5
Referências bibliográficas

BINGHAM, D. R.; SITTER, R. R. **Design Issues in fractional factorial split-plot experiments.** Journal of Quality Technology; Jan 2001; v.33, 1, ABI/INFORM Global pag. 39.

BISGAARD, S. **The Design and Analysis of 2^{k-p} x 2^{q-r} Split Plot Experiments.** Journal of Quality Technology; Jan 2000; 32, 1, ABI/INFORM Global pag. 39.

BISGAARD, S.; FULLER, H. T.; BARRIOS, E. **Quality quandaries: two-level factorials run as split plot experiments.** Quality Engineering; 1996; v.8, p. 705-708.

BOX, G. E. P. **Must we randomize our experiment**? Quality Engineering; 1990; v.2, p. 497-502.

CALADO, V.; MONTGOMERY, D. C. **Planejamento de experimentos usando *Statistica*.** E-papers Serviços Editoriais: Rio de Janeiro, 2003.

GALDÁMEZ, Edwin Vladimir Cardoza. **Aplicação das técnicas de planejamento e análise de experimentos na melhoria da qualidade de um processo de fabricação de produtos plásticos,** 2002**.** Disponível em: <http://www.teses.usp.br/teses >. Acesso em 28 out 2007.

GANJU, J.; LUCAS, J. M. **Detecting randomization restrictions caused by factors.** Journal of Statistical Planning and Inference; 1999; v.81, p. 129-140.

GANJU, J; LUCAS, J. M. **Randomized and random run order experiments** Journal of Statistical Planning and Inference; 2004; v.133, pag. 199-210.

GIL, Antonio C. **Como elaborar projetos de pesquisa**. Atlas: São Paulo, 1991.

GOMES, U. R.. **Otimização do processo de laminação a frio através de planejamentos de experimentos.** Dissertação (Mestrado em Engenharia Industrial) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2007.

JU, H. L.; LUCAS, J. M. L^k **Factorial experiments with hard-to-change and easy-to-change factors.** Journal of Quality Technology; Oct 2002; vol. 34, no 4, ABI/INFORM Global pag. 411.

LETSINGER, Jennifer D.; MYERS, Raymond H.; LENTNER, Marvin. **Response surface methods for bi-randomization structures.** Journal of Quality Technology; Oct 1996; vol. 28, no 4, ABI/INFORM Global pag. 381.

PUC-Rio - Certificação Digital Nº 0611755/CA

LOEPPKY, J.L.; SITTER, R. R. **Analyzing unreplicated blocked or split-plot fractional factorial designs.** Journal of Quality Technology; Jul 2002; v.34, pag. 229.

MARTINS, G. A. **Manual para elaboração de monografias e dissertações.** Atlas: São Paulo, 1994.

MONTGOMERY, Douglas C. **Design and Analysis of Experiments**. Wiley: New York, 2001.

MONTGOMERY, Douglas C.; RUNGER, George C. **Estatística Aplicada e Probabilidade para Engenheiros**. LTC: Rio de Janeiro, 2003.

MONTGOMERY, Douglas C.; MYERS, Raymond H. **Response Surface Methodology: Process and Product Optimization Using Designed Experiments**. Wiley-Interscience: 2002.

SKF/CSN. **Relatório de estudo sobre chatter nos laminadores LTF 1 e LTF 2.** SKF do Brasil Ltda: Rio de Janeiro, 2007.

VERGARA, Sylvia C. **Projetos e Relatórios de Pesquisa em Administração**. São Paulo: Atlas, 1997.

VIEIRA, Antonio F. C. **Análise da média e dispersão em experimentos fatoriais não replicados para otimização de processos industriais**. Tese (Doutorado em Engenharia Industrial) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2004.

VINING, G. Geoffrey; KOWALSKI, Scott M.; MONTGOMERY, Douglas C. **Response Surface Designs within a Spli-Plot structure.** Journal of Quality Technology; Apr 2005; v.37, p. 115.

WEBB, Derek F.; LUCAS, James M.; BORKOWSKI, John J. **Factorial experiments when factor levels are not necessarily reset.** Journal of Quality Technology; Jan 2004; vol. 36, pag. 1.

PUC-Rio - Certificação Digital Nº 0611755/CA

Apêndices

APÊNDICE A – **Etapas de execução de um projeto *split-plot* para o exemplo de força de tensão do papel.**

Elaborando o projeto

Utilizando um projeto fatorial geral, por se tratar de fatores com mais de dois níveis, conforme a Figura A.1, introduz-se o número de fatores a serem analisados no experimento. Sabe-se que o experimento apresenta dois fatores, entretanto, por sugestão do tutorial e para auxiliar na análise considera-se a replicação como um fator.

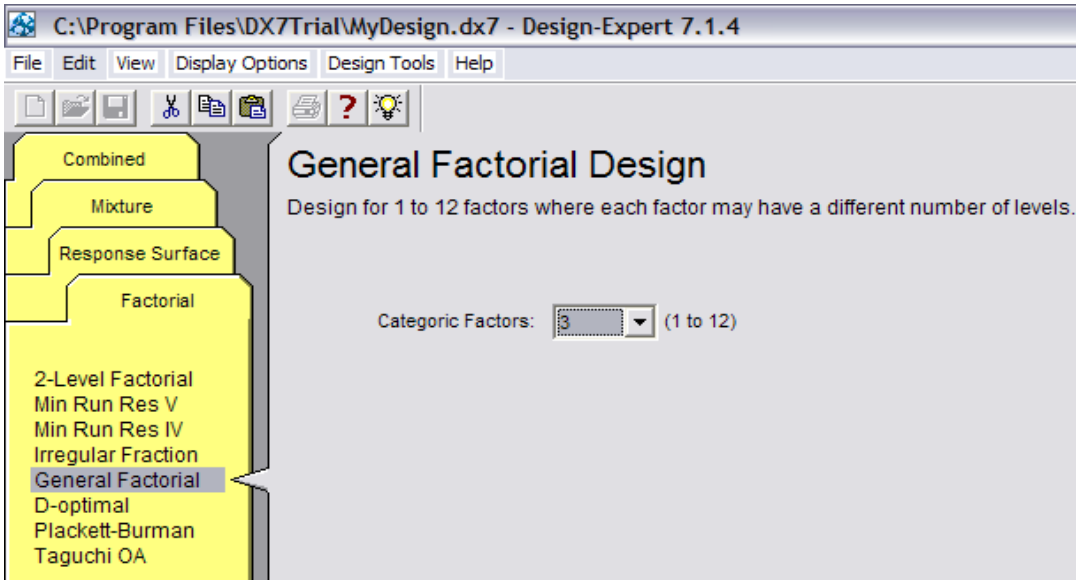


Figura A.1 – Seleção do tipo de projeto para a execução do experimento *split-plot*.

Dando continuidade ao projeto, na próxima tela introduz-se **réplica** para o nome do fator, com **3** níveis, denominados **dia 1**, **dia 2** e **dia 3**, conforme apresentado na Figura A.2.

PUC-Rio - Certificação Digital Nº 0611755/CA

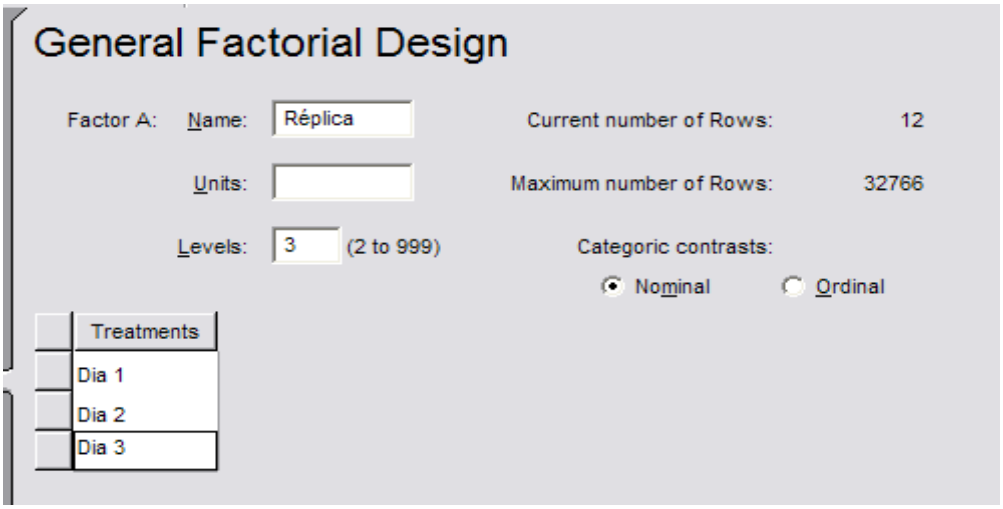
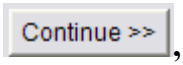


Figura A.2 – Inserção do nome e das características do fator A (Bloco).

Clicando novamente no botão ***Continue***, no canto direito inferior da tela 

, a próxima visualização refere-se ao fator *B*. Então se insere o nome do fator **Preparação da polpa (Prep polpa)**, com os **3** níveis **B1**, **B2** e **B3**, como na Figura A.3.

