

# Determinação do tamanho amostral para a validação de processo de produtos estéreis na indústria farmacêutica

Por Emerson Wruck e Fabio S. Julião

Este trabalho teve como objetivo propor uma metodologia para a determinação do tamanho amostral, que consiga inferir resultados para todo o lote inspecionado por amostragem, no processo de inspeção para a validação de processo de produtos estéreis na indústria farmacêutica, considerando o problema como um teste de hipóteses, sendo priorizada a especificação do erro tipo II. Foi feita a avaliação dos possíveis erros estatísticos envolvidos nos procedimentos farmacopéicos atuais de inspeção e foi sugerida uma adaptação desses procedimentos segundo a metodologia exposta no trabalho.

**Palavras-chave:** *amostragem de aceitação; teste de hipóteses; testes de esterilidade; validação de processos.*

This work aims to suggest a methodology to assign the size sampling which preview the overcome parameter for every sample checked by case. The procedure will apply on the inspection to validate the sterilized products process in pharmaceutical companies taking in account the problem as hypothesis test, privileging the error specification type II. The statistical errors were evaluated through the procedure of the pharmacopeia on the inspection, and it was suggested one small change on procedures as a result of the methodology proposed.

**Keywords:** *sampling acceptance; hypothesis tests; sterility tests; validation of processes.*

## Introdução

Para o teste de segurança biológica de produtos estéreis são estabelecidos procedimentos de inspeção, com base no sorteio aleatório de uma amostra, sendo esta utilizada nas análises microbiológicas (Método de Inoculação Direta ou Filtração em membrana) que vão determinar se há contaminação<sup>1</sup>. Se a mesma estiver contaminada, o lote é rejeitado e o processo é não validado. Apesar desse procedimento estar baseado em uma amostra, não é especificado qual o nível de significância estatística envolvido no problema. A Farmacopéia brasileira estabelece um tamanho padrão para esta amostragem que não garante a extrapolação dos resultados para todo o lote, mas apenas para a amostra inspecionada<sup>1</sup>.

Os níveis para os erros tipo I e tipo II, envolvidos no processo, são desconhecidos. Não é estabelecido um nível de qualidade aceitável *a priori* condizente com o processo de amostragem e, dessa forma, esse procedimento se torna objeto de grande discussão entre profissionais da indústria farmacêutica e técnicos da Agência Nacional de Vigilância Sanitária (ANVISA), pois a falta dessas especificações gera desconfiância quanto à validação do processo, uma vez que os resultados são garantidos apenas para a amostra inspecionada.

Esse trabalho teve como objetivo avaliar os possíveis erros envolvidos nesse processo de inspeção, considerando o problema como um teste estatístico de hipótese<sup>2</sup> e a partir desse contexto, estabeleceu-se uma proposta para a determinação do tamanho da amostra, baseado na pré-definição de parâmetros como o erro tipo II e na escolha adequada dos níveis de qualidade do lote<sup>3</sup>, por parte da indústria, que seja coerente com as necessidades da validação do processo de produção de produtos estéreis, garantindo os resultados, para todo o lote de onde é retirada a amostra.

## Materiais e Métodos

O procedimento utilizado na determinação do tamanho amostral consistiu em associar os conceitos do teste de hipótese e amostragem de aceitação com o intuito de propor um melhor suporte teórico para o problema e facilitar seu entendimento.

## TESTE DE HIPÓTESES

Em muitas situações temos interesse em tomar a decisão de aceitar ou rejeitar determinada afirmação baseados em um conjunto de evidências<sup>2</sup>. No caso específico do problema da validação de processo de produtos estéreis, o lote deverá ser aceito ou não baseado nas evidências geradas pela amostra retirada desse lote. Neste caso, pode-se formular duas hipóteses: a hipótese nula, definida por  $H_0$ : “O lote está conforme” e a alternativa  $H_1$ : “O lote não está conforme”. Com base nas evidências apresentadas na amostra, deve-se decidir entre rejeitar  $H_0$  (Aceitar  $H_1$ ), ou seja, rejeitar a condição de que o lote está conforme em favor da que o lote não está conforme, ou não rejeitar  $H_0$ , ou seja, aceitar a condição do lote conforme.

