

Amostragem por Conglomerados em Três Estágios para Estimação de Itens Defeituosos através de Métodos de Pesquisa Quantitativa

Helinton A. L. Barbosa¹; Fernando L. P. de Oliveira².

¹Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627 - CEP: 31270-901 - Belo Horizonte/ M.G.

²Universidade Federal de Ouro Preto, Departamento de Matemática - Campus Morro do Cruzeiro - Ouro Preto/MG

Resumo — Métodos de pesquisa quantitativa, realizados através de técnicas de amostragem probabilística, são ferramentas poderosas em processos de tomadas de decisão, sejam estas técnicas ou estratégicas. Neste estudo, através de um processo amostral com três estágios, são apresentados métodos de coleta dados para cálculo, construção e análise de estimativas do número de itens defeituosos em um conjunto de lotes de um processo de produção. Essa investigação visa orientar a tomada de decisão sobre a qualidade dos produtos ofertados. Depois de calculadas estimativas pontuais, construíram-se intervalos de confiança para que, assim, fosse decidida a recusa ou não dos itens. Para tal, a análise intervalar deveria indicar, em seu limite inferior de confiança, se o número de produtos defeituosos ultrapassava 3% do total de unidades. Os resultados obtidos na implementação da amostragem foram animadores e as estimativas intervalares nortearam estatisticamente qual decisão deveria ser tomada.

Palavras-Chave — Amostragem por Conglomerados, Coeficiente de Correlação Intraclasse, Pesquisa Quantitativa.

I. INTRODUÇÃO

O controle estatístico de qualidade tem como base a coleta de informações através de amostras aleatórias selecionadas de um processo ao longo de sua produção. Sabe-se que inspecionar toda linha de montagem ou fabricação tem um alto custo, além de consumir um tempo que, muitas vezes, não está disponível ou não é viável logisticamente. Em contrapartida, podem-se utilizar ferramentas para seleção de amostras do processo, como por exemplo, usando técnicas de amostragem de aceitação [1]. Neste artigo relata-se o uso de procedimentos de amostragem probabilística [2] usados para coletar dados de um processo, embasando, através das inferências obtidas, uma tomada decisão.

A preocupação das instituições de diversas áreas do conhecimento com a qualidade de produtos, consumidos ou produzidos, vem tendo destaque desde os primórdios da era industrial. Qualidade é um fator de importância quando se deseja adquirir produtos ou serviços que atendam as exigências do mercado. Como dito em [3]: "O fenômeno é geral, independente do fato do consumidor ser um indivíduo, uma organização industrial, uma loja de varejo, ou um programa militar de defesa". Isto é, a manutenção da qualidade é interessante a todos. Baseado nesse referencial, um órgão estatal

precisa identificar, em produtos de um fornecedor credenciado, falhas primárias, que podem comprometer todos os processos relacionados à aplicabilidade das unidades adquiridas. A motivação dessa investigação se deve a diversos registros anteriores, citando problemas relacionados a falhas de produtos novos por de defeitos de fabricação. Desejando resolver quaisquer pendências relacionadas à qualidade dos itens e validar ou não o contrato com o fornecedor, buscou-se, através de técnicas de amostragem probabilística em estudos quantitativos, em um processo amostral com três estágios, coletar e disponibilizar dados para construção e análise de estimativas do número de itens defeituosos.

Depois de coletadas informações e calculadas estimativas pontuais, construíram-se intervalos de confiança, para que assim, pudesse ser decidido pela recusa ou não dos produtos do fornecedor. Caso a análise intervalar indicasse, em seu limite inferior de confiança, que o número de produtos defeituosos ultrapassa 3% do total de itens, os lotes seriam devolvidos, o contrato cancelado e o fornecedor descredenciado.

De acordo com o sistema de logística definido para o órgão, cada lote é enviado a diferentes unidades subordinadas. Estes lotes possuem um número distinto de produtos, pois há em cada subunidade uma demanda específica em quantidades de itens. Esta característica singular foi considerada pelo plano amostral, promovendo probabilidades de seleção relacionadas às dimensões de cada lote.

