

Uma revisão da literatura acerca dos gráficos de controle para monitoramento de variáveis do tipo fração e proporção

André Felipe Berdusco Menezes

Departamento de Estatística, Universidade Estadual de Maringá, Brasil

Resumo

É muito comum em controle de qualidade o monitoramento de variáveis do tipo proporção ou fração. Na literatura estatística seguramente o gráfico p de Shewhart é a principal ferramenta utilizada para o monitoramento deste tipo de característica. O presente trabalho visa apresentar e comparar por meio de aplicações reais principais abordagens presentes na literatura para lidar com variáveis restritas ao intervalo $(0, 1)$. Nesse sentido, este artigo apresenta uma discussão de quatro refinamentos do gráfico p de Shewhart e o recente gráfico de controle com suposição de distribuição Beta (BCC). Ademais, uma extensão do BCC introduzida por Bayer, Tondolo e Müller (2018) que permite monitorar uma variável interesse na presença de outras variáveis de controle também é explorada. Todas as devidas implementações computacionais foram desenvolvidas no software R e os códigos estão disponíveis no apêndice deste artigo.

Palavras-chave: gráficos de controle, proporção de não conformes, distribuição Beta.

1 Introdução

Seguramente os gráficos de controle p e np são as principais ferramentas para monitorar variáveis do tipo proporção ou fração (MONTGOMERY, 2013). Nessas abordagens assumimos que a característica de interesse tem distribuição Binomial e os limites de controle para a proporção de interesse (p) são obtidos utilizando a aproximação pelo teorema do limite central para a distribuição Normal.

Uma alternativa mais elegante e flexível foi proposta por Sant’Anna e Caten (2012) assumindo que a proporção de interesse tem distribuição Beta. Como apontam os autores em várias situações práticas a proporção (p) é muito pequena ou grande, logo a distribuição Binomial é assimétrica e a aproximação para a Normal pode não ser satisfatória, permitindo valores fora do intervalo $(0, 1)$, os quais não tem significado prático. Vale ressaltar que Ryan (1989), Chen (1998) e Joeques e Barbosa (2013) propuseram correções para os limites do gráfico de controle p , no entanto por serem correções aditivas elas não interferem na forma da distribuição.

É muito comum em um processo de produção várias características serem observadas simultaneamente, ao invés de somente uma característica. Quando este fato ocorre e sabe-se que um conjunto de características influenciam em uma variável de interesse, então o

gráfico de controle de regressão (regression control chart) introduzido por Mandel (1969) pode ser uma ferramenta útil para a análise.

Recentemente, Bayer, Tondolo e Müller (2018) introduziram uma nova abordagem para monitorar uma variável de interesse quando a mesma está relacionada (associada) com outra(s) variável(is) de controle. Em sua proposta os autores utilizam o modelo de regressão Beta parametrizado introduzido por Cepeda-Cuervo (2001) e Ferrari e Cribari-Neto (2004) para monitorar a média e a dispersão de variáveis do tipo fração ou proporção na presença de variáveis de controle.

O presente artigo tem como objetivo apresentar e comparar por meio de aplicações reais os principais avanços e propostas presentes na literatura para monitoramento de variáveis do tipo fração e proporção. Nesse sentido, para melhor sistematização o trabalho foi dividido em três seções. A seção 2 discute as propostas e refinamentos dos gráficos de controle para variáveis do tipo fração e proporção. Na seção 3, 6 conjuntos de dados são analisados utilizando as metodologias descritas. Por fim, alguns comentários finais concluem este artigo.

2 Revisão da literatura

Esta seção é dedicada para discussão das abordagens para monitoramento de variáveis do tipo fração e proporção. Assim, na subseção 2.1 discuto o gráfico de controle p juntamente com três refinamentos propostos na literatura para redução de vício dos limites de controle. Enquanto que na subseção 2.2 é apresentada a proposta de Sant'Anna e Caten (2012) utilizando a distribuição Beta para obter os limites de controle. Por fim, estendendo o gráfico de controle Beta a subseção 2.3 aborda a recente proposta de Bayer, Tondolo e Müller (2018) para monitoramento de variáveis limitadas na presença de variáveis de controle.

2.1 Gráfico de controle p e seus refinamentos

Considere que uma amostra aleatória de n unidades de um determinado produto seja selecionada, se X representa o número de unidades do produto que não está em conformidade, então a variável aleatória X tem distribuição Binomial com função massa de probabilidade dada por

$$\Pr(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, \dots, n \quad (1)$$

em que p é a proporção de unidades não conformes. A média e variância de X são dadas por np e $np(1 - p)$, respectivamente.

A proporção amostral de não conformes é definida como a razão do número de unidades

não conformes na amostra pelo tamanho amostral, isto é, $\hat{p} = \frac{X}{n}$. A média e variância da variável aleatória \hat{p} são dadas, respectivamente, por

$$\mathbb{E}(\hat{p}) = p \quad \text{e} \quad \mathbb{V}(\hat{p}) = \frac{p(1-p)}{n}. \quad (2)$$

Se a proporção de não conformes é conhecida ou estimada com precisão então utilizamos o teorema do limite central para aproximar a distribuição de \hat{p} para a distribuição Normal com média e variância definidas em (2). Logo, os limites de controle são definidos por

$$\text{UCL} = p + 3\sqrt{\frac{p(1-p)}{n}} \quad \text{e} \quad \text{LCL} = p - 3\sqrt{\frac{p(1-p)}{n}}. \quad (3)$$

Estes limites são conhecidos na literatura como *3-sigma* ou $3\text{-}\sigma$ limites (MONTGOMERY, 2013). Entretanto, essa aproximação é considerada razoável quando $np(1-p) > 5$ e $0.1 < p < 0.9$ ou quando $np(1-p) > 25$.

Segundo Joeques e Barbosa (2013) em processos de alta qualidade os valores de p são usualmente pequenos e as amostras não são grandes o suficiente. Logo, os limites de controle baseado na aproximação Normal podem não ser adequados para descrever a real situação do processo.

Neste sentido, Ryan (1989) propôs um refinamento dos limites (3), o qual é definido como

$$\text{UCL}_1 = \text{UCL} + \frac{1.25}{n} \quad \text{e} \quad \text{LCL}_1 = \text{LCL} + \frac{1.25}{n}. \quad (4)$$

Chen (1998) estendeu a proposta de Winterbottom (1993) em relação ao refinamento dos limites de controle para o gráfico p utilizando a expansão Cornish–Fisher de quantis para corrigir a não normalidade. Para este caso, os limites de controle corrigidos são expressos por

$$\text{UCL}_2 = \text{UCL} + \frac{4}{3n}(1-2p) \quad \text{e} \quad \text{LCL}_2 = \text{LCL} + \frac{4}{3n}(1-2p). \quad (5)$$

No recente trabalho de Joeques e Barbosa (2013) novos limites de controle foram propostos refinando a correção de Winterbottom (1993) utilizando expansões dos cumulantes de terceira e quarta ordem. Os limites de controle aprimorados para o gráfico p são definidos por

$$\text{UCL}_3 = \text{UCL}_2 - \frac{p(1-p) + 2}{6n^2 [p(1-p)/n]^{1/2}} \quad \text{e} \quad \text{LCL}_3 = \text{LCL}_2 - \frac{p(1-p) + 2}{6n^2 [p(1-p)/n]^{1/2}}. \quad (6)$$

2.2 Gráfico de controle Beta

Na análise de dados restritos ao intervalo unitário, seguramente a distribuição Beta é a mais utilizada devido principalmente a flexibilidade de sua função densidade de pro-

bilidade bem como a existência de expressões para sua média e variância (GUPTA; NADARAJAH, 2004; JOHNSON; KOTZ; BALAKRISHNAN, 1995). Uma variável aleatória Y pertencente a família de distribuições Beta, indexada por dois parâmetros, tem função densidade de probabilidade (f.d.p.) definida por

$$f(y | \theta_1, \theta_2) = \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} y^{\theta_1-1} (1-y)^{\theta_2-1}, \quad 0 < y < 1 \quad (7)$$

em que $\theta_1, \theta_2 > 0$ são parâmetros de forma e $\Gamma(\theta_1) = \int_0^\infty u^{\theta_1-1} e^{-u} du$ é a função gama completa. A esperança e variância de Y são definidas, respectivamente, por

$$\mathbb{E}(X) = \frac{\theta_1}{(\theta_1 + \theta_2)} \quad \text{e} \quad \text{Var}(X) = \frac{\theta_1 \theta_2}{(\theta_1 + \theta_2)^2 (\theta_1 + \theta_2 + 1)}. \quad (8)$$

Apesar do intervalo da distribuição ser definido entre 0 e 1, é possível realizar uma generalização dessa característica para um Y que esteja restrito a um intervalo finito (m, n) , nos permitindo assim, modelar um conjunto de valores que estejam restritos a um intervalo identificável.

