

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

**FERRAMENTAS PARA MODELAGEM E
MONITORAMENTO DE CARACTERÍSTICAS DE
QUALIDADE DO TIPO FRAÇÃO**

ÂNGELO MÁRCIO OLIVEIRA SANT'ANNA

Porto Alegre
2009

ÂNGELO MÁRCIO OLIVEIRA SANT'ANNA

**FERRAMENTAS PARA MODELAGEM E
MONITORAMENTO DE CARACTERÍSTICAS DE
QUALIDADE DO TIPO FRAÇÃO**

Tese submetida ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul como requisito parcial à obtenção do título de Doutor em Engenharia de Produção, modalidade Acadêmica, na área de concentração em Sistemas de Qualidade.

Orientadora: Carla Schwengber ten Caten, Dr^a.

Porto Alegre
2009

ÂNGELO MÁRCIO OLIVEIRA SANT'ANNA

**FERRAMENTAS PARA MODELAGEM E MONITORAMENTO DE
CARACTERÍSTICAS DE QUALIDADE DO TIPO FRAÇÃO**

Esta tese foi julgada adequada para a obtenção do título de Doutor em Engenharia de Produção na modalidade Acadêmica e aprovada em sua forma final pela Orientadora e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul.

Prof^a. Carla Schwengber ten Caten, Dr^a.

PPGEP / UFRGS

Orientadora

Prof^a. Carla Schwengber ten Caten, Dr^a.

Coordenadora PPGEP / UFRGS

Banca Examinadora:

Guilherme Luís Roehe Vaccaro, Dr. (PPGEP / UNISINOS)

João Riboldi, Dr. (DEST / UFRGS)

José Luis Duarte Ribeiro, Dr. (PPGEP / UFRGS)

*"An approximate answer to the right problem
is worth a good deal more than an exact
answer to an approximate problem."*

John Wilder Tukey

DEDICATÓRIA

Dedico este trabalho à minha pessoa, pelos obstáculos superados, pelas escolhas, pelo esforço, dedicação e entusiasmo. Grato!

AGRADECIMENTOS

Agradeço, com muito reconhecimento, a todos que participaram direta ou indiretamente da concretização deste ideal e conclusão deste trabalho. Em particular, sou grato

Primeiramente a Deus, por me permitir mais essa vitória.

À minha orientadora, Prof^a. Carla Schwengber ten Caten, que dotada de apurada sensibilidade no trato com seus semelhantes, tem possibilitado o desenvolvimento científico e pessoal daqueles que a cercam. Grato, pelo estímulo, confiança, paciência e valiosas orientações ao longo de quase seis anos de mestrado e doutorado.

Aos funcionários e professores do PPGEP – UFRGS, por suas contribuições. Em especial, aos Profs. Cláudio Müller e Márcia Echeveste, pela simpatia, amizade e incentivo.

Agradeço aos membros da banca examinadora pelas valiosas contribuições e críticas ao trabalho. Suas opiniões, fundamentadas em suas experiências e conhecimentos, fortaleceram a relevância desta tese.

Aos meus pais e familiares, que mesmo à distância, sempre me transmitiram apoio, incentivo, carinho e amor incondicional.

Aos meus irmãos, em especial a Valmar e a Malu, pelo amor, carinho e estímulo.

Ao grande amigo Danilo Marcondes Filho, pela amizade, cumplicidade e parceria.

Ao grande amigo Alex Ramos, pela amizade, carinho e parceria, mesmo que distante.

A todos os amigos do LOPP – UFRGS, em especial a Camila Teles, Gilberto Tavares, Maria Tinoco, Rogério Miorando, Mara Diogenes, Morgana Pizzolato, Samanta Yang e Tomoe Gusberti, pelo companheirismo, bom convívio e alegrias compartilhadas.

SANT'ANNA, Ângelo M. O. *Ferramentas para modelagem e monitoramento de características de qualidade do tipo fração*. 2009. Tese (Doutorado em Engenharia) – Universidade Federal do Rio Grande do Sul, Brasil.

RESUMO

Esta tese apresenta a modelagem e o monitoramento de características de qualidade do tipo fração, que apresentam valores restritos ao intervalo $[0,1]$. A motivação inicial é que a característica de qualidade do tipo fração pode ser melhor aproximada pela distribuição de probabilidade Beta e, geralmente, a modelagem e o monitoramento utilizados não supõem tal distribuição. Desta forma, os objetivos desta tese são: (i) Propor uma sistemática para seleção de modelos paramétricos considerando a classificação das variáveis resposta e das variáveis explicativas; (ii) Propor uma modelagem da fração não conforme com enfoque no modelo de regressão Beta e no modelo linear generalizado; (iii) Propor uma nova carta de controle (CC), Carta Beta, para monitorar características de qualidade do tipo fração (p), com adaptação dos limites de controle à distribuição Beta; (iv) Propor uma CC baseada em modelos Beta para monitorar características de qualidade do tipo fração em função das variáveis de controle do processo; e (v) Propor uma CC baseada em modelos Beta por componentes principais para monitorar características de qualidade do tipo fração em função das variáveis de controle multicolineares. As cartas de controle propostas foram avaliadas através da aplicação de estudos de caso, exemplos retirados da literatura e comparação do número médio de amostras (NMA) via simulação de Monte Carlo. Os modelos de regressão propostos para modelagem de características de qualidade do tipo fração apresentam flexibilidade de escolha da função de ligação, bom desempenho no ajuste e precisão das estimativas dos efeitos das variáveis. As cartas de controle propostas detectaram mais rapidamente todas as alterações induzidas, apresentando melhor desempenho em comparação as outras cartas similares encontradas na literatura. Conclui-se que as ferramentas propostas são adequadas para a modelagem e o monitoramento de características de qualidade do tipo fração.

Palavras-chave: controle de qualidade, cartas de controle, fração, atributos.

SANT'ANNA, Ângelo M. O. *Tools for modelling and monitoring of the type fraction quality characteristics*. 2009. Thesi (Doctorate in Engineering) – Federal University of Rio Grande do Sul, Brazil.

ABSTRACT

This thesis presents the modeling and monitoring of the type fraction quality characteristics, which have values confined in the $[0,1]$ -interval. The main motivation is that the type fraction quality characteristics can be better approximated by the Beta probability distribution and generally used for modeling and monitoring not assumes this distribution. Thus, the objectives are: (i) propose a systematic for selecting parametric models considering the classification of the response and explanatory variables; (ii) propose a modeling of the fraction nonconforming items using the Beta regression model and generalized linear model; (iii) propose a new control chart (CC), Beta Charts, to monitor the type fraction quality characteristics (p), with adaptation of the control limits based in the Beta distribution; (iv) propose a Beta model-based to monitor the type fraction quality characteristics adjusted by control variables of the process, and (v) propose a Beta model-based control charts by principal components to monitor the type fraction quality characteristics adjusted by control variables multicollinearity. The proposed control charts were evaluated by application of case studies, numeric examples taken from literature and comparison the average run length (ARL) using Monte Carlo simulation. The proposed regression models for modeling of the type fraction quality characteristics have shown flexibility to choose link function, good performance on the fit and accuracy of estimates of the effects. The proposed control charts detected more quickly all changes induced, showing better performance compared to other similar control charts defined in literature. It is concludes that the proposed tools are appropriate for modeling and monitoring of the type fraction quality characteristics.

Key words: quality control, control charts, fraction, attributes.

LISTA DE FIGURAS

Figura 1: Formas da distribuição Binomial par $n = 50$ e $p = 0,001; 0,01; 0,05$ e $0,1$	18
Figura 2: Estrutura das etapas da pesquisa desenvolvida	22
Figura 3: Categorias do termo fração ou razão	32
Figura 4: Estrutura dos principais modelos de regressão paramétricos.....	38
Figura 5: Sistemática para seleção de modelos paramétricos	41
Figura 6: Gráficos de resíduos padronizados dos modelos Normal (a) e Quase-verossimilhança (b).....	47
Figura 7: Gráficos de envelope simulado com o ajuste dos modelos Normal (a) e Quase-verossimilhança (b).....	47
Figura 8: Distribuição da fração não conforme com distribuição Normal sobreposta.....	64
Figura 9: Gráficos da fração não conforme em função das variáveis seleção (a) e procedência (b).....	64
Figura 10: Gráficos da fração não conforme em função das variáveis classificador (a) e rebaixamento (b).....	65
Figura 11: Gráficos de resíduos Deviance para os dados com o ajuste dos Modelos de Quase-verossimilhança (a) e Modelo Beta (b).....	69
Figura 12: Gráficos de distância de Cook para os dados com o ajuste dos Modelos de Quase-verossimilhança (a) e Modelo Beta (b).....	69
Figura 13: Gráficos de alavanca generalizada para os dados com o ajuste dos Modelos de Quase-verossimilhança (a) e Modelo Beta (b).....	70
Figura 14: Gráficos de envelope simulado para os dados com o ajuste dos Modelos de Quase-verossimilhança (a) e Modelo Beta (b).....	70
Figura 15: Distribuição das CQ y_1 (a), y_2 (b) e y_3 (c) com distribuição densidade e distribuição Normal sobrepostas	85
Figura 16: Cartas de controle da fração de embalagens não conformes (y_1)	87
Figura 17: Carta de controle da fração de amendoins não contaminados (y_2).....	88
Figura 18: Carta de controle da fração de amônia não convertida (y_3).....	89
Figura 19: Categorias do termo fração ou razão	102
Figura 20: Cartas de controle proposta por Hawkins (1991), Haworth (1996), Logit Normal, Logística e CCMB das CQ y_1, y_2, y_3	114

Figura 21: Carta de controle CCMBCP (a) para os resíduos e carta CCCP para os CP ₁ (b) e CP ₂ (c).....	138
Figura 22: Carta de controle CCMBCP para os resíduos e carta CCCP para os CP ₁ e CP ₂ segundo alteração 1σ	140
Figura 23: Carta de controle CCMBCP para os resíduos e carta CCCP para os CP ₁ e CP ₂ segundo alteração 3σ	141

LISTA DE TABELAS

Tabela 1: Estimativas dos coeficientes e critérios de diagnóstico dos modelos ajustados.	46
Tabela 2: Caracterização dos níveis dos fatores controláveis	63
Tabela 3: Estimativas dos coeficientes, erros-padrões e critérios de adequação dos modelos de regressão finais	67
Tabela 4: Matriz resumo de dados experimentais utilizados na modelagem	73
Tabela 5: Estimativas dos coeficientes, erros-padrões e critérios de adequação dos modelos de regressão propostos	74
Tabela 6: Tabela da constante w_1 e w_2 para a carta de controle Beta	84
Tabela 7: Resumo estatístico das CQ dos três exemplos investigados	86
Tabela 8: Limites de controle das cartas para as CQ dos três exemplos investigados	86
Tabela 9: Comparação entre os Limites de Controle de Shewhart, Ryan (1989), Chen (1998) e Carta Beta para variadas condições de np	90
Tabela 10: Comparação do NMA entre as cartas de controle de Shewhart, Ryan (1989), Chen (1998) e Beta para $p_0 = 0,001$, $n = 1500$, $\lambda = 0,001$ e $p_0 = 0,01$, $n = 200$, $\lambda = 0,01$	92
Tabela 11: Comparação do NMA entre as cartas de controle de Shewhart, Ryan (1989), Chen (1998) e Beta para $p_0 = 0,10$, $n = 50$ e 300 , $\lambda = 0,02$	93
Tabela 12: Tabela de valores das constantes w_1 e w_2 para a Carta de controle Beta.....	96
Tabela 13: Resumo estatístico das CQ do estudo experimental.....	109
Tabela 14: Estimativas de média e variância dos resíduos padronizados, <i>student</i> e <i>deviance</i> dos modelos.....	113
Tabela 15: Valores da constante w (e erros-padrão) das cartas de controle CCMB, Hawkins (1991), Haworth (1996) baseado em 5.000 simulações.....	117
Tabela 16: Resultado dos NMA (e erros-padrão do NMA) das cartas de controle Hawkins (1991), Haworth (1996) e CCMB baseado em 5.000 simulações.....	118
Tabela 17: Estimativas dos coeficientes com respectivos erros padrões e critérios de diagnóstico dos modelos ajustados para a CQ y_1	121
Tabela 18: Estimativas dos coeficientes com respectivos erros padrões e critérios de diagnóstico dos modelos ajustados para a CQ y_2	122
Tabela 19: Estimativas dos coeficientes com respectivos erros padrões e critérios de diagnóstico dos modelos ajustados para a CQ y_3	123

Tabela 20: Matriz de correlação entre a CQ e as VC do estudo simulado.....	136
Tabela 21: Estimativas dos coeficientes, erros padrões e valores de FIV para as VC do MRB	136
Tabela 22: Estimativas da variabilidade e dos autovetores dos CPs	137
Tabela 23: Estimativas dos coeficientes, erros padrões para os CP do MRB	137
Tabela 24: Dados retirados da literatura: Montgomery 2005, p. 180 (a); Draper e Smith 1998, p. 101 (b) e Brownlee 1965, p. 454 (c) utilizados na aplicação da carta de controle Beta	156
Tabela 25: Matriz dos dados experimentais coletados no processo de manufatura de pneus radiais.....	157

SUMÁRIO

1	INTRODUÇÃO.....	14
1.1	TEMA E OBJETIVOS.....	15
1.2	JUSTIFICATIVAS DO TEMA E OBJETIVOS	16
1.3	DELINEAMENTO DO ESTUDO	20
1.4	DELIMITAÇÕES DO TRABALHO.....	25
2	ARTIGOS PROPOSTOS	27
2.1	ARTIGO 1 – Sistemática para Seleção de Modelos Paramétricos na Otimização Experimental.....	28
2.2	ARTIGO 2 – Modelagem da Fração não conforme em Processos Industriais.....	50
2.3	ARTIGO 3 – Carta de Controle Beta para Monitoramento de Características de Qualidade do tipo Fração	75
2.4	ARTIGO 4 – Carta de Controle baseada em Modelos Beta para Monitoramento de Características de Qualidade do tipo Fração	98
2.5	ARTIGO 5 – Carta de Controle baseada em Modelos Beta via Componentes Principais para Monitoramento de Características de Qualidade do tipo Fração..	124
3	CONSIDERAÇÕES FINAIS.....	145
3.1	CONCLUSÕES	145
3.2	SUGESTÕES PARA PESQUISAS FUTURAS	147
	REFERÊNCIAS.....	148
	APÊNDICE	155
A.1	Matriz de dados utilizados nos estudos desenvolvidos	155

1 INTRODUÇÃO

Para as empresas se manterem no mercado atual faz-se importante aprimorar seus processos e priorizar a qualidade dos seus produtos. Os esforços devem estar direcionados para controlar alterações no processo, assegurar as especificações dos seus produtos e reduzir os custos associados à manufatura, permitindo o aumento da produtividade. Um importante indicador de desempenho monitorado em uma empresa é a fração de produtos não conformes sobre o total de unidades produzidas, pois a produção destes produtos pode acarretar as empresas desperdícios, má qualidade e insatisfação de clientes.

Controle estatístico de processos (CEP) é uma técnica comumente utilizada no monitoramento e detecção de alterações no processo produtivo. Cartas de controle são importantes ferramentas do CEP e têm diversas aplicações no meio industrial graças a sua simplicidade operacional (KUME, 1993). Uma carta de controle é um registro gráfico da qualidade da característica particular de um produto. Desde sua introdução por Shewhart, as cartas de controle são aplicadas para distinguir em um processo produtivo as variações que são inerentes a tecnologia do processo (causas comuns) daquelas que resultam de falhas de operação (causas especiais).

Os limites de controle da carta de CEP são calculados considerando apenas as causas comuns de variabilidade. Quando um ponto é *plotado* fora dos limites de controle da carta, há uma evidência de causa especial no processo que deve ser investigada. Um aspecto importante no CEP é a análise de desempenho das cartas de controle, a qual é definida pelo número médio de amostras até a presença de um sinal (NMA). Esta medida de desempenho, por sua vez, depende dos parâmetros de projeto da carta de controle (MONTGOMERY, 2005).

As cartas de controle podem ser classificadas em duas categorias: cartas de controle para variáveis e para atributos. As cartas de controle para variáveis monitoram o comportamento das características de qualidade mensuradas numa escala numérica, sendo os limites de controle definidos através dos parâmetros da distribuição de probabilidade Normal. As cartas de controle para atributos monitoram o comportamento das características de qualidade mensuradas por comparação com um padrão de referência, podendo gerar dois tipos de indicadores: número de não conformes (defeituosos) e número de não conformidades (defeitos), sendo os limites de controle definidos pela distribuição Binomial e Poisson, respectivamente.

A seleção do tipo de carta de controle para atributos depende do tipo do indicador e do tamanho de amostra. As cartas de controle para atributos mais populares para monitorar a fração de produtos não conformes às especificações são Cartas p e np , desenvolvidas por Shewhart em 1924 (SHEWHART 1931). A média e a variância são calculadas assumindo que o número de produtos não conformes segue a distribuição de probabilidade Binomial, com parâmetros n e p , e os limites de controle e limite central são calculados aproximando-se pela distribuição Normal.

Desde a década de 30, as contribuições técnicas no CEP cresceram consideravelmente com o avanço da tecnologia e complexidade dos processos. Estas exigências conduziram a um grande número de inovações nos procedimentos de controle, por exemplo, uso de métodos como modelos de regressão, estatística multivariada e simulações.

Há processos em que a característica de qualidade é diretamente dependente de algumas variáveis de controle do processo, e estas podem assumir diferentes valores em função de alterações frequentes de ajuste do processo, modificando o modelo de referência da característica de qualidade. Nessa situação, a literatura recomenda utilizar cartas de controle baseadas em modelos, que consiste em ajustar a influência das variáveis de controle sobre a característica de qualidade do produto ou processo por meio de um modelo de regressão, e então *plotar* os resíduos ordinários gerados entre as características de qualidade observadas e previstas pelo modelo de regressão (MANDEL, 1969).

A carta de controle baseada em modelos é basicamente uma combinação das cartas de controle de Shewhart com modelos de regressão. Segundo Wade e Woodall (1993), a carta de controle de Shewhart objetiva monitorar um parâmetro de referência constante do processo e a carta de controle baseada em modelos objetiva monitorar uma característica de qualidade cujo parâmetro de referência varia conforme os valores das variáveis de controle.

1.1 TEMA E OBJETIVOS

O tema desta tese contempla cartas de controle estatístico (CC) para o monitoramento de processos que mensuram características de qualidade do tipo fração, que apresentam valores restritos ao intervalo $[0,1]$, considerando a distribuição de probabilidade e modelos de regressão Beta. Este tema está inserido na área de engenharia da qualidade, concentrando-se em cartas de controle e modelos de regressão.

O objetivo principal deste trabalho é modelar e monitorar características de qualidade do tipo fração, com mensurações restritas ao intervalo $[0,1]$. Como decorrência do objetivo principal, pretende-se alcançar os seguintes objetivos específicos:

- a) Propor uma sistemática para seleção de modelos paramétricos considerando a classificação das variáveis resposta e das variáveis explicativas;
- b) Propor uma modelagem da fração não conforme com enfoque no modelo de regressão Beta e no modelo linear generalizado;
- c) Propor uma nova CC, Carta Beta, para monitorar características de qualidade do tipo fração (p), com adaptação dos limites de controle à distribuição Beta;
- d) Propor uma CC baseada em modelos Beta para monitorar características de qualidade do tipo fração em função das variáveis de controle do processo;
- e) Propor uma CC baseada em modelos Beta por componentes principais para monitorar características de qualidade do tipo fração em função das variáveis de controle multicolineares.

1.2 JUSTIFICATIVAS DO TEMA E OBJETIVOS

O constante avanço da tecnologia empregada nos processos industriais, exige que ferramentas de controle estatístico da qualidade sejam capazes de monitorar, detectar e avaliar alterações nas características de qualidade, durante o processo de manufatura, o mais breve possível. A análise e investigação de possíveis efeitos das variáveis de controle na produção de produtos não conformes às especificações são importantes para a melhoria da qualidade dos processos industriais.

Os modelos de regressão relacionam o comportamento das características de qualidade, também chamadas de variáveis resposta ou dependentes, em função das variáveis de controle do processo, também chamadas de fatores controláveis ou variáveis explicativas ou independentes. Uma sistemática que oriente a escolha adequada de quais modelos de regressão podem ser utilizados na modelagem de dados experimentais, considerando as variáveis resposta e as variáveis explicativas, pode evitar inferências inadequadas e resultados errôneos. Conforme Hair et al. (1998), para o uso de abordagem estatística em análise de processos é necessário previamente identificar as variáveis em estudo, assim como, a classificação das variáveis em discreta ou contínua.

O modelo de regressão mais conhecido para modelar variáveis resposta contínuas é o modelo linear Normal, cujos pressupostos para seu uso são: (i) observações independentes e identicamente distribuídas (*i.i.d*); (ii) homoscedasticidade e (iii) normalidade dos erros. Segundo Kieschnick e McCulloch (2003), um modelo de regressão linear Normal utilizado na modelagem da fração não conforme é um modelo falho, pois não satisfaz as suposições, produzindo: (i) heteroscedasticidade, ou seja, não homogeneidade de variância das observações; (ii) não normalidade do termo de erro; e (iii) possibilidade da probabilidade estimada (fração predita) estar fora do limite do intervalo [0,1].

O modelo de regressão Beta apresenta características importantes na modelagem de características de qualidade do tipo fração, a saber: (i) distribuição de probabilidade que melhor se ajusta aos dados; (ii) não normalidade do termo de erro; (iii) variância dos dados dependente da média; (iv) probabilidade estimada (fração predita) contida no intervalo [0,1] e (v) modelagem adequada dos dados para pequenos tamanhos de amostra (KIESCHNICK e McCULLOGH, 2003; VASCONCELLOS e CRIBARI-NETO, 2005).

As empresas monitoram seus processos através das cartas de controle de Shewhart para variáveis ou para atributos. A carta de controle para atributos do tipo Carta p é utilizada para monitorar características de qualidade que mensuram a fração não conforme (p). A Carta p assume que a fração não conforme segue a distribuição Binomial e os limites de controle são calculados com aproximação pela distribuição Normal.

Em muitos estudos a Carta p é utilizada em situações onde o parâmetro p da distribuição Binomial é consideravelmente pequeno (i.e. $p = 0,001; 0,01; 0,05; 0,1; \dots$) e nestes casos a distribuição Binomial é assimétrica e a aproximação pela distribuição Normal não é satisfatória, pois os limites de controle assumem valores negativos ou maiores do que 1. Por exemplo, Seja Y uma variável aleatória com distribuição Binomial de parâmetros $p = 0,001; 0,01; 0,05; 0,1$ e $n = 100$ pode-se obter as formas da distribuição Binomial conforme a Figura 1.

Encontra-se na literatura diversos aprimoramentos da Carta p para monitorar a fração não conforme com objetivo de melhorar a aproximação pela distribuição Normal. Ryan e Schwertman (1997) sugeriram modificações para os limites de controle para valores de $p < 0,03$. McColl e Motley (1998) sugerem uma transformação potência ($x^{0,2777}$) no cálculo dos limites de controle para pequenos valores da fração não conforme (5ppm). Chen (1998) propôs um ajuste nos limites de controle da Carta p utilizando uma correção assintótica de primeira ordem.

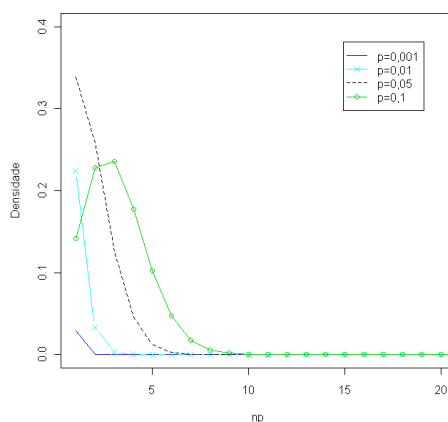


Figura 1: Formas da distribuição Binomial para $n = 50$ e $p = 0,001; 0,01; 0,05$ e $0,1$
 Fonte: Elaborado pelo autor

Nesta tese assume-se que no cálculo dos limites de controle, a distribuição Binomial seria melhor aproximada pela distribuição Beta do que pela distribuição Normal, o que melhoraria o desempenho da carta avaliada através do número médio de amostras (NMA) até a detecção do sinal. O NMA visa determinar a taxa de alarmes falsos da carta de controle investigada e a sensibilidade desta carta em detectar alterações pré-definidas no processo (MONTGOMERY, 2005). Com o cálculo dos valores do NMA comparado com outras cartas identificadas na literatura, espera-se validar as cartas de controle propostas nesta tese.

Ao utilizar cartas de controle, supõe-se que os dados provenientes do processo sejam independentes e identicamente distribuídos em torno de um modelo de referência com média constante (MONTGOMERY, 2005). Quando existir relação de causa e efeito entre as características de qualidade (CQ) monitorada e as variáveis de controle (VC) do processo e estas forem constantemente alteradas, o modelo de referência com média constante não será válido para todas as amostras do processo. Neste caso, os valores da CQ irão variar, conforme os valores das VC, aumentando desnecessariamente o número de alarmes falsos do processo.

As cartas de controle baseadas em modelos têm como principal função controlar uma variação da CQ, que ocorre devido à relação de dependência da CQ em função das VC do processo, ao invés de controlar um valor constante do processo como é usualmente feito pelas cartas de controle tradicionais. A carta de controle baseada em modelos realiza a previsão do valor da CQ por um modelo de regressão em função do seu relacionamento com as VC do processo. O procedimento é monitorar os resíduos entre as CQ observadas e as CQ previstas pelo modelo. Um detalhe relevante para utilização das cartas de controle baseadas em modelos é que exista uma relação de causa e efeito entre as CQ e as variáveis de controle,

sendo que essa relação deve ser modelada adequadamente para não conduzir a decisões errôneas acerca do processo.

Os primeiros trabalhos utilizando a abordagem de monitorar por carta de controle de Shewhart uma CQ em função de uma VC de processo utilizando regressão linear múltipla foram apresentados por Lyle (1944), DiPaola (1945), Weis (1957) e Mansfield e Wein (1958). Mandel (1969) descreve a ferramenta e apresenta similaridades e diferenças em relação à carta de controle de Shewhart (Carta \bar{x}). Em seguida, Zhang (1985), Rothschild e Roth (1986); Hawkins (1991, 1993), Spanos (1992) e Haworth (1996) expandem a aplicação da carta de controle utilizando modelos de regressão múltipla. Wade e Woodal (1993), Hauck et al. (1999), Jacobi et al. (2002), Loredó et al. (2002) e Shu et al. (2004) apresentam modificações da ferramenta para aplicações de monitoramento de processos em cascata e autocorrelacionados. As cartas de controle baseadas em modelos propostos pelos autores monitoram processos que mensuram CQ contínua utilizando o método de estimação por mínimos quadrados ordinários, com suposição de normalidade para as observações e para os resíduos e variância constante.

Em situações onde se deseja monitorar CQ do tipo atributos, essas suposições dificilmente podem ser satisfeitas. Skinner et al. (2003, 2004) monitoram a CQ com distribuição de Poisson em função das VC utilizando modelos lineares generalizados e utilizam o método de estimação por máxima verossimilhança. Cartas de controle univariada para monitorar simultaneamente características de qualidade do tipo frações constituem-se um desenvolvimento teórico recente, e conseqüentemente, ainda pouco explorado. A alternativa proposta nesta tese é uma carta de controle baseada em modelos Beta para monitorar os resíduos *deviance* de um processo que mensura características de qualidade do tipo fração em função de VC. Este modelo de regressão assume a distribuição de probabilidade Beta para a CQ do tipo fração, a função de ligação logit para relacionar com as VC e o método de máxima verossimilhança para estimar os coeficientes do modelo.

Algumas situações especiais dificultam a utilização das cartas de controle baseadas em modelos. Por exemplo, os modelos de regressão utilizados na modelagem da relação entre as variáveis investigadas pressupõem independência entre as VC. Quando há multicolinearidade entre as VC, os modelos de regressão podem apresentar efeitos errôneos nas estimativas do erro padrão dos coeficientes do modelo e prejudicar o teste de significância das VC acarretando na qualidade de adequação do modelo (NETER et al., 2005), sendo então necessário considerar a estrutura de correlação das VC multicolineares no relacionamento com as CQ.

A análise de componentes principais (ACP) permite eliminar a multicolinearidade entre as VC gerando componentes principais (CP) ortogonais, que são combinações lineares das VC multicolineares, sem acarretar perda significativa de informações. Segundo Jackson (1991), a regressão por componentes principais modela a CQ em função dos componentes principais e permite investigar a influência das VC originais. Através da proposta de utilizar a técnica multivariada de ACP, espera-se que a carta de controle baseada em modelos Beta também possa ser aplicada em processos que contenham VC multicolineares, aumentando a eficiência da carta de controle nesses processos.

Na aplicação de métodos estatísticos, como modelos de regressão e cartas de controle mais sofisticadas ou recentemente desenvolvidas, freqüentemente se encontra um obstáculo no que diz respeito aos recursos computacionais disponíveis. Os *softwares* estatísticos mais comumente usados em Engenharia como, STATGRAPHICS[®], SPSS[®] e MINITAB[®] não apresentam procedimentos para construir os modelos de regressão e as cartas de controle propostas neste trabalho. A construção dos modelos de regressão e das cartas de controle propostas foi realizada na linguagem computacional R[®] v.2.9.0, e no aplicativo Microsoft Excel[®] 2007.

Considera-se este estudo de caráter relevante para a área científica, por ser inovador em utilizar a distribuição de probabilidade Beta e os modelos de regressão Beta para a modelagem e monitoramento da CQ restrita ao intervalo [0,1] , assim como, na aplicação de técnicas como modelos de regressão e componentes principais no controle estatístico de processos que mensuram CQ do tipo fração.

1.3 DELINEAMENTO DO ESTUDO

Uma vez definidos os objetivos deste trabalho, torna-se necessário estabelecer o delineamento pelo qual os objetivos serão alcançados, os métodos abordados e a estrutura deste trabalho.

Método de Pesquisa

A sustentação metodológica para elaboração dos artigos propostos nesta tese segue a lógica hipotético-dedutiva. O método hipotético-dedutivo se inicia pela percepção de uma lacuna nos conhecimentos e possibilita construir hipóteses originadas de problemas teóricos/práticos existentes, que devem ser submetidas à verificação com propósito de serem

corroboradas (LAKATOS; MARCONI, 2005). O objetivo das hipóteses na pesquisa é sugerir explicações para o problema investigado e estas serão aceitas ou rejeitadas somente depois de devidamente testadas. A importância das hipóteses está em direcionar a investigação do pesquisador contribuindo no desenvolvimento do conhecimento científico (LAKATOS; MARCONI, 2005; GIL, 2008).

A caracterização do raciocínio hipotético-dedutivo aqui exposto configura-se por uma pesquisa aplicada, uma vez que objetiva gerar conhecimentos dirigidos para aplicação e à solução de problemas específicos (GIL, 2008). Os artigos a serem propostos referem-se à verificação empírica das hipóteses com base nas referências teóricas existentes e estudos experimentais investigados.

Gil (2002) discute a importância de caracterizar uma pesquisa científica de forma a delinear as etapas para a sua realização. A presente tese se caracteriza como uma pesquisa exploratória e experimental: (i) por possuir objetivos exploratórios, visando identificar características da relação entre variáveis em estudo e proporcionar um maior conhecimento sobre o problema estudado, contribuindo na compreensão da ocorrência do fenômeno; (ii) por haver o interesse em observar os efeitos das variáveis no objeto de estudo e a verificabilidade das hipóteses formuladas.

Método de Trabalho

A primeira etapa desta pesquisa envolve o desenvolvimento dos estudos apresentados na Figura 2. Esta etapa compreende a execução das atividades embasada na teoria e nos conceitos revisados a fim de alcançar os objetivos propostos. A etapa seguinte contemplou a validação das CC propostas através da: (i) análise comparativa com outras CC para média constante e média não-constante encontradas na literatura; e (ii) análise de sensibilidade das cartas de controle investigadas comparando-se o número médio de amostras (NMA) até a detecção do sinal em estudos de caso real e simulado. Na seqüência são descritos cada um dos estudos.

Estudos	Objetivos	Questões de Pesquisa	Revisão Teórica	Método de Pesquisa
Artigo 1	Propor uma sistemática para seleção de modelos paramétricos considerando a classificação das variáveis resposta e das explicativas	Q ₁ : A escolha adequada dos modelos paramétricos previne inferências inadequadas?	1. Classificação de Variáveis 2. Modelos de Regressão 3. Critérios de Diagnósticos	Pesquisa de caráter Explicativa e Estudo Experimental
Artigo 2	Propor uma modelagem da fração não conforme com enfoque no modelo de regressão Beta e no modelo linear generalizado	Q ₂ : Estes modelos de regressão são mais adequados para modelar fração?	1. Modelos Lineares Generalizados 2. Modelos de Regressão Beta 3. Critérios de Diagnósticos	Pesquisa de caráter Explicativa e Estudo Experimental
Artigo 3	Propor uma nova CC, Carta Beta, para monitorar características de qualidade do tipo fração (p), com adaptação dos limites de controle à distribuição Beta	Q ₃ : A aproximação da distribuição Binomial pela Beta é mais vantajosa?	1. Carta de Controle de Atributos 2. Distribuições de Probabilidade 3. Análise de Desempenho de CC	Pesquisa de caráter Explicativa e Estudo Experimental
Artigo 4	Propor uma CC baseada em modelos Beta para monitorar características de qualidade do tipo fração em função das variáveis de controle do processo	Q ₄ : O modelo Beta relaciona adequadamente a fração em função das VC?	1. CC baseada em Modelos 2. Modelo de Regressão Beta 3. Análise de Desempenho de CC	Pesquisa de caráter Explicativa e Estudo Experimental
Artigo 5	Propor uma CC baseada em modelos Beta por componentes principais para monitorar características de qualidade do tipo fração em função das variáveis de controle multicolineares	Q ₅ : O modelo Beta com CP relaciona adequadamente a fração em função das VC multicolineares?	1. CC baseada em Modelos 2. Análise de Componentes Principais 3. Análise de Desempenho de CC	Pesquisa de caráter Explicativa e Estudo Simulado

Figura 2: Estrutura das etapas da pesquisa desenvolvida

Fonte: Elaborado pelo autor

O artigo 1 – Sistemática para Seleção de Modelos Paramétricos na Otimização experimental – contempla o desenvolvimento de um estudo relacionado ao tema de pesquisa. Inicialmente buscou classificar os tipos de variáveis, considerando a natureza dos dados experimentais e revisar os modelos de regressão pertencentes à classe dos modelos lineares generalizados, a qual contempla uma gama de distribuições de probabilidade, i.e. Normal, Binomial, Gama, Exponencial, Geométrica, Weibull, entre outras. Em particular, estes modelos investigam a relação entre uma única variável resposta e diversas variáveis explicativas (modelagem univariada).

Em seguida, elaborou-se uma sistemática para seleção de modelos paramétricos, considerando a classificação da natureza das variáveis resposta e das variáveis explicativas. A sistemática para seleção de modelos foi aplicada a um estudo experimental com projeto fatorial fracionado (2^{k-1}) executado a partir de testes de detonação com explosivos em mina, onde se buscou avaliar a influência de variáveis explicativas na velocidade de partícula na vibração captada por um sismógrafo.

Os modelos paramétricos sugeridos para modelar o experimento apresentaram qualidade no ajuste e predição dos dados, conduzindo a resultados confiáveis. A sistemática desenvolvida mostrou-se como uma estratégia relevante e adicional para pesquisas em modelagem de dados e otimização experimental. Este artigo foi submetido para publicação na Revista Produção (ISSN 0377-2217) em Novembro 2008.

O artigo 2 – Modelagem da Fração não conforme em Processos Industriais – contempla o desenvolvimento de um estudo relacionado ao tema de pesquisa. A partir do estudo realizado anteriormente, buscou-se avaliar os modelos de regressão utilizados para modelar variáveis respostas do tipo fração de produtos não conformes, as quais apresentam valores restritos ao intervalo $[0,1]$. A modelagem dos dados foi realizada no processo industrial da empresa Bracol Couro, produtora de couro acabado e fornecedora para a indústria calçadista e de artefatos de couro da região do Vale do Rio dos Sinos. Para a empresa, um dos principais interesses era conhecer quais são as variáveis de controle de maior influência na produção de refugo por erros de classificação no produto acabado causados por erros de classificação no estágio *wet blue*, com o intuito de definir procedimentos para reduzir a ação das fontes de variabilidade.

A modelagem dos dados permitiu investigar uma relação não-linear entre a variável resposta e as variáveis de controle e obter estimativas precisas dos efeitos destas variáveis. Os critérios de diagnóstico de modelos verificaram a adequação dos modelos de regressão ajustados, permitindo realizar uma comparação entre os dois modelos utilizados e destacando vantagens e desvantagens do seu uso. Este artigo foi aceito para publicação na Revista Pesquisa Operacional (ISSN 0101-7438) em Junho 2009.

O artigo 3 – Carta de Controle Beta para Monitoramento de Características de Qualidade do tipo Fração – contempla o desenvolvimento de um estudo principal do tema de pesquisa. Inicialmente buscou categorizar as características de qualidade do tipo fração (percentual ou proporção) considerando a natureza de mensuração e distribuição de probabilidade. Realizou-se uma revisão sobre distribuições de probabilidade e cartas de controle para atributos utilizadas no monitoramento da fração de produtos não conformes em processos industriais. Em seguida, desenvolveu-se a construção da carta de controle Beta a partir dos cálculos para os limites de controle baseados na distribuição de probabilidade Beta.

Posteriormente, a validação da carta de controle proposta foi realizada comparando os resultados com outras cartas de controle encontradas na literatura, em duas etapas: (i) aplicação da carta de controle em exemplos numéricos retirados na literatura; e (ii) análise de sensibilidade da carta de controle via NMA. Estas etapas, embora sejam inter-relacionadas,

apresentam objetivos diferentes, já que a primeira visa provar a aplicabilidade e a praticidade da carta de controle proposta e a segunda objetiva verificar a eficiência na detecção de alterações no processo. Como resultado desta validação, conclui-se que a carta proposta apresenta melhor desempenho no monitoramento de processos produtivos que mensuram variáveis do tipo fração.

O artigo 4 – Carta de controle baseada em Modelos Beta para Monitoramento de Características de Qualidade do tipo Fração – contempla o desenvolvimento de um estudo principal do tema de pesquisa. Inicialmente realizou-se uma revisão sobre cartas de controle baseadas em modelos. Não foram encontrados trabalhos que apresentassem cartas de controle baseadas em modelos utilizadas no monitoramento de características de qualidade do tipo fração em sistemas produtivos. A partir do estudo anterior realizado sobre modelos de regressão Beta utilizados para modelar a fração não conforme, desenvolveu-se uma carta de controle baseada em modelos Beta para monitorar os resíduos *deviance* em processos que mensuram características de qualidade do tipo fração e que dependam de frequentes ajustes nas variáveis de controle do processo. Ademais, são descritas as etapas para implementação da carta de controle proposta.

Para validação da carta de controle proposta foi realizado um estudo em duas etapas: (i) aplicação da carta de controle em um estudo de caso real; e (ii) análise de sensibilidade da carta de controle via NMA. Os resultados foram comparados com os de outras cartas de controle similares encontrados na literatura. A carta de controle proposta foi aplicada em um processo de manufatura de pneus radiais na empresa multinacional Pirelli Pneus. A análise de sensibilidade consistiu num estudo de simulação de Monte Carlo para obtenção do NMA de cada carta de controle investigada para um processo sob controle e fora de controle considerando dois tipos e tamanhos de alterações induzidas nos coeficientes dos modelos de regressão. Como resultado desta validação, conclui-se que a carta proposta apresenta melhor desempenho no monitoramento de processos produtivos que mensuram características de qualidade do tipo fração e que dependem de alterações nas VC.

O artigo 5 – Carta de Controle baseada em Modelos Beta por Componentes Principais para Monitoramento de Características de Qualidade do tipo Fração – contempla o desenvolvimento de um estudo principal do tema de pesquisa. O intuito deste estudo é possibilitar o monitoramento de processos que mensuram características de qualidade do tipo fração em função das variáveis de controle multicolineares através de um procedimento metodológico que permita tal monitoramento e apresente eficiência no diagnóstico de alterações no processo. Inicialmente realizou-se uma revisão abordando os conceitos de

modelos de regressão Beta, análise de componentes principais, cartas de controle baseadas em modelos Beta (CCMB) e baseada em componentes principais (CCPC).

Para validação da carta de controle proposta foi realizado um estudo em duas etapas: (i) aplicação da carta de controle em um estudo de caso simulado; e (ii) análise de sensibilidade da carta de controle via NMA. A análise de sensibilidade consistiu num estudo de simulação de Monte Carlo para obtenção do NMA da carta de controle para um processo sob controle e fora de controle, considerando dois tamanhos de alteração aditiva induzida nas VC. Como resultado desta validação, conclui-se que a carta proposta apresenta bom desempenho no monitoramento de processos que mensuram características de qualidade do tipo fração com VC multicolineares.

Estrutura do Trabalho

Esta tese está organizada em três capítulos principais. No primeiro capítulo tem-se uma visão geral da pesquisa realizada, o contexto de variáveis do tipo fração e sua importância como característica de qualidade, o tema, os objetivos, as justificativas, o delineamento do estudo, a descrição dos artigos propostos e as delimitações da tese.

O segundo capítulo apresenta os artigos propostos e respectivos resultados obtidos conforme a Figura 2. O terceiro capítulo apresenta as considerações finais obtidas com o desenvolvimento dos artigos e sugestões para pesquisas futuras a partir dos assuntos abordados nesta pesquisa.

1.4 DELIMITAÇÕES DO TRABALHO

Esta pesquisa busca o desenvolvimento de novas propostas para a modelagem e o monitoramento de características de qualidade do tipo fração, que apresentam valores restritos ao intervalo $[0,1]$. As cartas de controle propostas são aplicadas e validadas em diferentes processos produtivos.

Os modelos de regressão e as cartas de controle propostos contemplam apenas uma única CQ restrita ao intervalo $[0,1]$, não sendo contemplados modelos de regressão e cartas de controle para processos multivariados. A análise multivariada utilizada se restringe a resolver os problemas de multicolinearidade entre as VC.

A utilização dos índices de capacidade, embora seja útil para avaliar a capacidade do processo em atender às especificações do produto, não será escopo do presente trabalho.

A avaliação do custo de implementação das cartas de controle propostas não será abordada nesta pesquisa.

A implementação das CC propostas em *software* não será escopo desta pesquisa.

2 ARTIGOS PROPOSTOS

- 2.1 **ARTIGO 1** – Sistemática para Seleção de Modelos Paramétricos na Otimização Experimental..... 28
- 2.2 **ARTIGO 2** – Modelagem da Fração não conforme em Processos Industriais..... 50
- 2.3 **ARTIGO 3** – Carta de Controle Beta para Monitoramento de Características de Qualidade do tipo Fração 75
- 2.4 **ARTIGO 4** – Carta de Controle baseada em Modelos Beta para Monitoramento de Características de Qualidade do tipo Fração 98
- 2.5 **ARTIGO 5** – Carta de Controle baseada em Modelos Beta via Componentes Principais para Monitoramento de Características de Qualidade do tipo Fração.. 124

2.1 ARTIGO 1

Sistemática para Seleção de Modelos Paramétricos na Otimização experimental

Systematic for select Parametric Models in Experimental Optimization

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

Enviado à Revista Produção (ISSN 0103-6513) em Novembro 2008

Sistemática para Seleção de Modelos Paramétricos na Otimização experimental

Ângelo Márcio Oliveira Sant'Anna

Carla Schwengber ten Caten

Universidade Federal do Rio Grande do Sul – UFRGS

Programa de Pós-graduação em Engenharia de Produção

End.: Av. Osvaldo Aranha, n. 99/ 5º andar, Bom Fim – CEP 90.035-190 – Porto Alegre/RS

Tel.: (51) 3308 3490 / (51) 3308 4293 / Fax.: (51) 3308 4007

E-mail.: angelo@producao.ufrgs.br, tencaten@producao.ufrgs.br

Resumo

É importante salientar que não é adequado tratar os diferentes tipos de variáveis com os mesmos modelos de regressão, sendo que há modelos de regressão específicos para cada conjunto de variáveis. O objetivo deste artigo é apresentar uma sistemática para seleção de modelos paramétricos, considerando a classificação das variáveis resposta e variáveis explicativas. A sistemática de seleção foi aplicada em um estudo de caso real em testes de detonação por explosivos, onde se buscou avaliar a influência de fatores controláveis na velocidade de partícula captada por um sismógrafo. Os modelos paramétricos sugeridos para modelar o experimento apresentaram qualidade no ajuste e na predição, conduzindo a resultados confiáveis. A sistemática desenvolvida se mostrou uma estratégia relevante e uma ferramenta adicional para pesquisas em modelagem de dados e otimização experimental.

Palavras-chave: Modelos de regressão, seleção de modelos e otimização experimental.

Systematic for Select Parametric Models in Experimental Optimization

Abstract

It's important to emphasize that is inappropriate to treat the different types of variables with the same parametric regression models, considering that there are specific regression models for each combination of variables. In this paper we propose a systematic for selection of parametric regression models, considering the classification of response and explanatory variables. The selection systematic was applied in a case study run using detonation with explosives tests, in which we evaluate the influence of controllable factors in the particle speed captured by a seismograph. The parametric models suggested for modeling the experiment shown goodness of fit and prediction, providing trustable results. The systematic developed showed as a relevant strategy and additional tools for modeling dates researches and experimental optimization.

Keyword: Regression models, selection models, experimental optimization.

1. Introdução

Projetos experimentais têm sido amplamente utilizados em análise de processos produtivos e fenômenos aleatórios a fim de identificar e controlar a influência de causas ou fatores sobre uma ou mais características de qualidade. Ademais, planejar um experimento contribui para estabelecer quais níveis dos fatores se obtêm os melhores resultados, permitindo definir o ajuste ótimo dos fatores controláveis de forma a maximizar o desempenho das características de qualidade, de maneira que os dados apropriados sejam coletados em tempo e custos mínimos.

Os projetos de experimentos podem ser classificados de acordo com os seguintes critérios: *(i)* número de fatores a serem estudados; *(ii)* pela estrutura do planejamento e *(iii)* pelo tipo de informação que o experimento pode fornecer. Existem vários tipos de projetos, por exemplo, o projeto fatorial completo, que testa todas as combinações de níveis dos fatores controláveis em estudo. Esta estratégia experimental aumenta a eficiência do experimento em detectar as possíveis diferenças entre os fatores, porque a variabilidade dos fatores investigados deixa de fazer parte do erro experimental. Há também, o projeto fatorial fracionado que consiste em realizar uma fração dos ensaios completos, em virtude de limitações experimentais ou de recursos (MONTGOMERY, 2004).

Uma vez que os fatores controláveis tenham sido identificados é geralmente importante modelar a relação existente entre os fatores e a característica de qualidade do produto ou processo, que muitas vezes é chamada de variável resposta. Segundo Montgomery et al. (2006), a análise de regressão é uma técnica estatística de investigação e modelagem que quantifica a influência dos fatores controláveis sobre a variável resposta. Estes fatores, também chamados de variáveis explicativas, são classificados em quantitativos ou qualitativos.

Conforme seja o experimento realizado, pode-se fazer uso de técnicas estatísticas de modelagem mais robustas à variação aleatória, aumentando a precisão da estimativa dos efeitos dos fatores e adequação dos modelos. Em recentes pesquisas, a utilização conjunta de modelos lineares generalizados (MLG) e projeto e análise de experimentos (PAE) têm permitido a modelagem da variável resposta em função das variáveis explicativas ou fatores controláveis e a identificação do ajuste ótimo de operação dos fatores controláveis que melhora o desempenho dos mesmos.

Segundo Larson e Farber (2005), a qualidade dos resultados de uma análise estatística está diretamente relacionada com a escolha adequada da técnica estatística a ser utilizada e esta com a classificação das variáveis respostas e explicativas em estudo.

