

**MONITORAÇÃO DO PROCESSO DE COLETAS DE
RESÍDUOS EM SANTA MARIA-RS USANDO GRÁFICOS
DE CONTROLE DE REGRESSÃO**

por

Luciane Flores Jacobi

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Engenharia de Produção, Área de concentração em Qualidade e Produtividade, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestre em Engenharia de Produção**

PPGEP

Santa Maria, RS - BRASIL

2001

Universidade Federal de Santa Maria

Centro de Tecnologia
Programa de Pós-Graduação em Engenharia de Produção

A Comissão Examinadora, abaixo assinada,
aprova a Dissertação de Mestrado.

**MONITORAÇÃO DO PROCESSO DE COLETAS DE
RESÍDUOS EM SANTA MARIA - RS, USANDO
GRÁFICOS DE CONTROLE DE REGRESSÃO**

elaborada por
Luciane Flores Jacobi

Como requisito parcial para obtenção do grau de
Mestre em Engenharia de Produção

COMISSÃO EXAMINADORA:

Adriano Mendonça Souza
(Presidente/Orientador)

João Eduardo da Silva Pereira
(Co-Orientador)

Anderson de Barros Dantas

Santa Maria, 17 de dezembro de 2001.

Ao meu esposo Elton, pelo acompanhamento, constante estímulo e compreensão em todas as horas.

Aos meus filhos, Otávio e Natália, perdão pela ausência nessa fase tão importante de suas vidas, todo o meu amor.

AGRADECIMENTO

Ao término deste trabalho, agradeço:

aos meus orientadores, Adriano e João Eduardo, pelo auxílio e sabedoria que souberam transmitir;

a todos os professores do PPGEP pela contribuição dada através de seus conhecimentos no decorrer do curso;

aos colegas do Departamento de Estatística que com muito carinho colaboraram, de alguma forma, para a realização deste trabalho;

a PRT pela disponibilidade;

aos membros da banca examinadora pelas sugestões;

a todos que, de alguma forma, compartilharam deste trabalho, o meu carinho.

SUMÁRIO

LISTA DE FIGURAS	vii
LISTA DE TABELAS	x
LISTA DE QUADROS	xi
LISTA DE ANEXOS	xii
LISTA DE SIGLAS	xiii
RESUMO	xiv
1 INTRODUÇÃO	01
1.1 Tema da Pesquisa	02
1.2 Justificativa e importância do trabalho	03
1.3 Objetivos	03
1.4 Metodologia	04
1.5 Delimitação da pesquisa	05
1.6 Organização do trabalho	05
2 REVISÃO DA LITERATURA	06
2.1 Introdução	06
2.2 Gráficos de controle	07
2.3 Análise de regressão	19
2.4 Comentários gerais do capítulo	45
3 GRÁFICO DE CONTROLE DE REGRESSÃO	46
3.1 Introdução	46
3.2 Previsão em regressão	47
3.3 Construindo o gráfico de controle de regressão	48
3.4 Comentários gerais do capítulo	53

4 APLICAÇÃO DA METODOLOGIA	44
4.1 Introdução	55
4.2 Descrição da PRT – empresa de prestação de serviço	57
4.3 Descrição do processo	59
4.4 Análise descritiva dos dados	63
4.5 Análise de correlação nos dados	70
4.6 Ajuste da equação de regressão e adequação do modelo	73
4.7 Estabelecendo o gráfico de controle de regressão	79
4.8 Comentários gerais do capítulo	83
5 CONCLUSÃO	85
6 REFERÊNCIAS BIBLIOGRÁFICAS	87
7 ANEXOS	91

LISTA DE FIGURAS

FIGURA 01 - Exemplos de gráfico de controle (a) Processo sob controle (b) Processo fora de controle	12
FIGURA 02 - Representação genérica do gráfico CUSUM sob a forma de máscara - V com $\alpha=\beta=0,05$ e $\delta=1$	16
FIGURA 03 - Representação genérica de um gráfico EWMA	17
FIGURA 04 - Representação genérica do gráfico de controle de regressão	18
FIGURA 05 - Modelo estatístico de uma regressão linear simples	24
FIGURA 06 - Representação das diferenças entre os valores estimados pela reta de regressão e os valores observados	25
FIGURA 07 - Representação da decomposição da soma de quadrados ..	32
FIGURA 08 - Bandas de confiança para ajuste de reta com $R^2 = 0,95$...	36
FIGURA 09 - Bandas de confiança para ajuste de reta com $R^2 = 0,25$...	37
FIGURA 10 - Derivação da distribuição amostral D_w , a estatística de Durbin-Watson	43
FIGURA 11 - Representação do gráfico de controle de regressão	50
FIGURA 12 - Logomarca da empresa	58

FIGURA 13 - Caminhões compactadores da empresa PRT - prestações de serviços de Santa Maria	60
FIGURA 14 - Gráfico "Box Plot" para as variáveis peso e quilometragem	66
FIGURA 15 - Gráfico representativo do volume líquido de resíduos recolhidos por dia, em Santa Maria, de julho a setembro de 2001	67
FIGURA 16 - Gráfico representativo da quilometragem percorrida pelos caminhões compactadores, por dia, em Santa Maria, de julho a setembro de 2001	69
FIGURA 17 - Diagrama de dispersão das variáveis peso e quilometragem para os meses de julho a setembro de 2001	71
FIGURA 18 - Diagrama de dispersão das variáveis peso e quilometragem, para os meses de julho a setembro, sem os pontos que desviam do padrão linear	72
FIGURA 19 - Gráfico de resíduos contra os valores ajustados, para os dados coletados	75
FIGURA 20 - Gráfico de resíduos ao longo do tempo para os dados analisados	76
FIGURA 21 - Gráfico de probabilidade normal para os resíduos dos dados estudados	78
FIGURA 22 - Gráfico de controle de regressão para as variáveis peso e quilometragem, nos meses de julho, agosto e setembro	80
FIGURA 23 - Gráfico de controle de regressão para as variáveis peso e quilometragem, nos meses de julho, agosto e setembro, sem os pontos fora de controle	81

FIGURA 24 - Gráfico de controle de regressão para as variáveis peso e quilometragem, nos meses de julho, agosto e setembro, com os limites padrão para o processo	82
---	----

LISTA DE TABELAS

TABELA 01 - Tabela das medidas descritivas para a variável peso	64
TABELA 02 - Tabela das medidas descritivas para a variável quilometragem	64

LISTA DE QUADROS

QUADRO 01 - Horário de recolhimento de resíduos por setor - Santa Maria	62
QUADRO 02 - Quadro demonstrativo do número de registros e o número de dias no mês que deveria haver coletas, por rota	68

LISTA DE ANEXOS

ANEXO A - Formulário preenchido pelos motoristas	92
ANEXO B - Limites superiores para o teste de Lilliefors	93
ANEXO C - Valores críticos d_L e d_S da estatística D de Durbin-Watson	94

LISTA DE SIGLAS

- CPE - Controle estatístico do processo
CUSUM - Somas cumulativas
EWMA - Média móvel exponencialmente ponderada
Km - Quilômetro
Kg - Quilograma
LM - Linha média
LIC - Limite inferior de controle
LSC - Limite superior de controle
MA - Média móvel
PRT - Pitaluga, Reis, Taschetto
QME – quadrado médio do erro
S – erro padrão estimado
SQE – soma do quadrado do erro

RESUMO

Dissertação de Mestrado
Programa de Pós-Graduação em Engenharia de Produção
Universidade Federal de Santa Maria

**MONITORAÇÃO DO PROCESSO DE COLETAS DE RESÍDUOS
EM SANTA MARIA-RS
USANDO GRÁFICOS DE CONTROLE DE REGRESSÃO**

AUTORA: Luciane Flores Jacobi
ORIENTADOR: Adriano Mendonça Souza
CO-ORIENTADOR: João Eduardo da Silva Pereira
Data e Local da Defesa: Santa Maria, 17 de dezembro de 2001.

Muitos dos sistemas utilizados nas empresas necessitam de processo de acompanhamento e controle de produção que aumentem a estabilidade de retorno financeiro e que permita investimentos de médio e longo prazo. Sabe-se hoje, que as variáveis envolvidas em um processo produtivo são, muitas vezes, correlacionadas, onde o controle individual dessas variáveis não é o mais indicado. Neste caso a qualidade depende do efeito comum destas variáveis, em lugar do efeito de cada uma separadamente. Os gráficos de controle convencionais, como X-barra, proporção e outros univariados, não são capazes de fazer uma análise quando se tem um conjunto de variáveis correlacionadas, mas o gráfico de controle de regressão é capaz de avaliar o efeito conjunto dessas variáveis. Esta pesquisa possui como objetivo empregar o gráfico de controle de regressão como ferramenta de controle estatístico para monitorar processos no qual possa ser identificado um sistema produtivo, onde uma variável de interesse possa ser expressa como função de uma variável de controle. Para se estabelecer o gráfico de controle de regressão foi necessário o estudo das teorias de análise de regressão linear simples e de gráficos de controle, as quais foram as bases para a realização deste trabalho. Após o estudo detalhado dessas teorias, buscou-se exemplificar essa metodologia, através de um estudo utilizando dados coletados nos meses de julho, agosto e setembro de 2001 na empresa PRT, no setor de Engenharia de Saneamento e Meio Ambiente, na área de recolhimento de resíduos domiciliares e comerciais, da cidade de Santa Maria, onde as variáveis analisadas foram quilometragem diária percorrida pelos caminhões e volume de resíduos coletados. Concluiu-se que, o sistema de coletas de resíduos não estava sendo realizado de forma satisfatória, pois alguns pontos estavam fora dos limites de controle. Sendo assim, sugere-se um acompanhamento maior pela empresa nas variáveis estudadas com o intuito de que, estando essas variáveis sob controle, a empresa seja capaz de prever com maior exatidão o gasto de combustível dos caminhões compactadores, utilizados para a coleta de resíduos.

ABSTRACT

Dissertação de Mestrado
Programa de Pós-Graduação em Engenharia de Produção
Universidade Federal de Santa Maria

**MONITORAÇÃO DO PROCESSO DE COLETAS DE RESÍDUOS
EM SANTA MARIA-RS
USANDO GRÁFICOS DE CONTROLE DE REGRESSÃO**

(Monitoring of the process of waste collecting in Santa Maria - RS, using of regression control charts)

AUTHOR: Luciane Flores Jacobi

ADVISOR: Adriano Mendonça Souza

CO-ADVISOR: João Eduardo da Silva Pereira

Date and place of defense: Santa Maria, 17 de dezembro de 2001.

Many of the systems used in the companies need a process to monitor and control production to increase the stability of the process and the financial return which allow medium and long term investments. Nowadays, it is known that the variables involved in a productive process are, many times, correlated, where the individual control of these variables is not the most indicated one. In this case, quality depends on the common effect of these variables, instead of the effect of each one separately. The conventional control charts, such as \bar{x} - bar, proportion and other univariate ones are not able to performing an analysis when a set of correlated variables should be monitored, but the regression control chart can evaluate the point effect of these variables. To put the regression control chart in effect, the regression analysis and control charts, were studied. The study was, in fact, the basis for the accomplishment of this work. After the detailed study of the theories used, examples of the referred methodology were sought, through a study with real data collected at the company PRT, at the Department Engineering of sanitation and Environment, in the area of dwelling and commercial refuse collecting, Santa Maria city. It follows that, the system of waste collecting was not being accomplished in a satisfactory way, for some points were out of control limit. In such case, it was suggested a more careful follow up by the company in the studied variables, with the thought that being these variables under control, the company is able of anticipating more accurately, the fuel expenses by the truck, used for the waste collecting.

1 INTRODUÇÃO

A velocidade das informações e as novas tecnologias estabeleceram um ambiente globalizado de alta concorrência, onde preço, prazo, qualidade e flexibilidade precisam ser atendidos.

Devido a isto, controlar as variáveis envolvidas no processo de produção, de modo a torná-lo mais eficiente, é uma das crescentes preocupações dos empresários, pois, desta maneira, pode-se reduzir os desperdícios e utilizar sua plena capacidade. Muitos dos sistemas utilizados nas empresas necessitam de processo de acompanhamento e controle de produção que aumentem a estabilidade de retorno financeiro e que permitam investimentos de médio e longo prazo.

A preocupação com a qualidade dos produtos teve origem com o surgimento da produção em maior escala, onde se constata o início da era da inspeção formal. Contudo, somente com a introdução da administração científica de Taylor, é que a qualidade foi reconhecida como área funcional dentro do gerenciamento de uma empresa (Levine, Berenson, Stephan, 2000).

Mas, com o crescimento da demanda e a intensificação da produção em massa, foi necessário substituir a inspeção 100% por uma inspeção por amostragem, surgindo, assim, os primeiros estudos sobre o controle das características de um processo produtivo. Estes estudos foram realizados por Walter Shewhart que, em 1924, desenvolveu um conjunto de conceitos que são as bases do moderno controle estatístico do processo. Shewhart, naquela época, preocupou-se em avaliar a qualidade, observando apenas uma característica do produto (Werkema, 1995; Siqueira, 1997).

Sabe-se, hoje, que as variáveis envolvidas em um processo produtivo são, muitas vezes, correlacionadas, e o controle individual destas variáveis não é o mais indicado. Neste caso, a qualidade depende do efeito comum destas variáveis, em lugar do efeito de cada variável separadamente. E, assim, fez-se necessário um outro conceito de gráficos de controle que pudessem se ajustar a esse tipo de processo.

Os gráficos de controle convencionais, como X-barra e proporção, não são capazes de fazer uma análise, quando se tem um conjunto de variáveis correlacionadas, mas o gráfico de controle de regressão é capaz de avaliar o efeito conjunto destas variáveis. Por este motivo, este gráfico é muito utilizado em processos em que o efeito de uma variável dependente é uma função linear de uma variável independente, fornecendo previsões de comportamento e revelando como a relação entre as variáveis está ocorrendo.

Nesta pesquisa, utiliza-se a técnica de regressão linear simples combinada com a de controle de qualidade, no controle simultâneo de variáveis correlacionadas. Autores, como DiPaola (1945), Mandel (1969) e Olin (1998), apresentaram estudos que mostram como o gráfico de controle de regressão pode ser construído a partir de dados históricos e, utilizando-se deste gráfico, são verificadas se as novas observações do processo estão dentro dos níveis estabelecidos.

1.1 Tema da pesquisa

O tema da presente pesquisa é a análise de regressão linear junto com gráficos de controle, pois aplica-se uma metodologia que possibilita a monitoração de variáveis correlacionadas num processo produtivo. O

problema proposto é como realizar o controle simultâneo das variáveis, pois o controle individual não revela a real situação do processo.

1.2 Justificativa e importância da pesquisa

O controle estatístico do processo (CEP) é uma metodologia utilizada no controle e melhoria da qualidade, podendo ser aplicado em todas as etapas produtivas, para o monitoramento do processo por meio de informações geradas por ele.

Considera-se esta pesquisa importante por disseminar conhecimento e trazer experiências reais para o meio acadêmico e por fazer uma monitoração simultânea das variáveis do processo. A metodologia estudada permite um controle eficaz, pois, muitas vezes, variáveis correlacionadas parecem estar sob controle estatístico, quando monitoradas separadamente, mas, na verdade, o sistema está fora de controle devido ao efeito da correlação. A implantação desta metodologia proporciona uma melhor informação a respeito do processo, permitindo uma tomada de decisão de forma correta e em tempo de implementar uma correção.

1.3 Objetivos

1.3.1 Objetivo Geral

Este trabalho possui como objetivo empregar o gráfico de controle de regressão como ferramenta de controle estatístico para monitorar processos produtivos, onde uma variável de estado que seja de interesse, possa ser expressa como função de uma variável de controle.

1.3.2 Objetivos Específicos

- Realizar uma revisão de literatura abrangente sobre os gráficos de controle de regressão;
- mostrar a importância dos gráficos de controle de regressão em processos com variáveis correlacionadas;
- fazer uma análise exploratória das variáveis em estudo;
- modelar a variável dependente (Y) em função da variável independente (X) através do modelo de regressão;
- monitorar o processo através dos limites de controle e identificar as causas que influenciam o processo;
- identificar os períodos de maior produção de volume de resíduos.

1.4 Metodologia

Como foi apresentado anteriormente, o objetivo deste trabalho é a monitoração de processos em que possam ser identificadas variáveis correlacionadas. Para que estes objetivos fossem atingidos, uma revisão da literatura foi realizada, a fim de que seja mostrado que a combinação destas duas teorias são importantes para o monitoramento deste processo.

Para se estabelecer o gráfico de controle de regressão, foi necessário o estudo das teorias de análise de regressão linear simples e de gráficos de controle que foram as bases para a realização deste trabalho.

Após o estudo detalhado das teorias utilizadas neste trabalho, buscou-se exemplificar esta metodologia através de um estudo, utilizando dados coletados na empresa PRT, no setor de Engenharia de Saneamento e Meio Ambiente. Os dados utilizados para a aplicação da metodologia foram coletados nas fichas preenchidas pelos motoristas dos caminhões

compactadores. Com a aquisição destas informações, construiu-se um banco de dados, em uma planilha eletrônica, que foram utilizados para a estimação da equação de regressão, a determinação dos limites inferiores e superiores de controle e a construção do gráfico de controle de regressão através dos programas computacionais Statistica 5.1 (Statsoft) e PcGive 8.0 (Chapman & Hall).

1.5 Delimitação da pesquisa

A pesquisa, aqui desenvolvida, é constituída da combinação de técnicas úteis para a monitoração de processos em que variáveis correlacionadas são identificadas.

Existem vários estudos sobre o controle de variáveis em processos produtivos, mas, na maioria das vezes, são em relação ao controle individual das variáveis e, portanto, não podendo ser utilizados para um estudo comparativo. Esta pesquisa, entretanto, apresenta uma técnica eficiente no controle simultâneo de variáveis correlacionadas.

1.6 Organização do trabalho

Esta pesquisa está organizada da seguinte forma: no capítulo 2 apresenta-se a revisão de literatura, no capítulo 3, a metodologia. No item 3.1, descreve-se a teoria dos gráficos de controle, passando ao item 3.2 que descreve a teoria de análise de regressão linear simples e, por fim, no item 3.3, combina-se estas teorias, apresentando-se, assim, o gráfico de controle de regressão. No capítulo 4, desenvolve-se a aplicação da metodologia e, no capítulo 5, apresenta-se as conclusões do estudo e as sugestões recomendadas.

2 REVISÃO DA LITERATURA

Neste capítulo, apresenta-se a revisão de literatura, dividida em dois itens, que servirá de suporte para o desenvolvimento deste trabalho. No item 2.2, aborda-se a metodologia dos gráficos de controle, mostrando como são construídos os limites de controle. No item 2.3, será discutida a análise de regressão, ressaltando as pressuposições para a realização desta análise e como são avaliadas. Desta forma, pretende-se mostrar como estas técnicas serão desenvolvidas e utilizadas neste trabalho.

2.1 Introdução

A implantação de um programa para melhoria da qualidade pode eliminar desperdícios, reduzir os índices de produtos defeituosos fabricados, diminuir a necessidade da realização de inspeção e aumentar a satisfação dos clientes, fatores que implicam em um aumento da produtividade e da competitividade das empresas.

As técnicas estatísticas representam ferramentas extremamente úteis no controle de qualidade de bens e serviços, devendo ser utilizadas ao longo de todo o processo de solução de problemas.

Duas destas técnicas serão abordadas nesta pesquisa, a dos gráficos de controle que é uma ferramenta muito útil para a avaliação do estado de controle de um processo e a análise de regressão que é uma técnica estatística de grande importância para a condução, de forma mais eficiente, das ações gerenciais de planejamento, manutenção e melhoria da qualidade.

2.2 Gráficos de controle

A história do controle de qualidade é tão antiga quanto a do próprio homem. Desde que o ser humano começou a manufaturar itens, a qualidade era controlada, principalmente, pela imensa experiência dos artesões que tinham que mostrar habilidade de produzir itens com qualidade (Siqueira 1997 ; Petenate 19-).