A função de distribuição acumulada da variável aleatória Y é definida como

$$F(y | \theta_1, \theta_2) = \int_0^y f(u | \theta_1, \theta_2) du = I_y(\theta_1, \theta_2) = \frac{B(y | \theta_1, \theta_2)}{B(\theta_1, \theta_2)} \quad (9)$$

em que $I_y(\theta_1, \theta_2)$ é a razão da função beta incompleta, $B(\theta_1, \theta_2)$ é a função beta e $B(y | \theta_1, \theta_2)$ é a função beta incompleta, para maiores detalhes ver Abramowitz e Stegun (1974).

A função quantil da distribuição Beta é dada por

$$\psi(\alpha | \theta_1, \theta_2) = F^{-1}(\alpha | \theta_1, \theta_2) = I_\alpha^{-1}(\theta_1, \theta_2), \quad (10)$$

que é a inversa da razão da função beta incompleta. Para maiores detalhes das propriedades da distribuição Beta menciono os livros de Johnson, Kotz e Balakrishnan (1995) e Gupta e Nadarajah (2004).

Sob o contexto de controle estatístico de qualidade Sant'Anna e Caten (2012) introduziram o gráfico de controle Beta (Beta Control Chart, BCC) considerando que a proporção de não conformes pode ser descrita pela distribuição Beta com densidade dada por (7). Assim, os limites de controle para o BCC são obtidos, respectivamente, por

$$\text{LCL} = \bar{p} - w_1 \sqrt{s^2(\bar{p})} \quad (11)$$

$$\text{UCL} = \bar{p} + w_2 \sqrt{s^2(\bar{p})} \quad (12)$$

em que \bar{p} e $s^2(\bar{p})$ são a média e a variância da proporção de não conformes, w_1 e w_2 são constantes que definem a largura dos limites de controle. Os valores de w_1 e w_2 podem

são obtidos por

$$w_1 = \frac{\bar{p} - \psi(\alpha/2 \mid \theta_1, \theta_2)}{\sqrt{s^2(\bar{p})}} \quad (13)$$

$$w_2 = \frac{\psi(1 - \alpha/2 \mid \theta_1, \theta_2)}{\sqrt{s^2(\bar{p})}} \quad (14)$$

em que $\alpha/2$ e $1 - \alpha/2$ definem a região de controle desejada, usualmente assume-se que $\alpha = 0.005$.

Por outro lado, Bayer, Tondolo e Müller (2018) mostraram que substituindo (13) em (11) e (14) em (12) os limites de controle podem ser obtidos diretamente pelo quantil da distribuição Beta, isto é,

$$\text{LCL} = \psi(\alpha/2 \mid \theta_1, \theta_2) \quad (15)$$

$$\text{UCL} = \psi(1 - \alpha/2 \mid \theta_1, \theta_2). \quad (16)$$

Na prática, não conhecemos os verdadeiros valores dos parâmetros θ_1 e θ_2 , portanto devemos estimar eles com base em uma amostra. Para estimação dos parâmetros o método da máxima verossimilhança será adotada, pois o mesmo desfruta de várias propriedades assintóticas (LEHMANN; CASELLA, 1998).

Seja $\mathbf{y} = (y_1, \dots, y_n)$ uma amostra de proporções não conformes, então os estimadores de máxima verossimilhança $\hat{\theta}_1$ e $\hat{\theta}_2$ de θ_1 e θ_2 , respectivamente são obtidos maximizando a função log-verossimilhança cuja expressão é dada por

$$\ell(\theta_1, \theta_2 \mid \mathbf{y}) = n \log \Gamma(\theta_1, \theta_2) - n \log \Gamma(\theta_1) + (\theta_1 - 1) \sum_{i=1}^n \log y_i + (\theta_2 - 1) \sum_{i=1}^n \log(1 - y_i).$$

Os estimadores de máxima verossimilhança não possuem expressão fechada e tem que ser computado numericamente, com o auxílio de um algoritmo iterativo, tais como Newthton-Raphson e quasi-Newton. No presente trabalho as estimativas de máxima verossimilhança foram obtidas por meio da biblioteca `fitdistrplus` do software R.

2.3 Gráfico de controle via regressão Beta

Embora o gráfico de controle Beta seja uma alternativa atraente para monitoramento de variáveis do tipo proporção e fração, o mesmo não possibilita situações onde variáveis de qualidade influenciam a característica de interesse, uma vez que a média e dispersão da variável é constante ao longo das observações. Para preencher essa insuficiência, Bayer, Tondolo e Müller (2018) introduziram o gráfico de controle via regressão Beta (*beta regression control chart*, BRCC), o qual permite monitorar a média e dispersão da variável do tipo proporção de fração na presença de variáveis de controle.

Os autores consideraram a parametrização da distribuição, onde $\mu = \frac{\theta_1}{\theta_1 + \theta_2}$ e $\sigma^2 = \frac{1}{1 + \theta_1 + \theta_2}$, isto é, $\theta_1 = \mu \left(\frac{1 - \sigma^2}{\sigma^2} \right)$ e $\theta_2 = (1 - \mu) \left(\frac{1 - \sigma^2}{\sigma^2} \right)$. Assumindo que a variável de interesse Y tem distribuição Beta com média μ e parâmetro de dispersão σ , então a função densidade de probabilidade de $Y \sim \text{Beta}(\mu, \sigma)$ é definida por

$$f(y | \mu, \sigma) = \frac{\Gamma \left(\frac{1 - \sigma^2}{\sigma^2} \right)}{\Gamma \left[\mu \left(\frac{1 - \sigma^2}{\sigma^2} \right) \right] \Gamma \left[(1 - \mu) \left(\frac{1 - \sigma^2}{\sigma^2} \right) \right]} y^{\mu \left(\frac{1 - \sigma^2}{\sigma^2} \right) - 1} (1 - y)^{(1 - \mu) \left(\frac{1 - \sigma^2}{\sigma^2} \right) - 1}, \quad (17)$$

em que $0 < y < 1$, $0 < \mu < 1$ e $0 < \sigma < 1$. Sob essa parametrização a média e variância de Y são dadas por

$$\mathbb{E}(X) = \mu \quad \text{e} \quad \text{Var}(X) = \mu(1 - \mu)\sigma^2. \quad (18)$$

A função de distribuição acumulada e quantil sob a parametrização (17) pode ser utilizando (9) e (10), isto é,

$$\mathcal{F}(y | \mu, \sigma) = F \left(y | \mu \left(\frac{1 - \sigma^2}{\sigma^2} \right), (1 - \mu) \left(\frac{1 - \sigma^2}{\sigma^2} \right) \right) \quad (19)$$

e

$$\mathcal{Q}(\alpha | \mu, \sigma) = \psi \left(\alpha | \mu \left(\frac{1 - \sigma^2}{\sigma^2} \right), (1 - \mu) \left(\frac{1 - \sigma^2}{\sigma^2} \right) \right). \quad (20)$$

Considere que uma amostra aleatória de tamanho n da característica de interesse y_i , $i = 1, \dots, n$ com distribuição Beta e densidade dada por (17). O modelo de regressão beta com dispersão variando é definido assumindo a seguinte estrutura

$$g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \quad (21)$$

$$h(\sigma_i) = \mathbf{z}_i^\top \boldsymbol{\gamma}, \quad (22)$$

em que $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)^\top$ e $\boldsymbol{\gamma} = (\gamma_0, \dots, \gamma_k)^\top$ são, respectivamente, os vetores de parâmetros associados a média e precisão do modelo de regressão, \mathbf{x}_i e \mathbf{z}_i são os vetores de covariáveis associados a média e dispersão do modelo na i -ésima observação, respectivamente e $g(\cdot)$ e $h(\cdot)$ são funções de ligações apropriadas, duas vezes diferenciáveis em relação aos parâmetros de regressão. Possíveis escolhas para as funções de ligação $g(\cdot)$ e $h(\cdot)$ são *logit*, *probit* ou *cloglog*.

Para um dado comprimento médio de amostras em controle (average run length under control, ARL_0) é possível determinar $\alpha = \frac{1}{\text{ARL}_0}$. Então, os limites do gráfico de controle via regressão Beta são definidos por

$$\text{LCL}_i = \mathcal{Q}(\alpha/2 | \mu_i, \sigma_i) \quad (23)$$

$$\text{UCL}_i = \mathcal{Q}(1 - \alpha/2 | \mu_i, \sigma_i), \quad (24)$$

em que μ_i e σ_i são obtidos pelas estruturas de regressão da média $\mu_i = g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$ e dispersão $\sigma_i = h^{-1}(\mathbf{z}_i^\top \boldsymbol{\gamma})$ e α é a probabilidade fixa de falso alarme.