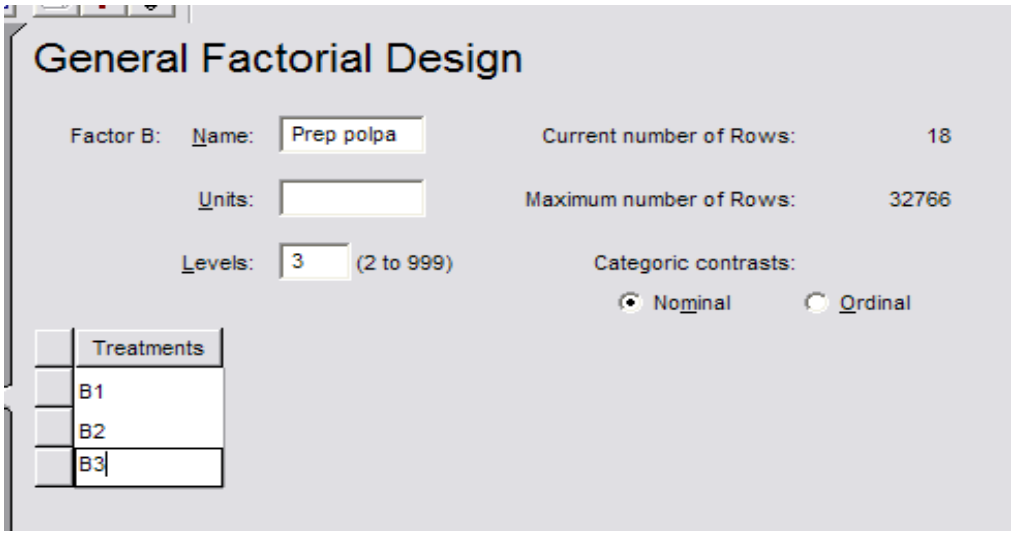


Figura A.3 – Inserção do nome e das características do fator B (whole plot).

O próximo passo é inserir o terceiro fator no projeto e nomeá-lo como **Temperatura (Temp),** com **graus Fahrenheit (graus F)** como unidades, **4** níveis, especificados como **200**, **225**, **250** e **275** para os tratamentos (Figura A.4). Para reconhecer a natureza numérica desse fator, deve-se mudar a categoria dos contrastes para **ordinal**.

PUC-Rio - Certificação Digital Nº 0611755/CA

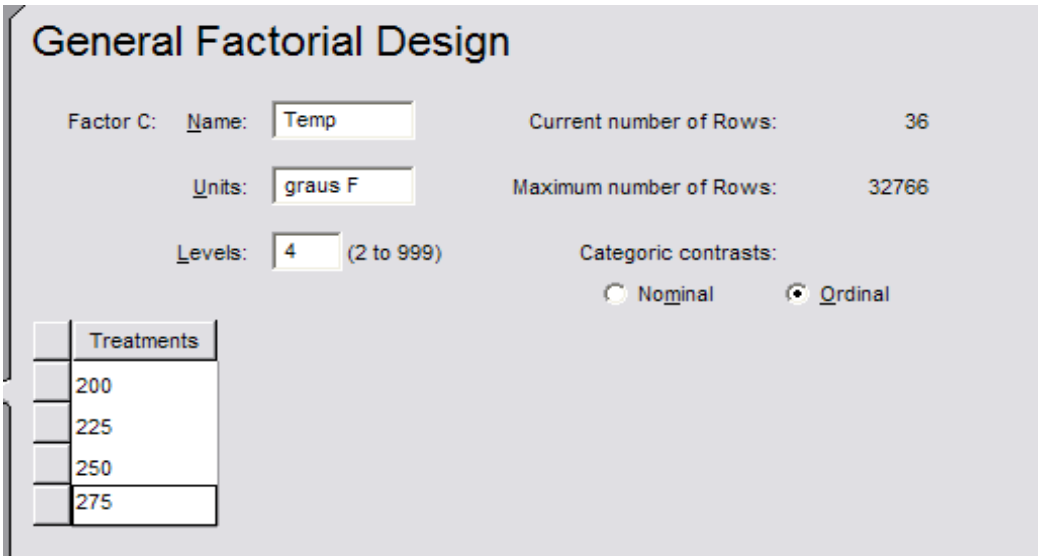


Figura A.4 – Inserção do nome e das características do fator C (Subplot).

A tela seguinte solicita o número de replicações para o experimento. Como, para auxiliar na análise, réplica foi considerada como um fator, sugere-se utilizar o valor "1" para produzir 36 corridas (3x3x4 níveis dos três fatores) e prosseguir. Nota-se, no próximo passo, a solicitação do número de respostas para o experimento. Apesar de se medir apenas uma resposta no experimento, para efeito de análise, esta resposta será particionada em **4** opções com as respectivas unidades referentes ao problema, conforme a Figura A.5. As demais informações solicitadas na tela não são requisitos para a execução do experimento.

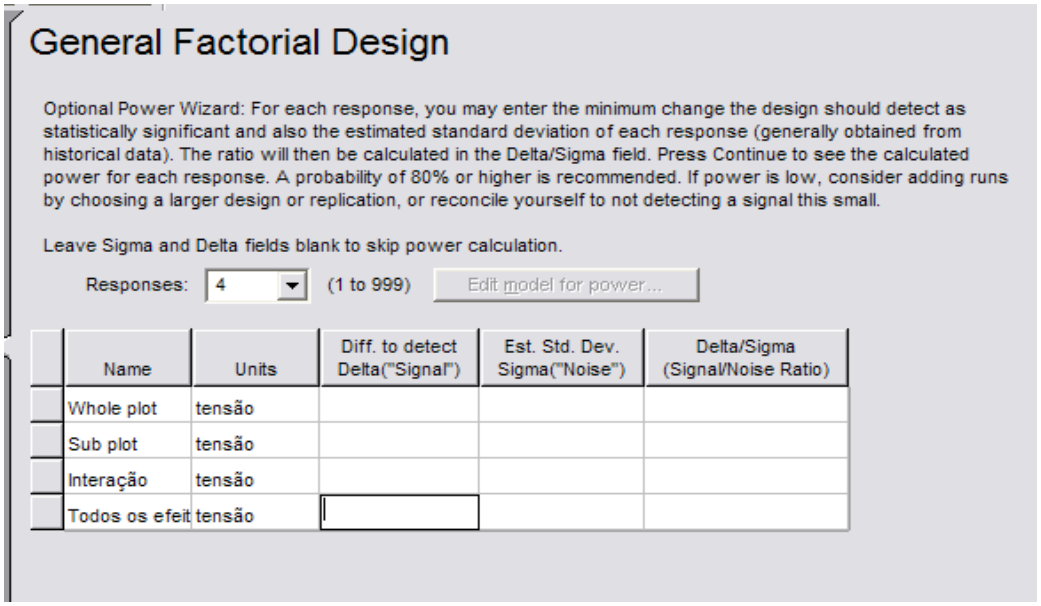


Figura A.5 – Definição das quatro respostas a serem analisadas.

PUC-Rio - Certificação Digital Nº 0611755/CA

A tela do projeto se apresenta agora em ordem completamente aleatória. Como não é o caso, deve-se configurar a ordem de execução das corridas conforme a alocação das observações em relação aos fatores. As três figuras a seguir ilustram como deve ser o arranjo no *software* para que o projeto reflita a realidade. Primeiramente, ordena-se as corridas com base no fator "preparação da polpa" (Figura A.6). Em seguida, faz-se o mesmo procedimento, só que desta vez ordenando as corridas com base no fator "réplica" (Figura A.7). Por fim, na coluna "*run*", atribui-se a reordenação das corridas com base na atual configuração (Figura A.8).

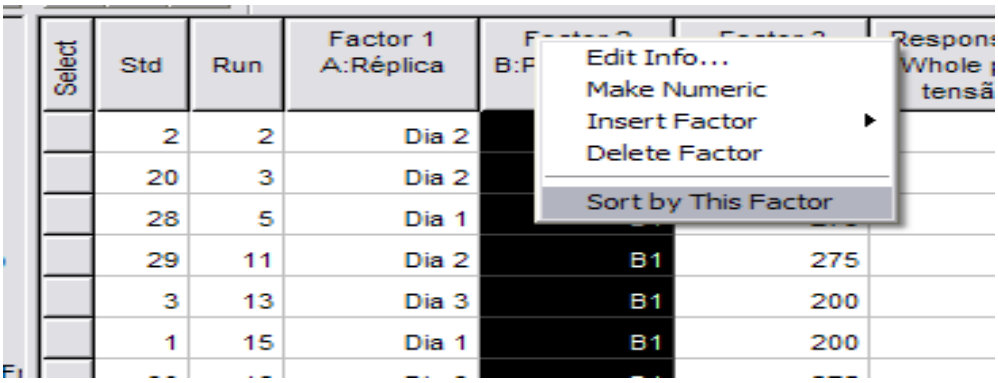


Figura A.6 – Re-organização dos dados com base no fator B.

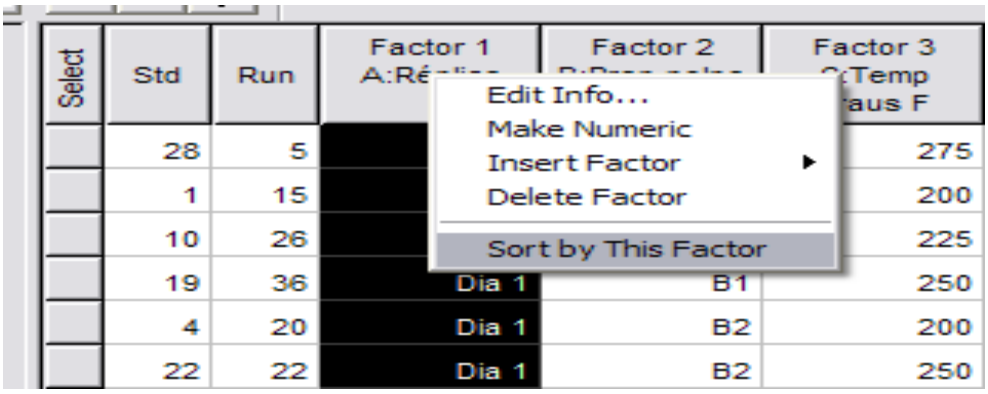


Figura A.7 – Re-organização dos dados com base nos blocos.

