas baixas, cria-se um estado de tensão residual nas cadeias moleculares, o que provoca um alto encolhimento residual no filme (BRUCKNER, 1990). Por este motivo a influência das temperaturas é predominantemente negativa, ou seja, quanto maior é a temperatura, menor é o encolhimento.

Quanto maior a razão de estiro, maior é a deformação sofrida pelas cadeias poliméricas, criando-se uma condição irreversível de redução do estiro, por se ultrapassar o limite de deformação elástica do material (LEROUX et al., 1992; BRUCKNER, 1990). Isto pode ser observado no modelo obtido, pela influência negativa dos estiros no encolhimento do filme.

A estabilidade na DT é influenciada principalmente pelo efeito de interação entre estiros (x_1 e x_2), e estiro com as temperaturas de estiro e pré-aquecimento (x_3 e x_4).

Este modelo tem comportamento semelhante ao da estabilidade na DM, que foi comentado anteriormente e seu comportamento se mostra bastante alinhado com as referências encontradas para esta propriedade.

Com os dados de processo e com os valores preditos com os modelos foi construída a Tabela 26.

Para validar os modelos obtidos foram reunidos dados sobre as propriedades medidas em laboratório de várias bobinas, bem como dados sobre as condições de processo utilizadas para sua produção. Estes resultados estão apresentados na Tabela 27.

Tabela 26 - Uma comparação entre valores preditos pelos modelos e valores observados para o projeto fatorial 2^4 , aplicado na fabricação de filmes de polipropileno transparente

x_1	x_2	x_3	x_4	Opacidade		TRDM		TRDT		ALDM		ALDT		EDDM		EDDT	
				predita	Observ.	predita	Observ.	predita	Observ.	predita	Observ.	predita	Observ.	predita	Observ.	predita	Observ.
-1	-1	-1	-1	1,0	1,1	152	148	285	283	189	177	55	49	2,2	2,2	1,0	1,2
1	-1	-1	-1	1,2	1,2	157	164	287	282	159	170	48	42	2,0	2,0	1,1	0,8
-1	1	-1	-1	1,2	1,3	133	129	287	285	159	151	48	51	3,1	3,0	2,1	1,5
1	1	-1	-1	1,1	1,1	147	146	285	280	170	158	55	58	2,3	2,3	0,8	1,0
-1	-1	1	-1	1,1	1,0	144	146	269	274	185	190	49	51	2,1	2,2	0,8	1,0
1	-1	1	-1	1,2	1,2	149	147	267	253	179	165	42	40	2,7	2,7	2,1	2,9
-1	1	1	-1	1,2	1,1	141	145	267	259	196	188	42	39	2,1	2,0	1,1	0,8
1	1	1	-1	1,0	1,1	155	155	269	275	149	166	49	54	2,0	2,2	1,0	1,0
-1	-1	-1	1	1,2	1,3	137	136	274	264	179	180	63	69	2,1	2,0	0,6	0,7
1	-1	-1	1	1,2	1,3	143	143	276	277	150	163	56	55	2,1	2,2	0,7	1,0
-1	1	-1	1	1,0	1,2	131	134	276	270	168	187	56	57	3,0	3,0	1,7	1,5
1	1	-1	1	1,1	1,2	145	146	274	277	179	167	63	59	2,4	2,5	0,4	0,2
-1	-1	1	1	1,1	1,0	129	135	258	251	176	189	57	53	2,0	1,8	0,4	0,2
1	-1	1	1	1,0	1,0	135	132	256	246	169	153	50	48	2,8	2,8	1,7	1,7
-1	1	1	1	1,2	1,1	139	135	256	258	206	196	50	56	2,0	2,0	0,7	1,0
1	1	1	1	1,2	1,2	153	152	258	258	159	172	57	51	2,1	2,0	0,6	0,3

Tabela 27 - Um teste de validação dos modelos para algumas bobinas produzidas após a melhoria aplicada na fabricação de filmes de polipropileno transparente

x_1	x_2	x_3	x_4	Opacidade		TRDM		TRDT		ALDM		ALDT		EDDM		EDDT	
				predita	Observ.	predita	Observ.	predita	Observ.	predita	Observ.	predita	Observ.	predita	Observ.	predita	Observ.
-0,52	0,56	1	-1	1,17	1,1	145	131	268	268	185	165	44	52	2,1	2,4	1,1	1,3
-0,52	0,56	1	0	1,18	1,1	142	144	262	249	187	164	48	39	2,1	2,3	0,9	1,2
-0,52	0,56	1	0	1,18	1,1	142	149	262	263	187	173	48	56	2,1	2,5	0,9	1,2
-0,52	0,56	1	0	1,18	1,1	142	143	262	258	187	169	48	46	2,1	2,5	0,9	1,2

5.1.5 Questões adicionais

O caso aqui relatado mostra a potencialidade que os planejamentos fatoriais em dois níveis oferece e ainda abre caminhos para outros estudos dentro do próprio processo. Empregando-se a função D (“*desirability function*”) aos modelos, foi possível chegar a uma condição ótima de operação, que resultou em uma acréscimo de 33% na produtividade da fábrica. Porém, muitas questões ainda podem ser respondidas a partir deste início da compreensão das interrelações entre as variáveis independentes. Com projetos adicionais, poder-se-ia incluir outras variáveis de interesse e até mesmo desenvolver outros produtos de interesse da empresa. Não se deve esquecer que o conhecimento adquirido com o uso dos métodos aqui aplicados irá servir de apoio em projetos futuros.

5.2 Aplicando EVOP (*Evolutionary Operation*) na melhoria do processo de fabricação de acrilamida

Este caso é parte do trabalho desenvolvido pelo aluno Nam Ting Kwok para obtenção do título de Mestre em Engenharia pela Escola Politécnica da USP.

5.2.1 O problema

A processo de fabricação de acrilamida de propriedade da empresa Nalco do Brasil LTDA opera em em batelada e historicamente apresentava problemas de qualidade do produto. A variabilidade excessiva no processo era uma preocupação da gerência e do pessoal de operação da fábrica. Eles necessitavam de um método estatístico capaz de diagnosticar a

causa dessa variabilidade causada provavelmente por certas no controle do mesmo, e que não requeresse nem investimento nem queda na produção.

5.2.2 A proposta de solução

A idéia de usar o programa EVOP na melhoria do processo de fabricação da acrilamida foi acolhida pela empresa, levando em conta duas das grandes vantagens oferecidas pelo uso do programa:

1. nenhum investimento financeiro (a melhoria foi obtida com os recursos existentes na própria fábrica); e,
2. a continuidade da produção sem prejuízo à empresa e aos seu clientes.

Além dessas duas vantagens, um ponto a ser destacado é o ganho de conhecimento sobre o processo com a implementação do programa. Então ficou decidido que a aplicação seria realizada como tema da dissertação do engenheiro Nam no LSTM do DEQ-EPUSP.

Como a aplicação do programa requer o conhecimento de algumas particularidades do EVOP, para uma melhor compreensão e acompanhamento do caso, será feita uma breve apresentação do que é efetivamente este procedimento de trabalho e de como é realizada a sua implementação.

5.2.3 Uma breve apresentação do programa EVOP

Embora seja uma aplicação do DOE 2^k, o programa EVOP tem certas particularidades, tais como:

1. classifica as variáveis independentes do processo por sua influência, porque seus efeitos podem ser quantificados numericamente, fazendo com que a seleção das mesmas seja feita de forma objetiva;
2. as variações a serem implementadas com as variáveis selecionadas podem ser feitas através de uma programação, utilizando a experiência e a vivência de operadores e de engenheiros da planta;
3. uma vez conhecida a magnitude dos efeitos, estes podem ser interpretados de forma a influenciar as decisões tomadas para melhoria do processo;
4. quando o ponto ótimo for atingido, o processo já estará adaptado a condições já incorporadas;
5. pequeno custo comparado aos métodos de simulação;
6. participação ativa de toda a equipe que atua no processo;
7. as mudanças são feitas de forma a não permitir perda de produtos, evitando assim a geração de sucatas indesejáveis na indústria química;
8. a simplicidade do método faz com que os resultados sejam altamente confiáveis e facilitem o entendimento de como estão sendo conduzidas as melhorias.

O programa EVOP tem por objetivo a obtenção das condições ótimas de operação com menor custo e atendimento aos requisitos de qualidade, sem alterar as condições de segurança de um processo industrial, com mínimo impacto nas condições processuais de cada passo do processo.

Segundo Box e Draper, (1998), o EVOP é uma ferramenta estatística para otimização de processos que cria um procedimento visando aumentar a produtividade. O EVOP é projetado para ser usado pelos operadores em condição de escala industrial. Portanto, em princípio deve

produzir produtos dentro da especificação. Esta metodologia está em contraste com técnicas experimentais típicas de desenvolvimento, que certamente, sob determinadas condições chegam a produtos fora da especificação desejada. Desta forma, a técnica EVOP é diferente dos métodos em escala de laboratório ou de planta piloto, onde é necessário um investimento adicional em dinheiro e tempo, bem como envolvimento de profissionais especializados em P&D.

Ao se fazer um delineamento experimental, procura-se utilizar o mais alto índice possível "variáveis/ensaios". Isto permite ao experimentador obter um número máximo de informações com mínimo de ensaios possível. Por exemplo, se são realizadas 16 corridas em um dado estágio de uma investigação, é usualmente melhor estudar 4 fatores (variáveis) em um projeto 2^4 fatorial, do que estudar 3 fatores em um projeto 2^3 fatorial repetido.

Ao contrário, considere o uso de um programa EVOP. Neste, poucas variáveis são estudadas ao mesmo tempo. Usualmente duas ou três variáveis, por vez, devido ao fato de a taxa sinal/ruído ser pequena. Outro ponto chave do EVOP é que sejam impostos pequenos intervalos de variação nas variáveis independentes, para que não haja grandes flutuações nas respostas, isto é, a qualidade do produto seja mantida. Assim, um grande número de corridas em regime de produção revelam os efeitos das variáveis estudadas, resultando em pequeno custo adicional.

Em ambiente produtivo recomenda-se que as mudanças de processo sejam relativamente simples, portanto, é prático variar somente dois ou três fatores em qualquer fase de um processo experimental. Em um programa EVOP, as estimativas das médias e dos efeitos são continuamente atualizadas. Esta informação é utilizada como guia para melhorar as condições de processo na fase seguinte.

Nos experimentos em escala de laboratório ou em plantas piloto obtém-se informações que muitas vezes não são reproduzidas na planta de produção, pois a mudança altera os efeitos das variáveis envolvidas no processo, causando às vezes grande diferenças econômicas. A Figura 25 ilustra este fato mostrando os resultados da mesma reação química em escala reduzida e em escala industrial. É visto que, devido a efeitos de variáveis não considerados na escala reduzida, o ponto ótimo do processo desta escala não é o mesmo do processo em escala industrial, apesar de as curvas de nível serem similares.

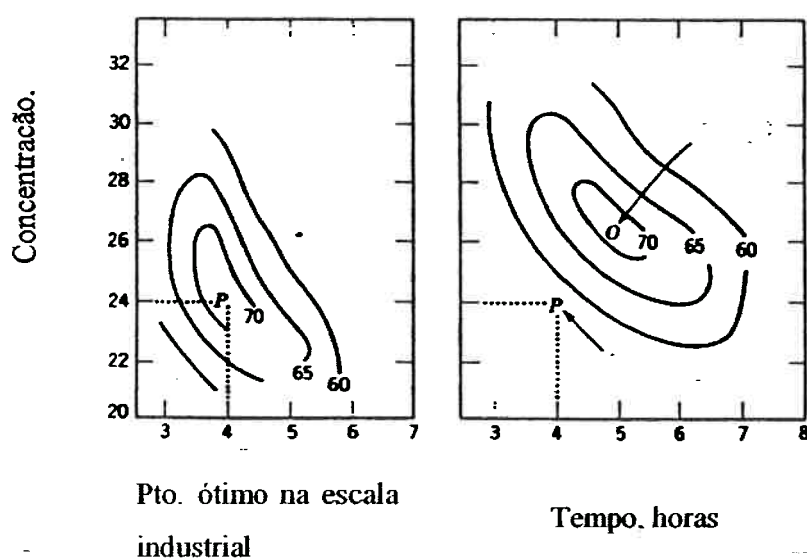


Figura 25. Linhas de nível das superfícies de rendimento, a) escala reduzida, e b) escala de produção. Fonte: (BOX e DRAPER, 1969).

Para que o ponto ótimo (P) da escala reduzida chegue ao ponto (O) da escala industrial serão necessários vários ajustes das variáveis de processo na planta industrial, pois nesta situação os efeitos das variáveis são diferentes.

Para iniciar o EVOP e mesmo durante a condução do mesmo assumem-se algumas variações baseadas na experiência, e muitas vezes, também é observada a influência de outras variáveis e até modificações não imaginadas no início. O uso do EVOP, freqüentemente, traz melhoria aos processos, com acréscimo de rendimento (redução de custo) e obtenção de produtos de

melhor qualidade, e mesmo no caso do processo já otimizado, poderá, no mínimo haver um melhor entendimento do processo.

Uma pesquisa feita por HAHN, 1976, mostrou que a maior relutância em se usar o EVOP foi justamente perturbar o processo. Por outro lado, sabe-se que poucas unidades industriais estão em condições otimizadas. Portanto, a necessidade de esclarecimento e treinamento deve ser aplicada antes de iniciar os testes do programa, a fim de se obter harmonia com todos os departamentos da planta de processos, fazendo com que os esforços sejam direcionados para otimização. O mesmo autor, HAHN, 1976, cita em seu artigo, que o mais crucial para o sucesso de um programa EVOP é o comprometimento ativo, a participação e a liderança com responsabilidade direta do dia a dia do processo de manufatura.

5.2.3.1 Uma comparação do EVOP com operação estática

A rotina de produção é geralmente conduzida com processos rígidos, com condições preestabelecidas e utilizada por longo período. Nesta rotina o operador sempre reproduz exatamente as mesmas condições, isto é chamada de OPERAÇÃO ESTÁTICA. Este método, embora seja o mais usual, impede a evolução do processo, que deveria ser considerado como um dos objetivos de uma planta industrial. Para que um programa do tipo EVOP não traga problemas, tal como, obtenção de produtos fora de especificação, as variações no processo devem ser estudadas previamente de forma a assegurar um mínimo risco de perdas de produto. A rotina é então, selecionar as variáveis a serem alteradas e fazer corridas repetidas vezes. A repetição persistente permite evidenciar os efeitos das variáveis no rendimento, na qualidade do produto e nas propriedades físicas em torno das condições estáticas do processo. Desta forma são obtidos não só o produto mas também as informações para melhoria do processo de manufatura (BOX; DRAPER, 1998).

Devido ao dinamismo imposto pelo método, as informações devem ser postas convenientemente resumidas num painel e devem ser automaticamente alteradas após o recebimento de novos resultados operacionais. Portanto, o EVOP é um método de otimização que demanda certa flexibilidade na variação dos fatores que influenciam o processo.

5.2.3.2 O planejamento EVOP

O EVOP utiliza séries de ensaios planejados segundo um projeto fatorial em dois níveis, cada série recebe o nome de réplica (*“replication”*) e pode ser repetida várias vezes. Os planejamentos mais comuns são com duas ou três variáveis, isto é, 2^2 e 2^3 , respectivamente. Os projetos fatoriais para o EVOP, em particular, são acompanhados de um ponto central que é comparado com a média dos resultados das condições dos pontos fatoriais. No programa EVOP um ciclo é definido como um conjunto completo de condições experimentais que devem ser ensaiadas (uma serie experimental planejada). Um conjunto de ciclos repetidos é definido como uma fase. A mudança de fase ocorre quando chega-se a alguma conclusão quanto a significância dos efeitos e/ou das interações. Algumas considerações devem ser observadas.

Considerações sobre a medição de variabilidade:

As pequenas variações no processo, que geralmente não são detectáveis em corridas individuais, são utilizadas como informações para melhoria. Uma planta química tem como característica as variações de processo consideradas como ruído, isto é, variações pequenas devido a variações de matéria primas ou desvios de instrumentação de controle. Este ruído aumenta com aumento de variabilidade dos fatores que compõe o processo. A variação final é a soma de todas as variações individuais. A magnitude da variação é medida por uma quantidade chamada de desvio padrão (σ), isto é, raiz quadrada da média dos quadrados dos desvios (BOX; DRAPER, 1998).

Numa planta, o tamanho do desvio padrão depende do grau de controle do processo e varia a medida que a planta se desgasta ao longo do tempo. Estes desvios são chamados de “ruídos”. Se a mudança de nível de uma variável produz um efeito na resposta do sistema que exceda ao nível de ruído, este pode ser explorado e detectado com facilidade. Caso contrário ficará encoberto pelo ruído.

Para descobrir os efeitos que estão encobertos pelos ruídos, deve-se aumentar o sinal do efeito ou reduzir os ruídos. No programa EVOP são aplicados os dois casos. O sinal do efeito é acrescido deliberadamente se são introduzidas mudanças nas variáveis cuidadosamente escolhidas. O efeito do ruído é reduzido por repetição dos ciclos fazendo-se a média dos resultados.

Sobre o esquema dos experimentos:

A fim de descrever os esquemas mais usuais para EVOP, serão apresentados esquemas com duas e com três variáveis. Contudo, é possível trabalhar com quaisquer número de variáveis. Nesse caso, o método perde o fator simplicidade, considerado como um dos pontos positivos (BOX e DRAPER, 1998).

5.2.3.3 O esquema EVOP para 2 variáveis

Como exemplo, será utilizado um estudo feito por Jenkins (1969) apresentado por Box, Hunter, W e Hunter, J. (1978), no qual o objetivo era diminuir o custo por tonelada de um certo produto numa planta petroquímica. Em um determinado estágio da investigação duas variáveis julgadas importantes foram:

1. a razão de refluxo da coluna de destilação;
2. a razão de vazão do reciclo pela vazão da purga.

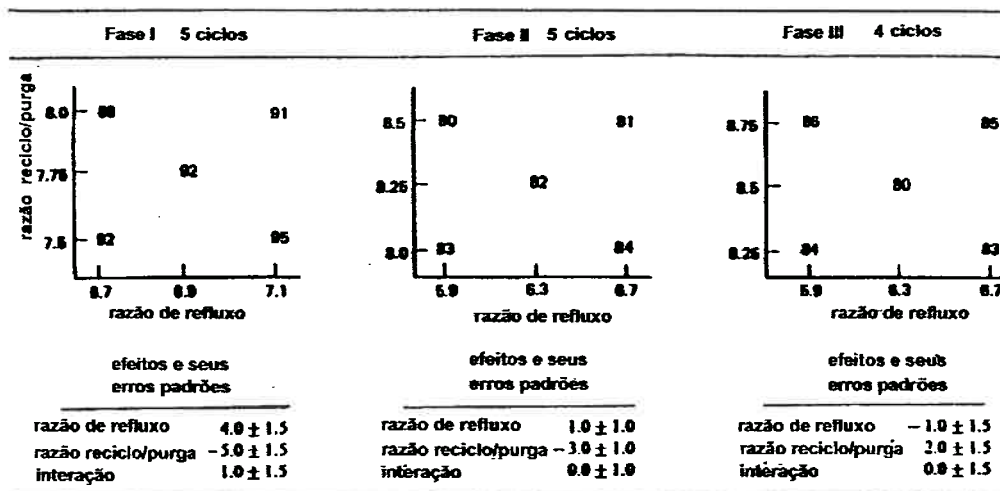


Figura 26. Informações obtidas de uma planta petroquímica de um programa EVOP, com duas variáveis ao final das fases I, II e III com objetivo de redução de custo por tonelada (fonte: BOX; HUNTER, W, HUNTER, J. 1978)

As três fases são mostradas na Figura 26. A medida do desempenho foi o custo médio por tonelada. O modelo empregado foi o de 2^2 fatorial com um ponto central. O resultado ao final da fase I, mostrado na Figura 26, são médias obtidas após 5 repetições (ciclos). Atingido este ponto foi decidido que havia evidência suficiente para justificar a alteração da razão de refluxo para menor e razão reciclo/purga para maior, iniciando, assim, a fase II. Os resultados da fase II, que terminou após 5 ciclos, confirmou que os custos são reduzidos quando as condições são alteradas neste sentido, e sugere que ainda é possível tentar aumentar a razão reciclo/purga, dando indicação da direção da fase III. Esta fase, finalizou após 4 ciclos, chegando a conclusão que o menor custo foi obtido com razão de refluxo de 6,3 e a razão reciclo/purga de 8,5, o custo por tonelada caiu em £12. O programa teve a duração de 4,5 meses a um custo de £6.000. A economia anual, com a redução do custo por tonelada de £92 para £80, foi de £100,000.

A medida do erro

A medida do erro ou do intervalo de confiança da média das corridas individuais é dado por 2 erros padrões, que abrange 95 % da área de confiança. Por exemplo, ao final da fase I (após 5 ciclos) a medida do erro ou limites de erro segundo Juran (1993) para os efeitos é de $\pm 1,5$. A cada ciclo completado, os desvios vão diminuindo cada vez mais, desde que o nível de ruído permaneça na mesma condição.

A análise das informações

A fim de detalhar a análise comumente feita para programas EVOP, tomou-se como ilustração a Figura 26 do exemplo. Ao final da fase I, e com os dados gerais de processo podem-se tomar as seguintes decisões:

1. continuar o teste mantendo-se os intervalos dos fatores e repetindo o ciclo com o objetivo de reduzir o erro padrão (para se verificar o comportamento dos efeitos);
2. modificar a operação na direção do ponto ótimo, iniciando nova fase.

Para iniciar nova fase existem algumas possibilidades:

1. adotar um dos pontos do ciclo da fase em questão como novo ponto central, e então eleger novos intervalos das variáveis para iniciar a nova fase sob estas condições;
2. explorar a indicação do sentido favorável e começar a nova fase sob estas condições. No exemplo da Figura 26, o ponto do vértice esquerdo superior da fase 1 foi eleito para o vértice direito inferior da fase 2, indicando o sentido favorável para melhoria do processo. Neste exemplo verifica-se uma tendência de aumento da razão reciclo/purga e diminuição da razão de refluxo;
3. substituir novas variáveis por uma ou mais variáveis antigas, lembrando que sempre há variáveis aguardando para serem testadas no esquema EVOP;

4. modificar o padrão de variação de forma que os efeitos sejam mais destacados (esta opção é feita quando se verifica que não há efeito apreciável com os limites de variação atuais. Geralmente, o acréscimo do intervalo de variação é necessário).

Vale ressaltar que, no EVOP, o julgamento do pessoal que opera, e que está envolvido, é muito importante, por isso deve ser tomado como um fator importante de mudança para novas fases. Embora se trate de um método científico, o fator subjetivo é utilizado por haver um número muito grande de variáveis num processo industrial, onde se quer identificar pequenos detalhes que se diferenciam das variações do tipo ruído, e que muitas vezes não é possível detectar por métodos mais simples.

5.2.4 O projeto experimental para o caso da acrilamida

5.2.4.1 A escolha das variáveis

O passo inicial da montagem do projeto experimental é a escolha das variáveis sob estudo. No caso da fabricação da acrilamida a partir da acrilonitrila descrito em Kwok (2000), diversas variáveis poderiam ser consideradas no programa de melhoria, a saber: temperatura de reação ou pressão de reação, concentração de reagentes, quantidade de catalisador, método de preparação do catalisador, granulometria do catalisador, pressão de destilação, agitação do reator, pH da mistura, e tipos diferentes de inibidores de polimerização.

Como o esquema proposto para essa melhoria foi o esquema empregando duas variáveis, a tarefa foi seleccionar duas variáveis do elenco mencionado, de modo que, essa implementação, implicasse em menor custo possível. Nesse contexto o número de variáveis independentes caiu drasticamente.

Para o estudo, a variável indicadora do desempenho do processo foi o tempo de reação, por ser de fácil medição. Basicamente, o investimento foi o tempo de estudo dos engenheiros e a atenção do pessoal de operação.

No contexto das restrições impostas ao investimento da implementação as duas variáveis escolhidas foram a pressão de reação e a concentração de reagentes estudada como a razão entre as quantidades de água e acrilonitrila, ou seja, água/acrilonitrila.

As demais variáveis foram mantidas nas condições usuais de processo.

5.2.4.2 O planejamento experimental

Foram consideradas como ponto central as condições de processos que estavam em rotina. Os intervalos das variáveis em torno do ponto central foram escolhidos conforme descrição a seguir:

1. em relação à pressão, o intervalo inicial foi de 3,75 a 3,85 kgf/cm². O controle de pressão foi feito por intervalos de trabalho levando-se em conta a precisão do equipamento de medição e a resposta dos controles acionados. Considerou-se a média do intervalo para efeito de cálculo e interpretação;
2. para a concentração de reagentes foi adotado um intervalo de variação de $\pm 2\%$ de acrilonitrila, considerado suficientemente pequeno para não trazer alteração à qualidade do produto.

A Figura 27 representa o esquema EVOP para esta primeira fase do estudo.

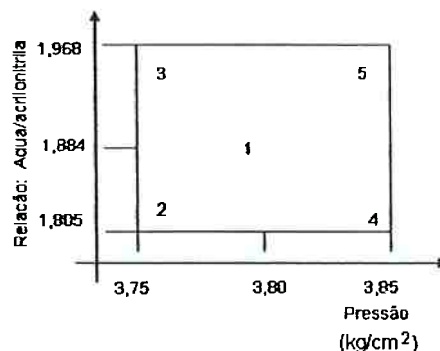


Figura 27. Esquema EVOP para fase 1 da otimização da produção de acrilamida a partir de acrilonitrila

Na Tabela 28 encontra-se o plano dos ensaios para cada ciclo, indicando as quantidades de reagentes e faixa de pressão de trabalho para a primeira fase. A diferença da quantidade de acrilonitrila entre as bateladas dos vértices é de 120L e de cada vértice para a batelada central 60 L.

Tomando como exemplo a batelada nº 1 da Tabela 28 tem-se 2115 L de acrilonitrila acrescentando 60 L de acrilonitrila temos 2175 L e tem-se as batelada nº 2 e 4, reduzindo-se 60 litros em relação a batelada nº1 tem-se 2055L e tem-se as bateladas nº 3 e 5. Para facilitar os trabalhos de compilação foi utilizado o índice água/acrilonitrila em volume, portanto na batelada nº1, tem-se $2115/2175 = 1,884$. Quanto à pressão durante a reação, para a batelada nº1, o intervalo de 3,6 a 3,9 kgf/cm^2 . O intervalo é explicado pela necessidade prática de controle durante a reação, onde o operador deve atuar na válvula de água de resfriamento, abrindo-a ou fechando-a, e o intervalo ocorrido na prática é de 0,3 kgf/cm^2 . A batelada nº2 deve ocorrer no intervalo de pressão de 3,5 a 3,8 kgf/cm^2 . O planejamento para as bateladas nº 3, 4 e 5 seguem a mesma regra. O planejamento é feito de forma que a pressão cubra o intervalo de 3,5 a 4,0 kgf/cm^2 , em intervalos de controle de 0,3 kgf/cm^2 , por limitação de

equipamento, tanto pela precisão de controle como pela pressão máxima permitida. A pressão máxima permitida do reator é de 4,0 kgf/cm².

Um ciclo corresponde a 5 bateladas, isto é, da batelada 1 a 5 corresponde ao ciclo 1 e de 6 a 10 corresponde ao ciclo 2 e assim por diante.

O controle da reação é feito manualmente com o operador acompanhando a pressão durante a reação, sendo que, conforme a necessidade, é injetado vapor na serpentina para aquecimento ou água de resfriamento para reduzir a temperatura e conseqüentemente a pressão. O carregamento é feito via um totalizador com preparador de batelada com bloqueio automático de quantidade, isto é, é feita a programação no equipamento e em seguida é feito o carregamento. Uma vez atingido o volume pré-determinado o totalizador bloqueia a passagem do reagente.

A folha de acompanhamento da reação é mostrada na Figura 28 e desta folha são extraídos os tempos para completar a folha de marcha do EVOP.

Tabela 28 - Condições operacionais para fase I do estudo EVOP, para otimização da reação de obtenção de acrilamida

batelada no.	total Acrilonitrila (litros)	total água (litros)	água/acn em volume	pressão (kgf/cm ²)	tempo de reação (min)	temperatura média (°C)	tempo de distribuição (min)	OBSERVAÇÕES
1	2115	3985	1,884	3,6 - 3,9				
2	2175	3925	1,805	3,5 - 3,8				
3	2055	4045	1,968	3,5 - 3,8				
4	2175	3925	1,805	3,7 - 4,0				
5	2055	4045	1,968	3,7 - 4,0				
6	2115	3985	1,884	3,6 - 3,9				
7	2175	3925	1,805	3,5 - 3,8				
8	2055	4045	1,968	3,5 - 3,8				
9	2175	3925	1,805	3,7 - 4,0				
10	2055	4045	1,968	3,7 - 4,0				
11	2115	3985	1,884	3,6 - 3,9				
12	2175	3925	1,805	3,5 - 3,8				
13	2055	4045	1,968	3,5 - 3,8				
14	2175	3925	1,805	3,7 - 4,0				
15	2055	4045	1,968	3,7 - 4,0				
16	2115	3985	1,884	3,6 - 3,9				
17	2175	3925	1,805	3,5 - 3,8				
18	2055	4045	1,968	3,5 - 3,8				
19	2175	3925	1,805	3,7 - 4,0				
20	2055	4045	1,968	3,7 - 4,0				
21	2115	3985	1,884	3,6 - 3,9				
22	2175	3925	1,805	3,5 - 3,8				
23	2055	4045	1,968	3,5 - 3,8				
24	2175	3925	1,805	3,7 - 4,0				
25	2055	4045	1,968	3,7 - 4,0				
26	2115	3985	1,884	3,6 - 3,9				
27	2175	3925	1,805	3,5 - 3,8				
28	2055	4045	1,968	3,5 - 3,8				
29	2175	3925	1,805	3,7 - 4,0				
30	2055	4045	1,968	3,7 - 4,0				
31	2115	3985	1,884	3,6 - 3,9				
32	2175	3925	1,805	3,5 - 3,8				
33	2055	4045	1,968	3,5 - 3,8				
34	2175	3925	1,805	3,7 - 4,0				
35	2055	4045	1,968	3,7 - 4,0				
36	2115	3985	1,884	3,6 - 3,9				
37	2175	3925	1,805	3,5 - 3,8				
38	2055	4045	1,968	3,5 - 3,8				
39	2175	3925	1,805	3,7 - 4,0				
40	2055	4045	1,968	3,7 - 4,0				
41	2115	3985	1,884	3,6 - 3,9				
42	2175	3925	1,805	3,5 - 3,8				
43	2055	4045	1,968	3,5 - 3,8				
44	2175	3925	1,805	3,7 - 4,0				
45	2055	4045	1,968	3,7 - 4,0				

PRODUÇÃO DE ACRILAMIDA						
PRODUTO		LOTE DO CATAL.		LOT. No.		
LOTE		BATCH No.				
DATA		OPERADOR				
1 - CARREGAMENTO						
Início:		h	Término:		h	Duração:
Não abra o respiro do reator enquanto o mesmo estiver sob vácuo. O Catalisador não deve entrar em contato com Ar !						
Matéria-Prima	AGUA	ACN	ACN Destil.	AGUA DEST	SOL INIB	Total
Quantidade						
2 - AQUECIMENTO						
Início:		h	Término:		h	
Duração:		h	Temperat. Inicial		oC	
3 - REAÇÃO (Considerar o início da reação quando a pressão atingir 3,2 Kg/cm ²)						
Início:		h	Término:		h	Duração:
Temperat. Inicial		oC	I.R.			
4 - DESTILAÇÃO						
Início:		h	I.R. Obtido:		Água Destilada:	L
					ACN Destilada:	L
Término:		h	Duração:		Destilado Total:	L
Passe R-97D na Tubulação e no Condensador para evitar polimerização						
REAÇÃO			DESTILAÇÃO			
Tempo (mm.)	Hora Real	Temperatura	Tempo (mm.)	Hora Real	Temperatura	
0			0			
10			10			
20			20			
30			30			
40			40			
50			50			
60			60			
70			70			
80			80			
90			90			
100			100			
110			110			
120			120			
130			130			
140			140			
150			150			
160			160			
170			170			
180			180			
190			190			
200			200			
Obs.:						

Figura 28. Ficha utilizada para coleta de dados de condições operacionais na produção de acrilamida

5.2.4.3 A folha de marcha e o seu manuseio

A folha de marcha do EVOP utilizada é vista na Figura 29. Os parâmetros estatísticos foram inseridos na folha para facilitar o trabalho de cálculo. Os cálculos foram feitos com a planilha Excel.

CÁLCULO DAS MÉDIAS						CALC DESVIO PADRAO			
Condições operacionais	1	2	3	4	5	Soma anterior S	180,60		
Soma do ciclo anterior	945,0	1035,0	1127,0	953,0	971,0	Nova S = Amplitude x K			
Média do ciclo anterior	78,8	86,3	93,9	79,4	80,9	= 28,8 x 0,41 = 11,82			
Novas observações	41,0	58,0	85,0	51,0	58,0	Nova soma S	192,42		
Diferenças (sinais observados)	37,8	28,3	8,9	28,4	22,9	Nova média SA = Nova soma S/(n-1)			
Novas Somas (NS)	986,0	1093,0	1212,0	1004,0	1029,0	Nova média SA = 192,4 / 12 = 16,04			
Novas médias (NSM)	A 75,8	B 84,1	C 93,2	D 77,2	E 79,2	Média anterior			
CÁLCULO DOS EFEITOS						n = 13 (numero de ciclos)			
EFEITO A		EFEITO B		EFEITO AB		CÁLCULO DE 2 EP LIMITES DE ERRO			
D 77,2	B 84,1	C 93,2	B 84,1	C 93,2		Para Novas Médias e Novos efeitos			
E 79,2	C 93,2	D 77,2	E 79,2	D 77,2		L 0,55 x SA 16,04 = 8,8			
F 156,4	G 177,3	F 172,4	G 161,3	F 163,2		Para Efeito de alteração na média			
G 177,3	H 177,3	H 161,3	I 170,5	G 170,5		M 0,50 x SA 16,04 = 8,0			
2 -20,9	2,0	11,1	2,0	-7,2	2,0				
-10,5		5,5		-3,6					
EFEITO DA MUDANÇA NA MÉDIA		MÉDIA DA FASE				FATORES			
F 156,4	A 75,8	H 333,7				n	K	L	M
G 177,3	A 4,0	A 75,8				2	0,30	1,41	1,26
H 333,7	303,4	409,5				3	0,35	1,15	1,85
I 30,3		81,9				4	0,37	1,00	0,89
5 6,1						5	0,38	0,86	0,80
						6	0,38	0,82	0,73
						7	0,40	0,76	0,68
						8	0,40	0,71	0,63
						9	0,41	0,67	0,60
						10	0,41	0,63	0,57
						11	0,41	0,60	0,54
						12	0,41	0,58	0,52
						13	0,41	0,55	0,50
						14	0,41	0,53	0,46

EVOP		Data:	09/12/97	RESULTADOS	
Otimização da produção de Acrilamida	RESPOSTA	Tempo de reação (min)		Efeito A	=> -10,5 ± 8,8
Elaborado por:	NTK			Efeito B	=> 5,5 ± 8,8
Catalizador:	K561+L578			Efeito de média	=> 6,1 ± 8,0
Fase:	1			Efeito AB	=> -3,6 ± 8,8
Ciclo:	13			Obs:	

Figura 29. Folha de marcha com os dados das corridas do ciclo 13 da fase 1

Os efeitos foram estimados no final do ciclo 1, mas como não houve repetição dos ensaios, nenhuma estimativa de variância do erro estaria disponível. Concluído o ciclo 2, tanto a estimativa dos efeitos como a estimativa da variância do erro foram possíveis.

Os cálculos para o preenchimento dos campos da folha de marcha de acordo com o modelo da Figura 29 é feita como segundo o procedimento a seguir:

Diferenças: subtraia as novas observações das médias anteriores. Anote o sinal algébrico da diferença. Exemplo: para ponto 1: $78,8 - 41,0 = 37,8$

Novas Somas: adicione as novas observações às somas anteriores. Exemplo: para ponto 1: $945,0 + 41,0 = 986,0$.

Novas médias: divida as novas somas por N, o número de ciclos. Exemplo: para ponto 1: $986,0/13 = 75,8$

Cálculo dos efeitos (por exemplo, o efeito A): escreva as novas médias para as condições operacionais 4 e 5 opostas (D) e (E). Some as duas de forma a obter o número do espaço (F). Execute a operação correspondente as condições operacionais 2 e 3, (B) e (C) para obter o número para o espaço (G).

A próxima operação é a subtração. copie (G) novamente em baixo de (F) e subtraia (F) de (G) e divida por 2. O sinal deve ser mantido conforme o resultado obtido. Este resultado é transportado para o quadro indicado "RESULTADOS"

No exemplo: $D+E = 77,2 + 79,2 = F = 156,4$; $B + C = 84,1 + 93,2 = G = 177,3$

$F - G = -20,9$; Efeito A = $-20,9/2 = -10,5$

O efeito B e o efeito AB seguem a mesma seqüência, porém nos seus respectivos espaços de cálculo mostrado na folha de marcha da Figura 29.

Efeito da mudança na média: copie (F) e (G) conforme mostrado e some os dois. Multiplique (A) por 4. A próxima operação é a subtração igual à anterior. Divida por 5 No exemplo: $F + G = 156,4 + 177,3 = 333,7$; $A \times 4 = 75,8 \times 4 = 303,4$

Efeito da alteração na média = $(333,7-303,4)/5 = 6,1$

Média da Fase: copie (H) do quadro da mudança na média e (A) do cálculo das médias, some os dois e divida por 5. No exemplo: $(H + A)/5 = (333,7 + 75,8)/5 = 81,9$.

Cálculo do desvio padrão: A amplitude é a diferença algébrica entre as diferenças mais positivas e mais negativas. A amplitude é sempre positiva. O desvio-padrão é estimado multiplicando a amplitude por K. No exemplo: Amplitude = $37,75 - 8,92 = 28,83$

K para 13 ciclos = 0,41, portanto = Desvio padrão S = $28,8 \times 0,41 = 11,85$.

Cálculo de Limites de Erro = 2 x EP: Soma dos desvios padrões anteriores = 180,60; Nova soma S = $180,60 + 11,85 = 192,42$; Limite de Erro para os efeitos A e B = L x Nova soma S/(n-1) = $0,55 \times 192,42/12 = 8,8$; Limite de erro para efeito da alteração na média = M x Nova soma S/(n-1) = $0,50 \times 192,42/12 = 8,0$

Constantes K, L e M: Leia os fatores K, L e M da tabela. K é uma constante que depende de números de ensaios por ciclo e do número de ciclos ensaiados. Este fator é proveniente de um método simplificado para estimar o desvio padrão (BOX e DRAPER, 1969).

$$L = \frac{2}{\sqrt{n}}, \text{ e } M = \frac{1,79}{\sqrt{n}} \quad (26)$$

Em que: n é o número de ciclos do programa até o momento do cálculo.

As instruções de cálculo apresentadas anteriormente são baseadas no tratamento de dados conforme descrito a seguir:

Considerando \bar{y}_i o valor médio da variável de resposta do ensaio nas condições operacionais i, as equações para obter os valores dos efeitos esquematizados na Figura 30 podem ser resumidas a seguir:

$$\text{Efeito } A = \frac{1}{2}(\bar{y}_4 + \bar{y}_5 - \bar{y}_3 - \bar{y}_2)$$

$$\text{Efeito } B = \frac{1}{2}(\bar{y}_3 + \bar{y}_5 - \bar{y}_2 - \bar{y}_4) \quad (27)$$

$$\text{Efeito } AB = \frac{1}{2}(\bar{y}_5 + \bar{y}_2 - \bar{y}_3 - \bar{y}_4)$$

O efeito da alteração na média, é dado por:

$$\text{Efeito da alteração da média} = \frac{1}{5}(\bar{y}_2 + \bar{y}_3 + \bar{y}_4 + \bar{y}_5 + \bar{y}_1) - \bar{y}_1 = \frac{1}{5}(\bar{y}_2 + \bar{y}_3 + \bar{y}_4 + \bar{y}_5 - 4\bar{y}_1) \quad (28)$$

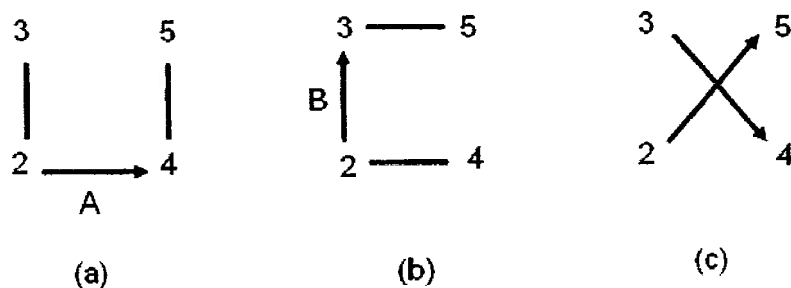


Figura 30. Os efeitos representados em diagramas para o DOE 2^2 , do estudo. No exemplo: (a) efeito da pressão; (b) efeito da relação volumétrica: água/acrilonitrila; (c) interação dos efeitos A e B

Para o exemplo da Figura 29, que corresponde ao final do ciclo 13 da fase 1 para o programa de melhoria da fabricação da acrilamida, verificam-se os resultados da Tabela 29.