Na prática, a estimação dos parâmetros μ_i e σ_i é realizada sob o paradigma da verossimilhança, com $\hat{\mu}_i = g^{-1}(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}})$ e $\hat{\sigma}_i = h^{-1}(\mathbf{z}_i^\top \hat{\boldsymbol{\gamma}})$, sendo que os coeficientes de regressão $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ são estimados maximizando a função log-verossimilhança, definida por

$$\ell(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{y}, \mathbf{x}, \mathbf{z}) = \sum_{i=1}^n \ell_i(\mu_i, \sigma_i)$$

em que

$$\begin{aligned} \ell_i(\mu_i, \sigma_i) &= \log \Gamma\left(\frac{1 - \sigma_i^2}{\sigma_i^2}\right) - \log \Gamma\left[\mu_i \left(\frac{1 - \sigma_i^2}{\sigma_i^2}\right)\right] - \log \Gamma\left[(1 - \mu_i) \left(\frac{1 - \sigma_i^2}{\sigma_i^2}\right)\right] \\ &+ \left[\mu_i \left(\frac{1 - \sigma_i^2}{\sigma_i^2}\right) - 1\right] \log y_i + \left[(1 - \mu_i) \left(\frac{1 - \sigma_i^2}{\sigma_i^2}\right) - 1\right] \log(1 - y_i). \end{aligned}$$

A maximização da função log-verossimilhança foi realizada no software R utilizando a função `optim` pelo método BFGS e com chutes iniciais obtidos seguindo a recomendação de Ferrari e Cribari-Neto (2004).

Finalmente Bayer, Tondolo e Müller (2018) propoem o seguinte esquema para implementação do gráfico de controle via regressão Beta:

1. Ajuste o modelo de regressão Beta obtendo as estimativas de máxima verossimilhança $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$ dos vetores de parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$, respectivamente;
2. Obtenha as estimativas da média μ e dispersão σ para cada amostra i , com $i = 1, \dots, n$. Ou seja, $\hat{\mu}_i = g^{-1}(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}})$ e $\hat{\sigma}_i = h^{-1}(\mathbf{z}_i^\top \hat{\boldsymbol{\gamma}})$;
3. Determine os limites de controle para um dado valor de ARL_0 , isto é,

$$\widehat{\text{LCL}}_i = \mathcal{Q}(\alpha/2 \mid \hat{\mu}_i, \hat{\sigma}_i) \tag{25}$$

$$\widehat{\text{UCL}}_i = \mathcal{Q}(1 - \alpha/2 \mid \hat{\mu}_i, \hat{\sigma}_i), \tag{26}$$

em que $\alpha = \frac{1}{\text{ARL}_0}$.

4. Plote cada amostra y_i junto com seu limite de controle $\widehat{\text{LCL}}_i$ e $\widehat{\text{UCL}}_i$, para $i = 1, \dots, n$.

Naturalmente, se existir uma observação y_i fora de seu limite de controle ($\widehat{\text{LCL}}_i$, $\widehat{\text{UCL}}_i$) então o processo não estará operando somente devido ao acaso, ou seja, o processo está fora de controle.

Bayer, Tondolo e Müller (2018) apontam que as principais vantagens do BRCC sobre o tradicional gráfico de controle via regressão são (i) acomoda variáveis assimétrica

e heterocedásticas, (ii) seus limites são restritos ao intervalo unitário e (iii) permite a modelagem simultânea da média e do parâmetro de dispersão.

3 Aplicações

Nesta seção 6 aplicações considerando dados reais com intuito de ilustrar as abordagens discutidas na seção anterior são apresentadas. Em específico as seguintes subseções consistem em (i) descrição dos dados, (ii) apresentação dos limites de controle estimados e (iii) discussão dos resultados encontrados.

3.1 Embalagens não conformes

Nesta primeira aplicação consideramos os dados apresentados em Montgomery (2013) relacionado ao processo de fabricação de embalagens de suco de laranja concentrado e congelado em 30 lotes de 50 embalagens. A característica que será monitorada é a proporção de embalagens não conforme.

Na Tabela 1 apresenta-se os limites dos gráficos de controle considerando as 5 abordagens discutidas na seção 2. Percebe-se que o método baseado na distribuição Beta apresentou maiores limites do que a abordagem de Shewhart e seus refinamentos.

Tabela 1: Limites dos gráficos de controle — embalagens não conformes.

Método	LCL	CL	UCL
BCC	0.0726	0.2318	0.4482
Shewhart	0.0524	0.2313	0.4102
Ryan	0.0774	0.2313	0.4352
Chen	0.0668	0.2313	0.4246
Joeke	0.0643	0.2313	0.4221

Esta diferença fica evidente na Figura 1 onde exibe os gráficos de controle para os diferentes métodos. Verifica-se que com exceção o gráfico BCC as demais abordagens indicam que duas observações estão fora de controle, especificamente as amostras 15 e 23, sendo que a amostra 15 não esta fora de controle segundo o gráfico BCC.

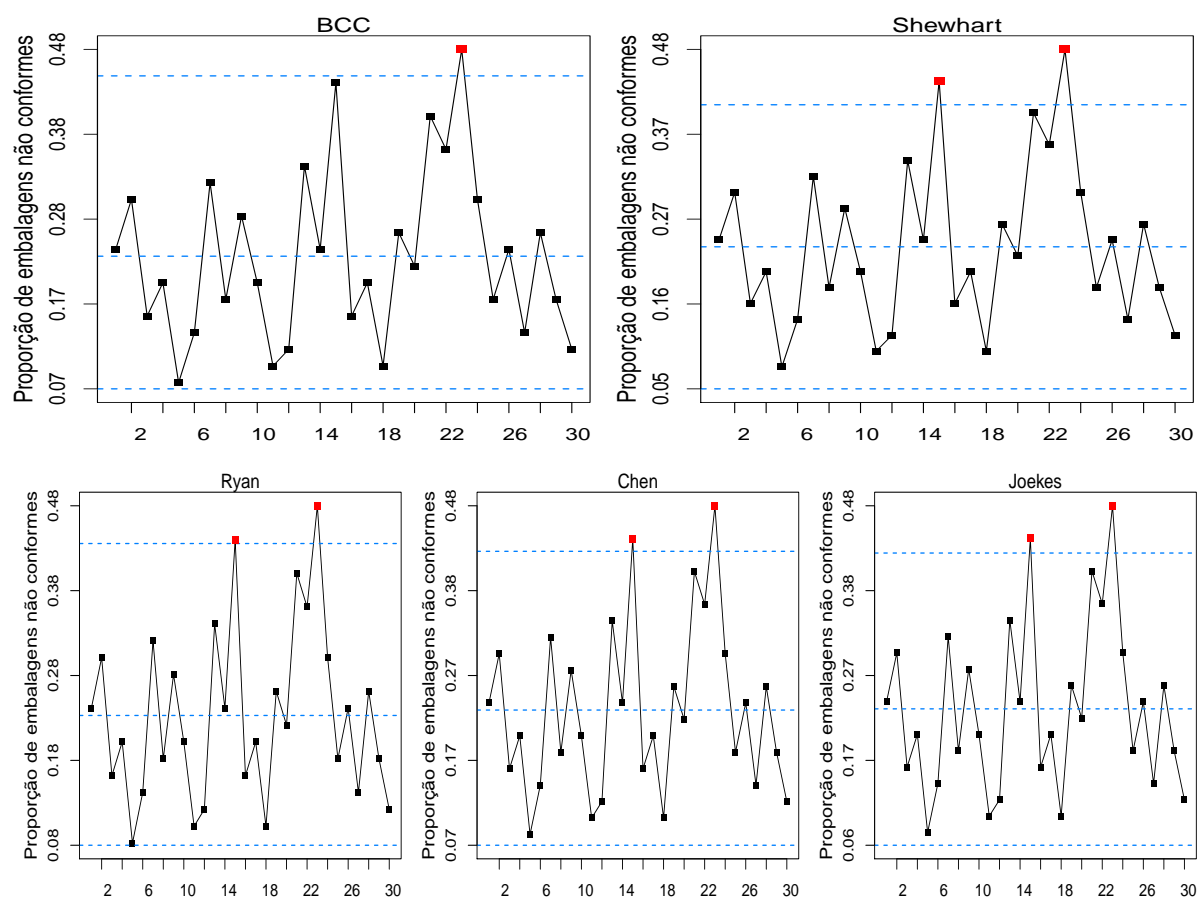


Figura 1: Gráficos de controle considerando diferentes abordagens — embalagens não conformes.