PUC-Rio - Certificação Digital Nº 0611755/CA

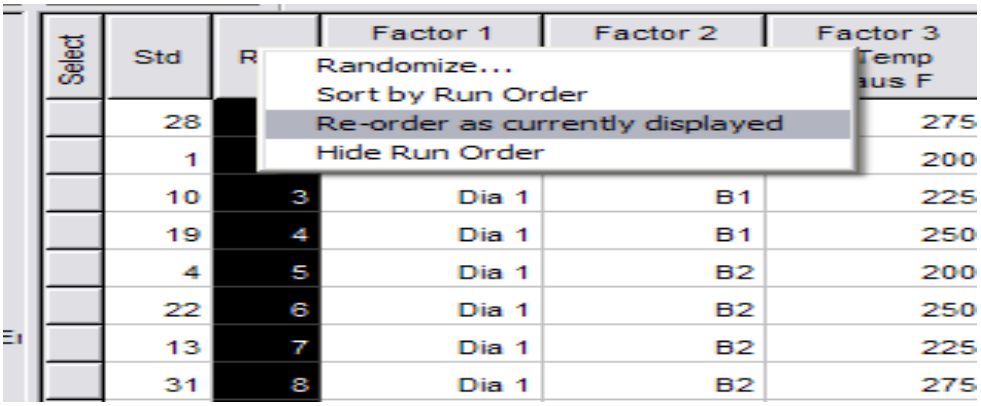


Figura A.8 – Re-organização da ordem das corridas experimentais.

Analisando os dados

Para que seja possível analisar um projeto *split-plot* no software *Design Expert* 7.1.4, é necessário criar ANOVA's separadas, e isto é feito manualmente. Cria-se uma ANOVA para o tratamento *whole plot* (preparação da polpa), outra para o tratamento *subplot* (temperatura) e, finalmente a interação *whole plot* e *subplot*, para que individualmente cada uma seja testada corretamente. Ainda ajusta-se um modelo completo para ter o diagnóstico de significância e os gráficos do modelo, mas ignora-se a ANOVA do modelo completo.

Select	Std	Run	Factor 1 A:Réplica	Factor 2 B:Prep polpa	Factor 3 C:Temp graus F	Response 1 Whole plot tensão	Response 2 Sub plot tensão	Response 3 Interação tensão	Response 4 Todos os efeit tensão
	28	1	Dia 1	B1	275	36	36	36	36
	1	2	Dia 1	B1	200	30	30	30	30
	10	3	Dia 1	B1	225	35	35	35	35
	19	4	Dia 1	B1	250	37	37	37	37
	4	5	Dia 1	B2	200	34	34	34	34
	22	6	Dia 1	B2	250	38	38	38	38
	13	7	Dia 1	B2	225	41	41	41	41
	31	8	Dia 1	B2	275	42	42	42	42
	34	9	Dia 1	B3	275	36	36	36	36
	16	10	Dia 1	B3	225	26	26	26	26
	7	11	Dia 1	B3	200	29	29	29	29
	25	12	Dia 1	B3	250	33	33	33	33
	2	13	Dia 2	B1	200	28	28	28	28
	20	14	Dia 2	B1	250	40	40	40	40
	29	15	Dia 2	B1	275	41	41	41	41
	11	16	Dia 2	B1	225	32	32	32	32
	32	17	Dia 2	B2	275	40	40	40	40
	23	18	Dia 2	B2	250	42	42	42	42
	14	19	Dia 2	B2	225	36	36	36	36
	5	20	Dia 2	B2	200	31	31	31	31
	35	21	Dia 2	B3	275	40	40	40	40
	17	22	Dia 2	B3	225	30	30	30	30
	8	23	Dia 2	B3	200	31	31	31	31
	26	24	Dia 2	B3	250	32	32	32	32
	3	25	Dia 3	B1	200	31	31	31	31
	30	26	Dia 3	B1	275	40	40	40	40
	21	27	Dia 3	B1	250	41	41	41	41
	12	28	Dia 3	B1	225	37	37	37	37

Figura A.9 – Arranjo do projeto *spli-plot*.

PUC-Rio - Certificação Digital Nº 0611755/CA

Para se analisar inicialmente o efeito *whole plot* (*B*), deve-se clicar no rótulo referente ao nome no diretório ***Analysis*** na janela à esquerda. Logo em seguida, deve-se clicar no botão ***Effects***, com o qual aparecerá a lista de efeitos do experimento. O *software* oferece quatro alternativas para designar os efeitos, dentre as quais: ***Model ("M"); Block ("b"); Error ("e"); Ignore ("X")***. O tratamento *whole plot* deverá ser testado em relação a **Réplica** através da interação *AB*. Dessa forma, será selecionado para o modelo o fator *B*; como bloco, a Réplica; como erro, a interação *AB*; e serão ignorados do modelo os demais efeitos (Figura A.10).

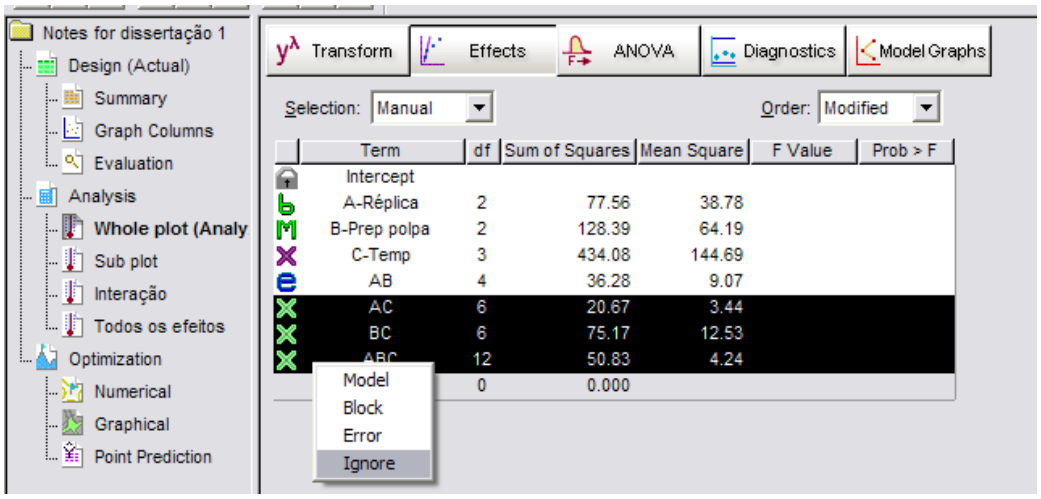


Figura A.10 – Lista de efeitos para o projeto *whole-plot*.

A seguir, seleciona-se o botão da ANOVA para analisar a significância dos tratamentos. Com base na Figura A.11, nota-se que o efeito *whole plot* (preparação da polpa) é significativo, uma vez que o *p-value* < 0,05. Aconselha-se, neste ponto, a não analisar o diagnóstico (*diagnostics*) e os gráficos do modelo, visto que os modelos ainda estão incompletos nesta etapa.

PUC-Rio - Certificação Digital Nº 0611755/CA

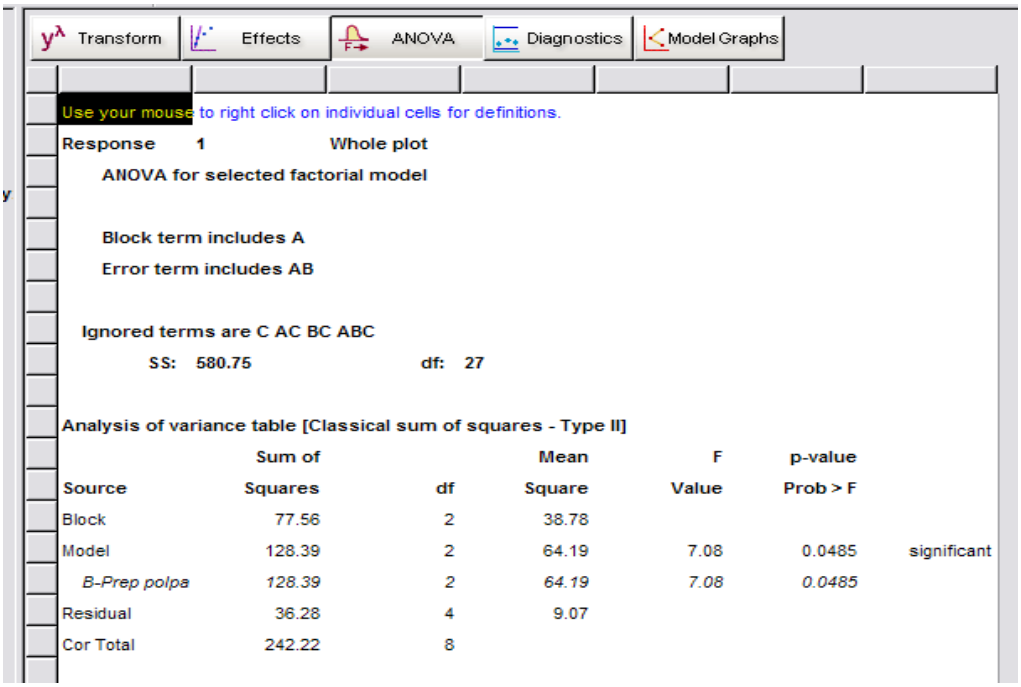


Figura A.11 – ANOVA para o projeto *whole plot*.

Para analisar o tratamento *subplot* (temperatura), seleciona-se a guia *subplot* no diretório de ***Analysis*** e depois, clica-se no botão ***Effects***. O tratamento *subplot* deve ser testado em relação à interação Réplica x Temperatura, isto é, a interação *AC*. Desse modo, a lista de efeitos deve estar configurada conforme a Figura A.12 (os fatores *A* e *B* definidos como bloco; o fator *C* selecionado para o modelo; a interação *AC* listada como erro; e as interações *AB*, *BC* e *ABC* ignoradas). Prosseguindo com a visualização da ANOVA (Figura A.13), é possível notar que o fator *subplot* (Temperatura) é significativo (*p-value* < 0,05). Novamente não se deve olhar o diagnóstico (*diagnostics*) e os gráficos do modelo, visto que, nesta etapa, os modelos ainda estão incompletos.

