

# Gráfico de controle de regressão aplicado na monitoração de processos

**LUCIANE FLORES JACOBI, MSc.**

Professora do Departamento de Estatística – UFSM.

E-mail: [lfjacobi@ccne.ufsm.br](mailto:lfjacobi@ccne.ufsm.br)

**ADRIANO MENDONÇA SOUZA, DR.**

Professor do Departamento de Estatística – UFSM.

E-mail: [amsouza@ccne.ufsm.br](mailto:amsouza@ccne.ufsm.br)

**JOÃO EDUARDO DA SILVA PEREIRA, MSc.**

Professor do Departamento de Estatística – UFSM.

E-mail: [jesp@ccne.ufsm.br](mailto:jesp@ccne.ufsm.br)

## Resumo

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

## Palavras-chave

Gráfico de controle de regressão, Variáveis correlacionadas, Controle estatístico do processo.

# Regression control chart applied in process monitoring

## Abstract

*The main purpose of this research is to apply the regression control chart as tool of statistical control to monitor productive processes, where a state variable that is of interest can be expressed as function of a control variable. Several studies exist to control variables in productive processes, but most of time they are separately in relation to the control of each variable, and however not could be used for a comparative study. This research, therefore, it presents an efficient technique to control simultaneous by correlated variables.*

## Key words

*Regression control chart, Correlated variables, Statistical process control.*

## INTRODUÇÃO

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

Devido a esse fato, 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, podem-se reduzir os desperdícios e utilizar sua plena capacidade. Muitos dos sistemas utilizados nas empresas necessitam de processos 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, necessitou-se substituir a inspeção 100% por uma inspeção por amostragem, surgindo, assim, os primeiros estudos sobre o controle das características de um processo produtivo. Esses 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 que o controle individual dessas variáveis não é o mais indicado. Nesse caso, a qualidade depende do efeito comum dessas 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 dessas variáveis. Por esse motivo, esse 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.

Neste estudo, 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 de como o gráfico de controle de regressão pode ser construído a partir de dados históricos e, utilizando-se desse gráfico, é verificado se as novas observações do processo estão dentro dos níveis estabelecidos.

**G** ráficos de controle convencionais, como X-barra e proporção, não são capazes de fazer uma análise quando se tem variáveis correlacionadas.

Após o estudo detalhado das teorias utilizadas neste trabalho, buscou-se exemplificar essa metodologia, através de uma aplicação, utilizando-se dados coletados na empresa PRT, no setor de Engenharia de Saneamento e Meio Ambiente. Os dados ocupados, para a aplicação da metodologia, foram coletados nas fichas preenchidas pelos motoristas dos caminhões compactadores. Com a aquisição dessas 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)*.

## GRÁFICO DE CONTROLE DE REGRESSÃO

Para controlar a qualidade de um produto, Montgomery (1997) diz que é necessária a identificação e a medição das variações ocorridas no processo de produção por meio dos gráficos de controle, que são ferramentas úteis para a avaliação do estado de controle estatístico do processo, pois servem para determinar 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).

### Construindo o gráfico de controle de regressão

Quando a equação de regressão for empregada, com o objetivo de controle, é preciso 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. Nesse caso, somente será necessário que as relações existentes no conjunto de dados original, usado para construir a equação de regressão, ainda sejam válidas (Werkema & Aguiar, 1996).

Supondo-se que se está interessado em fazer uma previsão para um valor futuro  $y_i$  – não observado – correspondente a  $X = x_i$ . Se se usar 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 se admite 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, a distribuição de probabilidade de  $\hat{y}$  é dada por:

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

tem-se que:

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

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

e, logo:

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

em que  $\sigma^2$  pode ser estimado por:

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

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:

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

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

Para o uso em gráficos de controle, sabe-se que a média do processo pode variar em múltiplos do desvio padrão, isto é,  $\bar{x} \pm 3S$ ,  $\bar{x} \pm 2S$ , ..., formando os limites inferior e superior de controle. Logo, utilizando-se o valor de  $S$  dado em (1), e substituindo-se em (2) o valor de  $t$  tabelado, por um valor múltiplo do erro padrão, obtêm-se os limites superior e inferior de controle para o gráfico de controle de regressão:

$$\begin{aligned} \text{LSC}_{\hat{y}} &= \hat{y} + 3S \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{XX}}} \\ \text{LMC}_{\hat{y}} &= a + bx = \hat{y} \\ \text{LIC}_{\hat{y}} &= \hat{y} - 3S \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{XX}}} \end{aligned} \quad (3)$$

Apesar de os gráficos de controle de regressão serem conhecidos desde 1969, quando Mandel fez sua primeira aplicação, Olin (1998) comparou três métodos para a construção de limites de controles, que se basearam nos seguintes aspectos:

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

em que 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.

O autor conclui que, embora os intervalos de predição e confiança pareçam ser tecnicamente mais corretos de um ponto de vista estatístico, os limites simples  $3\sigma$  são recomendados, pois são de fácil compreensão, simples e similares ao gráfico de Shewhart.

Mandel (1969), na época, aplicou 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, deveriam 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 fornecem 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 que pode ser utilizada para se organizar o gráfico de controle de regressão.

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 essa 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 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 detectar pontos atípicos. Logo após, estabelecem-se os limites de controle a partir dos dados históricos, que estão sob controle. Esses limites estáveis são úteis para se sobrepor um novo conjunto de dados, dessa maneira, faz-se a comparação se o processo atual concorda com o anterior. Isto é, se está entre os limites inferior e superior de controle. Um sinal fora de controle indica uma inconsistência entre o processo novo e o antigo.

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-se que um ponto fora do limite inferior de controle, embora sendo o indicador de um bom desempenho, deve merecer uma atenção especial e ser investigado.

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 devida a uma causa constante do sistema.

### 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 esse propósito, é importante que, ao se fazer previsões, seja considerado o intervalo relevante da variável independente. Esse 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):

- 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.
- A reta de regressão é estimada utilizando-se dados passados. Essa reta não poderá prever dados futuros se a relação entre X e Y se modificar.
- 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.

- O simples fato de existir uma forte associação entre duas variáveis não significa que haja entre elas uma relação de

**O** gráfico de controle de regressão é capaz de avaliar o efeito conjunto de variáveis correlacionadas.

causa e efeito. Se chegarmos a uma reta de regressão, que se ajuste bem à relação entre X e Y, então haverá quatro possibilidades:

1. Os valores de Y podem realmente depender dos valores de X, como supusemos até agora.
2. A relação observada pode ser completamente casual.
3. Pode haver uma terceira variável afetando X e Y.
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.

