



UNIVERSIDADE FEDERAL DO PARÁ
CENTRO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO
EM MATEMÁTICA E ESTATÍSTICA

José Gracildo de Carvalho Júnior

Gráfico de Controle de Regressão Estrutural

Orientadora: Profa. Sílvia dos Santos de Almeida, *Dra.*

Belém
2006

FICHA CATALOGRÁFICA

Dados Internacionais de catalogação na publicação Biblioteca do Centro de Ciências Sociais e Educação da UEPA.

Carvalho Júnior, José Gracildo de

Gráfico de controle de regressão estrutural / Maria de Fátima Cravo de Sousa; Orientadora, Silvia dos Santos de Almeida. Belém, 2006.

Dissertação (Mestrado em Matemática e Estatística) - Universidade Federal do Pará, Belém, 2006.

1. Prática de ensino 2. Professores - Estudo e Ensino 3. Métodos de ensino 4. Educação de adultos I. Título.

CDD: 21 ed. 370.733

José Gracildo de Carvalho Júnior

Gráfico de Controle de Regressão Estrutural

Dissertação de Conclusão de Curso apresentada ao Programa de Pós-Graduação em Matemática e Estatística da Universidade Federal do Pará para obtenção do Título de Mestre em Matemática e Estatística.

Orientadora: **Profa. Silvia dos Santos de Almeida, *Dra.***

Co-Orientador: **Prof. Edson M. L. S. Ramos, *Dr.***

Belém
2006

José Gracildo de Carvalho Júnior

Gráfico de Controle de Regressão Estrutural

Esta dissertação foi julgada e aprovada para a obtenção do título de Mestre em Matemática e Estatística, no Programa de Pós-Graduação em Matemática e Estatística, da Universidade Federal do Pará

Belém, 17 de Março de 2006

Prof. Marcus Pinto da Rocha, *Dr.*
(Coordenador do Programa de Pós-Graduação em Matemática e Estatística)

Banca Examinadora

Profa. Silvia dos Santos de Almeida, *Dra.*
Universidade Federal do Pará
Orientadora

Prof. Damiano Nóbrega da Silva, *Dr.*
Universidade Federal do Rio Grande do Norte
Examinador

Prof. Edson Marcos L. S. Ramos, *Dr.*
Universidade Federal do Pará
Co-Orientador
Examinador

DEDICATÓRIA

*A Meus Pais José Gracildo de Carvalho
e Maria Trindade Corrêa de Carvalho*

Agradecimentos

À Deus por ter me concedido a vida e a capacidade de realizar este trabalho;

À Universidade Federal do Pará;

A minha orientadora, Professora Silvia dos Santos de Almeida, integrante do Programa de Pós-Graduação em Matemática e Estatística, pela orientação competente, dedicando-se de forma marcante na elaboração deste trabalho;

Ao Professor Edson Marcos Leal Soares Ramos, integrante do Programa de Pós-Graduação em Matemática e Estatística, pela contribuição ao longo do curso de graduação, especialização e mestrado;

Ao Professor Joaquin Queiroz, integrante do programa de Pós-Graduação em Matemática e Estatística, pelas sugestões abalizadas e significativas para a realização deste trabalho;

Ao Professor Dr. Jorge Ferreira, integrante do Programa de Pós-Graduação em Matemática e Estatística, pela contribuição valiosa e significativa ao longo de todo o curso de mestrado;

Ao Professor Dr. João dos Santos Protázio, integrante do Programa de Pós-Graduação em Matemática e Estatística, pela contribuição valiosa e significativa ao longo do curso de mestrado;

Aos amigos, Marco Pollo, Paulo Fernando, Edilson Torres, Félix Lelis, Rogério Conceição, Vanilson Barradas, Rosa Lobo, Michelle Borges, Pablo Agria, Luis, Edney, Pedro, Heraldo, Diogo, Helena, Ana Paula, Silvia Helem, Alessandro Baldez, Renato, Irazel, Leandro e todos os demais amigos da UFPA pela amizade, incentivo e confiança em mim depositada ao longo de todo o curso de mestrado;

À minha Família, pela paciência e confiança atribuída à minha pessoa;

À minha eterna companheira, Glauce Glaebe Brandão, pelo amor, compreensão, apoio, carinho e por ter estado presente em todos os momentos que levaram a realização deste trabalho;

Finalmente, a todos que de alguma forma contribuíram para a realização deste trabalho.

*“Amo o Senhor Todo Poderoso,
porque ele ouve a minha voz
e as minhas súplicas.
Porque inclinou para mim
os seus ouvidos,
invocá-lo-ei enquanto eu viver ”.*

Salmos CXVI; cap.1,2.

Resumo

CARVALHO JR., José Gracildo de. Gráfico de Controle de Regressão Estrutural. 2006. Dissertação (Mestrado em Matemática e Estatística)-PPGME/UFPa, Belém - PA, Brasil.

O objetivo deste trabalho é propor uma ferramenta do controle estatístico da qualidade e de modelos de regressão com erro nas variáveis estrutural, baseado nos erros de medida das variáveis analisadas ao longo de um processo. Desenvolveu-se ao longo deste trabalho uma metodologia atribuída a variáveis sujeitas a erros de medida, além da proposição de três estimativas para σ utilizadas para a construção e desenvolvimento dos três respectivos gráficos de controle de regressão estrutural através dos resultados teóricos obtidos. A teoria desenvolvida mostrou-se eficaz e relevante para o monitoramento de processos onde haja erro de medição na variável independente. Verificou-se também, que as estimativas de mínimos quadrados para o modelo clássico de regressão são inconsistentes e viesadas, quando utilizadas no modelo com erro nas variáveis, fato pelo qual, apresentam-se estimativas consistentes e não viesadas para os parâmetros do modelo proposto. Dentre os gráficos de controle da qualidade desenvolvidos verificou-se que, quando as estimativas de σ utilizadas são baseadas no erro padrão sobre a reta de regressão ajustada, no erro padrão sobre a esperança condicional de Y dado X e no erro padrão sobre um valor predito da variável dependente, que estes gráficos monitoram satisfatoriamente processos relacionados as mais diversas áreas do conhecimento científico. Tabelas comparativas entre estimadores clássicos de regressão e estimadores admitindo-se erro de medição na variável independente são apresentadas, onde se verificou a incapacidade de tais estimadores clássicos em detectar de forma significativa os erros na variável independente quando o tamanho da amostra observada aumenta. Testou-se a metodologia proposta no escopo deste trabalho em dados da literatura, assim como em dados reais, constatando-se assim sua plena aplicabilidade. Finalmente apresentaram-se três novas ferramentas gráficas do controle estatístico da qualidade combinado com o modelo de regressão com variáveis sujeitas a erros de medida.

Palavra-Chave: Controle Estatístico da Qualidade, Gráficos de Controle de Regressão e Modelo Estrutural.

Abstract

CARVALHO Jr., José Gracildo de. Control Charts of Structural Regression. 2006. Thesis (Master in Matematic and Statistic) PPGME/UFPA, Belém - PA, Brazil.

The objective of this work is to suggest a statistical quality control tool and regression models with error in the structural variables based on the measure errors of the analyzed variables through the process. It was developed in this work an imputed methodology to variables submitted to measure errors besides the proposition of three estimates for used to the construction and development of three respective control charts of structural regression through the obtained theoretical results. The developed theory showed efficient and relevant to the processes accompaniment where there is measurement error in the independent variable. It was also verified that the minimum square estimates to the classic model of regression are inconsistent and biased where used in models with errors in the variables where presents consistent estimates and do not biased to the parameters of the proposal model. Among the developed quality control charts verified that when the used estimates are based on standard error on the adjusted regression line, in the standard error on the conditional mean of Y given X and in the standard error on a predict value of the dependent variable that this charts observe processes related to several areas of scientific knowing. Comparing tables between classic estimators of regression and estimators admitting measurement error in the independent variable are presented where it was verified the incapacity of such classic estimators in detecting of significant form the errors in the independent variable when the observed sample size increases. It was tested the propose methodology in this work in literature data, so as in real data. So, it was evidenced its complete applicability. Finally, it was presented three new graphic tolls of the statistical control of quality combined to the regression model with variables submitted to measurement errors.

Key Words: Statistical Quality Control, Regression Control Charts and Structural Model.

Sumário

Resumo	viii
Abstract	ix
Lista de Tabelas	xii
Lista de Figuras	xiii
1 Introdução	1
1.1 Principais Considerações	1
1.2 Importância e Justificativa do Trabalho	3
1.3 A Hipótese Básica da Dissertação	4
1.4 Definições dos Objetivos	4
1.4.1 Objetivo Geral	4
1.4.2 Objetivos Específicos	4
1.5 As Limitações do Trabalho	4
1.6 Estrutura do Trabalho	5
2 Revisão da Literatura	6
2.1 - Idéia Geral do Modelo com Erros nas Variáveis	6
2.1.1 - Modelo Funcional	14
2.1.2 - Modelo Estrutural	15
2.1.3 - Modelo Ultraestrutural	16
2.2 - Visão Geral sobre Gráficos de Controle	18
2.2.1 - A Relação entre Teste de Hipótese e Gráfico de Controle	19
2.2.2 - Tipos de Gráficos de Controle	21
2.2.3 - Gráfico de Controle por Variável	21
2.2.4 - Idéia Geral do Gráfico de Controle de Regressão	22
3 Modelo Estrutural com uma das Variâncias Conhecida	25
3.1 - Estimadores dos Parâmetros de Regressão	25
3.1.1 - Estimadores Consistentes de α e β	25
3.2 - Estimação do Verdadeiro Valor de U	32
3.3 - Estimador da Variância de Y ($\hat{\sigma}_y^2$)	33
3.4 - Estimador da Variância de \hat{Y} ($\hat{\sigma}_y^2$)	34
3.5 - Estimador da Variância de \hat{Y}_t ($\hat{\sigma}_{y_t}^2$)	35

3.6 - Erro Padrão da Reta de Regressão	36
3.7 - Erro Padrão da Reta de Regressão Estimada	36
3.8 - Erro Padrão do Valor Predito da Variável Dependente	37
3.9 - Intervalo de Confiança Assintótico	37
3.9.1 - Intervalos de Confiança para os Parâmetros da Regressão	37
3.10- Intervalo de Previsão Assintótico no Modelo Estrutural	38
3.10.1- Modelo de Previsão	38
3.10.2- Intervalo de Previsão no Modelo Estrutural	40
4 O Gráfico de Controle de Regressão	41
4.1 - Introdução	41
4.2 - Elementos do Gráfico de Controle de Regressão	42
4.3 - Construindo o Gráfico de Controle de Regressão	47
4.4 - Analisando o Gráfico de Controle de Regressão	52
5 O Gráfico de Controle de Regressão Para o Modelo Estrutural e Aplicação	54
5.1 - Introdução	54
5.2 - Passos para elaboração do Gráfico de Controle de Regressão Estrutural	57
5.3 Construindo o Gráfico de Controle de Regressão Estrutural	58
5.3.1 Gráfico de Controle de Regressão Estrutural para o Exemplo de Fuller (1987).	58
5.3.2 Gráfico de Controle de Regressão Estrutural para os Dados de uma Indústria Brasileira.	63
5.4 Comparação entre os Estimadores de MQO e do Modelo Estrutural	70
5.4.1 Comparação entre os EMQO e do Modelo Estrutural para o Exemplo de Fuller (1987).	70
5.4.2 Comparação entre os EMQO e do Modelo Estrutural para o Exemplo de Mandel (1969).	72
6 Considerações Finais e Recomendações	74
6.1 Sugestões Para Trabalhos Futuros	75
A ANEXO	76
A.1 Introdução	76
A.2 Expansão das Somas	76
A.3 Obtenção das Esperanças	77
B ANEXO	79
B.1 Ordens de Grandeza de Seqüências de Números Reais: $O(\cdot)$, $o(\cdot)$	79
B.1.1 - Introdução	79
B.1.2 - Definição	79
B.1.3 - Teorema	81
B.2 Expansões em Séries de Taylor.	82
Bibliografia	84

Lista de Tabelas

4.1	- Produção de milho e nível de nitrogênio no Solo de Iowa, nos Estados Unidos.	47
4.2	- Estatísticas necessárias à construção do gráfico de controle de regressão .	48
4.3	- Nível de Nitrogênio estimado para o Modelo Funcional com $\sigma_{\delta}^2 = 57$	49
5.1	- Estatísticas obtidas a partir da Tabela 4.1 para o Modelo Estrutural com $\sigma_{\delta}^2 = 57$	59
5.2	- Nível de Nitrogênio estimado para o Modelo Estrutural com $\sigma_{\delta}^2 = 57$. . .	60
5.3	- Dados reais fornecidos por uma Indústria Brasileira sobre a Densidade Aparente e Peso dos Eletrodos produzidos de Agosto/2001 à Janeiro/2003	64
5.4	- Estatísticas obtidas a partir da Tabela 5.3 para o Modelo Estrutural com $\sigma_{\delta}^2 = 10$	65
5.5	- Peso do Eletrodo estimado para o Modelo Estrutural com $\sigma_{\delta}^2 = 10$	66
5.6	- Estimativas obtidas por EMQO para os dados da Tabela 4.1.	70
5.7	- Estimativas obtidas a partir da Tabela 4.1 para o Modelo Estrutural com diferentes valores de σ_{δ}^2	71
5.8	- Dados sobre Horas e Volume de Processamento de Correspondência para as Agências de Correio A, B, C e D, nos anos fiscais de 1962 e 1963.	72
5.9	- Estimativas obtidas por EMQO para os dados da Tabela 5.8.	72
5.10	- Estimativas obtidas a partir da Tabela 5.8 para o Modelo Estrutural com diferentes valores de σ_{δ}^2	73

Lista de Figuras

2.1	- Gráfico de Controle	19
2.2	- Gráfico de Controle de Regressão	23
4.1	- Probabilidades Normais para $\mu \pm k\sigma$, $k = 1, 2, 3$	46
4.2	- Diagrama de Dispersão da Produção de milho e nível de nitrogênio no Solo de Iowa, nos Estados Unidos, 1987.	48
4.3	- Gráfico de Controle de Regressão Funcional com $\sigma_{\delta}^2 = 57$, da Produção de Milho e Nível de Nitrogênio no Solo - Estado de Iowa, 1969, com limites de $2EP(Y) = 2S_e$ e $2EP(\hat{Y}_i) = 2S_{\hat{Y}_i}$, respectivamente.	50
4.4	- Gráfico de Controle de Regressão Funcional com $\sigma_{\delta}^2 = 57$, da Produção de Milho e Nível de Nitrogênio no Solo - Estado de Iowa, 1969. com limites de $2EP(\hat{Y}_t) = 2S_{\hat{Y}_{pi}}$ e o Gráfico Comparativo de Regressão Funcional com $\sigma_{\delta}^2 = 57$, respectivamente.	51
5.1	- Fluxograma da Construção do Gráfico de Controle de Regressão Estrutural.	57
5.2	- Diagrama de Dispersão entre Produção de Milho e Nível de Nitrogênio no Solo - Estado de Iowa, 1987	58
5.3	- Gráfico de Controle de Regressão Estrutural com $\sigma_{\delta}^2 = 57$ da Produção de Milho e Nível de Nitrogênio no Solo - Estado de Iowa, 1987. Com Limites de $2EP(Y) = 2S_e$ e $2EP(\hat{Y}_i) = 2S_{\hat{Y}_i}$, respectivamente.	61
5.4	- Gráfico de Controle de Regressão Estrutural com $\sigma_{\delta}^2 = 57$ da Produção de Milho e Nível de Nitrogênio no Solo - Estado de Iowa, 1987. Com Limites de $2EP(\hat{Y}_t) = 2S_{\hat{Y}_{pi}}$ e o Gráfico Comparativo de Regressão Estrutural com $\sigma_{\delta}^2 = 57$, respectivamente.	62
5.5	- Diagrama de Dispersão entre Densidade Aparente e Peso	64
5.6	- Gráfico de Controle de Regressão Estrutural com $\sigma_{\delta}^2 = 10$ da Densidade Aparente e Peso do Eletrodo. Com Limites de $2EP(Y) = 2S_e$ e $2EP(\hat{Y}_i) = 2S_{\hat{Y}_i}$, respectivamente	67
5.7	- Gráfico de Controle de Regressão Estrutural com $\sigma_{\delta}^2 = 10$ da Densidade Aparente e Peso do Eletrodo. Com Limites de $2EP(\hat{Y}_t) = 2S_{\hat{Y}_{pi}}$ e o Gráfico Comparativo com $\sigma_{\delta}^2 = 10$, respectivamente.	69

Capítulo 1

Introdução

1.1 Principais Considerações

Há tempos os métodos estatísticos tornaram-se ferramentas relevantes nos estudos e levantamentos de fenômenos nas mais variadas áreas do conhecimento, como por exemplo: na saúde, na agricultura, na economia, nas engenharias e na administração, contribuindo de forma significativa para a solução de problemas e para a melhoria e aperfeiçoamento destas áreas. Em aplicações industriais, dentre os diversos métodos estatísticos destacam-se a Análise de Regressão, Planejamento de Experimentos e o Controle Estatístico da Qualidade (*CEQ*).

“O objetivo principal do *CEQ* é monitorar o desempenho de processos ao longo do tempo e detectar eventos incomuns que influenciam as propriedades determinantes da qualidade do produto final. Uma vez encontradas as causas especiais responsáveis pelo evento incomum, melhorias no processo e na qualidade dos produtos podem ser sugeridas” (Montgomery, 2001).

O monitoramento do processo é realizado observando-se a característica (variável) da qualidade, que é expressa na maioria das vezes em termos de uma medida, por exemplo, o diâmetro de uma peça, o tempo médio de determinado serviço, e assim por diante. Antecipar-se a diversas ocorrências que por ventura venham a ocasionar alterações em um processo, induzindo-o a uma estabilidade no sistema avaliado é de fundamental importância. Para isso, faz-se necessário, prever estes acontecimentos que certamente oneram os custos de produção. Tais previsões só poderão ser efetuadas, caso o processo analisado tenha alguma relação entre as características mensuradas. Uma vez detectado o relacionamento, deve-se então encontrar uma expressão quantitativa, que revele a relação numérica entre fatos, e que, conseqüentemente, possa explicar os fenômenos que geralmente ocorrem.

A definição do modelo matemático adequado que reflita a relação existente entre os diversos fenômenos pode propiciar a real possibilidade de interpretar a situação, além de significar a obtenção de estimativas e previsões de ocorrências futuras. Constata-se na prática a existência dessa relação entre duas ou mais variáveis, estabelecendo-se a partir disso um modelo entre as variáveis avaliadas no processo. O modelo de regressão linear simples é a forma mais simples de equacionar essas relações, a qual pode ser formalmente descrita em Neter *et al.* (1996), como

$$Y_i = \alpha + \beta X_i + \varepsilon_i; \quad i = 1, \dots, n, \quad (1.1)$$

onde X é uma variável independente, com valores fixados, Y é uma variável dependente, α e β são os parâmetros desconhecidos da regressão, $\varepsilon_1, \dots, \varepsilon_n$ são os erros estocásticos independentes e identicamente distribuídos com média zero e variância constante σ_ε^2 . Na prática, mesmo com os constantes avanços tecnológicos nos procedimentos de mensuração, não é realista supor que a variável X seja medida sem nenhum erro. Neste caso, têm-se os chamados Modelos com Erros nas Variáveis (*MEV*), que alguns autores preferem utilizar somente a nomenclatura: Modelo Funcional (quando X é considerado fixo), Modelo Estrutural (quando X é considerado uma variável aleatória) e Modelo Ultraestrutural (uma combinação dos dois modelos, dependendo da forma de apresentação da variável X).

Fuller (1987) apresenta diversos exemplos deste tipo de modelo, de onde destaca-se a relação entre o rendimento de milho (y) e nitrogênio disponível na terra (x), no Estado de Iowa, no Estados Unidos. Para estimar o nitrogênio disponível na terra (x), é necessário tirar uma amostra de terra do conjunto experimental e executar uma análise de laboratório na amostra selecionada. Como resultado da amostragem e da análise de laboratório, não se observa o valor de x , mas sim uma estimativa de x . Então, representa-se o nitrogênio observado pela variável X , onde o verdadeiro valor do nitrogênio é dado por $x = X + \delta_i$, e δ_i é denotado como o erro de medida introduzido na amostragem e na análise de laboratório. Como é observado por Fuller (1987, pg. 2-3), é importante destacar que o tipo de coleta dos dados de nitrogênio da terra permite duas interpretações dos verdadeiros valores de x .

“Numa primeira interpretação assume-se que os campos (setores), são conjuntos de terras experimentais administrados pela estação experimental, de forma que produzam níveis diferentes de nitrogênio da terra, nos diferentes campos. Por exemplo, se os campos fossem amostras aleatórias das fazendas, do Estado de *Iowa*, então os verdadeiros valores de nitrogênio na terra poderiam ser tratados como variáveis aleatórias. Caso contrário, poderiam tratar o verdadeiro, mas desconhecido, nível de nitrogênio nos diferentes campos como sendo fixo.” (Fuller, 1987, pg. 2-3)

Esses erros de medida tanto podem ocorrer nas variáveis dependentes como nas independentes. No caso deles ocorrerem apenas na variável dependente, podem ser tratadas por métodos convencionais, sendo detectados automaticamente através dos resíduos. Já nos casos em que os erros ocorrem também nas variáveis independentes, eles devem ser tratados através de métodos de estimação, diferentes dos convencionais. Segundo Almeida, (2003), pg. 26-32, em situações deste tipo, os efeitos dos erros de medida sobre os estimadores dos parâmetros de regressão tornam estes estimadores inconsistentes* e viciados†, afetando também os testes de hipóteses e os intervalos de confiança.

1.2 Importância e Justificativa do Trabalho

A importância deste trabalho caracteriza-se devido ao fato da metodologia de regressão com erro nas variáveis ser inovadora para a construção de gráficos de controle, diferentemente do que foi apresentado por Mandel (1969), com o gráfico de controle de regressão clássico e por Almeida (2003), com o gráfico de controle de regressão funcional, pelo fato, do gráfico de controle de regressão estrutural considerar a variável independente sujeita a erro de medida, como sendo uma variável aleatória, diferentemente dos dois gráficos de controle anteriores. Devido a este fato obtém-se assim, um monitoramento dos processos com maior eficiência, do que os gráficos de controle de regressão convencionais, e, no sentido de produzir uma menor probabilidade de alarmes falsos, (alarmes estes sendo detectados quando o processo não está fora de controle estatístico) que não levam em conta o erro de medida na variável independente.

* Veja mais detalhes no Capítulo 2, definição 2.2.

† Veja mais detalhes no Capítulo 2, definição 2.3.

Finalmente, neste trabalho será considerado o problema reportado anteriormente em Análise de Regressão Simples, com o uso do Modelo Aditivo Estrutural, em que tal teoria será combinada com a teoria de gráficos de controle, para a obtenção de uma nova proposta metodológica que identifique anormalidades nos processos, através de gráficos de controle de regressão.

1.3 A Hipótese Básica da Dissertação

Parte-se da hipótese de que o gráfico de controle de regressão estrutural apresentará resultados mais precisos e menos sujeitos as imprecisões causadas pelos erros de medida associado a variável independente.

1.4 Definições dos Objetivos

1.4.1 Objetivo Geral

O objetivo deste trabalho é desenvolver o Gráfico de Controle de Regressão Estrutural (GCRE).

1.4.2 Objetivos Específicos

Os objetivos específicos desta dissertação são:

- Apresentar o modelo com erros aditivos nas variáveis para o Modelo Funcional, Estrutural e Ultraestrutural;
- Apresentar estimadores consistentes para o Modelo Aditivo Estrutural;
- Apresentar o intervalo de previsão para o monitoramento da variação média de y , para o modelo aditivo Estrutural;
- Testar a metodologia gráfica proposta no escopo do trabalho, através de exemplos da literatura e de uma aplicação real.

1.5 As Limitações do Trabalho

As limitações encontradas ao longo do trabalho são apresentadas a seguir:

- Falta de aplicativos para modelos com erros nas variáveis e de Gráficos de Controle de Regressão;
- Escassez de material bibliográfico referente ao Gráfico de Controle de Regressão.

1.6 Estrutura do Trabalho

Esta dissertação encontra-se dividida em seis capítulos, a saber: capítulos:

- Capítulo 1: apresenta a introdução, importância do trabalho, objetivo e as limitações do trabalho;
- Capítulo 2: desenvolve-se um estudo teórico sobre, modelos com erro nas variáveis e de gráficos de controle, procurando destacar os principais autores e suas contribuições;
- Capítulo 3: dedica-se ao estudo do Modelo Estrutural com a suposição de normalidade e o conhecimento de uma das variâncias. Também são apresentados os estimadores consistentes dos parâmetros da regressão e seus limites de confiança;
- Capítulo 4: mostra o gráfico de controle de regressão, abordando-se a fundamentação teórica e um exemplo prático, publicado por Mandel (1969), é apresentado;
- Capítulo 5: traz a proposição de uma metodologia para a construção do Gráfico de Controle de Regressão na presença de erros nas variáveis, denominado de (GCRE) e aplicações em três bancos de dados desta teoria desenvolvida;
- Capítulo 6: apresenta as considerações finais e sugestões para trabalhos futuros.

No próximo capítulo se fará uma revisão da literatura sobre os modelos com erros nas variáveis e gráficos de controle, buscando-se dar referências e definições básicas sobre os assuntos pertinentes à dissertação.

Capítulo 2

Revisão da Literatura

Neste capítulo faz-se uma descrição da teoria do Modelo de Regressão com Erros nas Variáveis (*MEV*), em seguida apresenta-se uma visão geral sobre gráficos de controle e, conseqüentemente, o gráfico de controle de regressão é apresentado.

2.1 - Idéia Geral do Modelo com Erros nas Variáveis

Toda medição é passível de erro, e como tal, este é o argumento mais forte para se acreditar que eles estejam sempre presentes em qualquer processo de medição. Da teoria clássica de regressão, sabe-se que o modelo de regressão considera que as variáveis independentes X_i são medidas sem erros e, conseqüentemente, toda a sua formulação esta baseada nesta idéia, e não pode ser aplicada a modelos que contenham variáveis com erros de medida. Segundo Miazaki e Monteiro (1993), este tipo de problema é bastante antigo na análise de regressão linear (ou não linear) e ainda hoje ele continua merecendo atenção como mostram os trabalhos de Nascimento (1983), Fuller (1987), Cheng e Van Ness (1999), Carroll *et al.* (1985) e Almeida (2003), que tratam deste tópico, denominado de modelo com Erros nas Variáveis.

