

# Plano de amostragem para controle da qualidade de água na rede de distribuição do Município de São Paulo

Rodolfo José da Costa e Silva Júnior (1)  
José Everardo Van Den Brùle Matos (2)

## 1 Histórico

O Departamento de controle sanitário da Sabesp dentre outras atribuições controla a qualidade da água distribuída à população do Município de São Paulo.

Em anos anteriores este controle era feito com base nas reclamações da qualidade da água, oriundas de diversas fontes de informação que envolviam praticamente as seguintes variáveis: água suja, ausência de cloro e contaminação bacteriológica. As reclamações eram atendidas pontualmente, isto é, as equipes de controle se deslocavam até o local de cada reclamação, onde se procediam a inspeções sanitárias e decorrentes dessas inspeções se tomavam medidas corretivas tais como: descargas de rede para eliminação dos problemas de água suja, estabilização dos níveis de cloro e coleta para verificação da potabilidade bacteriológica.

Em anos subsequentes, outras variáveis relacionadas à qualidade da água foram agregadas ao controle exercido pelo Departamento de Controle Sanitário. Através do "Plano de Vigilância" Epidemiológica estabelecido pela Secretaria de Estado da Saúde, os casos de doenças hidrottransmissíveis passaram a ser notificados a este Departamento, gerando inspeções locais e colheita de água para exame de potabilidade bacteriológica.

A Cetesb-Companhia de Tecnologia de Saneamento Ambiental, em convênio já há algum tempo firmado com a Sabesp, desenvolveu o chamado "Programa de Vigilância da Qualidade da Água para Consumo Humano", notificando também a este Departamento as anomalias relacionadas à qualidade da água distribuída.

Diante do número e da diversidade das informações a serem controladas foi necessária uma revisão da sistemática de controle exercida pelo Departamento de Controle Sanitário e portanto foi criado um plano estatístico que

viabilizasse a melhor interpretação das variáveis controladas e o melhor desempenho quanto às ações corretivas.

A rede de distribuição de água do Município de São Paulo, que compreende aproximadamente 15.430 km de extensão, é definida por setores de abastecimento num total de 58 setores.

Os setores de abastecimento são áreas de influência de um ou mais reservatórios e estes, por sua vez, são alimentados pelos efluentes das Estações de Tratamento de Água que compõe os Sistemas Produtores da Sabesp.

O plano de amostragem desenvolvido por este departamento recebeu o nome de "Controle dos 100 pontos", pois se fundamentava na amostragem aleatória de 100 pontos dentro de cada um dos 58 setores de abastecimento e as variáveis controladas eram basicamente o cloro residual e a turbidez.

Após algum tempo em uso, verificaram-se falhas de ordem operacional com relação à frequência de amostragem de cada setor. Devido a problemas tais como: área do setor, dificuldades de coleta em determinadas áreas e o número de pontos a serem varridos, havia casos de um setor ser reamostrado após 6 meses. Com o fim de se eliminar tal distorção, pensou-se num controle mais efetivo, priorizando-se a frequência de amostragem. Para tal, foi dimensionado o tamanho da amostra com base nos dados de cloro e turbidez de cada setor.

## 2 Dimensionamento do tamanho da amostra

A amostragem tem como principal objetivo tirar conclusões de uma população de elementos sobre uma amostra que represente esses elementos e as conclusões assim obtidas constituem o que chamamos de inferência estatística. A similaridade entre uma amostra e a sua correspondente população é praticamente impossível, porém é perfeitamente viável se estimar o valor do erro ao di-

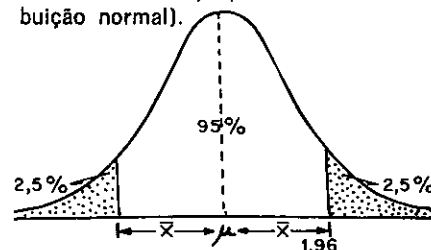
zermos que determinada amostra representa sua correspondente população.

A população é avaliada através da amostra, portanto, os parâmetros populacionais, como a média  $\mu$  e o desvio padrão  $\tau$ , são avaliados pelas estatísticas amostrais correspondentes, isto é, a média amostral  $\bar{x}$  e o desvio padrão amostral  $s$ . As estimativas amostrais são conhecidas como estimativas pontuais, porque originam uma única estimativa do parâmetro populacional.