### APLICAÇÃO DA METODOLOGIA

A degradação do meio ambiente tem sido alvo da preocupação de órgãos e instituições, sejam 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.

Sabe-se que lixo é qualquer resto que resulte de atividade humana, podendo ser de origem industrial, doméstica, hospitalar, comercial, agrícola, de serviços, entre muitas outras. Apresenta-se sob os estados: sólido, semi-sólido e líquido, podendo ser classificado em orgânico e inorgânico.

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

A reutilização do lixo, por meio da reciclagem, tornou-se um método eficiente para que não se tenha um acúmulo de resíduos e, assim, 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.

A estrutura dos órgãos públicos, em relação ao tratamento do lixo, é inadequada, desde a 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% contam 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,

**Q**uando a equação de regressão for empregada, com o objetivo de controle, é preciso que exista uma relação de causa e efeito.

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.

Nesse contexto, a Prefeitura Municipal de Santa Maria terceirizou esse serviço, contratando a Empresa PRT – Prestação de Serviços. A empresa encontra-se totalmente preparada para executar essa tarefa, mas também encontra dificuldades em dimensionar a frota de caminhões compactadores para realizar a coleta dos resíduos, tanto na zona urbana como na zona rural. Essa dificuldade não é encontrada apenas na cidade de Santa Maria, mas, também, em outras localidades em que a empresa atua. Dessa forma, pretende-se fornecer à empresa os períodos de maior, e menor, volume de resíduos a serem coletados, pois com a identificação desses períodos, e dos valores de volumes em quilogramas, a mesma poderá dimensionar a frota necessária para a realização dessa tarefa.

O processo de coleta é uma importante etapa no processo de tratamento de resíduos, pois é aí 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, sendo feita por caminhões compactadores que levam os resíduos ao aterro sanitário da cidade de Santa Maria – RS, localizado no Bairro Salgado Filho.

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 7.000 kg, cada, sendo, também, realizada 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 da 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, e aos domingos a coleta não é realizada.

O primeiro passo, junto à empresa, foi organizar os formulários preenchidos pelos motoristas e montar as planilhas que serviriam como banco de dados. 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. As variáveis do

processo, analisadas nesse trabalho, foram: a quilometragem diária percorrida pelos caminhões e o volume de resíduos recolhidos por dia.

A variável quilometragem é informada pelos próprios motoristas, que anotam o valor

inicial e o final da quilometragem do caminhão, tendo como ponto de referência a empresa. Para este procedimento, um formulário próprio é utilizado pela empresa.

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 na saída do aterro, encontrando-se, assim, o peso líquido de resíduos que foi depositado, servindo de base para o cálculo do pagamento mensal da empresa.

A distância percorrida, em cada setor, é sempre a mesma, pois os percursos das rotas são predeterminados. A variabilidade da quilometragem reside no fato de que, quando há aumento no volume de resíduos, os caminhões precisam ir mais vezes ao aterro sanitário, aumentando, dessa maneira, a quilometragem percorrida e, conseqüentemente, o gasto com combustível e o tempo de recolhimento. Isso poderia ser melhor dimensionado, se fosse possível identificar os períodos em que ocorresse excesso de resíduos. Dessa forma, um caminhão compactador, com maior capacidade, poderia ser destinado à rota onde este fato ocorresse, trazendo uma economia de combustível, mão-de-obra e tempo, beneficiando a empresa.

### **Análise descritiva dos dados**

Neste item, faz-se uma análise descritiva das variáveis envolvidas no estudo, nos meses de julho, agosto e setembro de 2001, pois, dessa maneira, será possível encontrar alguns pontos que mereçam uma maior atenção e que sejam passíveis de verificação. Nas Tabelas 1 e 2, apresentam-se as principais medidas descritivas das variáveis peso líquido e quilometragem, respectivamente.

Verificando-se a Tabela 1, percebe-se que no mês de julho o volume médio de resíduos coletados é inferior aos meses de agosto e setembro. Essa diferença gerou uma investigação ampla, no sentido de se estabelecer as possíveis causas desse declínio no volume de resíduos coletados, conforme mostrado no Quadro 2.

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 2.

Observa-se, na Tabela 2, 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. Essa mesma variação é observada para a variável peso, conforme mostra a Tabela 1.

As tabelas acima apresentadas 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 1.

**Tabela 1: Medidas descritivas para a variável peso.**

Meses	Medidas Descritivas					
	Valor de n <sup>**</sup>	Valor Mínimo	Valor Máximo	Média	Desvio Padrão	Coefficiente de Variação
Julho	26	15.610	135.285	64.019,65	31.182,41	0,4871
Agosto	27	87.583	201.535	145.528,6	28.714,89	0,1973
Setembro	25	58.330	165.765	123.641,4	31.054,80	0,2512

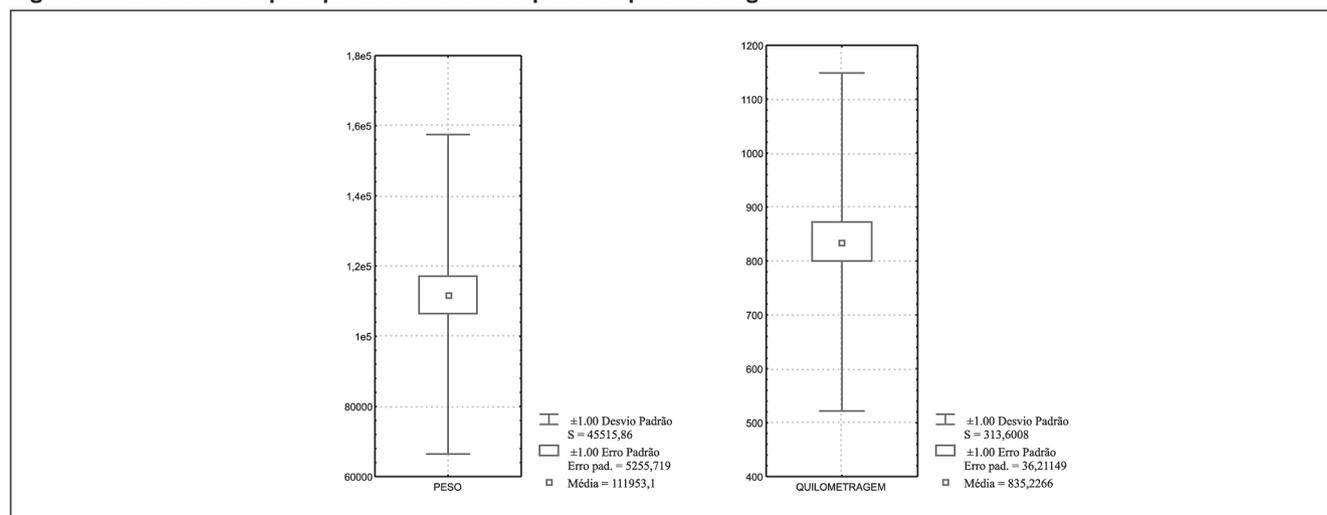
Notas: A variável é expressa em quilogramas (kg).  
n é o número de dias com recolhimento no mês.