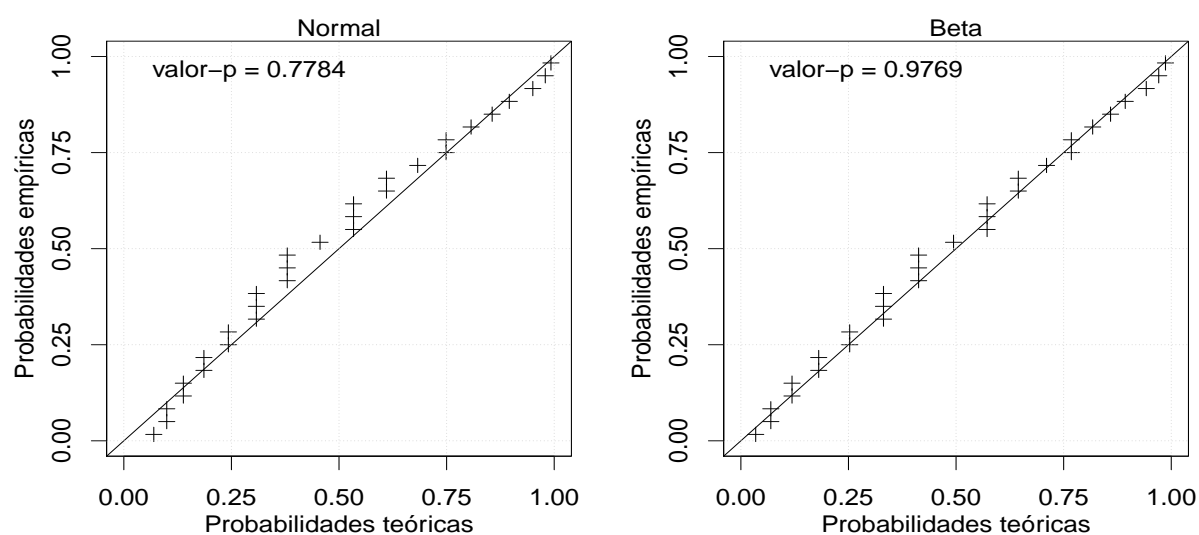


Figura 2: PP-Plots da característica de controle — embalagens não conformes.

3.2 Amendoins não contaminados

Nesta aplicação o conjunto de dados foi retirado de Draper e Smith (1998) e consiste em um estudo de contaminação de amendoins por substâncias tóxicas coletados em 34 lotes de 120 libras. Assim, a característica de qualidade que será monitorada é a proporção de amendoins não contaminados

Os resultados da Tabela 2 indicam que as correções do gráfico p de Shewhart não são condizentes com a característica de interesse uma vez que os limites inferiores e superiores não contém a proporção média estimada de amendoins não contaminados. Este fato comprova que mesmo os refinamentos do gráfico p de Shewhart podem ser inadequados quando a proporção está próxima dos limites do intervalo unitário.

Tabela 2: Limites dos gráficos de controle — amendoins não contaminados.

Método	LCL	CL	UCL
BCC	0.9966	0.9990	0.9999
Shewhart	0.9824	0.9990	1.0156
Ryan	1.0191	0.9990	1.0523
Chen	0.9432	0.9990	0.9764
Joeke	0.8911	0.9990	0.9243

Interessante observar na Figura 3 que o gráfico p de Shewhart aponta que o processo esta em controle, enquanto que o gráfico baseado na distribuição Beta (BCC) indica que as amostras 29 e 31 estão fora dos limites de controle, sugerindo que o processo esteja fora de controle. Nota-se que o BCC é mais sensível para detectar mudanças na média do processo.

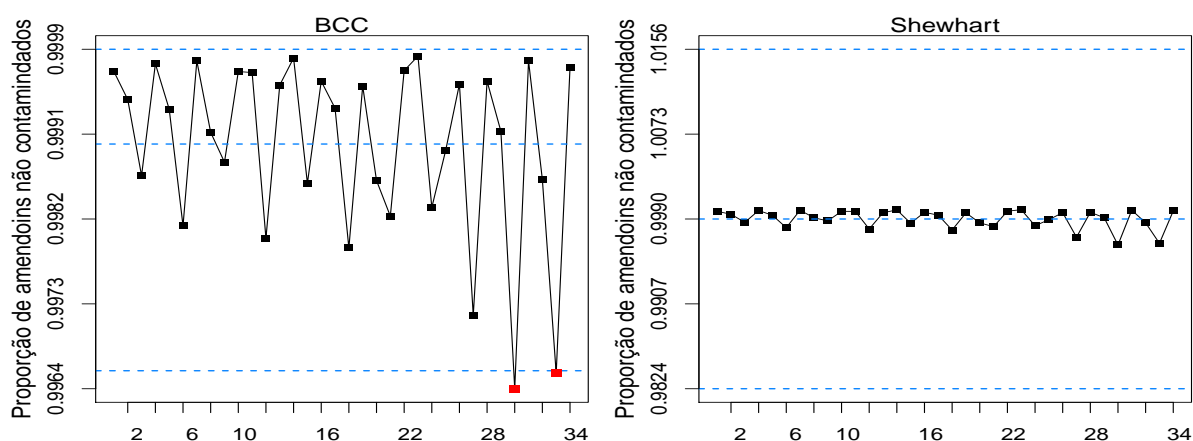


Figura 3: Gráficos de controle considerando diferentes abordagens — amendoins não contaminados.

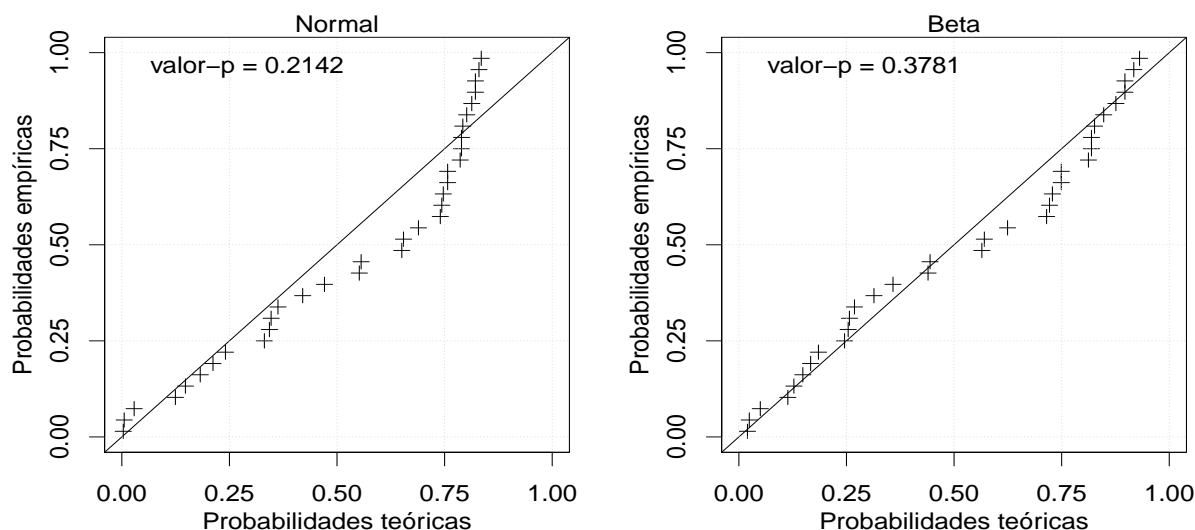


Figura 4: PP-Plots da característica de controle — amendoins não contaminados.

3.3 Oxidação da amônia

Este conjunto de dados foi obtido no livro de Brownlee (1965) e a característica de interesse é a proporção de amônia não convertida a qual foi obtida em um processo de fabricação de ácido nítrico por oxidação de amônia coletados em 21 dias.

A Tabela 3 fornece os valores para os limites de controle de acordo com a abordagem adotada. Observa-se que o gráfico p de Shewhart e seus refinamentos apresentaram um limite inferior para o gráfico de controle menor que 0, havendo assim nenhum sentido.

Tabela 3: Limites dos gráficos de controle — oxidação da amônia.

Método	LCL	CL	UCL
BCC	0.0045	0.0175	0.0390
Shewhart	-0.0684	0.0175	0.1034
Ryan	-0.0089	0.0175	0.1629
Chen	-0.0071	0.0175	0.1647
Joeke	-0.0337	0.0175	0.1381

Conforme ilustra a Figura 5 percebe-se que a o gráfico de controle baseado na suposição de distribuição Beta para a proporção de oxidação de amônia convertida é o único que indica que a amostra 1 esta fora de controle. Nos demais gráficos, observa-se limites bastante amplos sugerindo que o processo esta em controle.