Neste trabalho apresentam-se na Seção 2 todos os procedimentos amostrais utilizados, exemplificando e detalhando sua aplicação. A Seção 3 mostra os resultados obtidos pela coleta de dados, além das inferências intervalares visando aqui, à tomada de decisões. Por fim, na Seção 4, são apresentadas as conclusões e considerações finais do estudo.

II. POPULAÇÃO ALVO, TAMANHO AMOSTRAL E METODOLOGIA

Informações e bases iniciais

A população-alvo de estudo são todos os produtos disponibilizados pelo fornecedor através de 242 lotes

distintos, sendo estes divididos 68.669 caixas, totalizando 154.109 unidades possíveis de serem amostradas. O tamanho amostral total pesquisado foi de 800 itens, sendo que em cada uma das três fases do processo de coleta, houve um tamanho amostral associado. Ou seja, no estágio 1 foram amostrados 100 lotes de produtos, no segundo estágio da amostragem coletaram-se 8 caixas de produtos em cada lote e, no estágio 3, foi selecionada aleatoriamente uma unidade do produto por caixa. A seguir são dados detalhamentos sobre estes totais amostrais, além de outras informações técnicas.

Definições amostrais

A Teoria da Amostragem, área da Estatística responsável por estudos de pesquisa e sondagens em vários segmentos da sociedade, foi a referência usada para o delineamento do plano amostral probabilístico, que definiu quais foram os lotes, caixas e produtos amostrados. De forma ampla, o plano desenvolvido abrangeu três estágios:

- Estágio 1 – Nessa etapa foram identificados e selecionados os lotes que entraram na amostra;
- Estágio 2 – A segunda fase amostral selecionou aleatoriamente caixas dentro de cada lote sorteado na etapa 1;
- Estágio 3 – A parte final da amostragem foi definida como momento da escolha do produto na caixa. Cada unidade tinha uma probabilidade de seleção aleatória e diferente de zero.

Formalmente, o procedimento amostral usado na pesquisa foi a Amostragem por Conglomerados em três estágios [4], sendo realizada a Amostragem PPT-Sistemática¹ [5] no estágio 1; a Amostragem Sistemática no estágio 2 [6] e, no estágio 3, a Amostragem Aleatória Simples Sem Reposição [7].

Os processos de seleção nas etapas 1, 2 e 3 e a utilização das técnicas ou procedimentos amostrais foram realizados em *software* com módulo amostral específico² [8], evitando manipulações e minimizando a possibilidade de vícios na seleção das unidades amostrais e estimativas inferenciais.

Processo de amostragem e coleta de dados

1) *Introdução*: A mais recente remessa de produtos do fornecedor credenciado possuía exatos 242 lotes³ distintos. Desse total, foi preciso selecionar o percentual de alguns para a amostra. A literatura orienta que a quantidade de lotes selecionados depende do número de amostras coletadas em cada um deles. Desta forma, como o total amostral previsto para a pesquisa era de $n=800$ itens e sendo realizadas $k=8$ coletas por lote (sugere-se a partir de 5 amostras [5]), o número de lotes selecionados, calculado por (1), foi $m=100$, o que equivale a uma cobertura aproximada de 41,3% do total, ou seja:

$$m = \frac{n}{k} = \frac{800}{8} \cong 100 \quad (1)$$

Isso significa que, em números aproximados, 2 a cada 5 lotes, foram selecionados para o estudo.

2) *Estágio 1- Método PPT-Sistemático*: O processo de seleção de lotes levou em consideração o tamanho, em quantidade de caixas, de cada um. Isto é, um lote com um número de caixas superior a outro tem uma probabilidade de seleção maior. Esse esquema de amostral é chamado Amostragem PPT que, combinado a Amostragem Sistemática, selecionou os lotes para o estudo. Maiores detalhes sobre esse esquema podem ser consultados em [6].

