UM MODELO DE AMOSTRAGEM PARA O CONTROLE DA POTABILIDADE DE UM SISTEMA DE DISTRIBUIÇÃO DE ÁGUA

Engº ANTONIO CARLOS M. MATTOS (*)

SUMÁRIO

a) INTRODUÇÃO

Um sistema de controle da potabilidade da água em uma rede de distribuição deve ser constituído de cinco fases:

1) Coleta das amostras de água da rede, de acordo com algum critério científico.

2) Exames e análises físicos, quimicos e bacteriológicos das amostras.

3) Análise dos resultados, para determinar as amostras fora dos limites de potabilidade.

4) Inspeção das áreas onde foram coletadas as amostras contaminadas, para a determinação das causas da contaminação.

5) Estabelecimento de ações corretivas na rede de distribuição, para eliminar as causas da contaminação.

A primeira fase, onde é executada um plano de amostragem, é uma das mais importantes do Sistema de Controle, pois dela depende a confiabilidade de todo o Sistema. Um controle exercido a partir de um conjunto de amostras não suficientemente representativas da rede hidráulica, é um controle parcial e falho que não deve ser mantido.

O estabelecimento de um plano de amostragem confiável deve ser resultado de um estudo científico do problema, e do desenvolvimento de um modelo sustentável teoricamente e de fácil aplicação. E é exatamente isto que o presente trabalho se propõe a realizar.

(*) Engenheiro Eletricista (EPUSP), Pós-graduado em Administração de Empresas (FGV), Analista de Sistemas (IRIA – França), Professor da Fundação Getúlio Vargas (S. Paulo), Consultor da CETESB.

b) HIPÓTESES FUNDAMENTAIS DO MODELO DE AMOSTRAGEM

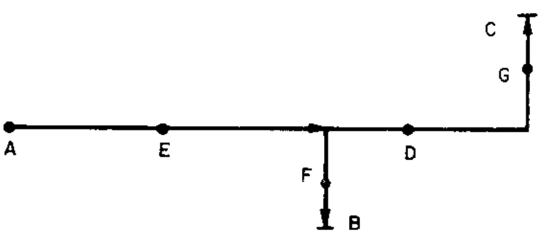
Qualquer modelo é sempre baseado em um conjunto de hipóteses que lhe garantem a validade. O não cumprimento na prática de uma dessas hipóteses, pode tornar o modelo completamente inadequado aos fins a que se destina.

Por essa razão, é importante destacar, inicialmente, que fatos serão admitidos como verdadeiros e reais.

— — — —

A rede urbana de distribuição de água possui infinitos pontos, onde pode ser coletada uma amostra para ser analisada, pois é contínua. No entanto, sendo uma rede **orientada**, ou seja, o sentido de percurso da água é único em cada instante, não é necessário considerar todos esses infinitos pontos como passíveis de serem amostrados; basta apenas um número limitado, para que se possa ter um conhecimento **total** do estado de potabilidade da água na rede.

Por exemplo, no sistema ABC abaixo, que contém infinitos pontos, apenas dois



são suficientes para se determinar o seu estado: os pontos de fim de linha B e C. Com efeito, qualquer contaminação introduzida na rede pelos pontos D, E, F ou G implicará na contaminação dos pontos B e/ou C. Por outro lado, se os pontos B e C não estiverem contaminados, então nenhum ponto da rede estará contaminado.

É claro que a determinação dos pontos que gozam da propriedade acima implica em um estudo detalhado das plantas da rede, pois inexistem regras gerais: cada trecho é um caso particular.

Esse conjunto de pontos significativos será chamado de **Cadastro dos Pontos de Amostragem.**

Podemos, então, estabelecer a

Hipótese n.º 1: «Existe um conjunto limitado de pontos da rede (Cadastro dos Pontos de Amostragem) tais que:

a) Se nenhum desses pontos estiver contaminado, então nenhum ponto da rede estará contaminado; e

b) Se pelo menos um desses pontos estiver contaminado, então pelo menos um ponto da rede estará contaminado».

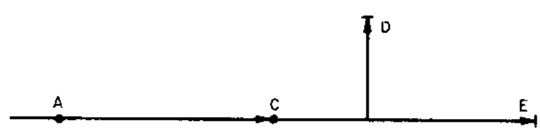
As condições da rede não são estáticas, mas estão em contínua mudança, ou devido às variações de consumo de água dos habitantes, ou às manobras da rede, ou às interrupções de fornecimento em alguns trechos. Em qualquer desses casos, a Hipótese 1 deve continuar válida.

Assim, é importante que os técnicos que forem realizar o cadastramento dos pontos, tenham em mente as características dinâmicas da rede, para evitar que, sob determinadas condições, algum trecho da rede fique fora de controle.

Chegamos, assim, à

Hipótese n.º 2: «A Hipótese n.º 1 deve ser válida quaisquer que sejam as condições hidráulicas da rede».

Consideremos o trecho mostrado abaixo:



Se, em um ponto A, houver introdução de material contaminado, a água que atingir esse ponto ficará também contaminada, mas apenas durante certo tempo, por exemplo duas horas, devido à ação do cloro. Supondo que, dada a vazão existente na linha, a água leve duas horas para atingir o ponto C, pode-se concluir que apenas o trecho AC estará contaminado. Ora, se os pontos de amostragem forem D e E, o trecho AC estará totalmente fora de controle. Para evitar tal eventualidade, o ponto C também deverá ser cadastrado, e a distância AC será a distância mínima entre dois pontos consecutivos (cerca de três quarteirões, v.g.).

Em outras palavras:

Hipótese n. 3: «Em qualquer linha da rede, dois pontos cadastrados não devem estar separados por uma distância superior a um certo limite, medida sobre essa linha».

