



COMPARAÇÕES DE MÉTODOS DE TRIAGEM POR GRUPOS

ELCIO DO NASCIMENTO CHAGAS

2003

ELCIO DO NASCIMENTO CHAGAS

**COMPARAÇÕES DE MÉTODOS DE
TRIAGEM POR GRUPOS**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

Orientador

Prof. Dr. Lucas Monteiro Chaves

LAVRAS

MINAS GERAIS – BRASIL

2003

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Chagas, Elcio do Nascimento

Comparações de métodos de triagem por grupos / Elcio do Nascimento
Chagas. – Lavras : UFLA, 2003.

61 p. : il.

Orientador: Lucas Monteiro Chaves.

Dissertação – UFLA.

Bibliografia.

1. Método de triagem. 2. Distribuição Binomial. 3. Número de testes
I. Universidade Federal de Lavras. II. Título.

CDD-519.532

ELCIO DO NASCIMENTO CHAGAS

**COMPARAÇÕES DE MÉTODOS DE
TRIAGEM**

Dissertação apresentada à
Universidade Federal de Lavras
como parte das exigências do
Programa de Pós-graduação em
Agronomia, área de concentração
em Estatística e Experimentação
Agropecuária, para a obtenção do
título de "Mestre".

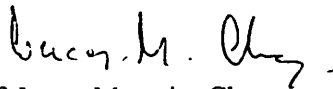
APROVADA em 26 de fevereiro de 2003

Prof. Daniel Furtado Ferreira

UFLA

Profª Adriana Mello Garcia Rabelo

UFLA



Prof. Lucas Monteiro Chaves
UFLA
(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

AGRADECIMENTOS

A realização deste trabalho somente foi possível graças aos esforços e colaboração diretos e concomitantes de várias pessoas e instituições, às quais profundamente agradeço, em especial:

À Escola Agrotécnica Federal de Alegre, à Universidade Federal de Lavras e à coordenação do Programa de Pós-graduação em Agronomia pela oportunidade de realizar este curso.

À coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão de bolsa de estudos.

Ao professor Lucas Monteiro Chaves pela orientação, confiança, amizade e constante apoio em todas as etapas deste trabalho.

Ao professor Daniel Furtado Ferreira pela co-orientação e sugestões.

Aos professores do Departamento de Ciências Exatas, pelos ensinamentos, sugestões, disponibilidade e amizade.

Aos colegas de turma pela amizade, companheirismo e troca de idéias.

Às funcionárias Maria, Maristela, Edila e Selma pela eficiência e atenção dispensada;

Este trabalho também foi possível graças ao auxílio indireto, mas não menos importante, de pessoas que me cercaram durante esse período, às quais muito devo:

À minha esposa Lana, pelo apoio, estímulo, sugestões, exemplo e, acima de tudo, paciência e companheirismo.

Às minhas filhas Liz Cristine e Maria, pelo tempo, compulsoriamente compartilhado com minhas atividades, pela paciência, pela responsabilidade, tão prematuramente adquirida.

Aos meus irmãos, pelos auxílios, amizade e por serem especiais.

Os frutos de hoje foram semeados há muito tempo... Por isso, agradeço, compartilho e dedico todas minhas conquistas aos meus pais, Pedro e Maria, que me deram tudo o que de melhor puderam e também dedico aos meus irmãos Anna e Hélio, por contribuírem efetivamente em minha formação.

SUMÁRIO

	Página
RESUMO.....	i
ABSTRACT.....	ii
1 INTRODUÇÃO.....	1
2 REFERENCIAL TEÓRICO.....	3
2.1 O Método de Triagem por Grupos.....	3
2.2 Uma hipótese simplificadora.....	12
2.2.1 Procedimento em três estágios.....	18
2.2.2 Procedimento em quatro estágios.....	19
2.2.3 Procedimento em cinco estágios.....	21
2.2.4 Procedimento em vários estágios.....	23
2.3 Método Básico de Matrizes (MBM).....	26
2.3.1 Custo do MBM.....	26
3 METODOLOGIA.....	30
3.1 Simulação dos dados.....	30
4 RESULTADOS E DISCUSSÃO.....	31
4.1 Procedimento divisão pela metade.....	31
4.2 Comparação de custos.....	33
4.2.1 MBM versus método de Watson.....	33
4.2.2 MBM versus método de Watson versus método divisão pela metade.....	35
4.3 Uma alternativa ao procedimento em três estágios.....	38

