



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA
DEPARTAMENTO DE FÍSICO-QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

Tese de Doutorado

OTIMIZAÇÃO DE PROCESSOS CONTENDO VARIÁVEIS DE
MISTURA PELO MÉTODO "SPLIT-PLOT"

João Alexandre Bortoloti

Orientador: Roy Edward Bruns

Campinas /SP
2006

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE
QUÍMICA DA UNICAMP

B648o Bortoloti, João Alexandre.
Otimização de processos contendo variáveis de
mistura pelo método "split-plot" / João Alexandre
Bortoloti. -- Campinas, SP: [s.n], 2006.

Orientador: Roy Edward Bruns.

Tese - Universidade Estadual de Campinas, Instituto
de Química.

1. *Split-plot*. 2. Quimiometria. 3. Otimização.
4. Análise de variância. I. Bruns, Roy Edward.
II. Universidade Estadual de Campinas. Instituto de
Química. III. Título.

Título em inglês: Split-plot process optimization containing mixture variables

Palavras-chaves em inglês: Split-plot, Chemometrics, Optimization, Variance analysis

Área de concentração: Físico-Química

Titulação: Doutor em Ciências

Banca examinadora: Roy Edward Bruns (orientador), Benicio de Barros Neto, Ieda Spacino Scarminio, Susanne Rath, Márcia Miguel Castro Ferreira, Ronei Jesus Poppi, Maria Isabel Rodrigues

Data de defesa: 22/06/2006

Aos meus pais Valdemar (*in memoriam*) e
Altina, a minha esposa Daniela, e minha
filha Laura.

Agradecimentos

Ao Prof. Roy E. Bruns que orientou esta tese e contribuiu muito para a minha formação.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela bolsa concedida.

Aos colegas de trabalho e amigos.

CURRICULUM VITAE

João Alexandre Bortoloti

Mestrado em Físico-Química

Ênfase em Quimiometria e Estatística para otimização de processos contendo variáveis de mistura. Universidade Estadual de Campinas (1999 - 2001)

Licenciatura em Química

Universidade Estadual de Campinas (1999 - 2002)

Bacharel em Química com atribuições Tecnológicas

Universidade Estadual de Campinas (1995 – 1998)

Técnico em Bioquímica

Escola Técnica Estadual Conselheiro Antônio Prado (1991 – 1994)

Experiência profissional

Professor de Estatística II e Matemática no Curso de Administração – ESAMC / Campinas

Professor de Técnicas de Análise Ambiental no Curso de Tecnologia em Gestão Ambiental – Universidade São Marcos / Paulínia

Professor em Cursos de Extensão do Instituto de Matemática e Estatística – UNICAMP / Curso de Extensão para especialização de professores formados na área de matemática e estatística (2001)

Desenvolvimento de projeto no Centro de pesquisa e desenvolvimento da TELEBRÁS “Ensaio Mecânicos com pultrudados de cabos ópticos ” / (CPqD - Campinas) - 1998

Estágio no Centro de Pesquisa da RHODIA – Paulínia / “ Formulações POUR-ON ” - 1994

Apresentação de trabalhos em Congressos

12º Encontro Nacional de Química Analítica / São Luis - 2003. “Aplicação de gráficos de probabilidade acumulada em planejamentos “split-plot” com e sem replicatas”

12º Encontro Nacional de Química Analítica / São Luis - 2003. “Determinação de chumbo por ASV em sistema ternário homogêneo de solventes pelo método split-plot”

11º Encontro Nacional de Química Analítica / Campinas - 2001. “Estudos para a diminuição de experimentos em planejamentos split-plot”

25º Reunião Anual da Sociedade Brasileira de Química. “ANOVA para planejamentos incompletos split-plot”

28º Reunião Anual da Sociedade Brasileira de Química. “A importância da aleatoriedade na execução de experimentos em química”

Artigos publicados

Bortoloti J.A., Andrade J.C., Bruns R.E. *J. Braz. Chem. Soc.* **2004**, 15, 241.

Bortoloti J.A., Bruns R.E., Andrade J.C., Vieira R.K. *Chemom. Intell. Lab. Syst.* **2004**, 70, 113.

Bortoloti J.A., Borges C.N., Bruns R.E. *Anal. Chim. Acta* **2005**, 544, 206.

Bortoloti J.A., Bruns R.E. *Química Nova* **2006** (no prelo).

Produção de Software

Otimização multivariada pelo método Split-plot – FORTRAN90 / MATLAB®

Resumo

O método “split-plot” é uma importante ferramenta para a otimização simultânea de sistemas contendo variáveis de processo e de mistura. Porém sua aplicação a problemas químicos ainda foi pouco explorada. Este método simplifica o trabalho no laboratório, mas o tratamento dos resultados é mais complexo. Neste trabalho foram realizados estudos de forma minuciosa sobre suas etapas, desde a elaboração do planejamento até a validação dos resultados. Pontos como a ANOVA (análise de variância), e métodos como ML (maximum likelihood), OLS (ordinary least squares) e REML (restricted maximum likelihood) foram discutidos e empregados em estudos comparativos com dados reais e simulados. Com o objetivo de permitir a popularização do método “split-plot” foram realizados estudos que permitissem a diminuição do número de experimentos executados. Também foram criados programas computacionais para a realização dos cálculos necessários, assim como a análise gráfica dos resultados. Um programa foi gerado para ser executado em ambiente Windows, enquanto outro foi desenvolvido para trabalhar em Matlab com grande flexibilidade para adaptações, ambos os programas estão disponíveis à comunidade. Os programas criados foram aplicados a estudos com dados reais e simulados, seus resultados foram comparados com programas como SAS e R.

O método “split-plot” também foi empregado em uma otimização conjunta das condições dos reagentes e solventes na determinação de Pb^{2+} por ASV (anodic stripping voltammetry). Três componentes de mistura, N,N-dimetilformamida (DMF), etanol e água, e o nível de duas variáveis de processo, acetato de amônio (eletrólito de suporte) e a concentração de ácido clorídrico, foram variados. Os cálculos das somas quadráticas da regressão e falta de ajuste da ANOVA para o “main-plot”, “sub-plot”, e da interação “main-sub-plot” são apresentados. Estes valores se mostram úteis para o desenvolvimento dos modelos.

A determinação dos graus de liberdade necessários para a validação dos modelos em planejamentos “split-plot” é feita de forma aproximada. Assim foram

empregados gráficos de probabilidade acumulada em vários estudos, para a validação dos modelos. Além disto, os gráficos de probabilidade acumulada também foram utilizados em estudos para reduzir o número de experimentos em planejamentos “split-plot”. O estudo foi realizado com três conjuntos de dados de três planejamentos “split-plot” da literatura: o primeiro conjunto se refere à otimização conjunta de três componentes de mistura de plasticida em diferentes condições de velocidade de extrusão e temperatura de secagem, o segundo, a otimização do preparo de croquete de peixe utilizando três ingredientes e diferentes tempos e temperatura de cozimento e fritura, e o terceiro a determinação catalítica de Cr (VI) empregando três reagentes de concentração variável e diferentes proporções de três solventes. Com o procedimento sugerido foram obtidos modelos que foram comparados com aqueles determinados utilizando-se a ANOVA com planejamentos “split-plot” completos.

Um método de simulação específico foi desenvolvido para casos de interesse. Isto possibilitou compreender como as diferentes fontes de erro afetam os componentes de variância e até mesmo os termos dos modelos ajustados por regressão. Vários conjuntos de dados foram simulados incluindo planejamentos fatoriais cuja simplicidade poderia estimular o emprego do método “split-plot”. Com a compreensão e domínio das técnicas de simulação novas aplicações e perspectivas foram geradas, ampliando as possibilidades do emprego real do método “split-plot”.

Abstract

The split-plot method is an important tool for the simultaneous optimization of systems affected by process and mixture variables. However its application to chemical problems is still little explored. This method permits simplification of laboratory work but the statistical treatment of the data is more difficult than for conventional methods. In this thesis a detailed study of each step of the split-plot method, from the design elaborating the experiments to be executed until the statistical validation of the final results, is reported. The ANOVA (analysis of variance) is described and the ML (maximum likelihood), OLS (ordinary least square) and REML (restricted maximum likelihood) methods are discussed and applied in comparative studies with real and simulated data. With the objective of showing the potential of the split-plot method studies were carried out that permit a reduction in the number of experiments that must be executed. Furthermore computer programs were created to execute the necessary calculations as well as to graphically display the results. One program was generated to be executed with Windows whereas another, with greater flexibility for user adaptations, was developed to work in the Matlab framework. The academic community can access both programs. These programs were tested using real and simulated data and their results compared with the SAS and R reference programs.

A split-plot design has been also used to simultaneously optimize reagent conditions and solvent medium for Pb^{2+} determination by anodic stripping voltammetry (ASV). Three mixture components, N,N-dimethylformamide (DMF), ethanol and water, and two process variable levels, ammonium acetate (supporting electrolyte) and hydrochloric acid concentrations, were varied. The calculations of main-plot, sub-plot and main-sub-plot interaction ANOVA sums of squares for regression and lack of fit are illustrated. These values are shown to be useful for model development.

The determination of the degrees of freedom necessary for the validation of the models in split-plot design is approximate. Graphs of cumulative probability

were used in several studies, for the validation of the models. Also the graphs cumulative probability graphs were also used in studies to reduce the number of experiments in split-plot design. The study was carried with three data sets of split-plot designs reported in the literature: the first refers to the simultaneous optimization of three plasticizer mixture components with different extrusion rates and drying temperatures, the second, three fish pattie ingredients at different cooking and frying temperature and times and, the third, Cr(VI) catalytic determinations employing three reagents of varying concentrations and three solvent components of varying proportions. Approximate models determined from the proposed procedure are compared with those determined using complete split-plot ANOVA analyses.

A simulation method was developed for certain interesting situations. The simulations permit a better understanding of how the different error sources affect the variance components and also the model coefficients and their standard errors. Several data sets for simple factorial designs executed according to the split-plot method were simulated. These simulations, besides increasing our understanding of the split-plot method, are useful in suggesting real applications of the split-plot method.

ÍNDICE

LISTA DE TABELAS	xxv
LISTA DE FIGURAS	xxxi
1 Introdução	1
2 Objetivos do projeto de doutorado	3
3 Metodologia	5
3.1 Planejamentos e modelos contendo somente variáveis de processo	5
3.2 Planejamentos e modelos de misturas	6
3.3 O Método Split-plot	8
3.3.1 Modelos e detalhamento da ANOVA no método split-plot	14
3.3.2 Falta de ajuste em modelos para planejamentos split-plot	18
3.3.3 ANOVA detalhada para planejamentos split-plot com número elevado de experimentos.	23
4 Estrutura dos planejamentos	29
4.1 Procedimentos para análise de dados	30
4.2 Componentes de variância	30
4.2.1. Dados Balanceados	31
4.2.2. Dados Desbalanceados	31
4.3 O método de máxima verossimilhança	32
4.4 O método de mínimos quadrados	35
4.5 O método de mínimos quadrados por máxima verossimilhança	38
4.6 O método da máxima verossimilhança restrita – REML	41
4.7 Comparação dos métodos ML, OLS e REML	41
5 Estudos com planejamentos incompletos	47
5.1 Estudos com planejamentos contendo duplicatas de misturas sorteadas por condição de processo (I)	47
5.1.1 Resultados do planejamento incompleto I	49
5.2 Estudos com planejamentos contendo o número completo de misturas, porém com o número de duplicatas reduzidas (II).	52
5.2.1 Resultados do planejamento incompleto II	53
6 Gráficos de Probabilidade Acumulada em planejamentos split-plot	59

6.1 Modelos aproximados para planejamentos split-plot	61
6.2 Aplicação de gráficos de probabilidade acumulada em estudos envolvendo dados reais	62
7 APLICAÇÃO E COMPARAÇÃO DO MÉTODO SPLIT-PLOT UTILIZANDO DADOS EXPERIMENTAIS	87
7.1 Análise de dados experimentais com tratamento split-plot e completamente aleatório.	93
7.2 Utilização do teste F como ferramenta de apoio para a escolha de parâmetros em modelos split-plot.	106
8 Simulações em planejamentos split-plot	111
9 Considerações finais	141
10 Conclusões	143
11 Estudos futuros	143
Referências bibliográficas	145

LISTA DE TABELAS

TABELA 1. FORMATO ANOVA PARA O MODELO CONVENCIONAL “SPLIT-PLOT” DA EQUAÇÃO 6.....	13
TABELA 2. ANOVA PARA O MODELO “SPLIT-PLOT” DA EQUAÇÃO 7. ⁵	14
TABELA 3. ANOVA PARA MODELO “SPLIT-PLOT”. ⁵	15
TABELA 4. FORMATO DETALHADO DA ANOVA PARA PLANEJAMENTOS QUE PERMITAM FALTA DE AJUSTE, ONDE A E B SÃO OS NÚMEROS DE PARÂMETROS DO MODELO REFERENTE AS VARIÁVEIS DE PROCESSO E DE MISTURA RESPECTIVAMENTE.....	16
TABELA 5. DADOS DE UMA PLANEJAMENTO “SPLIT-PLOT” COM DUAS VARIÁVEIS DE PROCESSO (Z_1 E Z_2) E TRÊS DE MISTURA (X_1 , X_2 E X_3) ^A	21
TABELA 6 . ANOVA DETALHADA PARA O MODELO DA EQUAÇÃO 13 AJUSTADO.	22
TABELA 7. RESULTADOS NÃO DETALHADOS DA ANÁLISE DE VARIÂNCIA OBTIDOS POR MEIO DAS EQUAÇÕES EQUAÇÃO 16 A EQUAÇÃO 21.	26
TABELA 8. PLANEJAMENTO FATORIAL PELO MÉTODO “SPLIT-PLOT” PARA OTIMIZAR A RESISTÊNCIA DE UM PLÁSTICO. OS DADOS SÃO COMPLETAMENTE BALANCEADOS.	43
TABELA 9. COEFICIENTES E ERROS DOS PARÂMETROS AJUSTADOS NO MODELO BILINEAR UTILIZANDO-SE OLS.....	44
TABELA 10. COEFICIENTES E ERROS DOS PARÂMETROS AJUSTADOS NO MODELO BILINEAR UTILIZANDO-SE ML E REML.....	44
TABELA 11. PLANEJAMENTO FATORIAL PELO MÉTODO “SPLIT-PLOT” PARA OTIMIZAR A RESISTÊNCIA DE UM PLÁSTICO. OS DADOS SÃO DESBALANCEADOS.	45
TABELA 12. COEFICIENTES E ERROS DOS PARÂMETROS AJUSTADOS NO MODELO BILINEAR UTILIZANDO-SE OLS, ML E REML PARA OS DADOS DESBALANCEADOS DA	46
TABELA 13. ANOVA PARA O PLANEJAMENTO I COM 8 COMPOSIÇÕES DE MISTURA POR CONDIÇÃO DE PROCESSO.	50
TABELA 14. ANOVA PARA O PLANEJAMENTO COMPLETO.....	50
TABELA 15. ANOVA PARA O PLANEJAMENTO I COM 3, 4, 5 E 6 REPLICATAS.....	52
TABELA 16. ANOVA PARA O PLANEJAMENTO II COM 4 DUPLICATAS E SEIS MEDIDAS SEM REPLICATAS POR CONDIÇÃO DE PROCESSO.....	54
TABELA 17. ANOVA PARA OS CONJUNTOS DE DADOS COM 3, 5, 6, 7, 8 E 9 DUPLICATAS NO PLANEJAMENTO II.....	56
TABELA 18. PLANEJAMENTO FATORIAL 2 ² COM ACETATO DE AMÔNIO E ÁCIDO CLORÍDRICO COMO VARIÁVEIS DE PROCESSO E DMF, ETANOL E ÁGUA COMO VARIÁVEIS DE MISTURA EM UM PLANEJAMENTO “SPLIT- PLOT”.	64
TABELA 19. DETERMINAÇÃO EXPERIMENTAL DE CHUMBO PARA 40 EXPERIMENTOS EM DUPLICATAS EM UM PLANEJAMENTO “SPLIT-PLOT” COM DUAS VARIÁVEIS DE PROCESSO E TRÊS DE MISTURAS. A RESPOSTA ANALÍTICA É DADA COMO VALORES DE CORRENTE DE PICO EM NA.....	65

TABELA 20. TABELA DE ANOVA “SPLIT-PLOT” PARA O MODELO LINEAR PARA VARIÁVEIS DE PROCESSO COMBINADO COM OS MODELOS LINEAR, QUADRÁTICO E ESPECIAL CÚBICO PARA VARIÁVEIS DE MISTURA.	66
TABELA 21. TABELA DE ANOVA “SPLIT-PLOT” PARA O MODELO BILINEAR PARA VARIÁVEIS DE PROCESSO COMBINADO COM OS MODELOS LINEAR, QUADRÁTICO E ESPECIAL CÚBICO PARA VARIÁVEIS DE MISTURA.	67
TABELA 22. PARÂMETROS, ESTIMATIVAS DO ERRO PADRÃO E RAZÃO PARA O TESTE t PARA O MODELO COMBINADO BILINEAR-CÚBICO-ESPECIAL.	69
TABELA 23 . RAZÕES COEFICIENTES / ERRO PADRÃO PARA VARIAÇÕES DE 50% NAS VARIÂNCIAS DAS REPLICATAS, “MAIN-PLOT” E “SUB-PLOT” PARA UM PLANEJAMENTO FATORIAL $2^{3.A}$	78
TABELA 24. ANOVA “SPLIT-PLOT” INCLUINDO AS SOMAS QUADRÁTICAS DE REGRESSÃO E FALTA DE AJUSTE DOS DADOS DO Cr(VI) DA REFERÊNCIA 36.	82
TABELA 25. COEFICIENTES DO MODELO PARA A DETERMINAÇÃO CATALÍTICA DE Cr(VI) USANDO REPLICATAS INDIVIDUAIS E DE FORMA CONJUNTA. PARÂMETROS EM NEGRITO SÃO SIGNIFICATIVOS NO NÍVEL DE 95% DE CONFIANÇA.	85
TABELA 26. COEFICIENTES E ERRO PADRÃO DOS PARÂMETROS AJUSTADOS NO MODELO BILINEAR. OS PARÂMETROS SIGNIFICATIVOS A 95% ESTÃO EM NEGRITO.	88
TABELA 27. ANOVA “SPLIT-PLOT” PARA OS DADOS DA TABELA 8.	89
TABELA 28. MODELO COM 31 PARÂMETROS AJUSTADO AOS DADOS DA TABELA 8.	91
TABELA 29. SOMAS QUADRÁTICAS REFERENTES AO MODELO DA EQUAÇÃO 31 PARA O MÉTODO “SPLIT-PLOT”.	91
TABELA 30. COEFICIENTES E ERROS PARA O MODELO “SPLIT-PLOT”.	92
TABELA 31. PLANEJAMENTO COM DUAS VARIÁVEIS DE PROCESSO, VELOCIDADE DE EXTRUSÃO (Z_1) E TEMPERATURA DE SECAGEM (Z_2) CUJOS NÍVEIS SÃO FIXADOS, FORMANDO O “MAIN-PLOT”, E TRÊS VARIÁVEIS DE MISTURA, X_1 , X_2 E X_3 , QUE CORRESPONDEM A DIFERENTES PROPORÇÕES DE PLASTIFICANTES ALEATORIZADAS NOS “SUB-PLOTS”.	94
TABELA 32. ANÁLISE DE VARIÂNCIA “SPLIT-PLOT” PARA OS DADOS DA TABELA 31.	95
TABELA 33. MODELO AJUSTADO AOS DADOS DA TABELA 31 E ERROS ASSOCIADOS AOS PARÂMETROS.	95
TABELA 34. MODELO AJUSTADO AOS DADOS DA TABELA 31 CONSIDERANDO OS EXPERIMENTOS REALIZADOS DE FORMA COMPLETAMENTE ALEATÓRIA.	99
TABELA 35. PLANEJAMENTO COM UM CONJUNTO DE DADOS EXTRAÍDO DA REFERÊNCIA 5. AS VARIÁVEIS DE MISTURA SÃO X_1 , X_2 E X_3 E AS DE PROCESSO Z_1 E Z_2 . VEJA REFERÊNCIA 5.	101
TABELA 36. ANOVA “SPLIT-PLOT” PARA OS DADOS DA TABELA 35	101
TABELA 37. MODELO LINEAR-BILINEAR AJUSTADO AOS DADOS DA TABELA 35. OS PARÂMETROS SIGNIFICATIVOS ESTÃO EM NEGRITO	102
TABELA 38. MODELO LINEAR-BILINEAR AJUSTADO AOS DADOS DA TABELA 35 CONSIDERANDO OS DADOS PROVENIENTES DE EXPERIMENTOS COMPLETAMENTE ALEATÓRIOS.	103

TABELA 39. PLANEJAMENTO “SPLIT-PLOT” UTILIZADO NA OTIMIZAÇÃO DA PRODUÇÃO DE VINIL. AS VARIÁVEIS DE MISTURA SÃO X_1 , X_2 E X_3 E AS DE PROCESSO Z_1 E Z_2 . VEJA REFERÊNCIA 5.	104
TABELA 40. MODELOS AJUSTADOS PELA ANÁLISE “SPLIT-PLOT” E COMPLETAMENTE ALEATÓRIA. OS PARÂMETROS EM NEGRITO SÃO CONSIDERADOS SIGNIFICATIVOS.	105
TABELA 41. AUMENTO NA SOMA QUADRÁTICA DEVIDA A REGRESSÃO (SQ_{REG}) PELA ADIÇÃO DE TERMOS NO MODELO AJUSTADO AOS DADOS DA TABELA 35.	108
TABELA 42. VALORES DE F PARA AS MÉDIAS QUADRÁTICAS (MQ) DOS PARÂMETROS DO MODELO LINEAR-BILINEAR. OS PARÂMETROS SIGNIFICATIVOS ESTÃO EM NEGRITO.	109
TABELA 43. ANOVA PARA OS DADOS DA TABELA 39 COM O MODELO LINEAR-BILINEAR CONTENDO O TERMO X_1X_2 AJUSTADO. AS FALTAS DE AJUSTE DO MODELO FORAM ADICIONADAS AO ERRO “SUB-PLOT”. ...	110
TABELA 44. SOMAS QUADRÁTICAS DOS PARÂMETROS AJUSTADOS POR REGRESSÃO. OS VALORES EM NEGRITO SÃO CONSIDERADOS SIGNIFICATIVOS A 95% DE CONFIANÇA.	110
TABELA 45. ANOVA PARA O MODELO “SPLIT-PLOT”.	112

LISTA DE FIGURAS

FIGURA 1. PLANEJAMENTO FATORIAL 2^3 . ¹	6
FIGURA 2. PLANEJAMENTO PARA MISTURAS FORMADAS POR TRÊS COMPONENTES, X_1 , X_2 E X_3	7
FIGURA 3. PLANEJAMENTOS PARA OTIMIZAR VARIÁVEIS DE PROCESSO E DE MISTURA SIMULTANEAMENTE ⁵	9
FIGURA 4. PLANEJAMENTO "SPLIT-PLOT" PARA TRÊS VARIÁVEIS DE PROCESSO EM DOIS NÍVEIS CADA E TRÊS VARIÁVEIS DE MISTURA FORMANDO DEZ COMPOSIÇÕES DIFERENTES PERFAZENDO 160 EXPERIMENTOS EM DUPLICATA.....	10
FIGURA 5 . DECOMPOSIÇÃO DO DESVIO DE UMA OBSERVAÇÃO EM RELAÇÃO À MÉDIA GLOBAL, $(y_i - \bar{y})$, NA SOMA DAS PARCELAS $(\hat{y}_i - \bar{y})$ E $(y_i - \hat{y}_i)$. ¹	17
FIGURA 6. DIVISÃO DA SOMA QUADRÁTICA DO "SUB-PLOT" (X) EM DUAS PARCELAS REFERENTES À REGRESSÃO (XREG) E FALTA DE AJUSTE (XLOF).	24
FIGURA 7. CONJUNTO DE PONTOS EXPERIMENTAIS {X,Y}	35
FIGURA 8. ILUSTRAÇÃO DO PLANEJAMENTO FATORIAL PELO MÉTODO "SPLIT-PLOT" PARA QUATRO VARIÁVEIS, (A) TEMPERATURA, (B) PORCENTAGEM DE ADITIVO, (C) VELOCIDADE DE AGITAÇÃO E (D) TEMPO DE PROCESSAMENTO.....	42
FIGURA 9: PLANEJAMENTO "SPLIT-PLOT" COM VARIÁVEIS DE PROCESSO REPRESENTADAS PELAS ARESTAS DO CUBO E EM SEUS VÉRTICES AS COMBINAÇÕES DE MISTURAS UTILIZADAS ³⁶	48
FIGURA 10 . REPRESENTAÇÃO DE UMA PLANEJAMENTO COMPLETO (A) E DE UM PLANEJAMENTO INCOMPLETO (B), ONDE "●" É UMA COMPOSIÇÃO DE MISTURA UTILIZADA E "○" É UMA COMPOSIÇÃO DE MISTURA NÃO UTILIZADA.	48
FIGURA 11. PLANEJAMENTO COMPLETO (A) E PLANEJAMENTO INCOMPLETO (B), ONDE "●" REPRESENTA COMPOSIÇÃO EM DUPLICATA E "○" COMPOSIÇÃO SEM DUPLICATA.	53
FIGURA 12. MÉDIA QUADRÁTICA DOS TERMOS DA ANOVA DA TABELA 7 VERSUS O NÚMERO DE REPLICATAS POR PLANEJAMENTO PARA O ESTUDO II.....	57
FIGURA 13. MÉDIA QUADRÁTICA DOS TERMOS DA ANOVA DA TABELA 7 VERSUS O NÚMERO DE REPLICATAS POR PLANEJAMENTO PARA O ESTUDO II COM ESCALA AMPLIADA.	57
FIGURA 14. PARÂMETROS DOS MODELOS VERSUS NÚMERO DE REPLICATAS POR PLANEJAMENTOS NO ESTUDO II.....	58
FIGURA 15. GRÁFICO DE PROBABILIDADE ACUMULADA PARA EFEITOS SIGNIFICATIVOS (●) E NÃO SIGNIFICATIVOS (○).	60
FIGURA 16. PLANEJAMENTO EXPERIMENTAL PARA DETERMINAÇÃO DE CHUMBO COM 3 COMPONENTES DE MISTURA EM UM ARRANJO FATORIAL 2^2 PARA VARIÁVEIS DE PROCESSO.	63
FIGURA 17. GRÁFICO NORMAL DE PROBABILIDADE COM VALORES DO TESTE-T DA TABELA 22. OS CÍRCULOS SÓLIDOS CORRESPONDEM A VALORES DE T MAIORES QUE 3 E OS TRIÂNGULOS SÓLIDOS CORRESPONDEM A VALORES DE T ENTRE 2 E 3.	68
FIGURA 18. GRÁFICO DE VALORES PREVISTOS PELA EQUAÇÃO 40 VERSUS VALORES EXPERIMENTAIS.....	70

FIGURA 19. SUPERFÍCIE DE RESPOSTA DAS MISTURAS PARA CADA CONDIÇÃO DETERMINADA PELAS VARIÁVEIS DE PROCESSO.	73
FIGURA 20. GRÁFICO DE PROBABILIDADE ACUMULADA DOS COEFICIENTES/ERRO PADRÃO ASSUMINDO-SE VARIÂNCIA UNITÁRIA PARA AS PRIMEIRAS REPLICATAS DE EXPERIMENTOS.	75
FIGURA 21. GRÁFICO DE PROBABILIDADE ACUMULADA PARA A RAZÃO COEFICIENTES / ERRO PADRÃO PARA O SEGUNDO GRUPO DE REPLICATAS DA REFERÊNCIA 5.	76
FIGURA 22. GRÁFICO DE PROBABILIDADE ACUMULADA PARA A RAZÃO COEFICIENTE / ERRO PADRÃO PARA DADOS COM AMBAS AS REPLICATAS. UTILIZOU-SE AS ESTIMATIVAS DE VARIÂNCIA FORNECIDAS PELA ANOVA.	77
FIGURA 23. SUPERPOSIÇÃO DOS GRÁFICOS DE PROBABILIDADE ACUMULADA PARA UM PLANEJAMENTO FATORIAL 2^3 VARIANDO OS VALORES DAS ESTIMATIVAS DAS VARIÂNCIAS DAS REPLICATAS, DO “MAIN-PLOT” E DO “SUB-PLOT”.	79
FIGURA 24. GRÁFICO DE PROBABILIDADE ACUMULADA DOS COEFICIENTES DO MODELO PRESENTE NA REFERÊNCIA 7.	80
FIGURA 25. GRÁFICO DE PROBABILIDADE ACUMULADA DA RAZÃO COEFICIENTES/ERRO PADRÃO DO MODELO FORNECIDO PELA REFERÊNCIA 7.	81
FIGURA 26. GRÁFICO DE PROBABILIDADE ACUMULADA PARA A RAZÃO COEFICIENTE/ERRO PADRÃO DE TODO O CONJUNTO DE DADOS DO Cr(VI). OS COEFICIENTES SIGNIFICATIVOS SÃO INDICADOS POR CÍRCULOS PREENCHIDOS.	83
FIGURA 27. GRÁFICO DE PROBABILIDADE ACUMULADA PARA A RAZÃO COEFICIENTES/ERRO PADRÃO PARA UMA REPLICATA DO CONJUNTO DE DADOS DO Cr(VI). ÀS VARIÂNCIAS DAS REPLICATAS, “MAIN-PLOT” E “SUB-PLOT” ATRIBUIU-SE O VALOR 1. O GRÁFICO UTILIZANDO O SEGUNDO CONJUNTO DE REPLICATAS É IDÊNTICO A ESTE. OS COEFICIENTES SIGNIFICATIVOS SÃO INDICADOS POR CÍRCULOS PREENCHIDOS.	84
FIGURA 28. GRÁFICO DE PROBABILIDADE ACUMULADA PARA OS COEFICIENTES DA TABELA 26. COMO OS ERROS SÃO TODOS IGUAIS NÃO HÁ DIFERENÇA EM UTILIZAR OS VALORES DOS COEFICIENTES OU A RAZÃO COEFICIENTE / ERRO NA ABSCISSA.	88
FIGURA 29. GRÁFICO DE PROBABILIDADE ACUMULADA PARA OS VALORES DAS RAZÕES COEFICIENTES / ERRO PADRÃO DA	92
FIGURA 30. GRÁFICO DE RESÍDUOS PARA O MODELO DESCRITO NA	96
FIGURA 31. HISTOGRAMA PARA A DISTRIBUIÇÃO DOS COEFICIENTES DO MODELO DESCRITO NA	97
FIGURA 32. RESPOSTAS PREVISTAS X RESPOSTA OBSERVADA PARA O MODELO DESCRITO NA	97
FIGURA 33. GRÁFICO DE PROBABILIDADE PARA VALORES DA RAZÃO COEFICIENTES/ERRO PADRÃO DA	98
FIGURA 34. GRÁFICO DE PROBABILIDADE ACUMULADA PARA OS VALORES DA RAZÃO COEFICIENTES/ERRO PADRÃO PRESENTES NA TABELA 34.	100
FIGURA 35. GRÁFICO DE PROBABILIDADE ACUMULADA COM OS VALORES OBTIDOS DA RAZÃO COEFICIENTES / ERRO PADRÃO CONTIDOS NA TABELA 37.	103

FIGURA 36. (A) COMPONENTE DE VARIÂNCIA PARA A SIMULAÇÃO CONTENDO 10000 REPLICATAS POR EXPERIMENTO EM QUE O ERRO PADRÃO DAS REPLICATAS FOI ELEVADO DE 0,5 A 5,0; (B) VALORES DOS DESVIOS PADRÃO.....	113
FIGURA 37. COEFICIENTES DOS PARÂMETROS DO MODELO OBTIDO POR REGRESSÃO LINEAR PELO MÉTODO DE MÍNIMOS QUADRADOS GENERALIZADOS PARA A SIMULAÇÃO CONTENDO 10000 REPLICATAS POR EXPERIMENTO EM QUE O ERRO PADRÃO DAS REPLICATAS FOI ELEVADO DE 0,5 A 5,0.....	114
FIGURA 38. RAZÃO COEFICIENTES/ERRO PADRÃO (T) PARA A SIMULAÇÃO CONTENDO 10000 REPLICATAS POR EXPERIMENTO EM QUE O ERRO PADRÃO DAS REPLICATAS FOI ELEVADO DE 0,5 A 5,0.....	116
FIGURA 39. VALORES DOS COMPONENTES DE VARIÂNCIA PARA A SIMULAÇÃO EM QUE O ERRO PADRÃO DAS REPLICATAS FOI ELEVADO DE 0,5 A 5,0, A, B, C, D, E, F REPRESENTAM SIMULAÇÕES REALIZADAS COM 2, 4, 8, 16, 32 E 64 REPLICATAS, RESPECTIVAMENTE, POR EXPERIMENTO.....	117
FIGURA 40. VALORES DE T PARA SIMULAÇÕES EM QUE O ERRO PADRÃO DAS REPLICATAS FOI ELEVADO DE 0,5 A 5,0; A, B, C, D, E, F REPRESENTAM SIMULAÇÕES REALIZADAS COM 2, 4, 8, 16, 32 E 64 REPLICATAS, RESPECTIVAMENTE, POR EXPERIMENTO.	118
FIGURA 41. GRÁFICOS DE PROBABILIDADE PARA A SIMULAÇÃO CONTENDO 10000 REPLICATAS POR EXPERIMENTO EM QUE O ERRO PADRÃO DAS REPLICATAS FOI ELEVADO DE 0,5 A 5,0 ATRAVÉS DE 50 ETAPAS. O GRÁFICO (A) INDICA A ETAPA 1 DA SIMULAÇÃO, O (B) A ETAPA 20, (C) E (D) INDICAM AS ETAPAS 40 E 50 DAS SIMULAÇÕES, RESPECTIVAMENTE.....	119
FIGURA 42. (A) COMPONENTE DE VARIÂNCIA PARA A SIMULAÇÃO CONTENDO 10000 REPLICATAS POR EXPERIMENTO EM QUE O ERRO PADRÃO DO “MAIN-PLOT” FOI ELEVADO DE 0,5 A 5,0 ATRAVÉS DE 50 ETAPAS, (B) VALORES DOS DESVIOS PADRÃO DOS COMPONENTES.	121
FIGURA 43. COEFICIENTES DOS PARÂMETROS DO MODELO OBTIDO POR REGRESSÃO LINEAR PELO MÉTODO DE MÍNIMOS QUADRADOS GENERALIZADOS PARA A SIMULAÇÃO CONTENDO 10000 REPLICATAS POR EXPERIMENTO. O ERRO PADRÃO DO “MAIN-PLOT” FOI ELEVADO DE 0,5 A 5,0 ATRAVÉS DE 50 ETAPAS.	121
FIGURA 44. VALORES DE T PARA OS MODELOS OBTIDOS POR REGRESSÃO LINEAR PELO MÉTODO DE MÍNIMOS QUADRADOS GENERALIZADOS PARA A SIMULAÇÃO CONTENDO 10000 REPLICATAS POR EXPERIMENTO. O ERRO PADRÃO DO “MAIN-PLOT” FOI ELEVADO DE 0,5 A 5,0.....	122
FIGURA 45. GRÁFICOS DE PROBABILIDADE PARA A SIMULAÇÃO CONTENDO 10000 REPLICATAS POR EXPERIMENTO EM QUE O ERRO PADRÃO DO “MAIN-PLOT” FOI ELEVADO DE 0,5 A 5,0. O GRÁFICO (A) INDICA A ETAPA 1 DA SIMULAÇÃO, O (B) A ETAPA 20, (C) E (D) INDICAM AS ETAPAS 40 E 50 DAS SIMULAÇÕES, RESPECTIVAMENTE.	123
FIGURA 46. (A) COMPONENTE DE VARIÂNCIA PARA A SIMULAÇÃO CONTENDO 10000 REPLICATAS POR EXPERIMENTO EM QUE O ERRO PADRÃO DO “SUB-PLOT” FOI ELEVADO DE 0,5 A 5,0, (B) VALORES DOS DESVIOS PADRÃO.....	124

FIGURA 47. GRÁFICOS DE PROBABILIDADE PARA A SIMULAÇÃO CONTENDO 10000 REPLICATAS POR EXPERIMENTO EM QUE O ERRO PADRÃO DO “SUB-PLOT” FOI ELEVADO DE 0,5 A 5,0. O GRÁFICO (A) INDICA A ETAPA 1 DA SIMULAÇÃO, O (B) A ETAPA 20, (C) E (D) INDICAM AS ETAPAS 40 E 50 DAS SIMULAÇÕES, RESPECTIVAMENTE.	125
FIGURA 48. VALORES DE T PARA OS MODELOS OBTIDOS POR REGRESSÃO LINEAR PELO MÉTODO DE MÍNIMOS QUADRADOS GENERALIZADOS PARA A SIMULAÇÃO CONTENDO 10000 REPLICATAS POR EXPERIMENTO EM QUE O ERRO PADRÃO DO “SUB-PLOT” FOI ELEVADO DE 0,5 A 5,0.	126
FIGURA 49. GRÁFICOS COM OS VALORES DOS COMPONENTES DE VARIÂNCIA DA REPLICATA, ERRO “MAIN-PLOT” E ERRO “SUB-PLOT”, A, B, C, D, E, F REPRESENTAM SIMULAÇÕES REALIZADAS COM 2, 4, 8, 16, 32 E 64 REPLICATAS, RESPECTIVAMENTE, POR EXPERIMENTO.....	128
FIGURA 50. COEFICIENTES DOS PARÂMETROS DOS MODELOS OBTIDOS POR REGRESSÃO LINEAR PARA SIMULAÇÕES CONTENDO 2, 4, 8, 16, 32 E 64 REPLICATAS POR EXPERIMENTO, REPRESENTADOS PELAS FIGURAS A, B, C, D, E, F RESPECTIVAMENTE. O ERRO PADRÃO DO “SUB-PLOT” FOI ELEVADO DE 0,5 A 5,0.	130
FIGURA 51. VALORES DE T PARA OS COEFICIENTES DOS MODELOS OBTIDOS POR REGRESSÃO LINEAR PARA A SIMULAÇÃO CONTENDO 2, 4, 8, 16, 32 E 64 REPLICATAS POR EXPERIMENTO, QUE CORRESPONDEM RESPECTIVAMENTE A A, B, C, D, E, F NA FIGURA. O ERRO PADRÃO DO “MAIN-PLOT” FOI ELEVADO DE 0,5 A 5,0.	131
FIGURA 52. GRÁFICOS DE PROBABILIDADE ACUMULADA PARA SIMULAÇÕES COM EXPERIMENTOS CONTENDO 10000 REPLICATAS. O ERRO PADRÃO DAS REPLICATAS FOI ELEVADO DE 0,5 ATÉ 5,0 EM 50 ETAPAS. O ERRO PADRÃO DO “MAIN-PLOT” E DO “SUB-PLOT” FOI MANTIDO EM 2,0. AS FIGURAS APRESENTAM NO TOPO A ETAPA QUE INDICAM DA SIMULAÇÃO: 1, 11 E 21.....	134
FIGURA 53. VALORES DE T PARA OS COEFICIENTES DOS PARÂMETROS DOS MODELOS GERADOS AO LONGO DO PROCESSO DE SIMULAÇÃO EM QUE O ERRO PADRÃO DAS REPLICATAS FOI ELEVADO DE 0,5 ATÉ 5,0. O ERRO PADRÃO DO “MAIN-PLOT” E DO “SUB-PLOT” FOI MANTIDO EM 2,0. AS SETAS INDICAM A MIGRAÇÃO DOS VALORES DE T EM DIREÇÃO AO ZERO NA MEDIDA EM QUE O ERRO DAS REPLICATAS É ELEVADO. OS DADOS FORAM TRATADOS COMO PROVENIENTES DE EXPERIMENTOS COMPLETAMENTE ALEATÓRIOS.....	135
FIGURA 54. VALORES DE T PARA OS COEFICIENTES DOS PARÂMETROS DOS MODELOS GERADOS AO LONGO DO PROCESSO DE SIMULAÇÃO EM QUE O ERRO PADRÃO DAS REPLICATAS FOI ELEVADO DE 0,5 ATÉ 5,0. O ERRO PADRÃO DO “MAIN-PLOT” E DO “SUB-PLOT” FOI MANTIDO EM 2,0. OS DADOS FORAM TRATADOS PELA ESTRATÉGIA “SPLIT-PLOT”.....	136
FIGURA 55. VALORES DE T PARA OS COEFICIENTES DOS PARÂMETROS DOS MODELOS GERADOS AO LONGO DO PROCESSO DE SIMULAÇÃO EM QUE O ERRO PADRÃO DO “MAIN-PLOT” FOI ELEVADO DE 0,5 ATÉ 5,0. O ERRO PADRÃO DAS REPLICATAS E DO “SUB-PLOT” FOI MANTIDO EM 2,0. AS SETAS INDICAM A MIGRAÇÃO DOS VALORES DE T EM DIREÇÃO AO ZERO NA MEDIDA EM QUE O ERRO DO “MAIN-PLOT” É ELEVADO. OS DADOS FORAM TRATADOS COMO PROVENIENTES DE EXPERIMENTOS COMPLETAMENTE ALEATÓRIOS.....	137

FIGURA 56. VALORES DE T PARA OS COEFICIENTES DOS PARÂMETROS DOS MODELOS GERADOS AO LONGO DO PROCESSO DE SIMULAÇÃO EM QUE O ERRO PADRÃO DO “MAIN-PLOT” FOI ELEVADO DE 0,5 ATÉ 5,0. O ERRO PADRÃO DAS REPLICATAS E DO “SUB-PLOT” FOI MANTIDO EM 2,0. OS DADOS FORAM TRATADOS PELA ESTRATÉGIA “SPLIT-PLOT”.....	138
FIGURA 57. VALORES DE T PARA OS COEFICIENTES DOS PARÂMETROS DOS MODELOS GERADOS AO LONGO DO PROCESSO DE SIMULAÇÃO EM QUE O ERRO PADRÃO DO “SUB-PLOT” FOI ELEVADO DE 0,5 ATÉ 5,0. O ERRO PADRÃO DAS REPLICATAS E DO “MAIN-PLOT” FOI MANTIDO EM 2,0. AS SETAS INDICAM A MIGRAÇÃO DOS VALORES DE T EM DIREÇÃO AO ZERO NA MEDIDA EM QUE O ERRO DAS REPLICATAS É ELEVADO. OS DADOS FORAM TRATADOS COMO PROVENIENTES DE EXPERIMENTOS COMPLETAMENTE ALEATÓRIOS.....	139
FIGURA 58. VALORES DE T PARA OS COEFICIENTES DOS PARÂMETROS DOS MODELOS GERADOS AO LONGO DO PROCESSO DE SIMULAÇÃO EM QUE O ERRO PADRÃO DO “SUB-PLOT” FOI ELEVADO DE 0,5 ATÉ 5,0. O ERRO PADRÃO DAS REPLICATAS E DO “MAIN-PLOT” FOI MANTIDO EM 2,0. AS SETAS INDICAM A MIGRAÇÃO DOS VALORES DE T EM DIREÇÃO AO ZERO NA MEDIDA EM QUE O ERRO DO SUB É ELEVADO. OS DADOS FORAM TRATADOS COMO PROVENIENTES DE EXPERIMENTOS DE UM PLANEJAMENTO “SPLIT-PLOT”.....	140

1 INTRODUÇÃO

Os métodos multivariados de otimização estão sendo cada vez mais aplicados em laboratórios químicos. Na indústria a utilização destes procedimentos já está se tornando comum no controle de qualidade e otimização de processos. Trabalhos no âmbito universitário aplicando estes métodos são encontrados na literatura científica com maior frequência.¹

Métodos multivariados de otimização baseados em conceitos estatísticos estão tendo bastante sucesso por várias razões. Primeiro, interações entre fatores somente podem ser descobertas usando-se estratégias multivariadas. O método clássico é univariado onde cada fator é otimizado separadamente dos outros. Experimentos são feitos mantendo-se todos os fatores constantes, variando somente o valor (nível) de um fator. Após isto os níveis dos outros fatores são variados da mesma maneira, um de cada vez, até o melhor resultado ser obtido. Este procedimento não é eficiente porque o valor otimizado de um fator depende dos valores dos outros fatores.^{1,2} Segundo, os parâmetros calculados para modelos multivariados são mais precisos do que as medidas individuais usadas para determinar o modelo. Esta constatação é baseada no teorema do limite central da estatística, o mesmo teorema que comprova que o erro no valor médio é menor do que o erro de uma observação individual. Como os parâmetros do modelo multivariado chamados efeitos são diferenças entre médias, estes são mais precisos do que as observações individuais.^{2,3}

Terceiro, planejamentos multivariados economizam experimentos¹. Otimizações são conseguidas com menos tempo, utilizando menos material e de uma maneira bem mais segura. O pesquisador pode sistematizar seu trabalho usando métodos multivariados de forma bem mais objetiva do que usando métodos convencionais de otimização. Isto acontece porque métodos multivariados tratam todos os efeitos, a serem otimizados, com a mesma importância eliminando possíveis preconceitos errados por razão de nossa intuição química que nem sempre é correta.

Em resumo métodos multivariados de otimização têm maior aplicabilidade, menor custo e maior precisão do que métodos univariados. Além disto, às vantagens de usar métodos multivariados em relação ao univariado aumentam quando o número de fatores a serem otimizados aumenta⁴.

Em geral existem dois tipos de variáveis ou fatores para serem otimizados. Um tipo de fator, chamado fator de processo, permite ajustes de qualquer fator independentemente dos valores dos outros fatores. A resposta ou resultado da otimização depende dos valores absolutos dos fatores empregados. O outro tipo de fator, chamado de variável de mistura, não pode ser ajustado independentemente dos outros fatores estudados, porque a resposta depende das proporções empregadas destes fatores.

Aplicações multivariadas que otimizam somente os valores de variáveis (fatores) de processos ou somente valores de variáveis de misturas estão ficando relativamente comuns em laboratórios de química e engenharia química. Mas estudos ajustando os dois tipos de fatores simultaneamente são bastante raros, especialmente em estudos químicos. Se existirem efeitos de interação entre variáveis de processo e variáveis de mistura as condições de otimização dificilmente serão descobertas usando procedimentos restritos a um tipo de variável. Por isto métodos gerais tratando os dois tipos de variáveis são importantes e planejamentos experimentais envolvendo ensaios em que todas as variáveis do processo e de mistura são ajustadas simultaneamente são necessários para achar as condições otimizadas para sistemas complexos^{5,6}.