PUC-Rio - Certificação Digital Nº 0611755/CA

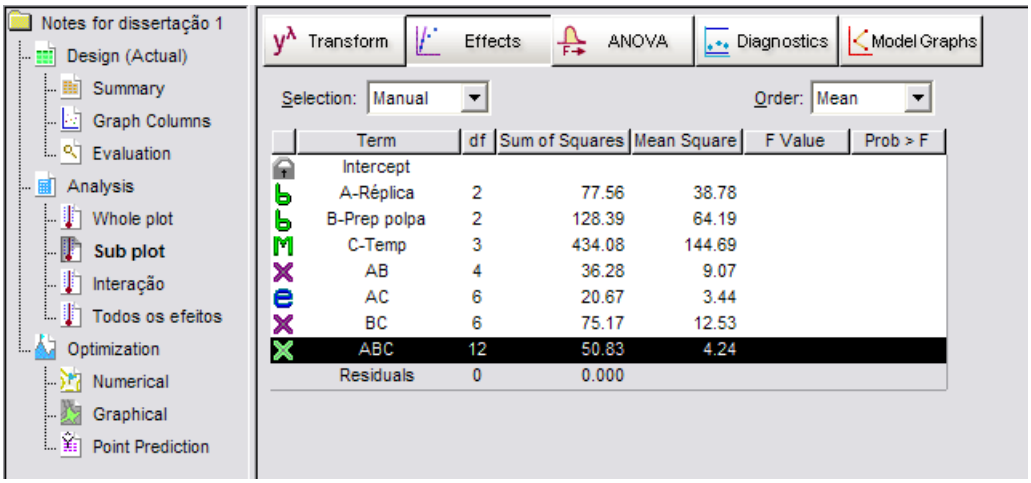


Figura A.12 – Lista de efeitos para o projeto *subplot*.

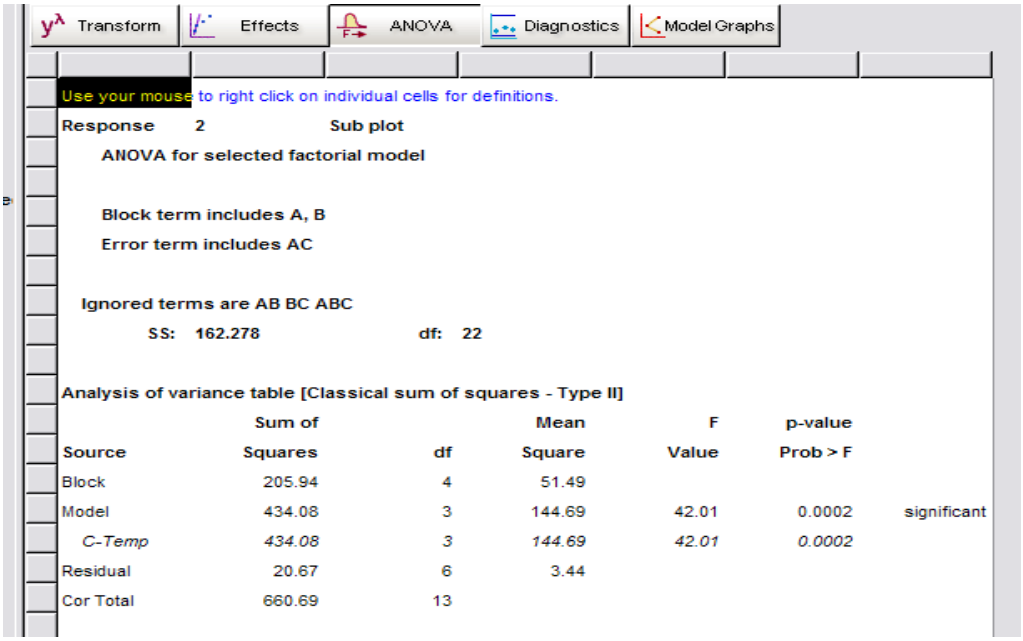


Figura A.13 – ANOVA para o projeto *subplot*.

No caso da interação *whole plot* x *subplot*, clica-se na opção interação no diretório de ***Analysis***. Semelhante aos demais, no botão ***Effects***, a interação *BC* deve ser testada em relação à interação *ABC*, logo esta é selecionada como erro. A réplica e os fatores *B* e *C* são considerados blocos. A interação *BC* é selecionada para o modelo, e os demais são ignorados, como mostrado na Figura A.14. Pela ANOVA (Figura A.15), observa-se que a interação *BC* é considerada não significativa pelo programa, visto que *p-value* > 0,05 em um valor limítrofe.

PUC-Rio - Certificação Digital Nº 0611755/CA

Dessa forma, pela proximidade do valor cabe ao experimentalista a decisão de ignorar a recomendação do software e considerar *BC* significativo.

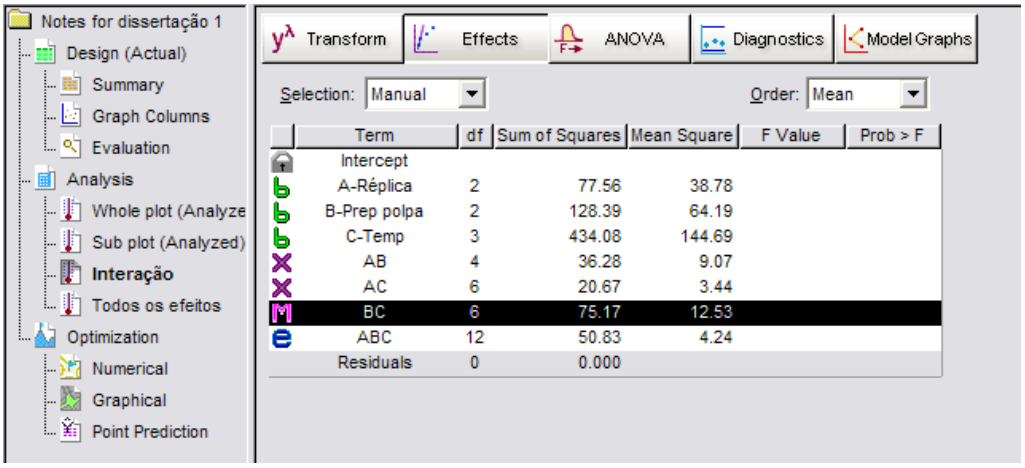


Figura A.14 – Lista de efeitos para o projeto de interação *whole plot x subplot.*

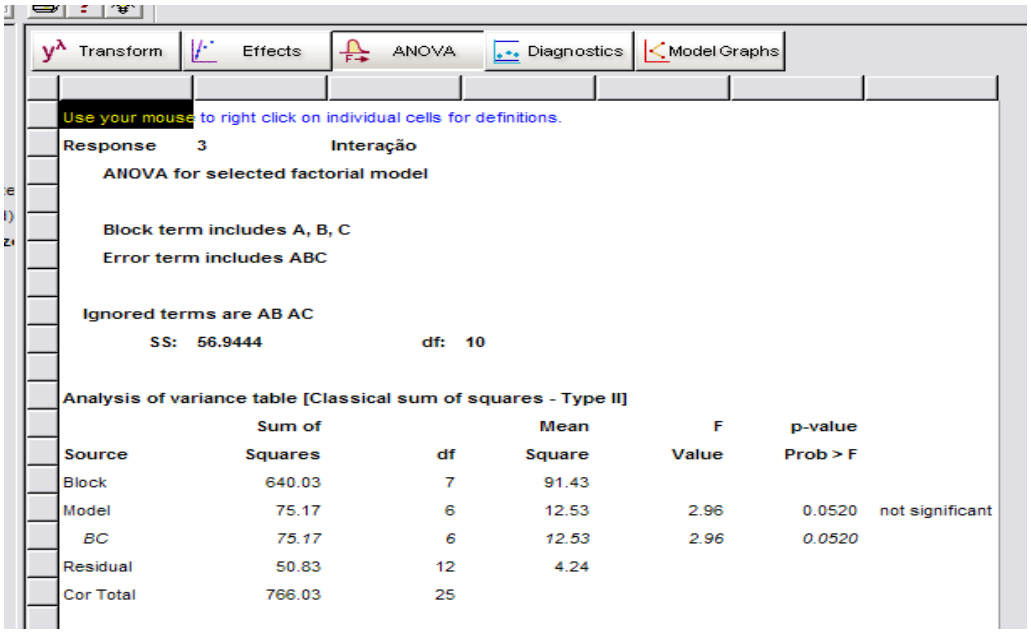


Figura A.15 – ANOVA para a interação *whole-plot x sub-plot.*

Antes que se avance para a fase final do projeto, devem ser levados em consideração alguns aspectos importantes, que justificam a análise estatística feita até o momento:

- Em um experimento blocado as interações bloco *versus* tratamento são usadas para estimar o erro.

PUC-Rio - Certificação Digital Nº 0611755/CA

- Em um projeto blocado completamente aleatorizado, todas as interações bloco *versus* tratamento são combinadas em uma única estimativa do erro.
- Em um projeto *split-plot*, devido às restrições quanto à aleatorização, determinadas interações bloco *versus* tratamento devem ser designadas para estimar o erro de um tratamento particular que esteja sendo testado.

Com base nestas justificativas, as análises de variância executadas até o momento neste projeto experimental foram: o tratamento *whole plot* (preparação da polpa) em relação à interação *AB* (bloco *versus* preparação da polpa); o tratamento *subplot* (temperatura) em relação à interação *AC* (bloco *versus* temperatura); e, por fim, a interação *BC* (polpa *versus* temperatura) em relação à interação *ABC* (bloco *versus* polpa *versus* temperatura).

Para que seja possível obter diagnósticos significativos e a análise gráfica do modelo, deve-se ajustar o modelo completo (Figura A.16), selecionando-se os fatores *B*, *C* e a interação *BC* para o modelo; *A* como bloco; e os demais fatores como erro. Por conseguinte, faz-se a análise do diagnóstico e dos gráficos do modelo, desconsiderando-se a ANOVA.