4.4 Comparação via simulação computacional dos casos dependente e independente.....	43
4.4.1 Procedimento MBM versus método de Watson.....	46
5 CONCLUSÕES.....	49
6 REFERÊNCIAS BIBLIOGRÁFICAS.....	50
ANEXOS.....	52

RESUMO

CHAGAS, Elcio do Nascimento. **Comparações de métodos de triagem por grupos**. LAVRAS: UFLA, 2003. 61 p. (Dissertação – Mestrado em Agronomia/ Área de concentração em Estatística e Experimentação Agropecuária)*

O método de triagem por grupos consiste em detectar elementos efetivos em uma população por meio do procedimento de dividir esta população em grupos de observação e testar estes grupos. Um dos aspectos da teoria é determinar o tamanho ótimo destes grupos tal que o número total de testes esperados seja mínimo. Um novo procedimento foi proposto e comparado com outros procedimentos já estabelecidos: procedimento em dois estágios, em vários estágios e método básico de matrizes. Para o caso em que existe dependência entre os indivíduos em relação ao fato de serem efetivos, ou não, foi feito um estudo via simulação computacional por meio de um programa desenvolvido com o uso do pacote estatístico SAS® (1990). Foram realizadas várias comparações em relação à esperança e desvio padrão do número total de testes e ao custo dos testes. Os novos procedimentos propostos se mostraram econômicos.

* Comitê Orientador: Lucas Monteiro Chaves - UFLA (Orientador), Daniel Furtado Ferreira - UFLA.

ABSTRACT

CHAGAS, Elcio do Nascimento. **Group screening methods comparisons**
LAVRAS: UFLA, 2003. 61 p. (Dissertation - Master in Agronomy / major in
Statistics and Agricultural Experimentation)*

The group screening method consists in detecting effective elements in a population by the procedure of dividing this population in observation groups and tests these groups. One of the aspects of the theory is to determine the optimum group size that minimizes the expected number of tests. A new procedure was proposed and compared with other procedures established: in two-stage procedure, in several stages and Basic Matrix Method. In the case of existing dependence among the individuals in relation to the fact that they are or not effective, a study of computer simulation was done by means of a program developed with the use of the statistical package SAS® (1990). Several comparisons were accomplished in relation to the expectation and standard deviation of the total number of tests, considering the cost of the tests. The new proposed procedures turned out to be economic.

* Guidance Committee: Lucas Monteiro Chaves - UFLA (Major Professor), Daniel Furtado Ferreira - UFLA.

1 INTRODUÇÃO

Quando se pretende classificar elementos como defeituosos ou não defeituosos, ou estimar a proporção de uma população infectada com uma doença (Sobel & Elashoff, 1975; Chen & Swallow, 1990), é possível, em certas situações, combinar um número de unidades em um grupo e executar um teste com este grupo. Com os resultados alcançados com os grupos pode-se detectar os indivíduos efetivos ou inferir sobre a taxa de ocorrência. Este método denomina-se Triagem por Grupos. Os grupos são denominados grupos de observação. Um grupo de observação pode ser, por exemplo, uma amostra combinada de sangue de diferentes indivíduos. Se o teste for positivo, então um ou mais indivíduos no grupo são efetivos.

O método de triagem por grupos surgiu na literatura estatística durante a Segunda Guerra Mundial com o intuito de identificar soldados infectados por sífilis (Dorfman, 1943; Wilks, 1962 e Feller, 1968). A idéia foi a de tomar amostras obtidas misturando o sangue de grupos de soldados. Se o resultado do teste fosse negativo, todos os soldados daquele grupo estariam livres da infecção. No caso de resultado positivo, um ou mais soldados do grupo estariam infectados e todos eles teriam que ser testados individualmente. Se o grupo possui K soldados, no primeiro caso apenas um teste substituiria K testes. No segundo caso seriam necessários $K + 1$ testes. Portanto haveria grupos com economia de $K - 1$ testes e outros com um gasto a mais de um teste. Se a taxa de infecção for alta, haverá predominância de grupos contaminados e, caso contrário, predominância de grupos não contaminados.