Como existem muitos modelos de regressão possíveis conforme o tipo de dados coletados surge a necessidade de um procedimento que oriente a escolha do melhor modelo, segundo a classificação das variáveis em estudo, possibilitando melhor resultado na modelagem de dados. Observou-se na literatura que não há uma orientação detalhada sobre quais modelos de regressão utilizar segundo um conjunto de informações coletadas.

Este artigo tem como objetivo apresentar uma sistemática para seleção de modelos paramétricos, considerando a classificação das variáveis resposta e explicativas. Para ilustrar esta sistemática foi aplicada em um estudo de caso real.

Na seção 2, é apresentada uma breve revisão sobre a classificação de variáveis, modelos de regressão paramétricos e critérios de diagnóstico. Na seção 3, é apresentada a sistemática para seleção dos modelos paramétricos. Na seção 4, é apresentada a aplicação da sistemática proposta. Finalmente, na seção 5, são apresentadas as conclusões sobre o estudo desenvolvido nesse artigo.

2. Modelagem de Dados Experimentais

Esta seção apresenta um referencial teórico da classificação de variáveis, dos modelos lineares generalizados e dos critérios de diagnóstico de modelos paramétricos que auxiliam na otimização experimental.

2.1 Classificação de Variáveis

O tratamento estatístico de dados inicia com a observação de que “tipo” de dados devem ser coletados para o estudo, pois as informações contidas nos dados são importantes e compõem a base do critério científico na tomada de decisão sobre um processo de produção. Segundo Montgomery (2004), na realização de uma abordagem estatística, ao projetar e analisar um experimento, é necessário previamente possuir uma idéia do que será estudado, de como os dados serão coletados, da natureza dos dados e um entendimento qualitativo de como serão analisados.

De acordo com a estrutura numérica, as variáveis podem ser classificadas em: (i) Qualitativas ou não-métricas – se os resultados das observações serão expressos através de categorias ou níveis que se distinguem por alguma característica não-numérica, apresentando

ou não ordenamento; (ii) Quantitativas ou métricas – se os resultados das observações forem expressos sempre através de números, que representam contagens ou medidas, pertencentes a um conjunto dos números reais. A Figura 3 apresenta a classificação das variáveis por tipo de mensuração.

As variáveis qualitativas apresentam dois níveis: nominal e ordinal. As nominais caracterizam-se por dados que consistem apenas em nomes, rótulos ou categorias. As ordinais apresentam ordenação das várias categorias, possibilitando verificar graus de intensidade entre elas.

As variáveis quantitativas apresentam dois níveis: discreta ou contínua. As discretas apresentam valores expressos como números inteiros, além da possibilidade desses números serem utilizados em forma de pontos de uma escala. E as contínuas, como valores expressos dentro de um intervalo de variação pertencente ao conjunto dos números reais, sendo subdivididas em intervalar e razão. A diferença entre as escalas intervalar e razão consiste no significado do valor zero, em que: (i) na escala intervalar, o valor zero é arbitrário, não indicando ausência da característica em estudo; (ii) na escala de razão, o valor zero indica ausência da característica em estudo.

Variável	Níveis	Subníveis	Exemplo
Qualitativa (Não-métrica)	Nominal		Fornecedor; etc.
	Ordinal		Grau de Satisfação; etc.
Quantitativa (Métrica)	Discreta		Nº de peças; etc.
	Contínua	Intervalar	Temperatura, etc.
		Razão	Tempo de vida útil, etc.

Figura 3: Categorias do termo fração ou razão
Fonte: Adaptado de Larson e Farber (2005) e Hair et al. (2006)

2.2 Modelos Lineares Generalizados

Segundo McCullagh e Nelder (1989), a proposta mais interessante e inovadora em modelagem foi apresentada por Nelder e Wedderburn em 1972, que utilizam a classe dos MLG's na modelagem da relação entre a variável resposta e as variáveis explicativas, abrindo o leque de opções para a distribuição de probabilidade da variável resposta, admitindo que a mesma pertença à família exponencial de distribuição. Esta família contempla as distribuições: Normal, Binomial, Poisson, Gama, Exponencial, entre outras. Ademais, estes modelos admitem variância não constante e dependente da média.

A forma das funções de distribuições que pertencem à família exponencial mais comumente vista é

$$f(y; \theta, \phi) = \exp\{[a(\phi)]^{-1}[y\theta - b(\theta)] + c(y; \phi)\} \quad (1)$$

onde $a(\cdot)$, $b(\cdot)$ e $c(\cdot)$ são funções conhecidas, θ é o parâmetro de localização e $\phi > 0$ é o parâmetro de dispersão da distribuição de probabilidade. O termo de dispersão ϕ define o formato (largo ou estreito) da curvatura da distribuição em relação ao parâmetro de localização θ (CORDEIRO, 1986; McCULLAGH e NELDER, 1989).

Qualquer processo industrial ou fenômeno aleatório cujas medições da variável resposta se distribuam segundo alguma das distribuições da família exponencial poderá ser modelada através da teoria de MLG. Os modelos de regressão pertencentes à classe dos MLG's apresentam uma estrutura com três componentes: a componente aleatória, a componente sistemática e a função de ligação. Sendo: (i) Componente aleatória – identifica a distribuição de probabilidade da variável resposta y pertencente à família exponencial; (ii) Componente sistemática – especifica a estrutura das variáveis explicativas, que é utilizada como preditor linear η ; (iii) Função de ligação – descreve a relação funcional entre a componente sistemática e o valor esperado da componente aleatória (a média μ da variável resposta y). A componente sistemática que compõem o modelo de regressão consiste na estrutura das variáveis explicativas como uma soma linear, definida por

$$\eta = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (2)$$

onde a função linear η é denominada de preditor linear, β s são coeficientes desconhecidos e k o número de variáveis explicativas ou fatores controláveis. Além disso, outra característica da componente sistemática de um MLG é que a média μ da variável resposta pode ser expressa por uma função conhecida $g(\cdot)$, denominada de função de ligação (MYERS et al., 2002), a qual descreve a relação funcional (ligação) entre a média μ e o preditor linear η , descrevendo a forma

$$g(\mu) = \eta = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (3)$$

Conforme Dobson (2002), a função de ligação $g(\cdot)$ é responsável pela transformação da média da variável resposta, e define a forma da relação entre as componentes aleatória e sistemática dos MLG's. Sant'Anna e Caten (2005) descrevem que a escolha da função de ligação depende do problema de modelagem em particular e os modelos de regressão podem apresentar uma função de ligação diferente conforme o conjunto de observações. Myers et al. (2002) ressaltam que, o uso da função de ligação na estrutura dos modelos contribui para a

estimação dos coeficientes desconhecidos e cálculos de intervalos de confiança. Entretanto, isso não garante a qualidade do ajuste do modelo aos dados.

A estimação dos coeficientes dos MLG's é realizada através do método clássico de máxima verossimilhança, em que os coeficientes β e ϕ são obtidos a partir da maximização do logaritmo da função de verossimilhança. O procedimento de maximizar é realizado por um algoritmo de otimização não-linear, tal como o algoritmo de Newton-Raphson, algoritmo Gauss-Newton ou algoritmo quasi-Newton (BFGS) (ver WEDDERBURN, 1974; CORDEIRO, 1986; McCULLAGH e NELDER, 1989).

Na seqüência são apresentados os principais modelos paramétricos contemplados no estudo e suas estruturas, conforme Figura 4.

2.2.1 Modelos Paramétricos Discretos

O modelo Poisson é empregado em estudos onde a variável resposta apresenta valores discretos (contagens), por exemplo, em acidentes de veículos com motoristas jovens (Lee et al., 2002), em infectados numa epidemia, em teorias de filas, dentre outras aplicações. O ajuste do modelo Poisson permite o uso das funções de ligação identidade, logaritmo e raiz quadrada ($g[\mu] = \eta = (\mu)^{1/2}$). A variância deste modelo é composta por um termo de dispersão e um termo dependente da média ($\text{Var}[y] = \phi\mu$) (McCULLAGH e NELDER, 1989).

O modelo Binomial Negativo pode ser utilizado em estudos similares nos quais o modelo Poisson é empregado, ou seja, quando a variável resposta são contagens. A vantagem deste modelo é investigar dados com presença de superdispersão, a qual consiste na variabilidade inter-unidades experimentais e variância maior que a média (HAMADA et al., 1997). O ajuste do modelo Binomial Negativo permite o uso das funções de ligação identidade, logaritmo e raiz quadrada ($g[\mu] = \eta = (\mu)^{1/2}$). A variância deste modelo é composta por um termo de dispersão e um termo dependente da média ($\text{Var}[y] = \mu + \mu^2/\phi$).

O modelo Logístico permite prever a relação entre uma variável resposta qualitativa e uma ou mais variáveis explicativas, que podem ser quantitativas ou qualitativas. No caso em que as variáveis qualitativas apresentem mais de dois níveis, sugere-se utilizar variáveis dummy. O nome deste modelo está relacionado a função de ligação que é utilizada na estrutura do modelo de regressão. Ou seja, a distribuição de probabilidade é Binomial e a função de ligação é a logística ($\ln[\mu/(1 - \mu)]$), também conhecida como “logit” (AGRESTI, 2007). Este modelo tem ampla utilização em estudos na área biomédica e mercado financeiro. O ajuste do modelo Logístico permite o uso das funções de ligação: “logit” ($g[\mu] = \eta =$

$\log(\mu/(1 - \mu))$) ou “Complemento loglog” ($g[\mu] = \eta = e^\mu / \{1 + e^\mu\}$), a fim de obter melhor ajuste aos dados. A variância deste modelo é composta por um termo de dispersão e um termo dependente da média ($\text{Var}[y] = \phi\mu\{1 - \mu\}$).

O modelo Probit pode ser utilizado nos mesmos estudos que o modelo Logístico, ou seja, quando existe uma variável resposta qualitativa e uma ou mais variáveis explicativas quantitativas. A restrição do modelo quanto às variáveis explicativas é devido a função de ligação “probit” ser uma função de distribuição normal acumulada $\Phi(\cdot)$, com propriedades inversa, monótona e diferenciável. O nome deste modelo está relacionado a função de ligação que é utilizada na estrutura do modelo de regressão (AGRESTI, 2007). O ajuste do modelo Probit permite o uso das funções de ligação: “probit” ($g[\mu] = \eta = \Phi^{-1}(\mu)$) ou “Complemento loglog” ($g[\mu] = \eta = e^\mu / \{1 + e^\mu\}$), a fim de obter melhor ajuste aos dados. A variância deste modelo é definida por $\text{Var}[y] = \mu\{1 - \mu\}/\phi$.

O modelo Log-linear pode ser utilizado em estudos similares nos quais os modelos Logístico e Poisson são empregados. No primeiro caso, em estudos com tabelas de contingência, a partir da contagem das observações nas categorias das variáveis resposta e explicativas ambas qualitativas. No segundo, em estudos que a variável resposta é quantitativa (discreta) e uma ou mais variáveis explicativas qualitativas (McCULLAGH e NELDER, 1989; AGRESTI, 2007). No caso em que as variáveis qualitativas apresentem mais de dois níveis, sugere-se utilizar variáveis *dummy*. O ajuste do modelo Log-linear permite o uso da função de ligação logaritmo. A variância deste modelo é composta por um termo de dispersão e um termo dependente da média ($\text{Var}[y] = \phi\mu$) (CORDEIRO, 1986).

2.2.2 Modelos Paramétricos Contínuos

O modelo Normal (ou Gaussiano) é utilizado na maioria dos processos de modelagem de fenômenos aleatórios e experimentos industriais, em virtude da distribuição de probabilidade Normal possuir importantes propriedades teóricas, consistência e eficiência. Este modelo é aplicado em estudos que apresentem variáveis respostas contínuas, com valores pertencentes ao conjunto dos números reais (\mathfrak{R}). Usualmente, o ajuste do modelo Normal requer o uso da função de ligação identidade, ou seja, $g[\mu] = \eta = \mu$, e variância constante ($\text{Var}[y] = \phi^2$). Conforme o problema de modelagem pode-se utilizar a função logaritmo ($g[\mu] = \eta = \log\mu$) ou “inversa” ($g[\mu] = \eta = 1/\mu$), a fim de obter melhor ajuste aos dados.

O modelo Log-normal é comumente utilizado em estudos econômicos e telecomunicações, pois as variáveis respostas nesta área são log-normalmente distribuídas. Este modelo apresenta estrutura similar ao modelo Normal, entretanto, possui flexibilidade de ajuste assimétrico aos dados. Ou seja, uma variável aleatória y segue uma distribuição log-normal, então $z^{-1} = \log(y)$ é normalmente distribuída, assim como, se z é normalmente distribuída, então $y = \exp(z)$ comporta-se como uma distribuição log-normal (JOHNSON et al, 1995a). O ajuste do modelo Log-Normal permite o uso das funções de ligação iguais ao modelo Normal, que são: identidade, logaritmo ou inversa.

Barndorff-Nielsen (1997) apresenta diversas aplicações do modelo Normal Inversa (ou Inversa Gaussiana) em estudos econométricos, a fim de modelar a volatilidade estocástica de retornos dos ativos das empresas inseridas no mercado acionário. Ademais, este modelo pode ser usado em estudos de confiabilidade, tempos de passeios aleatórios ou tempos de sobrevivência, em virtude da composição do termo de variabilidade da distribuição. O ajuste do modelo Normal Inversa permite o uso das funções de ligação iguais ao modelo Normal, que são: identidade, logaritmo, inversa ou μ^{-2} . A variância deste modelo é composta por um termo de dispersão e um termo dependente da média ($\text{Var}[y] = \phi\mu^3$) (McCULLAGH e NELDER, 1989).

O modelo Gama permite prever a relação entre uma variável resposta quantitativa contínua não-negativa, com uma longa assimetria à direita, e uma ou mais variáveis explicativas, que podem ser quantitativas ou qualitativas. Modelos Gama podem ser aplicados em estudos econométricos, como o modelo Normal Inversa. O ajuste do modelo Gama permite o uso das funções de ligação iguais ao modelo Normal, que são: identidade, logaritmo ou inversa. De maneira análoga, a escolha da função de ligação depende do processo de modelagem dos dados. A variância deste modelo é composta por um termo de dispersão e um termo dependente da média ($\text{Var}[y] = \phi^{-1}\mu^2$) (CORDEIRO, 1986).

Em virtude das propriedades da distribuição Weibull, o modelo Weibull é comumente utilizado em estudos de confiabilidade de equipamentos, tempo de vida útil e tempo de falha em diversas áreas do conhecimento. A distribuição de Weibull (2-parâmetros) é idêntica a distribuição Exponencial, quando o parâmetro de forma ($\nu = 1$) (JOHNSON et al, 1995a), e assim, esta pode ser descrita na forma da equação (1). O ajuste do modelo Weibull permite o uso das funções de ligação iguais ao modelo Normal, que são: identidade, logaritmo ou inversa. A variância deste modelo é composta por um termo de dispersão e um termo dependente da média ($\text{Var}[y] = \mu\phi - \mu^2$).

2.3 Extensão dos Modelos Paramétricos

O modelo de Quase-verossimilhança (MQV) permite prever a relação entre uma variável resposta qualitativa ou quantitativa e uma ou mais variáveis explicativas, que podem ser qualitativas ou quantitativas.

Weddeburn (1974) propôs os MQV's com base nos modelos de regressão pertencentes à classe dos MLG, em que os modelos apresentam apenas duas componentes: a componente sistemática (estrutura linear das variáveis explicativas) e a função de ligação, que relaciona a média μ da variável resposta à estrutura linear das variáveis explicativas.

A característica principal destes modelos de regressão, é que não há necessidade de assumir a princípio alguma distribuição de probabilidade para a variável resposta. Por conseguinte, a esperança matemática e a variância da variável resposta não são conhecidas a priori. Estes modelos podem adotar diversas funções de ligação, tais como: identidade, logaritmo, inversa, raiz quadrada ou potência, dentre outras; e de variância como: constant, $\mu(1 - \mu)$ ou μ^3 , dentre outras.

Assim, seja y uma variável resposta qualquer de interesse, que admite esperança e variância definidas por $E[y] = \mu$ e $\text{Var}[y] = \phi^*V(\mu)$, respectivamente. Em que a função de variância $V(\mu)$ é uma função dependente da média μ e ϕ é o parâmetro de dispersão. A função de quase-verossimilhança para um modelo de regressão é definida pela equação

$$Q(y; \mu) = \int_y^\mu \frac{y - t}{\phi \cdot V(t)} \partial t \quad (4)$$

Por exemplo, uma função de variância do tipo $\text{Var}[y_i] = \mu^2(1-\mu)^2$, apresenta o logaritmo da função de quase-verossimilhança da forma

$$Q(y; \mu) = \int_y^\mu \frac{y - \mu}{\phi \cdot \mu_i^2(1 - \mu_i)^2} \partial \mu \quad (5)$$

O modelo Beta permite prever a relação entre uma variável resposta quantitativa contínua restrita ao intervalo $[0,1]$ e uma ou mais variáveis explicativas através de uma estrutura de regressão. A família de distribuições Beta é composta de todas as distribuições de probabilidade que apresente uma variável resposta y cuja função densidade de probabilidade depende dos parâmetros θ_1 e θ_2 , e que sua função densidade pode ser escrita na forma,

$$f(y; \theta_1, \theta_2) = \frac{1}{B(\theta_1, \theta_2)} y^{\theta_1-1} (1 - y)^{\theta_2-1} \quad (6)$$

onde y é a variável aleatória, θ_1 e θ_2 , são parâmetros da função densidade de probabilidade e $B(\theta_1, \theta_2)$ uma função beta, θ_1 e $\theta_2 > 0$ (KIESCHNICK, R. e McCULLOGH, 2003).

A função $f(y; \theta_1, \theta_2)$ é efetivamente uma função densidade de probabilidade com parâmetros θ_1 e θ_2 . Nota-se ainda, que a função $f(y; \theta_1, \theta_2)$ assume valores estritamente positivos, pois para qualquer valor de y pertencente ao intervalo $[0,1]$, a função densidade descrita é crescente, ou seja, $f(y; \theta_1, \theta_2) \geq 0$.

Johnson et al. (1995b) apresentam diversos estudos em que a distribuição Beta apresenta melhor ajuste a distribuição dos dados em proporção que outras distribuições testadas. Segundo Kieschnick e McCulloch (2003), a distribuição de probabilidade Beta é uma função densidade de probabilidade que não pertence à família exponencial, ou seja, não pode ser escrita na forma da equação (1), pois a distribuição não pode ser escrita na forma canônica e apresentar um parâmetro de localização μ . A família de distribuições Beta contempla as distribuições Uniforme, Arco-seno e Dirichlet (JOHNSON et al, 1995b).

O ajuste do modelo Beta permite o uso das funções de ligação: “logit” ($g[\mu] = \eta = \log(\mu/(1 - \mu))$), “probit” ($g[\mu] = \eta = \Phi^{-1}(\mu)$) ou “Complemento loglog” ($g[\mu] = \eta = e^\mu / (1 + e^\mu)$), a fim de obter melhor ajuste aos dados.

Modelos Paramétrico	$a(\phi)$	θ	$b(\theta)$	$c(y; \phi)$	Função de Variância ($Var[y] = \phi V[\mu]$)	Função de Ligação ($g[\mu]$)
Modelo Normal	ϕ^2	μ	$\frac{\mu^2}{2}$	$-\frac{1}{2} \left(\ln 2\pi\phi^2 + \frac{y^2}{\phi^2} \right)$	ϕ^2	Identidade ($g[\mu] = \mu$) Logaritmo ($g[\mu] = \log(\mu)$) Inversa ($g[\mu] = 1/\mu$)
Modelo Log-Normal	ϕ^2	$\mu + \frac{\sigma^2}{2}$	$\frac{\mu^2}{2}$	$-\frac{1}{2} \left(\ln 2\pi\phi^2 + \frac{(\ln y)^2}{\phi^2} \right)$	ϕ^2	Identidade ($g[\mu] = \mu$) Logaritmo ($g[\mu] = \log(\mu)$) Inversa ($g[\mu] = 1/\mu$)
Modelo Normal Inversa	ϕ^2	$-\frac{1}{\mu^2}$	$\frac{1}{\mu}$	$-\frac{1}{2} \left(\ln 2\pi\phi^2 y^3 + \frac{1}{y\phi^2} \right)$	$\phi\mu^3$	Identidade ($g[\mu] = \mu$) Logaritmo ($g[\mu] = \log(\mu)$) Inversa ($g[\mu] = 1/\mu$) Potência ($g[\mu] = \mu^{-2}$)
Modelo Gama	ϕ^{-1}	$-\frac{1}{\mu}$	$-\ln\left(\frac{1}{\mu}\right)$	$-\ln\Gamma(\phi) + \phi \ln(\phi y) - \ln(y)$	$\phi^{-1}\mu^2$	Identidade ($g[\mu] = \mu$) Logaritmo ($g[\mu] = \log(\mu)$) Inversa ($g[\mu] = 1/\mu$)
Modelo Weibull	ϕ^{-1}	$\ln \frac{y}{\mu}$	$-\ln(\mu)$	$-\ln\left(\frac{y}{\mu}\right) - \left(\frac{y}{\mu}\right)^\phi$	$\mu\phi - \mu^2$	Identidade ($g[\mu] = \mu$) Logaritmo ($g[\mu] = \log(\mu)$) Inversa ($g[\mu] = 1/\mu$)
Modelo Poisson	1	$\ln(\mu)$	$-\mu$	$\ln(y!)$	$\phi\mu$	Identidade ($g[\mu] = \mu$) Logaritmo ($g[\mu] = \log(\mu)$) Raiz quadrada ($g[\mu] = \mu^{1/2}$)
Modelo Binomial Negativa	$\phi \ln(\phi)$	$\ln(\mu)$	$-(\phi + y)\ln(\mu + \phi)$	$\ln\left(\frac{\Gamma(\phi + y)}{\Gamma(\phi)y!}\right)$	$\mu + \frac{\mu^2}{\phi}$	Identidade ($g[\mu] = \mu$) Logaritmo ($g[\mu] = \log(\mu)$) Raiz quadrada ($g[\mu] = \mu^{1/2}$)
Modelo Log-Linear	1	$\ln(\mu)$	$-\mu$	$\ln(y!)$	$\phi\mu$	Identidade ($g[\mu] = \mu$) Logaritmo ($g[\mu] = \log(\mu)$) Raiz quadrada ($g[\mu] = \mu^{1/2}$)
Modelo Logístico	1	$\ln\left(\frac{\mu}{1-\mu}\right)$	$n \cdot \ln(1-\mu)$	$\ln\left(\frac{n}{y}\right)$	$\phi\mu(1-\mu)$	Logit ($g[\mu] = \log(\mu/(1-\mu))$) Cloglog ($g[\mu] = e^\mu / (1 + e^\mu)$)
Modelo Probit	1	$\ln\left(\frac{\mu}{1-\mu}\right)$	$n \cdot \ln(1-\mu)$	$\ln\left(\frac{n}{y}\right)$	$\phi\mu(1-\mu)$	Probit ($g[\mu] = \Phi^{-1}(\mu)$) Cloglog ($g[\mu] = e^\mu / (1 + e^\mu)$)
Modelo Quase-verossimilhança	---	---	---	---	$\phi V(\mu)$	Diversas
Modelo Beta					$\mu(1-\mu)\phi^{-1}$	Logit ($g[\mu] = \log(\mu/(1-\mu))$) Probit ($g[\mu] = \Phi^{-1}(\mu)$) Cloglog ($g[\mu] = e^\mu / (1 + e^\mu)$)

Figura 4: Estrutura dos principais modelos de regressão paramétricos

Fonte: Adaptado de Nelder e Wedderburn (1972), Cordeiro (1986), McCullagh e Nelder (1989) e Agresti (2007)

2.3 Critérios de Diagnóstico de Modelos

O desempenho de um modelo de regressão é avaliado pela sua qualidade de ajuste e definida a partir dos próprios dados utilizados na determinação do modelo. Modelos de regressão com boa adequabilidade apresentam pequena discrepância entre os dados reais e seus respectivos valores preditos. Ademais, conforme o estudo pode haver mais do que um modelo de regressão a ser utilizado para o mesmo conjunto de dados em estudo, entretanto, os critérios de diagnóstico serão os elementos de decisão para a escolha do melhor modelo.

Os critérios de diagnóstico utilizados neste artigo para análise foram *Deviance*, resíduos padronizados, critério de informação de Akaike e Envelope simulado.

- *Deviance*: A análise da *deviance* é feita através da comparação dos valores da medida *deviance* dos modelos ajustados. Segundo Atkinson e Riani (2000), a *deviance* é obtida como duas vezes a diferença entre o máximo do logaritmo da verossimilhança do modelo nulo e do modelo sob pesquisa,

$$D(y; -\hat{\mu}) = \sum_{i=1}^n 2[l_i(\tilde{\mu}, \phi) - l_i(\hat{\mu}, \phi)] \quad (7)$$

em que $\tilde{\mu}$ é solução de $\partial l_i / \partial \mu_i = 0$, isto é, $\phi(y_i^*, \mu_i^*) = 0$, $l_i(\tilde{\mu}, \phi)$ é a função de máxima verossimilhança do modelo sob pesquisa e $l_i(\mu, \phi)$ é a função de máxima verossimilhança do modelo nulo. Conforme Lee e Nelder (1998), usualmente costuma-se proceder a análise de deviance utilizando o ponto crítico $\chi^2(n-k)(\alpha)$ da distribuição qui-quadrado ao nível de significância igual a α , sendo n o número de observações e k o número de coeficientes do modelo.

Assim, se $D(y; \mu, \phi) \leq \chi^2_{(n-k)}(\alpha)$, pode-se considerar que há evidências de que o modelo sob pesquisa esteja bem ajustado aos dados, a um nível de α de significância, usualmente $\alpha < 0,05$, caso contrário deve-se descartar o modelo, pois o mesmo pode ser considerado inadequado. Esta estatística de decisão também pode ser utilizada para a função de quasi-deviance na adequação do modelo de Quase-verossimilhança (HURVICH e TSAI, 1995).

- *Critério de Informação de Akaike*: O AIC (Akaike Information Criterion) foi o primeiro critério baseado na informação de Kullback-Leibler (K-L) e assintoticamente não viesado para K-L. O critério AIC supõe que o modelo verdadeiro pertence ao conjunto de modelos candidatos e é definido por

$$AIC = -2l(\hat{\mu}, \phi) + 2(k + 1) \quad (8)$$

onde $l_i(\hat{\mu}, \hat{\phi})$ é a função de máxima verossimilhança do modelo ajustado e k o número de parâmetro do modelo. Segundo Rao e Wu (2005), o critério AIC foi desenvolvido através dos estimadores de máxima verossimilhança (EMV), para decidir qual o modelo mais adequado quando se utiliza muitos modelos com quantidades diferentes de coeficientes. A decisão quanto ao melhor modelo ajustado é realizado escolhendo o menor valor de AIC.

- *Resíduos padronizados de Pearson*: são os resíduos mais utilizados em análise gráfica de diagnóstico por se aproximarem da distribuição de probabilidade Normal através da padronização com a média e desvio-padrão. A análise destes resíduos permite fácil verificação da adequação do modelo aos dados, da aleatoriedade dos resíduos e de possíveis valores discrepantes. Para cada observação (i) da variável resposta y_i , pode-se definir o resíduo ordinário padronizado como

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\sigma}^2}} \quad (9)$$

Um gráfico desses resíduos contra a ordem das observações (i) não deveria mostrar nenhuma tendência, mas uma aleatoriedade. Através desse gráfico é possível verificar se existem pontos suspeitos de serem atípicos, ou seja, maiores que três desvios (RAO e WU, 2005).

- *Envelope simulado meio-Normal*: é uma ferramenta gráfica de diagnóstico muito útil em modelos lineares e não-lineares. A proposta desta medida é acrescentar ao gráfico de probabilidade normal usual um envelope simulado que pode ser usado para decidir se as observações são consistentes com o modelo ajustado.

Este gráfico é construído a partir da simulação de k valores (estatísticas de ordem) para cada valor previsto pelo modelo ajustado e gerado médias, valores mínimos e máximos de cada valor previsto. Esses valores mínimos e máximos das k estatísticas de ordem produzem o envelope. Desta forma, o gráfico apresentará um intervalo para cada valor previsto ordenadamente contra os escores meio-normais

$$\Phi^{-1} \left\{ \frac{(i + n - \frac{1}{8})}{2n + \frac{1}{2}} \right\} \quad (10)$$

onde $\Phi(\cdot)$ é a função de densidade acumulada da distribuição normal padrão e n é o número de observações. Segundo Atkinson e Riani (2000), caso ocorram tendências não aleatórias dos resíduos absolutos dentro do envelope simulado há indícios de escolha incorreta da distribuição de probabilidade para a variável resposta ou da função de ligação.

3. Sistemática para seleção de modelos paramétricos

Esta seção apresenta a sistemática para seleção dos principais modelos paramétricos, considerando a classificação das variáveis resposta e explicativas objetivando contribuir para uma adequada modelagem de dados e otimização experimental.

Segundo Sant'Anna (2006), qualquer parâmetro ou característica de qualidade de um produto ou processo que se interessa investigar, apresenta uma classificação correspondente a natureza de mensuração ou a forma como é mensurada. Um dos principais intuits em diferenciar os modelos de regressão, conforme a classificação das variáveis que se inserem no processo de pesquisa, está na busca pela eficiência da aplicação.

O sucesso de qualquer análise estatística depende, muitas vezes, da disponibilidade do modelo matemático se incorporar “corretamente” ao conjunto de dados. No caso de um ou mais modelos se ajustarem aos dados, surge o problema da escolha do “melhor modelo”. Esta escolha é realizada fundamentada nos critérios de diagnósticos de modelos de regressão lineares.

A sistemática proposta foi construída a partir de uma ampla revisão na literatura contemplando: (i) a classificação de variáveis em utilizadas nas áreas da Administração, Ciências Médicas, Economia, Engenharia e Estatística, com objetivo de estudar as nomenclaturas empregadas e criar uma padronização; (ii) os principais modelos de regressão paramétricos, suas estruturas, características e propriedades e (iii) e os critérios de diagnóstico empregados na análise de adequação de modelos de regressão. Os materiais bibliográficos pesquisados para desenvolvimento da sistemática proposta foram Nelder e Wedderburn (1972), Wedderburn (1974), Cordeiro (1986), McCullagh e Nelder (1989), Lee e Nelder (1998), Atkinson e Riani (2000), Myers et al. (2002) e Rao e Wu (2005).

O processo de análise experimental utilizando a sistemática proposta consiste nas seguintes etapas: (i) Classificação das variáveis envolvidas no experimento; (ii) Utilização da sistemática para seleção dos modelos paramétricos; (iii) Modelagem dos dados experimentais; (iv) Aplicação dos critérios de diagnóstico e (v) Definição do melhor ajuste dos fatores controláveis.

A Figura 5 apresenta a sistemática para seleção de modelos paramétricos, considerando a classificação das variáveis resposta e explicativas.

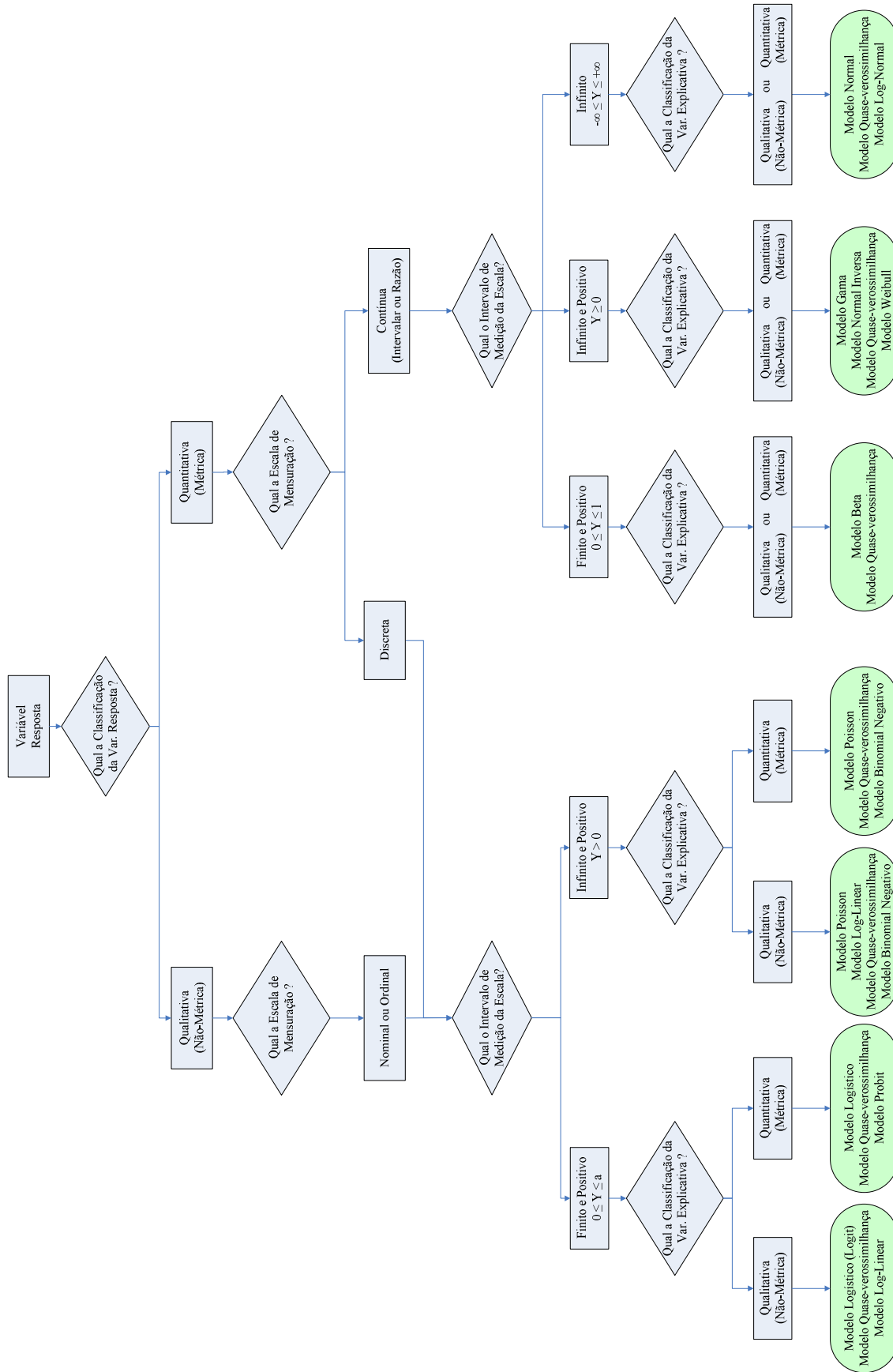


Figura 5: Sistemática para seleção de modelos paramétricos
 Fonte: Elaborado pelo autor

4. Aplicação da Sistemática de Seleção

A sistemática para seleção de modelos de regressão foi aplicado a um estudo experimental sobre testes de detonação com explosivos em mina, onde se buscou avaliar a influência de certos fatores na velocidade de partícula na vibração captada por um sismógrafo. De posse de informações técnicas elaborou-se um projeto fatorial 2^{k-1} fracionado onde um conjunto de fatores ($k = 6$) foi ensaiado a dois níveis sem replicações, sendo utilizada apenas a fração principal dos ensaios ($2^{k-1} = 2^{6-1} = 32$), pois a realização de todos os ensaios do delineamento no cenário de detonação resulta em perda de recursos, tempo e objetividade experimental.

4.1 Classificação das Variáveis

A variável resposta é a velocidade de partícula na vibração captada por um sismógrafo. As variáveis explicativas ou fatores controláveis testados a dois níveis no experimento foram:

1. Distância – a distância entre o local da detonação e sismógrafo, com os níveis 25m e 50m;
2. Carga – quantidade de explosivos em gramas, contemplando os níveis 1620g e 2720g;
3. Geologia – a camada da mina na qual é realizada a detonação, contemplando os níveis Siltito e Carvão;
4. Tipos de Explosivo – característica do material detonado (explosivo), contemplando os níveis ANWO e Emulsão;
5. Tipos de Furo – quantidade formato do local para colocação da carga para detonação, considerando os retardos para auxílio na detonação, com os níveis único e múltiplo;
6. Profundidade de Carga – profundidade do local para colocação da carga para detonação, com os níveis 1,10m e 1,70m.

4.2 Escolha dos modelos paramétricos

A escolha de um modelo de regressão adequada é realizada a partir do uso do fluxograma, onde se definiu como variável resposta “Velocidade de partícula na vibração captada por um sismógrafo”, e como variáveis explicativas, os fatores controláveis: distância, carga, geologia, tipo de explosivo, tipo de furo e profundidade de carga.

Os passos de orientação conforme a Figura 5, são:

1. Qual a Classificação da variável resposta?

A “Velocidade de partícula na vibração” é uma *variável Quantitativa*.

2. Qual a Escala de mensuração da variável resposta?

A “Velocidade de partícula na vibração” é uma *variável Quantitativa Contínua*.

3. Qual o Intervalo de medição da Escala?

A “Velocidade de partícula na vibração” está compreendida no intervalo $[-\infty, +\infty]$.

4. Qual a Classificação das variáveis explicativas?

As variáveis explicativas investigadas são *Variáveis Qualitativas*.

5. Quais os modelos sugeridos?

Os modelos escolhidos são: *Modelo Normal*, *Modelo Log-Normal* e *Modelo de Quase-verossimilhança*.

Pode-se notar que a sistemática de seleção apresenta três opções de modelos que podem ser usados de forma adequada na modelagem dos dados. A partir deste ponto, o processo de modelagem se inicia com certa confiança na estimação, análise e previsão de dados. Quanto à diferenciação entre estes três modelos, cabe verificar a melhor adequação dos modelos ao conjunto de dados considerando os critérios de diagnóstico de modelos.

4.3 Modelagem dos dados experimentais e Critérios de diagnóstico

A velocidade de partícula na vibração captada por um sismógrafo *Whiteseis*® após detonação por explosivos foi estimada segundo três modelos de regressão. A significância das estimativas dos coeficientes foi analisada pelo teste *Wald*. A análise da qualidade do ajuste foi baseada nos critérios de diagnóstico: Deviance, AIC, Resíduos padronizados e Envelope simulado.

O plano experimental foi constituído por 32 ensaios, coletadas à medida que ocorria a detonação por explosivos na mina, sem réplicas. A velocidade média de partícula após detonação foi de 38,44 e desvio-padrão de 27,70 m/s, com variação entre 2,82 e 94,20 m/s.

No primeiro estudo se utilizou um modelo Normal com método de estimação por máxima verossimilhança, função de ligação identidade e função de variância $V(\mu) = constante$. No segundo estudo, utilizou um modelo Log-normal com método de estimação por máxima verossimilhança, função de ligação identidade e função de variância $V(\mu) = constante$. No terceiro, um modelo linear considerando o método de estimação por máxima Quase-verossimilhança, função de ligação log e função de variância $V(\mu) = constante$.

A Tabela 1 mostra a qualidade de ajuste dos modelos de regressão, contendo: as estimativas dos coeficientes dos parâmetros ao nível de significância 5% e 10%, o valor do parâmetro de dispersão (ϕ), da *Deviance* e do AIC.

O teste de significância para os coeficientes dos fatores controláveis nos três modelos finais mostrou que o fator controlável “profundidade” não apresenta influência significativa na velocidade de partícula independente dos níveis. Para os outros fatores controláveis e as interações de segunda ordem, os três modelos de regressão propostos apresentaram diferenças no teste de significância.

O modelo Normal ajustado considerou os efeitos principais (distância, carga, geologia, explosivo e furo) e as interações de segunda ordem (distância \times carga, distância \times furo, distância \times geologia e geologia \times profundidade) como fatores significativos. Analisando a qualidade do ajuste do modelo, com base no critério de informação de Akaike (AIC = 224,95) e a *Deviance* = 1424,8 (19 graus de liberdade) que corresponde a um valor $p < 0,01$. Indicando um ajuste adequado aos dados.

O modelo Log-normal ajustado apresentou os efeitos principais (distância, carga, geologia e furo) e as interações de segunda ordem (distância \times geologia e geologia \times profundidade) foram significativas. Analisando a qualidade do ajuste do modelo LogNormal com AIC = 17,65 e a *Deviance* = 1,736 (22 graus de liberdade) que corresponde a um valor $p = 1,00$, indicando que o modelo não está com o ajuste adequado. Estes critérios indicam que a estrutura do modelo de regressão utilizado não está compatível aos dados, possivelmente, em virtude do desempenho da função densidade de probabilidade.

O modelo Quase-verossimilhança ajustado considerou os efeitos principais (distância, carga, geologia, explosivo e furo) e as interações de segunda ordem (distância \times furo, distância \times geologia e geologia \times profundidade) significativas. Analisando a qualidade do ajuste do modelo com AIC = 216,67 e a *Deviance* = 1155,6 (20 graus de liberdade) que corresponde a um valor $p < 0,01$, demonstrando que o modelo está bem ajustado.

Tabela 1: Estimativas dos coeficientes e critérios de diagnóstico dos modelos ajustados.

	Modelo Normal	Modelo LogNormal	Modelo Quase-verossimilhança
Função de Ligação	<i>Identidade</i>	<i>Identidade</i>	<i>Log</i>
Método de Estimação^Δ	<i>MV</i>	<i>MV</i>	<i>MQV</i>
Parâmetro	Estimativa	Estimativa	Estimativa
Intercepto	42,799** (1,6240)	3,375** (0,0523)	3,399** (0,1024)
Distância	-19,244** (1,6240)	-0,622** (0,0523)	-0,629** (0,1017)
Carga	7,104** (1,5890)	0,202** (0,0510)	0,168** (0,0320)
Geologia	-21,298** (1,6240)	-0,701** (0,0523)	-0,681** (0,1019)
Explosivo	-2,759* (1,5890)	-0,076 (0,0498)	-0,098** (0,0314)
Furo	3,836** (1,6240)	0,157** (0,0525)	0,070* (0,0380)
Profundidade	-1,080 (1,6240)	-0,015 (0,0509)	-0,009 (0,0410)
Distância*Carga	-3,974** (1,589)	-0,056 (0,0496)	-0,034 (0,0382)
Distância*Furo	-3,265** (1,6240)	-0,002 (0,0520)	-0,081** (0,0372)
Distância*Geologia	4,578** (1,6240)	-0,211** (0,0523)	-0,209** (0,1019)
Geologia*Profundidade	3,064* (1,6240)	0,094* (0,0523)	0,083** (0,0402)
Dispersão (ϕ)	74,991	0,079	57,778
<i>Deviance</i>	1424,8 (19 gl)	1,736 (22 gl)	1155,6 (20 gl)
AIC	224,95	17,65	216,67

^Δ MV, máxima verossimilhança; MQV, máxima quase-verossimilhança.

Erro padrão das estimativas são apresentados entre parênteses.

Nível de significância "Teste *Wald*": ($p < 0,05$)** e ($p < 0,10$)*

gl (graus de liberdade)

Fonte: Elaborado pelo autor (2009)

As Figuras 6(a) e (b) apresentam os gráficos de resíduos padronizados para os modelos Normal e Quase-verossimilhança melhor ajustados aos dados. Observa-se que os resíduos apresentam um comportamento aleatório e não apresentam valores residuais maiores que três desvios, para os dois modelos ajustados.

A Figura 7 apresenta os gráficos de envelope simulado meio-normal produzidos pelos modelos Normal e Quase-verossimilhança ajustados. Nota-se na Figura 7(a) que alguns resíduos se encontram próximos dos limites e dois resíduos fora dos limites do envelope, evidenciando uma possível falha na definição da distribuição de probabilidade ou da função de ligação, demonstrando que a dispersão não está controlada.

No envelope simulado meio-normal gerado pelo modelo de Quase-verossimilhança os resíduos se encontram dentro dos limites do envelope, evidenciando que a função de ligação “log” utilizada se adequou bem aos dados (ver Figura 7(b)).

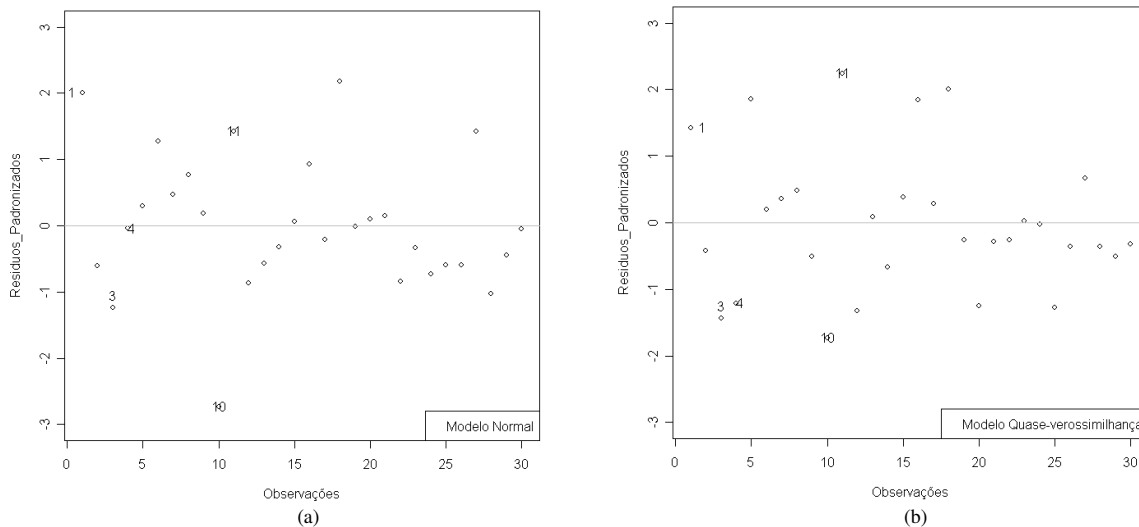


Figura 6: Gráficos de resíduos padronizados dos modelos Normal (a) e Quase-verossimilhança (b)
Fonte: Elaborado pelo autor (2009)

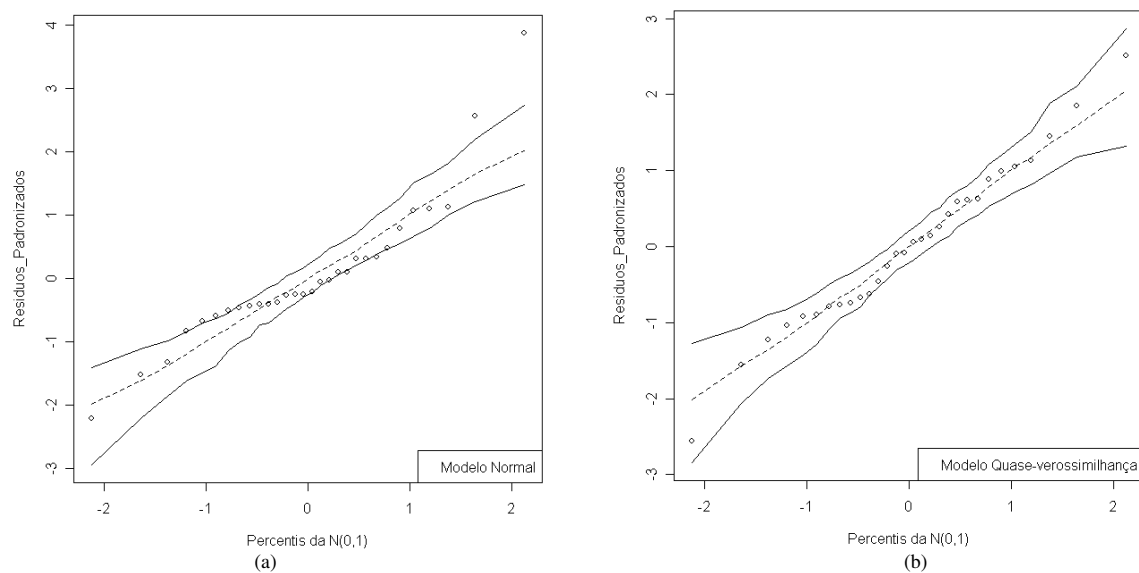


Figura 7: Gráficos de envelope simulado com o ajuste dos modelos Normal (a) e Quase-verossimilhança (b)
Fonte: Elaborado pelo autor (2009)

4.4 Definição do melhor ajuste dos Fatores Controláveis

No estudo de caso, a sistemática de seleção direcionou para três modelos de regressão paramétricos, e baseado nos critérios de diagnóstico utilizados, o modelo de Quase-verossimilhança foi o melhor modelo dentre eles, pois apresentou melhor adequação aos dados.