Um sistema de controle de potabilidade ideal seria aquele em que existisse, em cada ponto cadastrado, um sensor eletrônico que estaria analisando continuamente a qualidade da água passante por esse ponto, e que emitisse um sinal de alarme e um sinal de controle a um computador central, tão logo fosse detetada uma irregularidade (Sistema de Controle em Tempo Real). (*) Como tal sistema ainda não se acha implantado, seja por razões de custo, seja devido ao estágio tecnológico atual, um trecho da rede pode permanecer contaminado, no mínimo, 48 horas. Isto porque a realização das cinco fases do sistema de controle (V. Introd.) não é instantânea, levando esse tempo para serem concluídas.

Isso nos leva à:

Hipótese n.° 4: «Em não havendo um Sistema de Controle em Tempo-Real, qualquer ponto da rede pode permanecer contaminado durante pelo menos 48 horas, salvo em casos acidentais e raros».

Desconsiderando a existência de um sistema em tempo-real, a limitação de recursos a serem aplicados em um plano de amostragem irá implicar em um intervalo de tempo máximo de contaminação. Assim, por exemplo, se tivermos um total de 1.000 pontos cadastrados em uma cidade, e a limitação de verbas permitir uma amostra diária de, no máximo, 200 pontos, serão necessários 5 dias para que todos os 1.000 pontos sejam analisados (amostra aleatória sem reposição). Ou seja, um ponto pode permanecer contaminado durante 7 dias no máximo (Cf. Hipót. 4).

Desse modo, podemos estabelecer a

Hipótese n.° 5: «Qualquer ponto da rede pode permanecer contaminado durante um certo tempo máximo, que será tanto maior quanto menor for a disponibilidade de recursos financeiros para o processo de amostragem».

(*) Um tal sistema pode ser encontrado em «Hidrologic Information Systems», Bibliografia nº 2.

O controle da potabilidade da água é um problema de grande responsabilidade, que envolve uma série de implicações de saúde pública. Assim, como decorrência da Hipótese n.° 5 acima, seria de resultados imprevisíveis a permanência de pontos contaminados durante, digamos, 7 dias, como por exemplo pontos de entrada de água em hospitais, em escolas, em zonas de alto potencial epidemiológico, ou até mesmo em algumas residências. Deste modo, alguns pontos da rede serão de maior responsabilidade social que outros, devendo, portanto, ter uma freqüência de amostragem maior. Tais pontos serão chamados de **críticos.**

Por outro lado, existem pontos da rede onde a probabilidade de contaminação é maior que nos demais. Como exemplos, temos os pontos cuja pressão é muito variável, os de pressão inferior à atmosférica local, os com baixo teor de cloro residual livre (abaixo de 0,2 ppm), e os pertencentes a zonas de grande probabilidade de contaminação.

Tais pontos, que serão denominados **notáveis,** deverão, igualmente, ter uma freqüência de amostragem maior.

Por exclusão, os demais pontos serão chamados de **genéricos,** como por exemplo pontos de fim de linha, início de linha, bifurcações, etc.

Podemos resumir o acima exposto na

Hipótese n.° 6: «Os pontos críticos e notáveis da rede serão amostrados com uma freqüência maior que os pontos genéricos».

Uma vez constatado um ponto contaminado, torna-se necessário determinar as causas dessa irregularidade (Fase 4 da «Introdução»). Tal determinação implicará na amostragem de pontos **adicionais** da rede, de acordo com a experiência do responsável pelo Sistema de Controle. Ora, como existe uma limitação prática, que não permite o aumento indefinido de pontos por dia, é necessário prever uma «reserva» de pontos para tais eventualidades. Esses pontos serão chamados de **suspeitos.**

Temos, então, a

Hipótese n.º 7: «Os pontos suspeitos, amostrados diariamente, deverão ser determinados pela pessoa responsável pelo Sistema de Controle».

Desse modo, cada amostra diária será constituída de quatro estratos: um para os pontos críticos, outro para os notáveis, um terceiro para os genéricos, e um quarto para os suspeitos.

c) O MODELO DE AMOSTRAGEM

Uma vez estabelecidas as hipóteses fundamentais do plano de amostragem, torna-se necessário construir um modelo **operacional** que, satisfazendo às hipóteses, possa se constituir em uma rotina **prática.** Isto é o que será feito a seguir.

c 1) Composição da Amostra

O universo estatístico a ser estudado é o constituído pelo Cadastro dos Pontos de Amostragem. Seja N esse número de pontos.

Uma amostra **diária** (ou semanal, ou horária, etc.) será composta de quatro estratos:

n_c pontos críticos
n_n pontos notáveis
n_g pontos genéricos
n_s pontos suspeitos

sendo L o limite diário de pontos amostrados (obtido a partir de considerações econômicas, como veremos):

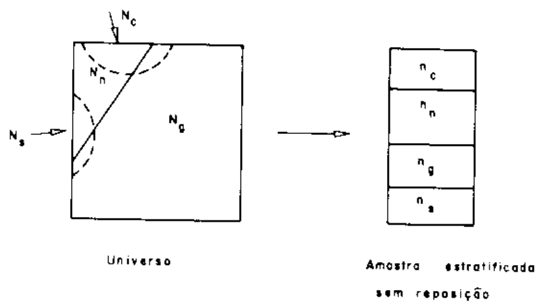
$$\boxed{L = n_c + n_n + n_g + n_s}$$

Por outro lado, sendo N_n o total dos pontos notáveis do universo estatístico, e N_g o total dos genéricos,

$$N = N_n + N_g$$

O esquema abaixo ilustra o processo de composição da amostra diária, onde N_c é o total dos pontos críticos e N_s o de suspeitos.

Notemos que a intersecção do conjunto dos pontos notáveis e genéricos é vazia, e que sua reunião é o universo.