O método de seleção utilizado no primeiro estágio do processo amostral foi então a amostragem PPT-Sistemática [9], em que se dividiu a população em intervalos de amostragem, para então se retirar uma amostra, considerando o tamanho de cada elemento na população. O intervalo de amostragem neste estágio foi determinado por (2):

$$I_a = \frac{N}{x} \quad (2)$$

Onde:

N = Total populacional de interesse;
 x = tamanho da amostra de interesse.

Segundo informações repassadas pelos interessados, o fornecedor disponibilizou ao todo 68.669 caixas de produtos, alocados em 242 lotes distintos, dos quais se pretende amostrar⁴ 100. O intervalo de amostragem deste estágio foi calculado (2) da seguinte maneira:

$$I_a = \frac{N}{x} = \frac{68669}{100} \cong 686$$

Para facilitar os cálculos e o entendimento da metodologia, o intervalo sistemático foi aproximado para 686. Em seguida, escolheu-se um número aleatório inicial, ou seja, um valor entre 1 e 686. A Tabela I demonstra o processo

⁴ O valor da medida máxima de tamanho (MOS) em todas as unidades da população não pode ser maior do que o valor MOS na população total dividido pelo tamanho da amostra. Exemplo:

Número máximo de MOS: Lote 72105204 com 683 caixas. O fornecedor disponibilizou ao todo 68.669 caixas (alocados em 242 lotes), ou seja, o tamanho amostral pode ser calculado pela razão: $683 \leq \frac{68669}{x}$, isto é,

$x \leq \frac{68669}{683}$, logo $x \leq 100,54$, em que x é o tamanho máximo permitido para a amostra.

Suponha um tamanho amostral $x=101$, ou seja, um valor que excede o tamanho amostral calculado anteriormente. Com esse número a amostragem não será realizada, pois não há como calcular o intervalo sistemático ponderado necessário ao método.

Imagine agora um tamanho amostral $x=100$, isto é, um valor que atende ao tamanho inicial estipulado, pois $100 \leq 100,54$. Com este tamanho amostral a fórmula é validada e o processo amostral pode acontecer, pois há como calcular o intervalo sistemático [8].

¹ Probabilidade Proporcional ao Tamanho-Sistemática.

² SPSS Complex Samples.

³ Cada lote possui quantidades de produtos diferenciadas, sendo cada um deles encaminhado a diferentes unidades.

de Amostragem Sistemática na Amostragem PPT [10], partindo da geração do número aleatório 1.

TABELA I - PROCESSO DE AMOSTRAGEM PPT-SISTEMÁTICA

Identificação do lote	Quantidade de caixas	Total acumulado de caixas	Número aleatório (entre 1 e 686)	Intervalo (não varia até que haja uma seleção)	Amostra selecionada (se o intervalo pertence ao acumulado)
72105001	335	335	1	1	sim (335>1)
72105002	350	685		1+686=687	não (685<687)
72105003	371	1056		687	sim (1056>687)
72005004	374	1430		687+686=1373	sim (1430>1373)
72005005	193	1623		1373+686=2059	não (1623<2059)
72105006	389	2012		2059	não (2012<2059)
72105007	142	2154		2059	sim (2154>2059)
72105008	300	2454		2059+686=2745	não (2454<2745)
72105009	408	2862		2745	sim (2862>2745)
72105010	108	2970		2745+686=3431	não (2970<3431)
72105011	360	3330		3431	não (3330<3431)
72105012	374	3704		3431	sim (3704>3431)
72105013	182	3886		3431+686=4117	não (3886<4117)
72105014	275	4161		4117	sim (4161>4117)
72105015	369	4530		1417+686=4803	não (4530<4803)
.
.
.
721015007	237	68669	.	.	.