Na análise das interações de segunda ordem distância \times carga, distância \times furo e distância \times geologia, observou-se que há uma redução na velocidade de partícula na vibração quando a variável explicativa “distância” está no nível alto (50m), independente dos níveis das variáveis explicativas carga, tipo de furo e geologia.

Com base no melhor modelo de regressão ajustado, o aumento da velocidade de partícula na vibração após detonação por explosivos está relacionado com a menor distância, maior volume de carga, múltiplos furos na área, agente explosivo emulsão e uma geologia menos densa, como carvão.

5. Considerações Finais

Este artigo propôs apresentar uma sistemática para seleção dos principais modelos paramétricos, considerando a classificação das variáveis resposta e explicativas e aplicar a um caso real. A análise detalhada em relação as variáveis envolvidas no experimento é importante no momento da modelagem dos dados e otimização experimental.

Conforme foi discutido, cada modelo de regressão tem suas características relacionadas às variáveis resposta e explicativas, e a especificação incorreta do modelo de regressão ou de componentes do modelo pode produzir inferências inadequadas e resultados experimentais errôneos. Portanto, uma orientação adequada e os critérios de diagnóstico contribuem para a decisão final do melhor modelo.

A sistemática desenvolvida forneceu uma orientação adequada, consistente e usabilidade satisfatória na investigação do experimento em detonação por explosivos, conduzindo a resultados seguros. Ademais, a sistemática destaca-se como uma estratégia relevante e adicional para pesquisas em modelagem de dados e otimização experimental.

6. Referências

- AGRESTI, A. *An Introduction to Categorical Data Analysis*, 2ª ed., New York: John Wiley & Sons, 2007.
- ATKINSON, A.C.; RIANI, M. *Robust Diagnostic Regression Analysis*, New York: Springer, 2000.

- BARNDORFF-NIELSEN, O.E. Normal Inverse Gaussian distribution and Stochastic volatility modelling, *Scandinavian Journal of Statistics*, v. 24, pp. 1–13, 1997.
- CORDEIRO, G.M. *Modelos Lineares Generalizados*, In: VII Simpósio Nacional de Probabilidade e Estatística, Campinas, São Paulo: ABE, 286 p., 1986.
- DOBSON, A.J. *An Introduction to Generalized Linear Models*, 3^a ed., London: Chapman & Hall, 2002.
- HAIR, J.F JR; BLACK, B.; ANDERSON, R.E; BABIN, B.; TATHAM, R.L. *Multivariate Data Analysis*, 6^a ed., New Jersey: Prentice-Hall Inc, 2006.
- HAMADA, M.; NELDER, J.A. Generalized linear models for quality-improvement experiments, *Journal of Quality Technology*, v. 29, pp. 292–304, 1997.
- HURVICH, C.M.; TSAI, C-L. Model selection for Quasi-likelihood models in Small Samples, *Biometrics*, v. 51, pp. 1077–1084, 1995.
- JOHNSON, N.L.; KOTZ, S.; BALAKRISHNAN, N. *Continuous Univariate Distributions*, vol. 1, New York: John Wiley & Sons, 1994.
- JOHNSON, N.L.; KOTZ, S.; BALAKRISHNAN, N. *Continuous Univariate Distributions*, vol. 2, New York: John Wiley & Sons, 1995.
- KIESCHNICK, R.; McCULLOUGH, B.D. Regression analysis of variates observed on [0,1]: percentages, proportions and fractions, *Statistical Modelling*, v. 3, pp. 193–213, 2003.
- LARSON, R.; FARBER, B. *Elementary Statistics*, 3^a ed., New Jersey: Prentice-Hall Inc, 2005.
- LEE, A.H.; STEVENSON, M.R.; WANG, K.; YAU, K.K.W. Modeling young driver motor vehicle crashes: data with extra zeros. *Accident Analysis and Prevention*, v. 34, n. 4, p.515-521, 2002.
- LEE, Y.; NELDER, J.A. Generalized linear models for the analysis of quality improvement experiments, *The Canadian Journal of Statistics*, v. 26, pp. 95–105, 1998.
- McCULLAGH, P.; NELDER, J.A. *Generalized Linear Models*, 2^a ed., London: Chapman & Hall, 1989.
- MONTGOMERY, D.C. *Design of Analysis of Experiments*, 6^a ed., New York: John Wiley & Sons, 2004.
- MONTGOMERY, D.C.; PECK, E.A.; VINING, G.G. *Introduction to Linear Regression Analysis*, 4^a ed., New York: John Wiley & Sons, 2006.
- MYERS, R.H.; MONTGOMERY, D.C.; VINING, G.G. *Generalized Linear Models with applications in Engineering and the Sciences*, New York: John Wiley & Sons, 2002.
- NELDER, J.A.; WEDDERBURN, R.W.M. Generalized Linear Models, *Journal of the Royal Statistical Society A*, v. 135, pp. 370–384, 1972.
- RAO, C.R.; WU, Y. Linear model selection by cross-validation, *Journal Statistical Planning and Inference*, v. 128, pp. 231–240, 2005.
- SANT'ANNA, A.M.O e CATEN, C.S. (2005) Modelagem da proporção de defeituosos usando Modelo de Quase-verossimilhança. XXV *ENEGEP*, Porto Alegre-RS
- SANT'ANNA, A.M.O. *Método de orientação à modelagem de dados mensurados em proporção*, Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil, 2006.
- WEDDERBURN, R.W.M. Quasi-likelihood functions, generalized linear models and the Gauss-Newton method, *Biometrika*, v. 61, pp. 439–447, 1974.

2.2 ARTIGO 2

Modelagem da Fração Não conforme em Processos Industriais

Modelling of the Fraction nonconforming in Industrial Processes

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

Enviado à Revista Pesquisa Operacional (ISSN 0101-7438) em Setembro 2008

Aceito na Revista Pesquisa Operacional (ISSN 0101-7438) em Junho de 2009

MODELAGEM DA FRAÇÃO NÃO CONFORME EM PROCESSOS INDUSTRIAIS

Ângelo Márcio Oliveira Sant'Anna

Carla Schwengber ten Caten

Programa de Pós-graduação em Engenharia de Produção

Universidade Federal do Rio Grande do Sul

Porto Alegre – RS

angelo@producao.ufrgs.br, tencaten@producao.ufrgs.br

Resumo

Em qualquer processo industrial, pode ser definido um conjunto de causas ou fatores que produzem determinado efeito sobre uma ou mais características de qualidade de um produto que pode ou não satisfazer às especificações do cliente, gerando a produção de produtos não conformes. A modelagem da fração de produtos não conformes utilizando-se o modelo de regressão linear não é adequada por pelo menos duas razões: (i) pressupõe que as proporções seguem a distribuição normal, que não é correto; e (ii) possibilita a previsão de valores fora do intervalo $[0,1]$. Alternativas à modelagem da fração de não conformes são os Modelos Lineares Generalizados e os Modelos de Regressão Beta. O objetivo deste artigo é modelar a fração não conforme às especificações de uma indústria curtidora com enfoque nos Modelos de Regressão Beta e Modelo Linear Generalizado. Esses modelos podem ser estendidos a processos industriais que envolvam a produção de produtos não conformes às especificações de manufatura.

Palavras-chave: Modelo linear generalizado; modelo regressão Beta; fração não conforme.

MODELLING OF THE FRACTION NONCONFORMING IN INDUSTRIAL PROCESSES

Abstract

In any industrial process, one can enumerate causes or factors that act on one or more quality characteristics of the resulting product such that they fail to meet customers' specifications, generating items deemed as nonconforming. Modeling the fraction or proportion of nonconforming items using linear regression models is not adequate for at least two reasons: (i) proportions are assumed to follow a Normal distribution, which is not correct, and (ii) predicted responses will not necessarily be confined in the $[0,1]$ -interval. Alternative approaches to the modeling of nonconforming proportions are based on Generalized Linear Models and Beta Regression Models. In this paper we present a case study where the objective is to model the nonconforming fraction of items emerging from a tanning process; our analysis uses Generalized Linear Models and Beta Regression Models. The developments presented in the paper may be extended to other industrial processes where the proportion of nonconforming items is available.

Keyword: Generalized linear model, Beta regression model; fraction nonconforming.

1. Introdução

O cenário mundial atual é de intensa competitividade devido ao crescente desenvolvimento tecnológico visando fabricar produtos que tenham características de qualidade conforme às especificações. Em processos de manufatura, a implementação de técnicas estatísticas permite eliminar desperdícios, reduzir os índices de produtos refugados, diminuir a necessidade da realização de inspeção e aumentar a satisfação dos clientes.

Em um processo industrial, pode ser definido um conjunto de causas ou fatores que produzem determinado efeito sobre uma ou mais características de qualidade de produtos que podem ou não satisfazer às especificações preestabelecidas pelo cliente. Muitas vezes não se consegue controlar todas as causas de variação, pois certas causas são inerentes ao processo (Montgomery, 2005). As causas de variação que podem ser controladas, também chamadas de fatores controláveis, podem interferir em um processo gerando a produção de produtos com características de qualidade não conformes às especificações preestabelecidas, os quais podem ser mensurados através da fração de produtos não conformes.

Segundo Montgomery et al. (2006), modelos de regressão consistem numa técnica estatística de modelagem e investigação que relaciona uma ou mais características de qualidade do produto, chamadas de variável dependente, com os fatores controláveis que podem afetá-las, chamados de variáveis independentes. Conforme Hamada e Nelder (1997), um modelo de regressão que apresenta um bom ajuste usualmente permite gerar boas estimativas dos efeitos dos fatores, consistindo numa estratégia eficiente de otimização, pois é possível prever a variável dependente em função do ajuste das variáveis independentes ou fatores controláveis.

Segundo Cox (1996), a modelagem da proporção em um determinado conjunto de observações, por meio de um modelo de regressão linear, nem sempre é recomendada, uma vez que este modelo requer a suposição de que as proporções seguem a distribuição normal. Segundo Kieschnick e McCulloch (2003), o uso do modelo de regressão linear na modelagem de proporções ou frações como variável dependente, é um modelo falho, pois possibilita a previsão de valores fora do limite do intervalo $[0,1]$. Alternativas para a modelagem da fração de produtos não conformes são os Modelos Lineares Generalizados (MLG) e os Modelos de Regressão Beta (MRB).

Conforme Myers et al. (2002), a teoria dos MLG apresenta opções para a distribuição da variável dependente, permitindo que dados provenientes de uma distribuição de

probabilidade Binomial possam ser modelados usando a distribuição original dos dados, sem a necessidade de realizar transformações nos dados.

Outra forma de relacionar a variável dependente e demais variáveis independentes, num processo de investigação e modelagem de dados, foi proposta por Ferrari e Cribari-Neto (2004), cuja estrutura do modelo de regressão baseia-se na suposição de que os dados mensurados em proporção seguem a distribuição de probabilidade Beta. Este procedimento é chamado de MRB.

O uso destes modelos de regressão permite aumentar a precisão da estimativa dos efeitos dos fatores controláveis e da previsão dos efeitos. Desta forma, permite-se identificar com mais precisão as condições ótimas de operação do processo de manufatura.

O objetivo deste artigo é modelar a fração não conforme às especificações de uma indústria curtidora com enfoque no Modelo de Regressão Beta e no Modelo Linear Generalizado.

Na seção 2, é apresentada uma breve revisão sobre os conceitos básicos de MLG e MRB. Na seção 3, são apresentadas as medidas de diagnóstico para análise de adequação dos modelos de regressão. Na seção 4, um estudo de aplicação prática. Finalmente, na seção 5, são feitas considerações finais sobre a modelagem desenvolvida nesse artigo.

2. Modelos de Regressão para Fração Não Conforme

Em muitas situações práticas em que se deseja realizar uma investigação entre uma variável dependente e demais variável independente, cuja variável dependente apresenta restrição nos valores mensurados como a fração de algum evento de interesse, é comum usar no processo de modelagem, o modelo de regressão linear. Contudo, segundo Cox (1996), a modelagem da proporção utilizando um modelo de regressão linear nem sempre é recomendada, pois este modelo requer a suposição de normalidade aos dados e homoscedasticidade (variâncias iguais). Assim, pelo fato dos dados serem mensurados em proporção dificilmente apresentarão normalidade. Portanto, deve-se buscar uma nova forma de relacionar as variáveis independentes à variável dependente.

Os modelos de regressão abordados como alternativas à modelagem da fração não conforme são: os modelos lineares generalizados, que foram propostos por Nelder e Wedderburn (1972), mais especificamente o modelo de Quase-verossimilhança e o modelo de regressão Beta, que foi proposto por Ferrari e Cribari-Neto (2004).

2.1. Modelos Lineares Generalizados

Uma classe de modelos conhecidos como MLG é apropriada para investigar o efeito de variáveis independentes sobre uma única variável dependente de comportamento não-normal. Estes modelos permitem estimar os efeitos relacionados com cada fator, analisar a influência e realizar previsões. Ademais, na construção destes modelos as variáveis independentes ou fatores podem ser de natureza quantitativa ou qualitativa.

Segundo Hamada e Nelder (1997), a classe dos MLG foi desenvolvida por Nelder e Wedderburn em 1972, e estes modelos se baseiam em distribuições de probabilidade, com um parâmetro de localização desconhecido (θ), admitindo que a mesma pertença à família exponencial. Esta família contempla as distribuições Normal, Binomial, Poisson, Gamma, Exponencial, Binomial Negativa. Ademais, estes modelos admitem variância não constante e dependente da média. Conforme Lee e Nelder (1998), esta classe de modelos é definida ainda por um conjunto de variáveis independentes que descreve a estrutura linear do modelo e uma função de ligação entre a média da variável dependente e a estrutura linear.

Para se admitir que uma distribuição de probabilidade pertença à família de distribuições exponencial é necessário que as funções de distribuições de probabilidade apresentem a forma

$$f(y; \theta) = \exp\{a(y)b(\theta) + c(\theta) + d(y)\} \quad (11)$$

onde $a(y)$, $b(\theta)$ e $c(\theta)$ são funções específicas, sendo θ o parâmetro natural da distribuição de probabilidade e y a variável dependente da distribuição (DOBSON, 2002).

Segundo Cordeiro (1986), uma variável dependente (y) que apresenta valores na forma de proporções segue a distribuição de probabilidade Binomial com parâmetros n e p . McCullagh e Nelder (1989) salientam que, em estudos com dados mensurados em proporção, a relação entre a variável dependente e as variáveis independentes é descrita formalmente através de um modelo de regressão Binomial.

A forma da função da distribuição de probabilidade Binomial pertencente à família exponencial é expressa por

$$f(y; \mu) = \exp\left\{y \log\left(\frac{\mu}{1-\mu}\right) + n \log(1-\mu) + \log\binom{n}{y}\right\} \quad (12)$$

observa-se que a forma da função de distribuição de probabilidade Binomial esta descrita como a forma da equação (11).

Os modelos de regressão pertencentes à classe dos MLG apresentam uma estrutura com três componentes: a componente aleatória, a componente sistemática e a função de

ligação. Sendo: (i) Componente aleatória – identifica a distribuição de probabilidade da variável dependente y pertencente à família exponencial; (ii) Componente sistemática – especifica a estrutura das variáveis independentes, que é utilizada como preditor linear η ; (iii) Função de ligação – descreve a relação funcional entre a componente sistemática e o valor esperado da componente aleatória (a média μ da variável dependente y).

A componente sistemática do modelo de regressão que consiste na estrutura das variáveis independentes é uma soma linear, definida como

$$\eta = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (13)$$

onde a função linear η é denominada de preditor linear, β 's coeficientes desconhecidos e k o número de variáveis independentes. Além disso, outra característica da componente sistemática de um modelo linear é que a média μ da variável dependente y pode ser expressa por uma função conhecida $g(\cdot)$, denominada de função de ligação (MYERS et al., 2002), a qual descreve a relação funcional (ligação) entre a média μ e o preditor linear η , descrevendo a forma

$$g(\mu) = \eta = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (14)$$

Conforme Dobson (2002), a função de ligação $g(\cdot)$ é responsável pela transformação da média da variável dependente, e não de cada observação. Sant'Anna e Caten (2005) descrevem que a escolha da função de ligação depende do problema de modelagem em particular e, que em muitas vezes os modelos de regressão podem apresentar uma função de ligação diferente conforme o conjunto de observações.

A estimação dos coeficientes para o MLG é realizada através do método clássico de máxima verossimilhança, em que os coeficientes β e ϕ (parâmetro de dispersão) são obtidos a partir da maximização do logaritmo da função de verossimilhança. O procedimento de maximizar a função de verossimilhança é realizado por um algoritmo de otimização não-linear, tal como o algoritmo de Newton-Raphson descrito por Cordeiro (1986) ou o algoritmo quasi-Newton (BFGS) descrito por Ferrari e Cribari-Neto (2004). Conforme McCullagh e Nelder (1989) o método clássico de estimação pode ser utilizado considerando qualquer função de distribuição de probabilidade para variável dependente.

Considerando uma amostra de n observações independentes, o logaritmo da função de verossimilhança (log-verossimilhança) dos modelos de regressão apresenta a forma geral,

$$l(\beta; \phi) = \sum_{i=1}^n l_i(\mu_i, \phi) \quad (15)$$

com μ_i definida de tal forma que satisfaz a equação $\mu_i = g^{-1}(\eta_i)$, que é função de β .

A diferenciação da função de log-verossimilhança em relação aos coeficientes desconhecidos (β) do modelo ajustado é definida pela derivada da função de log-verossimilhança, para t variáveis independentes, $t = 1, \dots, k$.

$$\frac{\partial l(\beta, \phi)}{\partial \beta} = \sum_{i=1}^n \frac{\partial l(\mu_i, \phi)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta} \quad (16)$$

Na abordagem de MLG, a estimativa dos coeficientes dos modelos usualmente é obtida pela maximização da função de log-verossimilhança (EMV). No entanto, conforme algumas situações experimentais a estimação dos coeficientes pelo método tradicional de EMV fica comprometida, em virtude de não cumprimento de certos pressupostos, assim, faz-se necessário o uso do método de estimação por quase-verossimilhança (conforme pode ser visto em Sant'Anna, 2006).

Modelo de Quase-verossimilhança

Weddeburn (1974) propôs os modelos de Quase-verossimilhança (MQV) com base nos modelos pertencentes à classe dos MLG, em que os modelos apresentam duas componentes: a componente sistemática (estrutura linear das variáveis independentes) e a função de ligação que relaciona a média (μ_i) da variável dependente à estrutura linear das variáveis independentes (\mathbf{x}_i).

A característica destes modelos de regressão, é que não há a necessidade de assumir a princípio alguma distribuição de probabilidade para a variável dependente. Por conseguinte, a esperança matemática e a variância da variável dependente não são conhecidas *a priori*.

Seja y_i uma variável dependente qualquer de interesse, que assume a $E[y_i] = \mu_i$ e uma variância definida por $\text{Var}[y_i] = \phi^* V(\mu_i)$, onde a função de variância $V(\mu_i)$ é uma função conhecida da média μ_i e ϕ é o parâmetro de dispersão constante. A função de quase-verossimilhança para um modelo de regressão é definida pela equação

$$Q(y; \mu) = \int_y^\mu \frac{y - v}{\phi \cdot V(v)} \partial v \quad (17)$$

O MQV utilizado na modelagem de um conjunto de dados mensurados em proporção não conforme é descrito a partir de uma variável dependente (y_i) que pode assumir qualquer função de variância. Por exemplo, seja uma função $V(\mu_i) = \mu(1-\mu)$ para a variável dependente descrita acima a forma se apresenta

$$Q(y; \mu) = \frac{1}{\phi} \int_y^\mu \frac{y - \mu}{\mu(1 - \mu)} \partial\mu \quad (18)$$

e o logaritmo da função de quase-verossimilhança fica nesse caso dado por

$$\log[Q(y; \mu)] = y \ln\left(\frac{\mu}{1 - \mu}\right) + \ln(1 - \mu) \quad (19)$$

que conforme McCullagh e Nelder (1989), a função acima corresponde a função de log-verossimilhança da distribuição de probabilidade Binomial com função de variância $V(\mu) = \mu(1 - \mu)$, a qual possui a seguinte forma

$$L(y; \mu) = y \ln\left(\frac{\mu}{1 - \mu}\right) + n \ln(1 - \mu) \quad (20)$$

Nota-se que a principal diferença entre como se formam as equações (19) e (20) está em que, quando se usa a função de quase-verossimilhança para estimar os coeficientes desconhecidos do modelo de regressão, apenas se define a relação da variância da variável dependente com a sua média, não sendo necessário definir anteriormente uma distribuição de probabilidade.

De acordo com Cox (1996), uma vantagem da flexibilidade de uso dos MQV's na modelagem de uma variável dependente de conjunto de dados, é que poderíamos utilizar uma função de variância que melhor se ajuste aos dados, sem assumir a priori uma distribuição de probabilidade para esta variável dependente. Além disso, esta função de variância pode não pertencer a nenhuma distribuição de probabilidade da família exponencial. Por exemplo, uma função de variância do tipo $\text{Var}[y_i] = \mu^2(1 - \mu)^2$, apresenta o logaritmo da função de quase-verossimilhança da forma

$$Q(y; \mu) = \int_y^\mu \frac{y - \mu}{\phi \cdot \mu_i^2 (1 - \mu_i)^2} \partial\mu \quad (21)$$

que pode ser reescrita como

$$Q(y; \mu) = \frac{1}{\phi} \left[(2y - 1) \log\left(\frac{\mu}{1 - \mu}\right) - \frac{y}{\mu} - \left(\frac{1 - y}{1 - \mu}\right) \right] \quad (22)$$

Para testar a significância dos coeficientes do modelo de regressão pelo teste da razão de quase-verossimilhança tem-se a estatística de quase-*deviance*. Pode-se dizer que a quase-*deviance* está para a modelagem pela função de quase-verossimilhança assim como a *deviance* está para a função de verossimilhança. Por analogia, a quase-*deviance* de um modelo qualquer é definida como o desvio deste modelo em relação ao modelo nulo, sendo,

$$D(y; \hat{\mu}) = -2\phi[Q(y; \hat{\mu}) - Q(y, y)] = -2\phi[Q(y; \hat{\mu})] = 2 \int_y^\mu \frac{y - \hat{\mu}}{V(\hat{\mu})} \quad (23)$$

em que $Q(y; \hat{\mu})$ é a função de máxima quase-verossimilhança do modelo sob pesquisa e $Q(y; y)$ é a função de máxima quase-verossimilhança do modelo nulo.

Para o modelo de Quase-verossimilhança definido pela equação (19), a estatística de quase-deviance é expressa da forma

$$D(y; \hat{\mu}) = -2\phi[Q(y; \hat{\mu})] = 2 \int_y^\mu \frac{y - \hat{\mu}}{\phi \cdot \hat{\mu}(1 - \hat{\mu})} \quad (24)$$

2.2. Modelo de Regressão Beta

Quando a variável dependente é mensurada em proporção, apresentando valores restritamente no intervalo unitário ($0 \leq y \leq 1$), a relação entre as variáveis dependente e independentes apresenta restrição no domínio da função ($0 < E(y) < 1$) (COX, 1996).

Partindo deste princípio, Ferrari e Cribari-Neto (2004) propuseram o MRB, que é um procedimento alternativo na modelagem de dados mensurados em proporção, cuja estrutura do modelo de regressão permite modelar as relações lineares e não-lineares entre as variáveis independentes e a variável dependente. Ademais, na construção destes modelos as variáveis independentes ou fatores controláveis podem ser de natureza quantitativa ou qualitativa.

A família de distribuições Beta é composta por todas as distribuições de probabilidade que apresente uma variável aleatória y cuja função densidade de probabilidade (fdp) depende dos parâmetros θ_1 e θ_2 , e que sua fdp pode ser escrita na forma,

$$f(y; \theta_1, \theta_2) = \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} y^{\theta_1-1} (1-y)^{\theta_2-1} \quad (25)$$

sendo $0 < y < 1$, θ_1 e $\theta_2 > 0$, $\Gamma(\theta)$ uma função *Gamma* avaliada no ponto θ , ou seja, com $\Gamma(\theta) = \int_0^\infty y^{\theta-1} e^{-y} dy$, $\theta > 0$.

A função $f(y) = f(y; \theta_1, \theta_2)$ é efetivamente uma função densidade de probabilidade com parâmetros θ_1 e θ_2 . Nota-se ainda, que a função $f(y; \theta_1, \theta_2)$ assume valores estritamente positivos, pois para qualquer valor de y pertencente ao intervalo $[0,1]$, a função densidade descrita é crescente, ou seja, $f(y) > 0$.

Segundo Kieschnick e McCullough (2003) e Ferrari e Cribari-Neto (2004), a distribuição de probabilidade Beta é uma função densidade de probabilidade que não pertence à família exponencial, pois a sua distribuição não pode ser escrita na forma canônica e não apresenta um parâmetro de localização θ . A família de distribuições Beta contempla as distribuições Uniforme, Arco-seno e Dirichlet.

O MRB proposto apresenta uma estrutura de regressão baseada em três componentes: (i) Componente aleatória – supõe-se que a variável dependente (y) segue unicamente a distribuição de probabilidade Beta; (ii) Componente sistemática – que especifica uma soma linear dos coeficientes desconhecidos das variáveis independentes, conhecido como preditor linear (η); (iii) Função de ligação – que permite modelar a média da variável dependente em relação às demais variáveis independentes através de uma função, conhecida como $g(\cdot)$.

A estrutura do modelo de regressão é expressa da forma,

$$g(\mu) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (26)$$

onde η é o preditor linear β 's coeficientes desconhecidos, k números de variáveis independentes, $g(\cdot)$ a função de ligação e μ a média da variável dependente.

Conforme Kieschnick e McCullough (2003), o MRB permite gerar estimativas precisas e seguras dos coeficientes, mesmo que o conjunto de dados coletados para a investigação seja consideravelmente pequeno ou os dados mensurados sejam próximos de zero e próximos de um.

Ferrari e Cribari-Neto (2004) se baseiam na suposição de que as proporções seguem uma distribuição de probabilidade Beta, para realizar uma parametrização da esperança matemática $E(y) = \frac{\theta_1}{\theta_1 + \theta_2}$ e da variância $Var(y) = \frac{\theta_1 \theta_2}{(\theta_1 + \theta_2)^2 (\theta_1 + \theta_2 + 1)}$ de uma variável dependente y , utilizando os parâmetros (θ_1 e θ_2) da distribuição de probabilidade Beta.

A estimação dos coeficientes para os MRB é obtida pela maximização da função de log-verossimilhança (EMV), em que os coeficientes β e ϕ são obtidos a partir da maximização do logaritmo da função de verossimilhança, conforme equação (15). O procedimento de maximizar a função de verossimilhança é realizado pelo algoritmo de otimização não-linear de *quasi*-Newton (BFGS).

3. Medidas de Diagnóstico

Uma etapa essencial na análise de ajuste dos modelos de regressão é a verificação da adequação dos modelos de regressão aos dados. Esta etapa, conhecida como medidas de diagnóstico, constitui um conjunto de critérios de adequação e ferramentas gráficas na avaliação da adequação do modelo aos dados. As medidas de diagnóstico fornecem subsídios para detectar: possível violação de alguma das suposições feitas para o modelo, especialmente para a componente aleatória (y) e a função de ligação ($g(\cdot)$), aleatoriedade dos dados, presença

de pontos extremos (*outliers*), adequação da distribuição de probabilidade proposta para a variável dependente e observação de pontos influentes. As medidas de diagnóstico propostas neste artigo para análise de adequação dos modelos de regressão são apresentadas a seguir.

3.1. Critérios de Adequação

(a) *Coefficiente de Determinação*: é uma medida global da qualidade do ajuste, utilizado como indicador numérico que permite comparar o desempenho de diferentes modelos, contudo, não é uma boa estratégia, pois o mesmo sempre aumenta com a inclusão de novas variáveis independentes. Para contornar esse problema, foi utilizado um coeficiente de determinação ajustado, denominado “pseudo” R^2 (R_p^2) que é definido como o quadrado do coeficiente de correlação amostral entre $g(\cdot)$ e $\hat{\eta}$. Segundo Rao e Wu (2005), este coeficiente se restringe a $0 \leq R_p^2 \leq 1$ e, quando $R_p^2 = 1$ existe uma concordância perfeita entre $\hat{\mu}$ e y , conseqüentemente, melhor será o ajuste.

(b) *Deviance*: A análise de *deviance* é feita através da comparação dos valores da medida deviance dos modelos ajustados. Segundo Atkinson e Riani (2000), esta medida é definida como duas vezes a diferença entre o máximo do logaritmo da verossimilhança do modelo nulo e do modelo sob pesquisa.

$$D(y; -\hat{\mu}) = \sum_{i=1}^n 2[l_i(\tilde{\mu}, \phi) - l_i(\hat{\mu}, \phi)] \quad (27)$$

em que $\tilde{\mu}$ é solução de $\partial l_i / \partial \mu_i = 0$, isto é, $\phi(y_i^*, \mu_i^*) = 0$, $l_i(\tilde{\mu}, \phi)$ é a função de máxima verossimilhança do modelo sob pesquisa e $l_i(\mu, \phi)$ é a função de máxima verossimilhança do modelo nulo. Conforme Lee e Nelder (1998), usualmente procede-se a análise de deviance utilizando o ponto crítico $\chi^2(n-k)(\alpha)$ da distribuição qui-quadrado ao nível de significância igual a α , sendo n o número de observações e k o número de coeficientes do modelo. Portanto se $D(y; \mu, \phi) \leq \chi^2_{(n-k)}(\alpha)$, pode-se considerar que há evidências que o modelo sob pesquisa esteja bem ajustado aos dados, a um nível de $\alpha\%$ de significância, usualmente $\alpha < 0,05$, caso contrário deve-se descartar o modelo, pois o mesmo pode ser considerado inadequado. Esta estatística de decisão também pode ser utilizada para a função de quasi-deviance na adequação do modelo de Quase-verossimilhança (HURVICH e TSAI, 1995).

(c) *Critério de Informação de Akaike*: O AIC (Akaike Information Criterion) foi o primeiro critério baseado na informação de Kullback-Leibler (K-L) e assintoticamente não viesado

para K-L. O critério AIC supõe que o modelo verdadeiro pertence ao conjunto de modelos candidatos e é definido por

$$AIC = -2l_i(\hat{\mu}, \phi) + 2(k + 1) \quad (28)$$

onde $l(\hat{\mu}, \hat{\phi})$ é a função de máxima verossimilhança do modelo ajustado e k o número de parâmetro do modelo. Segundo Rao e Wu (2005), o critério AIC foi desenvolvido através dos estimadores de máxima verossimilhança (EMV), para decidir qual o modelo mais adequado quando se utiliza muitos modelos com quantidades diferentes de coeficientes. A decisão quanto ao melhor modelo ajustado é realizado escolhendo o menor valor de AIC.

3.2. Ferramentas Gráficas

(a) *Resíduo Deviance*: é o resíduo mais recomendado em análise gráfica de diagnóstico, pois estes resíduos são os que mais se aproximam da distribuição de probabilidade Normal na verificação da adequação ao papel de probabilidade e aleatoriedade dos resíduos. Para cada observação (i) da variável dependente y_i , pode-se definir o desvio $r_i^d = D_i(y_i; \hat{\mu}_i)$, de tal modo que

$$D(y_i; \hat{\mu}_i, \phi) = \sum_{i=1}^n (r_i^d)^2 \quad (29)$$

em que

$$r_i^d = \text{sign}(y_i - \hat{\mu}_i) \{2[l_i(\tilde{\mu}_i, \phi) - l_i(\hat{\mu}_i, \phi)]\}^{1/2} \quad (30)$$

sendo que a i -ésima observação contribui com a quantidade $(r_i^d)^2$ para o desvio e uma observação com um valor absoluto grande de r_i^d , pode ser vista como discrepante (LEE e NELDER, 1998). Um gráfico desses resíduos contra o índice das observações (i) não deve mostrar nenhuma tendência e sim uma aleatoriedade.

(b) *Alavanca Generalizada*: tem-se mostrado uma ferramenta importante na análise gráfica de diagnóstico quanto a influência das observações em modelos de regressão, ou seja, avalia a importância individual de cada observação no próprio valor ajustado. A medida de alavancagem proposta por Wei et al. (1998), generaliza a definição de pontos de alavanca usados em modelos de regressão linear múltipla para outros modelos lineares pertencentes a classe dos MLG, sendo desenvolvida a partir dos elementos h_{ii} da matriz \mathbf{H} que é conhecida como matriz de projeção ou “matriz chapéu” ($\mathbf{H} = \mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}$). Supondo que todos os pontos exerçam a mesma influência sobre os valores ajustados, pode-se esperar que os

elementos h_{ii} da diagonal da matriz \mathbf{H} sejam definidos por w / n , onde w é o somatório dos elementos h_{ii} definido pelos coeficientes dos modelos e n é o número de observações.

A alavanca generalizada proposta por Wei et al. (1998) sugere que sejam examinados aqueles pontos tais que,

$$h_{ii} \geq \frac{3w}{n} \quad (31)$$

definidos como grandes pontos de alavanca. Ou seja, o valor de h_{ii} associado a i -ésima observação y_i é três vezes maior que a média de todos os h_{ii} da diagonal da matriz \mathbf{H} .

(c) *Distância de Cook*: é uma ferramenta gráfica muito utilizada para detectar a influência de cada observação nas estimativas dos coeficientes do modelo de regressão. Esta medida identifica a influência da retirada da i -ésima observação sobre as estimativas dos coeficientes do modelo, sendo definido por

$$D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})^t X^t X (\hat{\beta}_{(i)} - \hat{\beta})}{w\sigma^2} \quad (32)$$

onde D_i representa uma soma ponderada dos desvios entre as estimativas baseadas nos coeficientes $\hat{\beta}$ e $\hat{\beta}_{(i)}$, w é o posto da matriz diagonal das variáveis independentes, $\hat{\sigma}^2$ é a variância estimada. Assim, essa quantidade obtida pela soma mede a distância quadrática entre $\hat{\beta}$ e $\hat{\beta}_{(i)}$. Segundo Cook e Weisberg (1982), as observações serão consideradas influentes quando $D_i \geq F(k, n - k)(\alpha)$. Sabe-se que $F(k, n - k)(\alpha)$ é o valor crítico da distribuição F de Snedecor (n, k) ao nível $\alpha\%$ de significância. Geralmente, observações que apresentam $D_i > 0,05$, são consideradas influentes e devem ser investigadas.

(d) *Envelope simulado meio-Normal*: é uma ferramenta gráfica de diagnóstico muito útil em modelos lineares e não-lineares. A proposta desta medida é acrescentar ao gráfico de probabilidade normal usual um envelope simulado que pode ser usado para decidir se as observações são consistentes com o modelo ajustado.

Este gráfico é construído a partir da simulação de k valores (estatísticas de ordem) para cada valor previsto pelo modelo ajustado e gerado médias, valores mínimos e máximos de cada valor previsto. Esses valores mínimos e máximos das k estatísticas de ordem produzem o envelope. Desta forma, o gráfico apresentará um intervalo para cada valor previsto ordenadamente contra os escores meio-normais

$$\Phi^{-1} \left\{ \frac{(i + n - \frac{1}{8})}{2n + \frac{1}{2}} \right\} \quad (33)$$

onde $\Phi(\cdot)$ é a função de densidade acumulada da distribuição normal padrão e n é o número de observações. Segundo Atkinson e Riani (2000), caso ocorram tendências não aleatórias dos resíduos absolutos dentro do envelope simulado há indícios de escolha incorreta da distribuição de probabilidade para a variável dependente ou da função de ligação.

4. Estudo Aplicado

Esta seção apresenta um estudo para ilustrar a teoria descrita nas seções anteriores. A aplicação trata-se de uma pesquisa realizada numa empresa curtidora de couro, produtora de couro acabado e fornecedora para as indústrias de calçados e artefatos em couro.

A etapa *wet blue* do processo produtivo avaliado consiste em: o classificador recebe um lote de diferentes tamanhos contendo as matérias-primas e verifica se as características de qualidade satisfazem às especificações, por métodos cognitivos. As matérias-primas que não satisfazem às especificações são classificadas como produtos não conformes, e a fração de produtos não conformes às especificações, por lote, é considerada a variável dependente.

Os fatores controláveis definidos como variáveis independentes relevantes para a modelagem da fração não conforme foram: a seleção da matéria-prima em cinco diferentes estágios, conforme qualidade e preço (x_1); a procedência da matéria-prima adquirida pela empresa (x_2); o classificador que inspeciona as matérias-primas (x_3) e o estado de rebaixamento da matéria-prima (x_4). A Tabela 2 apresenta a caracterização dos níveis dos fatores controláveis do experimento. Os dados coletados contemplaram uma amostra de 754 lotes.

Tabela 2: Caracterização dos níveis dos fatores controláveis

Fatores Controláveis	Número de Níveis	Níveis Codificados				
Seleção de Couro	5	1	2	3	4	5
Procedência	5	1	2	3	4	5
Classificador	3	1	2	3	---	---
Rebaixamento	2	1	2	---	---	---

Fonte: Tabela elaborada pelo autor (2009)

4.1. Análise dos Efeitos

Numa análise preliminar da amostra coletada de 754 lotes, a fração média de produtos não conformes às especificações foi de 0,185, desvio-padrão de 0,170 e coeficiente

de variação de 91%. Nota-se que o histograma da fração de produtos não conformes ilustra claramente que os dados não seguem a distribuição Normal (ver Figura 8).

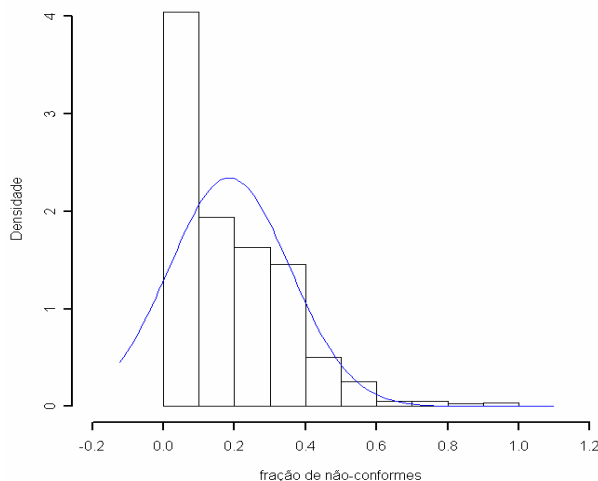


Figura 8: Distribuição da fração não conforme com distribuição Normal sobreposta
Fonte: Elaborado pelo autor (2009)

A análise dos efeitos das variáveis independentes sobre a variável dependente “fração de produtos não conformes”, mostra que, a variável independente “seleção” apresenta uma tendência crescente, ou seja, à medida que aumenta o nível da variável “seleção” aumenta a variável dependente fração de produtos não conformes às especificações (Figura 9(a)). Na Figura 9(b) os níveis da variável “procedência” não apresentam diferenças significativas ($p > 0,05$) para a fração de produtos não conformes às especificações, segundo a análise de variância (ANOVA).

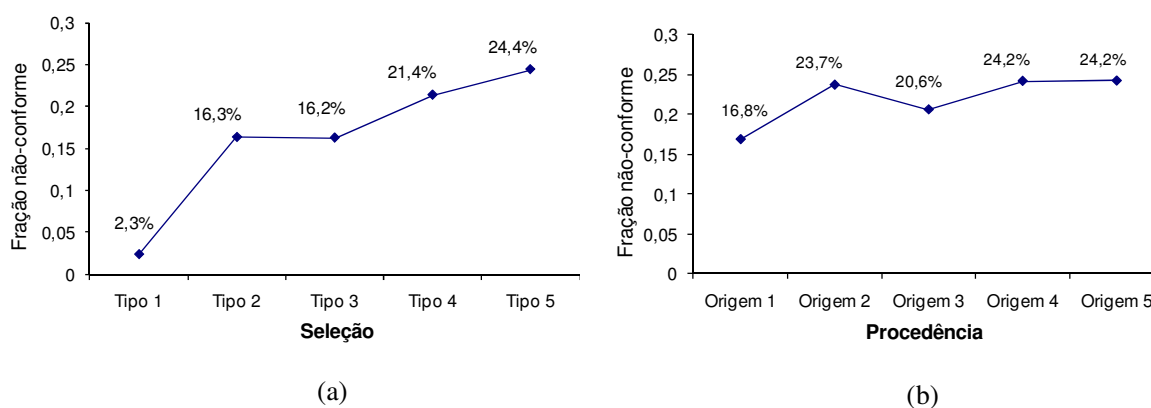


Figura 9: Gráficos da fração não conforme em função das variáveis seleção (a) e procedência (b)
Fonte: Elaborado pelo autor (2009)

Observa-se que o avaliador 3 da variável “classificador” apresentou maior fração não conforme que os outros dois avaliadores, sendo esta diferença estatisticamente significativa ($p < 0,01$) e que o nível rebaixado da variável “rebaixamento” apresenta significativamente ($p < 0,01$) menor fração não conforme, conforme Figura 10(a) e (b).

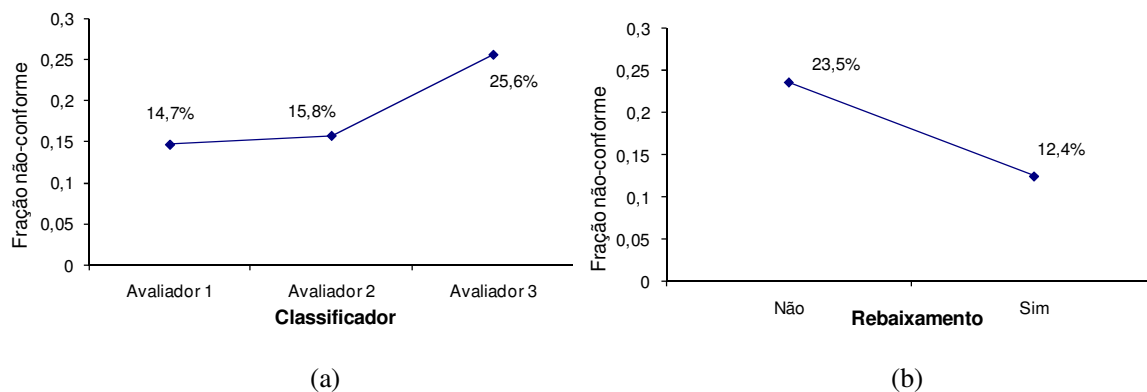


Figura 10: Gráficos da fração não conforme em função das variáveis classificador (a) e rebaixamento (b)
Fonte: Elaborado pelo autor (2009)

4.2. Estrutura dos Modelos Ajustados

Esta seção apresenta a estrutura dos modelos de regressão Beta e de Quase-verossimilhança utilizada na modelagem da fração não conforme às especificações na etapa do processo *wet blue* da empresa curtidora de couro.

No processo de modelagem, as variáveis independentes: seleção (x_1), procedência (x_2), classificador (x_3) e rebaixamento (x_4) foram substituídas pelas variáveis *dummy*, para construção dos modelos de regressão. As novas variáveis independentes foram definidas como: seleção tipo 2 (x_1), seleção tipo 3 (x_2), seleção tipo 4 (x_3), seleção tipo 5 (x_4), procedência 2 (x_5), procedência 3 (x_6), procedência 4 (x_7), procedência 5 (x_8), classificador 2 (x_9), classificador 3 (x_{10}) e rebaixado (x_{11}).

No ajuste do modelo pertencente à classe dos MLG, o modelo de Quase-verossimilhança foi considerado inicialmente conforme a equação reescrita a seguir,

$$g(\mu_i) = \beta_0 + \sum \beta_j x_{ij} + \varepsilon \quad \text{com } i = (1, \dots, 754); \quad j = (1, \dots, 11) \quad (34)$$

onde g é a função de ligação, β 's os coeficientes do modelo e ε o vetor de erro aleatório.

Observa-se que não é assumido a priori que a variável dependente (fração não conforme) possui uma distribuição de probabilidade específica. Segundo Sant'Anna e Caten (2005), para construção dessa classe de modelos outras funções de ligação e de variância

podem ser usadas, com objetivo de verificar qual melhor se ajusta aos dados em estudo. Nesta modelagem foram escolhidas a função de ligação “logito” e de variância do tipo $V(\mu) = \mu(1-\mu)$, pois forneceram o melhor ajuste aos dados, ou seja, produziram a menor quase-*deviance* .

Para a segunda modelagem foi ajustado um modelo de regressão Beta, assumiu-se que a variável dependente (y) segue uma distribuição de probabilidade Beta com média (μ) e utilizou-se as variáveis independentes ($x_1, x_2, x_3, \dots, x_{10}$ e x_{11}) como estrutura linear dos coeficientes, conforme a equação a seguir,

$$g(\mu_i) = \beta_0 + \sum \beta_j x_{ij} + \varepsilon \quad \text{com } i = (1, \dots, 754); \quad j = (1, \dots, 11) \quad (35)$$

onde g representa a função de ligação, ω 's os coeficientes do modelo e ε o vetor de erro aleatório. Para o modelo Beta, foi utilizada a função de ligação “logito” e função de variância do tipo $V(\mu) = \phi^* \{pq / [(p+q)^2 + (p+q+1)]\}$ por gerarem melhor ajuste do modelo aos dados.

4.3. Análise do Ajuste dos Modelos

Ao se modelar com variáveis independentes qualitativas, a estimativa de um dos níveis dessa variável é nula para que os estimadores dos demais níveis possam assumir valores positivos ou negativos, conforme a sua influência na variável dependente. Este nível base, que assume o seu estimador nulo, tornar-se um referencial passível de ser comparado com os outros níveis (SANT’ANNA e CATEN, 2005).

Na modelagem inicial dos dados, os modelos de regressão propostos apresentaram as variáveis independentes “seleção”, “classificador” e “rebaixamento” como estatisticamente significativas para explicar a variável dependente “fração não conforme às especificações”, ao nível de significância de 5%, baseado no teste de *Wald*. A variável independente “procedência” não apresentou significância estatística ao nível de 5% nos dois modelos de regressão ajustados, ver Apêndice A.

A Tabela 3 apresenta as estimativas dos coeficientes com respectivos erros padrões e critérios de adequação da qualidade do ajuste dos dois modelos de regressão ajustados.

Tabela 3: Estimativas dos coeficientes, erros-padrões e critérios de adequação dos modelos de regressão finais

Parâmetro	Modelo de Quase-verossimilhança		Modelo Beta	
	Estimativa	Erro padrão	Estimativa	Erro padrão
Intercepto	-3,7692*	0,3844	-2,4225*	0,2225
Seleção de Couro				
Tipo 1	-	-	-	-
Tipo 2	2,0475*	0,4219	1,5822*	0,2761
Tipo 3	2,2857*	0,3886	1,1202*	0,2270
Tipo 4	2,5226*	0,3878	1,2295*	0,2260
Tipo 5	2,7583*	0,3913	1,4469*	0,2308
Classificador				
Avaliador 1	-	-	-	-
Avaliador 2	0,2672*	0,1095	0,2883*	0,0900
Avaliador 3	0,4483*	0,0877	0,3740*	0,0744
Rebaixamento				
Não	-	-	-	-
Sim	-0,8357*	0,0917	-0,6807*	0,0749
Dispersão (ϕ)	0,15512*	0,5015	0,12337*	0,3942
R_p^2	0,467		0,581	
<i>Deviance</i>	115,65 (746 gl)		121,48 (746 gl)	
AIC	- 213,3		- 224,96	

*Nível de significância “Wald-test” ($p < 0,01$)
gl (graus de liberdade)

Fonte: Tabela elaborada pelo autor (2009)

A forma de regressão para o modelo de Quase-verossimilhança é descrito como $g(\hat{\mu}) = -3,7692 + 2,0475(\text{seleção 2}) + 2,2857(\text{seleção 3}) + 2,5226(\text{seleção 4}) + 2,7583(\text{seleção 5}) + 0,2672(\text{avaliador 2}) + 0,4483(\text{avaliador 3}) - 0,8357(\text{rebaixado})$. Para o modelo Beta a forma de regressão é apresentado como $g(\hat{\mu}) = -2,4225 + 1,5822(\text{seleção 2}) + 1,1202(\text{seleção 3}) + 1,2295(\text{seleção 4}) + 1,4469(\text{seleção 5}) + 0,2883(\text{avaliador 2}) + 0,3740(\text{avaliador 3}) - 0,6807(\text{rebaixado})$.