Tabela 29 - Resultados obtidos após 13 ciclos da fase 1 do programa EVOP para a melhoria da fabricação de acrilamida

Efeito A	→ $-10,5 \pm 8,8$ (pPressão durante a reação)
Efeito B	→ $5,5 \pm 8,8$ (concentração de reagente, relação água/acrilonitrila)
Efeito da média	→ $6,1 \pm 8,8$ (mudança na média)
Efeito A×B	→ $-3,6 \pm 8,8$ (efeito da interação da pressão × temperatura)
Melhor condição de operação → 1	

O efeito A (pressão durante a reação) foi estimado como $-10,5 \pm 8,8$. Isto significa que o aumento da pressão reduz o valor da função resposta, neste caso o tempo da reação. O efeito B (relação de volume de água /volume de acrilonitrila no carregamento) estimado em $5,5 \pm 8,8$, mostra tendência positiva, isto é, aumentando a relação de volume de água/volume de acrilonitrila o tempo de reação aumenta. Porém o efeito B é pouco significativo devido ao alto erro padrão encontrado em relação ao valor do efeito. O efeito da média (efeito da alteração na média), que foi estimado em $6,1 \pm 8,8$, mostra tendência positiva, indicando que a média encontrada está acima da média calculada dos pontos extremos do programa EVOP. Neste caso, a média encontrada não é favorável ao objetivo do experimento, ou seja, seria melhor se o valor fosse menor que a média calculada, indicando um mínimo próximo a este valor. Embora apenas o efeito da pressão seja altamente significativo, deve-se suspeitar dos efeitos de B e da mudança de média e não da interação A×B. A Figura 31 mostra o diagrama extraído da folha de marcha após o ciclo 13 da fase 1. O referido diagrama indica que o melhor tempo de reação é resultado das condições da reação do ponto 1. A Tabela 29 mostra que as condições da batelada nº1, se repete na batelada nº6 , 11, 16, e assim por diante.

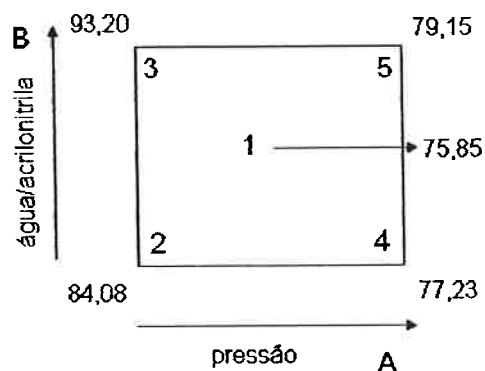


Figura 31. Resultado extraído da folha de marcha após o ciclo 13 da fase 1

O cálculo do erro padrão é baseado em Box e Draper (1969), em que o desvio padrão é estimado calculando-se a amplitude das diferenças observadas (a maior diferença menos a menor diferença, conservando os sinais) multiplicado por um fator K. Este fator depende do número de ciclos ocorrido. O valor de K é encontrado na tabela da folha de marcha da Figura 29, no lado direito inferior. A média do desvio padrão estimado é calculado somando aos desvios padrões calculados anteriormente a cada ciclo e dividindo por $(n-1)$. O cálculo de erro padrão (EP) para os efeitos A, B e da interação é feito multiplicando desvio médio padrão pelo fator L, sendo $L = 2/\sqrt{n}$, e para o efeito da alteração na média, o EP é encontrado multiplicando o desvio padrão médio por M, sendo $M = 1,79/\sqrt{n}$, em que n é o número de ciclos ocorridos.

Quando se encontra significância estatística do efeito, deve-se alterar o plano na direção da “descoberta”. Aumenta-se a amplitude para a variável não significativa, a fim de permitir a verificação da existência real do efeito devido a esta variável. Quando se encontra significância para as duas variáveis, o centro do plano é deslocado nas duas direções, proporcionalmente ao tamanho dos efeitos. Esta é a direção da ascensão mais acentuada. Sempre que o plano for alterado, tem-se o início de uma nova fase. Durante a segunda fase e

as posteriores, utiliza-se, com frequência, a estimativa anterior do erro padrão, uma vez que foi obtida sob o método operacional atual (JURAN, 1993).

5.2.4.4 Alterações

Para que o EVOP seja conservativo, as alterações devem ser contíguas, isto é, um ou mais dos pontos na fase anterior devem coincidir com os da nova fase. Isto condiciona as alterações aos tipos mostrados na Figura 32.

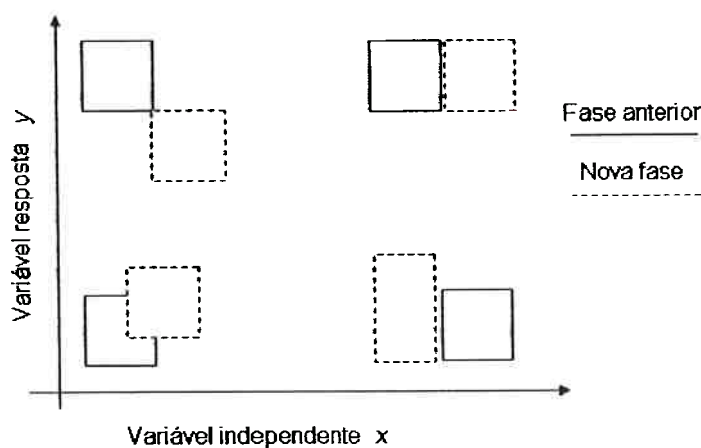


Figura 32. Esquemas para as alterações para fase seguinte de um plano EVOP, mostrando que os pontos devem ser contíguos.

5.2.4.5 Mudança na média

O efeito na mudança na média é a diferença entre a média da fase subtraída da média nas condições de referência. Constitui-se, assim um sinal de curvatura como mostra a Figura 33. Este efeito é usado em conjunto com os efeitos das variáveis para indicar quando um máximo (ou mínimo) for alcançado e, assim, indicar a sensibilidade da resposta às mudanças nos fatores independentes. Ainda segundo Box e Draper (1969) a mudança na média é a medida direta do custo incorrido para obter as informações em uma fase particular. Esta medida

evidencia se as condições experimentadas durante a fase fornece um resultado melhor ou pior do que o das condições de referência. Por isso, ela é uma indicação direta desse custo. Em casos raros, pode ocorrer que a primeira fase esteja localizada simetricamente em torno do máximo com relação aos dois fatores independentes escolhidos. Nesse caso, os fatores devem ser não significativos, mas a mudança na média pode ser significativa.

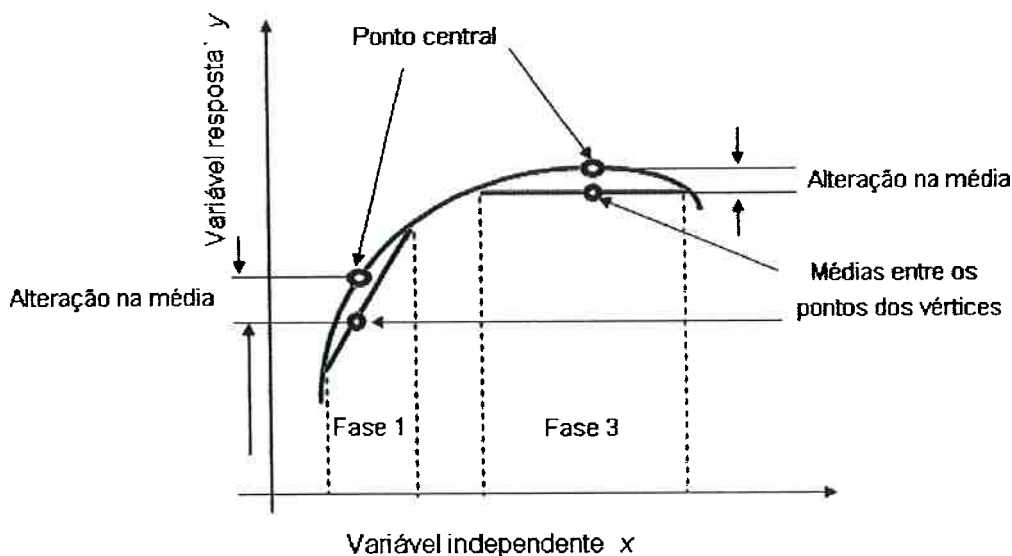


Figura 33. Seção transversal de uma superfície de resposta. Mudança na média indica curvatura (Fonte: JURAN, 1993)

5.2.5 Resultados da fase 1

Os experimentos foram realizados seguindo a metodologia descrita para o programa EVOP no item 5.2.4. Na fase 1 empregaram-se 3 lotes de catalisador, a saber: K561, I576 e A502 e foram realizados um total de 15 ciclos para o estudo EVOP. Isto significa que foram produzidas 75 bateladas de acrilamida com 4000 kg cada, isto é, 300.000 kg de solução de acrilamida a 45%. As folhas de marcha para com os resultados individuais de todos os lotes de catalisador, o procedimento de uso do catalisador bem como a compensação no tempo de

reação devido exaustão do catalisador estão descritos em detalhes na dissertação do Nam (KWOK, 2000).

5.2.5.1 Resultados e discussão com o lote de catalisador K561

A Tabela 30 mostra as condições operacionais obtidas com as 32 bateladas de acrilamida produzidas com o catalisados K561.

Os valores obtidos após a compensação matemática devido a exaustão do catalisador foram utilizadas na folha de marcha EVOP.

O catalisador K561, teve atividade produtiva até obter 33 bateladas, quando foi descartado. Foram consideradas 30 bateladas para EVOP, a fim de permitir a avaliação pela técnica, pois necessita que os ciclos sejam completos. Após 6 ciclos da fase 1, o resultado obtido está ilustrado na

Figura 34.

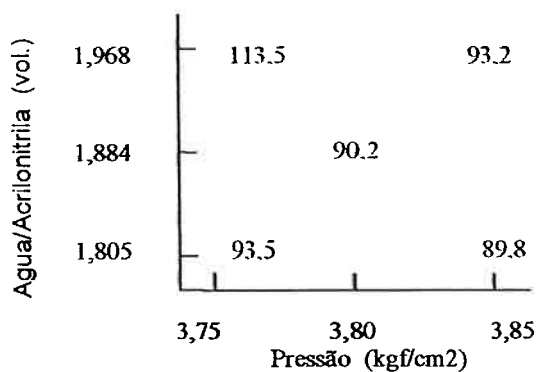
Observando-se a Figura 34 nota-se que os efeitos não são significativos em relação aos desvios. Por exemplo, para o efeito A, o resultado do efeito é $-12,0 \pm 17,2$. O valor de erro está muito acima do valor do efeito obtido, isto significa que as variações do processo são maiores que o efeito da pressão (efeito A). A mesma interpretação ocorre com os outros efeitos.

Existe uma tendência do efeito de pressão (efeito A), indicar que o aumento da pressão causa diminuição do tempo de reação, que é o objetivo do estudo. A tendência do tempo da reação aumentar com a relação água/acrilonitrila é vista pelo resultado $11,7 \pm 17,2$. Portanto para reduzir o tempo devemos reduzir esta relação, o que significa aumentar o carregamento de

acrilonitrila e reduzir o da água. Porém o efeito obtido ainda não apresenta a significância desejada.

Tabela 30 - Dados operacionais obtidos no estudo EVOP - fase 1, da reação de obtenção de acrilamida com o lote de catalisador K561

batch no.	total acrilonitrila (litros)	total de água (litros)	água/lacn em volume	pressão (kgf/cm ²)	tempo de reação (min)	temperatura máxima (°C)	tempo de destilação (min)	OBSERVAÇÕES	tempo de reação ajustado (min)
1	2115	3985	1,884	3,6 - 3,9	120	139	150		118
2	2175	3925	1,805	3,5 - 3,8	100	135	120		97
3	2055	4045	1,868	3,5 - 3,8	90	136	140		85
4	2175	3925	1,805	3,7 - 4,0	110	136	120		103
5	2055	4045	1,868	3,7 - 4,0	100	137	120		91
6	2115	3985	1,884	3,6 - 3,9	100	137	130		90
7	2175	3925	1,805	3,5 - 3,8	130	135	140		118
8	2055	4045	1,868	3,5 - 3,8	180	136	140		166
9	2175	3925	1,805	3,7 - 4,0	90	137	120		75
10	2055	4045	1,868	3,7 - 4,0	90	136	140		73
11	2115	3985	1,884	3,6 - 3,9	100	136	140		81
12	2175	3925	1,805	3,6 - 3,8	90	137	120		69
13	2055	4045	1,868	3,5 - 3,8	110	137	130		88
14	2175	3925	1,805	3,7 - 4,0	120	136	120		96
15	2055	4045	1,868	3,7 - 4,0	90	137	120		84
16	2115	3985	1,884	3,6 - 3,9	90	136	140		63
17	2175	3925	1,805	3,5 - 3,8	140	137	160		111
18	2055	4045	1,868	3,5 - 3,8	130	137	160		99
19	2175	3925	1,805	3,7 - 4,0	100	136	150		67
20	2055	4045	1,868	3,7 - 4,0	150	136	150		116
21	2115	3985	1,884	3,6 - 3,9	130	136	130		94
22	2175	3925	1,805	3,5 - 3,8	130	137	140		92
23	2055	4045	1,868	3,5 - 3,8	180	136	140	faltou energia	141
24	2175	3925	1,805	3,7 - 4,0	140	137	180		99
25	2055	4045	1,868	3,7 - 4,0	160	136	140		117
26	2115	3985	1,884	3,6 - 3,9	140	136	130		95
27	2175	3925	1,805	3,5 - 3,8	120	137	120		74
28	2055	4045	1,868	3,5 - 3,8	150	136	140		102
29	2175	3925	1,805	3,7 - 4,0	150	136	130		100
30	2055	4045	1,868	3,7 - 4,0	150	136	130		99
31	2115	3985	1,884	3,6 - 3,9	160	136	130		107
32	2175	3925	1,805	3,5 - 3,8	200	134	140		145



Efeitos e seus erros padrões

Pressão (A)	-12,0 ± 17,2
Água/acrilonitrila (vol) (B)	11,7 ± 17,2
Efeito da mudança na média	5,9 ± 15,3
Interação AB	- 8,3 ± 17,2

Figura 34. Resultado do EVOP da otimização de produção da acrilamida após 6 ciclos da fase 1, com o lote de catalisador K561. Os resultados são em tempo de reação expressos em minutos

Baseado nos valores obtidos podem-se tomar 2 decisões: a) aumentar o intervalo de variação, ou b) continuar os testes com os mesmos intervalos. Devido as limitações de pressão e verificação da turbidez da solução de acrilamida, optou-se por continuar os testes com os mesmos intervalos.

5.2.5.2 Resultados e discussão da fase 1 com catalisador L576

Utilizando a metodologia descrita anteriormente, foram feitas corridas com o lote L576 do catalisador, e os resultados obtidos estão mostrados na Tabela 31. A compensação pela perda da atividade do catalisador foi feita usando o modelo linear como feito no lote anterior.

O catalisador L576 teve atividade produtiva até obter 15 bateladas. Foram consideradas as 15 bateladas para EVOP, significando 3 ciclos adicionais ao programa. Portanto, tem-se no total

9 ciclos para avaliação. Ao final dos 9 ciclos da fase 1, o resultado obtido é ilustrado na Figura 35.

Tabela 31 - Dados operacionais obtidos no método EVOP, fase 1, da reação de acrilonitrila e água para produção de acrilamida, com catalisador lote L576

batch no.	total acrilonitrila (litros)	total de água (litros)	agua/acn em volume	pressão (kgf/cm ²)	tempo de reação (min)	temperatura máxima (°C)	tempo de destilação (min)	OBSERVAÇÕES	tempo de reação ajustado (min)
1	2115	3905	1,884	3,6 - 3,9	70	137	120		69
2	2175	3925	1,805	3,5 - 3,8	100	137	140		98
3	2055	4045	1,968	3,5 - 3,8	90	136	120		86
4	2175	3925	1,805	3,7 - 4,0	70	137	120		65
5	2055	4045	1,968	3,7 - 4,0	90	137	130		84
6	2115	3905	1,884	3,6 - 3,9	90	137	120		83
7	2175	3925	1,805	3,5 - 3,8	90	136	120		82
8	2055	4045	1,968	3,5 - 3,8	100	136	120		90
9	2175	3925	1,805	3,7 - 4,0	100	136	120		89
10	2055	4045	1,968	3,7 - 4,0	100	137	140		88
11	2115	3905	1,884	3,6 - 3,9	100	136	120		87
12	2175	3925	1,805	3,5 - 3,8	100	136	120		86
13	2055	4045	1,968	3,5 - 3,8	90	136	120		74
14	2175	3925	1,805	3,7 - 4,0	100	136	120		83
15	2055	4045	1,968	3,7 - 4,0	90	136	120		72

Os resultados mostram que os efeitos ainda estão menores do que o erro, isto é, as variações do processo ainda estão acima dos efeitos das variáveis selecionadas para otimização. Os sinais mantiveram-se os mesmos comparando com as bateladas feitas com catalisador anterior.

O efeito da pressão durante a reação (A) indica $-10,1 \pm 11,7$, mostrando tendência de que aumentando a pressão diminui o tempo de reação. O efeito A (pressão) é predominante no conjunto dos efeitos estudados, pois apresenta o maior valor relativo ao erro entre eles. O efeito B (razão água/acrilonitrila) ainda é pouco significativo, porém nota-se que existe uma tendência em sobressair em relação ao ruído, ao se comparar o resultado devido a este efeito com os dos lotes produzidos com o catalisador anterior.

A decisão foi continuar o esquema EVOP nesta mesma fase, pelos mesmos motivos anteriormente citado no catalisador K561.

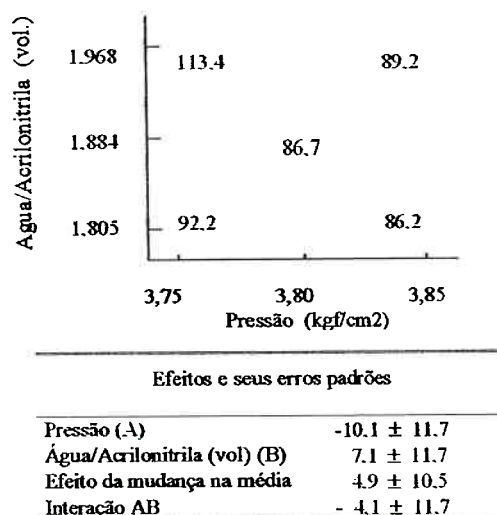


Figura 35. Resultado do EVOP da otimização de produção da acrilamida após 9 ciclos da fase 1; com o lote de catalisador 576. Os resultados são em tempo de reação expressos em minutos

5.2.5.3 Resultados e discussão da fase 1 com catalisador A502

Foram produzidos 6 ciclos nesta etapa. O número total de bateladas foi 36, porém parte dos lotes tiveram problemas de falta de vapor e conseqüentemente interrupção da reação. Foram aproveitados os dados de 30 bateladas, isto é 6 ciclos completos.

A Tabela 32 mostra os dados obtidos e os ajustes feitos devido a perda da atividade do catalisador.

Tabela 32 - Dados operacionais obtidos no estudo EVOP - obtenção de acrilamida com catalisador lote A502

batch no.	total acrilonitrila (litros)	total de água (litros)	água/acn em volume	pressão (kgf/cm ²)	tempo de reação (min)	temperatura média (°C)	tempo de destilação (min)	OBSERVAÇÕES	tempo de reação ajustado (min)
1	2115	3985	1,884	3,6 - 3,9	80	136	140		67
2	2176	3926	1,805	3,5 - 3,8	70	136	130		64
3	2066	4046	1,968	3,6 - 3,8	70	136	130		61
4	2175	3926	1,805	3,7 - 4,0	80	137	140		68
5	2066	4046	1,968	3,7 - 4,0	70	136	140		66
6	2115	3985	1,884	3,6 - 3,9	90	137	120		72
7	2176	3926	1,805	3,5 - 3,8	110	136	110		88
8	2066	4046	1,968	3,6 - 3,8	80	137	120		66
9	2176	3926	1,805	3,7 - 4,0	90	136	130		62
10	2066	4046	1,968	3,7 - 4,0	100	136	130		69
11	2115	3985	1,884	3,6 - 3,9	70	136	120		36
12	2175	3926	1,805	3,5 - 3,8	90	137	130		63
13	2066	4046	1,968	3,5 - 3,8	120	137	130		80
14	2175	3926	1,805	3,7 - 4,0	90	137	140		47
15	2066	4046	1,968	3,7 - 4,0	90	136	130		44
16	2115	3985	1,884	3,6 - 3,9	90	136	110		41
17	2176	3926	1,805	3,5 - 3,8	110	136	140		58
18	2066	4046	1,968	3,5 - 3,8	140	137	130		85
19	2176	3926	1,805	3,7 - 4,0	110	137	130		61
20	2066	4046	1,968	3,7 - 4,0	120	137	130		58
21	2115	3985	1,884	3,6 - 3,9	130	137	120	problemas de informação	65
22	2175	3926	1,805	3,6 - 3,8	170	136	170	"	102
23	2066	4046	1,968	3,5 - 3,8	150	137	130	"	79
24	2176	3926	1,805	3,7 - 4,0				"	
25	2066	4046	1,968	3,7 - 4,0				"	
26	2115	3985	1,884	3,6 - 3,9	150	137	130	"	70
27	2175	3926	1,805	3,5 - 3,8	140	137	150	"	57
28	2066	4046	1,968	3,5 - 3,8	150	137	130		64
29	2176	3926	1,805	3,7 - 4,0	150	136	120		61
30	2066	4046	1,968	3,7 - 4,0	140	137	140		48
31	2115	3985	1,884	3,6 - 3,9	140	136	140		46
32	2176	3926	1,805	3,6 - 3,8	150	137	100		51
33	2066	4046	1,968	3,5 - 3,8	180	136	130		78
34	2176	3926	1,805	3,7 - 4,0	180	137	120		76
35	2066	4046	1,968	3,7 - 4,0	180	136	160		72
36	2115	3985	1,884	3,6 - 3,9	200	136	130		

5.2.5.4 Discussão dos resultados ao final de 15 ciclos da fase 1 com os lotes: K561+L576+A502 do catalisador

Ao final de 15 ciclos com 3 lotes de catalisadores, obteve-se para programa EVOP 75 bateladas.

Os valores extraídos da Figura 36 foram transportados para Figura 37. As seguintes conclusões foram obtidas após uma análise detalhada dos resultados.

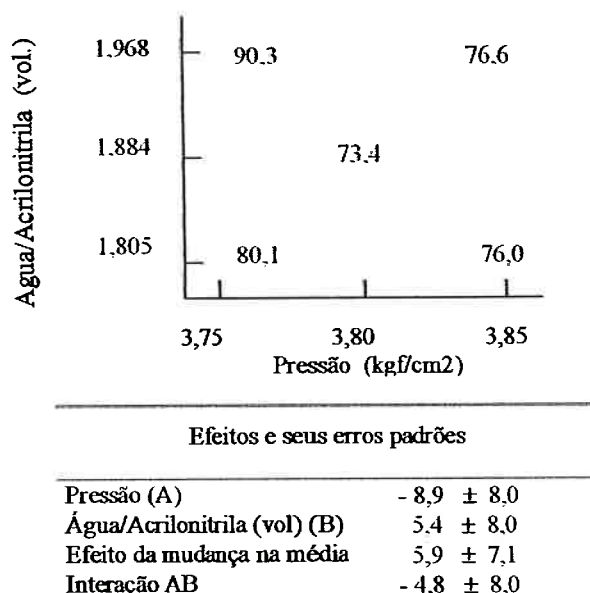


Figura 36. Resultado do EVOP da otimização de produção da acrilamida após 15 ciclos da fase 1. Os resultados são em tempo de reação expressos em minutos

As seguintes conclusões foram obtidas após uma análise detalhada dos resultados:

1. o limite de erro (2 x erro padrão) ficou reduzido a valores que fizeram os efeitos sobressaírem. O efeito da pressão durante a reação apresentou resultado de $-8,9 \pm 8,0$, indicando que o limite de erro está menor que o efeito. Para os outros efeitos não houve o mesmo resultado relativo;

2. efeito A, pressão: $-8,9 \pm 8,0$. Indica que aumentando a pressão, a produtividade aumenta, isto é, o tempo de reação diminui, pois o sinal negativo informa que a variável objetiva é de sentido oposto da variável causadora do efeito;
3. embora o efeito B não seja significativo, $5,4 \pm 8,0$, nota-se que a relação água/acrilonitrila aumenta o tempo de reação. ou seja, para diminuir o tempo de reação deve-se diminuir a relação acima. Este é um fator que pode ser avaliado melhor com intervalos maiores de variação;
4. O efeito da mudança na média: $5,9 \pm 7,1$, o efeito é positivo, isto é, o valor da média (tendência linear) está acima do ponto médio (ponto 1), significando que existe uma curvatura com um mínimo na superfície de resposta, o que favorece o objetivo do experimento;
5. O efeito AB, $-4,8 \pm 8,0$, interação do efeito da pressão com a relação de água/acrilonitrila, não é significativo. O limite do erro está muito acima do efeito obtido;
6. A qualidade ficou comprometida nesta fase, a discussão foi feita separadamente no item 5.2.6.7, por se tratar de um parâmetro de alta importância na fabricação da acrilamida. A consequência de um produto de baixa qualidade afeta a qualidade do polímero produzido com este monômero.

A qualidade da acrilamida produzida foi representada pelo teor de dímero ou polímero que indiretamente é medido pela turbidez na presença de isobutanol 1:1.

A medição da turbidez em FTU, sob 3 condições:

Condição A: Operação sem esquema EVOP, em condições estáticas, antes da aplicação do EVOP. A pressão média das bateladas é de $3,65 \text{ kgf/cm}^2$

Condição B: Fase 1 do EVOP. Pressão média de $3,75 \text{ kgf/cm}^2$

Condição C: Fase 2 do EVOP, será discutida posteriormente. A Pressão média do processo é de $3,70 \text{ kgf/cm}^2$.

A influência da pressão na formação do dímero ou polímero era esperada, pois é sabido que a temperatura maior promove formação de dímero ou polímero, tendo em vista que o aumento de pressão resulta em aumento de temperatura.

A medição da turbidez não foi feita a cada batelada e sim a cada 3 bateladas, isto é, a medição da turbidez é feita quando são misturadas 3 bateladas num tanque agitado mecanicamente. Após misturadas as 3 bateladas, é processada a filtração para eliminação de resíduos do catalisador. Esta fase do processo está fora do escopo do EVOP, porém é mencionada somente para que se possa tomar conhecimento da metodologia utilizada para amostragem do produto para obtenção dos resultados de controle de qualidade.

A análise dos resultados da fase 1, indicou que a média obtida antes do programa EVOP era 11,6 FTU e na fase 1 foi encontrada uma média de 17,6. A especificação máxima é de 30 FTU, e verifica-se que em alguns casos da fase I, a turbidez ultrapassou a especificação, colocando o produto sob risco de reprovação. Quanto mais próximo da exaustão do catalisador a turbidez ficava maior, indicando que a formação de dímero ou polímero é função também do tempo de residência no reator, pois, a perda da atividade do catalisador resulta em aumento de tempo de reação e consequentemente o tempo de residência.

Aos resultados finais da fase 1 permitiram a programação da fase 2.

5.2.6 Resultados e discussão da fase 2

Na fase 2 foram consideradas as limitações da pressão do reator, devido ao comprometimento com a qualidade da acrilamida produzida. A pressão média foi reduzida para $3,70 \text{ kgf/cm}^2$, no intervalo de $3,5$ a $3,9 \text{ kgf/cm}^2$ (na fase anterior era de $3,75 \text{ kgf/cm}^2$, no intervalo de $3,5$ a $4,0 \text{ kgf/cm}^2$). O índice (volume de água)/(volume de acrilonitrila) foi alterado, acrescentando 1% de acrilonitrila, uma vez que esta razão deve diminuir para melhorar a produtividade.

A Tabela 33 mostra as condições operacionais planejadas para a fase 2 do EVOP. A Figura 37 mostra uma representação esquemática comparando graficamente as condições planejadas para as duas fases de EVOP. Nota-se que o gráfico indica a direção em que os experimentos devem ser conduzidos

Os pontos 2 e 1' coincidem e a direção foi tomada no sentido de procurar menor tempo de reação.

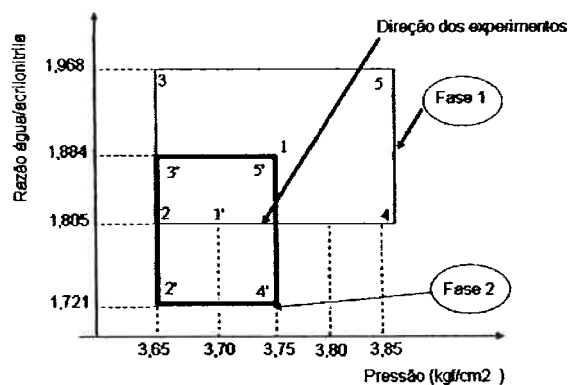


Figura 37. Representação esquemática do EVOP para processo de fabricação de acrilamida. As condições 1 a 5 são para fase 1 e as condições de 1' a 5' são para fase 2

A fase 2 foi realizada com o lote C609 do catalisador. Foram produzidas 33 bateladas, com 4000 kg de produto por batelada, sendo aproveitadas 30 para efeito de cálculo, ou seja, 6 ciclos completos. Os resultados operacionais estão apresentados na Tabela 34.

Tabela 33 - Tabela de acompanhamento com as condições operacionais para fase 2 do programa EVOP do processo de fabricação da acrilamida

batelada no.	total Acrilonitrila (litros)	total água (litros)	agua/acn em volume	pressão (kgf/cm2)	tempo de reação (min)	temperatura máxima (°C)	tempo de destilação (min)	OBSERVAÇÕES
1	2175	3925	1,805	3,6-3,8				
2	2242	3858	1,721	3,5-3,8				
3	2115	3985	1,884	3,5-3,8				
4	2242	3858	1,721	3,6-3,9				
5	2115	3985	1,884	3,6-3,9				
6	2175	3925	1,805	3,6-3,8				
7	2242	3858	1,721	3,5-3,8				
8	2115	3985	1,884	3,5-3,8				
9	2242	3858	1,721	3,6-3,9				
10	2115	3985	1,884	3,6-3,9				
11	2175	3925	1,805	3,6-3,8				
12	2242	3858	1,721	3,5-3,8				
13	2115	3985	1,884	3,5-3,8				
14	2242	3858	1,721	3,6-3,9				
15	2115	3985	1,884	3,6-3,9				
16	2175	3925	1,805	3,6-3,8				
17	2242	3858	1,721	3,5-3,8				
18	2115	3985	1,884	3,5-3,8				
19	2242	3858	1,721	3,6-3,9				
20	2115	3985	1,884	3,6-3,9				
21	2175	3925	1,805	3,6-3,8				
22	2242	3858	1,721	3,5-3,8				
23	2115	3985	1,884	3,5-3,8				
24	2242	3858	1,721	3,6-3,9				
25	2115	3985	1,884	3,6-3,9				
26	2175	3925	1,805	3,6-3,8				
27	2242	3858	1,721	3,5-3,8				
28	2115	3985	1,884	3,5-3,8				
29	2242	3858	1,721	3,6-3,9				
30	2115	3985	1,884	3,6-3,9				
31	2175	3925	1,805	3,6-3,8				
32	2242	3858	1,721	3,5-3,8				
33	2115	3985	1,884	3,5-3,8				
34	2242	3858	1,721	3,6-3,9				
35	2115	3985	1,884	3,6-3,9				
36	2175	3925	1,805	3,6-3,8				
37	2242	3858	1,721	3,5-3,8				
38	2115	3985	1,884	3,5-3,8				
39	2242	3858	1,721	3,6-3,9				
40	2115	3985	1,884	3,6-3,9				
41	2175	3925	1,805	3,6-3,8				
42	2242	3858	1,721	3,5-3,8				
43	2115	3985	1,884	3,5-3,8				
44	2242	3858	1,721	3,6-3,9				
45	2115	3985	1,884	3,6-3,9				

Quanto à exaustão do catalisador, o ajuste foi feito seguindo a mesma metodologia da fase 1, levando em conta a exaustão do referido lote.

Deve-se ressaltar que os desvios foram calculados utilizando a estimativa da fase anterior do erro padrão, por se considerar que a operação está sob as mesmas variações (ruidos) da fase 1. Os resultados referentes a fase 1, após 15 ciclos, e a fase 2 estão indicados na Figura 38.

Em relação à fase 2, as variações significativas foram o efeito da pressão ($-10,3 \pm 6,0$) e o da mudança da média ($-10,0 \pm 5,3$). Os efeitos da relação água/acrilonitrila ($1,9 \pm 6,0$) e a interação das duas variáveis estudadas ($-0,9 \pm 6,0$) não foram significativos.

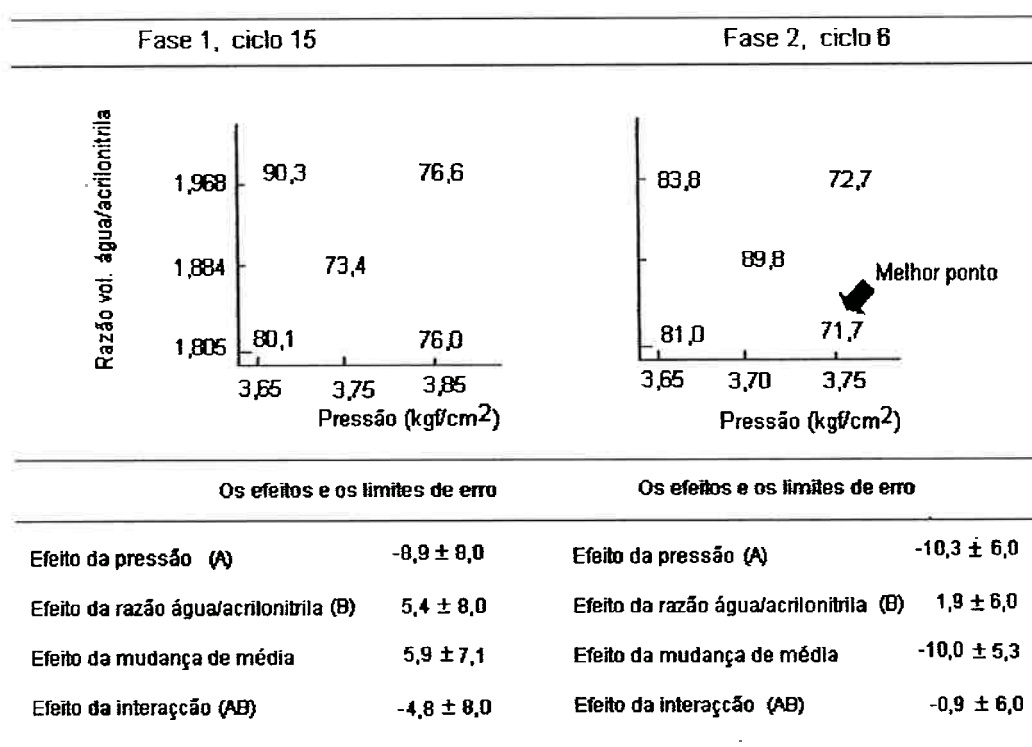


Figura 38. Resultado final do programa EVOP para otimização de produção de acrilamida. A variável objetiva é o tempo de reação em minutos

Tabela 34 - Resultados operacionais obtidos no estudo EVOP - fase 2, da reação de obtenção de acrilamida

batch no.	total acrilonitrila (litros)	total de água (litros)	água/ach em volume	pressão (kgf/cm ²)	tempo de reação (min)	temperatura máxima (°C)	tempo de destilação (min)	OBSERVAÇÕES	tempo de reação ajustado (min)
1	2176	3926	1,806	3,6-3,8	100	137	120		98
2	2242	3868	1,721	3,6-3,8	100	137	120		98
3	2116	3986	1,884	3,6-3,8	100	137	130		93
4	2242	3868	1,721	3,6-3,8	100	136	120		91
5	2116	3986	1,884	3,6-3,8	90	136	120		79
6	2176	3926	1,806	3,6-3,8	90	136	120		77
7	2242	3868	1,721	3,6-3,8	100	137	140	faltou vapor	84
8	2116	3986	1,884	3,6-3,8	100	136	120		82
9	2242	3868	1,721	3,6-3,8	90	137	120		70
10	2116	3986	1,884	3,6-3,8	100	136	120		78
11	2176	3926	1,806	3,6-3,8	100	136	160	problemas de falta de vapor	76
12	2242	3868	1,721	3,6-3,8	80	137	120		63
13	2116	3986	1,884	3,6-3,8	100	137	120		71
14	2242	3868	1,721	3,6-3,8	100	137	120		69
15	2116	3986	1,884	3,6-3,8	100	136	120		67
16	2176	3926	1,806	3,6-3,8	120	136	140		84
17	2242	3868	1,721	3,6-3,8	100	136	120		62
18	2116	3986	1,884	3,6-3,8	120	136	110		80
19	2242	3868	1,721	3,6-3,8	90	137	120		48
20	2116	3986	1,884	3,6-3,8	100	137	120		58
21	2176	3926	1,806	3,6-3,8	160	136	130		113
22	2242	3868	1,721	3,6-3,8	160	137	120		101
23	2116	3986	1,884	3,6-3,8	150	137	130		99
24	2242	3868	1,721	3,6-3,8	130	137	120		77
25	2116	3986	1,884	3,6-3,8	140	137	120		84
26	2176	3926	1,806	3,6-3,8	160	137	120		92
27	2242	3868	1,721	3,6-3,8	160	137	130		90
28	2116	3986	1,884	3,6-3,8	140	137	120		78
29	2242	3868	1,721	3,6-3,8	140	136	120		75
30	2116	3986	1,884	3,6-3,8	140	136	120		73
31	2176	3926	1,806	3,6-3,8	150	136	140		
32	2242	3868	1,721	3,6-3,8	150	136	120		
33	2116	3986	1,884	3,6-3,8	150	136	120		

O aumento da pressão causa redução no tempo de reação, porém existe uma restrição a esse aumento, devido à piora na qualidade do produto.

O efeito da mudança na média negativo ($-10,0 \pm 5,3$) indica que o tempo de reação obtido no ponto central (1) está maior que a média dos tempos obtidos nos 4 outros pontos (2, 3, 4 e 5). Isto indica que ainda existe possibilidade de redução do tempo no domínio delimitado pela fase 2. A mudança na média neste sentido não traria um efeito de redução de tempo de reação. Comparando o efeito da mudança na média da fase 1 ($5,9 \pm 7,1$) com o valor obtido na fase 2 ($-10,0 \pm 5,3$), verifica-se que houve inversão do sinal, indicando a inversão de um ponto de mínimo na superfície de resposta para um ponto de máximo. Com este tipo de variação, conclui-se que ainda não é possível definir uma tendência em relação a efeito da mudança na média. Esta tendência poderá ser verificada com a continuidade do experimento em uma nova fase.

Verificando-se o diagrama da Figura 38, o melhor ponto a ser operado é o ponto 4 da fase 2, indicado com uma flecha na figura (ou ponto 4 da Figura 37), devido ao menor tempo de reação obtido. As condições de operação são:

Carga de acrilonitrila = 2242 L

Carga de água = 3858 L

Pressão = 3,6 a 3,9 kgf/cm² (média = 3,75 kgf/cm²)

Quanto à qualidade, verificou-se uma melhora expressiva após a fase 2 do programa EVOP. O produto ficou dentro das especificações, indicando que é possível trabalhar em toda a gama de pressão do EVOP fase 2. Por este motivo, foi permitido escolher a pressão no limite máximo (3,9 kg/cm²), sem risco de obter um produto com qualidade reprovada.