Ao tomar qualquer uma dessas decisões, existe a possibilidade de erros, pois o procedimento está baseado em amostragem<sup>4</sup>. O erro tipo I, é o cometido quando rejeitamos a hipótese nula quando na verdade ela é verdadeira, ou seja,

$$\text{Erro Tipo I} = \alpha = P [ \text{Rejeitar } H_0 \mid H_0 \text{ é Verdadeira} ]$$

e o erro tipo II é cometido quando não rejeitamos  $H_0$ , quando  $H_0$  é falsa, ou seja,

$$\text{Erro Tipo II} = \beta = P [ \text{Não Rejeitar } H_0 \mid H_0 \text{ é Falsa} ]$$

A relação dos possíveis resultados do teste é ilustrada na Tabela 1:

**Tabela 1:** Resultados de um teste de hipóteses

Decisão	Situação	
	$H_0$ é Verdadeira	$H_0$ é Falsa
Rejeitar $H_0$	Erro Tipo I	Sem Erro
Não Rejeitar $H_0$	Sem Erro	Erro Tipo II

O teste de hipóteses ideal é aquele que minimiza de forma simultânea os dois tipos de erros. No entanto esse teste não existe, ambos dependem da regra de decisão adotada, de forma que a diminuição de um implica no aumento do outro. No geral, o pesquisador fixa um tipo de erro para a formulação da regra de decisão e o outro erro é apenas monitorado<sup>4</sup>. Normalmente é fixado o erro tipo I ( $\alpha$ ) e é construída a regra de decisão baseada nesse nível.

Controlar diretamente o erro tipo II fica difícil, dada a falta de especificação que gera a condição de  $H_0$  ser falsa (qualquer condição que faça o lote não conforme). Utiliza-se para tal a função poder do teste definida por  $1 - \beta$ , ou seja,

$$1 - \beta = P [ \text{Rejeitar } H_0 \mid H_0 \text{ é Falsa} ]$$

Neste caso, quanto maior for o valor da *função poder*, quando  $H_0$  é falsa, maior a eficiência do teste. Esta função pode ser generalizada como a probabilidade de rejeitar  $H_0$ .

Outra função usada para avaliação do teste, é a curva característica de operação, definida por  $1 - \alpha$ , ou seja,

$$1 - \alpha = P [ \text{Não Rejeitar } H_0 \mid H_0 \text{ é Verdadeira} ]$$

Também, neste caso, esta função pode ser generalizada como a probabilidade de aceitar  $H_0$ . Em geral, valores grandes obtidos da função quando  $H_0$  é verdadeira, sinaliza bom teste.

Dessa forma, procura-se buscar um procedimento de inspeção que apresente uma probabilidade alta de aceitação para níveis aceitáveis de fração de defeituosos no lote e uma probabilidade alta de rejeição (baixa aceitação) para níveis não aceitáveis de fração de defeituosos no lote.

## AMOSTRAGEM DE ACEITAÇÃO

Para os procedimentos de inspeção, normalmente a amostragem de aceitação<sup>3</sup> é empregada. Uma classe particular dos planos de aceitação que se enquadra no problema é o plano de amostragem única para atributos. Neste, seleciona-se aleatoriamente uma amostra de  $n$  unidades do lote e defini-se um número de aceitação  $c$ . Se há  $c$  ou menos defeituosos na amostra, aceita-se o lote, e se há mais de  $c$  itens defeituosos na amostra, rejeita-se o lote<sup>5</sup>.

Uma medida importante do desempenho de um plano de amostragem de aceitação é a curva característica de operação (CO) ou  $1 - \alpha$  a mesma usada para avaliar a qualidade de um teste de hipótese, só que aqui pensada para qualquer fração de defeituosos. No contexto da amostragem de aceitação, a curva plota a probabilidade de aceitação do lote *versus* a fração de defeituosos no

lote<sup>6</sup>. Aqui, a curva mede o poder discriminatório do plano amostral. Isto é, determina a probabilidade de que um lote, com certa fração de defeituosos, seja aceito. O complemento da curva CO é a probabilidade de rejeitar o lote para certos níveis de frações de defeituosos no lote, que pode ser vista como a função poder  $1 - \beta$  só que avaliada para qualquer nível de fração defeituosa.

Supondo o tamanho do lote ( $N$ ) grande, onde a relação entre tamanho da amostra e o tamanho do lote seja menor

que 0.1, ou seja,  $\frac{n}{N} < 0.1$  a distribuição do número de

defeitos  $X$  em uma amostra aleatória  $n$  de itens pode ser bem aproximada pelo modelo binomial com parâmetros  $n$  e  $p$ , onde  $p$  é a fração de itens defeituosos no lote<sup>7</sup>.