**Tabela 2: Medidas descritivas para a variável quilometragem.**

Meses	Medidas Descritivas					
	Valor de n	Valor Mínimo	Valor Máximo	Média	Desvio Padrão	Coefficiente de Variação
Julho	26	99	1.903	542,3846	366,6775	0,6760
Agosto	27	732	1.555	1.096,4440	153,8572	0,1403
Setembro	25	494	1182	939,7200	168,7697	0,1796

Notas: A variável é expressa em quilômetros (km).  
n é o número de dias com recolhimento no mês.

**Figura 1: Gráfico Box-plot para as variáveis peso e quilometragem.**



Constata-se, pela Figura 1, que as variáveis peso e quilometragem possuem o mesmo comportamento. Como a mediana está à 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 2. Essa variável mede a quantidade total de resíduos recolhida, por dia, na cidade de Santa Maria, sendo expressa em quilogramas (kg).

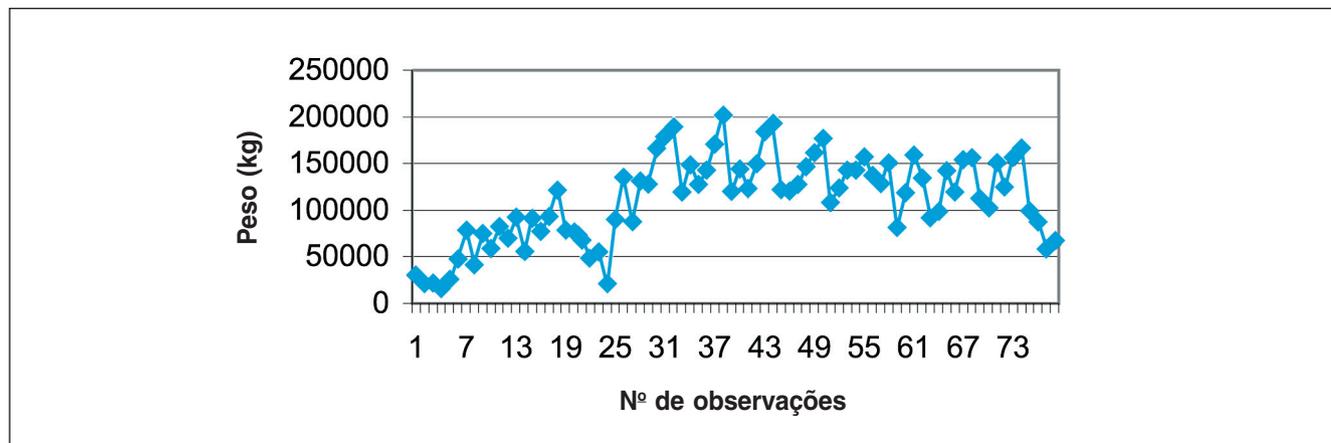
Essa variável foi analisada no período correspondente aos meses de julho, agosto e setembro de 2001. Observa-se que existe uma mudança de nível na média do processo, correspondendo ao mês de julho. Esse comportamento inspira uma investigação criteriosa, para tentar identificar as causas do ocorrido.

Um dos motivos para o baixo volume de resíduos coletados, é que o mês de julho corresponde ao mês das férias escolares. Além disso, foram identificadas muitas falhas nos relatórios, pois os motoristas que realizavam as coletas não preenchiam os formulários.

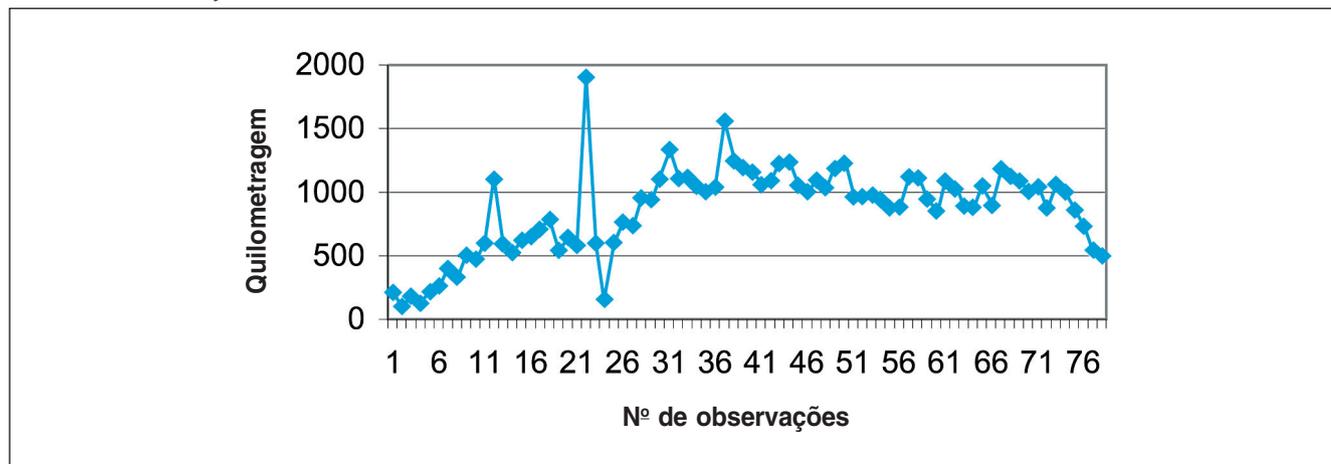
A quilometragem percorrida pelos caminhões é uma variável muito importante, pois dela depende o gasto com combustível. A quantidade total da quilometragem percorrida pelos caminhões, por dia, na cidade de Santa Maria, é mostrada no gráfico de seqüência na Figura 3.

Essa variável foi analisada no período de julho a setembro de 2001. Observa-se, na Figura 3, que houve uma mudança de nível na média do processo. Essa mudança ocorreu próxima à observação 30, que corresponde ao período final de julho e início de agosto. Observa-se, também, que a série apresenta quatro observações atípicas do restante da massa de dados, conhecida como *outlier*.