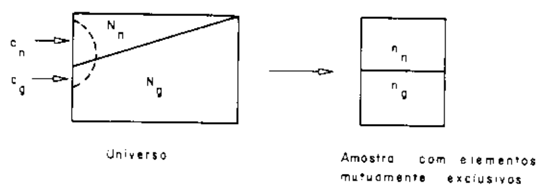
c 2) Confiabilidade do modelo

A confiabilidade de um sistema vem a ser a probabilidade desse sistema não falhar. (*) Em termos do presente modelo, sua confiabilidade R (de «reliability») será definida como «a probabilidade de se ter, em uma amostra, todos os pontos que estejam contaminados naquele instante, e que tenham sido cadastrados.»

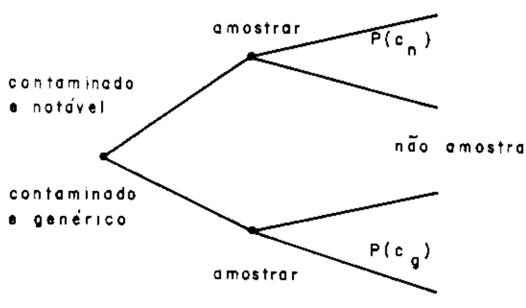
Assim, se em um certo dia, cinco pontos cadastrados estiverem contaminados, e se a probabilidade da amostra conter esses cinco pontos for de 60%, então a confiabilidade será de 60%.

Para determinar a expressão geral da confiabilidade do modelo em estudo, utilizamos as propriedades do «modelo de amostragem aleatória sem reposição.(**)

Dividindo o universo dos pontos em notáveis e genéricos, denominaremos de c_n o total dos pontos contaminados **e** notáveis (valores médios históricos), e de c_g o total dos contaminados **e** genéricos, como indicado no esquema abaixo:



Construimos, a seguir, a árvore das probabilidades para o caso em estudo:



(*) V., por ex., Machol (Bibl. 1), pg. 33-2.
(**) Este modelo se encontra no Apêndice.

As probabilidades associadas aos ramos da árvore são:

Probabilidade de	Valor
um ponto contaminado ser notável	$\frac{c_n}{c_n + c_g}$
um ponto contaminado ser genérico	$\frac{c_g}{c_n + c_g}$
serem amostrados os c_n pontos contaminados	$P(c_n) = \left(\frac{n_n}{N_n}\right)^{c_n}$
serem amostrados os c_g pontos contaminados	$P(c_g) = \left(\frac{n_g}{N_g}\right)^{c_g}$

Assim, a probabilidade de se amostrar todos os pontos contaminados será dado por:

$$R = \left(\frac{c_n}{c_n + c_g}\right)\left(\frac{n_n}{N_n}\right)^{c_n} + \left(\frac{c_g}{c_n + c_g}\right)\left(\frac{n_g}{N_g}\right)^{c_g}$$

Notemos que se $n_n = n_g = 0$, i.e., não existir amostra, a confiabilidade será zero, e se a amostra contiver todos os N pontos do universo estatístico, i.e., se $n_n = N_n$ e $n_g = N_g$, então a confiabilidade será total (100%). Por outro lado, se existir apenas um ponto contaminado e notável ($c_g = 0$ e $c_n = 1$), então $R = n_n/N_n$, e a confiabilidade será igual ao tamanho relativo da amostra dos pontos notáveis.

Observemos, também, que à medida que as amostras sem reposição vão sendo realizadas, a confiabilidade **dessas** amostras irá aumentando, até atingir 100%, ao fim de um período T_{max} (V. ítem c 3). Assim, a expressão acima pode ser interpretada como sendo a confiabilidade mínima do modelo, ou então a sua confiabilidade diária (semanal, etc.).

c 3) Períodos de amostragem dos pontos

Um ponto qualquer da rede não deve permanecer contaminado por um período maior que T_{max}. Ora, como a amostragem é aleatória **sem** reposição, isso implica em que todos os pontos devem ser amostrados durante esse período. Chamando

T_c = período de amostragem dos pontos críticos

T_g = idem, genéricos

T_n = ibidem, notáveis

teremos:

$T_c = \frac{N_c}{n_c}$	$T_n = \frac{N_n}{n_n}$	$T_g = \frac{N_g}{n_g}$

Estaremos supondo, ainda, que

$$T_{min} \leqslant T_c \leqslant T_n \leqslant T_g = T_{max}$$

que é a formalização das hipóteses 4, 5 e 6 (Cf. ítem b).

Desse modo, cada ponto crítico será amostrado uma vez a cada T_c dias, um notável a cada T_n dias, e um genérico a cada T_g dias. As freqüências de amostragem são dadas pelos inversos dos períodos.

c 4) Dimensionamento da amostra

Como já foi mencionado (V. ítem c 1), o tamanho global L da amostra é decorrência de considerações econômicas. Veremos, agora, como estabelecer tal relação.

A cada ponto amostrado estão associados os seguintes custos:

p_c = custo para coletar uma amostra de água em um ponto

p_b = custo de um exame bacteriológico em uma amostra

p_s = custo de uma inspeção sanitária em um local.

A amostragem de $L = n_c + n_n + n_g + n_s$ pontos por dia custará, então:

$$C = L \cdot (p_c + p_b) + (c_g + c_n) \cdot p_s$$

Daí poderemos obter o tamanho L da amostra, uma vez fixados os preços cobrados pelos exames e análises e o custo máximo que se pretende aceitar.

c 5) Dimensionamento dos estratos da amostra

Dado um custo máximo C, o tamanho L da amostra será obtido a partir da expressão acima.

A seguir, estabelecendo-se um valor para a confiabilidade R, tem-se uma relação entre n_n e n_g.

Ficam então determinados os tamanhos dos estratos dos pontos notáveis e genéricos, e respectivos períodos.

Podemos, também, estabelecer um valor para $T_{máx}$, calculando, então, n_n e n_g; neste caso, a confiabilidade será uma decorrência.