A American Society for Quality Control (ASQC) define qualidade como "a totalidade de características e peculiaridades de um produto ou serviço que levam a sua capacidade de satisfazer determinadas necessidades". Em outras palavras, a qualidade mede como um produto ou serviço atinge as necessidades do cliente. As organizações reconhecem que para ser competitivas na economia global atual, devem lutar por altos níveis de qualidade. Em consequência disto, aumentaram a ênfase em métodos para monitorar e manter a qualidade (Anderson, Sweeney, Willians, 2002).

Assegurar qualidade refere-se a todo sistema de políticas, procedimentos e pautas estabelecidas por uma organização para alcançar e manter a qualidade. Assegurar qualidade consiste de duas funções principais: engenharia e controle de qualidade. O objetivo da engenharia da qualidade é incluir qualidade no desenvolvimento de produtos e processos e identificar problemas potenciais de qualidade devidos à produção. O controle de qualidade consiste em fazer uma série de inspeções e medidas para determinar se os padrões da qualidade estão sendo atingidos. Se os padrões da qualidade não estiverem sendo atingidos, ações corretivas e/ou

preventivas podem ser tomadas para alcançar e manter a conformidade (Anderson, Sweeney, Willians, 2002).

Com o advento da produção em massa e a conseqüente intensificação da produção, a inspeção que antes era feita em todos os itens, passou a ser impraticável. Surge, então, a inspeção por amostragem. Mesmo assim, o problema da qualidade não foi resolvido, pois as atividades de inspeção são sempre limitadas, porque reagem ao passado, encontrando itens defeituosos depois que já foram produzidos, não contribuindo para a prevenção de defeitos e o aumento de produtividade (Petenate, 19-).

Como a simples inspeção final não melhorava a qualidade dos produtos fornecidos e a preocupação constante com os custos e com a produtividade, surgiu a necessidade de se utilizar as informações obtidas com a inspeção, para melhorar a qualidade dos produtos. Com isto, percebeu-se que a variabilidade era um fator inerente aos processos industriais e podia ser compreendida através da estatística e da probabilidade.

Logo, a partir de uma amostragem adequadamente dimensionada e planejada, não se precisaria esperar a conclusão do ciclo de produção para realizar as medições, pois poderiam ser feitas durante o processo de fabricação (Siqueira, 1997).

Em 1924, o Dr. Shewhart propôs o uso do gráfico de controle para a análise de dados resultantes de inspeção no processo, fazendo com que a importância dada à inspeção, um procedimento baseado na detecção e correção de produtos defeituosos, começasse a ser substituída por uma ênfase no estudo e prevenção dos problemas relacionados à qualidade, de

modo a impedir que os produtos defeituosos fossem produzidos (Werkema, 1995).

A partir daí, os problemas de controle e melhoria de qualidade foram tratados com enfoque na variabilidade do processo, abrindo um vasto campo de pesquisa e de implementação de técnicas que são utilizadas e aprimoradas constantemente (Zanini, 1999).

A variabilidade, também denominada variação ou dispersão, está presente em todos os processos de produção de bens e de fornecimento de serviços. Há muitas causas para a presença de variabilidade. De modo geral, podemos dizer que a variabilidade é o resultado de alterações nas condições sob as quais as observações são tomadas. Estas alterações podem refletir diferenças entre as matérias-primas, as condições dos equipamentos, os métodos de trabalho, as condições ambientais e os operadores envolvidos no processo (Werkema, 1995).

Sabe-se que todo produto possui um número de parâmetros que, em conjunto, descrevem sua adequação ao uso. Estes parâmetros são freqüentemente chamados de características da qualidade. Um método para efetuar o controle do processo através da avaliação das características, é o controle estatístico do processo. O método utiliza gráficos e outras ferramentas estatísticas para alertar da ocorrência de anormalidades originadas da variabilidade (Siqueira 1997 ; Petenate 19-).

As ferramentas que podem ser aplicadas no controle estatístico do processo (CEP) estão descritas abaixo e são conhecidas como as sete ferramentas. São úteis, especialmente, nos estágios iniciais de implementação de um programa de CEP (Petenate, 19_).

1. **Análise de Pareto:** esta ferramenta permite identificar e priorizar problemas e decidir onde concentrar os esforços de qualidade;
2. **Folhas de verificação:** são folhas para coleta de dados que ajudam a assegurar que os dados são coletados de forma acurada, rápida e completa;
3. **Diagrama de fluxo:** diagrama que estabelece os passos ou atividades de um processo na ordem em que ocorrem. Pode ser utilizado para identificar pontos no processo, onde o controle de atividades pode ser iniciado;
4. **Histograma:** são gráficos de uma distribuição. Ajudam a visualizar a distribuição da característica de qualidade em estudo, a determinar se o processo está operando da forma que gostaríamos e a identificar algumas causas influentes;
5. **Diagrama de causa e efeito:** esta técnica permite a utilização do conhecimento coletivo para a identificação de potenciais relações de causa e efeito, ajudando, assim, a encontrar soluções;
6. **Diagrama de dispersão:** é um gráfico que possibilita avaliar possíveis relações de causa e efeito entre duas variáveis. O gráfico ajuda a verificar a presença de relações e determinar a força e a direção da mesma;
7. **Cartas de controle:** são gráficos temporais sobre o comportamento de um processo. Ajudam a controlar o desempenho do processo ao longo do tempo.

Os gráficos de controle são uma ferramenta extremamente útil para identificar se as variações observadas num processo são decorrentes de

causas comuns de variação e, portanto, de pequena significância, ou decorrentes de causas especiais de variação e, portanto, de grande significância, que necessitam ser identificadas e eliminadas do processo (Siqueira, 1997; Montgomery, 1997).

A variação provocada por causas comuns, também conhecida como variabilidade natural do processo, é inerente ao processo considerado e estará presente, mesmo que todas as operações sejam executadas, empregando métodos padronizados. Quando apenas as causas comuns de variação estão atuando em um processo, a quantidade de variabilidade se mantém em uma faixa estável, conhecida como faixa característica do processo. Neste caso, diz-se que o processo está sob controle estatístico, apresentando um comportamento estável e previsível. Já as causas especiais de variação surgem esporadicamente, devido a uma situação particular que faz com que o processo se comporte de um modo completamente diferente do usual, o que pode resultar em um deslocamento do seu nível de qualidade.

Quando um processo está operando sob a atuação de causas especiais de variação, diz-se que está fora de controle estatístico e, neste caso, sua variabilidade, geralmente, é bem maior do que a variabilidade natural. As causas especiais de variação devem ser, de modo geral, localizadas e eliminadas e, além disto, devem ser adotadas medidas para evitar sua reincidência (Werkema, 1985; Levine, Berenson, Stephan, 2000).

Um gráfico de controle é uma representação visual de uma característica da qualidade medida ou calculada para uma amostra de itens, grafada em função do número da amostra ou de alguma outra variável indicadora do tempo (ordem cronológica). O gráfico consiste em uma linha

média (LM), um par de limites de controles, representados um abaixo (limite inferior de controle – LIC) e outro acima (limite superior de controle – LSC) da linha média e dos valores da característica da qualidade traçados no gráfico.

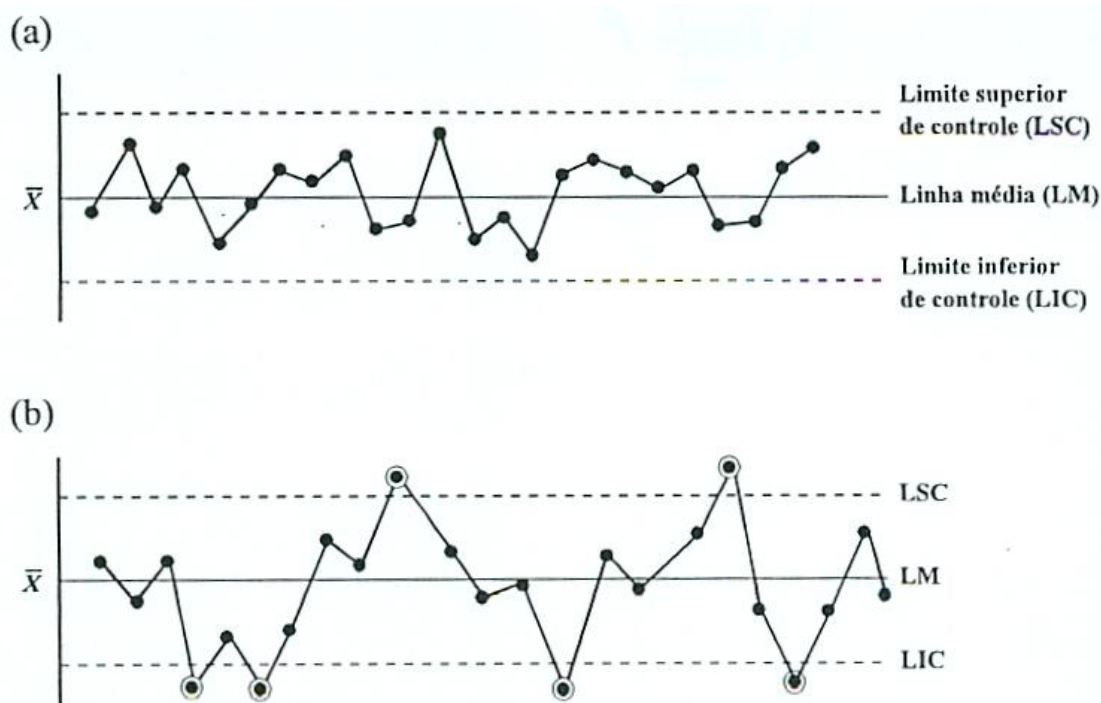


FIGURA 01 - Exemplos de gráfico de controle (a) Processo sob controle (b) Processo fora de controle.
(Figura adaptada de Werkema, 1995)

Pela Figura 01, pode-se perceber que a linha central representa o valor médio da característica da qualidade correspondente à situação do processo sob controle, isto é, sob a atuação de apenas causas de variação aleatórias. Os limites de controle LIC e LSC são determinados de forma que, se o processo está sob controle, praticamente, todos os pontos traçados no gráfico estarão entre estas linhas, formando uma nuvem aleatória de pontos distribuídos em torno da linha média. Os valores da característica da

qualidade traçados no gráfico indicam, então, a situação do processo no que diz respeito ao controle estatístico (Werkema, 1995; Siqueira, 1997).

A localização e o padrão dos pontos em um gráfico de controle possibilita determinar, com uma pequena probabilidade de erro, se o processo está sob controle estatístico. Uma indicação de que o processo possa estar fora de controle é um ponto fora dos limites de controle, ou seja, uma causa especial de variação está presente (Siqueira, 1997; Anderson, Sweeney, Willians, 2002).

Um processo também pode ser considerado fora de controle, mesmo quando todos os pontos encontram-se dentro dos limites de controle. Esta situação ocorre, quando um padrão de variação anormal está presente no processo. A probabilidade de ocorrência de um padrão anormal é, aproximadamente, igual a probabilidade de um ponto estar fora dos limites de $\pm 3\sigma$ (Anderson, Sweeney, Willians, 2002).

Para o cálculo dos limites dos gráficos de controle, Shewhart (1931) baseou-se na idéia de que, estando o processo sob controle estatístico e sendo as medidas individuais provenientes de uma mesma população, com distribuição normal, então, uma estatística w qualquer, calculada a partir dos valores amostrais e que tenha média μ_w e desvio padrão σ_w conhecidos, terá uma probabilidade próxima a um (ou 99,74%) de estar no intervalo de:

$$\mu_w \pm 3\sigma_w$$

Consequentemente, os limites do gráfico de controle, para esta estatística, serão:

$$LSC_w = \mu_w + 3\sigma_w$$

$$LM_w = \mu_w$$

$$LIC_w = \mu_w - 3\sigma_w$$

Diversos tipos de gráficos de controle têm sido desenvolvidos para analisar tanto variáveis como atributos. Entretanto, todos os gráficos de controle têm as mesmas duas funções básicas e são preparados e analisados de acordo com o mesmo roteiro. Segundo Petenate (19_), as funções são:

1. Emitir sinal sobre a presença de causas especiais de variação, de tal forma, que ações corretivas possam ser tomadas para trazer o processo para o estado de controle estatístico;
2. Fornecer evidências, se o processo está operando num regime de controle estatístico, de tal forma, que o cálculo da capacidade do processo de atender às especificações possa ser realizado.

Os gráficos de controle podem ser classificados de acordo com a característica que se supõe monitorar. Quando as medidas representadas resultam da contagem do número de itens dos produtos que apresentam uma característica particular de interesse, têm-se os gráficos de controle para atributos, por exemplo, os gráficos *c*, para o número de defeitos e os gráficos *p*, para a fração de defeitos produzidos pelo processo. Entretanto, quando a característica da qualidade é expressa por um número em uma escala contínua de medidas, têm-se os gráficos para variáveis, por exemplo, os gráficos *X*-barra para a média, *R* para a amplitude e *S* para o desvio padrão. Além disto, em alguns casos, pode-se construir gráficos de controle para observações individuais. Alguns gráficos aplicados neste caso são: gráfico de Somas Cumulativas do inglês "Cumulative Sum" (CUSUM), gráfico de Médias Móveis do inglês "Moving Average" (MA) e o gráfico de Médias Móveis Exponencialmente Ponderadas do inglês "Exponentially

Weighted Moving Average" (EWMA) (Werkema,1995; Zanini,1999; Vargas, 2001).

Uma desvantagem das cartas de Shewhart é que usam a informação sobre o processo contido no último ponto calculado. Se o ponto situa-se fora dos limites de controle, é um sinal de que o processo está fora de controle. Mas o ponto não incorpora informação sobre o passado do processo, sendo esta informação ignorada. Isto faz com que as cartas de Shewhart sejam pouco sensíveis para detectar pequenos deslocamentos do processo, da ordem de um desvio padrão (Petenate, 19_; Zanini, 1999; Lucas, 1976).

Duas alternativas efetivas ao gráfico de controle de Shewhart podem ser usadas: os gráficos de controle de Somas Acumuladas (CUSUM) e os gráficos de controle de Médias Móveis Exponencialmente Ponderadas (EWMA), que avaliam o processo como um todo, incorporando valores passados e atuais do processo (Epprecht, Ninio, Souza, 1998).

Por combinarem informações de muitas amostras, é que os gráficos CUSUM são mais eficientes do que os de Shewhart para detectar pequenas mudanças no processo, isto ocorre, em especial, quando $n=1$ (Zanini, 1999; Lucas, 1976).

Existem duas formas de representar o gráfico CUSUM: tabular e máscara - V. Uma máscara - V típica é apresentada na Figura 02.

Lucas (1976) comparou um gráfico CUSUM na forma de máscara-V com o gráfico de Shewhart e concluiu que é mais fácil de utilizar a máscara-V, quando os gráficos de Shewhart necessitam de critérios adicionais para detectar se o processo está fora de controle.

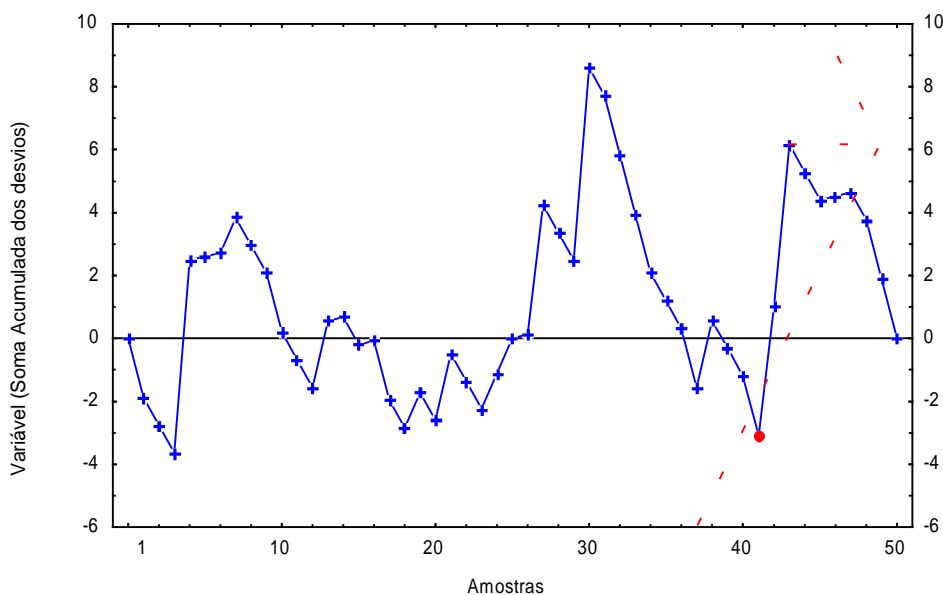


FIGURA 02 - Representação genérica do Gráfico CUSUM sob a forma de máscara - V com $\alpha=\beta=0,05$ e $\delta=1$.

O gráfico de controle de Média Móvel Exponencialmente Ponderada é também uma boa alternativa para os gráficos de controle de Shewhart, quando se tem interesse em detectar pequenas mudanças da média. O desempenho dos gráficos EWMA é, aproximadamente, equivalente a dos gráficos de controle de somas cumulativas (Zanini, 1999; Lucas & Saccucci, 1990).

Este gráfico surgiu, principalmente, para cobrir uma lacuna deixada pelos gráficos de Shewhart, pois é usado para descobrir pontos que estão fora de controle, onde existem pequenas variações na média esperada do processo, de 1σ ou $1,5\sigma$.

O procedimento possui um mecanismo que incorpora as informações de todo o subgrupo anterior mais as informações do subgrupo atual. Estas informações são obtidas através de ponderações, onde é possível atribuir aos valores passados um determinado grau de importância, conforme o

desejado, sendo um processo recursivo. Tem a vantagem de detectar pontos fora de controle mais rapidamente que os gráficos de Shewhart e, esta descoberta, pode ser realizada se estes pontos estiverem dentro dos limites de confiança de 3σ .

O gráfico EWMA, assim como o gráfico CUSUM, possui as linhas de limites diferentes das linhas de limites do gráfico de Shewhart, ou seja, não são paralelas ao eixo horizontal. O formato do gráfico EWMA pode ser visto, na Figura 03, a seguir:

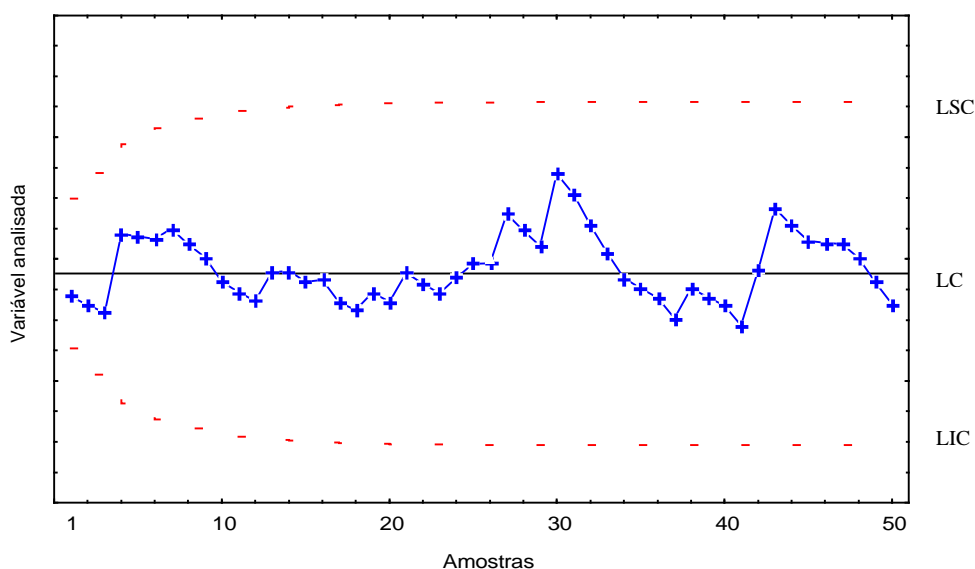


FIGURA 03 - Representação genérica de um gráfico EWMA.

Vargas (2001) realizou um estudo comparativo do desempenho dos gráficos de controle CUSUM e EWMA e concluiu que o gráfico de controle CUSUM mostrou-se mais eficiente em todas as análises realizadas com as mudanças na ordem de mais 1 desvio padrão para acima e para todas as alterações na ordem de menos 1,125 desvio padrão para baixo.