**Figura 2:** Gráfico representativo do peso de resíduos recolhidos por dia, em Santa Maria, de julho a setembro de 2001.



**Figura 3:** Gráfico representativo da quilometragem percorrida pelos caminhões compactadores, por dia, em Santa Maria, de julho a setembro de 2001.



### Análise de correlação nos dados

O interesse, neste trabalho, é construir um modelo de regressão que relacione a quilometragem diária, percorrida pelos caminhões compactadores (Y), à 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 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, um diagrama de dispersão foi construído a partir dos dados coletados e está apresentado na Figura 4. A configuração dos pontos, neste diagrama, sugere a existência de um relacionamento linear entre a quilometragem percorrida e a quantidade de resíduos coletados.

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. 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 4, revela três pontos que desviam de um padrão linear e, conforme Mandel (1969), esta situação é passível de uma investigação. Se esses pontos representarem uma condição não usu-

al, podem ser excluídos dos cálculos do gráfico de controle de regressão, pois, nessa etapa, está-se interessado em encontrar os limites de controle para o gráfico de regressão que revele uma situação de estabilidade do sistema. Dessa maneira, será possível uma comparação com etapas posteriores do processo de coleta dos resíduos. Assim, o processo estável servirá como padrão.

**Q**uando um modelo de regressão é escolhido, deve-se verificar se é adequado para os propósitos a que se destina.

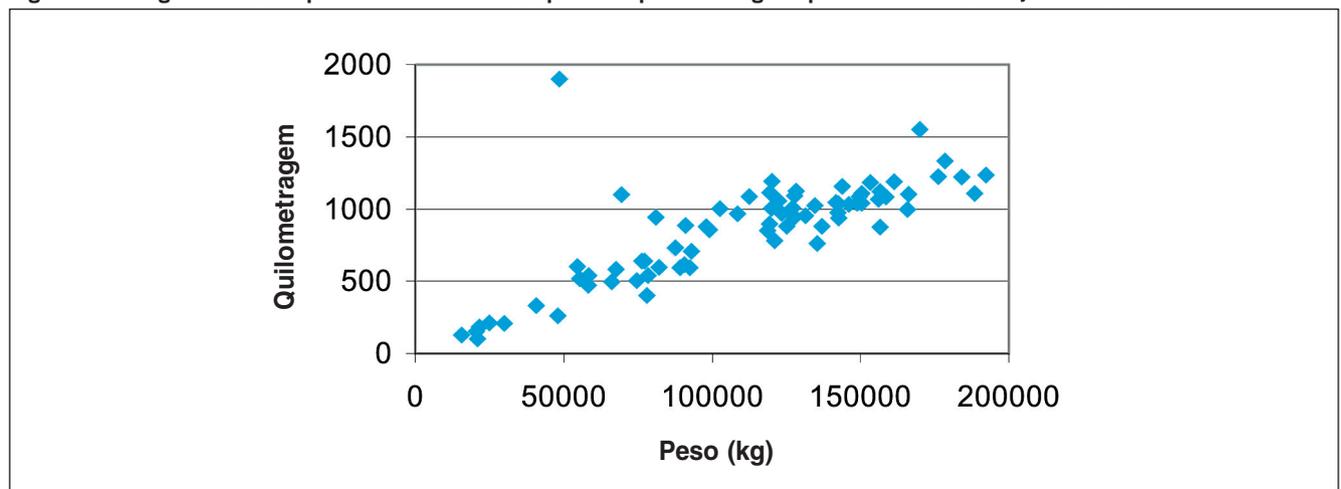
Investigando-se esses três pontos, verificou-se que a quilometragem rodada pelos caminhões estava muito alta para o volume de resíduos coletados nesses dias, não havendo, porém, nenhum motivo especial para isso, a não ser um possível engano nas anotações nas fichas de controle.

Como as causas dos três pontos que desviaram do padrão linear, mostrado na Figura 4, não são constantes no sistema, foram excluídos do cálculo do gráfico de controle de regressão. Dessa 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, construiu-se o diagrama de dispersão, mostrado na Figura 5, e calculou-se um novo coeficiente de correlação.

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.

Figura 4: Diagrama de dispersão das variáveis peso e quilometragem para os meses de julho a setembro de 2001.



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 essas variáveis.

**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 essas variáveis. A determinação dos parâmetros dessa reta é denominada ajustamento.

Estimaram-se os valores dos parâmetros, da equação de regressão linear simples, pelo método de mínimos quadrados, encontrando-se a seguinte equação:

$$\hat{y} = 124,82 + 0,0063x \quad (4)$$

(3, 29)    (20, 198)

em que, nos parênteses, estão os valores da estatística “t”, calculada para as estimativas dos parâmetros.

A equação (4) pode ser reescrita da seguinte forma:

$$\text{Quilometragem (estimada)} = 124,82 + 0,0063 (\text{peso}) \quad (5)$$

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 (VASCONCELLOS & ALVES, 2000; GUJARATI, 2000).

Para avaliar se a reta de regressão, ajustada em (5), pode ser empregada como base de futuras análises, foi realizada a análise de resíduos (Veja Neter, *et al.*, 1996). Os gráficos de resíduos que foram construídos são apresentados nas Figuras 6 a 8. A análise destes gráficos, e dos testes realizados, não indicou a presença de quaisquer inadequações do modelo de regressão.

O gráfico de resíduos contra os valores ajustados é apresentado na Figura 6. Neste gráfico, os resíduos estão situados, aproximadamente, em uma faixa horizontal centrada em  $e_i = 0$ .