5.2.7 Considerações finais sobre o programa EVOP

O método EVOP é uma das mais importantes aplicações do planejamento fatorial em dois níveis. É um método operacional de fácil tratamento e que pode ser usado em qualquer tipo de processo em que seja possível impor pequenas variações. Mesmo que inicialmente sejam verificadas variações que aparentemente estão encobertas por ruídos de processo, após um determinado número de corridas sob mesmas condições, os efeitos das variáveis controláveis (independentes) se sobressaem devido à possibilidade da repetição dos ciclos, diminuindo o erro padrão. No caso deste estudo foi necessário repetir 15 ciclos para que o efeito apresentasse um certo destaque. Aparentemente é um número muito grande de corridas. Porém levando-se em conta que tudo é feito em condições normais de produção e utilizando o produto obtido como se nada houvesse ocorrido, vê-se que é vantajosa a aplicação da técnica. No caso específico deste estudo, as expectativas foram superadas, porque além do ganho em produtividade houve um ganho adicional de qualidade do produto. Convém ressaltar que um programa dessa natureza envolve uma equipe que deve ser treinada para acompanhar toda a programação e, portanto o sucesso da aplicação do programa depende do envolvimento de todos da equipe.

Capítulo 6

Projetos fatoriais fracionados em dois níveis, DOE 2^{k-p}

Os projetos fatoriais em dois níveis são muito úteis para estimativas de efeitos principais e de interações. Eles indicam se o fenômeno tem ou não um comportamento linear. No entanto, quando se suspeita que 5 ou mais variáveis podem influenciar o fenômeno em questão, a série experimental (que cresce geometricamente) torna-se impossível de ser conduzida por razões de custos experimentais e disponibilidade de tempo. Além disso, a série completa traria informações que poderiam não apresentar interesse prático, como interações de quatro fatores em diante. Por esta razão, é recomendado o uso de planejamento fatorial fracionado, principalmente nos casos em que o número de variáveis independentes investigado é maior do que 4, e em casos extremos, realizam-se projetos fracionários com 3 variáveis. As frações são cuidadosamente definidas levando em consideração o princípio da hierarquia dos efeitos, isto é, os efeitos principais são mais importantes do que efeitos de interações de dois fatores, que são mais importantes do que as interações de três fatores, e assim por diante.

Assim, com base nesse princípio e em algumas restrições (decorrentes de suspeitas reais da influência de alguns efeitos principais e interações, principalmente de dois fatores), os projetos fatoriais fracionados podem ser usados em inúmeras situações.

Em sua forma saturada (quando todas as observações são usadas na estimativa da média e dos efeitos principais) eles são empregados na seleção de variáveis independentes (fatores, variáveis de processo, variáveis de entrada).

Existem várias categorias de projetos fracionados: os projetos em dois níveis, os projetos em 3 ou mais níveis, os projetos mistos, e assim por diante. No entanto, a fração eleita para a

investigação deve ser escolhida com base em princípios bem definidos, como explicado nas obras de Box, Hunter, W e Hunter, J. (2005), Montgomery (2005); Launsby (1996), entre outros.

Neste texto, a ênfase será dada aos projetos fatoriais fracionados em dois níveis como o grande mensageiro de informações para desdobramentos futuros. Para isto serão discutidos dois exemplos que esclarecem muito bem a capacidade que estes planejamentos têm de fornecer informações preciosas para a compreensão de problemas específicos.

6.1 A Construção

Box, Hunter, W e Hunter, J. (1978) iniciam o capítulo 12 falando da redundância que um planejamento completo com 7 ($2^7=128$ ensaios) variáveis, avaliando quais ele deveria selecionar para serem estimadas, e questionando sobre a importância de cada uma das 128 estatísticas. Por exemplo, o que as interações de 4 fatores em diante poderiam trazer de benefício ao estudo? Se, por exemplo, fossem realizados os 128 ensaios, e os valores desses 127 efeitos (1 grau de liberdade é consumido para a estimativa da média) fossem representados graficamente em escala de probabilidade normal, provavelmente 3, 4 ou 5 efeitos seriam considerados importantes. O que fazer com os 122 efeitos restantes? Eles teriam dimensões tão próximas do erro que não apresentariam interesse prático. Com base nessas evidências, e lembrando que o domínio experimental estudado é restrito, principalmente em experiências industriais (vide os casos dos capítulos 4 e 5 deste texto), a redundância de uma série experimental extensa não traria benefício ao estudo. É claro que, nos projetos fracionados, as informações julgadas a priori como não importantes, irão ser estimadas junto com as eleitas (as de maior interesse). Esses contrastes são denominados de “*alias*”, e neles dois ou mais efeitos são representados pela mesma comparação (diferentes

nomes para o mesmo efeito computado). As “alias” consideradas significativas são interpretadas pelo princípio da hierarquia dos efeitos. Desta forma a atribui-se, em princípio, a significância ao efeito de mais alta ordem. Portanto, pode-se considerar que interações de terceira ordem em diante são negligenciadas.

Lembrando que no projeto 2^7 a série completa pode estimar 128 estatísticas que são:

média	Interações						
	efeitos principais	2-fatores	3-fatores	4-fatores	5-fatores	6-fatores	7-fatores
1	7	21	35	35	21	7	1

O desdobramento do efeito aqui apresentado indica que é desnecessário, em uma investigação preliminar, estimar tantas quantidades sendo que apenas poucas sejam selecionadas.

6.1.1 A fração $\frac{1}{2}$

Para melhor compreensão do princípio de construção dos planejamentos fatoriais, aqui será apresentada a construção e a análise da fração $\frac{1}{2}$ de um planejamento fatorial 2^5 (série completa constituída por 32 observações), cuja fração $\frac{1}{2}$ é 2^{5-1} , isto é, $\frac{2^5}{2} = 16$.

Neste caso a série completa seria constituída de:

média	interações				
	efeitos principais	2-fatores	3-fatores	4-fatores	5-fatores
1	5	10	10	5	1

Ao se realizar apenas uma série de 16 experimentos, certamente estão sendo eleitos: a média, os efeitos principais e as interações de dois fatores, isto se constitui em 16 quantidades. E o que será feito com as outras 16 quantidades? As 10 interações de três fatores, as 5 interações de 4 fatores e a interação de 5 fatores? Elas serão estimadas em conjunto com as 16 primeiras quantidades relatadas na forma de “*alias*” também conhecido como “*counfunded*”.

6.1.2 A escolha das 16 observações

Inicialmente é importante lembrar que o projeto básico para esta geração é o completo 2^4 , como foi mencionado no início do capítulo 4, na lista de importância dos projetos fatoriais em dois níveis. O roteiro, para esta construção é:

1. construir matriz de sinais – e + , também conhecida como matriz de contrastes, do projeto básico 2^4 . Este planejamento tem 4 colunas de sinais correspondentes as variáveis, A, B, C e D (notação usada nos programas estatísticos);
2. definir a coluna de sinais – e + correspondente a quinta variável, E, a partir das colunas de sinais do projeto básico, multiplicando-se os sinais das colunas algebricamente;
3. a coluna 5 neste caso tem exatamente a seqüência de sinais da coluna ABCD, por exemplo.

Desta forma, ficou estabelecido que cada quantidade estimada é “*alias*”, para a fração $\frac{1}{2}$, de outra quantidade, correspondente a outra metade da fração (outros 16 ensaios) que estimariam interações de três, quatro e cinco fatores.

Como ilustração a Tabela apresenta a matriz do projeto 2^{5-1} , gerado pela definição da quinta coluna com o padrão: E=ABCD

Tabela 35 - Matriz de sinais do projeto fracionado 2^{4-1} , definindo a variável $E=ABCD$, e relação entre as colunas e o padrão representado pelos contrastes, /

#	projeto					interações de dois fatores									
	A	B	C	D	E	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
17	-	-	-	-	+	+	+	+	-	+	+	-	+	-	-
2	+	-	-	-	-	-	-	-	-	+	+	+	+	+	+
3	-	+	-	-	-	-	+	+	+	-	-	-	+	+	+
20	+	+	-	-	+	+	-	-	+	-	-	+	+	-	-
5	-	-	+	-	-	+	-	+	+	-	+	+	-	-	+
22	+	-	+	-	+	-	+	-	+	-	+	-	-	+	-
23	-	+	+	-	+	-	-	+	-	+	-	+	-	+	-
8	+	+	+	-	-	+	+	-	-	+	-	-	-	-	+
9	-	-	-	+	-	+	+	-	+	+	-	+	-	+	-
26	+	-	-	+	+	-	-	+	+	+	-	-	-	-	+
27	-	+	-	+	+	-	+	-	-	-	+	+	-	-	+
12	+	+	-	+	-	+	-	+	-	-	+	-	-	+	-
29	-	-	+	+	+	+	-	-	-	-	-	-	+	+	+
14	+	-	+	+	-	-	+	+	-	-	-	+	+	-	-
15	-	+	+	+	-	-	-	-	+	+	+	-	+	-	-
32	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Relação entre os pares de colunas			Padrão de confundimentos “(confounding)”		
A	=	BCDE	l_A	→	A + BCDE
B	=	ACDE	l_B	→	B + ACDE
C	=	ABDE	l_C	→	C + ABDE
D	=	ABCE	l_D	→	D + ABCE
E	=	ABCD	l_E	→	E + ABCD
AB	=	CDE	l_{AB}	→	AB + BCDE
AC	=	BDE	l_{AC}	→	AC + CDE
AD	=	BCE	l_{AD}	→	AD + BCE
AE	=	BCD	l_{AE}	→	AE + BCD
BC	=	ADE	l_{BC}	→	BC + ADE
BD	=	ACE	l_{BD}	→	BD + ACE
BE	=	ACD	l_{BE}	→	BE + ACD
CD	=	ABE	l_{CD}	→	CD + ABE
CE	=	ABD	l_{CE}	→	CE + ABD
DE	=	ABC	l_{DE}	→	DE + ABC

(I=ABCDE)

Para construir um planejamento desta natureza de uma forma mais concisa é comum usar a seguinte notação:

1. cada coluna de sinais – e + pode ser representada por uma ou mais letras maiúsculas, como já foi apresentado no início deste item. Assim, a variável 2 pode ser representada por B e a interação das quatro primeiras variáveis, ABCD;
2. nas operações entre colunas de sinais qualquer coluna multiplicada por ela mesma gera uma coluna de sinais +, aqui identificada como matriz coluna identidade I. Isto é: $A \times A = A^2 = I$. A identidade I é usada como elemento neutro da multiplicação de colunas. Ou seja, $A \times I = A$, $AB \times I = AB$ e assim por diante.

A partir deste princípio pode se construir qualquer projeto fracionado. Por outro lado, é conveniente lembrar que não se deve definir as colunas arbitrariamente, isto é, nunca gerar “*alias*” ou “*confounding*” de um efeito principal com outro efeito principal. Isto porque na hierarquia de efeitos eles seriam potencialmente os mais importantes.

6.1.3 O complemento da fração $\frac{1}{2}$

A série completa de ensaios do projeto 2^5 teria tamanho 32. Apresentou-se um projeto alternativo, com 16 ensaios, que será capaz de fornecer informações preciosas, já que os efeitos principais só seriam confundidos com interações de 4 fatores. Estas podem ser desprezadas sem risco para as conclusões sobre o estudo. O número que está na primeira parte coluna da matriz de sinais é o número correspondente ao do ensaio do projeto 2^5 completo, construído na ordem padrão. Agora surge uma nova questão: o que os outros 16 ensaios fariam se fossem eles os eleitos? Como seria a geração da quinta variável? Ela poderia ser gerada por $E = -ABCD$. Então quinta coluna fica com a seqüência de sinais dos 16 experimentos restantes, isto é, a fração complementar. Neste caso a matriz Identidade seria : $I = -ABCDE$. e os modelos confundidos seriam os apresentados na Tabela 36.

Tabela 36 - Matriz de sinais do projeto fracionado 24-1, definindo a variável E = ABCD, e relação entre as colunas e o padrão representado pelos contrastes ℓ'

Relação entre os pares de colunas	Padrão de confundimentos “(confounding)”
A = - BCDE	$\ell'_A \rightarrow A - BCDE$
B = - ACDE	$\ell'_B \rightarrow B - ACDE$
C = - ABDE	$\ell'_C \rightarrow C - ABDE$
D = - ABCE	$\ell'_D \rightarrow D - ABCE$
E = - ABCD	$\ell'_E \rightarrow E - ABCD$
AB = - CDE	$\ell'_{AB} \rightarrow AB - BCDE$
AC = - BDE	$\ell'_{AC} \rightarrow AC - CDE$
AD = - BCE	$\ell'_{AD} \rightarrow AD - BCE$
AE = - BCD	$\ell'_{AE} \rightarrow AE - BCD$
BC = - ADE	$\ell'_{BC} \rightarrow BC - ADE$
BD = - ACE	$\ell'_{BD} \rightarrow BD - ACE$
BE = - ACD	$\ell'_{BE} \rightarrow BE - ACD$
CD = - ABE	$\ell'_{CD} \rightarrow CD - ABE$
CE = - ABD	$\ell'_{CE} \rightarrow CE - ABD$
DE = - ABC	$\ell'_{DE} \rightarrow DE - ABC$

(I=ABCDE)

E os efeitos, estimados com os contrastes da fração complementar, seriam muito parecido com os do da fração anterior, porque certamente as interações de 3, 4 e 5 fatores teriam valores negligenciáveis.

A combinação das duas frações gera o projeto completo, que pelos contrastes ℓ e ℓ' , chega-se a:

$$\frac{1}{2}(\ell_A + \ell'_A) = \frac{1}{2}[(A + BCDE) + (A - BCDE)] = \frac{1}{2} \times 2A = A \quad (29)$$

Fazendo-se isso com todos os contrastes, é possível reconstruir novamente um projeto a partir da composição das frações até chegar ao projeto completo. É claro que isso traz alguns inconvenientes, porque cada fração pode funcionar como um bloco de experimentos, o que pode trazer algum efeito indesejável.

6.1.4 Resolução de um projeto fatorial fracionado

Para evitar enganos, existe um critério básico usado nos projetos fracionados chamado de **resolução do projeto**. Seria semelhante a se fazer uma análise cromatográfica na qual alguns componentes da mistura analisadas tivessem tempo de retenção na coluna cromatográfica muito próximos. O que aconteceria com os picos destes dois componentes? O analista não saberia discernir qual o teor de cada componente. Ele, na realidade, encontraria um teor que corresponderia a soma dos teores individuais de cada um deles. Esta mesma comparação pode ser usada para os projetos fracionados.

Por exemplo, sendo um dado efeito significativo, mas, por descuido o projeto foi construído, em um padrão que confunde dois efeitos principais, a quem atribuir a significância do efeito? As duas variáveis seriam igualmente importantes? Uma seria muito mais importante do que a outra? A qual delas afinal deve ser atribuído significado deste efeito? Para evitar estas impropriedades usa-se o conceito de resolução mencionado. A resolução escolhida para a construção do projeto deve levar em conta principalmente o custo e o tempo gastos na experimentação. Usam-se basicamente três tipos de resolução, assim designadas R-III, R-IV e R-V ou R-V+.

1. Um projeto que tem Resolução III não confunde efeito principal com outro, mas confunde efeitos principais com interações de dois fatores;
2. Um projeto que tem resolução IV não confunde efeitos principais com interações de dois fatores, mas confunde interações de dois fatores com outras interações de dois fatores.
3. Um projeto que tem resolução V não confunde efeitos principais nem interações de dois fatores umas com as outras, mas confunde interações de dois fatores com interações de três fatores e assim por diante.

A notação mais comum incluindo a resolução do projeto é colocar o número que corresponde a resolução em algarismos romanos como sub-índice. O projeto fracionado que estuda 5 variáveis com 16 ensaios, isto é, $25-1$ (fração $\frac{1}{2}$ do 25) é representado incluindo a sua resolução com: 2^{5-1}_V .

Os padrões ou modelos confundidos (*confounding pattern*) são escritos a partir da relação completa de definição, que também é conhecida como “*palavras*”, expressa a partir do(s) gerador(es) do projeto. O tamanho da menor *palavra* da definição completa do projeto também identifica a resolução do projeto. No caso da fração $\frac{1}{2}$ a relação completa de definição tem apenas uma identidade (uma única *palavra* com 5 caracteres, $I=ABCDE$), ele é um projeto fracionado com resolução V.

Pela outra definição de resolução, a resolução V, isto é, efeitos principais são confundidos com interações de 4 fatores e interações de dois fatores com interação de três, chega-se a mesma conclusão da definição de resolução pelo tamanho da menor *palavra*.

A matriz identidade, de um projeto fatorial fracionado, é a relação completa de definição porque a partir dela pode ser escrito todo o padrão apresentado na Tabela 36 usando a regra de operação mostrada anteriormente. Para exemplificar, serão estimados os contrastes I_D e I_{CE} do projeto 2^{5-1} .

Partindo-se da identidade:

$$I=ABCDE$$

As colunas com a seqüência idêntica de sinais serão dadas por:

$$D \times I = ABCDE \times D \text{ ou operado, } D = ABCE, \text{ porque } D \times I = D \text{ e } D \times D = I$$

Da mesma maneira:

$$CE \times I = ABCDE \times CE \text{ ou, } CE = ABD,$$

porque $CE \times I = CE$ e $ABCDE \times CE = ABC^2DE^2 = ABIDI = ABD$

Como foi dito anteriormente, o exemplo apresentado é do tipo 2^{k-1} , que apresenta apenas uma identidade geradora do projeto. Se a fração fosse $\frac{1}{4}$, ter-se-iam duas identidades, porque seriam necessárias definir duas variáveis e o projeto seria uma fração do tipo,

$$\frac{2^k}{2^2} = 2^{k-2} = \frac{1}{4} \times 2^5 \quad (30)$$

Genericamente falando, pode-se dizer que, $\frac{2^k}{2^p} = 2^{k-p} = \frac{1}{2^p} \times 2^k$, sendo k o número de variáveis independentes sob estudo e p o número de identidades geradoras que podem ser usadas na relação completa de definição do projeto.

6.1.5 A relação completa de definição de um projeto

Se em uma dada investigação desejam-se estudar 5 variáveis em 8 ensaios, ter-se-ia um projeto 2_{III}^{5-2} . Portanto, duas colunas de sinais teriam que ser construídas a partir do projeto básico, 2^3 . Por exemplo, a coluna 4 ($D=AB$) e a coluna 5 ($E=ABC$); No caso, ter-se-ia imediatamente duas identidades que seriam $I=ABD$ e $I=ABCE$, estas são as duas identidades geradoras do projeto. A relação completa de definição é obtida pelo produto destas gerando uma terceira, isto é: $I \times I = ABD \times ABCE = CDE$. Portanto, a relação completa de definição de um dado projeto fracionário seria construída a partir das identidades geradores, multiplicando-as, duas a duas, três a três, quatro a quatro e assim por diante, dependendo do número de colunas criadas. Para o projeto 2_{III}^{5-2} , a relação completa viria dos dois geradores mais a identidade gerada de uma única mutiplicação (só são 2 geradores), isto é: $I = ABD = CDE = ABCE$. A partir desta relação, pode-se chegar a duas importantes constatações:

1. a resolução do projeto, dada pela menor *palavra*, no caso com 3 caracteres (resolução III); e,
2. construir o padrão de confundimentos para os 8 contrastes.

Como exemplo, a construção de dois contrastes a partir da relação completa de definição.

$$I = ABD = CDE = ABCE$$

Para se chegar às colunas de sinais iguais, para a variável C, multiplica-se a relação completa por C, então:

$$I \times C = ABD \times C = CDE \times C = ABCE \times C \quad \text{então: } C = ABDC = DE = ABE$$

Da mesma forma para se chegar às colunas de sinais iguais para a interação AC, multiplica-se a relação completa por AC:

$$I \times AC = ABD \times AC = CDE \times AC = ABCE \times AC, \quad \text{logo: } AC = BCD = ADE = BE$$

Os 8 contrastes estimados desta maneira estão na Tabela 37

Tabela 37 - Relação entre colunas e padrão de confundimentos para a fração 2_{III}^{5-2}

Relação entre as colunas	Padrão de confundimentos “(confounding pattern)”
A = BD = BCE = ACDE	$\ell_A \rightarrow A + BD + BCE + ACDE$
B = AD = ACE = BCDE	$\ell_B \rightarrow B + AD + ACE + BCDE$
C = DE = ABE = ABCD	$\ell_C \rightarrow C + DE + ABE + ABCD$
D = AB = CE = ABCDE	$\ell_D \rightarrow D + AB + CE + ABCDE$
E = CD = ABC = ABDE	$\ell_E \rightarrow E + CD + ABC + ABDE$
BC = AE = ACD = BDE	$\ell_{BC} \rightarrow BC + AE + ACD + BDE$
AC = BE = ADE = BCD	$\ell_{AC} \rightarrow AC + BE + ADE + BCD$
(I=ABCDE)	

A matriz de sinais para este projeto seria a apresentada na Tabela 38.

Tabela 38 - Matriz de sinais para o projeto fracionado 2_{III}^{5-2}

#	Projeto				
	A	B	C	D=AB	E=ABC
1	-	-	-	+	-
2	+	-	-	-	+
3	-	+	-	-	+
4	+	+	-	+	-
5	-	-	+	+	+
6	+	-	+	-	-
7	-	+	+	-	-
8	+	+	+	+	+

6.1.6 A estrutura de um projeto fracionado saturado

Um projeto fracionado é saturado quando todas as observações são usadas para estimar os efeitos principais e a média. Por exemplo, o projeto $2^{7-4}=8$ ensaios, é um projeto saturado, porque, as oito observações são usadas na estimativa da média e dos 7 efeitos principais. Este é um dos projetos fracionados mais usado na seleção de variáveis, porque ele é capaz de avaliar simultaneamente 7 variáveis.

Pode-se observar que o projeto básico, para a construção desta fração, é o 2^3 . Portanto, é necessário definir mais 4 variáveis e obter 4 identidades geradoras da relação completa de definição. É claro que $\frac{1}{32} \times 128$ gera um projeto com resolução baixa, porque se tem que encontrar modelos onde os efeitos principais serão confundidos com interações de dois fatores, portanto um projeto com resolução III, ou 2_{III}^{7-4} , usando a notação apresentada.

Quando a fração é $\frac{1}{2^p}$, tem-se p geradores e 2^p relações de definição para o projeto.

Considerando um projeto 2^{7-4} , obtido com os seguintes geradores: $D=AB$, $E=AC$, $F=BC$ e $G=ABC$, e com as respectivas identidades

$$I=ABD, \quad I=ACE, \quad I=BCF \quad \text{e} \quad I=ABCG$$

Combinações tais como ABD e ACE podem ser referidas como “*palavras*”. A relação de definição inclui todas as *palavras* que são iguais a identidade I que originou as quatro colunas de sinais e as demais identidades construídas por multiplicações dos geradores juntos, isto é: para o exemplo, multiplicando-as dois a dois, três a três e as quatro relações, tem-se a relação completa de definição que geram todos os contrastes.

Assim, para o planejamento 2^{7-4}_{III} , a relação completa de definição é dada por:

$$I = ABD = ACE = BCF = ABCG = BCDE = ACDF = CDG = ABEF = BEG = AFG = DEF = ADEG = BDFG = CEFG = ABCDEFG$$

Tabela 39 - A matriz de sinais do projeto fracionado 2^{7-4}_{III} , cujas identidades são: $I=ABD$, $I=ACE$, $I=BCF$. e $I=ABCG$

	A	B	C	D= AB	E= AC	F= BC	G= ABC	Resposta
1	-	-	-	+	+	+	-	y_1
2	+	-	-	-	-	+	+	y_2
3	-	+	-	-	+	-	+	y_3
4	+	+	-	+	-	-	-	y_4
5	-	-	+	+	-	-	+	y_5
6	+	-	+	-	+	-	-	y_6
7	-	+	+	-	-	+	-	y_7
8	+	+	+	+	+	+	+	y_8

6.2 A estimativa dos efeitos

Os efeitos de um projeto fatorial fracionado são estimados com o mesmo procedimento aplicado aos fatoriais completos.

Antes de escolher qual a fração e a resolução do projeto que será realizado é de importância fundamental escolher os níveis da variável e escrever a matriz com as variáveis independentes em unidades originais. Um outro cuidado importante é realizar os ensaios em seqüência aleatória.

6.2.1 A interpretação dos efeitos

Também é feita seguindo a mesma regra dos fatoriais completos apresentada no item, 4.1.4.2.

Como, a maioria dos planejamentos fracionados são realizados em uma única replicata é comum usar a escala de probabilidade normal, complementado pelo PSE de Lenth, como base para o julgamento dos efeitos.

6.3 O efeito de blocos nos DOE 2^k e DOE 2^{k-p}

Uma fato corriqueiro na pesquisa é quando são usados produtos cujas características podem variar de lote para lote, como foi mostrado no caso do exemplo da penicilina no capítulo 2. Ao se iniciar os ensaios de um planejamento fatorial 2^3 , observa-se que a matéria prima principal não é do mesmo lote; assim, tem-se quantidade de cada lote para realizar apenas 4 experimentos. Neste caso, a matéria prima entra como um fator perturbador do projeto que pode alterar, sensivelmente, o resultado da análise. Então, uma alternativa interessante com o

objetivo de eliminar este efeito é considerar o efeito do fator matéria prima confundido com a interação ABC. Assim, ao gerar a matriz de ensaios do projeto 2^3 bem como as colunas de sinais referentes às interações, tomar-se-ia como bloco I os ensaios onde aparecem o sinal menos na interação ABC e o bloco II onde aparecem o sinal mais.

A Tabela 40 ilustra esta geração.

Tabela 40 – Exemplo de geração de blocos para DOE 2^k e DOE 2^{k-p} , bloco=ABC

#	A	B	C	AB	AC	BC	ABC	Bloco	A	B	C	#	
1	-	-	-	+	+	+	-	I	Bloco I	-	-	-	1
2	+	-	-	-	-	+	+	II		+	+	-	4
3	-	+	-	-	+	-	+	II		+	-	+	6
4	+	+	-	+	-	-	-	I		-	+	+	7
5	-	-	+	+	-	-	+	II	Bloco II	+	-	-	2
6	+	-	+	-	+	-	-	I		-	+	-	3
7	-	+	+	-	-	+	-	I		-	-	+	5
8	+	+	+	+	+	+	+	II		+	+	+	8

Os blocos podem ser representados geometricamente como na Figura 39

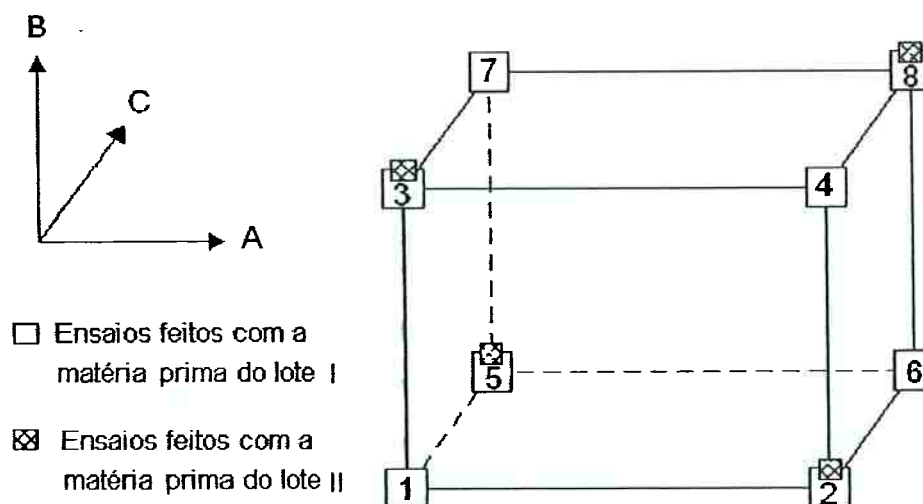


Figura 39. Arranjo de um projeto fatorial 2^3 em dois blocos de tamanho 4. Fonte: (BOX, HUNTER W., HUNTER, J. 2005)

A grande idéia deste arranjo em bloco é que a variável perturbadora não é confundida com os efeitos principais nem com interações de dois fatores, apenas com interações de três fatores. O mesmo princípio de construção de blocos é usado para o projeto fatorial fracionado. Lembrando sempre que jamais se pode confundir variável de bloco com efeitos principais. Um maior aprofundamento e uma tabela, com os principais arranjos de blocos para projetos fatoriais completos e fracionados, está em Box, Hunter W., Hunter, J. (2005).

6.4 Como usar a fração mais adequada à necessidade

Quando o pesquisador, técnico ou mesmo uma equipe está se preparando para uma nova pesquisa ou para realizar a melhoria de um processo já existente nenhuma variável ou interação entre variáveis suspeitas devem ser descartadas. Este elenco deve ser apreciado, discutido e dependendo do recurso disponível, deve-se construir o projeto mais adequado a investigação em questão.

Com o conhecimento básico de construção, análise e interpretação dos resultados dos projetos fracionados, apresentados até aqui, se pode obter informações que indiquem tendências claras de como prosseguir o estudo, com base nas questões iniciais propostas e na concepção do problema.

Muitas vezes escolhe-se uma fração com maior número de ensaios por insegurança ou medo de perder informações importantes, optando-se por uma fração com maior número de ensaios e com maior resolução. Isto pode onerar a experimentação e ampliar o tempo de execução do projeto. Às vezes é melhor arriscar-se trabalhando inicialmente com um projeto mais curto, e resolver a ambigüidade, se ela aparecer, com ensaios adicionais em um passo seguinte, fazendo uso da técnica mais conveniente à solução desejada.

A estratégia mais comum para solução de ambigüidades para frações com poucos experimentos é a inversão de sinais (*“foldover”*) de uma das colunas ou de múltiplas colunas de sinais da matriz original. Quando é invertida apenas uma coluna de sinais no projeto subsequente, o projeto resultante da adição dos dois conjuntos de dados, separa os efeitos desejados, mas o projeto resultante se mantém com a resolução da fração original. Se a necessidade for realizar mais uma fração de tal modo que o projeto fracionário resultante das duas frações adicionadas melhore a resolução, o *“foldover”* deve ser total, isto é, a próxima fração deve ser o espelho da original. Logo, as possibilidades, de se construir um projeto fracionado econômico e inteligente, dependem da habilidade do experimentador, que usa os métodos estatísticos. Os programas computacionais estatísticos geram as frações mais diretas que aparecem nas tabelas apresentadas na literatura. Mas, dependendo do problema, pode-se construir fração que melhor se ajusta à necessidade real, atendendo a critérios técnicos e econômicos.

Em resumo, os projetos fracionários podem ser úteis para:

1. construir o projeto segundo restrições impostas quer pelo fenômeno quer por suspeita do projeto gerar alguma ambigüidade;
2. eleger a fração na medida exata do recurso para o desenvolvimento, tendo em vista que poderá haver necessidade de realizar uma seqüência de projetos até atingir o resultado esperado.
3. construção de blocos, definindo a variável ou as variáveis de bloco em conjunto com as outras variáveis de processo.

6.4.1 Um exemplo de projeto que atende a uma necessidade específica

Um engenheiro químico está interessado no desenvolvimento experimental de um reator. As variáveis de interesse são: temperatura, concentração de um dado reagente, pH do meio, taxa de agitação e dois tipos de catalisador (*a* e *b*). Ele está particularmente interessado em duas interações de dois fatores que são: temperatura×concentração e catalisador×temperatura. O recurso disponível só permite a realização de 8 ensaios. É possível ajudá-lo construindo um projeto fracionado no qual estas duas interações não sejam confundidas com efeitos principais nem uma com a outra?

Observando a situação relatada pelo engenheiro, e olhando as frações armazenadas nos programas computacionais, nota-se que ela não se encaixa em nenhuma das frações propostas. Então, o que fazer? Será que uma fração tão específica existe? Esta situação será cuidadosamente analisada a seguir.

Variáveis envolvidas: A – temperatura, B – concentração, C – pH, D – agitação, E – catalisador. Das cinco variáveis 4 são contínuas (A, B, C e D) e uma é descontínua

(catalisador, a e b). Inicialmente devem-se fixar os níveis altos e baixos para as variáveis, a seguir escrever as restrições:

Lembrar que a fração procurada, com 8 ensaios é $\frac{1}{4}$ do projeto completo $2^5=32$, e deve obedecer as seguintes restrições: as interações AB e AE não podem ser confundidas entre si e nenhuma das duas com efeitos principais.

O projeto completo seria capaz de estimar a média, 5 efeitos principais, 10 interações de dois fatores, 10 interações de 3 fatores, 5 interações de quatro fatores e uma interação de 5 fatores. Portanto, 32 quantidades poderão ser estimadas com o projeto 2^5 .

O projeto objeto do engenheiro só pode estimar 8 quantidades. Então voltando aos princípios de construção de qualquer fração, serão lembradas algumas questões.

No caso, $p=2$, logo deverão ser escritas duas identidades geradoras, em um projeto com resolução III, porque das 10 interações de dois fatores possíveis com cinco variáveis, duas específicas, AB e AE, não podem ser confundidas entre si nem com os efeitos principais. Portanto, ainda sobram 8 interações de 2 fatores que deverão ser confundidas entre si e com efeitos principais. Por esta razão este projeto fracionário terá resolução III (2_{III}^{5-2}). Portanto, das 8 quantidades estimadas, uma será a média e as 7 restantes terão os padrões onde serão confundidos 5 efeitos principais com interações de dois, 3 e quatro fatores e as que serão confundidas e os dois contrastes restantes deverão ser reservados para as interações de dois fatores alvo do engenheiro, que também deverão ser confundidas com outras de dois fatores e de três fatores.

A construção do projeto fracionário usará como projeto básico o completo 2^3 , ou seja, as duas identidades geradoras nascerão das variáveis A, B e C. Como AB não pode ser confundida com efeitos principais, esta será imediatamente descartada. Então sobram, AC, BC e ABC.

Um projeto fracionário será construído com os seguintes geradores

$D=AC$ e $E=BC$.

A relação completa de definição para gerar os 8 contrastes será $I=ACD=BCE=ABDE$.

A matriz de sinais, a relação entre as colunas e os contrastes está apresentada na Tabela 41.

Tabela 41 - Projeto fracionado específico que não confunde efeitos principais com as interações AB e AE. Estas últimas não se confundem entre si

#	Matriz de sinais do projeto					Interações de dois fatores de interesse	
	A	B	C	D=AC	E=BC	AB	AE
17	-	-	-	+	+	+	-
2	+	-	-	-	+	-	+
3	-	+	-	+	-	-	+
20	+	+	-	-	-	+	-
5	-	-	+	-	-	+	+
22	+	-	+	+	-	-	-
23	-	+	+	-	+	-	-
8	+	+	+	+	+	+	+

Relação entre os pares de colunas	Padrão de confundimentos ("confounding pattern")
$A = CD = BDE = ABCE$	$l_A \rightarrow A + CD + BDE + ABCE$
$B = CE = ADE = ABCD$	$l_B \rightarrow B + CE + ADE + ABCD$
$C = AD = BE = ABCDE$	$l_C \rightarrow C + AD + BE + ABCDE$
$D = AC = ABE = BCDE$	$l_D \rightarrow D + AC + ABE + BCDE$
$E = BC = ABD = ACDE$	$l_E \rightarrow E + BC + ABD + ACDE$
$AB = DE = ACE = BCD$	$l_{AB} \rightarrow AB + DE + ACE + BCD$
$AE = BD = ABC = CDE$	$l_{AE} \rightarrow AE + BD + ABC + CDE$

$(I = ACD = BCE = ABDE)$

Observa-se pela tabela de contrastes que o engenheiro pode buscar as informações requeridas, com o projeto da Tabela 41, porque ele cumpre as exigências necessárias.

E agora vem uma outra pergunta. Seria possível construir um outro projeto fracionário que atendesse as mesmas necessidades com outras relações geradoras?

Um segundo projeto poderá ser criado definido $D = AC$ e $E = ABC$. A relação completa de definição seria dada por: $(I = ACD = BDE = ABCE)$.

A matriz de sinais e os padrões de confundimento deste segundo projeto estão na Tabela 42.

Tabela 42 - Projeto fracionado específico que não confunde efeitos principais com as interações AB e AE. Elas também não se confundem entre si.

#	Matriz de sinais do projeto					Interações de dois fatores de interesse	
	A	B	C	D=AC	E=ABC	AB	AE
17	-	-	-	+	-	+	+
2	+	-	-	-	+	-	+
3	-	+	-	+	+		-
20	+	+	-	-	-	+	-
5	-	-	+	-	+	+	-
22	+	-	+	+	-	-	-
23	-	+	+	-	-	-	+
8	+	+	+	+	+	+	+

Relação entre os pares de colunas	Padrão de confundimentos ("confounding pattern")
$A = CD = BCE = ABDE$	$l_A \rightarrow A + CD + BCE + ABDE$
$B = DE = ACE = ABCD$	$l_B \rightarrow B + DE + ACE + ABCD$
$C = AD = ABE = BCDE$	$l_C \rightarrow AD + ABE + BCDE$
$D = AC = BE = ABCDE$	$l_D \rightarrow D + AC + BE + ABCDE$
$E = BD = ABC = ACDE$	$l_E \rightarrow E + BD + ABC + ACDE$
$AB = CE = ADE = BCD$	$l_{AB} \rightarrow AB + CE + ADE + BCD$
$AE = BC = ABD = CDE$	$l_{AE} \rightarrow AE + BC + ABD + CDE$

$(I = ACD = BDE = ABCE)$

Então, vê-se que é possível resolver situações bem particulares usando o projeto que mais se ajustam a elas.

6.4.2 Uma solução como uma alternativa mais econômica de um projeto fatorial misto

Muitas vezes o pesquisador buscando maior segurança projeta um experimento com um número maior de ensaios pensando em não perder informações preciosas. Mesmo assim, após a realização de uma série extensa de ensaios, pode-se chegar a conclusão de que sejam necessários ensaios adicionais. Mais uma vez convém lembrar que os métodos seqüenciais são úteis para situações onde se quer investigar o domínio experimental desejado.

A seguir será feita uma comparação entre um projeto realizado por Montes e colaboradores (2005), para ilustrar que muitas vezes um projeto mais amplo, com mais do que 2 níveis, não melhora as informações de um completo em dois níveis e até mesmo de um fracionário bem ajustado.

Os autores tinham como objetivo otimizar as condições de extração para máximo rendimento de antocianinas da pele de Jaboticaba (*Myrcia Jaboticaba Berg*). Os fatores estudados foram A = solvente codificado (etanol e metanol); B = ácido codificado (HCl, ac.citrico e ac. acético); C = % solvente codificado (mistura solvente/água, 100% v/v, 90% v/v e 80% v/v); D = pH codificado (2, 3, e, 4) e as respostas foram: o rendimento de antocianina g/kg; e alguns atributos de cor L^* (Lightness), C_{ab}^* (Chroma) e h_{ab} (ângulo hue).

A discussão girará em torno apenas das informações estatísticas que podem ser estimadas a partir dos dados originais, usando dois projetos alternativos, a saber: um projeto fatorial completo em dois níveis, com 16 ensaios, DOE 2^4 , e um projeto fracionado com 8 ensaios DOE 2^{4-1}_{IV} .

A Tabela 43 Apresenta a transcrição dos dados do artigo original, que contém ao mesmo tempo os níveis de variação dos fatores e os resultados para as 4 respostas. Em toda a análise desse artigo serão usadas tabelas adicionais com base na original, para facilitar a comparação das análises.