Assim como os gráficos CUSUM e EWMA, outro gráfico que não segue o padrão do gráfico de controle de Shewhart, com linhas paralelas ao

eixo horizontal e que também usa observações individuais é o gráfico de controle de regressão. O formato do gráfico do controle de regressão pode ser observado na Figura 04.

Quando os pontos plotados em um gráfico de controle apresentam uma tendência ascendente ou descendente, isto pode ser atribuído a uma tendência anormal ou a uma tendência normal de variação, ou seja, uma mudança gradual na média pode ser esperada e considerada normal. Uma vez que o valor central do gráfico é uma reta inclinada, sua equação deve ser estabelecida de modo que se encontrem os coeficientes linear e angular da reta, utilizando-se o método dos mínimos quadrados (Siqueira, 1997).

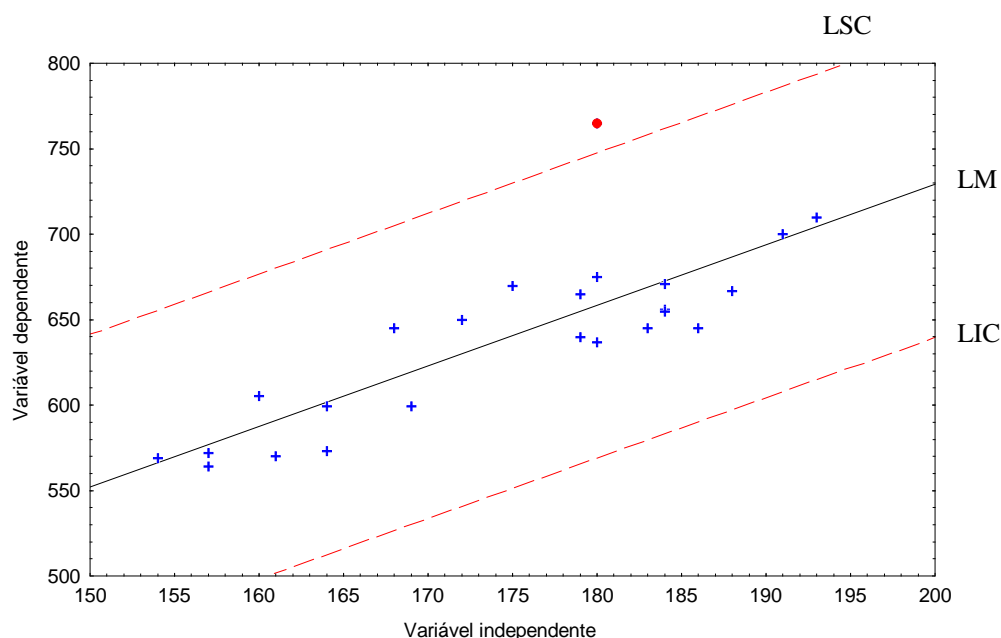


FIGURA 04 - Representação genérica do gráfico de controle de regressão.

No gráfico de controle de regressão, os limites superior e inferior de controle são paralelos à reta de regressão, em vez de serem paralelos ao eixo horizontal, como o gráfico tradicional de Shewhart.

O gráfico de controle de regressão é construído a partir de um conjunto de pares de valores e é utilizado para monitorar processos em que variáveis correlacionadas são identificadas.

2.3 Análise de Regressão

O termo regressão foi introduzido por Francis Galton. Em um famoso ensaio, Galton verificou que, embora houvesse uma tendência de pais altos terem filhos altos e de pais baixos terem filhos baixos, a altura média dos filhos de pais de uma dada altura tendia a se deslocar ou “regredir” até a altura média da população. A lei de regressão universal de Galton foi confirmada por Karl Person que coletou mais de mil registros das alturas dos membros de grupos de famílias. Ele verificou que a altura média dos filhos de um grupo de pais altos era inferior à altura de seus pais e que a altura média dos filhos de um grupo de pais baixos era superior à altura de seus pais. Os dois autores estavam interessados em provar a validade de uma suposta “Lei da Regressão Universal” que estabelecia que as características de uma pessoa eram transmitidas de pai para filho de forma amortecida (Gujarati, 2000; Vasconcellos & Alves, 2000).

Como o método se prestava à estimação de relações, passou a ser utilizado para o estudo da dependência de uma variável, a variável dependente em relação a uma ou mais variáveis, as variáveis explicativas, com o objetivo de estimar e/ou prever a média (da população) ou o valor médio da variável dependente em termos dos valores conhecidos ou fixos (em amostragem repetida) das variáveis explicativas (Gujarati, 2000).

A análise de regressão constitui um conjunto de métodos e técnicas para o estabelecimento de fórmulas empíricas que interpretam a relação funcional entre variáveis com boa aproximação. Esta análise é feita para que se possa encontrar alguma forma de medir a relação funcional entre as variáveis de cada conjunto, de tal forma, que essa medida possa mostrar que:

1. Se há relação entre as variáveis e, caso afirmativo, se é fraca ou forte;
2. Se esta relação existir, possa-se estabelecer um modelo que interprete a relação funcional entre as variáveis;
3. Constituído o modelo, usá-lo para fins de predição.

Considerando-se dois conjuntos de variáveis, X_i e Y_i , relacionadas por uma função matemática $Y_i = f(X_i)$.

Sejam Y_i um conjunto de n variáveis objeto de estudo e de previsão, espera-se que sofram influência de um número finito de variáveis X_1, \dots, X_k .

É comum, entretanto, que a variável dependente seja afetada por outros fatores, além dos considerados no modelo adotado. Admitindo-se que a variável dependente sofra a influência de $k + m$ variáveis, isto é, $Y = f(X_1, X_2, \dots, X_{k+m})$, e que, por vários motivos, não se considera a influência das variáveis X_{k+1}, \dots, X_{k+m} . Analisando Y_i como função das k primeiras variáveis permanece, então, um resíduo ou erro.

Admitindo-se que este modelo seja aditivo, o modelo estatístico fica:

$$Y_1, Y_2, \dots, Y_n = f(X_1, X_2, \dots, X_k) + H(X_{k+1}, \dots, X_{k+m})$$

Todas as influências das variáveis X_{k+1}, \dots, X_{k+m} , sobre as quais não se têm controle, serão consideradas como casuais e, sendo associada uma variável aleatória, obtém-se o seguinte modelo de regressão:

$$Y_1, Y_2, \dots, Y_n = f(X_1, X_2, \dots, X_k) + \varepsilon_i.$$

Outra justificativa para a existência do erro (ε_i), em um modelo estatístico, é dado pelos erros de mensuração da variável dependente. Em casos gerais, normalmente, existem tantos erros de mensuração como efeito de outras variáveis. Nestes casos, o erro residual do modelo será a soma destes dois tipos de erros.

Um diagrama de dispersão sugere a existência da relação funcional entre duas variáveis, mostrando esta relação através de uma linha construída no plano XY. Os pontos experimentais terão uma variação em torno da linha representativa da função, devido à existência de uma variação que chamaremos de variação residual.

Quando se aplica a análise de regressão ao estado da relação funcional entre duas variáveis, geralmente, encontram-se os seguintes problemas: especificação do modelo; estimação dos parâmetros; adaptação e significância do modelo adotado (Fonseca, Martins, Toledo, 1985; Hoffman, Vieira, 1977).

Especificação do Modelo

Sabe-se que K variáveis influenciam a variável dependente Y. O problema é encontrarmos a relação entre Y e X_1, X_2, \dots, X_k , detalhando a função: linear, polinomial, exponencial, logarítmica, etc.

A partir de uma especificação “a priori”, poderão ser indicadas certas condições parciais sobre o intercepto, declividade, presença de assíntota ou curvatura da função.

Outra opção seria utilizar o diagrama de dispersão, no caso de considerar-se apenas uma das K variáveis independentes.

Estimação dos Parâmetros

Após a determinação do modelo que seja representativo dos dados, deve-se estimar o valor dos diversos parâmetros. Por exemplo, se o modelo escolhido for da forma linear $Y = a + bX + e_i$, será necessário encontrar os parâmetros α e β , onde α é coeficiente linear e β é o coeficiente angular da reta.

Partindo da observação de uma amostra de “n” pares de valores (x_i, y_i) , $i = 1, 2, \dots, n$, encontra-se as estimativas “a” e “b”, obtendo dessa forma uma estimativa do modelo adotado:

$$\hat{y} = a + bx, \quad (2.1)$$

onde \hat{y} será o estimador de Y .

Para determinação de “a” e “b”, existem vários métodos. Será usado o método dos mínimos quadrados, por ser o que fornece os parâmetros que melhor ajustam os dados à reta de regressão, ou seja, que minimiza a soma dos quadrados dos erros (Maddala, 1992).

Adaptação e significância do modelo adotado

Nesta etapa, é verificado se a especificação adotada na primeira etapa, adapta-se convenientemente aos dados observados, através do cálculo da diferença entre os valores de Y observados e os valores de \hat{y} estimados

pela equação de regressão; $(Y - \hat{y})$, obtendo-se, assim, os erros de estimação.

Outra abordagem desta análise consiste no emprego dos testes de hipóteses e construção de intervalos de confiança para o modelo e seus parâmetros.

Pode-se, ainda, calcular o coeficiente de determinação R^2 , que será definido posteriormente, indicando quanto por cento a variação explicada pela regressão representa a variação total.

A natureza da relação existente entre as variáveis pode assumir diversas formas, abrangendo desde as funções matemáticas mais simples até as mais complexas. A relação mais simples consiste em uma relação linear ou retilínea.

Dados n pares de valores de duas variáveis dependentes X_i e Y_i ($i = 1, 2, \dots, n$) e, admitindo-se que Y é uma função linear de X :

$$F(X) = \alpha + \beta X$$

Pode-se estabelecer uma regressão linear simples, cujo modelo estatístico será:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (2.2)$$

Onde Y_i é a variável dependente, X_i é a variável independente, α e β são os parâmetros e ε_i representa a influência dos outros fatores, bem como os erros de medição da variável Y . O coeficiente angular da reta (β) é também denominado coeficiente de regressão e o coeficiente linear da reta (α) é também conhecido como termo constante da equação de regressão.

Ao estabelecer o modelo de regressão linear simples, pressupõe-se que:

1. A relação entre X e Y é linear;
2. Os valores de X são fixados (controlados), isto é, X não é uma variável aleatória, o que não ocorre nesta pesquisa;
3. A média dos erros é nula, isto é, $E(\varepsilon_i) = 0$;
4. Para um dado valor de X, a variância do erro ε_i , é sempre σ^2 , denominada variância residual, isto é, $V(\varepsilon_i) = \sigma^2$, ou $E [Y_i / X_i]^2 = \sigma^2$. Diz-se que o erro é homocedástico, ou que tem-se homocedasticidade (do erro ou da variável dependente);
5. O erro de uma observação é não correlacionado com o erro em outra observação, isto é, $E (\varepsilon_i, \varepsilon_j) = 0$, para $i \neq j$;
6. Os erros possuem distribuição normal;
7. Combinando as pressuposições 3,4, e 6, tem-se que: $\varepsilon_i \sim N (0, \sigma^2)$
8. Deve-se, ainda, verificar se o número de observações disponíveis é maior do que o número de parâmetros da equação de regressão.

Na Figura 05, representa-se o modelo estatístico de uma regressão linear simples, considerando-se as pressuposições de 1 a 4.

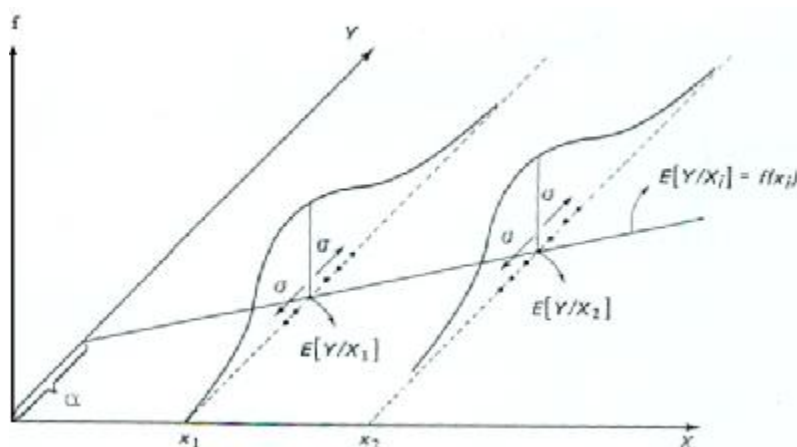


FIGURA 05 – Modelo estatístico de uma regressão linear simples.

(Figura adaptada de Fonseca, Martins, Toledo, 1985)

O primeiro passo na análise de regressão é obter os valores de “a” e “b” que são as estimativas dos parâmetros α e β da regressão. Os valores destas estimativas serão obtidos a partir de uma amostra de n pares de valores (x_i, y_i) que variam de $i = 1, 2, \dots, n$. Deseja-se que a reta \hat{y} seja tão próxima quanto possível do conjunto de pontos marcados, isto é, deseja-se minimizar a discrepância total entre os pontos marcados e a reta estimada, conforme ilustra a Figura 06 (Fonseca, Martins, Toledo, 1985; Intriligator, Bodkin, Hsiao, 1996).

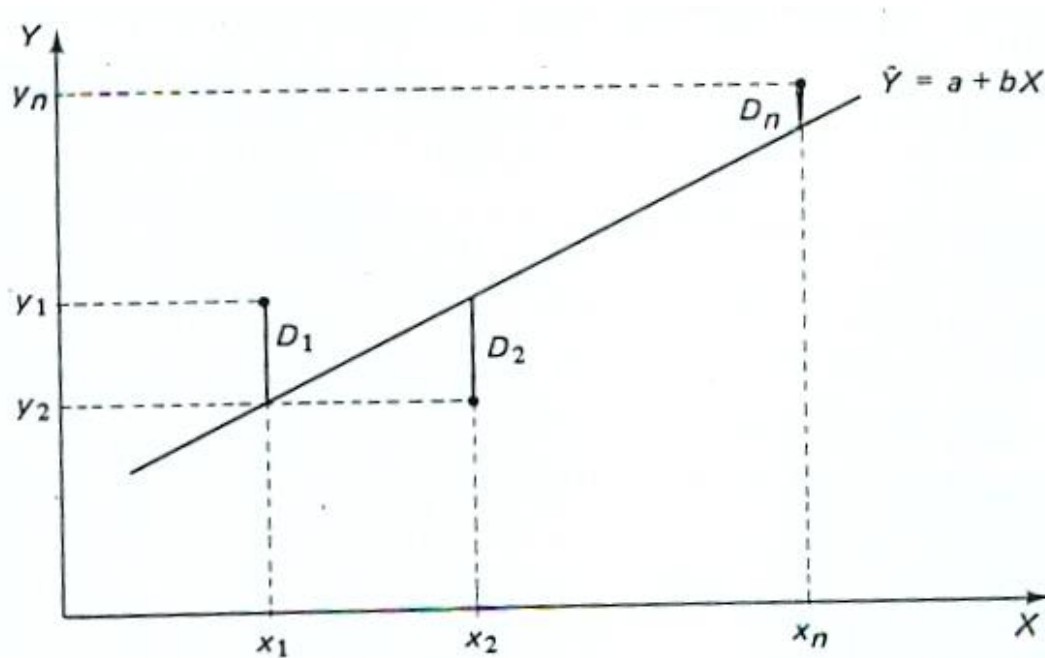


FIGURA 06 – Representação das diferenças entre os valores estimados pela reta de regressão e os valores observados.

(Figura adaptada de Fonseca, Martins, Toledo, 1985)

Observa-se que, para um dado x_i , existe uma diferença d_i entre o valor y_i observado e o seu correspondente \hat{y} , dado pela reta estimada. Os d_i são os erros ou desvios.

Para cada par de valores (x_i , y_i), pode-se estabelecer o desvio que é a diferença entre o valor real e o valor estimado por (2.1).

$$\varepsilon_i = y_i - \hat{y} = y_i - (a + bx_i)$$

A estimativa dos parâmetros através do Método dos Mínimos Quadrados (MMQ) consiste em adotar como estimativas dos parâmetros valores que minimizam a soma dos quadrados dos desvios. Assim, conforme (Neter, *et al*, 1996; Gujarati 2000; Charnet, *et al*, 1999), tem-se:

$$Z = \sum (y_i - \hat{y})^2 = \sum (y_i - a - bx)^2$$

Como Z depende dos valores de “a” e “b”, deriva-se Z em relação a “a” e “b”, e, para que Z seja mínimo, deve-se igualar $\frac{\partial Z}{\partial a}$ e $\frac{\partial Z}{\partial b}$ a zero.

Assim:

$$\frac{\partial Z}{\partial a} = -2 \sum [y_i - (a + bx_i)] = 0$$

$$\frac{\partial Z}{\partial b} = -2 \sum x [y_i - (a + bx_i)] = 0$$

Simplificando, obtém-se as duas equações a seguir, chamadas de equações normais:

$$\begin{cases} na - b \sum x_i = \sum y_i \\ a \sum x_i + b \sum x_i^2 = \sum x_i y_i \end{cases}$$

Isolando Y na primeira equação, tem-se:

$$\sum y = na + b \sum x$$

Efetuando a multiplicação na 2ª equação, tem-se:

$$\sum xy = a \sum x + b \sum x^2$$

Obtendo-se, então, o sistema abaixo:

$$\sum y = na + b \sum x \tag{2.3}$$

$$\sum xy = a \sum x + b \sum x^2 \quad (2.4)$$

Dividindo-se a equação (2.3) por n:

$$\frac{\sum y}{n} = \frac{na}{n} + \frac{b \sum x}{n}$$

Obtém-se que: $\bar{y} = a + b\bar{x}$

Isolando “a”, tem-se que:

$$a = \bar{y} - b\bar{x} \quad (2.5)$$

Substituindo o valor de “a” na equação (2.4):

$$\begin{aligned} \sum xy &= (\bar{y} - b\bar{x}) \sum x + b \sum x^2 \\ \sum xy &= \bar{y} \sum x - b \bar{x} \sum x + b \sum x^2 \\ \sum xy &= \frac{\sum y \sum x}{n} = b \left[\sum x^2 - \frac{(\sum x)^2}{n} \right] \end{aligned}$$

Isolando “b”:

$$b = \frac{\sum xy - \frac{\sum y \sum x}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} \quad (2.6)$$

Chamando de S_{xx} o denominador e de S_{xy} , o numerador de b, tem-se que:

$$b = \frac{S_{xy}}{S_{xx}} \quad (2.7)$$

que devem ser substituídas em $\hat{y} = a + bx$ na equação (2.1).

Considerando as pressuposições de 1 a 3 e as estimativas para "a" e "b", pode-se escrever que, se $\hat{y} = a + bx$ é o estimador de $f(x)$, ou seja, da componente funcional do modelo, então:

$\hat{y} = a + bx$ terá o seu valor esperado dado por:

$$E[\hat{y}] = E(a) + E(bx)$$

$$\text{Se } E[\hat{y}] = \alpha + \beta x$$

Logo, \hat{y} é um estimador justo de $f(x)$.