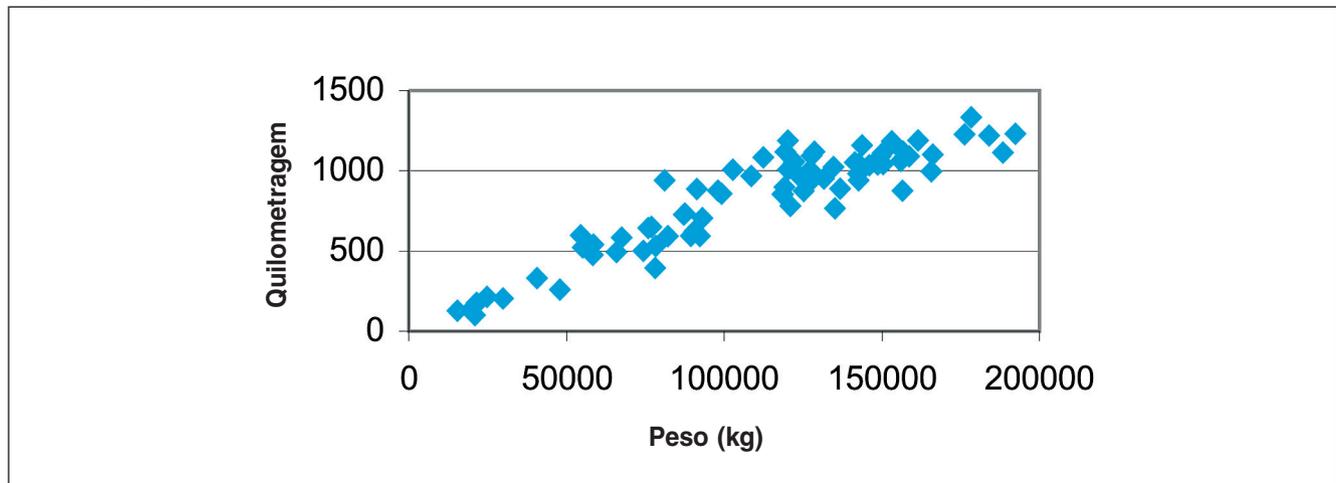
Analisando-se o gráfico da Figura 6 não são percebidas quaisquer indicações de que os pressupostos de linearidade e homocedasticidade tenham sido violados.

Outro pressuposto do modelo de regressão, que deve ser considerado, é a independência dos resíduos. Esse 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 (MADDALA, 1992; NETER, *et al.*, 1996).

Dado 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 7.

A partir da Figura 7, observa-se que os pontos não tendem a flutuar para cima e para baixo, em um padrão cíclico. Esse padrão indica que não há uma violação no pressuposto da independência dos resíduos. Além do gráfi-

**Figura 5: Diagrama de dispersão das variáveis peso e quilometragem, para os meses de julho a setembro, sem os pontos que desviam do padrão linear.**



co de resíduos, a autocorrelação também pode ser detectada e medida pela utilização da estatística de Durbin-Watson. Essa estatística mede a correlação entre cada resíduo para o período de tempo imediatamente antecedente àquele de interesse (VASCONCELLOS & ALVES, 2000).

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

A análise de regressão baseia-se, também, em 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 (NETER, 1996).

Figura 6: Gráfico de resíduos contra os valores ajustados, para os dados coletados.

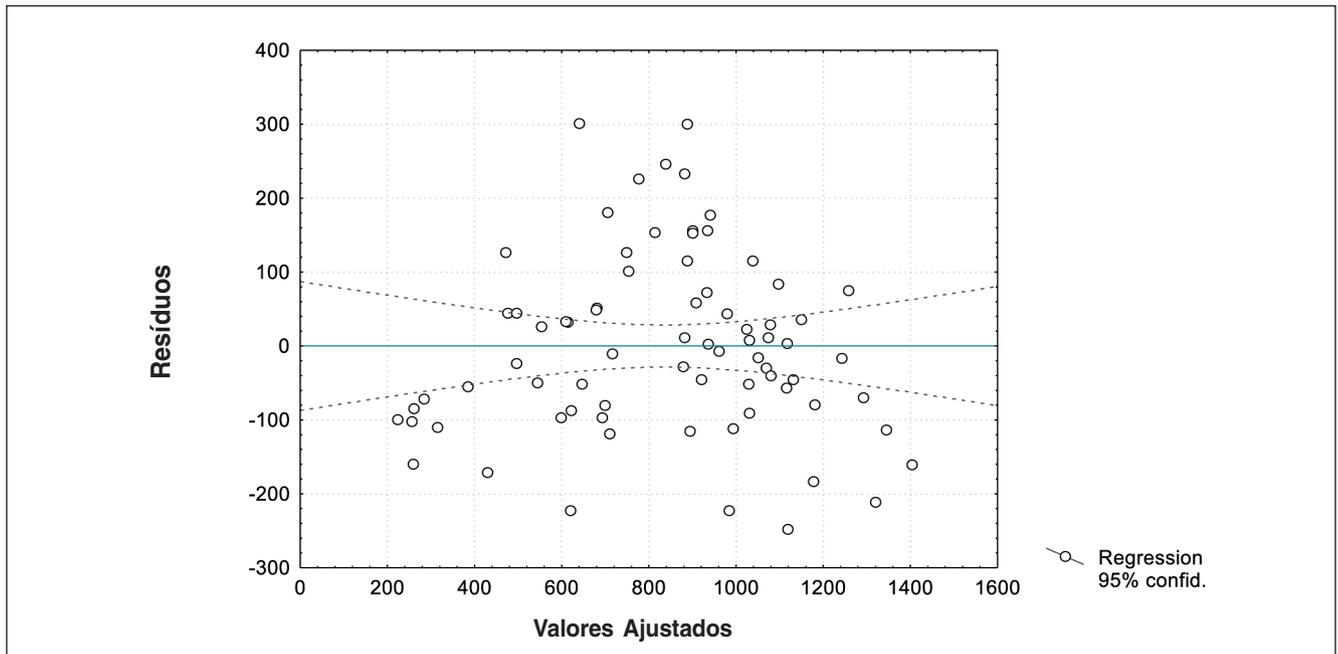
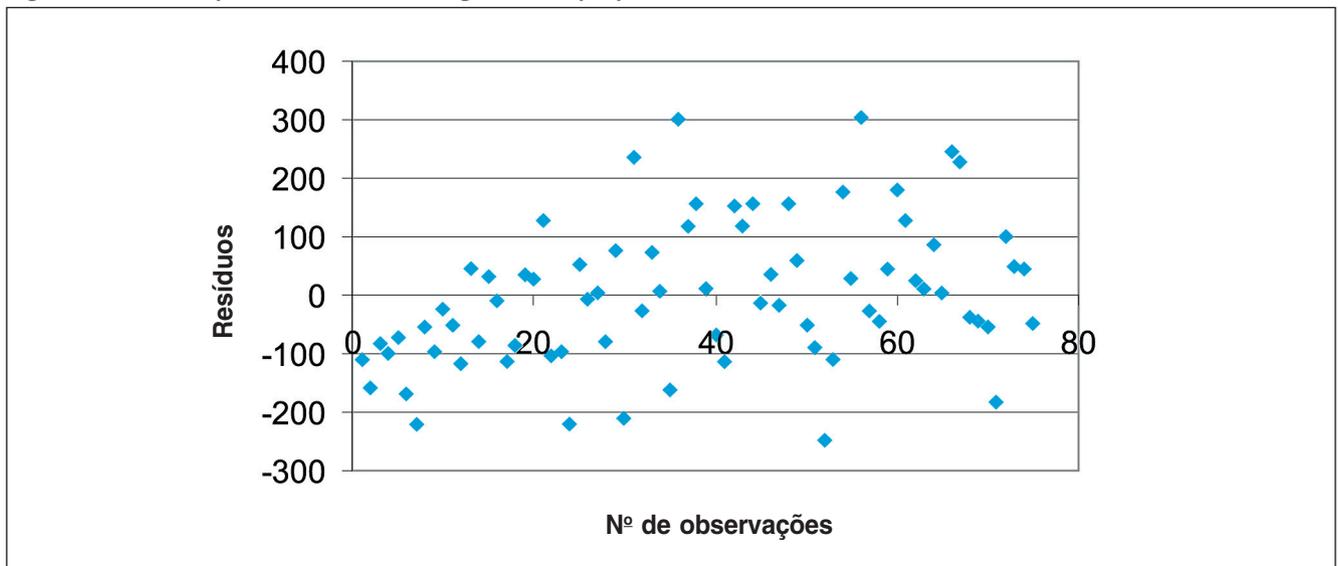