A probabilidade de aceitação é, a probabilidade do número de defeituosos ser menor ou igual a  $c$ , ou seja,

$$P_a = P[X \leq c] = \sum_{x=0}^c \binom{n}{x} p^x (1-p)^{n-x} \quad (1)$$

Dessa forma, a curva característica (CO) é construída para cada valor de  $c$ , variando o valor de  $p$ .

Dois pontos específicos da curva são o Nível de Qualidade Aceitável ou (NQA) e o Percentual Aceitável de Defeituosos do Lote ou (PADL)<sup>8</sup>. O Nível de Qualidade Aceitável representa o nível de qualidade aceitável pelo consumidor como média do processo, ou seja, o consumidor geralmente planejará o procedimento de amostragem de modo que a curva CO dê uma alta probabilidade de aceitação no NQA. O percentual aceitável de defeituosos do Lote é o mais baixo nível de qualidade que o consumidor está disposto a aceitar em um lote individual. Neste caso, o consumidor deverá estabelecer um plano que lhe dê proteção para lotes que tenham este nível de qualidade, ou seja, planos com baixa probabilidade de aceitação de lotes no PADL<sup>3</sup>.

Uma abordagem para construção do plano de inspeção, é exigir que a curva CO passe por dois pontos designados, ou seja, atenda às especificações para o erro tipo I, tipo II e níveis de qualidade. Suponha que se deseje construir um plano amostral de inspeção tal que a probabilidade de aceitação do lote a um nível de fração defeituosa  $p_1$  (que pode ser pensada como NQA) seja  $1 - \alpha$ , e que a

probabilidade de aceitação do lote a um nível  $p_2$  (que pode ser pensada com PADL) seja  $\beta$ . Supondo que o modelo Binomial seja apropriado, então este plano deverá ter  $n$  e  $c$  que satisfaçam:

$$1 - \alpha = \sum_{x=0}^c \binom{n}{x} p_1^x (1 - p_1)^{n-x}$$

$$\beta = \sum_{x=0}^c \binom{n}{x} p_2^x (1 - p_2)^{n-x} \quad (2)$$

No caso específico do problema da validação de processo de produtos estéreis através do teste de esterilidade, este pode ser visto como um problema de aceitação com  $c = 0$ , pois se pelo menos uma unidade estiver contaminada na amostra, o teste detecta a contaminação, e o lote é rejeitado, ou seja, o processo não é validado. Ainda, pode-se associar o problema a um teste de hipóteses, considerando  $\beta$  o erro tipo II, probabilidade de aceitar o lote a um nível não satisfatório (PADL), e  $\alpha$  o erro tipo I, probabilidade de não aceitar o lote a um nível satisfatório (NQA), assim considerando  $p_1$  como o NQA e  $p_2$  como o PADL, as Equações (2) são reescritas como:

$$1 - \alpha = (1 - NQA)^n$$

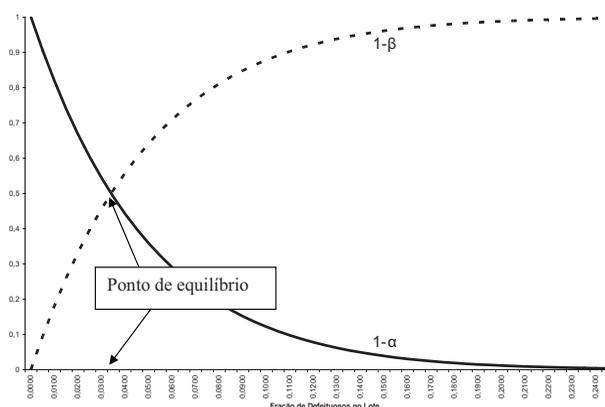
$$\beta = (1 - PADL)^n \quad (3)$$

Da Equação (3) podemos encontrar o tamanho amostral, que leva em consideração os erros tipo I e II, e os níveis de especificação NQA e PADL.

Para este tipo de problema, deve-se notar que o pior erro será o erro tipo II, pois se um lote com nível de qualidade não satisfatório for aceito, as conseqüências envolvem risco sanitário. Dessa maneira a abordagem do problema deve ser estabelecida controlando o erro tipo II ( $\beta$ ), e a especificação do PADL, que é de segurança para o consumidor final, e observando as implicações no erro tipo I e no NQA.