A variância de \hat{y} será:

$\text{var}[\hat{y}]$:

Se $\hat{y} = a + bx$, e $a = \bar{y} - b\bar{x}$, logo,

$$\hat{y} = \bar{y} - b\bar{x} + bx = \bar{y} + b(x - \bar{x})$$

$$\text{var}[\hat{y}] = \text{var}[\bar{y} + b(x - \bar{x})] = \text{var}(\bar{y}) + \text{var}[b(x - \bar{x})] =$$

$$\text{var}\left[\frac{\sum y}{n}\right] + (x - \bar{x})^2 \text{var}(b) =$$

$$\frac{1}{n^2} \text{var}(\sum y) + (x - \bar{x})^2 \text{var}(b) = \frac{\sum \text{var}(y)}{n^2} + (x - \bar{x})^2 \text{var}(b) =$$

$$\frac{\sum_{i=1}^n \sigma^2}{n^2} + (x - \bar{x})^2 \frac{\sigma^2}{S_{xx}} = \frac{n\sigma^2}{n^2} + (x - \bar{x})^2 \frac{\sigma^2}{S_{xx}} = \frac{\sigma^2}{n} + (x - \bar{x})^2 \frac{\sigma^2}{S_{xx}} = \sigma^2 \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right]$$

logo, a distribuição de probabilidade de \hat{y} , será:

$$\hat{y} \sim N \left[\alpha + \beta x; \sigma^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right) \right]$$

Observando-se as expressões dos parâmetros "a" e "b" da reta estimada, pode-se notar que ambos dependem de Y. Como foi visto anteriormente, Y é uma variável aleatória com distribuição normal de

média $\alpha + \beta X$ e variância σ^2 . Sendo os estimadores "a" e "b" funções lineares de uma variável aleatória normal. Portanto, deve-se encontrar as médias e variâncias dos estimadores "a" e "b". Primeiramente, determina-se o valor esperado de "b", isto é, $E[b]$. Assim:

$$b = \frac{\sum (x - \bar{x}) y}{\sum (x - \bar{x})^2}$$

$$E[b] = \frac{\sum (x - \bar{x}) E[y]}{\sum (x - \bar{x})^2}$$

$$E[b] = \frac{\sum (x - \bar{x}) E[y]}{\sum (x - \bar{x})^2} \quad E[y] = \alpha + \beta x$$

$$E[b] = \frac{\sum x(\alpha + \beta x) - \sum \bar{x}(\alpha + \beta x)}{\sum (x - \bar{x})^2} = \frac{\sum \alpha x + \sum \beta x^2 - \sum \bar{x} \alpha - \sum \bar{x} \beta x}{\sum (x - \bar{x})^2} =$$

$$\frac{\alpha \sum x + \beta \sum x^2 - \alpha \sum \bar{x} - \bar{x} \beta \sum x}{\sum (x - \bar{x})^2} = \frac{\beta \sum x^2 - \beta \frac{(\sum x)^2}{n}}{\sum (x - \bar{x})^2} = \frac{\beta \left[\sum x^2 - \frac{(\sum x)^2}{n} \right]}{\sum (x - \bar{x})^2}$$

Como, $E[b] = \beta$, logo "b" é um estimador não tendencioso ou imparcial de β .

A variância de "b" será dada por:

$$\text{var}[b] = \text{var}[\beta] + \text{var} \left[\frac{\sum \varepsilon_i (x - \bar{x})}{S_{xx}} \right] = 0 + \frac{\sum (x - \bar{x})^2}{(S_{xx})^2} \text{var}(\varepsilon_i)$$

Por hipótese do modelo, $\text{var}[b] = \sigma^2$, e $S_{xx} = \sum (x - \bar{x})^2$, conclui-se que:

$$\text{var}[b] = \frac{S_{xx} \sigma^2}{(\sum (x - \bar{x})^2)} = \frac{\sigma^2}{S_{xx}}$$

Logo, tem-se que a distribuição de probabilidade para “b” será:

$$b \sim N \left[\beta, \frac{\sigma^2}{S_{xx}} \right]$$

Quanto à distribuição da variável aleatória, “a” tem-se:

$$a = \bar{y} - b \bar{x}, \quad \text{como } \bar{y} = \frac{\sum y}{n}, \text{ então: } a = \frac{\sum y}{n} - b \bar{x}$$

$$Y = \alpha + \beta x + \varepsilon_i$$

$$a = \frac{\sum (\alpha + \beta x + \varepsilon)}{n} - b \bar{x}, \quad \text{como } \bar{x} = \frac{\sum x}{n}, \text{ então:}$$

$$a = \frac{na}{n} + \frac{\beta \sum x}{n} + \frac{\sum \varepsilon_i}{n} - b \bar{x} = \alpha + \beta \bar{x} + \frac{\sum \varepsilon_i}{n} + b \bar{x}$$

$$E[a] = E[a] + E[\beta \bar{x}] + \frac{E[\sum \varepsilon_i]}{n} - b \bar{x}, \quad \text{onde, } E[a] = a$$

Logo, “a” é um estimador justo de α .

Sua variância será:

$$\text{var}[a] = \text{var}[\alpha] + \text{var}[\beta \bar{x}] + \text{var} \left[\frac{\sum \varepsilon_i}{n} \right] + \text{var}[b \bar{x}]$$

$$0 + 0 + \frac{1}{n^2} \sum \text{var}[\varepsilon_i] + \bar{x}^2 \text{var}(b) = \frac{\sum \sigma^2}{n^2} + \frac{\bar{x}^2}{S_{xx}} s^2 = \frac{n\sigma^2}{n^2} + \frac{\bar{x}^2}{S_{xx}} s^2 = \frac{\sigma^2}{n} + \frac{\bar{x}^2}{S_{xx}} \sigma^2$$

Logo, tem-se que a distribuição de probabilidade de “a” será:

$$a \sim N \left[\alpha, \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right] \right]$$

Pelas demonstrações anteriores, observa-se que os estimadores dos parâmetros de uma regressão linear simples, obtidos pelo método dos mínimos quadrados, são estimadores lineares não tendenciosos de variância mínima.

Outra maneira de desenvolver o modelo de regressão básico e demonstrar seu maior uso, é considerar a relação no modelo de regressão diferentemente, ou seja, dividir a soma total de quadrados.

A forma para desenvolver esta divisão consiste em considerar $y_i - \bar{y}$, a quantidade fundamental que mede a variação das observações y_i , decomposta nos seguintes desvios:

$$\begin{array}{rcc}
 y_i - \bar{y} & = & \hat{y}_i - \bar{y} \quad + \quad y_i - \hat{y}_i \\
 \mathbf{1\ 2\ 3} & & \mathbf{1\ 2\ 3} \quad \quad \quad \mathbf{1\ 2\ 3} \\
 \text{Desvio} & & \text{Desvio do} & & \text{Desvio em} \\
 \text{total.} & & \text{valor da} & & \text{torno da linha} \\
 & & \text{regressão em} & & \text{de regressão.} \\
 & & \text{torno da média.} & &
 \end{array}$$

Assim, o desvio total $y_i - \bar{y}$ pode ser considerado como a soma de duas componentes:

1 - Os desvios dos valores ajustados \hat{y}_i em torno da média \bar{y} ;

2 - O desvio de y_i em torno da linha de regressão.

Pela Figura 07, pode-se facilmente observar a decomposição da soma total de quadrados para uma observação.

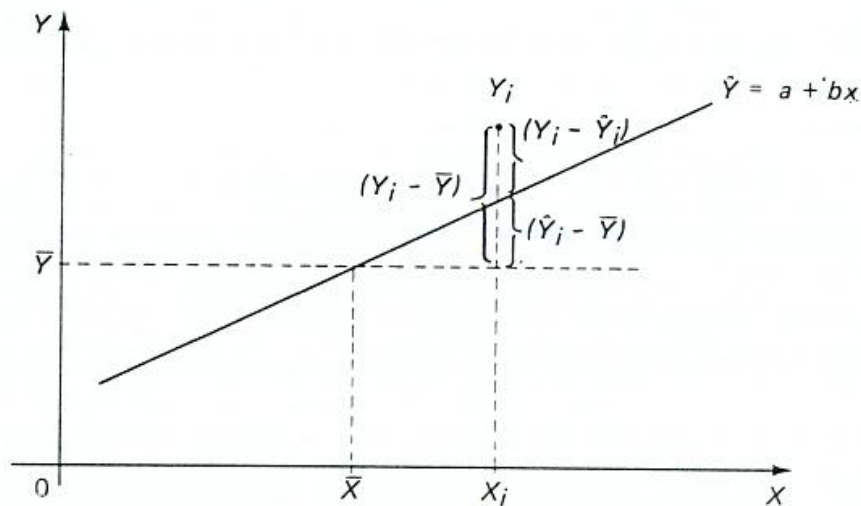


FIGURA 07 – Representação da decomposição da soma de quadrados.
(Figura adaptada de Fonseca, Martins, Toledo, 1985)

Os desvios totais resultam da soma dos desvios explicados, mais os resíduos, então:

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

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

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

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

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

A soma dos desvios calculados em torno da média (variação total), é igual à soma dos quadrados dos desvios em torno da linha de regressão (variação residual) mais a dos quadrados dos desvios da linha de regressão em torno da média (variação explicada) (Fonseca, Martins, Toledo, 1985).

Cálculo Prático das Variações

Variação total: (VT)

$$VT = \sum (y - \bar{y})^2 = S_{yy}$$

$$\sum (y^2 - 2y\bar{y} + \bar{y}^2) = \sum y^2 - 2\bar{y}\sum y + \sum \bar{y}^2 = \sum y^2 - 2\bar{y}\sum y + n\bar{y}^2$$

$$\sum y^2 - 2\bar{y}\sum y + n\bar{y} \frac{\sum y}{n} = \sum y^2 - 2\bar{y}\sum y + \bar{y}\sum y = \sum y^2 - \bar{y}\sum y =$$

$$\sum y^2 - \frac{\sum y}{n} \sum y = \sum y^2 - \frac{(\sum y)^2}{n} = S_{yy}$$

Variação Explicada: (VE)

$$VE = \sum (\hat{y}_i - \bar{y})^2 = \sum (a + bx - \bar{y})^2 = \sum (\bar{y} - b\bar{x} + bx - \bar{y})^2 = \sum [b(x - \bar{x})]^2$$

$$b^2 \sum (x - \bar{x})^2 = b^2 S_{xx} = \left[\frac{S_{xy}}{S_{xx}} \right]^2 S_{xx} = \frac{(S_{xy})^2}{S_{xx}}$$

Variação Residual: (VR)

$$VR = VT - VE$$

$$VR = S_{yy} - b^2 S_{xx}$$

O coeficiente de determinação ou explicação R^2 , é um indicador que nos fornece elementos para a análise do modelo adotado, indicando quanto

por cento a variação total explicada pela regressão representa da variação total. O R^2 é definido por:

$$R^2 = \frac{VE}{VT} = \frac{b^2 S_{xx}}{S_{yy}} = \frac{b S_{xy}}{S_{yy}}$$

Variando no seguinte intervalo : $0 \leq R^2 \leq 1$.

No caso de ter-se $R^2 = 1$, todos os pontos se situam “exatamente” sobre a reta de regressão. Diz-se, então, que o ajuste é perfeito. As variações de Y são 100% explicadas pelas variações de X, através da função especificada, não havendo, assim, desvios em torno da função estimada.

Por outro lado, se $R^2 = 0$, conclui-se que as variações de Y são exclusivamente aleatórias e a introdução da variável X no modelo não incorporará informação alguma sobre as variações de Y.

Neste caso, a associação entre X e Y é não linear, nos dados amostrais e a variável preditora X, de nada impede a relação da variação Y_i pela regressão linear. Na prática, é difícil encontrarmos $R^2 = 0$ ou $R^2 = 1$, e sim, valores entre estes limites.

O coeficiente R^2 pode ser visto como uma ferramenta para comparar diferentes modelos de regressão para uma mesma variável Y. Aquele que apresentar o maior R^2 deverá ser o melhor modelo. Há que se tomar certo cuidado, no entanto, quando os modelos comparados apresentarem diferentes números de variáveis (Vasconcelos & Alves, 2000).

Assim, para comparar modelos com diferentes números de variáveis, Theil propôs um coeficiente corrigido R^2 (Vasconcellos & Alves, 2000), em que:

$$\bar{R}^2 = 1 - \frac{SQR / (n - K)}{SQT / (n - 1)}$$

Desta forma, com base na fórmula original, corrigem-se as somas de quadrados pelos graus de liberdade.

Portanto, o \bar{R}^2 é o R^2 com ajustamento pelo tamanho da amostra e pelo número de graus de liberdade. Utiliza-se (n-1), porque a soma de quadrados total foi obtida com base nas n observações amostrais e da média de y, também calculada para a mesma amostra. Assim, para eliminar a interdependência, perde-se um grau de liberdade.

Os resultados da estimação de equações de regressão são, em geral, utilizados para prever valores da variável dependente associados a valores das variáveis explicativas que não fazem parte da amostra que deu origem à estimação. Assim, além das estimativas pontuais para a inclinação e o intercepto da reta de regressão, também é possível obter intervalos de confiança para estes parâmetros. Destaca-se que os intervalos de confiança são muito mais informativos que as estimativas pontuais, já que fornecem faixas dos possíveis valores que os parâmetros do modelo podem assumir, com um nível de confiança conhecido (Werkema & Aguiar, 1996).

Se para o modelo de regressão é válida a suposição que os ε_i são independentes e identicamente distribuídos $N(0, \sigma^2)$, então, um intervalo de $100(1-\alpha)\%$ de confiança para a inclinação β é dado por:

$$P\left(b - t_{\frac{\alpha}{2}, n-2} \sqrt{\frac{\sigma^2}{S_{xx}}} \leq \beta \leq b + t_{\frac{\alpha}{2}, n-2} \sqrt{\frac{\sigma^2}{S_{xx}}}\right) = 1 - \alpha$$

E, de modo análogo, um intervalo de $100(1-\alpha)\%$ de confiança para o intercepto α é:

$$P\left(a - t_{\frac{\alpha}{2}, n-2} \sqrt{\sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}}\right)} \leq \alpha \leq a + t_{\frac{\alpha}{2}, n-2} \sqrt{\sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}}\right)}\right) = 1 - \alpha$$

O intervalo de confiança fornece informação sobre a precisão das estimativas, no sentido de que quanto menor a amplitude do intervalo maior a precisão. Calculando-se intervalos de confiança para alguns valores de x , pode-se esboçar uma região em torno da reta estimada, indicando os limites superiores e inferiores destes intervalos. Esta região é também chamada de banda de confiança. As figuras 08 e 09 apresentam bandas de confiança correspondentes ao ajuste da reta de regressão com $R^2 = 0,95$ e $R^2 = 0,25$, respectivamente (Charnet, *et al*, 1999).

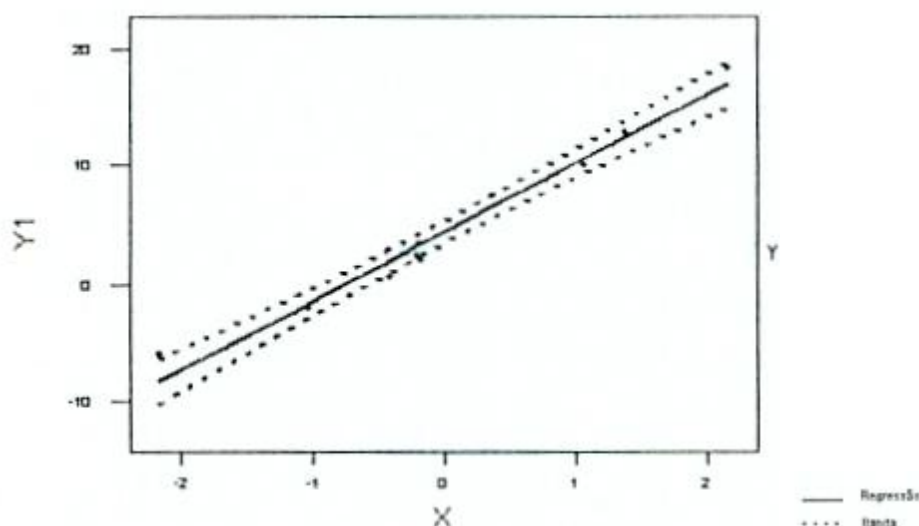


FIGURA 08 - Bandas de confiança para ajuste de reta com $R^2 = 0,95$.

(Figura adaptada de Charnet, *et al*, 1999)

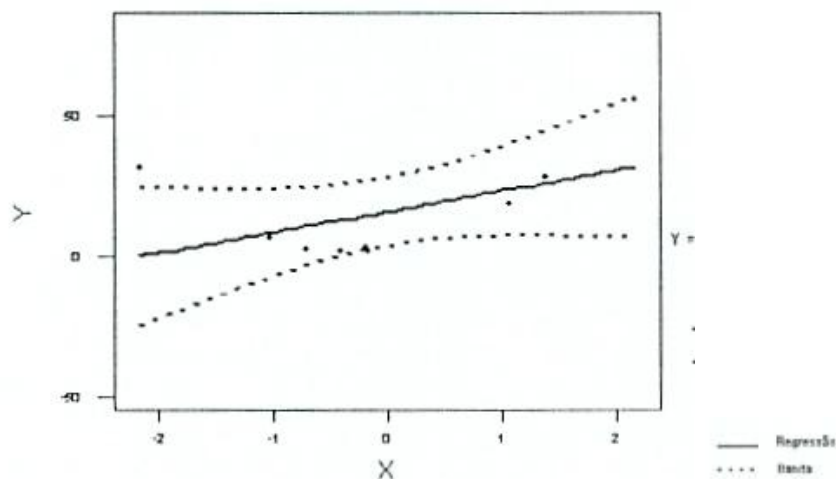


FIGURA 09 - Bandas de confiança para ajuste de reta com $R^2 = 0,25$.

(Figura adaptada de Charnet, *et al*, 1999)

Na Figura 08, pode-se perceber que, como o valor de R^2 é alto, as bandas de confiança têm amplitudes relativamente pequenas. Nota-se, também, que para intervalos correspondentes a valores centrais de x , as amplitudes são menores e, à medida que nos afastamos desta média, temos um aumento gradativo da amplitude.

Como já foi visto, para se estabelecer o modelo de regressão linear simples, é necessária a observação de alguns pressupostos. Os quatro principais pressupostos da regressão são: normalidade, homocedasticidade, independência dos erros e linearidade que serão descritos a seguir:

Normalidade: requer que os valores de Y sejam normalmente distribuídos para cada valor de x . Enquanto a distribuição dos valores de y_i em torno de cada nível de x não for extremamente diferente de uma distribuição normal, inferências sobre a linha de regressão e sobre

coeficientes de regressão não serão seriamente afetadas (Levine, Berenson, Stephan, 2000).

Os testes mais utilizados para verificar a normalidade de uma amostra de dados são o teste de χ^2 e o teste de Lilliefors. A vantagem do teste de Lilliefors sobre o χ^2 é que pode ser aplicado sem restrição, para pequenas amostras. Além disto, considera dados individualmente, não perdendo informação devido a agrupamentos, como ocorre no teste de χ^2 , sendo, na maioria das vezes, mais poderoso que aquele (Demétrio, 1978).

Teste de Lilliefors

O teste de Kolmogorov-Smirnov foi introduzido por Kolgomorov (1933) para verificar se uma série de dados pertence a uma determinada distribuição com média zero e variância conhecidas. Para se testar normalidade, Lilliefors (1967) introduziu uma modificação neste teste, ampliando o seu uso para os casos em que a média e a variância não são especificadas, mas sim, estimadas através dos dados da amostra (Demétrio, 1978).

Inicialmente, calculam-se a média \bar{X} , a variância S^2 dos dados e a variável $z_i = \frac{x_i - \bar{X}}{S}$.

Em seguida, ordena-se os z_i e considera-se:

$F(z_i)$ = proporção de valores esperados $\leq z_i$, valor obtido a partir da tabela de distribuição normal reduzida;

$S(z_i) = k/n$ proporção de valores obtidos $\leq z_i$;

onde k é o número de valores obtidos a partir dos valores observados $\leq z_i$, e n é o número de observações da amostra.

A estatística de Lilliefors (D) é definida da seguinte maneira:

$$D = \sup_{z_i} |F(z_i) - S(z_i)|$$

onde:

Sup = supremo em relação a z_i , isto é, a máxima distância vertical z_i entre $F(z_i)$ e $S(z_i)$.

Para a determinação de D, considera-se, em cada ponto z_i , as diferenças $|F(z_i) - S(z_i)|$ e $|F(z_i) - S(z_{i-1})|$ e toma-se apenas a maior delas.

O teste é bilateral onde se tem:

H_0 : É razoável estudar os dados através da distribuição normal;

H_1 : Não é razoável o estudo dos dados através da distribuição normal.

Rejeita-se a hipótese de nulidade, a um nível α de probabilidade, quando $D \geq d$. O valor d é encontrado no anexo A.

Homocedasticidade: requer que as variações em torno da linha de regressão sejam constantes para todos os valores de X. Isto significa que Y varia na mesma proporção, quando X for um valor baixo e quando X for um valor elevado. O pressuposto da homocedasticidade é importante na utilização do método dos mínimos quadrados, para determinar os coeficientes de regressão. Se houver sérios afastamentos deste pressuposto, pode-se aplicar transformações de dados ou métodos dos mínimos quadrados ponderados (Levine, Berenson, Stephan, 2000).