Os Modelos com Erros nas Variáveis são uma generalização dos Modelos de Regressão padrão. Supondo um Modelo de Regressão, onde as variáveis U_i e y_i são quantidades não observáveis e estão relacionadas através da equação linear apresentada por Moram (1971), como

$$y_i = \alpha + \beta U_i, \quad i = 1, \dots, n, \quad (2.1)$$

onde α e β são parâmetros desconhecidos. Porém nem y e nem U são observados diretamente, de modo que os valores observados são X_i e Y_i , onde

$$Y_i = y_i + \varepsilon_i, \quad i = 1, \dots, n, \quad (2.2)$$

e

$$U_i = X_i - \delta_i, \quad i = 1, \dots, n, \quad (2.3)$$

sendo os ε_i e δ_i , $i = 1, \dots, n$, conhecidos como erros de observação, e considerados mutuamente independentes e identicamente distribuídos com média zero e variâncias finitas σ_ε^2 e σ_δ^2 , respectivamente.

A modelagem apresentada pelas Equações (2.1)-(2.3) leva em conta o fato de que toda medida geralmente contém erros, quer sejam provocados pelo instrumento de medição, pelo operador ou por fatores externos (como por exemplo, variação na temperatura). Cheng e Van Ness (1999) mostram exemplos onde variáveis com erro de medida são apresentadas. Almeida *et al.* (2001) demonstram o uso do modelo com erros nas variáveis, mais especificamente o modelo Funcional, que é aplicado a um experimento de calibração, onde se propõem vários estimadores para obtenção do desconhecido valor de U . Almeida (2003) apresenta gráficos de controle aplicados ao Modelo com Erros nas Variáveis, mais especificamente ao Modelo Funcional, onde se propõem estimadores para obtenção do desconhecido valor de U , além de uma metodologia inédita quanto a utilização no Controle Estatístico da Qualidade, combinado com o Modelo de Regressão em variáveis sujeitas à erro de medida.

Outras denominações são utilizadas para o modelo de Moram (1971) apresentado nas Equações (2.1)-(2.3). Sprent (1990), por exemplo, os trata como “regressão com erros em x ” e “regressão com erros nas variáveis”. Atualmente alguns autores utilizam somente a nomenclatura: Modelo Funcional, Modelo Estrutural e Modelo Ultraestrutural, dependendo da forma como a variável U é apresentada.

Hwang (1986) substituindo $X_i = U_i\delta_i$, $i = 1, \dots, n$, na equação (2.3), isto é, os erros deixam de ser aditivos e passam a ser multiplicativos, apresenta o modelo multiplicativo.

Note que se pode reescrever os modelos dados nas Equações (2.1)-(2.3), como

$$Y_i = \alpha + \beta X_i + (\varepsilon_i - \beta\delta_i) \quad (2.4)$$

ou

$$Y_i = \alpha + \beta X_i + \nu_i, \quad (2.5)$$

onde $\nu_i = \varepsilon_i - \beta\delta_i$, $i = 1, \dots, n$.

Verifica-se que este não é um modelo de regressão tradicional, pois X_i é aleatório para qualquer modelo com erros, e também X está correlacionado com o erro ν_i , isto é, $Cov(X, \nu) = -\beta\sigma_\delta^2$. Caso sejam utilizadas estimativas de regressão ordinárias (mínimos quadrados) em variáveis com erros de medida, isto implicará em estimativas inconsistentes. Como comprovação desta afirmação, são necessárias algumas definições básicas da teoria assintótica, que podem ser encontradas em livros textos de Estatística, como por exemplo, Leite e Singer (1990) e James (1996), destacando-se

Definição 2.1 - Uma seqüência de variáveis aleatórias $\{X_n\}$, $n = 1, 2, 3, \dots$, converge em probabilidade para uma variável aleatória X , se para todo número real $\varepsilon > 0$,

$$P(|X_n - X| \geq \varepsilon) \longrightarrow 0; \quad (2.6)$$

quando $n \longrightarrow \infty$. Notação: $X_n \xrightarrow{p} X$ ou $P_{\lim} X_n = X$.

Alternativamente, a Equação (2.6) pode ser expressa como

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1 \quad (2.7)$$

ou

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0. \quad (2.8)$$

Essencialmente, $X_n \xrightarrow{p} X$ significa que, para todo “ n ” suficientemente grande, X_n e X são aproximadamente iguais com alta probabilidade.

Definição 2.2 - Estabelecendo $\hat{\theta}_n$ um estimador de um parâmetro, θ , baseado sobre uma amostra de tamanho n , isto é, $\hat{\theta}_n = \hat{\theta}_n(x_1, \dots, x_n)$. Quando o tamanho da amostra aumenta, a seqüência de estimadores, $\{\hat{\theta}_n\}$, é dita ser *consistente* se para qualquer constante $\varepsilon > 0$, tem-se que

$$P_\theta(|\hat{\theta}_n - \theta| > \varepsilon) \longrightarrow 0 \quad (2.9)$$

com $n \longrightarrow \infty$.

Segundo Almeida, (2003), Cap. 2, se θ é um vetor, então $|\hat{\theta}_n - \theta|$ é substituído por $\|\hat{\theta}_n - \theta\|$. Por conseguinte, um estimador é dito ser consistente, quando o tamanho da amostra cresce para ∞ , a estimativa converge em probabilidade para o verdadeiro parâmetro. Geralmente, considera-se consistência uma exigência mínima (condição necessária) para um estimador.

No caso da estimação de algum parâmetro de interesse a consistência implica uma crescente proximidade entre os valores do estimador e o verdadeiro valor do parâmetro com o aumento do tamanho amostral. Esta proximidade pode ser definida como uma distância conveniente entre o estimador e o parâmetro, o conceito de consistência está relacionado com a forma segundo a qual essa distância converge para zero, quando o tamanho da amostra cresce. “O mínimo que se pode exigir de uma estatística empregada em alguma regra de decisão é que suas flutuações aleatórias diminuam à medida que o tamanho amostral aumente ou, em outras palavras, que ela seja consistente” (Leite e Singer, 1990).

Portanto, no modelo de regressão com erros nas variáveis, as estimativas de α e β obtidas do modelo de regressão clássico, por mínimos quadrados dadas por

$$\hat{\alpha}_q = \bar{Y} - \hat{\beta}_q \bar{X} \quad (2.10)$$

e

$$\begin{aligned} \hat{\beta}_q &= \frac{S_{XY}}{S_{XX}} \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ &= \frac{\frac{\sum_{i=1}^n X_i Y_i}{n} - \left(\frac{\sum_{i=1}^n X_i}{n}\right) \left(\frac{\sum_{i=1}^n Y_i}{n}\right)}{\frac{\sum_{i=1}^n X_i^2}{n} - \left(\frac{\sum_{i=1}^n X_i}{n}\right)^2} \end{aligned} \quad (2.11)$$

são inconsistentes.

Para provar a inconsistência de (2.11), deve-se considerar o modelo definido pelas Equações (2.1)-(2.3), com a suposição de que a seqüência independente $\{U_i\}$, $i=1,2,\dots$ dos parâmetros incidentais (os valores de U_i são os próprios parâmetros associados a i -ésima observação) é tal que a seqüência de variáveis aleatórias independentes

$$\lim_{n \rightarrow \infty} \bar{U} = \mu < \infty, \quad (2.12)$$

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n U_i^2}{n} = \gamma < \infty \quad (2.13)$$

e

$$\lim_{n \rightarrow \infty} \left[\frac{\sum_{i=1}^n U_i^2}{n} - \frac{(\sum_{i=1}^n U_i)^2}{n} \right] = \eta < \infty, \quad (2.14)$$

onde $\bar{U} = \frac{\sum_{i=1}^n U_i}{n}$, com σ_δ^2 conhecido e sejam X_1Y_1, X_2Y_2, \dots variáveis aleatórias independentes com $E(X_iY_i) = \alpha U_i + \beta U_i^2$ e $Var(X_iY_i) = \sigma_\delta^2(\alpha^2 + 2\alpha\beta U_i + \beta^2 U_i^2 + \sigma_\varepsilon^2) + U_i^2 \sigma_\varepsilon^2$, $i = 1, 2, \dots$, de modo que $\sum_{n=1}^{\infty} \frac{Var(X_nY_n)}{n} < \infty$, e assintoticamente garante-se que $\omega = \frac{\sum_{i=1}^n X_i^2}{n} - \sigma_\delta^2$. Então pela lei forte dos grandes números de Kolmogorov, conclui-se que

$$\frac{\sum_{i=1}^n X_iY_i}{n} - \alpha \frac{\sum_{i=1}^n U_i}{n} - \beta \frac{\sum_{i=1}^n U_i^2}{n} \xrightarrow{p} 0. \quad (2.15)$$

Utilizando-se as Equações (2.12)-(2.14), juntamente com os resultados apresentados no Apêndice A.2, tem-se que

$$\frac{\sum_{i=1}^n X_iY_i}{n} \xrightarrow{p} \alpha\mu + \beta\gamma. \quad (2.16)$$

De forma similar mostra-se que

$$\frac{\sum_{i=1}^n Y_i}{n} \xrightarrow{p} \alpha + \beta\mu, \quad (2.17)$$

$$\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{p} \mu \quad (2.18)$$

e

$$\frac{\sum_{i=1}^n X_i^2}{n} \xrightarrow{p} \omega + \sigma_\delta^2. \quad (2.19)$$

Como

$$\frac{\sum_{i=1}^n X_i Y_i}{n} - \left(\frac{\sum_{i=1}^n X_i}{n} \right) \left(\frac{\sum_{i=1}^n Y_i}{n} \right) \xrightarrow{p} \alpha\mu + \beta\omega - (\alpha + \beta\mu)\mu \quad (2.20)$$

e

$$\frac{\sum_{i=1}^n X_i^2}{n} - \left(\frac{\sum_{i=1}^n X_i}{n} \right)^2 \xrightarrow{p} \omega + \sigma_\delta^2 - \mu^2, \quad (2.21)$$

portanto

$$S_{XY} \xrightarrow{p} \beta\ell \quad (2.22)$$

e

$$S_{XX} \xrightarrow{p} \sigma_\delta^2 + \ell. \quad (2.23)$$

Logo

$$\left(\hat{\beta}_q = \frac{S_{XY}}{S_{XX}} \right) \xrightarrow{p} \beta \left(\frac{\ell}{\ell + \sigma_\delta^2} \right), \quad (2.24)$$

onde $\ell = \frac{\sum_{i=1}^n (U_i - \bar{U})^2}{n-1}$, o que equivale a dizer que

$$\lim_{n \rightarrow \infty} (\hat{\beta}_q - \beta) \neq 0,$$

isto é $\hat{\beta}_q$ é inconsistente para estimar β , conforme se quer provar. De forma similar, prova-se a inconsistência de $\hat{\alpha}_q$.

Por conseguinte, visto que, as estimativas de mínimos quadrados não são consistentes, elas não são consideradas estimativas razoáveis para os parâmetros do modelo com erros nas variáveis. Assim, faz-se necessário garantir outra propriedade estatística desejável para os estimadores de um parâmetro populacional da regressão, a de que eles sejam não-viesados.

Definição 2.3 - Denota-se $\hat{\theta}_n$ um estimador de um parâmetro, θ , baseado em uma amostra de tamanho n , isto é, $\hat{\theta}_n = \hat{\theta}_n(x_1, \dots, x_n)$. O estimador $\hat{\theta}_n$ é dito um estimador **não-viesado** (não-tendencioso ou imparcial) do parâmetro populacional de θ se $E(\hat{\theta}_n) = \theta$, para todo θ .

Pode-se afirmar que, um estimador é dito ser não viciado, quando a esperança (ou média) deste estimador é igual ao seu parâmetro. É importante ressaltar que o estimador $\hat{\theta}_n$ é uma variável aleatória, isto é, ele representa uma dada fórmula de cálculo que fornecerá valores que podem variar de acordo com a amostra selecionada.

Segundo Almeida, 2003, pg. (30)-(31), para o modelo definido pelas Equações (2.1) a (2.3), os estimadores de mínimos quadrados dados por (2.10) e (2.11) são viesados. O estimador em (2.11) é função de (S_{XY}, S_{XX}) . Portanto, pode-se expandir $\hat{\beta}_q$ em série de Taylor até a ordem n^{-2} (ver, Apêndice B), em torno do ponto $a = (E(S_{XY}), E(S_{XX}))$, obtendo-se assim a seguinte expansão

$$\begin{aligned} \hat{\beta}_q &= f(a) + \frac{(S_{XY} - E(S_{XY}))}{1!} \frac{\partial \hat{\beta}_q}{\partial S_{XY}} \Big|_a + \frac{(S_{XX} - E(S_{XX}))}{1!} \frac{\partial \hat{\beta}_q}{\partial S_{XX}} \Big|_a \\ &\quad + \frac{(S_{XY} - E(S_{XY}))^2}{2!} \frac{\partial^2 \hat{\beta}_q}{\partial S_{XY}^2} \Big|_a + \frac{(S_{XX} - E(S_{XX}))^2}{2!} \frac{\partial^2 \hat{\beta}_q}{\partial S_{XX}^2} \Big|_a \\ &\quad + (S_{XY} - E(S_{XY}))(S_{XX} - E(S_{XX})) \frac{\partial^2 \hat{\beta}_q}{\partial S_{XY} \partial S_{XX}} \Big|_a + O_p(n^{-2}), \end{aligned} \quad (2.25)$$

em que

$$f(a) = \frac{S_{XY}}{S_{XX}} \Big|_a = \frac{E(S_{XY})}{E(S_{XX})} = \frac{\beta\ell}{\sigma_\delta^2 + \ell}, \quad (2.26)$$

$$\frac{\partial \hat{\beta}_q}{\partial S_{XX}} \Big|_a = -\frac{\beta\ell}{(\ell + \sigma_\delta^2)^2} + O(n^{-1}), \quad (2.27)$$

$$\frac{\partial^2 \hat{\beta}_q}{\partial S_{XX}^2} \Big|_a = \frac{2\beta\ell}{(\ell + \sigma_\delta^2)^3} + O(n^{-1}), \quad (2.28)$$

$$\frac{\partial \hat{\beta}_q}{\partial S_{XY}} \Big|_a = -\frac{1}{(\ell + \sigma_\delta^2)} + O(n^{-1}), \quad (2.29)$$

$$\frac{\partial^2 \hat{\beta}_q}{\partial S_{XY}^2} \Big|_a = 0 \quad (2.30)$$

e

$$\frac{\partial^2 \widehat{\beta}_q}{\partial S_{XX} \partial S_{XY}} \Big|_a = -\frac{1}{(\ell + \sigma_\delta^2)^2} + O(n^{-1}), \quad (2.31)$$

onde $\ell = \frac{\sum_{i=1}^n (U_i - \bar{U})^2}{n-1}$, σ_δ^2 , representa a variância de δ_i , $i = 1, \dots, n$, e

$$S_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}, \quad (2.32)$$

$$S_{XX} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \quad (2.33)$$

e

$$S_{YY} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}. \quad (2.34)$$

Logo, obtendo-se as esperanças dos momentos amostrais (ver Apêndice A.3), e aplicando-se a esperança na expansão, tem-se

$$E(\widehat{\beta}_q) = \frac{\beta \ell}{(\ell + \sigma_\delta^2)} + \frac{\beta \ell}{(\ell + \sigma_\delta^2)^3} \text{Var}(S_{XX}) - \frac{\text{Cov}(S_{XY}, S_{XX})}{(\ell + \sigma_\delta^2)^2} + O(n^{-2}). \quad (2.35)$$

Substituindo na Equação (2.35), as equações da variância e covariância dos momentos amostrais fornecidas no Apêndice A.3, chega-se a

$$E(\widehat{\beta}_q) = \beta \left(\frac{\ell}{\ell + \sigma_\delta^2} \right) + O(n^{-1}). \quad (2.36)$$

É possível verificar no Apêndice B, um estudo sobre as ordens de grandeza de seqüências de números reais, ou seja, os termos $O(\cdot)$ e $o(\cdot)$. Note que $E(\widehat{\beta}_q) \neq \beta$. Portanto, $\widehat{\beta}_q$ é assintoticamente um estimador viesado de β , conforme se quer demonstrar.

Definição 2.4 - Seja $\{X_n, n = 1, 2, \dots\}$ uma seqüência de variáveis aleatórias com função de distribuição F_n e seja X uma variável aleatória com função de distribuição F . Diz-se que X_n converge em distribuição para X se

$$\lim_{n \rightarrow \infty} F_n(X) = F(X), \quad (2.37)$$

em todos os pontos de continuidade de F .

Notação: $X_n \xrightarrow{d} X$.

Para Almeida (2003) a importância desta definição é caracterizada devido a necessidade de se estudar o comportamento da distribuição dos estimadores em questão, à medida que o tamanho da amostra aumenta, possibilitando com isto a construção de intervalos de confiança aproximados para o parâmetro. Porém, ressalta-se que no modelo em questão as variáveis U_i podem apresentar-se de duas formas diferentes: ou os valores destas variáveis são fixados como um conjunto de constantes ou supõe-se que sejam variáveis aleatórias independentes com alguma distribuição de probabilidade.

2.1.1 - Modelo Funcional

Cheng e Van Ness (1999) mostram que o modelo funcional caracteriza-se por considerar os U_i como constantes, resultando com isto em uma situação onde o número de parâmetros aumenta com o tamanho da amostra. Os U_i são denominados parâmetros incidentais e os demais parâmetros do modelo são chamados de “estruturais”. Conseqüentemente, neste modelo têm-se $n + 4$ parâmetros ao todo, sendo eles $\alpha, \beta, \sigma_\varepsilon^2, \sigma_\delta^2$ e U_i , com $i = 1, \dots, n$.

Considerando a suposição adicional de normalidade, o logaritmo (θ) da correspondente função de verossimilhança, segundo Bolfarine *et al.* (1992), é dado por

$$\theta = -\frac{n}{2} \log \sigma_\varepsilon^2 - \frac{n}{2} \log \sigma_\delta^2 - \frac{1}{2\sigma_\delta^2} \sum_{i=1}^n (X_i - U_i)^2 - \frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n (Y_i - (\alpha + \beta U_i))^2, \quad (2.38)$$

e utilizando a notação de Kendall e Stuart (1979), tem-se

$$-\frac{2}{n} \theta = \log \sigma_\delta^2 + \log \sigma_\varepsilon^2 + \frac{S_1}{\sigma_\delta^2} + \frac{S_2}{\sigma_\varepsilon^2}, \quad (2.39)$$

em que

$$S_1 = \frac{1}{n} \sum_{i=1}^n (X_i - U_i)^2 \quad (2.40)$$

e

$$S_2 = \frac{1}{n} \sum_{i=1}^n (Y_i - (\alpha + \beta U_i))^2. \quad (2.41)$$

Pode-se reescrever (2.39) como

$$-\frac{2}{n} \theta = h(\sigma_\delta, S_1) + h(\sigma_\varepsilon, S_2), \quad (2.42)$$

com $h(a, b) = \log a + (b/a)$. Vê-se que $h(a, 0) = \log(a)$ e quando $a \rightarrow 0$, $h(a, 0) \rightarrow -\infty$.

Aplicando esta idéia em (2.42) verifica-se que se tem $(-2/n)\theta \rightarrow -\infty$, fazendo $X_i = x_i$, com $\sigma_\delta \rightarrow 0$, $S_1 = 0$ e $(-2/n)\theta \rightarrow -\infty$, o que acarreta L (a função de verossimilhança) ser tal que $L \rightarrow \infty$, independentemente do valor de β . Então, não existe solução de verossimilhança para β . Esta situação foi estudada por Solari (1969), que mostrou que o ponto crítico da função de verossimilhança é, na verdade, um ponto de sela, e que somente com suposições adicionais feitas sobre os parâmetros do modelo a estimação por verossimilhança será possível.

As suposições mais comuns adotadas para contornar o problema da estimação de β no modelo funcional são

- (i) σ_ε^2 ou σ_δ^2 conhecido;
- (ii) σ_ε^2 e σ_δ^2 conhecidos;
- (iii) a razão $\frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$ conhecida;
- (iv) o α conhecido.

De acordo com Barnett (1969), a suposição (ii) recai na suposição (iii) na estimação de máxima verossimilhança dos parâmetros.

2.1.2 - Modelo Estrutural

Para Cheng e Van Ness (1999), a principal característica que o modelo estrutural apresenta é o fato de que os U_i são considerados variáveis aleatórias independentes entre si e também de ε_i e δ_i , para todo $i = 1, 2, \dots, n$. Além disso, elas devem possuir uma determinada função de densidade de probabilidade (distribuição), não necessariamente a distribuição normal, com média μ_u e variância σ_u^2 . Portanto, têm-se seis parâmetros no modelo $\alpha, \beta, \mu_u, \sigma_\varepsilon^2, \sigma_\delta^2$ e σ_u^2 .

Considerando válida a suposição de normalidade de U_i , ou seja, $U_i \sim N(\mu_u, \sigma_u^2)$, $i = 1, \dots, n$, tem-se que

$$\begin{pmatrix} U_i \\ \varepsilon_i \\ \delta_i \end{pmatrix} \sim N_3 \left(\begin{pmatrix} \mu_u \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & \sigma_\delta^2 \end{pmatrix} \right), i = 1, \dots, n, \quad (2.43)$$

onde N_3 denota a distribuição normal trivariada.

Note que o par (X_i, Y_i) possui distribuição normal bivariada com vetor de médias e matriz de covariância dados por

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2 \left[\begin{pmatrix} \mu_u \\ \alpha + \beta\mu_u \end{pmatrix}, \begin{pmatrix} \sigma_u^2 + \sigma_\delta^2 & \beta\sigma_u^2 \\ \beta\sigma_u^2 & \beta^2\sigma_u^2 + \sigma_\varepsilon^2 \end{pmatrix} \right]. \quad (2.44)$$

Segundo Bolfarine *et al.* (1992), tem-se aqui um modelo não identificável, já que possui seis parâmetros no modelo $(\alpha, \beta, \mu_u, \sigma_\varepsilon^2, \sigma_\delta^2$ e $\sigma_u^2)$ e o conjunto de estatísticas suficientes associado à distribuição normal bivariada (X, Y) estima os cinco parâmetros $E(X)$, $E(Y)$, $Var(X)$, $Var(Y)$ e $Cov(X, Y)$. Neste caso também, os estimadores de máxima verossimilhança não podem ser obtidos, pois como o modelo é não identificável, dois vetores de parâmetros diferentes podem levar ao mesmo valor da distribuição conjunta de (X, Y) (Fuller, 1987, pg. 30-34). Nesta situação, é impossível estimar consistentemente os parâmetros destes dados, pois o limite de um estimador consistente tem que ser único.

Em virtude disto, surge a necessidade de reduzir o número de parâmetros desconhecidos, colocando alguma suposição adicional, quando considera-se, por exemplo, uma das variâncias conhecidas (σ_u^2 ou σ_ε^2), ou as duas conhecidas (σ_u^2 e σ_ε^2), ou ainda, a razão das variâncias $\frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$, conhecida. Fuller (1987, pg. 5-9) mostra que o conhecimento do coeficiente de atenuação, $K_u = \frac{\sigma_u^2}{(\sigma_u^2 + \sigma_\delta^2)}$, também torna o modelo estrutural identificável. Este coeficiente representa o quanto da variabilidade de X_i é explicada pela variabilidade de U_i . Caso a suposição adicional não seja feita, a estimação dos parâmetros só é possível com a replicação de observações, como é visto na literatura em Bolfarine *et al.* (1992).

2.1.3 - Modelo Ultraestrutural

Dolby (1976) considera o modelo dado pelas Equações (2.1)-(2.3) com as suposições de que $\delta_i \sim N(0, \sigma_\delta^2)$, $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ e $U_i \sim N(\mu_i, \sigma_u^2)$, $i = 1, \dots, n$, como sendo o Modelo Linear Ultraestrutural. Este modelo é na verdade uma generalização dos Modelos Estrutural e Funcional apresentados por Kendall e Stuart (1979). Se as médias dos U_i forem constantes, ou seja, $\mu_1 = \dots = \mu_n$, o Modelo Ultraestrutural se reduz ao Modelo Estrutural, visto na Subseção 2.1.2, enquanto que se não existir variabilidade entre os U_i , ou seja, $\sigma_u^2 = 0$, o Modelo Ultraestrutural se reduz ao Modelo Funcional, visto na Subseção 2.1.1.

E ainda, pode-se verificar que os parâmetros do Modelo Ultraestrutural também são não identificáveis, de modo que suposições adicionais são necessárias para se obter um modelo identificável. Os procedimentos mais comumente utilizados consistem em supor que

- a) As razões das variâncias dos erros, $\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$ e $v = \frac{\sigma_u^2}{\sigma_\delta^2}$ são conhecidas;
- b) A variância dos erros (σ_ε^2 ou σ_δ^2), associados a variável X_i ou Y_i é conhecida;
- c) O intercepto α é conhecido.

Os procedimentos mencionados anteriormente serão levados em consideração para o caso de λ ser conhecido, porém v (item (a)) ser desconhecido. No caso em que σ_δ^2 é conhecido, os estimadores de máxima verossimilhança de α e β serão inconsistentes e não podem ser obtidos de forma explícita, conforme é mostrado em Dolby (1976). Contudo, a inconsistência dos estimadores de máxima verossimilhança de α e β não significa que estimadores consistentes não possam ser obtidos.