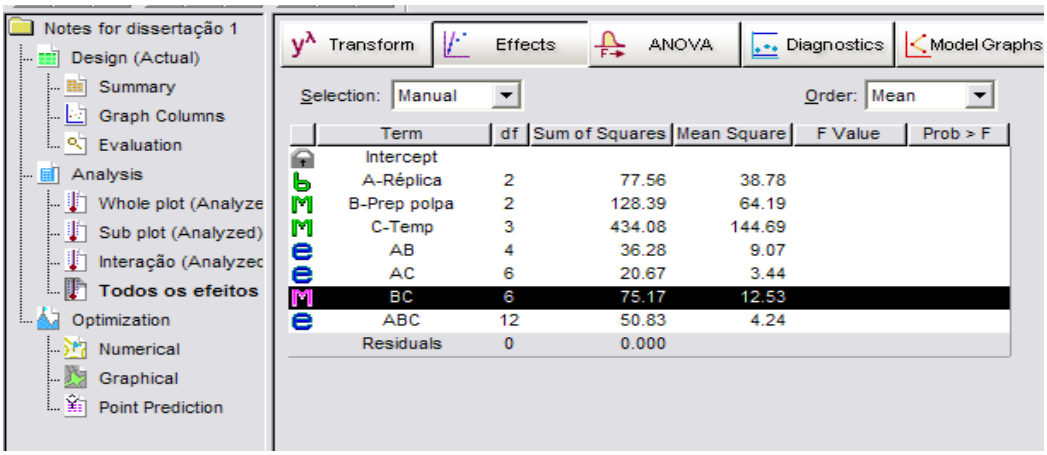


Figura A.16 – Lista de efeitos para o projeto completo.

Diagnóstico e influência dos resíduos

- Gráfico da Normal x Resíduos: este gráfico tem a finalidade de identificar alguma alteração no que tange a normalidade e a presença de observações

PUC-Rio - Certificação Digital Nº 0611755/CA

atípicas. Pode-se analisar, para o experimento em questão, segundo a Figura A.17, que não há indícios de observações atípicas e de perda de normalidade uma vez que os valores dos resíduos estão uniformemente dispostos sobre a reta.

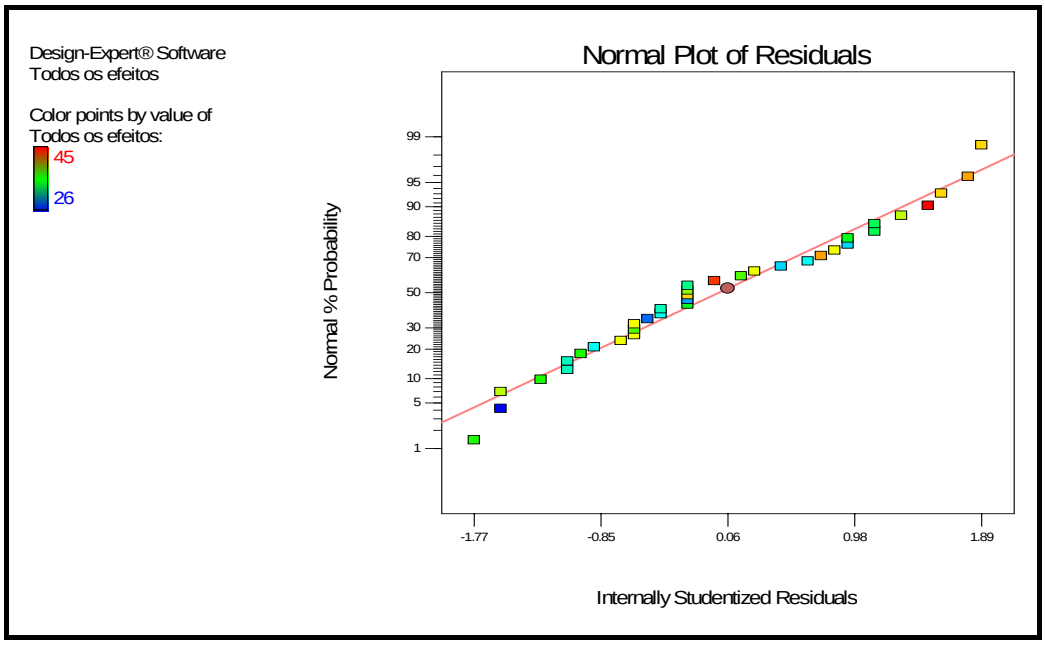


Figura A.17 – Gráfico de probabilidade normal para o experimento da polpa de papel.

- Gráfico de Resíduos x Diagnóstico (Valores previstos): As características importantes que este gráfico permite avaliar são: a adequação do modelo quanto à variância constante e a aditividade. Dessa forma, analisando o gráfico da Figura A.18 obtido para o experimento em questão, é possível identificar que não há indícios de padrão de comportamento estabelecido pelas observações, o que permite concluir que a variância é constante. No que tange a aditividade, pode-se observar que os valores estão distribuídos de modo uniforme em torno de zero, mas não se deve assegurar a aditividade dos efeitos.

PUC-Rio - Certificação Digital Nº 0611755/CA

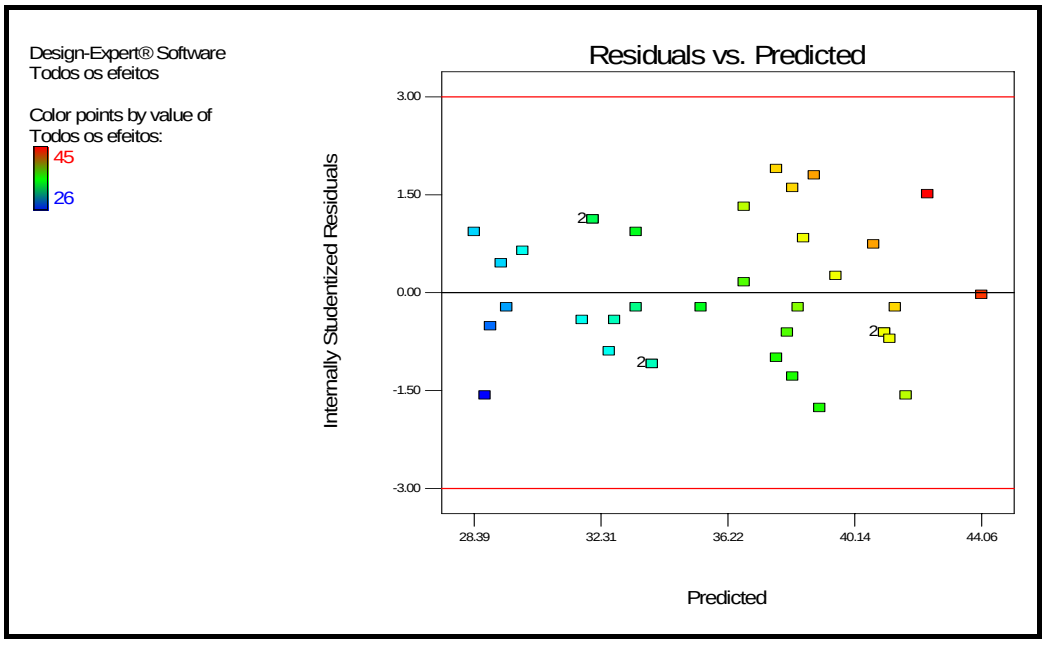


Figura A.18 – Gráfico dos resíduos *versus* os valores previstos.

- Gráfico da Distância de Cook (Cook's Distance - Di): Para analisar a influência dos resíduos é possível destacar na observação do gráfico abaixo que não há valores de resíduos maiores ou iguais a 0,5 (Di≥ 0,5), logo se confirma a ausência de observações influentes (*outlier*). Ressalta-se que quanto maior este resíduo (Valores de Di ≥ 0,5) mais influente a observação será para o modelo.