Tabela 43 - Tabela 1 do artigo original, apresentado níveis e códigos dos fatores e as 4 respostas para extração de antocianina da pele de jaboticaba. Fonte: (Montes et al. 2005)

#	Solvent	Acid	Solvent (% v/v)	pH	Anthocyanin (g/kg)	L*	C _{ab} [*] C	h _{ab}
1	Ethanol (-1)	HCl (-1)	100 (-1)	2 (-1)	0.14003	71.418	54.341	22.707
2	Ethanol (-1)	HCl (-1)	100 (-1)	3 (0)	0.13301	74.778	47.899	22.465
3	Ethanol (-1)	HCl (-1)	100 (-1)	4 (+1)	0.15052	75.772	45.578	22.619
4	Ethanol (-1)	HCl (-1)	90(0)	2 (-1)	0.14133	74.611	50.586	19.871
5	Ethanol (-1)	HCl (-1)	90 (0)	3 (0)	0.13205	76.649	45.297	19.521
6	Ethanol (-1)	HCl (-1)	90 (0)	4 (+1)	0.12056	75.770	46.906	20.538
7	Ethanol (-1)	HCl (-1)	80 (+1)	2 (-1)	0.13276	75.937	47.642	17.988
8	Ethanol (-1)	HCl (-1)	80 (+1)	3 (0)	0.11550	76.517	45.193	18.143
9	Ethanol (-1)	HCl (-1)	80 (+1)	4 (+1)	0.14103	76.680	44.941	18.900
10	Methanol (+1)	HCl (-1)	100 (-1)	2 (-1)	0.13531	78.799	41.380	15.654
11	Methanol (+1)	HCl (-1)	100 (-1)	3 (0)	0.16329	75.244	47.914	18.220
12	Methanol (+1)	HCl (-1)	100 (-1)	4 (+1)	0.14840	75.146	46.174	18.191
13	Methanol (+1)	HCl (-1)	90 (0)	2 (-1)	0.15357	75.781	47.528	17.780
14	Methanol (+1)	HCl (-1)	90 (0)	3 (0)	0.15369	76.283	45.973	17.058
15	Methanol (+1)	HCl (-1)	90 (0)	4 (+1)	0.15027	77.591	43.739	16.319
16	Methanol (+1)	HCl (-1)	80 (+1)	2 (-1)	0.11191	78.468	42.886	16.453
17	Methanol (+1)	HCl (-1)	80 (+1)	3 (0)	0.13436	77.188	44.657	16.933
18	Methanol (+1)	HCl (-1)	80 (+1)	4 (+1)	0.11946	80.853	38.254	15.197
19	Ethanol (-1)	Citric (0)	100 (-1)	2 (-1)	-	-	-	-
20	Ethanol (-1)	Citric (0)	100 (-1)	3 (0)	0.07485	80.006	40.133	20.363
21	Ethanol (-1)	Citric (0)	100 (-1)	4 (+1)	0.07960	83.801	30.984	22.027
22	Ethanol (-1)	Citric (0)	90 (0)	2 (-1)	-	-	-	-
23	Ethanol (-1)	Citric (0)	90 (0)	3 (0)	0.07411	80.821	39.346	17.228
24	Ethanol (-1)	Citric (0)	90 (0)	4 (+1)	0.07302	83.881	30.909	19.321
25	Ethanol (-1)	Citric (0)	80 (+1)	2 (-1)	-	-	-	-
26	Ethanol (-1)	Citric (0)	80 (+1)	3 (0)	0.08020	84.260	34.121	15.431
27	Ethanol (-1)	Citric (0)	80 (+1)	4 (+1)	0.08253	84.309	34.139	15.489
28	Methanol (+1)	Citric (0)	100 (-1)	2 (-1)	-	-	-	-
29	Methanol (+1)	Citric (0)	100 (-1)	3 (0)	0.08430	80.666	39.160	17.003
30	Methanol (+1)	Citric (0)	100 (-1)	4 (+1)	0.08706	83.436	2.852	14.660
31	Methanol (+1)	Citric (0)	90 (0)	2 (-1)	-	-	-	-
32	Methanol (+1)	Citric (0)	90 (0)	3 (0)	0.08736	82.059	35.458	15.569
33	Methanol (+1)	Citric (0)	90 (0)	4 (+1)	0.08662	83.587	32.498	14.534
34	Methanol (+1)	Citric (0)	80 (+1)	2 (-1)	-	-	-	-
35	Methanol (+1)	Citric (0)	80 (+1)	3 (0)	0.07390	83.374	32.958	14.145
36	Methanol (+1)	Citric (0)	80 (+1)	4 (+1)	0.08021	83.313	31.212	14.024
37	Ethanol (-1)	Acetic (+1)	100 (-1)	2 (-1)	0.06805	77.3	41.4	24.6
38	Ethanol (-1)	Acetic (+1)	100 (-1)	3 (0)	0.06443	83.750	32.251	20.730
39	Ethanol (-1)	Acetic (+1)	100 (-1)	4 (+1)	0.05484	85.921	26.784	23.954
40	Ethanol (-1)	Acetic (+1)	90 (0)	2 (-1)	0.04632	81.090	38.998	17.543
41	Ethanol (-1)	Acetic (+1)	90 (0)	3 (0)	0.06070	84.684	29.731	18.949
42	Ethanol (-1)	Acetic (+1)	90 (0)	4 (+1)	0.06005	84.499	30.066	22.442
43	Ethanol (-1)	Acetic (+1)	80 (+1)	2 (-1)	0.05115	81.705	37.919	16.118
44	Ethanol (-1)	Acetic (+1)	80 (+1)	3 (0)	0.06138	81.757	36.424	18.859
45	Ethanol (-1)	Acetic (+1)	80 (+1)	4 (+1)	0.05425	85.093	29.023	18.003
46	Methanol (+1)	Acetic (+1)	100 (-1)	2 (-1)	0.04399	83.606	33.529	15.602
47	Methanol (+1)	Acetic (+1)	100 (-1)	3 (0)	0.06779	80.840	37.747	17.356
48	Methanol (+1)	Acetic (+1)	100 (-1)	4 (+1)	0.04922	86.412	26.121	16.070
49	Methanol (+1)	Acetic (+1)	90 (0)	2 (-1)	0.05343	89.025	22.787	12.970
50	Methanol (+1)	Acetic (+1)	90 (0)	3 (0)	0.05053	87.112	25.929	14.126
51	Methanol (+1)	Acetic (+1)	90 (0)	4 (+1)	0.06061	87.939	23.752	14.084
52	Methanol (+1)	Acetic (+1)	80 (+1)	2 (-1)	0.04536	83.465	33.975	14.747
53	Methanol (+1)	Acetic (+1)	80 (+1)	3 (0)	0.06428	84.320	31.837	14.305
54	Methanol (+1)	Acetic (+1)	80 (+1)	4 (+1)	0.06314	85.881	28.616	14.065

Inicialmente, será feita uma discussão sobre os níveis apresentados para cada variável. Usar mais do que 2 níveis no primeiro projeto experimental é uma atitude contrária ao método seqüencial, onde as informações vão surgindo gradualmente, com projetos bem justos às questões propostas na concepção do problema. Muitas vezes se conhece pouco sobre a influência das variáveis independentes no espaço fator delimitado no projeto. Então, objetivamente, é interessante investir o menor esforço possível na fase inicial da pesquisa.

No caso do artigo aqui mencionado, os autores propuseram um projeto fatorial misto com uma replicata constituída por 54 ensaios.

A proposta da autora deste texto, seria iniciar com apenas dois níveis dos fatores e com uma série experimental de 8 ensaios, isto é, um planejamento fatorial fracionado. Se o resultado da análise indicasse alguma ambigüidade, a fração complementar seria realizada e a análise total poderia ser feita com 16 ensaios. Nesta proposta não seriam estudados os níveis do meio das variáveis, ácido, % de solvente e pH. Apenas os níveis mais altos e mais baixos da Tabela original.

A partir da Tabela 43 pode-se construir o projeto 2^4 (completo), que esta na

Tabela 44 e o fracionado 2^{4-1} na Tabela 45. Nas duas últimas tabelas, os números referentes ao ensaio são os mesmos que aparecem na tabela original.

Os dados dos dois projetos adicionais foram analisados e a estimativa dos efeitos dos dois projetos em dois níveis e do projeto misto como apresentado pelos autores, estão na Tabela 46.

Tabela 44 - Matriz de ensaios com variáveis originais e codificadas e resultados de 4 respostas do projeto 2^4 , retirados dos dados obtidos por Montes e colaboradores (2005) para extração de antocianina da pele de jaboticaba

#	Solvente	Ácido	Solvente (% v/v)	pH	Antocianina rend.(g/kg)	L*	C_{ab}^*	h_{ab}
1	Ethanol (-1)	HCl (-1)	100 (-1)	2 (-1)	0,14003	71,418	54,341	22,707
10	Methanol (+1)	HCl (-1)	100 (-1)	2 (-1)	0,13531	78,799	41,380	15,654
37	Ethanol (-1)	Acetic (+1)	100 (-1)	2 (-1)	0,06805	77,300	41,400	24,600
46	Methanol (+1)	Acetic (+1)	100 (-1)	2 (-1)	0,04399	83,606	33,529	15,602
7	Ethanol (-1)	HCl (-1)	80 (+1)	2 (-1)	0,13276	75,937	47,642	17,988
16	Methanol (+1)	HCl (-1)	80 (+1)	2 (-1)	0,11191	78,468	42,886	16,453
43	Ethanol (-1)	Acetic (+1)	80 (+1)	2 (-1)	0,05115	81,705	37,919	16,118
52	Methanol (+1)	Acetic (+1)	80 (+1)	2 (-1)	0,04536	83,465	33,975	14,747
3	Ethanol (-1)	HCl (-1)	100 (-1)	4 (+1)	0,15052	75,772	45,578	22,619
12	Methanol (+1)	HCl (-1)	100 (-1)	4 (+1)	0,14840	75,146	46,174	18,191
39	Ethanol (-1)	Acetic (+1)	100 (-1)	4 (+1)	0,05484	85,921	26,784	23,954
48	Methanol (+1)	Acetic (+1)	100 (-1)	4 (+1)	0,04922	86,412	26,121	16,070
9	Ethanol (-1)	HCl (-1)	80 (+1)	4 (+1)	0,14103	76,680	44,941	18,900
18	Methanol (+1)	HCl (-1)	80 (+1)	4 (+1)	0,11946	80,853	38,254	15,197
45	Ethanol (-1)	Acetic (+1)	80 (+1)	4 (+1)	0,05425	85,093	29,023	18,003
54	Methanol (+1)	Acetic (+1)	80 (+1)	4 (+1)	0,06314	85,881	28,616	14,065

Tabela 45 - Matriz de ensaios com variáveis originais e codificadas e resultados de 4 respostas do projeto fracionado 2^{4-1} , retirados dos dados obtidos por Montes e colaboradores (2005) para extração de antocianina da pele de jaboticaba

Run	Solvent	Acid	Solvent (% v/v)	pH	Anthocyanin yield (g/kg)	L*	C_{ab}^*	h_{ab}
1	Ethanol (-1)	HCl (-1)	100 (-1)	2 (-1)	0,14003	71,418	54,341	22,707
12	Methanol (+1)	HCl (-1)	100 (-1)	4 (+1)	0,14840	75,146	46,174	18,191
39	Ethanol (-1)	Acetic (+1)	100 (-1)	4 (+1)	0,05484	85,921	26,784	23,954
46	Methanol (+1)	Acetic (+1)	100 (-1)	2 (-1)	0,04399	83,606	33,529	15,602
9	Ethanol (-1)	HCl (-1)	80 (+1)	4 (+1)	0,14103	76,680	44,941	18,900
16	Methanol (+1)	HCl (-1)	80 (+1)	2 (-1)	0,11191	78,468	42,886	16,453
43	Ethanol (-1)	Acetic (+1)	80 (+1)	2 (-1)	0,05115	81,705	37,919	16,118
54	Methanol (+1)	Acetic (+1)	80 (+1)	4 (+1)	0,06314	85,881	28,616	14,065

Tabela 46 - Efeitos estimados para os dados de Montes e colaboradores (2005); para os projetos 2^4 , a fração 2^{4-1} e o projeto misto $2^1 \times 3^2$

Estimativa dos efeitos com o projeto completo 2^4						Estimativa dos efeitos com a fração 2^{4-1} , com D=ABC					
Termo ⁺	Efeito Antocianina	Efeito L^*	Efeito C_{ab}^*	Efeito h_{ab}		Termo	Efeito Antocianina	Efeito L^*	Efeito C_{ab}^*	Efeito h_{ab}	
Constante	0,09434	80,154	38,660	18,179		Constante	0,09431	79,853	39,399	18,249	
A	-0,00948	2,851	-4,587	-4,864		A + BCD	-0,00490	1,844	-3,195	-4,342	
B	-0,08118	7,039	-12,979	-0,569		B + ACD	-0,08206	8,850	-15,373	-1,628	
C	-0,00891	1,714	-1,506	-3,491		C + ABD	-0,00501	1,661	-1,617	-3,730	
D	0,00654	2,633	-5,948	0,391		D + ABC	0,01508	2,108	-5,540	1,058	
AB	0,00283	-0,514	1,365	-0,684		AB + CD	0,00547	-0,914	1,916	-0,860	
AC	-0,00035	-0,538	0,638	2,227		AC + BD	-0,00366	1,138	-2,484	2,092	
AD	0,00438	-1,645	2,796	-0,125		AD + BC	0,01274	-2,631	4,728	-0,957	
BC	0,00836	-0,988	1,931	-0,833		Efeitos lineares e quadráticos estimados: projeto fatorial misto					
BD	-0,00331	1,6758	-3,122	-0,135		Solvente (L)	0,00	1,83 ^b	-3,55 ^b	-4,01 ^b	
CD	0,00264	-0,400	0,551	-0,176		Ácido (L)	-0,08 ^b	7,83 ^b	-14,44 ^b	-1,11 ^a	
ABC	0,00854	-0,525	0,408	0,666		Ácido (Q) ^{††}	-0,02 ^b	1,94	-2,38	-1,43 ^a	
ABD	0,00391	-0,053	-0,110	-0,239		%Solvente(L)	-0,00 ^a	1,75 ^a	-2,06	-3,40 ^b	
ACD	-0,00088	1,812	-2,395	-1,059		%Solvente(Q) ^{†††}	0,00	0,64	-1,7	-0,47	
BCD	0,00458	-1,006	1,392	0,522		pH (L)	0,01	2,41 ^a	-6,15 ^b	0,52	
ABCD	-0,00005	-0,601	1,477	0,139		pH (Q) ^{††}	0,00	-0,44	0,78	0,04	

^a $p < 0,05$ e ^b $p < 0,001$ [†]A = solvente codificado; B = ácido codificado; C = %solvente codificado; D = pH codificado.^{†††}Os termos quadráticos só foram estimados por Montes e colaboradores (2005) pelo projeto misto cuja série experimental foi: $2^1 \times 3^2 = 54$ ensaios.

6.4.2.1 A interpretação dos efeitos da Tabela 46

Inicialmente, será apresentada a discussão do projeto fracionário 24-1. A tabela apresenta os contrastes padrões de confundimento e efeitos estimados para este projeto. Como se construiu apenas uma replicada do projeto, os efeitos foram julgados com o auxílio do papel de probabilidade normal e o gráfico de Pareto. A Figura 40 e a

Figura 43 ajudam na visualização, para uma posterior discussão comparativa.

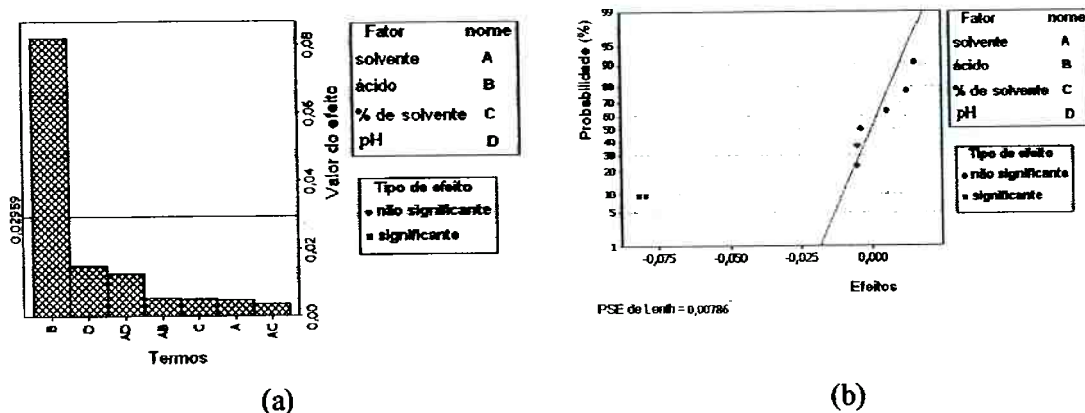


Figura 40. Representação gráfica dos efeitos gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta rendimento de fitocianina do projeto fatorial fracionado 2^{4-1}

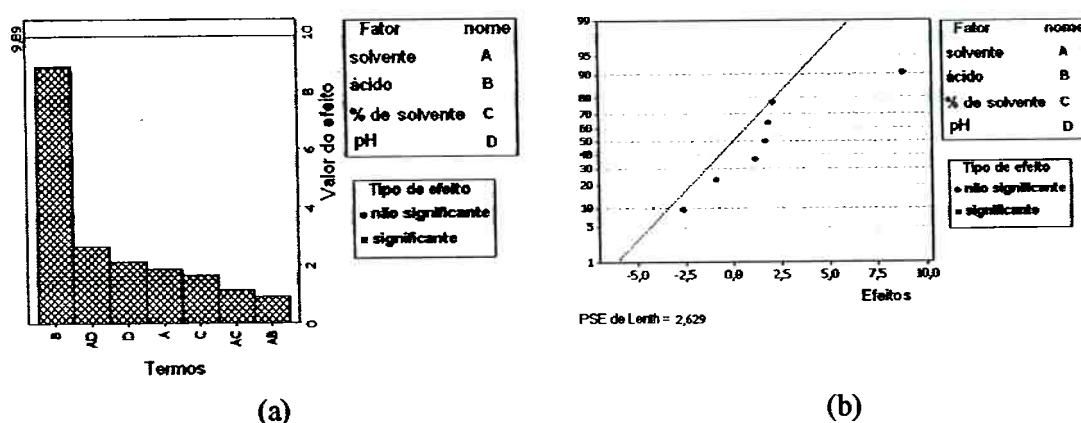


Figura 41. Representação gráfica dos efeitos em gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta L^* do projeto fatorial fracionado 2^{4-1}

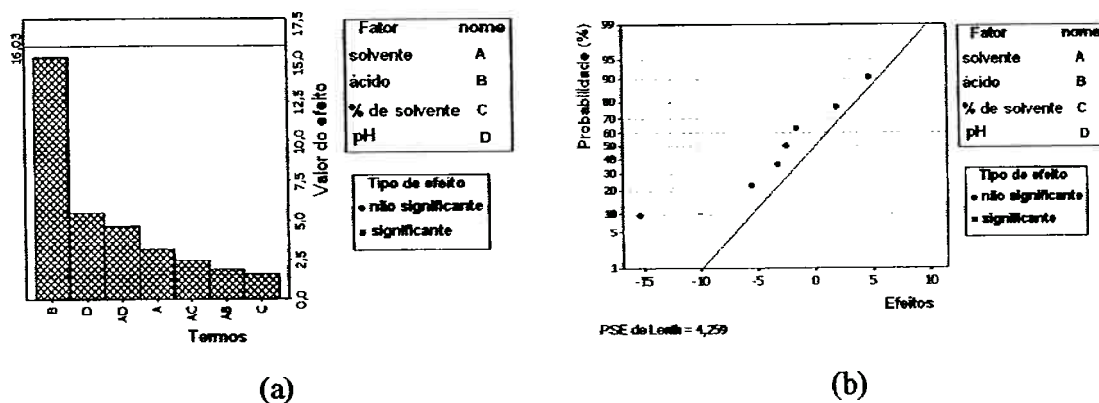


Figura 42. Representação gráfica dos efeitos em gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta do projeto fatorial fracionado 2^{4-1}

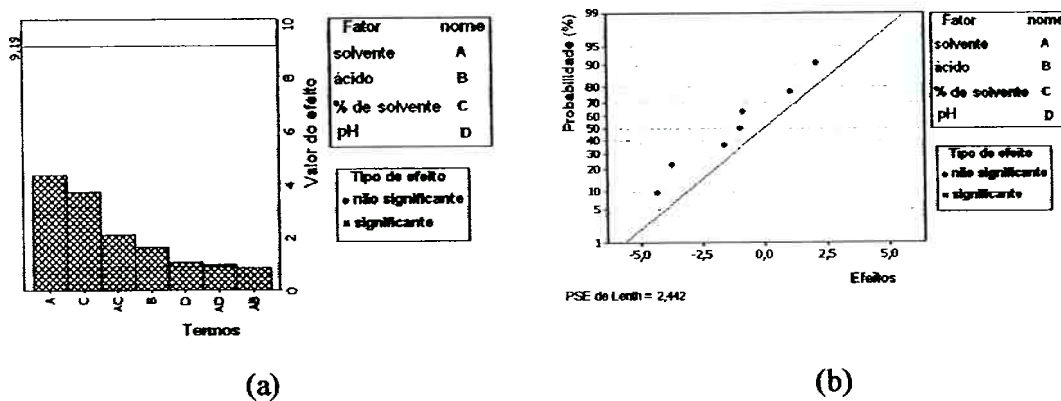


Figura 43. Representação gráfica dos efeitos em gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta hab do projeto fatorial fracionado 2^{4-1}

Os mesmos gráficos de visualização serão agora mostrados para o projeto fatorial completo 2^4 , analisado neste trabalho, na Figura 44 a e na Figura 47.

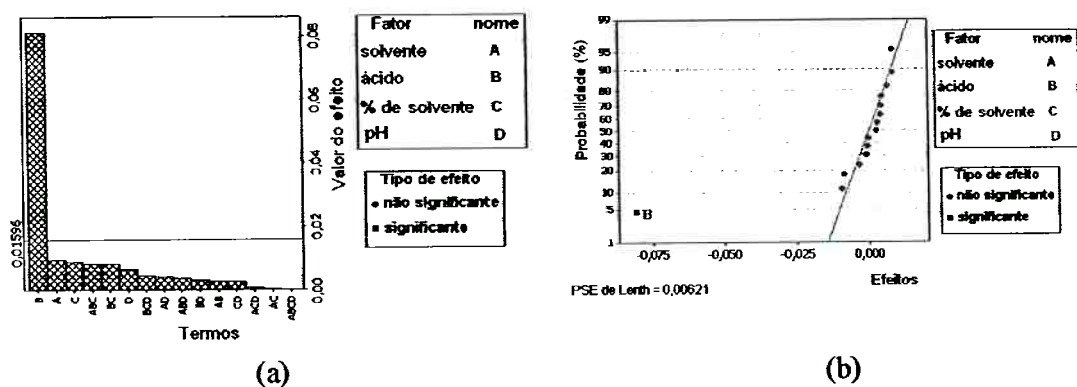


Figura 44. Representação gráfica dos efeitos gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta rendimento de fitocianina do projeto fatorial fracionado 2^4

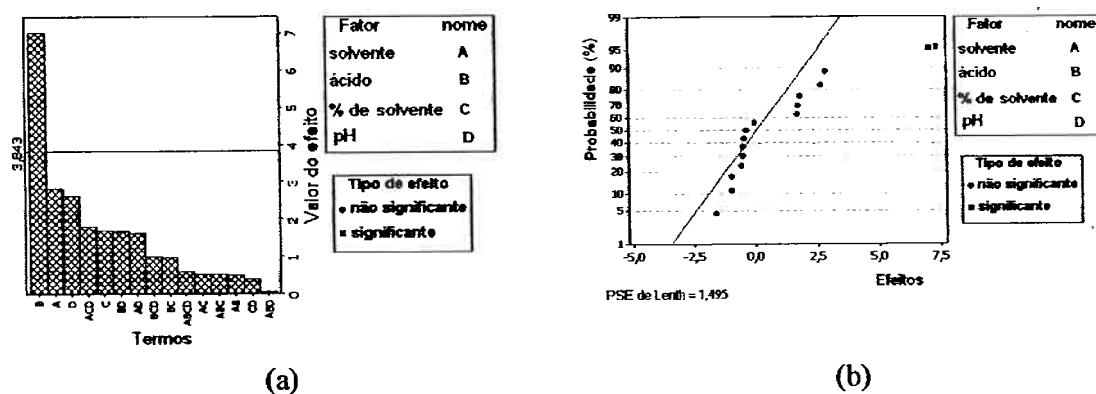


Figura 45. Representação gráfica dos efeitos em gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta L^* do projeto fatorial fracionado 2^4

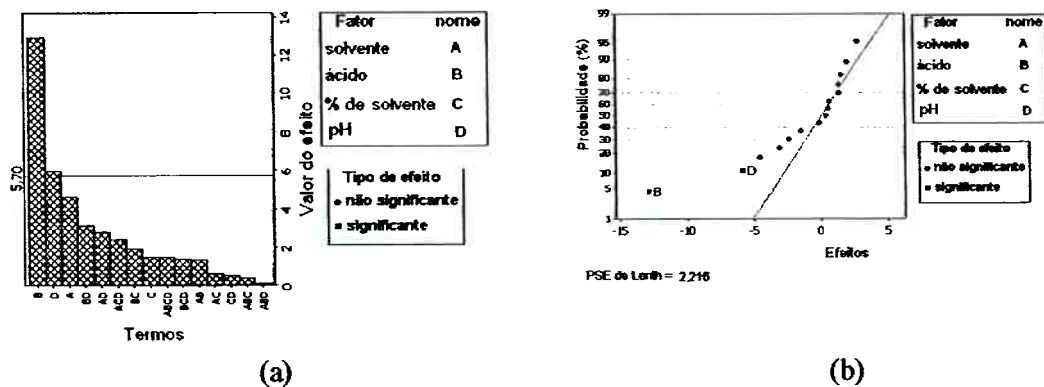


Figura 46. Representação gráfica dos efeitos em gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta do projeto fatorial fracionado 2^4

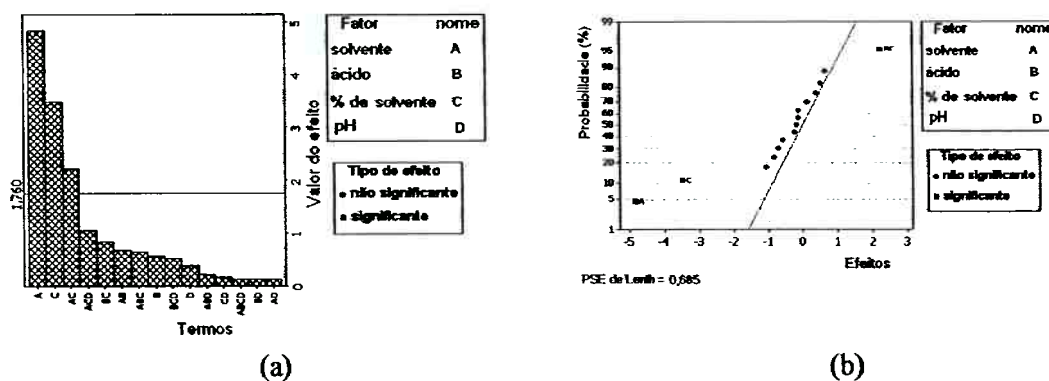


Figura 47. Representação gráfica dos efeitos em gráfico de Pareto (a) e em escala de probabilidade normal (b) para a resposta hab do projeto fatorial fracionado 2^4

Observando os efeitos quadráticos estimados pelo projeto $2^1 \times 3^2$, efetuado por Montes e colaboradores (2005), verifica-se que apenas o efeito quadrado da variável ácido teve significância com $p < 0,05$, na resposta h_{ab} . O valor desse efeito é 1,43 que é menor do 4,01 (efeito do tipo de solvente) e 3,4 (efeito da % de solvente). Já a variável rendimento antocianina (que é a principal, segundo os autores) não mostrou nenhuma significância dos efeitos quadráticos, o que justifica a realização de projetos mais econômicos.

Se compararmos os valores dos efeitos que sobressaem no projeto fracionado em relação aos efeitos obtidos pelos autores, pode-se ver se as tendências já eram claras em relação a esses valores. Os gráficos dos efeitos em escala de probabilidade não acusaram a significância de alguns, graças à incerteza na estimativa do PSE de Lenth, obviamente pelo limitado número de ensaios. Porém, um pesquisador mais envolvido com o estudo, ficaria com os efeitos entre o PSE (pseudo Standard error) e o ME (margin of error). Neste caso, as mesmas decisões tomadas com o projeto misto $2^1 \times 3^2$ poderiam ser tomadas com o fracionado 2^{4-1} , com respeito aos valores dos efeitos.

Para reforçar o que foi dito, será apresentada apenas a comparação dos valores dos efeitos lineares dos dois projetos mencionados, evidenciando a mesma tendência para a representação dos dados.

Tabela 47 - Comparação dos efeitos usando o projeto misto $2^1 \times 3^2$ e o projeto fracionado 2^{4-1}

Efeitos lineares estimados projeto misto, $2^1 \times 3^2$					Efeitos lineares estimados com a fração 2^{4-1}				
Termo	Antocia- nina	L^*	C_{ab}^*	h_{ab}	Termo	Antocia- nina	L^*	C_{ab}^*	h_{ab}
Solvente (L)	0,00	1,83	-3,55	-4,01	A + BCD	-0,00490	1,844	-3,195	-4,34 2
Ácido (L)	-0,08	7,83	-14,44	-1,11	B + ACD	-0,08206	8,850	-15,373	-1,62 8
%Solvente(L)	-0,00	1,75	-2,06	-3,40	C + ABD	-0,00501	1,661	-1,617	-3,73 0
pH (L)	0,01	2,41	-6,15	0,52	D + ABC	0,01508	2,108	-5,540	1,058

A comparação com dos efeitos usando o projeto misto $2^1 \times 3^2$ e o projeto completo 2^4 está na Tabela 48.

A única vantagem do projeto completo 24=16 ensaios é que os gráficos acusam a significância dos efeitos com maior clareza devido ao número de graus de liberdade na estimativa do PSE de Lenth.

Tabela 48 - Comparação dos efeitos usando o projeto misto $2^1 \times 3^2$ e o projeto completo 2^4

Efeitos lineares estimados projeto misto, $2^1 \times 3^2$					Efeitos lineares estimados com o projeto 2^4				
Termo	Antocia- nina	L^*	C_{ab}^*	h_{ab}	Termo	Antocia- nina	L^*	C_{ab}^*	h_{ab}
Solvente (L)	0,00	1,83	-3,55	-4,01	A	-0,00948	2,851	-4,587	-4,864
Ácido (L)	-0,08	7,83	-14,44	-1,11	B	-0,08118	7,039	-12,979	-0,569
%Solvente(L)	-0,00	1,75	-2,06	-3,40	C	-0,00891	1,714	-1,506	-3,491
pH (L)	0,01	2,41	-6,15	0,52	D	0,00654	2,633	-5,948	0,391

6.5 Comentários sobre o DOE 2^{k-p}

Os projetos experimentais são usados como catalisador do processo sequencial do aprendizado científico. Portanto, fazer uma investigação passo a passo tendo a oportunidade de conjecturar sobre os fatores e como variá-los, as respostas a serem medidas, a localização do domínio experimental e com tudo isso definido, verificar como proceder. É claro que diferentes pesquisadores ou técnicos farão diferentes escolhas, o importante é que qualquer que seja a estratégia escolhida ela deve convergir para uma solução útil e econômica para a situação. À medida que o processo sequencial avança, ganha-se mais e mais conhecimento sobre o fenômeno, vão sendo remodelados os projetos e alterados os domínios experimentais de modo a ir diminuindo as incertezas. Por que não se recomenda resolver um problema mais complexo com apenas um projeto experimental? Porque projetos menores, que vão paulatinamente eliminando as incertezas, são mais efetivos na investigação. Um número menor de ensaios cuidadosamente escolhidos podem não resolver todos os problemas, mas podem servir como base para conjecturas adicionais sobre o problema.

Uma complementação maior de cada item apresentado aqui neste capítulo poderá ser feita em

Box, Hunter W., Hunter, J. (2005) e Montgomery, (2005).

Capítulo 7

Metodologia da superfície de resposta, noções da regressão linear múltipla e estudo de caso

7.1 Introdução

A análise, otimização e predição de desempenho de processos na engenharia são baseadas na modelagem matemática. Se o pesquisador ou o técnico tem a sua disposição informações completas sobre o fenômeno (por exemplo, dados termodinâmicos, de cinética e de hidrodinâmica) sobre o processo ou sistema de interesse, ele pode desenvolver um modelo matemático mecanístico na forma de um sistema de equações diferenciais. Para calcular os coeficientes que acompanham o modelo e verificá-lo, podem ser realizados alguns experimentos planejados ou não (AKHNAZAROVA E KAFAROV, 1982).

Um modelo matemático típico é uma declaração do tipo:

$$\hat{y} = \phi(x_1, x_2, \dots, x_i) \quad (31)$$

Já o modelo estatístico que é construído a partir de dados experimentais é uma aproximação do tipo:

$$y = \phi(x_1, x_2, \dots, x_i) + \varepsilon, \text{ o que leva a } y = \hat{y} + \varepsilon \quad (32)$$

Em que y é o valor observado da resposta, \hat{y} é a resposta predita pelo modelo, x_i são as variáveis independentes ou fatores e ε é o erro experimental. Os fatores ocupam o que é considerado como **espaço do fator** que se constitui no domínio experimental a ser estudado, e a representação gráfica da função resposta é conhecida como superfície de resposta. Segundo

Box, Hunter, W. e Hunter J. (2005) os modelos são sempre uma aproximação, os dados é que são reais, ou seja, eles acontecem.

A Metodologia da Superfície de Resposta (MSR), em inglês conhecida como “*Response Surface Methodology (RSM)*”, é uma técnica muito adequada a construção de modelos empíricos, dado o seu grande poder de análise. Foi esquematizada segundo (KHURY, & CORNELL, 1987), por J. Wishart, C. P. Winsor, E. A. Mitscherlich, F. Yates e outros, por volta de 1930, mas, só a partir de 1951, a RMS foi formalmente desenvolvida por G.E.P. Box e K.B. Wilson e outros colegas do Imperial Chemical Industries na Inglaterra. O seu objetivo foi explorar relações existentes entre algumas variáveis que supostamente interferem no processo (conhecidas como variáveis de entrada, variáveis independentes ou variáveis controláveis) com uma variável resposta (conhecida como variável independente), tal como o rendimento do processo. Desde então, vem sendo amplamente empregada em diversos campos, desenvolvimentos industriais, otimização de processos, pesquisas na agricultura e na biologia e em simulação computacional. Pode, inclusive, ser aplicada em qualquer procedimento experimental.

A metodologia da superfície de resposta consiste, portanto, de um conjunto de técnicas usadas no estudo empírico das relações entre uma ou mais respostas (rendimento, viscosidade e etc) e outras variáveis independentes (tempo, temperatura, concentração, catalisador, tipo de solvente empregado na extração, etc).

A técnica vem sendo empregada para responder a questões do tipo:

1. como uma resposta particular é afetada por um dado conjunto de variáveis de entrada (independentes) no espaço do fator de interesse?

2. que conjunto de condições específicas das variáveis de entrada, produz simultaneamente, as especificações desejadas para uma ou múltiplas respostas estudadas?
3. que valores específicos das variáveis de entrada produz um rendimento máximo, para uma resposta específica, e qual a superfície de resposta que contém esse máximo?

Em outras palavras, a RSM é uma técnica de otimização simples e eficaz.

Além das questões enumeradas ela ainda apresenta vantagens tais como:

1. a metodologia é uma aproximação seqüencial, isto é, o resultado de cada etapa guia a experimentação a ser conduzida na próxima. Em cada etapa realizam-se um modesto número de experimentos, o que garante que o tempo e o esforço dos experimentadores não sejam desperdiçados em ensaios improdutivos;
2. pode ser usada, na simplificação de uma função resposta complexa por meio de dados transformados;
3. ajuda a entender o problema experimental como um todo em termos geométricos;
4. pode ser usado para múltiplas respostas;
5. ajuda a migrar do empirismo para o mecanismo.

De acordo com (KHURI e CORNELL, 1987), as variáveis independentes nos projetos fatoriais em dois níveis são codificadas para facilitar a construção dos projetos experimentais. A codificação adimensionaliza as variáveis, deixa os eixos ortogonais em k dimensões e ainda define todas as variáveis em uma mesma métrica. Existem ainda vantagens adicionais quando se trabalha com variáveis codificadas e contínuas, como, por exemplo, a possibilidade de se ajustar modelos de primeira ordem. As variáveis codificadas proporcionam facilidade computacional, com aumento da exatidão na estimativa dos coeficientes do modelo e

facilidade de interpretação dos coeficientes, pois todas as variáveis estão sendo estudadas dentro de um mesmo intervalo.

No projeto fatorial completo em dois níveis, pode-se trabalhar com variáveis independentes qualitativas (decontínuas) ou quantitativas (contínuas). Os níveis de cada variável podem ser codificados em nível baixo (-) e nível alto (+), como mostrado foi mostrado no capítulo 4. No caso de se trabalhar apenas com variáveis quantitativas, quando os códigos são substituídos por -1 e +1, para todas as variáveis independentes, podem ser relacionados às variáveis originais, conforme a relação linear mostrada na equação 33.

$$x = \frac{T - \bar{T}}{\frac{T_{\min} - T_{\max}}{2}} \quad (33)$$

Em que:

x é a variável codificada;

T é a variável em unidades originais;

\bar{T} é a média entre os valores do nível baixo e do nível alto da variável original;

T_{\min} é o valor do nível baixo da variável original;

T_{\max} é o valor do nível alto da variável original.

Portanto, por meio da relação linear estabelecida entre a métrica original da variável e a codificada pode-se usufruir das vantagens da codificação sem perder a relação com a métrica original.

7.2 Alguns tipos de delineamentos usados na RSM

Os projetos experimentais para análise de superfície de resposta são usualmente baseados em delineamentos fatoriais em dois ou mais níveis. A regressão dos dados obtidos com delineamentos fatoriais em dois níveis resulta em superfícies de primeira ordem (funções matemáticas de primeiro grau). Os projetos compostos e os completos em três níveis resultam em superfícies de segunda ordem (funções matemáticas de segundo grau).

Embora os fatoriais em dois níveis possam empregar variáveis qualitativas e quantitativas, a superfície de resposta emprega apenas variáveis quantitativas como variável independente. Recomenda-se o uso de variáveis quantitativas também para a variável dependente, porque, segundo Schmidt e Launsby, 1998, respostas contínuas proporcionam mais informações e requerem menos recurso do que as respostas não contínuas ou categóricas. Eles ainda recomendam que se a resposta for um conjunto de categorias ordenadas (por exemplo, ruim, regular, bom, muito bom, excelente), pode-se atribuir a cada categoria um número e conduzir a análise como se a resposta fosse contínua. Mas se a resposta for binária (por exemplo, passa/falha), seria necessário um grande número de réplicas de cada delineamento a fim de evitar resultados ambíguos.

O procedimento de uso mais objetivo da RSM é:

1. montar um projeto de fatorial completo ou fracionado, com ponto central;
2. realizar ensaios e verificar a significâncias dos efeitos, fazendo-se uma regressão para um modelo linear;
3. verificar se há tendência de curvatura do modelo, pela estimativa da combinação linear dos coeficientes b_{ii} ;
4. se há indicação de curvatura, compor o modelo, realizar os ensaios, fazer uma regressão linear múltipla dos resultados; seguida da análise do modelo com sua interpretação geométrica;

5. se ainda não há indicação de curvatura, fazer a busca por meio da técnica do gradiente ascendente.

7.2.1 O gradiente ascendente

Segundo Khuri e Cornell (1987), o procedimento de realização de uma sequência de experimentos para se obter a melhor resposta (condições ótimas do processo) é independente da escala para variáveis codificadas. O procedimento é iniciado aproximando-se a superfície de resposta, após a realização do projeto de primeira ordem, por um hiperplano. Com as informações fornecidas pelo hiperplano, determina-se a direção para se obter a resposta desejada (máxima ou mínima). Move-se, então, a superfície de resposta na direção conveniente, até a região onde a curvatura do modelo é significativa, mostrando, por exemplo, a presença de um máximo. Nesta região planeja-se um projeto de segunda ordem, isto é, um projeto composto capaz de ajustar um modelo quadrático.

Para descrever o método gradiente ascendente matematicamente, começa-se assumindo hiperplano (modelo de primeira ordem), do tipo:

$$\hat{y} = b_0 + \sum_{i=1}^k b_i x_i \quad (34)$$

A próxima etapa é mover, a partir do centro do domínio, na direção desejada, usando de preferência a distância r (unidades). A maximização da função resposta é obtida pelos multiplicadores de Lagrange. Uma descrição detalhada deste tipo de aplicação, encontra-se em, (KHURI e CORNELL, 1987).

7.2.2 O delineamento DOE 2^k , DOE 2^{k-p} e ambos com ponto central

Os delineamentos DOE 2^k e DOE 2^{k-p} foram explorados nos capítulos 4 e 6, respectivamente.