Detectando a heterocedasticidade

Uma forma de pesquisar a existência de heterocedasticidade consiste em estimar o modelo, utilizando mínimos quadrados e fazer o gráfico dos

resíduos de mínimos quadrados. Se os erros são homocedásticos, não deve haver qualquer padrão nos resíduos. Se os erros são heterocedásticos, podem tender a exibir maior variação.

Os resíduos são a diferença entre os valores observados Y_i e os correspondentes valores ajustados de \hat{y} . São, geralmente, denotados por ε_i e, definidos por:

$$\varepsilon_i = (y_i - \hat{y}_i)$$

Para o modelo de regressão $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ os resíduos são:

$$\varepsilon_i = y_i - (a + bx_i) = y_i - a - bx_i$$

Na sua magnitude, os resíduos são representados por um desvio vertical dos y_i observados e os pontos correspondentes a partir da função de regressão estimada, isto é, para os correspondentes valores ajustados \hat{y}_i . A forma envolve os desvios verticais de y_i para a desconhecida regressão linear.

Sob outro modelo, os resíduos são desvios verticais de y_i para os valores ajustados \hat{y}_i em relação a regressão linear estimada, e conhecida.

Os resíduos são muito utilizados para os estudos sobre a adequação do modelo de regressão aos dados.

Propriedades dos resíduos

a) Média: a média dos resíduos ε_i para o modelo de regressão linear simples é:

$$\bar{\varepsilon} = \frac{\sum_{i=1}^n \varepsilon_i}{n} = 0 \quad \text{onde } \bar{\varepsilon} \text{ é a média dos resíduos.}$$

A informação de que $\bar{\varepsilon}$ é sempre zero originou-se da informação que os erros verdadeiros ε_i , tem valor esperado $E(\varepsilon_i) = 0$.

b) Variância: a variância dos n resíduos ε_i é definida como:

$$\frac{\sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2}{n-2} = \frac{\sum_{i=1}^n (\varepsilon_i)^2}{n-2} = \frac{\text{SQE}}{n-2} = \text{QME}$$

Se o modelo for apropriado, o QME, como denotado anteriormente é um estimador não viesado da variância σ^2 dos termos dos erros.

c) Não dependência: os resíduos ε_i não são variáveis aleatórias independentes, porque envolvem valores ajustados \hat{y}_i , baseados na mesma função de regressão ajustada como resultado. Quando o tamanho da amostra é grande em relação ao número de parâmetros no modelo de regressão, o efeito da dependência sobre os resíduos ε_i não é muito importante e, para alguns fins, pode ser ignorado.

Independência de Erros: requer que o erro seja independente para cada valor de X . Este pressuposto, geralmente, se refere a dados que são coletados ao longo de um período de tempo. Quando os dados são coletados desta maneira, os resíduos para um determinado período de tempo são, freqüentemente, correlacionados com os do período de tempo anterior. Esta independência dos erros pode ser assegurada por um dos processos básicos da experimentação que é a casualização. Quando uma substancial correlação se encontra presente em um conjunto de dados, a validade de um modelo de regressão ajustado pode ficar seriamente comprometida (Levine, Berenson, Stephan, 2000; Demétrio, 1978).

Medindo a autocorrelação: a estatística de Durbin-Watson

A autocorrelação pode ser detectada e medida pela utilização de estatística de Durbin-Watson. Esta estatística mede a correlação entre cada resíduo e o resíduo para o período de tempo imediatamente antecedente àquele de interesse. A estatística de Durbin-Watson (D_w) é definida da seguinte maneira:

$$D_w = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

onde: e_i corresponde aos valores dos resíduos estimados pelo modelo;
 n é o número de observações.

Durbin e Watson também derivaram a distribuição amostral de D_w , que está apresentada na Figura 10. Os critérios para realização do teste são os seguintes:

1. Se o valor da estatística d for menor que d_l , então, rejeita-se H_0 , portanto existe autocorrelação, que é positiva;
2. Se o valor da estatística d for maior que $4-d_l$, então, rejeita-se H_0 , portanto existe autocorrelação, que é negativa;
3. Se o valor da estatística d se encontrar entre d_u e $4-d_l$, então, aceita-se H_0 , portanto, não existe autocorrelação;
4. Se o valor da estatística d se encontrar entre d_l e d_u ou entre $4-d_u$ e $4-d_l$, então, o teste é inconclusivo.

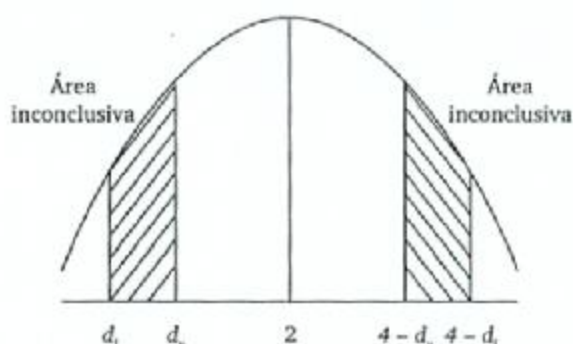


FIGURA 10 - Derivação da distribuição amostral D , a estatística de Durbin e Watson.
(Figura adaptada de Vasconcellos & Alves, 2000)

É importante destacar que o teste de Durbin e Watson apresenta algumas limitações:

1. Não é apropriado quando, entre as variáveis explicativas, está a variável dependente defasada;
2. Não é apropriado para testar a presença de autocorrelação decorrente de processos auto-regressivos de ordem superior a 1. Também não é adequado para processos que seguem um modelo de médias móveis ou um modelo ARMA de qualquer ordem;
3. É necessário que o modelo seja estimado com termo constante.

Linearidade: estabelece que a relação entre as variáveis deve ser linear. Duas variáveis poderiam perfeitamente ser relacionadas de uma maneira não-linear e o coeficiente de correlação linear seria igual a 0, indicando não haver qualquer relação.

Medindo a linearidade: a análise de correlação linear simples

Um problema de correlação simples surge toda vez que se quer verificar se há alguma relação entre variabilidades de duas variáveis

aleatórias. A medida de correlação entre duas variáveis é dada pelo coeficiente de correlação, definido por:

$$\rho_{xy} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

onde o campo de variação do coeficiente ρ_{XY} deve ser de: $-1 \leq \rho_{XY} \leq 1$

A interpretação do coeficiente de correlação como uma medida do grau de associação linear existente entre duas variáveis, é uma interpretação matemática, o que não implica, necessariamente, em uma relação de causa e efeito entre elas.

Estimativa do coeficiente de correlação

O estudo da correlação linear consiste em estimar o coeficiente de correlação através da disposição dos n pares (x_i, y_i) em torno de uma reta. Tal gráfico chama-se diagrama de dispersão. Por meio dele é possível determinar se as duas variáveis estão ou não relacionadas de forma linear, adotando r_{xy} como a estimativa de ρ_{xy} , que é o grau desta relação.

A estimativa do coeficiente de correlação linear, obtida a partir de uma amostra de pares (x_i, y_i) , é feita pelo coeficiente de correlação de Pearson, definido por:

$$r_{xy} = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{n} \right]} \sqrt{\left[\sum Y^2 - \frac{(\sum Y)^2}{n} \right]}} = \frac{S_{xy}}{\sqrt{S_{xx}} \sqrt{S_{yy}}}$$

onde: n = número de pares de dados;
 X e Y são as variáveis em estudo.

2.4 Comentários gerais do capítulo

Neste capítulo, desenvolveram-se as técnicas de controle de qualidade e análise de regressão que, combinadas, serão as ferramentas utilizadas para o monitoramento do processo.

È importante ressaltar que para o uso da análise de regressão seja válida, uma série de pressuposições devem ser atendidas, caso contrário, as conclusões poderão ficar comprometidas.

No capítulo seguinte, desenvolve-se a forma como o gráfico de controle de regressão deve ser implementado, onde se verifica de que modo é empregada cada técnica discutida neste capítulo.

3 GRÁFICO DE CONTROLE DE REGRESSÃO

No capítulo 2, fez-se a revisão de literatura, abordando-se os assuntos de gráficos de controle e análise de regressão. Estas técnicas serão utilizadas para o desenvolvimento do gráfico de controle proposto neste capítulo.

A metodologia, aqui proposta, será empregada para fazer o monitoramento de processos onde possa ser identificado um conjunto de variáveis correlacionadas. Para isto, são abordados assuntos como: previsão em regressão e construção do gráfico de controle de regressão.

3.1 Introdução

Para controlar a qualidade de um produto, Montgomery (1997) diz que é necessário a identificação e a medição das variações ocorridas no processo de produção por meio dos gráficos de controle. Eles são uma ferramenta útil para a avaliação do estado de controle estatístico de um processo, pois servem para diferenciar se as variações que ocorrem, são devidas a causas assinaláveis ou a causas aleatórias.

A necessidade de controle simultâneo de variáveis correlacionadas está no fato de que, muitas vezes, as variáveis parecem estar sob controle estatístico, quando analisadas separadamente, mas, na verdade, o sistema poderá estar fora de controle (Jackson, 1956).

3.2 Previsão em regressão

Uma das aplicações mais correntes dos modelos de regressão linear simples consiste na previsão do valor da variável dependente, Y , para um novo valor da variável independente, X . Utilizando-se o modelo de regressão com este propósito, é importante que, ao se fazer previsões, seja considerado o intervalo relevante da variável independente. Este intervalo relevante engloba todos os valores de X , desde o menor valor até o maior, utilizados no modelo de regressão.

Antes, porém, de começar a fazer previsões, deve-se atentar para algumas advertências importantes (Downing & Clark, 2000):

1. Qualquer previsão baseada em um modelo de regressão é uma previsão condicional, pois a previsão da variável dependente está sujeita ao valor da variável independente;
2. A reta de regressão é estimada, utilizando-se dados passados. Esta reta não poderá prever dados futuros, se a relação entre X e Y se modificarem;
3. Muitas previsões de regressão procuram prever valores de Y em situações em que o valor de X está fora do intervalo de valores de X , observados anteriormente. Tais previsões, conhecidas como extrapolações, são muito menos confiáveis do que previsões baseadas em valores da variável independente contidos no intervalo de valores, previamente observados;
4. O simples fato de existir uma forte associação entre duas variáveis não significa que haja, entre elas, uma relação de causa e efeito. Se chegarmos a

uma reta de regressão que se ajusta bem à relação entre X e Y, então, há quatro possibilidades:

- 4.1 Os valores de Y podem, realmente, depender dos valores de X, como suposto até agora;
- 4.2 A relação observada pode ser completamente casual;
- 4.3 Pode haver uma terceira variável, afetando X e Y;
- 4.4 Pode haver, efetivamente, uma ligação casual entre X e Y, mas é importante que X esteja causando Y. Talvez tenha-se determinado incorretamente qual seja a variável dependente e qual a variável independente.

3.3 Construindo o gráfico de controle de regressão

Quando a equação de regressão for empregada com o objetivo de controle, é necessário que a relação existente entre a variável de interesse e as variáveis utilizadas para seu controle sejam do tipo causa-e-efeito. No entanto, a existência de uma relação de causa-e-efeito não é necessária, se a equação for utilizada somente para predição. Neste caso, somente é necessário que as relações, que existiam no conjunto de dados original, usado para construir a equação de regressão, ainda sejam válidas (Werkema & Aguiar, 1996).

Supondo que se está interessado em fazer uma previsão para um valor futuro y_i – não observado – correspondente a $X = x_i$. Se usarmos o modelo de regressão linear simples ajustado, diz-se que $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$ é uma previsão de y_i . Nota-se que a previsão de y_i é igual ao estimador do valor esperado de y , para $X = x$, e o erro de previsão é definido por $y_i - \hat{y}_i$.

Como admite-se que cada nova observação, y é independente das observações anteriores (a partir das quais se obteve \hat{y}), tal observação será, também, independente de \hat{y} . Assim, como visto no capítulo 2, a distribuição de probabilidade de \hat{y} é dada por:

$$\hat{y}_i \approx N\left(\alpha + \beta x_i; \sigma^2 \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right]\right) \quad e, \quad Y_i \approx N(\alpha + \beta x_i; \sigma^2)$$

tem-se que:

$$E(y_i - \hat{y}_i) = 0 \quad e$$

$$V(y_i - \hat{y}_i) = V(y_i) + V(\hat{y}_i) = \sigma^2 + \sigma^2 \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right]$$

e, logo:

$$y_i - \hat{y}_i \approx N\left(0, \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right]\right)$$

onde σ^2 pode ser estimado por:

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2} = \text{QME} \quad (3.1)$$

Para cada estimativa particular da posição da reta de regressão e para cada valor particular de X , os limites do intervalo de previsão a $(1 - \alpha)100\%$ são dados por:

$$\hat{Y} \pm t_{\gamma} S \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}} \quad (3.2)$$

$$\text{onde: } \gamma = \left(n - 2, \frac{\alpha}{2} \right)$$

Para o uso em gráficos de controle, sabe-se que a média do processo pode variar em múltiplos do desvio padrão, como visto no item 2.1, do capítulo 2, ou seja, $\bar{x} \pm 3S$, $\bar{x} \pm 2S$, ... formando-se assim os LIC e LSC. Utiliza-se o valor de S dado em (3.1), substituindo-se em (3.2) o valor

de t tabelado, por um valor múltiplo do erro padrão, obtendo-se os limites superior e inferior de controle, para o gráfico de controle de regressão:

$$LSC_{\hat{y}} = \hat{y} + 3S \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}}$$

$$LMC_{\hat{y}} = a + bx = \hat{y}$$

$$LIC_{\hat{y}} = \hat{y} - 3S \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}}$$

A Figura 11, a seguir, representa o gráfico de controle de regressão, para o exemplo de Mandel (1969).

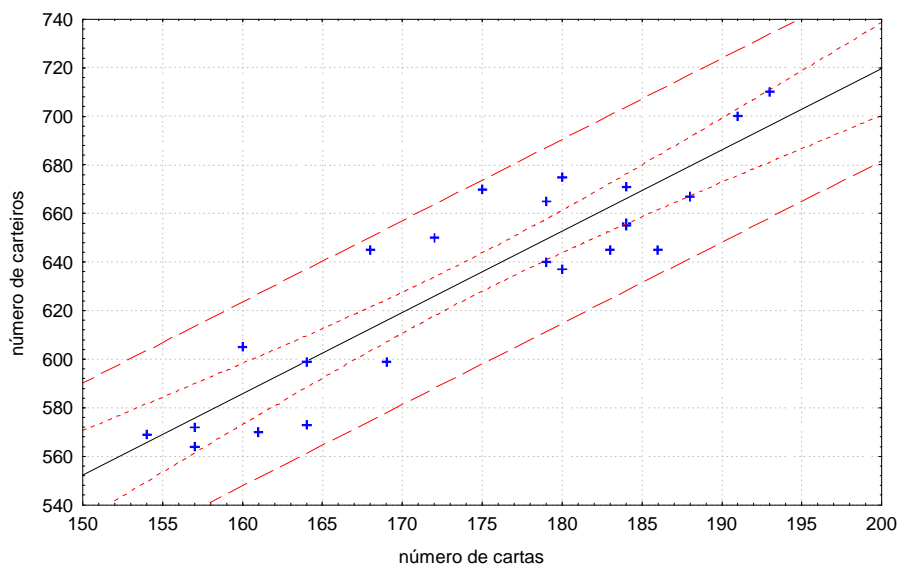


FIGURA 11 - Representação do gráfico de controle de regressão.

Fonte: Mandel (1969)

Olin (1998) comparou três métodos para a construção de limites de controle. Estes métodos foram os seguintes:

- Limites simples 3σ : $\hat{y} \pm 3\hat{\sigma}$;
- Intervalos de predição: $\hat{y} \pm k(x_0, \gamma) \hat{\sigma}$;
- Intervalos de confiança: $\hat{y} \pm c(x_0, \gamma, \alpha) \hat{\sigma}$

onde as constantes $k(x_0, \gamma)$ e $c(x_0, \gamma, \alpha)$ e os métodos usados para estimar $\mu(x, \beta)$ e $\sqrt{\text{var}(y)}$ dependem do modelo estatístico.

Ele concluiu que, embora os intervalos de predição e confiança pareçam ser tecnicamente mais corretos de um ponto de vista estatístico, os limites simples 3σ são recomendados, pois são de fácil compreensão, simples e similares ao gráfico de Shewhart.

Mandel (1969), em seu artigo, trabalhou com os limites simples considerando dois desvios em torno da linha de mínimos quadrados em vez de três desvios.

Embora sabendo-se que os limites de controle em torno da linha de regressão deveria alargar-se progressivamente ao longo da média, dois fatos fazem a aplicação do limite de controle paralelo possível. Primeiro, se a variação sobre a linha de regressão não for muito grande, os limites de controle não serão curvos. Segundo, as linhas de limites paralelas darão controle mais rígido a distâncias extremas da média (Mandel, 1969).

Quando se considera o conjunto de todos os valores possíveis de X , os intervalos formam uma banda de previsão e estas bandas de previsão podem ser utilizados para se organizar o gráfico de controle de regressão e, segundo Mandel (1969), o gráfico de controle de regressão difere do gráfico de controle tradicional nos seguintes aspectos :

- A média designada não é constante e esta variabilidade é dependente da magnitude da variável independente;
- As linhas dos limites de regressão são paralelas à linha de regressão, o que é melhor que uma linha paralela ao eixo x , pois os limites de controle acompanham a tendência existente entre as duas variáveis;

- A construção dos gráficos de controle de regressão é mais complexa que a construção do gráfico de controle convencional.

Para se estabelecer o gráfico de controle de regressão, a partir dos dados, é necessário, primeiramente, que se analise o diagrama de dispersão para verificar a linearidade da relação e para detectar pontos atípicos.

Após estabelecer os limites de controle a partir dos dados históricos, que estão sob controle, sobrepõe-se os dados novos contra os limites anteriores. Um sinal fora de controle indica uma inconsistência entre o processo novo e o processo antigo, desde que uma investigação do ponto fora de controle não revele nenhuma irregularidade no processo.

Neste estudo, quando um ponto situa-se acima do limite superior de controle indica que, provavelmente, a produtividade foi significativamente inferior ao nível de desempenho estabelecido no período.

Um ponto abaixo do limite inferior de controle indica uma possível melhoria na produtividade. Note que um ponto fora do limite inferior de controle, embora sendo um indicador de um bom desempenho, deve merecer também uma investigação.

Um novo gráfico de controle deve ser estabelecido, se houver uma mudança significativa no nível do desempenho. Se a tendência for um aumento na produtividade, um novo gráfico poderá ser construído a partir dos novos dados, mas, se a tendência for uma perda significativa em produtividade, então, um novo gráfico deverá ser estabelecido somente se essa mudança for causada por uma causa constante do sistema.

3.4 Comentários gerais do capítulo

Neste capítulo, desenvolveu-se o gráfico de controle de regressão, apresentando-se seus limites de controle e as interpretações para pontos fora de controle.

Mostra-se, também, que este gráfico pode ser usado para comparar o desempenho atual do sistema com o desempenho que mantinha anteriormente.

No próximo capítulo, exemplifica-se esta metodologia através de dados reais, coletados numa empresa de prestação de serviços.

4 APLICAÇÃO DA METODOLOGIA

Este capítulo tem por objetivo a aplicação da metodologia de gráficos de controle de regressão, desenvolvida no capítulo 3. Para a aplicação da metodologia, serão utilizados os conceitos e as técnicas de análise de regressão e controle de qualidade, apresentadas no capítulo 2.

Para esta aplicação, em primeiro lugar, far-se-á uma descrição da empresa, de onde os dados foram coletados e algumas definições que serão importantes para o bom entendimento da pesquisa. Em segundo, realizar-se-á uma análise exploratória dos dados, para que seja possível uma maior familiarização, sendo, também, possível a identificação de algumas causas que estão afetando o processo produtivo. Nesta etapa, são utilizadas as folhas de verificação das rotas feitas durante a etapa de coleta. E, finalmente, o gráfico de controle de regressão será utilizado para a monitoração conjunta das variáveis em estudo e, também, para que seja possível o conhecimento do comportamento das variáveis. Desta maneira, procurar-se-á identificar os períodos de maior e menor volume de demanda, fazendo com que a empresa esteja preparada para os possíveis picos de trabalho.

4.1 Introdução

A degradação do meio ambiente tem sido alvo da preocupação de órgãos e instituições, sejam eles governamentais ou não, bem como das comunidades, seus representantes e das pessoas em geral.