Tais considerações serão a seguir aplicadas a um caso real, onde os parâmetros serão submetidos a uma análise conjunta.

d) APLICAÇÕES PRÁTICAS DO MODELO

d 1) Controle de potabilidade em uma cidade de 1.000.000 de habitantes

Suponhamos que os preços cobrados sejam da ordem de

p_1 = Cr$ 6,00 — custo da coleta de uma amostra

p_2 = Cr$ 30,00 — custo de um exame bacteriológico

p_3 = Cr$ 50,00 — custo de uma inspeção sanitária em uma área.

Dados históricos fornecem os seguintes valores médios:

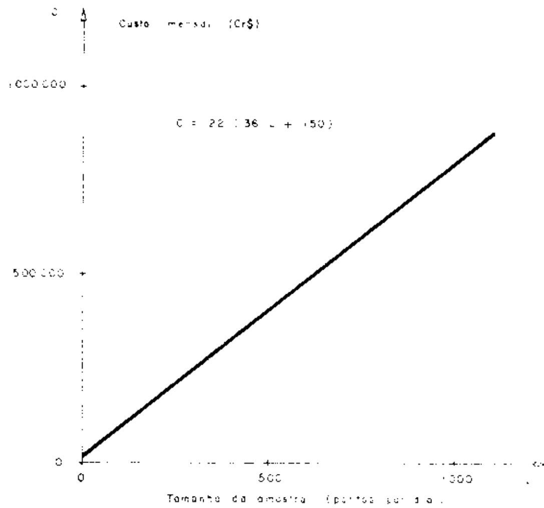
c_g = 1 ponto genérico contaminado por dia na rede

c_n = 2 pontos notáveis contaminados por dia na rede

O custo diário para amostrar L pontos e inspecionar três áreas suspeitas será, então:

$$C = 36 \cdot L + 150$$

O gráfico abaixo reflete a função acima:



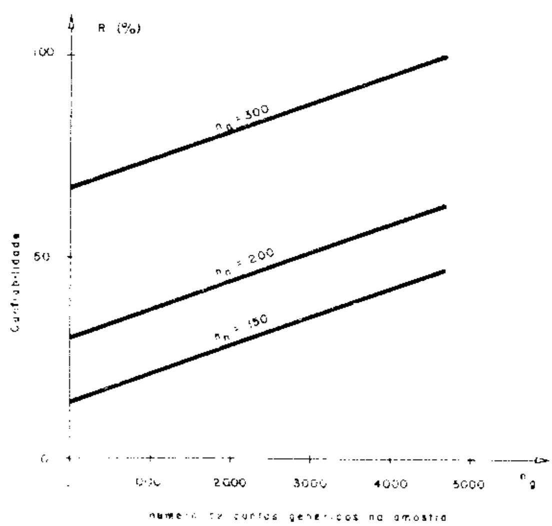
Para o cálculo da confiabilidade temos os valores

N_n	=	300 pontos notáveis na rede
N_g	=	4700 pontos genéricos
N	=	5000 pontos cadastrados

Temos daí que:

$$R = \frac{2}{3} \cdot \left(\frac{n_n}{300}\right) + \frac{1}{3} \cdot \left(\frac{n_g}{4700}\right)$$

Essa função está representada abaixo:



Para o cálculo dos periodos, temos:

N_c = 50 pontos criticos

T'_c = 7 dias

n_s = 5 pontos suspeitos por dia na rede.

O valor de T_c corrigido é obtido multiplicando o valor T'_c por 22/30, já que não há trabalho durante 8 dias por mês (no mínimo), e subtraindo-se 2 dias, que é o tempo mínimo decorrido entre a coleta da amostra e a obtenção dos resultados. Assim:

$$T_c = 22.7/30 - 2 \cong 3 \text{ dias}$$

e, portanto,

$$n_c = 50/3 \cong 17 \text{ pontos por dia}$$

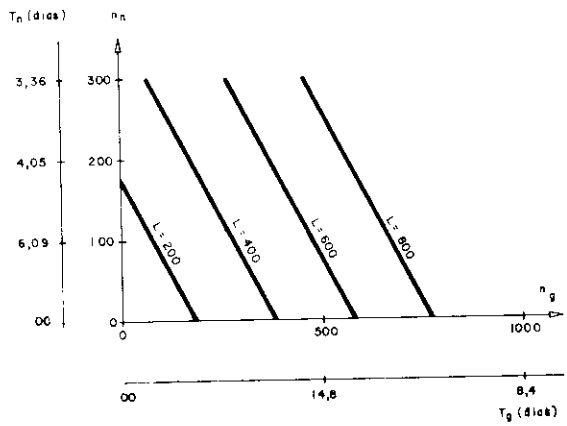
Como

$$L = n_c + n_s + n_n + n_g$$

vem

$$n_g = L - 22 - n_n$$

Abaixo se encontram esses pontos, com os respectivos periodos:



As considerações anteriores podem ser resumidas em um único gráfico, que poderá ser usado para **decidir** quais os valores de n_n e n_g que serão utilizados.

As relações usadas na construção desse gráfico foram as seguintes:

$$C = 22 \;.\; (36 \;.\; L + 150)$$

C: Cr$ L: pontos

$$n_g = L - (22 + n_n)$$

$$n_g = 141 \;.\; R - 0{,}1044444 \;.\; n_n^2$$

R: %

$$T_n = 2 + 409/n_n$$

T_n: dias

$$T_g = 2 + 6409/n_g$$

T_g: dias

Com base nesse gráfico (V. abaixo), podemos, agora, realizar uma série de considerações em torno do custo, períodos e confiabilidade, até localizar um **ponto de trabalho,** i.e., os parâmetros do modelo a serem utilizados.