## Resultados e Discussão

O procedimento atualmente adotado consiste, de forma simplificada, em selecionar uma amostra aleatória de unidades de tamanho 20 (vinte), e proceder a análise microbiológica – se ocorrer contaminação, rejeita-se o lote. Como a contaminação é observada se pelo menos uma unidade estiver contaminada na amostra o modelo de inspeção de aceitação é em único estágio com  $c = 0$ . A curva característica de operação e a função poder para este procedimento é ilustrada na Figura 1.



**Figura 1:** Curva característica de operação ( $1 - \alpha$ ) e função poder ( $1 - \beta$ ) para o procedimento com  $n = 20$

Para um tamanho amostral de 20 unidades, a curva característica de operação se mantém em um nível alto de aceitação para aumentos da fração de defeituosos no lote, fazendo com que o erro tipo II se mantenha alto. Perceba que considerando o PADL como uma fração de defeituosos de 0.01, a probabilidade de aceitar esse lote (erro tipo II) é de 0.8179 e o poder do teste é de 0.1821, ou seja, isso implica que para um lote com 1% de unidades contaminadas (1 a cada cem), a chance desse lote ser aceito, quando se usa uma amostra de tamanho 20 para inspeção, é de 81.79%. Perceba ainda que o ponto de equilíbrio para as funções poder e CO, é uma fração de defeituosos de 0.034, ou seja, somente para frações de defeituosos acima desse valor é que se rejeitaria lotes com maior chance que aceitaria. Com isto, pode-se concluir que o procedimento está muito propenso ao erro tipo II, ou seja, está muito propensa a aceitar lotes com níveis de qualidade baixa.

Considerando um tamanho amostral de 40 unidades, que seria a maior amostra de inspeção sugerida pela farmacopéia, temos a seguinte curva CO ilustrada na Figura 2. Perceba que o poder discriminatório do procedimento aumentou com  $n = 40$ , aqui o ponto de equilíbrio é de 0.017 e percebe ainda que para o mesmo patamar do PADL (0.01) a probabilidade de aceitar o lote cai para 0.669, ou seja o erro tipo II seria de 0.669. Apesar de cair, para este nível de PADL, tem-se ainda um erro tipo II muito alto, ou seja, o procedimento se mostra não adequado.

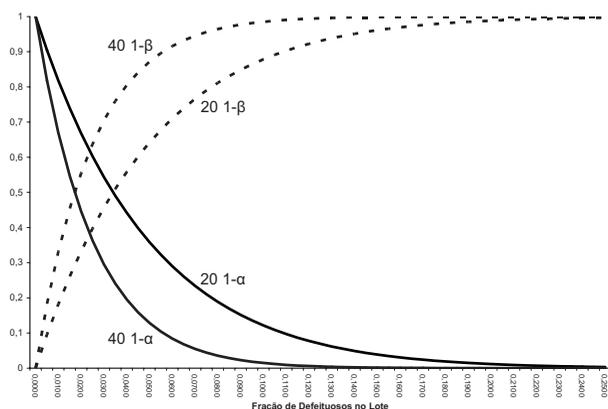


Figura 2: Curva característica de operação  $(1 - \alpha)$  e função poder  $(1 - \beta)$  para o procedimento com  $n = 20$  e  $n = 40$

A sugestão é desenvolver um plano amostral onde os erros tipo II e o PADL sejam controlados, ou seja, estabelecidos *a priori*, e a partir desses, considerando as Equações (3), encontrar um tamanho compatível com um NQA e  $\alpha$  aceitáveis.

Assim, das Equações (3) têm-se que,

$$n = \frac{\ln(\beta)}{\ln(1 - PADL)} \quad (4)$$

Encontrado o valor  $n$  da Equação (4), estabelecer um nível para o erro tipo I aceitável, e avaliar o NQA pela relação,

$$NQA = 1 - e^{\left( \frac{\ln(1-\alpha) \ln(1-PADL)}{\ln(\beta)} \right)} \quad (5)$$

Também, para avaliarmos um tamanho  $n$  especificado *a priori*, os demais parâmetros podem ser avaliados de forma recursiva, sempre tomando como base as Equações (3).