A solução mais comum é apresentada por Cheng e Van Ness (1991 e 1994), que afirmam que estes estimadores podem ser substituídos por um estimador obtido pelo método dos momentos. Os casos especiais do Modelo Ultraestrutural (Estrutural e Funcional), podem ser encontrados em várias literaturas, das quais destacam-se: Cheng e Van Ness (1999), Fuller (1987), Lindley (1947), Birch (1964), Barnett (1967), Solari (1969), Moberg e Sundberg (1978) e Gleser (1985). Visto que “ λ ” é conhecido pode-se aplicar o seguinte teorema:

Teorema: 2.1 Seja um Modelo Ultraestrutural com “ λ ” conhecido e μ^* e $s_{\mu\mu}^*$ são verdadeiras. Se $\sigma_u^2 + s_{yy}^* > 0$, quando os estimadores $\hat{\beta}$, $\hat{\alpha}$, e $\hat{\sigma}_\delta^2$ são consistentes, assintoticamente normais (para um “ n ” suficientemente grande a distribuição de probabilidade desses estimadores tende a um modelo de probabilidade normal), com uma covariância e distribuição assintótica de $n^{1/2}[(\hat{\alpha} - \alpha), (\hat{\beta} - \beta), (\hat{\sigma}_\delta^2 - \sigma_\delta^2)]'$ igual para

$$\Gamma(\hat{\alpha}, \hat{\beta}, \hat{\sigma}_\delta^2) \equiv \begin{bmatrix} \theta + (\mu^*)^2\phi & -\mu^*\phi & 0 \\ -\mu^*\phi & \phi & 0 \\ 0 & 0 & 2\sigma_\delta^4 \end{bmatrix}, \quad (2.45)$$

onde $\theta = \sigma_\varepsilon^2 + \beta^2\sigma_\delta^2$, $\phi = \tau^{-2}\{\theta(\tau + \sigma_\delta^2) - \beta^2\sigma_\delta^4\}$, $\tau = \sigma^2 + s_{\mu\mu}^*$,

$$\hat{\beta} = \frac{s_{yy} - \lambda s_{xx} + \{(s_{yy} - \lambda s_{xx})^2 + 4\lambda s_{xy}^2\}^{1/2}}{2s_{xy}}, \quad (2.46)$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}, \quad (2.47)$$

$$\lambda \equiv \frac{\sigma_\varepsilon^2}{\sigma_\delta^2} \quad (2.48)$$

e

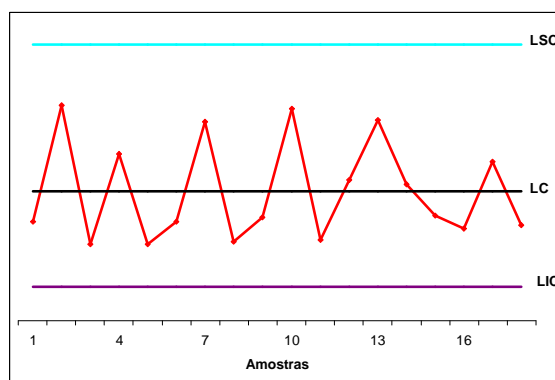
$$\hat{\sigma}_\delta^2 = \frac{2n}{n-2} \hat{\sigma}_\delta^2 = \frac{n}{(n-2)(\lambda + \hat{\beta}^2)} (s_{yy} - 2\hat{\beta}s_{xy} + \hat{\beta}^2 s_{xx}). \quad (2.49)$$

A prova desse Teorema pode ser verificada em Fuller, 1987, pg. (32)-(34).

2.2 - Visão Geral sobre Gráficos de Controle

Com a intenção de eliminar variações anormais em um determinado processo produtivo, W. A. Shewhart, em 1924, propôs originalmente o gráfico de controle. Um gráfico de controle consiste de três linhas paralelas: uma linha média que reflete o nível de operação do processo e duas linhas extremas denominadas limite superior de controle (*LSC*) e limite inferior de controle (*LIC*), obtidas em função da medida de variabilidade de alguma variável do processo (Shewhart, 1931).

Gráficos de controle são utilizados para obter-se um estado de controle estatístico e também para monitorar processos. Estes fatos aliados a simplicidade de construção tornam o gráfico de controle uma das ferramentas mais utilizadas do Controle Estatístico da Qualidade (*CEQ*). Para Montgomery (2001), um processo só está sob controle estatístico quando as causas especiais de variação são eliminadas do processo, e os pontos plotados do gráfico de controle permanecem dentro dos limites de controle. A Figura 2.1 apresenta um exemplo de um processo sob controle estatístico.

Figura 2.1 - *Gráfico de Controle*

2.2.1 - A Relação entre Teste de Hipótese e Gráfico de Controle

Somente quando este gráfico indicar que existe um estado de controle estatístico, a investigação pode começar com alguma garantia de que dados subsequentes serão confiáveis. Caso contrário, pode-se terminar com dados que são inúteis por serem contaminados com os efeitos de causas especiais que emanam do equipamento de medida (Nelson, 2002). Portanto, pode-se constatar que o principal objetivo de se aplicar um gráfico de controle é sugerir áreas de investigações. Para aplicar um gráfico de controle como um teste de hipóteses, as hipóteses nula e alternativa são definidas da seguinte maneira:

H_0 : O processo está sob controle estatístico

H_A : O processo não está sob controle estatístico.

Quando um valor observado da amostra é registrado dentro dos limites de controle, não se pode rejeitar H_0 (hipótese nula), e nenhuma ação corretiva deve ser adotada, pois o processo está sob controle estatístico. Por outro lado, quando um valor observado da amostra é registrado fora dos limites de controle, rejeita-se a hipótese nula, e uma ação corretiva deve ser registrada, visto que, o processo encontra-se fora de controle estatístico.

A região de não rejeição de H_0 será composta pelo conjunto de todos os resultados de uma estatística $g(x)$ que estejam dentro dos limites de controle. Ao se fazer uso do gráfico de controle, pode-se cometer duas espécies de erros, conhecidos na literatura como erro tipo I (α) e erro tipo II (β).

Formalmente define-se

$$(a) P(g(x) \notin (\text{LSC}, \text{LIC}) \mid \text{Processo está sob controle estatístico}) = \alpha;$$

$$(b) P(g(x) \in (\text{LSC}, \text{LIC}) \mid \text{Processo não está sob controle estatístico}) = \beta.$$

Ao se concluir que o processo está sob controle, quando ele não está, situação (b), comete-se o erro tipo II (β), também conhecido como erro negligenciado, e esse erro aumenta proporcionalmente à amplitude dos limites de controle.

Os dois erros podem trazer sérios problemas ao processo em estudo. Na situação (a), interromper um processo para identificar as causas de variações na qualidade quando essas não existem. E na situação (b) a não constatação de que o processo está fora de controle estatístico pode levar a perda de tempo, material e dinheiro.

Existe um terceiro instrumento utilizado para descrever o desempenho dos gráficos de controle denominado de função *ARL* (“Average Run Length”). Define-se o *ARL*, como sendo o número de períodos (ou tempo médio) necessário, até que eventualmente possa ser vista uma observação fora dos limites de controle, isto é, uma alteração no processo. Geralmente, deseja-se que o *ARL* seja grande quando nenhuma causa especial tenha ocorrido e seja pequeno caso o processo apresente problemas. A probabilidade de não se detectar uma alteração no processo decresce à medida que se aumenta o tamanho da amostra.

Em outras palavras, ao concluir que o processo está fora de controle estatístico, quando, de fato, ele está sob controle, situação (a), comete-se o erro tipo I, com probabilidade α , também conhecido como alarme falso. A probabilidade α de um ponto cair fora dos limites de controle 3σ é 0,27%. Quando a distância entre os limites de controle diminui, existe uma probabilidade maior de aumentar o número de alarmes falsos. Partindo-se da idéia estatística de que todo teste de hipótese pode ser completamente caracterizado através da função poder ou alternativamente pela curva característica de operação (*CCO*), geralmente utiliza-se uma destas funções para descrever o desempenho dos gráficos de controle.

2.2.2 - Tipos de Gráficos de Controle

Almeida (2003) mostra que há duas classes principais de gráficos de controle: a de controle por variáveis e a de controle por atributos. Os gráficos de controle por variáveis exigem medições em uma escala contínua, tais como: comprimento, peso, pH e resistência. Já os gráficos de controle por atributos, exigem somente uma classificação de medições categorizadas, tais como, sim ou não, conforme ou não conforme, defeituoso ou não defeituoso. Os dados oriundos de variáveis contínuas geralmente contém mais informações que os dados oriundos de atributos. Conseqüentemente, dados contínuos são preferidos no Controle Estatístico da Qualidade e são essenciais para o diagnóstico de um processo.

2.2.3 - Gráfico de Controle por Variável

De acordo com Siqueira (1997), o uso dos gráficos de controle por variável tem a finalidade de fornecer informações para a melhoria da qualidade, sobre a capacidade do processo, para a tomada de decisões relativas à especificação do produto, para tomada de decisões sobre o processo de produção e para a tomada de decisões sobre peças recém-produzidas. Os gráficos de controle por variável são extensivamente utilizados nas indústrias brasileiras.

Para construir um gráfico de controle, necessita-se seguir os seguintes passos:

1. Selecionar a característica da qualidade (variável) a ser controlada;
2. Definir o método de amostragem e o tamanho da amostra;
3. Coletar os dados;
4. Determinar o valor central e os limites de controle;

Os gráficos de controle para a média (\bar{X}) e amplitude (R) são os gráficos por variáveis mais utilizados na prática por sua simplicidade algébrica. Assim, suponha que uma variável que seja normalmente distribuída com média μ e desvio padrão σ . Seleciona-se uma amostra de tamanho n desta população, então a média desta amostra é dada por

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}, \quad (2.50)$$

Da Inferência Estatística sabe-se que \bar{X} tem distribuição normal com média μ e desvio padrão $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. Existe uma probabilidade dada por $1 - \alpha$ de que qualquer média amostral fique contida dentro do intervalo

$$\mu + (z_{\alpha/2}) \frac{\sigma}{\sqrt{n}} \quad (2.51)$$

e

$$\mu - (z_{\alpha/2}) \frac{\sigma}{\sqrt{n}} \quad (2.52)$$

Pode-se usar as Equações (2.51) e (2.52) como limite superior e inferior sobre um gráfico de controle para a média amostral, bastando para isso substituir $z_{\alpha/2}$ por $k = 1, 2$ ou 3 . E se uma média amostral cair fora desses limites, isto é uma indicação de que a média do processo é diferente de μ (Montgomery, 2001). Nesses intervalos, a probabilidade de um certo ponto cair fora dos limites é de 0,0027 (ou 0,00135 de cada lado), isto é, se X tem distribuição normal, tem-se a chance de 27 em 10.000 dos valores observados caírem fora dos limites estabelecidos. Sendo assim, tem-se

$$P(\bar{X} < \mu - 3\sigma_{\bar{x}}) = P(\bar{X} > \mu + 3\sigma_{\bar{x}}) = 0,00135 \quad (2.53)$$

ou

$$P(\mu - 3\sigma_{\bar{x}} < \bar{X} < \mu + 3\sigma_{\bar{x}}) = 0,9973. \quad (2.54)$$

A interpretação desse intervalo é que 99,73% dos itens distribuem-se em torno da média, isto é, espera-se obter itens fora desse intervalo somente 3 vezes em cada mil observações, ou ainda, para cada 1000 intervalos construídos dessa mesma maneira, 997 irão conter o verdadeiro valor da média.

2.2.4 - Idéia Geral do Gráfico de Controle de Regressão

Suponha a situação onde os pontos “plotados” em um gráfico apresentam-se com certa tendência ascendente ou descendente, isto pode ser atribuído a uma tendência normal do processo ou anormal de variação, tal como, por exemplo, é apresentado na Figura 2.2, que ilustra o desgaste de uma ferramenta de corte (a fabricação de produtos freqüentemente envolve processos que variam de dimensão de acordo com o número de unidades produzidas).

À medida que se aumenta a quantidade de unidades produzidas, a ferramenta em questão desgasta-se, e o valor por ela medido, também tende a aumentar gradualmente, podendo até alcançar o limite superior de rejeição, quando a ferramenta deve ser substituída.

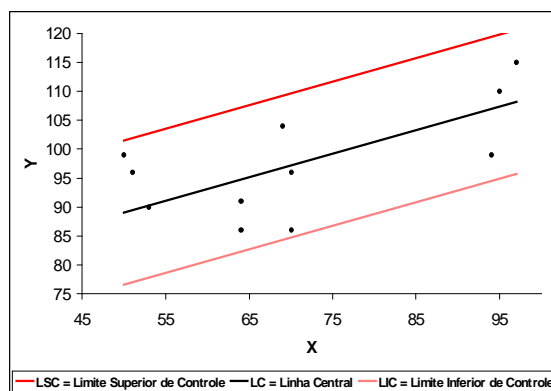


Figura 2.2 - Gráfico de Controle de Regressão

O estudo em detalhes da combinação da técnica convencional de gráficos de controle e a análise de regressão, além da apresentação de várias aplicações em processos administrativos, pode ser verificado em Mandel (1969). Alguns exemplos são a previsão da necessidade de trabalho e a programação de recursos para este trabalho; a medição dos ganhos (ou perdas) em horas trabalhadas; para a formulação de uma melhoria no orçamento anual; a avaliação dos desvios no valor orçado; o estabelecimento de padrões flexíveis de desempenho; e para analisar a eficiência das unidades organizacionais cobertas.

Alwan e Roberts (1988) propõem o uso de um gráfico de controle denominado gráfico de controle de causas especiais (GCCE). Este gráfico utiliza os resíduos do modelo *ARIMA* (Auto Regressivo Integrado Médias Móveis) (p, d, q) ao invés das observações do processo.

Segundo Siqueira (1997), uma vez que o valor central desse gráfico é uma reta inclinada, sua equação deve ser estabelecida usando-se o método dos mínimos quadrados. E de posse da equação da reta que representa o valor central, os limites de controle do gráfico podem ser determinados conforme é feito no gráfico de controle da média.

A discussão sobre o fato de que os limites de controle em torno da linha de regressão podem alargar-se progressivamente quando eles se afastam da média, pode ser visto em Draper e Smith (1998). Bowdem e Graybill (1966 e 1967) apresentam um método para a construção de faixas de confiança sobre um intervalo de comprimento finito para modelos lineares simples. Estas faixas de confiança, com coeficiente de confiança de $1 - \alpha$, são dadas por linhas retas em lugar de curvas e são também paralelas, se a largura da faixa for uniforme, ou trapezoidais em caso contrário. Com base nesta metodologia, Mandel (1969) justifica a utilização dos limites de controle paralelos, ou seja, de largura fixa. Ele ainda afirma que sempre será possível construir limites de controle paralelos, quando a variação sobre a linha de regressão não é muito grande, e também que os limites de controle paralelos irão fornecer um controle mais rígido para as distâncias extremas da média.

Mitra (1988) destaca que o valor central do gráfico de controle de regressão é uma reta inclinada e as linhas dos limites de controle são paralelas à linha de regressão, e os cálculos dos seus limites são um pouco mais complexos e mais demorados que os dos gráficos de controle convencionais.

Posteriormente, no capítulo 5, será discutido mais especificamente este tipo de gráfico, partindo de um exemplo real discutido por Mandel (1969). Todos os gráficos por variáveis apresentados anteriormente são utilizados para detectar grandes variações na média do processo. Entretanto, vale ressaltar que existem outros tipos de gráfico de controle que são utilizados para capturar pequenas mudanças na média do processo, como por exemplo: o gráfico de controle para soma acumulada (*CUSUM - Cumulative Sum Control Chart*) e o gráfico de controle para médias móveis ponderadas exponencialmente (*EWMA - Exponentially Weighted Moving Average Control Chart*).

Capítulo 3

Modelo Estrutural com uma das Variâncias Conhecida

Este capítulo dá ênfase ao estudo do modelo estrutural quando uma das variâncias do erro de medição é conhecida, e os erros de medição associados à variável independente são aditivos e normalmente distribuídos. Também são propostos estimadores consistentes para os parâmetros da regressão e posteriormente apresentar-se-ão os limites de confiança desses parâmetros.

3.1 - Estimadores dos Parâmetros de Regressão

No Capítulo 2 (Subseção 2.1.2), destaca-se que a estimação dos parâmetros da regressão no modelo estrutural só será possível se forem feitas suposições adicionais aos parâmetros. Uma delas é o conhecimento de uma das variâncias do erro, por exemplo σ_δ^2 .

Para Moberg e Sundberg (1978), no caso em que σ_δ^2 é conhecido, os estimadores de máxima verossimilhança de α e β não são consistentes e não são obtidos de forma explícita. Contudo, isso não significa que estimadores consistentes não possam ser obtidos. Cheng e Van Ness (1999) afirmam que no modelo de regressão estrutural a estimação de α e β é realizada, quando a variância dos erros de medição σ_δ^2 é conhecida, adotando-se os estimadores obtidos pelo método dos momentos no próprio modelo estrutural, pois os mesmos são estimadores consistentes.

3.1.1 - Estimadores Consistentes de α e β

Para Cheng e Van Ness (1999), sendo o modelo Estrutural formalmente definido como

$$Y_i = \alpha + \beta U_i + \varepsilon_i, \quad (3.1)$$

onde $U_i = X_i - \delta_i$ e $i = 1, \dots, n$. Sendo os U_i considerados variáveis aleatórias independentes entre si e também de ε_i e de δ_i , para todo $i = 1, \dots, n$; como sendo os erros de medida normais independentes e identicamente distribuídos, ambos com média zero e variâncias constantes σ_ε^2 e σ_δ^2 , respectivamente.

Considerando a suposição de normalidade de U_i , ou seja, $U_i \sim N(\mu_u, \sigma_u^2)$, $i = 1, \dots, n$, tem-se que

$$\begin{pmatrix} U_i \\ \varepsilon_i \\ \delta_i \end{pmatrix} \sim N_3 \left(\begin{pmatrix} \mu_u \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & \sigma_\delta^2 \end{pmatrix} \right), \quad (3.2)$$

onde N_3 denota a distribuição normal trivariada.

A partir do vetor (\mathbf{X}, \mathbf{Y}) dado por

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N_2 \left[\begin{pmatrix} \mu_u \\ \alpha + \beta\mu_u \end{pmatrix}, \begin{pmatrix} \sigma_u^2 + \sigma_\delta^2 & \beta\sigma_u^2 \\ \beta\sigma_u^2 & \beta^2\sigma_u^2 + \sigma_\varepsilon^2 \end{pmatrix} \right], \quad (3.3)$$

e sob as suposições anteriores, ou seja, que σ_δ^2 é conhecido, e $U_i \sim N(\mu_u, \sigma_u^2)$, o sistema de equações, é obtido através do método de estimação de Máxima Verossimilhança, como

$$\begin{cases} \bar{X} &= \hat{\mu}_u; \\ \bar{Y} &= \hat{\alpha} + \hat{\beta}\bar{X}; \\ S_{XX} &= \frac{\hat{\sigma}_u^2}{K_u}; \\ S_{XY} &= \hat{\beta}\sigma_u^2; \\ S_{YY} &= \hat{\beta}^2\hat{\sigma}_u^2 + \hat{\sigma}_\varepsilon^2. \end{cases} \quad (3.4)$$

ou ainda,

$$\begin{cases} S_{YY} = \sigma_Y^2 = \beta^2\sigma_u^2 + \sigma_\varepsilon^2; \\ S_{XY} = \sigma_{XY} = \beta\sigma_u^2; \\ S_{XX} = \sigma_X^2 = \sigma_u^2 + \sigma_\delta^2; \\ \mu_Y = E(Y) = \alpha + \beta\mu_u; \\ \mu_X = E(X) = \mu_u, \end{cases} \quad (3.5)$$

onde $K_u = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\delta^2}$. Assim, a partir das equações (3.4) - (3.5) e supondo $S_{XX} > \sigma_\delta^2$, têm-se os seguintes estimadores

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}; \quad (3.6)$$

$$\hat{\beta} = \frac{S_{XY}}{S_{XX} - \sigma_\delta^2}; \quad (3.7)$$

$$\hat{\sigma}_\varepsilon^2 = \widehat{Var}(\varepsilon) = S_{YY} - \hat{\beta}S_{XY} \quad (3.8)$$

e

$$\hat{\sigma}_u^2 = \widehat{Var}(u) = K_u S_{XX} \Rightarrow \hat{\sigma}_u^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\delta^2} \cdot S_{XX}, \quad (3.9)$$

onde as estimativas amostrais são definidas como

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}); \quad (3.10)$$

$$S_{XX} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2; \quad (3.11)$$

$$S_{YY} = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2; \quad (3.12)$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.13)$$

e

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (3.14)$$

Para que esses estimadores estejam definidos no espaço paramétrico, $\hat{\sigma}_\varepsilon^2$ deve ser positivo. O estimador de σ_ε^2 será positivo se $S_{YY} - \hat{\beta}S_{XY} > 0$, isto é, se $K_u S_{xx} S_{yy} - S_{xy}^2 > 0$, ou ainda,

$$S_{yy}(S_{xx} - \sigma_\delta^2) - S_{xy}^2 > 0. \quad (3.15)$$

Assintoticamente não há restrições, pois $\hat{\sigma}_\varepsilon^2 \rightarrow \sigma_\varepsilon^2$.

Estimadores amostrais onde a condição representada pela Equação (3.15) é violada são $\hat{\sigma}_\varepsilon^2 = 0$, $\hat{\beta} = \frac{S_{yy}}{S_{xy}}$ e $\hat{\alpha}$ definido por $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$.

Supondo que σ_δ^2 , a variância do erro de medida, tenha sido obtida baseado no conhecimento da variabilidade de processos anteriores. No entanto, Fuller (1987) apresenta uma relação para determinar $\hat{\sigma}_\delta^2$, sob a forma $\hat{\sigma}_\delta^2 = S_{xx} - S_{xy}^2 (S_{yy} - \hat{\sigma}_\varepsilon^2)^{-1}$.

Após a fixação de σ_δ^2 , deve-se então definir a equação de regressão estimada, \hat{Y} através da esperança ou média da equação de regressão populacional, Y , utilizando

$$\hat{Y} = \hat{\alpha} + \hat{\beta}\hat{U}_i, \quad i = 1, \dots, n, \quad (3.16)$$

onde os estimadores da regressão, $\hat{\alpha}$ e $\hat{\beta}$ são obtidos através de qualquer um dos métodos de estimação. “No caso do modelo Estrutural com uma das variâncias conhecidas, na presença de normalidade, os métodos de máxima verossimilhança e dos momentos se equivalem”.

Uma outra forma de representar o modelo de regressão com erro nas variáveis é através da substituição de X_i na equação de regressão $X_i = U_i + \delta_i$, obtendo-se, assim, um modelo aparentemente da forma do modelo de regressão clássico com X fixo, representado do seguinte modo

$$Y_i = \alpha + \beta X_i + \nu_i, \quad i = 1, \dots, n, \quad (3.17)$$

onde o erro $\nu_i = \varepsilon_i - \delta_i\beta$.

Portanto, o modelo estimado é

$$\hat{Y} = \hat{\alpha} + \hat{\beta}X, \quad (3.18)$$

onde os estimadores da regressão $\hat{\alpha}$ e $\hat{\beta}$ são obtidos por (3.6) e (3.7), respectivamente. Além disso, a $Cov(x, \nu) = -\beta\sigma_\delta^2$, que é a grande diferença em relação à análise de regressão clássica, na qual a referida covariância é nula. Um estimador não viesado da variância do erro σ_ν^2 é dado por

$$\hat{\sigma}_y^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y})^2}{n - 2} = S_\nu^2, \quad (3.19)$$

ou ainda

$$S_\nu^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y} - (X_i - \bar{X})\hat{\beta})^2}{n - 2}. \quad (3.20)$$

O modelo de regressão com erros nas variáveis Estrutural, caracteriza-se por apresentar um número fixo de parâmetros desconhecidos, ou seja, independe do tamanho da amostra observada. Essa definição é importante, pois geralmente, de forma assintótica os teoremas referentes aos estimadores de máxima verossimilhança possuem vetores de parâmetros, constantes mesmo com o aumento do tamanho amostral. Entretanto, as propriedades dos estimadores de máxima verossimilhança (ver, Stuart e Ord, (1991), Capítulo 18); resultam em que $E(X)$, $E(Y)$, $Var(X)$, $Var(Y)$ e $Cov(X, Y)$ são consistentes e assintoticamente Normais, ou seja, para um valor de n suficientemente grande (tendendo ao infinito), essas estimativas amostrais convergem para os próprios parâmetros populacionais e tendem à normalidade.

Pelas propriedades gerais das estimativas amostrais (ver, Lehmann, (1983), Capítulo 6); as estimativas de máxima verossimilhança de $E(X_i)$, $E(Y_i)$, $Var(X_i)$, $Var(Y_i)$ e $Cov(X_i, Y_i)$ são consistentes e assintoticamente normais, assim têm-se que

$$E(U) = \mu = \mu^*; \quad (3.21)$$

$$S_{\mu\mu}^* = 0; \quad (3.22)$$

$$\theta = \sigma_\varepsilon^2 + \beta^2 \sigma_\delta^2; \quad (3.23)$$

$$\phi = \sigma_u^{-4} \{ \theta(\sigma_u^2 + \sigma_\delta^2) - \beta^2 \sigma_\delta^4 \} = \frac{\theta}{\sigma_u^2} + \frac{\sigma_\varepsilon^2 \sigma_\delta^2}{\sigma_u^4} = \frac{\sigma_\varepsilon^2 \sigma_u^2 + \sigma_\varepsilon^2 \sigma_\delta^2 + \beta^2 \sigma_\delta^2 \sigma_u^2}{\sigma_u^4}, \quad (3.24)$$

onde: $\mu^* = \lim_{n \rightarrow \infty} \bar{\mu}_n = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mu_i$ e $S_{\mu\mu}^* = \frac{1}{n} \sum_{i=1}^n (\mu_i - \bar{\mu}_n)^2$.

Portanto, o vetor de estimativas de máxima verossimilhança implica em

$$n^{1/2} \left[\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\sigma}_\delta^2 \end{pmatrix} - \begin{pmatrix} \alpha \\ \beta \\ \sigma_\delta^2 \end{pmatrix} \right] \xrightarrow{d} N_3(\Gamma), \quad (3.25)$$

onde $n \rightarrow \infty$, N_3 representa a distribuição normal trivariada. \xrightarrow{d} indica uma convergência em distribuição. $N(m, \Gamma)$ representa respectivamente um vetor de médias “ m ” e matriz de covariâncias Γ , dada por

$$\Gamma(\alpha, \beta, \sigma_\delta^2) \equiv \begin{bmatrix} \theta + \mu^2 \phi & -\mu \phi & 0 \\ -\mu \phi & \phi & 0 \\ 0 & 0 & 2\sigma_\delta^4 \end{bmatrix}. \quad (3.26)$$

Portanto, conclui-se que $\hat{\sigma}_\delta^2$ é assintoticamente independente de $\hat{\alpha}$ e $\hat{\beta}$ e a covariância assintótica de $\hat{\alpha}$ e $\hat{\beta}$ é $-\mu\phi$. Conseqüentemente, pode-se mostrar que as estimativas amostrais S_{YY} , S_{XY} e S_{XX} são assintoticamente não viesadas (convergem em média para o próprio parâmetro) para as estimativas populacionais e todas têm variâncias que estão diminuindo a uma taxa n^{-1} .

