

5. Carta de controle e homogeneidade de variância

O desenvolvimento deste estudo faz menção a dois conceitos estatísticos: as cartas de controle, de amplo uso em controle estatístico de processo, e a homogeneidade de variâncias, que pode ser avaliada por um grande número de testes existentes. Este capítulo faz uma breve apresentação destes conceitos e de seus procedimentos de uso.

5.1. As cartas de controle

Segundo Montgomery (2001) carta de controle ou gráfico de controle é uma apresentação gráfica de uma característica da qualidade, que foi medida ou calculada a partir de uma amostra *versus* o número da amostra ou o tempo. É um tipo de gráfico (ASTM-c, 1976) comumente utilizado para o acompanhamento durante um processo. É construído baseado num histórico do processo e determina uma faixa chamada de tolerância limitada pela linha superior (limite superior de controle, LSC) e uma linha inferior (limite inferior de controle, LIC) e uma linha média do processo (limite central), que são estatisticamente determinadas. Já Moore *et. al.* (2006) afirma que os gráficos de controle são instrumentos estatísticos que monitoram um processo e alertam quando o mesmo foi perturbado, de modo a passar para um estado “fora de controle”. O uso de cartas de controle (ASTM-c, 1976) fornece critérios para a detecção de lacunas na execução do controle estatístico.

Montgomery (2001) diz que os gráficos de controle ajudam a determinar tanto se os dados do passado se originaram ou não de um processo que estava sob controle, quanto determinar se as amostras futuras desse processo indicam controle estatístico. Afirma, ainda, que tais gráficos possuem uma longa história de uso na indústria, devido a pelo menos cinco razões:

- a. são uma técnica comprovada para a melhoria da produtividade,
- b. são eficazes na prevenção de defeitos,
- c. evitam o ajuste desnecessário do processo,

- d. fornecem informação de diagnóstico, e
- e. fornecem informação sobre a capacidade do processo.

5.1.1. Gráficos de controle de Shewhart

Montgomery (2001) descreve o gráfico de controle de Shewhart como um modelo geral de gráfico de controle. Para ilustrar tal modelo, supõe uma estatística amostral w que meça algum parâmetro de qualidade de interesse. Supõe, também, que a média de w seja μ_w e o desvio padrão seja σ_w . Então, o limite central, o LSC e o LIC são dados por:

$$LSC = \mu_w + L\sigma_w \quad (5.1)$$

$$\text{Limite central} = \mu_w \quad (5.2)$$

$$LIC = \mu_w - L\sigma_w \quad (5.3)$$

sendo que L é a distância dos limites de controle ao limite central, expressa em unidades de desvio padrão.

Para Montgomery (2001) o gráfico de Shewhart é muito eficaz se a magnitude das mudanças que eventualmente ocorram for da ordem de uma vez e meia o desvio padrão ou mais. Infelizmente, para mudanças menores ele não é eficaz.

5.1.2. O gráfico de controle da soma cumulativa

Segundo Montgomery (2001) o gráfico de controle da soma cumulativa – cumsum – incorpora diretamente toda a informação na sequência dos valores da amostra, traçando as somas cumulativas dos desvios dos valores da amostra de um valor-alvo. A título de ilustração, sugere o seguinte exercício:

- a. coletar amostras de tamanho $n \geq 1$,
- b. supor que M_j seja a média da j -ésima amostra,
- c. atribuir o valor μ_0 ao alvo para a média do processo
- d. construir o gráfico de controle da soma cumulativa, traçando a quantidade dada por (5.4) em função da amostra i

$$C_i = \sum_{j=1}^i M_j - \mu_0 \quad (5.4)$$

C_i é a soma cumulativa até a i -ésima amostra, incluindo a própria.

Como combinam informações de várias amostras, os gráficos de somas cumulativas são mais eficazes que os gráficos de Shewhart para detectar pequenas mudanças no processo. Alves (2003) afirma (i) que o procedimento de soma acumulada proporciona um controle mais rigoroso do processo e (ii) que, se o processo permanecer sob controle para o valor médio desejado (μ_0), as somas acumuladas descrevem um percurso aleatório com média zero. Porém, se a média muda para um valor $\mu_1 > \mu_0$, uma tendência ascendente – ou positiva – se desenvolverá em C_i . Reciprocamente, se a média muda para algum valor $\mu_1 < \mu_0$, a soma acumulada C_i terá uma direção negativa. Assim, se nos pontos traçados aparecer uma tendência para cima ou para baixo, deve-se considerar isto como evidência de que a média do processo mudou.

5.2. Homogeneidade de variâncias

A homogeneidade de variâncias caracteriza os grupos de dados de forma a direcioná-los aos tratamentos estatísticos mais convenientes e, para muitos casos, é ponto de partida para se obter resultados estatisticamente válidos. Existem diversos testes de variância destinados à verificação da homogeneidade de dados. Para compreender seus princípios, consideram-se grupos de dados obtidos de um determinado estudo e formulam-se as seguintes hipóteses:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$$

$$H_1: \text{pelo menos um dos } \sigma^2\text{'s é diferente}$$

Onde σ_i^2 representa a variância de cada um dos grupos de dados. Os testes de variância verificam, através de cálculos pertinentes, qual das hipóteses acima é verdadeira. Comparam grupos grandes ou pequenos, de tamanhos iguais ou diferentes, e que obedeçam à distribuição normal ou não. Três testes são apresentados neste item: o teste de Cochran, o teste de Bartlett e o teste de Levene.