2 OBJETIVOS DO PROJETO DE DOUTORADO

- 1) Desenvolver um software para o método “split-plot” para estimular sua aplicação, permitindo entre outras coisas a realização de planejamentos mais complexos;
- 2) Desenvolver o software para o método “split-plot” em MATLAB para permitir adequações a novas situações de forma mais flexível;
- 3) Aplicar outros softwares estatísticos como o SAS^a, Statistica^b e o R^c em estudos comparativos com o programa desenvolvido;
- 4) Desenvolver a parte gráfica de softwares criados para que possam ser executados em ambiente Windows, tornando-os mais amigáveis aos usuários em geral;
- 5) Investigar como erros provocados por incertezas nos valores das variáveis afetam a qualidade dos resultados da otimização envolvendo variáveis de processo e de mistura;
- 6) Determinar os graus de liberdade envolvidos na análise de variância detalhada do método “split-plot”;
- 7) Avaliar a aplicação de gráficos de probabilidade como ferramenta de análise de planejamento para o método “split-plot”, com o objetivo de diminuir o número de experimentos;
- 8) Investigar a importância da aleatorização dos ensaios em problemas de otimização com variáveis de processo e de mistura;
- 9) Realizar estudos para diminuir o número de experimentos em planejamentos “split-plot”;
- 10) Usar os programas desenvolvidos em problemas com dados reais;
- 11) Realizar simulações para auxiliar na melhor compreensão do método “split-plot” em situações com diferentes erros.

^a SAS Institute (2001); SAS User’s Guide, Ver. 8.02. SAS Institute Inc., USA.

^b Statistica for Windows, Version 6.0, Statsoft Inc., Tulsa, OK, USA.

^c R: A language and environment for statistical computing, Foundation for Statistical Computing, Vienna, Austria Ver 2.1.1, <http://www.R-project.org/>

3 METODOLOGIA

3.1 Planejamentos e modelos contendo somente variáveis de processo

O planejamento fatorial é o mais básico usado em otimização multivariada. Ele permite determinar os efeitos principais e de interação entre as variáveis de processo ⁷. Em geral o modelo pode ser escrito como ⁵

$$\text{resposta} = \alpha_0 + \sum_{i=1}^n \alpha_i z_i + \sum_{i < j} \sum_{j=1}^n \alpha_{ij} z_i z_j$$

Equação 1

onde n é número de variáveis, α o coeficiente dos efeitos z . Um planejamento fatorial de dois níveis é adequado para determinar os coeficientes α deste modelo bilinear. Se existirem k fatores o planejamento fatorial é um fatorial 2^k . No caso de três fatores ou variáveis os experimentos a serem executados podem ser representados nos vértices de um cubo como mostrado na Figura 1. Os parâmetros α podem ser calculados por meio de regressão linear múltipla.

Os experimentos são realizados em ordem aleatória para minimizar os efeitos de qualquer erro sistemático nos valores dos efeitos, e permitir uma possível estimação de erro sem fazer replicatas ^{8,9}.

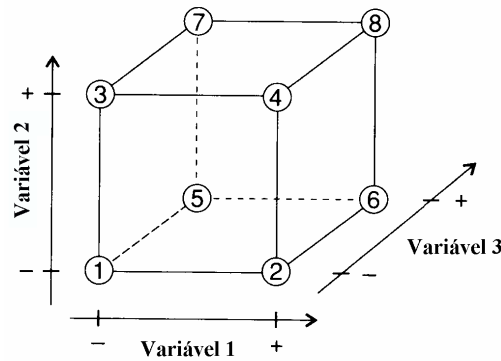


Figura 1. Planejamento fatorial 2^3 .¹

3.2 Planejamentos e modelos de misturas

Os planejamentos e modelos de mistura são caracterizados pelas condições $x_i \geq 0$, $i = 1, 2 \dots q$, onde x_i é a proporção do i -ésimo componente e havendo q componentes na mistura, $x_1 + x_2 + \dots + x_q = 1$.^{5,10}

Pela última relação pode se ver que uma mudança no valor de qualquer x_i provoca mudanças nos outros valores, x_j . Os modelos mais usados para modelar misturas são os de primeira ordem (linear).

$$\text{resposta} = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_q x_q = \sum_{i=1}^q \beta_i x_i ,$$

Equação 2

segunda ordem

$$\text{resposta} = \sum_{i=1}^q \beta_i x_i + \sum_{i < j} \sum \beta_{ij} x_i x_j ,$$

Equação 3

e o modelo cúbico especial

$$y = \sum_{i=1}^q \beta_i x_i + \sum_{i < j} \sum \beta_{ij} x_i x_j + \sum_{i < j < k} \sum \sum \beta_{ijk} x_i x_j x_k$$

Equação 4

os β 's são os coeficientes populacionais dos efeitos, sendo que β_{ij} e β_{ijk} representam coeficientes de efeitos de interação.

O planejamento que pode ser usado para resolver qualquer destes modelos está ilustrado na Figura 2.¹¹ Os valores das respostas obtidas para as misturas indicadas nesta figura são regressados sobre os valores das proporções x_i para obter os valores das estimativas, os b 's, dos parâmetros do modelo de mistura, os β 's. Como foi feito para o planejamento fatorial os ensaios com as diferentes misturas devem ser realizados em ordem aleatória. Nota-se que o modelo de mistura não tem um termo constante como existe na Equação 1 para um modelo envolvendo variáveis de processo.

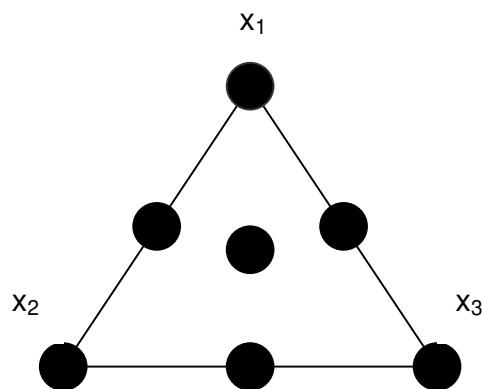


Figura 2. Planejamento para misturas formadas por três componentes, x_1 , x_2 e x_3 .

3.3 O Método Split-plot

Os planejamentos ilustrados na Figura 3 são chamados planejamentos estatísticos “split-plot”. Eles são usados quando se deseja otimizar variáveis de processo e de mistura simultaneamente.^{5,12} Neste exemplo com 3 variáveis de processo e três de mistura, têm-se $7 \times 8 = 56$ experimentos a serem executados, supondo que repetições não sejam feitas. Seguindo critérios estatísticos todos os experimentos devem ser feitos em ordem aleatória¹³. Neste caso a análise de variância convencional pode ser feita para avaliar os resultados da regressão.

Mas a execução em ordem aleatória de todos os ensaios pode apresentar problemas operacionais^{14,15} porque variáveis de processo e de mistura estão sendo ajustadas simultaneamente.^{5,16,17} Os problemas de otimização envolvendo variáveis dificilmente ajustáveis são bastante comuns em química. O método “split-plot” é ainda pouco encontrado em trabalhos nacionais, os casos mais freqüentes ocorrem em estudos sobre experimentos em agricultura^{18,19,20,21}. Muitos trabalhos encontrados na literatura tratam das peculiaridades matemáticas do método “split-plot” para situações específicas de planejamento^{22,23,24}, mas ainda há uma carência de material direcionado para pesquisadores interessados no método, mas sem formação estatística. Na literatura foram encontrados trabalhos que aplicam o método “split-plot”, mas o tratamento dos dados de forma geral se restringe apenas a análise de variância dos resultados, sem a geração de modelos que justifiquem porque alguns os fatores contribuem de forma significativa para a variância observada. Em aplicações químicas a geração dos modelos é normalmente de extrema importância.

Considere-se a simplificação na rotina de trabalho que é inserida pela Figura 3a. Neste caso são testadas sete misturas em blocos para as oito possíveis combinações das variáveis de processo indicado pelo planejamento fatorial 2^3 . Em lugar de fazer todos os ensaios em ordem aleatória as sete diferentes misturas são testadas simultaneamente para cada uma das combinações das variáveis de processo indicadas pelos vértices do cubo. Somente estas combinações de variáveis de processo são executadas aleatoriamente. Caso as mudanças nos

níveis das variáveis de processo sejam difíceis ou muito demoradas todas as formulações podem ser preparadas e testadas em paralelo resultando em economia considerável de tempo e esforço²⁵. Uma alternativa está ilustrada na Figura 3b. Neste caso as oito combinações das variáveis são feitas em blocos e as sete diferentes formulações de misturas são executadas aleatoriamente.

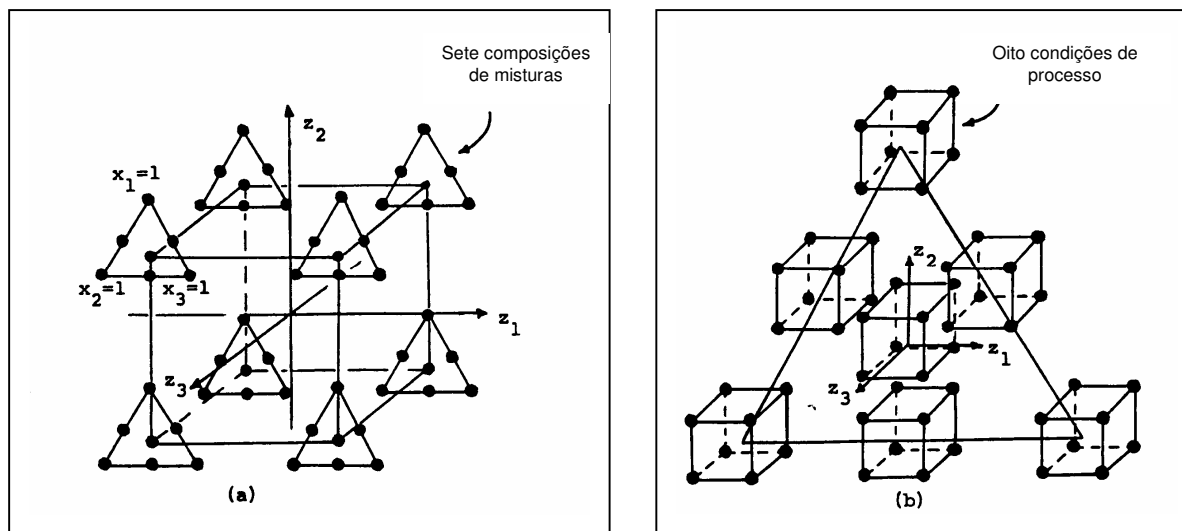


Figura 3. Planejamentos para otimizar variáveis de processo e de mistura simultaneamente⁵

Realizando-se as estratégias sugeridas pela Figura 3 há um custo: a análise estatística será mais complexa e um procedimento chamado análise “split-plot” precisa ser realizado. Trabalho no laboratório pode ser poupado, mas somente um maior conhecimento de estatística irá garantir que a qualidade dos resultados se mantenha^{4,26}.

Planejamentos estatísticos “split-plot” ainda são pouco usados em química uma vez que eles requerem a execução de um grande número de experimentos⁴⁻²⁶. O seu uso envolve um conjunto de experimentos para normalmente investigar um sistema seguido por um bloco de réplicas idênticas para estimar o erro experimental. Contudo a maioria dos químicos prefere executar uma quantidade maior de experimentos na região de interesse e limitar o número de replicatas.

Planejamentos “split-plot” na verdade simplificam o trabalho de laboratório uma vez que exigências de aleatorização completa são relaxadas, por outro lado, a análise de variância é muito mais complexa que a convencional. Em muitos casos encontrados na literatura os planejamentos possuem variáveis de processo e mistura organizadas em “main-plots” ou “sub-plots” dependendo do enfoque do estudo realizado. Um exemplo deste procedimento é ilustrado na Figura 4 onde há três variáveis de processo, z_1 , z_2 e z_3 , em dois níveis e três variáveis de mistura, x_1 , x_2 e x_3 , formando 10 composições diferentes. Ao todo são realizados 160 experimentos em duplicata.

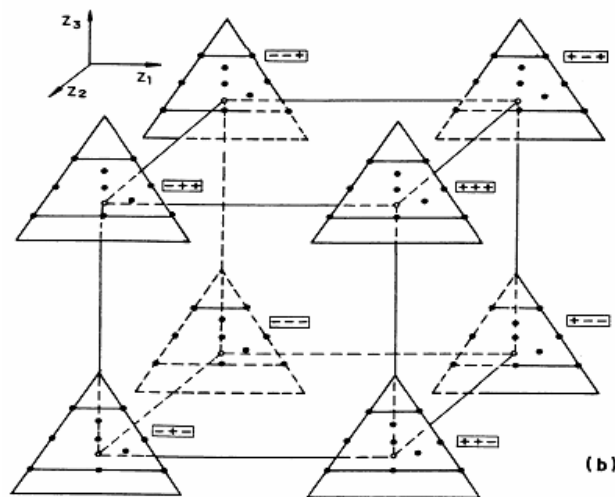


Figura 4. Planejamento “split-plot” para três variáveis de processo em dois níveis cada e três variáveis de mistura formando dez composições diferentes perfazendo 160 experimentos em duplicata.

A execução de 160 experimentos de forma completamente aleatória pode se tornar uma tarefa extremamente cansativa com enorme esforço no laboratório. Empregando o método “split-plot” o trabalho pode ser reduzido e otimizado⁴.

Um modelo para descrever o planejamento representado na Figura 4 que envolve variáveis de processo e mistura simultaneamente, pode ser obtido pela multiplicação do modelo bilinear para variáveis de processo por um modelo de mistura. Por exemplo, utilizando o modelo do planejamento fatorial com efeitos de

interação e o modelo linear de misturas, tem-se para duas variáveis de processo e três componentes de mistura

$$\begin{aligned} \text{resposta} &= \text{Equação 1} \times \text{Equação 2} = \sum_{i=1}^3 \beta_i (\alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \alpha_{12} z_1 z_2) x_i \\ &= \sum_{i=1}^3 [\beta_i^0 x_i + \beta_i^1 x_i z_1 + \beta_i^2 x_i z_2 + \beta_i^{12} x_i z_1 z_2] \end{aligned}$$

Equação 5

onde os termos cruzados em x_i e z_j representam efeitos de interação entre variáveis de processo e de mistura.

Quando descrito no contexto do procedimento “split-plot”, o erro envolvendo a combinação de variáveis de processo é o erro “main-plot” e o erro presente entre as composições de mistura (ou unidades “sub-plot”) é o erro “sub-plot”. Quando os papéis das composições de misturas e variáveis de processo são trocados, como na Figura 3b as composições de mistura são consideradas tratamentos de “main-plot” e as condições de processo são do “sub-plot”⁵.

Para experimentos realizados de forma totalmente aleatória utiliza-se uma ANOVA convencional, pois há apenas uma fonte de erro, contudo o procedimento “split-plot” necessita de uma análise de variância mais complexa, já que apresenta diferentes erros. A precisão na determinação do erro “main-plot” e “sub-plot” para um planejamento “split-plot” não é a mesma. As sub-unidades dentro da mesma unidade (“main-plot”) não são tratadas em ordem aleatória e por isto existe correlação entre seus erros. Por outro lado as unidades são tratadas aleatoriamente e as sub-unidades em diferentes unidades não têm erro correlacionado. Esta correlação normalmente é positiva em experimentos de química como, por exemplo, no caso de “drift” em instrumentos analíticos. O fato é que o efeito “main-plot” é calculado pela média total das unidades e a variância da soma dos erros contém uma contribuição positiva que depende da correlação da variância do efeito “main-plot”. Os efeitos “sub-plot” são calculados por diferenças e a contribuição dependendo da correlação é negativa, resultando em maior precisão.

Vale notar, que o ganho de precisão na determinação do erro “sub-plot” ocorre justamente pelo sacrifício da precisão no erro “main-plot”.

Muitos autores como Wooding e Montgomery^{4,27} definem o erro “main-plot” como interação de tratamento de replicatas por “main-plot” (RM_{ij}) em um modelo convencional de análise de variância.

$$Y_{ijkl} = \mu + R_i + M_j + RM_{ij} + S_k + RS_{ik} + MS_{jk} + RMS_{ijk} + \varepsilon_{l(ijk)}.$$

Equação 6

Na equação acima, Y_{ijkl} é a l -ésima observação na sub-unidade k que pertence j -ésimo unidade inteira (main) na i -ésima replicata.

μ = média global

R_i = efeito da i -ésima replicata, $i=1,2,\dots, r$ ($r = 2$ se foram feitas duplicatas)

M_j = efeito do j -ésimo tratamento “main-plot”; $j=1,2,\dots,P$ ($P = 4$ para um fatorial 2^2 , 8 para um fatorial 2^3 etc.)

S_k = efeito do k -ésimo tratamento “sub-plot”; $k=1,2,\dots,m$ (m é o número total de misturas)

RM_{ij} = efeito de interação replicata-“main-plot”

RS_{ik} = efeitos de interação replicata-“sub-plot”

MS_{jk} = interação de tratamentos “main-plot” por “sub-plot”

RMS_{ijk} = interação ternária replicata por “main-plot” por “sub-plot”

$\varepsilon_{l(ijk)}$ = erro associado com Y_{ijkl} ; $l = 1,2,\dots,L$

A análise de variância, ANOVA, na Tabela 1 corresponde ao modelo da Equação 6. Baseado nas expressões de esperanças para médias quadráticas, $E(MQ)$, testes de hipótese concernindo os efeitos dos tratamentos “main-plot”(Z), os tratamentos “sub-plot”(X) e as interações de tratamento “main-plot” por “sub-plot” (ZX) podem ser realizados e comparados com os valores de F tabelado ($F = MQZ/MQRZ$, $F = MQX/MQRX$ e $F = MQZX/MQRZX$). Vários autores como Snedecor e Cochran²⁸, Anderson e Bancroft³ e Steele e Torrie²⁹ questionam a presença do termo de interação replicata \times “sub-plot”, RS_{ik} , na Equação 6

particularmente quando o número de replicatas é pequeno. Estes autores recomendam que se retire os termos RS_{ik} e RMS_{ik} da Equação 6 e os agregam com $\varepsilon_{(ijk)}$ para formar um termo de erro “sub-plot”. Com isto um teste mais eficaz dos efeitos de tratamento “sub-plot” pode ser obtido²⁸. O novo modelo é

$$Y_{ijk} = \mu + R_i + Z_j + RZ_{ij} + X_k + ZX_{jk} + \varepsilon_{ijk} \quad ,$$

Equação 7

onde $i = 1, 2, \dots, r$, $j = 1, 2, \dots, p$ e $k = 1, 2, \dots, m$. Nesta equação os fatores “main-plot” e “sub-plot” na Equação 6 são substituídos por Z_j e X_k . Para este modelo a tabela de ANOVA está indicada na Tabela 2. Testes envolvendo as médias quadráticas podem ser feitos para avaliar as significâncias dos tratamentos “main-plot”, “sub-plot” e suas interações.

Tabela 1. Formato ANOVA para o modelo convencional “split-plot” da Equação 6.

Fonte de Variação	GL (graus de liberdade)	MQ (média quadrática)
Replicatas	$(r-1)^a$	MQR
Tratamentos main-plot	$(p-1)^b$	MQZ
Interações replicata-main-plot	$(r-1)(p-1)$	MQRZ
Tratamentos sub-plot	$(m-1)^c$	MQX
Interações replicatas-sub-plot	$(r-1)(m-1)$	MQRX
Interações main-plot-sub-plot	$(p-1)(m-1)$	MQZX
Interações main-plot-sub-plot-replicatas	$(r-1)(p-1)(m-1)$	MQRZX
Erro	0	MQE

^a r é o número de replicatas;

^b p é o número de condições de processo;

^c m é o número de composições de mistura.

Tabela 2. ANOVA para o modelo “split-plot” da Equação 7.⁵

Fonte de variação	GL	MQ	E (MQ) ^a
Replicatas (R)	(r-1)	MQR	$\sigma_e^2 + m\sigma_{RZ}^2 + mP\sigma_R^2$ ^b
Main-plot (Z)	(p-1)	MQZ	$\sigma_e^2 + m\sigma_{RZ}^2 + rm\theta_Z^2$
Erro main-plot (RZ)	(r-1)(p-1)	MQRZ	$\sigma_e^2 + m\sigma_{RZ}^2$
Sub-plot (X)	(m-1)	MQX	$\sigma_e^2 + rP\theta_X^2$
Interação main-plot-sub-plot (ZX)	(P-1)(m-1)	MQZX	$\sigma_e^2 + r\theta_{ZX}^2$
Erro sub-plot	P(r-1)(m-1)	MQE	σ_e^2

^a Expectativa ou esperança para média quadrática;

^b Os termos $\hat{\sigma}_R^2$, $\hat{\sigma}_{RZ}^2$ e $\hat{\sigma}_e^2$ são estimativas das variâncias das replicatas, do erro “main-plot” e “sub-plot”, respectivamente.

3.3.1 Modelos e detalhamento da ANOVA no método split-plot

Em toda situação em que um modelo matemático é ajustado a um conjunto de dados é necessário que se avalie a qualidade do ajuste. Caso o modelo gere resíduos grandes isto indica que o mesmo pode ser inadequado²⁷. É evidente que um modelo pode deixar resíduos, isto advém do fato dos dados estudados possuírem intrinsecamente erro em suas medidas. Somente um modelo que se ajuste perfeitamente aos dados experimentais não deixará resíduos, e na prática não se procuram modelos perfeitos, pois nestes casos normalmente eles são super ajustados a todas as informações e não somente a efeitos que possam ser fisicamente significativos.

A ANOVA é uma ferramenta que auxilia no estudo da qualidade dos modelos ajustados. Para planejamentos do tipo “split-plot” duas análises de variância são possíveis⁵. A primeira possui um caráter geral, que não depende de modelos ajustados aos resultados, permite a decomposição da soma quadrática total – que é a soma das diferenças entre cada medida e a média global elevadas ao quadrado – em parcelas referentes a efeitos como “main-plot”, “sub-plot” e a interação “main-plot-sub-plot”, além dos erros e a fonte de variância pelas replicatas. Este tipo de ANOVA é indicado na Tabela 3, que é similar a Tabela 2 mas possui a indicação das somas quadráticas (SQ),

Tabela 3. ANOVA para modelo “split-plot”.⁵

Fonte de variância	SQ	GL	MQ
Replicatas (R)	SQR	(r-1)	MQR
Main-plot (Z)	SQZ	(p-1)	MQZ
Erro Main-plot (RZ)	SQRZ	(r-1)(p-1)	MQRZ
Sub-plot (X)	SQX	(m-1)	MQX
Interação main-plot-sub-plot (ZX)	SQZX	(p-1)(m-1)	MQZX
Erro Sub-plot	SQE	p(r-1)(m-1)	MQE

onde r é o número de replicatas, p o número de condições de processo e m o número de composições de misturas. Mais adiante são apresentadas as equações que podem ser usadas para calcular as quantidades indicadas nesta Tabela. Com a variância total dividida em parcelas é possível avaliar a significância de cada efeito em relação ao seu próprio erro utilizando o teste F. A segunda ANOVA é mais específica, pois dependerá do modelo a ser ajustado. Ela permitirá avaliar o quanto o modelo explicará da variância de cada efeito. Esta segunda ANOVA é um detalhamento da primeira, subdividindo as fontes de variância referentes ao “main-plot”, “sub-plot” e a interação “main-plot-sub-plot” em duas partes: falta de ajuste e regressão explicada pelo modelo escolhido. Portanto, para a obtenção desta segunda ANOVA o modelo escolhido para ser ajustado aos dados deve possuir menos parâmetros do que o número de experimentos executados, permitindo assim obter os graus de liberdade para testar uma possível falta de ajuste⁵. Para um planejamento “split-plot” pode-se subdividir a ANOVA como indicado na Tabela 4.

Tabela 4. Formato detalhado da ANOVA para planejamentos que permitam falta de ajuste, onde a e b são os números de parâmetros do modelo referente às variáveis de processo e de mistura respectivamente.

Fonte de variância	Graus de liberdade (GL)	Média Quadrática (MQ)
Replicatas (R)	$(r-1)$	MQR
Main-plot (Z)	$(p-1)$	MQZ
Regressão (ZREG)	$(a-1)$	MQZREG
Falta de ajuste (ZLOF)	$(p-a)$	MQZLOF
Erro Main-plot ^a	$(r-1)(p-1)$	MQRZ
Sub-plot (X)	$(m-1)$	MQX
Regressão (XREG)	$(b-1)$	MQXREG
Falta de ajuste (XLOF)	$(m-b)$	MQXLOF
Interação main-sub-plot (ZX)	$(p-1)(m-1)$	MQZX
Regressão (ZXREG)	$(a-1)(b-1)$	MQZXREG
Falta de ajuste (ZXLOF)	$pm - ab - p - m + a + b$	MQZXLOF
Erro Sub-plot ^b	$(r-1)(m-1) + (r-1)(m-1)(p-1)$	MQE

^a Obtido da interação “main-plot”-replicata

^b Obtido da interação “sub-plot”-replicata e “main-sub-plot”-replicata.

Assim testes estatísticos poderão ser executados para avaliar a qualidade dos modelos gerados.

Observando a Tabela 4 nota-se que a variância relacionada ao “main-plot” pode ser explicada em três parcelas: uma devido ao erro, outra devido a regressão do modelo ajustado, e outra referente à falta de ajuste. O mesmo se aplica ao “sub-plot” e a interação “main-sub-plot”, com uma diferença: muitos estatísticos preferem inserir o erro referente à interação “main-sub-plot” dentro do erro “sub-plot” em planejamentos “split-plot”, como já explicado.

O fato de realizar experimentos em replicatas permite se obter uma estimativa do erro. Neste caso para cada fonte de variância (“main-plot”, “sub-plot” e a interação “main-plot-sub-plot”) têm-se suas interações com as replicatas. Isto permite obter o erro para cada fonte. Este erro associado a um efeito (como o “main-plot”) é análogo ao erro puro em uma modelagem convencional com replicatas. Mas não se deve esquecer que no planejamento “split-plot” há basicamente duas fontes de erros, uma devido ao “main-plot” e outra ao “sub-plot”. Isto não nos permite atribuir um erro puro único às medidas. Assim determinam-se erros diferentes para os blocos (“main-plot”) e para as medidas presentes no interior dos blocos (“sub-plot”).

A variância possui além do erro mais duas possíveis fontes: uma referente à regressão (variância explicada pelo modelo) e outra pela falta de ajuste do próprio modelo. Para compreender melhor estas fontes de variância será adotada a seguinte estratégia: decompõem-se os desvios referentes às respostas observadas em relação à média global. Observe a Figura 5 que mostra o caso análogo univariado para regressão linear convencional.

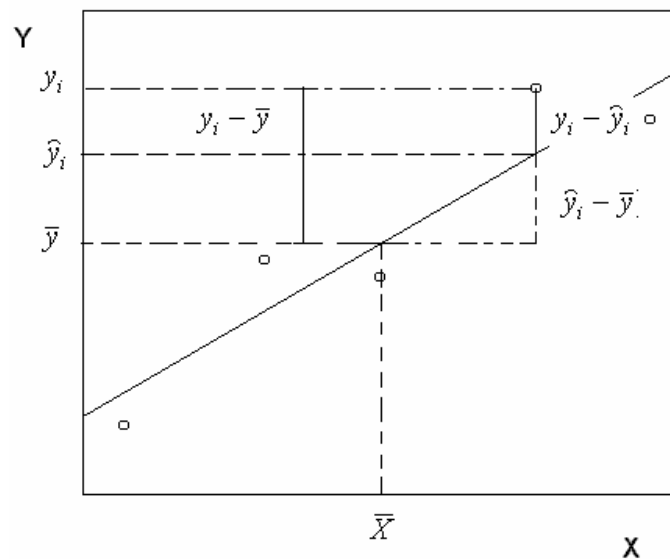


Figura 5 . Decomposição do desvio de uma observação em relação à média global, $(y_i - \bar{y})$, na soma das parcelas $(\hat{y}_i - \bar{y})$ e $(y_i - \hat{y}_i)$.¹

É possível notar que o desvio de uma resposta individual, y_i , em relação a média de todas as respostas observadas (\bar{y}), $(y_i - \bar{y})$, pode ser decomposto em duas parcelas

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

Equação 8

A primeira parcela da Equação 8, $(\hat{y}_i - \bar{y})$, representa o desvio da previsão feita pelo modelo para um ponto, \hat{y}_i , em relação à média global. A segunda parcela é a diferença entre o valor observado e o valor previsto. Em um modelo bem ajustado

o valor da segunda parcela deve ser pequeno. Como os valores dos dados reais oscilam de forma negativa e positiva em relação aos valores previstos pelo modelo, para que possa ser feita a soma destes desvios é necessário elevar os membros da Equação 8 ao quadrado, e fazer o somatório para todos os pontos.

$$\begin{aligned} \sum (y_i - \bar{y})^2 &= \sum [(\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)]^2 \\ &= \sum (\hat{y}_i - \bar{y})^2 + 2 \sum (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) + \sum (y_i - \hat{y}_i)^2 \end{aligned}$$

Equação 9

A segunda parcela da Equação 9 vale zero. Assim se obtém,

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$$

Equação 10

onde a primeira parcela se refere à soma quadrática devido a regressão (SQR) e a segunda parcela é a soma quadrática devido aos resíduos. Vale notar novamente, que em um modelo bem ajustado os valores dos resíduos devem ser pequenos.

3.3.2 Falta de ajuste em modelos para planejamentos split-plot

Para realizar um planejamento e análise tradicionais é necessária a realização de experimentos em replicatas³⁰. Isto permite obter uma estimativa do erro aleatório e dividir os resíduos em duas parcelas, uma devido ao erro puro, primeira parcela da Equação 11, e outra referente à falta de ajuste, segunda parcela da Equação 11,

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

Equação 11

onde j representa o número de replicatas dentro de uma condição experimental i.

A Equação 11 é rigorosamente correta para planejamentos tradicionais, mas existem equações análogas para cada fonte de variância no tratamento “split-plot”: “main-plot”, “sub-plot” e a interação “main-sub-plot”. Normalmente

planejamentos do tipo “split-plot” exigem a realização de uma grande quantidade de experimentos, uma vez que todas as composições de mistura do planejamento devem ser executadas em todas as condições de processo. Muito facilmente um planejamento pode exigir a realização de 30, 40, 50 experimentos ou mais. Assim mesmo utilizando modelos complexos com grande quantidade de parâmetros (entre efeitos individuais e cruzados) é comum restarem graus de liberdade para a falta de ajuste.

A determinação da falta de ajuste para modelos no planejamento “split-plot” é complexa porque há três fontes de variância: o “main-plot”, o “sub-plot” e a interação “main-sub-plot”. Utilizando estratégias comuns de ANOVA não seria possível determinar a falta de ajuste, pois estas são usadas para casos onde há uma fonte de variância apenas.

Para obter a ANOVA detalhada (Tabela 4) utilizou-se uma estratégia adotada na referência 5. Ao se ajustar um modelo do tipo “split-plot” deve-se adicionar termos que indiquem efeitos individuais do “main-plot” e do “sub-plot”, e termos cruzados de interações diversas, de modo a se obter um ajuste estatisticamente válido. Para exemplificar a estratégia adotada será utilizado um conjunto de dados contidos na Tabela 6 da referência 5, um planejamento que possui duas variáveis de processo, z_1 e z_2 , em dois níveis distintos cada, o que gera 4 condições de processo. Foram formuladas cinco composições de mistura utilizando-se três variáveis de mistura, x_1 , x_2 e x_3 . Ao todo foram executados 20 experimentos em duplicata, totalizando 40. Observe a Tabela 5. Se um modelo fosse ajustado a este planejamento com o objetivo de não possuir falta de ajuste matemática, o modelo deveria possuir 20 parâmetros. Com isto não haveria graus de liberdade para a falta de ajuste. Os graus de liberdade devem estar distribuídos corretamente entre cada fonte de variância (“main-plot”, “sub-plot” e “main-sub-plot”). Neste exemplo existem ao todo 4 condições de processo ($p = 4$), assim o “main-plot” possui 3 ($p - 1$) graus de liberdade para sua modelagem o que exige 4 parâmetros exclusivos no modelo para o “main-plot”. Por outro lado o “sub-plot” é formado por 5 misturas diferentes, exigindo 5 termos no modelo para ele. Como se sabe o modelo “split-plot” é obtido pelo produto entre o modelo “main-plot”

multiplicado pelo modelo para misturas (“sub-plot”)⁵. Desta forma um possível modelo (dentre outros possíveis) sem falta de ajuste matemática seria

$$\hat{y} = x_1 + x_2 + x_3 + x_1x_2 + x_1x_3 + (x_1 + x_2 + x_3 + x_1x_2 + x_1x_3) \cdot z_1 + (x_1 + x_2 + x_3 + x_1x_2 + x_1x_3) \cdot z_2 + (x_1 + x_2 + x_3 + x_1x_2 + x_1x_3) \cdot z_1z_2.$$

Equação 12

Observe que o modelo descrito na Equação 12 possui 20 termos (ou parâmetros), o que equivale ao número de experimentos e leva a uma falta de ajuste matemática igual a zero. Mas caso se desejasse ajustar um modelo mais simples, isto também seria possível. Contudo, o modelo deverá ser avaliado analisando-se seus resíduos (falta de ajuste e erro). Considere que agora o seguinte modelo tenha sido ajustado,

$$\hat{y} = x_1 + x_2 + x_3 + (x_1 + x_2 + x_3) \cdot z_1 + (x_1 + x_2 + x_3) \cdot z_2 + (x_1 + x_2 + x_3) \cdot z_1z_2.$$

Equação 13

percebe-se que o mesmo possui 12 termos o que permite avaliar o surgimento da falta de ajuste com 8 graus de liberdade. Mas a falta de ajuste não incide sobre tudo. O “main-plot” possui 4 parâmetros assim sua falta de ajuste é zero. Isto faz com que a falta de ajuste recaia sobre o “sub-plot” e a interação “main-sub-plot”. A Tabela 6 traz a ANOVA detalhada para o modelo da Equação 13. Observe que a soma quadrática referente a falta de ajuste para o “sub-plot” é de 81,225 e para “main-sub-plot” 12,275. Para se calcular os valores presentes na Tabela 6 foi utilizado o software SAS versão 8.02, com a opção de análise de regressão (PROC REG).

Tabela 5. Dados de um planejamento “split-plot” com duas variáveis de processo (z_1 e z_2) e três de mistura (x_1 , x_2 e x_3)^a

Experimento	z_1	z_2	x_1	x_2	x_3	Resposta
1	1	-1	0,850	0	0,150	8
2	1	-1	0,850	0	0,150	7
3	1	-1	0,720	0	0,280	6
4	1	-1	0,720	0	0,280	5
5	1	-1	0,600	0,250	0,150	10
6	1	-1	0,600	0,250	0,150	11
7	1	-1	0,470	0,250	0,280	4
8	1	-1	0,470	0,250	0,280	5
9	1	-1	0,600	0,125	0,215	11
10	1	-1	0,600	0,125	0,215	10
11	-1	1	0,850	0	0,150	12
12	-1	1	0,850	0	0,150	10
13	-1	1	0,720	0	0,280	9
14	-1	1	0,720	0	0,280	8
15	-1	1	0,600	0,250	0,150	13
16	-1	1	0,600	0,250	0,150	12
17	-1	1	0,470	0,250	0,280	6
18	-1	1	0,470	0,250	0,280	3
19	-1	1	0,600	0,125	0,215	15
20	-1	1	0,600	0,125	0,215	11
21	-1	-1	0,850	0	0,150	7
22	-1	-1	0,850	0	0,150	8
23	-1	-1	0,720	0	0,280	7
24	-1	-1	0,720	0	0,280	6
25	-1	-1	0,600	0,250	0,150	9
26	-1	-1	0,600	0,250	0,150	10
27	-1	-1	0,470	0,250	0,280	5
28	-1	-1	0,470	0,250	0,280	4
29	-1	-1	0,600	0,125	0,215	9
30	-1	-1	0,600	0,125	0,215	7
31	1	1	0,850	0	0,150	12
32	1	1	0,850	0	0,150	11
33	1	1	0,720	0	0,280	10
34	1	1	0,720	0	0,280	9
35	1	1	0,600	0,250	0,150	14
36	1	1	0,600	0,250	0,150	12
37	1	1	0,470	0,250	0,280	6
38	1	1	0,470	0,250	0,280	5
39	1	1	0,600	0,125	0,215	13
40	1	1	0,600	0,125	0,215	9

^a Dados da Tabela 6 da referência 5

Para poder calcular a falta de ajuste para o “sub-plot” foi utilizado inicialmente uma parte do modelo na Equação 13, contendo apenas os parâmetros referentes as variáveis de mistura (“sub-plot”),

$$\hat{y} = x_1 + x_2 + x_3.$$

Equação 14

Tabela 6 . Anova detalhada para o modelo da Equação 13 ajustado.

Fonte de variância	(GL)	(SQ)	(MQ)
Replicatas (R)	1	13,225	13,225
Main-plot (Z)	3	66,475	22,158
Regressão (Z)	3	66,475	22,158
Falta de ajuste (ZLOF)	0	0	0
Erro Main-plot	3	7,475	2,492
Sub-plot (X)	4	226,850	56,713
Regressão (XREG)	2	145,625	72,813
Falta de ajuste (XLOF)	2	81,225	40,613
Interação main-plot-sub-plot (ZX)	12	25,150	2,096
Regressão (ZXREG)	6	12,875	2,146
Falta de ajuste (ZXLOF)	6	12,275	2,046
Erro Sub-plot	16	12,800	0,800

Realizada a regressão o modelo gerado foi usado para calcular as respostas previstas. Assim foi possível determinar a soma quadrática explicada pela regressão para o “sub-plot” através da expressão

$$SQXREG = \sum_i^m \sum_j^{n_i} (\hat{y}_{.jk} - \bar{y}_{..k})^2, \text{ sendo } \bar{y}_{..k} \text{ a média das replicatas em todas as}$$

condições de processo para cada uma das k composições de mistura e $\hat{y}_{.jk}$ a resposta prevista para cada composição em cada condição de processo. O valor de SQXREG é o valor da soma quadrática de XREG na Tabela 6. Para calcular a falta de ajuste para o “sub-plot” completou-se o modelo da Equação 14 com mais dois termos, obtendo-se a equação

$$\hat{y} = x_1 + x_2 + x_3 + x_1x_2 + x_1x_3,$$

Equação 15

que é usada para gerar respostas que permitem a obtenção da soma quadrática pela regressão. Como não há mais graus de liberdade para a falta de ajuste seu valor será zero. Desta forma a diferença entre a soma quadrática gerada pela Equação 15 e pela Equação 14 fornece a falta de ajuste no “sub-plot” para o modelo da Equação 13.

Para obter a falta de ajuste para o “main-sub-plot” um procedimento análogo ao descrito anteriormente é usado. Inicialmente utiliza-se o modelo da Equação 13 para fazer o ajuste. Determinado os coeficientes do modelo isto permite obter a soma quadrática explicada pela regressão, é o valor de SQZXREG na Tabela 6. Mas note que neste modelo há somente 12 termos o que permite avaliar a falta de ajuste com 6 graus de liberdade. O modelo da Equação 13 é então completado com mais termos (x_1x_2 e x_2x_3 , e suas interações com z_1 , z_2 e z_1z_2) e se converte no modelo da Equação 12. Com o modelo completo contendo 20 parâmetros é feita a regressão e obtida a soma quadrática devido à regressão. Como agora a falta de ajuste é zero, isto permite afirmar que a diferença entre a soma quadrática devido à regressão para o modelo com 20 parâmetros e a soma quadrática devido à regressão para o modelo com 12 parâmetros expressa a falta de ajuste do modelo da Equação 13 em relação ao “main-sub-plot”.

O procedimento descrito para as determinações detalhadas da ANOVA é eficiente, porém possui uma importante limitação: o número de experimentos não deve ultrapassar o número de parâmetros no modelo considerado completo, ou seja, com todos os parâmetros possíveis. Assim se isto ocorrer, utilizando esta estratégia, não haveria como determinar as parcelas de falta de ajuste que formam a falta de ajuste total (soma de todas as faltas de ajuste).

3.3.3 ANOVA detalhada para planejamentos split-plot com número elevado de experimentos.

O planejamento “split-plot” comumente exige a realização de uma quantidade muito grande de experimentos em replicatas, como já exposto. Mas

adotando a estratégia da referência 5 não seria possível calcular a falta de ajuste nos casos onde o número de experimentos superasse a número de parâmetros no modelo considerado completo. Todavia para estes casos foi empregada uma estratégia nova que se assemelha em parte a da literatura, mas que permite realizar os cálculos.

Sabe-se que as informações contidas na Tabela 4 são apenas um detalhamento da Tabela 3. Em outras palavras, uma fonte de variância da Tabela 3 é apenas subdividida na Tabela 4 com o ajuste de um modelo. Porém a variância fornecida por uma fonte, como o “sub-plot”, por exemplo, independe do modelo, mas apenas das respostas dos experimentos realizados. Um modelo adequado ajustado tentará descrever ao máximo a fonte de variância, isto é o que se chama de soma quadrática explicada pela regressão. Contudo o modelo não explicará 100% da variância de uma determinada fonte, considerando, obviamente, que possua menos parâmetros do que experimentos executados. A parte da variância não explicada é assumida como falta de ajuste, uma vez que o erro é calculado separadamente. Assim a variância de uma fonte, como o “sub-plot”, poderá ser basicamente dividida em duas parcelas:

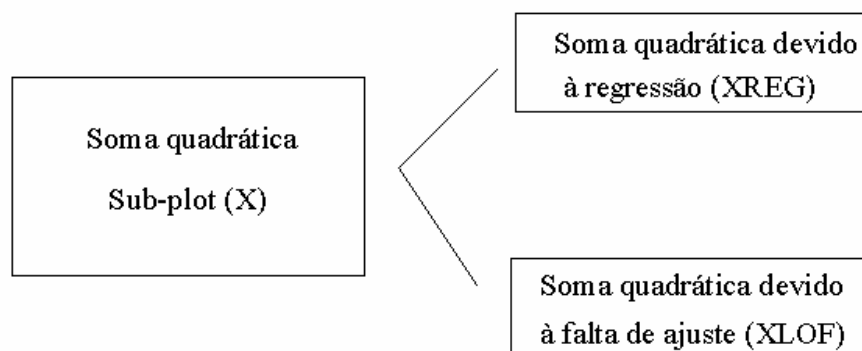


Figura 6. Divisão da soma quadrática do “sub-plot” (X) em duas parcelas referentes à regressão (XREG) e falta de ajuste (XLOF).

Portanto, determinando-se as somas quadráticas totais do “sub-plot” e devido a regressão, por diferença se obtém a falta de ajuste. A soma quadrática referente

ao “sub-plot” pode ser obtida pela Equação 16, que depende da resposta média para cada composição de mistura e da média global das respostas,

$$SQ_{\text{sub-plot}} = [(\bar{y}_{..1} - \bar{y}_{...})^2 + (\bar{y}_{..2} - \bar{y}_{...})^2 + \dots + (\bar{y}_{..m} - \bar{y}_{...})^2](r.p)$$

Equação 16

onde, $\bar{y}_{...}$ = média global das respostas;

$\bar{y}_{..m}$ = resposta média para cada composição de mistura;

r = número de replicatas;

p = número de condições de processo.

Esta média quadrática é igual SQXREG quando o modelo não tem graus de liberdade para testar a falta de ajuste. Na verdade todas as outras somas quadráticas com fontes descritas na Tabela 3 podem ser descritas pelas equações a seguir (eq.17-21). Note que estas equações independem do modelo a ser ajustado.

$$SQ_{\text{replicatas}} = [(\bar{y}_{1..} - \bar{y}_{...})^2 + (\bar{y}_{2..} - \bar{y}_{...})^2 + \dots + (\bar{y}_{r..} - \bar{y}_{...})^2](m.p)$$

Equação 17

$$SQ_{\text{main-plot}} = [(\bar{y}_{.1.} - \bar{y}_{...})^2 + (\bar{y}_{.2.} - \bar{y}_{...})^2 + \dots + (\bar{y}_{.p.} - \bar{y}_{...})^2](r.m)$$

Equação 18

$$SQ_{\text{interação rep-main-plot}} = \left[\sum_i^r \sum_j^p (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 \right] \cdot m$$

Equação 19

$$SQ_{\text{interação main-sub-plot}} = \sum_j^p \sum_h^m (\bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...})^2 \cdot r$$

Equação 20

$$SQ_{\text{total}} = \sum_i^r \sum_j^p \sum_k^m (y_{ijk} - \bar{y}_{...})^2$$

Equação 21

onde $\bar{y}_{.jk}$ = média das replicatas na j-ésima condição de processo e k-ésima composição de mistura; $\bar{y}_{.j.}$ = média de todas as composições e replicatas na j-ésima condição de processo; $\bar{y}_{..k}$ = média de todos os tratamentos e replicatas na k-ésima composição de mistura.

Para exemplificar esta estratégia será utilizado o conjunto de dados da Tabela 5. O primeiro passo é calcular a variância de cada fonte através da Equação 16 a Equação 21. Os resultados estão descritos na Tabela 7 e são iguais os valores dados na Tabela 6 que vieram da referência 5.

Tabela 7. Resultados não detalhados da análise de variância obtidos por meio das equações Equação 16 a Equação 21.