A coluna de Identificação do Lote correspondente ao código do lote, ordenada de forma crescente. Na coluna Quantidade de Caixas, tem-se o quantitativo de caixas em cada lote, sendo a coluna Total Acumulado a soma de cada lote com o anterior em número de caixas acumuladas. O número aleatório gerado neste exemplo (entre 1 e 686) foi 1, este representado na Tabela I pela coluna Número Aleatório, valor que foi a referência para a seleção dos lotes na coluna seguinte (Intervalo). Caso o intervalo estipulado seja menor que a quantidade de caixas acumuladas em cada lote, seleciona-se este lote que contém o valor acumulado, do contrário, passa-se a um novo acumulado até que a condição seja satisfeita. Após essa fase, um novo intervalo é calculado considerando o valor do intervalo anterior somado com 686. Perceba que a seleção do lote está ligada a quantidade de caixas que este possui, sendo que, quanto maior esse número, maior sua probabilidade de ser amostrado.

O cálculo do intervalo sistemático, em função do número total de caixas e a indexação dos lotes pela quantidade de caixas, permite que o método sistemático seja ponderado e faz com que a seleção amostral observe as características dimensionais de cada lote, o que não seria possível se a Amostragem Sistemática Simples fosse utilizada. Mais detalhes sobre a utilização do procedimento PPT-Sistemático podem ser consultados em [11] e [12].

3) *Estágio 2 - Amostragem Sistemática:* No segundo estágio amostral, foi necessário trabalhar o processo de amostragem nos lotes selecionados e, cada um deles, possuía um número distinto de caixas. O total de caixas no estudo chega a $N=68.669$. Por (3) a fração amostral do estágio 2 foi encontrada:

$$f_2 = \frac{n}{N} = \frac{800}{68669} \cong 0,01165 \quad (3)$$

Isto é, toda caixa teve igual probabilidade 0,01165 de pertencer à amostra [6].

O processo de seleção de caixas obedeceu ao esquema de Amostragem Sistemática [4]. Isto é, um lote com $B=208$ caixas foi selecionado no estágio 1 do processo amostral. Na segunda fase, foi necessário coletar $k=8$ caixas desse lote. A partir disso calculou-se por (4) um intervalo sistemático i de:

$$i = \frac{B}{k} = \frac{208}{8} = 26 \quad (4)$$

Ou seja, selecionou-se 1 caixa a cada 26. Com o uso desse procedimento, houve um espaçamento intencional entre as amostras, podendo ser coletados itens produzidos em momentos distintos da linha de produção. A primeira unidade amostrada foi selecionada através de um sorteio aleatório de números naturais no intervalo de 1 a 26, sendo 20 o número obtido nesse sorteio. Logo, as caixas selecionadas foram as de ordem:

$$20; 20+i; 20+2i; 20+3i; 20+4i; 20+5i; 20+6i; 20+7i$$

Numericamente, obtiveram-se as caixas:

$$\underline{20}; \underline{46}=(20+26); \underline{72}=(20+52); \underline{98}=(20+78); \underline{124}=(20+104); \underline{150}=(20+130); \underline{176}=(20+156); \underline{202}=(20+182)$$

Em situações em que o valor do intervalo i não foi múltiplo da quantidade de caixas do lote (ultrapassando a última caixa para seleção), foi utilizado o Método de Lahiri, [13]. Nesse esquema de Amostragem Sistemática as unidades amostrais são consideradas como se estivessem em círculo.

4) *Estágio 3 - Amostragem Aleatória Simples:* A seleção dos produtos nas caixas foi feita probabilisticamente com a aplicação da Amostragem Aleatória Simples Sem Reposição [7]. Ou seja, através de um número aleatório referencial e diferente para cada caixa, foi selecionado um produto dentro de cada embalagem, considerando sempre a ordem interna de disposição dos produtos se iniciando da esquerda para a direita⁵. Por questões logísticas, a quantidade de itens por caixa variava entre 2 e 3, fornecendo uma quantidade média de aproximadamente 2,24, valor necessário para estimação da fração amostral do estágio 3.