Como já foi relatado nos capítulos 4 e 6 os planejamentos fatoriais em dois níveis completos e fracionados só permitem estimativas de efeitos principais (lineares) e interações (indicação de não aditividade), porque em um planejamento de experimentos em dois níveis é assumida linearidade entre os níveis extremos dos fatores. Quando há presença forte de não aditividade no espaço do fator é conveniente verificar se existe tendência ao aparecimento de pontos extremos (máximos ou mínimos) dentro ou na vizinhança do espaço do fator. Esta verificação é feita por meio da adição de ensaios no ponto central do domínio experimental (espaço do fator) à matriz de sinais. Como o próprio nome indica, este ponto é codificado como ponto zero entre os níveis codificados -1 e $+1$ de cada fator. Estes pontos são em geral repetidos para se obter uma estimativa do erro experimental puro (erro estimado com repetições genuínas) e a indicação de curvatura do modelo, pela técnica do ponto central, é bem explorada pelo EVOP e foi relatada no capítulo 5.

Em um planejamento de experimentos em dois níveis é assumida linearidade entre os níveis extremos dos fatores. Como, para o uso da RSM, os fatores são variáveis contínuas uma verificação de curvatura pode ser feita para avaliar a validade da suposição de linearidade, pela adição de pontos centrais à matriz do planejamento.

Uma descrição detalhada dos procedimentos de cálculo dos efeitos das variáveis e avaliação de curvatura do modelo encontra-se em (BOX, HUNTER, W., HUNTER, J., 1978; KHURI e CORNELL, 1987 e MONTGOMERY, 2005).

Lembrar que os pontos fatoriais em dois níveis podem ser representados geometricamente por: um quadrado para $k = 2$, um cubo para $k = 3$ e um hipercubo para $k > 3$.

7.2.3 projetos e modelos de segunda ordem (projeto composto)

A região próxima a um ponto extremo (mínimo ou máximo) é também chamada região estacionária. (AKHMAZAROVA e KAFAROV, 1982). Um polinômio de segunda ordem pode ser utilizado para descrever este tipo de região. Os projetos compostos ou projetos seqüenciais, propostos por BOX e WILSON, são muito empregados na solução de problemas desta natureza.

O projeto composto consiste de um projeto fatorial completo ou uma fração do projeto fatorial, com k variáveis independentes, ao qual se adicionam $2k$ experimentos posicionados nos eixos coordenados do projeto fatorial, isto é, em: $(\pm \alpha, 0, \dots, 0)$, $(0, \pm \alpha, \dots, 0)$, $(0, 0, \dots, \pm \alpha)$, e assim por diante, dependendo do número de fatores, onde α é a distância do ponto central ao ponto estrela. O ponto central deve ser repetido n_0 vezes. O número de repetições, n_0 , e o valor de α dependem do número k de variáveis independentes, da ortogonalidade e da rotacionalidade do projeto. Para maiores detalhes sobre estes projetos consultar (BOX e DRAPER, 1987; AKHMAZAROVA e KAFAROV, 1982; KHURI e CORNELL, 1987, BREYFOGLE III, 1999).

Os projetos compostos permitem o ajuste de polinômios de segunda ordem, do tipo:

$$\hat{y} = b_0 + \sum_{i=1}^k b_i x_i + \sum_{i=1}^k b_{ii} x_i^2 + \sum_{i \neq j=1}^k b_{ij} x_i x_j \quad (35)$$

Em que \hat{y} é a variável dependente (resposta) predita; x_i são os fatores (variáveis independentes); e b_i , b_{ii} e b_{ij} são os coeficientes de regressão não conhecidos, para i e j variando de 1 a k .

Estes modelos explicam a grande maioria das regiões exploradas no estudo. As superfícies de resposta podem ser traçadas geometricamente quando se trabalha com $k = 2$ variáveis independentes, ou podem ser interpretadas matematicamente quando $k > 2$.

Atualmente existem programas estatísticos para microcomputador que auxiliam na construção dos projetos e no tratamento matemático dos dados, mas convém lembrar que o conhecimento prévio do assunto ajuda o pesquisador a não fazer interpretações incorretas a partir de cálculos corretos (BOX , HUNTER, W., HUNTER, J., 1978).

A estimativa dos coeficientes (b_i , b_{ii} e b_{ij}) do modelo, a partir dos dados obtidos com os planejamentos aqui relatados, pode ser feita mediante o uso da regressão linear múltipla.

7.3 Regressão linear múltipla

A regressão linear múltipla é um método estatístico poderoso que pode lidar com dados mais complexos do que os outros, por permitir determinar quantitativamente as relações entre variáveis. Ela pode ser usada com dados planejados ou não.

Para deixar o leitor consciente do uso desta técnica estatística serão enumeradas a seguir algumas virtudes, armadilhas e aplicações da regressão.

As virtudes:

1. é a única ferramenta estatística que permite solução de problemas com mais de duas variáveis independentes;
2. não requer necessariamente experimentos planejados, podendo analisar dados históricos existentes na planta, indicando tendências;
3. a regressão não apenas qualifica os efeitos, como também os quantifica, por meio de coeficientes numéricos.

As armadilhas:

1. nem sempre os modelos observados na regressão indicam relações de causa e efeito entre as variáveis estudadas, principalmente do ponto de vista fenomenológico;
2. pode ignorar correlações entre variáveis independentes;
3. a regressão de muitas variáveis pode gerar um grande número de equações, dificultando a determinação do modelo que melhor se ajusta aos dados;
4. os modelos gerados não permitem extrapolações seguras.

As principais aplicações:

1. descrição de um fenômeno;
2. predição em pontos não ensaiados;
3. controle de processos.

7.3.1 A regressão linear simples e múltipla

Se apenas uma variável independente está envolvida, a regressão é chamada de linear simples.

$$y = b_0 + b_1x + \varepsilon \quad (36)$$

Se os dados que originaram a regressão foram codificados, então b_0 é a grande média e b_1 é o coeficiente angular da reta. Por outro lado, se os dados não foram codificados, b_0 é o intercepto da reta e b_1 continua sendo o coeficiente angular da reta.

Se, duas ou mais variáveis independentes estão envolvidas, a regressão é chamada de **regressão linear múltipla**. O modelo geral da regressão linear múltipla é dado por:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_ix_i + \varepsilon, \quad (37)$$

para modelos lineares, ou,

$$y = b_0 + \sum b_ix_i + \sum b_{ii}x_ix_i + \sum b_{ij}x_ix_j + \varepsilon \quad \text{para modelos quadráticos.} \quad (38)$$

Na regressão múltipla os coeficientes b_0 , b_i , b_{ii} e b_{ij} tem o mesmo significado explicado para a regressão simples, com respeito aos dados codificados ou não, a única diferença é quanto ao número de regressores, isto é, i ou maior ou igual a 2.

As vantagens do uso dos níveis dos fatores codificados já foram mencionadas no item 7.1, mas duas delas são mais óbvias:

1. a estimativa dos coeficientes da regressão é mais fácil e mais exata;
2. a interpretação dos coeficientes é acentuada porque os fatores variam sob um mesmo intervalo, logo a comparação entre os valores do coeficiente é imediata.

Ambas as vantagens residem do fato que a matriz $X'X$ (X transposta vezes X) em variáveis codificadas tem uma forma mais simples (isto é, fora da diagonal existem muitos valores zero para os elementos c_{ij} da matriz, onde $i \neq j$, sendo o caso mais simples quando a matriz é diagonal). Isto não ocorre com a matriz $X'X$ obtida com os níveis originais dos fatores.

Além disso, com a forma mais simples da matriz $X'X$ obtida com os níveis codificados dos fatores quando alguns ou todos $c_{ij} = 0$ da matriz (para $i \neq j$), resulta que os correspondentes parâmetros estimados são não correlacionados. Isto significa que os numeradores das estatísticas associadas com o teste de hipótese dos parâmetros individuais são estatisticamente independentes.

A metodologia empregada na obtenção de modelos empíricos, a partir da regressão linear múltipla, pode ajustar modelos lineares, ou não, com relação as variáveis independentes, embora lineares com relação aos parâmetros.

7.3.2 A estimativa dos coeficientes do modelo e sua interpretação

Em linhas gerais, para modelar um fenômeno de forma empírica, deve-se seguir os seguintes passos:

1. obtenção de dados experimentais (de preferência planejados);
2. adotar um modelo (fenomenológico ou empírico);
3. regressão para obtenção dos parâmetros;
4. análise dos resíduos;
5. modelo ajustado.

7.3.2.1 Obtenção dos parâmetros da regressão por meio das equações normais

Cada ensaio experimental gerará um ponto que pode expressar uma equação. Logo a série experimental gerará um sistema de equações do tipo:

$$b_0 \cdot x_{1,0} + b_1 \cdot x_{1,1} + \dots + b_{p-1} \cdot x_{1,p-1} + \varepsilon_1 = y_1$$

$$b_0 \cdot x_{2,0} + b_1 \cdot x_{2,1} + \dots + b_{p-1} \cdot x_{2,p-1} + \varepsilon_2 = y_2$$

$$\begin{array}{cccc} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{array}$$

$$b_0 \cdot x_{n,0} + b_1 \cdot x_{n,1} + \dots + b_{p-1} \cdot x_{n,p-1} + \varepsilon_n = y_n$$

Em que n corresponde ao número de ensaios experimentais e p o número de coeficientes ajustados.

Na forma matricial o sistema pode ser expresso

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{1,0}x_{1,1}\dots x_{1,p-1} \\ x_{2,0}x_{2,1}\dots x_{2,p-1} \\ \vdots \\ x_{n,0}x_{n,1}\dots x_{n,p-1} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (39)$$

Ou ainda nas formas:

$$y = X\beta + \varepsilon \quad \text{ou} \quad \hat{y} = X\beta \quad (40)$$

Sendo a matriz y o valor real da resposta, a matriz \hat{y} o valor da resposta estimado pelo modelo, a matriz x_{ij} os valores dos níveis atribuídos aos fatores (i refere-se a linha e j refere-se a coluna da matriz) e a matriz ε o resíduo da corrida experimental i . O resíduo agrupa o erro puro e a falta de ajuste ao modelo.

$$\varepsilon_i = y_i - \hat{y}_i \quad (41)$$

Portanto:

1. quanto mais ajustado for o modelo ao conjunto de dados o resíduo tenderá ao erro puro;
2. qualquer modelo ajustado, estatisticamente, conterà resíduos, porque a equação ajustada não passa por todos os pontos;
3. para se chegar ao melhor ajuste do modelo é necessário minimizar a somatória dos desvios quadráticos associados a cada corrida experimental (método dos mínimos quadrados).

O vetor β é o vetor dos parâmetros que minimiza $|X\beta - y|$, isto é, minimiza a somatória dos desvios quadráticos.

Os valores do vetor β , cuja estimativa é o vetor dos coeficientes, é uma matriz coluna na qual os valores dos coeficientes b_i , se o modelo é linear, e b_{ii} e b_{ij} ($i \neq j$) se os modelos forem quadráticos, aparecem na mesma ordem das colunas da matriz X . As matrizes apresentadas foram para modelos lineares, mas o procedimento de geração da matriz X para modelos quadráticos é o mesmo usado nos planejamentos fatoriais em dois níveis, ou seja, as colunas dos parâmetros b_{ii} são obtidas pela multiplicação dos elementos da coluna por eles próprios (x_i^2) e as colunas dos elementos x_{ij} para as estimativas, b_{ij} é obtida pelo produto da coluna i e j correspondentes de modo que $i \neq j$.

7.3.2.2 A interpretação dos parâmetros da regressão

Após a estimativa dos efeitos é importante expressar os coeficientes e sua significância, em geral por meio do teste t de Student. Para isso estima-se a matriz variância - covariância (os elementos da diagonal são as variâncias e os demais elementos da matriz são as covariâncias) do planejamento por:

$$\text{VAR}(b) = \text{MSE}(X^T X)^{-1}, \text{ lembrando que } \text{VAR}(\beta) = \sigma^2(X^T X)^{-1}$$

Lembrando que a matriz B é uma estimativa da matriz β e que MSE (Mean Squared Error) é uma estimativa de σ^2 .

Ou ainda: VAR(b), na forma matricial é:

$$\text{VAR}(b) = \begin{bmatrix} s_{b_0}^2 & \text{cov } b_0 b_1 & \cdots & \text{cov } b_0 b_k \\ \text{cov } b_1 b_0 & s_{b_1}^2 & \cdots & \text{cov } b_1 b_k \\ \vdots & \vdots & s_{b_i}^2 & \vdots \\ \text{cov } b_k b_0 & \text{cov } b_k b_1 & \cdots & s_{b_k}^2 \end{bmatrix} \quad (42)$$

Esta matriz é também conhecida como matriz Variância - Covariância

Vale lembrar que a covariância é a medida correlação linear entre duas variáveis e pode ser expressa por:

$$\text{Cov}(y_1, y_2) = E(y_1 - \eta_1)(y_2 - \eta_2).$$

Como esta grandeza é dimensional, isto é, ela tem as unidades de $y_1 \times y_2$, ela é, normalmente, expressa em termos de uma grandeza adimensional, denominada coeficiente de correlação linear, definida como:

$$\rho(y_1, y_2) = \frac{\text{Cov}(y_1, y_2)}{\sigma_1 \times \sigma_2}, \text{ cuja estimativa para as correlações entre as variáveis do}$$

$$\text{planejamento duas a duas é conhecido como: } r(y_1, y_2) = \frac{\sum (y_1 - \bar{y}_1)(y_2 - \bar{y}_2)/(n-1)}{s_1 \times s_2}$$

Analogamente, para os coeficientes das variáveis, a matriz correlação assume a seguinte aparência:

$$\begin{bmatrix} 1 & \frac{\text{cov } b_0 b_1}{s_{b_0} s_{b_1}} & \cdots & \frac{\text{cov } b_0 b_k}{s_{b_0} s_{b_k}} \\ \frac{\text{cov } b_1 b_0}{s_{b_1} s_{b_0}} & 1 & \cdots & \frac{\text{cov } b_1 b_k}{s_{b_1} s_{b_k}} \\ \vdots & \vdots & 1 & \vdots \\ \frac{\text{cov } b_k b_0}{s_{b_k} s_{b_0}} & \frac{\text{cov } b_k b_1}{s_k s_{b_1}} & \cdots & 1 \end{bmatrix} \quad (43)$$

A partir dessas matrizes, duas tabelas para a regressão podem ser montadas:

1. a tabela dos parâmetros da regressão com seus respectivos erros padrões e suas significâncias estimadas pelo teste t com o número de graus de liberdade com o qual o resíduo foi estimado; e,
2. a matriz da análise de variância da regressão que julga pelo teste F a significância dos coeficientes do modelo, e pode até decompor o resíduo em duas parcelas (erro puro e falta de ajuste "*lack of fit*"). Esta última estimativa só é possível se houver repetições genuínas do ponto central ou réplicas do projeto experimental.

A tabela dos coeficientes tem a aparência exibida na Tabela 49.

Tabela 49 - Coeficientes obtidos por meio da regressão dos resultados de um planejamento genérico

Coeficiente	Erro padrão do coeficiente	Valor de t	probabilidade
b_0	$\sqrt{s_{b_0}^2}$	$\frac{ b_0 }{s_{b_0}}$	
b_1	$\sqrt{s_{b_1}^2}$	$\frac{ b_1 }{s_{b_1}}$	
\vdots	\vdots	\vdots	\vdots
b_{ij}	$\sqrt{s_{b_{ij}}^2}$	$\frac{ b_{ij} }{s_{b_{ij}}}$	
\vdots	\vdots	\vdots	\vdots
b_{ii}	$\sqrt{s_{b_{ii}}^2}$	$\frac{ b_{ii} }{s_{b_{ii}}}$	

A segunda matriz (ANOVA da regressão) tem a aparência exibida nas Tabela 50 e Tabela 51.

Tabela 50 - ANOVA da regressão quando não há repetições genuínas ou réplicas do projeto

Fonte de variação	G. L.*	Soma de quadrados	Desvios Quadráticos médios	Razão F	Valor de p
Regressão (ajuste ao modelo)	$p-1$	S_R	$S_R/(p-1)$	$\frac{S_R/(p-1)}{MSE}$	
Resíduo (erro)	$N-p$	S_E	$MSE = S_E/(N-p)$		

* quando a média, bo (níveis codificados) ou o intercepto (níveis não codificados) não estão incluídos na ANOVA, o ajuste ao modelo é estimado com $p-1$ graus de liberdade. Apenas os coeficientes, b_i , b_{ii} e b_{ij} são considerados.

Sendo:

$$S_R = \sum(\hat{y} - \bar{y})^2, \text{ soma dos desvios quadráticos devido à regressão;} \quad (44)$$

$$S_E = \sum(y - \hat{y})^2, \text{ soma de desvios quadráticos devido ao resíduo;} \quad (45)$$

$$S_T = \sum(y - \bar{y})^2, \text{ soma dos desvios quadráticos totais;} \quad (46)$$

MSE é o Mean Squared Error

Por outro lado, a tabela de análise da variância ainda pode ser apresentada com a decomposição do resíduo em falta de ajuste (lack of fit) e erro puro.

Tabela 51 - ANOVA da regressão quando há repetições genuínas ou réplicas do projeto

Fonte de variação	Soma de quadrados*	Graus de liberdade	Média e desvios quadráticos
Devido à regressão (modelo ajustado)	S_M	$p-1$	$F = \frac{S_L}{S_{PE}}$
Resíduo $\left\{ \begin{array}{l} \text{lack of fit} \\ \text{erro puro} \end{array} \right.$	$S_E \left\{ \begin{array}{l} S_L \\ S_{PE} \end{array} \right.$	$n-p \left\{ \begin{array}{l} n-p-r \\ r \end{array} \right.$	
Total	S_T	$n-1$	

* S_M é a soma de desvios devido a todos os coeficientes (modelo); S_E é a soma de desvios quadráticos devido ao resíduo; S_L é a soma de desvios quadráticos devida a falta de ajuste ao modelo ("lack of fit") e S_{PE} é a soma de desvios quadráticos devido ao erro puro (repetições ou réplicas genuínas).

Uma outra estatística que acompanha a estatística F da ANOVA da regressão é o coeficiente de determinação R^2 .

Ele é expresso pela relação: $R^2 = \frac{S_R}{S_T}$ (47)

R^2 é a medida da proporção da variação total de y em torno da média \bar{y} explicada pelo modelo ajustado. R^2 pode ser expresso em % quando o valor é multiplicado por 100. Um problema potencial com esta estatística é que ela sempre cresce quando são adicionados fatores ao modelo, mesmo que estes fatores não sejam significativos. Por esta razão, muitos pesquisadores preferem uma outra estatística denominada de R_a^2 (coeficiente de determinação ajustado).

O coeficiente de determinação ajustado para os graus de liberdade é calculado por:

$$R_a^2 = 1 - \frac{(n-1)}{(n-p)} \times \frac{S_E}{S_T} = 1 - \frac{MSE}{S_T / (n-1)} \quad (48)$$

R_a^2 é uma estatística que é ajustada para o tamanho do modelo, ou seja, o número de fatores.

O valor de R_a^2 pode decrescer se termos não significantes forem adicionados ao modelo.

E finalmente, gráficos de resíduos e reta de 45° ($y = x$) dos valores observados em função dos valores preditos dão, também, informações interessantes sobre o ajuste do modelo.

7.4 Estudo de caso: Influência das variáveis de processo na desidratação da glicerina por destilação azeotrópica.

Este caso é parte do trabalho desenvolvido pela aluna Evelyn Edith Gutiérrez Oppe para obtenção do título de Mestre em Engenharia (GUTIÉRREZ OPPE, E. E, 2008). O processo de purificação da glicerina oriunda do biodiesel está sendo estudado no LSTM e conta com a forte colaboração do Prof Wilson Miguel Salvagnini e alguns alunos de mestrado e doutorado. Aqui serão abordados apenas os resultados com a destilação azeotrópica. Foi depositado um pedido de patente do processo completo de purificação da glicerina oriunda do biodiesel, tendo como inventores, Maria Elena Santos Taqueda, Wilson Miguel Salvagnini e Evelyn Edith Gutierrez Oppe.

7.4.1 O problema

As políticas mundiais estão promovendo o uso de biocombustíveis, com objetivo de minimizar a dependência da extração de combustíveis fósseis e reverter as atuais mudanças climáticas. Destes biocombustíveis, o biodiesel vem sendo introduzido na matriz energética mundial. O processo corrente de produção deste combustível usa a reação de transesterificação dos triglicérides presentes nos óleos vegetais ou na gordura animal. Esta reação gera cerca de 10% de glicerina, que precisa ser tratada para uso em outras aplicações. Os processos convencionais de purificação da glicerina apresentam grande consumo energético. O LSTM está propondo uma nova rota de refino da glicerina substituindo a destilação a altos vácuo e/ou a evaporação, por destilação azeotrópica com tolueno. Este tipo de destilação explora a formação de azeótropo entre os componentes, para facilitar a separação de misturas difíceis de separar por uma destilação convencional, ou quando o custo

da separação é elevado. Então o estudo inicial foi averiguar tecnicamente a viabilidade da desidratação da glicerina por destilação azeotrópica usando tolueno como arrastador. Assim, poderá ser possível ter um método de purificação da glicerina oriunda do biodiesel com um consumo de energia inferior ao dos processos de purificação atuais.

7.4.2 A proposta de solução do problema

É evidente que a verificação dessa viabilidade depende do conhecimento das interrelações das variáveis de processo na concentração final da glicerina. Uma boa estratégia de ataque ao problema seria construir um modelo empírico preditivo da concentração final de glicerina, em função das variáveis que potencialmente podem interferir no processo e que seriam de fácil controle. Para este tipo de abordagem, a metodologia da superfície de resposta é reveladora. Então, neste estudo de caso, a proposta de solução foi usar a referida metodologia para avaliar o processo de desidratação da glicerina por meio da destilação azeotrópica com tolueno.

7.4.3 A seleção das variáveis e o projeto experimental

As variáveis de processo (variáveis independentes) selecionadas foram: vazão de alimentação de glicerina (g/s) (x_1 codificado), vazão de alimentação do vapor de tolueno (g/s) (x_2 codificado), concentração inicial da glicerina (% em massa) (x_3 codificado) e temperatura de alimentação da glicerina (°C) (x_4 codificado). Estas variáveis podem ou não interferir na concentração final de glicerina, medida por índice de refração da solução.

Para a obtenção do modelo do processo foram utilizadas soluções aquosas com glicerina P.A. (99,5 %). O grau de pureza da glicerina foi verificado mediante a medida do índice de refração (ASTM D1747-99).

O tolueno grau P.A. foi utilizado como solvente arrastador. O grau de pureza foi verificado mediante a determinação do ponto de ebulição do tolueno (107°C a 700 mm Hg), usando a norma ASTM D 1078 – 05.

O equipamento em estudo foi montado em escala de bancada, construído em vidro refratário e está esquematizado na Figura 48, juntamente com seus acessórios. O detalhamento de cada acessório do equipamento, o procedimento de operação do equipamento, o estabelecimento dos níveis de operação de cada variável de processo e a calibração da concentração mássica da glicerina (%) em função do índice de refração estão descritos em (GUTIÉRREZ OPPE, E. E., 2008).

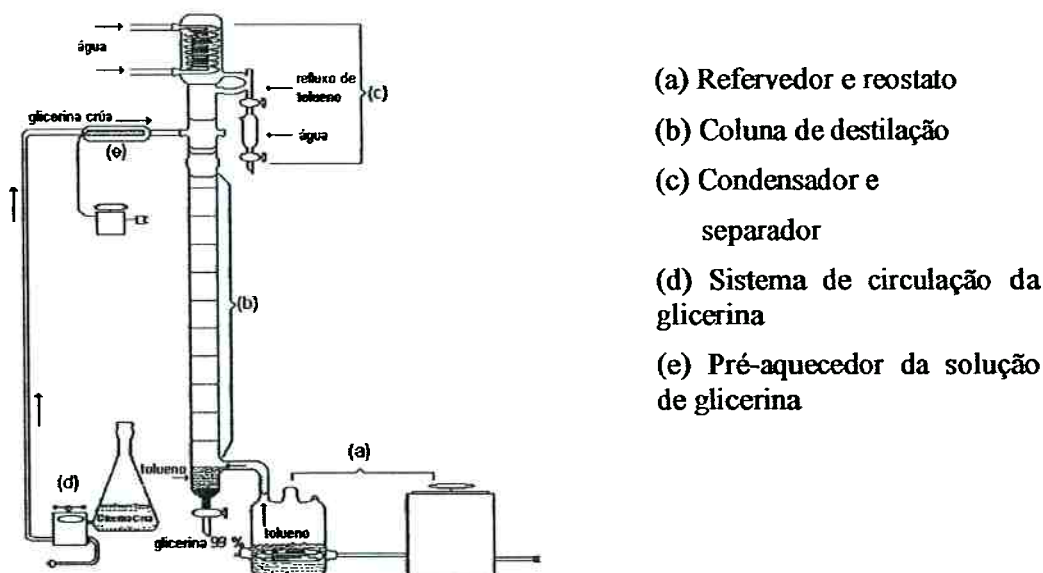


Figura 48. Esquema da aparelhagem utilizada na destilação azeotrópica da mistura glicerina-água montado no LSTM montado no LSTM

As variáveis de processo com seus níveis de variação originais e codificados estão apresentados na Tabela 52.

Tabela 52 - Apresentação das variáveis e seus níveis originais e codificados, para o estudo da destilação azeotrópica da solução de glicerina

Vazão de Alimentação da glicerina (g/s)	x_1	Vazão de alimentação Vapor de Tolueno (g/s)	x_2	Concentração inicial de glicerina (% em massa)	x_3	Temperatura de Entrada da Glicerina (°C)	x_4
0,12	-1	0,56	-1	50	-1	30	-1
0,51	0	0,83	0	70	0	60	0
0,89	+1	1,11	+1	90	+1	90	+1

A conversão das unidades originais das variáveis para a codificação foram feitas por meio das seguintes relações lineares:

$$x_1 = \frac{V_g - V_{g\text{médio}}}{(V_{g\text{máx}} - V_{g\text{mín}})/2} \quad (49)$$

em que: V_g = vazão de alimentação da glicerina (g/s), $V_{g\text{ médio}} = 0,51$ (g/s), $V_{g\text{ máximo}} = 0,89$ (g/s), $V_{g\text{ mínimo}} = 0,12$ (g/s).

$$x_2 = \frac{V_t - V_{t\text{médio}}}{(V_{t\text{máx}} - V_{t\text{mín}})/2} \quad (50)$$

em que: $V_{t\text{ inicial}} =$ vazão de alimentação de tolueno (g/s), $V_{t\text{ médio}} = 0,83$ (g/s), $V_{t\text{ máximo}} = 1,11$ (g/s), $V_{t\text{ mínimo}} = 0,56$ (g/s).

$$x_3 = \frac{Gli - Gli_{m\u00e9dio}}{(Gli_{m\u00e1x} - Gli_{m\u00edn})/2} \quad (51)$$

em que: Gli = % de glicerina alimentada, Gli_{m\u00e9dio} = 70 %, Gli_{m\u00e1x} = 90 %, Gli_{m\u00edn} = 50 %.

$$x_4 = \frac{T_i - T_{m\u00e9dia}}{(T_{m\u00e1x} - T_{m\u00edn})/2} \quad (52)$$

em que: T_i = temperatura inicial da glicerina, T_{m\u00e9dia} = 60 °C T_{m\u00e1x} = 90 °C, T_{m\u00edn} = 30 °C.

De posse das informa\u00e7\u00f5es apresentadas construiu-se um projeto experimental 2⁴ com 6 pontos centrais, perfazendo 22 ensaios que consistem de 16 pontos fatoriais e 6 pontos centrais, seguindo a codifica\u00e7\u00e3o apresentada na Tabela 52. A Tabela 53 apresenta os resultados destes experimentos e a Tabela 54 apresenta os efeitos e suas respectivas signific\u00e2ncias.

Tabela 53 - Matriz do planejamento fatorial com 6 pontos centrais indicando os resultados obtidos para concentração final de glicerina

Ordem padrão dos ensaios	x_1	x_2	x_3	x_4	Concentração final de glicerina (% massa) determinada por índice de refração
1	-1	-1	-1	-1	76,6
2	1	-1	-1	-1	51,6
3	-1	1	-1	-1	96,8
4	1	1	-1	-1	54,9
5	-1	-1	1	-1	99,7
6	1	-1	1	-1	94,6
7	-1	1	1	-1	96,2
8	1	1	1	-1	95,0
9	-1	-1	-1	1	84,5
10	1	-1	-1	1	56,6
11	-1	1	-1	1	88,2
12	1	1	-1	1	62,8
13	-1	-1	1	1	99,7
14	1	-1	1	1	93,5
15	-1	1	1	1	99,3
16	1	1	1	1	94,6
17	0	0	0	0	79,2
18	0	0	0	0	77,7
19	0	0	0	0	79,9
20	0	0	0	0	78,7
21	0	0	0	0	84,1
22	0	0	0	0	77,6

Observação: ordem aleatória de realização dos ensaios (20, 4, 22, 10, 17, 15, 8, 5, 19, 1, 13, 7, 6, 2, 12, 16, 11, 21, 9, 14, 3, 18)

Tabela 54 - Análise de variância mostrando a significância dos efeitos agrupados para o planejamento fatorial 24 com 6 pontos centrais

Fonte de variação	G. L.	Soma de desvios quadráticos	Média de desvios quadráticos	Razão F	Valor de p
Efeitos principais	4	3766,9	941,7	162,4	0,000
Interações de 2 fatores	6	766,3	127,7	22,0	0,002
Interações de 3 fatores	4	82,2	20,6	3,6	0,098
Interações de 4 fatores	1	29,7	29,7	5,1	0,073
Curvatura do modelo	1	88,5	88,5	15,3	0,011
Erro puro	5	28,9	5,8		

Observa-se pela ANOVA dos efeitos que há indicação de alta significância dos efeitos principais; das interações de dois fatores e da curvatura (valor de $p < 0,05$). Isso é indicativo de que existe um ponto extremo no espaço do fator ou mesmo na vizinhança do referido espaço. Então, seguindo a metodologia seqüencial, planejou-se o projeto de segunda ordem.

7.4.3.1 Planejamento fatorial composto

A composição do projeto foi feita acrescentando-se 8 pontos axiais ao projeto anterior, perfazendo agora um total de 30 ensaios. Para este projeto os níveis dos pontos axiais foram também -1 e $+1$, dada a limitação em ampliar os limites das variáveis, como é feito no projeto tipo estrela. O projeto composto realizado neste estudo é conhecido como projeto composto com de face centrada. Os pontos axiais bem como as concentrações finais de glicerina obtidas para os 30 ensaios (22 realizados com 2^4 com 6 pontos centrais) e mais os 8 pontos axiais estão apresentados na Tabela 55. Estes resultados foram analisados, com a finalidade de apresentação do modelo empírico, válido para a representação do referido conjunto de dados.

Tabela 55 - Matriz do planejamento fatorial com 6 pontos centrais e 8 pontos axiais indicando os resultados obtidos para concentração final de glicerina

Ordem padrão dos ensaios	x_1	x_2	x_3	x_4	Concentração final de glicerina (% massa) determinada por índice de refração
1	-1	-1	-1	-1	76,6
2	1	-1	-1	-1	51,6
3	-1	1	-1	-1	96,8
4	1	1	-1	-1	54,9
5	-1	-1	1	-1	99,7
6	1	-1	1	-1	94,6
7	-1	1	1	-1	96,2
8	1	1	1	-1	95,0
9	-1	-1	-1	1	84,5
10	1	-1	-1	1	56,6
11	-1	1	-1	1	88,2
12	1	1	-1	1	62,8
13	-1	-1	1	1	99,7
14	1	-1	1	1	93,5
15	-1	1	1	1	99,3
16	1	1	1	1	94,6
17	0	0	0	0	79,2
18	0	0	0	0	77,7
19	0	0	0	0	79,9
20	0	0	0	0	78,7
21	0	0	0	0	84,1
22	0	0	0	0	77,6
23	-1	0	0	0	96,3
24	1	0	0	0	73,9
25	0	-1	0	0	77,3
26	0	1	0	0	81,5
27	0	0	-1	0	60,3
28	0	0	1	0	96,9
29	0	0	0	-1	76,4
30	0	0	0	1	83,5

7.4.4 A análise e a interpretação do planejamento de 2ª ordem

Empregou-se o programa Estatística versão 7.0 para a realização dos cálculos estatísticos e a construção dos gráficos. Inicialmente estimaram-se os efeitos principais, as interações de dois fatores e os efeitos quadráticos. Os dois primeiros tipos de efeito já haviam indicado significância na análise do 2^4 com 6 pontos centrais. Vê-se agora, na Tabela 56, desdobrando-se termo a termo e verificando a significância de cada um, que o efeito da variável temperatura não foi significativo nem como efeito principal, nem como interação de dois fatores, nem como efeito quadrático. Portanto, pode-se prosseguir com a análise ignorando-se o efeito da temperatura.

A partir da constatação de que a temperatura inicial da glicerina não influencia no desempenho da destilação azeotrópica, no domínio experimental estudado, prosseguiu-se a análise e interpretação dos dados, ignorando o efeito da temperatura. Isto serviu para aumentar o número de graus de liberdade na estimativa dos resíduos.

Com os dados da Tabela 55 foi realizado a análise para a obtenção do modelo apresentado na equação 53 sem incluir o efeito da temperatura. Os coeficientes do modelo incluindo a temperatura (Tabela 56), os coeficientes do modelo sem incluir a temperatura (Tabela 57).

Tabela 56 - Efeitos principais; efeitos quadráticos e interações de dois fatores para o projeto composto de face centrada, com 30 experimentos

Termo	Efeito	Erro padrão do efeito	t (15 graus de liberdade)	Valor de p
Média	79,59	1,20	66,17	0,00000
x_1	-17,76	1,83	-9,73	0,00000
x_1^2	10,90	4,813	2,27	0,03868
x_2	3,91	1,83	2,147	0,04894
x_2^2	-0,50	4,81	-0,107	0,91889
x_3	26,36	1,83	14,44	0,00000
x_3^2	-2,10	4,81	-0,44	0,66896
x_4	2,32	1,83	1,27	0,22265
x_4^2	0,60	4,81	0,13	0,90212
$x_1 \times x_2$	-1,13	1,94	-0,58	0,56980
$x_1 \times x_3$	12,88	1,94	6,65	0,00000
$x_1 \times x_4$	1,130	1,94	0,58	0,56980
$x_2 \times x_3$	-4,48	1,94	-2,31	0,03543
$x_2 \times x_4$	-1,23	1,94	-0,63	0,53643
$x_3 \times x_4$	-1,33	1,94	-0,68	0,50416

x_1 = vazão de alimentação da glicerina (codificada); x_2 = vazão de alimentação de tolueno (codificada); x_3 = concentração de glicerina na alimentação (codificada) e x_4 = temperatura (codificada).

Tabela 57 - Efeitos e coeficientes para o projeto de face centrada, ignorando o efeito da temperatura na destilação azeotrópica da glicerina

Termo	Efeito	Erro Padrão	t(15)	p	Coefficiente	Erro Padrão	-95,%	+95,%
Média	79,61	1,04	76,46	0,00	79,61	1,04	77,39	81,83
x_1	-17,76	1,60	-11,13	0,00	-8,88	0,80	-10,58	-7,18
x_1^2	11,08	4,02	2,76	0,01	5,54	2,01	1,26	9,82
x_2	3,91	1,60	2,45	0,03	1,96	0,80	0,26	3,66
x_2^2	-0,32	4,02	-0,08	0,94	-0,16	2,01	-4,44	4,12
x_3	26,36	1,60	16,52	0,00	13,18	0,80	11,48	14,88
x_3^2	-1,92	4,02	-0,48	0,64	-0,96	2,01	-5,24	3,32
$x_1 \times x_2$	-1,13	1,69	-0,67	0,52	-0,56	0,85	-2,37	1,24
$x_1 \times x_3$	12,88	1,69	7,61	0,00	6,44	0,85	4,63	8,24
$x_2 \times x_3$	-4,48	1,69	-2,65	0,02	-2,24	0,85	-4,04	-0,43

A ANOVA para o modelo sem a inclusão da temperatura, com base no erro puro (decompondo o resíduo em erro puro e falta de ajuste) está na Tabela 58.

Observa-se com auxílio da tabela que não há falta de ajuste do modelo (valor de $p > 0,005$), e que os efeitos quadrático: x_2^2 (vazão de tolueno) e x_3^2 (concentração inicial de glicerol) e a interação entre $x_1 \times x_2$ não foram significativos ($p > 0,005$). Além disso, os coeficientes de determinação (ambas as definições) foram superiores a 0,9. Portanto, pode-se dizer que mais de 93% da variabilidade dos dados é explicado pelo modelo da equação 53. Antes de apresentar o modelo é conveniente verificar a sua adequação examinando, também, os gráficos de resíduos. Os gráficos estão nas Figura 49(a) e Figura 49(b).

Tabela 58 - ANOVA da regressão com a significância sendo estimada com base no erro puro, para o projeto de face centrada ignorando a variável temperatura

termo	Soma de quadrados	Graus de liberdade	Média de desv. quadráticos	Valor de F	Valor de p
x_1	1418,67	1	1418,67	123,91	0,00000
x_1^2	87,12	1	87,12	7,61	0,01464
x_2	68,84	1	68,84	6,01	0,02694
x_2^2	0,07	1	0,07	0,01	0,93746
x_3	3125,77	1	3125,77	273,01	0,00000
x_3^2	2,62	1	2,62	0,23	0,63945
$x_1 \times x_2$	5,06	1	5,06	0,44	0,51617
$x_1 \times x_3$	663,06	1	663,06	57,91	0,00000
$x_2 \times x_3$	80,10	1	80,10	7,00	0,01837
Falta de ajuste	95,74	5	19,15	1,67	0,20184
Erro puro	171,74	15	11,45		

$$R^2 = 0,954 \quad R_{adj}^2 = 0,933$$

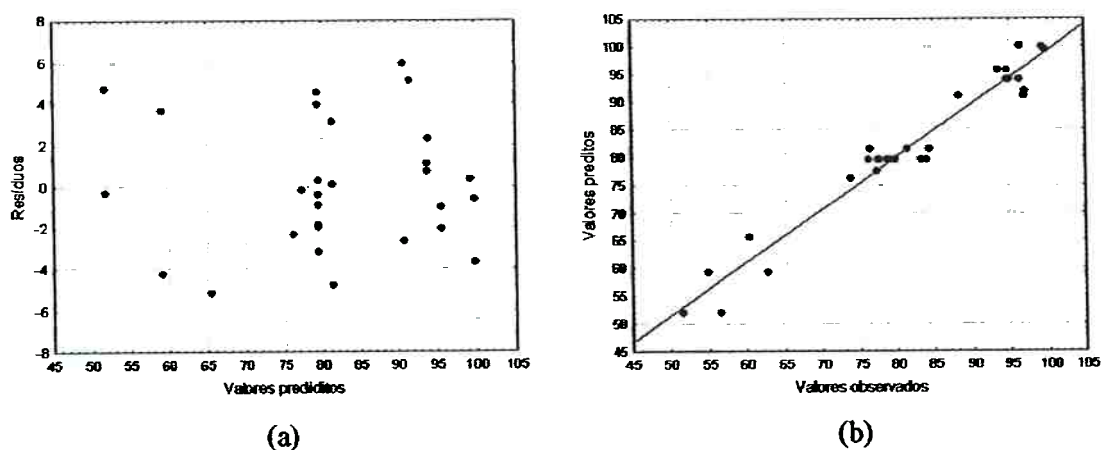


Figura 49. Representação gráfica (a) dos resíduos em função dos valores preditos e (b) valores preditos em função dos valores observados para verificação da adequação do modelo para a destilação azeotrópica do projeto de face centrada.