Fuller (1987) destaca que essas propriedades possibilitam obter a distribuição limite dos estimadores, conforme pode ser verificado a seguir

Seja o modelo definido pela equação $Y_i = \alpha + \beta U_i + \varepsilon_i$, $X_i = U_i + \delta_i$, $i = 1, \dots, n$; com σ_δ^2 conhecido, $\sigma_\varepsilon^2 > 0$ e $\ell > 0$. Então o vetor $n^{1/2}[(\hat{\alpha} - \alpha), (\hat{\beta} - \beta)]$, em que $(\hat{\alpha}, \hat{\beta})$ são definidos pelas Equações (3.6) e (3.7), convergem em distribuição para um vetor de variáveis aleatórias normais com média zero e matriz de covariâncias dada por

$$\Gamma = \begin{bmatrix} \mu_u(\sigma_u^2)^{-2} (S_{XX}\sigma_\nu^2 + \sigma_{X\nu}^2) + \sigma_\nu^2 & -\mu_u(\sigma_u^2)^{-2} (S_{XX}\sigma_\nu^2 + \sigma_{X\nu}^2) \\ -\mu_u(\sigma_u^2)^{-2} (S_{XX}\sigma_\nu^2 + \sigma_{X\nu}^2) & (\sigma_u^2)^{-2} (S_{XX}\sigma_\nu^2 + \sigma_{X\nu}^2) \end{bmatrix}, \quad (3.27)$$

onde $\nu_i = Y_i - \alpha - X_i\beta = \varepsilon_i - \delta_i\beta$ e $\sigma_{X\nu} = \sigma_{\delta\nu} = -\beta\sigma_\delta^2$. Além disso, $n\widehat{V}\{(\hat{\alpha}, \hat{\beta})\}$ converge em probabilidade para Γ , onde

$$\widehat{V}\{(\hat{\alpha}, \hat{\beta})'\} = \begin{bmatrix} \overline{X^2}\widehat{V}\{\hat{\beta}\} + S_\nu^2(n)^{-1} & -\overline{X}\widehat{V}\{\hat{\beta}\} \\ -\overline{X}\widehat{V}\{\hat{\beta}\} & \widehat{V}\{\hat{\beta}\} \end{bmatrix}, \quad (3.28)$$

$$\widehat{Var}(\hat{\alpha}) = n^{-1}S_\nu^2 + \overline{X^2}\widehat{Var}(\hat{\beta}) = \frac{n\overline{X^2}\widehat{Var}(\hat{\beta}) + S_\nu^2}{n} \quad (3.29)$$

e

$$\widehat{Var}(\hat{\beta}) = \frac{(S_{XX}S_\nu^2 + \widehat{\beta}^2\sigma_\delta^4)}{\ell^2(n-1)}, \quad (3.30)$$

onde S_ν^2 é expresso em (3.18),

$$\widehat{Cov}(\hat{\alpha}, \hat{\beta}) = -\overline{X}\widehat{V}\{\hat{\beta}\} \quad (3.31)$$

$$\widehat{\sigma}_{\delta\nu} = -\widehat{\beta}\sigma_\delta^2 \quad (3.32)$$

e

$$\ell = \frac{\sum_{i=1}^n (U_i - \overline{U})^2}{n-1}. \quad (3.33)$$

A normalidade do vetor de erros $(\varepsilon_i, \delta_i)$ permite apresentar expressões simplificadas, para a matriz de covariâncias dos estimadores, em termos do segundo momento da distribuição das variáveis originais. As aproximações baseadas em normalidades resultantes da utilização de distribuições com erros associados convergem fracamente para uma distribuição normal, proveniente da variância dos erros relativamente pequenos para a variância de X .

A variância da distribuição limite de $n^{1/2}(\hat{\beta} - \beta)$ é considerada grande em relação à correspondente variância do coeficiente de regressão de mínimos quadrados ordinários construídos com X_t . A diferença é a fonte de variação encontrada. O divisor, $(S_{XX} - \sigma_\delta^2)$ em (3.7) representa uma razão do estimador de mínimos quadrados ordinários e a variância de ν_t é relativamente grande, tal que, a variância sobre mínimos quadrados ordinários é expressa de forma linear.

Utiliza-se a quantidade $n^{1/2}$ para padronizar o estimador de (3.7) na obtenção da distribuição limite. Prefere-se usar o divisor $(n - 1)$ na estimativa da variância a qual é dada por

$$\widehat{Var}\{\hat{\beta}\} = \ell^{-2}(n - 1)^{-1}(S_{XX}S_\nu^2 + \hat{\beta}^2\sigma_\delta^4), \quad (3.34)$$

quando $\sigma_\delta^2 = 0$ o estimador da variância (3.34) se reduzir ao estimador da variância de mínimos quadrados ordinários. Baseado no mesmo argumento, utiliza-se o divisor $(n - 2)$ na definição do estimador (3.20). Dado $n\widehat{V}(\hat{\alpha}, \hat{\beta})$ o estimador da matriz de covariâncias da distribuição limite de $n^{1/2}(\hat{\alpha} - \alpha, \hat{\beta} - \beta)$. Não se pode considerar tecnicamente correto falar de $\widehat{Var}\{\hat{\beta}\}$, como sendo um estimador da variância de β . Na verdade ele é um estimador da variância da distribuição aproximada de β . E como $n\widehat{V}\{(\hat{\alpha}, \hat{\beta})'\}$ é um estimador consistente de Γ , segue que

$$t = \frac{(\hat{\beta} - \beta)}{(\widehat{V}(\hat{\beta}))^{1/2}}, \quad (3.35)$$

é aproximadamente normal. Na prática parece razoável aproximar a distribuição de “ t ” por uma *t-Student’s* com $n - 2$ graus de liberdade.

Fuller(1987) também destaca que, como as estimativas amostrais convergem em probabilidade para as estimativas populacionais, pode-se expandir $\hat{\beta}$ em uma série de Taylor sobre os valores populacionais, obtendo-se assim resultados similares aos das equações

$$\widehat{Var}(\hat{\alpha}) = \bar{X}^2\widehat{V}\{\hat{\beta}\} + S_\nu^2(n)^{-1}, \quad (3.36)$$

$$\widehat{Var}\{\hat{\beta}\} = \ell^{-2}(n - 1)^{-1}(S_{XX}S_\nu^2 + \hat{\beta}^2\sigma_\delta^4) \quad (3.37)$$

e

$$\widehat{Cov}(\hat{\alpha}, \hat{\beta}) = -\bar{X}\widehat{V}\{\hat{\beta}\}, \quad (3.38)$$

assegurando dessa forma a validade dessas equações.

Pode-se mostrar facilmente que

$$\text{Var}(\bar{X}) = \frac{\sigma_u^2 + \sigma_\delta^2}{n} \quad e \quad \text{Var}(\bar{Y}) = \frac{\beta^2 \sigma_u^2 + \sigma_\varepsilon^2}{n}.$$

3.2 - Estimação do Verdadeiro Valor de U

Considerando as suposições feitas para o modelo Estrutural, definido em (3.1); e seja σ_δ^2 conhecida e os estimadores de α , β e σ_ε^2 dados por (3.6), (3.7) e (3.8), respectivamente. Então, pode-se escrever,

$$\begin{pmatrix} Y_i - \alpha \\ X_i \end{pmatrix} = \begin{pmatrix} \beta \\ 1 \end{pmatrix} U_i + \begin{pmatrix} \varepsilon_i \\ \delta_i \end{pmatrix}, i = 1, \dots, n. \quad (3.39)$$

A estimação de U_i , onde assume-se que os U_i 's são valores de uma amostra aleatória, ou seja, U_i é considerado uma variável aleatória dos valores da população de U . Portanto, ao se assumir que U_i é aleatório, a matriz de variâncias e covariâncias de $(Y_i, X_i$ e $U_i)$ será dada por

$$\Sigma = \begin{bmatrix} \beta^2 \sigma_u^2 + \sigma_\varepsilon^2 & \beta \sigma_u^2 + \sigma_{\varepsilon\delta} & \beta \sigma_u^2 \\ \beta \sigma_u^2 + \sigma_{\varepsilon\delta} & \sigma_u^2 + \sigma_\delta^2 & \sigma_u^2 \\ \beta \sigma_u^2 & \sigma_u^2 & \sigma_u^2 \end{bmatrix} \quad (3.40)$$

Por conseguinte, sob a suposição de normalidade, o valor esperado de U_i , dado Z_i é

$$E\{U_i/Z_i\} = \mu_x + \gamma_2(Y_i - \mu_y) + \gamma_3(X_i - \mu_x) = \gamma_1 + \gamma_2 Y_i + \gamma_3 X_i, \quad (3.41)$$

onde $Z_i = (Y_i, X_i)$, $\gamma_1 = (1 - \gamma_3)\mu_x - \gamma_2 \mu_y$, $\mu_y = \alpha + \beta \mu_x$ e $(\gamma_2, \gamma_3)' = \Sigma_{ZZ}^{-1}(\beta \sigma_u^2, \sigma_u^2)'$.

Usando o fato de que $\Sigma_{ZZ} = (\beta, 1)' \sigma_u^2 (\beta, 1) + \Sigma_{\varepsilon\varepsilon}$, o valor esperado (3.34), pode ser escrito como

$$E\{U_i/Z_i\} = X_i - (Z_i - \mu_Z) \sigma_{ZZ}^{-1} \sigma_{Z\delta}, \quad (3.42)$$

onde $\sigma_{ZZ}^{-1} = \Sigma_{ZZ}^{-1} = \sigma_{Z\delta} = \Sigma_{Z\delta} = \Sigma_{\varepsilon\delta} = (\sigma_{\varepsilon\delta}, \sigma_\delta^2)'$. A variância condicional de U_i é dada por

$$V\{U_i/Z_i\} = \sigma_u^2 - (\gamma_2, \gamma_3)(\beta \sigma_u^2, \sigma_u^2)'. \quad (3.43)$$

É importante ressaltar a possibilidade de comparação entre o estimador

$$\hat{U}_i = a(Y_i - \alpha) + bX_i, \quad (3.44)$$

quando U_i é considerado como sendo uma quantidade fixa (parâmetro), em relação ao preditor (3.43), que é construído sob a suposição de U_i ser aleatório. Em ambos os casos o estimador (preditor) é obtido através da subtração entre um preditor de δ_i pelo de X_i . Se U_i é fixo (constante), o preditor de δ_i é construído, segundo uma variável aleatória ν_i , que contém uma componente não fixa. Se U_i é aleatório, o melhor preditor de δ_i é construído usando o vetor aleatório (Y_i, X_i) .

Considerando U_i como sendo aleatório e substituindo os parâmetros pelos respectivos estimadores, o verdadeiro valor de U_i , dado (Y_i, X_i) é expresso por

$$\tilde{U}_i = X_i - (Z_i - \bar{Z})S_{ZZ}^{-1}(\sigma_{\varepsilon\delta}, \sigma_\delta^2)' = \tilde{\gamma}_1 + \tilde{\gamma}_2 Y_i + \tilde{\gamma}_3 X_i, \quad (3.45)$$

onde $\tilde{\gamma}_1 = (1 - \tilde{\gamma}_3) \bar{X} - \tilde{\gamma}_2 \bar{Y}$, $(\tilde{\gamma}_2, \tilde{\gamma}_3)' = \begin{bmatrix} S_{YY} & S_{YX} \\ S_{XY} & S_{XX} \end{bmatrix}^{-1} \begin{bmatrix} S_{XY} \\ S_{XX} - \sigma_\delta^2 \end{bmatrix}$ e $(S_{ZZ}, S_{ZX}) = (n - 1)^{-1} \sum_{i=1}^n (Z_i - \bar{Z})'(Z_i - \bar{Z}, X_i - \bar{X})$. Ou equivalentemente, pode-se reescrever $(\tilde{\gamma}_2, \tilde{\gamma}_3)'$ de maneira que

$$\begin{bmatrix} \tilde{\gamma}_2 \\ \tilde{\gamma}_3 \end{bmatrix} = \begin{bmatrix} S_{YY} & S_{YX} \\ S_{XY} & S_{XX} \end{bmatrix}^{-1} \begin{bmatrix} S_{XY} \\ S_{XX} - \sigma_\delta^2 \end{bmatrix}. \quad (3.46)$$

Portanto, a obtenção dos estimadores, requer a introdução, no preditor, de um erro adicional com expansão em série de Taylor da ordem de $O_p(n^{-1/2})$. Caso se ignore a componente adicional, um estimador da variância do erro de predição é

$$\widehat{V}\{\tilde{U}_i - U_i/Z_i\} = \sigma_\delta^2 - (\sigma_{\varepsilon\delta}, \sigma_\delta^2)S_{ZZ}^{-1}(\sigma_{\varepsilon\delta}, \sigma_\delta^2)'. \quad (3.47)$$

3.3 - Estimador da Variância de Y ($\hat{\sigma}_y^2$)

Considera-se o modelo definido por (2.1) - (2.4), sendo os $U_i, i = 1, 2, \dots, n$, variáveis aleatórias e os erros de medição $(\varepsilon_i, \delta_i)$, normais, independentes identicamente distribuídos, ambos com média zero e variâncias constantes σ_ε^2 e δ_i , respectivamente. Portanto, o modelo estimado será dado por

$$\widehat{Y} = \widehat{\alpha} + \widehat{\beta}X, \quad (3.48)$$

onde os estimadores da regressão $\widehat{\alpha}$ e $\widehat{\beta}$ são obtidos por (3.6) e (3.7), respectivamente. Além disso a $Cov(x, \nu) = -\beta\sigma_\delta^2$.

Um estimador não viciado da variância do erro σ_ν^2 , segundo Almeida (2003), é dado por

$$\begin{aligned}\widehat{Var}(Y) &= \hat{\sigma}_y^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y})^2}{n-2} = \frac{1}{n-2} \sum_{i=1}^n \left(Y_i - \bar{Y} - (X_i - \bar{X})\hat{\beta} \right)^2 \\ &= S_\nu^2 = \frac{\sum_{i=1}^n \left(Y_i - \bar{Y} - (X_i - \bar{X})\hat{\beta} \right)^2}{n-2}\end{aligned}\quad (3.49)$$

onde $\hat{\beta}$, \bar{X} , e \bar{Y} são obtidos pelas equações (3.7), (3.13) e (3.14), respectivamente.

3.4 - Estimador da Variância de \hat{Y} ($\hat{\sigma}_y^2$)

O modelo definido pela Equação (3.1), sendo os $U_{i's}$, $i = 1, 2, \dots, n$, variáveis aleatórias e os erros de medição $(\varepsilon_i, \delta_i)$, normais, independentes identicamente distribuídos, ambos com média zero e variâncias constantes σ_ε^2 e δ_i , respectivamente. Considerando a suposição de normalidade de U_i , ou seja, $U_i \sim N(\mu_u, \sigma_u^2)$,

$i = 1, 2, \dots, n$, e o vetor (\mathbf{X}, \mathbf{Y}) dado por

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N_2 \left[\begin{pmatrix} \mu_u \\ \alpha + \beta\mu_u \end{pmatrix}, \begin{pmatrix} \sigma_u^2 + \sigma_\delta^2 & \beta\sigma_u^2 \\ \beta\sigma_u^2 & \beta^2\sigma_u^2 + \sigma_\varepsilon^2 \end{pmatrix} \right], \quad (3.50)$$

o modelo estimado é

$$\hat{Y} = \hat{\alpha} + \hat{\beta}\tilde{U}. \quad (3.51)$$

O estimador da variância de (3.51)

$$\hat{\sigma}_y^2 = \widehat{Var}(\hat{Y}_i) = \widehat{Var}(\hat{\alpha} + \hat{\beta}\tilde{U}_i), \quad (3.52)$$

onde, $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$, logo resulta que $\widehat{Var}(\hat{Y}_i) = \widehat{Var}(\bar{Y} - \hat{\beta}\bar{X} + \hat{\beta}\tilde{U}_i)$, segundo Fuller (1987), pode-se garantir que $\hat{\beta}\tilde{U}_i \xrightarrow{d} \hat{\beta}\bar{X}$, o quê resultará no estimador da variância expresso por

$$\widehat{Var}(\hat{Y}_i) = \widehat{Var}(\bar{Y}); \quad (3.53)$$

Finalmente, tem-se que

$$\widehat{Var}(\hat{Y}_i) = \widehat{Var}(\bar{Y})$$

$$\begin{aligned}
&= \text{Var} \left(\frac{\sum_{i=1}^n Y_i}{n} \right) \\
&= \frac{1}{n} \cdot \text{Var}(Y_i) \\
&= \frac{1}{n} (\beta^2 \sigma_u^2 + \sigma_\varepsilon^2). \tag{3.54}
\end{aligned}$$

onde β^2 , σ_u^2 , e σ_ε^2 são obtidos pelo sistema (3.5).

3.5 - Estimador da Variância de \hat{Y}_t ($\hat{\sigma}_{\hat{y}_t}^2$)

Dado o contínuo interesse em prever o valor de uma observação, por exemplo (Y_t), a qual refere-se ao valor de U_t , há necessidade de mensurar a qualidade de tal previsão, em outras palavras, a diferença entre o valor observado Y_i e o valor estimado \hat{Y}_i . Por conseguinte, caracteriza-se então, o erro de previsão \hat{Y}_t . Portanto, o estimador de $Y_i = \alpha + \beta U_i + \varepsilon_i$ é dado por $\hat{Y} = \hat{\alpha} + \hat{\beta} \tilde{U}_i$, a partir do qual pode-se concluir que o erro de previsão é expresso por

$$Y_i - \hat{Y}_i = e_i = Y_i - \bar{Y} - (\hat{Y}_i - \bar{Y}) = Y_i - \hat{Y}_i. \tag{3.55}$$

Porém, $e_i \sim N(0; \sigma_\varepsilon^2) \Rightarrow E(e_i) = E(Y_i - \hat{Y}_i) = 0$. No entanto, $E(e_i)^2 = E(Y_i - \hat{Y}_i)^2 \neq 0$, onde $E(Y_i - \hat{Y}_i)^2 = \widehat{\text{Var}}(e_i) = \widehat{\text{Var}}(Y_i - \hat{Y}_i)$. Portanto, pelas propriedades da variância,

$$\widehat{\text{Var}}(e_i) = \widehat{\text{Var}}(Y_i) + \widehat{\text{Var}}(\hat{Y}_i) - 2\text{Cov}(Y_i, \hat{Y}_i) = \widehat{\text{Var}}(Y_i) + \widehat{\text{Var}}(\hat{Y}_i),$$

visto que, a $\text{Cov}(Y_i, \hat{Y}_i) = 0$. Obtém-se então o estimador da variância de \hat{Y}_t , a partir de

$$\widehat{\text{Var}}(\hat{Y}_t) = \frac{\sum_{i=1}^n \left(Y_i - \bar{Y} - (X_i - \bar{X}) \hat{\beta} \right)^2}{n-2} + \frac{\beta^2 \sigma_u^2 + \sigma_\varepsilon^2}{n}, \tag{3.56}$$

sendo $(n-2)$ os graus de liberdade do modelo, $\hat{\beta}$, \bar{X} , e \bar{Y} sendo obtidos pelas Equações (3.7), (3.13) e (3.14), respectivamente. E β^2 , σ_u^2 , e σ_ε^2 são obtidos pelo sistema de Equações (3.5).

3.6 - Erro Padrão da Reta de Regressão

Quando se estuda apenas uma variável, por exemplo Y , têm-se que a variação amostral entre cada ponto observado e a média é dada pelo desvio padrão amostral (Ver, Ramos, *et al.*, 2002), o qual é expresso pela raiz quadrada da variância. Neste caso,

$$\hat{\sigma} = \sqrt{S_{YY}} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}}. \quad (3.57)$$

No entanto, quando se estuda duas ou mais variáveis, cuja relação é expressa por uma equação igual à apresentada em (3.1); O desvio padrão de Y (reta de regressão) passa a ser o afastamento médio e mínimo existente entre cada ponto observado e a reta estimada (Almeida, 2003) e (Mandel, 1969), e é chamado de “Erro Padrão da Reta de Regressão”, que no caso específico do modelo Estrutural, um estimador pode ser dado pela raiz quadrada de S_v^2 da seguinte forma

$$EP(Y) = \sqrt{S_v^2} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y} - (X_i - \bar{X})\hat{\beta})^2}{(n - 2)}}, \quad (3.58)$$

sendo $(n - 2)$ os graus de liberdade do modelo, $\hat{\beta}$, \bar{X} , e \bar{Y} sendo obtidos pelas Equações (3.7), (3.13) e (3.14), respectivamente.

3.7 - Erro Padrão da Reta de Regressão Estimada

Caso a intenção seja estimar o erro padrão da estimativa da reta de regressão, \hat{Y}_i , a raiz quadrada da variância de \hat{Y}_i , deverá ser calculada, ou seja, $\sqrt{\hat{\sigma}_{\hat{y}}^2}$, que é chamado de erro padrão da reta de regressão estimada, $\widehat{EP}(\hat{Y}_i)$, como verifica-se a seguir,

$$EP(\hat{Y}_i) = \sqrt{\frac{\beta^2 \sigma_u^2 + \sigma_\varepsilon^2}{n}}, \quad (3.59)$$

onde β^2 , σ_u^2 , e σ_ε^2 são obtidos pelo sistema de Equações (3.5).

Segundo Almeida (2003), em virtude da equação de regressão ter sido estimada com base em um número de observações, a variância sobre a regressão representa uma medida de erro com a qual qualquer valor de Y pode ser estimado de um valor observado de X .

3.8 - Erro Padrão do Valor Predito da Variável Dependente

Dado o contínuo interesse em prever o valor de uma observação (Y_t), relativa ao valor de (U_t), tal que, chega-se que o erro padrão de (Y_t), será expresso pela soma dos erros padrões entre Y_i e \hat{Y}_i , como mostra-se a seguir,

$$EP(\hat{Y}_t) = \sqrt{\widehat{Var}(Y_i) + \widehat{Var}(\hat{Y}_i)}. \quad (3.60)$$

Por conseguinte, a equação de previsão estimada, é dada com base em um erro de previsão estimado através de

$$EP(\hat{Y}_t) = \sqrt{\frac{\sum_{i=1}^n \left(Y_i - \bar{Y} - (X_i - \bar{X})\hat{\beta} \right)^2}{n-2} + \frac{\beta^2 \sigma_u^2 + \sigma_\varepsilon^2}{n}}. \quad (3.61)$$

sendo $(n-2)$ os graus de liberdade do modelo, $\hat{\beta}$, \bar{X} , e \bar{Y} sendo obtidos pelas Equações (3.7), (3.13) e (3.14), respectivamente, e β^2 , σ_u^2 , e σ_ε^2 são obtidos pelo sistema de Equações (3.5).

3.9 - Intervalo de Confiança Assintótico

Essa é uma área pouco explorada dentro dos modelos com erros nas variáveis. Na prática, o mais comum quando o modelo é identificável, é utilizar as estimativas consistentes, que são estimadores assintoticamente normais para construir os intervalos de confiança.

3.9.1 - Intervalos de Confiança para os Parâmetros da Regressão

Intervalos de confiança assintóticos para os parâmetros da regressão, α e β , utilizando as suas variâncias assintóticas, são obtidos por

$$P \left(\hat{\alpha} - (Z_{\lambda/2}) \widehat{EP}(\hat{\alpha}) \leq \alpha \leq \hat{\alpha} + (Z_{\lambda/2}) \widehat{EP}(\hat{\alpha}) \right) = 1 - \lambda \quad (3.62)$$

e

$$P \left(\hat{\beta} - (Z_{\lambda/2}) \widehat{EP}(\hat{\beta}) \leq \beta \leq \hat{\beta} + (Z_{\lambda/2}) \widehat{EP}(\hat{\beta}) \right) = 1 - \lambda, \quad (3.63)$$

onde a estimativa do erro padrão de α e β é dado pelos seus respectivos desvios padrões assintóticos, obtido através da raiz quadrada da estimativa da variância de $\hat{\alpha}$ e $\hat{\beta}$, apresentado por (3.29) e (3.30), respectivamente, ou seja,

$$EP(\hat{\alpha}) = \sqrt{\widehat{Var}(\hat{\alpha})} = \sqrt{\frac{n\bar{X}^2 \widehat{Var}(\hat{\beta}) + S_v^2}{n}} \quad (3.64)$$

e

$$EP(\hat{\beta}) = \sqrt{\widehat{Var}(\hat{\beta})} = \sqrt{\frac{S_{XX}S_v^2 + \hat{\beta}^2\sigma_\delta^4}{\ell^2(n-1)}}. \quad (3.65)$$

Para Cheng e Van Ness (1994 e 1999), para a construção do intervalo de confiança assintótico é necessário somente do conhecimento da probabilidade de cobertura total $[100(1 - \lambda)]$ e do comprimento ou largura (Z) esperada do intervalo.

3.10 - Intervalo de Previsão Assintótico no Modelo Estrutural

Muitas vezes constrói-se um modelo de regressão com a finalidade de prever uma variável, por exemplo Y , de outra variável U . E outras vezes se está interessado somente na relação entre as duas variáveis. No primeiro caso, o interesse está em prever um novo valor de Y , correspondente a um novo U , com certa precisão. E no último caso, o interesse é estimar α e β com precisão. Cheng e Van Ness (1999) enfatizam que nos dois casos são envolvidas funções perdas bastante diferentes, portanto, a escolha das estimativas dos parâmetros de regressão a serem adotadas, depende do objetivo a ser alcançado.