Fração amostral, pesos e erros amostrais

1) *Frações e Ponderadores Amostrais:* Calculando através de (5) a fração amostral f [6], que é uma função dos três estágios da Amostragem por Conglomerados, obteve-se:

$$f = \frac{n}{P} = \frac{800}{154109} \cong 0,0051911 \quad (5)$$

⁵ A escolha por esquerda ou direita foi definida por sorteio.

Onde P é a quantidade de itens acondicionados em todas as caixas e n o tamanho amostral. Isto é, qualquer produto constante em todos os lotes teve probabilidade 0,0051911 de pertencer à amostra selecionada [4]; [6]. Pode ser mostrado ainda que (5) pode ser escrito através de (6), isto é, como uma função de (7), (8) e (9):

$$f = f_1 \cdot f_2 \cdot f_3 \tag{6}$$

com

$$f_1 = \frac{mB_j}{N} \tag{7}$$

$$f_2 = \frac{k}{B_j} \tag{8}$$

$$f_3 = \frac{f}{f_1 \cdot f_2} = \frac{n}{Lmk} \tag{9}$$

Em que, m é a quantidade de lotes selecionados, B_j é tamanho do j -ésimo lote sorteado, N é o total de caixas em todos os lotes, k é a quantidade de caixas amostradas por lote, L representa a média geral de produtos por caixa e f_1, f_2 e f_3 são as frações amostrais dos estágios 1, 2 e 3 da Amostragem por Conglomerados. Os pesos amostrais w (10), w_i (11) são funções dos inversos das frações amostrais, ou seja:

$$w = \frac{1}{f} \tag{10}$$

$$w_i = \frac{1}{f_i}, \quad i = 1, 2, 3 \tag{11}$$

Eles serviram como ponderadores para os totais amostrais como, por exemplo, em (12):

$$\hat{\tau} = \sum_{i=1}^T w_i y_i = \sum_{i=1}^T \frac{N_i}{n_i} y_i \tag{12}$$

Nesse caso, $\hat{\tau}$ foi o estimador usado para total populacional da característica relevante, w_i foi o respectivo peso da variável (em T unidades totais) e y_i a quantidade com detalhes de interesse na amostra. N_i representa as unidades totais da população em foco e n_i é a quantidade de dessas unidades na amostra. A variância deste estimador [5] foi calculada por (13);

$$\text{var}(\hat{\tau}) = \sum_{i=1}^T \left(\frac{N_i}{n_i} \right)^2 \text{var}(y_i) = \sum_{i=1}^T \left(\frac{N_i}{n_i} \right)^2 \left(y_i - \frac{y_i}{n_i} \right) \tag{13}$$

Em geral, a obtenção de estimativas de variâncias (alternativamente, de desvios padrões ou ainda de coeficientes de variação) é necessária para que intervalos de confiança possam ser calculados e outras formas de inferência realizadas [2].

As três fases do processo amostral foram feitas probabilisticamente, garantindo assim margens de erro e estimativas de confiabilidade, as quais foram calculadas para se realizar inferências sobre as populações pesquisadas. Neste sentido, inicialmente tomou-se como estimativa do coeficiente de correlação intraclasse (em geral situado entre 0,15 e 0,20, [5]), necessário para cálculo do *deff* (Fator de Efeito do Desenho Amostral) e, conseqüentemente, das margens de erro, o valor 0,20, obtendo-se um *deff* igual a 2,0. Assumindo uma estimativa conservadora da proporção de interesse [14], ou seja, $\hat{p} = 0,5$, a margem de erro associada ao estudo seria de 4,9 pontos percentuais, para mais ou para menos, com uma estimativa de 95% de confiabilidade. A fórmula [5] para estimação do erro amostral está representada a seguir (14):