Observa-se pelos gráficos que o modelo explica muito bem a variabilidade dos resultados. Logo a equação 53 serve como boa aproximação para predizer valores da concentração de glicerina em pontos não ensaiados que estão dentro do domínio experimental.

$$\%GLI = 79,42 - 8,88x_1 + 5,54x_1^2 + 1,96x_2 + 13,18x_3 + 6,44x_1x_3 - 2,24x_2x_3 \quad (53)$$

Se $x_3 = -1$;

$$\%GLI = 77,46 - 8,88x_1 + 5,54x_1^2 + 15,42x_3 + 6,44x_1x_3$$

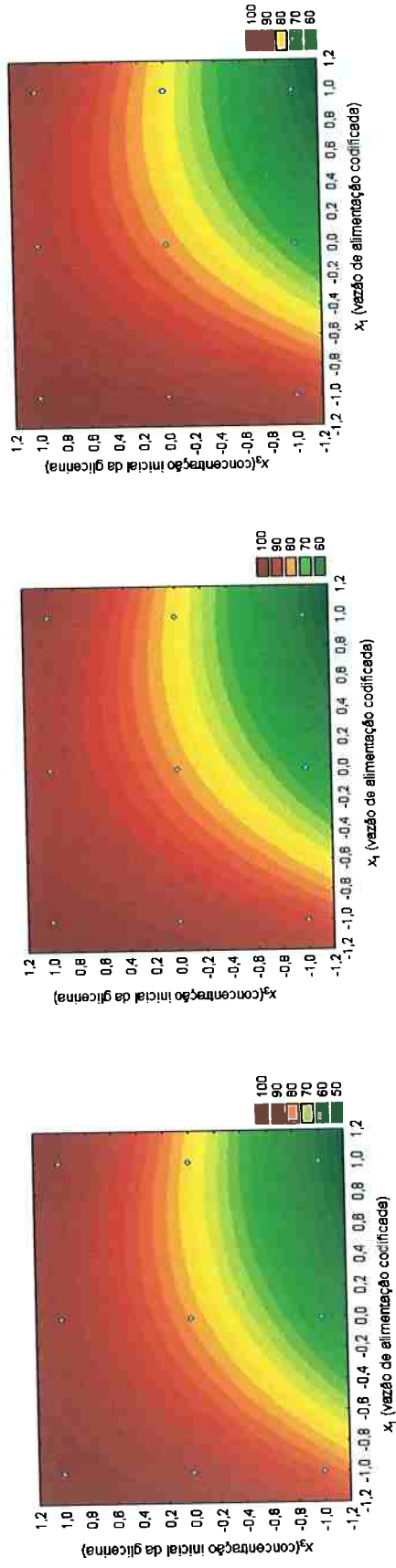
Se $x_3 = -0$

$$\%GLI = 79,42 - 8,88x_1 + 5,54x_1^2 + 13,18x_3 + 6,44x_1x_3$$

Se $x_3 = +1$;

$$\%GLI = 81,45 - 8,88x_1 + 5,54x_1^2 + 11,14x_3 + 6,44x_1x_3$$

Observando as equações geradas para valores crescentes da vazão de tolueno, nota-se que a constante cresce ligeiramente e o coeficiente linear de x_3 decresce também ligeiramente, de forma a compensar o aumento da constante. Por isso fica muito claro que quem governa a situação é a interação $x_1 \times x_3$. A influência da vazão de alimentação é maior com concentração inicial de glicerina mais baixa. O traçado das linhas de contorno apresentadas nas Figura 50 - Linhas de contorno da concentração final da glicerina predita pelo modelo da equação 53 obtido com os dados do projeto de face centrada (a), (b) e (c) indica a forte presença desta interação. Vê-se a influência da vazão de tolueno pelas maiores áreas avermelhadas, quando a vazão de tolueno cresce.



(a)

(b)

(c)

Figura 50 - Linhas de contorno da concentração final da glicerina prevista pelo modelo da equação 53 obtido com os dados do projeto de face centrada

Com este modelo empírico pode-se controlar com segurança a desidratação da glicerina no domínio experimental estudado.

7.4.5 Perspectivas da destilação azeotrópica para a desidratação da glicerina

A verificação da possibilidade de se obter glicerina com alto grau de pureza permitiu que o estudo nessa área dentro do LSTM progredisse.

O consumo energético na destilação azeotrópica é menor porque esta ocorre à pressão atmosférica (com a glicerina sendo sacada pelo fundo e a água pelo topo da coluna). Então, para que se tenha o fluido para ser processado por destilação azeotrópica, são necessárias algumas etapas adicionais no início do processo.

O modelo ainda pode ajudar a encontrar soluções com concentrações intermediárias de glicerina na saída, dependendo da aplicação a que se destina o produto.

O uso de processos alternativos requer cuidados especiais e análises criteriosas sobre as outras etapas do processo.

Capítulo 8

Considerações finais

A experiência adquirida durante a docência, lecionando e aplicando métodos estatísticos na pesquisa científica e tecnológica, fez com que a autora do texto selecionasse os métodos introdutórios ao projeto e análise de dados experimentais mais empregados na análise e melhoria de processos incluindo algumas aplicações desenvolvidas em colaboração com alunos e colegas. Objetivamente este texto tem a intenção de integrar algumas técnicas estatísticas nas práticas de engenharia, sem, no entanto, deixar de lembrar que o ensino de estatística para experimentadores (engenheiros, técnicos, cientistas) é fundamental como elemento motivador de seu uso.

Bisgaard (1989), em seu artigo sobre ensino de estatística, menciona dois pontos essenciais ao uso de métodos estatísticos na engenharia. Esse autor menciona que a estatística é encarada como um tópico separado, quase “ortogonal” à engenharia, a primeira não seria peça integral da segunda, uma vez que os métodos estatísticos auxiliariam na solução dos problemas propostos pela engenharia.

O autor, ainda, faz algumas comparações muito pertinentes, explicitando que, assim como a física é uma ciência distinta da matemática, a estatística, de maneira semelhante à física, usa a matemática para desenvolver teorias e métodos. Isto, no entanto, sem subestimar a importância da matemática na teoria estatística. Neste mesmo artigo, o autor coloca uma definição informal da estatística como sendo “a arte e a ciência da coleta e análise de dados”.

O renomado cientista Ronald Fisher (1948) também foi enfático ao explicitar o papel da estatística na pesquisa científica, acrescentando: “o fenômeno vem ao nosso conhecimento por meio da observação do mundo real”, e não é uma tarefa simples compreender, projetar e executar as formas de observação, estudo e experimentos, que possam suprir o conhecimento

necessário. Ou seja, o material observacional requer interpretação e análise. Nenhum progresso é esperado sem a experiência constante na análise e interpretação dos dados de diversos tipos. Ele ainda acrescenta que estamos prontos para planejar um experimento depois de realizá-lo”.

Nesta mesma linha de raciocínio, Box (1984) em seu artigo sobre “a importância da prática no desenvolvimento estatístico”, argumenta:

“para que o desenvolvimento ocorra são necessárias duas condições, ferramentas adequadas e motivação. As ferramentas são a matemática, a análise numérica e a computação. A motivação é a necessidade prática de resolver problemas”.

Ferramentas e motivação têm uma interação fortemente sinérgica. Ainda neste trabalho Box cita muitas contribuições à estatística aplicada de cientistas como Gosset, Fisher, Yates, Tukey, entre outros.

Sem ter a pretensão de autores envolvidos com o uso de estatística (como os mencionados anteriormente, sem esquecer Montgomery, cuja contribuição é incontestável em sua literatura sobre os métodos estatísticos aplicados), a autora deste texto apresentou a sua experiência como usuária de métodos estatísticos aplicáveis à pesquisa científica e tecnológica, na Universidade de São Paulo.

Os métodos aqui relatados foram cuidadosamente selecionados para apresentar sua eficácia quando usados adequadamente. Tentou-se também esclarecer que realizar planejamentos com séries experimentais curtas é essencial, mas é preciso ser cuidadoso na escolha do projeto.

Durante a experiência de uso efetivo dos métodos estatísticos aqui apresentados e de outros métodos que infelizmente não se teve a possibilidade de introduzir neste texto porque iria torná-lo muito mais longo, a autora acrescenta que contribuições adicionais podem ser vistas em trabalhos realizados em parceria com alunos e pesquisadores da comunidade científica.

Pelo menos três abordagens importantes, tratadas em outros trabalhos, são: o tratamento de dados históricos de processos industriais (Nasser Jr. 2005; Nasser Jr e Taqueda 2007); o uso de planejamento de misturas (AOKI, Guedes e Taqueda, 1998; Martinello, et al., 2006; Noriega, et al., 2008); a inclusão de variáveis de processo em planejamentos de misturas (Ferreira, et al. 2005); e a otimização de respostas múltiplas com a função desejabilidade, D (desirability function) (Altomani Jr, 1996).

Apesar da incontestável relevância do tema aqui apresentado, não se pode negar a dificuldade em tentar explicar a variabilidade dos processos por meio de uso dos métodos estatísticos com os recursos da linguagem escrita de forma compreensível. Em sala de aula, usam-se outras ferramentas, experiências, exercícios, e vários modos de abordagem que ajudam muito a conceituação da aplicação dos métodos.

A idéia do texto é fazer com que o leitor pense no uso dos métodos estatísticos como catalisador do conhecimento científico. As oportunidades de empregá-los em qualquer campo de atividade é cada vez maior, porque eles são capazes de acelerar a aquisição de conhecimento de forma robusta e com menor esforço econômico.

Durante a experiência acadêmica, as oportunidades de uso e comprovação da eficácia das aplicações de técnicas, quer com alunos do LSTM, quer em parceria com colegas do departamento ou de outras unidades da universidade e com empresas, são muitas e diversificadas.

Finalmente, é importante salientar que o texto não tem a abrangência que a autora gostaria, já que, se fosse esse o caso, o mesmo se tornaria inadequadamente extenso. Apesar disso, essa não deixa de ser a contribuição concreta a ser oferecida aos alunos e colegas que a prestigiaram, seja assistindo suas aulas, seja compartilhando na busca pelo conhecimento científico e tecnológico através dos métodos estatísticos.

Apêndice A

Conhecimentos Básicos de Estatística

A.1 Generalidades

A estatística é uma ciência preocupada com a organização, descrição, análise e interpretação de dados experimentais. Segundo Costa Neto, 1986, ela pode ser dividida em duas partes:

a estatística descritiva que trata da organização e descrição dos dados, isto é, extrair o máximo de informações contidas neles; e,

a estatística indutiva que cuida da análise e interpretação de elementos que se constituem, em geral, de uma amostra retirada de uma população que se deseja estudar.

População: é o número total de observações agrupadas que devem ocorrer como resultado do desempenho de uma medida. É sempre representada por uma distribuição.

Amostra: Um número limitado de observações, retiradas aleatoriamente de uma população, capaz de representar a população.

O conjunto de valores, também chamados de dados ou observações, obtidos da população ou da amostra em estudo são: para a população: $(x_1; x_2; x_3; x_4; \dots; x_N)$ e da amostra $(x_1; x_2; x_3; x_4; \dots; x_n)$. Com N o número total de valores observados para a população. N é um valor muito grande, muitas vezes tendendo a infinito e n um número limitado de observações da amostra.

Do ponto de vista da análise descritiva não é relevante se o conjunto de dados é uma amostra ou uma população inteira, já que, se os dados da amostra são obtidos a partir de uma tirada aleatória de uma população, a amostra é um subconjunto da população com um número limitado de elementos, onde todos eles serão examinados a fim de se realizar o estudo desejado. Portanto, a estatística indutiva permite tirar conclusões sobre a população a partir dos resultados observados nas amostras coletadas das respectivas populações.

É sabido que existe uma forte interação entre a aplicação dos métodos estatísticos na evolução de um desenvolvimento ou na melhoria de um processo ou de uma atividade.

Segundo BOX, et al 1978, esta interação pode ser bem ilustrada pela Figura 51. Onde, se pode confrontar o que foi dito sobre a estatística indutiva, isto é, um raciocínio que vem da prática para a teoria; do particular para o geral; dos resultados para as implicações; e é usado para gerar hipóteses. Já o raciocínio dedutivo vem da teoria à prática; do geral para o particular; das considerações para conseqüências lógicas; e é usado para testar hipóteses.

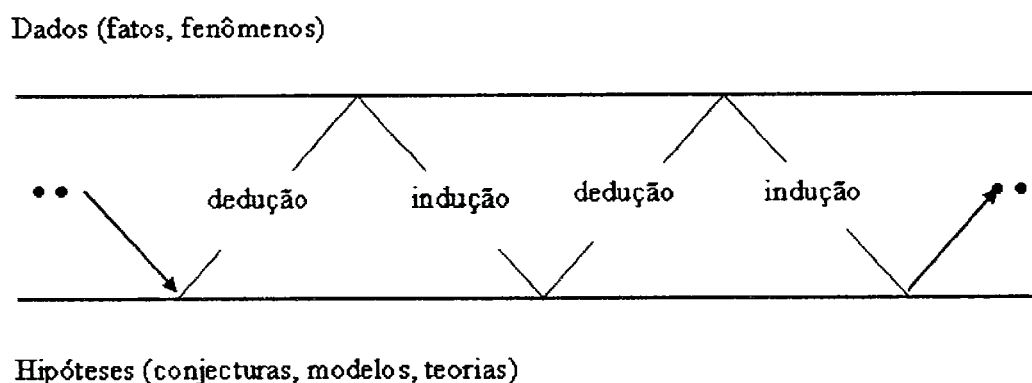


Figura 51. Ilustração da interação entre a aplicação de um método estatístico e o desenvolvimento ou a melhoria de um processo ou atividade. (fonte: BOX et al 1978)

Do exposto anteriormente é fácil perceber que o processo de indução não pode ser exato, isto é, se sujeita a cometer um erro que a própria estatística indutiva por meio de probabilidades dirá o quanto se está errando com as induções.

Pode-se dizer que a partir da Estatística Indutiva obtêm-se resultados sobre a população a partir de amostras, dizendo com que precisão e com que probabilidade se pode confiar nas conclusões obtidas.

É importante lembrar que quanto maior a amostra, mais precisas e mais confiáveis serão as induções observadas da população. O extremo seria dizer que os melhores resultados seriam obtidos examinando a população, o que não é factível na maioria dos casos.

Uma observação importante a ser feita é que antes de se analisar um conjunto de dados por meio de métodos da Estatística Indutiva, eles devem ser organizados por meio da Estatística Descritiva, lembrando que o método de Amostragem empregado na coleta dos dados é muito importante.

Costa Neto, 1986, diz que um estudo estatístico que recorre às técnicas da Estatística Indutiva deve envolver tópicos de Estatística Descritiva, Cálculo de Probabilidades e Amostragem, num esquema semelhante ao mostrado a seguir, ilustrado pelo próprio autor.

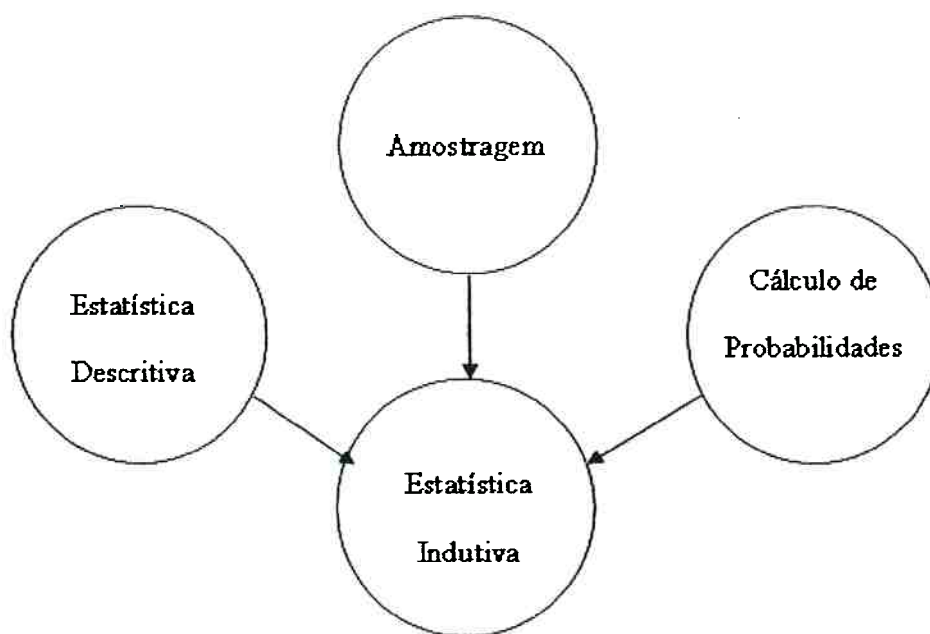


Figura 52. Esquema mostrando como a Estatística Descritiva, a Amostragem e o Cálculo de Probabilidades são importantes na análise por meio da Estatística Indutiva

Finalmente, a aplicação da estatística se estende a todas as áreas do conhecimento onde ocorre experimentação. Desde as ciências exatas, passando pelas ciências biológicas, sociais e administrativas.

A.2 Amostragem

A amostragem deve ser realizada de modo que a amostra seja representativa da população, isto é, a menos de algumas discrepâncias inerentes à aleatoriedade, a amostra deve conter todas as características básicas da população.

A.3 Estatística descritiva

Do ponto de vista da Estatística Descritiva que será apresentada neste item é indiferente se o conjunto de dados constitui-se em uma amostra ou em uma população. O mais importante é que a partir dos dados organizados, sejam bem definidas as características de interesse e obtidos o máximo possível de informações contidas neles. De uma maneira geral deseja-se reunir dados para obter informações a respeito de uma determinada área de interesse. Qualquer dado que varie é denominado de variável. Os dados podem assumir muitas formas, daí a necessidade de conhecê-las para decidir qual seria o teste estatístico mais adequado ao conjunto de dados.

As variáveis são de dois tipos: categóricas e numéricas. O diagrama da Figura 53 proposto por Patrie e Sabin, 2000 ilustra as diferenças básicas entre os tipos de variáveis.

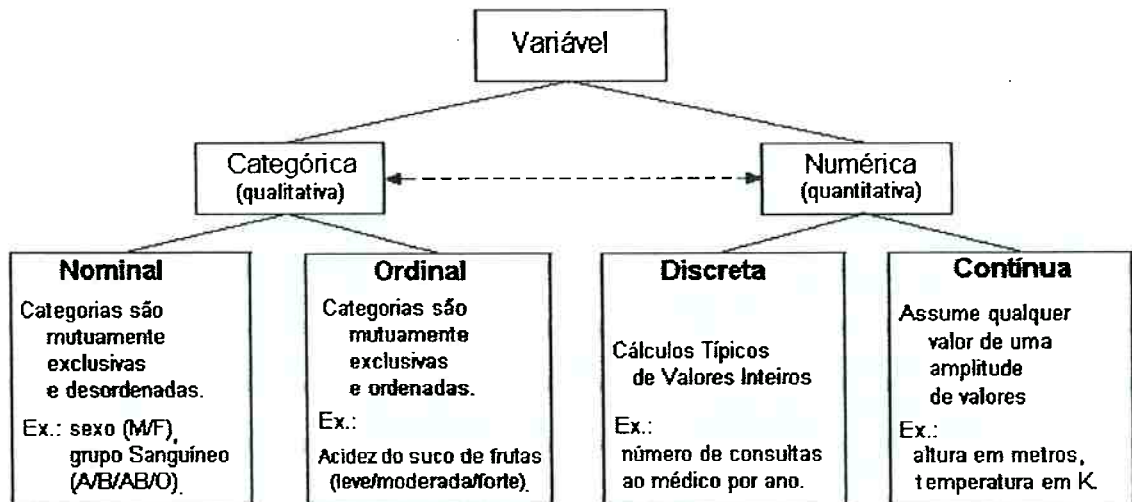


Figura 53 Diagrama de representação dos diversos tipos de variável

Deve-se salientar que em um processo têm-se as variáveis de entrada que são conhecidas como variáveis independentes (porque temos controle sobre elas) estas podem ser qualitativas ou quantitativas, e as variáveis de saída ou variáveis dependentes ou variáveis de resposta (porque dependem da influência das variáveis de entrada) que no caso dos processos da indústria de conversão, são quase sempre variáveis contínuas, por exemplo: o desempenho de um projeto medido pelo seu custo ou seu rendimento, ou a qualidade de um produto medido por alguma propriedade que lhe confere uma característica específica de qualidade. Portanto, a característica ou variável de interesse poderá ser qualitativa ou quantitativa.

Os valores das variáveis discretas são obtidos por alguma forma de contagem enquanto que, os valores de uma variável contínua é resultado de uma medição, expressa por uma unidade de medida característica da variável em questão.

Uma variável quantitativa discreta é interpretada pelos seus valores que são fixos, Ex: uma jogada de um dado apresenta um número fixo de pontos entre 1 e 6, enquanto que a interpretação de uma variável quantitativa contínua é de um valor aproximado, porque os instrumentos de medição são incapazes de oferecer precisão absoluta, Costa Neto, 1986.

Uma informação importante a ser acrescentada é que tanto para as variáveis quantitativas discretas quanto para as contínuas as técnicas da Estatística Descritiva serão idênticas. Assim, apenas na interpretação e descrição gráfica dos resultados é que haverá diferenças.

Na maioria dos processos da Engenharia Química (Indústria de Conversão) as variáveis estudadas são numéricas, logo é para este tipo de variável que será apresentada a Estatística Descritiva, lembrando que esta estatística descreve os dados por meio de gráficos, distribuições de frequência e parâmetros associados a estas distribuições.

A.3.1 Principais elementos da estatística descritiva

Introduzindo este item convém lembrar que a literatura representa os parâmetros da população por letras gregas e os da amostra por letras romanas, assim pode-se dizer que a população é representada pela amostra por meio das estatísticas: média da amostra, variância da amostra, desvio padrão da amostra, etc.

A.3.1.1 Descrição gráfica

A descrição gráfica de um conjunto de dados é feita mediante a verificação da frequência dos diversos valores da variável. A frequência é definida como o número de vezes que um dado valor foi observado, isto é, a frequência do i -ésimo valor será aqui denominada de f_i . Sendo n o número total de elementos observados, e k os diferentes valores assumidos pela variável, isto é: k varia de 1 a n .

$$\sum_{i=1}^k f_i = n \quad (\text{A. 1})$$

A associação das frequências fornece a distribuição de frequências que também é chamada de Diagrama de Frequências ou Histograma.

O histograma pode ser construído com as proporções ou freqüências relativas (usadas tanto para variável qualitativa quanto para variável quantitativa) como sendo a razão entre a freqüência e o número total de observações.

$$p_i = \frac{f_i}{n} \quad (\text{A. 2})$$

O histograma será a representação gráfica dos dados formados por uma sucessão de retângulos adjacentes, tendo cada um deles como base um intervalo de classe e como área a freqüência relativa dos valores que ocorrem naquele intervalo. Isto é equivalente a escolher uma escala vertical de modo que a área sob o histograma total seja igual a unidade. O diagrama de freqüências relativas para uma população é também chamado de distribuição de probabilidades.

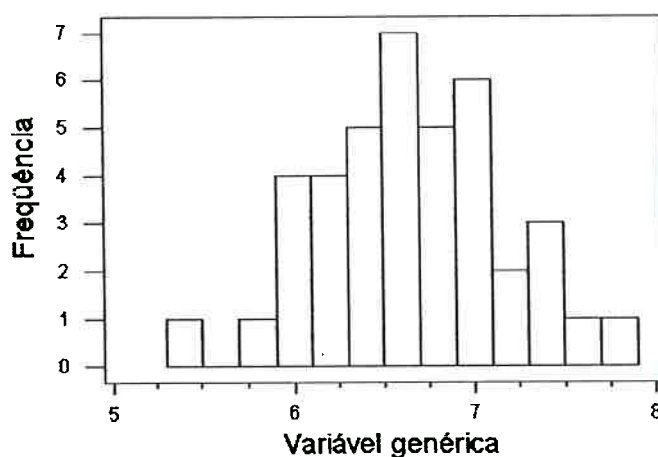


Figura 54 Exemplo de histograma, traçado para uma variável genérica

Além da descrição gráfica de um conjunto de dados, outras características numéricas da distribuição poderão ser sumarizadas por meio de certas quantidades que procuram quantificar os aspectos de interesse da distribuição. São elas as medidas de posição, de dispersão de

assimetria e de achatamento ou curtose. As medidas de posição e de dispersão são as de maior interesse, com grande aplicação na Estatística Indutiva.

A.3.1.2 Medidas de posição

As medidas de posição mais importantes de uma distribuição são: a média, a mediana e a moda. Elas servem para localizar a variável em seu eixo de variação. A média e a mediana indicam o centro da distribuição por critérios diferentes, portanto elas coincidem quando a distribuição é perfeitamente simétrica.

Média: é a medida da tendência central ou o centro de gravidade; é também o primeiro momento da distribuição.

Para a população,

$$\eta = \frac{\sum_{i=1}^N x_i}{N} \quad (\text{A. 3})$$

Sendo N um número muito grande de observações.

Ou, para a amostra:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (\text{A. 4})$$

sendo n um número limitados de observações.

Mediana, md : é uma quantidade que, como a média, também procura caracterizar o centro da distribuição de freqüências, porém com um critério diferente. Para os valores ordenados das observações, a mediana é dada pelo valor de ordem desse conjunto $(n+1)/2$ se n é ímpar ou o valor médio entre os valores de ordem $n/2$ e $(n/2)+1$ do conjunto de observações, se n for par. Portanto, a mediana divide as observações, previamente ordenadas, em duas metades. Quando o número de observações é ímpar, a mediana é o valor central da série ordenada de observações. Quando o número de observações é par, a mediana é igual à média dos dois valores centrais.

A idéia de a mediana dividir o conjunto ordenado em duas porções com igual número de observações pode ser interpretada como o ponto tal que, uma vertical por ele traçada divide a distribuição em duas partes iguais. A partir do exposto podem-se definir os quartis.

1º Quartil, Q_1 : é a medida abaixo da qual se situam 25% dos valores da série ordenada de observações, também chamado de quartil inferior;

2º Quartil, Q_2 : é a medida abaixo da qual se situam 50% dos valores da série ordenada de observações. O 2º quartil coincide com o valor da mediana

3º Quartil, Q_3 : é a medida abaixo da qual se situam 75% dos valores da série ordenada de observações, também chamado de quartil superior.

Estes três valores são freqüentemente usados para resumir os dados juntamente com o mínimo e o máximo. Eles são obtidos ordenando os dados do menor para o maior, e então conta-se o número apropriado de observações: ou seja é $\frac{n+1}{4}$, $\frac{n+1}{2}$ e $\frac{3(n+1)}{4}$ para o quartil inferior, mediana e quartil superior, respectivamente. Para um número par de observações, a mediana é a média dos valores do meio. Tem-se ainda os decis (dividem os dados em 10 partes iguais e

os percentis (dividem os dados em 100 partes iguais). Os quartis, decis e parcentis são genericamente chamados de fractis.

Moda ou modas, m_0 : Valor ou valores de máxima frequência das observações. A moda é também uma medida de posição.

Existe uma relação empírica entre a média a mediana e a moda expressa por:

$$\bar{x} - m_0 = 3(\bar{x} - md) \quad (\text{A. 5})$$

Esta expressão indica que geometricamente a mediana situa-se entre a media e a moda. Sua verificação, na prática, é mais realista para conjunto de dados maiores.

A.3.1.3 Medidas de dispersão

A média não dá uma informação completa sobre a distribuição. A amplitude, a variância (σ^2 da população e s^2 da amostra) e o desvio padrão (σ da população e s da amostra), erro padrão da média e o coeficiente de variação, completam as informações da distribuição.

Amplitude: é a diferença entre o valor máximo e o valor mínimo do conjunto de observações. A amplitude é também é chamada de intervalo. Por depender apenas dos valores extremos a amplitude não contém muitas informações sobre a distribuição, por isso contém pouca informação quanto à dispersão e não é muito utilizada como medida de dispersão.

$$R = x_{max} - x_{min} \quad (\text{A. 6})$$

Variância: representa a medida do espalhamento da distribuição. Ela diz quanto uma observação particular está afastada da média.

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \eta)^2}{N} \quad (\text{A. 7})$$

$$s^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{n-1} \quad (\text{A. 8})$$

Nota-se que, na equação da variância da amostra o denominador é N porque a média η é um parâmetro conhecido para a população, enquanto que a média da amostra é estimada a partir dos dados da própria amostra, esta é uma restrição que consome um grau de liberdade, fazendo com que o denominador da variância da amostra seja $(n-1)$.

Por outro lado, a dispersão indicada pela variância apresenta o inconveniente de possuir a escala (unidade de medida) quadrática em relação à média, enquanto que, a raiz quadrada do seu valor, denominada de *Desvio Padrão*, apresenta a vantagem de ter mesma unidade de medida da média e pode representar os desvios positivos e negativos da média.

Desvio padrão da população:

$$\sigma = \pm\sqrt{\sigma^2} = \pm\sqrt{\frac{\sum_{i=1}^N (x_i - \eta)^2}{N}} \quad (\text{A. 9})$$

Desvio padrão da amostra:

$$s = \pm\sqrt{s^2} = \pm\sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{n-1}} \quad (\text{A. 10})$$

Erro padrão da média, $s_{\bar{x}}$: é a raiz quadrada do quociente entre a variância, e o número de observações com o qual a referida média foi calculada, isto é:

$$s_{\bar{x}} = \sqrt{\frac{s^2}{n}} \quad (\text{A. 11})$$

Coeficiente de Variação (CV): é o quociente entre o desvio padrão, s , e a média \bar{x} expresso em percentagem.

Utilizado para comparação da dispersão, ou variabilidade, em termos relativos a seu valor médio.

$$CV = \frac{s}{\bar{x}} \times 100 \quad (\text{A. 12})$$

A.3.1.4 Medidas de assimetria

As medidas de assimetria indicam o quanto a distribuição é inclinada para a esquerda ou positivamente assimétricas ou inclinadas à direita ou negativamente assimétricas ou nulas.

Essas medidas procuram indicar o tipo de distribuição quanto a esse aspecto, embora nem sempre um valor zero indique simetria da distribuição. O valor numérico dessa medida é dado por:

$$\text{assimetria} = \frac{n}{(n-1)(n-2)} \times \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{s^3} \quad (\text{A. 13})$$

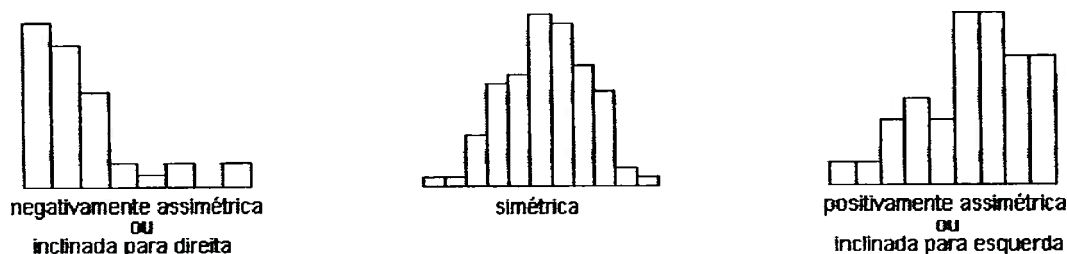


Figura 55. Exemplo de figuras com aspectos diferentes quanto à simetria

A.3.1.4 Medidas da achatamento ou curtose

Indica quão diferente é o achatamento da distribuição normal denominada de mesocúrtica, isto é a distribuição normal é tomada como referência e seu curtose é igual a 3. Assim distribuições mais achatadas do que a distribuição normal são denominadas platicúrticas e possuem curtose inferior a 3, e as mais alongadas do que a distribuição normal são denominadas de leptocúrticas e possuem curtose superior a 3. A figura a seguir indica o exposto.

O curtose é estimado por:

$$\text{curtose} = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \times \frac{\sum_{i=1}^{i=n} (x_i - \bar{x})^4}{s^4} - \frac{3(n-1)}{(n-2)(n-3)} \quad (\text{A. 14})$$

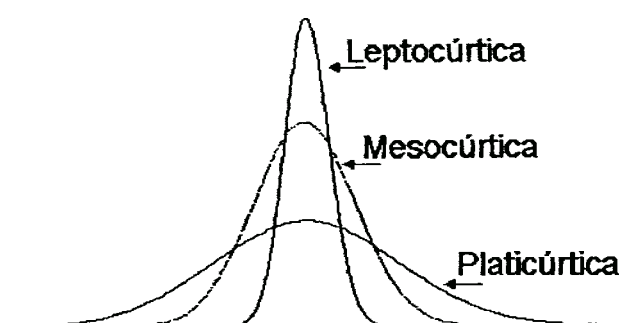


Figura 56. Exemplo de distribuições com diferentes curtoses

A.4 Distribuições

A.4.1 Distribuição Normal

Observações repetidas que diferem por causa do erro experimental variam em torno de um valor central na forma de uma distribuição grosseiramente simétrica na qual pequenos desvios ocorrem muito mais freqüentemente do que os grandes.

A distribuição contínua que está disponível para representar uma situação desta natureza, e que ocupa uma importante posição na teoria da estatística é a distribuição Gaussiana ou Normal.

A aparência dessa distribuição e sua função matemática estão apresentadas na Figura 6 e na equação 15, respectivamente.

O caráter simétrico da curva com sua mais alta ordenada no centro e suas caudas tendendo a zero, em ambos os sentidos, parecem indicar intuitivamente a expectativa do erro experimental. A distribuição normal tem a propriedade que o logaritmo de sua densidade de probabilidade, $p(x)$, é uma função quadrática do erro padronizado $(x - \eta) / \sigma$.

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\{(x-\eta)^2/2\sigma^2\}} \quad (\text{A. 15})$$

Existem duas razões para a importância da distribuição normal;

o efeito do limite central, que produz a tendência de o erro se distribuir semelhantemente a uma curva normal; e,

a robustez ou insensibilidade de muitos procedimentos estatísticos usados para desvios da normalidade teórica.

Em geral uma série de procedimentos estatísticos são derivados da consideração de normalidade das observações originais. Em muitos casos, aproximadamente normal é o que se requer para que esses métodos sejam utilizados. Neste contexto, os testes são ditos robustos a condição de não-normalidade.

A.4.1.1 Características importantes da distribuição normal

Uma vez conhecidas a média e a variância da distribuição, ela fica inteiramente caracterizada.

Box; Hunter; Hunter, 1978 apresentam resumidamente as seguintes declarações sobre a distribuição que são muito importantes como base para interpretação de resultados. São elas:

A probabilidade que um desvio positivo da média exceda 1 desvio padrão é 0,1587 (grosseiramente $\frac{1}{6}$). Esta fração está apresentada na área sombreada indicada na Figura 6;

Por causa da simetria da distribuição esta probabilidade é exatamente igual à chance de que um desvio negativo da média exceda um desvio padrão;

Das duas declarações anteriores fica claro que a probabilidade que um desvio em cada sentido (positivo e negativo) exceda um desvio padrão é de $2 \times 0,1587 = 0,3174$ (grosseiramente $\frac{1}{3}$), e

a probabilidade de tais desvios menores do que 1 desvio padrão é 0,6826 (grosseiramente $\frac{2}{3}$);

A chance que um desvio positivo da média exceda 2 desvios padrões é 0,02275 (grosseiramente $\frac{1}{40}$), que é representado pela área com o sombreado mais intenso da cauda da curva, também mostrado na Figura 57;

Novamente, exatamente igual à chance de ocorrer desvios da média que excedam dois desvios padrões.

Dessas duas declarações anteriores fica claro que a chance que um desvio da média em ambas as direções exceda dois desvios padrões é $2 \times 0,02275 = 0,0455$ (grosseiramente $\frac{1}{20}$ ou 0,05).

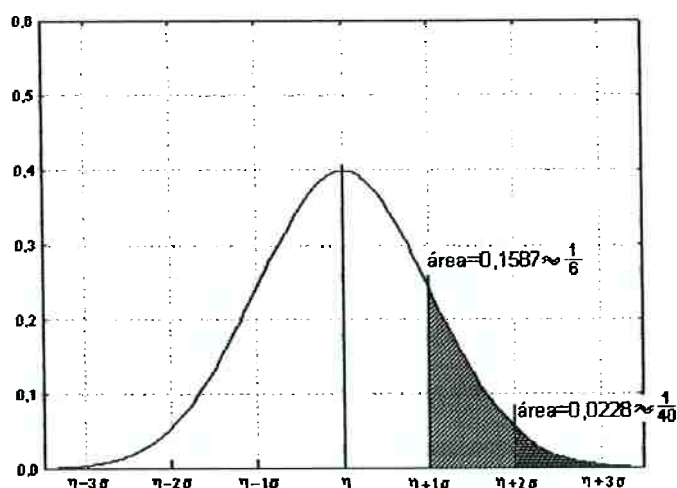


Figura 57. Distribuição normal com indicação das áreas das caudas com $x > \eta + 1\sigma$ e $x > \eta + 2\sigma$

A.4.1.2 Distribuição normal Padrão

A distribuição normal é apresentada por $N(\eta, \sigma^2)$, portanto esta distribuição é fortemente dependente da média e do desvio padrão. Esta distribuição na forma padronizada, também chamada de distribuição normal reduzida é apresentada como $N(0, 1)$, isto é, distribuição com média $\eta=0$ e $\sigma^2=1$ e, portanto, $\sigma=\pm 1$.

Esta distribuição tem densidade probabilidade expressa por:

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (\text{A. 16})$$

Qualquer distribuição normal pode ser transformada na distribuição normal padrão, z , mediante a relação linear:

$$z = \frac{x - \eta}{\sigma} \quad (\text{A. 17})$$

Esta declaração pode ser escrita:

$$Pr(x > \eta + \sigma) = Pr\{(x - \eta) > \sigma = Pr\left\{\left(\frac{x - \eta}{\sigma}\right) > 1\right\}$$

$$Pr(z > 1) = 0,1587$$

$$Pr(z < -1) = 0,1587$$

$$Pr(z > 2) = 0,0228$$

$$Pr(z < -2) = 0,0228$$

$$Pr(|z| > 2) = 0,0455$$

A distribuição normal padrão é tabelada. Em geral para determinar a probabilidade de x exceder algum valor x_0 , o desvio normal é computado por $z_0 = (x_0 - \eta) / \sigma$ e a probabilidade $Pr(z > z_0) = Pr(x > x_0)$, a partir da Tabela Convencional de Z, que dá $Pr(z > z_0)$ para vários valores de z_0 .

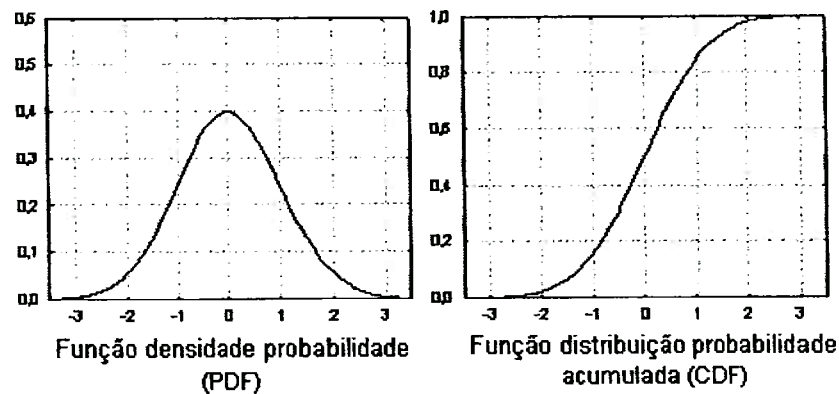


Figura 58. Distribuição normal padrão nas duas formas mais comuns de apresentação, função densidade probabilidade (PDF) e a função distribuição probabilidade acumulada (CDF)

A.4.1.3 Distribuição Lognormal

Uma variável x tem uma distribuição lognormal se $\text{Log}_e(x)$ ou $\ln(x)$ se distribui como uma normal. Conseqüentemente, com esta transformação, os dados podem ser analisados como se eles viessem de uma distribuição normal. Nesta distribuição, η e σ são determinados a partir dos dados transformados. A distribuição lognormal exibe várias formas PDF. Esta distribuição é comum em análises econômicas, biológicas, dados de vida, etc. Em outras palavras, é uma distribuição comum para dados que apresenta uma larga faixa de valores. A Figura 8 mostra a distribuição dos dados para triglicérides que apresentou este comportamento quando submetidos ao teste de Anderson Darling.