Ao estimarmos a média  $\mu$  de uma população através de sua estimativa pontual  $\bar{x}$ , observaremos que estes valores serão distintos e com certeza estaremos cometendo um erro de estimação. Com o fim de evitar isto, é usual se construir um intervalo com base na estimativa pontual, de tal forma que se possa, com uma probabilidade conhecida, afirmar que este intervalo conterá o parâmetro populacional  $\mu$ .

O intervalo, que com probabilidade conhecida deverá conter o valor real do parâmetro  $\mu$ , é chamado intervalo de confiança. A probabilidade conhecida de que um intervalo de confiança contenha o valor do parâmetro populacional  $\mu$  é chamada nível de confiança ou grau de confiança do intervalo, designado por  $1-\alpha$ . O valor  $\alpha$  será a probabilidade de erro na estimação por intervalo, isto é, a probabilidade de errarmos ao afirmar que o valor do parâmetro está contido no intervalo de confiança.

Em resumo, quando afirmamos que a média amostral  $\bar{x}$  está a menos de 1,96 desvio padrão a contar da média verdadeira  $\mu$ , podemos esperar estar certos 95% das vezes e errados 5% das vezes (propriedade da distribuição normal).



Logo, um intervalo de confiança de 95% leva consigo um risco de 5% de erro; 5% dos intervalos assim fixa-

(1) Eng. Civil Chefe da Divisão de Controle de Qualidade da Diretoria de Operação da Região Metropolitana de São Paulo — Sabesp.  
(2) Engenheiro Químico — Coordenador de Programa de Controle de Qualidade da Diretoria de Operação da Região Metropolitana de São Paulo — Sabesp.

dos não incluirão a média populacional  $\mu$ .

O erro num intervalo de estimação diz respeito ao desvio (diferença) entre a média amostra  $\bar{x}$  e a verdadeira média da população  $\mu$ . Logo, o intervalo

$$\bar{x} \pm z \cdot \frac{\tau}{\sqrt{n}}$$

pode ser escrito assim:  $\bar{x} \pm \text{erro}$ , o erro e sendo dado por

$$\text{erro} = z \cdot \frac{\tau}{\sqrt{n}} \quad (1)$$

onde:

$\bar{x}$  = média amostral

$z$  = número de desvios a contar da média e é função do nível de confiança.

$n$  = tamanho da amostra

$\tau$  = desvio padrão da população

Quando não se tem conhecimento do desvio padrão da população, usamos a sua estimativa amostral  $s$ ; assim o erro pode ser escrito:

$$\text{erro} = t \cdot \frac{s}{\sqrt{n}} \quad (2)$$

onde:

$t$  = Estatística estabelecida por Student.

É função do nível de significância e do  $n^\circ$  de graus de liberdade  $\Phi$  ( $\Phi = n-1$ )

$s$  = Desvio padrão calculado a partir da amostra.

$n$  = Tamanho da amostra.

Com as fórmulas (1) e (2), podemos determinar a quantidade de erro associada à dispersão de uma amostra, o tamanho da amostra e o nível de confiança. Entretanto, às vezes podemos querer especificar um erro tolerável e talvez achar o tamanho da amostra necessário para gerar aquele erro para determinado nível de confiança e um desvio padrão amostral determinado.

Das fórmulas (1) e (2) resulta:

$$n = \left[ z \cdot \frac{\tau}{\epsilon} \right]^2 \quad (3)$$

$$n = \left[ t \cdot \frac{s}{\epsilon} \right]^2 \quad (4)$$

(3) Tamanho da amostra quando se conhece o desvio padrão da população.

(4) Tamanho da amostra quando não se conhece o desvio padrão da população, mas sim a sua estimativa  $s$ , calculada a partir da amostra.

Uma vez que para o dimensionamento do tamanho da amostra dos setores de abastecimento não temos conhecimento do desvio padrão das populações estudadas, a fórmula a ser adotada é:

$$n = \left[ t \cdot \frac{s}{e} \right]^2$$

No caso em que desejamos assegurar uma certa precisão na estimativa da média, é necessário termos uma idéia sobre a ordem de grandeza do coeficiente de variação; portanto, a fórmula acima pode ser escrita assim:

$$n = \left[ t \cdot \frac{Cv}{e} \right]^2$$

onde:

$n$  = tamanho da amostra

$t$  = estatística de Student; função do nível de significância e do número de graus de liberdade.  $\Phi$  ( $\Phi = n-1$ )

$Cv$  = coeficiente de variação

$$Cv = \frac{s}{\bar{x}}$$

$e$  = erro relativo máximo.