3.10.1 - Modelo de Previsão

O modelo Estrutural definido por (3.1), com os erros $(\varepsilon_i, \delta_i)$ sendo independentes, identicamente distribuídos segundo uma distribuição normal. Então a distribuição condicional de Y dado X é

$$f(Y_i/X_i, U_i) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \exp \left\{ -\frac{(Y_i - \alpha - \beta U_i)^2}{2\sigma_\varepsilon^2} \right\}, \quad (3.66)$$

de modo que $E(Y/X, U) = \alpha + \beta\mu_u$, que envolve ambos os parâmetros, α e β . Sendo portanto, necessário utilizar-se dos estimadores dos parâmetros do modelo com erros nas variáveis, para predição de determinado valor de U_i , por exemplo um U_0 . Desta forma o modelo de previsão é dado por

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}U_0. \quad (3.67)$$

Se o interesse é um valor de Y que seja (3.67) correspondente à um valor de U , que não exista na amostra. Portanto, deve-se reservar o índice “ i ” para indicar os elementos pertencentes à amostra, e deve-se introduzir um outro índice, digamos “ t ”, para indicar outros valores de U . O novo valor, U_t , pode coincidir ou não com um dos valores de U_i da amostra. Desta forma, o modelo de previsão é dado por

$$\widehat{Y}_t = \widehat{\alpha} + \widehat{\beta}\widetilde{U}_t, \quad (3.68)$$

onde $\widehat{\beta}$, $\widehat{\alpha}$ e \widetilde{U} são os estimadores fornecidos pelas equações (3.6), (3.7) e (3.45), respectivamente e onde \widetilde{U} representa um “Vetor Aleatório”.

Devido a esse interesse constante em prever o valor de uma observação (Y_t), relativa ao valor de U_t , deve-se buscar uma medida da qualidade desta previsão, que normalmente é dada pela diferença entre o valor observado, Y_i , e o valor estimado, \widehat{Y}_t . Esta medida é comumente chamada de erro de previsão. Portanto, o estimador de $Y_i = \alpha + \beta U_i + \varepsilon_i$ é

$$\widehat{Y}_t = \widehat{\alpha} + \widehat{\beta}\widetilde{U}_t, \quad (3.69)$$

e o erro de previsão é

$$Y_i - \widehat{Y}_t = e_i = Y_i - \bar{Y} - (\widehat{Y}_t - \bar{Y}). \quad (3.70)$$

Nota-se que tanto \widehat{Y}_t como Y_i , são variáveis aleatórias e que de acordo com as definições feitas para o modelo (3.1); suas distribuições são independentes. Portanto, diz-se que \widehat{Y}_t é uma previsão não viesada do valor de Y_i , porque a esperança do erro de previsão é igual à zero, e o quadrado médio do erro de previsão, neste caso, a variância do erro de previsão, é

$$E(Y_i - \widehat{Y}_t)^2 = \widehat{Var}(e_i) = \widehat{Var}(Y_i) + \widehat{Var}(\widehat{Y}_t), \quad (3.71)$$

onde a $\widehat{Cov}(Y_i, \widehat{Y}_t) = 0$, logo, a variância do erro de previsão é composta de uma parcela apenas devido o erro ser aleatório. O estimador não-viesado desta variância é dado por

$$\widehat{Var}(\widehat{Y}_t) = \frac{\sum_{i=1}^n (Y_i - \widehat{Y}_t)^2}{n-2} + \frac{\beta^2 \sigma_u^2 + \sigma_\varepsilon^2}{n}, \quad (3.72)$$

que pode ser melhor expressa por

$$\widehat{Var}(\widehat{Y}_t) = \frac{n \sum_{i=1}^n (Y_i - \widehat{Y}_t)^2 + (n-2)(\beta^2 \sigma_u^2 + \sigma_\varepsilon^2)}{n^2 - 2n}. \quad (3.73)$$

E para avaliar essa precisão de \widehat{Y}_t como previsão do valor da nova observação, determina-se o intervalo de previsão, conforme apresentar-se-á a seguir.

3.10.2 - Intervalo de Previsão no Modelo Estrutural

O intervalo de previsão para a nova observação Y_t dado \tilde{U}_i é dado por

$$\hat{Y}_t \pm (Z_{\lambda/2})\widehat{EP}(\hat{Y}_t), \quad (3.74)$$

onde o erro padrão é obtido através da raiz quadrada da variância, como

$$EP(\hat{Y}_t) = \sqrt{\widehat{Var}(\hat{Y}_t)}, \quad (3.75)$$

e o $Z_{\lambda/2}$ é o percentil da distribuição normal, cujo valor mais usual é 1,96 que a maioria dos autores arredonda para 2, dando portanto, uma cobertura de 95% de confiança. Isto é, existe 95% de chance do verdadeiro valor de Y estar contido neste intervalo. Caso se deseje prever Y_t dado X , neste caso a equação de previsão estimada é obtida por (3.48) com erro de previsão estimado através de

$$EP(\hat{Y}_t) = \sqrt{\frac{n \sum_{i=1}^n (Y_i - \hat{Y})^2 + (n-2)(\beta^2 \sigma_u^2 + \sigma_\varepsilon^2)}{n^2 - 2n}}. \quad (3.76)$$

Portanto, o intervalo de previsão dado X é da forma

$$\hat{Y}_t \pm (Z_{\lambda/2})\widehat{EP}(\hat{Y}_t). \quad (3.77)$$

É importante ressaltar que o intervalo aumenta, à medida que U_t se afasta da sua média, e que ao se fazer uma extrapolação é necessário considerar que, o modelo linear que será ajustado só apresentará um ótimo comportamento dentro do intervalo coberto pela amostra. Portanto, a de se ter cuidado ao se extrapolar para valores muito distantes da amostra.

O conceito de intervalo de previsão é análogo ao de intervalo de confiança, com a diferença de que, enquanto o intervalo de confiança faz referência a uma constante (o parâmetro β , ou α , por exemplo), o intervalo de previsão faz referência a uma variável aleatória (neste caso, Y_t).

Capítulo 4

O Gráfico de Controle de Regressão

Neste capítulo apresenta-se o gráfico de controle de regressão, baseado na teoria proposta por Mandel (1969). Primeiramente a fundamentação teórica é abordada, e um exemplo prático publicado por Fuller (1987), sobre a relação entre a produção de milho e o nível nitrogênio no solo do estado de Iowa, é apresentado posteriormente.

4.1 - Introdução

Uma forma bastante eficiente de representar relações entre variáveis é através do Gráfico de Controle de Regressão, apresentado inicialmente por DiPaola (1945), que descreve o controle simultâneo de duas variáveis. Este tipo de gráfico ainda é pouco utilizado na literatura, por exigir do usuário a habilidade de saber relacionar as variáveis que possuem uma relação de causa e efeito. O estabelecimento errado das variáveis dependentes (efeitos) e independentes (causas), certamente implicará na tomada de decisões erradas. O objetivo deste tipo de gráfico é controlar uma variação média em lugar de uma constante (valor alvo nos gráficos do tipo *Shewhart*), onde a linha de regressão obtida através da equação de regressão clássica resume o relacionamento linear entre as duas variáveis de interesse. E, em torno desta linha de regressão, é estabelecido um intervalo de confiança, dentro do qual espera-se que o valor real da variável dependente esteja contido com uma certa probabilidade. Outra característica importante é que os *outliers* neste gráfico podem indicar as observações onde, por alguma razão, o relacionamento comum entre as duas variáveis de interesse não existe.

4.2 - Elementos do Gráfico de Controle de Regressão

No gráfico de controle convencional de Shewhart utiliza-se uma linha de desempenho médio, com limites de controle superior e inferior paralelos a esta linha média. Por exemplo, o gráfico \bar{X} tem a média das médias, $\bar{\bar{X}}$, como a linha central e $\bar{\bar{X}} \pm 3\sigma$, como as linhas de controle superior e inferior. A probabilidade de obtenção de um ponto fora dos limites de controle quando o processo está realmente sob controle é aproximadamente de 1 em 369. Essas três linhas são totalmente paralelas ao eixo “ x ”, indicando que uma média simples está sendo controlada.

No caso do gráfico de controle de regressão controla-se uma variação média, que é obtida através da linha de regressão estimada. Esta linha de regressão é chamada de linha central (LC), e é apresentada da seguinte forma

$$\hat{Y}_i = \hat{\alpha}_q + \hat{\beta}_q X_i, \quad i = 1, 2, 3, \dots, n, \quad (4.1)$$

onde Y é usado para representar a variável dependente, X é a variável independente e $\hat{\alpha}_q$ e $\hat{\beta}_q$ são os estimadores dos parâmetros desconhecidos do modelo de regressão obtidos por mínimos quadrados, a partir das Equações (2.10) e (2.11), respectivamente.

Quanto às linhas dos limites de controle, estas são paralelas à linha de regressão e não ao eixo horizontal, como é o caso do gráfico de controle do tipo *Shewhart*. Para obter-se a forma do gráfico de regressão, a relação entre as variáveis não pode ser desprezada, então, torna-se fundamental antes de construir o gráfico de controle de regressão, “plotar” os dados em um diagrama de dispersão, com o objetivo de confirmar se existe ou não a relação linear entre as variáveis, ou ainda, verificar a existência de pontos que fogem (*outliers*) do comportamento geral dos dados.

Assim, pontos que saem do padrão linear mostrado pelo diagrama de dispersão devem ser investigados. Se esses pontos são devidos a causas especiais, ou se normalmente não estiverem associados com o sistema de causas constantes, eles deverão ser excluídos dos cálculos utilizados para estabelecer o gráfico de controle. Os cálculos necessários a construção dos gráficos de controle de regressão são um pouco mais complexos e mais demorados que os dos gráficos de controle convencionais. O desvio padrão para o gráfico de regressão é o erro padrão estimado (\widehat{EP}) da linha de regressão. Ele não deve ser confun-

dido com o erro padrão de um valor predito da variável dependente. E segundo Mandel (1969), o erro padrão para o gráfico de controle de regressão (com ou sem erro de medida em X) é o desvio padrão estimado baseado nos desvios dos valores observados sobre a linha de regressão, ou seja,

$$EP(Y) = S_e = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2}}, \quad (4.2)$$

onde $(n-2)$ representa os graus de liberdade do modelo. Porém, como o objetivo do gráfico de controle é controlar uma variação média, que nada mais é do que uma esperança condicional de Y dado X , pode-se utilizar o erro padrão, para o modelo clássico de regressão, dado por

$$EP(\hat{Y}_i) = S_e \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}, \quad (4.3)$$

onde S_e é obtido pela Equação (4.2).

De acordo com Mandel (1969) uma outra maneira de se obter linhas de controle mais exatas é utilizando-se a estimativa de (\widehat{EP}) dada pelo erro padrão de um valor predito da variável dependente, dada por

$$EP(\hat{Y}_{Pi}) = S_e \sqrt{\frac{n+1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}, \quad (4.4)$$

onde o valor de S_e é obtido pela Equação (4.2) e X_0 é o valor particular de X no qual o erro padrão de \hat{Y}_{pi} é procurado e S_e é obtido pela Equação (4.2). Os limites que utilizam as Equações (4.3) e (4.4) como estimativas do erro padrão, são curvados, movendo-se para longe da linha de regressão nos extremos, para manter a probabilidade de alarme falso constante.

No gráfico de controle de regressão assume-se que os valores da variável dependente Y são linearmente relacionados com a variável independente X . Para cada valor específico X é assumido que os valores de Y são normalmente e identicamente distribuídos, com valor médio estimado pela linha de regressão e, com erro padrão que é independente do valor de X , sendo estimado pelos desvios das observações atuais e dos valores estimados de Y da linha de regressão.

Almeida (2003) com base na teoria apresentada por Mandel (1969) desenvolveu gráficos de controle de regressão aplicados a modelos com erros de medida na variável independente, mais especificamente o modelo de regressão funcional, onde a linha de regressão, (LC), que no caso do modelo clássico de regressão é representada pela Equação (4.1), no modelo com erro nas variáveis funcional é

$$\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta}\widehat{U}_i, \quad i = 1, 2, 3, \dots, n, \quad (4.5)$$

onde Y é usado para representar a variável dependente, U é a variável independente medida com erro e, $\widehat{\alpha}$ e $\widehat{\beta}$ são os estimadores dos parâmetros desconhecidos do modelo de regressão com erro de medida, apresentados nas Equações (3.6) e (3.7), respectivamente.

O erro padrão para o gráfico de controle de regressão, que é o desvio padrão estimado baseado nos desvios dos valores observados sobre (LC), para o gráfico de controle de regressão sem erro em X apresentado na Equação (4.2), no caso da variável ter sido medida com erro (modelo de regressão funcional) é dado por

$$EP(Y) = S_\nu = \sqrt{\frac{\sum_{i=1}^n (Y_i - \widehat{Y}_i)^2}{n - 2}}, \quad (4.6)$$

onde $(n - 2)$ representa os graus de liberdade do modelo. Porém, caso o interesse seja controlar a variação média de Y dado X , para o modelo funcional é dado por

$$EP(\widehat{Y}_i) = \sqrt{S_\nu^2 \left(\frac{1}{n} + \frac{S_{XX}(\widehat{U}_i - \overline{X})^2}{\ell^2(n-1)} \right) + \frac{\sigma_\delta^2 \widehat{\beta}}{\ell^2(n-1)}}, \quad (4.7)$$

onde S_ν é obtido pela Equação (4.6), $(n - 1)$ representa os graus de liberdade do modelo, σ_δ^2 , $\widehat{\beta}$, S_{XX} , \overline{X} , ℓ , são obtidos pelas Equações, (3.5), (3.7), (3.11), (3.13) e (3.33), respectivamente, e $\widehat{U}_i = \frac{\sigma_\delta^2 \widehat{\beta}(Y_i - \widehat{\alpha}) + \widehat{\sigma}_\varepsilon^2 X_i}{\widehat{\beta}^2 \sigma_\delta^2 + \widehat{\sigma}_\varepsilon^2}$, $i = 1, 2, 3, \dots, n$, onde $\widehat{\alpha}$ e $\widehat{\sigma}_\varepsilon^2$, são obtidos pelas Equações (3.6) e (3.8), respectivamente.

Se a intenção é obter linhas de controle mais exatas utiliza-se o *EP*, no caso do modelo de regressão funcional, dado por

$$EP(\widehat{Y}_{Pi}) = \sqrt{S_\nu^2 \left(1 + \frac{1}{n} + \frac{S_{XX}(\widehat{U}_i - \bar{X})^2}{\ell^2(n-1)} \right) + \frac{\sigma_\delta^2 \widehat{\beta}}{\ell^2(n-1)}}, \quad (4.8)$$

onde S_ν é obtido pela Equação (4.6), $(n-1)$ representa os graus de liberdade do modelo, σ_δ^2 , $\widehat{\beta}$, S_{XX} , \bar{X} , ℓ , são obtidos pelas Equações, (3.5), (3.7), (3.11), (3.13) e (3.33), respectivamente, e $\widehat{U}_i = \frac{\sigma_\delta^2 \widehat{\beta}(Y_i - \widehat{\alpha}) + \widehat{\sigma}_\varepsilon^2 X_i}{\widehat{\beta}^2 \sigma_\delta^2 + \widehat{\sigma}_\varepsilon^2}$, $i = 1, 2, 3, \dots, n$, onde $\widehat{\alpha}$ e $\widehat{\sigma}_\varepsilon^2$ são obtidos pelas Equações (3.6) e (3.8), respectivamente.

Para completar o gráfico de controle, é necessário decidir quão perto dos limites (superior e inferior) se permite que o processo varie. A pergunta em relação a proximidade dos limites pode inspirar uma variedade de respostas, nenhuma necessariamente correta. Existem vários fatores a serem considerados para tomar a decisão, porém todas são orientadas de uma maneira ou de outra, dentro dos custos da produção. Por exemplo, a utilização de 1σ que garante um limite estreito, com alto risco de alarmes falsos, ou 3σ que garante limites mais largos, com menor probabilidade de alarmes falsos, e este é um assunto de decisão gerencial baseado na economia e na experiência do processo em questão.

Assim, pode-se agora montar os limites do gráfico de controle, utilizando uma confiança fixada k , ou seja, $\pm k\widehat{\sigma}$, onde $k = 1, 2$ ou 3 e $\widehat{\sigma}$ é um estimador do erro padrão, que pode ser estimado pelas Equações (4.2), (4.3) ou (4.4), no caso do modelo de regressão clássica. Para o modelo de regressão com erro nas variáveis funcional $\widehat{\sigma}$ pode ser estimado pelas Equações (4.6), (4.7) ou (4.8). A seguir apresenta-se as possíveis variações de “ k ”, através da Figura 4.1.

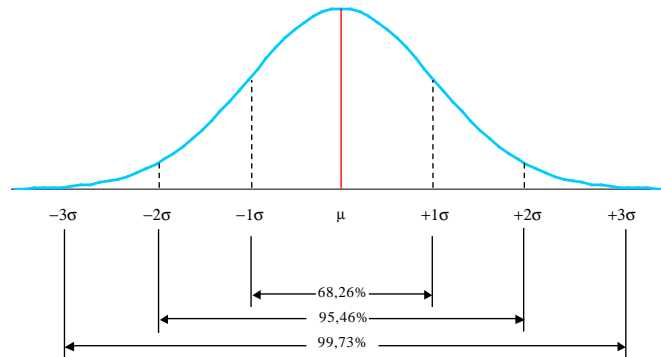


Figura 4.1 - Probabilidades Normais para $\mu \pm k\sigma$, $k = 1, 2, 3$.

Portanto, a linha central e os limites de controle do gráfico de controle de regressão clássico são obtidos por

$$\begin{aligned} LSC &= \hat{Y}_i + k\hat{\sigma}, \\ LC &= \hat{Y}_i = \hat{\alpha}_q + \hat{\beta}_q X_i, \\ LIC &= \hat{Y}_i - k\hat{\sigma}, \end{aligned} \quad (4.9)$$

onde $i = 1, 2, 3, \dots, n$, $\hat{\alpha}_q$ e $\hat{\beta}_q$ são os estimadores dos parâmetros desconhecidos do modelo de regressão obtidos por mínimos quadrados, pelas Equações (2.10) e (2.11), respectivamente.

Para o modelo de regressão com erros nas variáveis funcional, a linha central e os limites de controle do gráfico de controle de regressão são obtidos por

$$\begin{aligned} LSC &= \hat{Y}_i + k\hat{\sigma}, \\ LC &= \hat{Y}_i = \hat{\alpha} + \hat{\beta}\hat{U}_i, \\ LIC &= \hat{Y}_i - k\hat{\sigma}, \end{aligned} \quad (4.10)$$

onde $i = 1, 2, 3, \dots, n$, e \hat{U}_i é obtido a partir da Equação (4.8), $\hat{\alpha}$ e $\hat{\beta}$ são dados pelas Equações (3.6) e (3.7), respectivamente.

4.3 - Construindo o Gráfico de Controle de Regressão

Mandel (1969) apresenta gráficos de controle aplicados ao modelo clássico de regressão. Para ilustrar os passos necessários, a construção do gráfico de controle de regressão, utiliza-se uma aplicação mostrada por Almeida (2003), na qual o gráfico de controle é aplicado ao modelo funcional de regressão, ao conjunto de dados apresentado por Fuller (1987), o qual considera a relação entre a produção de milho (Y) e o nível de nitrogênio disponível no solo (X), coletada em 11 locais no estado de Iowa, nos Estados Unidos, conforme apresentado na Tabela 4.1

Tabela 4.1 - *Produção de milho e nível de nitrogênio no Solo de Iowa, nos Estados Unidos.*

Local	X(Nitrogênio no Solo)	Y(Produção)
1	70	86
2	97	115
3	53	90
4	64	86
5	95	110
6	64	91
7	50	99
8	70	96
9	94	99
10	69	104
11	51	96

Fonte: Fuller (1987), p. 18.

Entretanto, para se obter o nível de nitrogênio no solo se faz necessária uma análise química de uma pequena amostra do solo, selecionada de cada lote ou canteiro experimental. Como resultado da análise química, têm-se uma estimativa do nível de nitrogênio, X , e não o verdadeiro valor do nitrogênio, por exemplo, U . Neste caso, tem-se um modelo com erro nas variáveis, formalmente definido em (2.1), (2.2) e (2.3). A forma de coleta dos dados de nitrogênio permite duas interpretações dos verdadeiros valores de nitrogênio (aleatório ou fixo). Para o exemplo proposto Fuller (1987), considera que os desconhecidos valores de U_i , $i = 1, 2, \dots, n$ são fixos e que a variância do erro de medida associado a variável independente X é dada por $\sigma_x^2 = 57$.

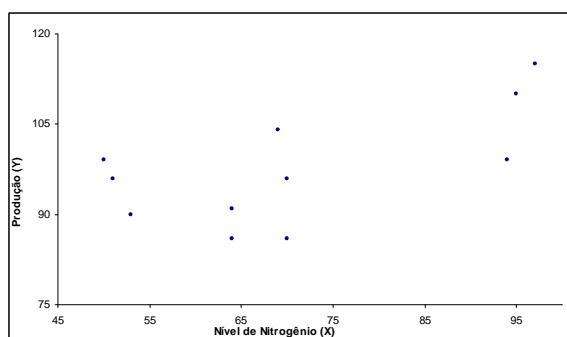


Figura 4.2 - Diagrama de Dispersão da Produção de milho e nível de nitrogênio no Solo de Iowa, nos Estados Unidos, 1987.

Observando-se a Figura 4.2 não se verifica nenhum ponto extremo (*outlier*). As estatísticas necessárias à construção do gráfico de controle de regressão funcional, apresentam-se na Tabela 4.2.

Tabela 4.2 - Estatísticas necessárias à construção do gráfico de controle de regressão funcional, com $\sigma_{\delta}^2 = 57$.

Símbolo	Nome	Valor Obtido
N	Número de pares de valores	11
EP(Y)	Erro padrão da produção de milho	7,710
\bar{Y}	Produção média de milho	97,450
S_{YY}	Desvio padrão da produção de milho	87,670
\bar{X}	Nível médio de nitrogênio no solo	70,64
S_{XX}	Desvio padrão do nível de nitrogênio	304,85
S_{XY}	Covariância entre produção de milho e nível de nitrogênio	104,88
ℓ	Resíduo médio quadrado de U_i	293,98
S_v^2	Variância ou resíduo médio quadrático do modelo	59,44
S_{ε}^2	Variância do erro ε	43,29
$\hat{\alpha}$	Intercepto	67,56
$\hat{\beta}$	Inclinação	0,420
$EP(\hat{\alpha})$	Erro padrão do intercepto	10,648
$EP(\hat{\beta})$	Erro padrão da inclinação	0,147

Com base nos valores apresentados na Tabela 4.2, verifica-se que as condições de regularidade quanto à $S_{XX} > \sigma_\delta^2$ e $S_{YY}(S_{XX} - \sigma_\delta^2) - S_{XY}^2 > 0$ são atendidas. Como $S_{XX} > \sigma_\delta^2$ e $S_{YY}(S_{XX} - \sigma_\delta^2) - S_{XY}^2 > 0$, pode-se obter as estimativas de β e α a partir das Equações (3.6) e (3.7), respectivamente. Para se estabelecer a linha central do gráfico de controle se faz necessário a obtenção dos valores estimados dos U_i , conforme é apresentado na Tabela 4.3.

Tabela 4.3 - *Nível de Nitrogênio estimado para o Modelo Funcional com $\sigma_\delta^2 = 57$.*

Local	X (Nitrogênio Observado)	Y (Produção Observada)	\hat{U}_i (Nitrogênio estimado)
1	70	86	64,96
2	97	115	99,88
3	53	90	53,00
4	64	86	60,10
5	95	110	96,01
6	64	91	62,36
7	50	99	54,64
8	70	96	69,47
9	94	99	90,24
10	69	104	72,26
11	51	96	54,09

Utilizando os dados da Tabela 4.3, pode-se estabelecer a equação estimada da linha de regressão para o modelo funcional, obtendo-se assim a linha central (LC) do Gráfico de Controle de Regressão Funcional, da seguinte forma

$$LC = \hat{Y}_{\hat{z}} = 67.56 + 0.42\hat{U}_i.$$

Com a linha central (LC) e fixando-se o valor de $k = 2$, ou seja, com limites de controle fixados em 95% de confiança, torna-se possível a obtenção do limite superior de controle (LSC) e do limite inferior de controle (LIC), utilizando-se a Equação (4.10). Os Gráficos de Controle de Regressão Funcional com limites de controle fixados em 95% de confiança, para os três estimadores de σ , apresentados nas Equações (4.6), (4.7) e (4.8), são mostrados nas Figuras (4.3) e (4.4), respectivamente.

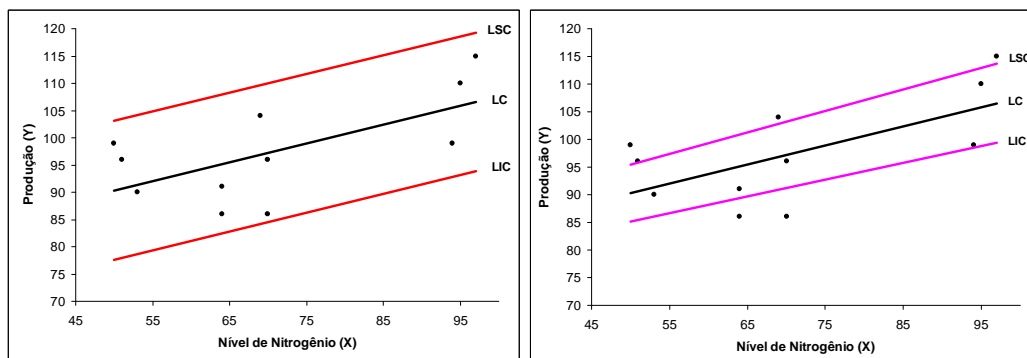


Figura 4.3 - Gráfico de *Controle de Regressão Funcional* com $\sigma_{\delta}^2 = 57$, da *Produção de Milho e Nível de Nitrogênio no Solo - Estado de Iowa, 1969*, com limites de $2EP(Y) = 2S_e$ e $2EP(\hat{Y}_i) = 2S_{\hat{Y}_i}$, respectivamente.

Os limites de um dos gráficos da Figura 4.3 foram obtidos utilizando-se o erro padrão da linha de regressão ($EP(Y)$), como estimador do desvio padrão σ . Nota-se que todos os pontos “plotados” no gráfico estão dentro dos limites de controle, o que mostra que o processo de produção de milho do estado de Iowa está sob controle estatístico. Por outro lado, quando utiliza-se o erro padrão da linha de regressão estimada $EP(\hat{Y}_i)$, como estimador do desvio padrão σ , obtém-se limites de controle bem diferentes, onde observa-se pontos fora dos limites de controle, indicando um processo fora de controle estatístico. Na Figura 4.4 pode-se novamente observar um processo sob controle estatístico. Porém, este primeiro gráfico foi elaborado utilizando como estimador do desvio padrão σ , o erro padrão da previsão ($EP(\hat{Y}_t)$) e também construiu-se um gráfico comparativo entre as três estimativas de σ .

