

DIMENSIONAMENTO DA AMOSTRA EM PESQUISA DE SATISFAÇÃO DE CLIENTES

José Luis Duarte Ribeiro

Programa de Pós-graduação em Engenharia de Produção – UFRGS
Praça Argentina 9, sala 402 – 90040-020 - Porto Alegre - RS; e-mail: ribeiro@vortex.ufrgs.br

Márcia Elisa Echeveste

Programa de Pós Graduação em Engenharia de Produção - UFRGS
Praça Argentina 9, sala 402 – 90040-020 - Porto Alegre - RS

Abstract

This paper describes a method for the definition of sample sizes to be used in market researches. The proposed method has seven phases and intends to facilitate the choice of the proper sample sizes to be used in a particular research. It contemplates three different possibilities: (i) samples of same size for all the stratum; (ii) samples sizes that are proportional to the size of the populations of each stratum, and (iii) samples sizes that minimizes the loss due to misjudgment. The proposed method was applied in a research for client satisfaction in the chemical sector. The results where very effective.

Área: 4. Engenharia do produto / 4.1. Pesquisa de Mercado

Key words: Market Research, Quality, Client Satisfaction

1. INTRODUÇÃO

Este trabalho tem como objetivos estabelecer e documentar uma sistemática para o dimensionamento da amostra em pesquisas de mercado. Isso será feito através da descrição de um caso prático, mais especificamente, uma Pesquisa de satisfação de clientes atualmente empregado por uma Indústria Química junto às suas vendas.

Na literatura existem várias publicações abordando o problema do dimensionamento do tamanho da amostra (ver por exemplo, Costa Neto, 1977; Snedecor e Cochran, 1980, Clark e Cook, 1983 e Nanni e Ribeiro, 1991), e as estratégias de dimensionamento diferem entre si. Isso acontece porque os autores partem de diferentes premissas. Essas premissas muitas vezes não são esclarecidas, e isso contribui para deixar o leitor com dúvidas a respeito de qual estratégia adotar. Esse trabalho vai mostrar que a questão de usar amostras de tamanho fixo ou variável para cada estrato depende dos objetivos de quem realiza a pesquisa. Uma vez que os objetivos da pesquisa tenham sido identificados, a estratégia de dimensionamento adequada pode ser escolhida com facilidade.

2. IDENTIFICAÇÃO DA POPULAÇÃO

O estudo de caso que é analisado neste artigo contempla uma pesquisa de opinião anteriormente realizada pela empresa junto às suas vendas, distribuídas em várias regiões do Brasil. As regiões possuem participação diferente no mercado global, dependendo do

volume de produtos comercializados. As vendas, por sua vez, são classificadas em pequena, média e grande, também de acordo com o volume de produtos comercializados mensalmente. (ver tabela 4).

3. PROPOSTA DE DIMENSIONAMENTO

O método proposto neste trabalho realiza o dimensionamento da amostra a partir das seguintes etapas:

Etapa 1: Estratificar a população e identificar subpopulações;

Etapa 2: Levantar um indicador numérico e verificar a sua variabilidade dentro das subpopulações;

Etapa 3: Estabelecer o nível de significância e um erro aceitável para o indicador escolhido;

Etapa 4: Dimensionar o tamanho da amostra por subpopulação, considerando a variabilidade, o erro aceitável e a significância desejada;

Etapa 6: Decidir por amostra de mesmo tamanho ou por amostras de tamanho variável;

Etapa7: Calcular os tamanhos das amostras por estrato e verificar a compatibilidade com os recursos disponíveis. Caso os recursos não sejam suficientes, voltar a etapa 3.

A figura 1 ilustra as etapas a serem percorridas para o dimensionamento do tamanho da amostra. A seguir cada etapa será comentada em maior detalhe.

Etapa 1: Estratificar a população e identificar subpopulações

Num estudo realizado por amostragem é importante verificar a variabilidade existente na população. A estratificação permite analisar um grupo heterogêneo através de subgrupos (estratos) homogêneos. Os estratos devem ser escolhidos de forma a possuírem homogeneidade interna e, por outro lado, apresentarem heterogeneidade entre si. Os fatores que levarão à classificação dos estratos depende do objeto do estudo, normalmente referem-se a regiões geográficas, turnos, dias de semana, modelo de produto, etc. Quando se trata de indivíduos, classe social, sexo, faixa etária, são exemplos típicos de fatores de estratificação.

Para este estudo, os fatores de estratificação foram a região e o porte da venda. Assim as subpopulações foram: região sul porte pequeno, região sul porte médio; região sul porte grande, região sudeste porte pequeno, região sudeste porte médio, região sudeste porte grande, e assim por diante. Totalizando 18 estratos = 6 regiões de 3 portes cada.