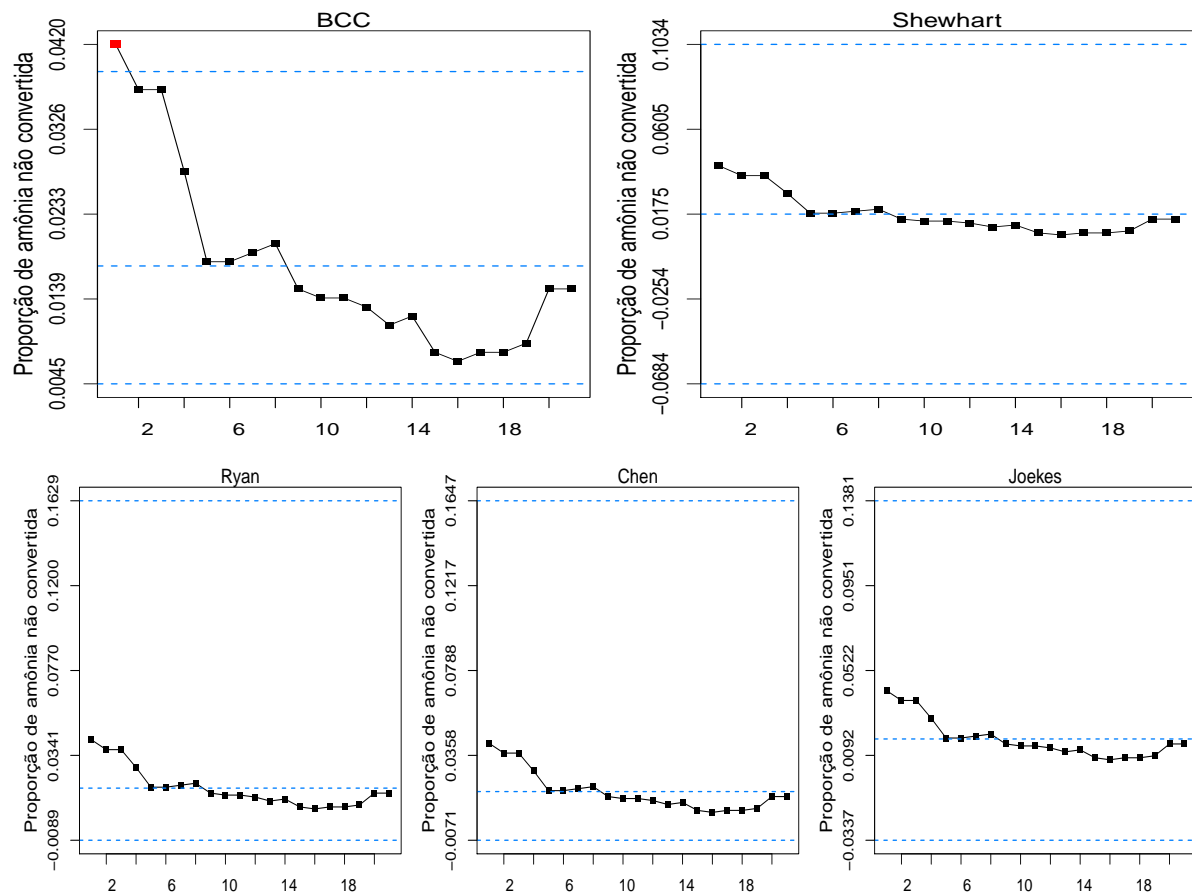


Figura 5: Gráficos de controle considerando diferentes abordagens — oxidação da amônia.

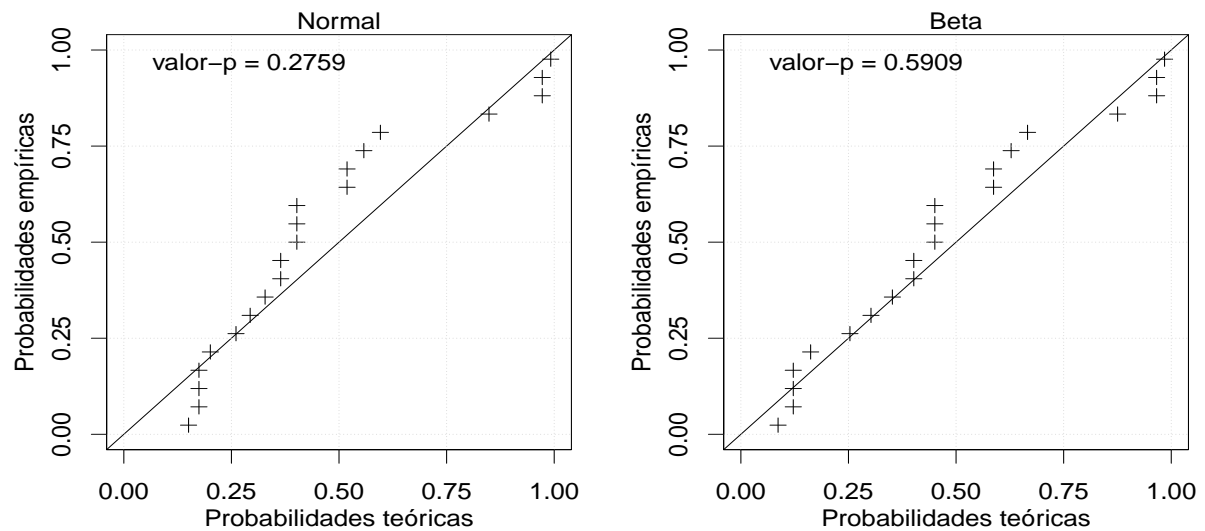


Figura 6: PP-Plots da característica de controle — oxidação da amônia.

3.4 Transistores

Os dados aqui analisados estão presentes no livro de Ryan (1989), sendo a característica de interesse a proporção de transistores não conformes em um total de 1000 inspecionados.

É possível notar na Tabela 4 que o limite inferior e superior de controle do BCC é superior ao de Shewhart e seus refinamentos, sugerindo então uma maior sensibilidade do BCC.

Tabela 4: Limites dos gráficos de controle – transistores

Método	LCL	CL	UCL
BCC	0.0061	0.0106	0.0163
Shewhart	0.0009	0.0106	0.0203
Ryan	0.0021	0.0106	0.0216
Chen	0.0022	0.0106	0.0216
Joekes	0.0021	0.0106	0.0215

Tal sensibilidade é exibida na Figura 7, onde percebe-se que somente o gráfico de controle Beta (BCC) indica que o processo esta fora de controle, sendo que as amostras 2 e 9 estão abaixo do limite inferior de controle.