Admitamos que tenha sido escolhido o ponto de trabalho

$\boxed{n_g = 200}$ e $\boxed{n_n = 280}$

Consequentemente, o modelo terá as seguintes características:

a) A amostra diária será constituída de 502 pontos:

$$L = 200 + 22 + 280 = 502$$

b) O custo mensal desse plano de controle será de Cr$ 400.884,00:

$$C = 22 \;.\; (36.502 + 150) = 400.884{,}00$$

c) A confiabilidade será de 59,5%:

$$R = 2 \;.\; n_n^2/270.000 + n_g/14.100 = 0{,}581 + 0{,}014 = 59{,}5\%.$$

Isto significa que para cada 100 amostras realizadas, em 59,5% estarão **todos** os três pontos contaminados (2 notáveis e 1 genérico). Em outras palavras, a cada 100 dias, a rede estará sob total controle durante 59,5 dias.

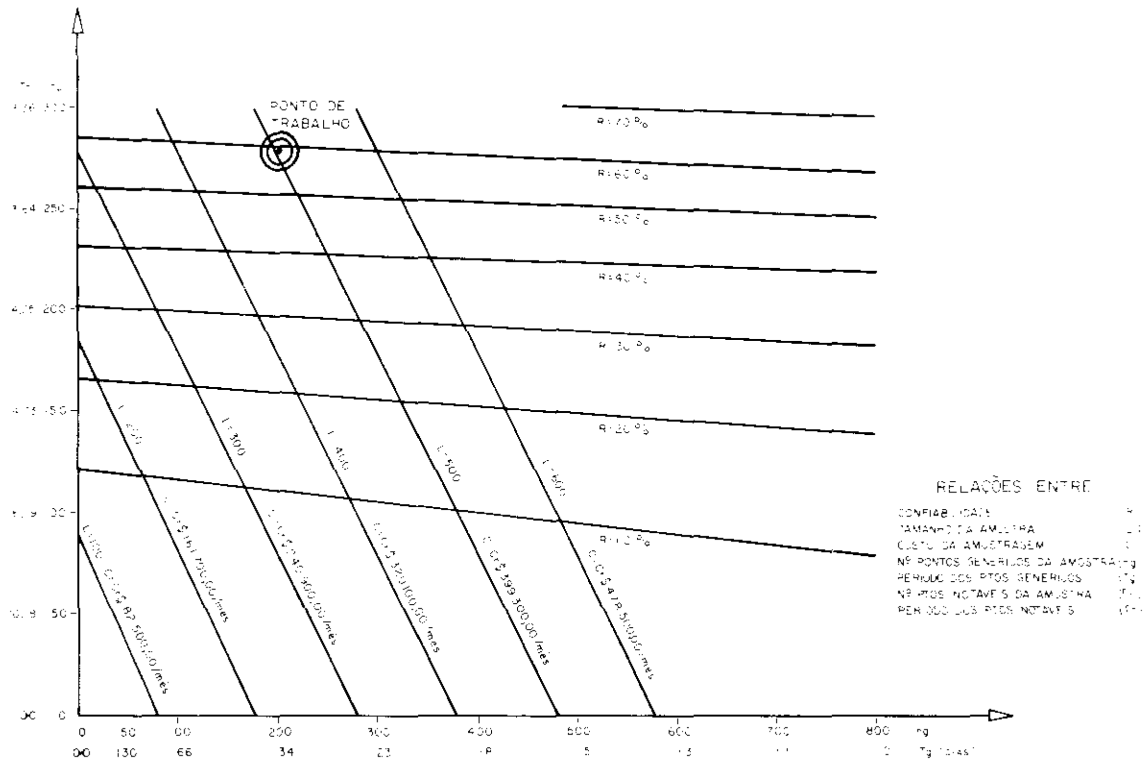
d) O período necessário para serem cobertos todos os pontos genéricos será de 34 dias:

$$T_g = 2 + 6.409/200 = 34$$

Assim, um grupo de pessoas poderá ingerir água contaminada durante, no máximo, 34 dias, que é o tempo decorrido entre duas amostragens sucessivas de um mesmo ponto. Entretanto, como a amostragem é aleatória, o período médio será de 18 dias (1 + 34/2); além do mais, como existe um ponto genérico contaminado para cada 3 contaminados, a probabilidade de ocorrência desses períodos é de 33,3%.

e) O período necessário para serem cobertos todos os pontos notáveis é de 3,46 dias:

$T_n = 2 + 409/208 = 3,46$ dias



sendo 2 dias o tempo necessário para a análise da amostra, e 0,39 dias a correção devida à existência de 22 dias de trabalho por mês; se fizéssemos 3,46 — 2 — 0,39 = 1,07, teríamos o número de dias mínimo para a detetação de um ponto contaminado, em condições ideais (300/280 = 1,07).

Assim, em 66,7% dos casos de contaminação na rede, o ponto será detetado em 3,46 dias no máximo, e em 2,73 dias em média (a amostra é aleatória, e 2,73 = = 1 + 3,46/2).

Esta é uma das grandes vantagens da estratificação da amostra: com apenas 6% dos pontos (300/5.000) conseguimos controlar 66,7% dos casos de contaminação em apenas 2,73 dias! É uma solução essencialmente econômica.

f) A composição diária da amostra será de

17	pontos críticos
280	pontos notáveis
200	pontos genéricos
5	pontos suspeitos
502	pontos por amostra diária

g) Quanto aos 5.000 pontos cadastrados, eles deverão ter sido obtidos de modo a serem satisfeitas as Hipóteses 1, 2 e 3 (V. ítem b) com um **mínimo** de pontos. Deverá, igualmente, estar indicado se o ponto é crítico, notável ou suspeito, além de sua localização física (rua, n.°, bairro, etc.). É importante que, ao ser feito o levantamento, esteja bem claro que os pontos serão amostrados com **freqüências diferentes e variáveis,** e a satisfação das Hipóteses 1, 2 e 3 deverá ser a única preocupação; desse modo, o cadastramento não será constituído de mero «espalhamento» ao acaso de pontos em um mapa.