Etapa 2: Levantar um indicador numérico e verificar a sua variabilidade dentro das subpopulações

Para obtermos uma indicação da variabilidade existente na população, escolhe-se um indicador numérico. Neste estudo o indicador utilizado foi a nota média dos questionários aplicados na pesquisa anterior. Para estimarmos a variabilidade, calculou-se o desvio padrão observado em dois estratos com características extremas no que se refere a atitude dos clientes: a região sudeste e a região nordeste. O desvio padrão foi calculado para todos os portes de venda: pequeno, médio e grande. A média e o desvio-padrão

estimado para esses estratos aparece na Tabela 1.

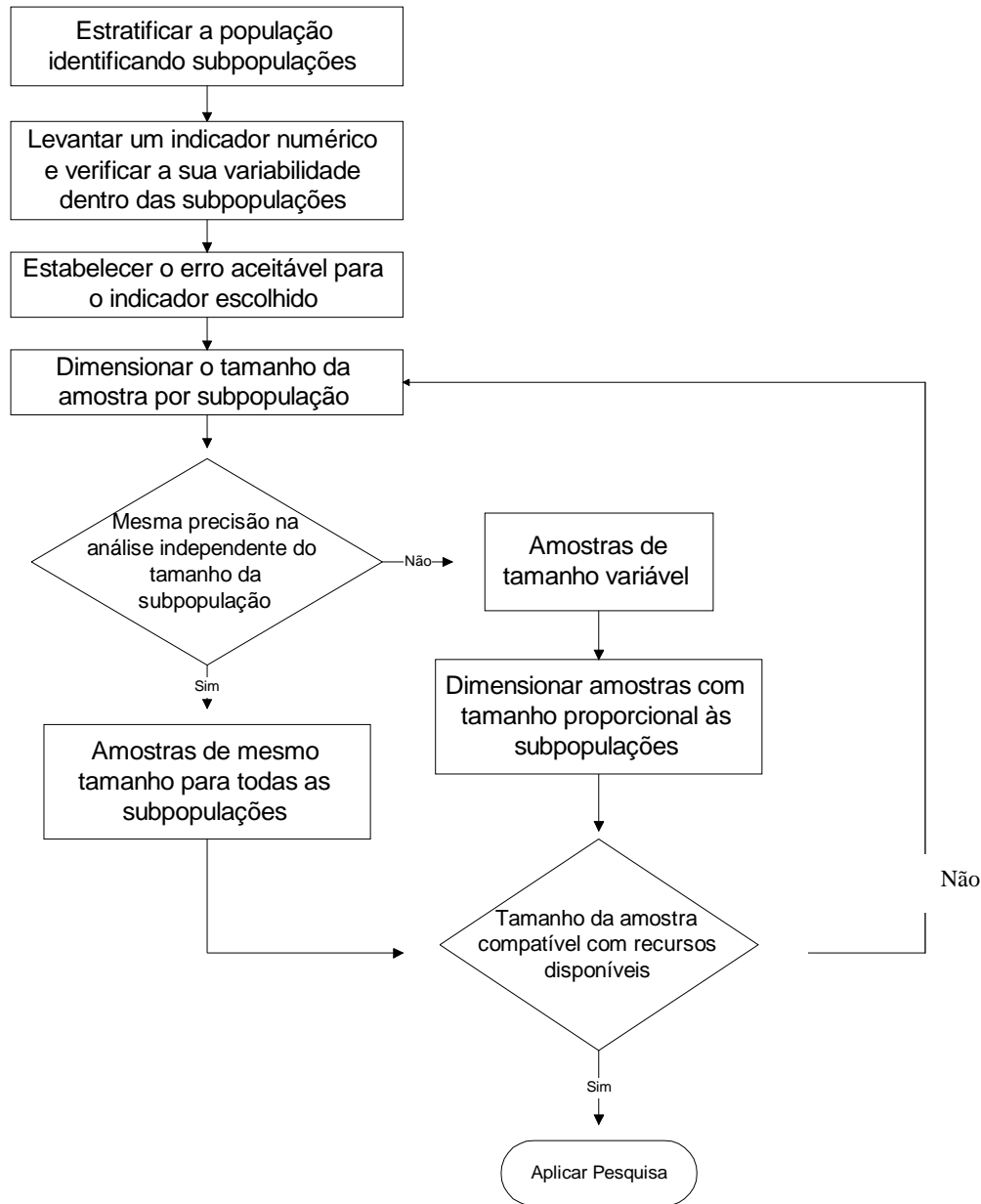


Figura 1: Método de Dimensionamento do tamanho da amostra.