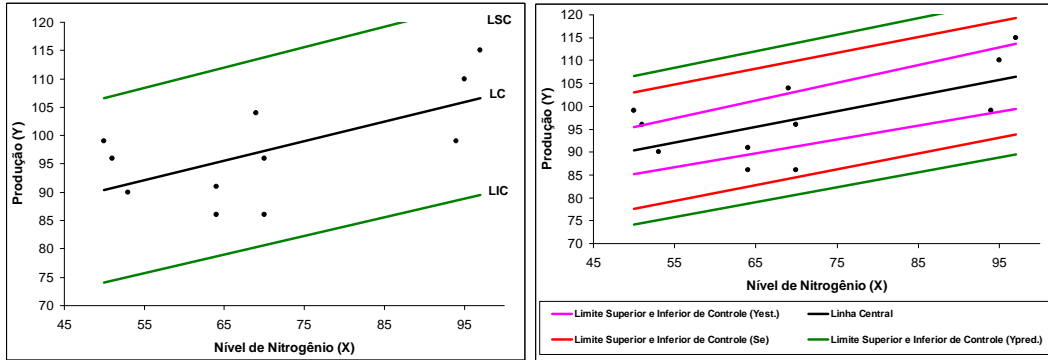


Figura 4.4 - Gráfico de Controle de Regressão Funcional com $\sigma_{\hat{y}}^2 = 57$, da Produção de Milho e Nível de Nitrogênio no Solo - Estado de Iowa, 1969. com limites de $2EP(\hat{Y}_t) = 2S_{\hat{Y}_{pi}}$ e o Gráfico Comparativo de Regressão Funcional com $\sigma_{\hat{y}}^2 = 57$, respectivamente.

Para mostrar a diferença causada pela escolha do estimador de variabilidade, no gráfico de controle de regressão funcional, a Figura 4.4, destaca os limites de controle para cada estimativa do desvio padrão σ . É importante destacar que a escolha da estimativa do desvio padrão depende muito do tipo de processo que se deseja controlar, por exemplo, no exemplo em questão, uma boa estimativa seria a utilização do erro padrão da previsão, pois conforme é mostrado na Figura 4.4 este limite é mais largo, possibilitando uma chance menor de ocorrência de alarmes falsos.

Pode-se perceber que a partir da Figura 4.4, que se está iniciando o processo de controle, deve-se optar pelos limites de controle baseados no erro padrão da linha de regressão ($EP(Y)$) ou no erro padrão da previsão $EP(\hat{Y}_t)$. Sendo este último bem mais preciso, já que se trata de um estimador da variabilidade da equação de regressão estimada, ou seja, de uma variabilidade da média, como é recomendado nos gráficos de controle de Shewhart. Por outro lado, se o processo já está sob controle há algum tempo, isto é, tem-se um processo conhecido como “processo refinado”, então recomenda-se utilizar o erro padrão da linha de regressão estimada $EP(\hat{Y}_i)$, como estimador do desvio padrão σ .

4.4 - Analisando o Gráfico de Controle de Regressão

O gráfico de controle de regressão, assim como o gráfico de controle do tipo *Shewhart* pode servir como uma ferramenta de controle e medida de desempenho. Para isto utiliza-se o seguinte princípio: quando as investigações sobre variações não forem para produzir decisões significativas para o processo (redução da variabilidade), aceite-as, pois estas provavelmente são devidas as causas comuns.

No exemplo em questão, o gráfico de controle de regressão foi utilizado para manter o controle do desempenho, isto é, o controle da produção de milho. Assim, imagina-se que exista um local 12, onde observa-se um nível de nitrogênio no solo igual à 80 e uma produção de 93. Pergunta-se:

- (1) Existiu um aumento ou diminuição na produtividade?
- (2) E este desempenho é aceitável?

Para responder estas perguntas, recorre-se ao desempenho histórico do processo. Quando o Nível de Nitrogênio ($X = 80$) é posto sobre a linha central da Figura 4.3, determina-se que a produção será de $Y = 99,67$. Neste caso responde-se a primeira pergunta (1): observa-se, que existiu um aumento de produtividade, visto que, a produção sofreu um acréscimo de 6,67 a mais, ou seja, $(99,67 - 93)$ a produção no local 12 ficou aquém da expectativa prevista. Logo, verifica-se que o desempenho está acima do normal, e uma decisão gerencial investigaria se a causa é especial ou aleatória.

Quanto a segunda pergunta (2), recorre-se aos limites de controle da Figura 4.3, como a produção de milho declarada esta dentro dos limites de controle, esta “falta” (produção inferior à esperada) pode ser aceita como normal em relação ao desempenho histórico do período.

É relevante para o processo, observar que sem a orientação oferecida pelos limites do gráfico de controle, o responsável pelo processo produtivo se preocuparia ao observar um *Déficit* na produção de 6,67 à menos, implicando contudo em uma atitude de revisão da situação, com o intuito de determinar as causas básicas do processo, o que acarretaria em um custo a mais neste processo. Ou seja, o responsável pelo processo produtivo ao deparar-se com uma produção inferior a estimada para aquele determinado local, julgaria que o processo estivesse fora de controle estatístico, quando na verdade ele não está.

Outro fato importante de se analisar são pontos que caem fora dos limites de controle superior e inferior. Por exemplo, quando um ponto cair acima do limite superior de controle, isto indica que a produtividade está significativamente mais alta do que o nível de desempenho estabelecido nos últimos lotes. O que poderia ser causado devido um desempenho irregular, coleta dos dados equivocada, fenômenos da natureza (tempestade, por exemplo), equipamento de medição do nível de nitrogênio com defeito, dentre outras causas. Por outro lado, se um ponto cair abaixo do limite inferior de controle, isto pode indicar uma deficiência real na produção de milho, ou pode ser devido à erros de superestimação na produção ou subestimação na informação do nível de nitrogênio. Observe que se um ponto cair fora do limite superior de controle, embora seja uma indicação de um excepcional desempenho, é também um bom motivo para investigação, desde que isto possa beneficiar o responsável pelo processo de produção a encontrar razões e adotar práticas que conduzam a um desempenho eficiente.

Pode-se utilizar o gráfico de controle para medir o efeito de novas técnicas de cultivo de milho sobre a produtividade. Como por exemplo, a efetividade de programas da melhoria da produtividade sobre o cultivo do milho, logo devido o processo de cultivo possibilitar que se compare a quantidade esperada da produção com a produção atual, ambos antes e depois da implementação de novas técnicas agrícolas, disponibilizando assim controle próprio a ser implementado sobre as variedades.

Os resultados alcançados com o gráfico de controle de regressão são relevantes para qualquer processo de monitoramento e aperfeiçoamento da qualidade, pois o mesmo garante

1. Uma descrição dos elementos básicos do gráfico de controle de regressão;
2. Uma revisão de estatísticas e técnicas de análises necessárias para estabelecer e usar este gráfico;
3. Fornecer aplicações ilustrativas para uma variedade de problemas de gerenciamento na produção.

Portanto, o gráfico demonstra-se bastante útil para uma variedade de problemas do gerenciamento da produção de milho e oferece possibilidades para aplicações em diversas instituições governamentais, empresariais e industriais.

Capítulo 5

O Gráfico de Controle de Regressão Para o Modelo Estrutural e Aplicação

Neste capítulo o Modelo Estrutural de regressão linear com erro nas variáveis, apresentado no Capítulo 3 e, a teoria de Gráficos de Controle de Regressão, mostrada no Capítulo 4 são combinados para produzir uma técnica metodológica efetiva de *Controle Estatístico da Qualidade (CEQ)*.

5.1 - Introdução

Quando os erros de medida são relativamente grandes em comparação com o erro inerente do processo, reduz-se o poder do gráfico de controle em detectar as possíveis mudanças no processo, isto é, as suas causas especiais (Kanazuka, 1986). Então se utiliza o Modelo Estrutural, em conjunto com o Gráfico de Controle de Regressão, para monitorar essas possíveis mudanças no processo com uma maior precisão. Se existe uma relação de causa-e-efeito entre as variáveis, então não há nenhuma razão para que quaisquer pares de variáveis altamente correlacionadas não possam ser “*plotadas*” em um Gráfico de Controle de Regressão. Em *CEQ* é freqüente o uso de variáveis que são resultantes de medições, como por exemplo, diâmetro e volume de uma peça ou dimensão de uma ferramenta do processo entre outros. Em resumo, qualquer que seja a forma de medição de uma variável, o erro de medida estará sempre presente, portanto deve ser levado em consideração. Isto é feito através do uso de modelos com erros nas variáveis, como o Modelo Estrutural combinado com gráficos de controle. A utilização de linhas de regressão é essencialmente idêntica aos procedimentos estatísticos usados na previsão de vendas ou de controle de produção de uma empresa. A única real diferença pode ser a utilização de linhas de limite de controle e a idéia de *CEQ*, de que deve haver uma causa assinalável (especial) se um ponto aparecer além destas linhas.

Neste Capítulo, dois exemplos são utilizados para ilustrar a aplicabilidade dos Gráficos de Controle de Regressão Estrutural (Fuller, 1987 *p.* 18 e os dados reais de uma indústria brasileira). A eficiência dos estimadores do Modelo de Regressão com Erro nas Variáveis Estrutural é avaliada em relação aos estimadores do Modelo clássico de Regressão, através de dois exemplos da literatura (Fuller, 1987 *p.* 18 e Mandel (1969, *v.* 1, N. 1)). Os exemplos ilustrarão a capacidade dos estimadores do Modelo com Erros nas Variáveis Estrutural, em detectar os erros de medida na variável independente quando a variância da amostra analisada aumenta, o que não ocorre com os estimadores do modelo clássico de regressão, pois os mesmos não conseguem detectar os erros de medições a medida que o tamanho da amostra aumente.

Um dos exemplos baseia-se em um estudo sobre a produção de milho no solo de *Iowa* retirado de Fuller (1987, *p.* 18). Este conjunto de dados consiste de 11 locais (amostras) onde foram verificados os níveis de nitrogênio e a produção de milho, ao qual foi aplicado a metodologia de Gráficos de Controle de Regressão Estrutural, com a intenção de monitorar esse processo de produção e comparou-se a eficiência dos estimadores do Modelo de Regressão com Erros nas Variáveis Estrutural, com as estimativas do Modelo clássico de Regressão.

O segundo exemplo baseia-se em um conjunto de dados que representa uma aplicação vista em Mandel (1969, *v.* 1, N. 1). Nela apresenta-se um estudo para controlar e processar correspondências nas 74 maiores agências de correios dos *EUA*. Comparou-se a eficiência dos estimadores do Modelo de Regressão com Erros nas Variáveis Estrutural, com as estimativas do Modelo clássico de Regressão.

Por fim, um conjunto de dados real foi utilizado neste trabalho o qual foi obtido através de um processo de produção de alumínio de uma indústria brasileira. Os dados referem-se as médias mensais da densidade aparente em g/cm^3 (Y), e do peso em Kg (X), dos blocos de carbono (eletrodos), necessários para a produção de alumínio, no período de agosto de 2001 à janeiro de 2003. A produção desta indústria baseia-se em alumínio primário, na forma de lingote de 22 Quilos, com uma especificação de pureza mínima igual a 99,7% de Al.

Ao longo do processo produtivo de alumínio (Al), através de eletrólise (processo que separa os elementos químicos através da utilização de eletricidade) da alumina (Al_2O_3), o carbono (C) dos eletrodos possui a importante função de associar-se as três moléculas de oxigênio da alumina, liberando assim as duas moléculas de alumínio. O alumínio que é liberado após este processo é o produto final. Os eletrodos são condutores de energia elétrica responsáveis pela eletrólise, nos fornos de produção do alumínio. A densidade aparente representa a característica da qualidade mais importante dos eletrodos, visto que, caso haja uma alta densidade aparente nos eletrodos isso acarretará um bloqueio na condução de corrente elétrica, o que irá impedir o processo de eletrólise necessário para a produção de alumínio. Em contra partida, caso a densidade aparente seja baixa, isso implicará em um alto consumo de eletrodos e um conseqüente aumento da contaminação do alumínio que se produzir.

Essa indústria possui um total de 985 fornos de produção de alumínio, cada forno do processo produtivo do alumínio da indústria “pesquisada” utiliza 18 eletrodos, cada um ao custo de U\$ 845,00 (Oitocentos e Quarenta e Cinco Dólares) e com uma vida média útil em torno de 22 dias. Portanto, torna-se imprescindível para a indústria que a densidade aparente dos blocos de carbono se mantenha sob controle, evitando assim um desperdício dessa matéria prima, o qual viria a aumentar os custos desse processo de produção. Aplicou-se a metodologia de Gráficos de Controle de Regressão Estrutural, a este conjunto de dados com a intenção de monitorar esse processo produtivo.

5.2 - Passos para elaboração do Gráfico de Controle de Regressão Estrutural

Para os modelos definidos pelas Equações (3.1)-(3.3), com a variância dos erros de medição σ_δ^2 conhecida. Então, para se “plotar” o Gráfico de Controle de Regressão Estrutural faz-se necessário os seguintes passos (Figura 5.1):

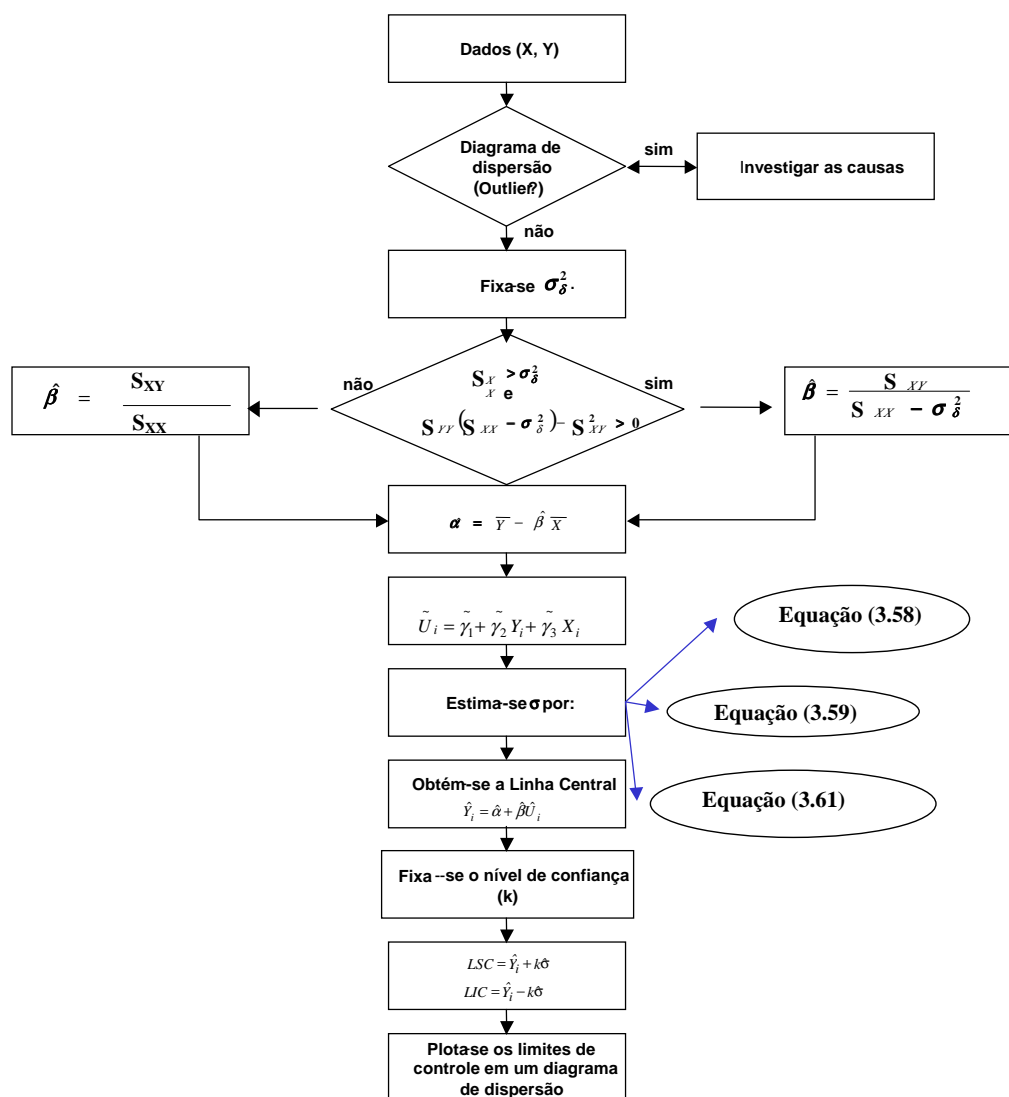


Figura 5.1 - Fluxograma da Construção do Gráfico de Controle de Regressão Estrutural.

5.3 Construindo o Gráfico de Controle de Regressão Estrutural

5.3.1 Gráfico de Controle de Regressão Estrutural para o Exemplo de Fuller (1987).

Para ilustrar os passos necessários a construção do Gráfico de Controle de Regressão Estrutural são utilizados os dados de produção de milho e nível de nitrogênio, apresentados na Tabela 4.1.

Para se obter o nível de nitrogênio no solo se faz necessária uma análise química de uma pequena amostra do solo, selecionada de cada lote ou canteiro experimental. Como resultado da análise química, têm-se uma estimativa do nível de nitrogênio, X , e não o verdadeiro valor do nitrogênio, por exemplo de U . Neste caso, tem-se um modelo com erro nas variáveis, formalmente definido em (3.1). A forma de coleta dos dados de nitrogênio permite duas interpretações dos verdadeiros valores de nitrogênio (aleatório ou fixo). Para o exemplo proposto, considera-se que os desconhecidos valores de U_i , $i = 1, 2, \dots, n$ são aleatórios (o que dá origem a um vetor aleatório) e que a variância do erro de medida associado a variável independente X é dada por $\sigma_\delta^2 = 57$.

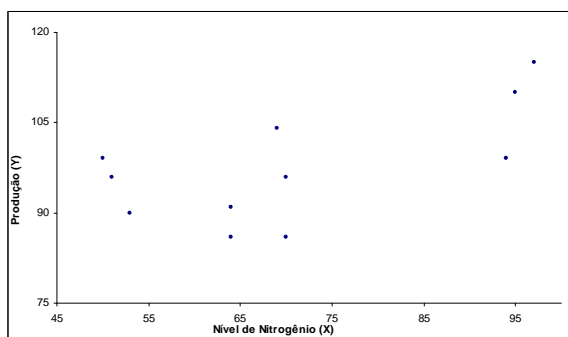


Figura 5.2 - Diagrama de Dispersão entre Produção de Milho e Nível de Nitrogênio no Solo - Estado de Iowa, 1987

Analizando-se a Figura 5.2 não se verifica nenhum ponto extremo (*outlier*).

Obtêm-se a seguir as estatísticas necessárias à construção do gráfico de controle de regressão Estrutural, as quais apresentam-se na Tabela 5.2.

Tabela 5.1 - Estatísticas obtidas a partir da Tabela 4.1 para o Modelo Estrutural com $\sigma_\delta^2 = 57$.

Símbolo	Nome	Valor Obtido
N	Número de pares de valores	11
EP(Y)	Erro padrão da produção de milho	7,710
\bar{Y}	Produção média de milho	97,450
S_{YY}	Variância da produção de milho	87,670
\bar{X}	Nível médio de nitrogênio no solo	70,640
S_{XX}	Variância do nível de nitrogênio	304,850
S_{XY}	Covariância entre produção de milho e nível de nitrogênio	104,880
ℓ	Resíduo médio quadrado de U_i	208,937
S_ν^2	Variância ou resíduo médio quadrático do modelo	59,440
σ_ε^2	Variância do erro ε	43,290
σ_u^2	Variância do erro U	247,854
$\hat{\alpha}$	Intercepto	67,560
$\hat{\beta}$	Inclinação	0,420
EP($\hat{\alpha}$)	Erro padrão do intercepto	10,648
EP($\hat{\beta}$)	Erro padrão da inclinação	0,147

Com base nos valores apresentados na Tabela 5.1, verifica-se que as condições de regularidade quanto à $S_{XX} > \sigma_\delta^2$ e $S_{YY}(S_{XX} - \sigma_\delta^2) - S_{XY}^2 > 0$, são atendidas. Como $S_{XX} > \sigma_\delta^2$ e $S_{YY}(S_{XX} - \sigma_\delta^2) - S_{XY}^2 > 0$, pode-se obter as estimativas de β e α a partir das equações (3.6) e (3.7), respectivamente.

Para a construção do preditor \tilde{U}_i (vetor aleatório), ou seja, a construção do verdadeiro valor de U_i utiliza-se a Equação (3.45), supondo que o mesmo é aleatório e com distribuição normal. A estimativa do vetor de médias é dada por

$$(\hat{\mu}_Y, \hat{\mu}_X)^T = (\bar{Y}, \bar{X})^T = (97, 45; 70, 64)^T.$$

Baseado em (3.46), chega-se à:

$$\begin{aligned} \begin{bmatrix} \tilde{\gamma}_2 \\ \tilde{\gamma}_3 \end{bmatrix} &= \begin{bmatrix} S_{YY} & S_{YX} \\ S_{XY} & S_{XX} \end{bmatrix}^{-1} \begin{bmatrix} S_{XY} \\ S_{XX} - \sigma_\delta^2 \end{bmatrix} \\ &= \begin{bmatrix} 87,6727 & 104,8818 \\ 104,8818 & 304,8545 \end{bmatrix}^{-1} \begin{bmatrix} 104,8818 \\ 304,8545 - 57 \end{bmatrix} \\ &= \begin{bmatrix} 0,3801 \\ 0,6822 \end{bmatrix}; \end{aligned}$$

Assim, obtém-se o termo independente do modelo de estimação do verdadeiro valor de U_i , isto é,

$$\tilde{\gamma}_1 = (1 - \tilde{\gamma}_3) \bar{X} - \tilde{\gamma}_2 \bar{Y} = (1 - 0,6822) 70,64 - 0,3801 \cdot 97,45 = -14,5942.$$

Portanto, concluí-se que a equação de estimação do verdadeiro valor de U_i , dado (Y_i, X_i) e considerando U_i como sendo aleatório é

$$\tilde{U}_i = -14,5942 + 0,3801Y_i + 0,6822X_i.$$

A partir de (3.48), obtém-se a variância estimada do erro de predição como

$$\hat{V}\{\tilde{U}_i - U_i/(Y_i, X_i)\} = 38,90.$$

Para se estabelecer a linha central do gráfico de controle se faz necessário a obtenção dos verdadeiros valores dos U_i , conforme é apresentado na Tabela 5.2.

Tabela 5.2 - *Nível de Nitrogênio estimado para o Modelo Estrutural com $\sigma_\delta^2 = 57$*

Local	X(Nitrogênio Observado)	Y(Produção Observada)	$\tilde{U}_{i,s}$ (Nitrogênio estimado)
1	70	86	65,85
2	97	115	95,29
3	53	90	55,77
4	64	86	61,75
5	95	110	92,03
6	64	91	63,66
7	50	99	57,14
8	70	96	69,65
9	94	99	87,16
10	69	104	72,01
11	51	96	56,69

A partir da Tabela 5.2, pode-se estabelecer a equação estimada da linha de regressão para o Modelo Estrutural, obtendo-se a linha central (LC) do Gráfico de Controle de Regressão Estrutural, dada por

$$LC = \hat{Y}_{\hat{\gamma}} = 67.56 + 0.42\tilde{U}_i.$$

Após a obtenção da linha central (LC) e fixando-se o valor de $k = 2$, ou seja, com limites de controle fixados em 95% de confiança, torna-se possível obter o limite superior de controle (LSC) e o limite inferior de controle (LIC).

Os Gráficos de Controle de Regressão Estrutural com limites de controle fixados em 95% de confiança, para os três estimadores de σ apresentados nas Equações (3.58), (3.59) e (3.61), são mostrados nas Figuras 5.3 e 5.4.

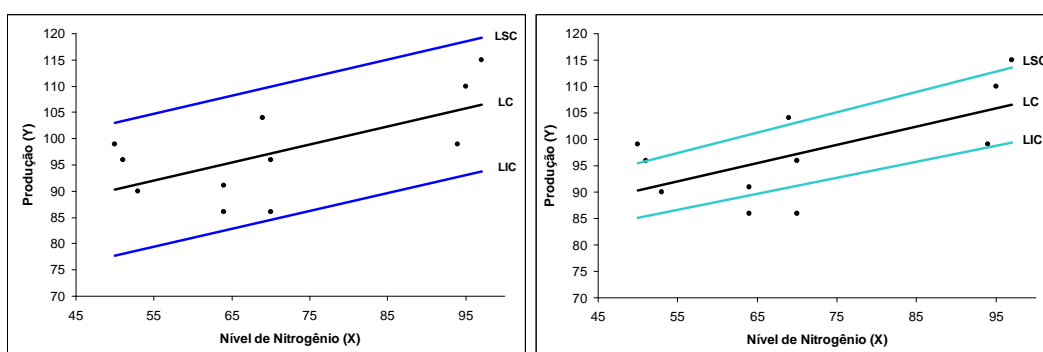


Figura 5.3 - Gráfico de Controle de Regressão Estrutural com $\sigma_0^2 = 57$ da Produção de Milho e Nível de Nitrogênio no Solo - Estado de Iowa, 1987. Com Limites de $2EP(Y) = 2S_e$ e $2EP(\hat{Y}_i) = 2S_{\hat{Y}_i}$, respectivamente.

Para a construção dos limites de controle do primeiro gráfico da Figura 5.3 utilizou-se o erro padrão da linha de regressão ($EP(Y)$), como estimador do desvio padrão de σ , ou seja, obtem-se parte da medida de variabilidade em torno da linha central, medida a qual, é multiplicada por uma constante “ k ” = 2 (95% de confiança). Pode-se perceber que todos os pontos “plotados” no gráfico estão dentro dos limites de controle, o que demonstra um indicativo de que o processo está sob controle estatístico. Entretanto, o segundo gráfico utiliza-se o erro padrão da linha de regressão estimada $EP(\hat{Y}_i)$, onde o objetivo do gráfico é controlar a variação média de um processo, ou seja, a esperança condicional da variável dependente (Y) dado uma variável independente (X), neste caso os limites de controle obtidos serão diferentes dos observados no primeiro gráfico. No segundo gráfico da Figura 5.3 nota-se, a presença de pontos fora dos limites de controle, superior e inferior, isto é, conclui-se que este processo não está sob controle estatístico e que esta estimativa de σ adotada ($2S_{\hat{Y}_i}$), não consegue gerar limites que compreendam os pontos observados neste processo de produção.