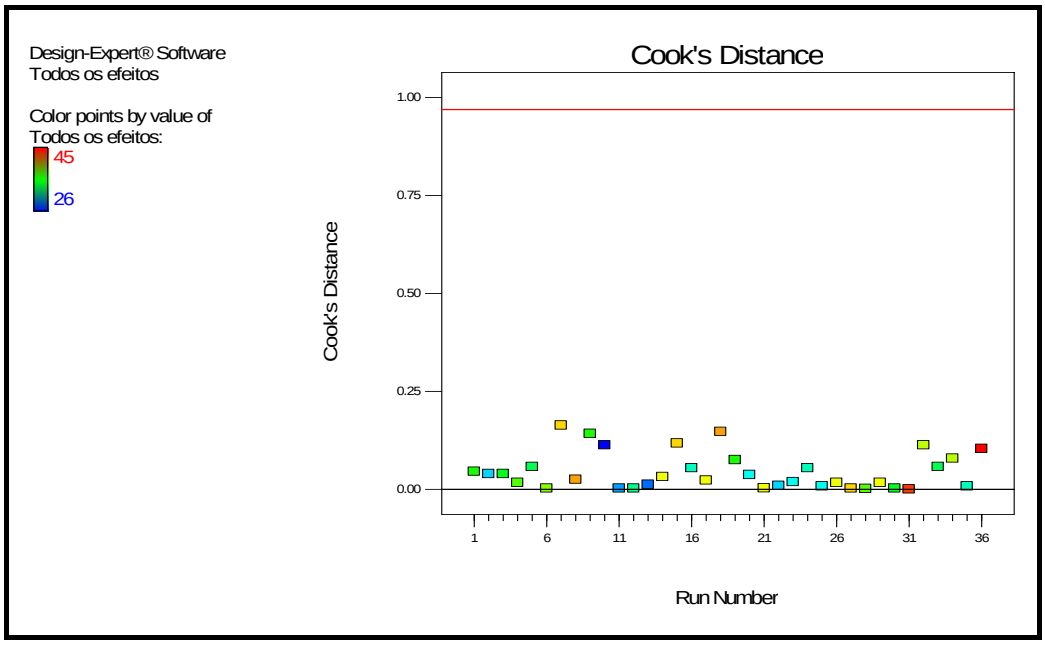
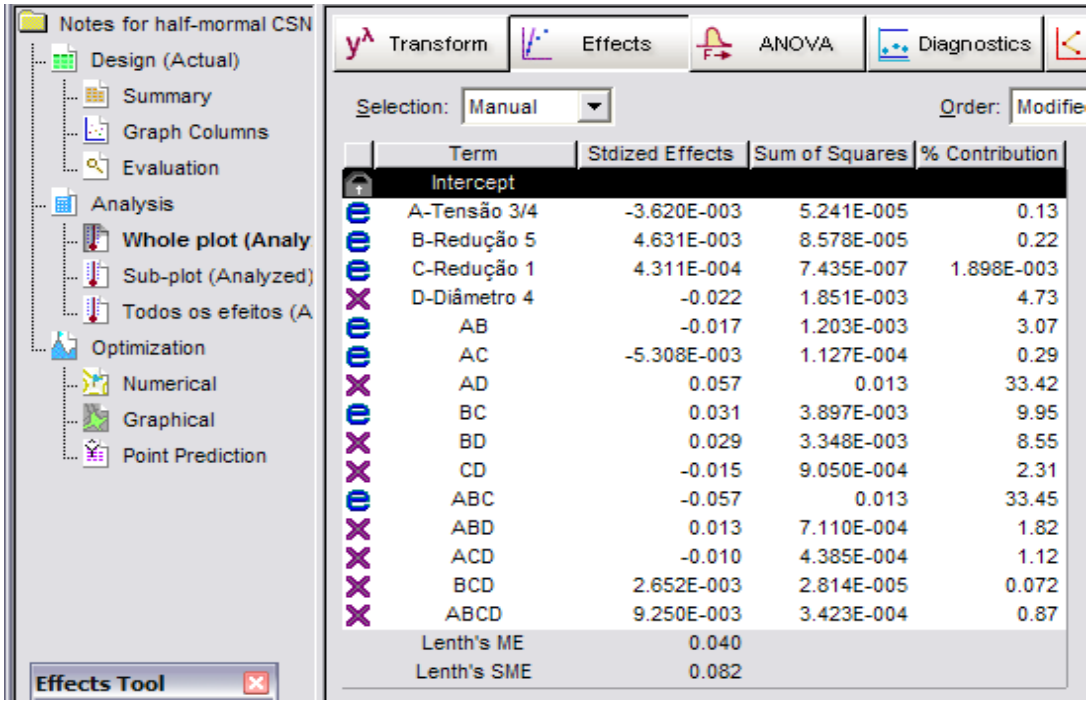


Figura A.19 – Gráfico da distância de *Cook* para o experimento *split-plot*.

PUC-Rio - Certificação Digital Nº 0611755/CA

APÊNDICE B **– Lista dos efeitos *whole-plot* para análise do gráfico *half-normal*.**

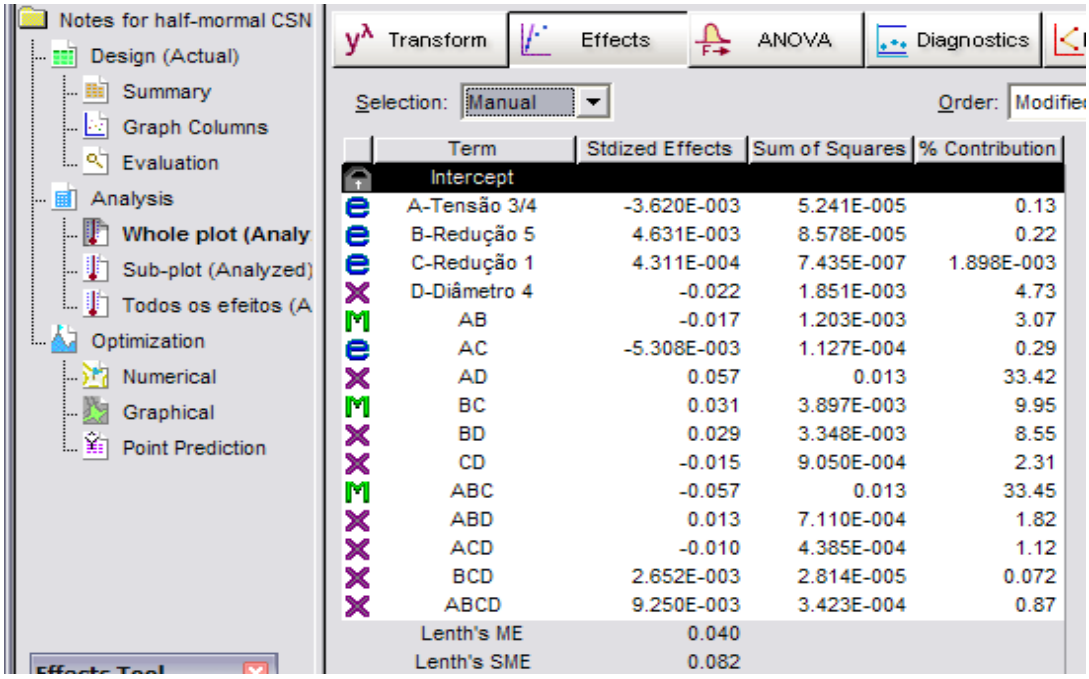


Term	Stdized Effects	Sum of Squares	% Contribution
Intercept			
A-Tensão 3/4	-3.620E-003	5.241E-005	0.13
B-Redução 5	4.631E-003	8.578E-005	0.22
C-Redução 1	4.311E-004	7.435E-007	1.898E-003
D-Diâmetro 4	-0.022	1.851E-003	4.73
AB	-0.017	1.203E-003	3.07
AC	-5.308E-003	1.127E-004	0.29
AD	0.057	0.013	33.42
BC	0.031	3.897E-003	9.95
BD	0.029	3.348E-003	8.55
CD	-0.015	9.050E-004	2.31
ABC	-0.057	0.013	33.45
ABD	0.013	7.110E-004	1.82
ACD	-0.010	4.385E-004	1.12
BCD	2.652E-003	2.814E-005	0.072
ABCD	9.250E-003	3.423E-004	0.87
Lenth's ME	0.040		
Lenth's SME	0.082		

Figura B.1 – Lista dos efeitos *whole-plot* para análise do gráfico *half-normal*.

PUC-Rio - Certificação Digital Nº 0611755/CA

APÊNDICE C – **Lista dos efeitos *whole-plot* para análise do gráfico *half-normal* após seleção dos efeitos significativos.**

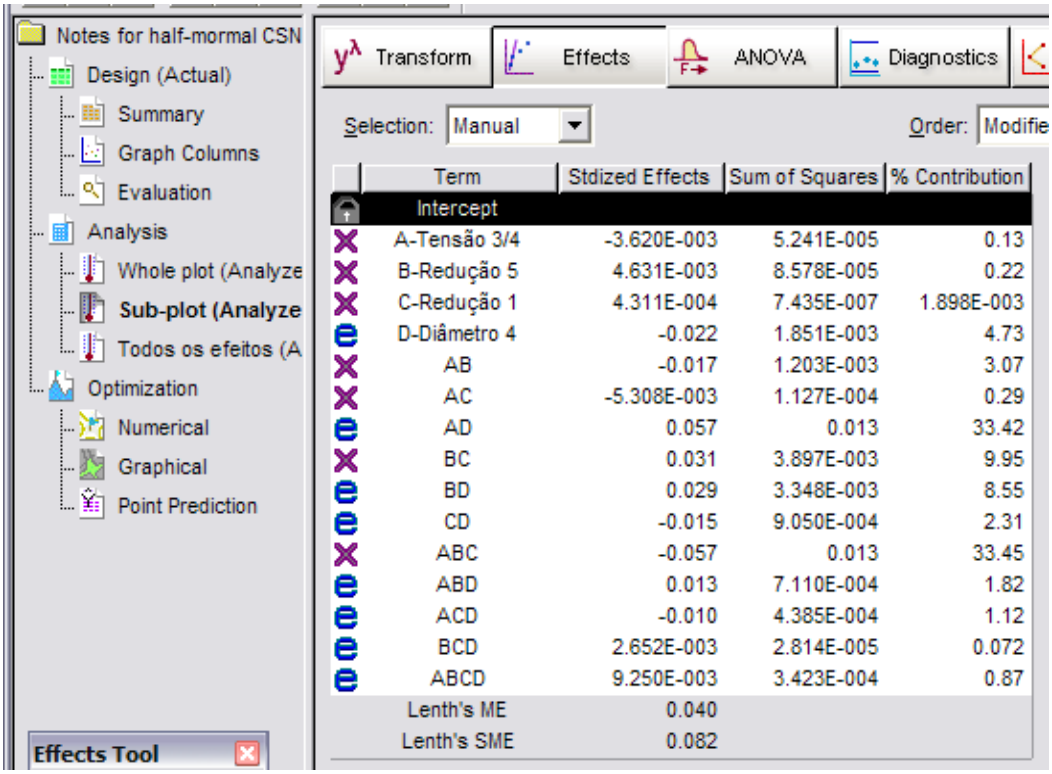


Term	Stdized Effects	Sum of Squares	% Contribution
Intercept			
A-Tensão 3/4	-3.620E-003	5.241E-005	0.13
B-Redução 5	4.631E-003	8.578E-005	0.22
C-Redução 1	4.311E-004	7.435E-007	1.898E-003
D-Diâmetro 4	-0.022	1.851E-003	4.73
AB	-0.017	1.203E-003	3.07
AC	-5.308E-003	1.127E-004	0.29
AD	0.057	0.013	33.42
BC	0.031	3.897E-003	9.95
BD	0.029	3.348E-003	8.55
CD	-0.015	9.050E-004	2.31
ABC	-0.057	0.013	33.45
ABD	0.013	7.110E-004	1.82
ACD	-0.010	4.385E-004	1.12
BCD	2.652E-003	2.814E-005	0.072
ABCD	9.250E-003	3.423E-004	0.87
Lenth's ME	0.040		
Lenth's SME	0.082		