$$e = \sqrt{\frac{p(1-p)(Z_{\alpha/2})^2}{n} \cdot [1 + \rho(k-1)]} \tag{14}$$

Em que, $Z_{\alpha/2}$ é o percentil da distribuição Normal com nível $\alpha = 0,05$ (95% de confiança), ρ é o coeficiente de correlação intraclasse, k a quantidades de caixas selecionadas por lote e n o tamanho amostral. O intervalo de confiança [5] para o total de produtos defeituosos, com confiabilidade de $(1-\alpha)100\%$, foi dado por (15):

$$\left[\hat{\tau} - Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n} \cdot [1 + \rho(k-1)]}; \quad \hat{\tau} + Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n} \cdot [1 + \rho(k-1)]} \right] \tag{15}$$

2) *Estimação do Coeficiente de Correlação Intraclasse e deff*: Nesta seção, são definidos formalmente os estimadores usados para o coeficiente de correlação intraclasse e *deff*, ambos necessários ao estudo. Procura-se aqui, apresentar de forma bem sucinta tais estimadores e seus parâmetros, indicando ao leitor interessado em maiores detalhes, as referências utilizadas.

Os valores para o coeficiente de correlação intraclasse ρ e para o *deff* foram estimados juntos aos conglomerados amostrais obtidos no estudo. Nessa linha, tem-se como estimadores de ρ as equações 16 a 20:

$$\hat{\rho} = r_{\text{int}} = \frac{s_{2ec}^2 - \frac{s_{2dc}^2}{\bar{B} - 1}}{s_{2ec}^2 + s_{2dc}^2} \tag{16}$$

sendo:

$$s_{2dc}^2 = \frac{1}{a} \sum_{\alpha=1}^a \left(\frac{B_\alpha}{\bar{B}} \right)^2 \frac{\psi}{B_\alpha} s_\alpha^2, \tag{17}$$

para

$$\psi = \frac{1}{A} \sum_{\alpha=1}^A B_{\alpha} \tag{18}$$

e

$$s_{2ec}^2 = \frac{1}{a-1} \sum_{\alpha=1}^a \left(\frac{B_{\alpha}}{\bar{B}} \right) (\bar{y}_{\alpha} - \bar{y}_{2c})^2, \tag{19}$$

com

$$\bar{y}_{2c} = \frac{\bar{B} \sum_{\alpha=1}^a \bar{y}_{\alpha}}{a\bar{B}} \tag{20}$$

onde:

- a: número de conglomerados no primeiro estágio⁶;
- B_{α} : número de unidades do conglomerado α no segundo estágio⁷;
- \bar{B} : número médio de unidades amostradas por conglomerado no estágio 2;
- ψ : tamanho médio (esperado) das amostras no segundo estágio;
- \bar{y}_{α} : média do conglomerado α no segundo estágio;
- \bar{y}_{2c} : média geral calculada a partir dos estágios 1 e 2;
- s_{2dc}^2 : estimador não viesado da variabilidade dentro dos conglomerados no estágio 2;
- s_{2ec}^2 : estimador não viciado para a variabilidade entre conglomerados no estágio 2.

Como a quantidade de amostras por lote no estágio 2 não variou, os conglomerados amostrados nessa fase foram considerados de tamanhos iguais, o que justifica os estimadores apresentados [7].

Algumas considerações podem ser feitas sobre o coeficiente de correlação intraclasse, segundo a literatura. Quando ρ é maior que zero, os conglomerados são menos precisos, para um determinado volume de amostra. Porém, quando ρ é menor que zero, como acontece às vezes, os conglomerados são mais precisos promovendo um ganho amostral [4]. Em geral, ρ é positivo, e isso leva a perda de eficiência em relação a Amostragem Aleatória Simples [7]. A eficiência do plano é calculada pelo Fator de Efeito do Desenho Amostral (*deff*), medido pela função (21):

$$deff[\bar{y}_{2c}] \cong 1 + \rho(\psi - 1) \tag{21}$$

⁶ Em nosso estudo poderemos considerar $a=m$ conglomerados amostrados no estágio 1, conforme notação anterior.