A medida geral de variabilidade foi estimada através de uma média ponderada, onde os pesos levaram em conta precisamente o percentual (volume) de vendas em cada estrato analisado. Esse percentual também aparece na Tabela 1, onde os dados foram modificados para preservar o sigilo das informações. A fórmula para obter a variabilidade é dada por:

$$\hat{\sigma}_e = \frac{\sum_{i=1}^k \hat{\sigma}_i \times w_i}{\sum_{i=1}^k w_i}, \quad (\text{e.q.1})$$

Onde:

$\hat{\sigma}_e$ = desvio padrão geral para a variável em análise (média dos questionários);

K = número de estratos;

$\hat{\sigma}_i$ = desvio padrão das médias dos questionários aplicados no estrato i ;

w_i = percentual de vendas associado ao estrato i ;

Tabela 1: Cálculo do desvio-padrão a partir da variabilidade observada nos resultados (médias) dos questionários anterior

Região	Porte		Importância relativa (W_i)	Média	Desvio-Padrão
Sul 33,5	Pequeno	63,5	0,213	8,23	0,75
	Médio	23,5	0,079	8,5	0,72
	Grande	13,0	0,044	8,08	0,97
Nordeste 14,5	Pequeno	63,0	0,091	8,12	0,62
	Médio	30,0	0,044	8,43	0,67
	Grande	7,0	0,010	8,17	0,88
			0,480		0,736

Assim,

$$\hat{\sigma}_e = \frac{0,213 \times 0,75 + 0,079 \times 0,72 + 0,044 \times 0,97 + 0,091 \times 0,62 + 0,044 \times 0,67 + 0,010 \times 0,88}{0,480} = 0,736$$

Etapa 3: Estabelecer o nível de significância e um erro aceitável para o indicador escolhido

O nível de significância foi estabelecido em 95%, ou seja, $\alpha = 0,05$. O erro máximo de estimação (ϵ) é a diferença máxima que se admite obter entre o valor estimado (amostra) e o valor real (população), considerando o nível de significância adotado. Esse erro é fixado antes de se iniciar a pesquisa e incide diretamente no tamanho final de amostra, isto é, quanto maior a precisão desejada (menor o erro aceitável) maior será o tamanho de amostra.

Inicialmente, partiu-se de um erro $\epsilon = 0,5$. Isso significa que as inferências a respeito da média de qualquer estrato poderão ser feitas através de um intervalo de confiança de $1-\alpha$ % com amplitude $\pm\epsilon$.

Etapa 4: Dimensionar o tamanho da amostra por subpopulação

O dimensionamento do tamanho da amostra depende de quatro fatores principais: (i) o nível de significância desejado; (ii) a variabilidade da população; (iii) erro máximo de estimação; (iv) custo da pesquisa.

O tamanho da amostra é dado pela seguinte fórmula:

$$n = \frac{z^2 \times \sigma_e^2}{\epsilon^2} \quad (\text{eq. 2})$$

O nível de significância indica a probabilidade que o valor real (da população) esteja contido no intervalo estimado a partir de dados amostrais. Utilizou-se nível de significância de 95% que está associado a $Z = 1,96$ desvios-padrões da distribuição Normal.

A variabilidade da população (σ_e) foi estimada na etapa 2. O erro máximo aceitável (ϵ) foi detalhado na etapa 3. Assim, neste caso, o tamanho de amostra recomendado seria:

$$n = \frac{(1,96)^2 \times (0,736)^2}{(0,5)^2} \cong 8$$

Como tem-se 18 estratos, o número total de questionários a ser aplicado é de $18 \times 8 = 144$ questionários.

Etapa 6: Decidir por amostra de mesmo tamanho ou por amostras de tamanho variável

O dimensionamento da amostra pode seguir três procedimentos distintos:

(i) *amostras iguais para todos os estratos*

Esse procedimento é recomendado quando se deseja a mesma precisão em todas as inferências, independentemente da importância, ou seja, do tamanho do estrato. Neste caso, o tamanho de amostra seria de 8 questionários para cada estrato, conforme calculado na etapa anterior.

(ii) *amostras proporcionais ao tamanho do estrato*

Esse procedimento é recomendado quando se deseja uma precisão muito maior nos estratos mais importantes. Nesse caso, o tamanho de amostra de 144 revendas seria distribuído de acordo com a participação de cada revenda no mercado. Por exemplo, revendas da região sul de porte pequeno são responsáveis por 12 % dos pedidos totais, o que daria: $0,12 \times 144 = 17,3$ questionários a serem aplicados neste estrato (ver tabela 4).