4.4. Análise de Adequabilidade dos Modelos

Avaliando as estimativas encontradas pelos modelos finais (ver Tabela 2), observa-se que para o modelo de Quase-verossimilhança a seleção do tipo 2 tem $\exp[\beta_1] = \exp[2,0475] = 7,75$ e para o modelo Beta $\exp[\beta_1] = \exp[1,5822] = 4,87$ o que significa estimar que, para o modelo de Quase-verossimilhança a seleção tipo 2 apresenta 7,75 vezes mais chances de produzir fração não conforme às especificações que a seleção tipo 1, já para modelo Beta a seleção tipo 2 apresenta 4,87 vezes mais chances de produzir fração não conforme às especificações que a seleção tipo 1. Bem como, para o modelo de Quase-verossimilhança, a

seleção tipo 5, apresenta 15,77 vezes mais chances ($\exp[\beta_4] = \exp[2,7583] = 15,77$) que a seleção tipo 1 e 425 vezes para o modelo Beta ($\exp[\beta_4] = \exp[1,4469] = 4,25$).

A variável independente “classificador”, no modelo de Quase-verossimilhança, os avaliadores 2 e 3 aumentam as chances de produzir fração não conforme em 31% ($\exp[\beta_5] = \exp[0,2672] = 1,31$) e 56% ($\exp[\beta_6] = \exp[0,4483] = 1,56$) respectivamente, em relação ao avaliador 1. Enquanto que, para o modelo Beta, as chances de produzir fração não conforme dos avaliadores 2 e 3 são de 33% ($\exp[\beta_5] = \exp[0,2883] = 1,33$) e 45% ($\exp[\beta_6] = \exp[0,3740] = 1,45$) respectivamente, comparando com o avaliador 1.

Para a variável independente “rebaixamento”, a estimativa do coeficiente é negativa ($\beta_7 = -0,8357$), indicando que o fato do estado de textura da matéria-prima estar rebaixado implica em que as chances de produzir frações de não conformes às especificações diminuem em 43% ($\exp[\beta_7] = \exp[-0,8357] = 0,43$) para o modelo de Quase-verossimilhança. Enquanto o modelo Beta estima que as chances diminuem em 51% ($\exp[\beta_7] = \exp[-0,6807] = 0,51$).

Analisando a qualidade do ajuste do modelo de Quase-verossimilhança, o coeficiente de determinação “pseudo” R_p^2 foi 0,467, a *Deviance* = 115,65 (746 graus de liberdade) que corresponde a um valor de probabilidade (*p-value*) $p < 0,01$ e um AIC = -213,3 demonstrando que o modelo se ajustou satisfatoriamente. Para o modelo Beta, o coeficiente de determinação foi de 0,581, a *Deviance* = 121,48 (746 graus de liberdade) e um AIC = -224,96 demonstrando que o modelo foi bem ajustado. Nota-se que as medidas de qualidade de ajuste para os modelos ajustados estão relativamente próximas, porém os critérios para o modelo Beta indicam que este é o melhor modelo (ver Tabela 3).

As Figuras 11(a) e (b) apresentam uma aleatoriedade dos resíduos *Deviance*, ou seja, os pontos não apresentam nenhuma tendência indicando que a função de ligação utilizada é adequada. Embora estas revelem que há dois pontos com maior valor residual, correspondendo às observações 685 e 702. Em relação à distância de Cook’s, nota-se que as mesmas observações se apresentam altamente influentes ($D_i > 0,05$) (Figura 12(a) e (b)).

Observa-se nas Figuras 13(a) e (b), que os Modelos de Quase-verossimilhança e Beta apresentaram similaridade na verificação dos pontos de alavanca no conjunto de dados. Os pontos de alavanca em destaque nas figuras referem-se às observações 685 e 702. Contudo, após um estudo realizado com as observações 685 e 702 se verificou que, embora estas aparecerem como pontos discrepantes nos gráficos de diagnóstico nos dois modelos ajustados, as mesmas não foram retiradas da modelagem, por não se constituírem observações discrepantes (*outliers*), não modificando as estimativas dos coeficientes dos modelos.

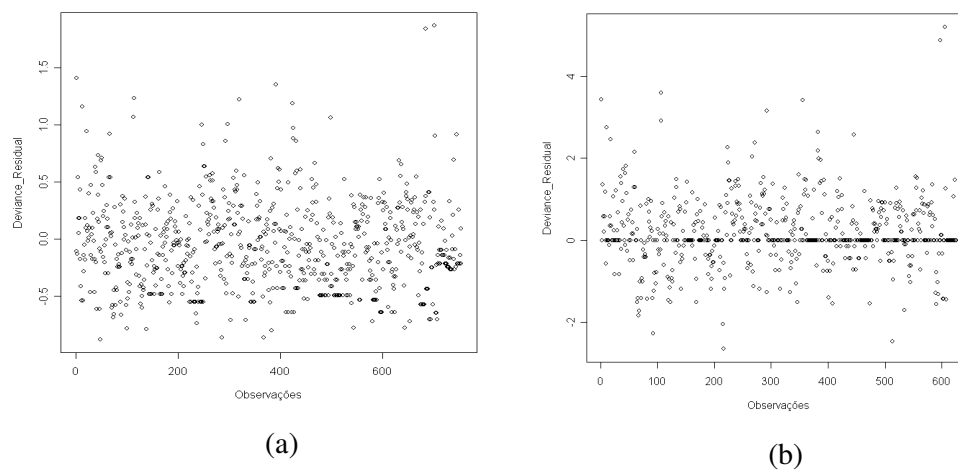


Figura 11: Gráficos de resíduos Deviance para os dados com o ajuste dos Modelos de Quase-verossimilhança (a) e Modelo Beta (b)

Fonte: Elaborado pelo autor (2009)

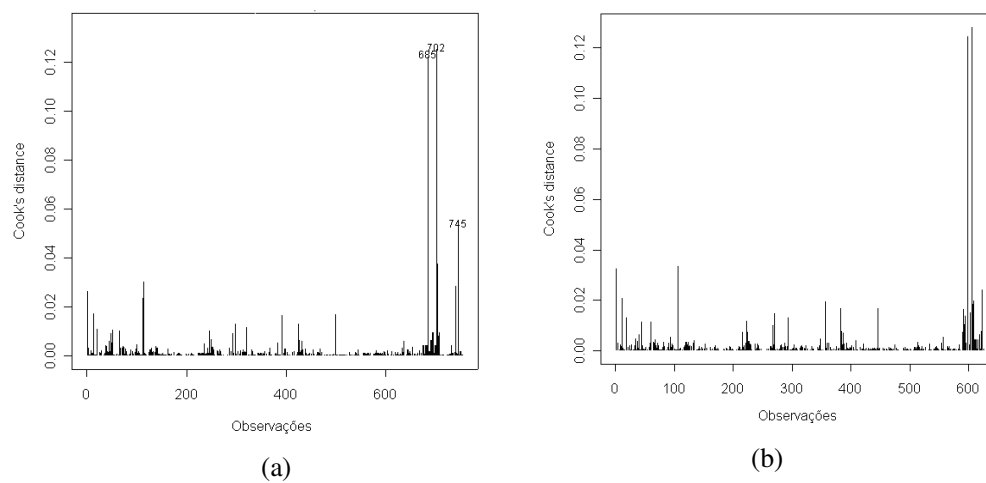


Figura 12: Gráficos de distância de Cook para os dados com o ajuste dos Modelos de Quase-verossimilhança (a) e Modelo Beta (b)

Fonte: Elaborado pelo autor (2009)

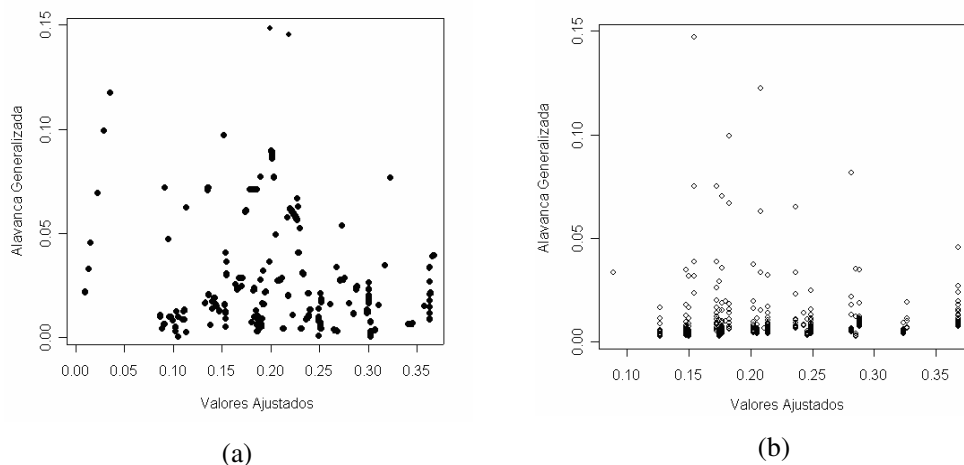


Figura 13: Gráficos de alavanca generalizada para os dados com o ajuste dos Modelos de Quase-verossimilhança (a) e Modelo Beta (b)
 Fonte: Elaborado pelo autor (2009)

Conforme Figura 14(a), o gráfico de envelope simulado meio-normal apresenta alguns resíduos fora do envelope, evidenciando a necessidade de um ajuste mais adequado. Uma opção para se buscar um melhor ajuste é investigar a possibilidade de haver outros fatores (variáveis independentes) que estejam influenciando na etapa *wet blue* do processo de produção da empresa, gerando a fração não conforme às especificações.

No envelope simulado meio-normal gerado pelo modelo Beta a maioria dos resíduos aderiu aos limites do envelope, embora haja alguns resíduos sobrepostos nos limites do envelope, e dois em destaque. Porém pode-se considerar que o modelo Beta se adequou perfeitamente aos dados (ver Figura 14(b)).

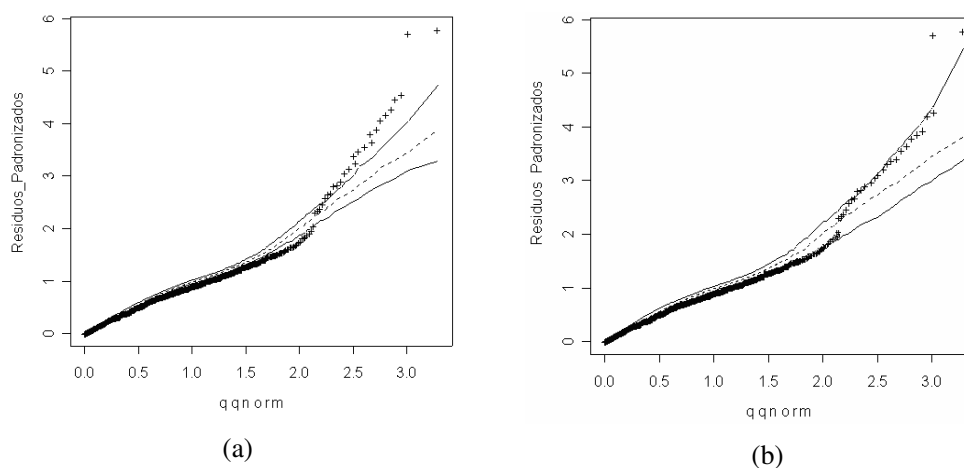


Figura 14: Gráficos de envelope simulado para os dados com o ajuste dos Modelos de Quase-verossimilhança (a) e Modelo Beta (b)
 Fonte: Elaborado pelo autor (2009)

5. Considerações Finais

A importância de conhecer e utilizar modelos de regressão vem da necessidade de investigar o efeito das variáveis independentes, chamadas de fatores controláveis, sobre a variável dependente, chamada de característica de qualidade. O foco deste estudo foi a modelagem da fração de produtos não conformes às especificações de um processo industrial em que a característica de qualidade é mensurada no intervalo $[0,1]$.

Observou-se que a aplicação de modelos de regressão mais corretos para o tipo de dados no processo de modelagem, como modelo de Quase-verossimilhança e modelo Beta, aumenta a precisão das estimativas e fornece uma análise de investigação mais consistente. Ademais, estes modelos consideram a relação não-linear entre as variáveis independentes e dependente e a variância dos erros dependente da média. Estas características são inerentes a dados mensurados em fração.

Os modelos de Quase-verossimilhança e Beta apresentam vantagem na modelagem dos dados mensurados em fração, por permitir flexibilidade de escolha da função de ligação e de variância que melhor se ajuste ao conjunto de dados. Com base na análise das medidas de diagnóstico os modelos de regressão propostos na modelagem da fração de produtos não conformes às especificações apresentaram bom desempenho no ajuste e precisão das estimativas dos efeitos dos fatores controláveis.

Espera-se que as comparações realizadas neste artigo forneçam alguma orientação na investigação e análise de dados em trabalhos futuros, pois a natureza dos dados deve ser levada em consideração. Com base nas análises realizadas recomenda-se o uso dos modelos de regressão Beta e de Quase-verossimilhança a processos industriais que mensurem a produção de produtos não conformes às especificações de manufatura.

Agradecimentos

Esta pesquisa foi apoiada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/Brasil). Os autores agradecem aos revisores anônimos pelos valiosos comentários e sugestões, aprimorando a qualidade deste trabalho.

Referências

- (1) Atkinson, A.C. & Riani, M. (2000) *Robust Diagnostic Regression Analysis*, Springer-Verlag, New York.
- (2) Cook, R.D. & Weisberg, S. (1982) *Residuals and Influence in Regression*, Chapman & Hall, New York.
- (3) Cordeiro, G.M. (1986) *Modelos Lineares Generalizados*. VII SINAPE, Campinas-SP.
- (4) Cox, C. (1996) Nonlinear quasi-likelihood models: applications to continuous proportions, *Computational Statistical & Data Analysis*, **21**, 449–461.
- (5) Dobson, A.J. (2002) *An Introduction to Generalized Linear Models*, 3ª ed., Chapman & Hall, London.
- (6) Ferrari, S.L.P & Cribari-Neto, F. (2004) Beta regression for modeling rates and proportions, *Journal of Applied Statistics*, **31**, 799–816.
- (7) Hamada, M. & Nelder, J.A. (1997) Generalized linear models for quality-improvement experiments, *Journal of Quality Technology*, **29**, 292–304.
- (8) Hurvich, C.M.; Tsai, C-L. Model selection for Quasi-likelihood models in Small Samples. *Biometrics*, Arlington, v. 51, n. 3, p. 1077–1084, 1995.
- (9) Kieschnick, R. & McCullough, B.D. (2003) Regression analysis of variates observed on [0,1]: percentages, proportions and fractions, *Statistical Modelling*, **3**, 193–213.
- (10) Lee, Y. & Nelder, J.A. (1998) Generalized linear models for the analysis of quality improvement experiments, *The Canadian Journal of Statistics*, **26**, 95–105.
- (11) McCullagh, P. & Nelder, J. A. (1989) *Generalized Linear Models*, 2ª ed., Chapman & Hall, London.
- (12) Montgomery, D.C. (2005) *Introduction Statistical Quality Control*, 5ª ed., John Wiley & Sons, New York.
- (13) Montgomery, D.C.; Peck, E.A & Vining, G.G. (2006) *Introduction to Linear Regression Analysis*, 4ª ed., John Wiley & Sons, New York.
- (14) Myers, R.H.; Montgomery, D.C. & Vining, G.G. (2002) *Generalized Linear Models with applications in Engineering and the Sciences*, John Wiley & Sons, New York.
- (15) Nelder, J.A. & Wedderburn, R.W.M. (1972) Generalized Linear Models, *Journal of the Royal Statistical Society A*, **135**, 370–384.
- (16) Prentice, R.L. (1986) Binary Regression using an extended Beta-Binomial distribution, with discussion of correlation induced by covariate measurement errors, *Journal of the American Statistical Association*, **81**, 321–327.
- (17) Rao, C.R. & Wu, Y. (2005) Linear model selection by cross-validation, *Journal Statistical Planning and Inference*, **128**, 231–240.
- (18) Sant’Anna, A.M.O & Caten, C.S. (2005) *Modelagem da proporção de defeituosos usando Modelo de Quase-verossimilhança*. XXV ENEGEP, Porto Alegre-RS.
- (19) Sant’Anna, A.M.O (2006) Método de orientação à modelagem de dados mensurados em proporção. Tese de M. Sc. PPGEP – UFRGS, Rio Grande do Sul, RS, Brasil.
- (20) Wedderburn, R.W.M. (1974) Quasi-likelihood functions, generalized linear models and the Gauss-Newton method, *Biometrika*, **61**, 439–447.

- (21) Wei, B-C.; Hu, Y-Q. & Fung, W-K. (1998) Generalized leverage and its applications, *Scandinavian Journal of Statistical*, **25**, 25–37.

Apêndice

A Tabela 4 apresenta um recorte da matriz de dados experimentais utilizados na construção dos modelos de regressão propostos, uma vez que totaliza 754 observações, se tornando inviável publicar. Os autores podem ceder gentilmente a matriz completa dos dados experimentais.

Tabela 4: Matriz resumo de dados experimentais utilizados na modelagem

BERTIN LTDA. Unidade EV (RS)							
RESULTADO DA CLASIFICAÇÃO DE COUROS WET BLUE							
Período: 02 à 06.06.2004							
Itens	Seleção	Procedência	Classificador	Rebaixamento	Volume	Conforme	Não Conforme
1	1	1	1	2	105	0,24	0,76
2	1	1	1	1	200	0,78	0,22
3	1	1	1	1	204	0,79	0,21
4	1	1	1	2	390	0,65	0,35
5	1	1	1	1	400	0,65	0,35
6	1	1	1	1	202	0,71	0,29
7	1	1	1	1	200	0,65	0,35
8	1	1	1	1	209	0,53	0,47
9	1	1	1	2	181	0,92	0,08
10	1	1	1	2	169	0,98	0,02
11	1	1	1	2	123	1	0
12	1	1	1	2	69	1	0
13	1	1	1	2	84	0,36	0,64
14	1	1	1	2	114	1	0
15	1	1	1	2	281	0,91	0,09
.
.
754	5	4	1	1	197	1	0

Fonte: Tabela elaborada pelo autor (2009)

A Tabela 5 apresenta as estimativas dos coeficientes, respectivos erros padrões e critérios de adequação da qualidade do ajuste dos modelos de regressão iniciais.

Tabela 5: Estimativas dos coeficientes, erros-padrões e critérios de adequação dos modelos de regressão propostos

Parâmetro	Modelo de Quase-verossimilhança		Modelo Beta	
	Estimativa	Erro padrão	Estimativa	Erro padrão
Intercepto	-3,7702*	0,3862	-3,3144*	0,2217
Seleção de Couro				
Tipo 1	-	-	-	-
Tipo 2	2,0507*	0,4233	1,6806*	0,2761
Tipo 3	2,2832*	0,3899	1,9123*	0,2267
Tipo 4	2,5236*	0,3892	2,1206*	0,2256
Tipo 5	2,7547*	0,3928	2,6724*	0,2304
Procedência				
Origem 1	-	-	-	-
Origem 2	0,0246	0,1360	0,0134	0,1146
Origem 3	0,0562	0,1069	0,0117	0,0889
Origem 4	0,0312	0,1598	0,0716	0,1352
Origem 5	0,1363	0,2185	0,1881	0,1886
Classificador				
Avaliador 1	-	-	-	-
Avaliador 2	0,2693*	0,1100	0,1813*	0,0899
Avaliador 3	0,4511*	0,0892	0,3953*	0,0734
Rebaixamento				
Não	-	-	-	-
Sim	-0,8317*	0,0961	-0,8183*	0,0749
Dispersão (ϕ)	0,15271*	0,4989	0,12483*	0,3817
R_p^2	0,361		0,429	
<i>Deviance</i>	113,10 (742 gl)		120,98 (742 gl)	
AIC	- 207,5		- 218,66	

*Nível de significância "teste *Wald-test*" ($p < 0,01$)
gl (graus de liberdade)

Fonte: Tabela elaborada pelo autor (2009)

2.3 ARTIGO 3

**Carta de Controle Beta para Monitoramento de Características de Qualidade do
tipo Fração**

Beta Control Charts for Monitoring of the type Fraction Quality Characteristics

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

Previsão de envio ao European Journal of Operational Research (ISSN 0377-2217)

Carta de Controle Beta para Monitoramento de Características de Qualidade do tipo Fração

Ângelo M. O. Sant'Anna

Universidade Federal do Rio Grande do Sul, Av. Osvaldo Aranha, 99 – 5° andar, Porto Alegre/
RS, 90035-190, Brasil.
e-mail: angelo@producao.ufrgs.br

Carla S. ten Caten

Universidade Federal do Rio Grande do Sul, Av. Osvaldo Aranha, 99 – 5° andar, Porto Alegre/
RS, 90035-190, Brasil.
e-mail: tencaten@producao.ufrgs.br

Resumo

As Cartas p e np são comumente empregadas no monitoramento de características de qualidade do tipo fração e assumem que os dados seguem a distribuição Binomial. Este artigo propõe uma nova carta de controle, Carta Beta, para monitorar variáveis do tipo fração (p). A Carta Beta apresenta os limites de controle baseada na distribuição de probabilidade Beta e foi aplicada no monitoramento da variável tipo fração em três estudos reais, comparando com os limites de controle propostos por Shewhart, Ryan (1989) e Chen (1998). A análise comparativa permitiu avaliar que a aproximação da distribuição Binomial pela Beta foi mais adequada, fornecendo valores contidos no intervalo $[0,1]$. Foi realizada uma análise de sensibilidade considerando o número médio de amostras para o processo em estado sob controle e fora de controle. Conclui-se que a Carta Beta apresentou melhor desempenho no monitoramento de variáveis do tipo fração entre as cartas de controle analisadas.

Palavras-chave: Controle estatístico de qualidade, fração, distribuição Beta

Beta Control Charts for monitoring of the type Fraction quality characteristics

Abstract

p -Charts and np -Charts are commonly used in the monitoring of the type fraction quality characteristics, which assume that the data are Binomial distributed. In this paper we propose a new control chart, *Beta Charts*, for monitoring of the type fraction quality characteristics (p). The *Beta Charts* presents the control limits based in the Beta probability distribution. *Beta Charts* was applied in monitoring of the quality characteristics in three real studies, comparing to the proposed control limits by Shewhart, Ryan (1989) and Chen (1998). The comparative analysis showed that the approach of the Binomial distribution by Beta distribution was more appropriated, with values confined in the $[0,1]$ -interval. The sensitivity analysis was realized considering the average run length for the process in-control and out of control. We may conclude *Beta Charts* shown the best performance in monitoring of the type fraction characteristics between the control charts analyzed.

Keyword: Statistical quality control, fraction, Beta distribution.

1. Introdução

O controle estatístico de processos (CEP) é comumente utilizado no monitoramento e na detecção de mudanças no processo produtivo. Cartas de controle de atributos são importantes ferramentas encontradas no CEP para monitorar processos com dados discretos. As cartas de controle para atributos mais populares para monitorar produtos não conformes às especificações são Cartas p e np , desenvolvidas por Shewhart em 1924. As estimativas de média e variância são calculadas assumindo que o número de produtos defeituosos segue a distribuição de probabilidade Binomial, com parâmetros n e p , e os limites de controle e limite central são calculados aproximando-se pela distribuição Normal.

As características de qualidade do tipo fração são observações expressas no intervalo $[0,1]$ e podem ser definidas em duas categorias. A primeira categoria compreende a razão entre dois números discretos, denominada de percentual; por exemplo, o número de peças defeituosas dividido pelo número total de peças do lote. A segunda categoria compreende a razão entre dois números contínuos, denominada de proporção; por exemplo, o volume de óleo bruto convertido em gasolina dividido pelo volume total.

A ocorrência da fração do tipo percentual (discreto/discreto) de um processo produtivo tem sido comumente modelada como distribuição Binomial, com parâmetros n e p . Conforme o teorema do limite central, se n for satisfatoriamente grande a aproximação da distribuição Binomial pela Normal é satisfatória. A fração do tipo proporção (contínuo/contínuo) pode ser modelada como distribuição Beta por ser uma distribuição contínua e seus parâmetros não dependerem de n .

Há várias regras que tratam das suposições de simetria e aproximação à distribuição Normal. Segundo Schader e Schmid (1989), a aproximação da distribuição Binomial pela distribuição Normal é satisfatória caso duas regras sejam satisfeitas: (i) $np(1-p) \geq 9$ e (ii) $np \geq 5$ quando $0 < p \leq 0,5 \leq (1-p)$. Conforme Fleiss et al. (2003), a aproximação é satisfatória quando o tamanho de p está no intervalo entre $(0,3 \leq p \leq 0,7)$ e n é suficientemente grande para $np \geq 5$ e $n(1-p) \geq n$, pois a variância $p(1-p)$ se mantém constante. Segundo Montgomery (2005), a aproximação da distribuição Binomial pela Normal é satisfatória quando $np \geq 10$ e p está no intervalo entre $(0,1 \leq p \leq 0,9)$.

Em muitos estudos a Carta p é utilizada em situações onde o parâmetro p é considerado pequeno (i.e. $p = 0,001; 0,01; 0,05; 0,1; \dots$) e nestes casos a distribuição Binomial é assimétrica e a aproximação pela distribuição Normal não é satisfatória, permitindo estimar valores negativos ou maiores do que 1.

Alguns estudos apresentam diversos aprimoramentos da Carta p para monitorar a característica de qualidade (CQ) do tipo fração. Quesenberry (1991) propõe a Carta Q Binomial para monitorar a fração não conforme utilizando uma transformação não-linear para os limites de controle. O autor demonstra que a transformação utilizada apresenta melhor aproximação da distribuição Binomial pela distribuição Normal. Heimann (1996) apresenta uma modificação aos limites de controle da Carta p para grandes tamanhos de amostras ($n > 10.000$), destacando que neste caso os limites de controle da Carta p são estreitos, aumentando a taxa de alarme falso. Ryan e Schwertman (1997) sugeriram modificações para os limites de controle da Carta np com objetivo de melhorar a aproximação pela distribuição Normal para valores de $p < 0,03$. McColl e Motley (1998) sugerem uma transformação potência ($x^{0,2777}$) para pequenos valores da fração não conforme (5ppm) como uma melhor aproximação da distribuição Binomial pela Normal.

Chen (1998) propôs um ajuste nos limites de controle da Carta p e comparou com a Carta p tradicional e a Carta Q Binomial através da taxa de alarme falso. Bourke (2008) comparou o desempenho de quatro cartas de controle no monitoramento de alterações na fração não conforme em processos industriais. O autor destaca a similaridade no desempenho da carta de controle *Synthetic* e Carta np em longo período de tempo em um processo sob controle. Sim e Lim (2008) adaptaram as cartas de controle de Shewhart para atributos para dados inflacionados de zero. Os autores utilizam o intervalo de Blyth-Still com limites de 3-*sigmas* para o cálculo dos limites de controle.

Bourke (1991) apresenta uma carta de controle para monitorar o número de ocorrência entre sucessivos produtos não conformes, considerando para tal a distribuição de probabilidade Geométrica. Kaminsky et al. (1992) propõem cartas de controle para monitorar o número de ocorrências e o número médio de ocorrências em unidades supondo a distribuição Geométrica modificada. Yang et al. (2002) propuseram a carta de controle com distribuição Geométrica para monitorar a fração não conforme em processos com alta qualidade ($p < 0,001$ e $n > 50.000$). Conforme os autores, a taxa de alarme falso é suavemente afetada quando o tamanho de amostra é pequeno ou há decréscimo da fração não conforme. Zhang et al. (2007) propuseram um monitoramento do tempo entre a ocorrência da fração de produtos não conformes, pressupondo distribuição Gama.

Há alguns estudos utilizando as cartas de somas cumulativas (CUSUM) e médias móveis exponencialmente ponderadas (EWMA) para detectar pequenas mudanças no parâmetro p , destacando melhor desempenho em comparação com a Carta p . Quesenberry (1995) propôs as cartas de controle Q CUSUM e Q EWMA para monitorar pequenas

mudanças na fração não conforme. O autor mostra que as cartas adaptadas à Estatística Q são mais sensíveis na detecção de pequenas mudanças. Reynolds e Stoumbos (1999, 2000) propuseram as cartas CUSUM Bernoulli e Binomial para monitorar a fração não conforme usando o método de Cadeias de Markov. Perry et al. (2007) propuseram a carta CUSUM para detectar mudança na fração de produtos não conformes em processos industriais via algoritmo não-linear.

Se a distribuição dos dados do processo é assimétrica, a taxa de alarme falso cresce com a assimetria devido à discrepância entre a forma da distribuição dos dados e a forma da distribuição Normal. Conforme Ferrel (1958) e Nelson (1979), uma solução nestes casos é assumir que a distribuição dos dados do processo seja conhecida e construir cartas de controle com limites exatos fornecendo taxas de alarmes falsos desejadas.

Este artigo propõe uma carta de controle Beta (Carta Beta) para monitorar processos que mensuram CQ do tipo fração. Esta carta de controle assume que as CQ do tipo fração pode ser aproximada pela distribuição Beta e propõe novos limites de controle construídos baseados nesta distribuição.

Com objetivo de realizar uma análise comparativa aplicaram-se os limites de controle propostos por Shewhart, Ryan (1989), Chen (1998) e Carta Beta em três estudos aplicados. Também foi calculado o número médio de amostras para avaliar a sensibilidade das cartas comparadas anteriormente

Na seção 2, é apresentada uma breve revisão sobre as distribuições de probabilidade. Na seção 3, são apresentadas as cartas de controle pesquisadas. Na seção 4, são apresentados os conceitos da Carta Beta. Na seção 5, é apresentado um estudo aplicado e a comparação gráfica entre as cartas de controle. A seção 6 apresenta a análise de sensibilidade das cartas de controle investigadas. Finalmente, na seção 7, são apresentadas as conclusões sobre o estudo desenvolvido nesse artigo.

2. Distribuição de Probabilidade

Seja Y uma variável aleatória que mensura o número de itens não conformes (y_i) em um tamanho de amostra (n_i) itens independentes, $i = 1, 2, \dots, m$. A probabilidade de Y ($P\{Y_i = y\}$) é definida pela distribuição Binomial,

$$P(Y_i = y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y} \quad (36)$$

Se o percentual de produtos não conformes for medido em um conjunto de dados ($\pi_i = y_i / n_i$) tem-se que, $Y_i \sim \text{Bin}(n_i ; \pi_i)$ e π_i define a fração de produtos não conformes. Se a ocorrência da fração de itens não conformes em um processo é independente e identicamente distribuída, e com base no teorema do limite central, pode-se assumir que a fração não conforme segue a distribuição de probabilidade Normal, para n suficientemente grande (MONTGOMERY, 2005).

Segundo Collett (2003) e Johnson et al. (2005), uma variável aleatória Y com distribuição Binomial apresenta a média e a variância dadas por, respectivamente,

$$E(Y) = n\pi \quad e \quad Var(Y) = n\pi(1 - \pi) \quad (37)$$

A aproximação da distribuição Binomial pela distribuição Beta pode ser mais apropriada, devido à variedade de forma que sua função densidade de probabilidade pode assumir. Desta forma, admite-se que a variável aleatória (Y_i) segue a distribuição Binomial e a fração (π_i) obtida a partir da variável (Y_i) em cada ocorrência ($i = 1, 2, \dots, m$) segue uma distribuição de probabilidade Beta, indexada pelos parâmetros (θ_1, θ_2), onde $\theta_1, \theta_2 > 0$.

Moitra (1990) relata que a distribuição de probabilidade Beta é extremamente versátil e encoraja sua aplicação em modelagem de uma variedade de CQ devido a sua flexibilidade. Rocke (1993) apresenta um estudo aplicando uma família de transformações para a distribuição Beta conforme diferentes formas assimétricas da densidade. Johnson et al. (1995b) discutem diversos estudos de diferentes ciências em que a distribuição Beta foi melhor ajustada frente diversas distribuições alternativas e sugerem que a distribuição Beta é uma distribuição satisfatória para representar dados em fração.

A família de distribuições Beta é composta por todas as distribuições de probabilidade que apresente uma variável aleatória Y cuja função densidade de probabilidade (FDP) depende dos parâmetros θ_1 e θ_2 , e que sua FDP pode ser escrita na forma,

$$f(y; \theta_1, \theta_2) = \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} y^{\theta_1-1} (1 - y)^{\theta_2-1} \quad (38)$$

sendo $\Gamma(\theta)$ uma função *Gama* avaliada no ponto θ , ou seja, com $\Gamma(\theta) = \int_0^\infty y^{\theta-1} e^{-y} dy$, $\theta > 0$.

Segundo Johnson et al. (1995b) uma variável aleatória Y com distribuição Beta de dois parâmetros possui média e variância dadas por, respectivamente,

$$E(Y) = \frac{\theta_1}{\theta_1 + \theta_2} \quad e \quad Var(Y) = \frac{\theta_1 \theta_2}{(\theta_1 + \theta_2)^2 \cdot (\theta_1 + \theta_2 + 1)} \quad (39)$$

Para a aproximação da distribuição Binomial pela distribuição Beta os parâmetros podem ser relacionados conforme as expressões (40),

$$\theta_1 = \pi \left[\frac{\pi(1-\pi)}{\sigma^2} - 1 \right] \quad e \quad \theta_2 = (1-\pi) \left[\frac{\pi(1-\pi)}{\sigma^2} - 1 \right] \quad (40)$$

A aproximação proposta da distribuição Binomial (discreta) pela distribuição Beta (contínua) apresenta uma restrição. Quando a variável aleatória Y (CQ do tipo fração) segue a distribuição Binomial e assume o valor zero ($y = 0$), a probabilidade associada a esse valor é maior do que zero ($\pi > 0$). Na aproximação pela distribuição Beta, quando a variável aleatória Y assume o valor zero ($y = 0$), a probabilidade associada a esse valor é igual a zero ($\pi = 0$). Ou seja, uma probabilidade teórica nula.

Na prática a ausência de uma CQ do produto ($Y = 0$) em um processo produtivo possui uma probabilidade maior do que zero ($\pi > 0$) de ocorrência e essa situação deve ser considerada. Neste caso, é necessário atribuir uma probabilidade $\pi > 0$ ao resultado $y = 0$, e então, se obtêm $P(Y = 0) = \pi$ e $P(Y > 0) = 1 - \pi$. Desta forma, π representaria a probabilidade de Y no ponto $y = 0$ e a FDP de Y representaria a probabilidade para valores de $Y > 0$, caracterizando-se neste caso uma distribuição mista (MOOD et al., 1986; SOONG, 2004).

3. Cartas de Controle

Os limites de controle da carta tradicional de Shewhart para monitorar a fração não conforme são determinados pela equação (41), assumindo que o tamanho de amostra n é grande suficiente de forma que a distribuição Binomial seja aproximadamente simétrica sob a média (\bar{p}). Isto implica que a distribuição pode ser aproximada a uma distribuição Normal.

$$LC = \bar{p} \pm w \cdot \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \quad (41)$$

onde w é uma constante que define a largura dos limites de controle correspondente a uma região de controle ($1 - \alpha$) e um número médio desejado de amostras até alarme falso (NMA_0). Habitualmente utiliza-se o valor de w igual a 3, devido à aproximação pela distribuição Normal, correspondendo a uma região 0,9973 de controle e um nível de $NMA = 370$.

Ryan (1989) propôs os limites de controle para a fração não conforme, segundo a equação (42), sugerindo melhorar a aproximação pela distribuição Normal devido à assimetria da distribuição Binomial quando a média (\bar{p}) é pequena.

$$LIC = \bar{p} - 3 \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} + \frac{1,25}{n} \quad (42)$$

$$LSC = \bar{p} + 3 \sqrt{\frac{\bar{p}(1 - \bar{p})}{n} + \frac{1,15}{n}}$$

Chen (1998) estendeu os limites de controle propostos por Winterbottom (1993) a partir da utilização da correção assintótica de primeira ordem de Cornish-Fisher para os limites de controle da Carta p , conforme a equação (43). Segundo Will (1968), a correção de Cornish-Fisher possui propriedades para aproximação assintótica para as distribuições F-Snedecor e Normal.

$$LIC = \bar{p} - 3 \sqrt{\frac{\bar{p}(1 - \bar{p})}{n} + \frac{4(1 - 2\bar{p})}{3n}} \quad (43)$$

$$LSC = \bar{p} + 3 \sqrt{\frac{\bar{p}(1 - \bar{p})}{n} + \frac{4(1 - 2\bar{p})}{3n}}$$

Nota-se que as aproximações propostas por Ryan (1989) e Chen (1998) aos limites de controle são aditivas com objetivo de melhorar a aproximação Binomial pela Normal sem interferir na forma da distribuição.

4. Carta de Controle Beta

A carta de controle Beta (Carta Beta) proposta neste trabalho tem por finalidade monitorar processos com características de qualidade (CQ) mensuradas em fração do tipo percentual ou proporção, as quais habitualmente seguem distribuições não-Normais e assimétricas. Esta carta de controle utiliza a distribuição de probabilidade Beta para o cálculo dos limites de controle.

A Carta Beta apresenta algumas vantagens em relação às outras CC: (i) monitora a variável fração de pequenos valores e de forma assimétrica; (ii) os limites de controle são restritos ao intervalo $[0,1]$, não estimando valores negativos ou maiores do um ($0 < p < 1$).

$$LIC = \bar{p} - w_1 \sqrt{s^2(\bar{p})} \quad (44)$$

$$LSC = \bar{p} + w_2 \sqrt{s^2(\bar{p})}$$

onde \bar{p} e $s^2(\bar{p})$ representam a média e a variância da variável fração estimado pela equação (39), w_1 e w_2 são constantes que definem a largura dos limites de controle correspondente a uma região de controle $(1 - \alpha)$ e um número médio desejado de amostras até um alarme falso (NMA_0).

Os limites de controle da Carta Beta são definidos a partir da função de distribuição acumulada (FDA) da variável aleatória Y cuja FDP segue a distribuição Beta, tal como,

$$F(y) = P(Y \leq y) = \int_0^y f(y; \theta_1, \theta_2) dy \quad (45)$$

sendo $F(y) = \Psi_{(\delta; \theta_1, \theta_2)}$ uma função contínua para todo valor y e acumulada quando $\Psi_{(\delta; \theta_1, \theta_2)} = \int_0^1 f(y; \theta_1, \theta_2) dy = 1$. Desta forma, os valores das constantes w_1 e w_2 dos limites de controle da Carta Beta podem ser estimados a partir das expressões,

$$w_1 = \frac{\bar{p} - \Psi_{\left(\frac{\alpha}{2}; \theta_1, \theta_2\right)}}{\sqrt{S^2(\bar{p})}} \quad (46)$$

$$w_2 = \frac{\Psi_{\left(\left[1 - \frac{\alpha}{2}\right]; \theta_1, \theta_2\right)} - \bar{p}}{\sqrt{S^2(\bar{p})}}$$

onde $\frac{\alpha}{2}$ e $\left[1 - \frac{\alpha}{2}\right]$ representam os pontos que limitam a FDA da variável aleatória Y , correspondendo a região de controle desejada.

Apesar da carta de controle proposta, Carta Beta, apresentar uma fundamentação teórica rigorosa, esta tem a desvantagem de não estar incluída em pacotes de *softwares* estatísticos disponíveis atualmente. Assim, este artigo apresenta alguns valores das constantes w_1 e w_2 para construção dos limites de controle da Carta Beta, conforme Tabela 6. Para outras combinações de n e p , os valores das constantes w_1 e w_2 podem ser vistos na Tabela 12 do apêndice.

A Carta Beta proposta apresenta a mesma praticidade e simplicidade operacional que a Carta p , proposta por Shewahrt (1924), além de não apresentar restrições quanto ao tamanho de amostra, ou seja, número de unidades a serem inspecionadas a cada amostra e a frequência de amostragem, para que seja possível uma análise eficiente na detecção de alterações no processo. Por exemplo, pode-se coletar amostras de tamanho 10, 15 ou 20 e a análise de alterações no processo será tão eficiente quanto coletar amostras de tamanho 50, 100 ou 200.

Os limites de controle da Carta Beta são calculados conforme a equação (44). Inicialmente mensura-se a média e o desvio-padrão da fração do tipo percentual ou proporção do processo, em seguida, verifica-se na Tabela 12 os valores das constantes w_1 e w_2 correspondentes a fração média em estudo e o tamanho de amostra coletado. Por fim, é calculado os limites de controle da Carta Beta, podendo apresentar assimetria, conforme a forma da distribuição dos dados.

Para o monitoramento de processos a longo prazo deve-se reavaliar periodicamente os limites de controle, mesmo que não ocorra evidência de alterações. Esse procedimento dinâmico permite que a carta de controle apresente eficiência no monitoramento e detecção de fontes de variabilidade.

Tabela 6: Tabela da constante w_1 e w_2 para a carta de controle Beta

α 0,0027		Tamanho de amostra (n)																		
p	$1-p$	10		30		50		100		150		200		300		500		1000		
		w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2	w_1	w_2	
0,0010	0,9990	0,10	13,48	0,17	11,96	0,22	10,96	0,32	9,52	0,39	8,69	0,45	8,13	0,55	7,38	0,71	6,54	1,00	5,60	
0,0020	0,9980	0,14	12,78	0,24	10,57	0,32	9,49	0,45	8,11	0,55	7,36	0,63	6,88	0,78	6,26	1,00	5,59	1,38	4,87	
0,0050	0,9950	0,22	10,88	0,39	8,58	0,50	7,63	0,71	6,49	0,87	5,92	1,00	5,57	1,22	5,12	1,51	4,66	1,89	4,19	
0,0100	0,9900	0,32	9,21	0,55	7,21	0,71	6,42	1,00	5,52	1,22	5,09	1,38	4,83	1,62	4,50	1,89	4,17	2,19	3,83	
0,0200	0,9800	0,45	7,62	0,78	6,04	1,01	5,44	1,39	4,77	1,63	4,46	1,78	4,28	1,99	4,05	2,20	3,81	2,43	3,58	
0,0500	0,9500	0,73	5,80	1,25	4,83	1,55	4,46	1,93	4,06	2,11	3,87	2,23	3,76	2,37	3,62	2,51	3,49	2,65	3,35	
0,1000	0,9000	1,05	4,69	1,70	4,14	1,97	3,91	2,27	3,67	2,40	3,55	2,49	3,48	2,58	3,40	2,68	3,31	2,77	3,22	
0,1500	0,8500	1,32	4,13	1,97	3,80	2,21	3,65	2,45	3,48	2,55	3,40	2,61	3,35	2,69	3,29	2,76	3,23	2,83	3,16	
0,2000	0,8000	1,55	3,75	2,16	3,57	2,37	3,47	2,56	3,36	2,65	3,30	2,70	3,26	2,76	3,22	2,81	3,17	2,87	3,12	
0,2500	0,7500	1,74	3,47	2,32	3,41	2,49	3,34	2,65	3,26	2,72	3,22	2,76	3,20	2,81	3,16	2,85	3,13	2,90	3,09	
0,3000	0,7000	1,92	3,24	2,44	3,27	2,59	3,24	2,73	3,19	2,78	3,16	2,81	3,14	2,85	3,12	2,89	3,10	2,92	3,07	
0,3500	0,6500	2,09	3,05	2,55	3,15	2,68	3,15	2,79	3,12	2,83	3,11	2,86	3,10	2,89	3,08	2,92	3,07	2,94	3,05	
0,4000	0,6000	2,25	2,87	2,66	3,04	2,76	3,06	2,85	3,06	2,88	3,06	2,90	3,05	2,92	3,05	2,94	3,04	2,96	3,03	
0,4500	0,5500	2,40	2,71	2,75	2,95	2,84	2,98	2,90	3,01	2,93	3,01	2,94	3,02	2,95	3,02	2,97	3,01	2,98	3,01	
0,5000	0,5000	2,55	2,55	2,85	2,85	2,91	2,91	2,95	2,95	2,97	2,97	2,98	2,98	2,99	2,99	2,99	2,99	3,00	3,00	
0,5500	0,4500	2,71	2,40	2,95	2,75	2,98	2,84	3,01	2,90	3,01	2,93	3,02	2,94	3,02	2,95	3,01	2,97	3,01	2,98	2,98
0,6000	0,4000	2,87	2,25	3,04	2,66	3,06	2,76	3,06	2,85	3,05	2,88	3,05	2,90	3,05	2,92	3,04	2,94	3,03	2,96	2,96
0,6500	0,3500	3,05	2,09	3,15	2,55	3,15	2,68	3,12	2,79	3,11	2,83	3,10	2,86	3,08	2,89	3,07	2,92	3,05	2,94	2,94
0,7000	0,3000	3,24	1,92	3,27	2,44	3,24	2,59	3,19	2,73	3,16	2,78	3,14	2,81	3,12	2,85	3,10	2,89	3,07	2,92	2,92
0,7500	0,2500	3,47	1,74	3,41	2,32	3,34	2,49	3,26	2,65	3,22	2,72	3,20	2,76	3,16	2,81	3,13	2,85	3,09	2,90	2,90
0,8000	0,2000	3,75	1,55	3,57	2,16	3,47	2,37	3,36	2,56	3,30	2,65	3,26	2,70	3,22	2,76	3,17	2,81	3,12	2,87	2,87
0,8500	0,1500	4,13	1,32	3,80	1,97	3,65	2,21	3,48	2,45	3,40	2,55	3,35	2,61	3,29	2,69	3,23	2,76	3,16	2,83	2,83
0,9000	0,1000	4,69	1,05	4,14	1,70	3,91	1,97	3,67	2,27	3,55	2,40	3,48	2,49	3,40	2,58	3,31	2,68	3,22	2,77	2,77
0,9500	0,0500	5,80	0,73	4,83	1,25	4,46	1,55	4,06	1,93	3,87	2,11	3,76	2,23	3,62	2,37	3,49	2,51	3,35	2,65	2,65
0,9800	0,0200	7,62	0,45	6,04	0,78	5,44	1,01	4,77	1,39	4,46	1,63	4,28	1,78	4,05	1,99	3,81	2,20	3,58	2,43	2,43
0,9900	0,0100	9,21	0,32	7,21	0,55	6,42	0,71	5,52	1,00	5,09	1,22	4,83	1,38	4,50	1,62	4,17	1,89	3,83	2,19	2,19
0,9950	0,0050	10,88	0,22	8,58	0,39	7,63	0,50	6,49	0,71	5,92	0,87	5,57	1,00	5,12	1,22	4,66	1,51	4,19	1,89	1,89
0,9980	0,0020	12,78	0,14	10,57	0,24	9,49	0,32	8,11	0,45	7,36	0,55	6,88	0,63	6,26	0,78	5,59	1,00	4,87	1,38	1,38
0,9990	0,0010	13,48	0,10	11,96	0,17	10,96	0,22	9,52	0,32	8,69	0,39	8,13	0,45	7,38	0,55	6,54	0,71	5,60	1,00	1,00

Fonte: Tabela elaborada pelo autor (2009)

5. Estudo Aplicado

Nesta seção são apresentados três exemplos de processos nos quais são monitoradas características de qualidade (CQ) mensuradas em fração do tipo percentual ou proporção. Os exemplos ilustram conjuntos de dados reais publicados pelos autores Montgomery (2005, p. 270), Draper e Smith (1998, p.101) e Brownlee (1965, p.454), que serão analisados utilizando os limites de controle propostos por Shewhart, Ryan (1989), Chen (1998) e Carta Beta proposta. Estes exemplos foram escolhidos a fim de avaliar as CQ do tipo fração com valores pequenos ($p \cong 0$), intermediários ($p \cong 0,5$) e grandes ($p \cong 1$).