Um critério prático para verificar se o cadastramento está correto é o seguinte: 1) escolher um ponto ao acaso na rede supondo que esse ponto se contamine, então pelo menos um ponto já cadastrado deverá também se contaminar (notar que a rede poderá sofrer manobras, e que após ter percorrido 3 ou 4 quarteirões a água poderá se descontaminar). 2) observar se existem pontos cadastrados que podem ser suprimidos sem que se altere a condição 1 acima.

Embora ao serem cadastrados, os pontos tenham sido classificados em críticos, notáveis e genéricos, nada impede que seja mudada sua classificação com o tempo; aliás, isto **deve** ser feito. O melhor critério de classificação são os resultados históricos. Por exemplo, podemos estabelecer que todos os pontos que tenham se contaminado por mais de três vezes nos últimos seis meses, serão pontos notáveis. Ou então, os 300 pontos de maior incidência nos últimos quatro meses serão notáveis. É claro que, a cada redefinição, os cálculos anteriores deverão ser refeitos, para que o modelo funcione bem próximo da realidade, isto é, seja **dinâmico.**

h) Veremos, agora, quando e quais pontos serão amostrados, ou seja, o programa de amostragem. Destaquemos, logo de início, que, dado o volume de dados e cálculos, esta programação é melhor realizada por computador, embora nada impeça que seja feita manualmente.

A partir dos valores

$$n = 17 \qquad N = 50$$

$$n = 280 \qquad N = 300$$

$$n = 200 \qquad N = 4700$$

iremos construir um plano de amostragem para um determinado número de dias. Admitamos que seja para 100 dias.

O cadastro dos pontos será dado pelos vetores (NP = n.° do ponto):

CRIT(NP) = código do ponto crítico, com NP = 1, 2, 3, ..., 50

NOT(NP) = código do ponto notável, com NP = 1, 2, 3, ..., 300

GEN(NP) = código do ponto genérico, com NP = 1, 2, ..., 4700

Notemos que os códigos de NOT, GEN e CRIT devem ser diferentes entre si, isto é (*)

$$(CRIT \cap NOT) = (CRIT \cap GEN) = (NOT \cap GEN) = \emptyset$$

Definamos a tabela abaixo:

ACRIT(NCD,ND) ANOT(NND,ND) AGEN(NGD,ND)

onde

ND = 1, 2, 3, ..., 100 é o n.° de ordem do dia em que serão amostrados os pontos;

ACRIT(NCD,ND) é o número do NCDésimo ponto crítico a ser amostrado no dia ND, com NCD = 1, 2, ..., 17;

ANOT(NND,ND) é o n.° do NNDésimo ponto notável a ser amostrado no dia ND, com NND = 1, 2, ..., 280;

AGEN(NGD,ND) é o n.° do NGDésimo ponto genérico a ser amostrado no dia ND, com NGD = 1, 2, ..., 200.

Isto posto, iniciemos a geração dos pontos pelo método de Monte-Carlo, com distribuição equiprovável em cada estrato. Começando com ACRIT, geramos 50 números ao acaso entre 1 e 50, sem repetição. Como temos 100 dias a 17 pontos por dia, necessitamos de 1.700 pontos gerados, ou seja, essa geração deve ser repetida por 34 ciclos (1.700/50). Com isso a matriz ACRIT(NCD,ND) é preenchida.

O mesmo raciocínio é repetido para ANOT (94 ciclos) e AGEN (5 ciclos).

Com isso obtemos, finalmente, o plano de amostragem, num total de 49.700 pontos, que é o seguinte:

(*) Embora no desenvolvimento do modelo, os pontos críticos estivessem incluídos nos genéricos e/ou notáveis, esta nova imposição, além de não alterar os resultados práticos, é uma medida simplificadora.

Dia	Códigos dos Pontos		
	Criticos	Notáveis	Genéricos
001	CRIT(ACRIT(1,1)) CRIT(ACRIT(2,1)) . . . . . . CRIT(ACRIT(17,1))	NOT(ANOT(1,1)) NOT(ANOT(2,1)) . . . . NOT(ANOT(280,1))	GEN(AGEN(1,1)) GEN(AGEN(2,1)) . . . . GEN(AGEN(200,1))
002 . . .	CRIT(ACRIT(1,2))	NOT(ANOT(1,2))	GEN(AGEN(200,1)) . . . . . .
100	CRIT(ACRIT(1,100)) . . . . . . CRIT(ACRIT(17,100)	NOT(ANOT(1,100)) . . . . NOT(ANOT(280,100))	GEN(AGEN(1,100)) . . . . GEN(AGEN(200,100))

Com esse plano em mãos, a **ordem** dos pontos em **cada** dia pode ser alterada, tendo em vista uma otimização dos percursos a serem seguidos para a coleta das amostras. Também deverão ser incluídos os pontos suspeitos, escolhidos pelo responsável pela coleta; se esses pontos já constarem do plano acima, deverão ser eliminados, para evitar repetições desnecessárias.

d 2) Controle na rede de uma cidade de 200.000 habitantes

Neste ítem, limitar-nos-emos a apresentar apenas os resultados, suprimindo a maioria dos comentários, já que a sistemática é a mesma que a do exemplo anterior.

Os dados históricos médios indicam:

$$c_g = 0,25 \qquad c_n = 0,75$$

num total de 1 ponto contaminado por dia, na cidade.