Por exemplo, considerando um PADL = 0.01 (que por sinal é muito grande, para procedimentos que envolvem risco sanitário) e um nível de erro tipo II ( $\beta$ ) de 0.05, ou seja, considerando a chance de aceitar um lote com 1% de fração de defeituosos que seja igual a 0.05 (5%), o tamanho amostral, necessário para garantir esta condição, será calculado pela Equação (4) como:

$$n = \frac{\ln(0.05)}{\ln(1 - 0.01)} \cong 299$$

Considerando um nível para o erro tipo I, ( $\alpha = 0.05$ ), o NQA é calculado da Equação (5) como:

$$NQA = 1 - e^{\left( \frac{\ln(1-0.05) \ln(1-0.01)}{\ln(0.05)} \right)} \cong 0.00017068$$

Assim, usando um tamanho amostral de 299, a probabilidade de aceitar um o lote com 1% de fração de defeituosos será de 0.05 (erro tipo II) e rejeitar um com 0.0017% de fração de defeituosos será de 0.05 (erro tipo I). Essa situação é ilustrada na Figura 3.

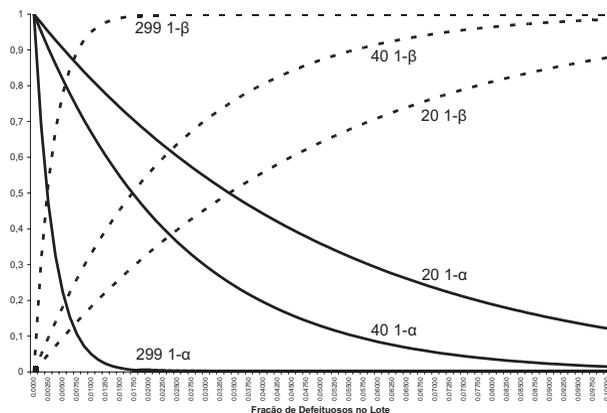


Figura 3: Curva característica de operação  $(1 - \alpha)$  e função poder  $(1 - \beta)$  para o procedimento com  $n = 20$ ,  $n = 40$  e  $n = 299$

Perceba que para  $n = 299$ , o poder discriminatório do plano amostral aumentou de forma considerável, e o ponto de equilíbrio neste caso é uma fração de defeituosos de aproximadamente 0.0025, sendo que estes resultados são garantidos para todo o lote.

Considerando alguns níveis para o PADL e fixados os erros tipo II ( $\beta$ ) e o tipo I ( $\alpha$ ) e utilizando as Equações (4) e (5) foram construídas tabelas indicativas para o tamanho amostral e o respectivo nível NQA, Tabela 2.

## Conclusões

Pela Tabela 2 e a Figura 3, percebe-se uma insuficiência no tamanho da amostra para os procedimentos adotados atualmente, quando aplicados à validação de processos, desejando inferir os resultados para todo o lote produzido, principalmente quando se considera um nível PADL compatível com a problemática envolvida a questões relacionadas ao risco sanitário. Dessa forma sugere-se adequar os procedimentos farmacopéicos existentes,

**Tabela 2:** Tamanhos Amostrais e NQA para níveis PADL especificados a priori.

$\beta$	=	0,05	$\alpha$	=	0,01
Defeitos por Lote (PADL)	PADL	n	NQA		
1 por 1000000	0,000001	2995731	0,0000000034		
1 por 500000	0,000002	1497865	0,0000000067		
1 por 100000	0,00001	299572	0,0000000335		
1 por 10000	0,0001	29956	0,0000003355		
1 por 5000	0,0002	14978	0,0000006710		
1 por 1000	0,001	2995	0,0000033566		
1 por 500	0,002	1497	0,0000067165		
1 por 100	0,01	299	0,0000337171		

$\beta$	=	0,1	$\alpha$	=	0,01
Defeitos por Lote (PADL)	PADL	n	NQA		
1 por 1000000	0,000001	2302584	0,0000000044		
1 por 500000	0,000002	1151292	0,0000000087		
1 por 100000	0,00001	230258	0,0000000436		
1 por 10000	0,0001	23025	0,0000004365		
1 por 5000	0,0002	11512	0,0000008730		
1 por 1000	0,001	2302	0,0000043670		
1 por 500	0,002	1151	0,0000087383		
1 por 100	0,01	230	0,0000438668		

$\beta$	=	0,05	$\alpha$	=	0,05
Defeitos por Lote (PADL)	PADL	n	NQA		
1 por 1000000	0,000001	2995731	0,0000000171		
1 por 500000	0,000002	1497865	0,0000000342		
1 por 100000	0,00001	299572	0,0000001712		
1 por 10000	0,0001	29956	0,0000017123		
1 por 5000	0,0002	14978	0,0000034248		
1 por 1000	0,001	2995	0,0000171305		
1 por 500	0,002	1497	0,0000342779		
1 por 100	0,01	299	0,0001720683		