Fonte	Soma quadrática (SQ)
Replicatas (R)	13,225
Main-plot (Z)	66,475
Erro main-plot (RZ)	7,475
Sub-plot (X)	226,85
Interação main-sub-plot (ZX)	25,150
Erro	12,800

O próximo passo será ajustar um modelo que permita explicar, através da regressão, parte da variância de cada fonte. O primeiro modelo a ser ajustado é o da Equação 14, e com ele se obtém as respostas por previsão ($\hat{y}_{.jk}$). Usando a

equação $\sum_i^m \sum_j^{n_j} (\hat{y}_{.jk} - \bar{y}_{..k})^2$ determina-se a soma quadrática devido à regressão

(SQXREG). Observe a resolução a seguir:

$$SQ_{Sub-plot} = SQXREG + SQXLOF$$

$$SQXLOF = SQ_{Sub-plot} - SQXREG$$

$$SQXLOF = 226,85 - 145,625$$

$$SQXLOF = 81,225.$$

Note que o resultado obtido é exatamente igual ao da Tabela 6. A seguir o modelo da Equação 13 é ajustado permitindo-se fazer a previsão de respostas. Com a segunda parcela da Equação 11, para o caso “split-plot”, é possível se determinar a falta de ajuste total. O modelo ajustado também permite obter a soma quadrática devido à regressão. A soma quadrática referente à regressão é a soma das parcelas de três fontes, uma referente ao “main-plot” e outras duas referentes ao “sub-plot” e a interação “main-sub-plot” ($SQZREG + SQXREG + SQZXREG$). A falta de ajuste possui só duas fontes ($SQXLOF$ e $SQZXLOF$), uma vez que não há graus de liberdade para a falta de ajuste do “main-plot” ($SQZLOF$). Assim,

$$SQZREG + SQXREG + SQZXREG = 224,975,$$

como

$$SQZREG = 66,475 \text{ e } SQXREG = 145,625$$

$$SQZXREG = 224,975 - (66,475 + 145,625)$$

$$SQZXREG = 12,875.$$

Mais uma vez o resultado obtido é exatamente igual ao contido na Tabela 6. A fonte de variância referente à interação “main-sub-plot” pode ser parcelada em duas partes: uma referente à regressão ($SQZXREG$) e outra à falta de ajuste ($SQZXLOF$). Os resultados da Tabela 6 permitem obter a soma quadrática da interação “main-sub-plot” ($SQZX$). Portanto, por diferença se obtém a $SQZXLOF$,

$$SQZXLOF = SQZX - SSZXREG$$

$$SQZXLOF = 12,275.$$

4 ESTRUTURA DOS PLANEJAMENTOS

Os planejamentos estatísticos para a realização de experimentos não possuem como objetivos apenas sistematizar o trabalho experimental e reduzir os custos com materiais e tempo. Há também a necessidade de garantir a qualidade dos resultados da análise estatística empregada para que efeitos das variáveis envolvidas nos planejamentos sejam obtidos de forma segura. Um importante ponto neste caso é a estrutura do planejamento experimental adotada, pois esta estrutura irá permitir a obtenção de resultados mais precisos.

Considerando inicialmente um caso simples de determinação de efeitos por regressão em um modelo, utiliza-se a equação de regressão

$$\mathbf{b} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y}$$

Equação 22a

onde \mathbf{X} representa a matriz de planejamento elaborada para a obtenção do modelo desejado, \mathbf{y} o vetor contendo os valores das respostas experimentais e o vetor \mathbf{b} fornecerá os coeficientes dos efeitos. Contudo há um importante detalhe, os coeficientes ajustados por regressão apresentam erros associados e com isto alguns efeitos, que na verdade simplesmente descrevem erro aleatório, devem ser descartados. Para casos em que a fonte de erro é única a determinação dos erros dos parâmetros é obtida facilmente extraindo-se a raiz quadrada dos elementos da diagonal principal da matriz obtida pela expressão

$$\mathbf{V}(\mathbf{b}) = (\mathbf{X}^t \mathbf{X})^{-1} \sigma^2$$

Equação 22b

onde σ^2 é a variância populacional de erro de medida.

Observando a expressão anterior nota-se que os erros associados aos parâmetros dependem de dois fatores. O primeiro é simplesmente o erro

experimental, $\sqrt{\sigma^2}$, que basicamente está relacionado à qualidade das medidas realizadas. O segundo, depende da estrutura da matriz de planejamento, \mathbf{X} . Um planejamento que maximize o produto $\mathbf{X}'\mathbf{X}$ irá por consequência minimizar a inversa do mesmo, assim os valores dos erros distribuídos para os parâmetros contidos no vetor \mathbf{b} serão minimizados. Ou seja, a estimativa dos efeitos será realizada de forma mais precisa e confiável. Portanto, a configuração da matriz de planejamento, \mathbf{X} , é fundamental para a qualidade do estudo realizado.

4.1 Procedimentos para análise de dados

Atualmente para se realizar a análise estatística de um conjunto de dados há muitas ferramentas. Contudo as ferramentas são normalmente específicas para certas condições e muitas vezes se aplicadas em situações inapropriadas poderão gerar resultados incoerentes. Softwares estatísticos como o SAS e o R disponibilizam estas ferramentas para os usuários. Mas cabe ao usuário a escolha do método adequado de acordo com seu planejamento experimental e o modelo adotado.

4.2 Componentes de variância

Existem vários métodos para estimar os componentes de variância, entre os mais importantes há o Método da Análise de Variância, Máxima Verossimilhança (ML)^a e Máxima Verossimilhança Restrita (REML)^b. Basicamente os dados a serem tratados podem ser divididos em dois grandes grupos, os dados balanceados e desbalanceados.

^a Do inglês Maximum Likelihood

^b Do inglês Restricted Maximum Likelihood

4.2.1. Dados Balanceados

É o caso mais simples para ser tratado. Normalmente para estimar os componentes de variância se utiliza a Análise de Variância (ANOVA). Em alguns softwares de estatística, como o SAS, o procedimento ANOVA é recomendado para situações completamente balanceadas com iguais replicações. Para situações em que o número de replicatas é diferente aconselha-se empregar o procedimento GLS (Generalized Least Squares).

A aplicação do método ANOVA para dados balanceados normalmente é simples. Como uma desvantagem este método não exclui a ocorrência de estimativas negativas. Normalmente uma estimativa negativa de um parâmetro, uma fonte de variância, que por definição é positiva é considerada um embaraço pelos estatísticos^{31,32}.

4.2.2. Dados Desbalanceados

Para o tratamento de dados desbalanceados o problema normalmente é a escolha do melhor método dentre os disponíveis. Normalmente os softwares mais empregados estão baseados nos métodos fundamentados na máxima verossimilhança o que os tornou mais populares. Os métodos de máxima verossimilhança são mais flexíveis, não exigindo delineamentos balanceados e geralmente conduzindo a maiores eficiências que o método de mínimos quadrados ordinários (OLS, ordinary least square). Além disso, os estimadores ML e REML podem ser robustos a desvios da normalidade segundo Harville³³. Assim, o método REML tornou-se, recentemente, o principal método de estimação de componentes de variância e tem sido amplamente empregado.

4.3 O método de máxima verossimilhança

Os princípios do método de máxima verossimilhança são antigos, contudo apenas com a popularização dos computadores é que o método ganhou espaço. Este método normalmente utiliza otimização numérica para a obtenção de parâmetros de interesse, o que justifica seu consumo computacional. Contudo ele apresenta propriedades assintóticas dos estimadores que são consistentes, isto é, a esperança do valor de um estimador converge ao valor real para casos em que o número de amostra é infinito.

O método de máxima verossimilhança adota como estimativa dos parâmetros, os valores que maximizam a probabilidade de ser obtida a amostra observada. Para obter os estimadores de máxima verossimilhança é necessário conhecer a distribuição da variável de estudo. Para exemplificar, considere que dois parâmetros de uma variável aleatória (x) queiram ser estimados, a média (μ) e a variância (σ^2). A amostra é considerada proveniente de uma população com distribuição normal. A densidade de probabilidade de obter um valor de x_i na amostra é

$$f(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i - \mu)^2}{2\sigma^2}\right\}.$$

Equação 23

Como os valores observados são independentes, a densidade de probabilidade de se obter os valores x_1, x_2, \dots, x_n da amostra é

$$L(x_1, x_2, \dots, x_n, \mu, \sigma^2) = f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i - \mu)^2}{2\sigma^2}\right\} =$$

$$(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{\sum(x_i - \mu)^2}{2\sigma^2}\right\}.$$

Equação 24

Esta é considerada a função de verossimilhança da amostra. Os estimadores de probabilidade máxima (ou verossimilhança) de μ e σ^2 são os valores que maximizam o valor de L . Prefere-se trabalhar com o logaritmo natural da função de verossimilhança ($\ln L$), já que maximizar o logaritmo natural de uma função é normalmente mais simples. Assim

$$\ln(L) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{\Sigma(x_i - \mu)^2}{2\sigma^2}$$

Equação 25

e calculando-se as derivadas parciais em relação a μ e σ^2 e as igualando a zero se obtém o sistema

$$\frac{\Sigma(x_i - \hat{\mu})}{\hat{\sigma}^2} = 0$$

Equação 26

$$-\frac{n}{2\hat{\sigma}^2} + \frac{\Sigma(x_i - \hat{\mu})^2}{2\sigma^4} = 0$$

Equação 27

Isolando $\hat{\mu}$ da Equação 26

$$\Sigma(x_i - \hat{\mu}) = 0$$

$$\Sigma(x_i) - \Sigma(\hat{\mu}) = 0$$

$$n\hat{\mu} = \Sigma(x_i)$$

$$\hat{\mu} = \frac{\Sigma(x_i)}{n} = \bar{x}$$

Equação 28

Desta forma \bar{x} é um estimador de máxima verossimilhança para a média.

Com a Equação 26 e Equação 27 obtém-se

$$\hat{\sigma}^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

que é considerado um estimador de máxima verossimilhança da variância tendencioso³³, uma vez que o estimador não tendencioso é

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}.$$

Para compreender um pouco mais do método de máxima verossimilhança considere um conjunto de pontos experimentais indicados na Figura 7. Um problema clássico consiste em obter a melhor função $f(x)$ para descrever um conjunto de pontos experimentais em que uma variável y , dependente, tem seus valores alterados por modificações em uma variável x independente. Este processo é chamado de regressão e é obtido normalmente por mínimos quadrados.

4.4 O método de mínimos quadrados

Em 1809, Carl F. Gauss demonstrou que a melhor maneira de determinar um parâmetro desconhecido de uma equação é minimizar a soma dos quadrados dos resíduos, mais tarde chamado de Mínimos Quadrados por Adrien M. Legendre. Em 1810, Pierre Laplace apresentou a generalização do método para problemas com vários coeficientes de parâmetros desconhecidos³¹.

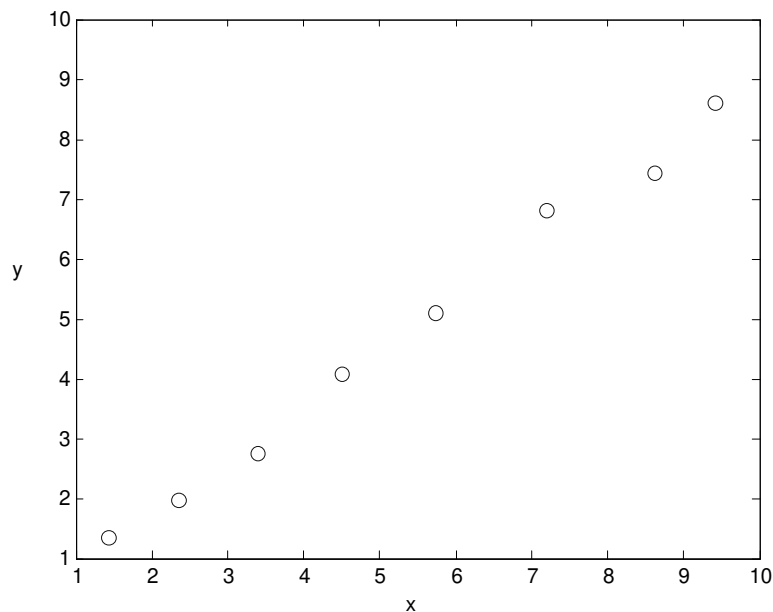


Figura 7. Conjunto de pontos experimentais $\{x,y\}$

Para tentar elucidar um pouco mais este método observe o conjunto de pontos experimentais representados na Figura 7. Considere que uma função que relaciona os valores de x e y , $f(x)=y$, seja desejada. A princípio muitas funções poderiam ser utilizadas para tentar prever valores de y utilizando valores de x . Mas, aparentemente, uma função, ou modelo, linear parece ser adequado. Contudo, a reta ajustada não irá se sobrepôr a todos os pontos. Assim o modelo irá prever respostas, \hat{y}_i , que não são exatamente iguais as experimentais, y_i , gerando um certo resíduo (ε_i) dado por $\varepsilon_i = y - \hat{y}$. O ideal seria se a reta contivesse todos os pontos, mas como isto não é possível o melhor que se espera

é que a reta se aproxime ao máximo de todos os pontos. Ou seja, a reta procurada seria aquela que minimiza a distância global de todos os pontos até ela. Portanto, para um modelo linear, $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, deve-se encontrar os valores de β_0 e β_1 que minimizam ε_i , mas como deseja-se minimizar a distância global minimiza-se $\sum \varepsilon_i^2$. Isto é obtido derivando-se $\sum \varepsilon_i^2$ em relação a β_0 e β_1 , e igualando-se as derivadas parciais a zero

$$\frac{\partial(\sum \varepsilon_i^2)}{\partial \beta_0} = 0$$

$$\frac{\partial(\sum \varepsilon_i^2)}{\partial \beta_1} = 0,$$

como $\sum \varepsilon_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - (\beta_0 + \beta_1 X_i))^2$, tem-se que

$$\frac{\partial(\sum \varepsilon_i^2)}{\partial \beta_0} = -2 \sum (y_i - \beta_0 - \beta_1 X_i) = 0$$

$$\frac{\partial(\sum \varepsilon_i^2)}{\partial \beta_1} = -2 \sum X_i (y_i - \beta_0 - \beta_1 X_i) = 0.$$

Isolando β_0 e β_1 , obtêm-se

$$\beta_0 = \frac{\sum y_i - \beta_1 \sum X_i}{n} = \bar{y} - \beta_1 \bar{X}$$

$$\beta_1 = \frac{\sum (X_i - \bar{X})(y_i - \bar{y})}{\sum (X_i - \bar{X})^2}$$

que são as estimativas conhecidas como mínimos quadrados ordinários, do inglês “ordinary least-squares” (OLS).

Os valores de β_0 e β_1 podem ser encontrados através da resolução da equação matricial (Equação 22a)

$$\mathbf{b} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y}$$

onde \mathbf{X} é a matriz planejamento, \mathbf{y} é o vetor resposta e \mathbf{b} é o vetor contendo os parâmetros do modelo obtido pela regressão. A equação anterior permite realizar a regressão por mínimos quadrados para qualquer modelo (linear, quadrático, cúbico etc) bastando organizar a matriz de entrada \mathbf{X} para o formato desejado. Contudo a equação anterior possui uma importante característica: deve ser aplicada a situações em que os experimentos são realizados de forma completamente aleatória. Neste caso as estimativas fornecidas não serão viesadas³². Para casos onde os experimentos não sejam realizados de forma completamente aleatórios aconselha-se⁵ utilizar os estimadores obtidos por mínimos quadrados generalizados dados por

$$\mathbf{b} = (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}^{-1} \mathbf{y}.$$

Equação 29

onde \mathbf{V} é dada por

$$\mathbf{V} = \{\mathbf{J}_n \otimes \mathbf{I}_r\} \hat{\sigma}_R^2 + \{\mathbf{J}_m \otimes \mathbf{I}_{rp}\} \hat{\sigma}_{RZ}^2 + \mathbf{I}_{mpr} \hat{\sigma}_e^2$$

Equação 30

sendo que \mathbf{J} é uma matriz com blocos diagonalizados com valores unitários para todos os elementos das diagonais dos blocos e valores nulos para os elementos restantes. \mathbf{I} são matrizes identidades e n , r , m e p são os números de experimentos, replicatas, misturas e condições de processo, respectivamente. Os termos $\hat{\sigma}_R^2$, $\hat{\sigma}_{RZ}^2$ e $\hat{\sigma}_e^2$ são estimativas das variâncias dos erros das replicatas, do “main-plot” e “sub-plot”, respectivamente sendo calculados a partir dos resultados da ANOVA para planejamentos contendo replicatas. A matriz de covariância de \mathbf{b} para determinação dos erros associados aos parâmetros do modelo, para casos com mais de uma fonte de erro, é dada por

$$\text{Cov}(\hat{\mathbf{b}}) = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \hat{\mathbf{V}} \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1}$$

já a matriz de covariância para mínimos quadrados generalizados é fornecida por

^a \otimes indica o operador do produto de Kronecker, veja referência 5

$$\text{Cov}(\hat{\mathbf{b}}) = (\mathbf{X}^t \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1}.$$

Apenas em casos onde o modelo é balanceado e para um conjunto de dados balanceados é possível assumir que

$$(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \hat{\mathbf{V}} \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} = (\mathbf{X}^t \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1}$$

e garantir que as estimativas de variância serão mínimas.

4.5 O método de mínimos quadrados por máxima verossimilhança

Na verdade muitas funções poderiam ser ajustadas aos pontos da Figura 7. Definir a melhor função irá depender de algum critério. O critério normalmente usado é o método de máxima verossimilhança que formula: *a melhor aproximação para descrever um conjunto de pontos experimentais é função $f(x)$ para a qual o particular conjunto de pontos obtidos é o mais verossímil possível*. Para o ajuste de função, o método consiste em determinar a função $f(x)$ para a qual é máxima a probabilidade de ocorrer o conjunto de pontos em estudo, isto é realizado por mínimos quadrados.

O método dos mínimos quadrados pode ser obtido do método de máxima verossimilhança, desde que as seguintes condições sejam obedecidas: primeiro, a distribuição do erro deve ser normal; segundo, a função com seus parâmetros evidenciados deve ser previamente escolhida. Para se estimar por verossimilhança os parâmetros α e β da equação

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

considerando que o erro ε_i é independente com média zero, variância σ^2 e distribuição normal, isto é,

$$\varepsilon \approx N(0, \sigma^2)$$

o valor da densidade de probabilidade para cada y_i será

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - (\alpha + \beta x_i))^2}{2\sigma^2}\right\}, \text{ com } \mu = \alpha + \beta x_i.$$

A probabilidade de o modelo ter gerado as observações y_i será

$$L = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - (\alpha + \beta x_i))^2}{2\sigma^2}\right\}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum (y_i - (\alpha + \beta x_i))^2\right\}$$

O logaritmo natural da função L é dado por

$$\ln(L) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{\sum (y_i - (\alpha + \beta x_i))^2}{2\sigma^2}.$$

Os estimadores de máxima verossimilhança de α , β e σ^2 serão encontrados igualando-se as derivadas, em relação à α , β e σ^2 , a zero.

$$\frac{\partial \ln(L)}{\partial \beta} = \frac{1}{2\sigma^2} \sum_{i=1}^n 2(y_i - (\alpha + \beta x_i))(-x_i) = 0$$

$$\frac{\partial \ln(L)}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (y_i - (\alpha + \beta x_i))^2 = 0$$

$$\frac{\partial \ln(L)}{\partial \alpha} = 2 \sum_{i=1}^n (y_i - (\alpha + \beta x_i))(-1) = 0$$

A resolução das equações anteriores para α , β e σ^2 fornece os estimadores

$$\hat{\beta} = \frac{\sum_{i=1}^n (y_i x_i)}{\sum_{i=1}^n x_i^2} = \hat{\beta}_{OLS} \text{ (ordinary least square)}$$

$$\sigma^2 = \frac{\sum_{i=1}^n (y_i - \beta x_i)^2}{n}$$

$$\alpha = \frac{\sum y_i}{n} - \beta \frac{\sum x_i}{n} = \bar{y} - \beta \bar{x}$$

Note que o estimador de máxima verossimilhança de β é igual ao estimador de mínimos quadrados ordinários e que o estimador de σ^2 é considerado viesado para amostras pequenas, pois a divisão é feita por n e não por $n-1$. Para amostras grandes, normalmente, a troca de n por $n-1$ não traz grandes conseqüências.

O método da máxima verossimilhança é iterativo e fornece estimativas viesadas (viciadas) porque o método não considera a perda de graus de liberdade resultante da estimação dos efeitos fixos do modelo. Como por exemplo, a perda de um grau de liberdade devido à média na determinação da variância de uma amostra.

4.6 O método da máxima verossimilhança restrita – REML

Este método é derivado do método da máxima verossimilhança, sendo utilizado inicialmente por Patterson e Thompson³⁴ para tratar dados em blocos com tamanhos diferentes. REML é equivalente a ML para um conjunto de dados que tenham sido padronizados para média zero. A modificação realizada conduz a estimativas idênticas àsquelas obtidas por análise de variância, se os experimentos forem balanceados e as restrições de não negatividade forem ignoradas³¹. Os métodos de máxima verossimilhança são considerados mais flexíveis, não exigindo balanceamento completo dos experimentos e modelos, e geralmente possuem maior eficiência que o método de mínimos quadrados. Além disso, Harville³³ sugere que os estimadores ML e REML podem ser robustos à desvios da normalidade por parte dos dados.

Vários algoritmos computacionais para a obtenção de componentes de variância por REML foram desenvolvidos³⁵. O método iterativo possui normalmente os seguintes passos: estima-se ou assume-se um valor inicial para as variâncias envolvidas, obtendo os valores para β e σ^2 , com os valores de β estima-se novamente valores para as variâncias. Este processo se repete até que certo critério de convergência seja atingido.

4.7 Comparação dos métodos ML, OLS e REML

A Figura 8 representa um planejamento fatorial 2^4 para compreender como quatro variáveis afetam a resistência de um plástico utilizando o método “split-plot” que possui restrições quanto a aleatorização dos experimentos⁴¹. O “split-plot” em planejamento fatorial é um caso especial e será explicado com maiores detalhes mais adiante. As variáveis estudadas foram temperatura (a), porcentagem de aditivo (b), velocidade de agitação (c) e tempo de processamento (d). Para facilitar o procedimento experimental a temperatura foi escolhida como “main-plot” e as outras três variáveis formaram o “sub-plot”. A temperatura foi fixada em dois níveis e aleatoriamente se sorteou os níveis do “sub-plot”. A Figura 8 mostra a estratégia

adotada. A Tabela 8 indica os resultados obtidos com os experimentos em duplicata.

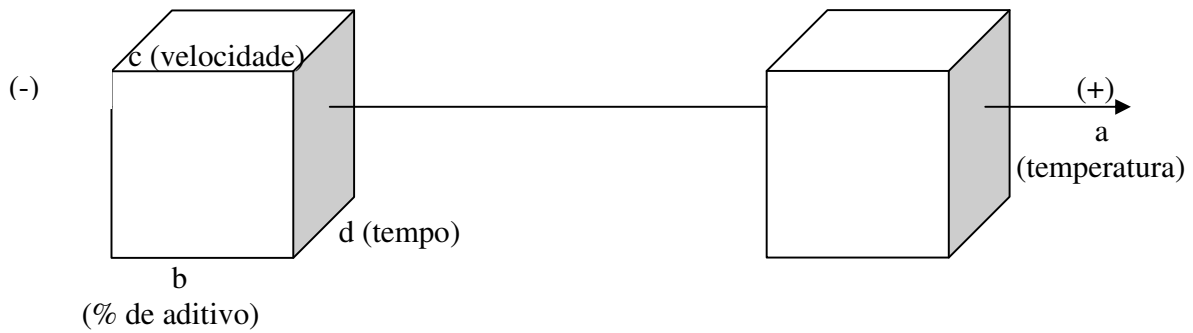


Figura 8. Ilustração do planejamento fatorial pelo método “split-plot” para quatro variáveis, (a) temperatura, (b) porcentagem de aditivo, (c) velocidade de agitação e (d) tempo de processamento.

Tabela 8. Planejamento fatorial pelo método “split-plot” para otimizar a resistência de um plástico. Os dados são completamente balanceados.

Temperatura	Aditivo	Velocidade	Tempo	Resposta
1	1	1	1	70,8
1	1	1	1	73,3
1	1	1	-1	66,2
1	1	1	-1	64,0
1	1	-1	1	66,8
1	1	-1	1	61,5
1	1	-1	-1	51,9
1	1	-1	-1	65,6
1	-1	1	1	68,5
1	-1	1	1	68,0
1	-1	1	-1	61,3
1	-1	1	-1	58,6
1	-1	-1	1	59,5
1	-1	-1	1	64,2
1	-1	-1	-1	58,5
1	-1	-1	-1	59,5
-1	1	1	1	63,9
-1	1	1	1	63,2
-1	1	1	-1	58,1
-1	1	1	-1	62,6
-1	1	-1	1	57,5
-1	1	-1	1	63,3
-1	1	-1	-1	57,4
-1	1	-1	-1	65,0
-1	-1	1	1	56,4
-1	-1	1	1	62,7
-1	-1	1	-1	56,5
-1	-1	1	-1	56,1
-1	-1	-1	1	53,2
-1	-1	-1	1	63,9
-1	-1	-1	-1	59,5
-1	-1	-1	-1	66,6

Inicialmente é necessário ajustar um modelo e neste caso o escolhido foi o bilinear,

$$\hat{y} = \alpha + b_1a + b_2b + b_3c + b_4d + b_{12}ab + b_{13}ac + b_{14}ad + b_{23}bc + b_{24}bd + b_{34}cd .$$

Equação 31

A primeira análise irá tratar os dados como se fossem provenientes de experimentos completamente aleatórios e, portanto, com uma única fonte de erro, para isto o método de mínimos quadrados ordinários (OLS) é empregado. Além disto, os dados são balanceados na medida em que todos os experimentos

possuem a mesma quantidade de replicatas. A Tabela 9 traz os coeficientes e erros dos coeficientes calculados.

Tabela 9. Coeficientes e erros dos parâmetros ajustados no modelo bilinear utilizando-se OLS.

Efeito	GL	Coeficientes	Erro padrão
Intercepto	1	62,003	0,666
a	1	1,634	0,666
b	1	1,191	0,666
c	1	1,134	0,666
d	1	1,541	0,666
ab	1	0,184	0,666
ac	1	1,566	0,666
ad	1	1,397	0,666
bc	1	0,934	0,666
bd	1	0,303	0,666
cd	1	1,172	0,666

Com os dados da Tabela 8 também se determinou os coeficientes do modelo empregando o método de máxima verossimilhança (ML) e o método de máxima verossimilhança restrita (REML). Estes métodos já consideram a existência de duas fontes de erros para calcular os erros associados aos parâmetros do modelo. Os resultados estão indicados na Tabela 10.

Tabela 10. Coeficientes e erros dos parâmetros ajustados no modelo bilinear utilizando-se ML e REML

Efeito	Coeficientes (REML)	Erro	Efeito	Coeficientes (ML)	Erro
Intercepto	62,003	1,628	Intercepto	62,003	1,151
A	1,634	0,928	a	1,634	0,653
B	1,191	0,553	b	1,191	0,455
C	1,134	0,553	c	1,134	0,455
D	1,541	0,553	d	1,541	0,455
ab	0,184	0,553	ab	0,184	0,455
ac	1,566	0,553	ac	1,566	0,455
ad	1,397	0,553	ad	1,397	0,455
bc	0,934	0,553	bc	0,934	0,455
bd	0,303	0,553	bd	0,303	0,455
cd	1,172	0,553	cd	1,172	0,455

Os resultados indicados na Tabela 10 mostram que os valores obtidos para os coeficientes tanto com ML e REML são idênticos, e, além disto, são exatamente

iguais aos valores obtidos utilizando OLS. Mas os erros calculados pelos métodos apresentam diferenças, pois o OLS está considerando dados completamente aleatórios, já o ML e o REML não. Entre ML e REML há diferenças, pois no cálculo dos componentes de variância o método ML não leva em consideração a perda dos graus de liberdade devido aos termos fixados no modelo gerado³¹. Contudo não se deve esperar que estes resultados, no cálculo dos coeficientes, sejam sempre concordantes, pois há a possibilidade de haver desbalanceamento dos dados. Uma situação de desbalanceamento seria retirar algumas replicatas de alguns experimentos, ou seja, para alguns experimentos não haveria medidas em duplicata, mas apenas uma medida única. Esta nova situação está indicada na Tabela 11.

Tabela 11. Planejamento fatorial pelo método “split-plot” para otimizar a resistência de um plástico. Os dados são desbalanceados.

Temperatura	Aditivo	Velocidade	Tempo	Resposta
1	1	1	1	70,8
1	1	1	-1	66,2
1	1	1	-1	64,0
1	1	-1	1	66,8
1	1	-1	1	61,5
1	1	-1	-1	51,9
1	1	-1	-1	65,6
1	-1	1	1	68,0
1	-1	1	-1	61,3
1	-1	1	-1	58,6
1	-1	-1	1	59,5
1	-1	-1	1	64,2
1	-1	-1	-1	58,5
-1	1	1	1	63,9
-1	1	1	1	63,2
-1	1	1	-1	58,1
-1	1	1	-1	62,6
-1	1	-1	1	57,5
-1	1	-1	1	63,3
-1	1	-1	-1	57,4
-1	-1	1	1	56,4
-1	-1	1	1	62,7
-1	-1	1	-1	56,5
-1	-1	-1	1	53,2
-1	-1	-1	1	63,9
-1	-1	-1	-1	59,5
-1	-1	-1	-1	66,6

Com os dados da Tabela 11 foram calculados novamente os parâmetros e erros para o modelo bilinear, notando que para o método OLS não se considera as

restrições de aleatoriedade, enquanto nos métodos ML e REML elas são consideradas. Os resultados são apresentados na Tabela 12.

Tabela 12. Coeficientes e erros dos parâmetros ajustados no modelo bilinear utilizando-se OLS, ML e REML para os dados desbalanceados da Tabela 11.

Variável	OLS		ML		REML	
	Coefficiente	Erro	Coefficiente	Erro	Coefficiente	Erro
Intercepto	61,799	0,816	62,008	1,352	62,022	1,910
A	1,665	0,821	1,555	0,631	1,529	0,931
B	1,024	0,811	1,056	0,516	1,052	0,660
C	1,300	0,821	1,246	0,523	1,252	0,669
D	1,597	0,823	1,373	0,527	1,360	0,674
AB	0,470	0,821	0,524	0,523	0,518	0,669
AC	1,278	0,837	1,203	0,534	1,211	0,682
Ad	1,218	0,829	1,313	0,530	1,341	0,679
Bc	1,085	0,821	1,083	0,528	1,045	0,679
Bd	0,563	0,823	0,731	0,529	0,712	0,678
Cd	0,858	0,829	0,897	0,528	0,893	0,675

Os valores da Tabela 12 mostram que no tratamento de dados desbalanceados os métodos OLS, ML e REML apresentam resultados diferentes. Os valores dos coeficientes obtidos apresentam diferenças, em virtude do desbalanceamento do conjunto de dados utilizados. Já o cálculo dos valores dos erros é influenciado pelo desbalanceamento e pelo fato dos métodos se comportarem de forma diferente frente a dados não completamente randomizados. Para situações como estas o método mais recomendado é o REML, e mesmo para situações mais simples sua aplicação é comum³¹.

Em resumo para casos em que os experimentos e replicatas são completamente balanceados utiliza-se normalmente OLS, GLS (general least square), ML e REML. Nos casos em que os dados sejam desbalanceados utiliza-se normalmente o método REML. Geralmente REML é empregado como método padrão em vários softwares, graças ao avanço computacional dos últimos anos. Contudo, é sempre importante consultar a documentação dos softwares para a escolha do método mais adequado para o estudo que se deseja realizar.

5 ESTUDOS COM PLANEJAMENTOS INCOMPLETOS

Como já explicado, em planejamentos “split-plot” é muito comum à necessidade de grandes quantidades de experimentos e isto pode ser ainda mais agravado se for considerada a realização de replicatas. Assim alguns estudos foram realizados com o objetivo de permitir que um planejamento “split-plot” seja executado com uma redução no número de experimentos.

5.1 Estudos com planejamentos contendo duplicatas de misturas sorteadas por condição de processo (I)

Utilizando o conjunto de dados da referência 36 que trata da otimização de um procedimento catalítico para a determinação de Cr (VI) foi montada uma série de planejamentos incompletos. O procedimento em questão foi baseado na reação de oxidação de o-dianisidina com peróxido de hidrogênio em meio pouco ácido sendo otimizado com respeito aos reagentes (HCl, H₂O₂ e o-dianisidina - variáveis de processo) e composição do solvente (mistura de água, acetona e N,N-dimetilformamida - variáveis de mistura), com a meta de alcançar alta sensibilidade. O modelo experimental “split-plot” permitiu variar simultaneamente fatores de processo e mistura, usando uma superfície de resposta aproximada.

No planejamento completo há 8 condições de processo com 10 composições de mistura em duplicatas perfazendo o total de 160 experimentos. A Figura 9 representa este planejamento. Para o planejamento incompleto foram testadas 5 possibilidades, com 3, 4, 5, 6 e 8 misturas com duplicatas por condição de processo. Na primeira, 3 composições de mistura foram sorteadas cujos resultados em duplicata foram incluídos no cálculo. Todas as composições, das dez possíveis, apareceram ao menos uma vez para evitar a perda de muitos graus de liberdade na modelagem, resultando em 48 experimentos. Este método foi

repetido para planejamentos onde 4, 5, 6 e 8 composições de misturas fossem sorteadas por condição de processo.

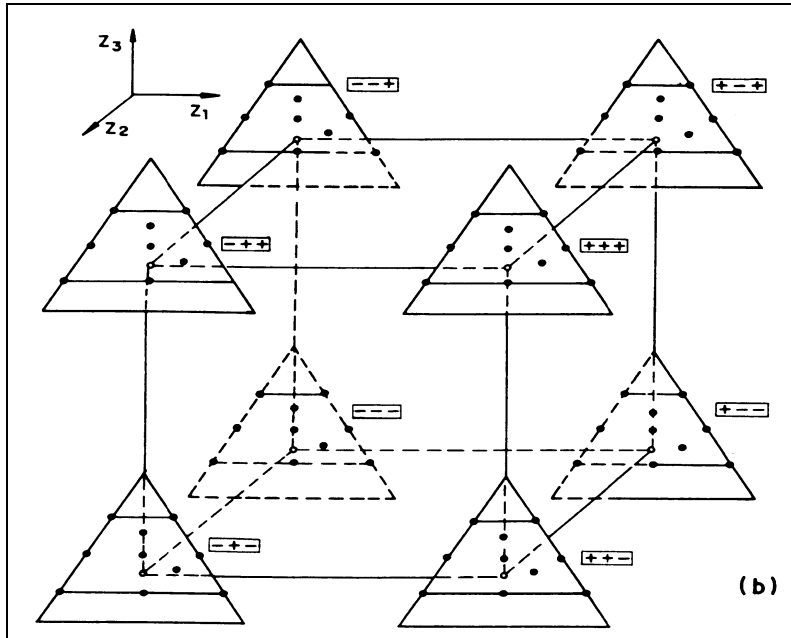


Figura 9: Planejamento "split-plot" com variáveis de processo representadas pelas arestas do cubo e em seus vértices as combinações de misturas utilizadas³⁶.

A Figura 10 mostra um planejamento simplificado com 2 variáveis de processo e 3 de misturas que exemplifica a estratégia adotada neste planejamento descrito.

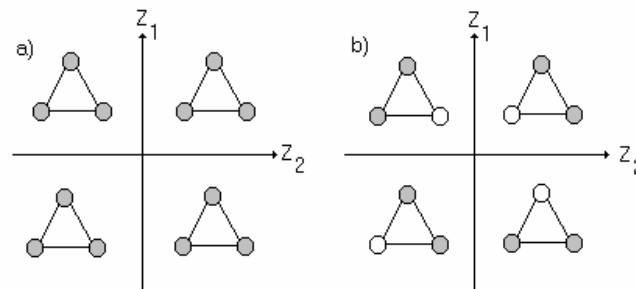


Figura 10 . Representação de um planejamento completo (a) e de um planejamento incompleto (b), onde "●" é uma composição de mistura utilizada e "○" é uma composição de mistura não utilizada.

O objetivo deste planejamento era descobrir se seria possível omitir a realização de experimentos referentes a algumas composições de misturas em certas condições de processo sem haver perda significativa de informações. Com este procedimento já se imaginava que poderia se estimular a falta de ajuste por diminuir o número de graus de liberdade na modelagem. Mas, por outro lado, com resultados que não indicassem perda significativa de informações na modelagem o método seria uma possibilidade muito interessante para reduzir o trabalho no laboratório.

Com o planejamento pronto foi realizada a regressão multivariada. Para cada planejamento foram gerados seis modelos combinados de “main-plot” e “sub-plot”: linear-linear, linear-quadrático, linear-cúbico especial, bilinear-linear, bilinear-quadrático e bilinear-cúbico especial. O modelo linear para o “main-plot” inclui os primeiros dois termos na Equação 1 e o bilinear inclui todos os termos desta equação. Os modelos linear, quadrático e cúbico especial do “sub-plot” são dados pelas Equações 2 – 4. Os modelos gerados foram comparados aos modelos gerados para o planejamento completo. A análise de variância para planejamentos incompletos “split-plot” é complexa exigindo que os cálculos sejam praticamente específicos para cada planejamento.

5.1.1 Resultados do planejamento incompleto I

A análise de variância para este planejamento incompleto, assim como para outros, necessita de equações mais gerais, pois ao calcular os termos da ANOVA nem todos receberão o mesmo peso previsto como no planejamento completo. A diminuição do número de experimentos em relação ao planejamento completo diminui o número de graus de liberdade na determinação da variância.

No planejamento completo o modelo matemático que apresentou melhores resultados foi o bilinear-quadrático, por isto dentre todos os modelos gerados para os planejamentos sempre se realizaram as comparações em relação a este modelo³⁷.

No planejamento I o modelo que apresentou maior concordância com o modelo obtido com o planejamento completo foi o que utilizou 8 composições de mistura em duplicatas por condição de processo, portanto, exigindo a execução de 128 experimentos. A Tabela 13 traz a ANOVA para este planejamento. O modelo foi validado pela determinação dos erros associados aos seus parâmetros. A Tabela 14 traz a ANOVA para o planejamento completo.

Tabela 13. ANOVA para o planejamento I com 8 composições de mistura por condição de processo.

Fonte	SQ	GL	MQ
Replicatas	0,005	1	0,005
Main-plot	1,220	7	0,174
Erro main-plot	0,003	7	0,0004
Sub-plot	5,766	7	0,823
Interação main-sub-plot	1,072	49	0,0219
Erro sub-plot	0,638	56	0,0114
Total	8,704		

Tabela 14. ANOVA para o planejamento completo

Fonte	SQ	GL	MQ
Replicatas	0,0064	1	0,006
Main-plot	17,224	7	2,461
Erro main-plot	0,0061	7	0,001
Sub-plot	69,219	9	0,769
Interação main-sub-plot	13,497	63	0,214
Erro sub-plot	0,0241	72	0,0003
Total	100,680	159	

A seguir são indicados os parâmetros significativos dos modelos obtidos pelo planejamento com 8 misturas em duplicatas, Equação 32, e com o planejamento completo, Equação 33.

$$\hat{y} = 0,909x_1 + 0,161 x_2 - 0,883 x_1x_3$$
$$(\pm 0,082) \quad (\pm 0,022) \quad (\pm 0,19)$$

Equação 32

$$\hat{y} = 0,914x_1 + 0,170 x_2 - 0,910 x_1x_3 + 0,2837 x_1z_1$$
$$(\pm 0,016) \quad (\pm 0,0074) \quad (\pm 0,033) \quad (\pm 0,017)$$

Equação 33

Para o cálculo dos erros utilizou-se a equação 8, porém como o planejamento não é completo realizou-se uma aproximação. As matrizes \mathbf{U}_1 e \mathbf{U}_2 utilizadas são de planejamentos completos, mas os termos σ_e^2 , σ_R^2 e σ_{RZ}^2 são provenientes da ANOVA específica para este planejamento. A aproximação de \mathbf{U}_1 e \mathbf{U}_2 se deu considerando que estes termos possuem como objetivo distribuir e dar pesos aos erros dos parâmetros, não alterando de forma significativa a dimensão dos erros associados.

A Tabela 15 traz as ANOVA obtidas para o planejamento I usando 3, 4, 5 e 6 replicatas e os seus respectivos modelos contendo apenas os termos significativos. Observa-se que os modelos se distanciam muito daquele obtido para o planejamento completo.

Tabela 15. ANOVA para o planejamento I com 3, 4, 5 e 6 replicatas.

Número de replicatas	Fonte	SQ	GL	MQ
3	Replicatas	0,0008	1	0,0008
	Main-plot	0,3636	7	0,0519
	Erro main-plot	0,0015	7	0,0002
	Sub-plot	2,1721	2	1,0860
	Interação mai-sub-plot	0,3306	14	0,0236
	Erro sub-plot	0,3837	16	0,0240
	Total	3,2523		
modelo	$\hat{y} = -1,227 x_1 - 1,247 x_2 + 3,296 x_1x_3 - 0,660 x_1z_1$			
4	Replicatas	0,0006	1	0,0006
	Main-plot	0,3842	7	0,0548
	Erro main-plot	0,0014	7	0,0002
	Sub-plot	2,3303	3	0,7767
	Interação mai-sub-plot	0,3937	21	0,0187
	Erro sub-plot	0,2276	24	0,0094
	Total	3,3379		
modelo	$\hat{y} = -3,034 x_1 + 1,438 x_2 + 3,386 x_1x_3 - 3,820 x_1z_1$			
5	Replicatas	0,0235	1	0,0235
	Main-plot	0,4548	7	0,0649
	Erro main-plot	0,0065	7	0,0009
	Sub-plot	2,9577	4	0,7394
	Interação mai-sub-plot	0,3411	28	0,0121
	Erro sub-plot	0,2027	32	0,0063
	Total	3,9867		
modelo	$\hat{y} = 1,006 x_1 + 0,482 x_2 + 2,745 x_1x_3 + 2,765 x_1z_1$			
6	Replicatas	0,0044	1	0,0044
	Main-plot	0,8452	7	0,1207
	Erro main-plot	0,0038	7	0,0005
	Sub-plot	2,2215	5	0,4443
	Interação mai-sub-plot	0,9763	35	0,0278
	Erro sub-plot	0,6955	40	0,0173
	Total	4,7468		
modelo	$\hat{y} = -1,368 x_1 - 0,626 x_2 + 3,704 x_1x_3 + 3,347 x_1z_1$			

5.2 Estudos com planejamentos contendo o número completo de misturas, porém com o número de duplicatas reduzidas (II).

Utilizando novamente o conjunto de dados da referência 36 uma outra estratégia de planejamento incompleto foi testada. Neste caso todas as composições de mistura são executadas, todavia apenas algumas em duplicata. O estudo compreendeu casos que possuíam de 3 a 9 duplicatas. Há a necessidade das duplicatas para que se possa realizar a ANOVA e obter as estimativas dos

erros presentes. Para o caso contendo 3 duplicatas existe a necessidade de executar 104 experimentos, pois são 6 experimentos correspondendo as duplicatas e 7 experimentos não duplicados para cada condição de processo que são ao todo 8.

A Figura 11 mostra um planejamento simplificado com 2 variáveis de processo e três de misturas que exemplifica a estratégia desta segunda forma de planejamento.

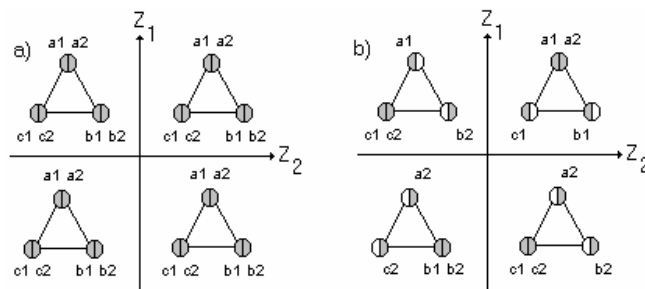


Figura 11. Planejamento completo (a) e planejamento incompleto (b), onde “1” representa composição em duplicata e “0” composição sem duplicata.

Depois de elaborar os detalhes do planejamento foi realizada a regressão multivariada. Isto foi feito para 3, 4, 5, 6, 7, 8 e 9 duplicatas. Para cada caso obtiveram-se seis modelos: linear-linear, linear-quadrático, linear-cúbico especial, bilinear-linear, bilinear-quadrático e bilinear-cúbico especial. Os modelos foram comparados aos obtidos com o planejamento completo.

Neste segundo planejamento evita-se a falta de ajuste do modelo, observado no estudo I, pois resultados para todas as composições das misturas estão sendo incluídos no cálculo para cada condição de processo.

5.2.1 Resultados do planejamento incompleto II

Por ser um planejamento incompleto a ANOVA recebe um tratamento diferenciado em relação ao do planejamento completo, pois nem todos os termos são calculados com todos os graus de liberdade inerentes. Como todas as composições de mistura são executadas ao menos uma vez a redução dos graus

de liberdade é menor se comparada ao planejamento I. Dos modelos gerados por regressão o modelo bilinear-quadrático foi escolhido para ser comparado com o obtido do planejamento completo. Dentre os planejamentos executados o que possuía 4 replicatas e 6 medidas sem replicatas já apresentou resultados comparáveis ao obtido com o planejamento completo, pois permitiu a obtenção de um modelo por regressão cujos termos significativos eram semelhantes aos do obtido pelo planejamento completo³⁷. Para este planejamento é necessária a execução de 112 experimentos. A Tabela 16 traz a ANOVA para este planejamento.

Tabela 16. ANOVA para o planejamento II com 4 duplicatas e seis medidas sem replicatas por condição de processo.

Fonte	SQ	GL	MQ
Replicatas	0,090	1	0,090
Main-plot	1,741	7	0,249
Erro main-plot	0,143	7	0,020
Sub-plot	5,175	9	0,575
Interação mai-sub-plot	0,337	63	0,005
Erro sub-plot	0,083	24	0,003
Total	7,569	111	

O modelo gerado foi validado com a determinação dos erros associados aos parâmetros. O modelo obtido contendo os parâmetros significativos é indicado a seguir, Equação 34, em comparação com o modelo do planejamento completo, Equação 35.

$$\hat{y} = 0,885x_1 + 0,169x_2 - 0,860x_1x_3 + 0,261x_1z_1$$

$$(\pm 0,052) \quad (\pm 0,027) \quad (\pm 0,11) \quad (\pm 0,059)$$

Equação 34

$$\hat{y} = 0,914x_1 + 0,170x_2 - 0,910x_1x_3 + 0,284x_1z_1$$

$$(\pm 0,016) \quad (\pm 0,0074) \quad (\pm 0,033) \quad (\pm 0,017)$$

Equação 35

A estratégia utilizada na determinação dos erros da Equação 34 foi a mesma do planejamento I. A Tabela 17 traz a ANOVA para os conjuntos de dados com 3, 5, 6, 7, 8 e 9 duplicatas, além dos modelos com parâmetros correspondentes ao do planejamento completo.