O custo mensal é

$$C = 22 \ . \ (36 \ . \ L + 50) \qquad C: Cr\$$$

Sendo

$$N_n = 25$$
$$N_g = 475$$
$$N = 500 \text{ pontos cadastrados}$$

a confiabilidade se torna

$$R = 0,75 \left(\frac{n_n}{25}\right)^{0,75} + 0,25 \left(\frac{n_g}{475}\right)^{0,25}$$

Admitiremos

$$N_c = 0 \quad e \quad n_s = 4$$

donde

$$n_g = L - 4 - n_n$$

Para a construção do «gráfico decisório», usaremos as relações:

$$C = 22 \ . \ (36 \ . \ L + 50)$$

$$n_g = L - (4 + n_n)$$

$$n_g = 475 \ . \left| 4 \ . \ R - 3 \left(\frac{n_n}{25}\right)^{0,75} \right|^4$$

$$T_n = 2 + 34/n_n$$

$$T_g = 2 + 648/n_g$$

Como «ponto de trabalho», escolhemos o ponto (V. gráf. abaixo)

$$\boxed{n_g = 16} \quad e \quad \boxed{n_n = 19}$$

com as seguintes características:

L = 16 + 19 + 5 = 40
pontos amostrados por dia

C = Cr$ 32.780,00
custo mensal do plano

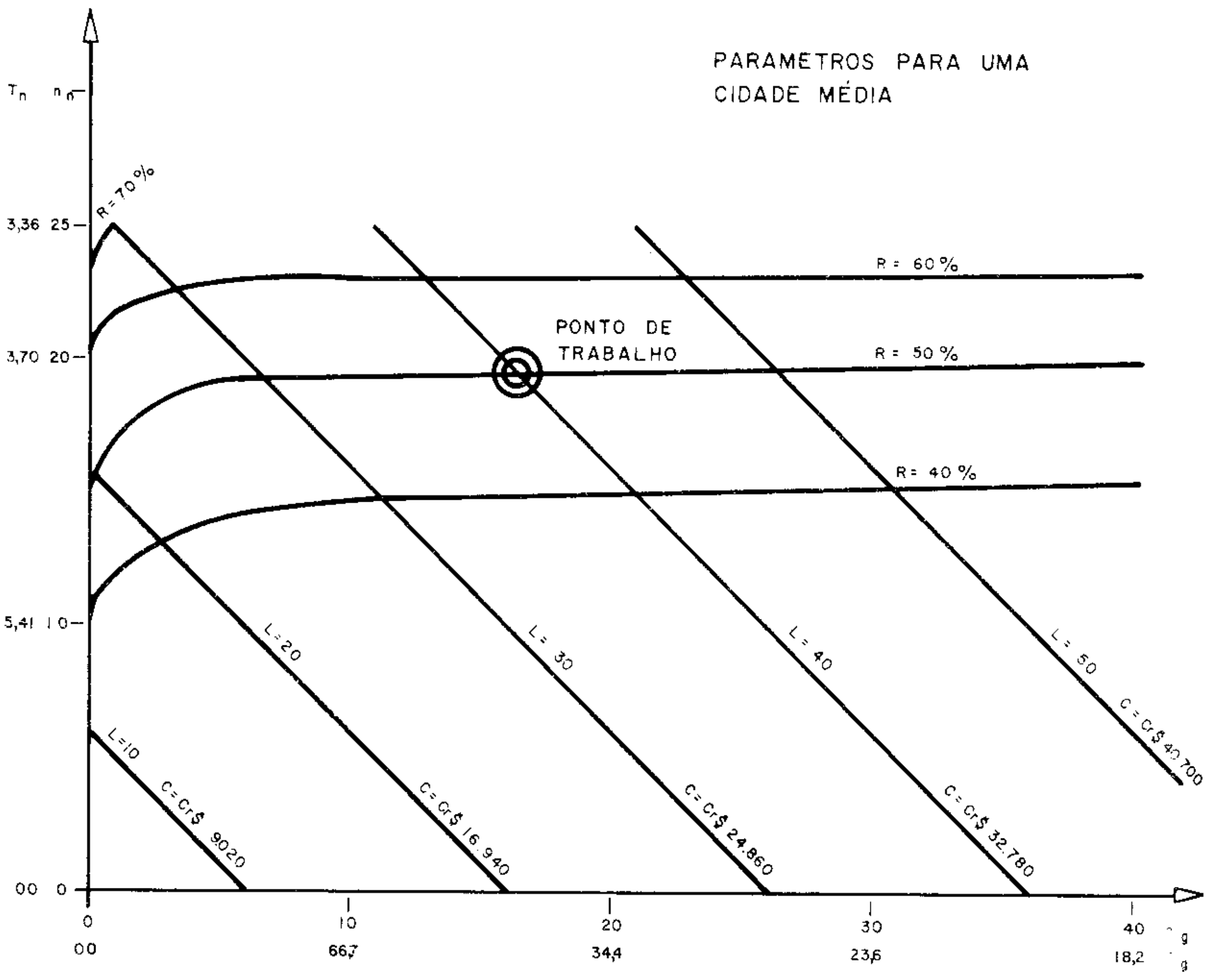
R = 50%
confiabilidade

T. = 2 + 648/16 = 42,5
dias

T.. = 2 + 34/19 = 3,79
dias

A composição da amostra será, então

0	pontos críticos
19	pontos notáveis
16	pontos genéricos
5	pontos suspeitos
40	pontos amostrados por dia



e) APÊNDICE: O MODELO DA AMOSTRAGEM SEM REPOSIÇÃO(*)

Consideremos um universo estatístico constituído dos cinco pontos A, B, C, D e E (n = 5).

Se tomarmos uma amostra de dois pontos (a = 2), teremos as possibilidades seguintes:

AB, AC, AD, AE, BC, BD, BE, CD, CE e DE,

com um total de 10 amostras; assim

$$\binom{n}{a} = 10$$

Por outro lado, se estivermos interessados em ter um (ou mais) determinado ponto na amostra (c = 1), como por exemplo o ponto A, deveremos ter uma das amostras

(*) V. bibliogr. 4 (Feller), pg. 59.

AB, ou AC, ou AD, ou AE,
num total de 4 amostras possíveis, dado pela relação

$$\binom{n-c}{a-c} = 4$$

Assim, a probabilidade de, com uma amostra de **a** pontos, tomados aleatoriamente em **n** pontos, se obter **c** pontos dados, será

$$P = \frac{\binom{n-c}{a-c}}{\binom{n}{a}} \quad \text{com} \quad n \geqslant a \geqslant c$$

que, no exemplo acima é de 4/10 = 40%.