O exemplo 1 consiste em um conjunto de dados de um processo de fabricação de embalagens de suco de laranja concentrado e congelado em 30 lotes de 50 embalagens. O exemplo 2 consiste em um conjunto de dados do estudo de contaminação de amendoins por substâncias tóxicas coletados em 34 lotes de 120 libras (± 54 kg). O exemplo 3 consiste em

um conjunto de dados de um processo de fabricação de ácido nítrico por oxidação de amônia coletados em 21 dias.

As características de qualidade a serem monitoradas no estudo aplicado são: percentual de embalagens não conformes (y_1), proporção de amendoins não contaminados (y_2) e proporção de amônia não convertida (y_3). A variável y_1 é obtida a partir da razão entre o número de embalagens não conformes às especificações e o número total de embalagens do lote. As variáveis y_2 e y_3 são obtidas pela razão entre o volume de matéria-prima não conforme e o volume total.

A Figura 15 ilustra a distribuição das CQ y_1 , y_2 e y_3 através de histogramas, distribuição densidade e distribuição Normal. É possível observar a forma assimétrica e bimodal da distribuição dos dados. A Tabela 7 apresenta um resumo estatístico dos dados investigados. Estas estimativas são utilizadas para estabelecer os limites de controle das cartas pesquisadas.

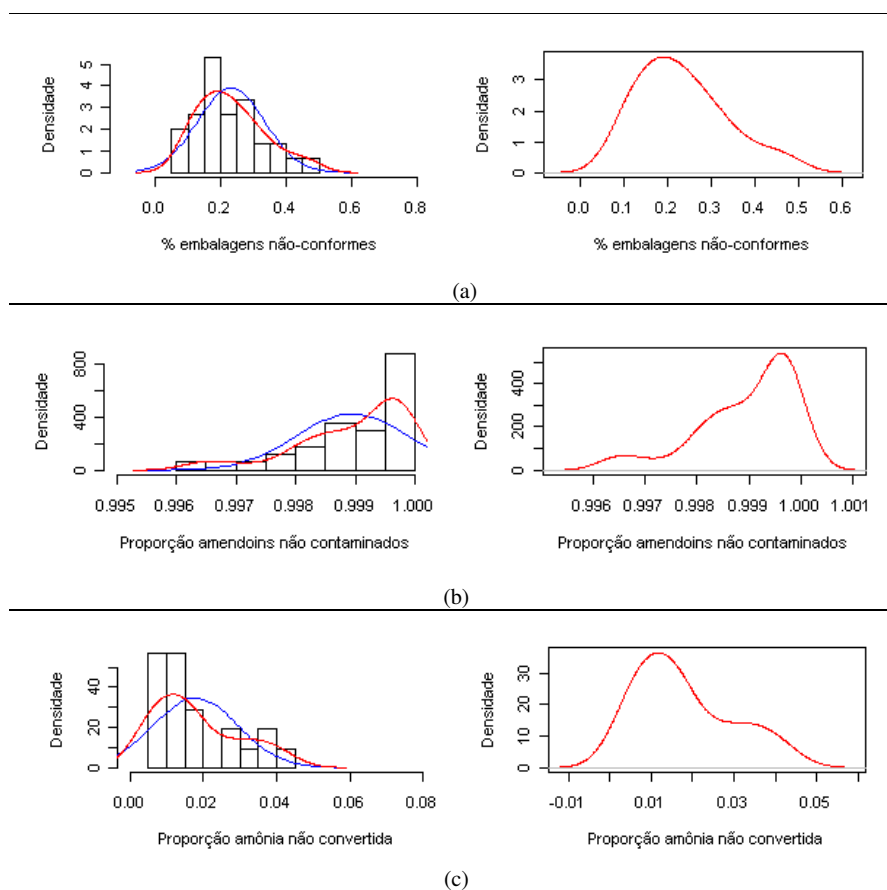


Figura 15: Distribuição das CQ y_1 (a), y_2 (b) e y_3 (c) com distribuição densidade e distribuição Normal sobrepostas

Fonte: Elaborado pelo autor (2009)

Tabela 7: Resumo estatístico das CQ dos três exemplos investigados

Variáveis	Mínimo	Máximo	Média	Variância
y_1	0,080	0,480	0,2313	0,1778
y_2	0,99642	0,99987	0,99896	0,000001
y_3	0,0050	0,0420	0,0179	0,00013

Fonte: Tabela elaborada pelo autor (2009)

5.1. Análise comparativa entre as Cartas de Controle

Esta seção apresenta uma comparação entre os limites de controle propostos para a Carta Beta e os limites de controle propostos por Shewhart, Ryan (1989) e Chen (1998) para as três CQ do exemplo numérico.

O valor da constante w adotado para o cálculo dos limites de controle das cartas de Shewhart, Ryan (1989) e Chen (1998) foi $w = 3$, admitindo a probabilidade $\alpha = 0,0027$ de alarmes falsos no monitoramento do processo, baseado na distribuição Normal. Para a Carta Beta os valores das constantes w_1 e w_2 foram determinados admitindo a probabilidade $\alpha = 0,0027$ de alarmes falsos no monitoramento do processo, baseado na distribuição Beta. A adoção de valores da constante w referentes a mesma probabilidade e alarmes falsos das cartas de controle permitiram o estudo de comparação gráfica entre as cartas de controle estudadas

A partir das estimativas de média e de variância amostrais apresentadas na Tabela 7, são calculados os limites de controle das cartas pesquisadas, conforme a Tabela 8.

Tabela 8: Limites de controle das cartas para as CQ dos três exemplos investigados

	y_1		y_2		y_3	
	LIC	LSC	LIC	LSC	LIC	LSC
Shewhart	0,0525	0,4102	0,99596	1,00196	-0,0385	0,0743
Ryan	0,0775	0,4331	0,99726	1,00316	-0,0135	0,0973
Chen	0,0668	0,4244	0,99726	1,00066	-0,0128	0,0999
Beta	0,0882	0,4399	0,99419	0,99995	0,0007	0,0685

Fonte: Elaborado pelo autor (2009)

As Figuras 16, 17 e 18 apresentam os limites de controle propostos por Shewhart, Ryan (1989), Chen (1998) e Carta Beta para as CQ y_1 , y_2 e y_3 respectivamente. Verifica-se que os limites de controle da Carta Beta apresentam similaridade em relação aos outros

limites de controle. Observa-se que, segundo os limites de controle investigados, o processo está fora de controle estatístico, havendo duas causas especiais na CQ y_1 investigada (Figura 16). Segundo Montgomery (2005), estas causas especiais identificadas se referem a introdução de novos lotes de matéria-prima e de novo operador na linha de produção, respectivamente.

Observa-se na Figura 17 que os limites de controle propostos por Shewhart, Ryan (1989) e Chen (1998) extrapolam o limite superior da distribuição Binomial, que é restrito ao intervalo $[0,1]$, com valores acima de 1. No entanto, os limites de controle da Carta Beta não extrapolam esta região. Conforme os limites de controle investigados o processo pode ser considerado sob controle estatístico, não havendo causas especiais na CQ y_2 , corroborando com o estudo realizado pelos autores Draper e Smith (1998).

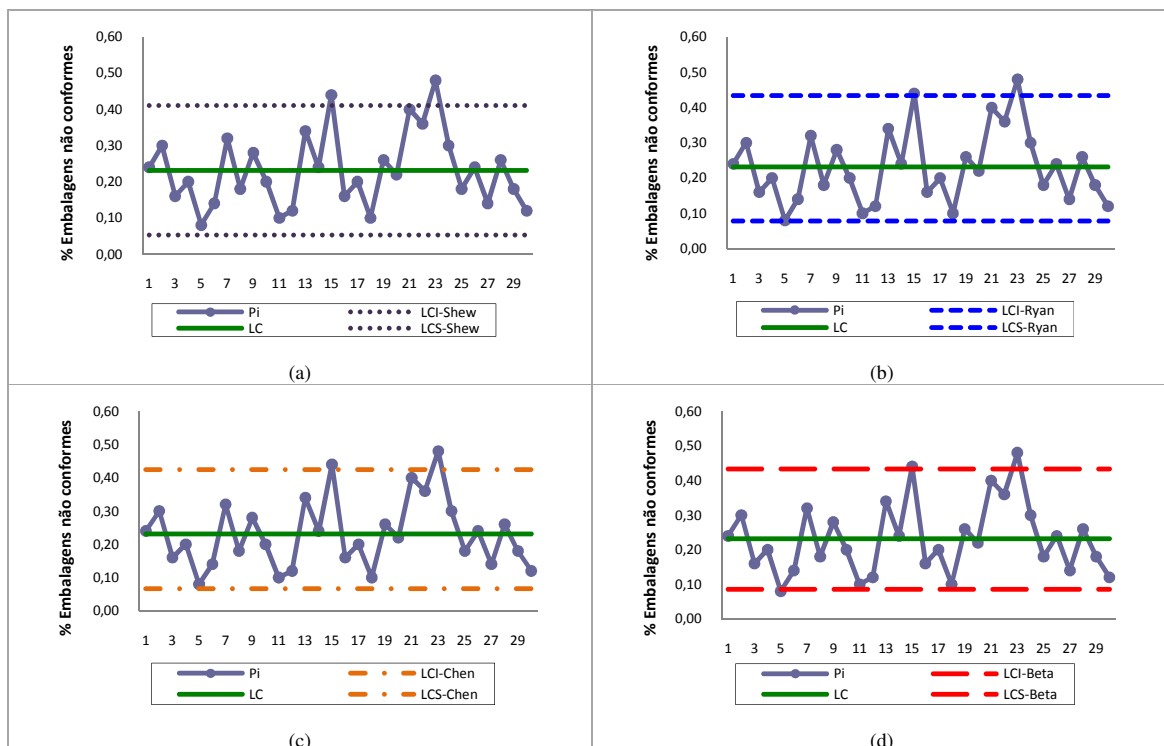


Figura 16: Cartas de controle da fração de embalagens não conformes (y_1)
Fonte: Elaborado pelo autor (2009)

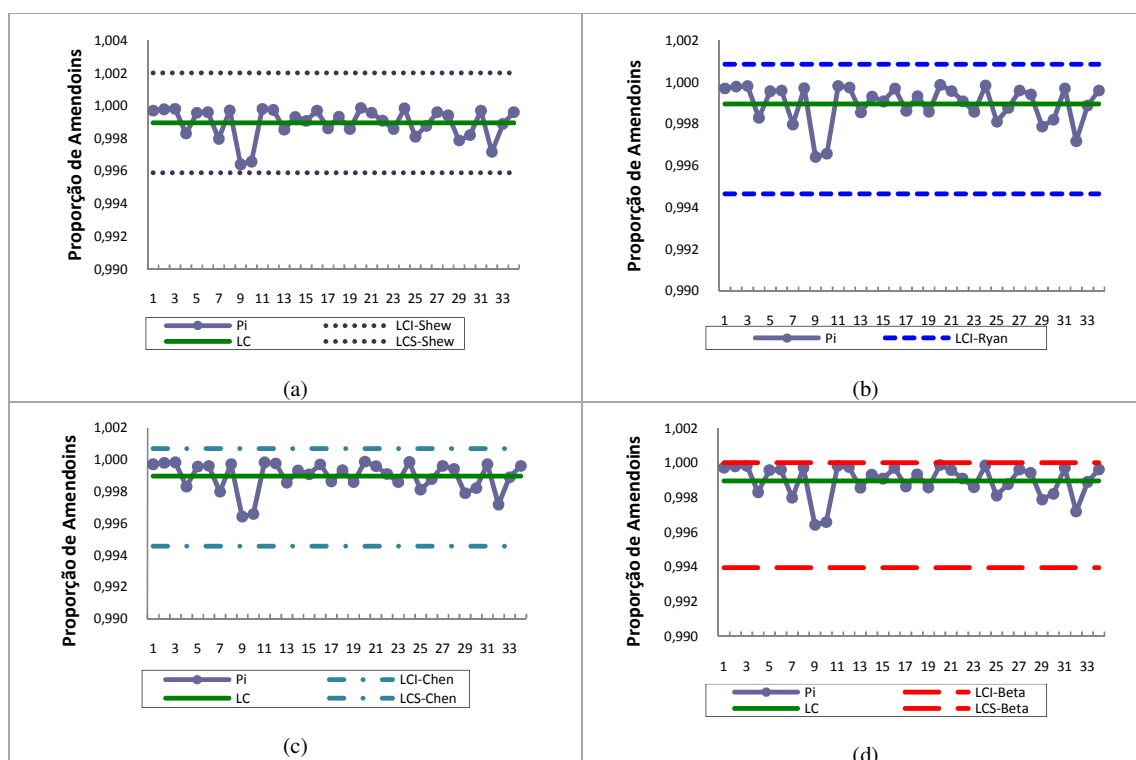


Figura 17: Carta de controle da fração de amendoins não contaminados (y_2)
 Fonte: Elaborado pelo autor (2009)

Na Figura 18, os limites de controle propostos por Shewhart, Ryan (1989) e Chen (1998) extrapolam o limite inferior da distribuição Binomial, que é restrito ao intervalo $[0,1]$, apresentando valores negativos. No entanto, os limites de controle da Carta Beta não extrapolam esta região. Conforme os limites de controle investigados o processo pode ser considerado sob controle estatístico, não havendo causas especiais na CQ y_3 investigada.

Concluindo a comparação gráfica apresentada, os limites de controle propostos para a Carta Beta assumindo a distribuição Beta apresentaram estimativas satisfatórias contidas na região de restrição $[0,1]$ da variável aleatória Y . Entretanto, as cartas de controle investigadas com aproximação pela distribuição Normal apresentaram limites de controle com valores negativos e maiores do 1. Assim, salienta-se uma melhor adequação da aproximação da distribuição Binomial pela distribuição Beta.

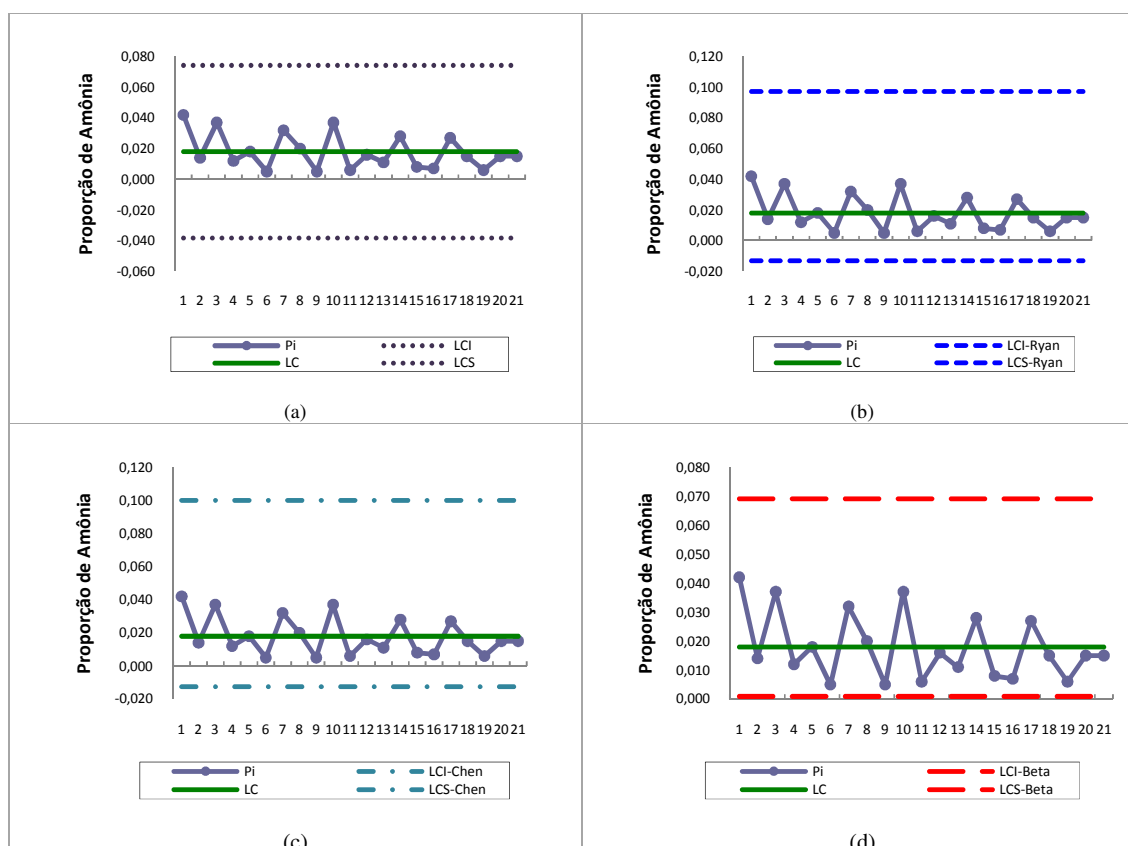


Figura 18: Carta de controle da fração de amônia não convertida (y_3)
Fonte: Elaborado pelo autor (2009)

6. Análise de Sensibilidade das Cartas de Controle

Esta seção apresenta a análise de sensibilidade entre os limites de controle propostos por Shewhart, Ryan (1989), Chen (1998) e Carta Beta observando a taxa de alarme falso. A Tabela 9 apresenta os limites inferiores e superiores de controle da carta de Shewhart, Ryan (1989), Chen (1998) e Carta Beta para quatro pequenos valores de $p = 0,001, 0,01, 0,05$ e $0,10$ e variadas condições de n , contemplando np de 1 a 10. Como ilustração, considere o valor de $p = 0,01$ e $n = 200$ ($np = 2$), nota-se que o limite inferior das cartas de controle propostos Shewhart, Ryan (1989) e Chen (1998) estimam valores negativos (LIC-Shewhart = -0,01111, LIC-Ryan = -0,00486 e LIC-Chen = -0,00457). Enquanto que a Carta Beta apresenta valores contidos no intervalo $[0,1]$ (LIC-Beta = 0,00026).

Conforme se pode verificar na Tabela 9, as aproximações propostas por Shewhart, Ryan (1989) e Chen (1998) apresentam limites de controle inferiores positivos apenas quando $np \geq 9$, $np \geq 6 | p > 0,03$ e $np \geq 6 | p > 0,015$ respectivamente. A Carta Beta apresenta

limites de controle adequados para qualquer condição de n e p , pois seus limites de controle não dependem de n .

Como análise comparativa destes resultados, os limites propostos por Ryan (1989) e Chen (1998) com aproximação pela distribuição Normal apresentam melhor desempenho que os de Shewhart, e os limites de controle da Carta Beta com aproximação pela distribuição Beta apresentam desempenho superior a todas as cartas analisadas.

Tabela 9: Comparação entre os Limites de Controle de Shewhart, Ryan (1989), Chen (1998) e Carta Beta para variadas condições de np .

p	n	np	Shewhart		Ryan		Chen		Beta	
			LIC	LSC	LIC	LSC	LIC	LSC	LIC	LSC
0,001	1000	1	-0,0020	0,0040	-0,000748	0,005148	-0,000668	0,005329	0,000001	0,006597
0,01	100	1	-0,0198	0,0398	-0,007350	0,051350	-0,006783	0,052916	0,000013	0,064949
0,05	20	1	-0,0962	0,1962	-0,033702	0,253702	-0,036202	0,256202	0,000052	0,301819
0,1	10	1	-0,1846	0,3846	-0,059605	0,499605	-0,077938	0,491272	0,000077	0,545013
0,001	2000	2	-0,0011	0,0031	-0,000495	0,003695	-0,000455	0,003786	0,000026	0,004445
0,01	200	2	-0,0111	0,0311	-0,004857	0,036857	-0,004574	0,037640	0,000262	0,043962
0,05	40	2	-0,0534	0,1534	-0,022130	0,182130	-0,023380	0,183380	0,001274	0,209015
0,1	20	2	-0,1012	0,3012	-0,038746	0,358746	-0,047913	0,354579	0,002450	0,390961
0,001	3000	3	-0,0007	0,0027	-0,000315	0,003115	-0,000288	0,003175	0,000071	0,003620
0,01	300	3	-0,0072	0,0272	-0,003067	0,031067	-0,002878	0,031589	0,000705	0,035875
0,05	60	3	-0,0344	0,1344	-0,013576	0,153576	-0,014410	0,154410	0,003524	0,172241
0,1	30	3	-0,0643	0,2643	-0,022650	0,302650	-0,028761	0,299872	0,007045	0,326645
0,001	4000	4	-0,0005	0,0025	-0,000187	0,002787	-0,000167	0,002832	0,000116	0,003167
0,01	400	4	-0,0049	0,0249	-0,001800	0,027800	-0,001658	0,028191	0,001165	0,031432
0,05	80	4	-0,0231	0,1231	-0,007476	0,137476	-0,008101	0,138101	0,005873	0,151779
0,1	40	4	-0,0423	0,2423	-0,011052	0,271052	-0,015636	0,268969	0,011873	0,290115
0,001	5000	5	-0,0003	0,0023	-0,000091	0,002571	-0,000075	0,002607	0,000158	0,002876
0,01	500	5	-0,0033	0,0233	-0,000849	0,025649	-0,000736	0,025962	0,001588	0,028567
0,05	100	5	-0,0154	0,1154	-0,002883	0,126883	-0,003383	0,127383	0,008030	0,138485
0,1	50	5	-0,0273	0,2273	-0,002279	0,250279	-0,005946	0,248613	0,016300	0,266098
0,001	6000	6	-0,0002	0,0022	-0,000016	0,002416	-0,000002	0,002446	0,000196	0,002671
0,01	600	6	-0,0022	0,0222	-0,000103	0,024103	-0,000008	0,024364	0,001964	0,026541
0,015	400	6	-0,0032	0,0332	-0,000108	0,036108	0,000000	0,036466	0,002951	0,039674
0,02	300	6	-0,0042	0,0442	-0,000082	0,048082	0,000018	0,048515	0,003941	0,052715
0,025	240	6	-0,0052	0,0552	-0,000025	0,060025	0,000044	0,060511	0,004934	0,065664
0,03	200	6	-0,0062	0,0662	0,000063	0,071937	0,000080	0,072454	0,005930	0,078522
0,04	150	6	-0,0080	0,0880	0,000333	0,095667	0,000178	0,096178	0,007932	0,103958
0,05	120	6	-0,0097	0,1097	0,000730	0,119270	0,000313	0,119687	0,009947	0,129032
0,1	60	6	-0,0162	0,2162	0,004644	0,235356	0,001588	0,233967	0,020226	0,248871
0,001	7000	7	-0,0001	0,0021	0,000045	0,002298	0,000057	0,002323	0,000229	0,002516
0,01	700	7	-0,0013	0,0213	0,000504	0,022925	0,000585	0,023149	0,002297	0,025019
0,05	140	7	-0,0053	0,1053	0,003669	0,113473	0,003312	0,113831	0,011640	0,121899
0,1	70	7	-0,0076	0,2076	0,010287	0,223999	0,007668	0,222809	0,023685	0,235802
0,001	8000	8	-0,0001	0,0021	0,000096	0,002204	0,000106	0,002226	0,000259	0,002395
0,01	800	8	-0,0006	0,0206	0,001009	0,021991	0,001080	0,022187	0,002593	0,023826
0,05	160	8	-0,0017	0,1017	0,006122	0,108878	0,005810	0,109190	0,013141	0,116291
0,1	80	8	-0,0006	0,2006	0,015002	0,214998	0,012710	0,213956	0,026745	0,225471
0,001	9000	9	0,0000	0,0020	0,000139	0,002127	0,000148	0,002147	0,000285	0,002297
0,01	900	9	0,0001	0,0199	0,001439	0,021228	0,001502	0,021402	0,002857	0,022860
0,05	180	9	0,0013	0,0987	0,008210	0,105123	0,007933	0,105401	0,014478	0,111740
0,1	90	9	0,0051	0,1949	0,019021	0,207646	0,016984	0,206720	0,029465	0,217068
0,001	10000	10	0,0001	0,0019	0,000177	0,002063	0,000185	0,002081	0,000309	0,002216
0,01	1000	10	0,0006	0,0194	0,001811	0,020589	0,001867	0,020746	0,003094	0,022059
0,05	200	10	0,0038	0,0962	0,010017	0,101983	0,009767	0,102233	0,015677	0,107960
0,1	100	10	0,0100	0,1900	0,022500	0,201500	0,020667	0,200667	0,031900	0,210060

Fonte: Elaborado pelo autor (2009)

A análise de sensibilidade contempla um processo sob controle e outro fora de controle. Para o processo sob controle será calculado o número médio de amostras até a presença de uma falsa causa especial (NMA_0). Um maior valor de NMA_0 indica uma menor probabilidade de alarmes falsos no processo. O NMA_0 pode ser escrito em função da

probabilidade de erro tipo I (α), que é a probabilidade da carta de controle detectar indevidamente uma alteração no parâmetro de referência monitorado quando o processo está sob controle estatístico (MONTGOMERY, 2005).

$$NMA_0 = \frac{1}{\alpha} \quad (47)$$

onde α é a probabilidade de alarmes falsos, e definido como,

$$\alpha = 1 - P[LIC < x < LSC | p = p_0] \quad (48)$$

onde p_0 é a média da variável do tipo fração de um processo sob controle.

Para o processo fora de controle será calculado o número médio de amostras até a detecção de causas especiais (NMA_1). O menor valor de NMA_1 representa um menor número médio de amostras coletadas até a detecção da alteração induzida. O NMA_1 pode ser escrito em função da probabilidade de erro tipo II (γ), que é a probabilidade da carta de controle não detectar uma alteração no parâmetro de referência monitorado quando o processo está fora de controle estatístico (MONTGOMERY, 2005).

$$NMA_1 = \frac{1}{[1 - \gamma]} \quad (49)$$

em que γ é a probabilidade de alarmes verdadeiros.

$$\gamma = P[LIC < x < LSC | p = p_1] , \text{ com } p_1 = p_0 + \lambda \quad (50)$$

onde λ é a alteração induzida no processo, p_1 é a média da variável do tipo fração de um processo fora de controle.

As Tabelas 10 e 11 apresentam uma comparação entre as cartas de controle de Shewhart, Ryan (1989), Chen (1998) e Carta Beta para alguns valores dos parâmetros p e n , as probabilidades de alarmes α e γ e os valores de NMA correspondentes. É desejável que uma carta de controle apresente um NMA_0 grande a fim de evitar alarmes falsos e um NMA_1 pequeno a fim de detectar rapidamente alterações no processo.

A sensibilidade das cartas de controle investigadas para $p = 0,001$ e $n = 1500$ mostra similaridade entre as cartas de Shewhart, Ryan (1989) e Chen (1998) para um processo sob controle ($NMA_0 = 226$). A carta proposta por Chen (1998) apresenta maior rapidez na detecção de alterações no valor do parâmetro p ($p_1 = 0,002, \dots, p_4 = 0,01$). A Carta Beta apresenta melhor sensibilidade para a probabilidade de alarme falso e desempenho na detecção de alterações que as outras cartas investigadas.

Para $p = 0,01$ e $n = 200$ as cartas de Shewhart, Ryan (1989) e Chen (1998) apresentam similaridades para um processo sob controle e na detecção de alterações no valor do parâmetro p ($p_1 = 0,002, \dots, p_4 = 0,01$). A carta Beta apresenta melhor sensibilidade para a probabilidade de alarme falso e desempenho na detecção de alterações, conforme Tabela 10.

Como a distribuição Binomial é discreta, os limites inferiores de controle com 3-sigma são truncados no valor zero, o valor nominal de α não é igual a 0,0027. Assim, as cartas de controle com aproximação pela distribuição Normal apresentam limites estreitos, gerando NMA_0 menor que 370 e alarmes falsos com maior frequência que o ideal. Enquanto que, com aproximação pela distribuição Beta é possível definir exatamente o valor nominal de interesse.

Tabela 10: Comparação do NMA entre as cartas de controle de Shewhart, Ryan (1989), Chen (1998) e Beta para $p_0 = 0,001, n = 1500, \lambda = 0,001$ e $p_0 = 0,01, n = 200, \lambda = 0,01$

	$n = 1500$						$n = 200$					
	p	LIC	LSC	α	γ	NMA	p	LIC	LSC	α	γ	NMA
Shewhart	0,001	-0,0014	0,0034	0,0044	-	226	0,01	-0,0111	0,0311	0,0043	-	233
	0,002	-0,0015	0,0055	-	0,916	12	0,02	-0,0097	0,0497	-	0,891	9
	0,003	-0,0012	0,0072	-	0,703	3	0,03	-0,0062	0,0662	-	0,606	3
	0,005	-0,0005	0,0105	-	0,241	1	0,05	0,0038	0,0962	-	0,124	1
	0,01	0,0023	0,0177	-	0,003	1	0,10	0,0364	0,1636	-	0,000	1
Ryan	0,001	-0,0006	0,0042	0,0044	-	226	0,01	-0,0049	0,0369	0,0033	-	303
	0,002	-0,0006	0,0062	-	0,916	12	0,02	-0,0034	0,0554	-	0,891	9
	0,003	-0,0004	0,0080	-	0,703	3	0,03	0,0001	0,0719	-	0,606	3
	0,005	0,0004	0,0112	-	0,241	1	0,05	0,0100	0,1020	-	0,124	1
	0,01	0,0031	0,0185	-	0,003	1	0,10	0,0426	0,1694	-	0,000	1
Chen	0,001	-0,0006	0,0043	0,0044	-	226	0,01	-0,0046	0,0376	0,0043	-	233
	0,002	-0,0006	0,0063	-	0,670	3	0,02	-0,0033	0,0561	-	0,891	9
	0,003	-0,0004	0,0081	-	0,548	2	0,03	0,0001	0,0725	-	0,606	3
	0,005	0,0004	0,0113	-	0,167	1	0,05	0,0098	0,1022	-	0,124	1
	0,01	0,0032	0,0186	-	0,000	1	0,10	0,0417	0,1690	-	0,000	1
Beta	0,001	0,00001	0,0052	0,0027	-	370	0,01	0,00026	0,0440	0,0027	-	370
	0,002	0,00014	0,0072	-	0,000	1	0,02	0,00234	0,0623	-	0,850	7
	0,003	0,00041	0,0090	-	0,000	1	0,03	0,00593	0,0785	-	0,552	2
	0,005	0,00122	0,0122	-	0,000	1	0,05	0,01568	0,1080	-	0,079	1
	0,01	0,00400	0,0195	-	0,000	1	0,10	0,04728	0,1739	-	0,000	1

Fonte: Elaborado pelo autor (2009)

A análise de sensibilidade para $p = 0,1$ e $n = 50$ mostra que as cartas de Ryan (1989) e Chen (1998) apresentam para um processo sob controle, um desempenho inferior a Carta Beta na detecção de alterações, conforme Tabela 11. Isto pode ser devido à modificação proposta pelos autores possuir propriedade de correção com assimetria positiva (WILL et al., 1968). Estas cartas de controle apresentam similaridade e melhor desempenho para um processo sob controle e na detecção de alterações para $p = 0,1$ e $n > 200$, conforme Tabela 11. Segundo Chen (1998, p. 150), os resultados utilizando os limites da correção de Cornish-

Fisher são similares com os resultados de Ryan (1989), entretanto, deve ser preferível devido a sua base teórica mais consistente e maior dependência de p .

Observa-se nas Tabelas 10 e 11, que a carta Beta apresentou $NMA_0 = 370$ para a probabilidade 0,9973 de um processo sob controle ($\alpha = 0,0027$) e maior rapidez na detecção de alterações, com menores valores de NMA_1 para os dois cenários testados.

A Carta Beta apresenta sensibilidade satisfatória, pois permite estimar limites desejáveis para o processo sob controle e apresenta os menores valores de NMA_1 em todas as alterações de descontrole, quando comparado com as outras cartas investigadas.

Tabela 11: Comparação do NMA entre as cartas de controle de Shewhart, Ryan (1989), Chen (1998) e Beta para $p_0 = 0,10$, $n = 50$ e 300 , $\lambda = 0,02$

	$n = 50$						$n = 300$					
	p	LIC	LSC	α	γ	NMA	p	LIC	LSC	α	γ	NMA
Shewhart	0,10	-0,0273	0,2273	0,0032	-	311	0,10	0,0480	0,1520	0,0030	-	335
	0,12	-0,0179	0,2579	-	0,986	74	0,12	0,0637	0,1763	-	0,951	20
	0,15	-0,0015	0,3015	-	0,937	16	0,15	0,0882	0,2118	-	0,540	2
	0,20	0,0303	0,3697	-	0,711	3	0,20	0,1307	0,2693	-	0,016	1
	0,30	0,1056	0,4944	-	0,139	1	0,30	0,2206	0,3794	-	0,000	1
	Ryan	0,10	-0,0023	0,2503	0,0010	-	995	0,10	0,0522	0,1558	0,0027	-
0,12		0,0071	0,2809	-	0,995	195	0,12	0,0679	0,1801	-	0,938	16
0,15		0,0235	0,3245	-	0,970	33	0,15	0,0923	0,2157	-	0,503	2
0,20		0,0553	0,3927	-	0,814	5	0,20	0,1349	0,2731	-	0,023	1
0,30		0,1306	0,5174	-	0,223	1	0,30	0,2248	0,3832	-	0,000	1
Chen		0,10	-0,0059	0,2486	0,0010	-	995	0,10	0,0516	0,1555	0,0027	-
	0,12	0,0024	0,2781	-	0,995	195	0,12	0,0671	0,1797	-	0,938	16
	0,15	0,0172	0,3202	-	0,970	33	0,15	0,0913	0,2150	-	0,503	2
	0,20	0,0463	0,3857	-	0,814	5	0,20	0,1334	0,2719	-	0,023	1
	0,30	0,1162	0,5051	-	0,223	1	0,30	0,2224	0,3812	-	0,000	1
	Beta	0,10	0,01630	0,2661	0,0027	-	370	0,10	0,05530	0,1588	0,0027	-
0,12		0,02444	0,2942	-	0,972	35	0,12	0,07066	0,1828	-	0,856	7
0,15		0,03859	0,3342	-	0,907	11	0,15	0,09461	0,2178	-	0,325	1
0,20		0,06620	0,3965	-	0,660	3	0,20	0,13635	0,2743	-	0,002	1
0,30		0,13221	0,5098	-	0,105	1	0,30	0,22456	0,3825	-	0,000	1

Fonte: Elaborado pelo autor (2009)

7. Conclusão

O objetivo deste artigo foi propor uma carta de controle Beta (Carta Beta) para monitorar processos que mensuram características de qualidade do tipo fração. Esta carta de controle assume a distribuição de probabilidade Beta para a fração do tipo percentual ou proporção, a qual apresenta valores restritos ao intervalo $[0,1]$.

A Carta Beta foi aplicada no monitoramento de características de qualidade de três exemplos aplicados. A fim de realizar uma análise comparativa, aplicaram-se também as cartas propostas por Shewhart, Ryan (1989) e Chen (1998) nas características de qualidade

monitoradas pela Carta Beta. Esta análise permitiu avaliar que a aproximação da distribuição Binomial pela distribuição Beta foi mais adequada, fornecendo valores restritos ao intervalo [0,1].

A análise de sensibilidade entre as cartas de controle investigadas consistiu no cálculo do NMA_0 e NMA_1 de cada carta de controle para um processo sob controle e outro fora de controle, respectivamente. Os resultados da análise indicaram que a Carta Beta apresenta desempenho satisfatório, pois gerou maior valor de NMA_0 para a probabilidade de alarme falso conforme o valor nominal desejado e menor valor de NMA_1 quando comparado com as outras cartas de controle. A Carta Beta detecta mais rapidamente todas as alterações induzidas, apresentando melhor desempenho entre as cartas analisadas.

A análise dos limites de controle propostos por Shewhart, Ryan (1989) e Chen (1998) apresenta limites de controle aceitáveis quando $np \geq 9$, $np \geq 6 | p > 0,03$ e $np \geq 6 | p > 0,015$ respectivamente. A análise comparativa permitiu verificar que os limites de controle da Carta Beta apresentam desempenho superior às demais cartas, comprovando que a aproximação pela distribuição Beta é superior quando comparada com a aproximação pela distribuição Normal.

Conclui-se que a Carta Beta pode ser uma ferramenta adequada para o monitoramento de características de qualidade do tipo fração restritas ao intervalo [0,1], permitindo limites de controle assimétricos e taxa de alarme falso desejada.

Agradecimentos

Esta pesquisa foi apoiada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/Brasil).

Referências

- (1) BOURKE, P.D. Detecting a shift in fraction nonconforming using Run-Length control charts with 100% Inspection, *Journal of Quality Technology*, Milwaukee, v. 23, n. 3, p. 225–238, 1991.
- (2) BOURKE, P.D. Performance comparisons for the Synthetic control charts for detecting increases in fraction nonconforming, *Journal of Quality Technology*, Milwaukee, v. 40, n. 4, p. 461–475, 2008.
- (3) BROWNEE, K.A. *Statistical theory and methodology in science and engineering*, New York: John Wiley & Sons, 1965, 590p.
- (4) CHEN, G. An improved p Chart through simple adjustments, *Journal of Quality Technology*, Milwaukee, v. 30, n. 2, p. 142–151, 1998.

- (5) COLLETT, D. *Modelling binary data*, 2^a ed., London: Chapman-Hall, 2003, 387p.
- (6) DRAPER, N.R.; SMITH, H. *Applied Regression Analysis*, New York: John Wiley & Sons, 1998, 706p.
- (7) FERRELL, E. B. Control charts for log-normal universe, *Industrial Quality Control*, Québec, v. 15, n. 1, p. 4–6, 1958.
- (8) FLEISS, J.L., LEVIN, B.; PAIK, M.C. *Statistical methods for Rates and Proportions*, 3^a ed., New Jersey: John Wiley & Sons, 2003, 760p.
- (9) HEIMANN, P.A. Attributes control charts with large sample size, *Journal of Quality Technology*, Milwaukee, v. 28, n. 4, p. 451–459, 1996.
- (10) JOHNSON, N.L.; KOTZ, S.; BALAKRISHNAN, N. *Continuous Univariate Distributions*, v. 2, 2^a ed., New York: John Wiley & Sons, 1995, 752p.
- (11) KAMINSKY, F.C.; BENNEYAN, J.C. DAVIS, R.D.; BURKE, R. Statistical control charts based on a Geometric distribution, *Journal of Quality Technology*, Milwaukee, v. 24, n. 2, p. 63–69, 1992.
- (12) McCOLL, J.I.; MOTLEY, T.J. Control charts applicable when the fraction nonconforming is small, *Journal of Quality Technology*, Milwaukee, v. 30, n. 3, p. 240–247, 1998.
- (13) MOITRA, S.D. Skewness and the Beta distribution, *Journal of the Operational Research Society*, Houndmills, v. 41, n. 10, p. 953–961, 1990.
- (14) MONTGOMERY, D.C. *Introduction to Statistical Quality Control*, 5^a ed., New York: John Wiley & Sons, 2005, 732p.
- (15) MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. *Introduction to the Theory of Statistics*, London: McGraw-Hill, 1986, 564p.
- (16) NELSON, P. R. Control charts for Weibull processes with standards given, *IEEE Transactions on Reliability*, Knoxville, v. 28, n. 3, p. 383–387, 1979.
- (17) PERRY, M. B.; PIGNATIELLO, J.J Jr.; SIMPSON, J.R. Estimating the change point of the Process Fraction nonconforming with a Monotonic change Disturbance in SPC. *Quality and Reliability Engineering International*, Chichester, v. 23, n. 3, p. 327–339, 2007.
- (18) QUESENBERY, C.P. On properties of Binomial Q Charts for Attributes, *Journal of Quality Technology*, Milwaukee, v. 27, n. 3, p. 204–213, 1995.
- (19) QUESENBERY, C.P. SPC Q charts for a Binomial parameter p : short and long runs, *Journal of Quality Technology*, Milwaukee, v. 23, n. 3, p. 239–246, 1991.
- (20) R Development Core Team. *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Disponível em <http://www.r-project.org>, ISBN 3-900051-07-0, 2009.
- (21) REYNOLDS, M.R.; STOUMBOS, Z.G. A CUSUM charts for monitoring a proportion when inspecting continuously, *Journal of Quality Technology*, Milwaukee, v. 31, n. 1, p. 87–108, 1999.
- (22) REYNOLDS, M.R.; STOUMBOS, Z.G. A general approach to modeling CUSUM charts for a proportion, *IIE Transactions*, Philadelphia, v. 31, n. 1, p. 87–108, 2000.
- (23) ROCKE, D.M. On the Beta transformation family, *Technometrics*, Washington, v. 35, n. 1, p. 72–81, 1993.
- (24) RYAN, T.P. *Statistical Methods for Quality Improvement*, New York: John Wiley & Sons, 1989, 348p.

- (25) RYAN, T.P.; SCHWERTMAN, N.C. Optimal limits for attributes control charts, *Journal of Quality Technology*, Milwaukee, v. 29, n. 1, p. 86–98, 1997.
- (26) SCHADER, M.; SCHMID, F. Two rules of thumb for the approximation of the Binomial distribution by Normal distribution, *The American Statistician*, Alexandria, v. 43, n. 1, p. 23-24, 1989.
- (27) SIM, C.H.; LIM, M.H. Attribute charts for zero-inflated process, *Communications in Statistics – Simulations and Computation*, Hamilton, v. 37, n. 7, p. 1440–1452, 2008.
- (28) SOONG, T.T. *Fundamentals of Probability and Statistics for Engineers*, New York: John Wiley & Sons, 2004, 391p.
- (29) WILL, G.W.; DAVIS, A.W. Generalized asymptotic expansions of Cornish-Fisher type. *The Annals of Mathematical Statistics*, Chapel Hill, v. 39, n. 4, p. 1264–1273, 1968.
- (30) WINTERBOTTOM, A. Simple adjustments to improve control limits on attribute charts, *Quality and Reliability Engineering International*, Chichester, v. 9, n. 2, p. 105–109, 1993.
- (31) YANG, Z.; KURALMANI, V.; XIE, M.; TSUI, K-L. On the performance of geometric charts with estimated control limits, *Journal of Quality Technology*, Milwaukee, v. 34, n. 4, p. 448–458, 2002.
- (32) ZHANG, C.W.; XIE, M.; GOH, T.N.; LIU, J.Y. A control chart for the Gamma distribution as a model of time between events, *International Journal of Production Research*, London, v. 45, n. 23, p. 5649–5666, 2007.

2.4 ARTIGO 4

**Carta de Controle baseada em Modelos Beta para Monitoramento de Características de
Qualidade do tipo Fração**

**Beta Model-based Control Charts for monitoring of the type Fraction quality
characteristics**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

Previsão de envio ao International Journal of Production Research (ISSN 0020-7543)

Carta de Controle baseada em Modelos Beta para monitoramento de Características de Qualidade do tipo Fração

Ângelo Márcio Oliveira Sant'Anna
Carla Schwengber ten Caten

Universidade Federal do Rio Grande do Sul, Brasil
angelo@producao.ufrgs.br
tencaten@producao.ufrgs.br

Resumo

As cartas de controle baseada em modelos são ferramentas úteis para monitorar processos quando as características de qualidade (CQ) variam em função de freqüentes alterações de ajuste em uma ou mais variáveis de controle (VC) do processo. Estas cartas de controle comumente utilizam o modelo de regressão múltipla, no entanto, este não é adequado para a modelagem de CQ do tipo fração por pelo menos duas razões: (i) pressupõe que as frações seguem a distribuição Normal, o que não é correto; e (ii) possibilita a previsão de valores fora do intervalo $[0,1]$. Este artigo propõe uma carta de controle baseada em modelos Beta (CCMB) para monitorar CQ do tipo fração em função das VC do processo. Esta carta de controle utiliza o modelo de regressão Beta e monitora os resíduos *deviance* do modelo Beta. A CCMB foi aplicada no monitoramento das CQ de um processo de fabricação de pneus. Desenvolveu-se um estudo comparativo e uma análise de sensibilidade da CCMB e das cartas propostas por Hawkins (1991) e Haworth (1996). Conclui-se que a CCMB apresenta melhor desempenho para monitorar CQ do tipo fração entre as cartas de controle analisadas.

Palavras-chave: Controle estatístico de qualidade, Carta de controle baseada em modelos e distribuição Beta .

Beta Model-based Control Charts for monitoring of the type Fraction Quality Characteristics

Abstract

Model-based control charts are useful tools for monitoring processes when the quality characteristics (QC) frequently vary with adjustments in one or more control variables (CV) of the process. These control charts commonly use multiple regression model, however, this model is not adequate for modeling the type fraction QC for at least two reasons: (i) fractions are assumed to follow a Normal distribution, which is not correct, and (ii) predicted responses will not necessarily be confined in the $[0,1]$ -interval. In this paper we propose a Beta model-based control charts (BMCC) to monitor the type fraction QC according to the VC. This control chart uses the Beta regression model and monitors *deviance* residuals from the Beta model. The BMCC was applied in monitoring the QC of a tire manufacturing process. A comparative study and sensitivity analysis of BMCC and an analysis of the charts proposed by Hawkins (1991) and Haworth (1996) were conducted. It is concludes that the BMCC performs best at monitoring of the type fraction QC among the control charts analyzed.

Keyword: Statistical quality control, model-based control charts, Beta distribution.

1. Introdução

Observa-se na literatura que o controle estatístico de processo (CEP) vem sendo utilizado para monitorar a qualidade dos produtos em diversos processos industriais, devido principalmente a sua simplicidade operacional. Tradicionalmente utilizam-se cartas de controle para monitorar as características de qualidade (CQ) do processo.

Segundo Montgomery (2005), a utilização das cartas de controle requer que os dados monitorados sejam independentes e identicamente distribuídos em torno de um parâmetro de referência com média constante. No entanto, muitas vezes as CQ variam frequentemente em função de alterações do ajuste de uma ou mais variáveis de controle (VC) do processo, fazendo com que o parâmetro de referência seja diferente para cada novo ajuste, o que implica no uso de cartas de controle específicas para monitorar cada ajuste das VC, dificultando a operacionalização do CEP.

Uma alternativa proposta na literatura são as cartas de controle baseadas em modelos, onde inicialmente é realizada a previsão do valor da CQ por um modelo de regressão em função do seu relacionamento com as VC do processo. Ao invés de monitorar a média das CQ, as cartas de controle baseadas em modelos monitoram os resíduos entre as CQ observadas e as CQ previstas pelo modelo.

O monitoramento das CQ utilizando as cartas de controle baseadas em modelos apresenta algumas vantagens: (i) operacionalização da carta de controle é relativamente simples; (ii) identificação da influência das variáveis de controle; (iii) o controle estatístico é baseado nos resíduos do modelo de regressão que usualmente são não-correlacionados no tempo; e (iv) permite monitorar alterações na média da CQ e nos coeficientes das variáveis de controle (SKINNER et al., 2003; JEARKPAPORN et al., 2005; SHU et al., 2008).

A estratégia de monitorar uma CQ relacionada a uma VC foi inicialmente proposta por Mandel (1969), chamada de carta de controle de regressão. Hawkins (1991, 1993) propôs a carta de controle de ajustamento de regressão estendendo a proposta de Mandel para monitorar processos com múltiplas variáveis de controle. Haworth (1996) propôs limites de controle baseados na estatística *t-student* para a carta de controle proposta por Hawkins. Wade e Woodal (1993) apresentam uma revisão sobre carta de controle de seleção de causas desenvolvida por Zhang (1985) e propõem intervalos de predição para os limites de controle. Hauck et al. (1999) estendem a idéia de Zhang numa aplicação de processos em cascata. Loredó et al. (2002) propuseram o uso da carta de controle de ajustamento de regressão em processos autocorrelacionados utilizando para os limites de controle a amplitude móvel dos

resíduos padronizados. Shu et al. (2004) propõe um tamanho mínimo de amostra necessário para o monitoramento eficiente das cartas de controle de regressão, ajustamento de regressão e seleção de causas.

As cartas de controle propostas pelos autores utilizam os modelos de regressão linear múltipla para a previsão das CQ, o que requer que as mesmas sejam normalmente distribuídas e homoscedásticas, suposições muitas vezes não satisfeitas em processos industriais, quando se monitora CQ do tipo atributos. Skinner et al. (2003) propuseram a carta de controle baseada em modelos para monitorar processos com CQ que seguem a distribuição de Poisson. Skinner et al. (2004) estendem a aplicação do método para monitorar CQ com distribuição de Poisson e superdispersão.