Para evitar que alguns estratos ficassem com um número muito pequeno ou muito grande de questionários, fixou-se que o tamanho de amostra deveria ser no mínimo 3 e no máximo 30 questionários. Estes valores foram arbitrários e podem ser alterados sempre que for desejado, assim como o erro máximo de estimação e o nível de confiança.

(iii) *amostras que minimizem a perda imposta por erros de julgamentos*

A terceira alternativa busca minimizar a perda associada a erros de julgamento. Essa perda está associada a dois fatores principais: (i) a magnitude dos erros de inferência, os quais dependem dos tamanhos de amostra em cada estrato; e (ii) ao próprio tamanho dos estratos (erros de julgamento irão resultar em uma perda maior no caso de estratos maiores).

Considerando que a perda seja proporcional ao quadrado dos erros de inferência (seguindo as idéias de Taguchi et al., 1990), pode ser demonstrado que o tamanho das amostras deve ser diretamente proporcional a raiz quadrada do tamanho dos estratos. Têm-se a seguinte formulação matemática do problema:

$$\text{Perda total} \sim \sum p_i \times \epsilon_i^2$$

$$\sim \sum p_i \times (\sigma / \sqrt{n_i})^2$$

$$\sim \sigma^2 \sum p_i / n_i$$

onde:

- p_i representa o tamanho do estrato i ;
- $\varepsilon_i = \sigma / \sqrt{n_i}$ representa o erro de inferência;
- σ é o desvio padrão do indicador em estudo;
- n_i é o tamanho da amostra a ser aplicada neste estrato.

Para minimizar o somatório de p_i / n_i , dado que a soma de n_i é fixa e igual a N (número total de questionários a serem distribuídos), é necessário que os valores de n_i sejam proporcionais à raiz quadrada de p_i . Isso será demonstrado através de um exemplo onde tem-se três estratos de tamanhos $p_i = 25, 9$ e 1 , e $N = 30$ questionários a serem distribuídos. Os cálculos aparecem na Tabela 2.

$p_i =$	25	9	1	Σn_i (fixo)	$\Sigma p_i / n_i$ (perda)
$n_i =$	10	10	10	30	3,500
	12	10	8	30	3,108
	14	10	6	30	2,852
	16	10	4	30	2,712
	16,67	10,00	3,33	30	2,697
	17	10	2	30	2,703
	18	10	1	30	3,289

==> mínimo quando os $n_i \sim \sqrt{p_i}$

Tabela 2: cálculo da perda para diversas situações hipotéticas.

Assim, por exemplo, as vendas da região sul de porte pequeno que são responsáveis por 12 % das vendas irão gerar um tamanho de amostra que é proporcional à raiz quadrada deste percentual.

A motivação para o uso da perda proporcional ao quadrado do erro de inferência é a seguinte: pequenos erros de inferência vão afetar muito pouco a tomada de decisão e, portanto, a perda devida a erros de julgamento será pequena; por outro lado, grandes erros de inferência podem afetar dramaticamente a tomada de decisão, podendo conduzir a perdas muito grandes, que não crescem linearmente, mas sim quadraticamente na medida que o erro aumenta.

Etapa7: Calcular os tamanhos das amostras por estrato e verificar a compatibilidade com os recursos disponíveis. Caso os recursos não sejam suficientes, voltar a etapa 3.

Para facilitar o cálculo do tamanho de amostra utilizou-se um aplicativo de apoio à decisão desenvolvido no Microsoft Excel com as fórmulas necessárias. Os campos hachurados são editáveis e podem ser alterados sempre que necessário. São eles: o nível de significância, o erro máximo de estimação, o tamanho mínimo e máximo para cada estrato (ver Tabela 3) e as porcentagens de participação de cada estrato que foram utilizadas para atribuir a importância relativa de cada estrato dentro da população (ver Tabela 4).

Assim, cada vez que um destes campos forem alterados, o aplicativo automaticamente recalcula os tamanhos de amostras utilizando os métodos (1), (2) e (3) destacados no topo da Tabela 4.

A decisão de qual procedimento adotar dependerá dos objetivos do estudo,

conforme mencionado anteriormente. Além disso, o tamanho da amostra deve ser compatível com os recursos disponíveis para a pesquisa. Caso os recursos não sejam suficientes, recalcula-se o tamanho da amostra retornando a etapa 3 e revendo o nível de significância e/ou o erro aceitável.

Tabela 3: Dados preliminares necessários para o dimensionamento das amostras.