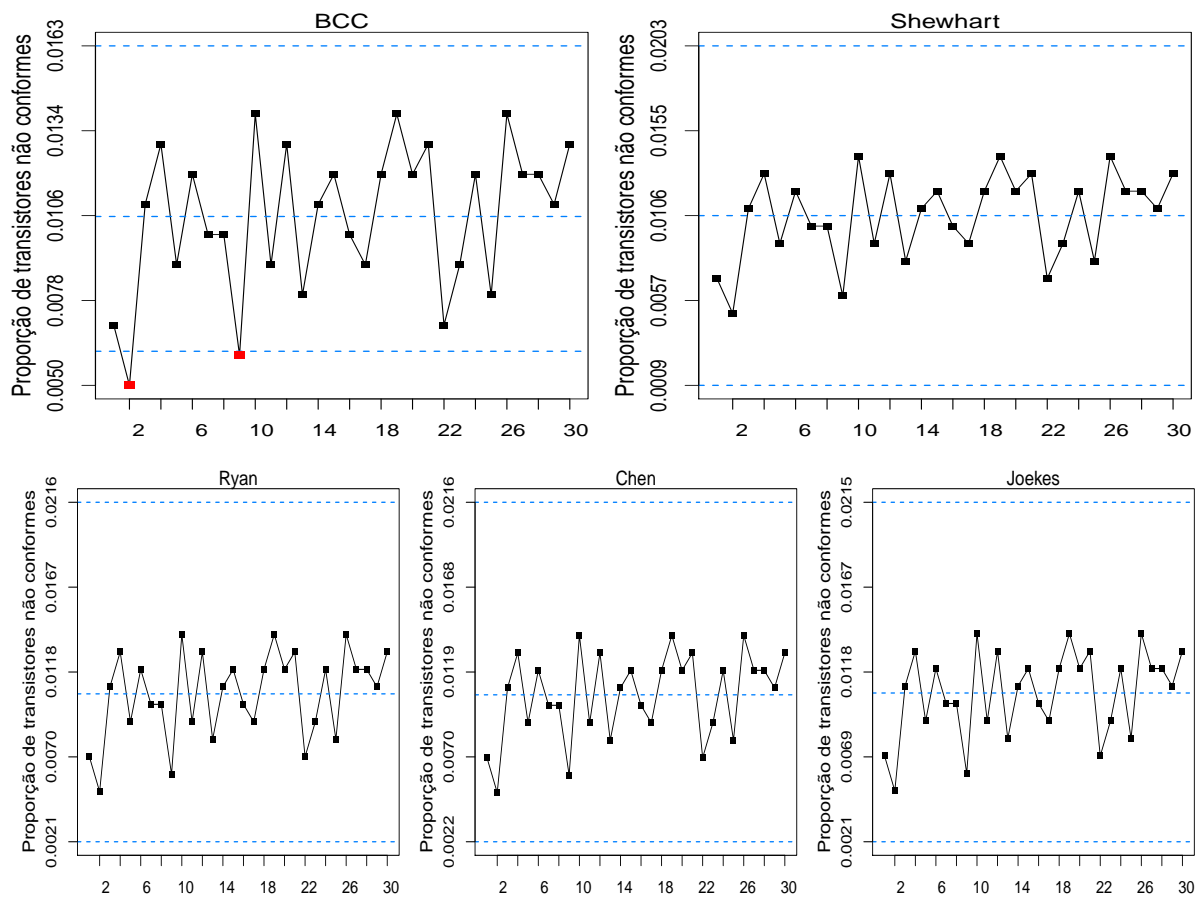


Figura 7: Gráficos de controle considerando diferentes abordagens — transistores.

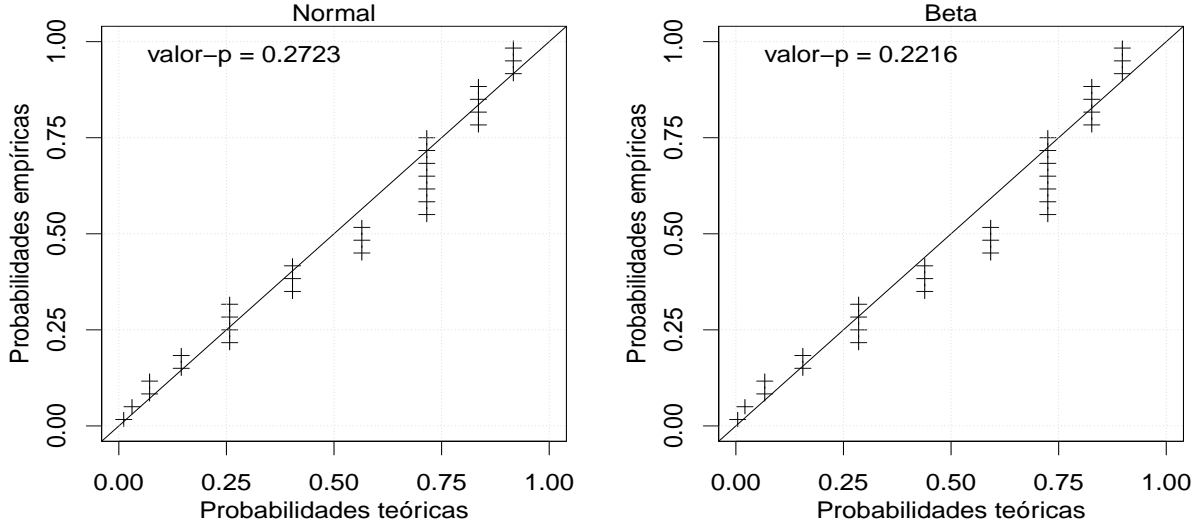


Figura 8: PP-Plots da característica de controle — transistores.

3.5 Manufatura de pneus

Nesta subseção apresento a aplicação realizada por Bayer, Tondolo e Müller (2018) considerando o conjunto de dados analisado por Sant’Anna (2009) relacionado a fabricação de pneus radiais de uma empresa multinacional de borrachas. A característica de qualidade monitorada é a proporção de massa não convertida, obtida pela razão entre o volume de matéria-prima não convertida em produto dividida pelo volume total (SANT’ANNA, 2009). As variáveis de controle no processo são comprimento da banda de rodagem (x_1), comprimento da seção (x_2), comprimento do flanco (x_3), postamento do friso (x_4) e postamento da cova (x_5).

Os dados foram obtidos através de um experimento fatorial fracionário com as variáveis de controle ajustadas em dois níveis mais dois pontos centrais resultando em 18 ensaios experimentais, com 30 pneus em cada ensaio. Assim, as variáveis de controle estão classificadas em três distintos níveis, assumindo os valores $-1, 0$ e 1 (SANT’ANNA, 2009).

Seguindo Bayer, Tondolo e Müller (2018) o modelo de regressão considerado pode ser expresso por

$$\begin{aligned} \log\left(\frac{\mu_i}{1 - \mu_i}\right) &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 I_{1i} + \beta_4 I_{2i} + \beta_5 I_{3i} \\ \log\left(\frac{\sigma_i}{1 - \sigma_i}\right) &= \gamma_0 + \gamma_1 x_{1i} + \gamma_2 I_{1i}, \quad i = 1, \dots, 18 \end{aligned} \tag{27}$$

em que I_1, I_2 e I_3 são interações entre as variáveis de controle, definidas por $I_1 = x_1 \times x_2$, $I_2 = x_1 \times x_4$ e $I_3 = x_2 \times x_5$.

Os resultados do ajuste estão expostos na Tabela 5. Verifica-se que ao nível de 5% de significância todas as covariáveis são importantes para o modelo.

Tabela 5: Resultados inferências do modelo (27).

Parâmetro	Estimativa	Erro padrão	IC 95%	valor- p
β_0	-3.5896	0.2126	(-4.0063, -3.1729)	0.0000
β_1	0.4599	0.2229	(0.0231, 0.8967)	0.0391
β_2	0.4751	0.2291	(0.0261, 0.9240)	0.0381
β_3	-0.6807	0.2198	(-1.1115, -0.2499)	0.0020
β_4	0.3055	0.0178	(0.2707, 0.3403)	0.0000
β_5	0.2106	0.0178	(0.1757, 0.2455)	0.0000
γ_0	-3.0944	0.2552	(-3.5946, -2.5942)	0.0000
γ_1	-0.8731	0.3664	(-1.5911, -0.1551)	0.0172
γ_2	0.8750	0.3660	(0.1576, 1.5923)	0.0168

A partir das estimativas dos parâmetros do modelo foi possível obter o gráfico de controle via regressão Beta, apresentado na Figura 9. Além do BRCC também apresenta-se o gráfico de controle via regressão linear (RCC) e o BCC. Interessante observar que o RCC apresenta limites inferior a 0, evidenciando que a suposição de normalidade não é adequada. Além disso, somente o BRCC indica que o processo está fora de controle, pois a observação seis está abaixo do limite inferior do gráfico de controle. Analisando a característica desta amostra, foi possível notar que sua proporção de massa não convertida está abaixo do valor esperado quando todas as covariáveis são mantidas constantes.