Se tanto o valor de **n** como o valor de **a** forem grandes, poderemos usar a propriedade:

Se $n \rightarrow \infty$ e $a \rightarrow \infty$

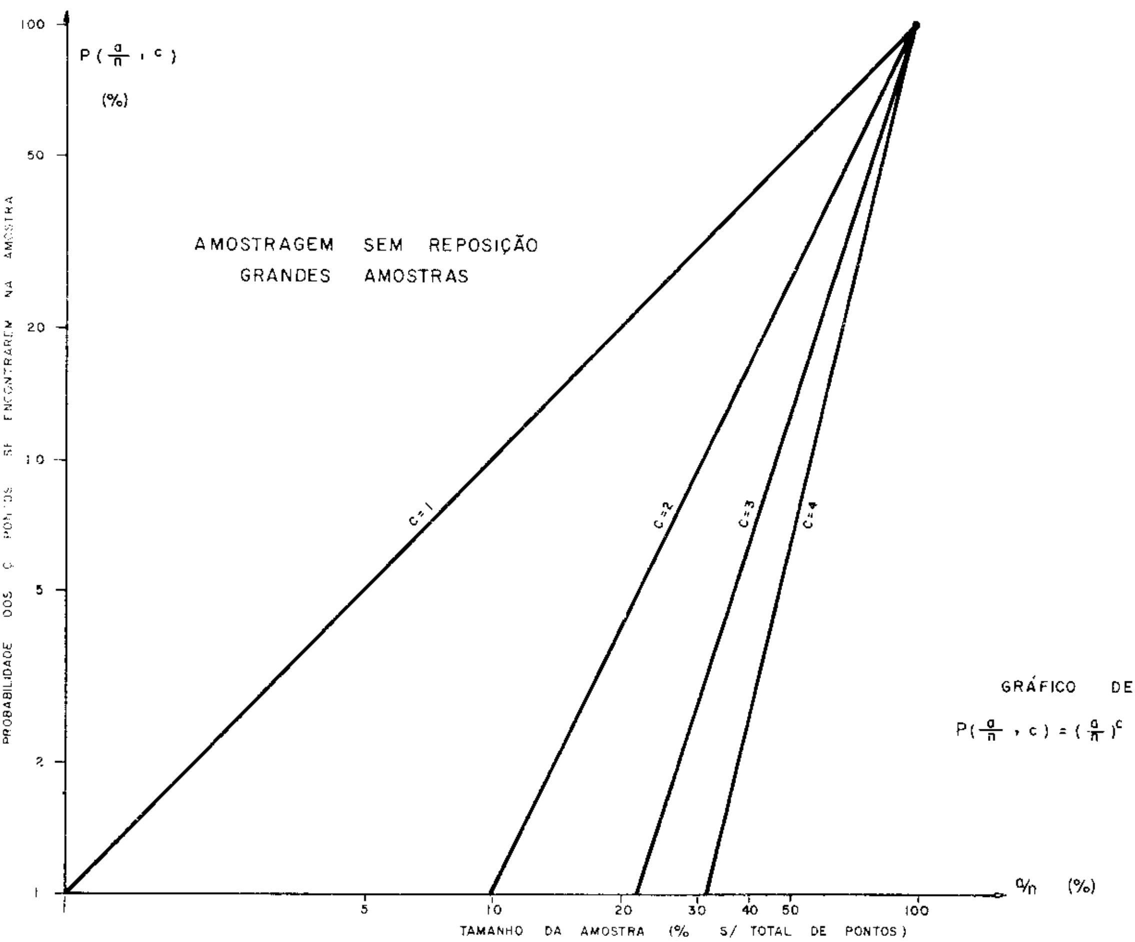
com

$$a/n \rightarrow K,$$

então

$$\boxed{\lim P \rightarrow K^c}$$

No gráfico a seguir mostramos essa função.



f) BIBLIOGRAFIA UTILIZADA

1. MACHOL, Robert E., ed. — «System Engineering Handbook», McGraw-Hill Co., New York, 1965.

2. WHETSTONE e GRIGORIEV, ed. — «Hydrologic Information Systems», UNESCO-WMO, Paris, 1972.

3. GARCEZ, L. Nogueira — «Elementos de Engenharia Hidráulica e Sanitária», E. Blücher edit., S. Paulo, 1960.

4. FELLER, W. — «An Introduction to Probability Theory and Its Applications», 3ª edição, Wiley, N.Y., 1968.

5. KAUFMANN, FUSTIER, DREVET — «L'Inventique», Entreprise Moderne d'Edition, Paris, 1970.

6. MEIER, NEWELL, PAZER — «Simulation in Business and Economics», Prentice-Hall, Inc., N.Y., 1969.

7. NAYLOR et al. — «Técnicas de Simulação em Computadores», Edit. Vozes Ltda., Rio, 1971.

8. «Plano de Controle de Potabilidade do Sistema de Abastecimento de Água de Guarulhos», public. interna do CETESB, S. Paulo, 1972.

9. MEICHES, J. — «Alguns problemas atuais de Saneamento Básico em São Paulo», Revista Politécnica, Ed. Especial, S. Paulo, 1973, p. 8.

10. VICTORETTI, B. A. — «Politica de Controle de Poluição das águas no Estado de S. Paulo», Rev. Politécnica, Ed. Especial, S. Paulo, 1973, p. 14.

11. «Controle de Potabilidade da água do Sistema distribuidor de S. Paulo», publ. int. do CETESB, S. Paulo.

12. BRANCO, Samuel M. — «Poluição», Ao Livro Técnico S/A., GB, 1972.

13. LIPSCHUTZ, S. — «Theory and Problems of Probability», McGraw-Hill, N.Y., 1968.

14. COCHRAN, W. — «Técnicas de amostragem», Ed. Fundo de Cultura, Rio, 1965.