Comparando-se os resultados obtidos com o planejamento I e II observa-se que o segundo possui mais vantagens. A estratégia I exigiu a execução de 128 experimentos enquanto a segunda exigiu a realização de 112, para obter modelos semelhantes, o que significa uma diferença de 16 experimentos a menos para serem executados no planejamento. Segundo, o modelo obtido com a estratégia II possui os mesmos parâmetros significativos do modelo para o planejamento completo, enquanto o planejamento I possui um modelo com um parâmetro significativo a menos referente à interação x_1z_1 .

Os valores das médias quadráticas na Tabela 17 foram plotados no gráfico da Figura 12. Observando a Figura 12 é possível notar que o valor da média quadrática “sub-plot” aumenta com a elevação do número de replicatas por planejamentos. Os valores referentes ao “main-plot” permanecem relativamente constantes. O gráfico indicado na Figura 13, é uma ampliação da Figura 12 na escala MQ. Observa-se que a média quadrática das replicatas diminui com o aumento do número de replicatas, a média quadrática do erro “sub-plot” e do erro “main-plot” apresentam o mesmo comportamento. Por outro lado o efeito de interação “main-plot–sub-plot” possui sua média quadrática elevada com o aumento de replicatas. Nota-se que os dados utilizados a partir de 7 replicatas geram somas sem grandes variações entre elas.

Tabela 17. ANOVA para os conjuntos de dados com 3, 5, 6, 7, 8 e 9 duplicatas no planejamento II.

Número de replicatas	Fonte	SG	GL	MQ
3	Replicatas	0,0575	1	0,0574
	Main-plot	1,3397	7	0,1913
	Erro main-plot	0,2002	7	0,0286
	Sub-plot	5,0614	9	0,5623
	Interação main-sub-plot	0,2605	63	0,0041
	Erro sub-plot	0,2391	16	0,0149
	Total	7,1585	103	
modelo	$\hat{y} = 0,9191 x_1 + 0,168 x_2 - 0,915 x_1x_3 + 0,257 x_1z_1$			
5	Replicatas	0,0953	1	0,0952
	Main-plot	1,0787	7	0,1540
	Erro main-plot	0,2019	7	0,0288
	Sub-plot	4,829	9	0,5366
	Interação main-sub-plot	0,3426	63	0,0054
	Erro sub-plot	0,4527	32	0,0141
	Total	6,9999	119	
modelo	$\hat{y} = 0,951 x_1 + 0,171 x_2 - 1,031 x_1x_3 + 0,334 x_1z_1$			
6	Replicatas	0,0013	1	0,0013
	Main-plot	1,2081	7	0,1725
	Erro main-plot	0,0451	7	0,0064
	Sub-plot	4,8727	9	0,5414
	Interação main-sub-plot	0,4186	63	0,0066
	Erro sub-plot	0,4218	41	0,0102
	Total	6,9678	128	
modelo	$\hat{y} = 0,930 x_1 + 0,174 x_2 - 0,908 x_1x_3 + 0,293 x_1z_1$			
7	Replicatas	0,0265	1	0,0265
	Main-plot	1,6324	7	0,2332
	Erro main-plot	0,0581	7	0,0083
	Sub-plot	6,4071	9	0,7119
	Interação main-sub-plot	0,8277	63	0,0131
	Erro sub-plot	0,1538	48	0,0032
	Total	9,1058	135	
modelo	$\hat{y} = 0,943 x_1 + 0,167 x_2 - 0,977 x_1x_3 + 0,275 x_1z_1$			
8	Replicatas	1,88E-05	1	1,88E-05
	Main-plot	1,5364	7	0,2194
	Erro main-plot	0,0419	7	0,0060
	Sub-plot	6,4356	9	0,7150
	Interação main-sub-plot	1,0069	63	0,0159
	Erro sub-plot	0,2756	56	0,0049
	Total	9,2967	143	
modelo	$\hat{y} = 0,968 x_1 + 0,172 x_2 - 1,059 x_1x_3 + 0,225 x_1z_1$			
9	Replicatas	0,0051	1	0,0051
	Main-plot	1,6718	7	0,2388
	Erro main-plot	0,0102	7	0,0014
	Sub-plot	6,7642	9	0,7515
	Interação main-sub-plot	1,2091	63	0,0191
	Erro sub-plot	0,1720	64	0,0027
	Total	9,8325	151	
modelo	$\hat{y} = 0,96220 x_1 + 0,16968 x_2 - 1,03456 x_1x_3 + 0,23102 x_1z_1$			

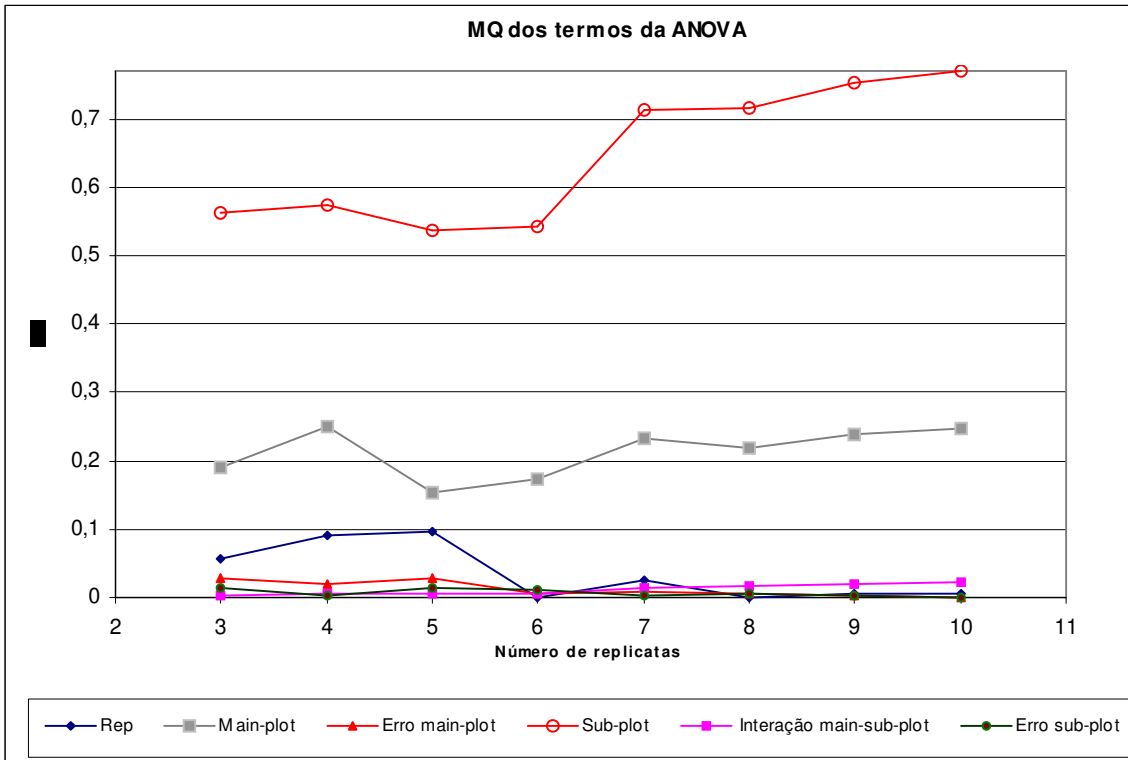


Figura 12. Média quadrática dos termos da ANOVA da Tabela 7 versus o número de replicatas por planejamento para o estudo II

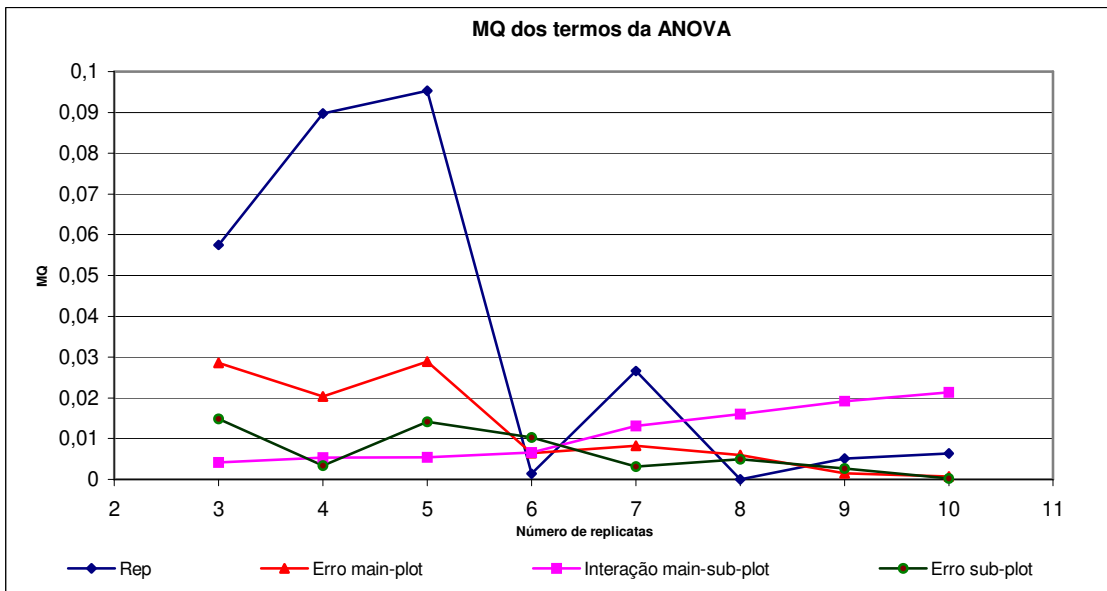


Figura 13. Média quadrática dos termos da ANOVA da Tabela 7 versus o número de replicatas por planejamento para o estudo II com escala ampliada.

Os parâmetros dos modelos obtidos por regressão para os planejamentos com 3, 4, 5, 6, 7, 8 e 9 replicatas foram plotados no gráfico da Figura 14. É possível observar que as flutuações dos valores dos parâmetros são relativamente pequenas, mesmo variando muito o número de replicatas utilizadas de 3 a 10.³⁷

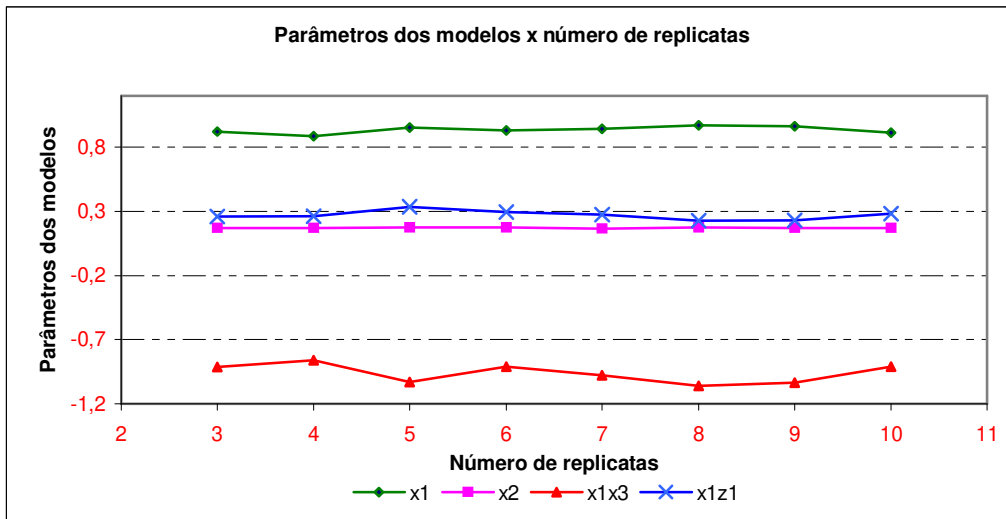


Figura 14. Parâmetros dos modelos versus número de replicatas por planejamentos no estudo II

Assim o planejamento II mostrou ser mais adequado que o I, além de indicar que a redução de experimentos é possível³⁷.

6 GRÁFICOS DE PROBABILIDADE ACUMULADA EM PLANEJAMENTOS SPLIT-PLOT

Devido ao erro experimental uma dispersão natural ocorre nas medidas. Ao se modelar um sistema provavelmente alguns parâmetros do modelo estarão descrevendo somente as variações ocasionadas por este erro. Mas obviamente estes parâmetros não estão descrevendo efeitos significativos das variáveis de interesse e, portanto, devem ser descartados. Uma das formas de avaliar a significância de um parâmetro é compará-lo com seu erro. Assim um teste t com graus de liberdade adequados poderá indicar os valores de corte dos parâmetros. Contudo em planejamentos do tipo “split-plot” muitas vezes a determinação dos graus de liberdade para o teste t não é simples e até imprecisa⁵. É neste tipo de situação que os gráficos de probabilidade acumulada podem ser uma ferramenta extremamente útil.

Sabe-se que estatisticamente os parâmetros que descrevem erros aleatórios são exemplos de hipótese nula, uma vez que o valor verdadeiro de cada um deles seria zero. Portanto, estes parâmetros, se colocados num gráfico em papel de probabilidade normal, devem seguir uma reta centrada em zero. Os efeitos significativos não se incluem na reta, pois não fazem parte da mesma distribuição. A Figura 15 é um exemplo de gráfico de probabilidade acumulada, que será discutido mais adiante.

Para se utilizar o gráfico de probabilidade acumulada deve-se tomar alguns cuidados. Primeiro, tentar plotar simplesmente os valores dos parâmetros no gráfico, como feito para planejamentos fatoriais em variáveis de processo, trará provavelmente resultados incorretos, pois muitos parâmetros que possam ter valores altos podem também possuir um grande erro associado, assim sua significância pode ser nula. Segundo, diferentes fontes de erro são agregadas ao cálculo dos erros dos parâmetros, desta forma a dimensão do erro pode variar muito entre os diferentes tipos de parâmetros (como os do “sub-plot”, “main-sub-plot” e “main-plot”). Assim na estratégia adotada se assumiu que os valores dos parâmetros não significativos simplesmente estão modelando erro e, portanto,

devem seguir uma distribuição normal. Todavia devido as diferentes ordens de grandeza dos parâmetros os mesmos devem ser “reescalados” para serem comparados. Isto é feito dividindo os valores dos parâmetros pelos valores dos erros. Com isso se obtém a razão que indica o quanto um parâmetro supera seu erro. Obviamente os parâmetros que mais superarem seus erros serão os mais significativos, e aqueles que não são significativos seguirão uma distribuição normal, cujo gráfico de probabilidade acumulada é uma reta centrada em zero¹.

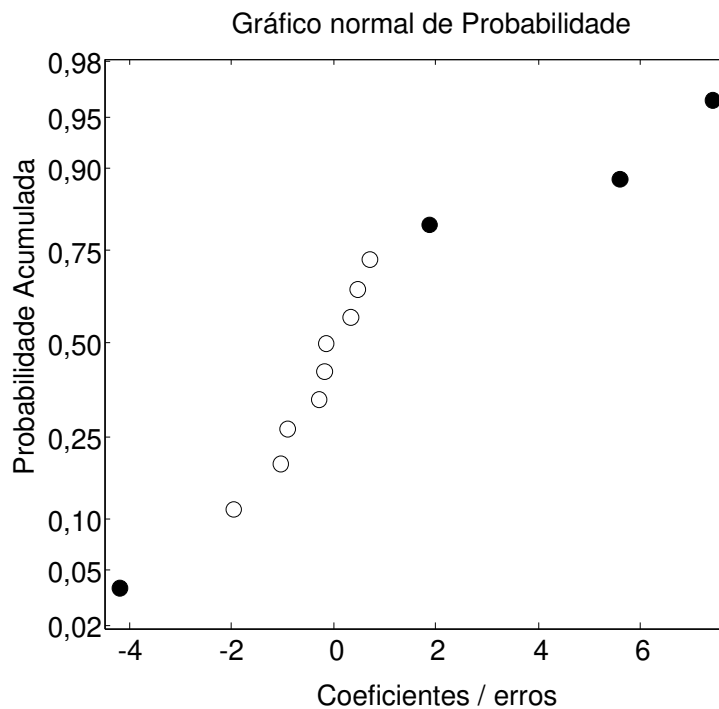


Figura 15. Gráfico de probabilidade acumulada para efeitos significativos (●) e não significativos (○).

6.1 Modelos aproximados para planejamentos split-plot

Normalmente planejamentos “split-plot” envolvem um grande número de variáveis e os modelos ajustados podem possuir um número elevado de termos. Um modelo fatorial trilinear, por exemplo, para três variáveis de processo^{7,8}.

$$\hat{y} = \alpha_0 + \sum_{i=1}^3 \alpha_i z_i + \sum_{i < j} \sum_{j=1}^3 \alpha_{ij} z_i z_j + \alpha_{123} z_1 z_2 z_3$$

Equação 36

e especial cúbico para três variáveis de mistura^{9,10}

$$\hat{y} = \sum_{i=1}^3 \beta_i x_i + \sum_{i < j} \sum_{j=1}^3 \beta_{ij} x_i x_j + \beta_{123} x_1 x_2 x_3$$

Equação 37

podem ser multiplicados gerando um modelo com $7 \times 8 = 56$ termos envolvendo variáveis de processo e de mistura. Este modelo descreverá efeitos de interação mostrando como os níveis dos fatores podem influenciar as propriedades das misturas ou como diferentes composições de misturas produzem diferentes efeitos quando se alteram os níveis dos fatores de processo. Outros modelos também podem ser ajustados. Como já exposto, muitos termos podem não ser significativos e, portanto, diferenciados em um gráfico de probabilidade acumulada. Porém em planejamentos “split-plot” há mais de uma fonte de erro afetando os parâmetros além dos mesmos serem de ordens diferentes. Portanto, há a necessidade de corrigir os parâmetros para que todos possam ser plotados em um mesmo gráfico de probabilidade acumulada. Nos estudos realizados a correção foi feita dividindo-se cada parâmetro por seu respectivo erro ou por uma quantidade relacionada com este erro. As correções dos parâmetros podem ser obtidas pela raiz quadrada dos elementos da diagonal da matriz de covariância⁴

$$Cov = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$$

Equação 38

6.2 Aplicação de gráficos de probabilidade acumulada em estudos envolvendo dados reais

Vários estudos foram realizados com dados reais. O planejamento “split-plot” foi utilizado em um projeto de pós-doutoramento³⁸ em um estudo para modelar e otimizar um sistema de extração líquido-líquido por fase única, além de ter sido aplicado em estudos para a determinação de chumbo por ASV (voltametria de redissolução anódica) em um projeto de doutoramento³⁹. Em ambos os projetos o programa desenvolvido em nosso laboratório foi aplicado no tratamento dos dados.

Através do método “split-plot” foi realizada a otimização da determinação de chumbo por ASV. O chumbo é tóxico mesmo em baixas quantidades, devido ao seu efeito cumulativo nos organismos vivos. Dentre muitos métodos, a voltametria de redissolução anódica (ASV), tem sido muito utilizada em análises de traços de metais, pelo seu baixo custo, simplicidade de uso e baixos limites de detecção.

A aplicação da ASV em amostras de matrizes complexas é complicada devido ao efeito da matriz e aos deslocamentos dos potenciais dos íons estudados. Visando minimizar estes problemas, resolveu-se aplicar o Sistema Ternário Homogêneo de Solventes (STHS) composto por N,N'-DMF/etanol/água, para determinar Pb(II) em matrizes agroambientais. A Figura 16 mostra o planejamento “split-plot” com variáveis de misturas formando o “sub-plot” e as variáveis de processo o “main-plot”. Três variáveis de mistura: DMF (x_1), etanol (x_2) e água (x_3) e duas variáveis de processo: concentração de acetato de amônio, Z_1 , (eletrólito suporte) e ácido clorídrico, Z_2 , foram conjuntamente ajustadas.

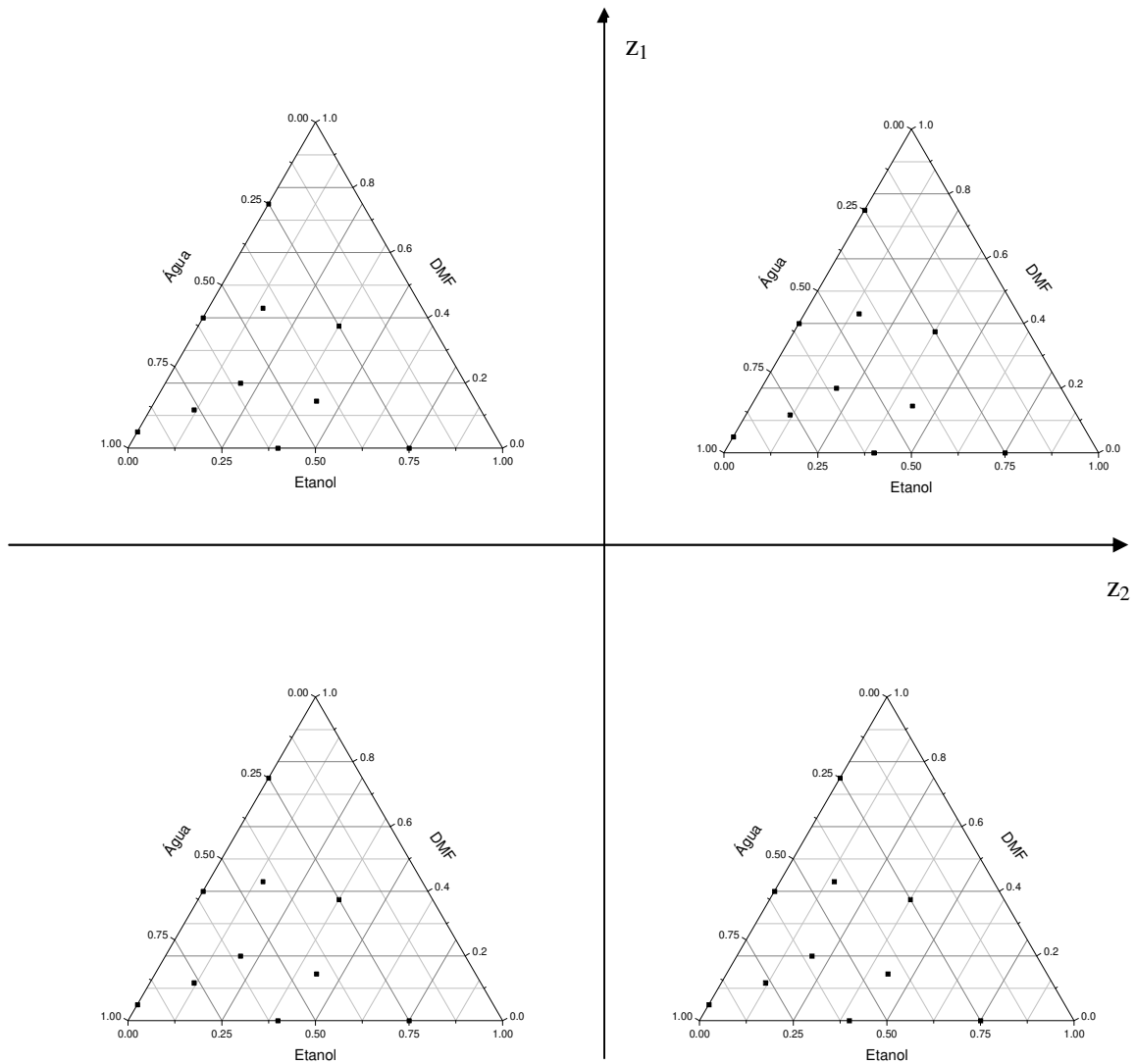


Figura 16. Planejamento experimental para determinação de chumbo com 3 componentes de mistura em um arranjo fatorial 2^2 para variáveis de processo.

A Tabela 18 indica as 10 composições de misturas e as variáveis de processo escaladas.

Tabela 18. Planejamento fatorial 2^2 com acetato de amônio e ácido clorídrico como variáveis de processo e DMF, etanol e água como variáveis de mistura em um planejamento “split-plot”³⁹.

Main-plot	z_1^a	z_2^a	$C_{\text{acetato}} / \text{mol L}^{-1}$	$C_{\text{HCl}} / \text{mol L}^{-1}$
1	-1	-1	0,1	$8,8 \times 10^{-3}$
2	+1	-1	0,3	$8,8 \times 10^{-3}$
3	-1	+1	0,1	$2,1 \times 10^{-1}$
4	+1	+1	0,3	$2,1 \times 10^{-1}$

a) $z_1 = \frac{C_{\text{acetato}} - 0.2}{0.10}$; $z_2 = \frac{C_{\text{HCl}} - 0.1094}{0.1006}$ onde os valores tem unidade mol L⁻¹.

Sub-plot	DMF (x_1)	Etanol (x_2)	Água (x_3)
1	0,750	0	0,250
2	0	0,750	0,250
3	0,050	0	0,950
4	0,400	0	0,600
5	0	0,400	0,600
6	0,375	0,375	0,250
7	0,200	0,200	0,600
8	0,113	0,113	0,775
9	0,144	0,431	0,425
10	0,431	0,144	0,425

A resposta analítica observada nos experimentos (corrente absoluta em nA) é indicada na Tabela 19.

Tabela 19. Determinação experimental de chumbo para 40 experimentos em duplicatas em um planejamento “split-plot” com duas variáveis de processo e três de misturas. A resposta analítica é dada como valores de corrente de pico em nA.

Número da formulação ^a	$z_1 = -1^b$		$z_1 = +1$		$z_1 = -1$		$z_1 = +1$	
	$z_2 = -1$		$z_2 = -1$		$z_2 = +1$		$z_2 = +1$	
	R_1^c	R_2	R_1	R_2	R_1	R_2	R_1	R_2
1	42,69	44,60	0,00	0,00	0,00	0,00	0,00	0,00
2	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
3	97,66	70,42	84,87	105,00	96,68	122,00	89,37	90,13
4	61,07	61,61	38,71	57,30	66,64	40,00	48,55	41,74
5	38,90	61,12	46,18	50,58	37,01	35,73	6,89	9,99
6	80,59	114,00	51,05	35,73	75,74	64,31	73,87	28,06
7	33,08	34,75	48,70	38,38	46,71	41,81	48,51	31,39
8	67,60	68,95	74,53	96,08	53,87	62,88	25,63	22,29
9	35,29	29,17	28,93	32,44	44,30	40,38	15,27	17,31
10	38,76	34,21	28,10	42,75	42,67	55,92	38,39	40,34

^a Número da formulação da Tabela 18.

^b O valor $z_1 = z_2 = 0$ equivale a 0,2 mol kg⁻¹ de acetato de amônio e 0,120 mol kg⁻¹ de HCl.

^c Número da replicata.

Diferentes modelos foram ajustados combinando-se os modelo linear e bilinear para processo com os modelos linear, quadrático, cúbico especial e cúbico para misturas. Seis modelos combinando variáveis de processo e de mistura foram investigados. A Tabela 20 traz a ANOVA para o modelo de processo linear, e os modelos linear, quadrático e cúbico especial para misturas.

Tabela 20. Tabela de ANOVA “split-plot” para o modelo linear para variáveis de processo combinado com os modelos linear, quadrático e especial cúbico para variáveis de mistura.

Modelo	Fonte	GL	SQ	MQ
	Reps (R)	1	2,65	2,65
	Main-plot (Z)	3	4110,78	1370,26
Linear	ZREG	2	3852,40	1926,20
	ZLOF	1	258,39	258,39
	Erro (RZ)	3	400,42	133,47
	Sub-plot (X)	9	52770,00	5863,40
Linear	XREG	2	29375,00	14697,50
	XLOF	7	23375,00	3339,29
Quadrático	XREG	5	49820,00	9964,00
	XLOF	4	2950,00	737,50
Cúbico especial	XREG	6	52682,00	8780,33
	XLOF	3	88,00	29,33
	Main x Sub-plot int.	27	10513,44	389,39
Linear	ZXREG	4	679,00	169,75
	ZXLOF	23	9834,00	427,57
Quadrático	ZXREG	10	2960,00	296,00
	ZXLOF	17	7553,00	444,29
Cúbico especial	ZXREG	12	4234,00	352,83
	ZXLOF	15	6279,00	418,60
	Erro sub-plot	36	3830,59	106,41
	Total	79	71627,50	

A Tabela 21 traz os mesmos modelos de misturas combinados com o modelo bilinear para processo. Analisando a falta de ajuste de cada modelo elegeu-se o modelo bilinear-cúbico-especial como o preferido, pois ele não apresenta falta de ajuste na parte do modelo que se refere ao “main-plot” e entre os modelos de mistura o modelo cúbico-especial apresenta a menor falta de ajuste. O erro padrão dos parâmetros do modelo, indicados na Tabela 22, foram calculados utilizando os resultados da ANOVA da Tabela 21 e o método descrito por Cornell⁵.

Tabela 21. Tabela de ANOVA “split-plot” para o modelo bilinear para variáveis de processo combinado com os modelos linear, quadrático e especial cúbico para variáveis de mistura.

Modelo	Fonte	GL	SQ	MQ
	Reps (R)	1	2,65	2,65
	Main-plot (Z)	3	4110,78	1370,26
Bilinear	ZREG	3	4110,78	1370,26
	ZLOF	0	-	-
	Error (RZ)	3	400,42	133,47
	Sub-plot (X)	9	52770,00	5863,40
Linear	XREG	2	29375,00	14697,50
	XLOF	7	23375,00	3339,29
Quadrático	XREG	5	49820,00	9964,00
	XLOF	4	2950,00	737,50
Cúbico especial	XREG	6	52682,00	598,66
	XLOF	3	88,00	29,33
	Main x Sub-plot int.	27	10513,44	389,39
Linear	ZXREG	6	3708,00	618,00
	ZXLOF	21	6805,00	324,05
Quadrático	ZXREG	15	6250,00	416,67
	ZXLOF	12	4263,00	355,25
Cúbico especial	ZXREG	18	7760,44	431,14
	ZXLOF	9	2753,00	305,89
	Erro sub-plot	36	3830,59	106,41
	Total	79	71627,50	

Não há um método exato para determinar o valor de corte pelo teste t , $t = b_k^l / se(b_k^l)$, onde $se(b_k^l)$ representa o erro padrão envolvendo a k -ésima variável de mistura e a l -ésima variável de processo. Assim, a investigação para se obter o valor crítico de t foi realizada com o gráfico de probabilidade acumulada. O gráfico é indicado na Figura 17. Para uma aplicação apresentada na referência 5, Cornell usa o valor de t igual ou maior que 3 para indicar os coeficientes mais significativos do modelo. No gráfico normal de probabilidade os pontos correspondentes a estes valores de t são indicados por círculos sólidos, enquanto que os triângulos sólidos correspondem a valores de t entre 2 e 3.

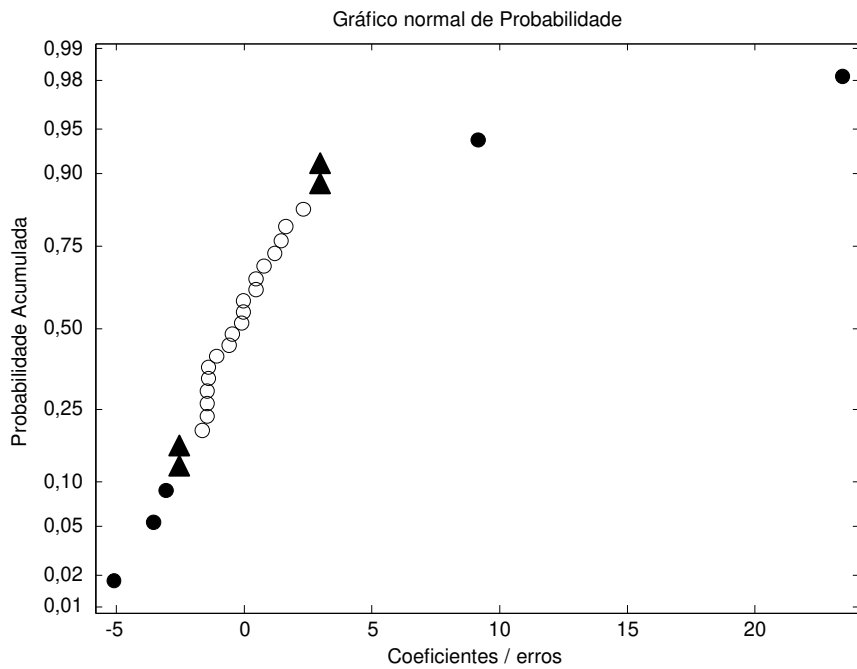


Figura 17. Gráfico normal de probabilidade com valores do teste-t da Tabela 22. Os círculos sólidos correspondem a valores de t maiores que 3 e os triângulos sólidos correspondem a valores de t entre 2 e 3.

Aplicando este valor de corte para o modelo bilinear-cubico-especial os termos significativos são x_3 , x_1x_2 , $x_1x_2x_3$, $x_1x_2z_1$ e $x_2x_3z_1$. O ajuste por mínimos quadrados apenas destes termos resulta em

$$\hat{y} = 99,94x_3 + 868,58x_1x_2 - 1711,02x_1x_2x_3 - 108,13x_1x_2z_1 - 30,99x_2x_3z_2.$$

$(\pm 4,30)$ $(\pm 107,72)$ $(\pm 333,17)$ $(\pm 31,49)$ $(\pm 13,44)$

Equação 39

Repetindo este procedimento com termos que possuem os valores de t maior que 2 na Tabela 22 fornece

$$\hat{y} = 98,59x_3 + 800,86x_1x_2 - 68,24x_2x_3 - 1481,58x_1x_2x_3 - 201,79x_1x_2z_1 + 239,97x_1x_2x_3z_1$$

$(\pm 4,03)$ $(\pm 100,02)$ $(\pm 20,06)$ $(\pm 310,57)$ $(\pm 89,10)$ $(\pm 264,81)$

$$- 38,28x_2x_3z_2 + 86,58x_1x_2x_3z_2 - 6,73x_3z_1z_2$$

$(\pm 15,10)$ $(\pm 106,13)$ $(\pm 2,77)$

Equação 40

Tabela 22. Parâmetros, estimativas do erro padrão e razão para o teste t para o modelo combinado bilinear-cúbico-especial.

Variáveis	Parâmetros	Erro padrão	t calculado
x1	-17,46	11,54	-1,51
x2	-12,58	10,84	-1,16
x3	99,38	4,25	23,41
x1x2	786,59	87,43	9,00
x1x3	-2,63	37,25	-0,07
x2x3	-81,58	33,28	-2,45
x1x2x3	-1433,00	278,16	-5,15
x1z1	-16,88	11,44	-1,48
x2z1	13,03	11,13	1,17
x3z1	-2,96	4,45	-0,67
x1x2z1	-270,42	86,88	-3,11
x1x3z1	15,73	36,39	0,43
x2x3z1	-50,22	34,61	-1,45
X1x2x3z1	715,55	279,16	2,56
x1z2	-16,55	11,25	-1,47
x2z2	31,47	10,77	2,92
x3z2	-0,70	4,36	-0,16
x1x2z2	-148,16	86,61	-1,71
x1x3z2	14,61	36,04	0,41
x2x3z2	-119,91	33,09	-3,62
x1x2x3z2	633,91	276,29	2,29
x1z1z2	18,10	11,25	1,61
x2z1z2	7,61	10,77	0,71
x3z1z2	-9,89	4,36	-2,27
x1x2z1z2	122,63	86,61	1,42
x1x3z1z2	-2,21	36,039	-0,06
x2x3z1z2	-17,20	33,089	-0,52
x1x2x3z1z2	-411,49	276,29	-1,49

A Figura 18 contém um gráfico de valores experimentais versus valores previstos da Equação 40. O desvio corresponde a um erro quadrático médio de 15,92. Os pontos estão distribuídos aleatoriamente sem evidências de falta de ajuste. Um gráfico similar com a Equação 39 possui características semelhantes, mas com um erro de 20,49.

Vale notar que a Equação 39 não contém qualquer termo significativo cruzado z_1z_2 , enquanto a Equação 40 tem. Estas observações são consistentes com os resultados da ANOVA na Tabela 20. A adição do termo cruzado z_1z_2 -variável de mistura melhora levemente a falta de ajuste do modelo. Isto pode ser visto comparando os valores das médias quadráticas de falta de ajuste e regressão da interação “main-sub-plot” na Tabela 20 e Tabela 21.

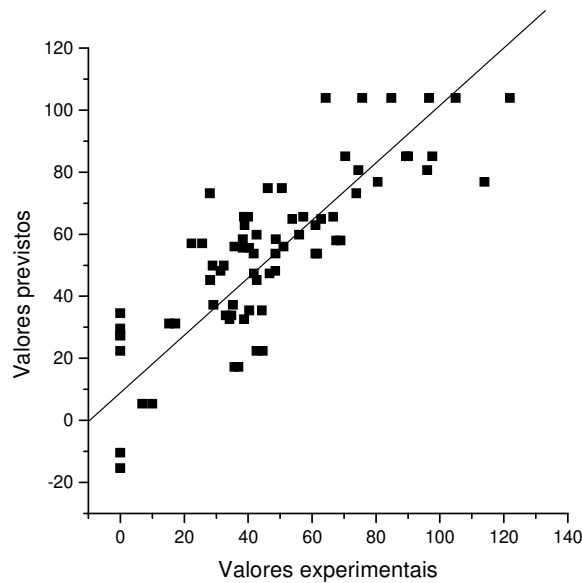


Figura 18. Gráfico de valores previstos pela Equação 40 versus valores experimentais.

A presença de termos cruzados significativos xz no modelo mostra que a superfície de resposta das variáveis de mistura depende dos níveis de concentração do acetato e ácido. Realizando o ajuste por mínimos quadrados dos dados das misturas por condição de processo resulta nos seguintes modelos de mistura

$$z_1 = z_2 = -1$$

$$\hat{y} = 34,07x_1 - 49,47x_2 + 93,15x_3 + 1327,8x_1x_2 - 35,18x_1x_3 + 71,35x_2x_3 - 3193,95x_1x_2x_3$$

Equação 41

$$z_1 = +1, z_2 = -1$$

$$\hat{y} = -35,89x_1 - 38,63x_2 + 107,01x_3 + 541,7x_1x_2 + 0,7x_1x_3 + 5,31x_2x_3 - 939,87x_1x_2x_3$$

Equação 42

$$z_1 = -1, z_2 = +1$$

$$\hat{y} = -35,23x_1 - 1,75x_2 + 111,53x_3 + 786,22x_1x_2 - 1,54x_1x_3 - 134,07x_2x_3 - 1103,15x_1x_2x_3$$

Equação 43

$$z_1 = z_2 = +1$$

$$\hat{y} = -32,79x_1 + 39,53x_2 + 85,83x_3 + 490,64x_1x_2 + 25,5x_1x_3 - 268,91x_2x_3 - 495,03x_1x_2x_3$$

Equação 44

Estas equações podem ser comparadas com as equações obtidas em cada uma das quatro condições de processo usando as Equação 39 e Equação 40. Substituindo os valores apropriados de z_1 e z_2 na Equação 40 resulta nas equações

$$z_1 = z_2 = -1$$

$$\hat{y} = 91,86x_3 + 1002,65x_1x_2 - 29,96x_2x_3 - 1862,13x_1x_2x_3$$

Equação 45

$$z_1 = +1, z_2 = -1$$

$$\hat{y} = 105,30x_3 + 599,07x_1x_2 - 29,96x_2x_3 - 1274,19x_1x_2x_3$$

Equação 46

$$z_1 = -1, z_2 = +1$$

$$\hat{y} = 105,32x_3 + 1002,65x_1x_2 - 106,52x_2x_3 - 1688,97x_1x_2x_3$$

Equação 47

$$z_1 = z_2 = +1$$

$$\hat{y} = 91,86x_3 + 599,07x_1x_2 - 106,52x_2x_3 - 1101,03x_1x_2x_3$$

Equação 48

Os coeficientes nestas equações são consistentes com os da Equação 41 a Equação 44 obtidos por regressão linear dos dados nos “main-plots” individuais. Os dois conjuntos de equações têm coeficientes x_3 mais positivos para $z_1 = +1, z_2 = -1$ e $z_1 = -1, z_2 = +1$ nas condições do “main-plot”. Isto é devido a significância do termo $x_3z_1z_2$ na Equação 40. Os coeficientes x_1x_2 mais positivos ocorrem nas equações para as condições $z_1 = z_2 = -1$ e $z_1 = -1, z_2 = +1$ no “main-plot”. Este comportamento pode ser atribuído ao termo $x_1x_2z_1$ na Equação 40. Além disso, os valores relativos dos coeficientes $x_1x_2x_3$ da Equação 45 a Equação 48 são muito concordantes com os da Equação 41 a Equação 44³⁹.

Como há termos cruzados xz significativos na Equação 39 e Equação 40, pode-se esperar diferentes condições de otimização para cada quadrante do “main-plot”. A Figura 19 mostra as superfícies de resposta para cada quadrante. De acordo com a Figura 19 a mistura de solventes mais adequados depende da concentração do eletrólito e do ácido. Entre os pontos no quadrante (- -), na Figura 19, a detecção máxima do Pb^{2+} é observada para o sistema ternário homogêneo de solventes dado por 40% m/m de DMF, 35% m/m de etanol e 25% m/m de água, correspondendo à mistura de 8,0 g de DMF, 7,0 g de etanol e 5,0 g de água. Em torno deste ponto, há uma tendência clara para o aumento do sinal voltamétrico para o chumbo na medida em que a proporção de água diminui. Para os outros três quadrantes do planejamento “split-plot”, quando um dos dois ou ambos, eletrólito suporte e ácido possuem concentração alta, uma solução aquosa quase pura deve ser usada para maximizar a resposta analítica.

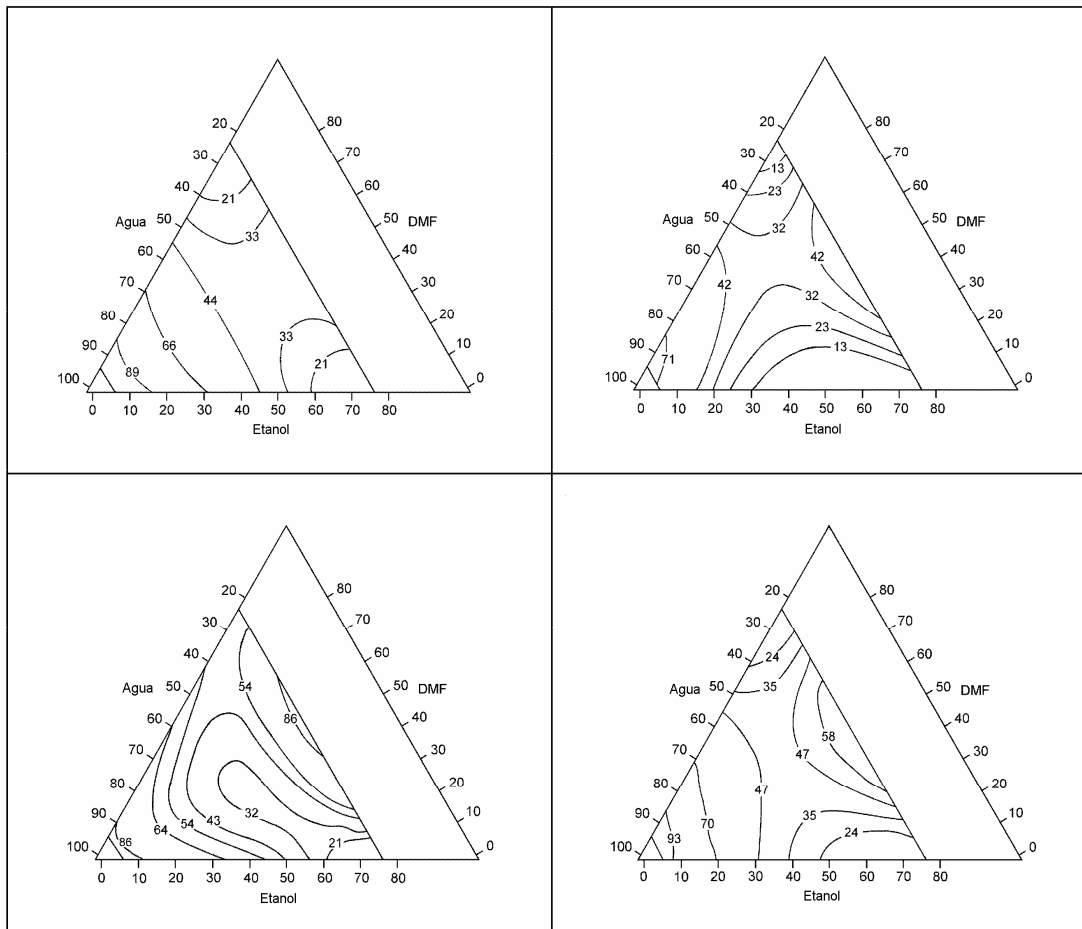


Figura 19. Superfície de resposta das misturas para cada condição determinada pelas variáveis de processo.

Para comprovar a robustez e confiabilidade do uso dos gráficos normais de probabilidade outros conjuntos de dados reais foram utilizados em vários estudos⁴⁰.

Na Tabela 6 da referência 5 Cornell lista valores de respostas para a espessura do vinil usando um planejamento “split-plot” para três plasticidas em duas diferentes velocidades de extrusão e duas temperaturas de secagem. Cinco composições de misturas foram investigadas em quatro condições de processos diferentes. O planejamento “split-plot” foi executado em dois blocos de replicatas cuja análise permitiu obter o modelo

$$\hat{y} = 11,48 x_1 - 69,54 x_2 - 2,63 x_3 + 148,64 x_1 x_2 + 0,26 x_1 z_1 + 1,26 x_2 z_1 - 0,70 x_3 z_1 + 2,96 x_1 z_2 - 0,55 x_2 z_2 - 2,82 x_3 z_2 - 0,57 x_1 z_1 z_2 - 1,57 x_2 z_1 z_2 + 2,31 x_3 z_1 z_2$$

$\begin{matrix} * & * & & * & & & * \\ (1,00) & (9,86) & (2,83) & (17,23) & (0,69) & (1,45) & (2,25) & (0,69) \end{matrix}$

Equação 49

Os termos marcados com um asterisco têm a razão coeficiente/erro maior que três e foram considerados significativos na referência 5.