Figura C.1 – Lista dos efeitos *whole-plot* para análise do gráfico *half-normal* após seleção dos efeitos significativos

PUC-Rio - Certificação Digital Nº 0611755/CA

APÊNDICE D **- Lista dos efeitos *sub-plot* para análise do gráfico *half-normal***

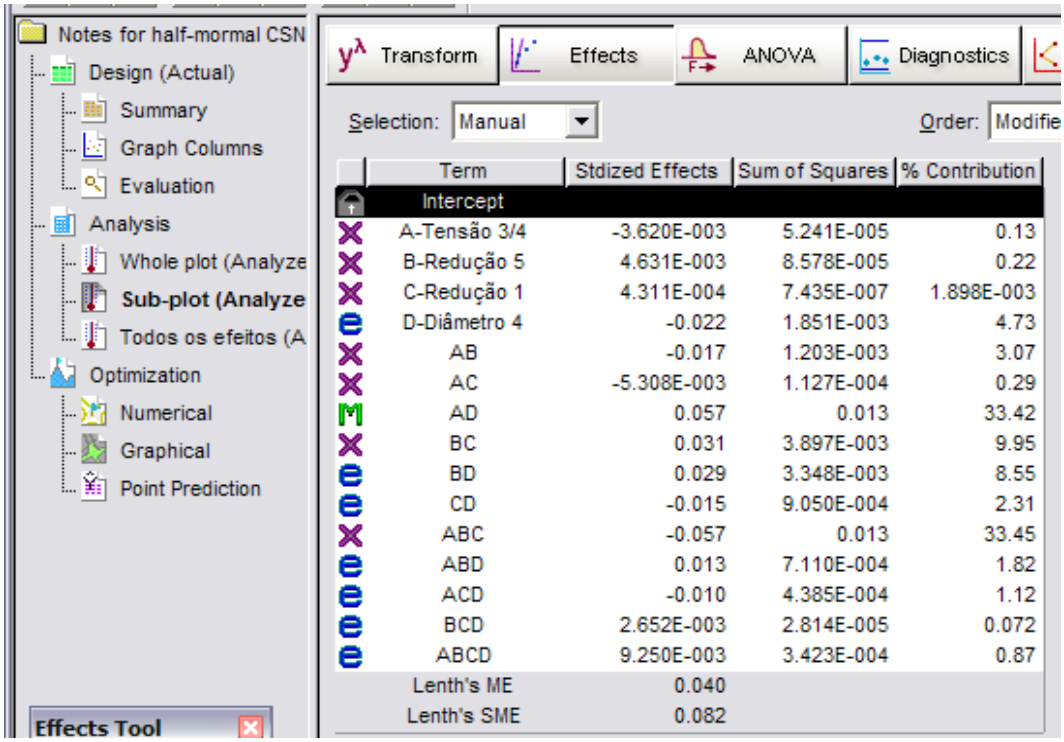


	Term	Stdized Effects	Sum of Squares	% Contribution
	Intercept			
X	A-Tensão 3/4	-3.620E-003	5.241E-005	0.13
X	B-Redução 5	4.631E-003	8.578E-005	0.22
X	C-Redução 1	4.311E-004	7.435E-007	1.898E-003
e	D-Diâmetro 4	-0.022	1.851E-003	4.73
X	AB	-0.017	1.203E-003	3.07
X	AC	-5.308E-003	1.127E-004	0.29
e	AD	0.057	0.013	33.42
X	BC	0.031	3.897E-003	9.95
e	BD	0.029	3.348E-003	8.55
e	CD	-0.015	9.050E-004	2.31
X	ABC	-0.057	0.013	33.45
e	ABD	0.013	7.110E-004	1.82
e	ACD	-0.010	4.385E-004	1.12
e	BCD	2.652E-003	2.814E-005	0.072
e	ABCD	9.250E-003	3.423E-004	0.87
	Lenth's ME	0.040		
	Lenth's SME	0.082		

Figura D.1 – Lista dos efeitos *sub-plot* para análise do gráfico *half-normal*.

PUC-Rio - Certificação Digital Nº 0611755/CA

APÊNDICE E **- Lista dos efeitos *sub-plot* para análise do gráfico *half-normal* após seleção dos efeitos significativos.**



	Term	Stdized Effects	Sum of Squares	% Contribution
	Intercept			
X	A-Tensão 3/4	-3.620E-003	5.241E-005	0.13
X	B-Redução 5	4.631E-003	8.578E-005	0.22
X	C-Redução 1	4.311E-004	7.435E-007	1.898E-003
e	D-Diâmetro 4	-0.022	1.851E-003	4.73
X	AB	-0.017	1.203E-003	3.07
X	AC	-5.308E-003	1.127E-004	0.29
M	AD	0.057	0.013	33.42
X	BC	0.031	3.897E-003	9.95
e	BD	0.029	3.348E-003	8.55
e	CD	-0.015	9.050E-004	2.31
X	ABC	-0.057	0.013	33.45
e	ABD	0.013	7.110E-004	1.82
e	ACD	-0.010	4.385E-004	1.12
e	BCD	2.652E-003	2.814E-005	0.072
e	ABCD	9.250E-003	3.423E-004	0.87
	Lenth's ME	0.040		
	Lenth's SME	0.082		

Figura E.1 – Lista dos efeitos *sub-plot* para análise do gráfico *half-normal* após seleção dos efeitos significativos.

PUC-Rio - Certificação Digital Nº 0611755/CA

Anexos

ANEXO A – Revisão de regressão linear e procedimento para cálculo de intervalo de confiança para a média da resposta e de intervalo de previsão para uma futura resposta (Vieira, 2004).

A.1 Regressão Linear

A análise de regressão linear procura estabelecer a relação entre uma variável de resposta y e um conjunto de variáveis independentes x_1, x_2, ..., x_k. Entende-se por variável de resposta uma medida de desempenho ou uma característica da qualidade de um produto ou um processo produtivo, os quais são influenciados pelas variáveis independentes, que também podem ser chamadas de variáveis de regressão. O termo "linear" é utilizado pelo fato de a equação da variável de resposta y representar uma função linear dos parâmetros desconhecidos β_0, β_1, β_2, ..., β_k, como será visto a seguir.

A.1.1 Função de Resposta

Quando um produto ou um processo possui uma resposta y que depende das variáveis de entrada controláveis x_1, x_2, ..., x_k, diz-se que há uma relação funcional entre y e x_1, x_2, ..., x_k, que pode ser representada da seguinte forma:

$$y = f(\beta_1, \beta_2, \ldots, \beta_k, x_1, x_2, \ldots, x_k) + \varepsilon \qquad \text{(A.1)}$$

sendo o termo ε o erro, que representa outras fontes de variabilidade não consideradas em f, como erros de medição da resposta ou outras variações inerentes ao processo ou sistema.

Geralmente, a relação funcional apresentada na Equação (A.1) não é conhecida, o que propicia a utilização de modelos lineares de regressão, que podem ter a seguinte representação:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \qquad \text{(A.2)}$$

onde os parâmetros desconhecidos, β_0, β_1, β_2, ..., β_k, são chamados de coeficientes de regressão.

Os modelos aparentemente mais complexos também podem ser representados pelo modelo apresentado na Equação (A.2). Como exemplo,

PUC-Rio - Certificação Digital Nº 0611755/CA

considere a adição de um termo de interação a um modelo de primeira ordem com duas variáveis, ou seja,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon \quad \text{(A.3)}$$

Se for feita uma substituição do tipo $x_3 = x_1 x_2$ e $\beta_3 = \beta_{12}$, o modelo apresentado na equação (A.3) pode ser escrito da seguinte forma:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \quad \text{(A.4)}$$

que tem a mesma forma da Equação (A.2).

Da mesma forma, considerando um modelo de segunda ordem com duas variáveis, tem-se

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon \quad \text{(A.5)}$$

Fazendo uma substituição do tipo $x_3 = x_1^2$, $x_4 = x_2^2$, $x_5 = x_1 x_2$, $\beta_3 = \beta_{11}$, $\beta_4 = \beta_{22}$ e $\beta_5 = \beta_{12}$, o modelo da Equação (I.5) fica da seguinte forma:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon \quad \text{(A.6)}$$

que também é um modelo de regressão linear.

A seguir, serão apresentados métodos para estimar os parâmetros dos modelos de regressão linear e para testar a significância dos coeficientes, o que é freqüentemente chamado de ajuste do modelo.

A.1.2 Estimação dos Parâmetros nos Modelos de Regressão Linear

O método dos mínimos quadrados é o método clássico de estimação dos parâmetros dos modelos de regressão linear.

Considere $n > k$ observações da variável de resposta, ou seja, $y_1, y_2, ..., y_n$, sendo que, para cada resposta, tem-se as observações das variáveis de regressão, como mostrado na Tabela A.1.

Escrevendo a Equação (A.2) utilizando os dados apresentados na Tabela A.1, tem-se

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i \,, \quad i = 1, 2, \ldots, n \quad \text{(A.7)}$$

Assume-se que o erro ε_i são variáveis aleatórias independentes, com média zero, ou seja, $E(\varepsilon_i) = 0$, e variância constante igual a σ^2.

PUC-Rio - Certificação Digital Nº 0611755/CA

Tabela A.1 – Dados para o Modelo de Regressão Linear

x_1	x_2	...	x_k	y
x_{11}	x_{12}	...	x_{1k}	y_1
x_{21}	x_{22}	...	x_{2k}	y_2
.	.		.	.
.	.		.	.
.	.		.	.
x_{n1}	x_{n2}	...	x_{nk}	y_n

Na forma matricial, a Equação (A.7) é representada da seguinte forma:

$$\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} \tag{A.8}$$

onde

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad e \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Nesse caso, $\mathbf{y}$ é um vetor $n \times 1$ das observações, $\mathbf{X}$ é uma matriz $n \times p$ ($p = k + 1$) dos níveis das variáveis independentes, $\boldsymbol{\beta}$ é um vetor $p \times 1$ dos coeficientes de regressão e $\boldsymbol{\varepsilon}$ é um vetor $n \times 1$ dos erros aleatórios.