#### Exemplo de aplicação

Foram colhidas 61 amostras num setor de abastecimento para análise dos níveis de cloro residual, obtendo-se os seguintes resultados:

média amostral,  $\bar{x} = 0,30$  ppm

desvio padrão amostral,  $s = 0,20$  ppm

Deseja-se saber o tamanho da amostra para se avaliar a verdadeira média do cloro residual, com uma confiança de 95% e um erro relativo inferior a 30% no valor da média.

Para 95% de confiança e  $\Phi = n-1 = 60-1 = 60$  graus de liberdade temos,  $t = 2$ .

$$Cv = \frac{s}{\bar{x}} = \frac{0,20}{0,30} \cong 0,67$$

$e = 0,30$

$$n = \left[ t \cdot \frac{Cv}{e} \right]^2 = \left[ 2 \cdot \frac{0,67}{0,30} \right]^2 \cong 20$$

O tamanho ideal da amostra é 20.

Observar que quanto maior a dispersão dos valores de cloro, maior será

o tamanho da amostra para um erro relativo fixado.

Posto o que acabamos de ver, foi estabelecido um plano-piloto para dimensionamento do tamanho da amostra de cada um dos 58 setores de abastecimento, levando-se em conta primeiramente o parâmetro cloro residual.

A metodologia adotada foi a seguinte:

(1) Foi considerado um histórico de aproximadamente 4 anos, de todas as amostragens realizadas em cada setor de abastecimento;

(2) Admitimos como média e desvio padrão amostral a média das médias e dos desvios-padrões de todas as amostragens compreendidas no período de 1982 a 1985;

(3) Devido à dificuldade de se levantar o número de pontos amostrados em cada setor de abastecimento no período mencionado, considerou-se para efeito do dimensionamento do tamanho da amostra a estatística  $Z$  em substituição à estatística  $t$ . Embora teoricamente a distribuição  $t$  seja sempre correta quando não se conhece o desvio padrão da população, independentemente do tamanho da amostra sabe-se que para grandes amostras é razoável usar valores  $Z$  para aproximar valores  $t$ ;

(4) Foi fixada em 95% a confiança desejada e portanto  $Z = 1,96$  (tabela da distribuição normal);

(5) Foi fixado em erro relativo inferior a 30% no valor da média.

Os resultados, isto é, os tamanhos das amostras para cada setor de abastecimento, foram listados na tabela A, coluna 1.

Os tamanhos obtidos constituíram, portanto, o ponto de partida para a amostragem destes setores. Observamos que, dependendo do erro na estimativa da média do parâmetro em estudo, o tamanho da amostra aumenta ou diminui conforme o erro a ser tolerado. Mesmo com um erro relativo inferior a 30% no valor da média do cloro residual, os tamanhos das amostras obtidos foram consideravelmente elevados e isto se deveu principalmente aos seguintes fatores:

— Alguns setores foram pouco amostrados no período considerado;

— A frequência de amostragem era muito baixa, cada setor em média era amostrado a cada 5 meses e portanto nestes intervalos as variações registradas em termos de desvio padrão eram muito altas;

— Considerando o próprio setor de abastecimento, observamos também elevada variação dos níveis de cloro, provavelmente pela não representatividade dos pontos amostrados. Esta falta de representatividade estaria relacionada à falta de controle na distri-

Tabela A

Nº DE ORDEM	SETOR DE ABASTECIMENTO	TAMANHO DA AMOSTRA (n)	
		COLUNA 1	COLUNA 2
01	Água Branca	75	19
02	Americanópolis	54	16
03	Anel do Brás	52	18
04	Araçá	219	20
05	Arthur Alvim	39	20
06	Avenida	57	10
07	Brooklin	33	20
08	Butantã	46	16
09	Cangaíba	106	10
10	Capão Redondo	95	14
11	Capela do Socorro	48	15
12	Casa Verde	58	10
13	Chácara Flora	47	10
14	Cidade Vargas	64	10
15	Consolação	71	50
16	Ermelino Matarazzo	87	92
17	Freguesia do Ó	58	10
18	Guianazes	108	20
19	Guarau	54	10
20	Interlagos	42	18
21	Ipiranga	94	10
22	Itaim	25	20
23	Itaquera	64	16
24	Jabaquara	45	20
25	Jaçanã	62	10
26	Jaceguava	60	16
27	Jaguare	98	20
28	Jaraguá	64	20
29	Jardim Arpoador	41	20
30	Jardim Popular	51	21
31	Lapa	48	10
32	Mirante	49	10
33	Mombaca	136	78
34	Mooça	27	21
35	Morumbi	49	15
36	Parque Edu Chaves	84	10
37	Penha	63	100
38	Pirajussara	53	20
39	Pirituba	80	10
40	Sacomã	33	19
41	Santana	74	11
42	Santo Amaro	40	10
43	São Miguel	41	19
44	Sapopemba	48	20
45	Shangrilá	23	23
46	Vila Alpina	98	54
47	Vila América	65	10
48	Vila Brasilândia	84	10
49	Vila Deodoro	58	20
50	Vila Formosa	55	20
51	Vila Jaguara	57	10
52	Vila Madalena	81	10
53	Vila Maria	96	10
54	Vila Mariana	47	10
55	Vila Mascote	62	10
56	Vila Medeiros	67	10
57	Vila Nova Cachoeirinha	75	20
58	Vila Romana	59	10