A Figura 20 e Figura 21 contêm os gráficos de probabilidade acumulada para cada bloco de replicatas. A variância do “main-plot” e “sub-plot” da Equação 30 não pode ser calculado para blocos sem replicatas, assim valores iguais a 1 foram atribuídos arbitrariamente para os cálculos preliminares. Mais tarde estes valores foram alterados para testar a robustez do método. A Figura 20 corresponde ao primeiro bloco de replicatas. Os pontos que indicam os três termos significativos estão bem separados do zero da abscissa. Os pontos a extrema direita da figura se referem aos termos x_1 e $x_1 x_2$ e um ponto a extrema esquerda da figura corresponde ao coeficiente de x_2 .

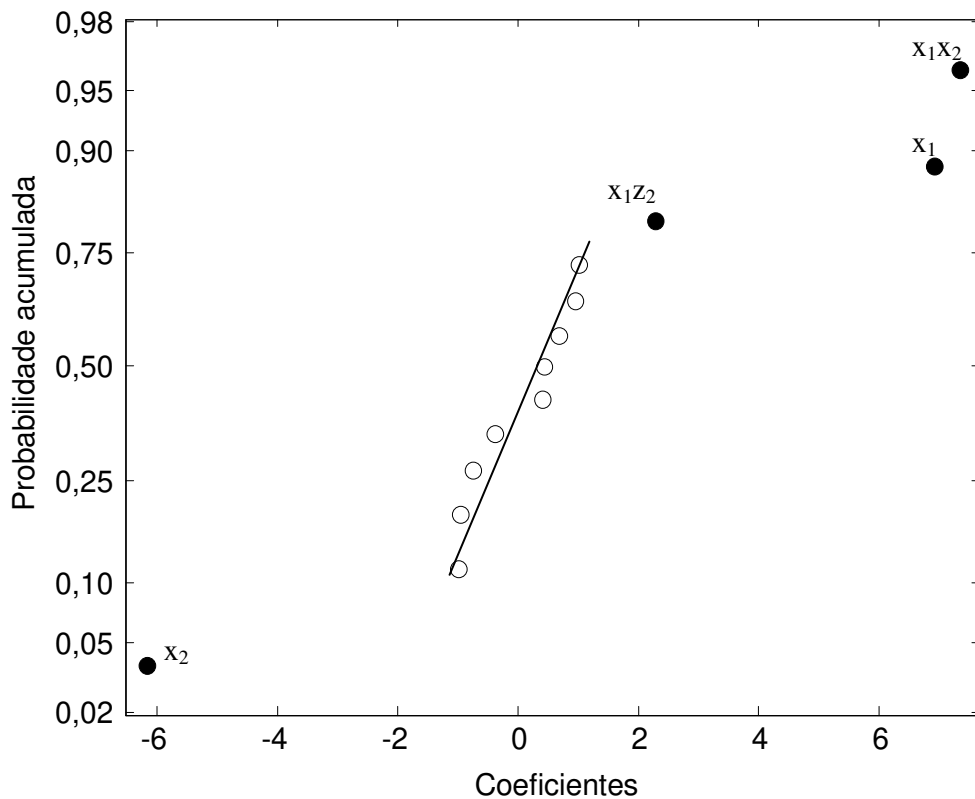


Figura 20. Gráfico de probabilidade acumulada dos coeficientes/erro padrão assumindo-se variância unitária para as primeiras replicatas de experimentos.

O ponto que indica o outro termo significativo, $2,96 x_1 z_2$, também aparece um pouco a direita do zero da abscissa. O gráfico de probabilidade acumulada para o segundo bloco de replicatas indicado na Figura 21 é muito parecido com o da Figura 20 exceto que o ponto correspondente ao termo $2,96 x_1 z_2$ aparece mais próximo do zero da abscissa. Em qualquer caso, os três pontos correspondentes aos mais importantes termos na Equação 49 sempre caem afastados da linha central sendo classificados como significativos. Certamente o químico pode estar interessado no termo $x_1 z_2$ que sugere que as propriedades da composição do plasticida 1, x_1 , muda dependendo do nível da temperatura de secagem, z_2 . O gráfico de probabilidade acumulada obtido para o modelo gerado com dados em replicatas, indicado na Figura 22 conduz a escolha dos mesmos quatro parâmetros significativos sugeridos pela referência 5, usando os resultados da

ANOVA. A ANOVA neste planejamento “split-plot” forneceu as seguintes estimativas de variâncias: $\sigma_R^2 = 0,537$ para as replicatas, $\sigma_{RZ}^2 = 0,28$ para o “main-plot” e $\sigma_e^2 = 1,092$ para o “sub-plot”. A robustez dos gráficos de probabilidade acumulada foram investigados variando cada uma das estimativas em $\pm 50\%$ segundo um planejamento fatorial 2^3 . Assim oito combinações de variâncias de replicatas, “main-plot” e “sub-plot” foram geradas e substituídas na Equação 30 para calcular de forma aproximada os erros dos parâmetros e encontrar os quocientes coeficiente/erro padrão para serem plotados no gráfico de probabilidade acumulada⁴⁰.

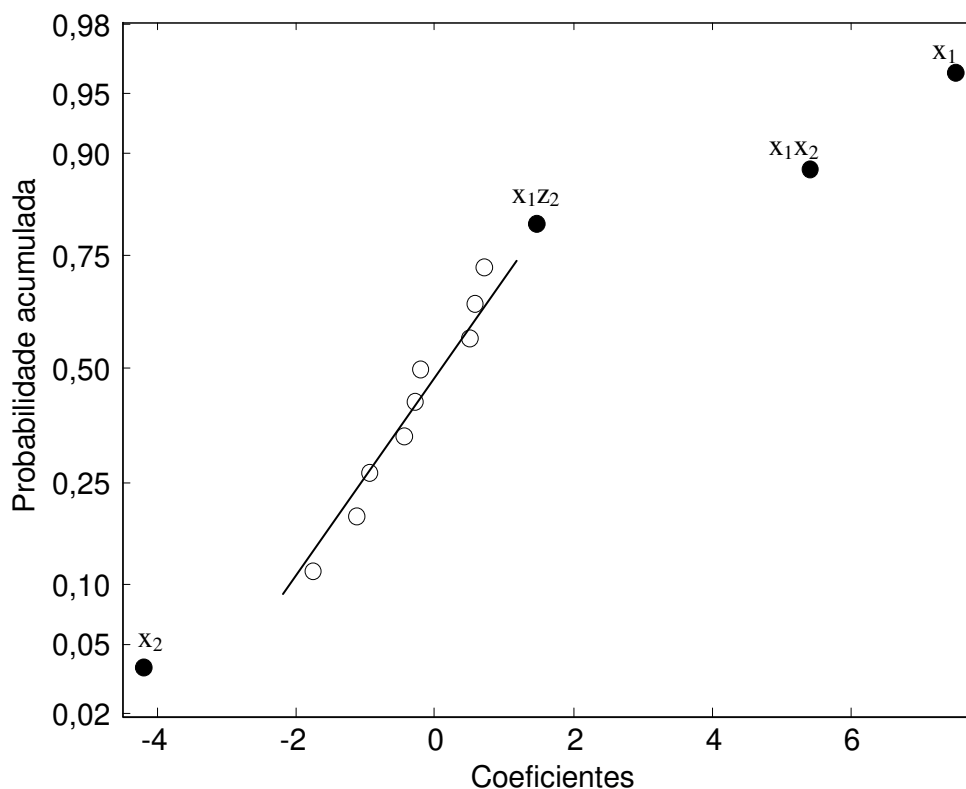


Figura 21. Gráfico de probabilidade acumulada para a razão coeficientes / erro padrão para o segundo grupo de replicatas da referência 5.

As razões coeficientes/erro padrão para o planejamento fatorial 2^3 estão indicadas na Tabela 23. Todos os parâmetros dos modelos se mostraram mais sensíveis a

alterações na variância do “sub-plot” do que nas outras variâncias. Os coeficientes significativos possuem a razão parâmetro/erro maior que os não significativos. Assim os valores das razões para os parâmetros não significativos permanecem concentrados próximos a zero no gráfico de probabilidade acumulada.

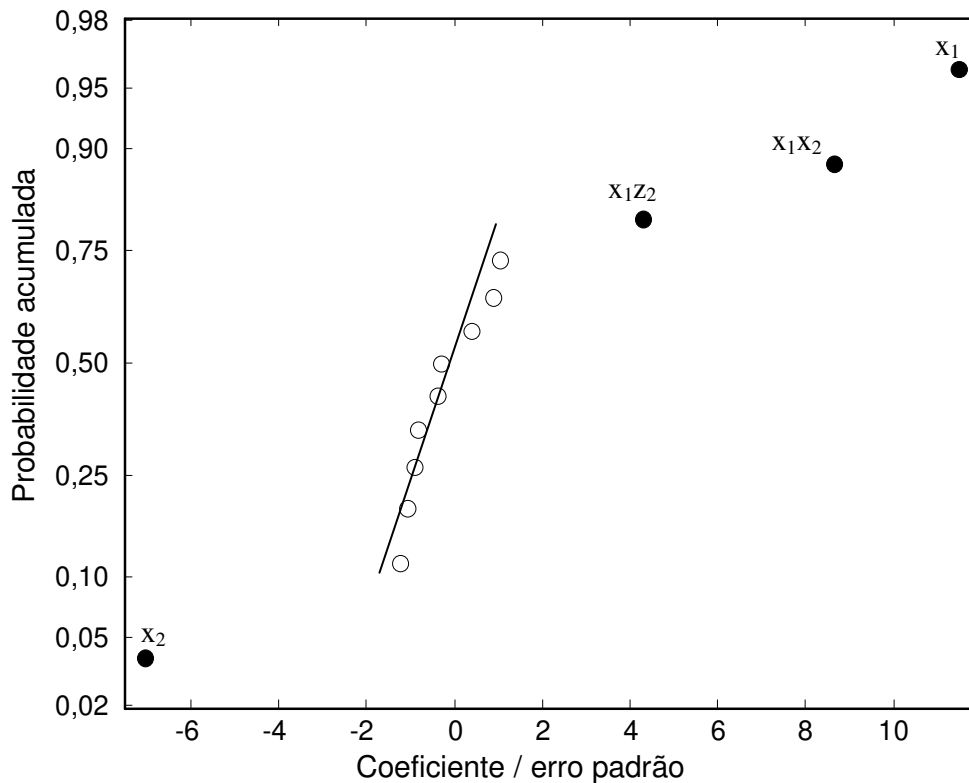


Figura 22. Gráfico de probabilidade acumulada para a razão coeficiente / erro padrão para dados com ambas as replicatas, utilizou-se as estimativas de variância fornecidas pela ANOVA.

Tabela 23 . Razões coeficientes / erro padrão para variações de 50% nas variâncias das replicatas, “main-plot” e “sub-plot” para um planejamento fatorial $2^{3,a}$

	- - -	+ - -	- + -	- + +	+ + -	- - +	+ - +	+ + +	
b_1	16,188	13,072	15,652	10,310	12,785	10,458	9,458	9,348	x1
b_2	-9,974	-9,946	-9,970	-5,764	-9,942	-5,764	-5,759	-5,758	x2
b_3	-1,313	-1,271	-1,307	-0,767	-1,266	-0,768	-0,759	-0,758	x3
b_{12}	12,195	12,195	12,195	7,041	12,195	7,041	7,041	7,041	x1x2
b_1^1	0,541	0,541	0,536	0,313	0,536	0,316	0,315	0,313	x1z1
b_2^1	1,114	1,114	1,054	0,643	1,054	0,6668	0,6668	0,643	x2z1
b_3^1	-0,443	-0,443	-0,443	-0,256	-0,443	-0,256	-0,256	-0,256	x3z1
b_1^2	4,692	4,692	3,758	2,709	3,758	3,196	3,196	2,709	x1z2
b_2^2	-0,496	-0,496	-0,457	-0,287	-0,457	-0,300	-0,301	-0,287	x2z2
b_3^2	-1,318	-1,318	-1,057	-0,761	-1,057	-0,897	-0,897	-0,761	x3z2
b_1^{12}	-1,202	-1,202	-1,190	-0,694	-1,190	-0,701	-0,701	-0,694	x1z1z2
b_2^{12}	-1,392	-1,392	-1,317	-0,804	-1,317	-0,847	-0,847	-0,804	x2z1z2
b_3^{12}	1,456	1,456	1,455	0,841	1,455	0,841	0,841	0,841	x3z1z2

- a) Variância das replicatas, (-) 0,269 ; (+) 0,806;
 Variância do “main-plot”, (-) 0,14; (+) 0,42;
 Variância do “sub-plot”, (-) 0,546 ; (+) 1,638.

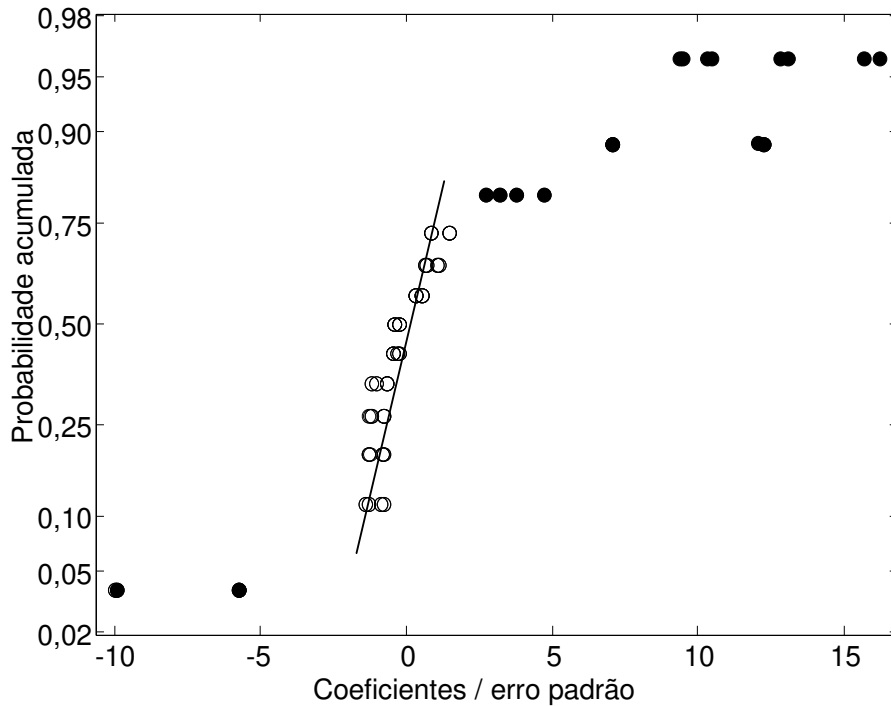


Figura 23. Superposição dos gráficos de probabilidade acumulada para um planejamento fatorial 2^3 variando os valores das estimativas das variâncias das replicatas, do “main-plot” e do “sub-plot”.

A Figura 23 mostra a sobreposição dos gráficos de probabilidade acumulada para todos os parâmetros calculados no fatorial 2^3 . As razões para todos os parâmetros significativos, exceto x_1z_2 , são claramente separadas da linha central.

Outro conjunto de dados utilizado se encontra na referência 7, onde medidas da textura de croquetes de peixe preparado pela mistura de três espécies de peixe em diferentes condições de temperatura, tempo de cozimento e tempo de fritura. As variáveis de mistura foram às proporções das três espécies de peixes utilizadas, e as variáveis de processo o tempo e temperatura de cozimento e o tempo de fritura. Os parâmetros e erro padrão do modelo com 56 termos,

$$\hat{y} = (\alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \alpha_3 z_3 + \alpha_{12} z_1 z_2 + \alpha_{13} z_1 z_3 + \alpha_{23} z_2 z_3 + \alpha_{123} z_1 z_2 z_3) \times (\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3),$$

foram determinados e treze coeficientes do modelo foram considerados significativos com um nível de confiança de 99% usando um teste t aproximado.

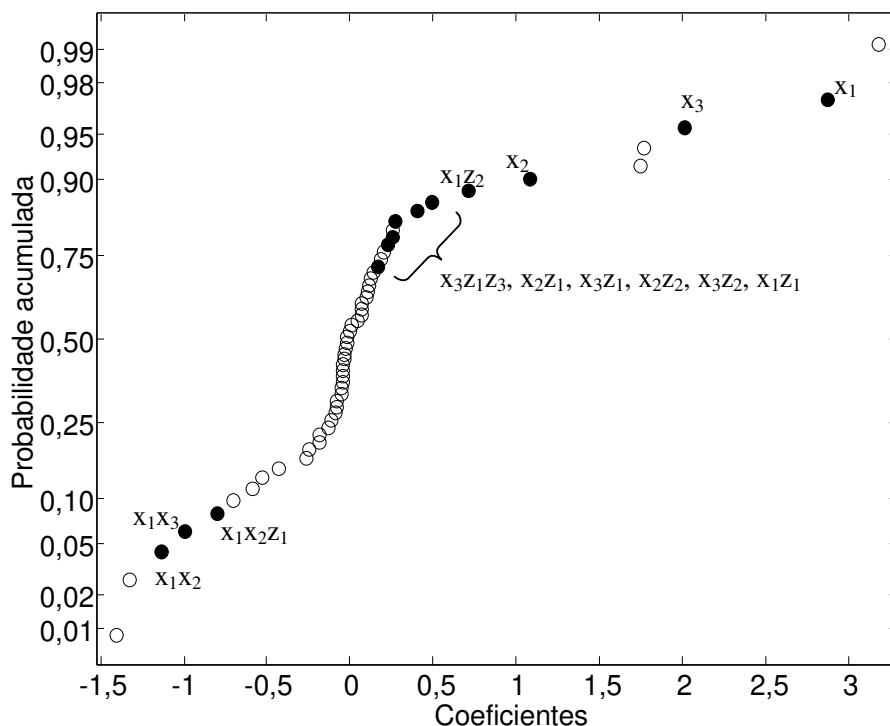


Figura 24. Gráfico de probabilidade acumulada dos coeficientes do modelo presente na referência 7.

A Figura 24 apresenta o gráfico de probabilidade acumulada de todos os 56 parâmetros com os termos significativos indicados por círculos sólidos. Este gráfico não é útil para separar os parâmetros significativos dos não significativos⁴⁰. Por outro lado, a Figura 25 contém um gráfico de probabilidade acumulada para a razão parâmetro/erro padrão dos 56 termos. Neste caso os treze parâmetros que se encontram distanciados dos outros centrados em zero correspondem aos significativos indicados na referência 7.

Em outro trabalho publicado por Reis et al.³⁶ o método “split-plot” foi aplicado a um planejamento para otimizar um procedimento catalítico para a determinação de Cr(VI). Três concentrações de reagentes foram variadas como

variáveis de processo de acordo com um fatorial 2^3 enquanto três solventes foram tratados como variáveis de misturas em dez composições. Cada bloco de replicatas foi constituído de 80 experimentos. Um modelo com dezesseis termos foi ajustado aos 160 dados. A referência 36 traz a tabela de ANOVA representada aqui pela Tabela 24 de forma mais detalhada, contendo as faltas de ajuste do modelo.

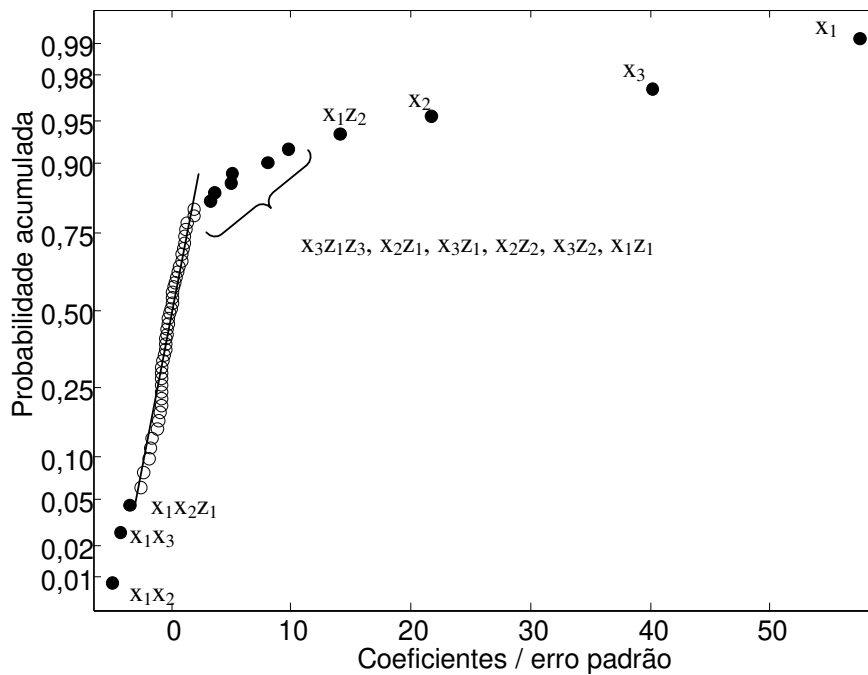


Figura 25. Gráfico de probabilidade acumulada da razão coeficientes/erro padrão do modelo fornecido pela referência 7.

Tabela 24. ANOVA “split-plot” incluindo as somas quadráticas de regressão e falta de ajuste dos dados do Cr (VI) da referência 36.

Fonte de variância	SQ	GL	MQ
Replicatas	0,0064	1	0,0064
Main-plot (Z)	1,7224	7	0,2461
Regressão (ZREG)	1,7218	6	0,2870
Falta de ajuste (ZLOF)	0,0006	1	0,0006
Erro main-plot (RZ)	0,0061	7	0,0009
Sub-plot (X)	6,9923	9	0,7769
Regressão (XREG)	6,8855	6	1,1476
Falta de ajuste (XLOF)	0,1068	3	0,0356
Interação Main-sub-plot (ZX)	1,3497	63	0,0214
Regressão (ZXREG)	1,2808	36	0,0356
Falta de ajuste (ZXLOF)	0,0689	27	0,0026
Erro sub-plot	0,0241	72	0,0003
Total	10,1010	159	

Embora não haja uma falta de ajuste significativa, ao nível de 95%, para a parte do modelo que trata do “main-plot”, para o “sub-plot” foi determinado

$$\frac{(SSXLOF + SSXZLOF) / (v_{XLOF} + v_{XZLOF})}{sub - plot \ error} = \frac{(0,1068 + 0,0689) / (3 + 27)}{0,024 / 72} = 17,6$$

contra um valor crítico tabelado de $F_{30,72} = 2$ ao nível de 99% de confiança.

Um gráfico de probabilidade acumulada contendo a razão coeficiente/erro padrão para um modelo dos 160 experimentos é indicado na Figura 26.

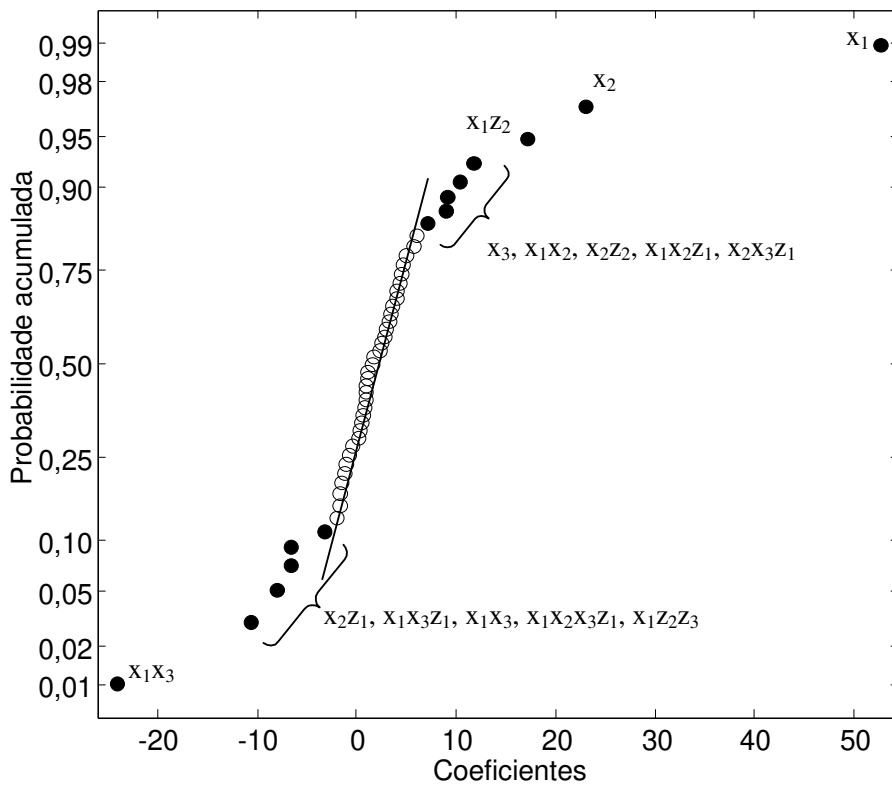


Figura 26. Gráfico de probabilidade acumulada para a razão coeficiente/erro padrão de todo o conjunto de dados do Cr (VI). Os coeficientes significativos são indicados por círculos preenchidos.

As estimativas de variância das replicatas, do “main-plot” e “sub-plot” foram obtidas da ANOVA. Observando a Figura 26 nota-se que 14 pontos estão separados dos outros cuja distribuição se aproxima de zero. Estes foram considerados estatisticamente significativos ao nível de 99% de confiança na referência 36. Todavia, os pontos que indicam parâmetros não significativos não estão exatamente centrados em zero. Assim o gráfico de probabilidade acumulada pode ser útil na detecção de falta de ajuste dos modelos.

A Figura 27 foi obtida plotando os valores das razões dos parâmetros/erros, sendo que neste caso os erros foram estimados atribuindo-se aos valores das variâncias, na Equação 30, o valor 1. Nesta figura há as informações mais

importantes que são encontradas na Figura 26. Os pontos considerados significativos na Figura 26 e Figura 27 são plenamente concordantes. Além do fato de que as distribuições dos pontos próximos a zero possuem um deslocamento na mesma direção. Assim as informações obtidas pela Figura 26 são também obtidas da Figura 27 proveniente de um conjunto de dados sem replicatas. Isto é confirmado por uma análise da Tabela 25.

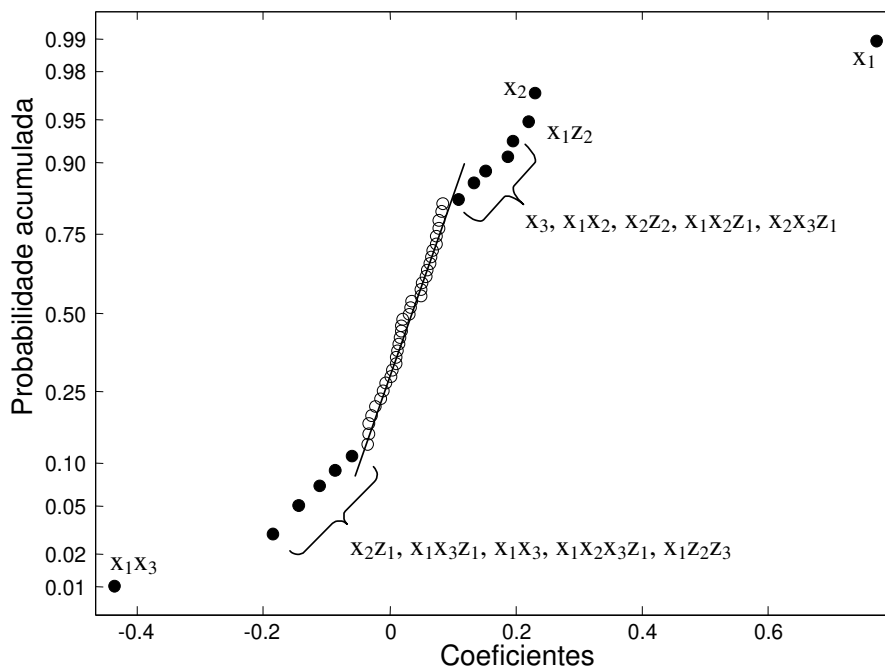


Figura 27. Gráfico de probabilidade acumulada para a razão coeficientes/erro padrão para uma replicata do conjunto de dados do Cr (VI). Às variâncias das replicatas, “main-plot” e “sub-plot” atribuiu-se o valor 1. O gráfico utilizando o segundo conjunto de replicatas é idêntico a este. Os coeficientes significativos são indicados por círculos preenchidos.

O gráfico de probabilidade acumulada para o segundo conjunto de replicatas fornece resultados essencialmente idênticos⁴⁰.

Tabela 25. Coeficientes do modelo para a determinação catalítica de Cr (VI) usando replicatas individuais e de forma conjunta. Parâmetros em negrito são significativos no nível de 95% de confiança.

Parâmetros	Todas as replicatas	1° bloco ^a	2° bloco
b_1	0,9469	0,9609	0,9329
b_2	0,1743	0,1757	0,1729
b_3	0,0536	0,0554	0,0519
b_{12}	0,3752	0,375	0,3753
b_{13}	-1,0103	-1,0194	-1,0012
b_{23}	0,0136	0,0003	0,0269
b_{123}	0,6144	0,9131	0,3158
b_1^1	0,0834	0,0599	0,1068
b_2^1	-0,0529	-0,0539	-0,0518
b_3^1	0,0273	0,0247	0,0299
b_{12}^1	0,4347	0,4628	0,4066
b_{13}^1	-0,3406	-0,3009	-0,3803
b_{23}^1	0,256	0,2593	0,2528
b_{123}^1	-0,9265	-0,9886	-0,8645
b_1^2	0,2896	0,281	0,2982
b_2^2	0,044	0,0398	0,0481
b_3^2	0,0187	0,0181	0,0193
b_{12}^2	0,1377	0,1704	0,1049
b_{13}^2	-0,2798	-0,2503	-0,3092
b_{23}^2	0,0029	0,0184	-0,0125
b_{123}^2	0,1112	0,2526	-0,0301
b_1^3	-0,0341	-0,0244	-0,0439
b_2^3	0,0023	0,0004	0,0042
b_3^3	0,0139	0,0147	0,013
b_{12}^3	0,1342	0,0989	0,1695
b_{13}^3	0,1924	0,1777	0,207
b_{23}^3	-0,0399	-0,0181	-0,0618
b_{123}^3	0,1217	0,218	0,0253
b_1^{12}	0,0395	0,0332	0,0459
b_2^{12}	0,0052	0,0052	0,0053
b_3^{12}	0,0008	-0,0023	0,004

b_{12}^{12}	0,249	0,2427	0,2553
b_{13}^{12}	-0,0373	-0,0269	-0,0477
b_{23}^{12}	-0,0118	-0,0475	0,0238
b_{123}^{12}	-0,153	0,0894	-0,3954
b_1^{13}	-0,028	-0,0192	-0,0367
b_2^{13}	0,0119	0,0093	0,0144
b_3^{13}	0,0081	0,0074	0,0088
b_{12}^{13}	0,1662	0,1559	0,1764
b_{13}^{13}	0,0298	0,0027	0,0569
b_{23}^{13}	-0,0279	-0,0278	-0,0279
b_{123}^{13}	0,5828	0,6644	0,5012
b_1^{23}	-0,0564	-0,0796	-0,0331
b_2^{23}	0,0075	0,0056	0,0095
b_3^{23}	0,0044	0,0031	0,0057
b_{12}^{23}	0,1452	0,1627	0,1277
b_{13}^{23}	0,1165	0,1429	0,09
b_{23}^{23}	-0,0399	-0,0266	-0,0532
b_{123}^{23}	0,1502	0,1896	0,1108

^a Corresponde aos parâmetros do gráfico de probabilidade acumulada da Figura 27

7 APLICAÇÃO E COMPARAÇÃO DO MÉTODO SPLIT-PLOT UTILIZANDO DADOS EXPERIMENTAIS

Em um trabalho de Kowalski⁴¹ o método “split-plot” foi usado em um planejamento fatorial 2^4 para compreender como quatro variáveis afetam a resistência de um plástico. As variáveis estudadas foram temperatura (a), porcentagem de aditivo (b), velocidade de agitação (c) e tempo de processamento (d). Para facilitar o procedimento experimental a temperatura foi escolhida como “main-plot” e as outras três variáveis formaram o “sub-plot”. A temperatura foi fixada em dois níveis e aleatoriamente se sorteou os níveis do “sub-plot”. A Figura 8 mostra a estratégia adotada. A Tabela 8 indica os resultados obtidos com os experimentos em duplicata.

Devido ao procedimento adotado é necessário que a análise de variância seja específica, ou seja, a análise “split-plot”. Uma análise convencional, considerando experimentos executados de forma completamente aleatória, pode levar a conclusões erradas sobre os efeitos significativos nesta otimização. Para ilustrar estas situações as duas análises de variância foram realizadas.

Inicialmente é necessário ajustar um modelo e neste caso o escolhido foi o bilinear,

$$\hat{y} = \alpha + b_1a + b_2b + b_3c + b_4d + b_{12}ab + b_{13}ac + b_{14}ad + b_{23}bc + b_{24}bd + b_{34}cd .$$

(Equação 31)

A primeira análise trata os dados como provenientes de experimentos completamente aleatórios e, portanto, com uma única fonte de erro. A Tabela 26 traz os coeficientes, o erro padrão e em negrito os parâmetros considerados significativos pelo teste t. Temperatura, tempo e a interação temperatura-velocidade e temperatura-tempo são significativos a 95% de confiança, e com 90% inclui-se as variáveis aditivo, velocidade e a interação velocidade-tempo. Com os coeficientes dos parâmetros foi construído um gráfico de probabilidade acumulada indicado na Figura 28. Os parâmetros significativos a 95% estão

preenchidos em preto, os preenchidos em cinza são significativos a 90% de confiança.

Tabela 26. Coeficientes e erro padrão dos parâmetros ajustados no modelo bilinear. Os parâmetros significativos a 95% estão em negrito.

Efeito	GL	Coeficientes	Erro padrão	razão Coef./erro
a	1	1,634	0,667	2,45
b	1	1,190	0,667	1,79
c	1	1,134	0,667	1,70
d	1	1,541	0,667	2,31
ab	1	0,184	0,667	0,28
ac	1	1,566	0,667	2,35
ad	1	1,397	0,667	2,10
bc	1	0,934	0,667	1,40
bd	1	0,303	0,667	0,46
cd	1	1,172	0,667	1,76

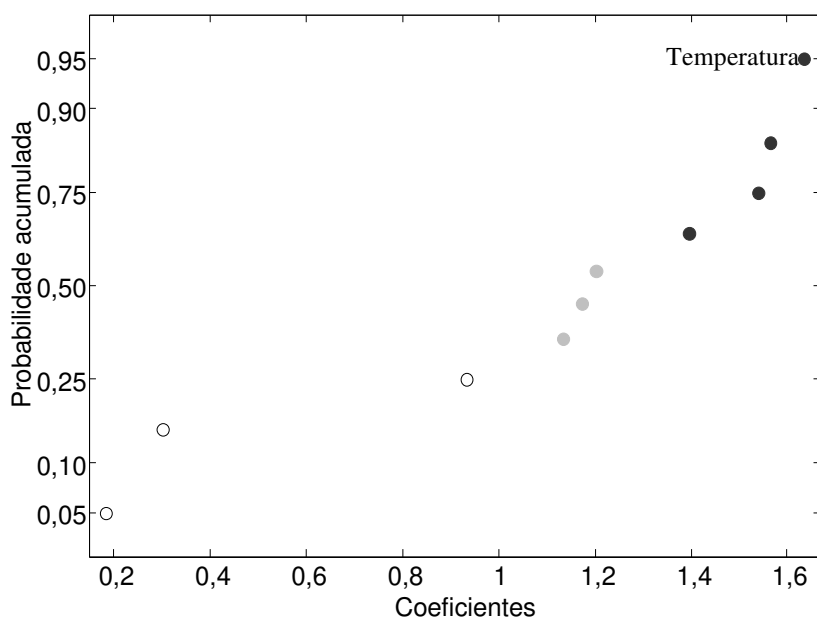


Figura 28. Gráfico de probabilidade acumulada para os coeficientes da Tabela 26. Como os erros são todos iguais não há diferença em utilizar os valores dos coeficientes ou a razão coeficiente / erro na abscissa.

Pela Figura 28 e Tabela 26 nota-se que a variável temperatura apresenta um efeito pronunciado se destacando entre os termos significativos.

Todavia não se deve esquecer que os experimentos não foram realizados de forma completamente aleatória, mas com um procedimento “split-plot”. Portanto, não há apenas uma fonte de erro afetando os resultados. Na verdade, duas fontes de erro estão presentes: o erro “main-plot” e o erro “sub-plot”. Assim a variável temperatura, que constitui o “main-plot”, é afetada por uma fonte de erro diferente daquela que afeta as variáveis do “sub-plot”. Então considerando o tratamento adequado para o planejamento foi refeita a análise estatística.

Inicialmente foi calculada a ANOVA para os resultados dos experimentos, apresentada na Tabela 27.

Tabela 27. ANOVA “split-plot” para os dados da Tabela 8.

Fonte	SQ	GL	MQ
Replicatas	84,83	1	84,83
Main-plot	85,48	1	85,48
Erro main-plot	27,56	1	27,56
Sub-plot	244,63	7	34,95
Interação main-sub-plot	145,70	7	20,81
Erro sub-plot	174,81	14	12,70

A ANOVA da Tabela 27 permite separar as fontes de erro do “main-plot”, “sub-plot” e a soma quadrática devido às replicatas. Kowalski sugere que as somas quadráticas devido a replicata e ao erro “main-plot” podem ser somadas. Desta forma o valor da média quadrática devida ao erro “main-plot” seria de 56,20 com 2 graus de liberdade. Pode-se então fazer um teste F entre o efeito do “main-plot” e o erro presente em suas medidas pela razão $85,48/56,20 = 1,52$ que não é significativo comparado ao valor de $F_{1; 2; 95\%} = 18,51$ ou $F_{1; 2; 90\%} = 8,53$. O que nos permite concluir que o efeito “main-plot” não é significativo. Por outro lado, autores como Cornell⁵ adicionam ao erro “sub-plot” as possíveis faltas de ajuste do modelo escolhido como adequado obtendo mais graus de liberdade para tratar o erro, além de adicionar ao erro “sub-plot” o erro proveniente da interação “main-sub-

plot” com o mesmo objetivo. Para exemplificar a maneira como as somas quadráticas são tratadas um modelo com 31 termos foi gerado, o que esgota toda a falta de ajuste e toda variância é explicada pelos parâmetros ajustados. A Tabela 28 traz os 31 parâmetros do modelo e a soma quadrática explicada por cada um na regressão. O valor da soma quadrática devido à replicata (r) é de 84,83 e a interação da replicata com o “main-plot”, erro “main-plot” (ra), é de 27,56 plenamente concordantes com os valores da Tabela 27. O valor do efeito “main-plot” (a) é igual a 85,48 e o efeito do “sub-plot” pode ser obtido pela somatória das somas quadráticas dos termos b, c, bc, d, bd, cd e bcd e possuindo 7 graus de liberdade e equivalendo a 244,64. Para determinar o erro “sub-plot” basta somar os termos de interação entre (b, c, d) e replicata (r), $rb + rc + rbc + rd + rbd + rcd + rbcd = 117,30$, já o erro “main-sub-plot” pode ser determinado pelas interações de a com b, c, d e r, cuja soma vale 57,51. O valor do erro “sub-plot” e “main-sub-plot” são somados no método “split-plot” o que equivale a 174,81 com 14 graus de liberdade.

Para comparar a ANOVA “split-plot” com a que considera todos os experimentos completamente aleatórios ajustou-se o mesmo modelo da Equação 31 que considera apenas interações binárias entre os efeitos. Assim os termos de interação superior não participam do modelo sendo equivalentes a falta de ajuste e, portanto, adicionados ao erro “sub-plot”: erro “sub-plot” + soma quadrática de termos não ajustados = $174,81 + 11,06 = 185,87$, contendo 19 graus de liberdade. A média quadrática do erro será, portanto, igual a 9,78. Com as somas quadráticas dos efeitos da Tabela 28 pode-se realizar um teste F para determinar os parâmetros significativos a 95%, $F_{1,19,95\%} = 4,38$, e 90%, $F_{1,19,90\%} = 2,99$. A Tabela 29 indica as razões entre as médias quadráticas obtidas por regressão dos parâmetros e a média quadrática do erro. Comparando os valores da Tabela 26 e Tabela 29 percebe-se algumas alterações. Primeiro, a temperatura que inicialmente era um fator muito significativo, na segunda análise não é. Segundo, o aditivo que foi considerado significativo na primeira análise com apenas 90% de confiança passou a ser significativo com 95%. Terceiro, a interação velocidade-tempo teve um efeito significativo apenas com 90% de confiança na análise

considerando os experimentos completamente aleatórios, mas na análise “split-plot” o efeito velocidade-tempo é significativo já em 95% de confiança. A Figura 29 mostra o gráfico de probabilidade acumulada para as razões os coeficientes / erro padrão para o modelo da Equação 31 pelo método “split-plot”. Os valores dos coeficientes e seus erros são indicados na Tabela 30.

Tabela 28. Modelo com 31 parâmetros ajustado aos dados da Tabela 8.

Fonte	GL	SS	Fonte	GL	SS
r	1	84,83	rd	1	0,81
a	1	85,48	ad	1	62,44
ra	1	27,56	rad	1	4,28
b	1	45,36	bd	1	2,94
rb	1	0,002	rbd	1	43,95
ab	1	1,09	abd	1	0,75
rab	1	5,04	rabd	1	0,26
c	1	41,18	cd	1	43,95
rc	1	46,32	rcd	1	14,99
ac	1	78,44	acd	1	2,82
rac	1	0,63	racd	1	15,82
bc	1	27,94	bcd	1	7,32
rbc	1	0,30	rbcd	1	10,93
abc	1	0,17	abcd	1	0,003
rabc	1	0,07	rabcd	1	31,40
d	1	75,95			

Tabela 29. Somas quadráticas referentes ao modelo da Equação 31 para o método “split-plot”.

Efeito	SQ	GL	MQ	MQ _{erro}	Razão MQ/MQ _{erro}
a	85,48	1	85,48	56,20	1,52
b	45,36	1	45,36	9,78	4,63 *
c	41,18	1	41,18	9,78	4,21 [∇]
d	75,95	1	75,95	9,78	7,77 *
ab	1,08	1	1,08	9,78	0,11
ac	78,44	1	78,44	9,78	8,02 *
ad	62,44	1	62,44	9,78	6,38 *
bc	27,94	1	27,94	9,78	2,86
bd	2,94	1	2,94	9,78	0,30
cd	43,95	1	43,95	9,78	4,49 *

* parâmetro significativo a 95%

[∇] parâmetro significativo a 90%

Tabela 30. Coeficientes e erros para o modelo “split-plot”.

Efeito	coeficiente	erro	razão coef./erro
a	1,6344	0,9281	1,76
b	1,1906	0,5529	2,15
c	1,1344	0,5529	2,05
d	1,5406	0,5529	2,79
ab	0,1844	0,5529	0,33
ac	1,5656	0,5529	2,83
ad	1,3969	0,5529	2,53
bc	0,9344	0,5529	1,69
bd	0,3031	0,5529	0,55
cd	1,1719	0,5529	2,12

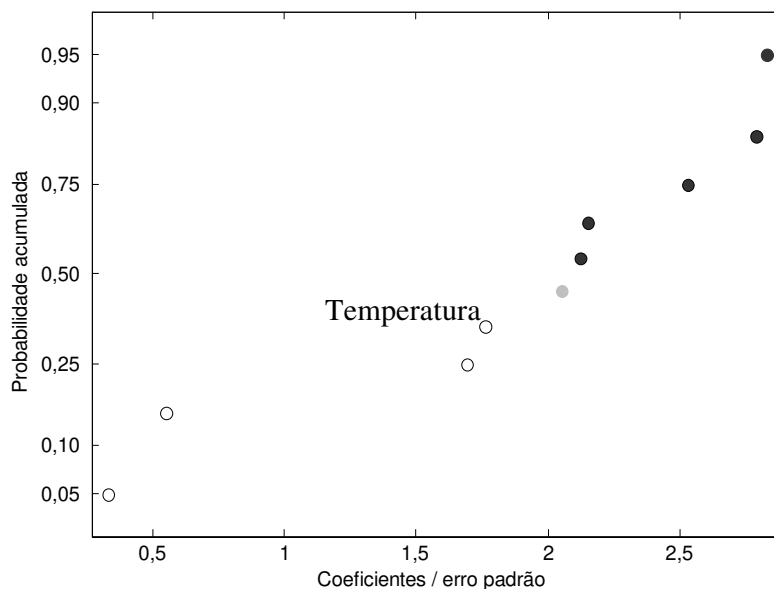


Figura 29. Gráfico de probabilidade acumulada para os valores das razões coeficientes / erro padrão da Tabela 30, os círculos preenchidos pretos são significativos a 95% e em cinza a 90%.

A forma de representar em gráfico a razão coeficiente / erro padrão em gráficos de probabilidade acumulada como ferramenta para determinação de parâmetros significativos de modelos já vem sendo aplicado em nossos estudos com grande eficiência. Comparando a Figura 28 com a Figura 29 nota-se claramente que a variável temperatura, na análise “split-plot”, após ser corrigida por seu erro deixa de ser significativa como antes. Isto deixa evidente a

necessidade de realizar uma análise de variância adequada ao planejamento adotado. Vale notar que a aplicação do método “split-plot” a planejamentos fatoriais pode significar um importante passo para a popularização do método entre pesquisadores das universidades e indústrias.

7.1 Análise de dados experimentais com tratamento split-plot e completamente aleatório.

O objetivo deste estudo é determinar se um conjunto de experimentos realizados segundo um procedimento “split-plot” pode receber um tratamento convencional considerando que todos os experimentos são aleatórios sem levar (ou levar) a conclusões incorretas.

Anteriormente foi discutido o conjunto de dados de um procedimento “split-plot” em fatorial onde a temperatura consistia o “main-plot”. Mas será que em qualquer situação o fato de realizar os experimentos em agrupados em bloco possui conseqüências tão grandes?

O segundo conjunto de dados tratados está presente na referência 42. No planejamento indicado há duas variáveis de processo, velocidade de extrusão (z_1) e temperatura de secagem (z_2) cujos níveis são fixados, formando o “main-plot”, e três variáveis de mistura, x_1 , x_2 e x_3 , que correspondem a diferentes proporções de plastificantes aleatorizadas nos “sub-plots”. O planejamento é descrito na Tabela 31.