Entre os objetivos do método de triagem por grupos estão: reduzir o número esperado de testes e encontrar o tamanho ótimo do grupo que minimize o número esperado de testes. Uma economia considerável será obtida se a taxa

de ocorrência for baixa. Dorfman (1943) obteve na prática uma economia acima de 80%.

O método de triagem por grupos tornou-se uma área de pesquisa com aplicações em vários segmentos, como pode-se encontrar em citações de vários autores.

Patel (1962) utilizou o método com mais de dois estágios, em delineamentos experimentais. Mauro & Smith (1982) aplicaram o método em dois estágios, com variância do erro igual a zero, em delineamentos experimentais. Mauro (1984) aplicou o método com a variância do erro diferente de zero nas observações. Hepwort (1996) aplicou o método na avaliação de doenças de plantas (indústria de cravos na Austrália). Ivanova & Malone (1999) compararam, via simulação computacional, a eficiência na fabricação de semicondutores, utilizando ou não o método de triagem. Johnson & Gastwirth (2000) incorporaram o controle de qualidade no método de triagem ao estudarem epidemia de HIV, devido à possibilidade de ocorrerem falsos negativos. Hung & Swallow (2000) incluíram covariáveis em estudo com o objetivo de aumentar o poder dos testes.

O presente trabalho tem como objetivo estudar situações particulares dos métodos de triagem por grupos e fazer comparações entre eles. Simulações computacionais foram realizadas por meio de um programa desenvolvido com o uso do pacote estatístico SAS[®] (1990).

2 REFERENCIAL TEÓRICO

2.1. O Método de Triagem por Grupos

Na investigação de uma variável resposta, um grande número de indivíduos pode estar envolvido, porém apenas alguns deles podem de fato afetar a resposta. Tais indivíduos são denominados efetivos. Métodos para encontrar os poucos indivíduos efetivos em um grande número de indivíduos são denominados métodos de triagem. A idéia de colocar os indivíduos em grupos, testá-los e então fazer testes individuais apenas para aqueles que contenham indivíduos efetivos foi sugerida, segundo Watson (1961), por W.S.Connor. Métodos desse tipo são denominados métodos de triagem por grupos. Um exemplo clássico foi proposto por Dorfman (1943) para detectar sífilis em soldados durante a Segunda Guerra Mundial.

Supõe-se que N indivíduos serão testados. Assume-se que:

1. Todos os indivíduos tem, a priori, a mesma probabilidade p de serem efetivos.
2. Não existem interações entre os indivíduos.
3. Os indivíduos são divididos em grupos.

Neste trabalho será considerado apenas o caso em que não existe a possibilidade de falsos negativos ou falsos positivos, ou seja:

4. Grupos positivos são os que contêm pelo menos um indivíduo efetivo.
5. Grupos negativos são os que não contêm indivíduos efetivos.

O objetivo do método é:

- a) Detectar todos os indivíduos efetivos.
- b) Realizar esse objetivo com o menor número de testes possível.

Pela pressuposição (1), um grupo que contém K indivíduos é efetivo com probabilidade $1 - q^K$, em que $q = 1 - p$.

Seja $N = mK$, em que:

N : tamanho da população;

m : número de grupos;

K : número de indivíduos por grupo.

Define-se a variável de Bernoulli.:

X_i : variável aleatória número de testes para o grupo i .

$$X_i = \begin{cases} 1 & \text{se o teste for negativo} \\ K+1 & \text{se o teste for positivo} \end{cases}$$

$$P(X_i = 1) = q^K$$

$$P(X_i = K + 1) = 1 - q^K$$

Logo, a esperança do número de testes para o grupo i é:

$$E(X_i) = 1 \cdot q^K + (K + 1)(1 - q^K)$$

$$\Rightarrow E(X_i) = q^K + K - Kq^K + 1 - q^K$$

Simplificando, tem-se que:

$$E(X_i) = K(1 - q^K) + 1$$

Considerando agora a variável aleatória X , número total de testes, tem-

se que $X = \sum_{i=1}^m X_i$. A esperança de X é dada por

$$E(X) = \sum_{i=1}^m E(X_i) = \sum_{i=1}^m [K(1-q^K) + 1]$$

$$\Rightarrow E(X) = [K(1-q^K) + 1] \frac{N}{K}$$

$$\Rightarrow E(X) = \frac{NK}{K} - \frac{NK}{K} q^K + \frac{N}{K}$$

$$E(X) = N - Nq^K + \frac{N}{K}$$

$$\Rightarrow E(X) = N(1-q^K) + \frac{N}{K} \quad (1)$$

Nota-se que quando $q = 0$, todos os indivíduos na população são efetivos e $E(X) = N + \frac{N}{K}$. Em caso contrário ($q = 1$), nenhum indivíduo na população é efetivo e $E(X) = \frac{N}{K} = m$.