buição dos pontos a serem amostrados e também à interferência de pontos de outros setores.

Surgiu então a necessidade de se aferir o plano de amostragem, implantando-se em campo as plantas dos setores de abastecimento com a marcação dos seus respectivos limites, e se passou a controlar a distribuição dos pontos dentro de cada setor. Em amostragens subsequentes, verificamos que, fixado o mesmo erro relativo (30%), os tamanhos das amostras diminuíram conforme os valores listados na tabela A, coluna 2.

### 3 Sistemática atual do plano de amostragem

Atualmente, o Plano de Amostragem para a rede de distribuição do município de São Paulo, englobando os 58 setores de abastecimento, é fundamentado no controle dos parâmetros cloro residual e turbidez. Como foi mencionado anteriormente, os setores de abastecimento constituem áreas de influência de um ou mais reservatórios, que por sua vez são alimentados por um ou mais efluentes de Estações de Tratamento de Água, de tal forma que estes setores podem ser considerados distintamente. Os 58 setores de abastecimento cobrem praticamente toda a Capital e cada um deles é amostrado com frequência mínima mensal da seguinte forma:

As plantas que constituem os setores são distribuídas às equipes de campo e estas escolhem ao acaso os pontos que deverão ser amostrados, tomando-se cuidado para que a distribuição dos mesmos, dentro do setor, venha a representá-lo da melhor forma possível;

O tamanho da amostra de cada setor é determinado através de computador, de acordo com o procedimento já explicado anteriormente, levando-se sempre em conta os dados da última amostragem. O n, isto é, o tamanho da amostra, é determinado tanto para os resultados de cloro residual como para a turbidez e no final adota-se o maior valor.

Por questão de limitação de recursos pertinentes à operacionalização do plano, o tamanho máximo da amostra a ser considerado é igual a 100. No caso de setores em que se obtêm tamanhos de amostras muito pequenos, o mínimo a ser considerado é igual a 20.

No período de um mês, todos os setores são amostrados com frequência variável, porém, no mínimo com frequência mensal.

Os resultados de cloro residual e de turbidez são planilhados em cor-

respondência aos setores de abastecimento e a rede como um todo. No final, com uso de computador, é dado um tratamento estatístico a estas informações, obtendo-se uma avaliação da qualidade da água em termos de cloro e turbidez para cada setor e para a rede como um todo.

O tratamento estatístico destas informações se baseia no cálculo das probabilidades de atendimento a padrões estabelecidos, tanto para o cloro como para a turbidez.

O cálculo da probabilidade é dado pela seguinte fórmula:

$$Z = \frac{x - \bar{x}}{s}$$

onde:

$Z$  = n.º de desvios padrões a contar da média da variável em estudo função da confiança desejada (no caso 95%).

$x$  = Valor arbitrário, no caso o padrão estabelecido da variável em estudo.

$\bar{x}$  = Média amostral

$s$  = Desvio padrão amostral

A estatística  $Z$  é lida numa tabela normal padronizada fornecendo a área sob a curva normal entre a média e valores obtidos de  $Z$ . Esta área corresponde à probabilidade de um valor cair naquele intervalo.

Do exposto, são tiradas as seguintes informações:

#### Para o cloro residual

$P(x < 0,2)$ : Probabilidade de aparecerem valores de cloro abaixo de 0,2 ppm (caracterizando uma situação anômala, isto é, ausência de cloro).