O uso do modelo de regressão linear múltipla na modelagem de atributos como fração de produtos não conformes não é adequado, pois pressupõe que a CQ segue a distribuição Normal, o que não é correto, e possibilita a previsão de valores fora do limite do intervalo $[0,1]$. Uma alternativa para modelagem da QC do tipo fração é o modelo de regressão Beta (KIESCHNICK e McCULLOGH, 2003). Este modelo de regressão baseia-se na suposição de que a fração segue a distribuição Beta, pois suas mensurações assumem valores restritos ao intervalo $[0,1]$.

Este artigo propõe uma carta de controle baseada em modelos Beta (CCMB) para monitorar CQ do tipo fração em função das variáveis de controle do processo. Esta carta de controle utiliza o modelo de regressão Beta e monitora os resíduos *deviance* do modelo Beta.

A CCMB proposta foi aplicada no monitoramento de CQ do tipo fração em um processo de fabricação de pneus. A fim de realizar uma análise comparativa, aplicaram-se também as cartas propostas por Hawkins (1991) e Haworth (1996) nas CQ monitoradas pela CCMB. Também foi realizado um estudo de simulação para avaliar a sensibilidade das cartas comparadas anteriormente.

Na seção 2, é apresentada uma breve revisão sobre terminologias empregadas na literatura, distribuição de probabilidade Beta e modelos de regressão Beta. Na seção 3, são apresentados os conceitos da CCMB. Na seção 4, é apresentada a aplicação da CCMB e a comparação entre as cartas de controle. A seção 5 apresenta um estudo de simulação do desempenho do NMA. Finalmente, na seção 6, são apresentadas as conclusões sobre o estudo desenvolvido nesse artigo.

2. Referencial teórico

2.1 Categorias do termo fração ou razão

Muitos estudos em diferentes áreas do conhecimento avaliam características de qualidade com observações expressas no intervalo $[0,1]$ sem definir adequadamente a terminologia da variável em análise. A Figura 19 apresenta a classificação das terminologias empregadas para definir o termo fração ou razão. A primeira categoria percentual compreende a razão entre dois números discretos; por exemplo, o número de peças defeituosas dividido pelo número total de peças do lote. A segunda categoria proporção compreende a razão entre dois números contínuos; por exemplo, o volume de óleo bruto convertido em gasolina dividido pelo volume total.

Termo	Categoria	Forma	Exemplo
Fração (ou Razão)	Percentual	discreto	$\frac{\text{n}^\circ \text{ de peças defeituosas}}{\text{n}^\circ \text{ total de peças}}$
		discreto	$\frac{\text{n}^\circ \text{ total de peças}}{\text{n}^\circ \text{ total de peças}}$
	Proporção	contínuo	$\frac{\text{volume sob pesquisa}}{\text{volumetotal}}$
		contínuo	$\frac{\text{volumetotal}}{\text{volumetotal}}$

Figura 19: Categorias do termo fração ou razão

Fonte: Elaborado pelo Autor

2.2. Distribuição Beta

A ocorrência da fração do tipo percentual (discreto/discreto) de um processo produtivo tem sido comumente modelada como distribuição Binomial, com parâmetros n e p . Conforme o teorema do limite central, se n for satisfatoriamente grande a aproximação da distribuição Binomial pela Normal é satisfatória. Segundo Schader e Schmid (1989), a aproximação da distribuição Binomial pela distribuição Normal é satisfatória caso duas regras sejam satisfeitas: (i) $np(1 - p) \geq 9$ e (ii) $np \geq 5$ quando $0 < p \leq 0,5 \leq (1 - p)$. Conforme Montgomery (2005), a aproximação da distribuição Binomial pela Normal é satisfatória quando $np \geq 10$ e p está no intervalo entre $(0,1 \leq p \leq 0,9)$.

Entretanto, há processos em que a fração do tipo percentual de itens é muito pequena ou a distribuição dos dados é assimétrica, tornando inadequada a aproximação.

A fração do tipo percentual é aquela onde y_i representa o número de itens não conformes em um tamanho de amostra n_i itens, $i = 1, 2, \dots, m$. Assim, $Y_i \sim \text{Bin}(n_i; p_i)$; onde $p_i = y_i / n_i$ é o percentual de itens não conformes. Para Moitra (1990); Rocke (1993) e Johnson et al. (1995a), a fração de sucessos de um referido evento é uma variável aleatória com

mensurações positivas e restritas ao intervalo $[0,1]$, a qual pode ser aproximada por uma distribuição de probabilidade Beta, indexada por dois parâmetros $\theta_1, \theta_2 > 0$. A aproximação da distribuição Binomial pela Beta pode ser mais apropriada, devido à variedade de forma que sua função densidade de probabilidade pode assumir.

A fração do tipo proporção (contínuo/contínuo) pode ser modelada como distribuição Beta por esta ser uma distribuição contínua e seus parâmetros não dependerem de n . Seja Y_i a proporção de sucesso de um referido evento dado um espaço amostral. Assim, $Y_i \sim \text{Beta}(p_i; q_i)$; onde $p_i = y_i$ é a proporção de volume sob pesquisa em relação ao volume total. Moitra (1990) relata que a distribuição de probabilidade Beta é extremamente versátil e encoraja sua aplicação em modelagem de uma variedade de CQ devido a sua flexibilidade. Rocke (1993) apresenta um estudo aplicando uma família de transformações para a distribuição Beta conforme diferentes formas assimétricas da densidade. Johnson et al. (1995b) discutem diversos estudos de diferentes ciências em que a distribuição Beta foi melhor ajustada frente a diversas distribuições alternativas, e sugerem que a distribuição Beta é uma boa distribuição para representar dados em fração.

A família de distribuições Beta é composta por todas as distribuições de probabilidade que apresente uma variável aleatória Y cuja função densidade de probabilidade (FDP) depende dos parâmetros θ_1 e θ_2 , e que sua FDP pode ser escrita na forma,

$$f(y; \theta_1, \theta_2) = \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} y^{\alpha-1} (1 - y)^{\beta-1} \quad (51)$$

sendo $\Gamma(\theta)$ uma função *Gama* avaliada no ponto θ , ou seja, com $\Gamma(\theta) = \int_0^\infty y^{\theta-1} e^{-y} dy$, $\theta > 0$.

O foco do artigo é a distribuição Beta de dois parâmetros, pois é a distribuição mais comumente utilizada para modelar dados em fração (JOHNSON et al., 1995b).

Segundo Johnson et al. (1995b) uma variável aleatória Y com distribuição Beta de dois parâmetros possui média e variância dadas por, respectivamente,

$$E(Y) = \frac{\theta_1}{\theta_1 + \theta_2} \quad e \quad Var(Y) = \frac{\theta_1 \theta_2}{(\theta_1 + \theta_2)^2 \cdot (\theta_1 + \theta_2 + 1)} \quad (52)$$

2.3. Modelo de Regressão Beta

O modelo de regressão Beta (MRB), proposto por Ferrari e Cribari-Neto (2004), pode ser utilizado para modelar a relação linear ou não-linear de uma CQ do tipo fração em função das VC do processo.

O MRB assume que a CQ tem distribuição Beta de dois parâmetros, e a estrutura do modelo de regressão apresenta os parâmetros de média e dispersão definidos por: $\mu = \frac{\theta_1}{\theta_1 + \theta_2}$ e $\phi = \theta_1 + \theta_2$, respectivamente, isto é, $\theta_1 = \mu\phi$ e $\theta_2 = (1 - \mu)\phi$. Então, a equação (52), fica escrita como,

$$E(Y) = \mu \quad e \quad Var(Y) = \frac{Var(\mu)}{(1 + \phi)} \quad (53)$$

em que $Var(\mu) = \mu(1 - \mu)$ denota a função de variância do modelo. A densidade da variável aleatória Y descrita na equação (51) pode ser reescrita como,

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1 - \mu)\phi)} y^{\mu\phi - 1} (1 - y)^{(1 - \mu)\phi - 1} \quad (54)$$

em que $0 < y, \mu < 1$; $\phi > 0$.

O MRB é representado por uma estrutura de regressão na forma,

$$g(\mu) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (55)$$

ou

$$\mu = E(Y) = g^{-1}(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) \quad (56)$$

sendo μ a média da variável resposta Y , β s os coeficientes desconhecidos, x_1, \dots, x_k as variáveis de controle e $g(\cdot)$ a função de ligação. Para dados com distribuição Beta, as funções de ligação mais comumente disponíveis são a logit, a probit, a log-log e a complemento log-log. Essas quatro funções de ligações apresentadas são contínuas e estritamente crescentes no intervalo unitário $[0,1]$.

Os coeficientes β s e ϕ são obtidos através do procedimento iterativo de estimação por máxima verossimilhança. Este procedimento é realizado pelo algoritmo de otimização não-linear de Broyden–Fletcher–Goldfarb–Shanno (BFGS), também conhecido como *quasi-Newton method*. Segundo Ferrari e Cribari-Neto (2004), o algoritmo BFGS fornece melhor desempenho na estimação dos parâmetros desconhecidos para diferentes tamanhos de amostra e formas da distribuição Beta.

Seja Y uma variável aleatória com distribuição Beta, o logaritmo da estatística de verossimilhança apresenta a forma geral,

$$l(\beta, \phi) = \sum_{i=1}^n l_i(\mu_i, \phi) \quad (57)$$

onde $l_i(\mu_i, \phi) = \log \Gamma(\phi) - \log \Gamma(\mu_i \phi) - \log \Gamma((1 - \mu_i) \phi) + (\mu_i \phi - 1) \log y_i + [(1 - \mu_i) \phi] \log(1 - y_i)$

com μ_i definida de tal forma que satisfaz a equação $\mu_i = g^{-1}(\beta_t \mathbf{X}_t)$, que é a média da variável Y . A estatística de verossimilhança é definida pela derivada da função de log-verossimilhança em relação aos parâmetros desconhecidos (β, ϕ) do modelo ajustado, isto é,

$$\frac{\partial l_i(\beta, \phi)}{\partial \beta} = \sum_{i=1}^s \frac{\partial l_i(\mu_i, \phi)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta} \quad (58)$$

e

$$\frac{\partial l_i(\beta, \phi)}{\partial \phi} = \sum_{j=1}^s \frac{\partial l_i(\mu_i, \phi)}{\partial \phi} \quad (59)$$

Note que $\frac{\partial \mu_i}{\partial \eta_i} = \frac{1}{g'(\mu_i)}$; $\frac{\partial \eta_i}{\partial \beta} = x_t$ e também,

$$\frac{\partial l_i(\mu_i, \phi)}{\partial \mu_i} = \phi \left\{ \log \frac{y_i}{1 - y_i} - [\delta(\mu_i \phi) - \delta(1 - \mu_i) \phi] \right\} \quad (60)$$

em que $\delta(\cdot)$ é uma função diferencial *Gama*, i.e. $\delta(z) = \frac{\partial \log \Gamma(z)}{\partial z}$, $z > 0$. Fazendo $y_i^* = \log \frac{y_i}{1 - y_i}$ e $\mu_i^* = [\delta(\mu_i \phi) - \delta(1 - \mu_i) \phi]$, pode-se reescrever como,

$$\frac{\partial l_i(\mu_i, \phi)}{\partial \mu_i} = \phi \{ y_i^* - \mu_i^* \} \quad (61)$$

A estatística de verossimilhança do modelo de regressão é utilizada para obtenção dos resíduos *deviance*. Os resíduos *deviance* são usados na carta de controle baseada em modelos para o monitoramento e controle do processo a fim de detectar a mudança na média ou na variabilidade.

Uma etapa essencial na análise do ajuste de modelos de regressão é a verificação da violação das suposições feitas para o modelo, por exemplo, (i) avaliar a adequação da distribuição de probabilidade proposta para a CQ; (ii) avaliar a capacidade preditiva; e (iii) analisar os resíduos para detectar a presença de pontos extremos (*outliers*). Modelos com bom desempenho estatístico apresentam pequena discrepância entre os dados reais e seus respectivos valores preditos.

Neste trabalho serão utilizados como critérios de diagnóstico de modelos, o coeficiente de determinação R^2 , o critério de informação de Akaike (AIC), que foi desenvolvido através dos estimadores de máxima verossimilhança (EMV) para decidir o melhor modelo ajustado com um número reduzido de parâmetros e o resíduo *deviance*, que é baseado na estatística de verossimilhança. Mais detalhes sobre critérios de diagnóstico podem ser vistos em Sant'Anna e Caten (2009).

3. Carta de controle baseada em modelos Beta

A carta de controle baseada em modelos Beta (CCMB) proposta neste trabalho tem por finalidade monitorar processos com CQ mensuradas em fração ou razão, as quais habitualmente seguem distribuições não-Normais e de forma assimétricas. Esta carta de controle utiliza o MRB para estimar a CQ do tipo fração em função das VC. Os resíduos *deviance* gerados pelo método de estimação por máxima verossimilhança são *plotados* na carta de controle a fim de monitorar o comportamento do processo e detectar mudanças devido a causas especiais.

A CCMB apresenta algumas vantagens em relação às outras CC: (i) a CQ prevista é restrita no intervalo [0,1]; (ii) os resíduos *deviance* monitorados são não-correlacionados; e (iii) o procedimento de estimação dos coeficientes admite relação não-linear entre as variáveis de interesse.

Hawkins (1991) salienta que, conforme o modelo de regressão é ajustado, o desempenho das cartas de controle é aperfeiçoado. Segundo Jearkpaporn et al. (2005), o ajuste inadequado do modelo de regressão pode afetar o desempenho das cartas de controle no monitoramento do processo.

Suponha um conjunto de s variáveis aleatórias Beta distribuídas Y_1, Y_2, \dots, Y_s em função de um conjunto de k variáveis de controle x_1, x_2, \dots, x_k com $k+1$ coeficientes desconhecidos $\beta_1, \beta_2, \dots, \beta_k$. Assumindo que uma função de ligação logit relaciona tais variáveis, tem-se que,

$$Y_j \sim \text{BETA}[\text{logit}(\mu_j) = \boldsymbol{\beta}_{jt} \mathbf{X}_t]; \quad \text{com } j = 1, \dots, s; \quad t = 0, \dots, k \quad (62)$$

em que cada Y_j são variáveis aleatórias Beta distribuídas com média $\mu_j = E(Y_j)$ e função de ligação inversa $g^{-1}(\cdot) = \log[\mu / (1 - \mu)]$ que modela a média de cada variável aleatória j , e

$$\mu_j = E(Y_j) = g^{-1}(\boldsymbol{\beta}_{jt} \mathbf{X}_t) = \frac{e^{\beta_{j0} + \beta_{j1}x_1 + \dots + \beta_{jk}x_k}}{1 + e^{\beta_{j0} + \beta_{j1}x_1 + \dots + \beta_{jk}x_k}} \quad (63)$$

assim, os coeficientes $\boldsymbol{\beta}$ s do modelo de regressão tem uma importante interpretação que é a razão de chances (*odds ratio*), definida por $e^{\beta_{jt}}$.

Observa-se que a relação entre as variáveis Y_j e um dado conjunto de VC X_t é independente e muitas vezes não-linear. O método proposto monitora de forma univariada a mudança na média das variáveis aleatórias Y_j dado uma mudança em qualquer ajuste das VC X_t através do resíduo *deviance* baseado na estatística de verossimilhança.

Pierce e Schafer (1986) descrevem que os resíduos *deviance* seguem a distribuição Normal assintótica. Ferrari e Cribari-Neto (2004) discutem que a distribuição dos resíduos *deviance* do modelo Beta possui propriedades similares aos resíduos *deviance* dos modelos lineares generalizados. Skinner et al. (2004) afirmam que os resíduos *deviance* são sensíveis aos efeitos de diferentes tipos de mudança na média ou na variabilidade imposta aos dados. Esta característica dos resíduos proporciona a carta de controle baseada em modelos vantagens no monitoramento e controle de processos.

O resíduo *deviance* (r_d) para os j modelos de regressão Beta é estimado por,

$$r_{d_{ij}} = \text{sign}(y_{ij} - \hat{\mu}_{ij}) \left\{ 2 \left[\phi \left(\log \frac{y_{ij}}{1 - y_{ij}} - [\delta(\hat{\mu}_{ij}\phi) - \delta(1 - \hat{\mu}_{ij})\phi] \right) \right] \right\}^{1/2} \quad (64)$$

em que a estatística de verossimilhança é utilizada como uma medida de ponderação para os resíduos ordinários ($y_{ij} - \hat{\mu}_{ij}$).

Os resíduos *deviance* são *plotados* em j cartas de controle univariadas de Shewhart, com limites de controle estatísticos calculados a partir da média \bar{r}_{d_j} e variância $s^2(\bar{r}_{d_j})$ amostral dos resíduos. Desta forma, os limites de controle e linha central da CCMB são dados por,

$$\begin{aligned} LSC_j &= \bar{r}_{d_j} + w \sqrt{s^2(\bar{r}_{d_j})} \\ LC_j &= \bar{r}_{d_j} \\ LIC_j &= \bar{r}_{d_j} - w \sqrt{s^2(\bar{r}_{d_j})} \end{aligned} \quad (65)$$

onde w é uma constante que define a largura dos limites de controle correspondente ao nível desejado do número médio de amostras até alarme falso (NMA_0) e $j = 1, \dots, s$. O NMA_0 determina o número de amostras até a ocorrência de um alarme falso quando o processo está sob controle. Uma vez determinado os limites de controle, as CCMB com resíduos *deviance* podem ser utilizadas para monitorar as CQ mensuradas em frações restritas no intervalo $[0,1]$.

3.1. Etapas para aplicação da CCMB

O procedimento para aplicação da CCMB consiste em *plotar* os resíduos *deviance* em j cartas de controle univariadas de Shewhart para monitorar j CQ do processo ao longo do tempo. Conforme Woodall (2000), muitos artigos apresentam o monitoramento do processo por cartas de controle negligenciando a Fase I e as suposições vitalmente importantes das CQ investigadas.

A implantação da CCMB deve ser dividida em duas fases. A fase de análise retrospectiva (Fase I) e a fase de monitoramento do processo (Fase II). Na Fase I busca-se analisar o processo e estimar os parâmetros a serem usados para determinar os limites de controle para a Fase II. Na Fase II, deve-se assumir que os valores dos parâmetros são conhecidos e os dados a serem monitorados seguem a mesma distribuição que os da Fase I.

As etapas de execução da Fase I consistem em:

- (1) Coleta das j CQ(Y_j) e VC (x_t) do processo industrial;
- (2) Análise exploratória dos dados históricos;
- (3) Ajuste dos j modelos de regressão Beta (MRB) que relacionam as j CQ com as t VC. O ajuste dos j MRB deve considerar a escolha da função de ligação mais adequada e dos critérios de diagnóstico;
- (4) Obtenção dos resíduos *deviance* dos j MRB a partir da equação (64) ;
- (5) Cálculo dos limites de controle das j CCMB segundo estimativas de média e variância amostral dos resíduos *deviance* dos j MRB conforme equação (65);

As etapas de execução da Fase II consistem em:

- (6) Coleta de novas observações das j CQ (Y_j) e t VC (x_t).
- (7) Estimção dos novos resíduos *deviance* dos j MRB a partir da equação (64);
- (8) Monitoramento dos resíduos *deviance* em j CCMB. Qualquer ponto fora dos limites de controle indica a presença de causas especiais no processo.

4. Aplicação da CCMB

A CCMB foi aplicada a um processo de manufatura de pneus radiais de uma empresa multinacional de borrachas. As características de qualidade a serem monitoradas no processo investigado são: percentual da variação força radial não conforme (y_1), percentual da espessura radial não conforme (y_2) e proporção de massa não convertida (y_3), que são mensuradas no final do processo de manufatura. As variáveis y_1 e y_2 são obtidas a partir da razão entre o número de pneus não conformes às especificações e o número total de pneus do lote. A variável y_3 é obtida pela razão entre o volume de matéria-prima não convertida em produto dividida pelo volume total, estando diretamente relacionada à perda de matéria-prima.

As variáveis de controle deste processo são: comprimento da banda de rodagem (x_1); comprimento da seção (x_2); comprimento do flanco (x_3); postamento do friso (x_4); postamento da cova (x_5).

Os dados utilizados na estimação dos j MRB foram obtidos através de um experimento realizado na empresa. Neste experimento, utilizou-se um projeto fatorial fracionado com as variáveis de controle ajustadas em dois níveis (2^{5-1}) mais dois pontos centrais resultando em 18 ensaios experimentais, com 30 pneus em cada ensaio. Este projeto experimental objetivou estimar a variabilidade (erro puro) do processo sem interferir nas estimativas dos coeficientes do modelo de regressão. O experimento possibilitou conhecer a relação entre as j CQ com as k VC e construir MRB que descrevem tal relação. A Tabela 13 apresenta um resumo estatístico dos dados experimentais.

Tabela 13: Resumo estatístico das CQ do estudo experimental

Variáveis	Mínimo	Máximo	Média	Variância	Coefficiente Variação
y_1	0,100	0,500	0,258	0,019	53%
y_2	0,100	0,800	0,516	0,025	30%
y_3	0,011	0,083	0,039	0,0005	57%

Fonte: Tabela elaborada pelo autor (2009)

As CQ y_1 e y_2 são frações do tipo percentual e apresentam np igual a 7,7 e 15,5 respectivamente, demonstrando a priori que y_1 e y_2 poderiam ser modeladas pela distribuição Normal de forma satisfatória.

A expressão (66) apresenta os modelos MRB estimados para as CQ y_1 , y_2 e y_3 contendo as VC que apresentaram significância estatística para explicar as CQ y_1 , y_2 e y_3 , ao nível de significância de 5%, baseado no teste estatístico de *Wald*. Os MRB y_1 , y_2 e y_3 estimados apresentam distribuição Beta, função de ligação logit e método de estimação por máxima verossimilhança. Os coeficientes de determinação R^2 e critério de informação de Akaike (AIC): $R^2 = 0,824$ e $AIC = -36,161$, $R^2 = 0,932$ e $AIC = -33,475$ e $R^2 = 0,735$ e $AIC = -93,493$ respectivamente, demonstram que os três modelos se ajustaram satisfatoriamente, conforme as Tabelas 17, 18 e 19.

$$\begin{aligned}
 y_1 &= BETA[\exp(-1,165 - 0,174x_2 - 0,392x_1x_2 + 0,474x_1x_4 + 0,358x_2x_3 \\
 &\quad + 0,230x_2x_5)] \\
 y_2 &= BETA[\exp(0,068 - 0,136x_3 - 0,271x_1x_2 + 0,379x_1x_4 + 0,274x_2x_4 \\
 &\quad + 0,419x_2x_5 + 0,500x_3x_4)] \\
 y_3 &= BETA[\exp(-3,310 + 0,158x_1 + 0,194x_2 - 0,392x_1x_2 + 0,210x_1x_4 \\
 &\quad + 0,222x_2x_5)]
 \end{aligned} \tag{66}$$

A partir dos modelos de regressão Beta estimados para as CQ y_1 , y_2 e y_3 obtêm-se os resíduos *deviance* (r_d) de cada modelo, a partir das equações (64), e as respectivas estimativas de média e desvio-padrão amostral dos resíduos.

As estimativas de médias e variâncias dos resíduos *deviance* apresentadas na Tabela 14 são utilizadas para calcular os limites de controle e linha central da CCMB proposta. Ademais, o valor da constante w adotada para este estudo foi 3, admitindo a probabilidade $\alpha = 0,0027$ de alarmes falsos no monitoramento do processo. A escolha do valor da constante w foi para permitir o estudo de comparação gráfica entre as cartas de controle propostas por Hawkins (1991) e Haworth (1996).

4.1. Análise comparativa entre as Cartas de Controle

A CCMB proposta para as três CQ é comparada com as cartas de controle baseadas em modelos propostas por Hawkins (1991) e Haworth (1996). Foram construídos modelos de regressão para cada CQ y_1 , y_2 e y_3 , observando-se a qualidade do ajuste dos modelos aos dados. Em seguida foram gerados os resíduos padronizados para a carta de controle proposta por Hawkins (1991) e os resíduos *student* para as cartas de controle proposta por Haworth (1996) utilizando o modelo Normal.

A expressão (67) apresenta os modelos de regressão Normal estimados contendo as VC que apresentaram significância estatística para explicar as CQ y_1 , y_2 e y_3 , ao nível de significância de 5%, baseado no teste *t-Student*. Os modelos Normais y_1 , y_2 e y_3 estimados apresentam distribuição Normal, função de ligação identidade e método de estimação por mínimos quadrados ordinários (MQO).

Os coeficientes de determinação R^2 e critério de informação de Akaike (AIC): $R^2 = 0,556$ e $AIC = -25,66$, $R^2 = 0,753$ e $AIC = -25,99$ e $R^2 = 0,453$ e $AIC = -48,23$ respectivamente, demonstram que os três modelos apresentaram um ajuste moderado, conforme as Tabelas 17, 18 e 19.

$$\begin{aligned}
 y_1 &= NORMAL[0,261 - 0,070x_1x_2 + 0,081x_1x_4 + 0,060x_2x_3] \\
 y_2 &= NORMAL[0,519 - 0,057x_1x_2 + 0,081x_1x_4 + 0,059x_2x_4 + 0,091x_2x_5 \\
 &\quad - 0,111x_3x_4] \\
 y_3 &= NORMAL[0,039 - 0,013x_1x_2 + 0,008x_1x_4 + 0,008x_2x_5]
 \end{aligned} \tag{67}$$

A expressão (68) apresenta os modelos de regressão Logit Normal estimados contendo as VC que apresentaram significância estatística para explicar as CQ y_1 , y_2 e y_3 , ao

nível de significância de 5%, baseado no teste *t-Student*. Os modelos de regressão Logit Normal y_1 , y_2 e y_3 estimados apresentam distribuição Normal, função de ligação logit e método de estimação por mínimos quadrados ordinários (MQO).

Segundo Kieschnick e McCullough (2003), este modelo de regressão utiliza a função de ligação logit para relacionar as variáveis e estima seus parâmetros por mínimos quadrados ordinários. Conforme Cox (1996), o método de mínimos quadrados é uma abordagem simples para estimar parâmetros de modelos de regressão e pode ser utilizado em modelagem de dados contínuos com variâncias conhecidas.

Os coeficientes de determinação e critério de informação de Akaike (AIC): $R^2 = 0,544$ e $AIC = 24,24$, $R^2 = 0,811$ e $AIC = 24,40$ e $R^2 = 0,367$ e $AIC = -34,68$ respectivamente, demonstrando que os modelos apresentaram um ajuste moderado, conforme as Tabelas 17, 18 e 19.

$$\begin{aligned} y_1 &= NORMAL[\text{logit}(-1,173 - 0,357x_1x_2 + 0,467x_1x_4 + 0,336x_2x_3)] \\ y_2 &= NORMAL[\text{logit}(0,070 - 0,289x_1x_2 + 0,401x_1x_4 + 0,276x_2x_4 + 0,426x_2x_5 \\ &\quad - 0,516x_3x_4)] \\ y_3 &= NORMAL[\text{logit}(-3,360 - 0,378x_1x_2 + 0,229x_2x_5)] \end{aligned} \quad (68)$$

A expressão (69) apresenta os modelos de regressão Logística estimados contendo as VC que apresentaram significância estatística para explicar as CQ y_1 , y_2 e y_3 , ao nível de significância de 5%, baseado no teste de *Wald*. Os modelos de regressão Logística y_1 , y_2 e y_3 estimados apresentam distribuição Binomial, função de ligação logit e método de estimação por máxima verossimilhança. Os coeficientes de determinação e critério de informação de Akaike (AIC): $R^2 = 0,710$ e $AIC = -30,59$, $R^2 = 0,777$ e $AIC = -28,03$ e $R^2 = 0,581$ e $AIC = -55,51$ respectivamente, demonstrando que os modelos apresentaram um ajuste moderado, conforme as Tabelas 18, 19 e 20.

$$\begin{aligned} y_1 &= [Logística(-1,163 - 0,396x_1x_2 + 0,479x_1x_4 + 0,361x_2x_3 + 0,231x_2x_5)] \\ y_2 &= [Logística(0,068 - 0,272x_1x_2 + 0,380x_1x_4 + 0,280x_2x_4 + 0,420x_2x_5 \\ &\quad - 0,499x_3x_4)] \\ y_3 &= [Logística(-3,310 - 0,381x_1x_2 + 0,228x_1x_4 + 0,237x_2x_5)] \end{aligned} \quad (69)$$

As Tabelas 17, 18 e 19 apresentam os resultados do ajuste dos modelos de regressão Beta, Normal, Logit Normal e Logística aplicados na modelagem das frações não conforme y_1 , y_2 e y_3 (ver Apêndice). Esta comparação permite avaliar a melhor adequação dos modelos aos dados, as estimativas e as inferências de cada modelo. Observa-se que os coeficientes das

VC dos modelos de regressão Logística e Beta são similares, diferindo nos valores da estimativa dos erros padrões e nos critérios de adequação.

A partir dos modelos de regressão Normal estimados para as CQ y_1 , y_2 e y_3 obtêm-se os resíduos padronizado (r_p) e *student* (r_s) de cada modelo, a partir das equações (70) e (71), e as respectivas estimativas de média e desvio-padrão amostral dos resíduos.

$$r_{p_i} = \frac{(y_i - \hat{\mu})}{\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{\mu})^2}{n - t}}} \quad (70)$$

$$r_{s_i} = \frac{(y_i - \hat{\mu})}{\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{\mu})^2}{n - t} \cdot (1 - [X_t'(X'X)^{-1}X_t])}} \quad (71)$$

em que X_t' um vetor de VC da $t^{\text{ésima}}$ observação e X a matriz de vetores de VC.

As estimativas de médias e variâncias dos resíduos padronizados e *student* apresentadas na Tabela 14 são utilizadas para calcular os limites de controle e linha central das cartas propostas por Hawkins (1991) e Haworth (1996), conforme as expressões (72) e (73),

$$\begin{aligned} LSC_j &= \bar{r}_{p_j} + 3\sqrt{s^2(\bar{r}_{p_j})} \\ LC_j &= \bar{r}_{p_j} \\ LIC_j &= \bar{r}_{p_j} - 3\sqrt{s^2(\bar{r}_{p_j})} \end{aligned} \quad (72)$$

$$\begin{aligned} LSC_j &= \bar{r}_{s_j} + t_{(\frac{\alpha}{2}; n-t)}\sqrt{s^2(\bar{r}_{s_j})} \\ LC_j &= \bar{r}_{s_j} \\ LIC_j &= \bar{r}_{s_j} - t_{(\frac{\alpha}{2}; n-t)}\sqrt{s^2(\bar{r}_{s_j})} \end{aligned} \quad (73)$$

onde $t_{(\frac{\alpha}{2}; n-t)}$ é uma constante que define a largura dos limites de controle correspondente ao nível desejado do número médio de amostras até alarme falso (NMA₀).

Após estabelecer os limites de controle das cartas propostas por Hawkins (1991), Haworth (1996), CCMB e limites para os modelos Logit Normal e Logística, coletou-se 60 novas amostras de tamanho 50 unidades das CQ y_1 , y_2 e y_3 , e das VC provenientes do processo de manufatura de pneus radial. Os resíduos padronizados, *student* e *deviance*, são calculados e monitorados a partir da nova amostra.

Tabela 14: Estimativas de média e variância dos resíduos padronizados, *student* e *deviance* dos modelos

Modelos de regressão		y_1		y_2		y_3	
		Média	Variância	Média	Variância	Média	Variância
Normal	Resíduos padronizados	0,0482	0,985	0,0893	1,016	0,0572	0,943
Normal	Resíduos <i>student</i>	0,0395	1,131	0,1332	1,425	0,0668	1,109
Logit Normal	Resíduos padronizados	0,0483	0,983	0,0841	1,024	0,0680	0,935
Logística	Resíduos <i>deviance</i>	0,0381	1,056	0,0625	1,083	0,0312	1,074
Beta	Resíduos <i>deviance</i>	0,0487	1,053	0,0197	1,048	0,0471	1,040

Fonte: Tabela elaborada pelo autor (2009)

A Figura 20 apresenta as cartas de controle propostas por Hawkins (1991), Haworth (1996) e CCMB, além dos limites de controle para os modelos Logit Normal e Logística para as CQ y_1 , y_2 e y_3 . Verifica-se que as cartas de controle apresentam similaridade no monitoramento da variabilidade de cada CQ. As cartas de controle propostas por Hawkins (1991) apresentam 1 ponto (amostra 38) fora dos limites de controle para a variável y_1 e 1 ponto (amostra 03) fora dos limites de controle para a variável y_3 , caracterizando um processo fora de controle estatístico, para as duas CQ (Figura 20(a), (b) e (c)).

Observa-se nas Figura 20(d), (e) e (f) que as cartas de controle propostas por Haworth (1996) apresentam variabilidade similar as cartas de Hawkins (1991) e indicam os mesmos pontos como fora dos limites de controle, caracterizando um processo fora de controle estatístico, para as CQ y_1 e y_3 . As Figura 20(g), (h) e (i) apresentam o monitoramento das CQ utilizando os limites de controle calculados a partir dos resíduos do modelo Logit Normal e nota-se 1 ponto (amostra 38) fora dos limites de controle para a variável y_1 e 1 ponto (amostra 25) fora dos limites de controle para a variável y_3 , sugerindo um processo fora de controle estatístico.

Os resíduos *deviance* da CCMB e da carta de controle utilizando o modelo de regressão Logística apresentam um comportamento similar e não há pontos fora dos limites de controle, demonstrando um ajuste mais adequado dos modelos de regressão Beta e Logística. Observa-se que, segundo as cartas de controle com resíduos *deviance*, o processo pode ser considerado sob controle estatístico, não havendo causas especiais para as três CQ investigadas (Figuras 20(j) à (o)).

Na análise de investigação das CQ do processo, detectou-se que não há causas especiais, evidenciando que as cartas propostas por Hawkins (1991), Haworth (1996) e Logit Normal identificaram alarmes falsos. Assim, acredita-se que a CCMB e a carta de controle utilizando o modelo de regressão Logística são mais adequados para modelar monitorar a fração não conforme.

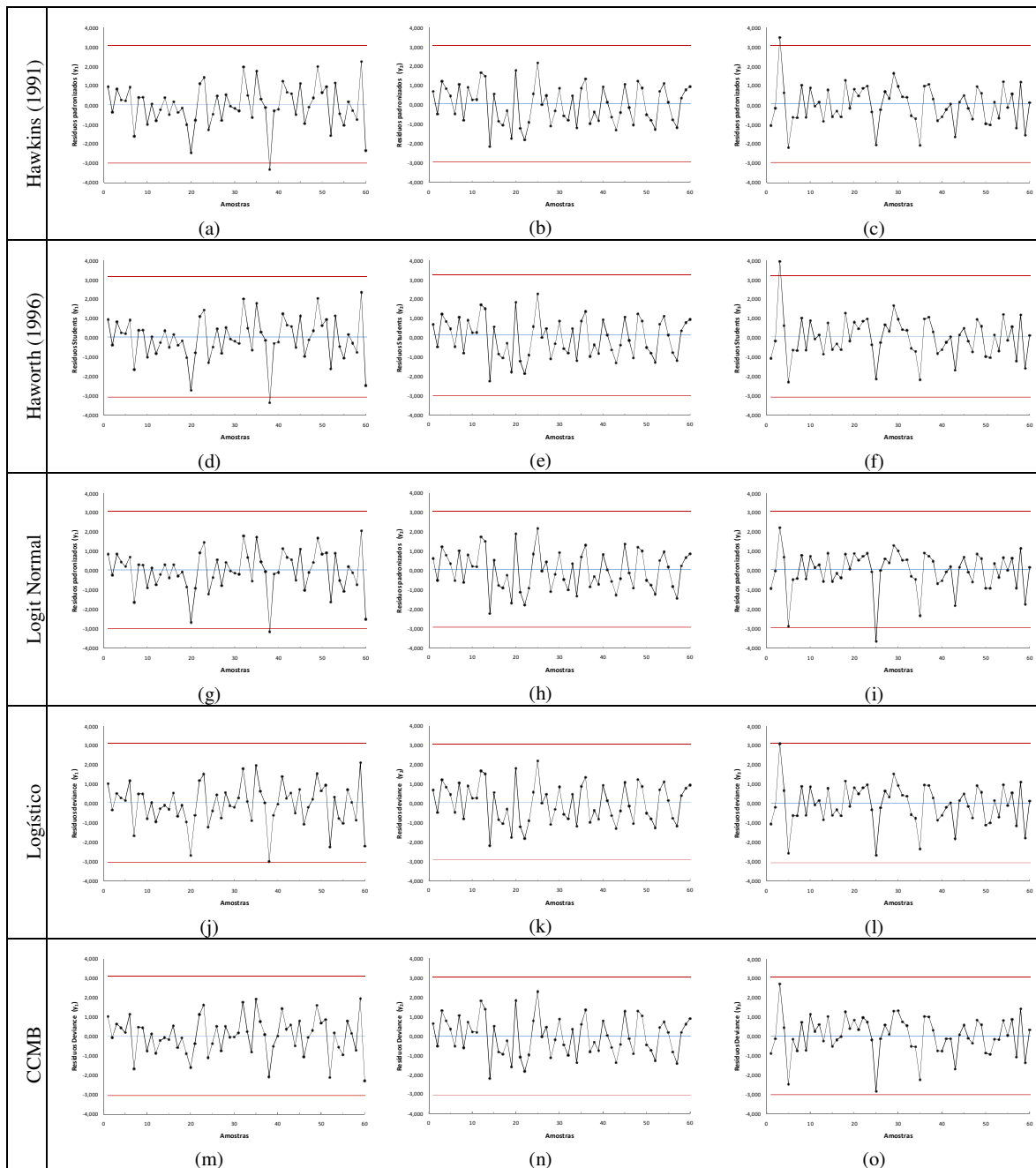


Figura 20: Cartas de controle proposta por Hawkins (1991), Haworth (1996), Logit Normal, Logística e CCMB das CQ y_1 , y_2 , y_3

Fonte: Elaborado pelo autor (2009)

5. Análise de Sensibilidade das Cartas de Controle

Foi realizado um estudo de simulação para analisar a sensibilidade das cartas de controle propostas por Hawkins (1991), Haworth (1996) e CCMB, considerando um processo sob controle e outro fora de controle. Para o processo sob controle foi calculado o número médio de amostras até a presença de uma falsa causa especial (NMA_0). O maior valor de NMA_0 indica uma menor probabilidade de alarmes falsos no processo. O NMA_0 pode ser escrito em função da probabilidade de erro tipo I (α), que é a probabilidade da carta de controle detectar uma alteração no modelo de referência monitorado quando o processo está sob controle estatístico (MONTGOMERY, 2005).

Para o processo fora de controle foram induzidas dois tipos de alterações:

(i) uma alteração aditiva na média das CQ $\mu_j = \beta_{j0} + \varphi$;

$$\begin{aligned}\mu_j &= \beta_{j0} + \beta_{j1}x_{j1} + \dots + \beta_{jk}x_{jk} \\ \mu_j &= (\beta_{j0} + \varphi) + \beta_{j1}x_{j1} + \dots + \beta_{jk}x_{jk}\end{aligned}\quad (74)$$

A alteração aditiva considera dois tamanhos: um sigma (1σ) e três sigmas (3σ). Por exemplo, uma alteração aditiva com tamanho 1σ indica que a média de y_j é modificada de μ_{y_j} para $\mu_{y_j} + 1\sigma_{y_j}$ onde σ_{y_j} é o desvio-padrão amostral da $j^{\text{ésima}}$ CQ.

(ii) uma alteração multiplicativa em algum dos coeficientes dos modelos de regressão,

$$\mu_j = (\varphi) \cdot \beta_{jk}x_{jk};$$

$$\begin{aligned}\mu_j &= \beta_{j0} + \beta_{j1}x_{j1} + \dots + \beta_{jk}x_{jk} \\ \mu_j &= (\beta_{j0} \cdot \varphi) + \beta_{j1}x_{j1} + \dots + \beta_{jk}x_{jk}\end{aligned}\quad (75)$$

e

$$\mu_j = \beta_{j0} + (\beta_{j1} \cdot \varphi)x_{j1} + \dots + \beta_{jk}x_{jk}\quad (76)$$

A alteração multiplicativa considerou-se dois tamanhos de constante 1,5 e 3. Por exemplo, uma alteração multiplicativa com tamanho 1,5 indica que o coeficiente do modelo modifica de μ_{y_j} para $\mu_{y_j} \cdot 1,5$. Estas alterações representam formas de alterações no processo produtivo, tanto na CQ quanto na VC e objetivam avaliar o poder de detecção das cartas de controle no monitoramento.

Para o processo fora de controle foi calculado o número médio de amostras até a detecção de causas especiais (NMA_1). O menor valor de NMA_1 representa um menor número médio de amostras coletadas até a detecção da alteração induzida. O NMA_1 pode ser escrito em função da probabilidade de erro tipo II (γ), que é a probabilidade da carta de controle não

detectar uma alteração no modelo de referência monitorado quando o processo está fora de controle estatístico (MONTGOMERY, 2005).

O método utilizado para calcular os valores de NMA_0 e NMA_1 foi simulação de Monte Carlo usando o *software* livre R[®] v.2.9.0. A simulação consistiu na geração de 10.000 observações das VC com distribuição Uniforme e das CQ segundo as estimativas amostrais identificadas na fase de análise descritiva do experimento industrial. Estes dados são utilizados para estimar os valores preditos a partir das equações (66) e (67) e respectivos resíduos deviance (r_d), padronizados (r_p) e *student* (r_s) a partir das equações (64) e (70). Estes resíduos estimados são *plotados* nas respectivas cartas de controle.

A simulação do processo sob controle foi realizada 5.000 vezes, sendo computada a média e o erro-padrão dos percentuais de pontos fora dos limites de controle estatístico ($p_f(\%)$). O número médio de amostras até o falso alarme (NMA_0) para as j cartas CCMB, pode ser calculado como:

$$NMA_0 = \frac{1}{[1 - (1 - \alpha)^j]} \quad (77)$$

em que α é a probabilidade de alarmes falsos definido por $\bar{p}_f(\%)$ (SKINNER, 2003).

A simulação de um processo fora de controle foi realizada 5000 vezes, induzindo os dois tipos e tamanhos de alterações propostas, sendo calculada a média e erro-padrão do percentual de pontos entre os limites de controle até a detecção da causa especial ($p_s(\%)$). O número médio de amostras até a detecção da causa especial (NMA_1) para as j cartas CCMB, pode ser expresso por:

$$NMA_1 = \frac{1}{[1 - \beta^j]} \quad (78)$$

em que β é a probabilidade de alarmes verdadeiros definido por $\bar{p}_f(\%)$. Esta fórmula foi gerada a partir da extensão da fórmula do NMA_0 proposta por Skinner (2003).

Para realizar a análise de sensibilidade das cartas de controle, inicialmente foi utilizada a simulação de Monte Carlo para obter o valor da constante w que equaliza o NMA_0 entre as cartas de controle investigadas. Sullivan e Woodall (1996) utilizam esta simulação para obter os valores de w apropriados para comparar diferentes tipos de cartas de controle multivariadas.

A simulação foi desenvolvida iniciando com um valor arbitrário para w ($w = 3$) e alterando até obter um NMA_0 desejado. Os valores da constante w foram ajustados para $NMA_0 = 333,7$ ($\alpha = 0,001$ e $j = 3$) para cada CQ. A Tabela 15 apresenta os valores da constante w de cada CQ para o cálculo dos limites de controle das cartas investigadas.

Tabela 15: Valores da constante w (e erros-padrão) das cartas de controle CCMB, Hawkins (1991), Haworth (1996) baseado em 5.000 simulações.

Característica Qualidade	Tipo	Carta Hawkins (1991)	Carta Haworth (1996)	Carta CCMB
y_1	<i>Percentual</i>	3,78 (0,013)	3,78 (0,013)	3,15 (0,011)
y_2	<i>Percentual</i>	3,23 (0,015)	3,23 (0,016)	2,96 (0,014)
y_3	<i>Proporção</i>	4,50 (0,016)	4,50 (0,016)	3,24 (0,012)

Fonte: Tabela elaborada pelo autor (2009)

A Tabela 16 apresenta os valores dos NMA_0 , NMA_1 e respectivos erros-padrão das cartas de controle CCMB, Hawkins (1991) e Haworth (1996) segundo o tipo e tamanho da alteração. Nota-se que a CCMB apresentou o maior valor de NMA_0 (337, 335 e 384) para um processo sob controle. Enquanto que as outras cartas comparadas apresentaram maiores valores de NMA_0 .

Analisando os valores de NMA_1 para as situações de alteração “ 1σ + aditiva” para as CQ y_1 , y_2 e y_3 , a CCMB apresenta o menor valor de NMA_1 (1,77, 1,23 e 18,21) quando comparado com as outras cartas de controle estudadas. Assim como, para a situação de alteração de “ 3σ + aditiva” em y_1 , y_2 e y_3 , os menores valores de NMA_1 (1,34, 1,17 e 7,03) foram obtidos pela CCMB. Para alterações multiplicativas a CCMB também apresenta menores valores de NMA_1 dentre as cartas de controle investigadas.

A CCMB apresenta desempenho satisfatório, pois gerou o menor valor de NMA_1 em todas as alterações induzidas de descontrole, quando comparado com as outras cartas estudadas. Nota-se que a carta CCMB detecta mais rapidamente todos os tamanhos e tipos de alterações induzidas.

Tabela 16: Resultado dos NMA (e erros-padrão do NMA) das cartas de controle Hawkins (1991), Haworth (1996) e CCMB baseado em 5.000 simulações.

Tamanho Mudança	Tipo Mudança	Variável afetada	Carta Hawkins (1991)	Carta Haworth (1996)	Carta CCMB
–	–	–	330,12 (3,232)	332,72 (3,870)	337,48 (3,304)
–	–	–	322,17 (3,227)	328,90 (3,610)	335,67 (3,286)
–	–	–	221,44 (3,241)	224,67 (3,200)	384,67 (3,105)
1 σ	<i>Aditiva</i>	y_1	51,89 (0,732)	59,94 (0,841)	1,77 (0,017)
1,5	<i>Multiplicativa (β_0)</i>	y_1	20,32 (0,182)	14,12 (0,370)	6,10 (0,100)
1,5	<i>Multiplicativa (β_1)</i>	y_1	30,17 (0,271)	18,25 (0,310)	11,51 (0,208)
1 σ	<i>Aditiva</i>	y_2	30,24 (0,470)	26,78 (0,531)	1,23 (0,013)
1,5	<i>Multiplicativa (β_0)</i>	y_2	22,45 (0,321)	20,23 (0,187)	3,92 (0,074)
1,5	<i>Multiplicativa (β_1)</i>	y_2	11,59 (0,307)	13,63 (0,218)	6,09 (0,192)
1 σ	<i>Aditiva</i>	y_3	126,63 (3,208)	127,85 (3,330)	18,21 (0,173)
1,5	<i>Multiplicativa (β_0)</i>	y_3	128,38 (3,102)	125,32 (3,103)	12,12 (0,210)
1,5	<i>Multiplicativa (β_1)</i>	y_3	122,21 (3,200)	125,74 (3,187)	9,17 (0,207)
3 σ	<i>Aditiva</i>	y_1	7,00 (0,089)	7,69 (0,100)	1,85 (0,034)
3	<i>Multiplicativa (β_0)</i>	y_1	8,24 (0,100)	6,15 (0,063)	1,65 (0,014)
3	<i>Multiplicativa (β_1)</i>	y_1	9,02 (0,108)	7,23 (0,102)	4,14 (0,052)
3 σ	<i>Aditiva</i>	y_2	17,41 (0,198)	15,34 (0,175)	1,17 (0,006)
3	<i>Multiplicativa (β_0)</i>	y_2	13,15 (0,074)	12,19 (0,058)	1,94 (0,021)
3	<i>Multiplicativa (β_1)</i>	y_2	10,48 (0,086)	8,05 (0,084)	4,03 (0,127)
3 σ	<i>Aditiva</i>	y_3	123,75 (3,248)	120,66 (3,065)	7,03 (0,090)
3	<i>Multiplicativa (β_0)</i>	y_3	130,13 (3,261)	126,12 (3,100)	6,10 (0,074)
3	<i>Multiplicativa (β_1)</i>	y_3	129,63 (3,150)	125,82 (3,022)	5,32 (0,063)

Fonte: Tabela elaborada pelo autor (2009)

6. Conclusão

O objetivo deste artigo foi propor uma carta de controle baseada em modelos Beta (CCMB) para monitorar processos que mensuram características de qualidade do tipo fração ou razão com valores restritos ao intervalo [0,1] e que dependem de freqüentes alterações de ajustes nas variáveis de controle. Inicialmente foi apresentado um referencial teórico sobre distribuição Beta e modelos de regressão Beta. Na seqüência foi proposta a CCMB e os passos para a sua implantação.