⁷ Como B_{α} e \bar{B} são iguais, as fórmulas se tornam simplificadas.

Fazendo $\psi =k$, foi estimado o valor do *deff* no plano para os dados coletados. A partir desse cálculo, as margens de erro associadas ao estudo foram estimadas (14).

III. RESULTADOS

Após a coleta de dados nos três estágios amostrais e estimadas as estatísticas necessárias para o cálculo do coeficiente de correlação intraclasse, *deff* e das margens de erro, apresentam-se, na Tabela II, os principais resultados e informações relacionados ao estudo.

TABELA II – PRINCIPAIS INFORMAÇÕES, ESTIMATIVAS E RESULTADOS.

Estágios		3
População (itens)		154.109
Tamanho amostral		800
Fração amostral (<i>f</i>)		0,0051911
Peso (<i>w</i>)		192,636
Itens defeituosos na amostra		46
	Rho	0,0017
Estimativas para cálculo de margem de erro	<i>Deff</i>	1,012
	$\sigma^2_{conglomerados}$	0,253
	Margem de erro	0,0349
	Estimativa Pontual	5,75%
Estimativa para % de produtos defeituosos	Intervalo Inferior (95%)	2,26%
	Intervalo Superior (95%)	9,24%
	Estimativa Pontual	8.861
Estimativa para total de produtos defeituosos	Intervalo Inferior (95%)	3.483
	Intervalo Superior (95%)	14.240
Máximo aceitável de % de produtos defeituosos	Pelo intervalo inferior	3,0%
Máximo aceitável de total de produtos defeituosos	Pelo intervalo inferior	4.624

Os cálculos da Tabela II mostraram uma baixa correlação intraclasse estimada, situada em torno de $\rho=0,0017$, isto é, bem próxima de zero. Essa correlação forneceu um *deff*=1,012, apresentando uma margem de erro de 3,49 pontos percentuais, para mais ou para menos, sendo considerada uma confiabilidade de 95%. Perceba que houve um ganho considerável em relação à estimativa inicial de margem de erro (4,9 pontos percentuais) devido ao valor de ρ ser próximo a zero.

O peso *w*, calculado por (10) a partir da fração amostral *f* (5) relacionada ao processo, indica que os produtos na

amostra tiveram peso 192,637. Esse peso foi usado para se calcular através de (12) a população relacionada ao total de itens defeituosos.

Os resultados da pesquisa indicaram que o total de produtos falhos foi de 46 em 800 itens amostrados, isto é 5,75%, o que forneceu um total estimado de acordo com (12) de 8.862 produtos com defeito. Ou seja, há uma estimativa pontual de que haja um total de quase nove mil itens com algum tipo de falha na fabricação.

A margem de erro associada ao estudo, isto é, os 3,49 pontos percentuais de erro, para mais ou para menos, sobre a estimativa pontual, forneceu um limite inferior ou patamar mínimo de 2,26% de produtos defeituosos, chegando a 9,24% no limite superior. Estes percentuais apresentaram, em números totais através de (15), um intervalo de confiança de 95% para o total de produtos defeituosos, as estimativas de 3.483 a 14.240 itens falhos. O percentual de 3% de erros máximos esperados⁸ para todos os produtos totaliza 4.624 itens. Pode ser observado na Tabela II que esse valor encontra-se inserido no intervalo de confiança calculado anteriormente.

IV. COMENTÁRIOS E CONCLUSÕES FINAIS

Os resultados encontrados foram baseados em técnicas de Amostragem por Conglomerados em três estágios, considerando os fatores externos, como o coeficiente de correlação intraclasse e o *deff*, na estimação da margem de erro, garantido que não foram ignoradas possíveis fontes de inflação de variabilidade, que não estariam mensurados se um processo de amostragem simples fosse considerado.