Figura 7: Gráfico para resíduos ao longo do tempo para os dados analisados.



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 8. 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.

Usualmente, a visualização da aproximação dos pontos, por uma linha reta, é subjetiva e, então, para completar a avaliação da normalidade, deve ser realizado o teste de normalidade, baseado na estatística de Lilliefors.

Pela visualização do gráfico da Figura 8, 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 se obtendo, dessa 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. Nessa análise, não foram encontradas indicações de falta da adequação do modelo de regressão e das suposições a ele associadas.

**Estabelecendo o gráfico de controle de regressão**

Como o modelo de regressão adapta-se bem aos dados coletados, e nenhuma das pressuposições foi violada, estabelece-se o gráfico de controle de regressão para os dados

coletados nos meses de julho, agosto e setembro de 2001.

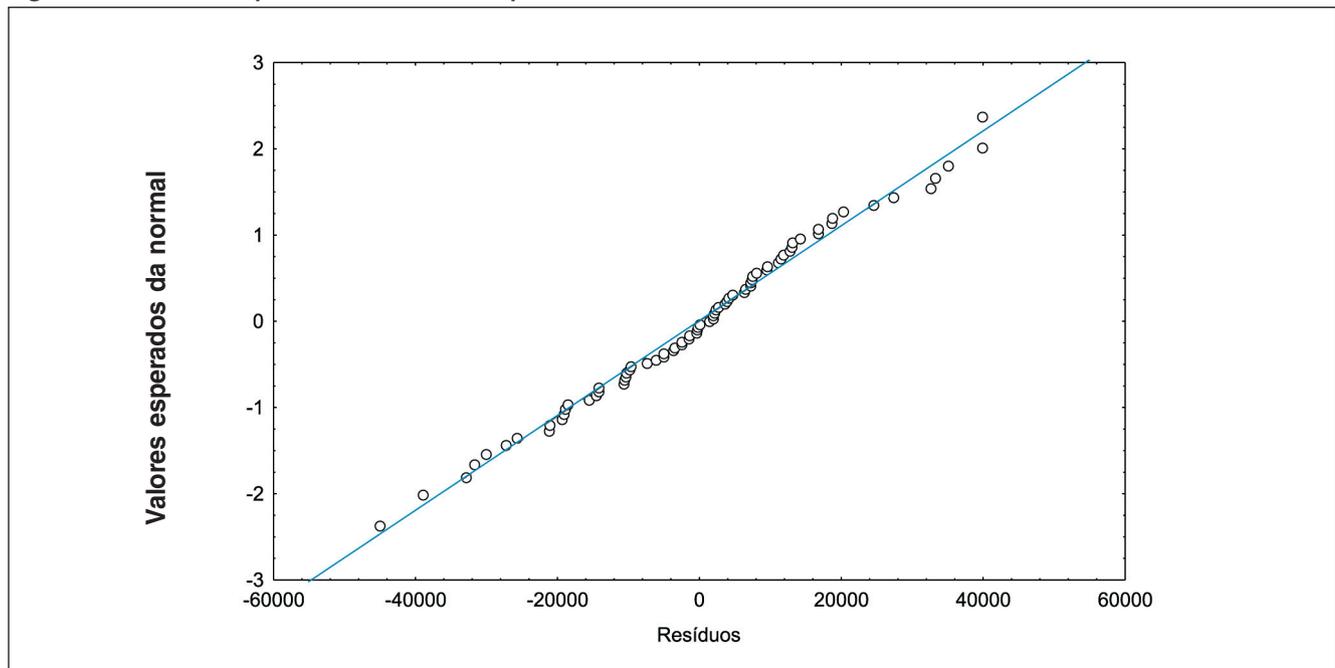
Para o gráfico de controle de regressão, a linha central é a linha de regressão  $\hat{y} = 124,82 + 0,0063x$ . Usando-se essa linha de regressão, e duas vezes o erro padrão estimado da reta de regressão, em 123,01 (isto é,  $\pm 2S$ ), definido em (1), o gráfico de controle, com limites de controle colocados a 95% de confiança, estabeleceu-se como mostrado na Figura 9.

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 nesses dias, e o ponto abaixo, limite inferior de controle, sugere que houve uma melhoria 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 de controle, para o gráfico de controle de regressão, é necessário que se excluam os pontos fora de controle, desde que esses pontos sejam devidos a causas assinaláveis. Como para os dados coletados não se estabeleceu uma causa comum no sistema para os problemas ocorridos, esses pontos foram retirados e um novo gráfico foi construído, conforme é mostrado na Figura 10.

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, a qual será utilizada para monitorar

**Figura 8: Gráfico de probabilidade normal para os resíduos dos dados estudados**



etapas futuras do processo, como sendo:

$$\hat{y} = 71,9 + 0,0068x \quad (6)$$

(2,69) (29,83)

Os valores da estatística “t” estão entre parênteses, 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-se essa linha de regressão, e duas vezes o erro padrão estimado da reta de

regressão, e duas vezes o erro padrão estimado da reta de

Figura 9: Gráfico de controle de regressão para as variáveis peso e quilometragem, nos meses de julho, agosto e setembro.