A CCMB foi aplicada no monitoramento das características de qualidade: percentual da variação força radial não conforme, percentual da espessura radial não conforme e proporção de massa não convertida em matéria-prima, mensuradas ao final do processo de manufatura de pneus radial de uma empresa multinacional de borrachas.

A fim de realizar uma análise comparativa, aplicaram-se também as cartas propostas por Hawkins (1991), Haworth (1996), CCMB e limites de controle a partir dos modelos Logit

Normal e Logística nas características de qualidade monitoradas pela CCMB. Assim, acredita-se que a CCMB e a carta de controle utilizando o modelo de regressão Logística são mais adequados para modelar monitorar características de qualidade do tipo fração.

Posteriormente, foi realizado um estudo de simulação de Monte Carlo para analisar a sensibilidade da carta CCMB e das cartas comparadas. As simulações realizadas consistiram em um processo sob controle e outro fora de controle, induzindo dois tipos e tamanhos de alterações. Para cada processo, calculou-se respectivamente o NMA_0 e NMA_1 (e erros-padrão) para cada uma das cartas de controle analisadas.

Os resultados da análise de sensibilidade indicaram que a CCMB apresenta desempenho satisfatório, pois apresenta maior valor de NMA_0 para a probabilidade de alarme falso conforme o valor nominal desejado e menor número de amostras até a detecção do alarme verdadeiro (NMA_1) dado as alterações induzidas, quando comparado com as outras cartas de controle. A CCMB detecta mais rapidamente todos os tamanhos e tipos de alterações induzidas, apresentando melhor desempenho entre as cartas analisadas.

Destacam-se algumas vantagens do monitoramento utilizando a carta CCMB: (i) as frações podem ser melhor aproximadas pela distribuição Beta; (ii) o modelo de regressão Beta utilizado permitiu modelar a relação não-linear entre as característica de qualidade e variáveis de controle; (iii) o método de estimação por máxima verossimilhança permitiu obter resíduos *deviance* mais precisos.

Conclui-se que CCMB pode ser uma ferramenta adequada para o monitoramento de características de qualidade do tipo fração, que apresentam valores restritos ao intervalo [0,1], que não seguem a distribuição Normal e que dependam de freqüentes ajustes nas variáveis de controle do processo.

Agradecimentos

Esta pesquisa foi apoiada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/Brasil).

Referências

- (1) Ferrari, S.L.P & Cribari-Neto, F. (2004) Beta regression for modeling rates and proportions, *Journal of Applied Statistics*, **31**, 799–816.
- (2) Hauck, D.J.; Runger, G.C. & Montgomery, D.C. (1999) Multivariate statistical process monitoring and diagnosis with grouped regression-adjusted variables, *Communication in Statistics: Simulation and Computation*, **28**, 309–328.

- (3) Hawkins, D.A. (1991) Multivariate quality control based on regression-adjusted variables, *Technometrics*, **33**, 61–75.
- (4) Hawkins, D.M. (1993) Regression adjustment for variables in multivariate quality control, *Journal of Quality Technology*, **25**, 170–182.
- (5) Haworth, D.A. (1996) Regression control charts to manage software maintenance, *Software Maintenance Research and Practice*, **8**, 35–48.
- (6) Jearkpaporn, D.; Montgomery, D.C.; Runger, G.C. & Borrór, C.M. (2005) Model-based process using robust generalized linear models, *International Journal of Production Research*, **43**, 1337–1354.
- (7) Johnson, N.L.; Kemp, A.W.; Kotz, S. (2005) *Univariate Discrete Distributions*. 3^a ed., New York: John Wiley & Sons, 646p.
- (8) Johnson, N.L.; Kotz, S. & Balakrishnan, N. (1995) *Continuous Univariate Distributions*, vol.2, 2^a ed., John Wiley & Sons, New York.
- (9) Kieschnick, R. & McCullough, B.D. (2003) Regression analysis of variates observed on [0,1]: percentages, proportions and fractions, *Statistical Modelling*, **3**, 193–213.
- (10) Loredó, E.N.; Jearkpaporn, D. & Borrór, C.M. (2002) Model-based control chart for autocorrelated data, *Quality and Reliability Engineering International*, **18**, 489–496.
- (11) Mandel, B.J. (1969) The regression control charts, *Journal of Quality Technology*, **1**, 1–9.
- (12) Moitra, S.D. (1990) Skewness and the Beta distribution, *Journal of the Operational Research Society*, **41**, 953–961.
- (13) Montgomery, D.C. (2005) *Introduction to Statistical Quality Control*, 5^a ed., John Wiley & Sons, New York.
- (14) Pierce, D.A. & Schafer, D.W. (1986) Residuals in generalized linear model, *Journal of the American Statistical Association*, **81**, 977–986.
- (15) R Development Core Team. *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, ISBN 3-900051-07-0, 2009.
- (16) Rocke, D.M. (1993) On the Beta transformation family, *Technometrics*, **35**, 72–81.
- (17) Sant’Anna, A.M.O. & Caten, C.S. (2009) Modelagem da fração de não conforme em processos industriais, *Pesquisa Operacional*, no prelo.
- (18) Schader, M.; Schmid, F. Two rules of thumb for the approximation of the Binomial distribution by Normal distribution, *The American Statistician*, Alexandria, v. 43, n. 1, p. 23-24, 1989.
- (19) Shu, L.; Tsui, K.; Tsung, F. (2004) Run-length performance of regression control charts with estimated parameters, *Journal of Quality Technology*, **36**, 280–292.
- (20) Shu, L.; Tsui, K.; Tsung, F. (2008) A review of regression control charts, *Encyclopedia of Statistics in Quality and Reliability*, v. **260**, 1-9, 2008.
- (21) Skinner, K.R.; Montgomery, D.C. & Runger, G.C. (2003) Process monitoring for multiple count data using generalized linear model-based control charts, *International Journal of Production Research*, **41**, 1167–1180.

- (22) Skinner, K.R.; Montgomery, D.C. & Runger, G.C. (2004) Generalized linear model-based control charts for discrete Semiconductor process data, *Quality and Reliability Engineering International*, **20**, 777–786.
- (23) Sullivan, J.H. & Woodall, W.H. (1996) A comparison of multivariate control charts for individuals observations, *Journal of Quality Technology*, **28**, 398–408.
- (24) Wade, W. & Woodall, W.H. (1993) A review and analysis of cause-selecting control charts, *Journal of Quality Technology*, **25**, 161–169.
- (25) Woodall, W.H. (2000) Controversies and Contradictions in Statistical Process Control, *Journal of Quality Technology*, **32**, 341–378.
- (26) Zhang, G.X. (1985) Cause-selecting control charts – A new type of quality control charts, *The QR Journal*, **12**, 221–225.
- (27) Cox, C. (1996) Nonlinear quasi-likelihood models: applications to continuous proportions, *Computational Statistical & Data Analysis*, **21**, 449–461.

Apêndice

Tabela 17: Estimativas dos coeficientes com respectivos erros padrões e critérios de diagnóstico dos modelos ajustados para a CQ y_i

	Modelo Normal	Modelo Logit Normal	Modelo Logístico	Modelo Beta
Distribuição	<i>Normal</i>	<i>Normal</i>	<i>Binomial</i>	<i>Beta</i>
Função de Ligação	<i>Identidade</i>	<i>Logit</i>	<i>Logit</i>	<i>Logit</i>
Método de Estimação^Δ	<i>MQO</i>	<i>MQO</i>	<i>EMV</i>	<i>EMV</i>
Parâmetro	Estimativa	Estimativa	Estimativa	Estimativa
Intercepto	0,261** (0,0227)	-1,173** (0,1137)	-1,163** (0,1030)	-1,165** (0,0743)
X1	0,006 (0,0241)	-0,020 (0,1206)	-0,015 (0,1069)	-0,012 (0,0767)
X2	-0,034 (0,0241)	-0,153 (0,1206)	-0,178 (0,1068)	-0,174* (0,0767)
X3	0,001 (0,0241)	-0,030 (0,1206)	-0,017 (0,1075)	-0,015 (0,0771)
X4	0,011 (0,0241)	0,090 (0,1206)	0,072 (0,1050)	0,069 (0,0755)
X5	-0,009 (0,0241)	-0,093 (0,1206)	-0,111 (0,1082)	-0,109 (0,0776)
X1*X2	-0,070* (0,0241)	-0,357* (0,1206)	-0,396** (0,1043)	-0,392** (0,0751)
X1*X4	0,081** (0,0241)	0,467** (0,1206)	0,479** (0,1082)	0,474** (0,0776)
X2*X3	0,060* (0,0241)	0,336* (0,1206)	0,361** (0,1076)	0,358** (0,0772)
X2*X5	0,042 (0,0241)	0,235 (0,1206)	0,231* (0,1061)	0,230* (0,0763)
R_p^2	0,556	0,544	0,710	0,824
<i>AIC</i>	-25,66	24,24	-30,59	-34,16

^Δ MQO, mínimo quadrados ordinários; EMV, máxima verossimilhança.

Erro padrão das estimativas são apresentados entre parênteses.

Nível de significância “Teste Wald”: (p < 0,01)** e (p < 0,05)*

Tabela 18: Estimativas dos coeficientes com respectivos erros padrões e critérios de diagnóstico dos modelos ajustados para a CQ y_2

	Modelo Normal	Modelo Logit Normal	Modelo Logístico	Modelo Beta
Distribuição	<i>Normal</i>	<i>Normal</i>	<i>Binomial</i>	<i>Beta</i>
Função de Ligação	<i>Identidade</i>	<i>Logit</i>	<i>Logit</i>	<i>Logit</i>
Método de Estimação^Δ	<i>MQO</i>	<i>MQO</i>	<i>EMV</i>	<i>EMV</i>
Parâmetro	Estimativa	Estimativa	Estimativa	Estimativa
Intercepto	0,519** (0,0197)	0,070** (0,0944)	0,068** (0,0954)	0,068* (0,0548)
X1	0,006 (0,0209)	0,041 (0,1002)	0,031 (0,1020)	0,033 (0,0586)
X2	-0,021 (0,0209)	-0,068 (0,1002)	-0,076 (0,1020)	-0,076 (0,0585)
X3	-0,025 (0,0209)	-0,138 (0,1002)	-0,136 (0,1026)	-0,136* (0,0589)
X4	0,009 (0,0209)	0,001 (0,1002)	0,011 (0,1028)	0,010 (0,0590)
X5	0,027 (0,0209)	0,095 (0,1002)	0,097 (0,1028)	0,102 (0,0588)
X1*X2	-0,057* (0,0209)	-0,289* (0,1002)	-0,272* (0,1025)	-0,271** (0,0588)
X1*X4	0,081** (0,0209)	0,401** (0,1002)	0,380* (0,1030)	0,379** (0,0591)
X1*X5	0,038 (0,0209)	0,205 (0,1002)	0,194 (0,1029)	0,188** (0,0590)
X2*X4	0,059* (0,0209)	0,276* (0,1002)	0,280* (0,1027)	0,274** (0,0589)
X2*X5	0,091** (0,0209)	0,426** (0,1002)	0,420** (0,1031)	0,419** (0,0591)
X3*X4	-0,111** (0,0209)	-0,516** (0,1002)	-0,499** (0,1020)	-0,500** (0,0588)
R_p^2	0,753	0,811	0,777	0,932
<i>AIC</i>	-25,99	24,40	-28,03	-31,47

^Δ MQO, mínimo quadrados ordinários; EMV, máxima verossimilhança.

Erro padrão das estimativas são apresentados entre parênteses.

Nível de significância "Teste Wald": (p < 0,01)** e (p < 0,05)*

Fonte: Elaborado pelo autor (2009)

Tabela 19: Estimativas dos coeficientes com respectivos erros padrões e critérios de diagnóstico dos modelos ajustados para a CQ y_3

	Modelo Normal	Modelo Logit Normal	Modelo Logístico	Modelo Beta
Distribuição	<i>Normal</i>	<i>Normal</i>	<i>Binomial</i>	<i>Beta</i>
Função de Ligação	<i>Identidade</i>	<i>Logit</i>	<i>Logit</i>	<i>Logit</i>
Método de Estimação^Δ	<i>MQO</i>	<i>MQO</i>	<i>EMV</i>	<i>EMV</i>
Parâmetro	Estimativa	Estimativa	Estimativa	Estimativa
Intercepto	0,039** (0,0042)	-3,360** (0,1279)	-3,310** (0,1140)	-3,312** (0,0871)
X1	0,003 (0,0044)	0,130 (0,1357)	0,137 (0,1188)	0,155* (0,0885)
X2	0,005 (0,0044)	0,171 (0,1357)	0,193 (0,1068)	0,193** (0,0885)
X3	0,0001 (0,0044)	-0,041 (0,1357)	-0,044 (0,1121)	-0,030 (0,0835)
X4	0,004 (0,0044)	0,139 (0,1357)	0,105 (0,1098)	0,095 (0,0823)
X5	0,0006 (0,0044)	0,012 (0,1357)	-0,014 (0,1103)	0,002 (0,0825)
X1*X2	-0,013** (0,0044)	-0,378* (0,1357)	-0,381** (0,1159)	-0,393** (0,0869)
X1*X4	0,008* (0,0044)	0,144 (0,1357)	0,228* (0,1125)	0,217** (0,0839)
X2*X5	0,008* (0,0044)	0,229* (0,1357)	0,237* (0,1129)	0,228** (0,0840)
R_p^2	0,453	0,367	0,581	0,616
<i>AIC</i>	-48,23	32,24	-55,51	-69,80

^Δ MQO, mínimo quadrados ordinários; EMV, máxima verossimilhança.

Erro padrão das estimativas são apresentados entre parênteses.

Nível de significância "Teste Wald": (p < 0,01)** e (p < 0,05)*

Fonte: Elaborado pelo autor (2009)

2.5 ARTIGO 5

**Carta de Controle baseada em Modelos Beta por Componentes Principais para
Monitoramento de Características de Qualidade do tipo Fração**

**Beta Model-based Control Charts using Principal Components for monitoring of the
type fraction quality characteristics**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

Previsão de envio ao Quality and Reliability Engineering International (ISSN 0748-8017)

CARTA DE CONTROLE BASEADA EM MODELOS BETA POR COMPONENTES PRINCIPAIS PARA MONITORAMENTO DE CARACTERÍSTICAS DE QUALIDADE DO TIPO FRAÇÃO

Ângelo Márcio Oliveira Sant'Anna.

Carla Schwengber ten Caten.

Laboratório de Otimização de Produtos e Processos, Escola de Engenharia / UFRGS;

angelo@producao.ufrgs.br.; tencaten@producao.ufrgs.br

Resumo

As cartas de controle baseada em modelos são ferramentas úteis para monitorar processos quando as características de qualidade (CQ) variam em função de freqüentes ajustes em uma ou mais variáveis de controle (VC) do processo. No entanto, quando há multicolinearidade entre as VC do processo, os modelos de regressão podem apresentar erro no teste de significância dos coeficientes do modelo. Este artigo propõe uma carta de controle baseada em modelos Beta por componentes principais (CCMBCP) para monitorar CQ do tipo fração em função das VC multicolineares e uma análise de diagnóstico das alterações identificadas pela CCMBCP. A CCMBCP foi validada em um estudo de caso simulado. O resultado da análise de sensibilidade, via simulação de Monte Carlo, indicaram que a CCMBCP apresenta desempenho satisfatório para monitorar variáveis do tipo fração em função de VC multicolineares.

Palavras-chave: Controle estatístico de qualidade, Carta de controle baseada em modelos, Carta de controle por componentes principais.

BETA MODEL-BASED CONTROL CHARTS USING PRINCIPAL COMPONENTS FOR MONITORING OF THE TYPE FRACTION QUALITY CHARACTERISTICS

Abstract

Model-based control charts are useful tools for monitoring processes when the quality characteristics (QC) frequently vary with adjustments in one or more control variables (CV) of the process. However, when there is high multicollinearity among the VC processes, regression models may have shown error to significance test of the coefficients model. In this we paper propose a Beta model-based control charts by principal components (BMCCPC) to monitor the type fraction quality characteristics in function of multicolineares CV and a diagnostic analysis of the changes identified by CCMBCP. The CCMBCP was validated in a simulated case study. The sensitivity analysis by Monte Carlo simulation indicates that the CCMBCP presents satisfactory performance for monitoring of the type fraction quality characteristics in function of multicolineares CV.

Keyword: Statistical quality control, model-based control charts, principal components control charts.

1. Introdução

O controle estatístico do processo (CEP) vem sendo utilizado para monitorar a qualidade dos produtos em diversos processos industriais, devido principalmente a sua capacidade em distinguir entre causas comuns e causas especiais de variação. Tradicionalmente utilizam-se cartas de controle (CC) para monitorar as características de qualidade (CQ) do processo.

As CC propostas inicialmente por Shewhart requerem que os dados monitorados sejam independentes e identicamente distribuídos em torno de um modelo de referência com média constante. Entretanto, é freqüente observar processos onde as CQ variam em função de alterações do ajuste das variáveis de controle (VC) do processo, gerando assim um modelo de referência distinto para cada novo ajuste.

Uma alternativa proposta na literatura são as cartas de controle baseadas em modelos, onde inicialmente é realizada a previsão do valor da CQ por um modelo de regressão em função do seu relacionamento com as VC do processo. Ao invés de monitorar a média das CQ, as cartas de controle baseadas em modelos monitoram os resíduos entre as CQ observadas e as CQ previstas pelo modelo.

Trabalhos propondo aplicações de CC para monitorar CQ mensuradas em fração do tipo percentual ou proporção utilizando modelos de regressão são escassos na literatura. Uma abordagem recente desenvolvida por Sant'Anna e Caten (2009) propõem uma CC baseada em modelos Beta (CCMB). A CCMB é uma CC para os resíduos *deviance* do modelo de regressão Beta que modela CQ do tipo fração em função das VC do processo.

No entanto, a adequação dos modelos de regressão da CQ em função das VC está restrita a um cenário onde a correlação linear entre as VC não é significativa. Quando há multicolinearidade (i.e. correlação linear acentuada entre as VC), os modelos de regressão podem apresentar efeitos errôneos nas estimativas da variância e covariância de seus coeficientes e prejudicar o teste de significância das VC e a conseqüente adequação do modelo (NETER et al., 2005).

Uma alternativa usual consiste em utilizar regressão por componentes principais (RCP). A análise de componentes principais (ACP) é uma técnica de análise multivariada que visa transformar um grupo de variáveis correlacionadas em variáveis latentes, denominadas componentes principais (CPs). Os CPs representam variáveis ortogonais e são obtidos através da combinação linear das VC. Geralmente, poucos CPs são suficientes para representar as

principais estruturas de correlação nos dados. A RCP é uma regressão da CQ em função dos CPs ortogonais ao invés das VC originais. Aplicações de RCP também são escassas na literatura e estão restritos aos modelos de regressão com CQ contínua.

Este artigo propõe uma CC baseada em modelos Beta por componentes principais (CCMBCP) para monitorar CQ do tipo fração em função das VC multicolineares. A CCMBCP proposta aplica inicialmente a ACP nas VC para eliminar a multicolinearidade entre as VC gerando os CPs, e no segundo momento ajusta o modelo de regressão Beta em função dos CPs e monitorar os resíduos *deviance* do modelo Beta.

Adicionalmente, este trabalho propõe uma abordagem posterior para diagnóstico de alterações apontadas pela CCMBCP, que consiste em utilizar CC Shewhart para monitorar os escores dos CPs (CCCP) que são resultantes de alterações em uma ou mais VC.

A carta de controle proposta será validada por um estudo de caso simulado. A análise de sensibilidade da carta será realizada por simulação de Monte Carlo, com alterações nas VC. A simulação será realizada na linguagem computacional R[®] v.2.9.0.

Na seção 2, é apresentada uma breve revisão sobre as CCMB, a ACP e sobre as CCCP. A seção 3 apresenta a carta de controle baseada em modelos por componentes principais (CCMBCP). A seção 4, apresenta a aplicação da CCMBCP em um estudo simulado. Finalmente, na seção 5, são apresentadas as conclusões sobre o estudo desenvolvido nesse artigo.

2. Referencial teórico

2.1. Modelo de Regressão Beta

Quando a característica de qualidade é mensurada como fração, apresentando valores restritamente no intervalo $[0,1]$, a relação entre a CQ e as VC apresenta restrição no domínio da função ($0 < E(Y) < 1$) (COX, 1996). A fração de sucessos de um referido evento é uma variável aleatória com mensurações positivas e restritas ao intervalo ($0 \leq y \leq 1$), a qual pode ser aproximada por uma distribuição de probabilidade Beta, indexada por dois parâmetros θ_1 , $\theta_2 > 0$ (JOHNSON et al., 1995). Moitra (1990) relata que a distribuição de probabilidade Beta é extremamente versátil e encoraja sua aplicação em modelagem de uma variedade de CQ devido a sua flexibilidade.

Rocke (1993) apresenta um estudo aplicando uma família de transformações para a distribuição Beta conforme diferentes formas assimétricas da densidade. Johnson et al. (1995b) discutem diversos estudos de diferentes ciências em que a distribuição Beta foi melhor ajustada frente a diversas distribuições alternativas, e sugerem que a distribuição Beta é uma boa distribuição para representar dados do tipo fração.

A família de distribuições Beta é composta por todas as distribuições de probabilidade que apresente uma variável aleatória Y cuja função densidade de probabilidade (FDP) depende dos parâmetros θ_1 e θ_2 , e que sua FDP pode ser escrita na forma,

$$f(y; \theta_1, \theta_2) = \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} y^{\theta_1-1} (1-y)^{\theta_2-1} \quad (79)$$

sendo $0 < y < 1$, θ_1 e $\theta_2 > 0$, $\Gamma(\theta)$ uma função *Gama* avaliada no ponto θ , ou seja, com $\Gamma(\theta) = \int_0^\infty y^{\theta-1} e^{-y} dy$, $\theta > 0$.

A função $f(y) = f(y; \theta_1, \theta_2)$ é efetivamente uma função densidade de probabilidade com parâmetros θ_1 e θ_2 . Nota-se ainda, que a função $f(y; \theta_1, \theta_2)$ assume valores estritamente positivos, pois, para qualquer valor de y pertencente ao intervalo $[0,1]$, a função densidade descrita é crescente, ou seja, $f(y) > 0$.

O MRB apresenta uma estrutura de regressão baseada em três componentes: (i) Componente aleatória – supõe-se que a variável resposta (Y) segue unicamente a distribuição de probabilidade Beta; (ii) Componente sistemática – que especifica uma soma linear dos coeficientes desconhecidos das variáveis independentes, conhecido como preditor linear (η); (iii) Função de ligação – que permite modelar a média da variável dependente em relação às demais variáveis independentes através de uma função, conhecida como $g(\cdot)$.

A estrutura do modelo de regressão é expressa da forma,

$$g(\mu) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (80)$$

ou

$$\mu = E(Y) = g^{-1}(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) \quad (81)$$

sendo μ a média da variável resposta Y , γ 's os coeficientes desconhecidos, x_1, x_2, \dots, x_k as variáveis de controle e $g(\cdot)$ a função de ligação. Para dados com distribuição Beta, as funções de ligação mais comumente disponíveis são a logit, a probit, a log-log e a complemento log-log. Essas quatro funções de ligações apresentadas são contínuas e estritamente crescentes no intervalo unitário $[0,1]$.

Conforme Kieschnick e McCullough (2003), o MRB permite gerar estimativas precisas e seguras dos coeficientes, mesmo que o conjunto de dados coletados para a investigação seja consideravelmente pequeno ou os dados mensurados sejam próximos de zero ou próximos de um.

Segundo Johnson et al. (1995b) uma variável aleatória Y com distribuição Beta de dois parâmetros possui média e variância dadas por, respectivamente,

$$E(Y) = \frac{\theta_1}{\theta_1 + \theta_2} \quad e \quad Var(Y) = \frac{\theta_1 \theta_2}{(\theta_1 + \theta_2)^2 \cdot (\theta_1 + \theta_2 + 1)} \quad (82)$$

Ferrari e Cribari-Neto (2004) se baseiam na suposição de que as frações seguem uma distribuição de probabilidade Beta, para realizar uma parametrização da esperança matemática e da variância de uma variável aleatória Y com distribuição Beta, utilizando os parâmetros (θ_1 e θ_2) da distribuição de probabilidade Beta, sendo definidos por: $\mu = \frac{\theta_1}{\theta_1 + \theta_2}$ e $\phi = \theta_1 + \theta_2$, respectivamente, isto é, $\theta_1 = \mu\phi$ e $\theta_2 = (1 - \mu)\phi$. Então, a equação (82), fica escrita como,

$$E(Y) = \mu \quad e \quad Var(Y) = \frac{Var(\mu)}{(1 + \phi)} \quad (83)$$

em que $Var(\mu) = \mu(1 - \mu)$ denota a função de variância do modelo.

Os coeficientes β s e ϕ são obtidos através do procedimento iterativo de estimação por máxima verossimilhança. Este procedimento é realizado pelo algoritmo de otimização não-linear de Broyden–Fletcher–Goldfarb–Shanno (BFGS), também conhecido como *quasi-Newton method*. Segundo Ferrari e Cribari-Neto (2004), o algoritmo BFGS fornece melhor desempenho na estimação dos parâmetros desconhecidos para diferentes tamanhos de amostra e formas da distribuição Beta.

Seja Y uma variável aleatória com distribuição Beta, o logaritmo da estatística de verossimilhança apresenta a forma geral,

$$l(\beta, \phi) = \sum_{i=1}^n l_i(\mu_i, \phi) \quad (84)$$

onde $l_i(\mu_i, \phi) = \log \Gamma(\phi) - \log \Gamma(\mu_i \phi) - \log \Gamma((1 - \mu_i)\phi) + (\mu_i \phi - 1) \log y_i + [(1 - \mu_i)\phi] \log(1 - y_i)$

com μ_i definida de tal forma que satisfaz a equação $\mu_i = g^{-1}(\beta, \mathbf{X}_i)$, que é a média da variável Y . A estatística de verossimilhança é definida pela derivada da função de log-verossimilhança em relação aos parâmetros desconhecidos (β, ϕ) do modelo ajustado, isto é,

$$\frac{\partial l_i(\beta, \phi)}{\partial \beta} = \sum_{i=1}^s \frac{\partial l_i(\mu_i, \phi)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta} \quad (85)$$

e

$$\frac{\partial l_i(\beta, \phi)}{\partial \phi} = \sum_{j=1}^s \frac{\partial l_i(\mu_i, \phi)}{\partial \phi} \quad (86)$$

Note que $\frac{\partial \mu_i}{\partial \eta_i} = \frac{1}{g'(\mu_i)}$; $\frac{\partial \eta_i}{\partial \beta} = X_t$ e também,

$$\frac{\partial l_i(\mu_i, \phi)}{\partial \mu_i} = \phi \left\{ \log \frac{y_i}{1 - y_i} - [\delta(\mu_i \phi) - \delta(1 - \mu_i) \phi] \right\} \quad (87)$$

em que $\delta(\cdot)$ é uma função diferencial *Gama*, i.e. $\delta(z) = \frac{\partial \log \Gamma(z)}{\partial z}$, $z > 0$. Fazendo $y_i^* = \log \frac{y_i}{1 - y_i}$ e $\mu_i^* = [\delta(\mu_i \phi) - \delta(1 - \mu_i) \phi]$, pode-se reescrever como,

$$\frac{\partial(\mu_i, \phi)}{\partial \mu_i} = \phi \{ y_i^* - \mu_i^* \} \quad (88)$$

A estatística de verossimilhança do modelo de regressão é utilizada para obtenção dos resíduos *deviance*. Os resíduos *deviance* são usados na carta de controle baseada em modelos para o monitoramento e controle do processo a fim de detectar a mudança na média ou na variabilidade.

Uma etapa essencial na análise do ajuste de modelos de regressão é a verificação da violação das suposições feitas para o modelo, por exemplo, (i) avaliar a adequação da distribuição de probabilidade proposta para a CQ; (ii) avaliar a capacidade preditiva; e (iii) analisar os resíduos para detectar a presença de pontos extremos (*outliers*). Modelos com bom desempenho estatístico apresentam pequena discrepância entre os dados reais e seus respectivos valores preditos.

2.2. Carta de controle baseada em modelos Beta

A carta de controle baseada em modelos Beta (CCMB) proposta neste trabalho tem por finalidade monitorar processos com CQ mensuradas em fração ou razão, as quais habitualmente seguem distribuições não-Normais e de forma assimétricas. Esta carta de controle utiliza o MRB para estimar a CQ do tipo fração em função das VC. Os resíduos *deviance* gerados pelo método de estimação por máxima verossimilhança são *plotados* na carta de controle a fim de monitorar o comportamento do processo e detectar mudanças devido a causas especiais.

A CCMB apresenta algumas vantagens em relação às outras CC: (i) a CQ prevista é restrita no intervalo [0,1]; (ii) os resíduos *deviance* monitorados são não-correlacionados; e

(iii) o procedimento de estimação dos coeficientes admite relação não-linear entre as variáveis de interesse.

Hawkins (1991) salienta que, conforme o modelo de regressão é ajustado, o desempenho das cartas de controle é aperfeiçoado. Segundo Jearkpaporn et al. (2005), o ajuste inadequado do modelo de regressão pode afetar o desempenho das cartas de controle no monitoramento do processo.

Suponha uma variável aleatória Beta distribuída y em função de um conjunto de k VC x_1, x_2, \dots, x_k com $k+1$ coeficientes desconhecidos $\beta_0, \beta_1, \beta_2, \dots, \beta_k$. Assumindo que uma função de ligação logit relaciona tais variáveis, tem-se que,

$$Y \sim BETA[\text{logit}(\mu) = \boldsymbol{\beta}_t \mathbf{X}_t]; \quad \text{com } t = 0, \dots, k \quad (89)$$

em que cada Y são variáveis aleatórias Beta distribuídas com média $\mu = E(Y)$ e função de ligação inversa $g^{-1}(\cdot) = \log[\mu / (1 - \mu)]$ que modela a média da variável aleatória Y , e

$$\mu = E(y) = g^{-1}(\boldsymbol{\beta}_t \mathbf{X}_t) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}} \quad (90)$$

assim, os coeficientes $\boldsymbol{\beta}$ s do modelo de regressão tem uma importante interpretação que é a razão de chances (*odds ratio*), definida por e^{β_t} .

Observa-se que a relação entre a variável aleatória Y e um dado conjunto de t VC X_t é independente e muitas vezes não-linear. O método proposto monitora de forma univariada a mudança na média da variável aleatória Y dado uma mudança em quaisquer ajuste das VC X_t através do resíduo *deviance* baseado na estatística de verossimilhança.

Pierce e Schafer (1986) descrevem que os resíduos *deviance* seguem a distribuição Normal assintótica. Ferrari e Cribari-Neto (2004) discutem que a distribuição dos resíduos *deviance* do modelo Beta possui propriedades similares aos resíduos *deviance* dos modelos lineares generalizados. Skinner et al. (2004) afirmam que os resíduos *deviance* são sensíveis aos efeitos de diferentes tipos de mudança na média ou na variabilidade imposta aos dados. Esta característica dos resíduos proporciona a carta de controle baseada em modelos vantagens no monitoramento e controle de processos.

O resíduo *deviance* (r_d) para o modelo de regressão Beta é estimado por,

$$r_{d_i} = \text{sign}(y_i - \hat{\mu}_i) \left\{ 2 \left[\phi \left(\log \frac{y_i}{1 - y_i} - [\delta(\mu_i \phi) - \delta(1 - \mu_i) \phi] \right) \right] \right\}^{1/2} \quad (91)$$

em que a estatística de verossimilhança é utilizada como uma medida de ponderação para os resíduos ordinários ($y_i - \hat{\mu}_i$).

Os resíduos *deviance* são *plotados* na carta de controle univariadas de Shewhart, com limites de controle estatísticos calculados a partir da média \bar{r}_d e variância $s^2(\bar{r}_d)$ amostral dos resíduos. Desta forma, os limites de controle e linha central da CCMB são dados por,

$$\begin{aligned} LSC &= \bar{r}_d + w\sqrt{s^2(\bar{r}_d)} \\ LC_j &= \bar{r}_d \\ LIC &= \bar{r}_d - w\sqrt{s^2(\bar{r}_d)} \end{aligned} \quad (92)$$

onde w é uma constante que define a largura dos limites de controle correspondente ao nível desejado do número médio de amostras até alarme falso (NMA_0). O NMA_0 determina o número de amostras até a ocorrência de um alarme falso quando o processo está sob controle. Uma vez determinado os limites de controle, a CCMB com resíduos *deviance* podem ser utilizadas para monitorar a CQ mensurada em fração restrita ao intervalo $[0,1]$.

2.3. Análise de Componentes Principais

A análise de componentes principais (ACP) consiste numa técnica estatística multivariada que busca capturar informação sobre a estrutura de correlação linear de um grupo de variáveis correlacionadas sob análise (Rencher, 2002). Esta informação é condensada em um número menor de variáveis não-correlacionadas, denominadas Componentes Principais (CPs), que representam as projeções das variáveis originais em novos eixos ortogonais. Os desenvolvimentos da ACP aplicada ao Controle Estatístico Multivariado de Processo devem-se a Jackson & Mudholkar (1979), Jackson (1991), Kourti & MacGregor (1996) e MacGregor (1997). A seguir são apresentados, de maneira resumida, os fundamentos da ACP.

Considere uma matriz $\mathbf{X}_{(n \times k)}$ de dados centrados acerca de k variáveis de processo correlacionadas, onde cada linha contém uma observação k -variada, representada por um vetor linha $\mathbf{x}'_{j(1 \times p)}$. A estrutura de correlação dos dados de matriz \mathbf{X} é obtida na matriz de covariâncias (ou correlações) amostrais $\mathbf{S}_{(k \times k)}$. Como tal matriz é simétrica e não singular, existe uma matriz $\mathbf{U}_{(k \times k)}$ ortogonal, que diagonaliza \mathbf{S} . Desta forma, tem-se $\mathbf{U}'\mathbf{S}\mathbf{U} = \mathbf{S}_c$, onde \mathbf{S}_c é uma matriz diagonal que contém os k autovalores λ_t positivos de \mathbf{S} . A matriz \mathbf{U} apresenta nas suas colunas os k autovetores \mathbf{u}_t que trazem as cargas da combinação linear que projeta as variáveis originais no $t^{\text{ésimo}}$ eixo ortogonal dado pelo $t^{\text{ésimo}}$ CP, para $t=1,\dots,k$. O autovetor λ_t

descreve a variância do $t^{\text{ésimo}}$ CP. Para fins aplicados, cada vetor \mathbf{u}_t é normalizado, isto é, utiliza-se $\frac{\mathbf{u}_t}{\|\mathbf{u}_t\|^2}$.

Como os CPs são não-correlacionados, cada um descreve uma estrutura exclusiva de variabilidade dos dados. Normalmente, como as variáveis de processo apresentam forte correlação linear, um número $l < k$ de CPs trazem parte substancial da estrutura de variabilidade contida na matriz \mathbf{X} .

O vetor $\mathbf{z}_{t(n \times 1)}$, traz os escores referentes ao $t^{\text{ésimo}}$ CP das n observações preliminares do processo, obtidos através de $\mathbf{z}_t = \mathbf{X}\mathbf{u}_t$, para $t = 1, \dots, l$. Considerando que cada variável de processo seguem uma distribuição Normal, o $t^{\text{ésimo}}$ CP segue uma distribuição Normal com média 0 (pois os dados de \mathbf{X} são padronizados) e variância λ_t .

A projeção de uma nova observação k variada dada pelo vetor $\mathbf{x}_{(k \times 1)}$, nos eixos ortogonais definidos pelos l CPs, é obtida através de,

$$\mathbf{z} = \mathbf{U}'\mathbf{x} \quad (93)$$

onde $\mathbf{z} = [z_1, z_2, \dots, z_l]$ é o vetor contendo os l escores referentes a nova observação; a matriz $\mathbf{U} = [\mathbf{u}_1 | \mathbf{u}_2 | \dots | \mathbf{u}_l]$ contém nas suas colunas os l autovetores associados, e \mathbf{U}' representa a sua transposta.

2.4. Carta de Controle via Componentes Principais

Esta seção apresenta a base teórica para construção das CC para os CPs. As cartas de controle por componentes principais (CCCP) propostas constituem-se numa adaptação da tradicional CC de Shewhart para monitoramento dos escores dos CPs ao invés das VC originais (JACKSON, 1991).

Considere a VC X_t de interesse a ser monitorada em um determinado processo e n observações preliminares independentes (x_{1t}, \dots, x_{nt}) retiradas do processo sob controle estatístico. Suponha também que X_t siga uma distribuição Normal, com média $\mu_t = \mu_{0t}$ e desvio-padrão $\sigma_t = \sigma_{0t}$. Na prática os valores de μ_{0t} e σ_{0t} são estimados a partir da amostra preliminar através da média \bar{x}_t e do desvio padrão s_t amostrais.

Considerando que as observações da VC X_t estejam centradas (i. e., $\mu_{0t} = 0$), os limites de controle da CC de Shewhart são dados por $LC = \pm w \cdot \sqrt{s_t^2}$, onde w é o número de desvios, determinado a partir da probabilidade desejada de alarme falso para a CC. Como o desvio padrão do processo é estimado, w é obtido através da distribuição t com $n-1$ graus de

liberdade. O monitoramento da observação da $t^{\text{ésima}}$ VCs X_t é então realizado incluindo cada nova observação na CC_t de Shewhart, para $t=1, \dots, k$.

As CCCPs são derivadas da ACP descrita na seção anterior e aplicadas no monitoramento simultâneo de k VC correlacionadas. Considere agora uma matriz \mathbf{X} , de dados padronizados, contendo em suas linhas n amostras preliminares k -variadas das VC X_t , num processo em controle estatístico, para $t=1, \dots, k$. É construída uma CCCP para monitorar os escores médios de cada CP. Obtém-se $\mathbf{z}_t = \mathbf{X}\mathbf{u}_t$ e considerando que as n observações (x_{1t}, \dots, x_{nt}) da VC X_t estejam padronizadas (i. e., $\mu_{0t(n \times t)} = 0$ e $\sigma_{0t(t \times 1)} = 1$), os limites de controle da CCCP são dados por,

$$LC = \pm w \cdot \sqrt{\lambda_t} \quad (94)$$

onde λ_t representa a variância do $t^{\text{ésimo}}$ CP e w o número de desvios, determinado a partir da probabilidade desejada de alarme falso para a CC. Como λ_t é estimado a partir da amostra preliminar, w é obtido através da distribuição t com $n-1$ graus de liberdade.

O monitoramento dos escores do $t^{\text{ésimo}}$ CP para cada nova amostra das VCs é realizado a partir da CCCP _{t} , colocando na CC o escore $\mathbf{z}_t (z_{1t}, \dots, z_{nt})$ devido ao $t^{\text{ésimo}}$ CP, para $i=1, \dots, n$ e $t=1, \dots, l$, calculado através da equação (93).

3. Carta de Controle baseada em Modelos Beta por Componentes Principais

A carta de controle baseada em modelos Beta por componentes principais (CCMBCP) proposta neste trabalho tem por finalidade monitorar processos que mensuram CQ do tipo fração em função de VC multicolineares. Inicialmente aplica-se a ACP para estimar os CPs das VCs multicolineares. Esta carta de controle utiliza o MRB para estimar a CQ do tipo fração em função dos CPs ortogonais. Os resíduos *deviance* gerados pelo método de estimação por máxima verossimilhança são *plotados* na carta de controle a fim de monitorar o comportamento do processo e detectar alterações devido a causas especiais. Adicionalmente, utilizam-se as cartas de controle baseada em componentes principais (CCCP) para diagnosticar quais VC influenciaram na alteração do processo.

3.1. Etapas para Implementação da CCMBCP

Segundo Woodall (2000), muitos artigos apresentam o monitoramento do processo por cartas de controle negligenciando a Fase I e as suposições vitalmente importantes das CQ investigadas. A implantação da CCMBCP deve ser dividida em três fases. A fase de análise

retrospectiva (Fase I), a fase de monitoramento do processo (Fase II) e fase de diagnóstico (Fase III). Na Fase I busca-se analisar o processo e estimar os parâmetros a serem usados para determinar os limites de controle para a Fase II. Na Fase II, deve-se assumir que os valores dos parâmetros são conhecidos e os dados a serem monitorados seguem a mesma distribuição da Fase I. Atendida esta suposição, monitoram-se os resíduos *deviance* para identificação de causas especiais. Na Fase III, busca-se diagnosticar através das cartas dos escores dos CPs as alterações identificadas na fase anterior.

As etapas de execução da Fase I consistem em:

- (1) Coleta da CQ(Y) e VC (X_t) do processo;
- (2) Análise exploratória dos dados históricos;
- (3) Aplica a técnica de ACP nas t VCs multicolineares gerando os componentes principais (CPs) ortogonais;
- (4) Ajuste do modelo de regressão Beta (MRB) que relacionam a CQ com os CPs gerados. Este ajuste define o número l de CPs pela análise de significância dos coeficientes do MRB;
- (5) Obtenção dos resíduos *deviance* do modelo de regressão Beta por componentes principais a partir da equação (91);
- (6) Cálculo dos limites de controle da CCMBCP segundo estimativas de média e variância amostral dos resíduos *deviance* do MRB por CP conforme equação (92);
- (7) Cálculo dos limites de controle da CCCP a partir da equação (94);

As etapas de execução da Fase II consistem em:

- (8) Coleta de novas observações da CQ (Y) e t VC (X_t).
- (9) Obtenção dos escores dos l CPs referentes às novas observações através da equação (93);
- (10) Estimativa dos novos resíduos *deviance* do MRB por CP a partir da equação (91);
- (11) Monitoramento dos resíduos *deviance* na CCMBCP. Qualquer ponto fora dos limites de controle indica a presença de causas especiais no processo.

A etapa de execução da Fase III consiste em:

- (12) Monitoramento dos escores dos CPs através das (CCCPs) quando a CCMBCP sinalizar uma causa especial.

4. Estudo de Caso Simulado

O estudo de caso simulado representa um processo de manufatura com a CQ fração de produtos não conformes definida por Y e as VC definidas por X_1 ; X_2 ; X_3 e X_4 .

A simulação consistiu na geração de 200 observações das quatro VC com distribuição Normal 4-variada e da CQ com distribuição Beta.

A Tabela 20 apresenta a matriz de correlação entre a CQ Y e as VC do estudo simulado. Nota-se que todas as VC são correlacionadas com a CQ, evidenciando que há relação de causa e efeito entre as VC e a CQ. As variáveis de controle X_1 e X_2 e as variáveis X_3 e X_4 são significativamente correlacionadas entre si, evidenciando multicolinearidade entre as VCs.

Na modelagem inicial das CQ em função das VC, apresentada na Tabela 22, o modelo de regressão Beta apresenta as VC (X_1 ; X_2 ; X_3 e X_4) como não significativas para explicar a CQ fração não conforme (Y), baseado no teste de *Wald* com 5% significância. Conforme a Tabela 21, o fator de inflação da variância (FIV) confirma a presença de multicolinearidade existente entre as VC ($FIV > 5$), evidenciando a interferência na modelagem dos dados do processo.

Tabela 20: Matriz de correlação entre a CQ e as VC do estudo simulado

	Y	X_1	X_2	X_3	X_4
Y	1	0,818 (0,000)	0,812 (0,000)	0,641 (0,000)	0,452 (0,000)
X_1	0,818 (0,000)	1	0,916 (0,000)	0,096 (0,240)	0,160 (0,236)
X_2	0,812 (0,000)	0,916 (0,000)	1	0,035 (0,110)	0,027 (0,182)
X_3	0,641 (0,000)	0,096 (0,240)	0,035 (0,110)	1	0,884 (0,000)
X_4	0,452 (0,000)	0,160 (0,236)	0,027 (0,182)	0,884 (0,000)	1

Significante quando $p < 0,05$.

Fonte: Tabela elaborada pelo autor (2009)

Tabela 21: Estimativas dos coeficientes, erros padrões e valores de FIV para as VC do MRB

Variável	Estimativa	Erro padrão	Wald-test	Pr (> wl)	FIV
<i>Intercepto</i>	-2,2573	0,01288	-175,19	< 0,0001	---
X_1	0,3277	0,30405	1,078	0,281	26,3
X_2	-0,1140	0,30587	-0,373	0,709	26,3
X_3	-0,1510	0,32668	-0,462	0,644	26,3
X_4	0,2471	0,32745	0,755	0,450	26,3
R_p^2	0,300				
AIC	-207,87				

Fonte: Tabela elaborada pelo autor (2009)

A aplicação da técnica de ACP objetiva eliminar a multicolinearidade existente e capturar a direção da variabilidade das VC do processo. A técnica permite obter l componentes principais (CPs) ortogonais, sendo cada CP formado por uma combinação linear distinta das 4 VC originais. Aplicando ACP nas VC correlacionadas do estudo, obtêm-se as estimativas de variabilidade, os percentuais de variação individual e acumulada, e os autovetores dos CPs, conforme a Tabela 22. Os autovetores dos CPs representam a carga e a direção da variabilidade das VC em cada CP.

Tabela 22: Estimativas da variabilidade e dos autovetores dos CPs

Parâmetro	CP₁	CP₂	CP₃	CP₄
Desvio-padrão	1,401	1,328	0,295	0,052
Percentual de Variação	0,526	0,392	0,058	0,024
Percentual acumulado	0,526	0,918	0,976	1,000
Variáveis				
X_1	0,70752	-0,00739	-0,03548	-0,05387
X_2	0,70660	-0,00561	-0,01021	-0,17002
X_3	0,00691	0,76507	-0,08980	0,06045
X_4	0,00810	0,77553	0,14964	-0,28583

Fonte: Tabela elaborada pelo autor (2009)

Na modelagem das CQ em função dos CPs, o modelo de regressão Beta apresenta os dois primeiros CPs (CP_1 e CP_2) como significativos para explicar a CQ fração (Y), baseado no teste de *Wald* com 5% de significância, conforme a Tabela 23.

Tabela 23: Estimativas dos coeficientes, erros padrões para os CP do MRB

Parâmetro	Estimativa	Erro padrão	Wald-test	Pr (> w)	FIV
<i>Intercepto</i>	-2,2574	0,01290	-174,973	< 0,0001	---
CP_1	0,1526	0,00928	16,440	< 0,0001	1,0
CP_2	0,0658	0,00942	6,988	< 0,0001	1,0
CP_3	0,2787	0,70911	0,5119	0,6091	1,0
CP_4	0,4564	1,78326	0,8173	0,4148	1,0
R_p^2	0,7204				
AIC	-509,25				

Fonte: Tabela elaborada pelo autor (2009)

A partir do modelo de regressão Beta ajustado por CPs obtêm-se os resíduos *deviance* (r_d) do modelo, a partir da equação (91), e as respectivas estimativas de média e variância amostral dos resíduos. Estas estimativas são utilizadas para calcular os limites de

controle e linha central da CCMBCP, conforme equação (92). Ademais, são utilizadas também as estimativas de média e variância os escores dos CPs (CP_1 e CP_2) para calcular os limites de controle e linha central da CCCP a partir da equação (94).

Para ilustrar o monitoramento de novas observações em um processo sob controle, foram geradas 100 novas observações das quatro VC com distribuição Normal 4-variada multicolineares e da CQ com distribuição Beta. A partir das novas observações foram obtidos os escores dos dois CPs significativos e os valores preditos pelo MRB por CP estabelecido na Fase I.