Inicialmente os dados foram tratados corretamente considerando o procedimento experimental “split-plot”. A Tabela 32 traz a ANOVA realizada que permite observar que o efeito do “main-plot”, do “sub-plot” e da interação “main-sub-plot” são significativos, já que todos superam o valor de F tabelado a 95% de confiança.

O próximo passo é avaliar o ajuste de um modelo, no caso foi escolhido o quadrático para o “sub-plot” e bilinear para o “main-plot”. Este modelo apresenta uma característica interessante, ele é composto por 24 parâmetros o que esgota os graus de liberdade para testar as faltas de ajuste. A Tabela 33 indica os

coeficientes e erros obtidos para o modelo quadrático-bilinear ajustado. A Figura 30 mostra o gráfico de resíduos para o modelo quadrático-bilinear. Pode-se notar uma distribuição simétrica e aleatória dos resíduos.

Tabela 31. Planejamento com duas variáveis de processo, velocidade de extrusão (z_1) e temperatura de secagem (z_2) cujos níveis são fixados, formando o “main-plot”, e três variáveis de mistura, x_1 , x_2 e x_3 , que correspondem a diferentes proporções de plastificantes aleatorizadas nos “sub-plots”.

z_1	z_2	x_1	x_2	x_3	Resposta	z_1	z_2	x_1	x_2	x_3	Resposta
-1	-1	1	0	0	7	1	-1	1	0	0	10
-1	-1	1	0	0	8	1	-1	1	0	0	12
-1	-1	0	1	0	4	1	-1	0	1	0	5
-1	-1	0	1	0	4	1	-1	0	1	0	8
-1	-1	0	0	1	5	1	-1	0	0	1	9
	-1	0	0	1	7	1	-1	0	0	1	8
-1	-1	0,5	0,5	0	7	1	-1	0,5	0,5	0	14
-1	-1	0,5	0,5	0	8	1	-1	0,5	0,5	0	15
-1	-1	0,5	0	0,5	8	1	-1	0,5	0	0,5	12
-1	-1	0,5	0	0,5	10	1	-1	0,5	0	0,5	11
-1	-1	0	0,5	0,5	4	1	-1	0	0,5	0,5	8
-1	-1	0	0,5	0,5	3	1	-1	0	0,5	0,5	7
-1	1	1	0	0	10	1	1	1	0	0	6
-1	1	1	0	0	13	1	1	1	0	0	5
-1	1	0	1	0	8	1	1	0	1	0	7
-1	1	0	1	0	8	1	1	0	1	0	4
-1	1	0	0	1	3	1	1	0	0	1	6
-1	1	0	0	1	7	1	1	0	0	1	7
-1	1	0,5	0,5	0	12	1	1	0,5	0,5	0	5
-1	1	0,5	0,5	0	16	1	1	0,5	0,5	0	5
-1	1	0,5	0	0,5	9	1	1	0,5	0	0,5	4
-1	1	0,5	0	0,5	13	1	1	0,5	0	0,5	6
-1	1	0	0,5	0,5	7	1	1	0	0,5	0,5	7
-1	1	0	0,5	0,5	10	1	1	0	0,5	0,5	8

Tabela 32. Análise de variância “split-plot” para os dados da Tabela 31.

Fonte	GL	SQ	MQ	F	F _{Tab. (95%)}
Replicatas (R)	1	14,0833	14,083		
Main-plot (Z)	3	170,167	56,722	10,80	5,39
Regressão (ZREG)	3	170,167	56,722		
Erro main-plot (RZ)	3	15,750	5,250		
Sub-plot (X)	5	118,917	23,783	18,90	2,16
Regressão (XREG)	5	118,917	23,783		
Interação main-sub-plot	15	125,583	8,372	6,65	1,84
Regressão (ZXREG)	15	125,583	8,372		
Erro sub-plot	20	25,167	1,258		
Total	47		469,667		

Tabela 33. Modelo ajustado aos dados da Tabela 31 e erros associados aos parâmetros.

Efeito	Estimativa	Erro padrão	Razão estimativa/erro padrão
x ₁	8,875	0,652	13,622
x ₂	6,000	0,652	9,210
x ₃	6,500	0,652	9,977
x ₁ x ₂	11,250	1,943	5,790
x ₁ x ₃	5,750	1,943	2,959
x ₂ x ₃	2,000	1,943	1,029
x ₁ z ₁	-0,625	0,490	-1,274
x ₂ z ₁	0,000	0,490	0,000
x ₃ z ₁	1,000	0,490	2,039
x ₁ x ₂ z ₁	-0,750	1,943	-0,386
x ₁ x ₃ z ₁	-4,250	1,943	-2,187
x ₂ x ₃ z ₁	1,000	1,943	0,515
x ₁ z ₂	-0,375	0,490	-0,765
x ₂ z ₂	0,750	0,490	1,529
x ₃ z ₂	-0,750	0,490	-1,529
x ₁ x ₂ z ₂	-3,750	1,943	-1,930
x ₁ x ₃ z ₂	-2,250	1,943	-1,158
x ₂ x ₃ z ₂	5,000	1,943	2,573
x ₁ z ₁ z ₂	-2,375	0,490	-4,843
x ₂ z ₁ z ₂	-1,250	0,490	-2,549
x ₃ z ₁ z ₂	-0,250	0,490	-0,510
x ₁ x ₂ z ₁ z ₂	-8,750	1,943	-4,504
x ₁ x ₃ z ₁ z ₂	-3,250	1,943	-1,673
x ₂ x ₃ z ₁ z ₂	-2,000	1,943	-1,029

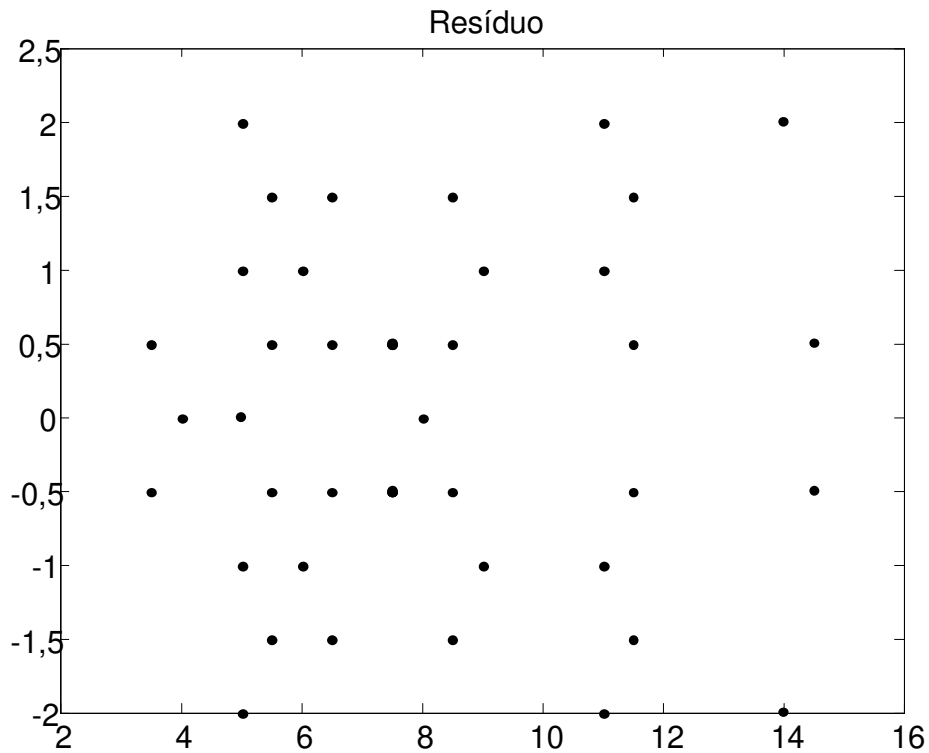


Figura 30. Gráfico de resíduos para o modelo descrito na Tabela 33.

A Figura 31 mostra um histograma construído com os parâmetros ajustados no modelo quadrático-bilinear, aparentemente há uma distribuição normal dos parâmetros não significativos. A Figura 32 indica as respostas previstas \times respostas observadas de forma relativamente simétrica. Assim pela análise gráfica realizada não há motivos para rejeitar o modelo. Contudo, não há uma maneira exata de determinar o valor de corte para o teste t. Assim aplicou-se como ferramenta de decisão o gráfico de probabilidade acumulada indicado na Figura 33.

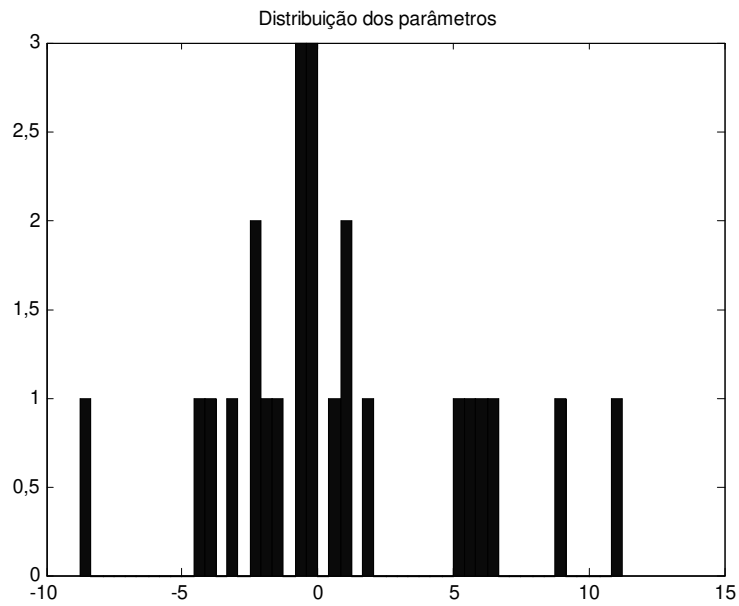


Figura 31. Histograma para a distribuição dos coeficientes do modelo descrito na Tabela 33.

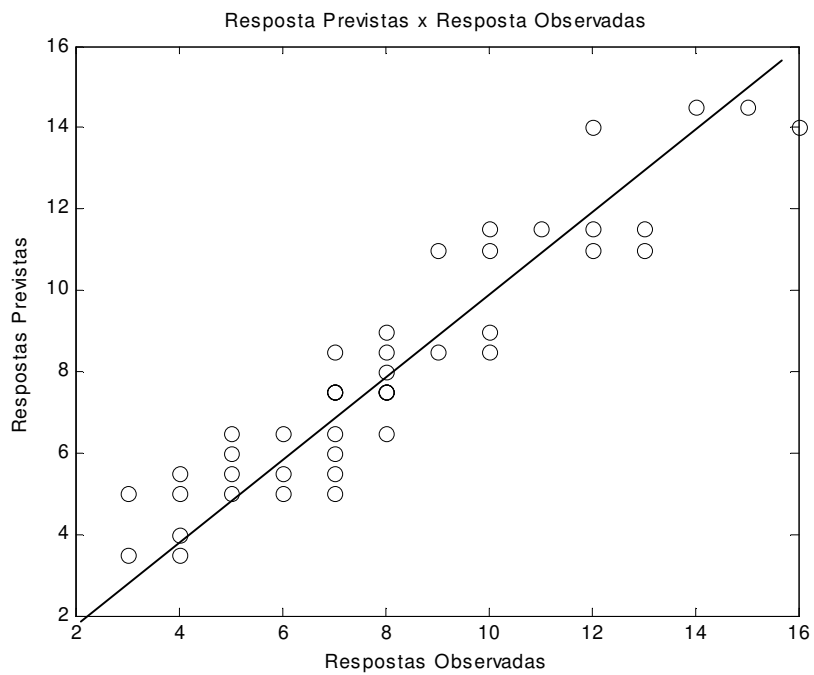


Figura 32. Respostas previstas x resposta observada para o modelo descrito na Tabela 33.

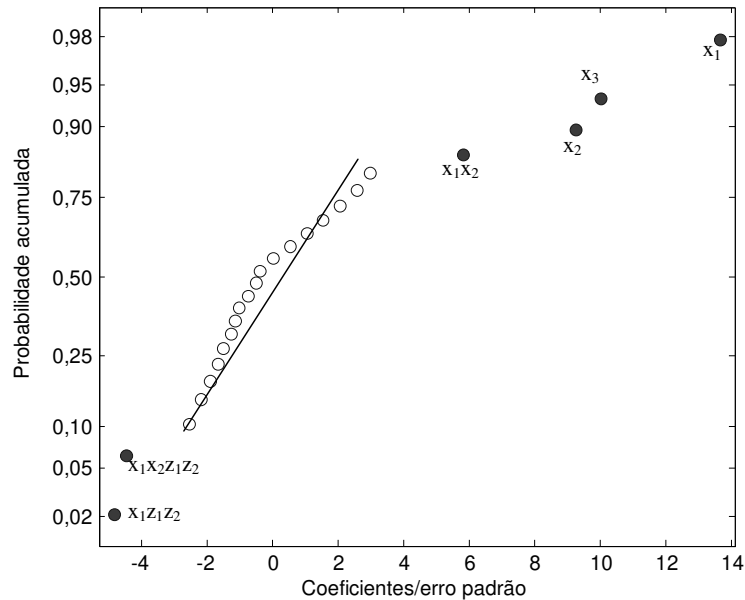


Figura 33. Gráfico de probabilidade para valores da razão coeficientes/erro padrão da Tabela 33.

Analisando a Figura 33 nota-se a direita quatro pontos, x_1 , x_2 , x_3 , x_1x_2 , afastados do conjunto de pontos pertencentes a reta centrada em zero. À esquerda há também dois pontos, $x_1x_2z_1z_2$ e $x_1z_1z_2$, afastados do zero. Assim, realizando-se a análise estatística adequada ao procedimento experimental adotado, “split-plot”, os termos preenchidos na figura anterior representam os efeitos significativos.

O outro procedimento adotado trata os dados como provenientes de experimentos completamente aleatórios. O mesmo modelo, quadrático-bilinear, foi ajustado e a Tabela 34 indica as estimativas dos efeitos (coeficientes), os erros dos coeficientes e a razão dos coeficientes pelos erros.

Tabela 34. Modelo ajustado aos dados da Tabela 31 considerando os experimentos realizados de forma completamente aleatória.

Efeito	Estimativa	Erro padrão	Razão Coef./erro padrão
x_1	8,875	0,535	16,580
x_2	6,000	0,535	11,210
x_3	6,500	0,535	12,140
x_1x_2	11,250	2,622	4,290
x_1x_3	5,750	2,622	2,190
x_2x_3	2,000	2,622	0,760
x_1z_1	-0,625	0,535	-1,170
x_2z_1	0,000	0,535	0,000
x_3z_1	1,000	0,535	1,870
$x_1x_2z_1$	-0,750	2,622	-0,290
$x_1x_3z_1$	-4,250	2,622	-1,620
$x_2x_3z_1$	1,000	2,622	0,380
x_1z_2	-0,375	0,535	-0,700
x_2z_2	0,750	0,535	1,400
x_3z_2	-0,750	0,535	-1,400
$x_1x_2z_2$	-3,750	2,622	-1,430
$x_1x_3z_2$	-2,250	2,622	-0,860
$x_2x_3z_2$	5,000	2,622	1,910
$x_1z_1z_2$	-2,375	0,535	-4,440
$x_2z_1z_2$	-1,250	0,535	-2,340
$x_3z_1z_2$	-0,250	0,535	-0,470
$x_1x_2z_1z_2$	-8,750	2,622	-3,340
$x_1x_3z_1z_2$	-3,250	2,622	-1,240
$x_2x_3z_1z_2$	-2,000	2,622	-0,760

Para analisar os parâmetros significativos foi construído o gráfico de probabilidade acumulada indicado na Figura 34 . Os pontos preenchidos neste gráfico correspondem a x_1 , x_2 , x_3 , x_1x_2 , $x_1x_2z_1z_2$ e $x_1z_1z_2$ e são os mais distantes do zero da abcissa, assim foram considerados significativos.

Vale notar que considerando as duas formas de tratar os dados, “split-plot” e completamente aleatório, obteve-se os mesmos parâmetros significativos. Ao que parece o efeito de “blocar” os experimentos, neste caso, não trouxe alterações significativas na escolha dos parâmetros.

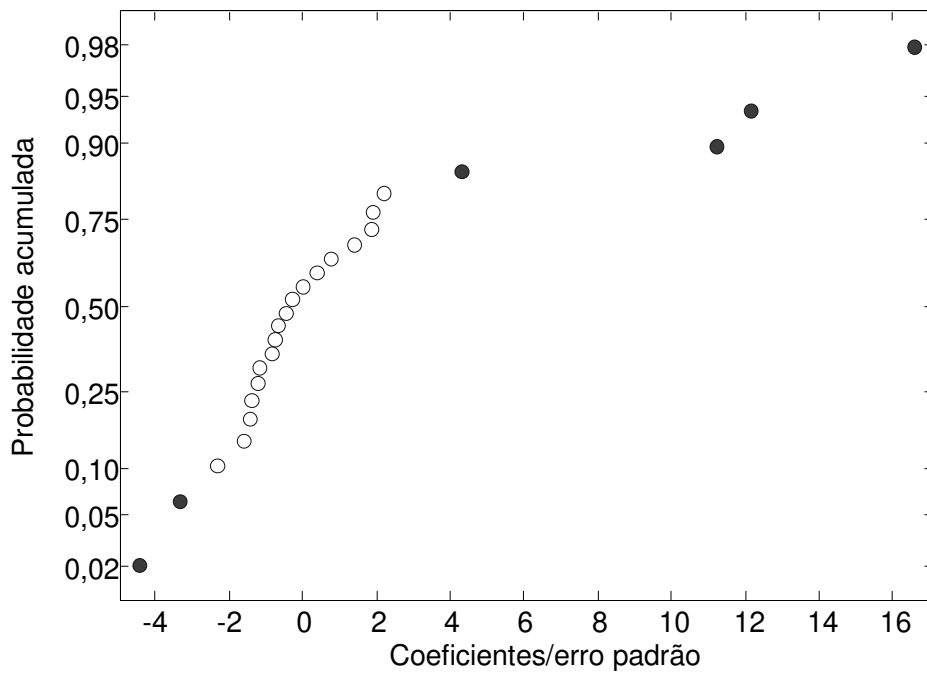


Figura 34. Gráfico de probabilidade acumulada para os valores da razão coeficientes/erro padrão presentes na Tabela 34.

O terceiro e o quarto conjunto de dados, para a análise considerando tratamento “split-plot” e completamente aleatório, foram retirados da referência 5. A Tabela 35 traz o planejamento adotado. Neste caso, z_1 e z_2 em dois níveis formam o “main-plot”. O “sub-plot”, realizado de forma aleatória, é constituído por x_1 , x_2 e x_3 . Os dados foram inicialmente tratados como “split-plot”, de acordo com o planejamento experimental adotado. A Tabela 36 traz a ANOVA realizada. Um teste F indica que os efeitos do “main-plot”, “sub-plot” e da interação “main-sub-plot” são significativos a 95% de confiança.

Tabela 35. Planejamento com um conjunto de dados extraído da referência 5. As variáveis de mistura são x_1 , x_2 e x_3 e as de processo z_1 e z_2 . Veja referência 5.

z_1	z_2	x_1	x_2	x_3	Respostas
-1	1	1	0	0	5
-1	1	1	0	0	6
-1	1	0	1	0	6
-1	1	0	1	0	8
-1	1	0	0	1	9
-1	1	0	0	1	10
1	1	1	0	0	6
1	1	1	0	0	7
1	1	0	1	0	9
1	1	0	1	0	10
1	1	0	0	1	9
1	1	0	0	1	11
-1	-1	1	0	0	4
-1	-1	1	0	0	4
-1	-1	0	1	0	7
-1	-1	0	1	0	7
-1	-1	0	0	1	6
-1	-1	0	0	1	7
1	-1	1	0	0	3
1	-1	1	0	0	4
1	-1	0	1	0	6
1	-1	0	1	0	5
1	-1	0	0	1	9
1	-1	0	0	1	8

Tabela 36. ANOVA “split-plot” para os dados da Tabela 35

Fonte	GL	SQ	MQ	F	$F_{Tab(95\%)}$
Replicatas (R)	1	2,6667	2,6667		
Main-plot (Z)	3	33,5001	11,1667	11,17	9,28
Erro main-plot (RZ)	3	3,0000	1,0000		
Sub-plot (X)	2	57,5834	28,7917	98,70	4,46
Int. main-sub-plot (ZX)	6	8,7498	1,4583	5,00	3,58
Erro Sub-plot	8	2,3336	0,2917		

O modelo linear para mistura e bilinear para processo foi ajustado aos dados da Tabela 35. Os coeficientes do modelo e os erros associados aos parâmetros estão indicados na Tabela 37. Os valores da razão coeficientes / erro padrão foram

utilizados para a elaboração do gráfico de probabilidade acumulada indicado na Figura 35.

Tabela 37. Modelo linear-bilinear ajustado aos dados da Tabela 35. Os parâmetros significativos estão em negrito

Efeitos	Coefficientes	Erro padrão	Coef./erro padrão
x_1	4,875	0,368	13,247
x_2	7,250	0,368	19,701
x_3	8,625	0,368	23,438
x_1z_1	0,125	0,2569	0,487
x_2z_1	0,250	0,2569	0,973
x_3z_1	0,625	0,2569	2,433
x_1z_2	1,125	0,2569	4,379
x_2z_2	1,000	0,2569	3,893
x_3z_2	1,125	0,2569	4,379
$x_1z_1z_2$	0,375	0,2569	1,460
$x_2z_1z_2$	1,000	0,2569	3,893
$x_3z_1z_2$	-0,375	0,2569	-1,460

Os parâmetros x_1 , x_2 , x_3 , x_1z_2 , x_2z_2 , x_3z_2 e $x_2z_1z_2$ foram considerados significativos pelo teste t e estão preenchidos no gráfico da Figura 35.

O mesmo conjunto de dados foi tratado como se todos os experimentos tivessem sido realizados de forma completamente aleatória. O modelo linear-bilinear foi ajustado aos dados da Tabela 35, os resultados são apresentados na Tabela 38. O teste t indica que os parâmetros significativos são x_1 , x_2 , x_3 , x_1z_2 , x_2z_2 , x_3z_2 e $x_2z_1z_2$, ou seja, os mesmos da análise “split-plot”. Portanto, a submissão dos dados à blocagem não apresentou efeitos significativos na seleção dos parâmetros em relação à análise apropriada “split-plot”. O gráfico de probabilidade acumulada para este tratamento é similar ao da Figura 35.

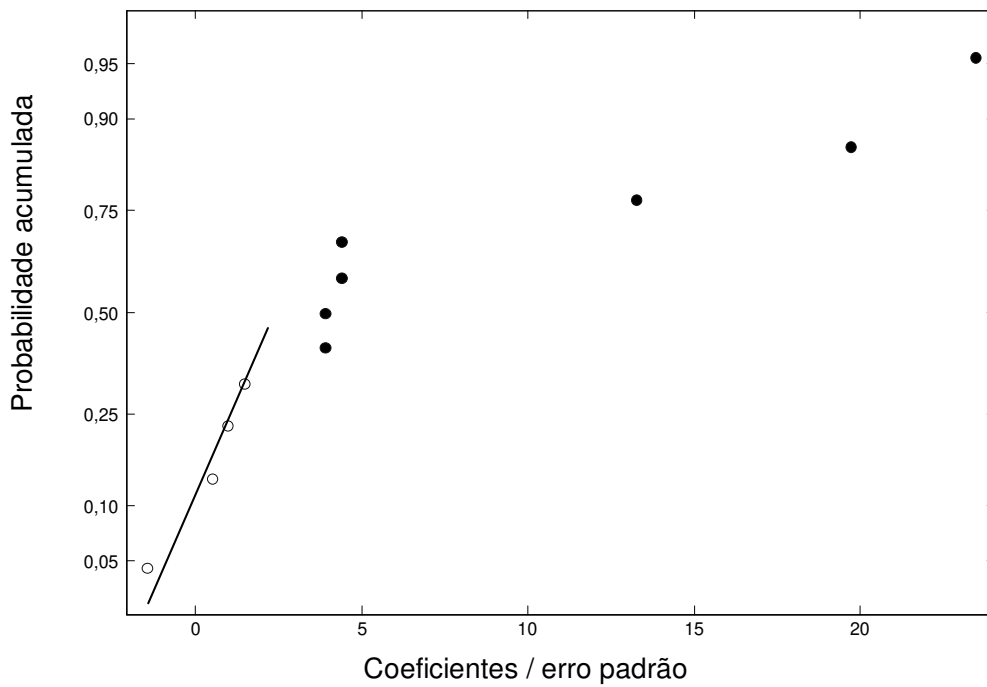


Figura 35. Gráfico de probabilidade acumulada com os valores obtidos da razão coeficientes / erro padrão contidos na Tabela 37.

Tabela 38. Modelo linear-bilinear ajustado aos dados da Tabela 35 considerando os dados provenientes de experimentos completamente aleatórios.

Efeito	Coefficiente	Erro padrão	Coef./erro padrão
x_1	4,875	0,2887	16,89
x_2	7,250	0,2887	25,11
x_3	8,625	0,2887	29,88
x_1z_1	0,125	0,2887	0,43
x_2z_1	0,250	0,2887	0,87
x_3z_1	0,625	0,2887	2,17
x_1z_2	1,125	0,2887	3,90
x_2z_2	1,000	0,2887	3,46
x_3z_2	1,125	0,2887	3,90
$x_1z_1z_2$	0,375	0,2887	1,30
$x_2z_1z_2$	1,000	0,2887	3,46
$x_3z_1z_2$	-0,375	0,2887	-1,30

O quarto conjunto de dados analisado é proveniente de um planejamento “split-plot” que foi utilizado na otimização da produção de vinil para aplicações automobilísticas. As variáveis de mistura x_1 , x_2 e x_3 representam proporções de agentes plastificantes utilizados. As variáveis de processo z_1 e z_2 são a taxa de extrusão e temperatura de secagem, respectivamente, testados em dois níveis. O planejamento é representado na Tabela 39.

Tabela 39. Planejamento “split-plot” utilizado na otimização da produção de vinil. As variáveis de mistura são x_1 , x_2 e x_3 e as de processo z_1 e z_2 . Veja referência 5.

x_1	x_2	x_3	z_1	z_2	Resposta da 1 ^a replicata	Resposta da 2 ^a replicata
0,85	0	0,15	+1	-1	8	7
0,72	0	0,28	+1	-1	6	5
0,60	0,25	0,15	+1	-1	10	11
0,47	0,25	0,28	+1	-1	4	5
0,66	0,125	0,125	+1	-1	11	10
0,85	0	0,15	-1	+1	12	10
0,72	0	0,28	-1	+1	9	8
0,60	0,25	0,15	-1	+1	13	12
0,47	0,25	0,28	-1	+1	6	3
0,66	0,125	0,125	-1	+1	15	11
0,85	0	0,15	-1	-1	7	8
0,72	0	0,28	-1	-1	7	6
0,60	0,25	0,15	-1	-1	9	10
0,47	0,25	0,28	-1	-1	5	4
0,66	0,125	0,125	-1	-1	9	7
0,85	0	0,15	+1	+1	12	11
0,72	0	0,28	+1	+1	10	9
0,60	0,25	0,15	+1	+1	14	12
0,47	0,25	0,28	+1	+1	6	5
0,66	0,125	0,125	+1	+1	13	9

Os dados foram tratados pela análise “split-plot” e como se fossem provenientes de experimentos completamente aleatórios. A Tabela 40 traz os modelos ajustados considerando as duas formas de análise. É possível observar que, independente do tratamento dado aos resultados, os parâmetros considerados significativos permanecem nesta condição.

Tabela 40. Modelos ajustados pela análise “split-plot” e completamente aleatória. Os parâmetros em negrito são considerados significativos.

Efeito	Análise para experimentos aleatórios			Análise split-plot		
	Coef.	Erro	Coef./erro	Coef.	Erro	Coef./erro
x₁	11,482	1,044	10,990	11,480	1,000	11,480
x₂	-69,540	12,270	-5,670	-69,540	9,860	-7,053
x ₃	-2,630	3,463	-0,760	-2,630	2,830	-0,929
x₁x₂	148,640	21,490	6,920	148,640	17,230	8,627
x ₁ z ₁	0,257	0,822	0,310	0,260	0,690	0,377
x ₂ z ₁	1,257	1,794	0,700	1,260	1,450	0,869
x ₃ z ₁	-0,705	2,798	-0,250	-0,700	2,250	-0,311
x₁z₂	2,953	0,822	3,590	2,960	0,690	4,290
x ₂ z ₂	-0,547	1,794	-0,300	-0,550	1,450	-0,379
x ₃ z ₂	-2,816	2,798	-1,010	-2,820	2,250	-1,253
x ₁ z ₁ z ₂	-0,570	0,822	-0,690	-0,570	0,690	-0,826
x ₂ z ₁ z ₂	-1,570	1,794	-0,880	-1,570	1,450	-1,083
x ₃ z ₁ z ₂	2,314	2,798	0,830	2,310	2,250	1,027

7.2 Utilização do teste F como ferramenta de apoio para a escolha de parâmetros em modelos split-plot.

A determinação dos parâmetros significativos em um modelo ajustado para as respostas de um planejamento “split-plot” não é uma tarefa simples, pois, primeiro, duas fontes de erro afetam a qualidade das respostas do “main-plot” e do “sub-plot”. Porém ao ajustar o modelo os parâmetros não apresentam apenas efeitos isolados do “main-plot” ou do “sub-plot”. Há vários níveis de interação entre efeitos que descrevem o “main-plot” e o “sub-plot”, além de interações cruzadas entre variáveis do próprio “main-plot”, ou do “sub-plot”. Em um modelo como o cúbico especial-bilinear para três variáveis de mistura e duas de processo o termo de interação de maior ordem é o $z_1z_2x_1x_2x_3$. Este termo ao ser ajustado possui uma incerteza associada a seu valor, contudo esta incerteza não provém de uma única fonte, mas de duas. Para determinar o erro associado a cada parâmetro utiliza-se a expressão

$$Cov(\hat{\beta}) = (\mathbf{X}^t \hat{\mathbf{V}} \mathbf{X})^{-1}$$

Equação 50

onde $\hat{\beta}$ é o vetor que contém os parâmetros ajustados por regressão, e a matriz $\hat{\mathbf{V}}$ é obtida da Equação 30. Contudo a Equação 50 é aplicada para casos onde tanto os experimentos quanto o modelo são balanceados. Para situações contendo modelos desbalanceados, que possuem graus de liberdade para testar a falta de ajuste, há uma expressão mais geral para determinar os erros⁵:

$$Cov(\hat{\mathbf{b}}) = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \hat{\mathbf{V}} \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1}.$$

Além da complexidade algébrica descrita há ainda um outro fator importante: depois de ajustado os parâmetros de um modelo quanto eles devem superar o seu erro para serem considerados significativos? A resposta a esta questão é obtida realizando-se um teste t com graus de liberdade adequados.

Contudo em um modelo “split-plot” a determinação dos graus de liberdade é uma tarefa complexa, específica para cada caso e obtida de forma aproximada⁴³.

Um método que já vem sendo testado com grande êxito em nossos estudos é a aplicação de gráficos de probabilidade acumulada como ferramenta para identificar os parâmetros significativos nos modelos “split-plot”. A técnica foi empregada em várias situações com ótimos resultados.

Uma nova possibilidade em estudo para auxiliar na escolha dos parâmetros significativos é a utilização do teste F para avaliar se um parâmetro contribui de forma significativa ou não para a soma quadrática devida a regressão. Para exemplificar o método será utilizado o planejamento da Tabela 35.

Inicialmente ajusta-se um modelo com apenas um parâmetro, além do intercepto α , $\hat{y} = \alpha + b_1x_1$, onde \hat{y} é o valor previsto e x_1 o efeito da primeira variável de mistura. Determinando-se as respostas previstas é possível calcular a soma quadrática devida à regressão. Para o modelo $\hat{y} = \alpha + b_1x_1$ o valor obtido é de 50,02. A seguir o modelo é completado com mais um termo, $\hat{y} = \alpha + b_1x_1 + b_2x_2$, onde x_2 representa o efeito da segunda variável de mistura. A nova soma quadrática fornecida pela regressão é de 57,58. Todavia não se pode esquecer que 50,02 dos 57,58 já eram explicados pelo efeito x_1 . Portanto, o que cabe a x_2 é a diferença, $57,58 - 50,02 = 7,56$. Como 7,56 é a soma quadrática resultante da adição de um termo extra ao modelo a mesma possui um grau de liberdade, e o valor da média quadrática será igual ao valor da própria soma quadrática. Usando o valor da média quadrática pode-se fazer um teste F com a média quadrática dos erros, pois um parâmetro significativo deve contribuir de forma significativa para a soma quadrática devida a regressão. Por outro lado, um parâmetro não significativo oriundo provavelmente da modelagem do erro experimental não deve, a priori, contribuir de fato para uma melhor modelagem dos efeitos. Assim sua contribuição para a soma quadrática devida a regressão deve ter pouca significância sendo detectada pela realização de um teste F.

O mesmo procedimento pode ser seguido para todo termo extra adicionado ao modelo. A Tabela 41 mostra o aumento da soma quadrática devido à regressão para adição de termos até completar o modelo linear-bilinear.

Tabela 41. Aumento na soma quadrática devido à regressão (SQ_{REG}) pela adição de termos no modelo ajustado aos dados da Tabela 35.

Modelo	SQ_{REG}	Aumento na SQ_{REG} com a adição do parâmetro
$\hat{y} = \alpha + b_1x_1$	50,02	50,02
$\hat{y} = \alpha + b_1x_1 + b_2x_2$	57,58	7,56
$\hat{y} = \alpha + b_1x_1 + b_2x_2 + b_3x_3 + b_1^1x_1z_1$	57,71	0,13
$\hat{y} = \alpha + b_1x_1 + b_2x_2 + b_3x_3 + b_1^1x_1z_1 + b_2^1x_2z_1$	58,21	0,50
$\hat{y} = \alpha + b_1x_1 + b_2x_2 + b_3x_3 + b_1^1x_1z_1 + b_2^1x_2z_1 + b_3^1x_3z_1$	61,34	3,13
$\hat{y} = \alpha + b_1x_1 + b_2x_2 + b_3x_3 + b_1^1x_1z_1 + b_2^1x_2z_1 + b_3^1x_3z_1 + b_1^2x_1z_2$	71,47	10,13
$\hat{y} = \alpha + b_1x_1 + b_2x_2 + b_3x_3 + b_1^1x_1z_1 + b_2^1x_2z_1 + b_3^1x_3z_1 + b_1^2x_1z_2 + b_2^2x_2z_2$	79,47	8,00
$\hat{y} = \alpha + b_1x_1 + b_2x_2 + b_3x_3 + b_1^1x_1z_1 + b_2^1x_2z_1 + b_3^1x_3z_1 + b_1^2x_1z_2 + b_2^2x_2z_2 + b_3^2x_3z_2$	89,6	10,13
$\hat{y} = \alpha + b_1x_1 + b_2x_2 + b_3x_3 + b_1^1x_1z_1 + b_2^1x_2z_1 + b_3^1x_3z_1 + b_1^2x_1z_2 + b_2^2x_2z_2 + b_3^2x_3z_2 + b_1^{12}x_1z_1z_2$	90,73	1,13
$\hat{y} = \alpha + b_1x_1 + b_2x_2 + b_3x_3 + b_1^1x_1z_1 + b_2^1x_2z_1 + b_3^1x_3z_1 + b_1^2x_1z_2 + b_2^2x_2z_2 + b_3^2x_3z_2 + b_1^{12}x_1z_1z_2 + b_2^{12}x_2z_1z_2$	98,73	8,00
$\hat{y} = \alpha + b_1x_1 + b_2x_2 + b_3x_3 + b_1^1x_1z_1 + b_2^1x_2z_1 + b_3^1x_3z_1 + b_1^2x_1z_2 + b_2^2x_2z_2 + b_3^2x_3z_2 + b_1^{12}x_1z_1z_2 + b_2^{12}x_2z_1z_2 + b_3^{12}x_3z_1z_2$	99,86	1,13

A Tabela 36 indica a ANOVA com a contribuição das diferentes fontes de variância, o que permite obter a soma e a média quadrática devido ao erro “sub-plot”. Assim pode-se realizar um teste F entre a média quadrática fornecida por regressão para cada parâmetro com a média quadrática do erro “sub-plot”. O valor

tabelado de $F_{1; 8}$ para 95% de confiança é de 5,32, portanto, a razão que superar este valor será considerada significativa. A Tabela 42 indica os parâmetros considerados significativos pelo teste F. Pela referência 5 os parâmetros x_1 , x_2 , x_3 , x_1z_2 , x_2z_2 e x_3z_2 também são considerados significativos, sendo que x_3z_1 não é significativo a 95% mas é a 94%.

Tabela 42. Valores de F para as médias quadráticas (MQ) dos parâmetros do modelo linear-bilinear. Os parâmetros significativos estão em negrito.

Fonte	MQ_{REG}	MQ erro sub-plot	$MQ_{REG} / MQ \text{ erro}$
x_1	50,02	0,2917	171,4803
x_2	7,56	0,2917	25,9256
x_1z_1	0,13	0,2917	0,4285
x_2z_1	0,50	0,2917	1,7141
x_3z_1	3,13	0,2917	10,7130
x_1z_2	10,13	0,2917	34,7103
x_2z_2	8,00	0,2917	27,4254
x_3z_2	10,13	0,2917	34,7103
$x_1z_1z_2$	1,13	0,2917	3,8567
$x_2z_1z_2$	8,00	0,2917	27,4254
$x_3z_1z_2$	1,13	0,2917	3,8567

Um segundo conjunto de dados presente na referência 5 foi analisado da mesma forma que a descrita anteriormente. O planejamento contendo duas variáveis de processo, z_1 e z_2 , em dois níveis e três variáveis de mistura formando cinco composições diferentes em um planejamento “split-plot” está indicado na Tabela 39. Foi realizada a ANOVA com o modelo linear-bilinear contendo o termo cruzado x_1x_2 apresentada na Tabela 43. Alguns autores como Cornell⁵ e Kowalski⁴¹ adicionam ao erro “sub-plot” as faltas de ajustes do modelo, o mesmo procedimento foi utilizado neste caso.

Tabela 43. ANOVA para os dados da Tabela 39 com o modelo linear-bilinear contendo o termo x_1x_2 ajustado. As faltas de ajuste do modelo foram adicionadas ao erro “sub-plot”.

Fonte de variância	GL	SQ	MQ
Replicatas (R)	1	13,225	13,225
Main-plot (Z)	3	66,475	22,158
Erro main-plot (RZ)	3	7,475	2,492
Regressão sub-plot (XREG)	3	226,819	75,606
Regressão main-sub-plot (ZXREG)	6	12,875	2,146
Erro Sub-plot	23	25,160	1,092

A soma quadrática referente a cada parâmetro ajustado na regressão foi determinada e comparada com o valor do erro “sub-plot”. Um teste F foi realizado e comparado com o valor tabelado de $F_{1; 23; 0,95} = 4,28$, os valores estão indicados na Tabela 44. Os termos considerados significativos estão em negrito e correspondem exatamente aos mesmos determinados pelo teste t indicado na referência 5.

Tabela 44. Somas quadráticas dos parâmetros ajustados por regressão. Os valores em negrito são considerados significativos a 95% de confiança.

Efeito	MQ _{REG}	MQ _{erro sub-plot}	MQ _{REG} /MQ _{erro sub-plot}
x_1	42,08	1,092	38,53
x_2	103,55	1,092	94,82
x_1x_2	81,19	1,092	74,35
x_1z_1	1,00	1,092	0,92
x_2z_1	0,74	1,092	0,68
x_3z_1	0,11	1,092	0,10
x_1z_2	72,52	1,092	66,41
x_2z_2	1,41	1,092	1,29
x_3z_2	1,72	1,092	1,57
$x_1z_1z_2$	0,19	1,092	0,18
$x_2z_1z_2$	0,50	1,092	0,45
$x_3z_1z_2$	1,16	1,092	1,06

Portanto, nos estudos realizados o uso do teste F como ferramenta de apoio para a seleção de parâmetros significativos apresentou resultados satisfatórios.

8 SIMULAÇÕES EM PLANEJAMENTOS SPLIT-PLOT

O estudo de planejamentos “split-plot” com simulações é extremamente importante, pois tem como objetivo esclarecer alguns pontos complexos da modelagem, análise de variância, cálculo dos erros dos parâmetros e determinação da robustez dos métodos utilizados em diferentes situações de planejamentos.

O tipo de simulação desejada com o método “split-plot” não é comum na literatura. Assim o primeiro passo foi descobrir como realizar as simulações de forma correta. Como há mais de uma fonte de erro envolvida a simulação não é trivial. Para gerar os dados antes é necessário entender quais os princípios estatísticos envolvidos, qual o modelo matemático a ser utilizado e como os dados devem se comportar já que o método “split-plot” possui algumas pré-suposições.

O método “split-plot” pode considerar que há seis fontes de variância afetando uma resposta e seu modelo pode ser escrito como

$$y_{ijk} = \mu + R_i + Z_j + RZ_{ij} + X_k + ZX_{jk} + \varepsilon_{ijk} , \quad (\text{Equação 7})$$

onde y_{ijk} representa uma resposta. Observa-se que a resposta depende da média, μ , e de mais seis fontes, sendo

R_i o efeito da i -ésima replicatas;

Z_j o efeito da j -ésima condição de processo;

X_k o efeito da k -ésima composição de mistura;

ZX_{jk} o efeito da k -ésima composição de mistura na j -ésima condição de processo;

RZ_{ij} o efeito do erro do “main-plot” e;

ε_{ijk} o erro do “sub-plot”.

Vale notar que três fontes são originadas dos erros das medidas experimentais e seus valores possuem uma distribuição normal com média zero. Estes fatores correspondem a R_i cuja distribuição deve possuir variância σ_R^2 , RZ_{ij} com variância σ_{RZ}^2 e ε_{ijk} com variância σ_e^2 .

Utilizando o modelo anterior pode-se prever a composição das variâncias para cada um dos seis fatores. A expectativa, ou esperança, da média quadrática está descrita na Tabela 45.⁵

Tabela 45. ANOVA para o modelo “split-plot”.

Fonte de variação	GL	MQ	E (MQ)
Replicatas (R)	(r - 1)	MQR	$\sigma_e^2 + m\sigma_{RZ}^2 + mP\sigma_R^2$
Main-plot (Z)	(P - 1)	MQZ	$\sigma_e^2 + m\sigma_{RZ}^2 + rm\theta_Z^2$
Erro main-plot (RZ)	(r - 1)(P - 1)	MQRZ	$\sigma_e^2 + m\sigma_{RZ}^2$
Sub-plot (X)	(m - 1)	MQX	$\sigma_e^2 + rP\theta_X^2$
Interação main-sub-plot (ZX)	(P - 1)(m - 1)	MQZX	$\sigma_e^2 + r\theta_{ZX}^2$
Erro sub-plot	P(r - 1)(m - 1)	MQE	σ_e^2

Analisando a Tabela 45 nota-se que o erro “sub-plot” afeta todas as médias quadráticas, uma vez que a composição da variância de todos os efeitos possui o termo σ_e^2 . Já a variância devida ao erro “main-plot”, σ_{RZ}^2 , aparece na composição do termo RZ, Z e R, não afetando as variâncias envolvendo o “sub-plot”. Por fim a variância das replicatas, σ_R^2 , influencia apenas o termo R não afetando os demais. Portanto, uma simulação conduzida de forma correta deverá ser coerente com as previsões descritas na Tabela 45.

A primeira etapa da simulação consiste na criação de um planejamento experimental. Neste caso foi adotado um planejamento fatorial, equivalente ao indicado na Figura 8 e Tabela 8, contendo quatro variáveis, a, b, c e d, sendo que a variável a constitui o “main-plot” e as outras formam o “sub-plot”. As respostas foram geradas pelo modelo pré-estabelecido

$$y = 15 + 11a + 13b + 0c - 8d + 3,5bc + 7bd + 0cd + 5ab + 0ac + 0ad$$

Equação 51

e nelas foram adicionados os erros gerados.

O primeiro estudo envolveu aumentos nos valores dos erros de replicatas sobre as respostas, mantendo-se as outras fontes de variância fixadas. Erro

padrão crescente cujo valor variou no intervalo de 0,5 até 5,0 foi aplicado às respostas. Este intervalo foi dividido em 50 partes, sendo que cada parte equivale a uma etapa da simulação. Para cada experimento foram simuladas 10000 replicatas.

Com as respostas geradas pela simulação foi realizada a ANOVA, obtendo-se os novos coeficientes para o modelo por regressão, os componentes de variância, σ_R^2 , σ_{RZ}^2 e σ_e^2 , os valores para a razão coeficiente/erro padrão = t, e foram construídos os gráficos de probabilidade acumulada para os valores de t.

A Figura 36 indica os valores de variância (a) e desvio padrão (b) para as simulações com a elevação do erro das replicatas. Pode-se notar que os valores de erro padrão controlados adicionados às respostas são recuperados na análise de variância pelos componentes de variância. Isto é um indicativo claro de que a simulação foi conduzida de forma correta.