Na Figura 5.4 pode-se observar um processo sob controle estatístico. Para a sua construção utilizou-se como estimador da variabilidade (estimador do desvio padrão σ), o erro padrão da previsão ($EP(\hat{Y}_t) = S_{\hat{Y}_{pi}}$), objetivando se obter linhas de controle mais exatas. Para se verificar a diferença gerada pela escolha, de um dos três estimadores de σ apresentados, durante a construção do Gráfico de Controle de Regressão Estrutural, construiu-se o segundo gráfico da Figura 5.4, o qual sintetiza as Figuras (5.3), e o primeiro gráfico da Figura (5.4) apresentados anteriormente.

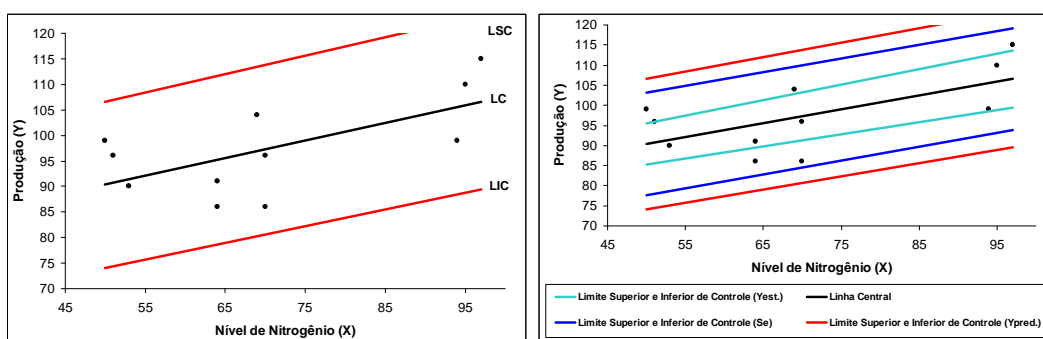


Figura 5.4 - Gráfico de Controle de Regressão Estrutural com $\sigma_{\delta}^2 = 57$ da Produção de Milho e Nível de Nitrogênio no Solo - Estado de Iowa, 1987. Com Limites de $2EP(\hat{Y}_t) = 2S_{\hat{Y}_{pi}}$ e o Gráfico Comparativo de Regressão Estrutural com $\sigma_{\delta}^2 = 57$, respectivamente.

É importante destacar que a escolha da estimativa do desvio padrão será influenciada pelo tipo de processo que se tem como objetivo controlar. Verifica-se que, ao se analisar o segundo gráfico da Figura 5.4, que o gráfico construído utilizando-se o erro padrão de um valor predito da variável dependente, é o que melhor se aplica a este processo de produção, visto que, dentre as três estimativas de erro utilizadas, além deste conseguir compreender todos os pontos observados, também produz os limites de controle mais largos, com isto, uma eventual situação de anormalidade (causa especial ou aleatória), será mais facilmente controlada por estes limites, conseqüentemente a chance de alarmes falsos ocorrerem ao longo do monitoramento deste processo serão menores, por tanto, a opção por este terceiro estimador de σ é a mais adequada para o processo de produção de milho, em relação ao nível de nitrogênio disponível no solo.

Na Figura 5.4, caso o processo de controle esteja em seu início, faz-se a opção pelos limites de controle baseado nos desvios dos valores observados sobre a linha de regressão ($EP(Y) = S_e$) ou ainda na estimativa do erro padrão dada pelo erro padrão de um valor predito da variável dependente ($EP(\hat{Y}_t)$). Por conseguinte, caso o processo apresente-se sob controle estatístico há um bom tempo, isto é, caso se verifique que o processo é “refinado”, faz-se necessário a utilização do erro padrão da linha de regressão estimada $EP(\hat{Y}_t)$, como estimador do desvio padrão σ , ou seja, uma estimativa da variabilidade da reta de regressão (variabilidade da média), o que objetiva-se em gráficos de controle que monitoram o nível do processo, tal qual, é recomendado nos gráficos de controle de Shewhart. Utilizando a estimativa do erro padrão sobre um valor predito de Y , verifica-se que o mesmo é mais preciso, visto que, é a estimativa que produz limites de controle mais largos e portanto, menos sujeitos a alarmes falsos (supor que o processo está fora de controle estatístico quando na verdade ele não está), pois trata de um estimador da variabilidade da equação de regressão de um valor predito da variável dependente.

5.3.2 Gráfico de Controle de Regressão Estrutural para os Dados de uma Indústria Brasileira.

Considera-se a variável densidade aparente (Y) e peso (X) do eletrodo como sendo relacionadas de forma linear. Devido ao fato de que essas variáveis são oriundas de processos de medições, então imagina-se que as mesmas apresentem erros de medidas. Logo, a relação entre essas duas variáveis pode ser expresso através do Modelo de Regressão Estrutural, o qual, é definido pela Equação (3.1). A partir de informações obtidas junto a gerência de processos dessa indústria, fixa-se a variância do erro de medida da variável X , em $\sigma_s^2 = 10kg$. Apresenta-se as medidas da Densidade (Y) e do Peso (X), na Tabela 5.3.

Tabela 5.3 - Dados reais fornecidos por uma Indústria Brasileira sobre a Densidade Aparente e Peso dos Eletrodos produzidos de Agosto/2001 à Janeiro/2003

Ordem	Meses/Ano	X(Peso em kg)	Y(Densidade Aparente em g/cm^3)
1	Agosto/2001	976	1,623
2	Setembro/2001	1004	1,685
3	Outubro/2001	973	1,611
4	Novembro/2001	982	1,659
5	Dezembro/2001	973	1,611
6	Janeiro/2002	975	1,623
7	Fevereiro/2002	993	1,676
8	Março/2002	974	1,612
9	Abril/2002	976	1,625
10	Mai/2002	1000	1,683
11	Junho/2002	977	1,613
12	Julho/2002	989	1,645
13	Agosto/2002	995	1,677
14	Setembro/2002	984	1,643
15	Outubro/2002	985	1,626
16	Novembro/2002	979	1,619
17	Dezembro/2002	1001	1,655
18	Janeiro/2003	980	1,621

A partir dos dados de medidas da Densidade Aparente (Y) e do Peso (X), apresentados na Tabela 5.3, gerou-se um diagrama de dispersão, dado na Figura 5.5

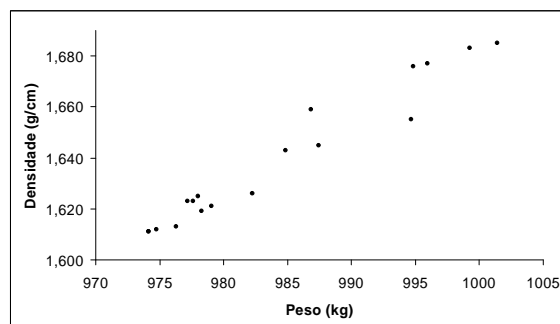


Figura 5.5 - Diagrama de Dispersão entre Densidade Aparente e Peso

Observando-se a Figura 5.5 não se verifica nenhum ponto extremo (*outlier*).

Obtêm-se a seguir as estatísticas necessárias à construção do gráfico de controle de regressão Estrutural, as quais apresentam-se na Tabela 5.4.

Tabela 5.4 - Estatísticas obtidas a partir da Tabela 5.3 para o Modelo Estrutural com $\sigma_\delta^2 = 10$.

Símbolo	Nome	Valor Obtido
N	Número de pares de valores	18
EP(Y)	Erro padrão da Densidade Aparente	0,0012
\bar{Y}	Densidade Aparente Média	1,639
S_{YY}	Desvio padrão da Densidade Aparente	0,000714
\bar{X}	Peso Médio do Eletrodo	984,22000
S_{XX}	Desvio padrão do Peso	106,88890
S_{XY}	Covariância entre Densidade e Peso do Eletrodo	0,24963
ℓ	Resíduo médio quadrado de U_i	85,45990
S_ν^2	Variância ou resíduo médio quadrático do modelo	0,000041
σ_ε^2	Variância do erro ε	0,000071
σ_u^2	Variância do erro U	96,88890
$\hat{\alpha}$	Intercepto	-0,89843
$\hat{\beta}$	Inclinação	0,00258
EP($\hat{\alpha}$)	Erro padrão do intercepto	0,19820
EP($\hat{\beta}$)	Erro padrão da inclinação	0,00020

Com base nos valores apresentados na Tabela 5.4, verifica-se que as condições de regularidade quanto à $S_{XX} > \sigma_\delta^2$ e $S_{YY}(S_{XX} - \sigma_\delta^2) - S_{XY}^2 > 0$, são atendidas. Como $S_{XX} > \sigma_\delta^2$ e $S_{YY}(S_{XX} - \sigma_\delta^2) - S_{XY}^2 > 0$, pode-se obter as estimativas de β e α a partir das equações (3.6) e (3.7), respectivamente. Para a construção do preditor \tilde{U}_i (vetor aleatório), ou seja, a construção do verdadeiro valor de U_i utiliza-se a Equação (3.45), supondo que o mesmo é aleatório e com distribuição normal. A estimativa do vetor de médias é dada por

$$(\hat{\mu}_Y, \hat{\mu}_X)^T = (\bar{Y}, \bar{X})^T = (1, 639; 984, 22)^T.$$

Baseado em (3.46), chega-se à:

$$\begin{aligned} \begin{bmatrix} \tilde{\gamma}_2 \\ \tilde{\gamma}_3 \end{bmatrix} &= \begin{bmatrix} S_{YY} & S_{YX} \\ S_{XY} & S_{XX} \end{bmatrix}^{-1} \begin{bmatrix} S_{XY} \\ S_{XX} - \sigma_\delta^2 \end{bmatrix} \\ &= \begin{bmatrix} 0,00071 & 0,24963 \\ 0,24963 & 106,8889 \end{bmatrix}^{-1} \begin{bmatrix} 0,24963 \\ 106,8889 - 10 \end{bmatrix} \\ &= \begin{bmatrix} 179,635 \\ 0,452 \end{bmatrix}; \end{aligned}$$

Assim, obtém-se o termo independente do modelo de estimação do verdadeiro valor de U_i , isto é,

$$\tilde{\gamma}_1 = (1 - \tilde{\gamma}_3) \bar{X} - \tilde{\gamma}_2 \bar{Y} = (1 - 0,452) 984,22 - 179,635 \cdot 1,639 = 244,932.$$

Portanto, concluí-se que a equação de estimação do verdadeiro valor de U_i , dado (Y_i, X_i) e considerando U_i como sendo aleatório é

$$\tilde{U}_i = 244,932 + 179,635Y_i + 0,452X_i.$$

A partir de (3.48), obtém-se a variância estimada do erro de predição como

$$\hat{V}\{\tilde{U}_i - U_i/(Y_i, X_i)\} = 8,19.$$

Para se estabelecer a linha central do gráfico de controle se faz necessário a obtenção dos verdadeiros valores dos U_i , conforme é apresentado na Tabela 5.5.

Tabela 5.5 - *Peso do Eletrodo estimado para o Modelo Estrutural com $\sigma_\delta^2 = 10$*

Local	X(Peso do Eletrodo em kg)	Y(Densidade em g/cm^3)	$\tilde{U}_{i,s}$ (Peso estimado)
1	976	1,623	977,632
2	1004	1,685	1001,425
3	973	1,611	974,120
4	982	1,659	986,810
5	973	1,611	974,120
6	975	1,623	977,180
7	993	1,676	994,836
8	974	1,612	974,752
9	976	1,625	977,991
10	1000	1,683	999,258
11	977	1,613	976,287
12	989	1,645	987,459
13	995	1,677	995,920
14	984	1,643	984,840
15	985	1,626	982,238
16	979	1,619	978,269
17	1001	1,655	994,680
18	980	1,621	979,080

A partir da Tabela 5.5, pode-se estabelecer a equação estimada da linha de regressão para o Modelo Estrutural, obtendo-se a linha central (LC) do Gráfico de Controle de Regressão Estrutural, dada por

$$LC = \hat{Y}_i = -0.898 + 0.002578\tilde{U}_i.$$

Após a obtenção da linha central (LC) e fixando-se o valor de $k = 2$, ou seja, com limites de controle fixados em 95,46% de confiança, torna-se possível obter o limite superior de controle (LSC) e o limite inferior de controle (LIC).

Os Gráficos de Controle de Regressão Estrutural com limites de controle fixados em 95,46% de confiança, para os três estimadores de σ apresentados nas Equações (3.58), (3.59) e (3.61), são mostrados nas Figuras 5.6 e 5.7.

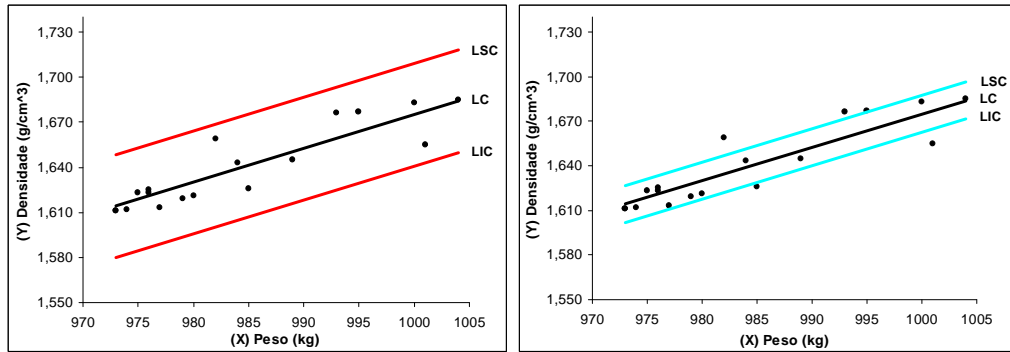


Figura 5.6 - Gráfico de Controle de Regressão Estrutural com $\sigma_{\hat{Y}_i}^2 = 10$ da Densidade Aparente e Peso do Eletrodo. Com Limites de $2EP(Y) = 2S_e$ e $2EP(\hat{Y}_i) = 2S_{\hat{Y}_i}$, respectivamente

Para a construção dos limites de controle do primeiro gráfico da Figura 5.6 utilizou-se o erro padrão da linha de regressão ($EP(Y)$), como estimador do desvio padrão de σ , ou seja, obteve-se parte da medida de variabilidade em torno da linha central, medida a qual, é multiplicada por uma constante “ k ” = 2 (95,46% de confiança). Pode-se perceber que todos os pontos “plotados” no primeiro gráfico estão dentro dos limites de controle, o que demonstra um indicativo de que o processo está sob controle estatístico. Entretanto, ao utilizar-se o erro padrão da linha de regressão estimada $EP(\hat{Y}_i)$, no segundo gráfico da Figura 5.6 onde o objetivo é controlar a variação média de um processo, ou seja, a esperança condicional da variável dependente (Y) dado uma variável independente (X),

neste caso os limites de controle obtidos são diferentes dos obtidos no primeiro gráfico da Figura 5.6, como pode ser observado e, diferentemente do primeiro gráfico, esta estimativa de σ não conseguiu monitorar de forma satisfatória este processo, visto que, observa-se pontos fora dos limites superior e inferior de controle, fato o qual não ocorre com o primeiro gráfico da Figura 5.6. No primeiro gráfico da Figura 5.7 não é notada a presença de pontos fora dos limites de controle, superior e inferior, como no gráfico anterior onde se utilizou outra estimativa de σ , o que levá-nos a concluir que este processo está sob controle estatístico e que esta estimativa da variabilidade adotada ($2S_{\hat{Y}_{pi}}$), assim como a estimativa adotada no gráfico anterior ($EP(Y)$), conseguem monitorar de forma satisfatória este processo.

Na Figura 5.7 pode-se observar um processo sob controle estatístico. Para a sua construção utilizou-se como estimador da variabilidade (estimador do desvio padrão σ), o erro padrão da previsão ($2EP(\hat{Y}_t) = 2S_{\hat{Y}_{pi}}$), objetivando-se obter linhas de controle mais exatas. Para se verificar a diferença gerada pela escolha, de um dos três estimadores de σ apresentados, durante a construção do Gráfico de Controle de Regressão Estrutural, construiu-se na Figura 5.7, um gráfico comparativo o qual sintetiza a Figura 5.6 e o primeiro gráfico da Figura 5.7 apresentadas anteriormente. Faz-se necessário destacar que a escolha da estimativa do desvio padrão será influenciada pelo tipo de processo que se tem como objetivo controlar. Verifica-se que, ao se analisar a Figura 5.7, que o gráfico construído utilizando-se o erro padrão de um valor predito da variável dependente, é o que melhor se aplica a este processo de produção, visto que, dentre as três estimativas de erro utilizadas, além deste conseguir compreender todos os pontos observados (assim como o primeiro gráfico da Figura 5.6), também produz os limites de controle mais largos, com isto, uma eventual situação de anormalidade (causa especial ou aleatória), será mais facilmente controlada por estes limites, conseqüentemente a chance de alarmes falsos ocorrerem ao longo do monitoramento deste processo serão menores, por tanto, a opção por este terceiro estimador de σ é a mais adequada para o processo da Densidade Aparente, em relação ao Peso do Eletrodo.

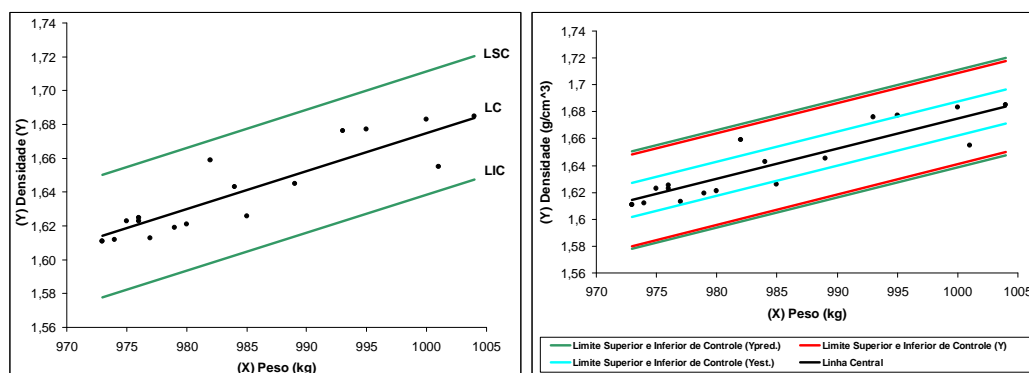


Figura 5.7 - Gráfico de Controle de Regressão Estrutural com $\sigma_{\hat{\delta}}^2 = 10$ da Densidade Aparente e Peso do Eletrodo. Com Limites de $2EP(\hat{Y}_t) = 2S_{\hat{Y}_t}$ e o Gráfico Comparativo com $\sigma_{\hat{\delta}}^2 = 10$, respectivamente.

No segundo gráfico da Figura 5.7, percebe-se que caso o processo de controle esteja em seu início, pode fazer-se a opção pelos limites de controle baseado nos desvios dos valores observados sobre a linha de regressão ($EP(Y) = S_e$), ou ainda na estimativa do erro padrão dada pelo erro padrão de um valor predito da variável dependente ($EP(\hat{Y}_t)$).

Na Figura 5.7, ao utilizar-se a estimativa do erro padrão sobre um valor predito de Y , faz-se a opção pelos limites de controle baseado na estimativa do erro padrão dada pelo erro padrão de um valor predito da variável dependente ($EP(\hat{Y}_t)$). Por conseguinte, caso o processo apresente-se sob controle estatístico há um longo período de tempo, isto é, caso se verifique que o processo é “refinado”, faz-se necessário a utilização do erro padrão da linha de regressão estimada $EP(\hat{Y}_i)$, como estimador do desvio padrão σ , ou seja, uma estimativa da variabilidade da reta de regressão (variabilidade da média), o que objetiva-se em gráficos de controle que monitoram o nível do processo, tal qual, é recomendado nos gráficos de controle de Shewhart. Utilizando a estimativa do erro padrão sobre um valor predito de Y , verifica-se que o mesmo é mais preciso, visto que, é a estimativa que produz limites de controle mais largos e portanto, menos sujeitos a alarmes falsos, pois trata de um estimador da variabilidade da equação de regressão de um valor predito da variável Y .

Comprovou-se assim, que o gráfico de controle de regressão estrutural mostrou-se uma ferramenta de controle capaz de medir o desempenho, utilizando-se para isto do fato de que, a partir do momento em que as variações ocorram, certamente em virtude de causas consideradas comuns ou “aleatórias” (ambas originárias da variabilidade natural do processo), estas devem ser aceitas pois considera-se que o processo está sob controle estatístico, ou seja, desde que as investigações não sejam provavelmente para produzir tomadas de decisões incorretas. Isto caracteriza o princípio do gráfico de controle convencional e é conhecido como princípio do gerenciamento pela execução.

5.4 Comparação entre os Estimadores de MQO e do Modelo Estrutural

5.4.1 Comparação entre os EMQO e do Modelo Estrutural para o Exemplo de Fuller (1987).

Esta seção tem a finalidade de comparar os Estimadores de Mínimos Quadrados Ordinários (EMQO), com os estimadores do Modelo Estrutural, os quais são necessários à obtenção dos limites de Controle no Gráfico de Controle de Regressão. A partir da Tabela 4.1 obtêm-se as seguintes estimativas pelo Método do Mínimos Quadrados Ordinários.

Tabela 5.6 - *Estimativas obtidas por EMQO para os dados da Tabela 4.1.*

Estimativas	$\hat{\alpha}$	$\hat{\beta}$	$EP(\hat{\alpha})$	$EP(\hat{\beta})$	$EP(Y)$
Valor	73,153	0,344	9,951	0,137	7,570

A Tabela 5.7 apresenta estimativas de α , β , $EP(\alpha)$, $EP(\beta)$ e $EP(Y)$, para diferentes valores de σ_{δ}^2 , para o Modelo Estrutural utilizando-se o conjunto de dados da Tabela 4.1.

Tabela 5.7 - Estimativas obtidas a partir da Tabela 4.1 para o Modelo Estrutural com diferentes valores de σ_δ^2 .

σ_δ^2	$\hat{\alpha}$	$\hat{\beta}$	$EP(\hat{\alpha})$	$EP(\hat{\beta})$	$EP(Y)$
1	73,073	0,345	9,951	0,137	7,571
5	72,748	0,349	9,955	0,137	7,572
10	72,329	0,356	9,966	0,137	7,574
15	71,895	0,362	9,986	0,138	7,578
20	71,447	0,368	10,016	0,138	7,584
25	70,982	0,375	10,056	0,139	7,592
30	70,500	0,381	10,107	0,139	7,603
35	70,001	0,389	10,418	0,148	7,615
40	69,483	0,396	10,457	0,144	7,631
45	68,944	0,404	10,503	0,145	7,650
50	68,385	0,411	10,557	0,146	7,672
57	67,560	0,420	10,648	0,147	7,710
60	67,198	0,428	10,693	0,148	7,728
65	66,567	0,437	10,777	0,149	7,763
70	65,909	0,446	10,872	0,150	7,803
75	65,223	0,456	10,980	0,152	7,848
80	64,507	0,466	11,102	0,154	7,899
85	63,757	0,477	11,239	0,156	7,957
90	62,973	0,488	11,394	0,158	8,022
95	62,152	0,500	11,566	0,160	8,095
100	61,290	0,512	11,757	0,163	8,178

As Tabelas 5.6 e 5.7 mostram o efeito dos erros de medida sobre os estimadores de mínimos quadrados. Nota-se que o *EMQO* é afetado pelo erro de medida, isto é, a medida que aumenta o erro associado à variável independente (X), diminui a capacidade dos estimadores de mínimos quadrados em detectá-los, implicando com isso em estimativas inconsistentes, ou seja, estimativas da variabilidade cada vez mais distantes das verdadeiras estimativas dos dados analisados. Por conseguinte, essa inconsistência dos estimadores de mínimos quadrados, implica que ao construir-se Gráficos de Controle de Regressão utilizando essas estimativas, os mesmos apresentarão limites muito estreitos, o que pode vir a gerar a ocorrência de alarmes falsos, ou seja, imaginar-se que o processo em análise está fora de controle estatístico, quando na verdade ele está sob controle estatístico.

5.4.2 Comparação entre os EMQO e do Modelo Estrutural para o Exemplo de Mandel (1969).

Mandel (1969), apresenta um estudo para controlar e processar correspondências nas 74 maiores agências de correios dos EUA, analisando a relação entre o número de homens-hora trabalhados e o volume de correspondência processada. Os dados são mostrados na Tabela 5.8; Onde X : é a quantidade de correspondência manipulada (em milhões) e Y : é o número de homens-hora (em milhares).

Tabela 5.8 - *Dados sobre Horas e Volume de Processamento de Correspondência para as Agências de Correio A, B, C e D, nos anos fiscais de 1962 e 1963.*

Período de 4 semanas	X (Ano Fiscal de 1962)	Y (Ano Fiscal de 1962)	Período de 4 semanas	X (Ano Fiscal de 1963)	Y (Ano Fiscal de 1962)
1	157	572	1	154	569
2	161	570	2	157	564
3	168	645	3	164	573
4	186	645	4	188	667
5	183	645	5	191	700
6	184	671	6	180	765
7	268	1053	7	270	1070
8	180	675	8	180	637
9	175	670	9	172	650
10	193	710	10	184	655
11	184	656	11	179	665
12	179	640	12	169	599
13	164	599	13	160	605

Fonte: Mandel (1969).

A partir da Tabela 5.8 obtêm-se as seguintes estimativas pelo método do mínimos quadrados ordinários.

Tabela 5.9 - *Estimativas obtidas por EMQO para os dados da Tabela 5.8.*

Estimativas	$\hat{\alpha}$	$\hat{\beta}$	$EP(\hat{\alpha})$	$EP(\hat{\beta})$	$EP(Y)$
Valor	50,44	3,345	59,46	0,34	18,92

E com o mesmo conjunto de dados, apresentado na Tabela 5.8, para o Modelo Estrutural obtêm-se as estimativas mostradas na Tabela 5.10.

Tabela 5.10 - Estimativas obtidas a partir da Tabela 5.8 para o Modelo Estrutural com diferentes valores de σ_δ^2 .