Nível de significância	0,95
Valor de z associado ao nível de significância.	1,96
Desvio padrão natural do processo	0,736
Erro aceitável	0,50
Número de questionários médio por estrato	8
Número de estratos	18
Número total de questionários	144
Número mínimo de questionários por estrato	3
Número máximo de questionários por estrato	30

Tabela 4: Dimensionamento das amostras, conforme os três critérios enunciados.

Regiões	Clientes		(2)			(3)		(1)		
			% sobre total	Proporcional ao %	Raiz do %	Raiz do % ajustada	Proporcional a raiz do %	Tamanho fixo		
Sul	Peq	59,0	0,121	17,4	17	0,348	0,089	12,8	13	8
	Med	27,5	0,056	8,1	8	0,237	0,061	8,7	9	8
	Grd	13,5	0,028	4,0	4	0,166	0,043	6,1	6	8
Sudeste	Peq	63,5	0,213	30,6	30	0,461	0,118	17,0	17	8
	Med	23,5	0,079	11,3	11	0,281	0,072	10,3	10	8
	Grd	13,0	0,044	6,3	6	0,209	0,053	7,7	8	8
Centro-Oeste	Peq	61,5	0,049	7,1	7	0,222	0,057	8,2	8	8
	Med	22,0	0,018	2,5	3	0,133	0,034	4,9	5	8
	Grd	16,5	0,013	1,9	3	0,115	0,029	4,2	4	8
Nordeste	Peq	63,0	0,091	13,2	13	0,302	0,077	11,1	11	8
	Med	30,0	0,044	6,3	6	0,209	0,053	7,7	8	8
	Grd	7,0	0,010	1,5	3	0,101	0,026	3,7	4	8
Norte	Peq	74,5	0,078	11,3	11	0,280	0,072	10,3	10	8
	Med	13,5	0,014	2,0	3	0,119	0,030	4,4	4	8
	Grd	12,0	0,013	1,8	3	0,112	0,029	4,1	4	8
Exterior	Peq	21,0	0,027	3,9	4	0,165	0,042	6,1	6	8
	Med	30,5	0,040	5,7	6	0,199	0,051	7,3	7	8
	Grd	48,5	0,063	9,1	9	0,251	0,064	9,2	9	8
100			1,000	144	147	4	1	144	143	144

4. CONSIDERAÇÕES FINAIS

Este artigo apresenta um método de dimensionamento da amostra que pretende atender a qualquer pesquisa de opinião. Numa primeira aplicação, o dimensionamento da amostra foi direcionado para uma pesquisa de satisfação do cliente realizada em todo território nacional, com uma periodicidade anual. Entretanto, as informações de entrada podem ser alteradas tornando o método flexível ao dimensionamento da amostra em pesquisas com propósitos diferentes.

O método de dimensionamento da amostra contempla 7 etapas que cobrem desde a definição dos estratos até o dimensionamento final e distribuição do esforço de amostragem por estrato.

O método apresentado propõe três procedimentos para o cálculo do tamanho da amostra: (i) quando se deseja tamanhos de amostras iguais por estrato, (ii) quando se deseja tamanhos de amostras proporcionais ao tamanho do estrato na população e (iii) quando se deseja tamanhos de amostra que minimizem a perda imposta por erros de julgamentos.

Para facilitar e direcionar o uso do método proposto, um aplicativo de apoio à decisão utilizando o Microsoft Excel foi desenvolvido e contém o formulário correspondente a todos os procedimentos detalhados..

A decisão final dependerá da equipe de pesquisa que a partir do aplicativo, poderá testar diferentes cenários e distribuir os questionários através dos estratos, avaliando simultaneamente os tamanhos de amostra, ou seja, os custos de amostragem, e a confiabilidade das respectivas estimativas.

REFERÊNCIAS BIBLIOGRAFIAS

Costa Neto, P. L. O. (1977) - Estatística. Edgar Blücher, São Paulo.

Clarke, G. M. & Cook, D. (1983), A Basic Course in Statistics. 2nd ed., Edward Arnold Ltda, London.

Nanni, L.F. & Ribeiro, J.L. (1991) - Planejamento e avaliação de experimentos. Caderno de Engenharia 17/87, 2^a ed., CPGEC/UFRGS, Porto Alegre, Brasil.

Snedecor, G. W. & Cochran, W. G. (1980), Statistical Methods, 7th ed., The Iowa State Univ. Press, Iowa, USA.

Taguchi, G., Elsayed, E. & Hsiang, T. (1990) - Engenharia da Qualidade em Sistemas de Produção. McGraw-Hill, São Paulo, Brasil.