$\beta$	=	0,1	$\alpha$	=	0,05
Defeitos por Lote (PADL)	PADL	n	NQA		
1 por 1000000	0,000001	2302584	0,0000000223		
1 por 500000	0,000002	1151292	0,0000000446		
1 por 100000	0,00001	230258	0,0000002228		
1 por 10000	0,0001	23025	0,0000022277		
1 por 5000	0,0002	11512	0,0000044557		
1 por 1000	0,001	2302	0,0000222873		
1 por 500	0,002	1151	0,0000445964		
1 por 100	0,01	230	0,0002238602		

$\beta$	=	0,05	$\alpha$	=	0,1
Defeitos por Lote (PADL)	PADL	n	NQA		
1 por 1000000	0,000001	2995731	0,0000000352		
1 por 500000	0,000002	1497865	0,0000000703		
1 por 100000	0,00001	299572	0,0000003517		
1 por 10000	0,0001	29956	0,0000035172		
1 por 5000	0,0002	14978	0,0000070347		
1 por 1000	0,001	2995	0,0000351872		
1 por 500	0,002	1497	0,0000704084		
1 por 100	0,01	299	0,0003534099		

$\beta$	=	0,1	$\alpha$	=	0,1
Defeitos por Lote (PADL)	PADL	n	NQA		
1 por 1000000	0,000001	2302584	0,0000000458		
1 por 500000	0,000002	1151292	0,0000000915		
1 por 100000	0,00001	230258	0,0000004576		
1 por 10000	0,0001	23025	0,0000045760		
1 por 5000	0,0002	11512	0,0000091524		
1 por 1000	0,001	2302	0,0000457793		
1 por 500	0,002	1151	0,0000916024		
1 por 100	0,01	230	0,0004597724		

aumentando o tamanho da amostra para a inspeção, utilizando a metodologia proposta, no sentido de melhorar a eficácia do procedimento.

Uma nova discussão deve ser feita quanto ao estabelecimento dos níveis PADL, erro tipo II ( $\beta$ ), erro tipo I ( $\alpha$ ) e o NQA. Deve-se levar em consideração que os erros podem prejudicar os consumidores finais, quanto se aceita lotes com PADL altos, ou seja, se valida um processo não eficiente. E para a indústria, quando rejeitamos lotes bons, ou seja, não se valida processos eficientes. Nesse sentido, trabalhos voltados para a estimação da fração de defeituosos produzidos, devem ser propostos, auxiliando dessa forma no estabelecimento desses índices, e por consequência no aprimoramento dos processos de controle de qualidade existentes.

## REFERÊNCIAS

1. Brasil. Comissão Permanente de Revisão da Farmacopéia (Org.). *Farmacopéia Brasileira: Parte I*. 4. ed. Brasília: Atheneu, 1988.
2. Bolfarine, H. E Sandoval, C. *Introdução à Inferência Estatística*. Rio de Janeiro, IMPA. Coleção Matemática Aplicada, 2001.
3. Montgomery, D.C. *Introdução ao Controle Estatístico de Qualidade*, 4º edição. Rio de Janeiro, RJ, LTC, 2004.
4. Bussab, W.B. E Moretin, P.A. *Estatística Básica*. São Paulo, Editora Atual, 1987
5. Hald, A. *Statistics Theory of Sampling Inspection by Attributes*. New York, Academic Press, 1981.
6. DUNCAN, A.J. *Quality Control and Industrial Statistics*. 5th edition, Chicago, Richard D. Irwin, Inc, 1986.
7. Johnson, N.L.; Kotz, S. & Balakrishnan, N. *Univariate Discrete Distributions*. 2 ed., New York, John Wiley & Sons, Inc (1992).
8. Costa, A.F.B.; Epprecht, E.K.; Carpinetti, L.C.R. *Controle Estatístico de Qualidade*. São Paulo: Atlas, 2004.

---

Emerson Wruck<sup>1,2\*</sup>, Fabio S. Julião<sup>2</sup>

<sup>1</sup> Unidade Universitária de Ciências Exatas e Tecnológicas, UEG, CP 459, CEP 75001-970, Anápolis, GO, Brasil

<sup>2</sup> Faculdade de Tecnologia SENAI Roberto Mange, CEP 75113-630, Anápolis, GO Brasil.

\* E-mail: wruck.senai@sistemafieg.org.br