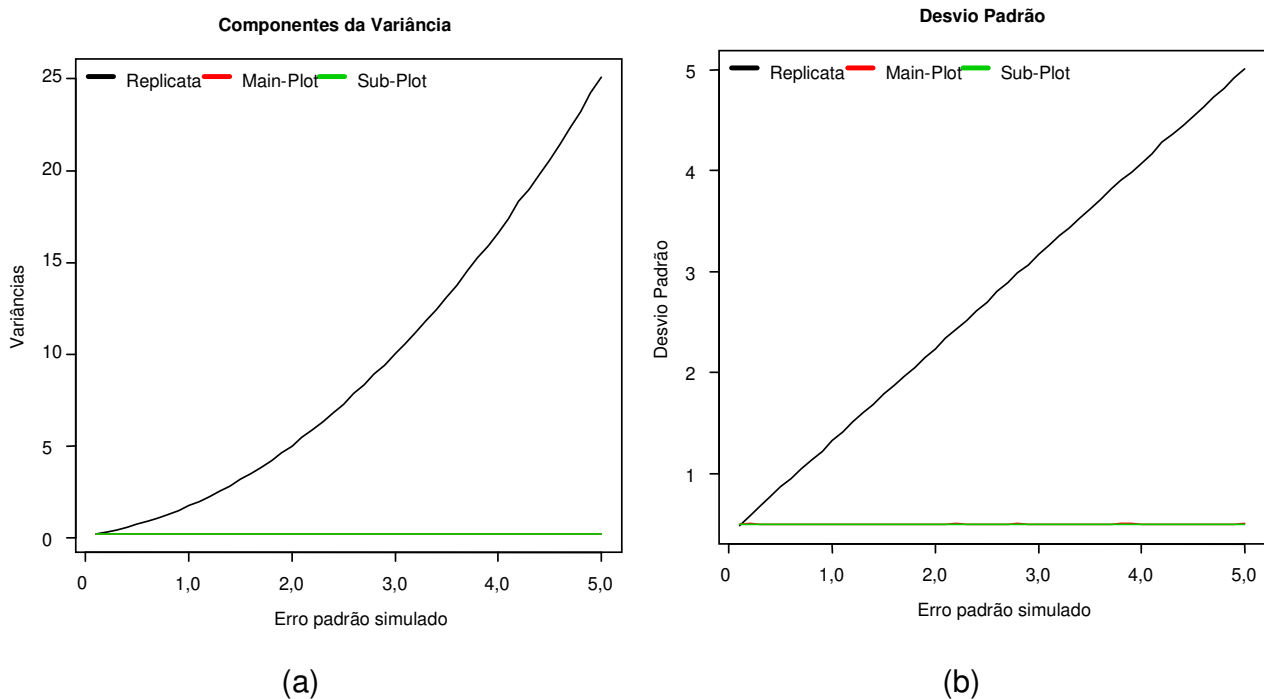


Figura 36. (a) Componente de variância para a simulação contendo 10000 replicatas por experimento em que o erro padrão das replicatas foi elevado de 0,5 a 5,0; (b) valores dos desvios padrão.

A Figura 36(b) também indica que realmente foi possível controlar os acréscimos de erro às respostas. Pequenas perturbações são observadas no componente de variância do “main-plot”, mas o valor oscila em torno do desvio padrão pré-fixado para este componente, assim como para o “sub-plot”, de 0,5.

A Figura 37 indica os valores dos coeficientes dos parâmetros dos modelos obtidos por regressão pelo método de mínimos quadrados generalizados. O valor de cada coeficiente, em quase todos os casos, sofre poucas alterações e oscila claramente em torno do valor inicialmente gerado. Isto indica que o método de mínimos quadrados generalizados é eficiente, neste caso, para recuperar os valores dos coeficientes com respostas alteradas por erros.

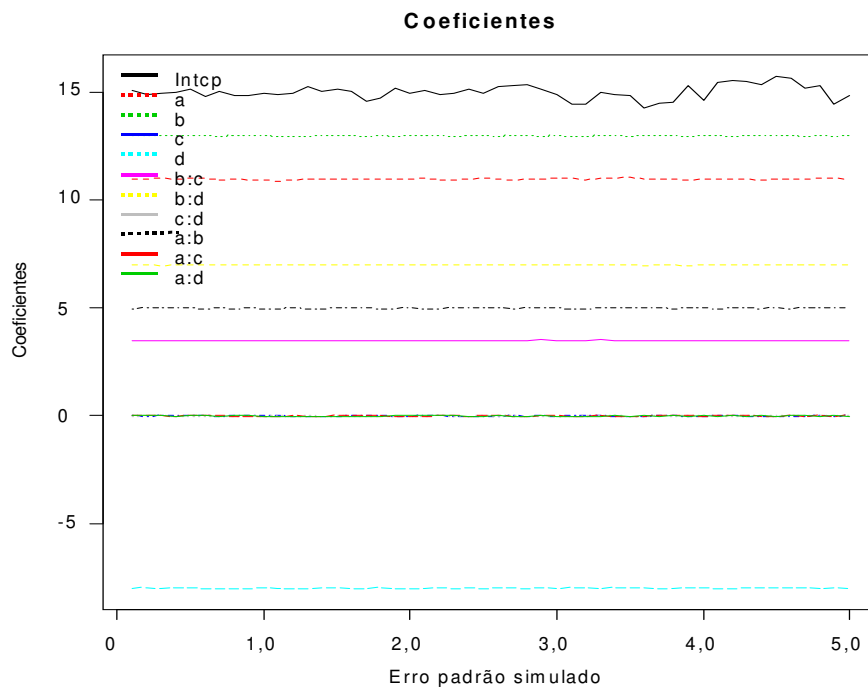


Figura 37. Coeficientes dos parâmetros do modelo obtido por regressão linear pelo método de mínimos quadrados generalizados para a simulação contendo 10000 replicatas por experimento em que o erro padrão das replicatas foi elevado de 0,5 a 5,0.

A Figura 38 contém os valores da razão coeficientes/erro padrão (t) para os 50 modelos gerados com elevação controlada do erro das replicatas. É possível

observar que quase todos os valores oscilam em torno de um valor médio. Contudo o valor de t para o intercepto (t_{intcp}) cai com a evolução das etapas da simulação, ou seja, a elevação do erro das replicatas. Este fato, que aparentemente, não é tão simples de se interpretar pode ser compreendido com a dedução analítica do erro associado a cada parâmetro. Para este caso em particular a expressão para o cálculo do erro padrão associado aos coeficientes dos parâmetros é escrito como

$$s_{intcp} = \sqrt{\frac{1}{r}\sigma_R^2 + \frac{1}{rp}\sigma_{RZ}^2 + \frac{1}{rmp}\sigma_e^2} \quad (\text{erro padrão associado ao intercepto, com } r \text{ replicatas, } p \text{ condições de processo e } m \text{ composições de mistura})$$

Equação 52

$$s_{main-plot} = \sqrt{\frac{1}{rp}\sigma_{RZ}^2 + \frac{1}{rmp}\sigma_e^2} \quad (\text{erro padrão associado ao coeficiente do efeito “main-plot”})$$

Equação 53

$$s_{sub-plot} = \sqrt{\frac{1}{rmp}\sigma_e^2} \quad (\text{erro padrão associado aos coeficientes dos efeitos do “sub-plot”}),$$

Equação 54

assim é possível notar que a variância associada as replicatas, σ_R^2 , só aparece na expressão do erro do intercepto, s_{intcp} , e não influencia os erros do “main-plot” ou “sub-plot”. Portanto, isto justifica plenamente o fato de que a elevação no erro das replicatas diminui a significância do valor do intercepto e não afeta os valores de t dos outros parâmetros. As equações anteriores (54 – 56) foram obtidas utilizando-se um programa escrito em Matlab para obtenção de expressões simbólicas.

Outras simulações contendo um número menor de replicatas foram conduzidas para avaliar as implicações nestes casos. A Figura 39 indica os valores dos componentes de variância obtidos em simulações realizadas com 2, 4, 8, 16, 32 e 64 replicatas por experimento. Com um número muito pequeno de replicatas, 2 ou 4, a variância das replicatas sofre grandes oscilações no decorrer dos passos das simulações, incluindo reduções de variância em passos que deveriam indicar valores crescentes. Além disto, o componente de variância do

“main-plot” apresenta oscilações mais pronunciadas em situações com poucas replicatas. Isto está basicamente associado ao fato de que os números aleatórios quando gerados em quantidades muito reduzidas pouco refletem a distribuição normal. A normalidade dos erros gerados é fundamental na realização das simulações. Com a elevação do número de replicatas as oscilações nos valores crescentes do componente de variância das replicatas são menores, assim como as perturbações apresentadas pelo componente associado à variância do “main-plot”. As simulações com 64 replicatas já apresentam valores muito próximos aos obtidos com 10.000 replicatas.

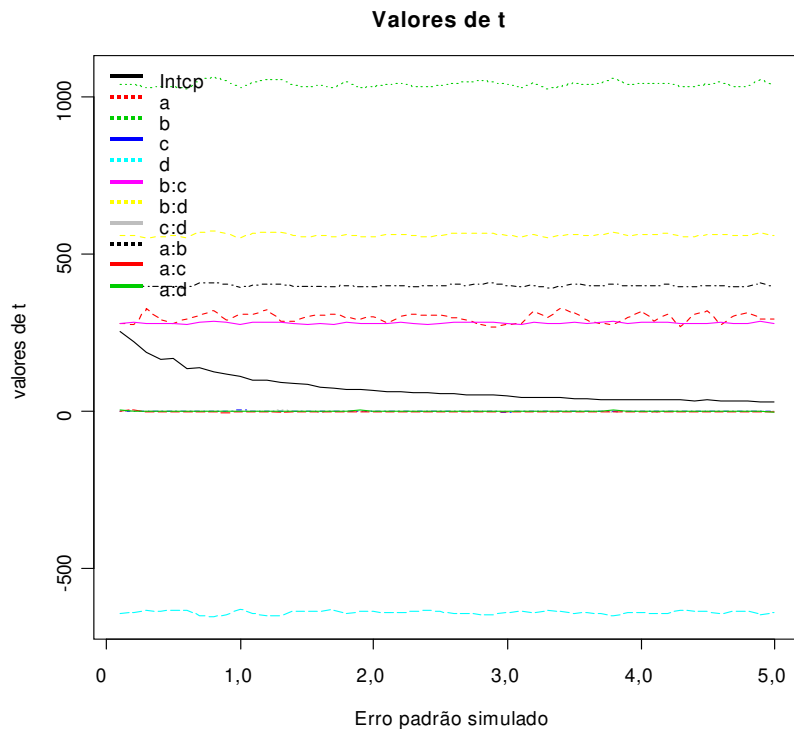


Figura 38. Razão coeficientes/erro padrão (t) para a simulação contendo 10000 replicatas por experimento em que o erro padrão das replicatas foi elevado de 0,5 a 5,0.

A Figura 40 indica os valores para as razões coeficientes/erro padrão, t , para experimentos simulados com replicatas que variaram de 2 até 64. Para os casos contendo duas e quatro replicatas há uma grande variação nos valores de t

para praticamente todos os coeficientes, com exceção dos que possuem valor zero. Com o aumento do número de replicatas para 16, Figura 40 (d), as oscilações apresentadas pelos valores de t diminuem sensivelmente, e a diminuição gradativa da significância do valor do intercepto se torna mais evidente.

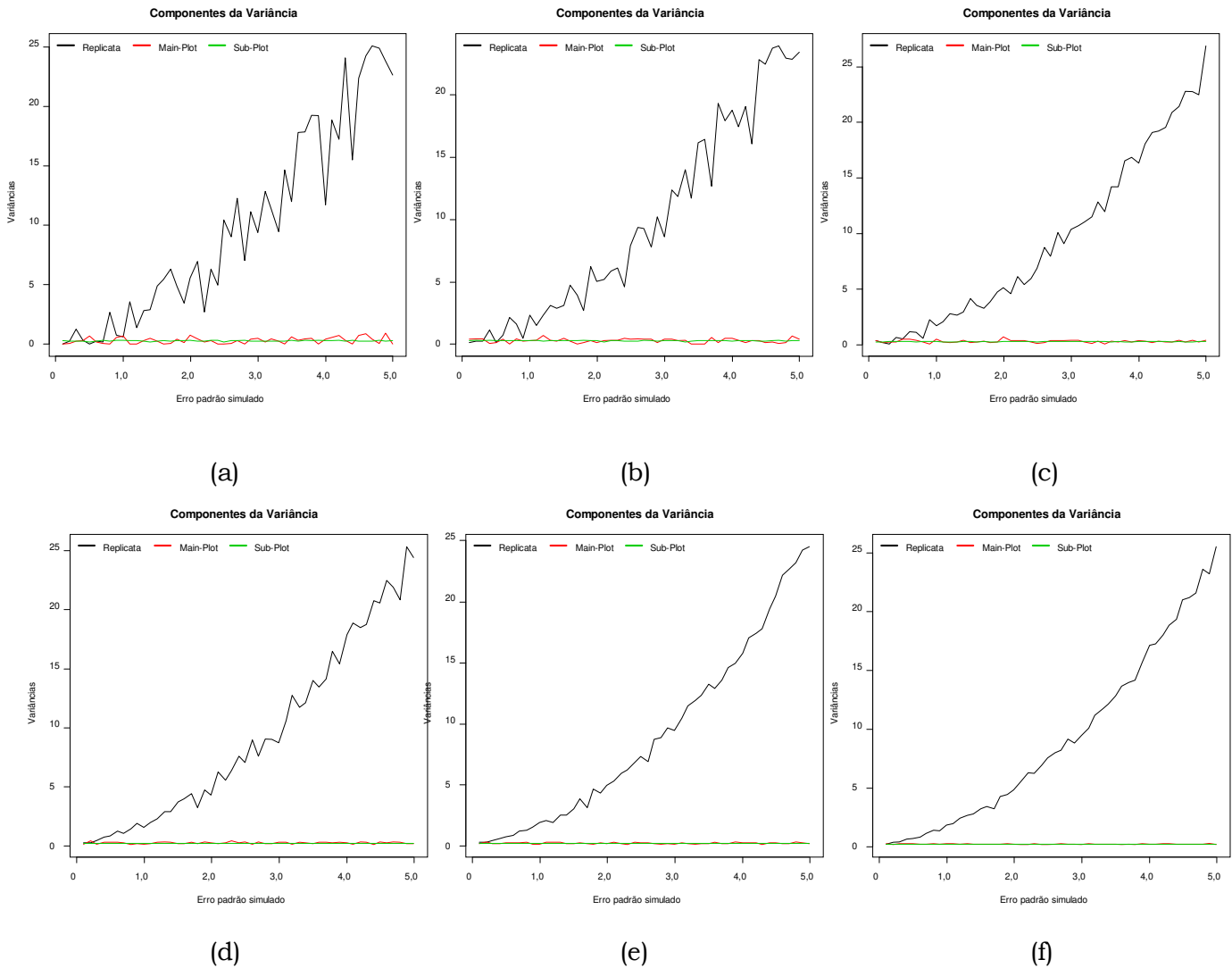


Figura 39. Valores dos componentes de variância para a simulação em que o erro padrão das replicatas foi elevado de 0,5 a 5,0, a, b, c, d, e, f representam simulações realizadas com 2, 4, 8, 16, 32 e 64 replicatas, respectivamente, por experimento.

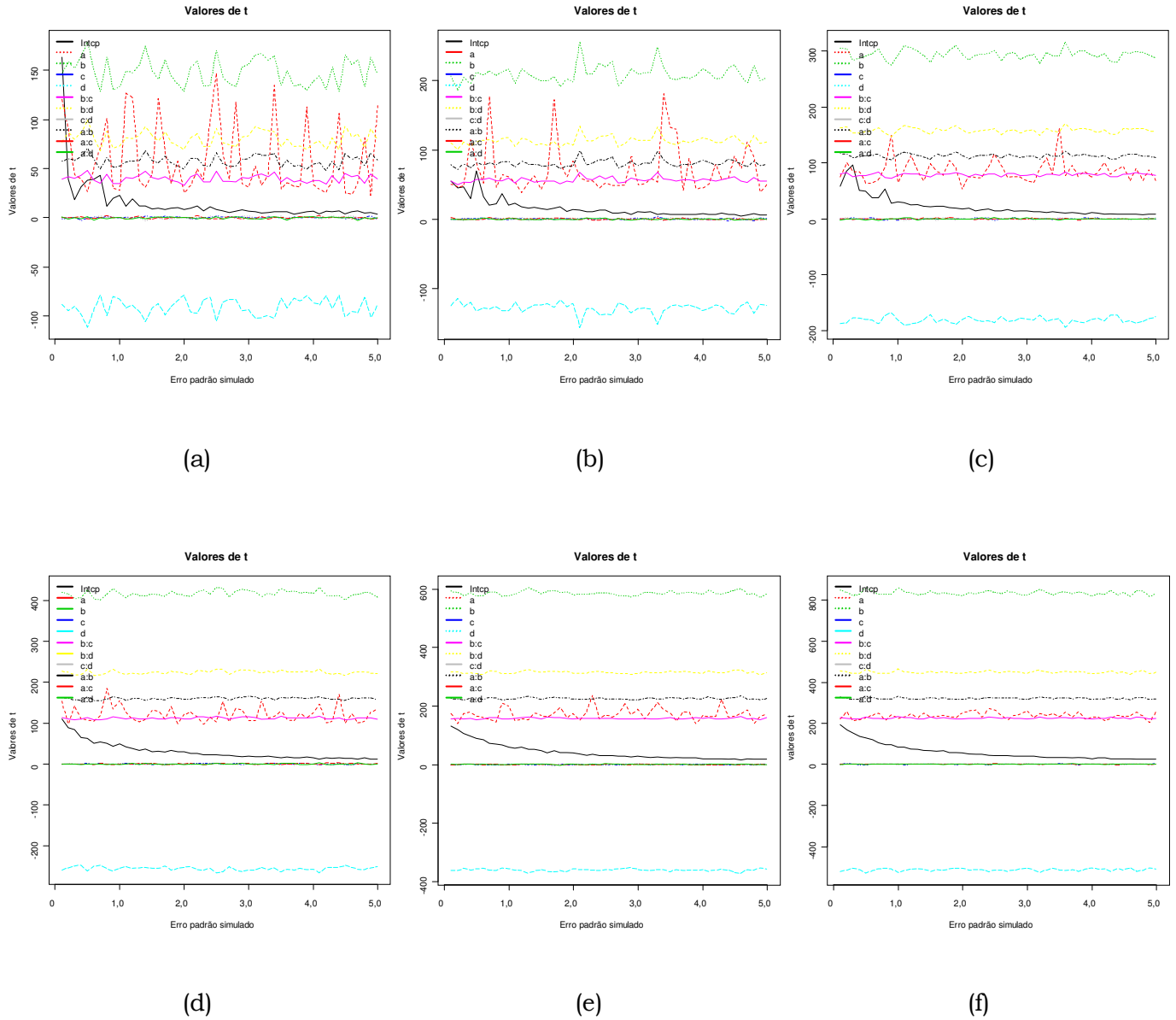


Figura 40. Valores de t para simulações em que o erro padrão das replicatas foi elevado de 0,5 a 5,0; a, b, c, d, e, f representam simulações realizadas com 2, 4, 8, 16, 32 e 64 replicatas, respectivamente, por experimento.

Pela Figura 40 pode-se observar ainda que mesmo em condições com poucas replicatas, de forma geral, os coeficientes pré-definidos no modelo como diferentes de zero permanecem com valores de t também nesta condição.

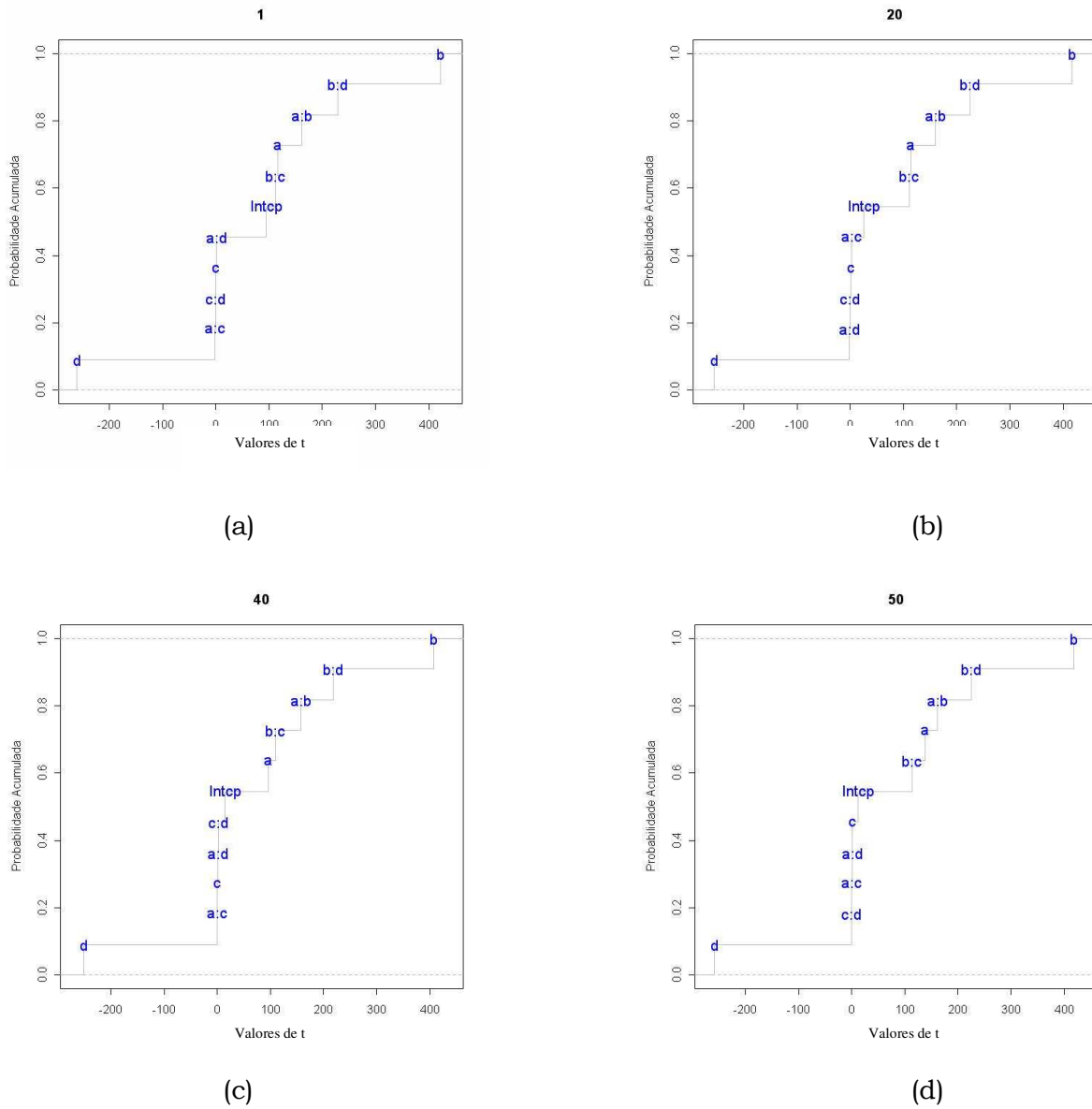


Figura 41. Gráficos de probabilidade para a simulação contendo 10000 replicatas por experimento em que o erro padrão das replicatas foi elevado de 0,5 a 5,0 através de 50 etapas. O gráfico (a) indica a etapa 1 da simulação, o (b) a etapa 20, (c) e (d) indicam as etapas 40 e 50 das simulações, respectivamente.

A Figura 41 mostra os gráficos de probabilidade acumulada para os valores de t . Pode-se constatar que os valores de t para o intercepto (Intcp) vão se aproximando de zero na medida em que as etapas das simulações prosseguem com a elevação da variância das replicatas. Esta constatação está de acordo com o previsto pela Equação 52. Vale notar também que os coeficientes dos parâmetros a , b , ab , bc , bd e d mantêm-se distante do zero do eixo da abscissa conservando a significância ao longo do processo de simulação. Por outro lado, os coeficientes dos parâmetros c , ac , ad e cd permanecem centrados em zero com o decorrer da simulação. Estas constatações estão de acordo com as características do modelo indicado pela Equação 51.

O segundo estudo envolveu aumentos nos valores dos erros do “main-plot” sobre as respostas, mantendo-se as outras fontes de variância constantes com desvio padrão igual a 0,5. Foram novamente aplicadas cinquenta faixas de erro crescentes sobre as respostas que variaram com um desvio padrão de 0,5 até 5,0. Foram geradas simulações contendo 10000 replicatas por experimento.

O primeiro fato a ser verificado é se a simulação é realizada de forma correta, coerente com as expectativas fornecidas pela Tabela 45. A Figura 42 indica a evolução dos valores dos componentes de variância e os desvios padrão. O erro “main-plot” cresce de forma linear de acordo com o planejamento de simulação adotado, isto indica que a variância dos valores aleatórios gerados de forma controlada é recuperada após a análise de variância. Isto comprova novamente que o modelo estatístico utilizado é eficiente e reforça a consistência das simulações realizadas. Pela Figura 42 é possível notar que o componente de variância da replicata é mais sensível as alterações do erro “main-plot” se comparado ao componente de variância do “sub-plot”. A Figura 43 apresenta os coeficientes dos parâmetros do modelo ajustados a cada etapa da simulação, verifica-se que as oscilações dos valores são pequenas ao longo das etapas.

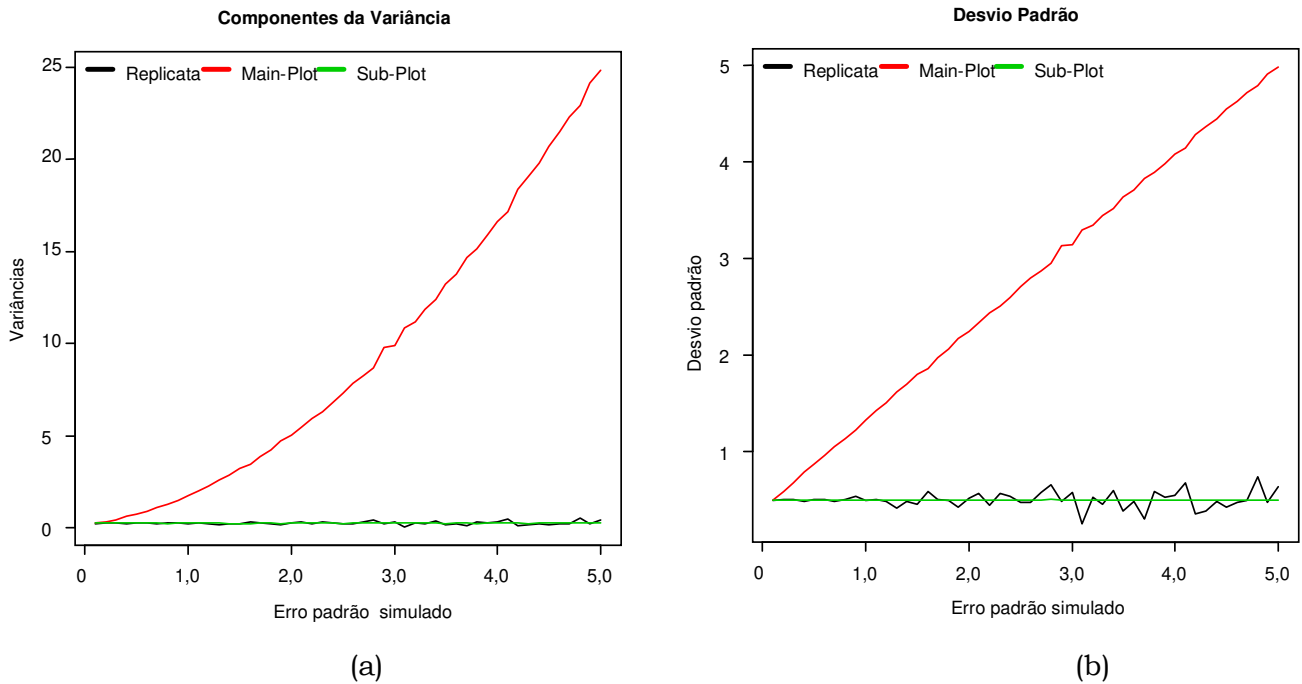


Figura 42. (a) Componente de variância para a simulação contendo 10000 replicatas por experimento em que o erro padrão do “main-plot” foi elevado de 0,5 a 5,0 através de 50 etapas, (b) valores dos desvios padrão dos componentes.

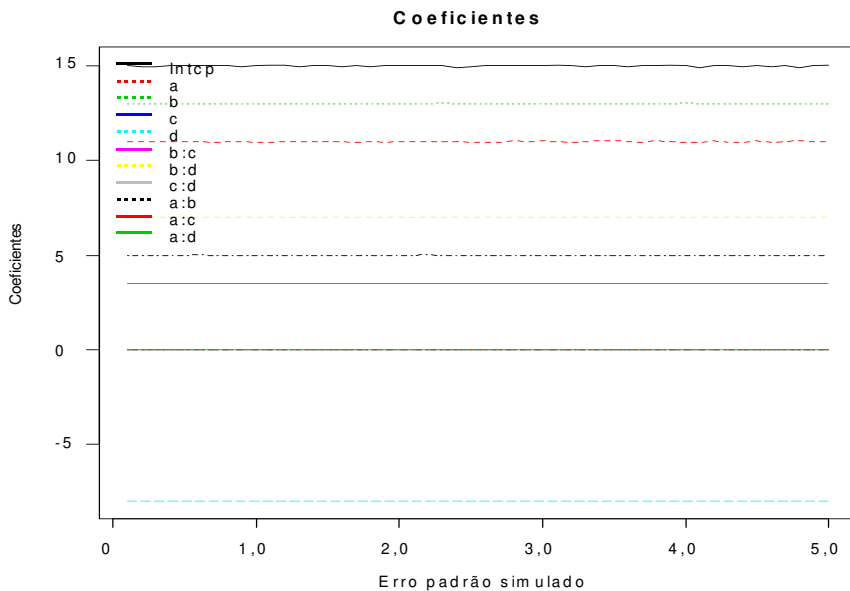


Figura 43. Coeficientes dos parâmetros do modelo obtido por regressão linear pelo método de mínimos quadrados generalizados para a simulação contendo 10000 replicatas por experimento. O erro padrão do “main-plot” foi elevado de 0,5 a 5,0 através de 50 etapas.

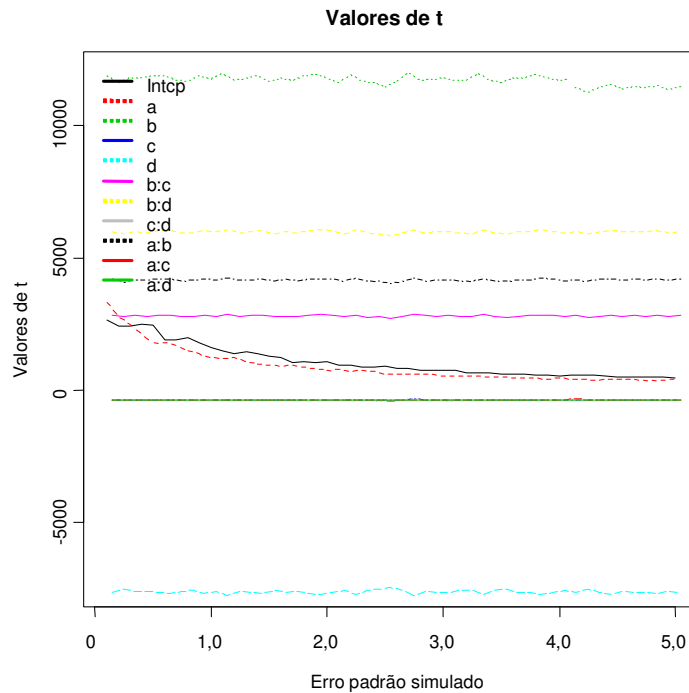


Figura 44. Valores de t para os modelos obtidos por regressão linear pelo método de mínimos quadrados generalizados para a simulação contendo 10000 replicatas por experimento. O erro padrão do “main-plot” foi elevado de 0,5 a 5,0.

Observando a Figura 44, que indica os valores de t para os parâmetros dos modelos obtidos ao longo da realização das simulações, é possível notar que os valores associados ao intercepto e ao efeito “main-plot” diminuem com a elevação da variância do erro “main-plot”. Esta observação é coerente com o previsto pela Equação 52 e Equação 53. A Figura 45 apresenta os gráficos de probabilidade acumulada para os valores de t dos parâmetros dos modelos gerados ao longo do processo de simulação. Nota-se que os valores de t para o intercepto (Intcp) e para o “main-plot” (a) são os que sofrem alterações claras e tendem em direção ao zero da abscissa, havendo até uma alteração na ordem dos mesmos da figura (a) para a (b). Os outros parâmetros significativos permanecem ocupando as mesmas posições, e os não significativos se mantêm centrados em zero. Novamente este comportamento já era esperado pela avaliação da Equação 52 e Equação 53.

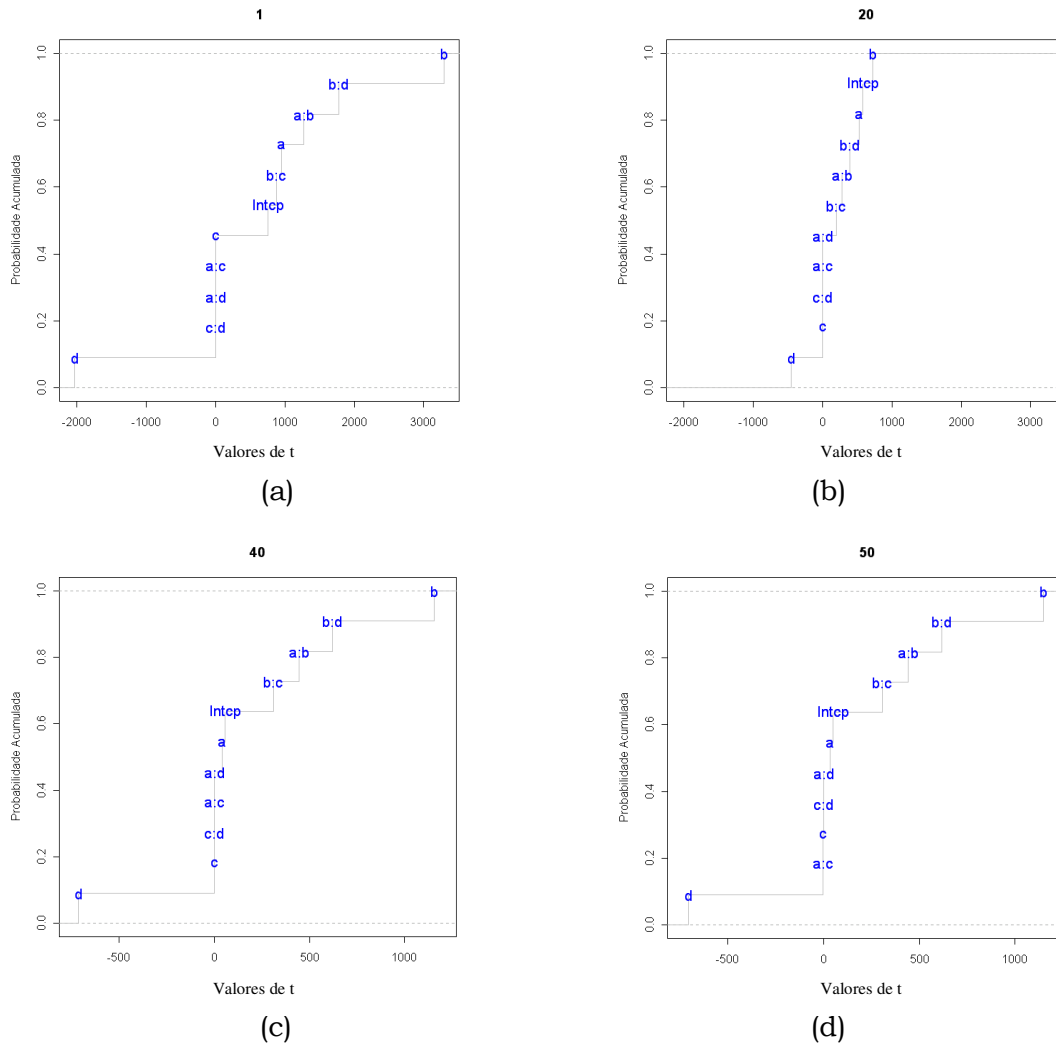


Figura 45. Gráficos de probabilidade para a simulação contendo 10000 replicatas por experimento em que o erro padrão do “main-plot” foi elevado de 0,5 a 5,0. O gráfico (a) indica a etapa 1 da simulação, o (b) a etapa 20, (c) e (d) indicam as etapas 40 e 50 das simulações, respectivamente.

O terceiro estudo teve como objetivo elevar gradativamente o valor do erro “sub-plot” sobre as respostas geradas pelo modelo. As outras fontes de variância foram mantidas constantes com um valor de desvio padrão de 0,5. Simulações com 10000 replicatas por experimento foram executadas elevando-se o erro padrão do “sub-plot” de 0,5 até 5,0. Como nos casos anteriores o primeiro ponto a

se observar é se a simulação está realmente sendo conduzida de forma correta, ou seja, de acordo com o previsto pela Tabela 45.

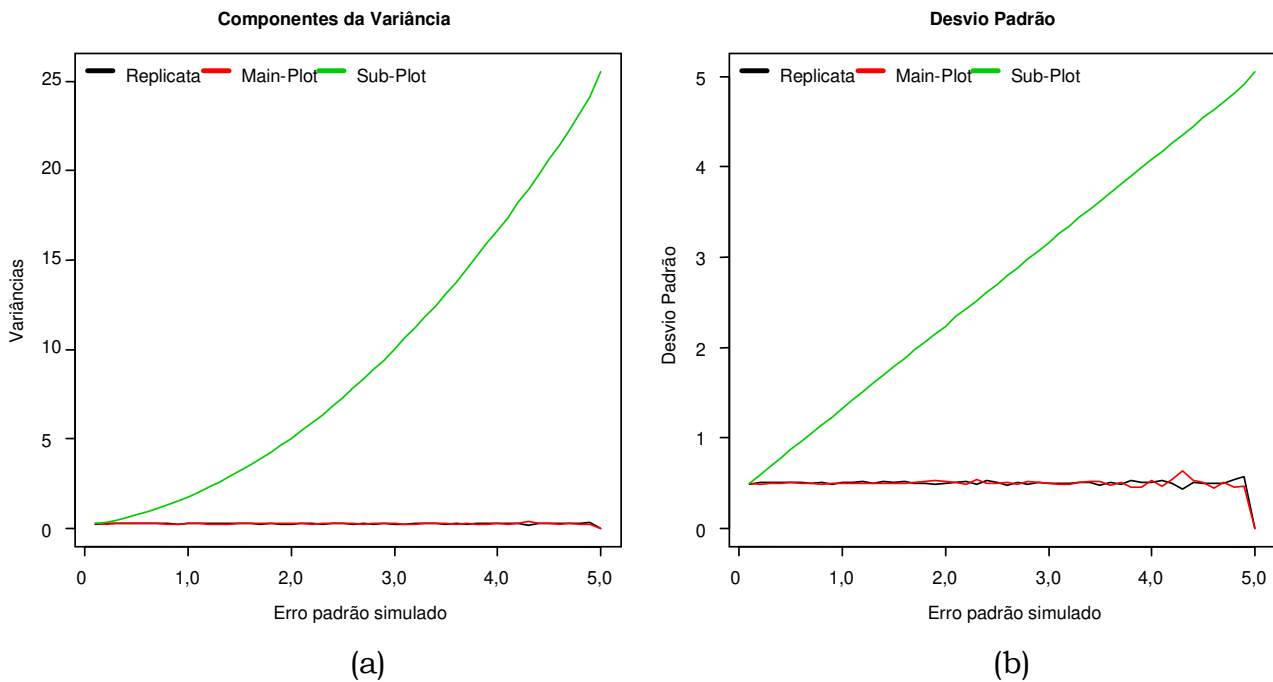


Figura 46. (a) Componente de variância para a simulação contendo 10000 replicatas por experimento em que o erro padrão do “sub-plot” foi elevado de 0,5 a 5,0, (b) valores dos desvios padrão.

Observando a Figura 46, que apresenta os valores dos componentes de variância (a) e dos desvios padrão (b), percebe-se que os resultados são coerentes com o desejado, ou seja, uma elevação linear dos desvios padrão do “sub-plot” com o decorrer da simulação. Os valores do erro “main-plot” e do erro das replicatas também apresentam oscilações ao longo das etapas da simulação, mas tendem a ficar centrados em 0,5. Contudo, nas etapas finais da simulação quando o erro padrão está próximo de 5,0 o erro “main-plot” e das replicatas se torna mais influenciado e as oscilações se intensificam.

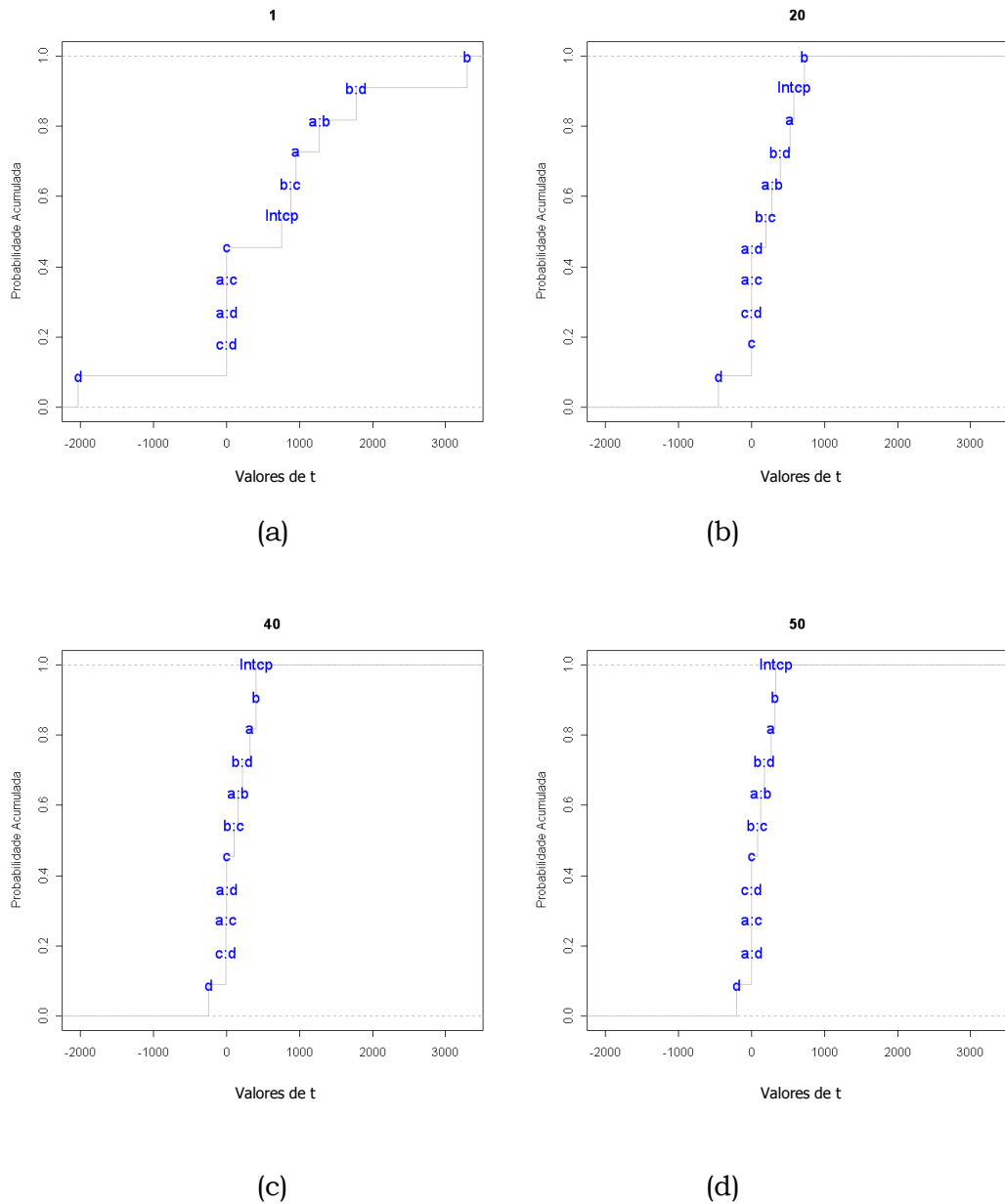


Figura 47. Gráficos de probabilidade para a simulação contendo 10000 replicatas por experimento em que o erro padrão do “sub-plot” foi elevado de 0,5 a 5,0. O gráfico (a) indica a etapa 1 da simulação, o (b) a etapa 20, (c) e (d) indicam as etapas 40 e 50 das simulações, respectivamente.

A Figura 47 mostra a evolução dos gráficos de probabilidade acumulada durante a simulação com o aumento do erro “sub-plot” para os modelos criados nas etapas 1, 20, 40 e 50 que correspondem as figuras a, b, c e d, respectivamente. Comparando o modelo indicado na figura a com o b observa-se que os

parâmetros que indicam o efeito do “sub-plot” sofrem grandes alterações. Estes parâmetros se aproximam muito do zero indicando elevada perda de significância. Na medida em que o erro “sub-plot” é elevado nas etapas seguintes da simulação este processo se intensifica. Valores para a razão coeficiente/erro padrão de alguns parâmetros como b , bc e ab aproximam-se claramente do zero. Os valores de t para os coeficientes do efeito “main-plot” (a) e do intercepto (Intcp) também são influenciados, mas de maneira menos efetiva, pois o erro padrão associado a estes parâmetros depende de outros fatores como indicado na Equação 52 e Equação 53.

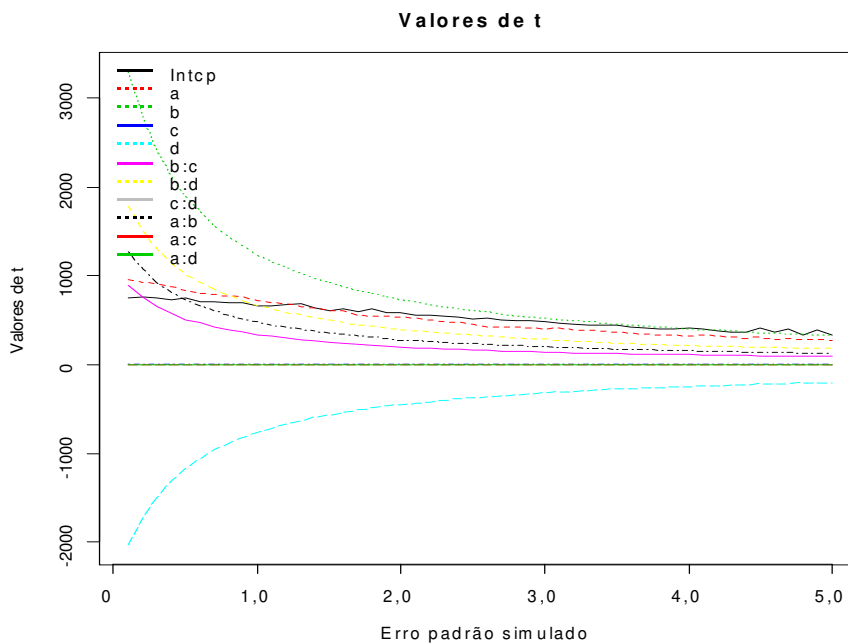


Figura 48. Valores de t para os modelos obtidos por regressão linear pelo método de mínimos quadrados generalizados para a simulação contendo 10000 replicatas por experimento em que o erro padrão do “sub-plot” foi elevado de 0,5 a 5,0.