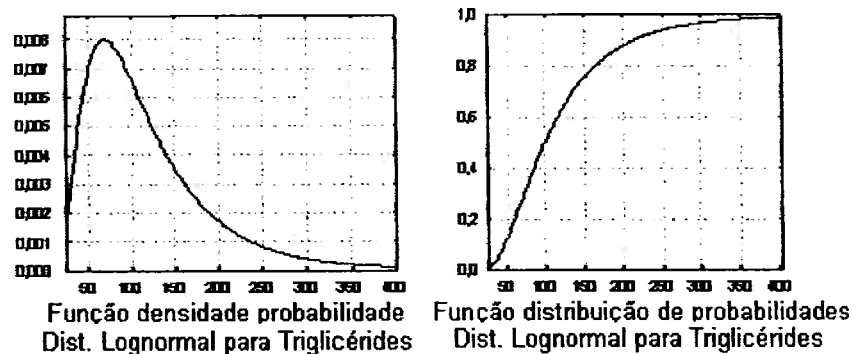


Figura 59. Ilustração da distribuição lognormal base e. O exemplo é a distribuição ajustada para triglicérides séricos em cálculo biliar

A.4.2 Outras Distribuições

A.4.2.1 Distribuição t

No caso da distribuição normal o desvio padrão σ é conhecido. Na prática, quase sempre, usa-se a amostra, logo o desvio padrão, σ , é desconhecido, e, portanto substituído pelo desvio padrão da amostra, s . Se a média, η , é conhecida e tem-se uma estimativa para o desvio padrão, pode-se estimar uma estatística denominada t_0 de Student, mostrada na equação 15. Não se pode estimar z_0 por não se conhecer o valor de σ .

$$t_0 = \frac{x_0 - \eta}{s} \quad (\text{A. 18})$$

Com base em algumas considerações, a quantidade $t_0 = (x - \eta) / s$, tem uma distribuição conhecida, cujos pontos estão tabelados em função do número de graus de liberdade com que foi estimado o desvio padrão.

Sabe-se que quando o número de graus de liberdade cresce o desvio padrão da amostra s tende ao desvio padrão da população σ , a distribuição t_0 de Student tende a distribuição normal. Portanto, a forma precisa da distribuição t_0 depende do grau de incerteza em s^2 , que é medido pelo número de graus de liberdade ν no qual a estatística s^2 está baseada. A Figura 9 mostra o que acaba de se relatado. Por exemplo:

$\nu = 1$	$\Pr(t > 2) = 0,148$
$\nu = 9$	$\Pr(t > 2) = 0,038$
$\nu = \infty$ (distribuição normal)	$\Pr(t > 2) = 0,023$

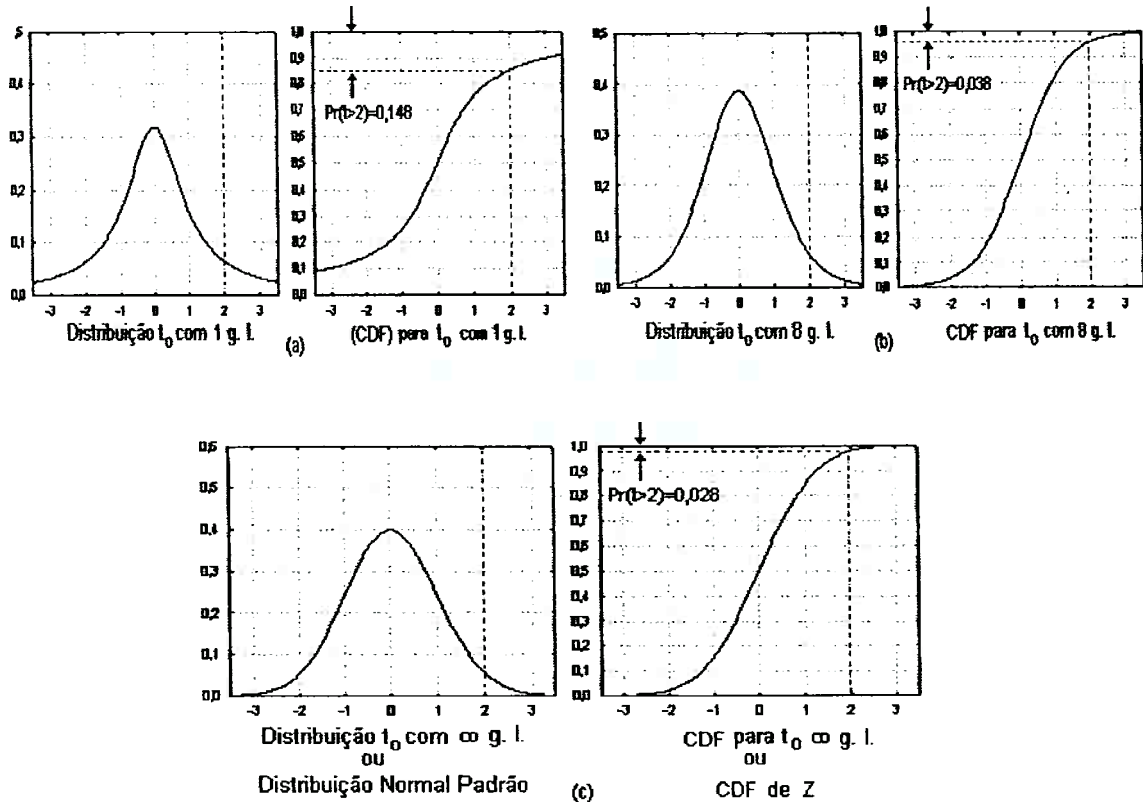


Figura 60. Evolução da aproximação da distribuição t_0 da distribuição normal padrão em função do número de graus de liberdade, (a) com 1 g. l.; (b) com 8 g.l. e (c) com ∞ g. l. ou distribuição normal padrão z.

Pela figura pode-se observar que exceto para as caudas maiores (baixos valores do número de g.l., ν) a distribuição t_0 proporciona uma importante aproximação para se fazer estimativas da população a partir da amostra. No entanto, algumas considerações devem ser respeitadas para a validade da distribuição t . São elas:

x é distribuída normalmente em torno de σ^2 ;

s é distribuída independentemente de x ;

a quantidade s^2 , que tem ν graus de liberdade, é calculada a partir da consideração de que x se distribui como uma curva normal com variância σ^2 .

A.4.2.1.1 Graus de liberdade de uma estatística

Foi visto anteriormente que a variância de uma amostra é calculada com o valor, $n-1$, no denominador e não n . A necessidade dessa correção está relacionada com o número de graus de liberdade, v , dessa estatística. Isto é, 1 grau de liberdade foi consumido, ao ser determinada a média \bar{x} a partir dos dados da própria amostra. Isto é, ao se estimar a variância da amostra, pressupõe-se que a média já tenha sido estimada anteriormente e agora tem-se $v=n-1$ graus de liberdade para variar. Dessa restrição pode-se escrever que:

$$x_1 + x_2 + x_3 + \dots + x_n = n\bar{x} \quad (\text{A. 19})$$

É importante lembrar que a introdução de outras restrições levaria a perda de mais graus de liberdade.

A.4.3 Distribuição χ^2

A distribuição amostral da estatística s^2 , conforme definida pela equação 8 está relacionada a uma família de distribuições de probabilidades de grande importância em diversos problemas da Estatística Indutiva. Elas são chamadas de χ^2 (*chi-quadrado*).

Essa estatística é definida pela a equação:

$$\chi_v^2 = \sum_{i=1}^v \left(\frac{x_i - \eta}{\sigma} \right)^2 = \sum_{i=1}^v z_i^2 \quad (\text{A. 20})$$

Observa-se que a distribuição χ^2 é função do número de graus de liberdade. Ela é uma distribuição assimétrica para poucos graus de liberdade e tende a uma distribuição normal a medida que o número de graus de liberdade aumenta.

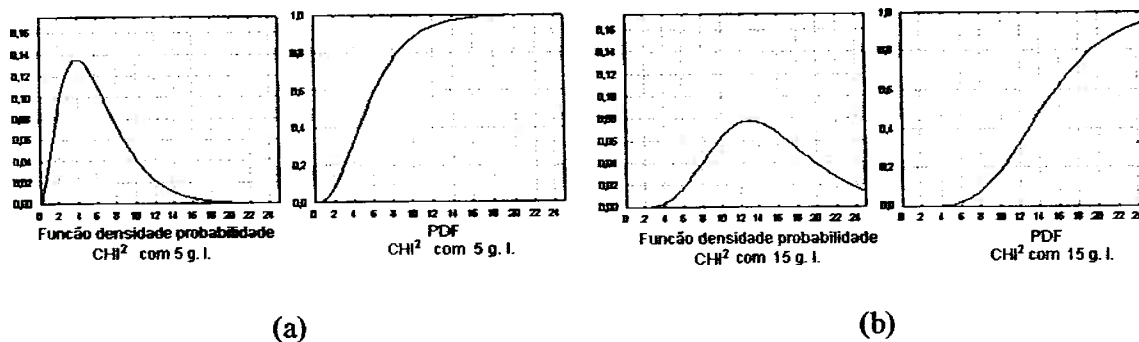


Figura 61. (a) distribuição χ^2 com 5 g. l. e (b) distribuição χ^2 com 15 g. l

A.4.4 Distribuição F de Snedecor

Supondo que uma amostra com n_1 observações é uma tirada aleatória de uma população normal com variância σ_1^2 , uma segunda amostra com n_2 observações é uma tirada aleatória de uma população normal com variância σ_2^2 , e que estimativas s_1^2 e s_2^2 das variâncias das duas populações sejam calculadas com ν_1 e ν_2 graus de liberdade. Então s_1^2 / σ_1^2 é distribuída como $\chi_{\nu_1}^2 / \nu_1$, e s_2^2 / σ_2^2 é distribuída como $\chi_{\nu_2}^2 / \nu_2$.

A razão $(\chi_{\nu_1}^2 / \nu_1) / (\chi_{\nu_2}^2 / \nu_2)$ tem uma distribuição assimétrica denominada F, com ν_1 e ν_2 graus de liberdade, cujos pontos de probabilidade estão tabelados. A distribuição F pode ser então aproximada por:

$$\frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} \approx F_{\nu_1 \nu_2} \text{ ou, equivalentemente, por } \frac{s_1^2}{s_2^2} \approx \frac{\sigma_1^2}{\sigma_2^2} \times F_{\nu_1 \nu_2}$$

A figura a seguir ilustra esta distribuição.

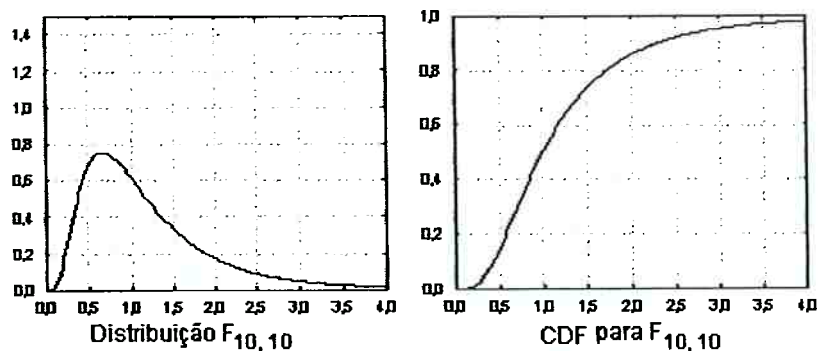


Figura 62. Distribuição F_{ν_1, ν_2} , no caso $\nu_1 = 10$ e $\nu_2 = 10$

A.5 Estatística Indutiva

O objetivo principal da estatística indutiva é tirar conclusões probabilísticas sobre aspectos das populações, com base na observação de amostras tiradas de populações de interesse. Essas amostras servirão de base para as inferências que serão feitas acerca das respectivas populações.

Os problemas da estatística indutiva podem ser tratados como os problemas de estimação de parâmetros e os testes de hipóteses. Genericamente falando a estimação de parâmetros pode ser feita por ponto ou por intervalo.

A estimação por ponto consiste em fornecer a melhor estimativa possível do parâmetro. Ele apresentará um único valor sobre o eixo de variação da variável. Se a determinação de um dado parâmetro é a meta final do estudo estatístico em pauta, a estimação por ponto é muitas vezes insuficiente, porque a probabilidade de que a estimativa por ponto venha a coincidir com o valor verdadeiro do parâmetro é, em geral, muito baixa. Isso decorre de os estimadores serem variáveis contínuas e aleatórias. Portanto, pode-se estar cometendo um erro de estimação, quando se está estimando um parâmetro populacional. Daí a idéia de se construir um intervalo em torno da estimativa por ponto, de modo que este intervalo tenha uma probabilidade conhecida de conter o verdadeiro valor do parâmetro. Essa é a idéia da

estimação por intervalo que se constitui em um problema típico da Estatística Indutiva. Este intervalo é denominado de Intervalo de Confiança.

A.5.1 Intervalo de confiança para a média da população quando σ é desconhecido

Neste item será mostrado como se constrói o intervalo de confiança da média da população quando o desvio padrão da população é desconhecido, o que ocorre na maioria dos casos práticos.

Quando σ é desconhecido, estima-se o seu valor por meio do desvio padrão da amostra, isto é, pela equação 9. Como s é uma variável aleatória, a substituição do parâmetro σ por s só é justificável para amostras com tamanho grande.

Logo para amostras grandes o intervalo de confiança para média pode ser estimado pela equação:

$$x \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (\text{A. 21})$$

Por outro lado, se a amostra for pequena, torna-se necessário a introdução de uma correção feita por meio da utilização da variável t_0 de Student ao invés de z , no cálculo do intervalo de confiança, e a equação torna-se:

$$x \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \quad (\text{A. 22})$$

O valor de t é obtido com o número de graus de liberdade com o qual foi estimado o valor do desvio padrão da amostra, para o nível de significância especificado.

A.5.2 Intervalo de confiança para a variância e o desvio padrão da população

O intervalo de confiança, para a variância da população, σ^2 , ao nível de $1-\alpha$ de confiança, pode ser construído a partir do conhecimento da distribuição χ^2 com v graus de liberdade, conforme as respectivas desigualdades:

Para a variância

$$\frac{(n-1)s^2}{\chi_{n-1, \alpha/2}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{n-1, 1-\alpha/2}^2} \quad (\text{A. 23})$$

Para o desvio padrão:

$$\sqrt{\frac{(n-1)s^2}{\chi_{n-1, \alpha/2}^2}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi_{n-1, 1-\alpha/2}^2}} \quad (\text{A. 24})$$

A.5.3 Testes de hipóteses

Os testes de hipóteses que serão tratados neste item serão referentes a parâmetros populacionais, portanto testes paramétricos.

Sabe-se que um dos objetivos da pesquisa científica é a formulação de hipóteses que são empregadas para adicionar base às decisões que venham a ser tomadas. Quando se formula uma hipótese acerca de um parâmetro populacional, ela é considerada válida até que se prove o contrário.

Uma hipótese estatística é uma declaração sobre alguma população quando se necessita tomar uma decisão a partir da amostra. Ela consiste de duas declarações complementares sobre a população. Por exemplo:

H_0 : a resposta média do índice de colesterol no sangue de pacientes acometidos de cálculo biliar é igual à resposta média desse mesmo índice nos pacientes acometidos de pólipos;

H_1 : a resposta media do índice de colesterol no sangue de pacientes acometidos de cálculo biliar é diferente da resposta média desse mesmo índice nos pacientes acometidos de pólipos.

H_0 e H_1 são chamadas hipótese nula e hipótese alternativa, respectivamente. Em geral para manter uniformidade, o resultado final é sempre anunciado em termos da hipótese nula, isto é, aceitar ou rejeitar H_0 .

Como a população é raramente conhecida com 100% de certeza, as duas declarações servem de apoio para o teste, com base nos dados da amostra. Neste caso, a probabilidade e a estatística são combinadas com os dados da amostra para inferir alguma coisa sobre a população inteira com uma quantidade de incerteza mensurável.

Portanto, dois riscos podem estar envolvidos dentro de um teste de hipóteses, risco α (erro do tipo I) e risco β (erro do tipo II). Quando conduzimos uma análise estatística e concluímos pela rejeição de H_0 , nós o fazemos com um risco α de erro. Este é o risco de rejeitar a hipótese nula erroneamente. Já o risco β é o risco de não rejeitar a hipótese nula erroneamente. Esta declaração também pode ser escrita como:

Erro tipo I: rejeitar H_0 , sendo H_0 Verdadeira;

Erro tipo II: aceitar H_0 . Sendo H_0 Falsa.

A probabilidade, α , do erro tipo I é denominada de *nível de significância* do teste.

Costa Neto, 1986, apresenta condensado em uma tabela que será transcrita como Figura 59.

Ilustração da distribuição lognormal base e. O exemplo é a distribuição ajustada para triglicérides séricos em cálculo biliar os resultados da aplicação de um teste de hipóteses e as respectivas probabilidades.

Tabela 59 - Resultados da aplicação de um teste de hipóteses, com as respectivas probabilidades

		Realidade	
		H_0 verdadeira	H_0 falsa
Decisão	Aceitar H_0	Decisão correta ($1 - \alpha$)	Erro tipo II β
	Rejeitar H_0	Erro tipo I α	Decisão correta ($1 - \beta$)

Pela tabela, o erro tipo I só poderá ser cometido se H_0 for verdadeira, e o erro tipo II, se H_0 for falsa. Da mesma forma o erro tipo I só poderá ser cometido se se rejeitar H_0 , e o erro tipo II, se se aceitar. A faixa de valores da variável de teste que leva a rejeição de H_0 é denominada de região crítica do teste, a faixa restante se constitui na região de aceitação. Em vista da dificuldade de se entender os erros tipo I e tipo II, pode-se em geral pensar nesses erros como sendo:

α o risco de rejeitar H_0 quando ela é verdadeira e β o risco de aceitar H_0 quando ele é falsa.

A.5.4 Teste de normalidade de Anderson Darling

A estatística de Anderson-Darling é uma medida de quão afastados os pontos da distribuição estão de uma reta traçada em papel de probabilidade normal. Neste teste é atribuído um peso maior aos pontos que estão nas caudas da distribuição. O programa usa a estatística Anderson Darling ajustada, porque a estatística varia quando um número diferente de pontos são usados para compor a distribuição (ROMEU, 2003).

$$A^2 = \sum_{i=1}^n \frac{1-2i}{n} \{ \ln(F_0[Z_{(i)}]) + \ln(1-F_0[Z_{n+1-i}]) \} - n \quad (\text{A. 25})$$

onde F_0 é a probabilidade da distribuição $Z_{(i)}$ [calculada com os parâmetros da distribuição $(\eta, \sigma; Z_{(i)})$] assumida normal, n é o número de observações (tamanho da amostra), \ln é o logaritmo natural (base e) e o subscrito i é a observação 1 a n , com os dados da amostra dispostos em ordem crescente.

Ao interpretar a hipótese nula, que a distribuição é normal, com os parâmetros assumidos, é então rejeitada, a um nível de significância $\alpha=0,05$, para uma amostra de tamanho n se A^2 for maior do que o valor crítico, VC :

$$VC = 0,752/(1 + 0,75/n + 2,25/n^2) \quad (\text{A. 26})$$

Isto é, a regra de rejeição é:

$$A^2 > VC = 0,752/(1 + 0,75/n + 2,25/n^2) \quad (\text{A. 27})$$

Em geral os programas de computador não estimam o valor crítico, expressam o nível de significância da estatística A^2 , que será julgada pelo analista, conforme o nível de significância estabelecido pela análise, como se julga um teste de hipótese convencional.

A.5.5 Teste t de Student

É o teste mais empregado para julgar a hipótese de comparação de médias de duas amostras. Ele é usado também para estimar o intervalo de confiança da diferença de médias quando o desvio padrão da população, σ , é desconhecido, que é o caso do estudo em questão. Serão comparadas médias de duas amostras e estimado os intervalo de confiança 95% da diferença. O teste é bilateral e as hipóteses testadas serão assim declaradas:

$$H_0 : \eta_1 - \eta_2 = \delta = \delta_0 = 0$$

$$H_1 : \delta \neq \delta_0 = 0$$

Onde η_1 e η_2 são as médias das populações e δ_0 é a diferença definida, por hipótese, entre as duas populações.

O teste t de Student ora apresentado pode ser aplicado em duas situações: com variâncias iguais ou variâncias desiguais.

A.5.5.1 Teste t para variâncias iguais

Se as variâncias são consideradas estatisticamente iguais (pelo teste F), o desvio padrão combinado a partir dos desvios padrões de cada amostra, levando em conta os graus de liberdade com os quais eles foram estimados, respectivamente, deve ser utilizado para o cálculo do valor de t.

Ou seja:

$$s = \pm \sqrt{\frac{\nu_1 s_1^2 + \nu_2 s_2^2}{\nu_1 + \nu_2}} \quad (\text{A. 28})$$

onde:

s_1 e s_2 são os desvios padrões das amostras e ν_1 e ν_2 são os graus de liberdade com que s_1 e s_2 foram estimados.

Portanto, considerando a diferença $\bar{x}_1 - \bar{x}_2$ o desvio padrão da diferença é dado por:

$$s_{\bar{x}} = \pm s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (\text{A. 29})$$

Onde n_1 e n_2 são os números de observações das amostras 1 e 2 respectivamente.

O valor de t é dado por:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{s_{\bar{x}}} \quad (\text{A. 30})$$

Os graus de liberdade, *g.l.*, com os quais a estatística *t* é estimada é dado por:

$$\nu = n_1 + n_2 - 2 \quad (\text{A. 31})$$

O intervalo de confiança para δ é:

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} s_{\bar{x}} < \delta < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2} s_{\bar{x}} \quad (\text{A. 32})$$

Com nível de confiança $1 - \alpha$

O teste *t* para variâncias iguais é levemente mais poderoso que o teste para variâncias desiguais.

A.5.5.2 Teste *t* para variâncias desiguais

Quando é assumido que as variâncias são desiguais pelo teste *F*, o desvio padrão da diferença de médias é estimado como:

$$s_{\bar{x}} = \pm \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (\text{A. 33})$$

Os graus de liberdade, *g.l.*, com os quais a estatística *t* é estimada é dado por:

$$\nu = \frac{(VAR_1 + VAR_2)^2}{[(VAR_1)^2 / (n_1 - 1)] + [(VAR_2)^2 / (n_2 - 1)]} \quad (\text{A. 34})$$

Onde:

$$VAR_1 = s_1^2 / n_1 \text{ e } VAR_2 = s_2^2 / n_2$$

Se o valor de ν , estimado pela equação 31 for um número decimal, assume-se o valor inteiro mais próximo.

A.5.6 Teste F

A distribuição F pode ser aproximada, como já indicado, pelas equações a seguir,

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \approx F_{v_1, v_2} \quad \text{ou, equivalentemente,} \quad \frac{s_1^2}{s_2^2} \approx \frac{\sigma_1^2}{\sigma_2^2} \times F_{v_1, v_2} \quad (\text{A. 35})$$

Esta distribuição é usada para checar hipóteses concernentes a razão de variâncias de duas populações normais. Em particular, a hipótese de que duas variâncias são iguais ($\sigma_1^2 = \sigma_2^2$) pode ser testada referindo-se a razão das variâncias da amostras diretamente pelo valor de F tabelado. O esquema da análise de variância idealizado por R. A. Fisher usa o teste F para avaliar a significância de uma ou mais variáveis. Esta análise conhecida como ANOVA indica qualitativamente se o efeito da variável ou das variáveis são significativos ou não, isto é, se a variabilidade explicada por esta variável é substancialmente maior do que a variância que ela não explica (erro ou resíduo). O resultado da análise é disposto em forma de tabela, conhecida como tabela de análise de variância. O teste com esta disposição é muito comum quando se deseja comparar mais do que duas médias para cada fator, assim têm-se: ANOVA one-way, Two-way, etc.

$$H_o : \eta_1 = \eta_2 = \eta_3 = \dots, \text{ ou}$$

H_1 : pelo menos uma media é considerada diferente.

A interpretação é feita por meio do valor da probabilidade, p , na saída do programa ou por comparação do F tabelado com o F calculado, para os respectivos graus de liberdade com os quais foram estimadas as variâncias dos tratamentos e do resíduo, respectivamente. Rejeita-se a hipótese nula se p é menor do que o nível de significância, α , crítico adotado ou se F tabelado é maior do que F calculado. A análise de variância será detalhada no capítulo 2.

A.6 Visualização de dados

A variabilidade de um processo ou atividade pode ser entendida pelos métodos estatísticos. Sucessivas observações de um sistema ou fenômeno não produzem o mesmo resultado. Então o julgamento estatístico com base em uma referência relevante é útil para incorporar esta variabilidade no processo de tomada de decisão. Uma variável aleatória qualquer pode ser pensada como o seguinte modelo:

$$X = \eta + \varepsilon \quad (\text{A. 36})$$

Em que, η é a constante (média dos resultados obtidos) e ε é a perturbação aleatória. Em outras palavras a constante permanece, porém ε muda com pequenas alterações do ambiente, equipamento, etc., se o resultado for tomado seguindo o mesmo procedimento. Os resultados do desempenho de um ensaio podem ser apresentados na forma de gráficos.

Esta visualização dá muitas idéias sobre como a variável se comporta, se existe diferença entre amostras, como duas ou mais variáveis se correlacionam, etc. Isso ajuda o pesquisador tomar decisões e formular hipóteses. A seguir serão mostrados vários tipos de gráficos que são frequentemente usados, além do histograma já mencionado anteriormente.

A.6.1 Digrama de pontos

O diagrama de pontos é um gráfico útil para exibir um pequeno número de observações até cerca de 20 observações.

A Figura 1 do capítulo 2 ilustra esta idéia.

A.6.1 Box and Whisker plots

Box-and-Whisker plots ou simplesmente box-plots são simples representações diagramáticas dos cinco números sumários: [mínimo, 1º quartil (quartil inferior), 2º quartil (mediana), 3º quartil (quartil superior), máximo]. Eles são úteis para mostrar as características distributivas

dos dados. Um box-plot consiste de uma caixa (box), bigodes e outliers, como mostra o esquema da Figura 12. Eles podem ser apresentados vertical ou horizontalmente.

A linha que divide a caixa é a mediana, as linhas a partir da caixa são os bigodes (whiskers) elas limitam a distribuição e os pontos indicados pelos asteriscos são os pontos fora da curva (outliers). O fundo da caixa é o primeiro quartil e o topo da caixa é o terceiro quartil. Os limites dos bigodes são estimados pelas equações a seguir:

Limite inferior:

$$L.I. = Q_1 - 1,5(Q_3 - Q_1) \quad (A. 37)$$

Limite superior:

$$L.S. = Q_3 + 1,5(Q_3 - Q_1) \quad (A. 38)$$

Os pontos que ficarem fora desses dois limites são os pontos considerados fora da distribuição, também conhecidos largamente como “outliers”.

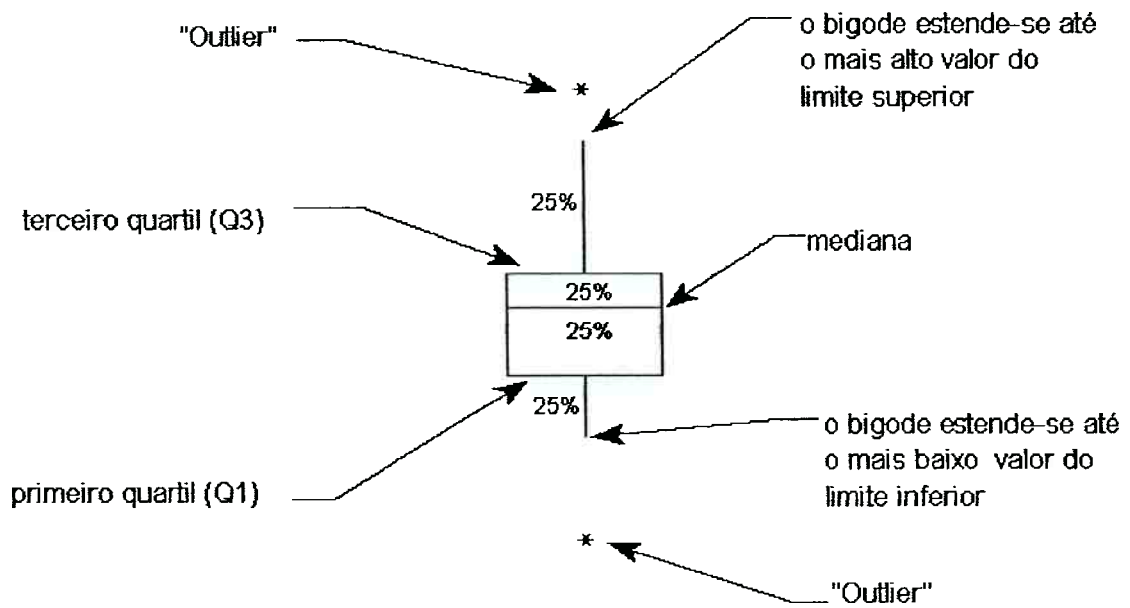


Figura 63. Esquema ilustrativo apresentando as principais partes do Box-and-Wisker plots ou simplesmente box plot

A.6.2 Matriz de correlações

É uma matriz bi-dimensional e simétrica composta por gráficos correlacionando as diversas variáveis entre si, sendo útil para visualizar de forma imediata e qualitativa as potenciais correlações entre todas as variáveis envolvidas, como mostra a Figura 64.

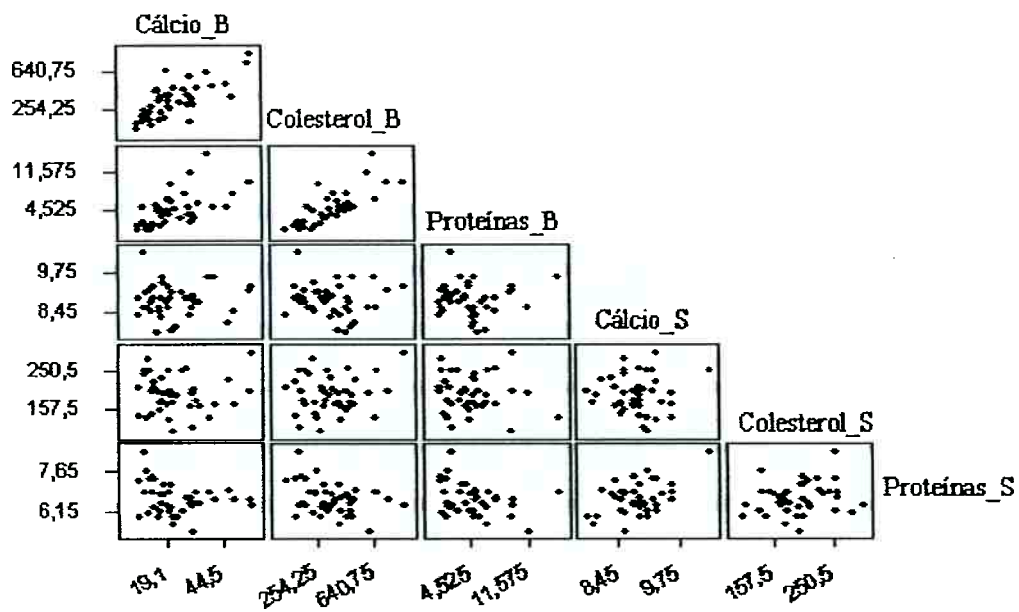


Figura 64. Ilustração da matriz correlação com algumas variáveis tratadas neste trabalho

Capítulo 9

Referências Bibliográficas

REFERÊNCIAS

- AKHNAZAROVA, S.; KAFAROV, V. **Experiment optimization in chemistry and chemical engineering**. Moscou, Mir publishers, 1982. 312p.
- ALTOMANI Jr, W.. **Otimização do Processo de Produção de Filmes de Polipropileno Com A Técnica Evop e A Função Desejabilidade**. 1996. 1v. Dissertação (Mestrado) Escola Politécnica, Universidade de São Paulo, São Paulo, 1996.
- AOKI, I. V.; GUEDES, I. C.; TAQUEDA, M. E. S. . Polarization curves and a experiment design as tools in the search of optimized inhibitors mixture formulation for HSLA steel in hydrochloric acid. **Materials Science Forum Trans Tech Publications**, v. 289-92, p. 1237-1244, 1998.
- AUSTIN, T. E. **Application of Six Sigma Methodologies to Improve Requirements Management for Customer Programs**. SAE International. ISSN 0148-719. 2006. p. 1-8.
- BISGAARD, S. Teaching Statistics to engineers. Invited paper presented at American Statistical Association Annual meeting. Washington, D. C. August, 1989
- BOX, G. E. P. The importance of practice in the development of statistics. **Technometrics**. V. 26,1. February, 1984.
- BOX, G. E. P.; HUNTER, W. G.; HUNTER, J. S. **Statistic for experimenters, an introduction to design, data analysis and model building**. 2ª ed. New York: John Wiley, 2005. 672 p.

- BOX, G. E. P.; HUNTER, W. G.; HUNTER, J. S. **Statistics for experimenters. An introduction to design, data analysis and model building.** New York, John Wiley & Sons, 1978. 653 p.
- BOX, G.E.P.; DRAPER, N. R. **Evolutionary Operation.** New York, John Wiley & Sons, 1969. 237 p.
- BOX, G.E.P.; DRAPER, N. R. **Evolutionary Operation.** New York, John Wiley & Sons, 1998. 237 p.
- BREYFOGLE III, F. W. **Implementing six sigma: smart solutions using statistical methods.** New York, John Wiley & Sons, 1999. 791 p.
- BRÜCKNER-MASCHINENBAU **General description.** Service manual, v. 1 do projeto SAO 1138, Siegsdorf, 1990.
- BRÜCKNER-MASCHINENBAU **Process information.** Descrição do projeto SAO 1138, Siegsdorf, 1989.
- COSTA NETO, P. L. O. **Estatística.** São Paulo, Edgard Blucher, 1986.
- DANIEL, C. Use of half-normal plot in interpreting factorial two-level experiments. **Technometrics**, V. 8, 1774. 1959.
- FERREIRA, L. F. P.; TAQUEDA, M. E. S.; VITOLO, M.; CONVERTI, A.; PESSOA JR, A.. Liquid-liquid extraction of commercial glucooxidase by reversed micelles. **Journal of Biotechnology**, v. 116, p. 411-416, 2005.
- FISHER, R. A. Biometry. **Biometrics** . V.4,1-9. 1948.
- GUTIERREZ OPPE, E. E. Desidratação por destilação azeotrópica da glicerina obtida como subproduto da produção do biodiesel. 2008. 1v. Dissertação (Mestrado) Escola Politécnica, Universidade de São Paulo, São Paulo, 2008.
- HAHN, G.J. : Process Improvement Using Evolutionary Operation. **Chemtech**, March 1976, p. 204-206
- INDUSTRIAL VENTILATION: a manual of recommended. twentieth third ed. Cincinnati, Ohio, ACGIH, 1998.

- JURAN, J.M.; GRZYNA, F.M. : **Controle da Qualidade**. Handbook, Makron Books do Brasil Editora Ltda , Editora McGraw-Hill Ltda. São Paulo, 1991. v. vii.
- KHURI, A. I.; CORNELL, J. A. **Response Surfaces: designs and analysis**. New York, Marcel Dekker, Inc., 1987. 405p
- KOCH, W. H; LICHT, W. New design approach boosts cyclone efficiency. **Chemical Engineering**. v 84, p. 80-88. 1977.
- KWOK, N. T. Otimização do processo de fabricação de acrilamida aplicando a técnica EVOP (EVOLUTIONARY OPERATION).. 2000. 1v. Dissertação (Mestrado) Escola Politécnica, Universidade de São Paulo, São Paulo, 2000.
- LENTH, R. V. Quick and easy analysis of unreplicated factorial. *Technometrics*, V. 31, 4. 1989.
- LEROUX, B.; ELMES, T.; MILLS, P. Study of shrinkage in biaxially oriented isotactic polypropylene. **Journal of materials science**, v. 27, p. 1475-1478. 1992.
- MAJEWSKI, R. ; ELIAS, R. B.; TAQUEDA, M. E. S. . Cyclone Washer Adapted from an American-Type Cyclone Separator. *Chemical Engineering and Technology*, v. 31, p. 1015-1022. 2008.
- MAJEWSKI R. Projeto, Construção e Avaliação Experimental de um Lavador de Gases Industrial. 2006. 1v. Tese (Doutorado) Escola Politécnica, Universidade de São Paulo, São Paulo, 2006.
- MAJEWSKI R. Concepção, desenvolvimento e teste de um lavador de gases. 2002. 1v. Dissertação (Mestrado) Escola Politécnica, Universidade de São Paulo, São Paulo, 2002.
- MARTINELLO, T.; KANEKO, T. M. S.; VELASCO, M. V.; TAQUEDA, M. E. S.; CONSIGLIERE, V. O. Optimization of poorly compactable drug tablets manufactured by direct compression using the mixture experimental design. **International Journal of Pharmaceutics**, v. 322, p. 87-95, 2006.

- MONTES ,C.; Vicário, I.M.;Raymundo, M.; Fett, R.; Heredia, F.J. Application of tristimulus colorimetry to optimize the extraction of anthocyanins from Jaboticaba (*Myrcia Jaboticaba* Berg.). **Food Research International**. v.38, p. 983–988, 2005.
- MONTGOMERY, D. C., *Design and analysis of experiments*, sixth ed., John Wiley & Sons, Inc., Hoboken, NJ, 2005.
- MONTGOMERY, D. C.; Runger, G. C.; Hubele, N. F. *Estatística Aplicada à Engenharia*. Rio de Janeiro, LTC, 2004. 335 p.
- MOPLEFAN S.P.A. **Propriedades gerais do polipropileno - considerações estruturísticas**. /Apostila interna publicada pelo Centro de Pesquisas, Terni, 1982.
- NASSER Jr, R. ; TAQUEDA, M. E. S. Improvement of Filter Tow Quality through Statistical Analysis. **Chemical Engineering and Technology**, v. 30, p. 64-70, 2007.
- NASSER Jr., R. Impacto das variáveis de produção na qualidade do Filter Tow. 2005. 1v. Dissertação (Mestrado) Escola Politécnica, Universidade de São Paulo, São Paulo, 2005.
- NORIEGA, P.; ROPKE, C. D.; CONSIGLIERE, V. O.; TAQUEDA, M. E. S.; TAVARES, L. C.; WASICKY, A.; BACCHI, E. M.; BARROS, S. B. M. Optimization of Pothomorphe umbrella (l.) Miquel topical formulations using experimental deseign. **International Journal of Pharmaceutics**, v. 353, p. 149-159, 2008.
- PETRIE, A.; SABIN, C. **Compêndio de estatística médica**. Lisboa, Instituto Piaget, 2000. 342 p.
- ROMEU, J. L. Anderson Darling: a Goodness of Fit Test for small samples Assumptions. A publication of DOD reability Analysis center. 2003, v. 10, 5 (Internet), 6p.
- TOMASELLO, M.L. *Origens culturais da aquisição do conhecimento humano (cap. 1 e 7)*. São Paulo: Martins Fontes,1999.

TUKEY, J. W. One degree of freedom for non-additivity, **Biometrics**, 1949. v.5, n. 3, p. 232-242.

SCHIMIDT, S. R. LAUNSBY, R. G. Understanding industrial designed experiments. 4 e. Colorado Springs, Air Academic Press, 1998. 280p.

ZAJĄCZKOWSKI, J; ODPYLANIE w przemyśle. Arkady, Warszawa, 1971.

ANEXO A

(Artigo utilizado no capítulo 6)



Application of tristimulus colorimetry to optimize the extraction of anthocyanins from Jaboticaba (*Myrcia Jaboticaba* Berg.)

C. Montes ^a, I.M. Vicario ^a, M. Raymundo ^b, R. Fett ^b, F.J. Heredia ^{a,*}

^a Department of Food Science and Nutrition, Faculty of Pharmacy, University of Seville, CIP. García González 2, 41012 Seville, Spain

^b Área do Bromatologia, Universidade Federal de Santa Catarina, SC, Brazil

Received 1 June 2004; accepted 22 January 2005

Abstract

This paper was aimed at optimizing the extraction conditions for the maximum yield of anthocyanins from Jaboticaba (*Myrcia Jaboticaba* Berg.) skins. Tristimulus colorimetry was applied to evaluate quantitatively and qualitatively the process. A mixed 2 and 3 full factorial design was used ($2^1 \cdot 3^3$). The factors and levels studied were type of solvent: ethanol or methanol; solvent/water mixture: 100% v/v, 90% v/v, 80% v/v; type of acid: chlorhidric acid, citric acid or acetic acid and pH: 2, 3 or 4. The estimated effect and ANOVA parameters were calculated. The yield of anthocyanins, quantified following a spectrophotometric method based on the absorbance at 535 nm was $92.92 \pm \text{mg/kg}$ of skin. The yield of anthocyanins showed a high and significant correlation with chroma (C_{ab}^*) ($p < 0.05$; $r = 0.84$) and lightness (L^*) ($p < 0.05$; $r = -0.85$) and it was significantly ($p < 0.001$) affected by the type of acid used and the solvent/water% ($p < 0.05$).