σ_δ^2	$\hat{\alpha}$	$\hat{\beta}$	$EP(\hat{\alpha})$	$EP(\hat{\beta})$	$EP(Y)$
1	46,265	3,369	59,488	0,340	18,932
5	28,953	3,469	60,052	0,344	18,989
10	5,826	3,601	61,948	0,354	19,183
15	-19,141	3,744	65,507	0,375	19,540
20	-42,175	3,899	71,227	0,408	20,089
25	-75,543	4,068	79,815	0,457	20,864

Os *EMQO*, referentes aos dados da Tabela 4.1, assim como à Tabela 5.8, vieram ratificar o que já havia sido discutido no “Capítulo 2” deste trabalho, isto é, comprovaram a teoria citada anteriormente quanto à inconsistência dos estimadores de mínimos quadrados, em virtude da incapacidade dos mesmos em detectar os erros associados à variável independente X , a medida que o tamanho da amostra analisada aumenta, ou seja, a partir de que, a variância do erro associado a variável X aumenta, diminui a capacidade dos *EMQO*, em detectá-los, fato o qual não ocorre com os estimadores do Modelo de Regressão Estrutural, logo, concluí-se que os *EMQO* não funcionam de forma satisfatória na presença de erros de medida sobre tudo na variável independente.

Capítulo 6

Considerações Finais e Recomendações

Objetivou-se neste trabalho a apresentação de uma nova metodologia com o intuito de elaborar gráficos de controle aplicados ao modelo de regressão com variáveis sujeitas a erros de medida, no caso a variável independente. Para tal, combinou-se a teoria proposta por Mandel (1969) no que se refere a gráficos de controle aplicados ao modelo clássico de regressão, com a teoria do modelos Estrutural de regressão, apresentada por Cheng e Van Ness (1999).

Com base nos resultados apresentados neste trabalho, pode-se comprovar que os estimadores obtidos através do método de mínimos quadrados, mostraram-se influenciados pelo erro de medida associado a variável independente, resultando com isso em uma diminuição dos valores estimados da variabilidade do processo e, contudo, comprometendo as estimativas dos limites dos gráficos de controle (inferior e superior), gerando com isso limites de controle mais estreitos do que realmente deveriam ser, ocasionando portanto, uma baixa eficiência dos limites de controle do gráfico de regressão, o qual apresentará uma alta probabilidade de ocorrência do erro tipo I (supor que o processo está fora de controle estatístico quando na verdade ele não está).

Os limites do Gráfico de Controle de Regressão, tanto com a variável independente medida com erro (Modelo Estrutural), tal qual, no Modelo Clássico de Regressão, são influenciados pela escolha do estimador da variabilidade ($\hat{\sigma}$), o qual se utiliza para a construção dos limites de controle como foi possível concluir. No caso dos modelos de regressão erro nas variáveis (Estrutural e Funcional) apresentados neste trabalho, a estimativa do erro padrão de previsão dada pela Equação (3.61), produz limites de controle mais largos do que as demais estimativas para $\hat{\sigma}$, implicando com isso em em uma diminuição da chance de ocorrência de alarmes falsos.

Conseguiu-se constatar ao longo da execução deste trabalho que, a aplicação em dados reais ilustra a potencialidade do gráfico de controle de regressão estrutural proposto como ferramenta eficaz de controle da qualidade e de medida de desempenho de um processo.

6.1 Sugestões Para Trabalhos Futuros

Como sugestões para trabalhos futuros, pode-se destacar:

- Propor uma metodologia de obtenção dos limites de controle no caso de modelos de calibração com e sem erros na variável independente;
- Criar índices de capacidade do processo para os modelos com erros nas variáveis;
- Investigar e propor alternativas em modelos de regressão sem a suposição de normalidade dos erros de medida, para o estabelecimento de gráficos de controle;
- Estender os resultados apresentados neste trabalho para modelos multivariados;
- Desenvolver uma metodologia de gráficos de controle em regressão quantílica.

Apêndice A

ANEXO

A.1 Introdução

Neste apêndice, será apresentado a obtenção das esperanças das somas dos erros ε_i e δ_i , bem como a obtenção das esperanças dos produtos destas somas, com o intuito de se obter as esperanças e variâncias das variáveis aleatórias X_i, Y_i e $X_i Y_i$, para todo $i=1, 2, \dots, n$, onde as variáveis X_i e Y_i estão definidas no modelo (3.1).

A.2 Expansão das Somas

Pode-se observar que os somatórios a seguir podem ser expressos de forma que:

$$\begin{aligned} & \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n} = \\ & = \frac{1}{n} (\alpha \sum_{i=1}^n U_i + \beta \sum_{i=1}^n U_i^2 + \sum_{i=1}^n U_i \varepsilon_i + \alpha \sum_{i=1}^n \delta_i + \beta \sum_{i=1}^n U_i \varepsilon_i + \sum_{i=1}^n \delta_i \varepsilon_i) \quad (\text{A.1}) \end{aligned}$$

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n (\sum_{i=1}^n U_i^2 + \sum_{i=1}^n \delta_i^2 - (\sum_{i=1}^n U_i)) \quad (\text{A.2})$$

$$\frac{1}{n} \sum_{i=1}^n X_i Y_i = \frac{1}{n} (\alpha \sum_{i=1}^n U_i + \beta \sum_{i=1}^n U_i^2) \quad (\text{A.3})$$

$$\frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} (\alpha + \beta \sum_{i=1}^n U_i + \sum_{i=1}^n \varepsilon_i) \quad (\text{A.4})$$

$$\frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} (\sum_{i=1}^n U_i + \sum_{i=1}^n \delta_i) \quad (\text{A.5})$$

$$\frac{1}{n} \sum_{i=1}^n X_i^2 = \frac{1}{n} (\sum_{i=1}^n U_i^2 + 2 \sum_{i=1}^n U_i \delta_i + \sum_{i=1}^n \delta_i^2) \quad (\text{A.6})$$

$$(X_1 + \dots + X_k)^n = \sum \frac{n!}{n_1!n_2!n_3!\dots n_k!} X_1^{n_1} X_2^{n_2} X_3^{n_3} \dots X_k^{n_k}. \quad (\text{A.7})$$

onde $n_1 + \dots + n_k = n$.

$$\left(\sum_{i=1}^n U_i \varepsilon_i\right)^2 = \sum_{i=1}^n U_i^2 \varepsilon_i^2 + 2 \sum_{i < j=1}^n U_i \varepsilon_i U_j \varepsilon_j. \quad (\text{A.8})$$

A.3 Obtenção das Esperanças

Para obter-se as esperanças, variâncias e covariâncias, necessita-se utilizar as propriedades das somas das esperanças. Portanto, como já havia sido citado anteriormente no capítulo 3, os erros são definidos como independentes e identicamente distribuídos com média zero e variâncias constantes. E sejam os quatro primeiros momentos finitos definidos como:

$$E(\varepsilon_i^1) = 0; \quad E(\varepsilon_i^2) = \sigma_\varepsilon^2; \quad E(\varepsilon_i^3) = \sigma_\varepsilon^3; \quad E(\varepsilon_i^4) = \sigma_\varepsilon^4;$$

$$E(\delta_i^1) = 0; \quad E(\delta_i^2) = \sigma_\delta^2; \quad E(\delta_i^3) = \sigma_\delta^3; \quad E(\delta_i^4) = \sigma_\delta^4;$$

$$E(U_i^1) = \sigma_u^1; \quad E(U_i^2) = \sigma_u^2; \quad E(U_i^3) = \sigma_u^3; \quad E(U_i^4) = \sigma_u^4;$$

Conseqüentemente, verifica-se os seguintes resultados:

$$Var(\varepsilon_i) = E(\varepsilon_i^2) - (E(\varepsilon_i))^2 = \sigma_\varepsilon^2;$$

$$Var(\delta_i) = E(\delta_i^2) - (E(\delta_i))^2 = \sigma_\delta^2;$$

$$Cov(\varepsilon_i, \delta_i) = E(\varepsilon_i \delta_i) - E(\varepsilon_i)E(\delta_i) = E(\varepsilon_i)E(\delta_i) - E(\varepsilon_i)E(\delta_i) \Rightarrow Cov(\varepsilon_i \delta_i) = 0.$$

Portanto, utilizando-se as expressões dos somatórios da seção anterior e as suposições acima tem-se que:

$$E\left(\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}\right) = \beta\ell = E(S_{XY});$$

$$E\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}\right) = \sigma_\delta^2 + \ell = E(S_{XX});$$

$$E\left(\frac{\sum_{i=1}^n X_i Y_i}{n}\right) = \alpha\bar{U} + \beta\frac{\sum_{i=1}^n U_i^2}{n};$$

$$E\left(\frac{\sum_{i=1}^n Y_i}{n}\right) = \alpha + \beta\bar{U};$$

$$E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n}E\left(\sum_{i=1}^n (U_i + \delta_i)\right) = \frac{1}{n}\sum_{i=1}^n E(U_i) + \frac{1}{n}\sum_{i=1}^n E(\delta_i) = \sigma_u^1;$$

$$E\left(\frac{\sum_{i=1}^n X_i^2}{n}\right) = \frac{\sum_{i=1}^n U_i^2}{n} + \frac{\sum_{i=1}^n \delta_i^2}{n};$$

$$Var(S_{XX}) = 4\ell\frac{\sigma_\delta^2}{n} + \frac{2\sigma_\delta^4}{n} + O(n^{-2});$$

$$Var(S_{XY}) = \frac{\sigma_\varepsilon^2\sigma_\delta^2}{n} + \ell\left(\frac{\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2}{n}\right) + O(n^{-2})$$

$$\text{e } Cov(S_{XY}, S_{XX}) = 2\beta\ell\frac{\sigma_\delta^2}{n},$$

$$\text{onde } \ell = \frac{\sum_{i=1}^n (U_i - \bar{U})^2}{n-1}.$$

Apêndice B

ANEXO

B.1 Ordens de Grandeza de Seqüências de Números Reais: $O(\cdot)$, $o(\cdot)$.

B.1.1 - Introdução

Neste apêndice, será apresentado um problema frequente no estudo dos métodos assintóticos consiste na comparação de seqüências de variáveis aleatórias. “A base para o estudo desse problema é a análise comparativa do *comportamento* de duas Seqüências de números reais $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$, quando n tende ao infinito. Nesta Seção apresenta-se alguns resultados sobre esse tema. O caso típico é aquele onde os termos gerais das seqüências de interesse não são dados explicitamente, mas uma ou algumas de suas propriedades é (são) conhecida (s). Por exemplo, seja $a_n = b_n^2$, $n \geq 1$ e supõem-se que $\{b_n\}_{n \geq 1}$ seja limitada para todo n suficientemente grande, ou mais especificamente, que existam um número real $K > 0$ e o número inteiro positivo $n_0 = n_0(K)$, tal que, $|b_n| \leq K, \forall n \geq n_0$. O problema é o de deduzir propriedades da seqüência $\{a_n\}_{n \geq 1}$. Neste caso, é claro que, $|a_n| \leq K^2, \forall n \geq n_0$ e, conseqüentemente, pode-se deduzir que $\{a_n\}_{n \geq 1}$, também é limitada para todo n suficientemente grande. Para se estudar situações mais complexas convém apresentar a seguinte definição:

B.1.2 - Definição

Sejam $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$ seqüências de números reais, logo diz-se que

- (a) $a_n = O(b_n)$ se existirem um número real $K > 0$ e um número inteiro positivo $n_0 = n_0(K)$ tal que $|a_n/b_n| \leq K, \forall n \geq n_0$,
- (b) $a_n = o(b_n)$ se para todo $\epsilon > 0$ existir um número inteiro positivo $n_0 = n_0(\epsilon)$ tal que $|a_n/b_n| < \epsilon, \forall n \geq n_0$.

Em outras palavras, diz-se que $a_n = O(b_n)$ se a razão $|a_n/b_n|$ for limitada para todo n suficientemente grande e que $a_n = o(b_n)$ se $a_n/b_n \rightarrow 0$ quando $n \rightarrow \infty$. Em particular, $a_n = O(1)$ se existir um número real $K > 0$ tal que $|a_n| \leq K$ para todo n suficientemente grande e $a_n = o(1)$ se $a_n \rightarrow 0$ quando $n \rightarrow \infty$. Essencialmente, a Definição B.3 está relacionada com a comparação das *ordens de grandeza* (ou de magnitude) de $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$. Portanto, afirmar que $a_n = O(b_n)$ corresponde a dizer que a ordem de grandeza de $\{a_n\}_{n \geq 1}$ é, *no máximo*, igual à de $\{b_n\}_{n \geq 1}$ para todo n suficientemente grande, de maneira similar, afirmar que $a_n = o(b_n)$ corresponde a dizer que a ordem de grandeza de $\{a_n\}_{n \geq 1}$ é menor que a de $\{b_n\}_{n \geq 1}$, para todo n suficientemente grande.” (Leite e Singer, 1990).

Exemplo:

(1) $n = o(n^2)$, pois $\frac{n}{n^2} = \frac{1}{n} \rightarrow 0$, quando $n \rightarrow \infty$.

(2) $n^{-1} = o(1)$, pois $\frac{n^{-1}}{1} = \frac{1}{n} \rightarrow 0$, quando $n \rightarrow \infty$.

(3) $10n^2 + n = O(n^2)$, pois $\frac{10n^2 + n}{n^2} = 10 + \frac{1}{n} \leq 11, \forall n \geq 1$.

(4) $n^2 = O(6n^2 + n)$, pois $\frac{n^2}{6n^2 + n} < \frac{n^2}{6n^2} \leq \frac{1}{6}, \forall n \geq 1$.

(5) $n = O(n^2)$, pois $\frac{n}{n^2} = \frac{1}{n} \leq 1, \forall n \geq 1$.

(6) $3n = o(n^2)$, pois $\frac{3n}{n^2} = \frac{3}{n} \rightarrow 0$, quando $n \rightarrow \infty$.

Leite e singer (1990) apresentam o Teorema B.5 com alguns resultados envolvendo a notação $O(\cdot)$ e $o(\cdot)$ os quais são bastante úteis nas aplicações.

B.1.3 - Teorema

Sejam $\{a_n\}_{n \geq 1}$, $\{b_n\}_{n \geq 1}$, $\{c_n\}_{n \geq 1}$ e $\{d_n\}_{n \geq 1}$ seqüências de números reais,

(1) Se $a_n = o(b_n)$, então $a_n = O(b_n)$,

(2) Se $a_n = O(b_n)$ e $c_n = O(d_n)$, então:

(a) $a_n c_n = O(b_n d_n)$,

(b) $|a_n|^s = O(|b_n|^s)$, para todo $s > 0$,

(c) $a_n + c_n = O(\max\{|b_n|, |d_n|\})$,

(3) Se $a_n = o(b_n)$ e $c_n = o(d_n)$, então:

(a) $a_n c_n = o(b_n d_n)$,

(b) $|a_n|^s = o(|b_n|^s)$, para todo $s > 0$,

(c) $a_n + c_n = o(\max\{|b_n|, |d_n|\})$,

(4) Se $a_n = O(b_n)$ e $c_n = o(d_n)$, então $a_n c_n = o(b_n d_n)$,

(5) Se $a_n = O(b_n)$ e $b_n = o(c_n)$, então $a_n = o(c_n)$.

Pode-se verificar as provas para os itens 2, 4 e 5 do Teorema B.5 em Leite e Singer, 1990, *pg.* (11).

Uma extensão do estudo das ordens de grandeza de seqüências de números reais pode ser verificado na forma de exemplos e exercícios, assim como ordens de grandeza de seqüências de vetores em Leite e Singer 1990, *pg.* (9)-(14).

B.2 Expansões em Séries de Taylor.

Para Leite e Singer (1990), uma das ferramentas utilizadas com frequência no estudo de métodos assintóticos é a aproximação de uma função real f de variável real por um polinômio. Embora a expansão de f em série de potências em torno de ponto x_0 seja importante sob o ponto de vista teórico, o que realmente interessa nas aplicações é o estudo do resto proveniente da aproximação de f por um polinômio.

Dentre as possíveis aproximações de uma função f por um polinômio, segundo Leite e Singer (1990), a que se tornar relevante corresponde ao caso em que

- (a) f coincide com o polinômio em um determinado ponto x_0 ,
- (b) A k -ésima derivada de f coincide com a k -ésima derivada do polinômio no ponto x_0 , $k = 1, 2, \dots, n$. Por exemplo, se f for derivável no ponto x_0 , o polinômio P_1 de grau no máximo 1, tal que, $P_1(x_0) = f(x_0)$ e $P_1'(x_0) = f'(x_0)$ é dado por $P_1 = f(x_0) + f'(x_0)(x - x_0)$. Geometricamente, P_1 corresponde à reta tangente à curva associada à função f no ponto $(x_0, f(x_0))$. Se D_f for o domínio de f , então para todo $x \in D_f$, o resto ou erro $R_1(x)$ da aproximação de $f(x)$ por $P_1(x)$ é dado por

$$R_1(x) = f(x) - P_1(x) = f(x) - f(x_0) - f'(x_0)(x - x_0).$$

Então, para todo $x \neq x_0$, temos

$$\frac{R_1(x)}{(x - x_0)} = \frac{f(x) - f(x_0)}{(x - x_0)} - f'(x_0),$$

e portanto,

$$\lim_{x \rightarrow x_0} \frac{R_1(x)}{(x - x_0)} = f'(x_0) - f'(x_0) = 0,$$

isto é, o resto $R_1(x)$ tende a zero mais rapidamente que $x - x_0$ quando $x \rightarrow x_0$. Neste sentido o polinômio P_1 é uma aproximação de f em torno de x_0 e é denominado de **Polinômio de Taylor**, de ordem 1, de f em torno de x_0 . Quando $f''(x_0)$ existe, pode-se aproximar f por um polinômio P_2 tal que, $P_2(x_0) = f(x_0)$, $P_2'(x_0) = f'(x_0)$ e $P_2''(x_0) = f''(x_0)$, ou seja,

$$P_2(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2.$$

Nesse caso, a reta tangente associada a P_2 coincide com a reta tangente associada a f , no ponto $(x_0, P_2(x_0)) = (x_0, f(x_0))$, bem como P_2 e f tem a mesma “curvatura” no referido ponto. Nesse sentido, P_2 é uma melhor aproximação que P_1 para a função f no ponto x_0 , e é denominado **Polinômio de Taylor**, de ordem 2, de f em torno de x_0 .

Para todo $x \in D_f$ o resto da aproximação de f por P_2 é dado por

$$R_2(x) = f(x) - P_2(x) = f(x) - f(x_0) - f'(x_0)(x - x_0) - \frac{x_0}{2}(x - x_0)^2.$$

Então, para todo $x \neq x_0$,

$$\frac{R_2(x)}{(x - x_0)^2} = \frac{f(x) - f(x_0) - f'(x_0)(x - x_0) - \frac{f''(x_0)(x - x_0)^2}{2}}{(x - x_0)^2}$$

e como f é contínua em x_0 , temos

$$\lim_{x \rightarrow x_0} \frac{R_2(x)}{(x - x_0)^2} = \frac{0}{0}.$$

Pela regra de *L'Hospital* segue que

$$\begin{aligned} \lim_{x \rightarrow x_0} \frac{R_2(x)}{(x - x_0)^2} &= \lim_{x \rightarrow x_0} \frac{f'(x) - f'(x_0) - f''(x_0)(x - x_0)}{2(x - x_0)} \\ &= \frac{1}{2} \lim_{x \rightarrow x_0} \left[\frac{f'(x) - f'(x_0)}{(x - x_0)} - f''(x_0) \right] = \frac{1}{2} [f''(x_0) - f''(x_0)] = 0. \end{aligned}$$

Conseqüentemente, pode-se concluir que $R_2(x) \rightarrow 0$ mais rapidamente que $(x - x_0)^2$ quando $x \rightarrow x_0$.

Bibliografia

- ALMEIDA, S. S., *Desenvolvimento de Graficos de Controle Aplicados ao Modelo Funcional de Regressao*, Tese de Doutorado: UFSC, Florianópolis, 2003.
- ALMEIDA, S. S., LIMA, C.R.O.P. e SANDOVAL, M. C. *Linear calibration in functional models without the normality assumption*, Relatório Técnico (RT-MAE 2001-04), Departamento de Estatística IME-USP, p. 1-18, Fevereiro, 2001.
- ALWAN, L. C. e ROBERTS, H. V. *Times series modeling for statistical process control*, Journal of Business Economics Statistics, v. 16, p. 83-89, 1988.
- BARNETT, V. D. *A note on linear structural relationships When both residual variances are know*, Biometrics, v. 54, p. 670-672, 1967.
- BARNETT, V. D. *Simultaneous pairwise linear structural relationships*, Biometrics, v. 25, p. 129-142, 1969.
- BIRCH, M. W. *A note on the maximum likelihood estimation of a linear structural relationship*, Journal of the American Statistical Association, v. 59, p. 1175-1178, 1964.
- BOLFARINE, H., J. RODRIGUES e CORDANI, L. K., *O modelo de regressao com erros nas variaveis*. Texto do Minicurso do X SINAPE, São Paulo, 1992.
- BOWDEN, C. D. e GRAYBILL, A. F. *Confidence bands of uniform and proportional width for linear models*, American Statistical Association Journal, March, p. 182-198, 1966.
- BOWDEN, C. D. e GRAYBILL, A. F. *Linear segment confidence bands for simple linear models*, American Statistical Association Journal, June, p. 403-408, 1967.
- Carroll, R. J., Ruppert, D. e Stefanski, L. A. *Measurement Error in Nonlinear Models*, Chapman e Hall/CRC, 1995.
- CHENG, C. L. e VAN NESS, J. W. *On the unreplicated ultrastructural model*, Biometrika, v. 78, p. 442-445, 1991.
- CHENG, C. L. e VAN NESS, J. W. *On estimating linear relationships when both variables are subject to errors*, Royal Statistical Society, v. 56, n^o1, p. 167-183, 1994.
- CHENG, C. L. e VAN NESS, J. W. *Statistical Regression With Measurement Error*, New York, ARNOLD, 1999.
- DRAPER, N. R. e SMITH, H. *Applied Regression Analysis*, JOHN WILEY, New York, 3rd edition, 1998.

- DIPAOLA, P. P. *Use of correlation in quality control*, Industrial Quality Control, v. 2, n^o1, July, p. 10-14, 1945.
- DOLBY, G. R. *The ultrastructural relation: a synthesis of the functional and structural relations*, Biometrics, v. 63, p. 39-50, 1976.
- FULLER, W. A., *Measurement Error Models*. JOHN WILEY, New York, 1987.
- GLESER, L. J. *A note on G. R. Dolby's unreplicated model*, Biometrics, v. 72, p. 117-124, 1985.
- HWANG, J. T. *Multiplicative errors-in-variables models with applications to recent data released by the U. S. department of energy*, Journal of the American Statistical Association, v. 81, p. 680-688, 1986.
- JAMES, B. R. *Probabilidade: Um curso em nível intermediário*. 2^a edição, Instituto de Matemática Pura e Aplicada - IMPA, Rio de Janeiro, 1996.
- KANAZUKA, T. *The effect of measurement error on the power of $\bar{X} - R$ charts*, Journal of Quality Technology, v. 18, n^o2, April, p. 91-95, 1986.
- KENDALL, M. G. e STUART, A. *The Advanced Theory of Statistics*. v. 1, 2^a edição, London, GRIFFIN, 1979.
- LEITE, J. G. e SINGER, J. M. *Métodos Assintóticos em Estatística - Fundamentos e Aplicações*, In: 9^o Simpósio Brasileiro de Probabilidade e Estatística, ABE - IME - USP, São Paulo, 1990.
- LINDLEY, D. V. *Regression lines and the linear functional relationship*, Supplement to the Journal Royal Statistical Society, v. 9, p. 218-244, 1947.
- MANDEL, B. J. *The regression control chart*, Journal of Quality Technology, vol. 1, n^o1, p.1-9, January, 1969.
- MIAZAKI, E. S. e MONTEIRO, M. A. *Comparação de variáveis*, ATLAS da 3^a. Escola de Regressão - ABE, p.58-63, 1993.
- MITRA, A. *Quality Control and Improvement*, 2^a edição, PRENTICE HALL, Upper Saddle River, New Jersey, 1988.
- MOBERG, L. e SUNDBERG, R. *Maximum likelihood estimation of a linear functional relationship when one of the departure variances is known*, Scandinavian Journal of Statistics, v. 5, p. 61-64, 1978.
- MONTGOMERY, D. C. *Introductions to Statistical Quality Control*. New York, JOHN WILEY & SONS, 4th ed., 2001.
- MORAN, P. A. P. *Estimating Structural and functional relationships*. Journal of Multivariate Analysis, vol. 1, p. 232-255, 1971.
- NASCIMENTO, J. A. *Métodos de estimação consistente não bayesiano para o modelo linear de erro nas variáveis*, Dissertação de Mestrado, UFPE, Recife, 1983.

-
- NELSON, L. S. *Statistical Consulting in Manufacturing/Engineering*. Journal of Quality Technology, vol. 34, n^o4, October, 2002.
- NETER, KUTNER, NACHTSHEIM, and WASSERMAN, *Applied Linear Statistical Models*. RICHARD D. IRWIN, BURR RIDGE, Illinois, 4th ed., 1996.
- RAMOS, E. M. L. S.; SILVIA, L. S. C. V.; SAMOHYL, R. W.; ALMEIDA, S. S. *Utilização da amplitude amostral como estimador do desvio padrão: discussão e revisão teórica* In: 15^o Simpósio Nacional de Probabilidade Estatística, Águas de Lindóia, São Paulo, 29/07 a 02/08/2002. Anais do 15^o SINAPE, v. 1, p. 207, 2002.
- SHEWHART, W. A. *Economic Control of Quality of Manufactured Product*. New York, D. VAN NOSTRAND COMPANY, INC., 1931.
- SIQUEIRA, L. G. P. *Controle Estatístico do Processo*, Equipe Grifo, são Paulo, 1997.
- SOLARI, M. E. *The maximum likelihood solution to the problem of estimating a linear functional relationship*, Journal Royal Statistical Society, B, v. 31, p. 372-375, 1969.
- SPRENT, P. *Some historical of functional and structural relationships*, Contemporary Mathematics, v. 112, p. 3-15, 1990.
- STUART, A. e ORD, J. K. *Kendall's Advanced Theory of Statistics*, EDWARD ARNOLD, London, v. 1, 6th. ed., 1991