$P(0,2 \leq x \leq 1,3)$ : Probabilidade de aparecerem valores de cloro na faixa de 0,2 a 1,3 ppm.

$P(x > 1,3)$ : Probabilidade de aparecerem valores de cloro acima de 1,3 ppm.

Com relação à turbidez segue-se o mesmo raciocínio, variando apenas os padrões estabelecidos.

#### Para a turbidez

$P(x < 1,0)$ : Probabilidade de aparecerem valores de turbidez abaixo de 1,0 NTU.

$P(1,0 \leq x \leq 5,0)$ : Probabilidade de aparecerem valores de turbidez na faixa de 1,0 a 5,0 NTU.

Tabela B

OR-DEM	SETOR	CLORO RESIDUAL (ppm)					TURBIDEZ (NTU)						
		PROBABILIDADES			MÉDIA	DES-VIO PADRÃO	n	PROBABILIDADES			MÉDIA	DES-VIO PADRÃO	n
		<0,2	0,2-1,3	>1,3				<1	1-5	>5			
1	Água Branca	0,13	99,82	0,00	0,46	0,09	20	23,00	76,93	0,06	1,75	1,01	20
2	Americanópolis	3,91	96,09	0,00	0,37	0,10	20	97,95	2,05	0,00	0,56	0,21	20
3	Anel do Brás	0,03	99,97	0,00	0,61	0,12	40	78,63	21,37	0,00	0,77	0,29	40
4	Araça	0,91	99,09	0,00	0,39	0,08	20	43,78	56,22	0,00	1,09	0,55	20
53	Vila Romano	0,13	99,57	0,00	0,29	0,07	10	1,00	0,00	0,00	0,55	0,00	10
REDE		5,57	94,43	0,00	0,65	0,23	1015	50,40	41,60	0,00	0,33	0,52	1015

$P(x > 5,0)$ : Probabilidade de aparecerem valores de turbidez acima de 5,0 NTU (caracterizando uma situação anômala, isto é, água suja).

Além dessas informações, são fornecidas para cada setor e para a rede, como um todo, a média, o desvio padrão e o tamanho da amostra coletada. Ver tabela B.

Algumas informações são plotadas em gráfico com o fim de se ter melhor visualização de seu comportamento ao longo do ano.

Para efeito de operacionalização do plano, adotou-se o seguinte critério: Um setor é considerado problemático quando a sua probabilidade em termos de ausência de cloro ( $< 0,20$  ppm) e/ou turbidez ( $> 5,0$  NTU) for superior a 10%. Dessa forma, são tomadas medidas corretivas, isto é, são acionadas descargas na rede e em alguns casos reforço na cloração, passando o setor a ter uma frequência de amostragem prioritária. O controle de potabilidade bacteriológica é feito somente nos casos onde se caracteriza ausência de cloro e as reclamações relativas a doenças hidrotansmissíveis são atendidas no local da reclamação gerando inspeções sanitárias cujos resultados são repassados à Secretaria da Saúde.

## 4 Conclusões

A sistemática de controle de qualidade da água de rede adotada pelo Departamento de Controle Sanitário favoreceu benefícios significantes com relação aos seguintes aspectos:

#### (a) Dimensionamento do tamanho da amostra

Os benefícios alcançados se devem principalmente à frequência de amostragem dos setores e à adoção de critérios mais científicos com relação ao número de pontos a ser amostrado, consequentemente obtendo-se maior confiabilidade no plano de amostragem.

#### (b) Controle da distribuição dos pontos amostrados

A marcação dos limites dos setores em planta, bem como o controle da distribuição dos mesmos, contribuíram significativamente para maior representatividade das informações obtidas.

#### (c) Tratamento estatístico dos dados

O controle dos níveis de cloro e de turbidez, que era feito com base no atendimento de reclamações recebidas, ganhou maior importância quando se passou a estudar as populações setoriais através de suas respectivas amostras.

## 5 Bibliografia

- DIXON, W. J. & MASSEY F. J. *Introduction to Statistical Analysis*, New York, MacGraw, 1978.
- LEVIN JACK, *Estatística Aplicada a Ciências Humanas*, Harper & Row do Brasil Ltda., 1978.
- STEVENSON, WILLIAN J., *Estatística Aplicada à Administração*, Harper & Row do Brasil Ltda., 1981.
- COSTA NETO, PEDRO LUIZ DE OLIVEIRA, *Estatística*, São Paulo, Egdard Blücher, 1977.