5.2.1.**O teste de Cochran**

Segundo Leão (2007), o teste de Cochran é usado para comparar a maior variância de um grupo com as outras deste mesmo grupo. Consiste em calcular todas as variâncias envolvidas em um estudo e dividir a maior delas pela soma de todas, conforme mostra a equação (5.5).

$$C = \frac{\text{maior variância}}{\text{soma de todas as variâncias}} \quad (5.5)$$

O resultado para C obtido em (5.5) é comparado com os valores críticos de uma tabela estatística, que leva em conta o número de variâncias envolvidas e o número de graus de liberdade. O teste de Cochran compara a homogeneidade das variâncias uma a uma, e só se aplica a dados com distribuição normal.

5.2.2.**O teste de Bartlett**

De acordo com Freitas *et. al.* (2003), um dos testes de variância mais utilizados é o teste de Bartlett. Na execução dos cálculos correspondentes ao teste, toma-se o seguinte procedimento:

Primeiro, calcula-se as k variâncias amostrais $s_1^2, s_2^2, s_3^2, \dots, s_k^2$, das amostras de tamanho $n_1, n_2, n_3, \dots, n_k$, com:

$$\sum_{i=1}^k n_i = N \quad (5.6)$$

Depois disto, combinam-se as variâncias amostrais para fornecer a estimativa:

$$s_p^2 = \frac{\sum_{i=1}^k (n_i - 1) s_i^2}{N - k} \quad (5.7)$$

Por fim, calcula-se:

$$b = \frac{(N - k) \ln s_p^2 - \sum_{i=1}^k (n_i - 1) \ln s_i^2}{1 + \frac{1}{3(k-1)} \left[\sum_{i=1}^k \frac{1}{n_i - 1} - \frac{1}{N - k} \right]} \quad (5.8)$$

O valor de b é corresponde ao valor da variável aleatória B que possui uma distribuição χ^2 com $k-1$ graus de liberdade. O denominador é grande quando as variâncias amostrais diferem significativamente, e igual a zero quando todas as variâncias amostrais são iguais. Assim, rejeita-se H_0 ao nível de significância α quando:

$$b \geq \chi_{\alpha, k-1}^2 \quad (5.9)$$

Almeida *et.al.* (2008) afirma que o teste de Bartlett para homogeneidade de variâncias é uma ferramenta eficiente somente se as variáveis possuem distribuição normal. Quando a suposição de normalidade é violada, o tamanho do teste pode ser muito maior do que o nível de significância fixado.

5.2.3. O teste de Levene

Segundo Mazucheli e Borges (2009), o teste conhecido na literatura como teste de Levene é usado para testar a igualdade de variâncias de populações normais univariadas. Segundo Almeida *et.al.* (2008), o teste de Levene é um procedimento relativamente insensível a desvios da normalidade. É um teste robusto porque na ausência de normalidade seu tamanho real é próximo do nível de significância fixado para uma grande variedade de distribuições de probabilidade.

Na execução dos cálculos correspondentes ao teste de Levene, toma-se o seguinte procedimento:

Seja Y_{ij} a j -ésima observação no i -ésimo grupo, onde:

$$i = 1, \dots, a, e$$

$$j = 1, \dots, n_i,$$

E o interesse em testar as seguintes hipóteses:

$$H_0 = \sigma_1^2 = \dots = \sigma_n^2$$

$$H_1 = \sigma_i^2 \neq \sigma_j^2 \text{ para pelo menos um par de } (i, j), (i \neq j)$$

A estatística do teste de Levene é definida na forma:

$$W = \frac{N-k}{k-1} \cdot \frac{\sum_{i=1}^k N_i (Z_{i.} - Z_{..})^2}{\sum_{i=1}^k \sum_{j=1}^{N_i} (Z_{ij} - Z_{i.})^2} \quad (5.10)$$

Nesta expressão tem-se que:

W = resultado do teste,

k = número de grupos aos quais os dados pertencem,

N = número total de dados,

N_i = número de dados no i -ésimo grupo,

$Z_{i.}$ = média de Z_{ij} para o grupo i

$Z_{.j}$ = média de todos Z_{ij}

$Z_{ij} = |Y_{ij} - \bar{Y}_{i.}|$ = módulo da subtração do j -ésimo dado no i -ésimo grupo pela média robusta do i -ésimo grupo

O valor de W é comparado com o valor de $F(\alpha, k-1, N-k)$, onde F é o valor encontrado no teste de distribuição F com $k-1$ e $N-k$ graus de liberdade, sendo α o nível de significância escolhido. Para os propósitos deste estudo, $\alpha = 0,01$.