O método dos mínimos quadrados escolhe os β's na Equação (A.7) de tal forma que a soma dos quadrados dos erros ε_i sejam minimizados. Para isso, define-se primeiramente a função de mínimos quadrados L.

$$L = \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X\beta})'(\mathbf{y} - \mathbf{X\beta}) = \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X\beta}$$

Sabendo que $\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}$ é uma matriz 1×1, a sua transposta $\boldsymbol{\beta}\mathbf{X}\mathbf{y}'$ é ela própria. Com isso, a função L pode ser expressa da seguinte forma:

$$L = \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X\beta} \tag{A.9}$$

Como a função L deve ser minimizada em relação a $\boldsymbol{\beta}$, os estimadores de mínimos quadrados, ou seja, $\hat{\boldsymbol{\beta}}$, deve satisfazer a

PUC-Rio - Certificação Digital Nº 0611755/CA

$$\left.\frac{\partial L}{\partial \boldsymbol{\beta}}\right|_{\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = 0$$

Simplificando,

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{A.10}$$

A Equação (A.10) é o conjunto de equações normais de mínimos quadrados representados na forma matricial. Desde que $\mathbf{X}'\mathbf{X}$ seja positiva definida, pode-se resolver a Equação (A.10) multiplicando ambos os seus membros por $(\mathbf{X}'\mathbf{X})^{-1}$. Dessa forma, os estimadores de mínimos quadrados de $\boldsymbol{\beta}$ são

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{A.11}$$

e o modelo de regressão ajustado é

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \tag{A.12}$$

Na forma escalar, o modelo ajustado é

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j x_{ij}\ , \quad i = 1, 2, \ldots, n \tag{A.13}$$

A diferença entre a observação y_i e o valor ajustado $\hat{y}_i$ é o resíduo. Com isso, o vetor $n \times 1$ dos resíduos é

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \tag{A.14}$$

Nos modelos de regressão linear, o método dos mínimos quadrados produz estimadores não enviesados dos parâmetros $\boldsymbol{\beta}$. Portanto, $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ (Myers e Montgomery, 2002, pág. 25).

A variância de $\hat{\boldsymbol{\beta}}$ pode ser obtida a partir da matriz de variância-covariância:

$$\text{var}(\hat{\boldsymbol{\beta}}) = E\left[\left(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})\right)\left(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})\right)'\right] \tag{A.15}$$

que é uma matriz simétrica $p \times p$, cujo i-ésimo elemento da diagonal principal é a variância de β_i e o elemento (ij) é a covariância entre β_i e β_j. A matriz de covariância de $\hat{\boldsymbol{\beta}}$ é (Myers e Montgomery, 2002, pág. 27):

$$\text{var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \tag{A.16}$$

O estimador dos mínimos quadrados de $\boldsymbol{\beta}$ é um estimador linear não enviesado e de variância mínima, o que lhe confere o titulo de melhor estimador linear não enviesado.

PUC-Rio - Certificação Digital Nº 0611755/CA

Pode-se demonstrar (Myers e Montgomery, 2002, pág. 27) que a estimativa da variância σ^2 do erro ε é relação entre a soma dos quadrados dos resíduos SS_E pelo numero de graus de liberdade ($n - p$), ou seja:

$$\hat{\sigma}^2 = \frac{SS_E}{n-p} \tag{A.17}$$

sendo a soma dos quadrados dos resíduos obtida da seguinte forma:

$$SS_E = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 \tag{A.18}$$

A.2 Intervalos para a Média e para a Previsão da Resposta

Dado um ponto $x_{01}, x_{02}, \ldots, x_{0k}$, no espaço das variáveis regressoras, tem-se o vetor

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ x_{01} \\ x_{02} \\ \vdots \\ x_{0k} \end{bmatrix}$$

pode-se obter um intervalo de confiança para a média da resposta e o intervalo de previsão para uma futura observação de um valor individual da resposta.

A.2.1 Intervalo de Confiança para a Média da Resposta

A média da resposta no ponto $\mathbf{x}_0$ é

$$\mu_{y|\mathbf{x}_0} = \beta_0 + \beta_1 x_{01} + \beta_2 x_{02} + \ldots + \beta_k x_{0k}$$

O estimador da média da resposta neste ponto é

$$\hat{\mu}_{y|\mathbf{x}_0} = \hat{y}(\mathbf{x}_0) = \mathbf{x}_0' \hat{\boldsymbol{\beta}} \tag{A.24}$$

O estimador não é enviesado, pois

$$E\left[\hat{\mu}_{y|\mathbf{x}_0}\right] = E\left(\mathbf{x}_0' \hat{\boldsymbol{\beta}}\right) = \mathbf{x}_0' \boldsymbol{\beta} = \mu_{y|\mathbf{x}_0}$$

A variância do estimador da média da resposta é

$$\text{var}\left[\hat{\mu}_{y|\mathbf{x}_0}\right] = \text{var}\left[\hat{y}(\mathbf{x}_0)\right] = \text{var}\left(\mathbf{x}_0' \hat{\boldsymbol{\beta}}\right) \tag{A.25}$$

Na Equação (2.16) tem-se que:

PUC-Rio - Certificação Digital Nº 0611755/CA

$$\text{var}\left(\hat{\boldsymbol{\beta}}\right) = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

Então,

$$\text{var}\left[\hat{\mu}_{y|\mathbf{x}_0}\right] = \sigma^2 \mathbf{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0$$

Como a distribuição de *y* é normal, o quociente

$$t = \frac{\hat{\mu}_{y|\mathbf{x}_0} - \mu_{y|\mathbf{x}_0}}{\text{var}\left(\hat{\mu}_{y|\mathbf{x}_0}\right)} \tag{A.26}$$

tem distribuição *t* com $(n - p)$ graus de liberdade.

Portanto, para um intervalo de confiança de $100(1 - \alpha)\%$ tem-se que:

$$-t_{\alpha/2,n-p} \leq \frac{\hat{\mu}_{y|\mathbf{x}_0} - \mu_{y|\mathbf{x}_0}}{\sqrt{\text{var}\left(\hat{\mu}_{y|\mathbf{x}_0}\right)}} \leq t_{\alpha/2,n-p} \tag{A.27}$$

Nas Equações (A.24) e (A.25) tem-se que $\hat{\mu}_{y|\mathbf{x}0} = \hat{y}\left(\mathbf{x}_0\right)$ e $\text{var}\left[\hat{\mu}_{y|\mathbf{x}_0}\right] = \sigma^2 \mathbf{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0$.

Substituindo em (A.27) tem-se que:

$$-t_{\alpha/2,n-p} \leq \frac{\hat{y}\left(\mathbf{x_0}\right) - \mu_{y|\mathbf{x}_0}}{\sqrt{\sigma^2 \mathbf{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0}} \leq t_{\alpha/2,n-p}.$$

O que é equivalente a

$$\begin{aligned} &\hat{y}\left(\mathbf{x_0}\right) - t_{\alpha/2,n-p} \sqrt{\sigma^2 \mathbf{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0} \leq \mu_{y|\mathbf{x}_0} \\ &\leq \hat{y}\left(\mathbf{x_0}\right) + t_{\alpha/2,n-p} \sqrt{\sigma^2 \mathbf{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0} \end{aligned} \tag{A.28}$$

que é o intervalo de confiança de $100(1 - \alpha)\%$ para a média da resposta no ponto $x_{01}, x_{02}, \ldots, x_{0k}$.

A.2.2 Intervalo de Previsão para uma Futura Resposta

O modelo no ponto $\mathbf{x}_0$ é

$$y\left(\mathbf{x}_0\right) = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02} + \ldots + \hat{\beta}_k x_{0k} + \varepsilon \text{ ou}$$

$$y\left(\mathbf{x}_0\right) = \mathbf{x}_0' \hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}$$

A estimativa de uma nova resposta neste ponto é a mesma estimativa da média:

$$E\left[y\left(\mathbf{x}_0\right)\right] = \mathbf{x}_0' \hat{\boldsymbol{\beta}}$$

A variância de uma nova resposta neste ponto é

PUC-Rio - Certificação Digital Nº 0611755/CA

$$\text{var}\left[y(\mathbf{x}_0)\right] = \text{var}\left(\mathbf{x}_0'\hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}\right) = \text{var}\left(\mathbf{x}_0'\hat{\boldsymbol{\beta}}\right) + \text{var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 + \sigma^2$$

ou

$$\text{var}\left[y(\mathbf{x}_0)\right] = \sigma^2\left(\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 + 1\right) \tag{A.29}$$

e o intervalo de 100(1 – α)% de probabilidade para uma nova resposta no ponto $x_{01}, x_{02}, \ldots, x_{0k}$ é

$$\begin{aligned} &\hat{y}(\mathbf{x_0}) - t_{\alpha/2,n-p}\sqrt{\text{var}\left[y(\mathbf{x}_0)\right]} \le y(\mathbf{x}_0) \le \hat{y}(\mathbf{x_0}) + t_{\alpha/2,n-p}\sqrt{\text{var}\left[y(\mathbf{x}_0)\right]} \\ &\hat{y}(\mathbf{x_0}) - t_{\alpha/2,n-p}\sqrt{\sigma^2\left[\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 + 1\right]} \le y(\mathbf{x}_0) \\ &\le \hat{y}(\mathbf{x_0}) + t_{\alpha/2,n-p}\sqrt{\sigma^2\left[\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 + 1\right]} \end{aligned} \tag{A.30}$$

PUC-Rio - Certificação Digital Nº 0611755/CA