A Figura 48 indica os valores das razões coeficientes/erro padrão para os parâmetros dos modelos obtidos no processo de simulação. Nota-se uma queda acentuada nos valores associados ao efeito “sub-plot”, e uma queda mais suave nos valores associados ao “main-plot” (a) e ao intercepto.

Também foram realizadas simulações com um número menor de replicatas para avaliar as conseqüências sobre os componentes da variância, os valores de t e os coeficientes obtidos nos modelos por regressão linear. Simulações contendo 2, 4, 8, 16, 32 e 64 replicatas foram realizadas. O erro padrão do “main-plot” e das replicatas foram mantidos constantes, enquanto o erro do “sub-plot” foi elevado ao longo de 50 etapas de 0,5 até 5,0. A Figura 49 apresenta os gráficos dos componentes de variância para simulações realizadas com 2, 4, 8, 16, 32 e 64 replicatas por experimento, indicados por a, b, c, d, e, f, respectivamente. Com um número reduzido de replicatas, 2 ou 4, os componentes de variância, incluindo o do “main-plot” e das replicatas, apresentam muitas oscilações em torno do valor desejado. Isto está relacionado, como já mencionado, ao fato de que ao gerar quantidades reduzidas de replicatas há a possibilidade dos valores não refletirem precisamente a distribuição normal. As simulações que utilizaram maiores quantidades de replicatas por experimento, 32 e 64, geraram dados com distribuição mais próxima da desejada. Assim as replicatas ao serem analisadas geram valores na ANOVA compatíveis com o pré-determinado na geração dos dados simulados.

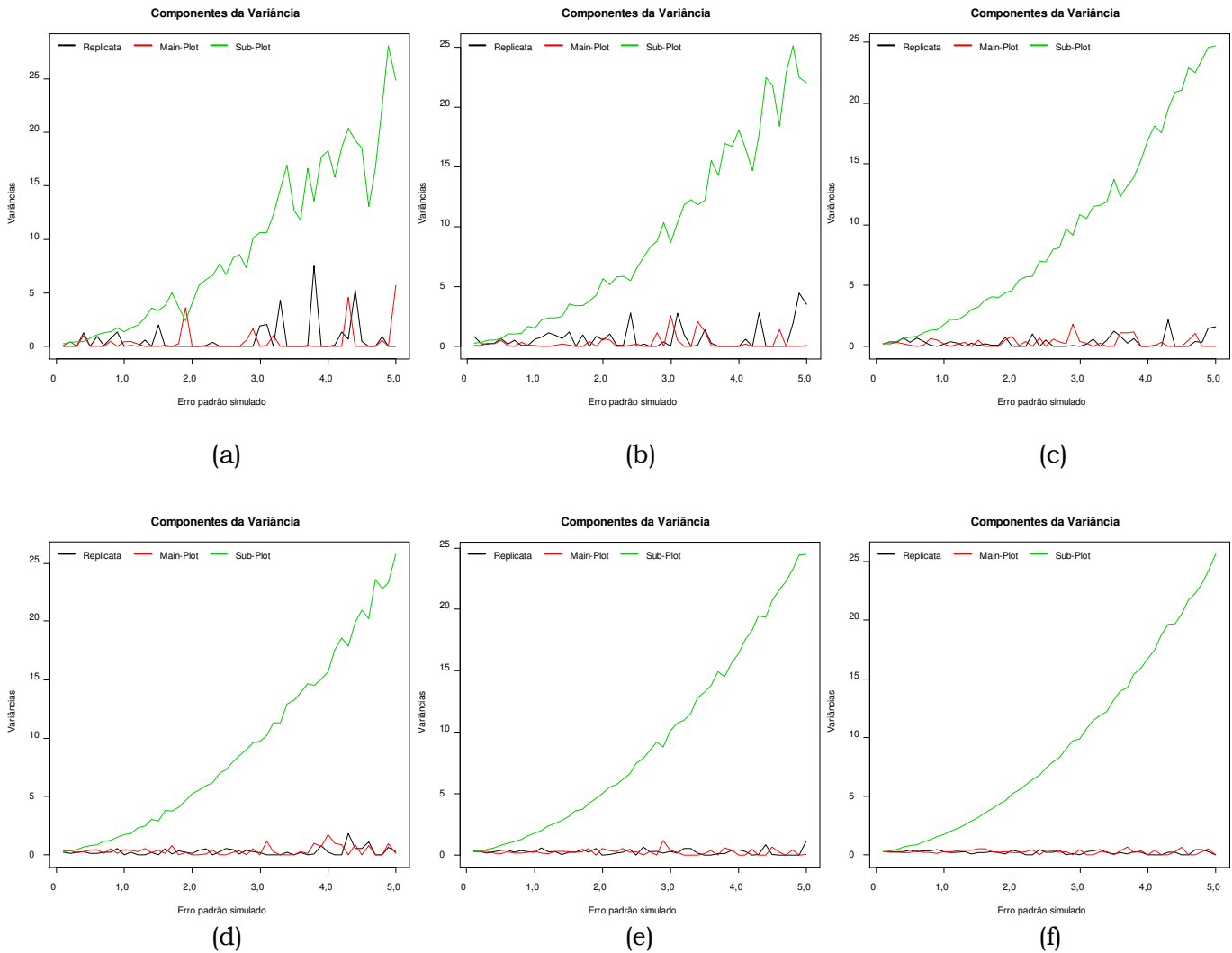


Figura 49. Gráficos com os valores dos componentes de variância da replicata, erro “main-plot” e erro “sub-plot”, a, b, c, d, e, f representam simulações realizadas com 2, 4, 8, 16, 32 e 64 replicatas, respectivamente, por experimento.

Simulações com poucas replicatas também indicam que os valores dos coeficientes dos modelos podem sofrer grandes oscilações. Isto poderia acarretar uma interpretação incorreta sobre a significância dos efeitos em algum estudo realizado. A Figura 50 (a) e (b) mostra os valores dos coeficientes para os parâmetros dos modelos obtidos no decorrer das etapas da simulação. Os valores dos coeficientes significativos apresentam grandes variações na medida em que o valor do erro “sub-plot” é elevado. Acima da etapa quarenta da simulação (Figura

50) o coeficiente bc que é considerado significativo pelo modelo inicial da simulação poderá ser interpretado como não significativo, uma vez que seu valor se aproxima do zero. Por outro lado o parâmetro ad , cujo valor no modelo inicial da simulação vale zero, poderá assumir valores próximos aos dos parâmetros significativos. Estes fenômenos são frutos da utilização do baixo número de replicatas e de uma distribuição de variância de erro que não reflete precisamente a distribuição normal. Esta observação é interessante, pois reforça a necessidade de se realizar experimentos sempre evitando erros sistemáticos cuja distribuição não é normal. Por outro lado, a Figura 50 (e, f) apresenta coeficientes com baixa oscilação em torno de seus valores médios, mesmo na etapas finais em que o erro “sub-plot” se aproxima de 5,0, os coeficientes significativos e não significativos permanecem na mesma condição.

Todavia a simples observação dos valores dos coeficientes não é suficiente para a seleção dos efeitos significativos ou não. A análise correta para este caso é a utilização dos valores de t .

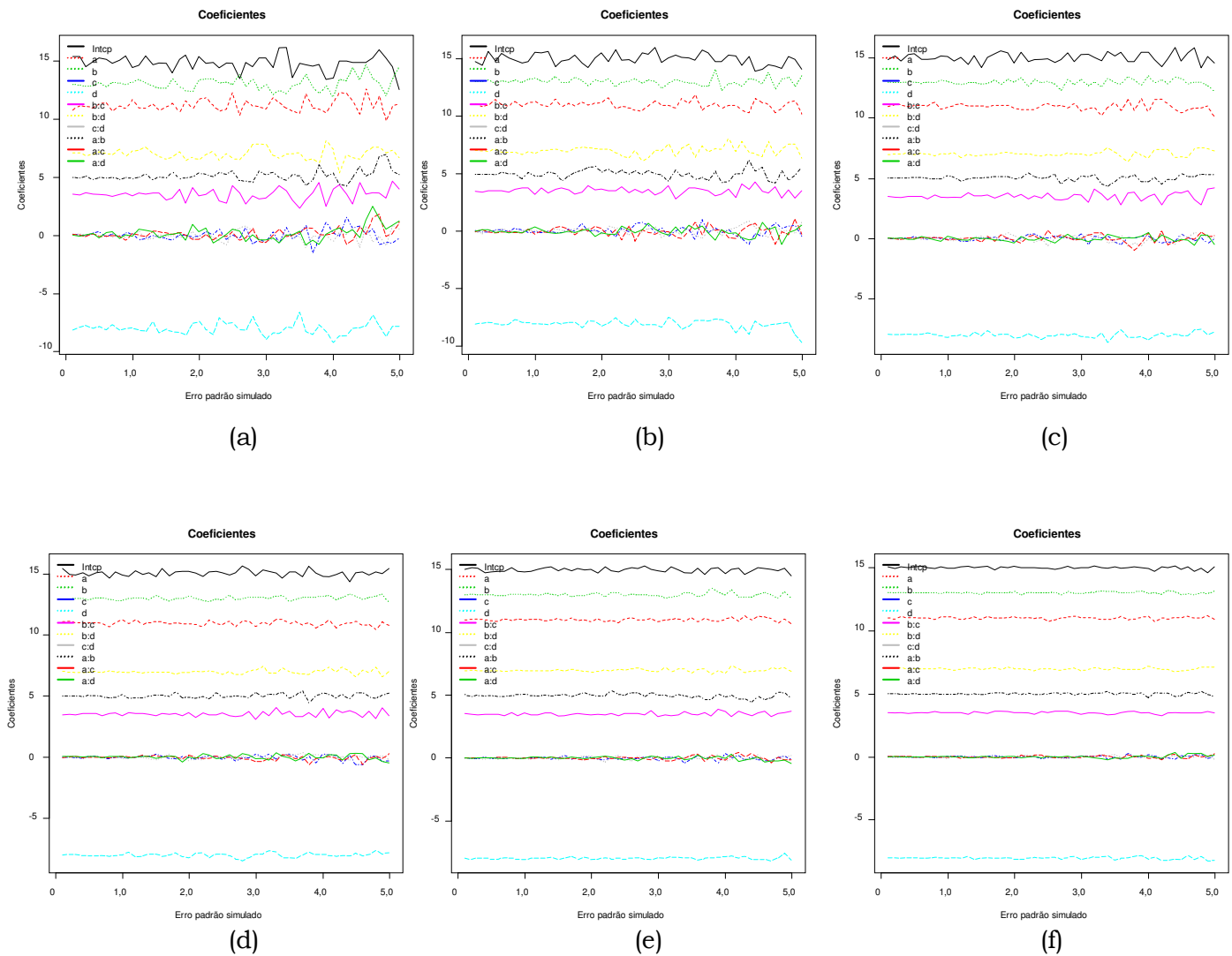


Figura 50. Coeficientes dos parâmetros dos modelos obtidos por regressão linear para simulações contendo 2, 4, 8, 16, 32 e 64 replicatas por experimento, representados pelas figuras a, b, c, d, e, f respectivamente. O erro padrão do “sub-plot” foi elevado de 0,5 a 5,0.

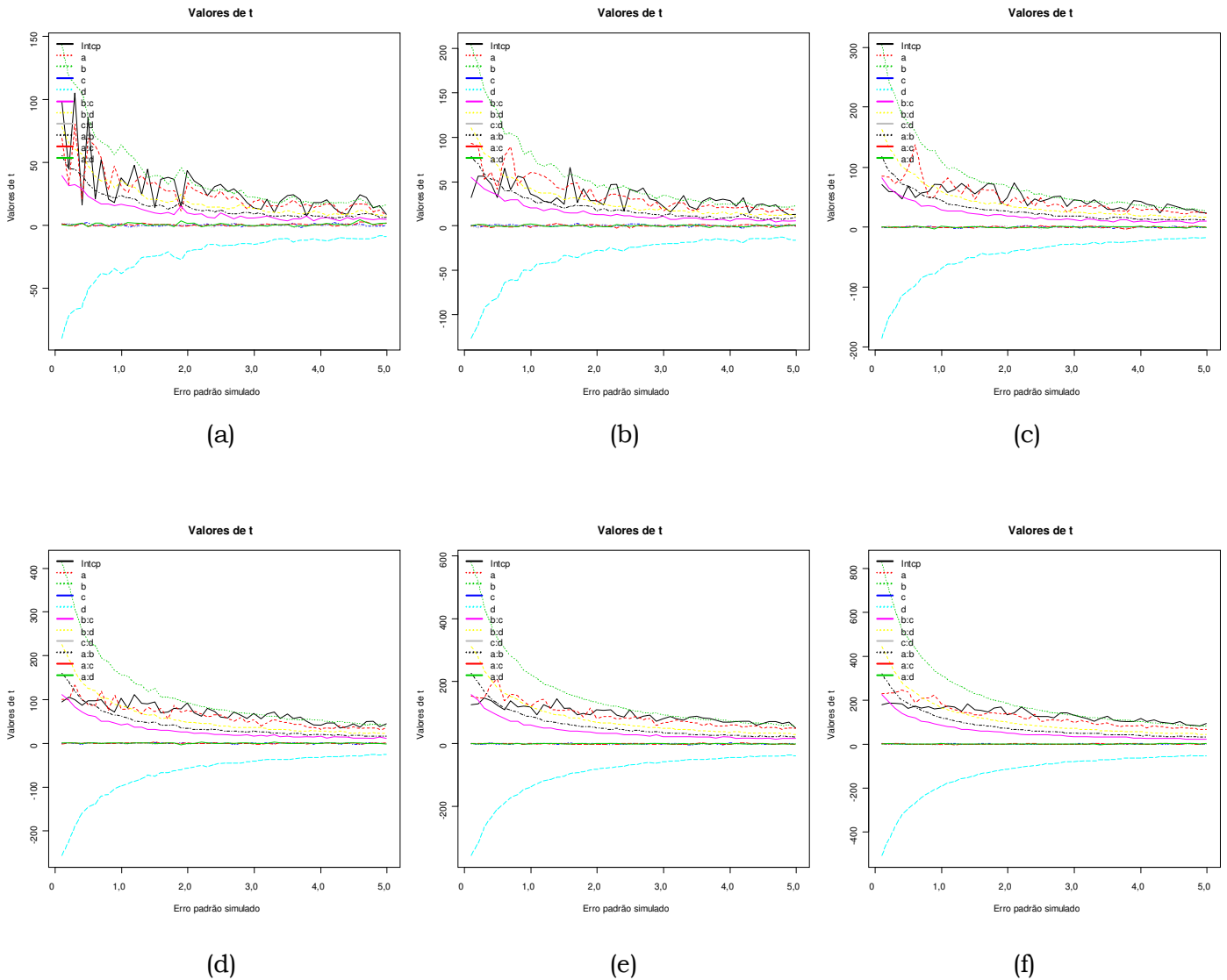


Figura 51. Valores de t para os coeficientes dos modelos obtidos por regressão linear para a simulação contendo 2, 4, 8, 16, 32 e 64 replicatas por experimento, que correspondem respectivamente a a, b, c, d, e, f na figura. O erro padrão do “main-plot” foi elevado de 0,5 a 5,0.

As simulações contendo poucas replicatas, 2 ou 4, mostram que os valores de t para o intercepto e o efeito “main-plot” apresentam grandes oscilações, como indicado na Figura 51 (a e b). Por outro lado, as simulações contendo um número maior de replicatas, 32 e 64, Figura 51 (e e f) respectivamente, fornece valores de

t cuja significância é coerente com o modelo inicial, Equação 51. Portanto, planejamentos “split-plot” que envolvam valores altos de erro “sub-plot” com número reduzido de replicatas podem apresentar complicações para a determinação de coeficientes significativos. Além disto, há também a dificuldade relacionada à determinação de graus de liberdade com precisão para o teste t. A ferramenta que apresentou melhores resultados para estes casos foram os gráficos de probabilidade acumulada.

Simulações para diferentes tratamentos de dados provenientes de planejamentos split-plot

Muitas vezes os químicos ou quaisquer outros experimentalistas blocam seus experimentos para facilitar o trabalho no laboratório, mas não empregam o tratamento adequado nos dados provenientes destes experimentos. O problema está no fato de tratar dados que sofreram restrições quanto a aleatorização completa como se os mesmos fossem autenticamente aleatórios. Os tratamentos incorretos podem levar à conclusões errôneas em muitos casos, assim um dos objetivos desta etapa das simulações é determinar em quais situações o tratamento incorreto pode ser mais prejudicial. Paralelamente a este fato os gráficos de probabilidade acumulada serão mais uma vez testados em novos casos.

O planejamento utilizado é equivalente ao descrito na Figura 8 e Tabela 8, contendo quatro variáveis, *a*, *b*, *c* e *d*, sendo que a variável *a* constitui o “main-plot” e as outras formam o “sub-plot”. Foram gerados dados em que um dos componentes de variância, replicata, “main-plot” ou “sub-plot”, teve seu erro padrão alterado 0,5 até 5,0, enquanto os outros dois componentes tiveram seus valores de variância mantidos com erro padrão igual a 2,0. Os dados gerados por simulação foram tratados por duas estratégias, a primeira, considera os dados

como se fossem obtidos de experimentos realizados de forma completamente aleatória. Na segunda análise os dados foram submetidos ao tratamento correto, “split-plot”. A avaliação dos modelos empregou o gráfico de probabilidade acumulada. O modelo ajustado foi

$$y = 15 + 11a + 13b + 0c - 8d + 3,5bc + 7bd + 0cd + 5ab + 0ac + 0ad$$

(Equação 51).

A Figura 52 mostra os gráficos de probabilidade acumulada para os valores da razão coeficientes/erros das simulações em que o erro das replicatas foi elevado de 0,5 a 5,0, enquanto os erros “main-plot” e “sub-plot” foram mantidos em 2,0. Há uma alteração de posições entre o valor de t para o intercepto, e para os parâmetros a e bc , devido a mudança de significância do intercepto ao longo das etapas da simulação. A Figura 53 apresenta um gráfico de probabilidade acumulada para os valores de t dos coeficientes dos parâmetros ao longo das 50 etapas da simulação. Neste caso a análise dos dados foi realizada como se os mesmos fossem provenientes de experimentos completamente aleatórios. Os valores dos parâmetros significativos b , bd , ab e d migram em direção ao zero na medida em que o erro das replicatas é elevado, conforme as setas indicadas na Figura 53.

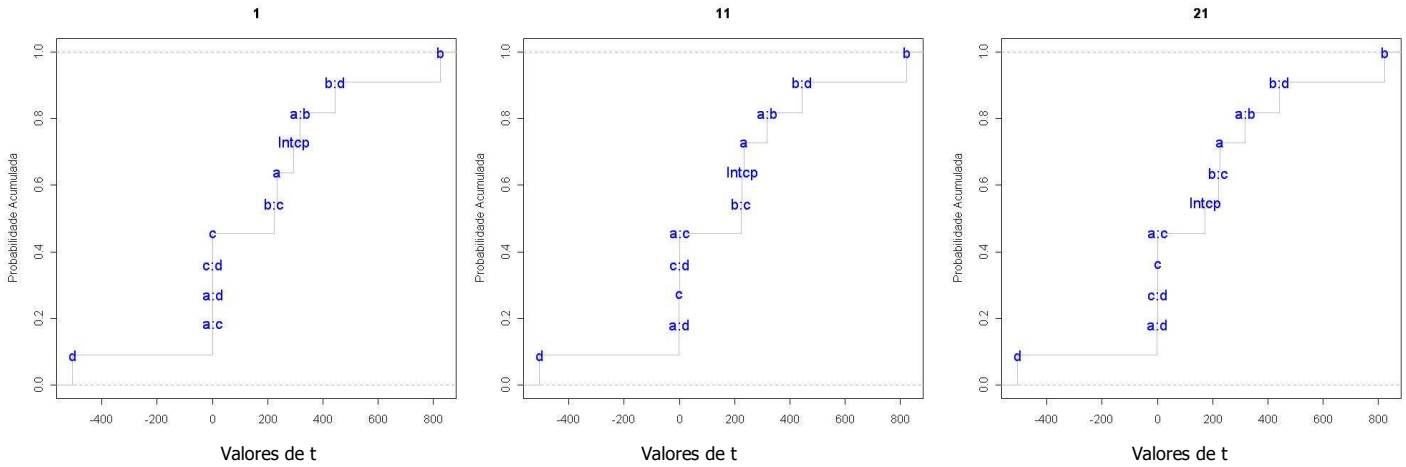


Figura 52. Gráficos de probabilidade acumulada para simulações com experimentos contendo 10000 replicatas. O erro padrão das replicatas foi elevado de 0,5 até 5,0 em 50 etapas. O erro padrão do “main-plot” e do “sub-plot” foi mantido em 2,0. As figuras apresentam no topo a etapa que indicam da simulação: 1, 11 e 21.

Os parâmetros a, bc e o intercepto foram excluídos da Figura 53 para não serem confundidos, pois mudam de posição ao longo das etapas de simulação. Os valores de t para os coeficientes c, ac, ad e cd, que são não significativos, permanecem centrados em zero sofrendo alterações de posição sem nenhuma significância.

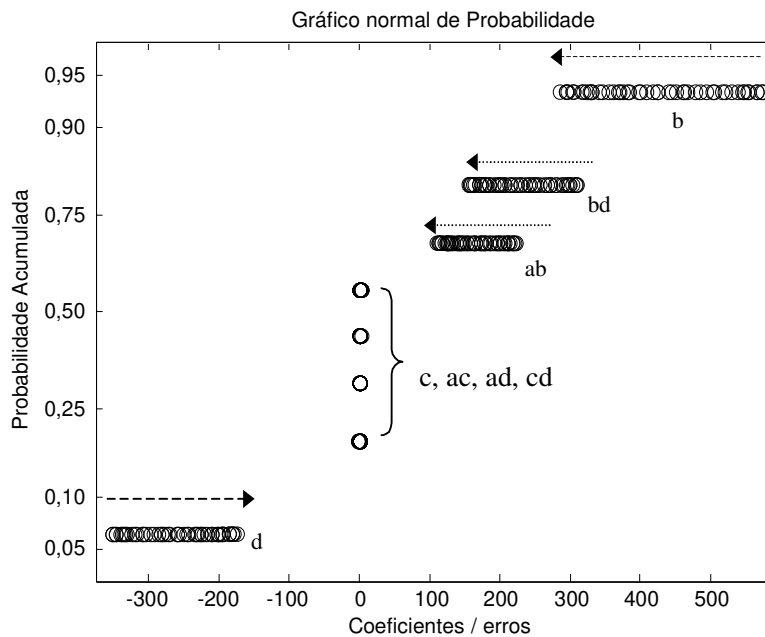


Figura 53. Valores de t para os coeficientes dos parâmetros dos modelos gerados ao longo do processo de simulação em que o erro padrão das replicatas foi elevado de 0,5 até 5,0. O erro padrão do “main-plot” e do “sub-plot” foi mantido em 2,0. As setas indicam a migração dos valores de t em direção ao zero na medida em que o erro das replicatas é elevado. Os dados foram tratados como provenientes de experimentos completamente aleatórios.

Contudo ao tratar os dados de forma correta, como provenientes de planejamentos “split-plot”, várias alterações são observadas. Conforme ilustrado na Figura 54 os valores de t para os coeficientes sofrem pouquíssimas alterações. Este comportamento é esperado, pois elevações no erro padrão das replicatas devem influenciar o valor de t para o intercepto. O valor de t para o intercepto foi excluído do gráfico por alterar sua posição com os parâmetros a e bc . Ou seja, a análise correta dos dados permitirá diferenciar os parâmetros significativos dos não significativos. Já a análise convencional dos dados, como provenientes de experimentos completamente aleatórios, poderá comprometer os resultados, uma vez que elevados erros de replicatas, nesta análise, comprometem incorretamente a significância dos parâmetros.

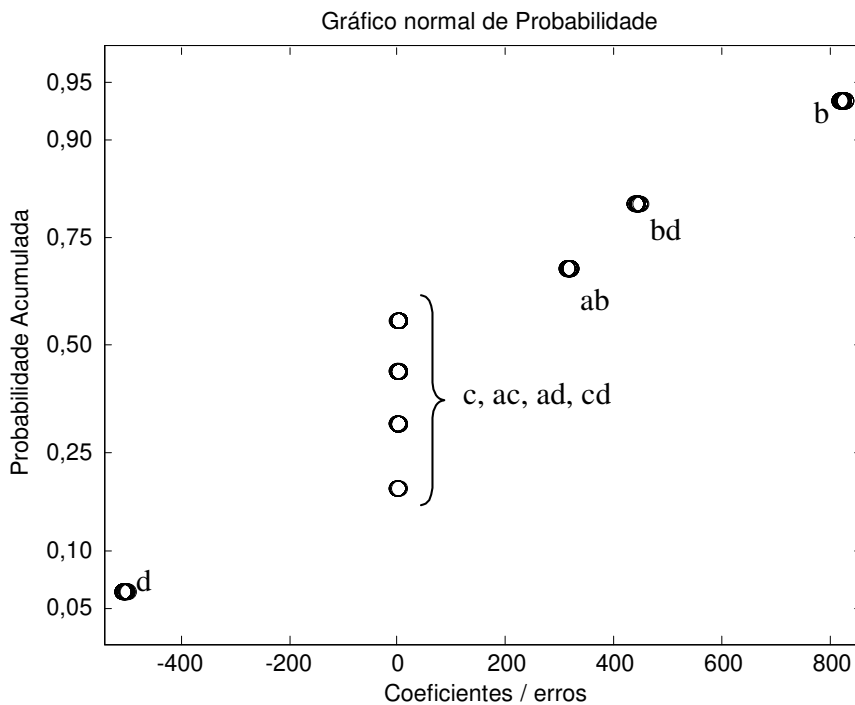


Figura 54. Valores de t para os coeficientes dos parâmetros dos modelos gerados ao longo do processo de simulação em que o erro padrão das replicatas foi elevado de 0,5 até 5,0. O erro padrão do “main-plot” e do “sub-plot” foi mantido em 2,0. Os dados foram tratados pela estratégia “split-plot”.

Simulações em que o erro “main-plot” foi gradualmente elevado também foram conduzidas para avaliar as conseqüências das duas formas de análise de dados. Enquanto o erro padrão do “main-plot” foi elevado de 0,5 até 5,0 o erro das replicatas e do “sub-plot” foi mantido em 2,0. A análise convencional, que considera dados de experimentos executados de forma completamente aleatória, indica uma grande perda de significância dos parâmetros b , bd e d , na medida em que o erro “main-plot” é elevado, observe a Figura 55. Neste caso os parâmetros a , ab , bc e o intercepto foram omitidos, pois há uma alternância de posições entre eles. As setas apresentadas na Figura 55 mostram como os valores de t migram em direção ao zero na medida em que o erro “main-plot” é elevado. Caso o erro da “main-plot” seja maior, os parâmetros apresentarão significância mais baixa, segundo esta análise.

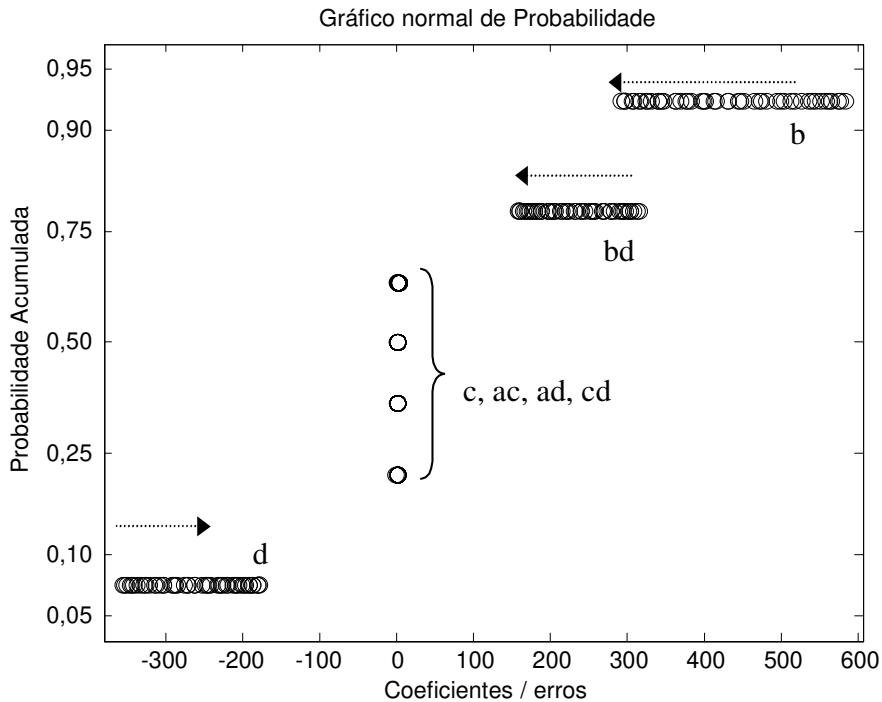


Figura 55. Valores de t para os coeficientes dos parâmetros dos modelos gerados ao longo do processo de simulação em que o erro padrão do “main-plot” foi elevado de 0,5 até 5,0. O erro padrão das replicatas e do “sub-plot” foi mantido em 2,0. As setas indicam a migração dos valores de t em direção ao zero na medida em que o erro do “main-plot” é elevado. Os dados foram tratados como provenientes de experimentos completamente aleatórios.

Por outro lado, ao realizar a análise correta para os dados fica evidente que os parâmetros indicados não perdem a significância, uma vez que o erro “main-plot” afeta apenas o intercepto e o efeito do “main-plot”, segundo o modelo estatístico adotado para o “split-plot”. Estas observações estão presentes na Figura 56. Portanto, mais uma vez a utilização de uma análise incorreta poderá claramente comprometer os resultados de um estudo realizado.

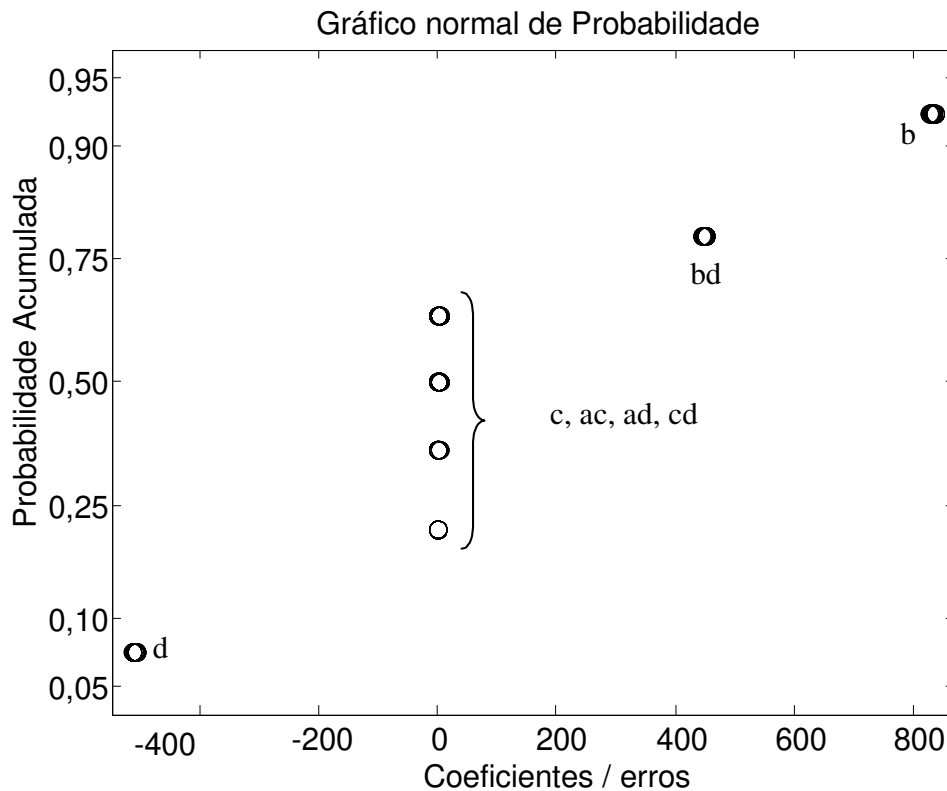


Figura 56. Valores de t para os coeficientes dos parâmetros dos modelos gerados ao longo do processo de simulação em que o erro padrão do “main-plot” foi elevado de 0,5 até 5,0. O erro padrão das replicatas e do “sub-plot” foi mantido em 2,0. Os dados foram tratados pela estratégia “split-plot”.

Por fim foram executadas simulações em que o erro padrão “sub-plot” foi elevado de 0,5 até 5,0 enquanto o erro padrão “main-plot” e das replicatas foi mantido em 2,0. A Figura 57 apresenta o gráfico normal de probabilidade para os valores de t dos parâmetros dos modelos gerados, os termos a , ab , bd , bc , e o intercepto foram omitidos, pois trocam de posições ao longo das etapas de simulação. Neste caso a análise realizada foi a convencional. O erro “sub-plot” influencia todos os parâmetros, uma vez que está presente em todas as expressões de erro padrão associado a efeitos. Assim, à medida que o erro “sub-plot” aumenta a significância dos parâmetros diminui, aproximando-se do zero. As setas indicadas na Figura 57 indicam a tendência dos valores de t em se aproximar do zero.

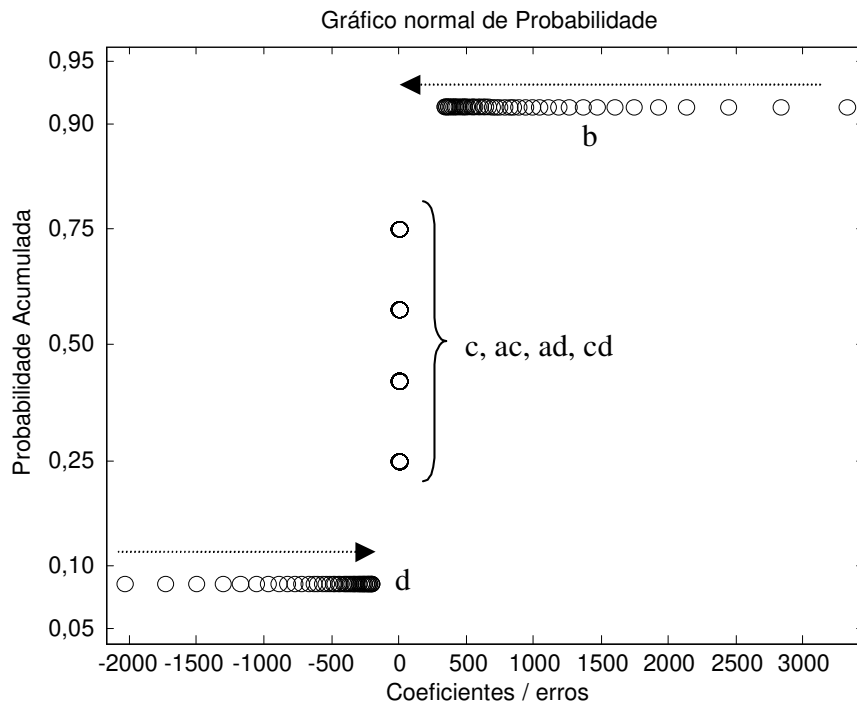


Figura 57. Valores de t para os coeficientes dos parâmetros dos modelos gerados ao longo do processo de simulação em que o erro padrão do “sub-plot” foi elevado de 0,5 até 5,0. O erro padrão das replicatas e do “main-plot” foi mantido em 2,0. As setas indicam a migração dos valores de t em direção ao zero na medida em que o erro das replicatas é elevado. Os dados foram tratados como provenientes de experimentos completamente aleatórios

Por outro lado, a análise “split-plot” dos dados revela a mesma tendência. Isto é coerente com o modelo, ou seja, neste caso em que o erro “sub-plot” aumenta há uma relação muito forte com o próprio erro aleatório envolvido na análise convencional. O erro “sub-plot” é aquele que na análise “split-plot” possui característica completamente aleatória, e sua influência está presente em todas as outras fontes de variância. A Figura 58 apresenta o gráfico de probabilidade acumulada com setas que mostram a migração dos valores de t dos parâmetros em direção ao zero com o aumento do erro “sub-plot”.

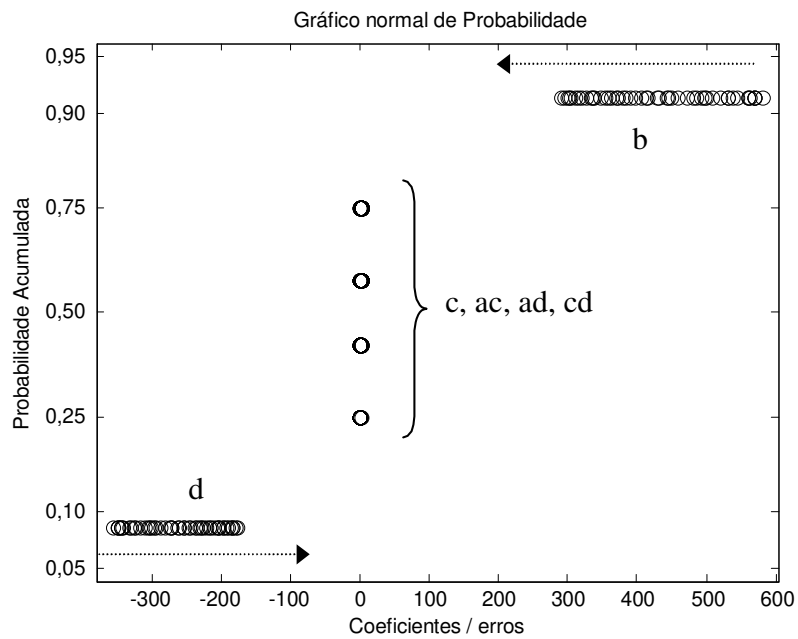


Figura 58. Valores de t para os coeficientes dos parâmetros dos modelos gerados ao longo do processo de simulação em que o erro padrão do “sub-plot” foi elevado de 0,5 até 5,0. O erro padrão das replicatas e do “main-plot” foi mantido em 2,0. As setas indicam a migração dos valores de t em direção ao zero na medida em que o erro do sub é elevado. Os dados foram tratados como provenientes de experimentos de um planejamento “split-plot”.

9 CONSIDERAÇÕES FINAIS

Com o desenvolvimento deste projeto de doutorado uma nova linha de pesquisa foi criada em nosso grupo de trabalho, pois agora há o embasamento teórico necessário, além de muitos pontos complexos terem sido tratados e compreendidos. Novos estudos e projetos estão sendo realizados baseados no método “split-plot”. Novas ferramentas computacionais para o tratamento dos dados foram criadas e estão disponíveis.

O aluno Cleber N. Borges aluno de doutorado e bolsista do CNPq está desenvolvendo um projeto no qual o método “split-plot” está sendo aplicado para modelar um sistema mistura-mistura em cromatografia.

A aluna Márcia C. Breitz, bolsista da FAPESP (processo 04/03934-5), está desenvolvendo um projeto de mestrado em que o método “split-plot” é empregado para modelar um sistema cromatográfico contendo variáveis de processo e mistura.

O método “split-plot” também foi empregado em estudos realizados pelo aluno Raimundo K. Vieira em um projeto de doutoramento para otimizar a determinação de chumbo. O planejamento “split-plot” foi aplicado em um estudo realizado em um projeto de pós-doutorado da aluna Aline Renée Coscione, bolsista da FAPESP, com o objetivo de modelar e otimizar um sistema de extração líquido-líquido por fase única.

10 CONCLUSÕES

Neste projeto de doutorado o método “split-plot” foi estudado detalhadamente. Devido à necessidade de se realizar uma grande quantidade de cálculos para a análise de variância, geração e validação dos modelos e análise gráfica foi desenvolvido um programa computacional. Com o software gerado deseja-se estimular o emprego deste método de otimização na comunidade científica do país. Vários estudos foram realizados para a melhor compreensão de como as diferentes fontes de erro afetam os dados e os modelos gerados por regressão. Os graus de liberdade, adequados para os casos de otimização conjunta de variáveis de processo e mistura, foram determinados. A aplicação de gráficos de probabilidade acumulada permitiu a seleção de efeitos significativos em planejamentos “split-plot”, uma vez que não há um método exato para a seleção dos parâmetros significativos dos modelos obtidos por regressão. Os gráficos de probabilidade acumulada também apresentaram ótimos resultados em estudos que visavam à redução do número de experimentos em planejamentos “split-plot”. Um método de simulação para o planejamento “split-plot” foi desenvolvido e aplicado em vários estudos, permitindo a compreensão das particularidades envolvidas.

11 ESTUDOS FUTUROS

O método “split-plot” poderá ser empregado em vários sistemas contendo variáveis de processo-processo, mistura-mistura, processo-mistura e mistura-processo. Os gráficos de probabilidade acumulada representam uma ferramenta com grande potencialidade para o estudo de problemas químicos. O domínio das técnicas de simulação para o planejamento “split-plot” também permite que vários estudos sejam realizados em situações específicas. Além de contribuir de forma muito significativa para a compreensão do método de forma pormenorizada.

REFERÊNCIAS BIBLIOGRÁFICAS

- ¹Barros Neto B., Scarminio I.S., Bruns R.E. Como Fazer Experimentos: Pesquisa e Desenvolvimento na Ciência e na Indústria, Editora da Unicamp, **2001**, Campinas.
- ²Box G.E.P, Hunter W.G., Hunter J.S. Statistics for experimenters An introduction to design, data analysis and model building, John Wiley & Sons, **1978**, New York.
- ³Anderson R.L., Bancroft T.A. Statistical Theory in Research, McGraw-Hill, **1952**, New York.
- ⁴Wooding W.M. *J. Qual. Tech.* **1973**, 5, 16.
- ⁵Cornell J.A. *J. Qual. Tech.* **1988**, 20, 2.
- ⁶Box G. E. P., Jones S. *J. Applied Stat.* **1992**, 19, 3.
- ⁷Cornell J.A., Gorman J.W. *J. Qual. Tech.* **1984**, 16,20.
- ⁸Cornell J.A. *Technometrics* **1975**, 17, 25.
- ⁹Cornell J.A. *Technometrics* **1977**, 19, 237.
- ¹⁰Cornell J.A. *J. Am. Stat. Assoc.* **1971**, 66, 42.
- ¹¹Cornell J.A. Experiments with Mistures: Design, Models and the analysis of Mixture data, John Wiley & Sons, **1990**, New York.
- ¹²Prescott P. *Qual. Tech. & Quant. Manag.* **2004**, 1, 87.
- ¹³Nobile A., Green P.J. *Biometrika* **2000**, 87, 15.
- ¹⁴Tack L., Vandebroek M. *J. Qual. Tech.* **2002**, 34, 422.
- ¹⁵Kulahci M., Ramirez J., Tobias R. *J. Qua. Tech.* **2006**, 38,56.
- ¹⁶Cornell J.A. *Technometrics* **1973**, 15, 437.
- ¹⁷Goos P., Vandebroek M. *J. Qual. Tech.* **2001**, 33,436.
- ¹⁸Martin T. N., Storck L., Lúcio A., Lorentz L. H. *Ciência Rural* **2005**, 35, 1257.
- ¹⁹Morais A. R., Nogueira M.C. S. *Sci. Agric.* **1995**, 52, 249.
- ²⁰Pinho S. Z., Mischan M.M. *Sci. agric.* **1996**, 53, 131.
- ²¹Morais A. R., Nogueira M. C. S. *Sci. agric.* **1996**, 53, 138.
- ²²Goos P., Vandebroek M. *J. Qual. Tech.* **2004**, 36, 12.
- ²³Raghavarao D., Xie Y. *J. Stat. Plan.* **2003**, 116, 197.
- ²⁴Mukerjee R., Fang K. *Stat. Sin.* **2002**, 12, 885.
- ²⁵Kowalski, S. M. *J. Qual. Tech.* **2002**, 34, 399.
- ²⁶Potcner K.J., Kowalski S.M. *Qual. Prog.* **2004**, 37,67.
- ²⁷Montgomery D.C. Design and Analysis of Experiments, John Wiley & Sons, **1976**, New York.
- ²⁸Snedecor G.W., Cochran W.G, Statistical Methods, The Iowa State University Press, **1980**, Ames.
- ²⁹Steel R. G. D., Torrie J. H. Principles and Procedures of Statistics With Special Reference to the Biological Sciences, Mc Graw-Hill, **1960**, New York.
- ³⁰Vuolo J.H. Fundamentos da Teoria de Erros, Edgard Blücher LTDA, **1992**, São Paulo.
- ³¹Searle S.R. Linear Models for Unbalanced Data, John Wiley & Sons, **1987**, New York.
- ³²Searle, S. R. Linear Models, John Wiley & Sons, **1971**, New York.
- ³³Harville D.A. *J. Am. Stat. Assoc.* **1977**, 72, 320.
- ³⁴Patterson H.D. Thompson R. *Biometrika* **1971**, 58, 545.

-
- ³⁵Harvey A.C . The Econometric Analysis of Time Series, Philip Allan, **1990**, Londres.
- ³⁶Reis C., Andrade J.C., Bruns R.E., Moran R.C.C.P. *Anal. Chim. Acta* **1998**, 369,269.
- ³⁷Bortoloti J.A., Andrade J.C., Bruns R.E. *J. Braz. Chem. Soc.* **2004**, 15, 241.
- ³⁸Coscione A.R., Modelamento e otimização de um sistema de extração líquido-líquido por fase única usando o planejamento split-plot, Relatório Fapesp, Instituto de Química, Unicamp, Campinas, Dezembro de 2003.
- ³⁹Bortoloti J.A., Bruns R.E., Andrade J.C., Vieira R.K. *Chemom. Intell. Lab. Syst.* **2004**, 70, 113.
- ⁴⁰Bortoloti J.A., Borges C.N., Bruns R.E. *Anal. Chim. Acta* **2005**, 544, 206.
- ⁴¹Kowalski S.M., Potcner K.J. *Qual. Prog.* **2003**, 36, 60.
- ⁴²Piepel G.F., Cornell J.A. *J. Qual. Tech.* **1994**, 26, 177.
- ⁴³Satterthwaite F.E. *Psychometrika* **1941**, 6, 309.