Tal preocupação recai, fatalmente, sobre a geração do lixo e o impacto que este processo causa ao meio ambiente, alterando a qualidade de vida no planeta, principalmente, no perímetro urbano.

Lixo é qualquer resto que resulta de atividade humana, podendo ser de origem industrial, doméstica, hospitalar, comercial, agrícola, de serviços, entre muitos outros. Apresenta-se sob os estados sólido, semi-sólido e líquido e pode ser classificado em orgânico e inorgânico.

O lixo é gerado há muito tempo, em grande quantidade e continuamente. A reutilização e reciclagem são práticas bastante antigas. "Sucateiros" da antigüidade recolhiam espadas nos campos de batalha para fazer novas armas.

Até a metade do século passado, a composição do lixo era, predominantemente, de matéria orgânica, ou seja, de resto de comida. Com o avanço da tecnologia, materiais como plásticos, isopor, pilhas, baterias de celulares e de filmadoras, resíduos hospitalares e lâmpadas são presenças, cada vez mais constantes, nos lixos em todo o planeta. E, desta forma, a sua reutilização por meio da reciclagem se tornou um método eficiente, para que não se tenha um acúmulo de resíduos. Desta forma, o que era considerado lixo passou a ser entendido como resíduos que, na maioria das

vezes, são reutilizados, tornando os impactos ambientais cada vez menores e protegendo, de certa forma, os recursos naturais existentes no planeta.

Os ciclos naturais de decomposição e reciclagem da matéria podem reaproveitar o lixo humano. Contudo, uma grande parte deste lixo sobrecarrega o sistema. O problema se agrava, porque muitas das substâncias manufaturadas pelo homem não são biodegradáveis, isto é, não se decompõem facilmente como: vidros, latas e alguns plásticos que levam muitos anos para se decompor. Estes resíduos provocam a poluição.

A estrutura dos órgãos públicos em relação ao tratamento do lixo é inadequada, passando pela forma de coleta até o tratamento final, que são os aterros sanitários, compostagem e incineração. As cidades brasileiras produzem cerca de 90 mil toneladas de lixo por dia, sendo que 34 milhões de habitantes urbanos não dispõem de coleta domiciliar. Na zona urbana, apenas 35%, conta com rede pública de esgoto, enquanto que 34% da população lança dejetos a céu aberto.

De um modo geral, nas cidades, o lixo é coletado por uma empresa contratada pelo governo ou por particulares, pois, devido à crescente produção de resíduos por parte da população, a terceirização da coleta de resíduos sólidos compactáveis apresentou-se como a melhor alternativa para as prefeituras e órgãos responsáveis pelo serviço.

Neste contexto, a Prefeitura Municipal de Santa Maria terceirizou este serviço, contratando a Empresa PRT - prestação de serviços. A empresa encontra-se totalmente preparada para executar esta tarefa, embora encontre dificuldades em dimensionar a frota de caminhões compactadores para realizar a coleta dos resíduos, tanto na zona urbana como na zona rural. Esta dificuldade não é encontrada apenas na cidade de Santa Maria, mas

também em outras localidades onde a empresa atua. Desta forma, pretende-se fornecer à empresa informações sobre os períodos de maior e menor volume de resíduos a serem coletados, pois com a identificação destes períodos e valores de volumes em quilogramas, a empresa poderá dimensionar a frota necessária para a realização desta tarefa.

4.2 Descrição da PRT - empresa de prestação de serviço

O grupo PRT surgiu da lacuna criada pelo mercado das empresas de prestação de serviços, ciente de sua potencialidade e investindo em qualidade e agilidade na administração dos contratos, buscando afastar a nefasta opinião formada sobre as empresas neste ramo.

O Grupo PRT é constituído pela PRT – Prestação de Serviços Ltda, PRT – Engenharia de Saneamento e Meio Ambiente e Vigillare Serviços de Segurança e Monitoramento de Sistemas Ltda, que contam com mais de 1.500 funcionários.

Na Figura 12, apresenta-se a logomarca que identifica a empresa nos diversos setores de atuação.



FIGURA 12 - Logomarca da empresa.

Nesta primeira etapa, o objetivo foi atingido, haja vista, que, num período de cinco anos, tornou-se uma das três maiores empresas do estado no segmento de prestação de serviços, atuando com parceiros públicos e privados, desenvolvendo um trabalho eficaz e reconhecido no estado do Rio Grande do Sul.

Atualmente, a PRT agrega novos serviços a sua sistemática operacional, destacando-se a engenharia ambiental, através da implantação de projetos, execução e operação de aterros sanitários, assim como a coleta de resíduos sólidos compactáveis, a coleta seletiva e, mais recentemente, a coleta de resíduos de saúde.

Hoje, a PRT, com matriz em Santa Maria e filiais nas cidades de Ijuí, São Borja, Jaguarão, Canoas, São Gabriel e Porto Alegre, presta serviços em mais de 100 municípios, situando-se entre as três maiores empresas de limpeza e entre as duas maiores empresas de coleta de resíduos do estado.

A empresa não possuía banco de dados, pois recebia um valor fixo da Prefeitura Municipal de Santa Maria pelos serviços prestados, independente da quantidade de resíduos coletados.

A partir de 2001, com a nova administração municipal, a empresa passou a receber pelo peso de resíduos coletados, fazendo-se, então, necessário o controle desta variável.

A coleta de dados junto à empresa começou no mês de março. O primeiro passo foi organizar os formulários preenchidos pelos motoristas e montar as planilhas que serviriam como banco de dados.

Verificou-se que, no mês de março, havia muitos dados incompletos. No mês de abril, a balança estava quebrada e o peso era estimado pela

última pesagem do caminhão, o que acarretou um peso superestimado e, no mês de maio, foram perdidos os registros da primeira quinzena.

Após todos estes problemas serem detectados, enquanto da realização deste trabalho, houve uma reorganização na empresa com a contratação de funcionários encarregados da organização e coleta destes dados.

Estes funcionários, no mês de junho, realizaram um treinamento junto com os motoristas de como deveriam preencher o formulário adequadamente. Como no mês de julho ainda ocorreram muitos problemas, foi realizado outro treinamento e, com isto, percebeu-se uma melhora no preenchimento dos formulários nos meses de agosto e setembro.

4.3 Descrição do processo

O processo de coleta de resíduos é uma importante etapa no processo de tratamento de resíduos, pois é nesta etapa que os materiais orgânicos e inorgânicos devem ser coletados separadamente e a coleta seletiva deve ser implementada para que o processo de reciclagem tenha o seu início. Neste trabalho, apenas a coleta não seletiva será abordada, pois é feita por caminhões compactadores, mostrados na Figura 13, e os resíduos coletados são destinados ao aterro sanitário de Santa Maria, localizado no Bairro Salgado Filho.



FIGURA 13 - Caminhões compactadores da empresa PRT - prestações de serviços de Santa Maria.

Para o recolhimento dos resíduos domiciliares e comerciais, da cidade de Santa Maria, a Empresa PRT disponibiliza oito (08) caminhões compactadores, com capacidade média de 7000 Kg. A empresa realiza, também, a coleta do lixo hospitalar e, no centro da cidade, existe a coleta seletiva.

A empresa dividiu a cidade em 22 setores, a fim de obter um melhor desempenho na atividade de coleta e um melhor aproveitamento do número de caminhões compactadores. Dos 22 setores, 5 possuem coletas diárias, enquanto que os outros possuem coletas alternadas, três vezes por semana, sendo que, no domingo não há coleta em nenhum dos setores.

No Quadro 01, apresenta-se a divisão da cidade nos 22 setores de coleta, bem como seus horários de recolhimento.

As variáveis envolvidas no processo de recolhimento de resíduos são a quilometragem percorrida pelo caminhão compactador, o volume líquido de resíduos recolhido, o tempo gasto no recolhimento e o consumo de combustível.

A variável quilometragem é informada pelos próprios motoristas que anotam o valor inicial e o final da quilometragem do caminhão, tendo como

referência a empresa. Para este procedimento, um formulário próprio é utilizado pela empresa, encontrado no Anexo A.

Os valores do peso líquido são fornecidos por um funcionário da Prefeitura Municipal de Santa Maria que é encarregado da pesagem dos caminhões na entrada e saída do aterro, encontrando-se, assim, o peso líquido de resíduos que foi depositado. Desta forma, é efetuado o pagamento mensal à empresa, baseado no peso líquido depositado no aterro.

QUADRO 01 - Horário de recolhimento de resíduos por setor - Santa Maria

Setor 01 - Itararé 02 - Schirmer 03 - Km 3 04 - Camobi 05 - Ferrari 06 - UFSM	Segunda - Quarta - Sexta Turno da manhã
Setor 07 - Urlândia 08 - Dom Antônio	Segunda - Quarta - Sexta Turno da tarde
Setor 09 - Centro 10 - Centro 11 - Rosário 12 - Medianeira 13 - Leste	Diário Turno da noite
Setor 14 - Tancredo Neves 15 - Santa Marta 16 - Prado 17 - Cooteto 18 - Lameira 19 - Patronato	Terça - Quinta - Sábado Turno da manhã
Setor 20 - Oliveira 21 - Carolina 22 - Casa de Saúde	Terça - Quinta - Sábado Turno da tarde

A distância percorrida em cada setor é sempre a mesma, pois os percursos das rotas são pré-determinados. A variabilidade da quilometragem está no fato de que, quando há aumento no volume de resíduos, os caminhões precisam ir mais vezes ao aterro sanitário, localizado no Bairro Salgado Filho, aumentando, desta maneira, a quilometragem percorrida e, conseqüentemente, o gasto com combustível e o tempo de recolhimento. Isto poderia ser melhor dimensionado, se fosse possível identificar os períodos em que um excesso de resíduos poderá ocorrer. Desta forma, um caminhão compactador com maior capacidade, poderá ser destinado à rota onde este fato poderá ocorrer, o mesmo se aplica no caso inverso. Desta maneira, uma economia de combustível e tempo seria realizada, proporcionando benefícios para a empresa.

4.4 Análise descritiva dos dados

Neste item, faz-se uma análise descritiva das variáveis envolvidas no estudo, pois, desta maneira, será possível encontrar alguns pontos que mereçam uma maior atenção.

As variáveis do processo, analisadas neste trabalho, foram a quilometragem diária percorrida pelos caminhões e o volume de resíduos recolhidos por dia.

Primeiramente, realizou-se uma análise descritiva dos dados nos meses de julho, agosto e setembro, para a verificação do comportamento das variáveis. Na Tabela 01, apresentam-se as principais medidas descritivas da variável peso líquido e, na Tabela 02, apresentam-se as principais medidas descritivas da variável quilometragem.

TABELA 01 - Tabela das medidas descritiva para a variável peso*.

Meses	Medidas Descritivas					
	Valor de n**	Valor Mínimo	Valor Máximo	Média	Desvio Padrão	Coefficiente de Variação
Julho	26	15610	135285	64019,65	31182,41	0,4871
Agosto	27	87583	201535	145528,6	28714,89	0,1973
Setembro	25	58330	165765	123641,4	31054,80	0,2512

* A variável é expressa em quilogramas (Kg)

** n é o número de dias com recolhimento no mês

Pela Tabela 01, percebe-se que, no mês de julho, o volume médio de resíduos coletados é inferior aos meses de agosto e setembro. Esta diferença gerou uma investigação ampla, no sentido de se estabelecer as possíveis causas deste declínio no volume de resíduos coletados, conforme mostrado no Quadro 02.

TABELA 02 - Tabela das medidas descritiva para a variável quilometragem*.

Meses	Medidas Descritivas					
	Valor de n**	Valor Mínimo	Valor Máximo	Média	Desvio Padrão	Coefficiente de Variação
Julho	26	99	1903	542,3846	366,6775	0,6760
Agosto	27	732	1555	1096,4440	153,8572	0,1403
Setembro	25	494	1182	939,7200	168,7697	0,1796

* A variável é expressa em quilômetros (Km)

** n é o número de dias com recolhimento no mês

Como o volume de resíduos coletados no mês de julho é inferior em relação aos outros meses, conseqüentemente, a quilometragem percorrida pelos caminhões também é menor para este mês, conforme mostrado na Tabela 02.

Pode-se, também, observar, na Tabela 02, que o coeficiente de variação para o mês de julho é o mais alto, mostrando que ocorreu uma

variação muito grande de quilometragem rodada pelos caminhões dia-a-dia. Esta mesma variação é, também, observada para a variável peso, conforme mostra a Tabela 01.

As tabelas apresentadas anteriormente podem ser melhor visualizadas por meio do gráfico de "Box-plot" que é muito útil para a comparação entre as variáveis, fornecendo uma boa visualização de suas variabilidades, conforme pode ser visto na Figura 14.

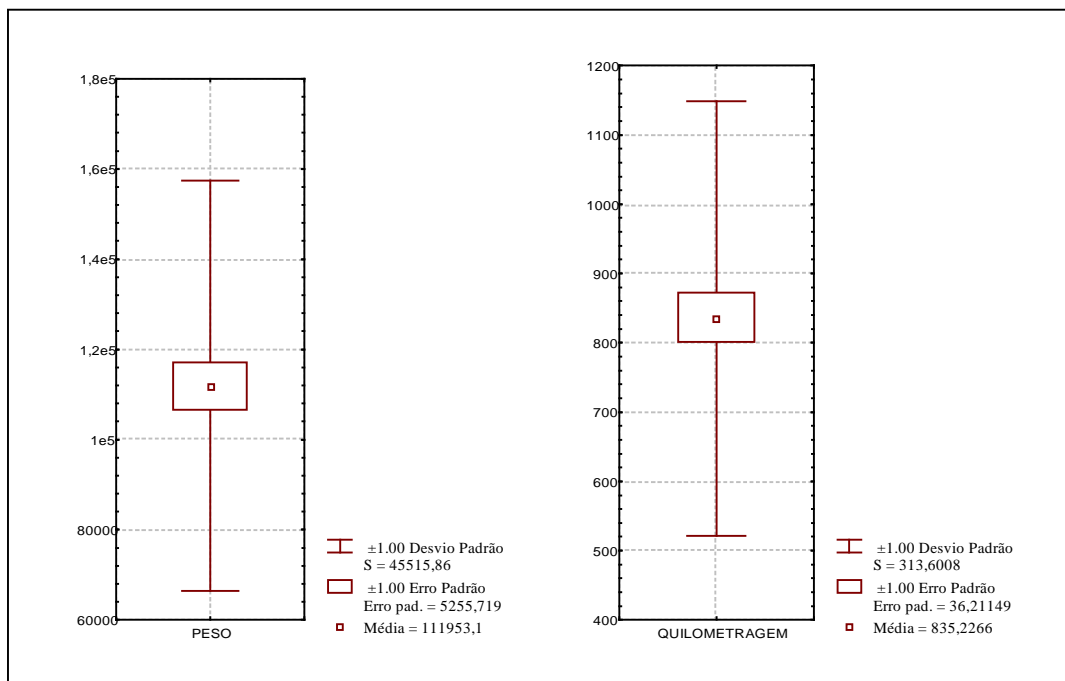


FIGURA 14 - Gráfico Box Plot para as variáveis peso e quilometragem.

Observando a Figura 14, percebe-se que as variáveis peso e quilometragem possuem o mesmo comportamento. Como a mediana está a mesma distância do 1º e do 3º quartil, existem fortes indicações de simetria nos dados, além disso, sendo a amplitude bem maior que a distância interquartílica, conclui-se que há grande dispersão das observações.

Para melhor observar o comportamento da variável volume líquido (peso), traça-se um gráfico de seqüência, mostrado na Figura 15. Esta variável mede a quantidade total de resíduos recolhida, por dia, na cidade de Santa Maria, sendo expressa em quilogramas (Kg).

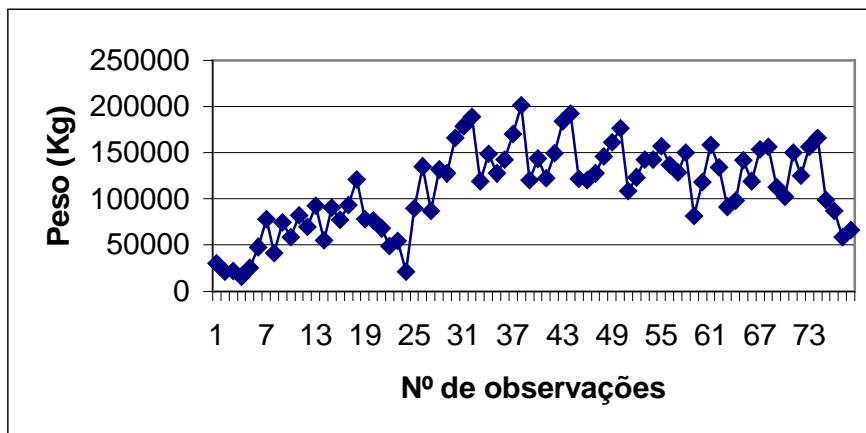


FIGURA 15 - Gráfico representativo do peso de resíduos recolhidos por dia, em Santa Maria, de julho a setembro de 2001.

Esta variável foi analisada no período correspondente aos meses de julho, agosto e setembro de 2001. Observa-se que existe uma mudança de nível na média do processo, no mês de julho. Este comportamento inspira uma investigação criteriosa, para tentar identificar as causas do ocorrido.

Um dos motivos para o reduzido volume de resíduos coletados é o fato do mês de julho coincidir com o mês de férias escolares. Além disso, houve muitas falhas nos relatórios, sendo que os motoristas que realizavam as coletas não preenchiam os formulários, veja o Quadro 02.

QUADRO 02 - Quadro demonstrativo do número de registros e o número de dias no mês que deveria haver coletas, por rota.

Rota	Nº de coletas registradas			Nº de dias de coletas		
	Julho	Agosto	Setembro	Julho	Agosto	Setembro
01	09	16	13	13	14	12
02	06	14	13	13	14	12
03	10	11	12	13	14	12
04	00	13	12	13	14	12
05	12	14	12	13	14	12
06	00	14	09	13	14	12
07	06	14	08	13	14	12
08	04	09	01	13	14	12
09	09	23	18	26	27	25
10	07	26	17	26	27	25
11	10	27	20	26	27	25
12	19	27	26	26	27	25
13	11	25	15	26	27	25
14	00	13	08	13	13	13
15	06	11	13	13	13	13
16	10	12	07	13	13	13
17	00	12	11	13	13	13
18	11	13	12	13	13	13
19	07	12	12	13	13	13
20	00	03	07	13	13	13
21	06	12	12	13	13	13
22	07	10	03	13	13	13
Sem rotas* / Nº de dias registrados	06	10	23	26	27	25

* Número de registros preenchidos pelos motoristas em que não havia a indicação de que rota havia sido realizada a coleta.

A quilometragem percorrida pelos caminhões é uma variável muito importante, pois o gasto com combustível depende de seus valores. Para observar seu comportamento, traça-se um gráfico de seqüência, mostrado na Figura 16. Esta variável expressa a quantidade total da quilometragem percorrida pelos caminhões, por dia, na cidade de Santa Maria.

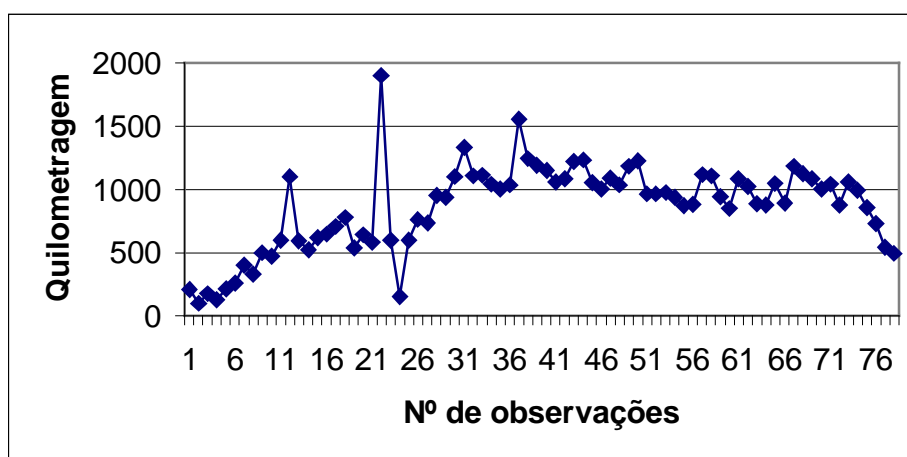


FIGURA 16 - Gráfico representativo da quilometragem percorrida pelos caminhões compactadores, por dia, em Santa Maria, de julho a setembro de 2001.