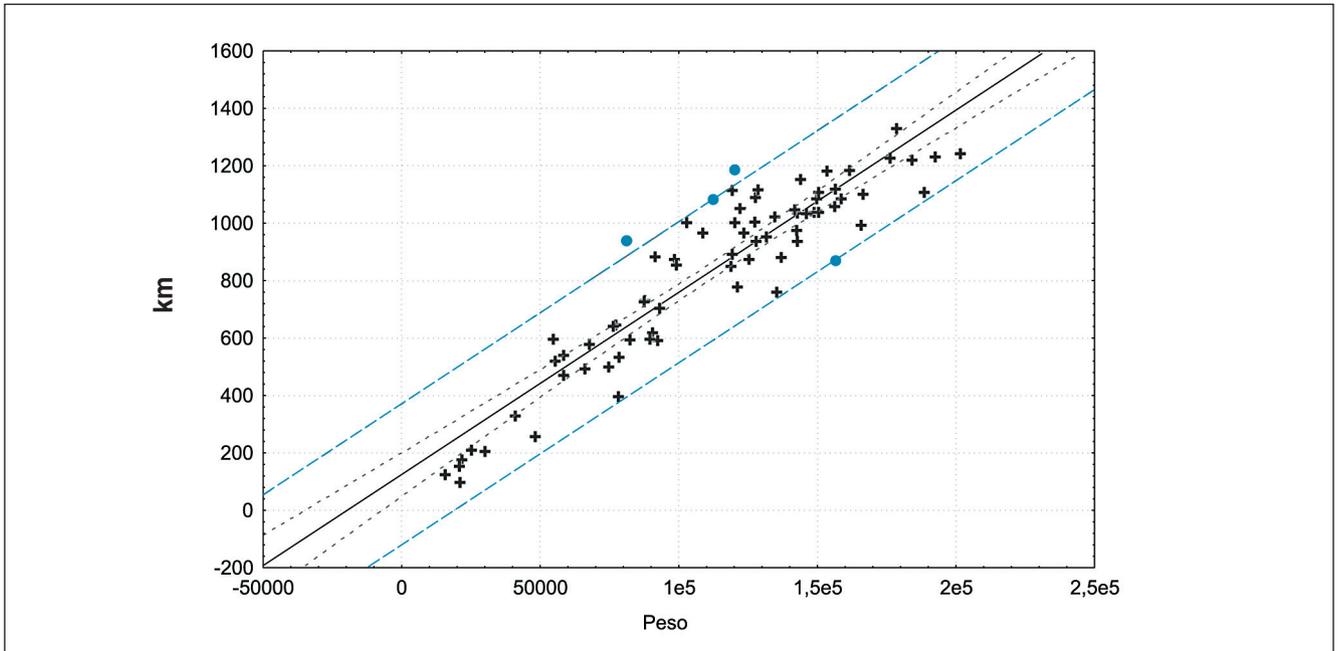
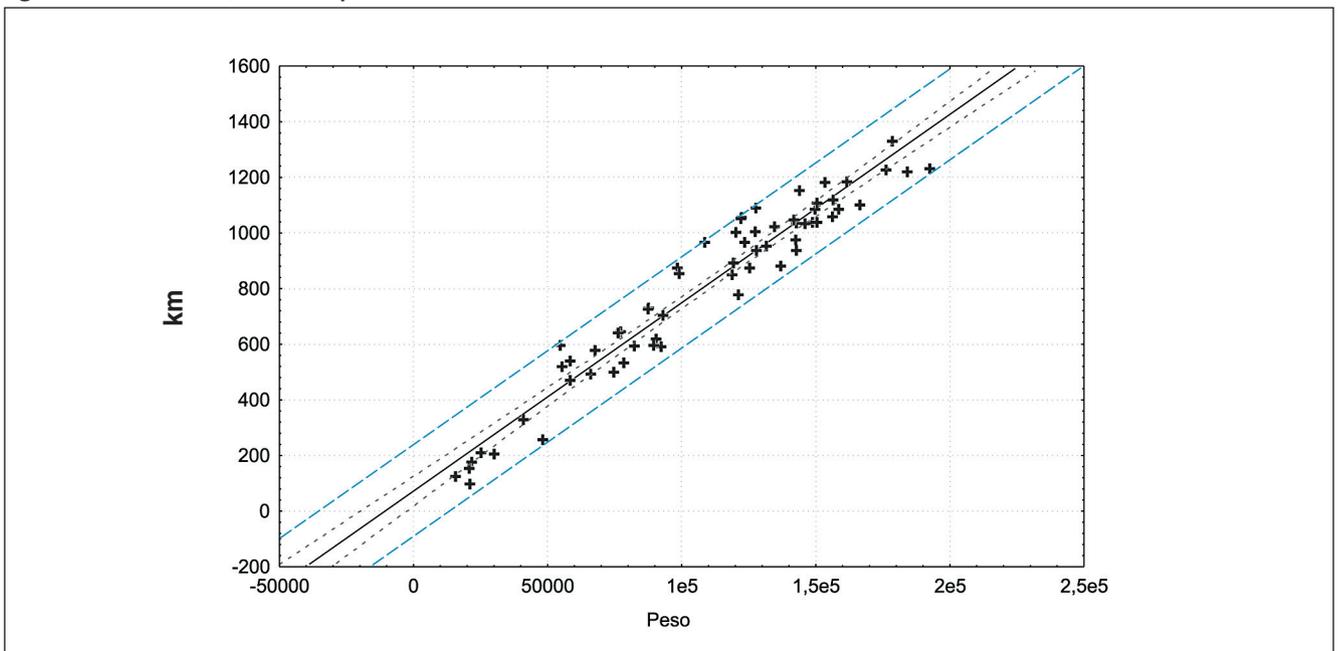


Figura 10: Gráfico de controle de regressão para as variáveis peso e quilometragem, nos meses de julho, agosto e setembro, sem os pontos fora de controle.



regressão em 81,91 (isto é,  $\pm 2S$ ), o gráfico de controle, com limites de controle colocados a 95% de confiança, foi novamente estabelecido, como mostrado na Figura 11.

O gráfico de controle de regressão detectou treze pontos fora de controle, sendo que, destes, sete no mês de setembro, dos quais 5 estão acima do limite superior de controle. Isto significa que o caminhão coletor percorreu uma quilometragem maior do que deveria, indicando que houve um deslocamento extra do caminhão ao aterro sanitário, devido a um excesso de resíduos. Este problema seria resolvido se o caminhão compactador fosse substituído por um caminhão de maior capacidade. Nesse caso, a empresa está perdendo em tempo, pois o caminhão faz mais viagens ao aterro

**A** técnica foi adequada e trouxe uma melhoria para a empresa, pois identificaram-se os pontos que estavam fora de controle.

sanitário com capacidade incompleta, sem levar em consideração custos com combustível, desgaste com pneus e, muitas vezes, horas extras para os funcionários.