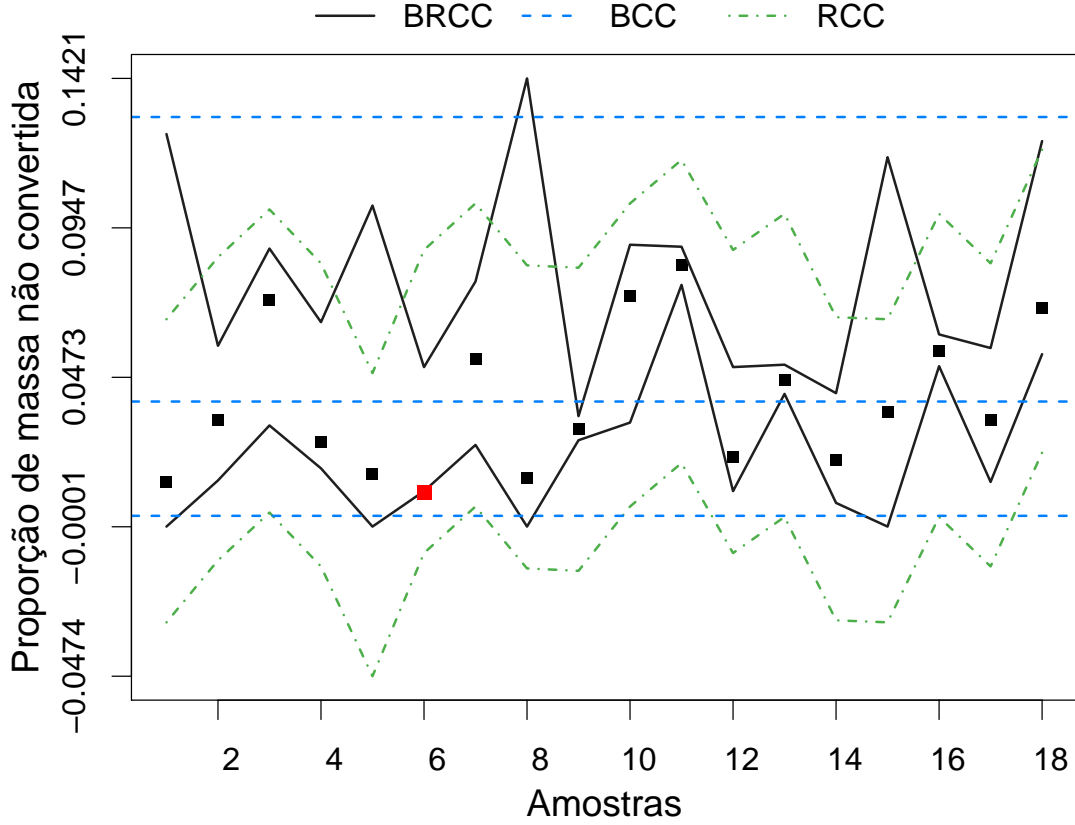


Figura 9: Gráficos de controle considerando diferentes abordagens — manufatura de pneus.

3.6 Umidade relativa

Nesta última aplicação será apresentada a análise dos dados de umidade relativa do ar na cidade de Maringá. Os dados foram coletados pela estação meteorológica 83767 durante o período de 10/05/2016 a 10/06/2018, totalizando um total de 727 observações diárias. Os dados estão disponíveis no site do INMET pelo seguinte endereço <<http://www.inmet.gov.br>>. O objetivo da análise é monitorar a umidade relativa ao longo deste período afim de detectar se em determinado dia a umidade estava fora de controle. Ressalta-se que o monitoramento da umidade relativa pode ser importante por exemplo para diagnósticos ambientais e avaliação de risco.

Nesse sentido variáveis *dummy* representando cada estação do ano foram utilizadas como variáveis de controle (covariáveis), sendo que a primavera foi considerada como estação de referência. Assim, o modelo de regressão considerado é dado por

$$\begin{aligned} \log \left(\frac{\mu_i}{1 - \mu_i} \right) &= \beta_0 + \beta_1 \text{Inverno}_i + \beta_2 \text{Verão}_i + \beta_3 \text{Outono}_i \\ \log \left(\frac{\sigma_i}{1 - \sigma_i} \right) &= \gamma_0 + \gamma_1 \text{Inverno}_i + \gamma_2 \text{Verão}_i + \gamma_3 \text{Outono}_i \end{aligned} \quad (28)$$

para $i = 1, \dots, 727$.

Na Tabela 6 são apresentados os resultados do ajuste do modelo (28). Percebe-se que todas as covariáveis foi significativa para descrever a média da umidade relativa. Em contrapartida, no que tange o parâmetro de dispersão somente a covariável que representa a estação verão (γ_2) foi significativa.

Tabela 6: Resultados inferências do modelo (28).

Parâmetro	Estimativa	Erro padrão	IC 95%	valor- p
β_0	0.6027	0.0536	(0.4976, 0.7078)	0.0000
β_1	-0.2389	0.0723	(-0.3807, -0.0971)	0.0010
β_2	0.5209	0.0704	(0.3828, 0.6589)	0.0000
β_3	0.4600	0.0806	(0.3021, 0.6180)	0.0000
γ_0	-0.6289	0.0668	(-0.7597, -0.4980)	0.0000
γ_1	-0.1011	0.0935	(-0.2844, 0.0822)	0.2795
γ_2	-0.3968	0.0932	(-0.5795, -0.2141)	0.0000
γ_3	0.0348	0.0956	(-0.1525, 0.2220)	0.7161

A Figura 10 apresenta os limites dos gráficos de controle BCC, RCC e BRCC para monitorar a umidade relativa de Maringá. Percebe-se que os gráficos BCC e RCC não detectaram nenhuma amostra fora de controle. Além disso, nota-se que os limites do RCC são largos e extrapolam o limite superior do intervalo $(0, 1)$. Em contrapartida, o BRCC exibe uma boa sensibilidade identificando oito atípicas observações, sendo quatro (113, 119, 463 e 464) durante o inverno, duas (27 e 400) no outono, uma (539) na primavera e uma (603) no verão. Essas observações apresentam os maiores valores de umidade relativa para a sua estação em que ocorrem. Esta conclusão é confirmada na Tabela 7 onde apresento os valores observados da umidade relativa junto com medidas descritivas conforme a estação.

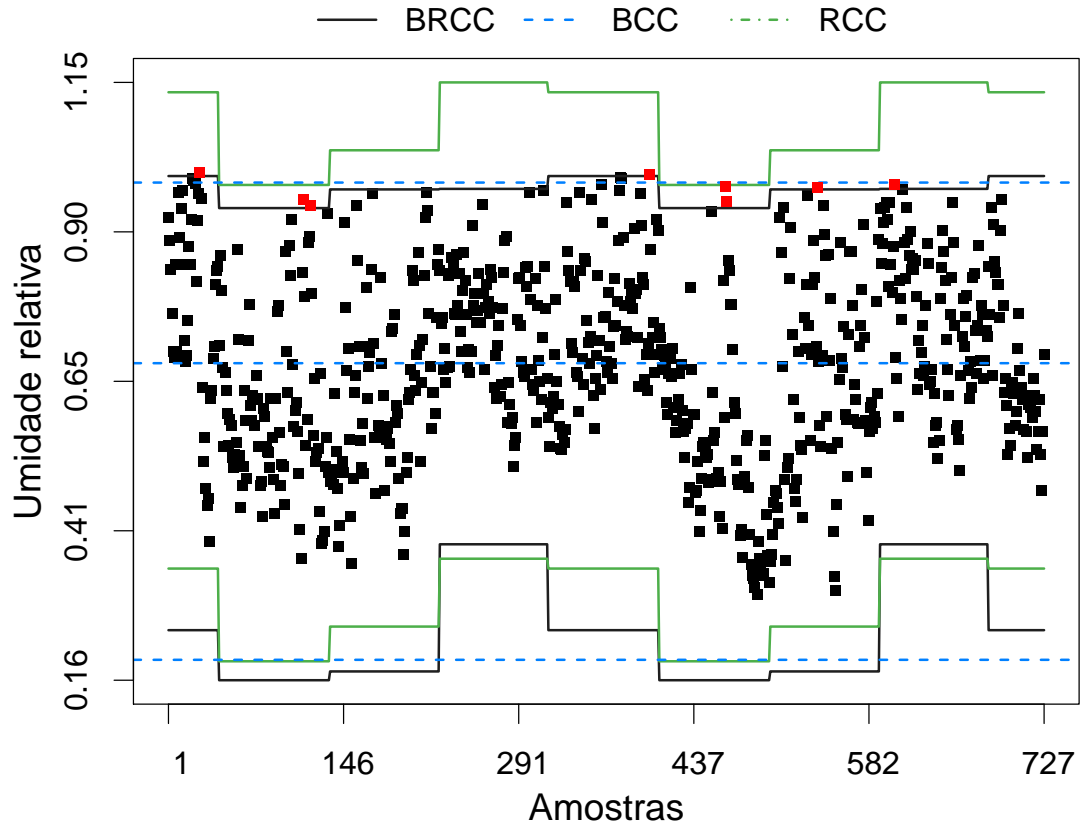


Figura 10: Gráficos de controle considerando diferentes abordagens — umidade relativa.

Tabela 7: Observações fora de controle e estatísticas descritivas conforme a estação.

Obs.	Data	Estação	Umidade relativa	Média	Mediana	Max.	Desv. Padrão
113	2016-08-30	Inverno	95.25	58.38	56.50	97.50	15.00
119	2016-09-05	Inverno	94.25				
463	2017-08-15	Inverno	97.50				
464	2017-08-16	Inverno	95.00				
27	2016-06-05	Outono	99.75	73.68	72.00	99.75	13.75
400	2017-06-13	Outono	99.50				
539	2017-10-30	Primavera	97.25	64.10	62.38	97.25	15.91
603	2018-01-02	Verão	97.75	75.31	76.12	97.75	10.75

4 Considerações Finais

O objetivo deste trabalho foi apresentar uma revisão teórica com aplicações a conjunto de dados reais das principais abordagens para monitoramento de variáveis do tipo proporção e fração. Em um primeiro momento foi discutiu-se o gráfico p de Shewhart e quatro propostas cujo objetivo são melhorar a perspicácia do gráfico p de Shewhart quando a característica de interesse apresenta valores próximo dos limites do intervalo $(0, 1)$. Em

seguida, foi apresentado e discutido em detalhes a recente proposta de Sant'Anna e Caten (2012) do gráfico de controle Beta. Por fim, apresentou-se os detalhes teóricos do gráfico de controle via regressão Beta introduzido por Bayer, Tondolo e Müller (2018).