Uma observação importante se refere ao coeficiente ρ , ou seja, como este foi bem próximo a zero, houve um ganho na precisão da pesquisa, isto é, a amostragem por conglomerados está bem próxima a amostragem aleatória simples (AAS), em termos de margens de erro. Por exemplo, a variabilidade da AAS é 0,25, enquanto pela Tabela II, observa-se que a variância da amostragem por conglomerados é 0,253, permitindo o cálculo de margens de erro com estimativas bem próximas.

Considerando que o critério de aceitação dos lotes do produto definiu que o limite inferior do intervalo de confiança não poderia exceder 3% de itens defeituosos, concluiu-se pela não rejeição do fornecedor, pois o intervalo de confiança estimado de 2,26% a 9,24% de produtos defeituosos contém o valor de aceitação dos lotes, ou seja, 3%.

Por fim, os procedimentos aqui relatados podem servir como referência para que sejam investigados, com a utilização de outras metodologias estatísticas, como o controle estatístico de processos, métodos de controle de qualidade, ou ainda, técnicas de planejamentos de experimentos, o índice de falhas desses produtos relacionados a subdivisões que possam indicar problemas operacionais na linha de produção. Pode-se verificar também alguma outra característica que mereça ser considerada para qualificação dos itens junto aos consumidores finais.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] A. C. A. Maximiano, *Teoria Geral da Administração*. São Paulo: Atlas, 1997.
- [2] D. G. C. Pessoa, P. L. N. Silva, *Análise de Dados Amostrais Complexos*. São Paulo: Associação Brasileira de Estatística, 1998.
- [3] D. C. Montgomery, *Introdução ao Controle Estatístico da Qualidade*. 4ed. LTC, 2004.
- [4] W.G. Cochran, *Sampling Techniques*, 3ª ed., New York: John Wiley & Sons, 1977.
- [5] H. D. Neder, *Amostragem em Pesquisas Socioeconômicas*. Campinas: Editora Alínea, 2008.
- [6] N. N. Silva, *Amostragem Probabilística: Um Curso Introductório*. São Paulo: EdUSP, 2004.
- [7] H. Bolfarine, & W. O. Bussab, *Elementos de Amostragem*, 1º ed., São Paulo: Edgard Blücher, 2005.
- [8] *SPSS Complex Samples: User's Guide*, Chicago: SPSS Inc., 2007.
- [9] P. R. Cunha, "Técnicas de amostragem aplicadas pelas empresas de auditoria independente em Santa Catarina". Dissertação de Mestrado. Universidade Regional de Blumenau. 2005. Disponível em: http://proxy.furb.br/tede/tde_arquivos/1/TDE-2007-01-19T121950Z232/Publico/xDiss%20Paulo%20Roberto%20da%20Cunha.pdf
- [10] K. M. Sullivan, S. May, G. Maberly, "Urinary iodine assessment: a manual on survey and laboratory methods", 2nd ed. UNICEF, PAMM, 2000. Em: http://www.iccid.org/media/assessment%20tools/survey_sites.pdf
- [11] M. S. Azevedo, "Nota técnica: o índice do medo do desemprego". IPEA, Instituto de Pesquisa Econômica Aplicada, 2009. Disponível em: http://www.ipea.gov.br/sites/000/2/boletim_mercado_de_trabalho/mt40/03_NT_indice.pdf
- [12] P. L. N. Silva, D. G. C. Pessoa, M. F. Lila, "Análise estatística de dados da PNAD: incorporando a estrutura do plano amostral". *Revista Ciência e Saúde Coletiva*. Vol.7, No.4, Rio de Janeiro. 2002. Em: http://www.scielo.br/scielo.php?pid=S141381232002000400005&script=sci_arttext
- [13] D. B. Lahiri, "A method for sample selection providing unbiased ratio estimates". *Bulletin of International Statistics Institute* 33, 24-47, 1951.
- [14] M. N. Magalhaes & A. C. P. Lima, *Noções de Estatística e Probabilidade*. São Paulo: EdUSP, 2002.

⁸ Em relação à estimativa intervalar inferior.