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

O que poderia ter ocasionado tais pontos, é o fato de, muitas vezes, alguns caminhões saírem em socorro de outros que

estão realizando a coleta diária, para que o recolhimento seja mais rápido. Mas, como a carga não é completada, e uma vez que o caminhão coletor só poderá ir até o aterro sanitário se estiver com, no mínimo, a metade de sua capacidade, o veículo retorna à empresa com carga, saindo no dia seguinte para o recolhimento normal sem sua total capacidade de coleta.

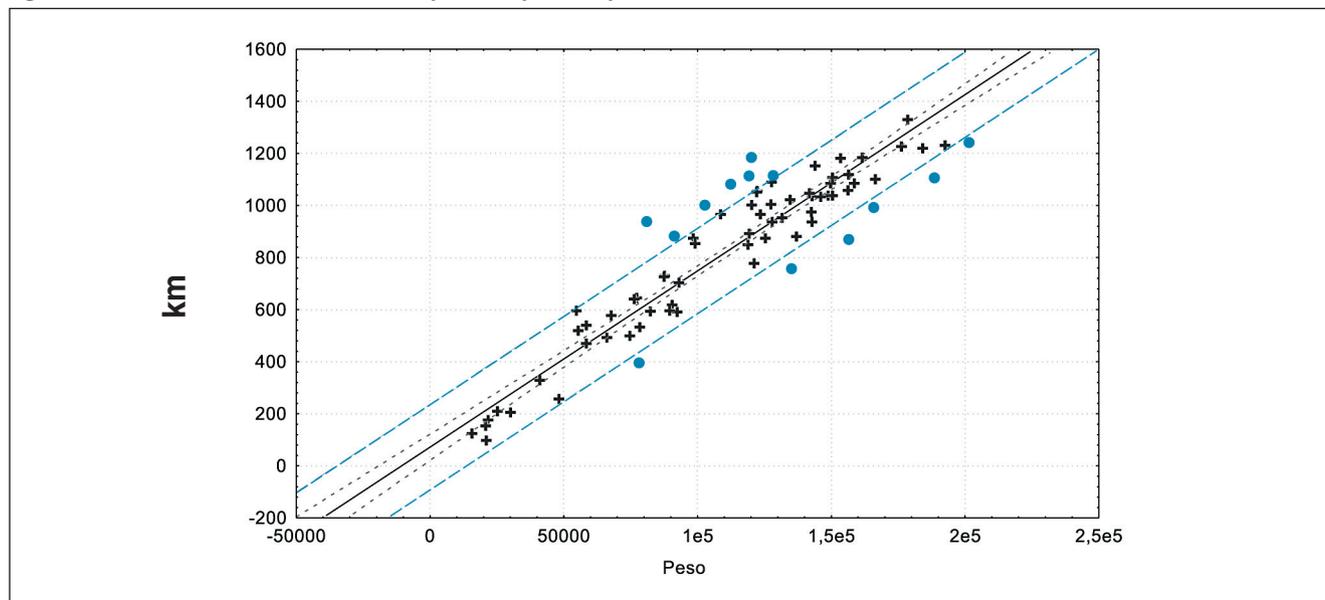
**CONCLUSÕES**

O gráfico de controle de regressão permitiu observar que o sistema de coleta de resíduos não estava sendo realizado de forma satisfatória, pois alguns pontos estavam fora dos limites de controle. Isso indica que havia dias em que a

quilometragem percorrida pelos caminhões estava muito alta para o volume de resíduos recolhidos, ou seja, havia uma baixa produtividade. Assim como havia dias em que a quilometragem estava muito baixa para o volume de resíduos

coletados, sugerindo, dessa forma, um aumento de produtividade ou erro nos registros. Vale ressaltar que a distância percorrida, em cada setor, é sempre a mesma, pois os percursos das rotas são predeterminados. A variabilidade da quilometragem reside no fato de que, quando há aumento no volume de resíduos, os caminhões precisam ir mais vezes ao aterro sanitário, aumentando, dessa maneira, a quilometragem percorrida e, conseqüentemente, o gasto com combustível e o tempo de recolhimento.

**Figura 11: 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.**



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.

Deve-se ressaltar, ainda, que durante o período analisado alguns fatos ocorridos podem ter influenciado o comportamento das variáveis, 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 estejam fora dos limites de controle no gráfico de regressão, isso não significa que o recolhimento dos resíduos não tenha sido realizado, pois a empresa não foi notificada nem pela popu-

lação, nem pela contratante do serviço.

Salienta-se que a técnica foi adequada e trouxe uma melhoria para a empresa, pois identificaram-se os pontos que estavam fora de controle. Além disso, transformações internas ocorreram na empresa, havendo uma melhor organização do banco de dados e contratação de novos funcionários para atenderem ao monitoramento deste setor.

Em consequência dessa medida, houve redução no gasto com combustível. Além disso, detectou-se, no período de férias, uma redução no volume de resíduos coletados, pois tendo a cidade como sua principal vocação a universitária, contando com mais de quatro universidades, é de se entender que no período de férias haja um menor fluxo de pessoas e, conseqüentemente, um menor acúmulo de resíduos.

## ■ Referências bibliográficas

- DIPAOLA, P.P. Use of correlation in quality control. *Industrial Quality Control*, v. 2, n. 1, p. 10-14, July, 1945.
- DRAPER, N.R.; SMITH, H. *Applied regression analysis*. New York: John Wiley, 1966. 407 p.
- GUJARATI, D.N. *Econometria básica*. 3. ed. São Paulo: Makron Books, 2000. 846 p.
- JACKSON, J.E.: Quality control methods for two related variables. *Industrial Quality Control*, v. 12, n. 7, p. 4-8, Jan. 1956.
- LEVINE, D.M., BERENSON, M.L., STEPHAN, D. *Estatística: teoria e aplicações*. Rio de Janeiro: Livros Técnicos e Científicos, 2000. 811 p.
- 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.
- 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. In: *Annual Fall Technical Conference, 42º*, New York, 1998. Proceedings, New York, 1998. P. 1-17.
- 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.
- VASCONCELLOS, M.A.S.; ALVES, D. (Coord.) *Manual de econometria*. São Paulo: Atlas, 2000. 308 p.
- WERKEMA, M.C.C. *Ferramentas da qualidade: ferramentas estatísticas básicas para o gerenciamento de processos*. Belo Horizonte: QFCO, 1995. v. 3.
- 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, 1996. v. 7.

## ■ Agradecimentos

Agradecemos à FAPERGS (Fundo de Amparo à Pesquisa do Estado do Rio Grande do Sul) pelo auxílio financeiro que viabilizou o desenvolvimento deste projeto (processo 01513175) e à empresa PRT – Prestação de serviço pela disponibilidade dos dados.