No que tange as aplicações, seis conjuntos de dados reais foram analisados utilizando as abordagens apresentadas neste trabalho. Em geral, foi possível concluir empiricamente que os gráficos BCC e BRCC são mais sensíveis na detecção de amostras fora de controle do que a abordagem de Shewhart e o RCC. Ademais, ficou evidente que a abordagem baseada na distribuição Normal apresenta limites fora do intervalo unitário, não possuindo nenhum significado prático.

Ressalta-se que as análises foram conduzidas no software R e as funções para cada gráfico de controle estão disponíveis no apêndice deste trabalho.

Referências

- ABRAMOWITZ, M.; STEGUN, I. A. *Handbook of Mathematical Functions with Formulas, graphs, and Mathematical Tables*. New YorkDover Publications, Incorporated, 1974. (National Bureau of Standards Applied Mathematics Series).
- BAYER, F. M.; TONDOLO, C. M.; MÜLLER, F. M. Beta regression control chart for monitoring fractions and proportions. *Computers & Industrial Engineering*, v. 119, p. 416–426, 2018.
- BROWNLEE, K. A. *Statistical theory and methodology in science and engineering*. John Willey & Sons, Inc, 1965.
- CEPEDA-CUERVO, E. *Variability modeling in generalized linear models*. Tese (Doutorado) — Mathematics Institute, Universidade Federal do Rio de Janeiro, 2001.
- CHEN, G. An improved p chart through simple adjustments. *Journal of Quality Technology*, v. 30, n. 2, p. 142–151, 1998.
- DRAPER, N. R.; SMITH, H. *Applied Regression Analysis*. John Willey & Sons, Inc, 1998.
- FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, Taylor & Francis, v. 31, n. 7, p. 799–815, 2004.
- GUPTA, A. K.; NADARAJAH, S. *Handbook of Beta Distribution and Its Applications*. CRC Press, 2004.
- JOEKES, S.; BARBOSA, E. P. An improved attribute control chart for monitoring non-conforming proportion in high quality processes. *Control Engineering Practice*, v. 21, n. 4, p. 407–412, 2013.
- JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. *Continuous Univariate Distributions*. Second. New YorkJohn Wiley & Sons Inc., 1995. v. 2.
- LEHMANN, E. J.; CASELLA, G. *Theory of Point Estimation*. Springer Verlag, 1998.

MANDEL, B. J. The regression control chart. *Journal of Quality Technology*, v. 1, n. 1, p. 1–9, 1969.

MONTGOMERY, D. C. *Introduction to Statistical Quality Control*. 7th. ed. John Wiley & Sons, Inc, 2013.

RYAN, T. P. *Statistical Methods for Quality Improvement*. New York John Wiley & Sons, 1989.

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

SANT'ANNA, Â. M. O.; CATEN, C. S. ten. Beta control charts for monitoring fraction data. *Expert Systems with Applications*, v. 39, n. 11, p. 10236–10243, 2012.

WINTERBOTTOM, A. Simple adjustments to improve control limits on attribute charts. *Quality and Reliability Engineering International*, v. 9, n. 2, p. 105–109, 1993.

Apêndice

Neste apêndice encontra-se os códigos em R utilizados para a análise dos conjuntos de dados.

Listing 1: Implementação do gráfico p de Shewhart e seus refinamentos.

```
1 pcontrol <- function(prop, n, correction = c("none", "seno", "ryan", "chen", "joeke"))
2 {
3   correction <- match.arg(correction)
4   p <- mean(prop)
5   CL.shewhart <- p + c(-1, 1) * 3 * sqrt(p * (1 - p) / n)
6   CL.ryan <- CL.shewhart + 1.25 / n
7   CL.chen <- CL.shewhart + 4 / (3 * n) * (1 - 2 * p)
8   CL.joeke <- CL.chen - (p * (1 - p) + 2) / (6 * n ^ 2 * sqrt((p * (1 - p) / n)))
9
10  CL <- switch(correction,
11    none = CL.shewhart,
12    ryan = CL.ryan,
13    chen = CL.chen,
14    joeke = CL.joeke)
15
16  CL <- c(CL[1], p, CL[2])
17  return(CL)
18 }
```

Listing 2: Implementação do Beta Control Chart (BCC).

```
1 bcc <- function(y, alpha = 0.005)
2 {
3   fit <- fitdist(data = y, distr = "beta")
4   thetas <- coef(fit)
5   LCL <- qbeta(p = alpha / 2, shape1 = thetas[1], shape2 = thetas[2])
6   UCL <- qbeta(p = 1 - (alpha / 2), shape1 = thetas[1], shape2 = thetas[2])
7   CL <- as.vector(thetas[1] / (thetas[1] + thetas[2]))
8   limites <- c(LCL, CL, UCL)
9   return(limites)
10 }
```

Listing 3: Implementação do Regression Control Chart (RCC).

```
1 rcc <- function(model, data, alpha = 0.005)
2 {
3   n <- nrow(data)
4   fit <- lm(formula = model, data = data)
5   mut <- fit$fitted.values
6   k <- length(fit$coefficients)
7   sigma <- sqrt(1 / (n - k - 1) * sum(fit$residuals^2))
8   LCL <- qnorm(p = alpha / 2, mean = mut, sd = sigma)
9   UCL <- qnorm(p = 1 - alpha / 2, mean = mut, sd = sigma)
10  limites <- data.frame(LCL = LCL, UCL = UCL)
11  return(limites)
12 }
```

Listing 4: Implementação do Beta Regression Control Chart (BRCC).

```

1 brcc <- function(y, X, Z, alpha = 0.005, diagnostic = FALSE, npdf)
2 {
3   y <- as.vector(y)
4   X <- as.matrix(cbind(1, X))
5   Z <- as.matrix(cbind(1, Z))
6
7   n <- length(y)
8   r <- ncol(X)
9   s <- ncol(Z)
10
11  loglike <- function(parms, y)
12  {
13    beta <- parms[1:r]
14    gamma <- parms[(r + 1):(r + s)]
15    eta1 <- X %*% beta
16    eta2 <- Z %*% gamma
17    mu <- exp(eta1) / (1 + exp(eta1))
18    sigma <- exp(eta2) / (1 + exp(eta2))
19    aux1 <- (1 - sigma ^ 2) / sigma ^ 2
20    theta1 <- mu * aux1
21    theta2 <- (1 - mu) * aux1
22    ll <- sum(dbeta(x = y, shape1 = theta1, shape2 = theta2, log = TRUE))
23    return(ll)
24  }
25  ystar <- log(y / (1 - y))
26  betaols <- lm.fit(X, ystar)$coefficients
27  etaols <- X %*% betaols
28  muols <- exp(etaols) / (1 + exp(etaols))
29  varols <- as.numeric(t(ystar - etaols) %*% (ystar-etaols)) / ((n - r) * ((1 / (muols
30    * (1 - muols))) ^ 2))
31  phiols <- (muols * (1 - muols) / varols) - 1
32  sigmaini <- sqrt(1 / (1 + phiols))
33  gamaols <- lm.fit(Z, log(sigmaini / (1 - sigmaini)))$coefficients
34  ini <- c(betaols, gamaols)
35
36  opt <- optim(ini, loglike, method = "BFGS", control = list(fnscale = -1, maxit = 500,
37    reltol = 1e-9), hessian = TRUE, y = y)
38
39  if(opt$conv != 0) return(warning('FUNCTION DID NOT CONVERGE'))
40
41  beta <- opt$par[1:r]
42  gamma <- opt$par[(r + 1):(r + s)]
43  eta1 <- X %*% beta
44  eta2 <- Z %*% gamma
45  mu <- as.vector(exp(eta1) / (1 + exp(eta1)))
46  sigma <- as.vector(exp(eta2) / (1 + exp(eta2)))
47  aux1 <- (1 - sigma ^ 2) / sigma ^ 2
48  theta1 <- mu * aux1
49  theta2 <- (1 - mu) * aux1
50
51  LCL <- qbeta(p = alpha / 2, shape1 = theta1, shape2 = theta2)
52  UCL <- qbeta(p = 1 - (alpha / 2), shape1 = theta1, shape2 = theta2)
53  limites <- data.frame(LCL = LCL, UCL = UCL)
54  return(limites)
55 }
```