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Anthocyanins; Jaboticaba; CIELAB; Color; Tristimulus colorimetry

1. Introduction

Consumers' concern over the use of synthetic additives in food is increasing and in recent years there has been a growing interest in the food industry for new sources of natural pigments (Giusti & Wrolstad, 1996). Anthocyanins are natural pigments, responsible for a wide range of colors present in vegetables, flowers, fruits and derived products. Grape pomace, a sub-product originated in the red wine vinification process, is one of the main sources of enocyanin, a pigment usually employed as a food colorant, which was first isolated in 1879 (Marakakis, 1982). Anthocyanins can be located in different tissues of the berry, including the skin, the seeds or/and the pulp. Several extraction methods have been proposed to obtain extracts rich in anthocyanins,

usually based on solvents as methanol, ethanol, acetone, water or mixtures (Gonnet & Fennet, 2000; Metivier, Francis, & Clydesdale, 1980) in an acid medium (Hebrero, Santos-Buelga, & Rivas-Gonzalo, 1988; Ortega-Meder, Rivas-Gonzalo, Vicente, & Santos-Buelga, 1994; Rivas-Gonzalo, Bravo-Haro, & Santos-Buelga, 1995), or in neutral solvents (Revilla, Ryan, & Martín-Ortega, 1998). Antioxidants have also been added to the extraction solvent to prevent oxidation (Escribano-Bailón, Gutiérrez-Fernández, Rivas-Gonzalo, & Santos-Buelga, 1992).

Some factors such as pH, temperature, light, oxygen, metals, etc., may limit the usage of anthocyanins as food colorants. It is well known that the highest stability of these compounds is achieved in an acidic medium; however, they may undergo denaturalization when they are extracted from a natural source. The extraction process involves a loss of color followed by formation of brownish degradation products and insoluble compounds.

* Corresponding author. Tel.: +34 95 4556761; fax: +34 95 4557017.
E-mail address: heredia@us.es (F.J. Heredia).

This phenomenon has been evaluated by means of spectrophotometric indexes of degradation which are related to the changes in the visible spectrum of these compounds (Fuleki & Francis, 1968a).

The method for identification of anthocyanins is based on the absorption maximum in the visible and ultraviolet regions. In acidic media, their absorption spectra are characterized by two separate bands, one in the visible region between 465 and 550 nm and a less intense one in the UV region between 270 and 280 nm. The absorption maxima in the visible region vary with the nature of the solvent and its pH value. However, the absorption in the ultraviolet region is due to phenols grouping and it is not affected by pH changes. The wavelength of maximum absorption (λ_{\max}) in the visible region is around 520 nm, although it depends on the substituents (Marakakis, 1982; Somers & Ziemelis, 1985). The effect of pH on the chromatic properties of individual anthocyanins (extracted from red grapes) in model solutions has been established by Tristimulus Colorimetry (Heredia, Francia-Aricha, Rivas-Gonzalo, Vicario, & Santos-Buelga, 1998).

Anthocyanins extracted from grape fruits vary widely from 250 to 2600 mg/kg depending on the grape variety (Arozarena et al., 2002; Ortega-Meder et al., 1994). It is interesting to note the relevance of some under utilized tropical fruits as a potential source of anthocyanins for food-colorant purpose. One of these fruits is Jaboticaba (*Myrcia jaboticaba* Berg.), also known as guaperu, hipavuru, sabara or ybapuru. The tree is mainly grown in Brazil, where there is an important market for the fresh fruit and the derived products as jellies, jams and wine and, to a limited extent, in other parts of South America (Popenoe, 1974).

Jaboticaba fruit is grape-like in appearance and texture, although its skin is thicker and tougher. This fruit is dark purple to almost black color due to a high content of anthocyanins (Trevisan, Bobbio, & Bobbio, 1972).

The application of colorimetric systems, based on uniform color spaces (CIELUV and CIELAB) and non-uniform color spaces (CIEXYZ), is of great value in the quantification and characterization of the color properties of pigments and foods such as the cochineal pigment (Méndez, González, Lobo, & Carnero, 2004) or orange juices (Meléndez-Martínez, Vicario, & Heredia, 2003), among others.

In this study, the main factors (solvent and acid medium) related with the extraction of anthocyanins from a non-conventional source, the jaboticaba fruit skin, were investigated. The color properties of the anthocyanins extracts were established by Tristimulus Colorimetry. The aim was to choose the best conditions for a quantitative and qualitative (related to color properties) extraction of anthocyanins from Jabotica skin in order to be used as food colorant.

2. Material and methods

2.1. Samples

Around 3 kg of Jaboticaba fruits (*Myrcia jaboticaba* Berg.) were harvested from three different trees in the surroundings of Florianópolis (Brazil). Immediately after harvesting, the fruits were frozen and kept at -18°C in the dark until analyses. Before extraction, the frozen fruits were manually peeled.

2.2. Extraction procedure

The extraction conditions were based on that proposed by Francis (1982, 1989). Portions of 10 g of skin of Jaboticaba fruit were extracted with 20 ml of a mixture solvent/water for 24 h at 4°C (Francis, 1982). Subsequently, the skin bulk was pressed and the extract filtered through Whatman no. 1 filters. The different extracts were kept at 4°C in the dark until analyses.

2.3. Experimental design

A mixed 2 and 3 full factorial design was used ($2^1 \cdot 3^3$) in order to find the optimized setting for the different factors that affect the final yield of anthocyanins. The factors studied were: type of solvent, solvent concentration (% solvent/water, v/v), type of acid, and pH. Two solvents, ethanol and methanol, were added fixed proportions of water and assayed at three levels of concentration 100% v/v, 90% v/v, 80% v/v. Chlorhidric acid, citric acid, and acetic acid were selected to obtain three levels of solvent acidity, pH 2, pH 3 and pH 4. The total number of runs for this design should be 54, however, citric acid ($\text{p}K_{\text{a}} = 3.15$) only provided $\text{pH} > 2$, so only 48 runs were done. The order of the experiments was fully randomized to avoid the effects of lurking variables. The results obtained were analyzed by means of analysis of variance (ANOVA). The multivariate statistical analyses were carried out using the software Statistica[®]v 5.5 (Statsoft, 1999).

2.4. Quantitative analysis

Anthocyanins were quantified following the spectrophotometric method proposed by Francis (1982, 1989). One hundred microliters aliquots of the filtered extract (obtained following the procedure mentioned above) were made up to 10 ml with a mixture ethanol:HCl (85:15 v/v). The concentration of pigment was determined applying the Lambert–Beer law. The factor 98.2 is the molar absorptivity value for the acid–ethanol solvent (Fuleki & Francis, 1968b) and it refers to the absorption of a mixture of cranberry anthocyanins in acid–ethanol, measured in a 1-cm cell at 535 nm, at a concentration of 1% (w/v) ($E_{1\text{cm},535}^{1\%}$).

The spectra, recorded in a HP8452-UV diode array spectrophotometer (Hewlett–Packard) were measured at 25 °C and 535 nm, against the solvent. For that purpose 10-mm quartz cells were used.

The anthocyanin yield (g/kg) was calculated using the following equation:

$$\text{Total anthocyanins(TA)} = (A_{535} \cdot \text{dilution factor}) / E_{1\text{cm},535}^{1\%}$$

where A_{535} is the absorbance in the diluted sample and dilution factor, 20

2.5. Color coordinates

Aliquots of 500 μL of the extract were made up to 10 mL with distillate water in a volumetric flask. The spectral transmittances in the visible spectra (380–780 nm) were measured with a HP8452 UV diode array spectrophotometer (Hewlett–Packard) using cells with a path length of 10 mm. To obtain the tristimulus values, the weighted-ordinated method (constant intervals, $\Delta\lambda = 2$ nm) was applied, using as references the CIE Standard Illuminant D_{65} , the CIE 1964 Standard Observer, and water as reference blank

Following the most recent recommendations made by the CIE, CIELAB System (the variables related with psychometric color attributes: lightness, L^* , chroma, C_{ab}^* and hue, h_{ab}) for color specification was applied (CIE, 1986; Wyszecki & Stiles, 1982). An specific software was used for calculations: PCROM[®] (Álvarez-Moro & Heredia-Mira, 1994).

3. Results and discussion

A mixed 2 and 3 full factorial design ($2^1 \cdot 3^3$) was applied to study the effect of the different variables affecting the extraction of anthocyanins from Jaboticaba skin. The evaluated parameters were the extraction yield, defined as mass of anthocyanins/mass of jaboticaba skins (% w/w) for the different extraction conditions assayed, together with the chromatic parameters L^* , C_{ab}^* and h_{ab} . Table 1 list the levels of each factor and their coding, as well as the response for the different parameters measured in each experiment. The yield of anthocyanins was relatively high, being the mean value 92.92 ± 37.59 (43.99–163.29 mg/kg skin). Jaboticaba fruit could be considered a good source of anthocyanins when compared with other traditional sources as grapes, for which contents between 300 and 2000 mg/kg of berry have been reported (Arozarena et al., 2002; Ortega-Meder et al., 1994). In relation to the color properties of the different extracts, lightness (L^*) is the attribute related with the transmission of light observed, and, therefore to the concentration of the solution; it ranged from 71.4%, for the

darker solution, to 89.2% for the lighter one. Chroma (C_{ab}^*) is the quantitative component of chromaticity, and hue (h_{ab}) is its qualitative expression; chromaticity is the bidimensional parameter that correlates with the visual sensation attribute colorfulness. Hue (h_{ab}) values ranged from 13° to 24°. The main differences among the extracts were found in terms of chroma (C_{ab}^*) with values ranging from 22.8 to 54.3 CIELAB units, which indicate that the differences observed among the color characteristics of the different extracts of anthocyanins were mainly quantitative.

Correlations between the yield of anthocyanins and the color parameters were also explored. A high and significant ($p < 0.05$; $r = 0.84$) correlation was found between the yield of anthocyanins and chroma. This positive correlation with C_{ab}^* indicates that higher C_{ab}^* values are related to higher yields of anthocyanins. Also a negative and significant correlation ($p < 0.05$; $r = -0.85$) was found with L^* , indicating that higher L^* values are related to lower yields.

ANOVA analysis was applied in order to highlight the most influential parameters in the extraction process. To estimate the ANOVA parameters, the factor levels were internally recoded to test the linear (L) and quadratic (Q) components in the relationship between the factors and depended variables. Thus, regardless of the original metric of factor setting, the values were always recoded as $-1, 0, +1$ to perform the computation (Table 1). With the use of this transformation, each parameter is independent of the measure units. The linear effect was considered as the difference between the average responses obtained at the low and high settings for each factors. In the same way, the quadratic effect could be interpreted as the difference between the average response at the medium setting and the combined high and low settings for the respective factors. The dependent variables of interest considered were the yield of anthocyanins and the CIELAB color coordinates L^* , C_{ab}^* and h_{ab} . Table 2 shows the effect estimated for the different factors studied. As it can be observed, among the factors included in this study the type of acid showed the most significant ($p < 0.001$) effect on the yield of anthocyanins. Thus, the highest yields were obtained for the lowest setting (-1) of this factor: hydrochloric acid followed by citric acid and acetic acid. As for hydrochloric acid and citric acid, the best results were obtained for pH values between 3 and 4 units. In the case of acetic acid, pH 3 resulted in a higher yield of anthocyanins. In any case, the lowest yields obtained were those based on extracting mixtures at pH 2. Some authors (Anderson, Gueffroy, Webb, & Kepner, 1970) have reported that an excess of mineral acid can lead to hydrolysis of anthocyanins acetates during extraction, however acidic solvent are more efficient in extracting anthocyanins in comparison to neutral ones (Revilla et al., 1998).

Table 1

Factors, levels and coding of each factor in the experimental design and the responses to the different parameters obtained for Jaboticaba fruit at each of the extraction conditions assayed

Run	Solvent	Acid	Solvent (% v/v)	pH	Anthocyanin yield (g/kg)	<i>L</i> *	<i>C</i> _{ab} *	<i>h</i> _{ab}
1	Ethanol (-1)	HCl (-1)	100 (-1)	2 (-1)	0.14003	71.418	54.341	22.707
2	Ethanol (-1)	HCl (-1)	100 (-1)	3 (0)	0.13301	74.778	47.899	22.465
3	Ethanol (-1)	HCl (-1)	100 (-1)	4 (+1)	0.15052	75.772	45.578	22.619
4	Ethanol (-1)	HCl (-1)	90 (0)	2 (-1)	0.14133	74.611	50.586	19.871
5	Ethanol (-1)	HCl (-1)	90 (0)	3 (0)	0.13205	76.649	45.297	19.521
6	Ethanol (-1)	HCl (-1)	90 (0)	4 (+1)	0.12056	75.770	46.906	20.538
7	Ethanol (-1)	HCl (-1)	80 (+1)	2 (-1)	0.13276	75.937	47.642	17.988
8	Ethanol (-1)	HCl (-1)	80 (+1)	3 (0)	0.11550	76.517	45.193	18.143
9	Ethanol (-1)	HCl (-1)	80 (+1)	4 (+1)	0.14103	76.680	44.941	18.900
10	Methanol (+1)	HCl (-1)	100 (-1)	2 (-1)	0.13531	78.799	41.380	15.654
11	Methanol (+1)	HCl (-1)	100 (-1)	3 (0)	0.16329	75.244	47.914	18.220
12	Methanol (+1)	HCl (-1)	100 (-1)	4 (+1)	0.14840	75.146	46.174	18.191
13	Methanol (+1)	HCl (-1)	90 (0)	2 (-1)	0.15357	75.781	47.528	17.780
14	Methanol (+1)	HCl (-1)	90 (0)	3 (0)	0.15369	76.283	45.973	17.058
15	Methanol (+1)	HCl (-1)	90 (0)	4 (+1)	0.15027	77.591	43.739	16.319
16	Methanol (+1)	HCl (-1)	80 (+1)	2 (-1)	0.11191	78.468	42.886	16.453
17	Methanol (+1)	HCl (-1)	80 (+1)	3 (0)	0.13436	77.188	44.657	16.933
18	Methanol (+1)	HCl (-1)	80 (+1)	4 (+1)	0.11946	80.853	38.254	15.197
19	Ethanol (-1)	Citric (0)	100 (-1)	2 (-1)	-	-	-	-
20	Ethanol (-1)	Citric (0)	100 (-1)	3 (0)	0.07485	80.006	40.133	20.363
21	Ethanol (-1)	Citric (0)	100 (-1)	4 (+1)	0.07960	83.801	30.984	22.027
22	Ethanol (-1)	Citric (0)	90 (0)	2 (-1)	-	-	-	-
23	Ethanol (-1)	Citric (0)	90 (0)	3 (0)	0.07411	80.821	39.346	17.228
24	Ethanol (-1)	Citric (0)	90 (0)	4 (+1)	0.07302	83.881	30.909	19.321
25	Ethanol (-1)	Citric (0)	80 (+1)	2 (-1)	-	-	-	-
26	Ethanol (-1)	Citric (0)	80 (+1)	3 (0)	0.08020	84.260	34.121	15.431
27	Ethanol (-1)	Citric (0)	80 (+1)	4 (+1)	0.08253	84.309	34.139	15.489
28	Methanol (+1)	Citric (0)	100 (-1)	2 (-1)	-	-	-	-
29	Methanol (+1)	Citric (0)	100 (-1)	3 (0)	0.08430	80.666	39.160	17.003
30	Methanol (+1)	Citric (0)	100 (-1)	4 (+1)	0.08706	83.436	32.852	14.660
31	Methanol (+1)	Citric (0)	90 (0)	2 (-1)	-	-	-	-
32	Methanol (+1)	Citric (0)	90 (0)	3 (0)	0.08736	82.059	35.458	15.569
33	Methanol (+1)	Citric (0)	90 (0)	4 (+1)	0.08662	83.587	32.498	14.534
34	Methanol (+1)	Citric (0)	80 (+1)	2 (-1)	-	-	-	-
35	Methanol (+1)	Citric (0)	80 (+1)	3 (0)	0.07390	83.374	32.958	14.145
36	Methanol (+1)	Citric (0)	80 (+1)	4 (+1)	0.08021	83.313	31.212	14.024
37	Ethanol (-1)	Acetic (+1)	100 (-1)	2 (-1)	0.06805	77.3	41.4	24.6
38	Ethanol (-1)	Acetic (+1)	100 (-1)	3 (0)	0.06443	83.750	32.251	20.730
39	Ethanol (-1)	Acetic (+1)	100 (-1)	4 (+1)	0.05484	85.921	26.784	23.954
40	Ethanol (-1)	Acetic (+1)	90 (0)	2 (-1)	0.04632	81.090	38.998	17.543
41	Ethanol (-1)	Acetic (+1)	90 (0)	3 (0)	0.06070	84.684	29.731	18.949
42	Ethanol (-1)	Acetic (+1)	90 (0)	4 (+1)	0.06005	84.499	30.066	22.442
43	Ethanol (-1)	Acetic (+1)	80 (+1)	2 (-1)	0.05115	81.705	37.919	16.118
44	Ethanol (-1)	Acetic (+1)	80 (+1)	3 (0)	0.06138	81.757	36.424	18.859
45	Ethanol (-1)	Acetic (+1)	80 (+1)	4 (+1)	0.05425	85.093	29.023	18.003
46	Methanol (+1)	Acetic (+1)	100 (-1)	2 (-1)	0.04399	83.606	33.529	15.602
47	Methanol (+1)	Acetic (+1)	100 (-1)	3 (0)	0.06779	80.840	37.747	17.356
48	Methanol (+1)	Acetic (+1)	100 (-1)	4 (+1)	0.04922	86.412	26.121	16.070
49	Methanol (+1)	Acetic (+1)	90 (0)	2 (-1)	0.05343	89.025	22.787	12.970
50	Methanol (+1)	Acetic (+1)	90 (0)	3 (0)	0.05053	87.112	25.929	14.126
51	Methanol (+1)	Acetic (+1)	90 (0)	4 (+1)	0.06061	87.939	23.752	14.084
52	Methanol (+1)	Acetic (+1)	80 (+1)	2 (-1)	0.04536	83.465	33.975	14.747
53	Methanol (+1)	Acetic (+1)	80 (+1)	3 (0)	0.06428	84.320	31.837	14.305
54	Methanol (+1)	Acetic (+1)	80 (+1)	4 (+1)	0.06314	85.881	28.616	14.065

The use of methanol or ethanol as extracting solvents did not show a significant effect in the final yield. Metivier et al. (1980) reported that methanol was 20% more effective than ethanol and 73% more effective than water in recovering anthocyanins from grape

pomaces. According to final yield of anthocyanins, the efficiency of the procedure was more affected ($p < 0.05$) by the solvent concentration. In relation to the effect estimated, a slight reduction in the yield of the anthocyanins would be expected when setting the

Table 2
Linear (L) and quadratic (Q) effect estimated and ANOVA data for each parameter studied

Factors (effect)	Anthocyanin's yield (g/kg)	L^*	C_{ab}^*	h_{ab}
Solvent (L)	0.00	1.83 ^b	-3.55 ^b	-4.01 ^b
Acid (L)	-0.08 ^b	7.83 ^b	-14.44 ^b	-1.11 ^a
Acid (Q)	-0.02 ^b	1.94	-2.38	-1.43 ^a
% Solvent (L)	-0.00 ^a	1.75 ^a	-2.06	-3.40 ^b
% Solvent (Q)	0.00	0.64	-1.07	-0.47
pH (L)	0.01	2.41 ^a	-6.15 ^b	0.52
pH (Q)	0.00	-0.44	0.78	0.04

^a $p < 0.05$.

^b $p < 0.001$.

value for solvent concentration to its lowest (-1) setting (80% v/v).

The estimated effect for the color parameters are also shown in Table 2. The type of solvent and type of acid were the most significant factors ($p < 0.001$) affecting lightness (L^*). Acetic acid would improve L^* value by 7.83. In the same way, using methanol would improve by 1.83 the expected value for L^* . The highest lightness values were obtained when acetic acid and methanol were used as solvent mixture. Likewise, higher pH values significantly increased ($p < 0.05$) the expected value for L^* by 2.41. However, L^* is related with light transmission, so lower values for L^* are desirable, since they are related to more efficient anthocyanin extractions. This happened when the solvent mixture was based on ethanol acidify with chlorhidric acid. C_{ab}^* was significantly affected by the type of solvent ($p < 0.001$), type of acid ($p < 0.001$) and pH ($p < 0.001$). High C_{ab}^* values are desirable, since this parameter is the quantitative expression of colorfulness. According to the estimated effect calculated for C_{ab}^* , changing the setting of the factors type of solvent, type of acid and pH from the high (1) to the low (-1) setting would improve C_{ab}^* values by 3.55, 14.44 and 6.15, respectively.

On the other hand, h_{ab} values were significantly affected by the type of solvent ($p < 0.001$) and its concentration ($p < 0.001$) and, to a lower extend, by the type of acid ($p < 0.05$). All the estimated effects were negative, indicating that changing those factors from the high (+1) to low (-1) setting would improve the expected value for hue by 4.01, 1.11 and 3.4, respectively.

According to the quantitative extraction of anthocyanins from Jaboticaba and to the qualitative properties of the color of the extracted pigments, ethanol acidify with HCl and pH > 2 would be the best option to obtain a pigment with the highest C_{ab}^* and h_{ab} values, and the lowest L^* values.

Tristimulus Colorimetry allows a qualitative analysis of the extracts, which is an advantage over the quantitative method based in the measurement of absorbance at 535. When hue is considered, it is possible to establish the best extraction conditions to keep the blue-red color

of anthocyanins. This color is related with the structural integrity of petunidin 3-monoglucoside, the major anthocyanin present in Jaboticaba extracts.

References

- Álvarez-Moro, C., & Heredia-Mira, F. J. (1994). *PCROM: Cálculos colorimétricos*, v. 2.0. [SE-2374]. 1994. Sevilla: Registro de la Propiedad Intelectual.
- Anderson, D. W., Gueffroy, D. E., Webb, A. D., & Kepner, R. E. (1970). Identification of acetic acid as an acylating agent of anthocyanin pigments in grapes. *Phytochemistry*, 9(7), 1579.
- Arozarena, I., Ayestarán, B., Cantalejo, M. J., Navarro, M., Vera, M., Abril, I., et al. (2002). Anthocyanin composition of Tempranillo, Garnacha and Cabernet Sauvignon grapes from high- and low-quality vineyards over two years. *European Food Research and Technology*, 214(4), 303–309.
- CIE. (1986). *Colorimetry* (2nd ed.). Publication CIE No. 15.2, Vienna: CIE Central Bureau.
- Escribano-Bailón, T., Gutiérrez-Fernández, Y., Rivas-Gonzalo, J., & Santos-Buelga, C. (1992). Characterization of procyanidins of *Vitis vinifera* variety Tinta del País grape seeds. *Journal of Agricultural and Food Chemistry*, 40, 1794–1799.
- Francis, F. J. (1982). Analysis of anthocyanins. In P. Markakis (Ed.), *Anthocyanins as food colors* (pp. 1982). New York: Academic Press [Chapter 7].
- Francis, F. J. (1989). Food colorants: anthocyanins. *Critical Reviews in Food Science and Nutrition*, 28, 273–314.
- Fuleki, T., & Francis, F. J. (1968a). Quantitative methods for anthocyanins II. Determination of total anthocyanins and degradation index for cranberry. *Journal of Food Science*, 33, 78–83.
- Fuleki, T., & Francis, F. J. (1968b). Quantitative methods for anthocyanins I. Extraction and determination of total anthocyanins in cranberry. *Journal of Food Science*, 33, 72–76.
- Giusti, M. M., & Wrolstad, R. E. (1996). Characterization of red radish anthocyanins. *Journal of Food Science*, 61(2), 322–326.
- Gonnet, J. F., & Fenet, B. (2000). "Cyclamen red" colors based on a macrocyclic anthocyanin in carnation flowers. *Journal of Agricultural and Food Chemistry*, 48, 22–26.
- Hebrero, E., Santos-Buelga, C., & Rivas-Gonzalo, J. C. (1988). High performance liquid chromatography–diode array spectroscopy identification of anthocyanins of *Vitis vinifera* variety Tempranillo. *American Journal of Enology and Viticulture*, 39(3), 227–233.
- Heredia, F. J., Francia-Aricha, E., Rivas-Gonzalo, J. C., Vicario, I. M., & Santos-Buelga, C. (1998). Chromatic characterization of anthocyanins from red grapes-I. pH effect. *Food Chemistry*, 63, 491–498.
- Marakakis, P. (1982). *Anthocyanins as food colors*. London: Academic Press, pp. 248–250.
- Meléndez-Martínez, A. J., Vicario, I. M., & Heredia, F. J. (2003). Application of tristimulus colorimetry to estimate the carotenoids content in ultrafrozen orange juices. *Journal of Agricultural and Food Chemistry*, 51(25), 7266–7270.
- Méndez, J., González, M., Lobo, M. G., & Carnero, A. (2004). Color quality of pigments in cochineals (*Dactylopius coccus* Costa). Geographical origin characterization using multivariate statistical analysis. *Journal of Agricultural and Food Chemistry*, 52(5), 1331–1337.
- Metivier, R. P., Francis, F. J., & Clydesdale, F. M. (1980). Solvent-extraction of anthocyanins from wine pomace. *Journal of Food Science*, 45(4), 1099–1100.
- Ortega-Meder, M. D., Rivas-Gonzalo, J. C., Vicente, J. L., & Santos-Buelga, C. (1994). Diferenciación de variedades de uvas tintas por su composición antocianica. *Revista Española de Ciencia y Tecnología de los Alimentos*, 34(4), 409–426.

- Popenoe, W. (1974). *Manual of tropical and subtropical fruits*. Hafner Press.
- Revilla, E., Ryan, J. M., & Martín-Ortega, G. (1998). Comparison of several procedures used for the extraction of anthocyanins from red grapes. *Journal of Agricultural and Food Chemistry*, 46, 4592–4597.
- Rivas-Gonzalo, J. C., Bravo-Haro, S., & Santos-Buelga, C. (1995). Detection of compounds formed through the reaction of malvidin 3-monoglucoside and catechin in the presence of acetaldehyde. *Journal of Agricultural and Food Chemistry*, 43, 1444–1449.
- Somers, T. C., & Ziemelis, G. (1985). Spectral evaluation of total phenolic components in *Vitis-vinifera*-grapes and wines. *Journal of the Science of Food and Agriculture*, 36(12), 1275–1284.
- Statsoft. (1999). *Statistica for windows*.
- Trevisan, L. V., Bobbio, P. A., & Bobbio, F. O. (1972). Anthocyanins organic acids and carbohydrates of *Mirciaria jaboticaba*, Berg. *Journal of Food Science*, 37(6), 818–819.
- Wyszecki, G., & Stiles, W. S. (1982). *Color science. Concepts and methods. Quantitative data and formulae*. New York: Wiley, pp. 117–248.

ANEXO B
(Tabelas estatísticas)

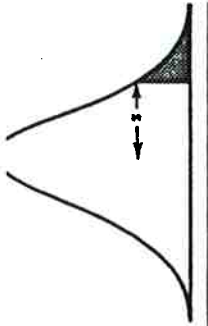


TABLE A. Tail area of unit normal distribution

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

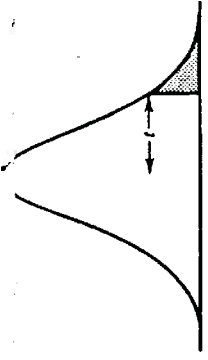


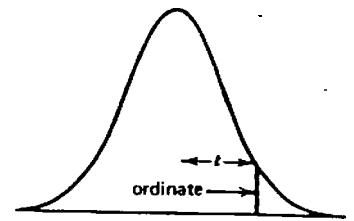
TABLE B1. Probability points of the t distribution with v degrees of freedom

v	tail area probability										
	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62	31.998
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.326	31.998	12.924
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924	8.610
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610	6.869
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869	5.959
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959	5.408
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408	5.041
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041	4.781
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781	4.587
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587	4.437
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437	4.318
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318	4.221
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221	4.140
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140	4.073
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073	4.015
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015	3.965
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965	3.922
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922	3.883
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883	3.850
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850	3.819
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819	3.792
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792	3.767
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767	3.745
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745	3.725
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725	3.707
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707	3.690
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690	3.674
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674	3.659
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659	3.646
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646	3.551
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551	3.460
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460	3.373
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373	3.291
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291	

Source: Taken with permission from E. S. Pearson and H. O. Hartley (Eds.) (1958). *Biometrika Tables for Statisticians*, Vol. 1, Cambridge University Press.

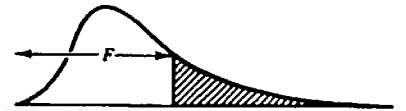
Parts of the table are also taken from Table III of Fisher and Yates: *Statistical Tables for Biological, Agricultural and Medical Research*, published by Longman Group Ltd., London (previously published by Oliver and Boyd, Edinburgh), by permission of the authors and publishers.

TABLE B2. Ordinates of t distribution with ν degrees of freedom



ν	value of t												
	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00
1	0.318	0.300	0.255	0.204	0.159	0.124	0.098	0.078	0.064	0.053	0.044	0.037	0.032
2	0.354	0.338	0.296	0.244	0.193	0.149	0.114	0.088	0.068	0.053	0.042	0.034	0.027
3	0.368	0.353	0.313	0.261	0.207	0.159	0.120	0.090	0.068	0.051	0.039	0.030	0.023
4	0.375	0.361	0.322	0.270	0.215	0.164	0.123	0.091	0.066	0.049	0.036	0.026	0.020
5	0.380	0.366	0.328	0.276	0.220	0.168	0.125	0.091	0.065	0.047	0.033	0.024	0.017
6	0.383	0.369	0.332	0.280	0.223	0.170	0.126	0.090	0.064	0.045	0.032	0.022	0.016
7	0.385	0.372	0.335	0.283	0.226	0.172	0.126	0.090	0.063	0.044	0.030	0.021	0.014
8	0.387	0.373	0.337	0.285	0.228	0.173	0.127	0.090	0.062	0.043	0.029	0.019	0.013
9	0.388	0.375	0.338	0.287	0.229	0.174	0.127	0.090	0.062	0.042	0.028	0.018	0.012
10	0.389	0.376	0.340	0.288	0.230	0.175	0.127	0.090	0.061	0.041	0.027	0.018	0.011
11	0.390	0.377	0.341	0.289	0.231	0.176	0.128	0.089	0.061	0.040	0.026	0.017	0.011
12	0.391	0.378	0.342	0.290	0.232	0.176	0.128	0.089	0.060	0.040	0.026	0.016	0.010
13	0.391	0.378	0.343	0.291	0.233	0.177	0.128	0.089	0.060	0.039	0.025	0.016	0.010
14	0.392	0.379	0.343	0.292	0.234	0.177	0.128	0.089	0.060	0.039	0.025	0.015	0.010
15	0.392	0.380	0.344	0.292	0.234	0.177	0.128	0.089	0.059	0.038	0.024	0.015	0.009
16	0.393	0.380	0.344	0.293	0.235	0.178	0.128	0.089	0.059	0.038	0.024	0.015	0.009
17	0.393	0.380	0.345	0.293	0.235	0.178	0.128	0.089	0.059	0.038	0.024	0.014	0.009
18	0.393	0.381	0.345	0.294	0.235	0.178	0.129	0.088	0.059	0.037	0.023	0.014	0.008
19	0.394	0.381	0.346	0.294	0.236	0.179	0.129	0.088	0.058	0.037	0.023	0.014	0.008
20	0.394	0.381	0.346	0.294	0.236	0.179	0.129	0.088	0.058	0.037	0.023	0.014	0.008
22	0.394	0.382	0.346	0.295	0.237	0.179	0.129	0.088	0.058	0.036	0.022	0.013	0.008
24	0.395	0.382	0.347	0.296	0.237	0.179	0.129	0.088	0.057	0.036	0.022	0.013	0.007
26	0.395	0.383	0.347	0.296	0.237	0.180	0.129	0.088	0.057	0.036	0.022	0.013	0.007
28	0.395	0.383	0.348	0.296	0.238	0.180	0.129	0.088	0.057	0.036	0.021	0.012	0.007
30	0.396	0.383	0.348	0.297	0.238	0.180	0.129	0.088	0.057	0.035	0.021	0.012	0.007
35	0.396	0.384	0.348	0.297	0.239	0.180	0.129	0.088	0.056	0.035	0.021	0.012	0.006
40	0.396	0.384	0.349	0.298	0.239	0.181	0.129	0.087	0.056	0.035	0.020	0.011	0.006
45	0.397	0.384	0.349	0.298	0.239	0.181	0.129	0.087	0.056	0.034	0.020	0.011	0.006
50	0.397	0.385	0.350	0.298	0.240	0.181	0.129	0.087	0.056	0.034	0.020	0.011	0.006
∞	0.399	0.387	0.352	0.301	0.242	0.183	0.130	0.086	0.054	0.032	0.018	0.009	0.004

TABLE D. Percentage points of the F distribution: upper 25% points



$v_1 \backslash v_2$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32	9.41	9.49	9.58	9.63	9.67	9.71	9.76	9.80	9.85
2	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.41	3.43	3.43	3.44	3.45	3.46	3.47	3.48
3	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.46	2.46	2.46	2.47	2.47	2.47	2.47	2.47
4	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
5	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.88	1.88	1.88	1.88	1.87	1.87	1.87
6	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.77	1.76	1.76	1.75	1.75	1.75	1.74	1.74	1.74
7	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.69	1.69	1.68	1.68	1.67	1.67	1.66	1.66	1.65	1.65	1.65
8	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.63	1.62	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.58
9	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.55	1.54	1.54	1.53	1.53	1.53
10	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46
11	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43
12	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41
13	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.40	1.39
14	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37
15	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46	1.45	1.44	1.43	1.41	1.41	1.40	1.39	1.38	1.37	1.36
16	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33
17	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33
18	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32
19	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.33	1.32	1.30
20	1.40	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31	1.29
21	1.40	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28
22	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.37	1.36	1.34	1.33	1.32	1.31	1.30	1.29	1.28
23	1.39	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.29	1.27
24	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.36	1.35	1.33	1.32	1.31	1.30	1.29	1.28	1.26
25	1.39	1.47	1.46	1.44	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.32	1.31	1.29	1.28	1.27	1.25
26	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.35	1.34	1.32	1.31	1.30	1.29	1.28	1.26	1.25
27	1.38	1.46	1.45	1.43	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.33	1.32	1.31	1.30	1.28	1.27	1.26	1.24
28	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.31	1.30	1.29	1.27	1.26	1.25	1.23
29	1.38	1.45	1.45	1.43	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.32	1.31	1.30	1.29	1.27	1.26	1.25	1.23
30	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.34	1.32	1.30	1.29	1.28	1.27	1.26	1.24	1.23
40	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.31	1.30	1.28	1.26	1.25	1.24	1.22	1.21	1.19
60	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.27	1.25	1.24	1.22	1.21	1.19	1.17	1.15
120	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.26	1.24	1.22	1.21	1.19	1.18	1.16	1.13	1.10
∞	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.22	1.19	1.18	1.16	1.14	1.12	1.08	1.00

Source: M. Merrington and C. M. Thompson (1943), Tables of percentage points of the inverted beta (F) distribution. *Biometrika*, 33, 73. Used by permission.

TABLE D (continued). Percentage points of the F distribution: upper 10% points

$v_1 \backslash v_2$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67						

TABLE D (continued). Percentage points of the F distribution: upper 5% points

v_1 v_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.11	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

TABLE D (continued). Percentage points of the F distribution: upper 1% points

v_1 v_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999.50	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
4																			

TABLE D (continued), Percentage points of the F distribution: upper 0.1% points

v_1 v_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4053*	5000*	5404*	5625*	5764*	5859*	5929*	5981*	6023*	6056*	6107*	6158*	6209*	6235*	6261*	6287*	6313*	6340*	6366*
2	998.5	999.0	999.2	999.2	999.3	999.3	999.4	999.4	999.4	999.4	999.4	999.4	999.4	999.4	999.5	999.5	999.5	999.5	999.5
3	167.0	148.5	141.1	137.1	134.6	132.8	131.6	130.6	129.9	129.2	128.3	127.4	126.4	125.9	125.4	125.0	124.5	124.0	123.5
4	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47	48.05	47.41	46.76	46.10	45.77	45.43	45.09	44.75	44.40	44.05
5	47.18	37.12	33.20	31.09	29.75	28.84	28.16	27.64	27.24	26.92	26.42	25.91	25.39	25.14	24.87	24.60	24.33	24.06	23.79
6	35.51	27.00	23.70	21.92	20.81	20.03	19.46	19.03	18.69	18.41	17.99	17.56	17.12	16.89	16.67	16.44	16.21	15.99	15.75
7	29.25	21.69	18.77	17.19	16.21	15.52	15.02	14.63	14.33	14.08	13.71	13.32	12.93	12.73	12.53	12.33	12.12	11.91	11.70
8	25.42	18.49	15.83	14.39	13.49	12.86	12.40	12.04	11.77	11.54	11.19	10.84	10.48	10.30	10.11	9.92	9.73	9.53	9.33
9	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11	9.89	9.57	9.24	8.90	8.72	8.55	8.37	8.19	8.00	7.81
10	21.04	14.91	12.55	11.28	10.48	9.92	9.52	9.20	8.96	8.75	8.45	8.13	7.80	7.64	7.47	7.30	7.12	6.94	6.76
11	19.69	13.81	11.56	10.35	9.58	9.05	8.66	8.35	8.12	7.92	7.63	7.32	7.01	6.85	6.68	6.52	6.35	6.17	6.00
12	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48	7.29	7.00	6.71	6.40	6.25	6.09	5.93	5.76	5.59	5.42
13	17.81	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98	6.80	6.52	6.23	5.93	5.78	5.63	5.47	5.30	5.14	4.97
14	17.14	11.78	9.73	8.62	7.92	7.43	7.08	6.80	6.58	6.40	6.13	5.85	5.56	5.41	5.25	5.10	4.94	4.77	4.60
15	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	5.81	5.54	5.25	5.10	4.95	4.80	4.64	4.47	4.31
16	16.12	10.97	9.00	7.94	7.27	6.81	6.46	6.19	5.98	5.81	5.55	5.27	4.99	4.85	4.70	4.54	4.39	4.23	4.06
17	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75	5.58	5.32	5.05	4.78	4.63	4.48	4.33	4.18	4.02	3.85
18	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56	5.39	5.13	4.87	4.59	4.45	4.30	4.15	4.00	3.84	3.67
19	15.08	10.16	8.28	7.26	6.62	6.18	5.85	5.59	5.39	5.22	4.97	4.70	4.43	4.29	4.14	3.99	3.84	3.68	3.51
20	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.82	4.56	4.29	4.15	4.00	3.86	3.70	3.54	3.38
21	14.59	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11	4.95	4.70	4.44	4.17	4.03	3.88	3.74	3.58	3.42	3.26
22	14.38	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99	4.83	4.58	4.33	4.06	3.92	3.78	3.63	3.48	3.32	3.15
23	14.19	9.47	7.67	6.69	6.08	5.65	5.33	5.09	4.89	4.73	4.48	4.23	3.96	3.82	3.68	3.53	3.38	3.22	3.05
24	14.03	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80	4.64	4.39	4.14	3.87	3.74	3.59	3.45	3.29	3.14	2.97
25	13.88	9.22	7.45	6.49	5.88	5.46	5.15	4.91	4.71	4.56	4.31	4.06	3.79	3.66	3.52	3.37	3.22	3.06	2.89
26	13.74	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.64	4.48	4.24	3.99	3.72	3.59	3.44	3.30	3.15	2.99	2.82
27	13.61	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.57	4.41	4.17	3.92	3.66	3.52	3.38	3.23	3.08	2.92	2.75
28	13.50	8.93	7.19	6.25	5.66	5.24	4.93	4.69	4.50	4.35	4.11	3.86	3.60	3.46	3.32	3.18	3.02	2.86	2.69
29	13.39	8.85	7.12	6.19	5.59	5.18	4.87	4.64	4.45	4.29	4.05	3.80	3.54	3.41	3.27	3.12	2.97	2.81	2.64
30	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	4.00	3.75	3.49	3.36	3.22	3.07	2.92	2.76	2.59
40	12.61	8.25	6.60	5.70	5.13	4.73	4.44	4.21	4.02	3.87	3.64	3.40	3.15	3.01	2.87	2.73	2.57	2.41	2.23
60	11.97	7.76	6.17	5.31	4.76	4.37	4.09	3.87	3.69	3.54	3.31	3.08	2.83	2.69	2.55	2.41	2.25	2.08	1.89
120	11.38	7.32	5.79	4.95	4.42	4.04	3.77	3.55	3.38	3.24	3.02	2.78	2.53	2.40	2.26	2.11	1.95	1.76	1.54
∞	10.83	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10	2.96	2.74	2.51	2.27	2.13	1.99	1.84	1.66	1.45	1.00

* Multiply these entries by 100.