Esta variável foi analisada no período de julho a setembro de 2001. Pela Figura 16, observa-se que houve uma mudança de nível na média do processo. Esta mudança ocorreu próxima à observação 30 que corresponde ao período final de julho e início de agosto.

Observa-se, também, que a série apresenta 04 observações atípicas do restante da massa de dados, conhecida como "outlier".

4.5 Análise de correlação nos dados

Neste item, uma análise do comportamento conjunto dos dados deve ser conduzida para, posteriormente, ser ajustada à equação de regressão aos dados.

O interesse, neste trabalho, é construir um modelo de regressão que relacione a quilometragem diária percorrida pelos caminhões compactadores (Y) com a quantidade de resíduos coletados por dia (X). Observe que a quantidade diária de resíduos coletados não pode ser controlada. O procedimento adotado consistiu em observar as variáveis resíduos coletados (x_i) e a quilometragem percorrida (y_i) em cada um dos dias do período analisado. Portanto, (x_i, y_i) são variáveis aleatórias conjuntamente distribuídas e as principais inferências de interesse que serão estabelecidas, estarão condicionadas aos valores observados para X.

Como o interesse está em conhecer o tipo de relacionamento existente entre a quilometragem percorrida pelos caminhões compactadores e a quantidade de resíduos coletados, foi construído um diagrama de dispersão a partir dos dados coletados e que está apresentado na Figura 17. A configuração dos pontos, neste diagrama, sugere a existência de um relacionamento linear entre a quilometragem percorrida e a quantidade de resíduos coletados.

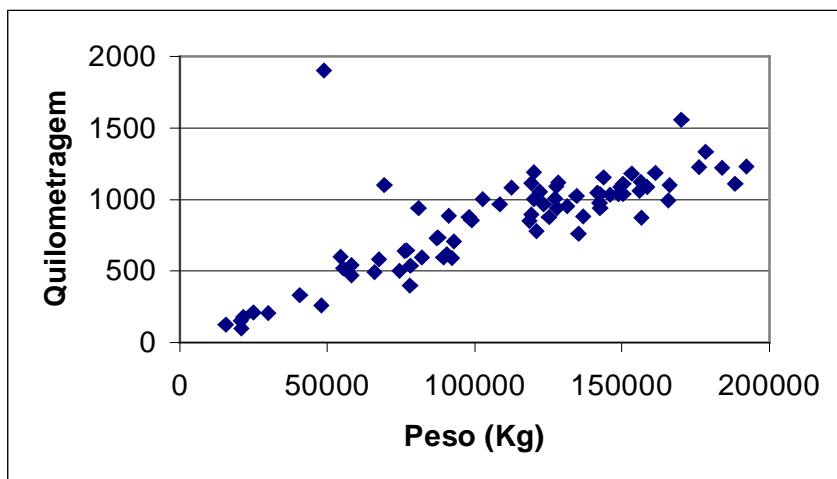


FIGURA 17 - Diagrama de dispersão das variáveis peso e quilometragem para os meses de julho a setembro de 2001.

Com o objetivo de quantificar o relacionamento linear entre X e Y, visualizado a partir da análise do diagrama de dispersão, calculou-se o coeficiente de correlação amostral entre X e Y, visto no Capítulo 2 (item 2.2). O coeficiente de correlação amostral foi de $r = 0,7773$, o que revela que existe uma forte correlação positiva entre as variáveis.

O diagrama de dispersão, apresentado na Figura 17, revela três pontos que se desviam de um padrão linear e, conforme Mandel (1969), esta situação é passível de uma investigação. Se estes pontos representarem uma condição não usual, podem ser excluídos dos cálculos do gráfico de controle de regressão, pois, nesta etapa, se está interessado em encontrar os limites de controle para o gráfico de regressão que revele uma situação de estabilidade do sistema. Desta maneira, será possível uma comparação com etapas posteriores do processo de coleta dos resíduos. Assim, o processo estável servirá como padrão.

Investigando-se estes três pontos, verificou-se que a quilometragem rodada pelos caminhões estava muito alta para o volume de resíduos coletados nestes dias, não havendo, porém, nenhum motivo especial para isto, a não ser um possível engano nas anotações nas fichas de controle.

Como as causas dos três pontos que se desviaram do padrão linear, mostrado na Figura 17, não são constantes no sistema, estes pontos foram excluídos do cálculo do gráfico de controle de regressão. Desta maneira, utiliza-se apenas o processo considerado estável para se encontrar os limites de controle que servirão para monitorar o processo futuro.

Com a exclusão dos três pontos, novamente, construiu-se o diagrama de dispersão, mostrado na Figura 18, e calculou-se um novo coeficiente de correlação.

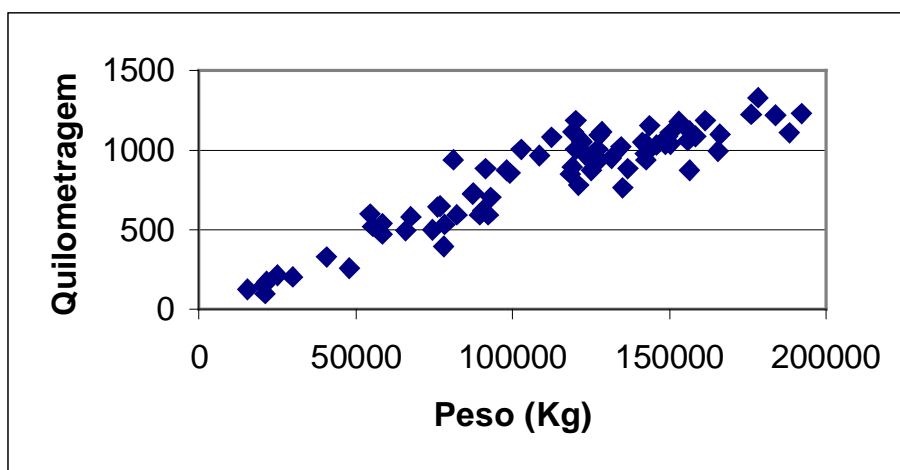


FIGURA 18 - Diagrama de dispersão das variáveis peso e quilometragem, para os meses de julho a setembro, sem os pontos que desviam do padrão linear.

O novo coeficiente de correlação amostral calculado, desconsiderando os três dias que desviaram do padrão linear, foi de $r = 0,9210$, mostrando que a correlação entre as variáveis é, ainda, mais forte.

Calculou-se, também, o coeficiente de determinação que mede a proporção da variação que é explicada pela variável independente no modelo de regressão. O resultado foi $R^2 = 0,8482$, o que significa que 84,82% da quilometragem percorrida pelos caminhões compactadores pode ser explicada pela variabilidade no volume de resíduos coletados.

Como as variáveis aqui estudadas são linearmente correlacionadas e o seu alto nível de correlação sugere que possa haver uma relação funcional entre elas, uma reta de regressão será ajustada, determinando as relações entre estas variáveis.

4.6 Ajuste da equação de regressão e adequação do modelo

Construir um modelo de regressão linear de Y sobre X consiste em obter, a partir dos valores observados, uma reta que melhor represente a relação verdadeira entre estas variáveis. A determinação dos parâmetros dessa reta é denominada ajustamento.

Estimou-se os valores dos parâmetros da equação de regressão linear simples, pelo método de mínimos quadrados, como visto no Capítulo 02 (item 2.2), encontrando-se a seguinte equação:

$$\hat{y} = 124,82 + 0,0063x \quad (4.1)$$

(3,29) (20,198)

Onde, entre parênteses, estão os valores da estatística “t” calculada para as estimativas dos parâmetros.

A equação (4.1) pode ser rescrita da seguinte forma:

$$\text{Quilometragem (estimada)} = 124,82 + 0,0063 (\text{peso}) \quad (4.2)$$

Quando um modelo de regressão é escolhido em uma pesquisa, deve-se verificar se é adequado para os propósitos a que se destina. Uma ou mais características do modelo podem não se ajustar aos dados da amostra, então, é importante investigar a aptidão do modelo.

Para avaliar se a reta de regressão ajustada em 4.2 pode ser empregada como base para futuras análises, foi conduzido um estudo de sua adequação aos dados coletados, por meio de análise de resíduos. Os gráficos de resíduos que foram construídos, são apresentados nas Figuras 19 a 21. A análise destes gráficos e dos testes realizados, ao nível de significância de 5%, ou seja, $\alpha = 0,05$, não indicou inadequação do modelo de regressão.

O gráfico de resíduos contra os valores ajustados é apresentado na Figura 19. Neste gráfico, os resíduos estão situados, aproximadamente, em uma faixa horizontal centrada em $e_i = 0$.

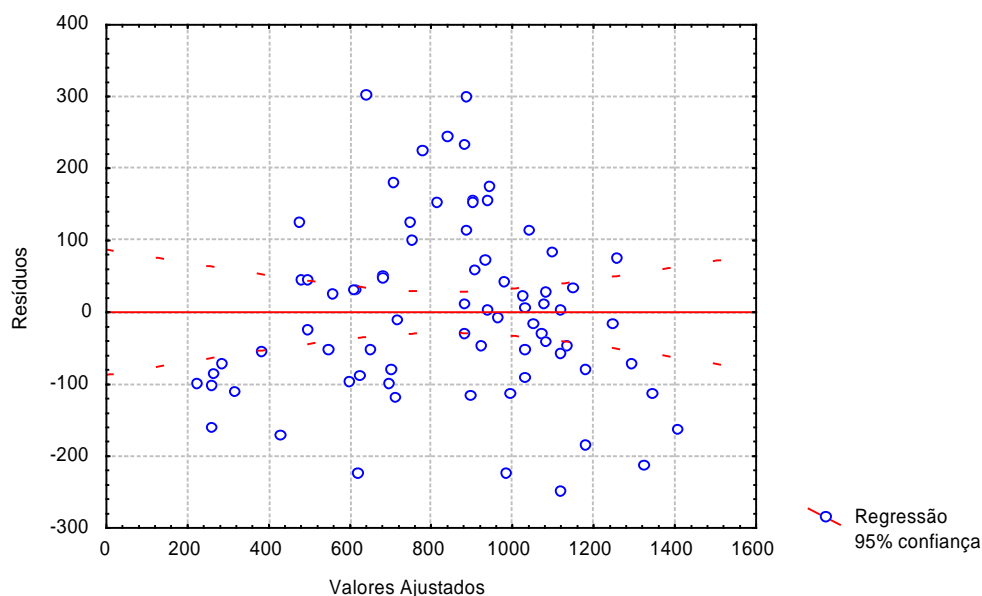


FIGURA 19 - Gráfico de resíduos contra os valores ajustados, para os dados coletados.

Analisando-se o gráfico da Figura 19, não são percebidas indicações de que os pressupostos de linearidade e homocedasticidade tenham sido violados.

Outro pressuposto do modelo de regressão que deve ser considerado, é a independência dos resíduos. Este pressuposto é, geralmente, violado, quando os dados são coletados ao longo de períodos sequenciais de tempo, uma vez que um resíduo, em qualquer ponto no tempo, pode tender a ser idêntico a resíduos em pontos adjacentes no tempo.

Uma vez que os dados foram coletados diariamente num período consecutivo de 3 meses, os resíduos foram plotados ao longo do tempo para a verificação da existência de um padrão, o que nos mostra a Figura 20.

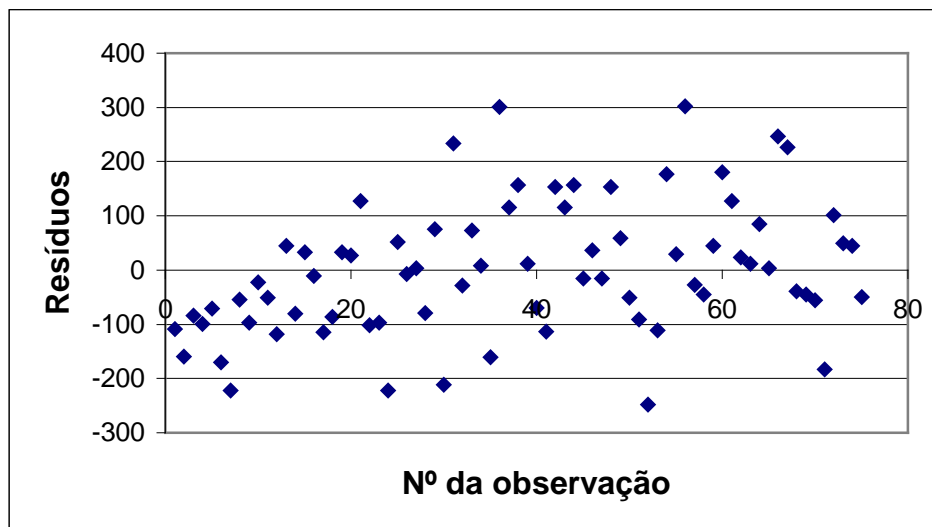


FIGURA 20 - Gráfico para resíduos ao longo do tempo para os dados analisados.

A partir da Figura 20, observa-se que os pontos não tendem a flutuar para cima e para baixo, em um padrão cíclico. Este padrão indica que não há uma violação no pressuposto da independência dos resíduos.

Além do gráfico de resíduos, a autocorrelação também pode ser detectada e medida pela utilização da estatística de Durbin-Watson. Esta estatística mede a correlação entre cada resíduo para o período de tempo imediatamente antecedente àquele de interesse. A estatística de Durbin-Watson foi definida no Capítulo 2 (item 2.2).

Deste modo, para os dados coletados a respeito da empresa PRT, com uma variável independente e 75 observações, obteve-se a probabilidade exata do teste $p = 0,7751$, pode-se concluir que não existe autocorrelação entre os resíduos.

A análise de regressão pressupõe, também, que os erros seguem uma distribuição Normal. A condição de normalidade dos resíduos não é necessária para a obtenção dos estimadores de mínimos quadrados, mas é fundamental para a definição de intervalos de confiança e testes de

significância. Ou seja, na falta de normalidade, os estimadores são não-tendenciosos, mas os testes não têm validade, principalmente, em amostras pequenas. Entretanto, pequenas fugas da normalidade não causam grandes problemas.

O teste mais simples e rápido é o teste gráfico que compara a frequência acumulada dos resíduos padronizados com a curva normal, conforme mostra a Figura 21. A suposição de normalidade será válida, se os pontos do gráfico estiverem localizados, aproximadamente, ao longo de uma linha reta. Na visualização da linha reta, devem ser enfatizados os valores centrais do gráfico e não os extremos.

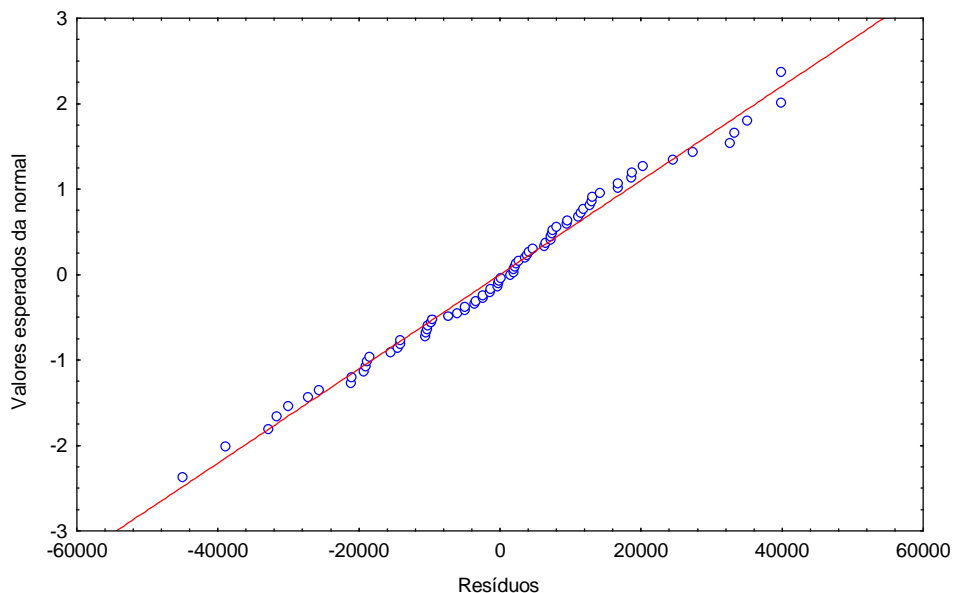


FIGURA 21 – Gráfico de probabilidade normal para os resíduos dos dados estudados

Usualmente, a visualização da aproximação dos pontos por uma linha reta é subjetiva e, então, para completar a avaliação da normalidade, pode ser realizado o teste de normalidade, baseado na estatística de Lilliefors, visto no Capítulo 2 (item 2.2).

Pela visualização do gráfico da Figura 21, percebeu-se que os pontos centrais estavam localizados, de forma aproximada, ao longo de uma reta. Assim, para confirmar esta indicação, para $n = 75$ observações, encontrou-se a probabilidade exata do teste $p > 0,20$, não obtendo-se, desta forma, indicações de violação de normalidade.

A avaliação da validade da suposição de normalidade foi a última etapa da análise de resíduos para os dados da empresa. Nesta análise, não foram encontradas indicações de falta da adequação do modelo de regressão e das suposições a ele associadas.

4.7 Estabelecendo o gráfico de controle de regressão

Como o modelo de regressão adapta-se bem aos dados coletados e nenhuma das pressuposições foram violadas, se estabelece este gráfico de controle de regressão, na empresa PRT, nos meses de julho, agosto e setembro.

Para o gráfico de controle de regressão, a linha central é a linha de regressão $\hat{y} = 124,82 + 0,0063x$. Usando esta linha de regressão e duas vezes o erro padrão, estimado em 123,01 (isto é, $\pm 2S$), definido em 3.1, o gráfico de controle, com limites colocados a 95% de confiança, foi estabelecido como mostrado na Figura 22.

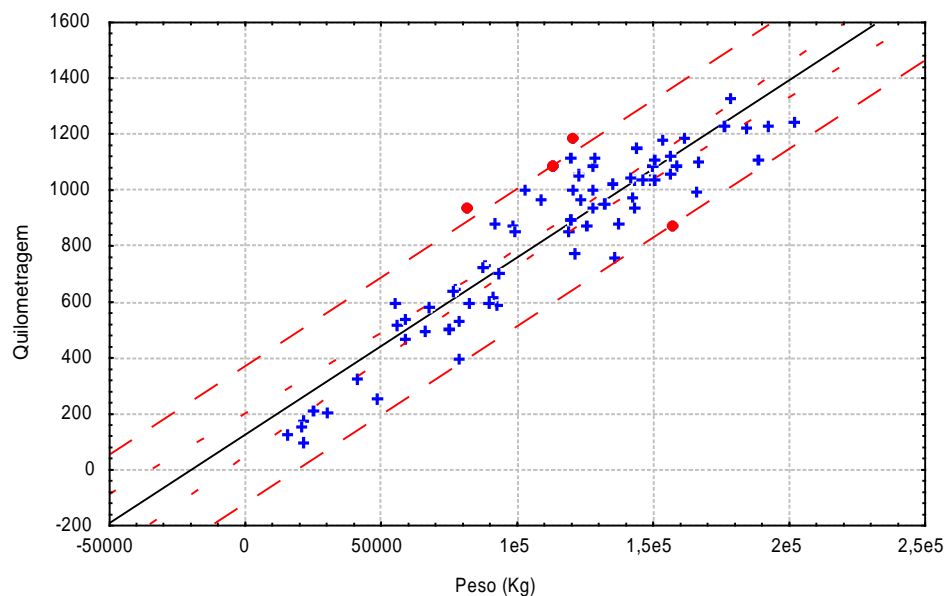


FIGURA 22 - Gráfico de controle de regressão para as variáveis peso e quilometragem, nos meses de julho, agosto e setembro.