Na Fase II, são estimados os novos resíduos *deviance* referentes às novas observações. Estes resíduos *deviance* são *plotados* na CCMBCP a fim de monitorar o comportamento do processo e detectar possíveis alterações devido a causas especiais. Na Fase III, os escores dos CPs são *plotados* nas respectivas CCCPs a fim de diagnosticar quais VC influenciaram na possível alteração do processo.

A Figura 21 ilustra o comportamento dos resíduos *deviance* do MRB por CP e dos escores dos CP_1 e CP_2 para as 100 novas observações. Nota-se que a CCMBCP para os resíduos *deviance* e as CCCPs para os escores dos CP_1 e CP_2 apresentam comportamento aleatório, dentro dos limites de controle, caracterizando um processo sob controle.

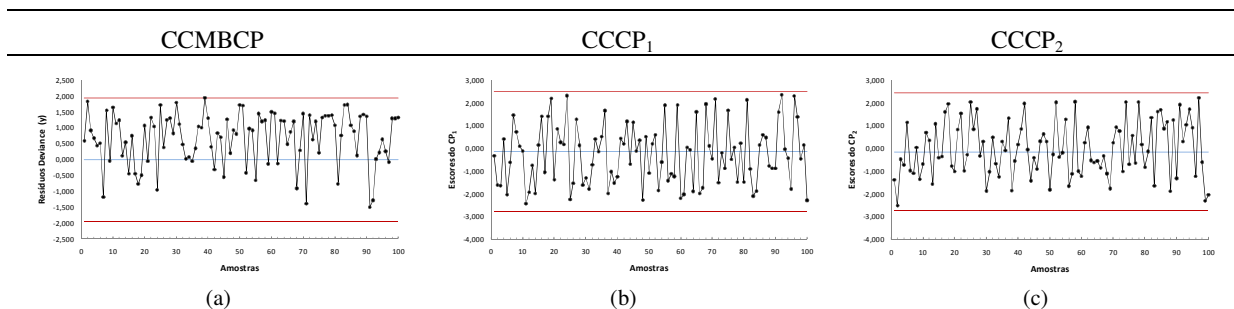


Figura 21: Carta de controle CCMBCP (a) para os resíduos e carta CCCP para os CP_1 (b) e CP_2 (c)

Fonte: Elaborado pelo autor (2009)

Para o processo fora de controle foram induzidas alterações aditivas nas VC em estudo, sendo, $\mathbf{X}_t = \mathbf{X}_t + \varphi$, $t = 1, \dots, l$;

$$\mathbf{z} = \mathbf{U}'\mathbf{x}$$

$$\begin{aligned} z_1 &= a_1(X_1 + \varphi) + a_2X_2 + \dots + a_kX_k \\ z_2 &= b_1(X_1 + \varphi) + b_2X_2 + \dots + b_kX_k \\ &\vdots \\ z_l &= l_1(X_1 + \varphi) + l_2X_2 + \dots + l_kX_k \end{aligned} \quad (95)$$

A alteração aditiva considerou dois tamanhos: um sigma (1σ) e três sigmas (3σ). Por exemplo, uma alteração aditiva com tamanho 1σ indica que a média de X_t é modificada de μ_{x_t} para $\mu_{x_t} + 1\sigma_{x_t}$, onde σ_{x_t} é o desvio-padrão amostral da $t^{\text{ésima}}$ VC, isto implica que, a média dos escores dos l CPs serão também alteradas.

As Figuras 22 e 23 ilustram o comportamento dos resíduos *deviance* do MRB por CP e dos escores dos CP₁ e CP₂ para as 100 novas observações com as alterações propostas. As alterações induzidas consistiram em: (i) alteração em X_1 ; (ii) alteração em X_4 ; e (iii) alteração em ambos X_1 e X_4 . As alterações nas VC X_1 e X_4 objetivaram avaliar a estrutura de correlação proposta e as altas cargas em CP₁ e CP₂, respectivamente.

Nas Figuras 22 e 23(a), (b) e (c), observa-se que as cartas CCMBPC para os resíduos *deviance* detectam pontos fora dos limites de controle, sinalizando causas especiais. As cartas CCCP₁ mostram as alterações induzidas na VC X_1 , evidenciando o deslocamento da média do processo. As cartas CCCP₂ apresentam o comportamento dos escores dentro dos limites de controle, embora, com uma leve alteração na variabilidade.

As Figuras 22 e 23(d), (e) e (f) mostram que as cartas CCMBPC detectam os resíduos *deviance* fora dos limites de controle, sinalizando causas especiais, as cartas CCCP₁ apresenta o comportamento dos escores dentro dos limites de controle, embora, com uma leve alteração na variabilidade e as cartas CCCP₂ mostram as alterações induzidas na VC X_4 , evidenciando o deslocamento da média do processo.

Nota-se que as cartas CCMBPC detectam os resíduos *deviance* fora dos limites de controle, sinalizando causas especiais e as cartas CCCP₁ e CCCP₂ mostram as alterações induzidas nas VC X_1 e X_4 , evidenciando o deslocamento da média do processo, sem alteração na variabilidade (ver Figuras 22 e 23(g), (h) e (i)).

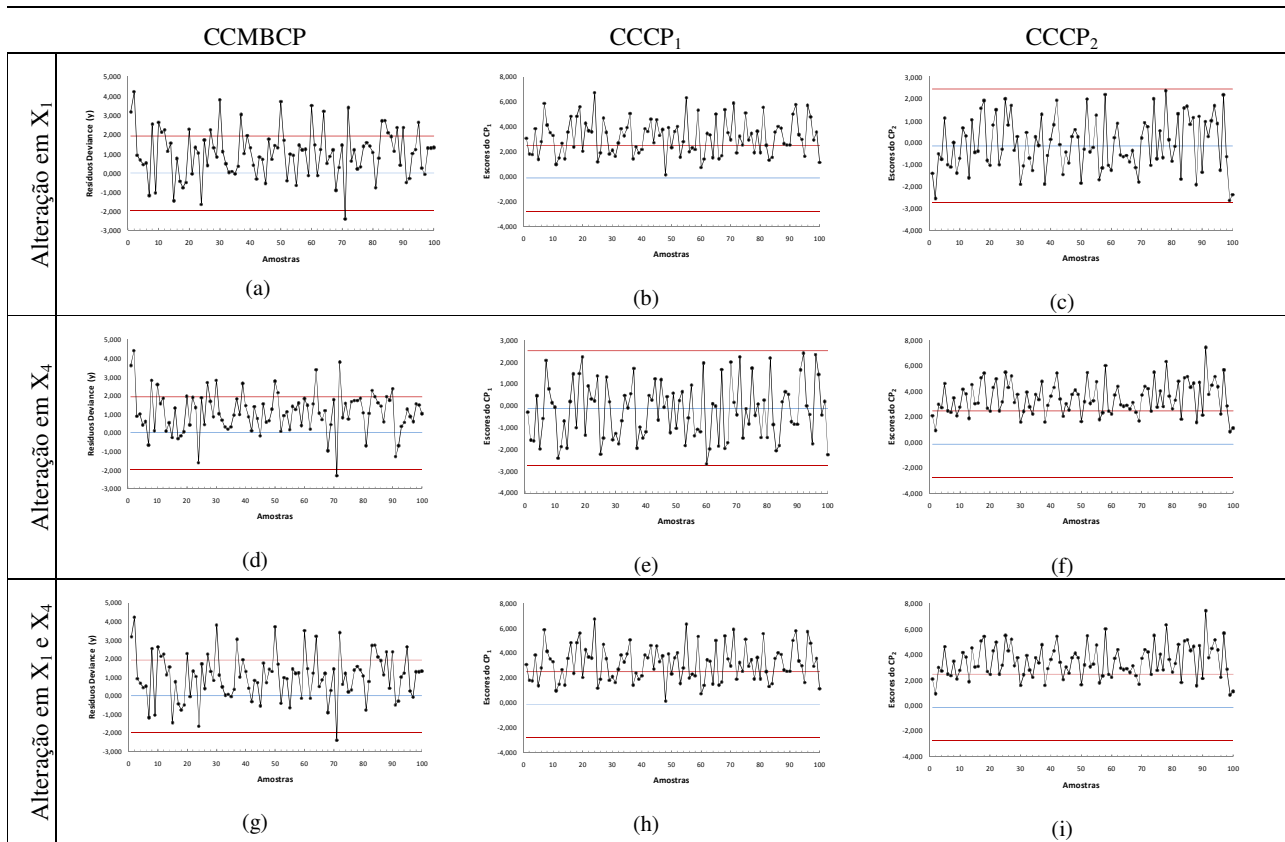


Figura 22: Carta de controle CCMBCP para os resíduos e carta CCCP para os CP₁ e CP₂ segundo alteração 1σ
 Fonte: Elaborado pelo autor (2009)

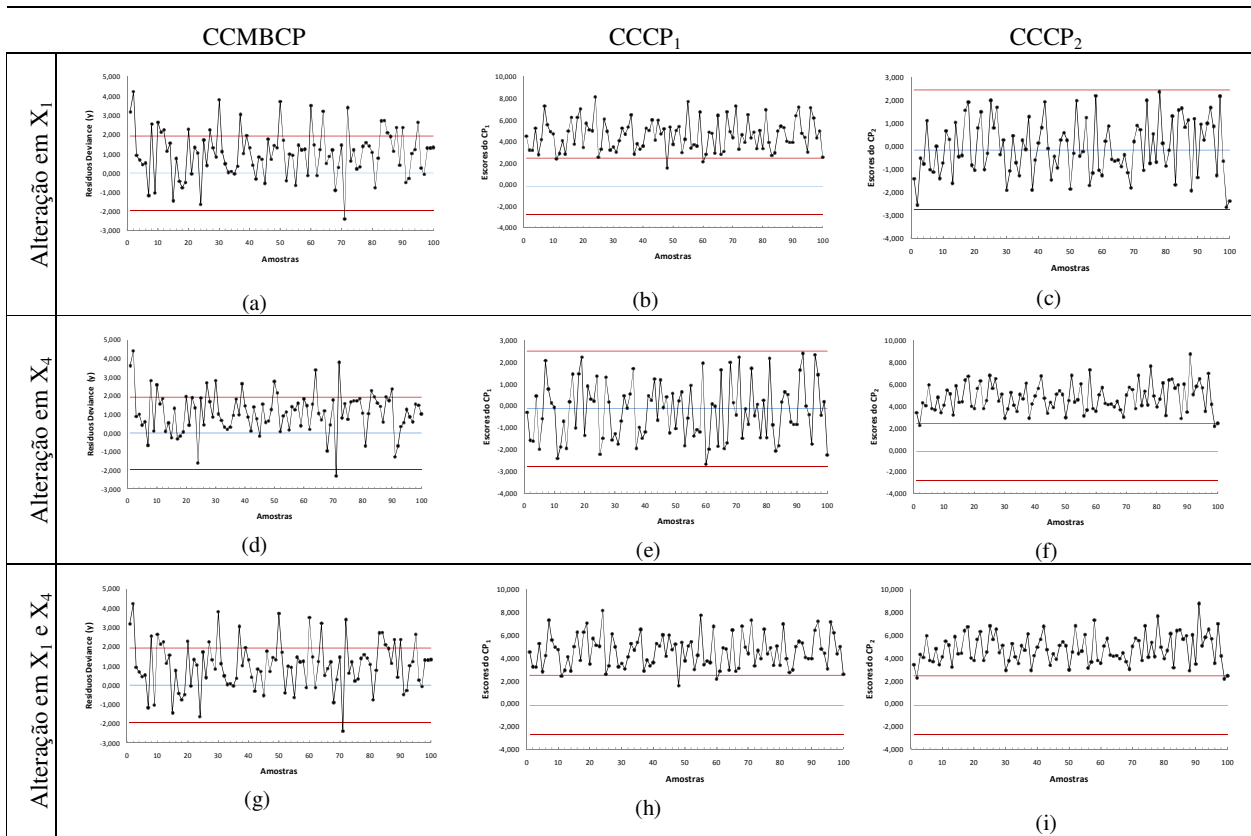


Figura 23: Carta de controle CCMBCP para os resíduos e carta CCCP para os CP₁ e CP₂ segundo alteração 3σ

Fonte: Elaborado pelo autor (2009)

4.1. Análise de Sensibilidade

Foi realizado um estudo de simulação de Monte Carlo, utilizando o *software* livre R[®] v.2.9.0, para analisar a sensibilidade da carta CCMBCP em um processo sob controle e outro fora de controle. Para o processo sob controle foi calculado o número médio de amostras até a presença de uma falsa causa especial (NMA_0). O maior valor de NMA_0 indica uma menor probabilidade de alarmes falsos no processo. O NMA_0 pode ser escrito em função da probabilidade de erro tipo I (α), que é a probabilidade da carta de controle detectar uma alteração no modelo de referência monitorado quando o processo está sob controle estatístico (MONTGOMERY, 2005).

Para o processo fora de controle foi calculado o número médio de amostras até a detecção de causas especiais (NMA_1). O menor valor de NMA_1 representa um menor número médio de amostras coletadas até a detecção da alteração induzida. O NMA_1 pode ser escrito em função da probabilidade de erro tipo II (γ), que é a probabilidade da carta de controle não

detectar uma alteração no modelo de referência monitorado quando o processo está fora de controle estatístico (MONTGOMERY, 2005).

O cálculo dos valores de NMA_0 e NMA_1 consistiu na geração de 10.000 observações das quatro VC com distribuição Normal 4-variada multicolineares e da CQ com distribuição Beta e na estimação dos escores dos CPs, dos valores preditos e dos resíduos *deviance* pelo MRB por CP. A simulação do processo sob controle foi realizada 5.000 vezes, sendo computada a média e o erro-padrão dos percentuais de pontos fora dos limites de controle estatístico ($p_f(\%)$). O número médio de amostras até o falso alarme (NMA_0) para a carta CCMBCP foi calculado conforme:

$$NMA_0 = \frac{1}{\alpha} \quad (96)$$

em que α é a probabilidade de alarmes falsos definido por $\bar{p}_f(\%)$.

A simulação de um processo fora de controle foi realizada 5000 vezes induzindo alterações aditivas nas VC propostas, sendo calculada a média e erro-padrão do percentual de pontos entre os limites de controle até a detecção da causa especial ($p_s(\%)$). O número médio de amostras até a detecção da causa especial (NMA_1) para a carta CCMBCP, pode ser expresso por:

$$NMA_1 = \frac{1}{(1 - \beta)} \quad (97)$$

em que β é a probabilidade de alarmes verdadeiros definido por $\bar{p}_f(\%)$.

A simulação foi desenvolvida com o valor de $w = 3$ para o cálculo dos limites de controle, obtendo para um processo sob controle um $NMA_0 \cong 357 (1,782)$ e para um processo fora de controle com mudança em X_1 um $NMA_1 \cong 3,218(0,024)$, com mudança em X_4 um $NMA_1 \cong 3,038(0,039)$ e com mudança em X_1 e X_4 um $NMA_1 \cong 1,819(0,017)$.

A CCMBCP proposta apresentou desempenho satisfatório no monitoramento da fração não conforme em função dos CPs, pois obteve um valor grande de NMA_0 para um processo sob controle e um valor pequeno de NMA_1 conforme as alterações induzidas de descontrole.

5. Conclusão

O objetivo deste artigo foi propor uma carta de controle baseada em modelos Beta por componentes principais (CCMBCP) para monitorar características de qualidade do tipo fração, com valores restritos ao intervalo $[0,1]$, em função das variáveis de controle multicolineares. Inicialmente foi apresentado um referencial teórico sobre modelos de regressão Beta, carta de controle baseada em modelos Beta, análise de componentes principais e carta de controle baseada em componentes principais. Na seqüência foi proposta a CCMBCP e os passos para a sua implantação.

A CCMBCP foi aplicada em um estudo de caso simulado de um processo de manufatura com uma CQ com distribuição Beta e quatro VC com distribuição Normal 4-variada multicolineares. A simulação de Monte Carlo foi realizada para um processo sob controle e outro fora de controle, induzindo diferentes alterações aditivas. Este estudo permitiu analisar o desempenho da carta CCMBCP na detecção de alterações induzidas no processo e a análise de diagnóstico utilizando as cartas CCCP.

Os resultados da análise de sensibilidade indicaram que a CCMBCP apresenta desempenho satisfatório no monitoramento de CQ do tipo fração em função de VC multicolineares. Adicionalmente, a análise de diagnóstico via CCCP proposta contribuiu na identificação das fontes de alteração do processo.

Conclui-se que CCMBCP pode ser uma ferramenta adequada para o monitoramento de CQ do tipo fração, com valores restritos ao intervalo $[0,1]$, que não seguem a distribuição Normal e que dependam de freqüentes ajustes nas VC, sendo estas multicolineares.

Agradecimentos

Esta pesquisa foi apoiada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/Brasil).

Referências

- (1) SANT'ANNA, A.M.O.; CATEN, C.S. Modelagem da fração de não conformes em processos industriais. **Pesquisa Operacional**, Rio de Janeiro, 2009, no prelo.
- (2) NETER, J.; KUTNER, M H.; NACHTSHEIM, C. J.; LI, W. **Applied Linear Statistical Models**. 5ª ed., New York: Mc Graw-Hill/Irwin, 2005, 1396 p.
- (3) COX, C. Nonlinear quasi-likelihood models: applications to continuous proportions. **Computational Statistical & Data Analysis**, Los Angeles, v. 21, n. 4, p. 449–461, 1996.

- (4) JOHNSON, N.L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous Univariate Distributions**. v. 2, 2^a ed., New York: John Wiley & Sons, 1995, 717p.
- (5) MOITRA, S.D. Skewness and the Beta distribution. **Journal of the Operational Research Society**, Houndmills, v. 41, n. 10, p. 953–961, 1990.
- (6) ROCKE, D.M. On the Beta transformation family. **Technometrics**, Washington, v. 35, n. 1, p. 72–81, 1993.
- (7) KIESCHNICK, R.; McCULLOUGH, B.D. Regression analysis of variates observed on [0,1]: percentages, proportions and fractions. **Statistical Modelling**, Amsterdam, v. 3, n. 3, p. 193–213, 2003.
- (8) FERRARI, S.L.P.; CRIBARI-NETO, F. Beta regression for modeling rates and proportions. **Journal of Applied Statistics**, Sheffield, v. 31, n. 7, p.799–816, 2004.
- (9) HAWKINS, D. M. Multivariate Quality Control Based on Regression-Adjusted Variables. **Technometrics**, Washington, v. 33, n. 1, p. 61-75, 1991.
- (10) PIERCE, D.A.; SCHAFER, D.W. Residuals in generalized linear model. **Journal of the American Statistical Association**, Durham, v. 81, n. 396, p. 977–986, 1986.
- (11) SKINNER, K. R.; MONTGOMERY, D. C.; RUNGER, G. C. Generalized Linear Model-based Control Charts for Discrete Semiconductor Process Data. **Quality and Reliability Engineering International**, Chichester, v. 20, n. 8, p. 777-786, 2004.
- (12) RENCHER, A.C. **Methods of Multivariate Analysis**, 2^a ed., New York: John Wiley & Sons, 2002, 708p.
- (13) JACKSON, J.E. **A User's guide to Principal Components**. New York: John Wiley & Sons, 1991, 570p.
- (14) JACKSON, J. E.; MUDHOLKAR, G. S. Control Procedures for Residuals Associated with Principal Component Analysis. **Technometrics**, Washington, v.21, n. 3, p. 341-349, 1979.
- (15) KOURTI, T. & MACGREGOR, J. F. Multivariate SPC Methods for Process and Product Monitoring. **Journal of Quality Technology**, Milwaukee, v. 28, n. 4, p. 409-428, 1996.
- (16) WOODALL, W.H. Controversies and Contradictions in Statistical Process Control. **Journal of Quality Technology**, Milwaukee, v. 32, n. 4, p.341–378, 2000.
- (17) MACGREGOR, J.F. Using on-line process data to improve quality: challenges for statisticians. **International Statistical Review**, New York, v. 65, n. 3, p. 309-323, 1997.

3 CONSIDERAÇÕES FINAIS

Este capítulo apresenta inicialmente as conclusões da tese, e na seqüência, são apresentadas sugestões para pesquisas futuras advindos de idéias e delimitações desta tese.

3.1 CONCLUSÕES

A tese versou sobre a modelagem e monitoramento de características de qualidade do tipo fração em processos industriais.

Em relação ao objetivo específico **propor uma sistemática para seleção de modelos paramétricos considerando a classificação das variáveis resposta e das variáveis explicativas** o artigo 1 classificou os tipos de variáveis, considerando a natureza dos dados experimentais e elaborou um fluxograma orientativo com os principais modelos de regressão paramétricos pertencentes à classe dos modelos lineares generalizados.

Os modelos paramétricos sugeridos para modelar o experimento de detonação por explosivos apresentaram ajuste adequado aos dados e identificação de variáveis explicativas que influenciam a velocidade de partícula captada por um sismógrafo, conduzindo a resultados seguros. A sistemática desenvolvida se mostrou uma estratégia relevante e adicional para pesquisas em modelagem de dados e otimização experimental

Em relação ao objetivo específico **propor uma modelagem da fração não conforme com enfoque no modelo de regressão Beta e no modelo linear generalizado** o artigo 2 analisou a fração de produtos não conformes em um processo de manufatura em um curtume de couro utilizando os modelos de regressão Beta e de Quase-verossimilhança.

Observou-se que a aplicação de modelos de regressão mais adequados para o tipo de dados no processo de modelagem, como modelo de regressão Beta e de Quase-verossimilhança, aumenta a precisão das estimativas e fornece uma análise de investigação mais consistente. Ademais, estes modelos consideram a relação não-linear entre as variáveis resposta e as variáveis explicativas e a variância dependente da média.

Com base nas análises realizadas recomenda-se o uso dos modelos de regressão Beta e de Quase-verossimilhança a processos que mensurem a fração de produtos não conformes às especificações de manufatura.

Em relação ao objetivo específico **propor uma nova CC, Carta Beta, para monitorar características de qualidade do tipo fração (p), com adaptação dos limites de controle à distribuição Beta**, o artigo 3 categorizou as características de qualidade do tipo fração (percentual ou proporção) considerando a natureza de mensuração e distribuição de probabilidade e revisou as distribuição de probabilidade e as cartas de controle para atributos utilizadas no monitoramento da fração não conforme.

A carta de controle Beta (CCB) desenvolvida forneceu estimativas dos limites de controle contidos no intervalo $[0,1]$ e os resultados da análise de sensibilidade indicaram que a CCB apresenta desempenho satisfatório, pois apresenta maior valor de NMA_0 para a probabilidade de alarme falso conforme o valor nominal desejado e menor número de amostras até a detecção do alarme verdadeiro (NMA_1) dado as alterações induzidas, em comparação com as outras cartas de controle identificadas na literatura.

Em relação ao objetivo específico **propor uma CC baseada em modelos Beta para monitorar características de qualidade do tipo fração em função das variáveis de controle do processo**, o artigo 4 categorizou as características de qualidade do tipo fração (percentual ou proporção) considerando a natureza de mensuração e distribuição de probabilidade e revisou as cartas de controle baseada em modelos e modelos de regressão utilizados na modelagem de características de qualidade do tipo fração.

A carta de controle CCMB desenvolvida utiliza o modelo de regressão Beta e monitora os resíduos *deviance* do modelo Beta em processos que mensuram características de qualidade do tipo fração e que dependam de freqüentes ajustes nas variáveis de controle do processo. Esta carta foi aplicada no monitoramento do processo de uma indústria de borracha e comparada a outras cartas sugeridas na literatura. Os resultados da análise de sensibilidade indicaram que a CCMB apresenta desempenho satisfatório, pois apresenta maior valor de NMA_0 para a probabilidade de alarme falso conforme o valor nominal desejado e menor número de amostras até a detecção do alarme verdadeiro (NMA_1) dado as alterações induzidas, em comparação com as outras cartas de controle identificadas na literatura. Ademais, foram apresentadas as etapas para implementação da carta de controle proposta.

Em relação ao objetivo específico **propor uma CC baseada em modelos Beta por componentes principais para monitorar características de qualidade do tipo fração em função das variáveis de controle multicolineares** o artigo 5 contemplou o uso do modelo de regressão Beta, da análise de componentes principais e da regressão por componentes principais na modelagem de características de qualidade do tipo fração.

A carta de controle CCMBCP desenvolvida utiliza o modelo de regressão Beta por componentes principais e monitora os resíduos *deviance* do modelo Beta em um estudo simulado que mensura características de qualidade do tipo fração e que depende de freqüentes ajustes nas variáveis de controle multicolineares. Os resultados da análise de sensibilidade indicaram que a CCMBCP apresenta desempenho satisfatório, pois apresenta um valor alto de NMA_0 para a probabilidade de alarme falso e um menor número de amostras até a detecção do alarme verdadeiro (NMA_1) dado as alterações aditivas induzidas. Adicionalmente, a análise de diagnóstico via CCCP proposta contribuiu na identificação das fontes de alteração do processo. Ademais, foram apresentadas as etapas para implementação da carta de controle proposta.

3.2 SUGESTÕES PARA PESQUISAS FUTURAS

Este trabalho apresentou a possibilidade de utilizar a distribuição de probabilidade Beta na modelagem e no monitoramento de características de qualidade do tipo fração com valores restritos ao intervalo $[0,1]$, como a fração de produtos não conformes.

Os estudos de modelagem e de monitoramento aqui propostos não se esgotam com a conclusão desta tese. As ferramentas desenvolvidas para monitoramento de características de qualidade do tipo fração se fazem válidas, sugere-se como complemento ao trabalho desenvolvido a realização de pesquisas futuras relacionadas aos seguintes tópicos:

- a) Propor uma carta de controle de somas cumulativas (CUSUM) com limites de controle baseados na CCB e na CCMB;
- b) Propor uma carta de controle de médias móveis exponencialmente ponderadas (EWMA) com limites de controle baseados na CCB e CCMB;
- c) Incorporar uma carta de controle de extrapolação para as VC nas etapas do método de aplicação da CCMB e da CCMBCP.
- d) Obter índices de capacidade adaptados às CCB e CCMB.
- e) Aproximar a distribuição Binomial pela Poisson e comparar com a carta CCB.
- f) Aplicação da técnica de mínimos quadrados parciais, do inglês *partial least squares* (PLS), para o monitoramento de processos multivariados, em abordagem semelhante a carta de controle baseada em modelos Beta;
- g) Elaboração de um aplicativo para operacionalização das cartas CCB, CCMB e CCMBCP;

REFERÊNCIAS

AGRESTI, A. **An Introduction to Categorical Data Analysis**. 2^a ed., New York: John Wiley & Sons, 2007, 372p.

ATKINSON, A.C.; RIANI, M. **Robust Diagnostic Regression Analysis**. New York: Springer, 2000, 327p.

BARNDORFF-NIELSEN, O.E. Normal Inverse Gaussian distribution and stochastic volatility modelling. **Scandinavian Journal of Statistics**, Oslo, v. 24, n. 1, p. 1–13, 1997.

BOURKE, P.D. Detecting a shift in fraction nonconforming using Run-Length control charts with 100% Inspection. **Journal of Quality Technology**, Milwaukee, v. 23, n. 3, p. 225–238, 1991.

BOURKE, P.D. Performance comparisons for the Synthetic control charts for detecting increases in fraction nonconforming. **Journal of Quality Technology**, Milwaukee, v. 40, n. 4, p. 461–475, 2008.

BROWNLEE, K.A. **Statistical theory and methodology in science and engineering**. New York: John Wiley & Sons, 1965, 590p.

CHEN, G. An improved p Chart through simple adjustments. **Journal of Quality Technology**, Milwaukee, v. 30, n. 2, p. 142–151, 1998.

COLLETT, D. **Modelling binary data**. 2^a ed., London: Chapman-Hall, 2003, 387p.

COOK, R.D.; WEISBERG, S. **Residuals and Influence in Regression**, New York: Chapman & Hall, 1982, 230p.

CORDEIRO, G.M. **Modelos Lineares Generalizados**. In: VII Simpósio Nacional de Probabilidade e Estatística, Campinas, São Paulo: ABE, 1986, 286p.

COX, C. Nonlinear quasi-likelihood models: applications to continuous proportions. **Computational Statistical & Data Analysis**, Los Angeles, v. 21, n. 4, p. 449–461, 1996.

COX, C. Nonlinear quasi-likelihood models: applications to continuous proportions. **Computational Statistical & Data Analysis**, Los Angeles, v. 21, n. 4, p. 449–461, 1996.

DIPAOLA, P. P. Use of Correlation in Quality Control. **Industrial Quality Control**, Milwaukee, v. 2, n. 1, p. 10-14, 1945.

DOBSON, A.J. **An Introduction to Generalized Linear Models**. 3^a ed., London: Chapman & Hall, 2002, 225p.

DRAPER, N.R.; SMITH, H. **Applied Regression Analysis**. New York: John Wiley & Sons, 1998, 706p.

FERRARI, S.L.P.; CRIBARI-NETO, F. Beta regression for modeling rates and proportions. **Journal of Applied Statistics**, Sheffield, v. 31, n. 7, p.799–816, 2004.

FERRELL, E. B. Control charts for log-normal universe. **Industrial Quality Control**, Québec, v. 15, n. 1, p. 4–6, 1958.

FLEISS, J.L.; LEVIN, B.; PAIK, M.C. **Statistical methods for Rates and Proportions**. 3^a ed., New Jersey: John Wiley & Sons, 2003, 760p.

GIL, A.C. **Como elaborar projetos de pesquisa**. 4^a ed. São Paulo: Atlas, 2002, 175 p.

GIL, A.C. **Métodos e Técnicas de Pesquisa Social**. 6^a ed. São Paulo: Atlas, 2008, 200 p.

HAIR, J.F JR.; BLACK, B.; ANDERSON, R.E; BABIN, B.; TATHAM, R.L. **Multivariate Data Analysis**. 6^a ed., New Jersey: Prentice-Hall Inc, 2006, 899p.

HAMADA, M.; NELDER, J.A. Generalized linear models for quality-improvement experiments. **Journal of Quality Technology**, Milwaukee, v. 29, n. 3, p. 292–304, 1997.

HAUCK, D.J.; RUNGER, G.C.; MONTGOMERY, D.C. Multivariate statistical process monitoring and diagnosis with grouped regression-adjusted variables. **Communication in Statistics: Simulation and Computation**, Hamilton, v. 28, n. 2, p.309–328, 1999.

HAWKINS, D. M. Multivariate Quality Control Based on Regression-Adjusted Variables. **Technometrics**, Washington, v. 33, n. 1, p. 61-75, 1991.

HAWKINS, D. M. Regression Adjustment for Variables in Multivariate Quality Control. **Journal of Quality Technology**, Milwaukee, v. 25, n.3, p. 170-182, 1993.

HAWORTH, D. A. Regression Control Charts to Manage Software Maintenance. **Journal of Software Maintenance**, Sussex, v. 8, n. 1, p 35-48, 1996.

HEIMANN, P.A. Attributes control charts with large sample size, **Journal of Quality Technology**, Milwaukee, v. 28, n. 4, p. 451–459, 1996.

HURVICH, C.M.; TSAI, C-L. Model selection for Quasi-likelihood models in Small Samples. **Biometrics**, Arlington, v. 51, n. 3, p. 1077–1084, 1995.

JACKSON, J. E.; MUDHOLKAR, G. S. Control Procedures for Residuals Associated with Principal Component Analysis. **Technometrics**, Washington, v.21, n. 3, p. 341-349, 1979.

JACKSON, J.E. **A User's guide to Principal Components**. New York: John Wiley & Sons, 1991, 570p.

JACOBI, L. F.; SOUZA, A. M.; PEREIRA, J. E. S. Gráfico de Controle de Regressão Aplicado na Monitoração de Processos. **Revista Produção**, São Paulo, v. 12, n. 1, p. 46-59, 2002.

JEARKPAPORN, D.; MONTGOMERY, D. C.; RUNGER, G. C.; BORROR, C. M. Model-based Process Monitoring using Robust Generalized Linear Models. **International Journal of Production Research**, London, v. 43, n. 7, p. 1337-1354, 2005.

JOHNSON, N.L.; KEMP, A.W.; KOTZ, S. **Univariate Discrete Distributions**. 3^a ed., New York: John Wiley & Sons, 2005, 646p.

JOHNSON, N.L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous Univariate Distributions**. v. 1, 2^a ed., New York: John Wiley & Sons, 1995, 756p.

JOHNSON, N.L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous Univariate Distributions**. v. 2, 2^a ed., New York: John Wiley & Sons, 1995, 717p.

KAMINSKY, F.C.; BENNEYAN, J.C.; DAVIS, R.D.; BURKE, R. Statistical control charts based on a Geometric distribution. **Journal of Quality Technology**, Milwaukee, v. 24, n. 2, p. 63-69, 1992.

KIESCHNICK, R.; McCULLOUGH, B.D. Regression analysis of variates observed on [0,1]: percentages, proportions and fractions. **Statistical Modelling**, Amsterdam, v. 3, n. 3, p. 193-213, 2003.

KOURTI, T. & MACGREGOR, J. F. Multivariate SPC Methods for Process and Product Monitoring. **Journal of Quality Technology**, Milwaukee, v. 28, n. 4, p. 409-428, 1996.

KUME, H. **Métodos Estatísticos para Melhoria da Qualidade**, São Paulo: Gente, 1993, 245p.

LAKATOS, E.M.; MARCONI, M.A. **Fundamentos de metodologia científica**. 6^a ed. São Paulo: Atlas, 2005, 315p.

LARSON, R.; FARBER, B. **Elementary Statistics**. 3^a ed., New Jersey: Prentice-Hall Inc, 2005, 615p.

LEE, A.H.; STEVENSON, M.R.; WANG, K.; YAU, K.K.W. Modeling young driver motor vehicle crashes: data with extra zeros. **Accident Analysis and Prevention**. Barrington, v. 34, n. 4, p.515-521, 2002.

LEE, Y.; NELDER, J.A. Generalized linear models for the analysis of quality improvement experiments. **The Canadian Journal of Statistics**, Ottawa, v. 26, pp. 95-105, 1998.

LOREDO, E. N.; JEARKPAPORN, D.; BORROR, C. M. Model-based Control Chart for Autoregressive and Correlated Data. **Quality and Reliability Engineering International**, Chichester, v. 18, n. 6, p. 489-496, 2002.

LYLE, P. **Regression Analysis of Production Costs and Factory Operations**, 2^a ed., New York: Hafner Publishing Company, 1944. 204 p.

MACGREGOR, J.F. Using on-line process data to improve quality: challenges for statisticians. **International Statistical Review**, New York, v. 65, n. 3, p. 309-323, 1997.

MANDEL, B. J. The Regression Control Chart. **Journal of Quality Technology**, Milwaukee, v. 1, n. 1, p. 1-9, 1969.

MANSFIELD, E.; WEIN, H. H. A Regression Control Chart for Costs. **Journal of Royal Statistical Society Series C – Applied Statistics**, London, v. 7, n. 1, p. 48-57, 1958.

McCOLL, J.I.; MOTLEY, T.J. Control charts applicable when the fraction nonconforming is small. **Journal of Quality Technology**, Milwaukee, v. 30, n. 3, p. 240–247, 1998.

McCULLAGH, P.; NELDER, J.A. **Generalized Linear Models**, 2^a ed., London: Chapman & Hall, 1989, 511p.

MOITRA, S.D. Skewness and the Beta distribution. **Journal of the Operational Research Society**, Houndmills, v. 41, n. 10, p. 953–961, 1990.

MONTGOMERY, D. C. **Design and Analysis of Experiments**. 6^a ed., New York: John Wiley & Sons, 2004, 684 p.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to Linear Regression Analysis**. 4^a ed., New York: John Wiley & Sons, 2006, 641 p.

MONTGOMERY, D.C. **Introduction to Statistical Quality Control**. 5^a ed., New York: John Wiley & Sons, 2005, 732p.

MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. **Introduction to the Theory of Statistics**, London: McGraw-Hill, 1986, 564p.

MYERS, R.H.; MONTGOMERY, D.C.; VINING, G.G. **Generalized Linear Models with applications in Engineering and the Sciences**. New York: John Wiley & Sons, 2002, 512p.

NELDER, J.A.; WEDDERBURN, R.W.M. Generalized Linear Models. **Journal of the Royal Statistical Society A**, London, v. 135, n. X, p. 370–384, 1972.

NELSON, P. R. Control charts for Weibull processes with standards given. **IEEE Transactions on Reliability**, Knoxville, v. 28, n. 3, p. 383-387, 1979.

NETER, J.; KUTNER, M H.; NACHTSHEIM, C. J.; LI, W. **Applied Linear Statistical Models**. 5^a ed., New York: Mc Graw-Hill/Irwin, 2005, 1396 p.

PERRY, M. B.; PIGNATIELLO, J.J Jr.; SIMPSON, J.R. Estimating the change point of the Process Fraction nonconforming with a Monotonic change Disturbance in SPC. **Quality and Reliability Engineering International**, Chichester, v. 23, n. 3, p. 327-339, 2007.

PIERCE, D.A.; SCHAFER, D.W. Residuals in generalized linear model. **Journal of the American Statistical Association**, Durham, v. 81, n. 396, p. 977–986, 1986.

PRENTICE, R.L. Binary Regression using an extended Beta-Binomial distribution, with discussion of correlation induced by covariate measurement errors. **Journal of the American Statistical Association**, Durham, v. 81, n. 396, p.321–327, 1986.

QUESENBERRY, C.P. On properties of Binomial Q Charts for Attributes. **Journal of Quality Technology**, Milwaukee, v. 27, n. 3, p. 204–213, 1995.

QUESENBERRY, C.P. SPC Q charts for a Binomial parameter p : short and long runs. **Journal of Quality Technology**, Milwaukee, v. 23, n. 3, p. 239–246, 1991.

R Development Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Disponível em <http://www.r-project.org>, ISBN 3-900051-07-0, 2009.

RAO, C.R.; WU, Y. Linear model selection by cross-validation. **Journal Statistical Planning and Inference**, Hamilton, v. 128, n. 1, p. 231–240, 2005.

RENCHE, A.C. (2002) **Methods of Multivariate Analysis**, 2ª ed., New York: John Wiley & Sons, 708p.

REYNOLDS, M.R.; STOUMBOS, Z.G. A CUSUM charts for monitoring a proportion when inspecting continuously. **Journal of Quality Technology**, Milwaukee, v. 31, n. 1, p. 87–108, 1999.

REYNOLDS, M.R.; STOUMBOS, Z.G. A general approach to modeling CUSUM charts for a proportion. **IIE Transactions**, Philadelphia, v. 31, n. 1, p. 87–108, 2000.

ROCKE, D.M. On the Beta transformation family. **Technometrics**, Washington, v. 35, n. 1, p. 72–81, 1993.

ROTHSCHILD, B. F.; ROTH, S. R. Statistical Process Control of Plating Solutions with Regression Control Charts. **The SAMPE Journal**, Covina, v. 22, n. 5, p. 37-41, 1986.

RYAN, T. P. **Statistical Methods for Quality Improvement**. New York: John Wiley & Sons, 1989, 446p.

RYAN, T.P.; SCHWERTMAN, N.C. Optimal limits for attributes control charts. **Journal of Quality Technology**, Milwaukee, v. 29, n. 1, p. 86–98, 1997.

SANT'ANNA, A.M.O. **Método de orientação à modelagem de dados mensurados em proporção**. 2006. 104f. Dissertação (Mestrado em Engenharia de Produção) – Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil, 2006.

SANT'ANNA, A.M.O.; CATEN, C.S. Modelagem da fração de não conformes em processos industriais. **Pesquisa Operacional**, Rio de Janeiro, 2009, no prelo.

SANT'ANNA, A.M.O.; CATEN, C.S. Modelagem da proporção de defeituosos usando Modelo de Quase-verossimilhança. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 25, 2005, Porto Alegre. **Anais...** Porto Alegre: ABEPRO, 40p., 2006.

SCHADER, M.; SCHMID, F. Two rules of thumb for the approximation of the Binomial distribution by Normal distribution. **The American Statistician**, Alexandria, v. 43, n. 1, p. 23-24, 1989.

SHEWHART, Walter A. **Economic Control of Quality of Manufactured Product**, Princeton: Van Nostrand Reinhold, 1931, 170p.

SHU, L.; TSUI, K.; TSUNG, F. A review of regression control charts, **Encyclopedia of Statistics in Quality and Reliability**, Chichester, v. 260, p. 1-9, 2008.

SHU, L.; TSUI, K.; TSUNG, F. Run-length performance of regression control charts with estimated parameters. **Journal of Quality Technology**, Milwaukee, v. 36, n. 3, p.280–292, 2002.

SHU, L; TSUNG, F; TSUI, K. L. Run-length Performance of Regression Control Charts with Estimated Parameters. **Journal of Quality Technology**, Milwaukee, v. 36, n. 3, p. 280-292, 2004.

SIM, C.H.; LIM, M.H. Attribute charts for zero-inflated process. **Communications in Statistics – Simulations and Computation**, Hamilton, v. 37, n. 7, p. 1440–1452, 2008.

SKINNER, K. R.; MONTGOMERY, D. C.; RUNGER, G. C. Generalized Linear Model-based Control Charts for Discrete Semiconductor Process Data. **Quality and Reliability Engineering International**, Chichester, v. 20, n. 8, p. 777-786, 2004.

SKINNER, K. R.; MONTGOMERY, D. C.; RUNGER, G. C. Process Monitoring for Multiple Count Data using Generalized Linear Model-based Control Charts. **International Journal of Production Research**, London, v. 41, n. 6, p. 1167-1180, 2003.

SOONG, T.T. **Fundamentals of Probability and Statistics for Engineers**, New York: John Wiley & Sons, 2004, 391p.

SPANOS, C. J. Statistical Process Control in Semiconductor Manufacturing. **Proceedings of the IEEE**, New York, v. 80, n. 6, p. 819-830, 1992.

SULLIVAN, J.H.; WOODALL, W.H. A comparison of multivariate control charts for individuals observations. **Journal of Quality Technology**, Milwaukee, v. 28, n. 4, p. 398–408, 1996.

VASCONCELLOS, K.L.P.; CRIBARI-NETO, F. Improved maximum likelihood estimation in a new class of Beta regression models. **Brazilian Journal of Probability and Statistics**, São Paulo, v. 19, n. 1, p.13–31, 2005.

WADE, M. R.; WOODALL, W. H. A Review and analysis of cause-Selecting Control Charts. **Journal of Quality Technology**, Milwaukee, v. 25, n. 3, p. 161-169, 1993.

WEDDERBURN, R.W.M. Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. **Biometrika**, Oxford, v. 61, n. 3, p. 439–447, 1974.

WEI, B-C.; HU, Y-Q.; FUNG, W-K. Generalized leverage and its applications, **Scandinavian Journal of Statistical**, Oslo, v. 25, n. 1, p. 25–37, 1998.

WEIS, P. E. An Application of Two-way X-Bar Chart. **Industrial Quality Control**, Milwaukee, v. 14, n. 6, p. 23-27, 1957.

WILL, G.W.; DAVIS, A.W. Generalized asymptotic expansions of Cornish-Fisher type. **The Annals of Mathematical Statistics**, Chapel Hill, v. 39, n. 4, p. 1264–1273, 1968.

WINTERBOTTOM, A. Simple adjustments to improve control limits on attribute charts, **Quality and Reliability Engineering International**, Chichester, v. 9, n. 2, p. 105–109, 1993.

WOODALL, W.H. Control charts based on Attribute data: Bibliography and Review. **Journal of Quality Technology**, Milwaukee, v. 29, n. 2, p.172–183, 1997.

WOODALL, W.H. Controversies and Contradictions in Statistical Process Control. **Journal of Quality Technology**, Milwaukee, v. 32, n. 4, p.341–378, 2000.

WOODALL, W.H. The use of control charts in health-care and public health surveillance. **Journal of Quality Technology**, Milwaukee, v. 38, n. 2, p.89–104, 2006.

YANG, Z.; KURALMANI, V.; XIE, M.; TSUI, K-L. On the performance of geometric charts with estimated control limits. **Journal of Quality Technology**, Milwaukee, v. 34, n. 4, p. 448–458, 2002.

ZHANG, C.W.; XIE, M.; GOH, T.N.; LIU, J.Y. A control chart for the Gamma distribution as a model of time between events. **International Journal of Production Research**, London, v. 45, n. 23, p. 5649–5666, 2007.

ZHANG, G.X. Cause-selecting control charts – A new type of quality control charts, **The QR Journal**, Milwaukee, v. 12, n. 4, p.221–225, 1985.

APÊNDICE

A.1 Matriz de dados utilizados nos estudos desenvolvidos

A Tabela 24 apresenta os conjuntos de dados de processos retirados da literatura para ilustrar a aplicação da carta de controle Beta proposta no Artigo 3.

Tabela 24: Dados retirados da literatura: Montgomery 2005, p. 180 (a); Draper e Smith 1998, p. 101 (b) e Brownlee 1965, p. 454 (c) utilizados na aplicação da carta de controle Beta

Amostra	Tamanho da amostra [n]	Peças defeituosas [d]
1	50	12
2	50	15
3	50	8
4	50	10
5	50	4
6	50	7
7	50	16
8	50	9
9	50	14
10	50	10
11	50	5
12	50	6
13	50	17
14	50	12
15	50	22
16	50	8
17	50	10
18	50	5
19	50	13
20	50	11
21	50	20
22	50	18
23	50	24
24	50	15
25	50	9
26	50	12
27	50	7
28	50	13
29	50	9
30	50	6

(a)

Amostra	Prop Amendoins (120 libras)
1	0,99971
2	0,99979
3	0,99982
4	0,99830
5	0,99957
6	0,99961
7	0,99798
8	0,99972
9	0,99642
10	0,99658
11	0,99982
12	0,99975
13	0,99855
14	0,99932
15	0,99908
16	0,99970
17	0,99863
18	0,99933
19	0,99858
20	0,99987
21	0,99958
22	0,99909
23	0,99859
24	0,99985
25	0,99811
26	0,99877
27	0,99961
28	0,99942
29	0,99788
30	0,99821
31	0,99971
32	0,99718
33	0,99889
34	0,99961

(b)

Amostra	Perda Conversão
1	0,042
2	0,014
3	0,037
4	0,012
5	0,018
6	0,005
7	0,032
8	0,020
9	0,005
10	0,037
11	0,006
12	0,016
13	0,011
14	0,028
15	0,008
16	0,007
17	0,027
18	0,015
19	0,006
20	0,015
21	0,015

(c)

A Tabela 25 apresenta a matriz de dados experimentais utilizados na construção dos modelos de regressão Beta e aplicação da carta de controle baseada em Modelos Beta proposta no Artigo 4, totalizando 18 ensaios.

Tabela 25: Matriz dos dados experimentais coletados no processo de manufatura de pneus radiais

Obs	Fatores Controláveis					Var. Respostas		
	CROD X1	CESC X2	CFLAN X3	PFRIS X4	PCV X5	RFV Y1	ERV Y2	MSS Y3
1	-1	-1	-1	-1	1	0,333	0,444	0,0140
2	-1	1	1	1	-1	0,300	0,400	0,0339
3	1	1	1	1	1	0,300	0,600	0,0719
4	1	1	-1	1	-1	0,200	0,600	0,0267
5	-1	-1	1	1	1	0,100	0,100	0,0167
6	0	0	0	0	0	0,200	0,500	0,0108
7	-1	1	-1	1	1	0,200	0,700	0,0532
8	-1	-1	1	-1	-1	0,300	0,800	0,0155
9	1	-1	1	-1	1	0,100	0,600	0,0311
10	-1	1	-1	-1	-1	0,200	0,300	0,0730
11	1	-1	1	1	-1	0,500	0,500	0,0828
12	0	0	0	0	0	0,100	0,400	0,0220
13	1	-1	-1	-1	-1	0,400	0,500	0,0464
14	1	1	1	-1	-1	0,111	0,222	0,0210
15	-1	-1	-1	1	-1	0,200	0,600	0,0362
16	1	-1	-1	1	1	0,500	0,800	0,0558
17	1	1	-1	-1	1	0,100	0,400	0,0337
18	-1	1	1	-1	1	0,500	0,800	0,0692