Observa-se que quatro pontos são detectados fora dos limites de controle, dois pontos no mês de agosto e dois pontos no mês de setembro. Os três pontos acima do limite superior de controle indicam que houve uma quilometragem muito alta para o volume de resíduos coletados nestes dias e o ponto abaixo do limite inferior de controle sugere que houve uma melhora na produtividade do sistema, ou seja, para o volume de resíduos coletados a quilometragem percorrida foi baixa.

Mas, para se estabelecer os limites inferiores e superiores, para o gráfico de controle de regressão, é necessário que se exclua os pontos fora de controle, desde que estes pontos sejam devidos a causas assinaláveis.

Como, para os dados coletados não se estabeleceu uma causa comum no sistema para os problemas ocorridos, estes pontos foram retirados e um novo gráfico foi construído, conforme é mostrado na Figura 23.

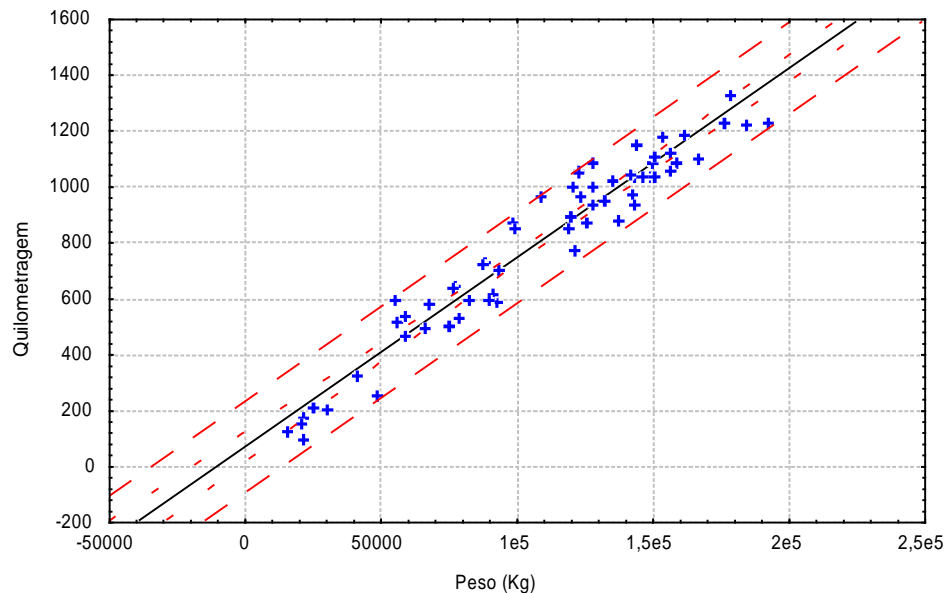


FIGURA 23 - Gráfico de controle de regressão para as variáveis peso e quilometragem, nos meses de julho, agosto e setembro, sem os pontos fora de controle.

Como um novo gráfico foi construído, com os dados sem os pontos que estavam fora dos limites de controle, uma nova equação de regressão foi estabelecida, para representar o processo sob controle, que será utilizada para monitorar etapas futuras do processo, como sendo:

$$\hat{y} = 71,9 + 0,0068x \quad (4.3)$$

(2,69) (29,83)

Onde, entre parênteses, estão os valores da estatística “t” calculada para as estimativas dos parâmetros.

Logo, o gráfico de controle de regressão, para as variáveis peso e quilometragem é determinado a partir da linha central $\hat{y} = 71,9 + 0,0068x$. Utilizando esta linha de regressão e duas vezes o erro padrão estimado em

81,91(isto é, $\pm 2S$), definido em 3.1, o gráfico de controle, com limites colocados a 95% de confiança, foi novamente estabelecido, como mostrado na Figura 24.

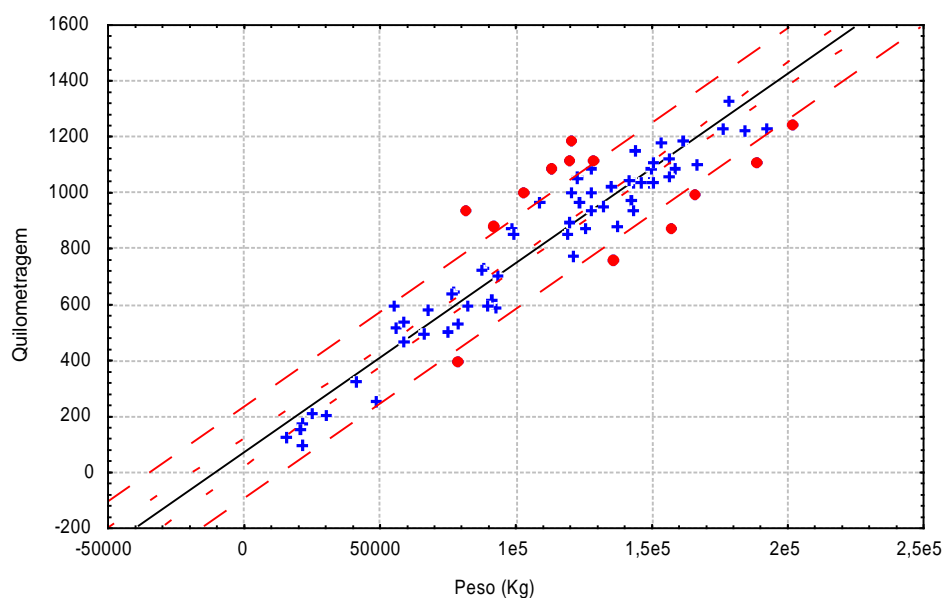


FIGURA 24 - Gráfico de controle de regressão para as variáveis peso e quilometragem, nos meses de julho, agosto e setembro, com os limites padrão para o processo.

O gráfico de controle de regressão detectou 13 pontos fora de controle, sendo que destes, 07 no mês de setembro, dos quais 05 estão acima do limite superior de controle. Isto significa que o caminhão percorreu uma quilometragem maior do que deveria, indicando que houve um deslocamento extra ao aterro sanitário, devido a um excesso de resíduos. Este problema seria resolvido se o caminhão compactador fosse substituído por outro, de maior capacidade. Neste caso, a empresa está perdendo em tempo, pois o caminhão faz mais viagens ao aterro com capacidade incompleta, sem levar em consideração os custos com

combustível, o desgaste com pneus e, muitas vezes, horas extras pagas aos funcionários.

Os pontos abaixo do limite inferior de controle, que significam um aumento na produtividade, mostram que uma grande quantidade de resíduos foi coletada, utilizando quilometragem menor do que a habitual.

O que poderia ter ocasionado tais pontos é o fato de, muitas vezes, alguns caminhões saírem ao socorro de outros que estão realizando a coleta diária, para que o recolhimento seja mais rápido. Porém, como a carga não é completada, o veículo retorna à empresa com carga, saindo no dia seguinte para o recolhimento normal sem sua total capacidade.

4.8 Comentários gerais do capítulo

Neste capítulo, desenvolveu-se o gráfico de controle de regressão, exemplificando-se a metodologia proposta no capítulo 3, através de dados reais, coletados numa empresa de prestação de serviços.

Mostra-se que as variáveis volume de líquido coletado e quilometragem percorrida pelos caminhões compactadores são altamente correlacionadas e que a análise de regressão pode ser empregada neste estudo.

Observa-se que o sistema encontra-se fora de controle, necessitando de um acompanhamento mais detalhado, para a verificação do que está causando a instabilidade do processo.

Alguns fatores observados na realização desta pesquisa contribuíram para que o gráfico de controle de regressão detectasse pontos fora de

controle. Estes fatores são coletas de dados incorretas, problemas com a balança no aterro e odômetro dos caminhões quebrados.

A coleta de dados incorreta significa falta de anotações nos formulários da quilometragem de saída ou chegada do caminhão, data da coleta ou peso líquido recolhido. Isto provoca um prejuízo para a empresa, pois o pagamento dos serviços prestados é feito baseado na planilha de coletas de dados.

Logo, sugere-se um treinamento para os motoristas a fim de diminuir estes erros e aumentar a lucratividade da empresa.

5 CONCLUSÃO

Na análise descritiva das variáveis estudadas, foi possível perceber que o volume de resíduos produzidos no mês de julho foi menor em relação aos meses de agosto e setembro.

A análise de correlação entre as variáveis mostrou que havia um alto grau de associação entre elas e que uma análise de regressão linear poderia ser conduzida. Então, a análise de regressão foi estabelecida e seus pressupostos foram analisados para se saber se o modelo era adequado para os propósitos aos quais se destina. Como nenhum dos pressupostos da análise de regressão foi violado, o modelo de regressão ajustado foi considerado adequado.

Com o gráfico de controle de regressão, foi possível observar que o sistema de coletas de resíduos não estava sendo realizado de forma satisfatória, pois alguns pontos estavam fora dos limites de controle. Isto indicava que havia dias em que a quilometragem percorrida pelos caminhões estava muito alta para o volume de resíduos recolhidos, ou seja, havendo uma baixa produtividade. Assim como, havia dias em que a quilometragem estava muito baixa para o volume de resíduos coletados, sugerindo, desta forma, um aumento de produtividade ou erro nos registros.

Sendo assim, sugere-se um acompanhamento maior pela empresa nas variáveis estudadas com o intuito de que, estando estas variáveis sob controle, a empresa seja capaz de prever com maior exatidão o gasto de combustível dos caminhões compactadores.

Deve-se ressaltar ainda que, durante o período analisado, alguns fatos ocorridos podem ter influenciado o comportamento do gráfico, como por exemplo: velocímetro estragado, balança quebrada, falta de anotações nos

formulários do volume coletado, da quilometragem de saída ou de chegada do caminhão.

É importante lembrar que, embora alguns pontos tenham ficado fora dos limites de controle no gráfico de regressão, isto não significa que o recolhimento dos resíduos não tenha sido realizado, pois a empresa não foi notificada nem pela população, nem pela Prefeitura Municipal de Santa Maria que é a contratante do serviço.

Salienta-se que a técnica foi adequada e trouxe uma melhoria para a empresa, pois identificou os pontos que estavam fora de controle, além disso, transformações internas ocorreram, pois houve uma maior organização do banco de dados e contratação de funcionários para atenderem no monitoramento deste setor.

Deixa-se, como sugestão para estudos futuros, o estabelecimento dos limites de controle para um modelo de regressão linear múltiplo, pois muitos sistemas produtivos não são explicados apenas por uma variável, ou, muitas vezes, é necessário o emprego de variáveis defasadas ou variáveis “dummy”.

Sugere-se, também, o uso das variáveis quilometragem percorrida pelos caminhões, volume líquido recolhido e tempo gasto no recolhimento para prever o consumo de combustível, levando em consideração períodos como férias escolares, eventos turísticos e feiras que se realizam no município.

Finalizando este estudo, sugere-se que o sistema seja implementado em uma planilha eletrônica para facilitar o uso da metodologia pelas empresas.

6 BIBLIOGRAFIA

ANDERSON, D. R., SWEENEY, D. J., WILLIAMS, T. A. **Estatística aplicada à administração e economia**. São Paulo: Pioneira Thomson Learning, 2002. 642 p.

CHARNET, R. *et al.* **Análise de modelos de regressão linear com aplicações**. São Paulo: Ed. UNICAMP, 1999. 356 p.

CAMPOS, H. de. **Estatística experimental não paramétrica**. 4. ed. Piracicaba: ESALQ, 1983. 349 p.

DEMÉTRIO, C. G. B. **Transformação de dados: efeitos sobre a análise da variância**. Piracicaba, 1978. 112 f. Dissertação (Mestrado em Experimentação e Estatística) - Escola Superior de Agricultura "Luiz de Queiroz", 1978.

DIPAOLA, P. P. Use of correlation in quality control. **Industrial Quality Control**, v. 2, n. 1, July, p. 10-14, 1945.

DOWNING, D. & CLARK, J. **Estatística aplicada: um modo fácil de dominar os conceitos básicos**. São Paulo: Saraiva, 2000. 455 p.

DRAPER, N. R. & SMITH, H. **Applied regression analysis**, New York: John Wiley, 1966. p. 1 - 43.

EPPRECHT, E. K., NINIO, A. L., SOUZA, M. O. de. Projeto ótimo de gráficos de médias móveis ponderadas exponencialmente (EWMA) para controle estatístico de processo, **Pesquisa Operacional**, v. 18, n. 2, p. 109 - 130, 1998.

FONSECA, J. S. da, MARTINS, G. de A., TOLEDO, G. L. **Estatística aplicada**. São Paulo: Atlas, 1985. 267 p.

GUJARATI, D. N. **Econometria Básica**. 3. ed. São Paulo: Makron Books, 2000. 846 p.

HILL, C. ; GRIFFITHS, W. ; JUDGE, G. **Econometria**. São Paulo: Saraiva, 1999. 408 p.

HOFFMANN, R. & VIEIRA, S. **Análise de regressão**: uma introdução à econometria. São Paulo: HUCITEC-EDUSP, 1977. 339 p.

INTRILIGATOR, M. D., BODKIN, R. G., HSIAO, C. **Econometric models, techniques, and applications**. Upper Saddle River: Prentice Hall, 1996. 654 p.

JACKSON, J. E.; Quality control methods for two related variables, **Industrial Quality Control**, v. 12, n. 7, Jan., p. 4-8, 1956.

KENETT, R. & ZACKS, S. **Modern industrial statistics**: design and control of quality and reliability. Belmont: Duxbury Press, 1998. 621 p.

KUME, H. **Métodos estatísticos para melhoria da qualidade**. São Paulo: Gente, 1993. 245 p.

LEVINE, D. M. ; BERENSON, M. L. ; STEPHAN, D. **Estatística**: teoria e aplicações. Rio de Janeiro: Livros Técnicos e Científicos, 2000. 811 p.

LUCAS, J. M. The design and use of V-mask control schemes, **Journal of Quality Technology**, v. 8, n. 1, p. 1 - 12, 1976.

LUCAS, J. M. & SACCUCCI, M. S. Exponentially weighted moving average control schemes: properties and enhancements, **Technometrics**, v. 32, n. 1, p. 1 - 29, 1990.

MADDALA, G. S., **Introduction to econometrics**. 2nd ed. Englewood Cliffs: Prentice Hall, 1992. 631 p.

MANDEL, B. J., Statistical programs of the United States post office department, **Industrial Quality Control**, p. 535 - 538, 1967.

_____ The regression control chart, **Journal of Quality Technology**, v. 1, n. 1, p. 1 - 9, 1969.

MONTGOMERY, D. C. **Introduction to statistical quality control**. 3rd ed. New York: John Wiley, 1997. 677 p.

MYERS, R. H. **Classical and moder regression with applications**. 2nd ed. Belmont: Duxbury Press, 1989, cap. 2, p. 20-67.

NETER, J. *et al.* **Applied linear regression models**. 3rd ed. Irwin, 1996. 720 p.

OLIN, Bryan D., Regression Control Charts Revisited: Methodology and Case Studies, **42º Annual Fall Technical Conference**, Corning, New York, 1998.

PETENATE, A. J.; **Controle estatístico de processos**. São Paulo: ABEQ - Associação Brasileira de engenharia química, 19_.

RAMOS, A. W. **Controle estatístico de processo para pequenos lotes**. São Paulo: Edgard Blucher, 1995. 151 p.

SCHNEIDER, P. R. **Análise de regressão aplicada à engenharia florestal**. Santa Maria: CEPEF, 1998. 236 p.

SHEWHART, W. A., **Economic control of quality of quality of manufactured product**. New York: D. Van Nostrand, 1931.

SIQUEIRA, L. G. P., **Controle estatístico do processo**, São Paulo: Pioneira, 1997. 130 p.

SOUZA, A. M. **Monitoração e ajuste de realimentação em processos produtivos multivariados**. Florianópolis, 2000. 166 f. Tese (Doutorado em Engenharia de Produção) - Universidade Federal de Santa Catarina, 2000.

VANNI, S. M. **Modelos de regressão: estatística aplicada**. São Paulo: Legnar, 1998. 177 p.

VARGAS, V. do C. C. de **Estudo comparativo do desempenho das cartas de controle CUSUM e EWMA**. Santa Maria, 2001. 156 p. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal de Santa Maria, 2001.

VASCONCELLOS, M. A. S. & ALVES, D. **Manual de econometria**. São Paulo: Atlas, 2000. 308 p.

VIEIRA, S. **Estatística para a qualidade**. Rio de Janeiro: Campus, 1999. 198 p.

WALLIS, W. A. & ROBERTS, H. V. **Statistics: a new Approach**. Chicago: The Free Press, 1956. p. 549-553.

WERKEMA, M. C. C. **Ferramentas da qualidade: ferramentas estatísticas básicas para o gerenciamento de processos**. Belo Horizonte: QFCO, v.3, 1995. 384 p.

WERKEMA, M. C. C. & AGUIAR, S. **Ferramentas da qualidade: análise de regressão: como entender o relacionamento entre as variáveis de um processo**. Belo Horizonte: QFCO, v.7, 1996. 311 p.

ZANINI, R. R.. **Análise dos dados da qualidade da água em Santa Maria - RS através de gráficos de controle**. Santa Maria, 1999. 129 f. Dissertação (Mestrado em Métodos Quantitativos) - Universidade Federal de Santa Maria, 1999.

ANEXOS

ANEXO A - Formulário preenchido pelos motoristas.

PRT - Prestação de Serviços Ltda.

Ficha de Controle

Cidade Placa

Data Nº de Cargas

Rota Nº Descrição resumida

Nome do Motorista

Coletores

H Saída: KM	H Chegada: KM
H Início: KM	H Final: KM
H Início: KM	H Final: KM

ÍTEMS QUE DEVEM SER VERIFICADOS

A) Testar sistema elétrico (Lâmpadas, faróis, etc..)

Situação:

B) Sistema de compactação

Situação:

C) Nível de combustível

Situação:

D) Instrumentos do painel

Situação:

E) Verificar a drenagem do reservatório de ar

Situação:

F) Verificar algum vazamento

Situação:

G) Pressão dos pneus

Situação:

H) Feixe de molas

Situação:

I) Limpeza do veículo e equipamentos necessários (vassoura, pá etc..)

Situação:

VISTO MOTORISTA

VISTO ENCARREGADO

ANEXO B - Limites superiores para o teste de Lilliefors.

n	α				
	0,20	0,15	0,10	0,05	0,01
4	0,300	0,319	0,352	0,381	0,417
5	0,285	0,299	0,315	0,337	0,405
6	0,265	0,277	0,294	0,319	0,364
7	0,247	0,258	0,276	0,300	0,348
8	0,233	0,244	0,261	0,285	0,331
9	0,223	0,233	0,249	0,271	0,311
10	0,215	0,224	0,239	0,258	0,294
11	0,206	0,217	0,230	0,249	0,284
12	0,199	0,212	0,223	0,242	0,275
13	0,190	0,202	0,214	0,234	0,268
14	0,183	0,194	0,207	0,227	0,261
15	0,177	0,187	0,201	0,220	0,257
16	0,173	0,182	0,195	0,213	0,250
17	0,169	0,177	0,189	0,206	0,245
18	0,166	0,173	0,184	0,200	0,239
19	0,163	0,169	0,179	0,195	0,235
20	0,160	0,166	0,174	0,190	0,231
25	0,142	0,147	0,158	0,173	0,200
30	0,131	0,136	0,144	0,161	0,187
> 30	$\frac{0,738}{\sqrt{n}}$	$\frac{0,768}{\sqrt{n}}$	$\frac{0,805}{\sqrt{n}}$	$\frac{0,886}{\sqrt{n}}$	$\frac{1,031}{\sqrt{n}}$

Tabela adaptada de: CAMPOS, H. de, **Estatística experimental não paramétrica**, 4 ed, Piracicaba, ESALQ, 1983.

ANEXO C - Valores críticos de d_l e d_u da estatística D_w de Durbin-Watson (os valores críticos unilaterais)*

$\alpha = 0,05$

n	P = 1		P = 2		P = 3		P = 4		P = 5	
	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

* n = número de observações; p = número de variáveis independentes

ANEXO C – Continuação

$\alpha = 0,01$										
	P - 1 = 1		P - 1 = 2		P - 1 = 3		P - 1 = 4		P - 1 = 5	
	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

Fonte : LEVINE, D. M. ; BERENSON, M. L. ; STEPHAN, D. **Estatística: Teoria e Aplicações**. Livros Técnicos e Científicos, Rio de Janeiro, 1998.