Exemplo: Supõe-se $N = 10000$ pessoas, divididas em grupos de 1000.

O valor esperado do número de testes em função de q será:

$$\begin{cases} f(q) = N(1-q^K) + \frac{N}{K} = 10000(1-q^{1000}) + 10 \\ f(0) = N + \frac{N}{K} = 10000 + 10 = 10010 \\ f(1) = \frac{N}{K} = \frac{10000}{1000} = 10 \end{cases}$$

Para um valor fixo do tamanho do grupo K , a esperança do número de testes em função de q é uma função decrescente, pois:

$$\frac{dE(X)}{dq} = \frac{d}{dq} \left(N(1-q^K) + \frac{N}{K} \right) = -KNq^{K-1} < 0.$$

Neste caso, o método somente é interessante para valores de q maiores que a solução da equação $N = E(X) = N(1-q^K) + \frac{N}{K}$.

No exemplo:

$$10000 = N(1-q^K) + \frac{N}{K}$$

$$10000 = 10000(1-q^{1000}) + 10$$

$$\frac{9900}{10000} = 1 - q^{1000}$$

$$q^{1000} = 0,01$$

$$1000 \cdot \ln q = \ln 0,01$$

$$1000 \cdot \ln q = -4,6052$$

$$\ln q = -0,0046052$$

$$q = 0,9954$$

A variância do número total de testes X é $\sigma_X^2 = E(X^2) - [E(X)]^2$

$$= E \left[\left(\sum_{i=1}^{N/K} X_i \right)^2 \right] - \left[E \left(\sum_{i=1}^{N/K} X_i \right) \right]^2$$

$$\begin{aligned}
&= E\left(\sum_{i=1}^{N/K} X_i X_j\right) - \left[\sum_{i=1}^{N/K} (EX_i)\right]^2 \\
&= E\left(\sum_{i=1}^{N/K} X_i^2\right) + \sum_{i \neq j}^{N/K} E(X_i X_j) - \left[\sum_{i=1}^{N/K} (EX_i)\right]^2 - \sum_{i \neq j}^{N/K} E(X_i)E(X_j).
\end{aligned}$$

Como os X_i são independentes,

$$E(X_i X_j) = E(X_i)E(X_j) \quad i \neq j.$$

Tem-se que $\sigma_x^2 = E\left(\sum_{i=1}^{N/K} X_i^2\right) - \left[\sum_{i=1}^{N/K} (EX_i)\right]^2$

$$\begin{aligned}
&= \sum_{i=1}^{N/K} E(X_i^2) - \sum_{i=1}^{N/K} [(EX_i)]^2 \\
&= \frac{N}{K} \left\{ E(X_i^2) - [(EX_i)]^2 \right\} \\
&= \frac{N}{K} \left\{ q^K + (K+1)^2 (1-q^K) - [q^K + (K+1)(1-q^K)]^2 \right\} \\
&= \frac{N}{K} \left\{ q^K + (K+1)^2 (1-q^K) - [q^{2K} + 2q^K (K+1)(1-q^K) \right. \\
&\quad \left. + (K+1)^2 (1-q^K)^2] \right\} \\
&= \frac{N}{K} \left\{ q^K + (K^2 + 2K + 1)(1-q^K) - [q^{2K} + 2q^K K - 2Kq^{2K} \right. \\
&\quad \left. + 2q^K - 2q^{2K} + (K^2 + 2K + 1)(1-2q^K + q^{2K})] \right\}
\end{aligned}$$

$$= \frac{N}{K} \left\{ q^K + K^2 - K^2 q^K + 2K - 2Kq^K + 1 - q^K - [q^{2K} + 2q^K K - 2Kq^{2K} + 2q^K - 2q^{2K} + 1 - 2q^K + q^{2K} + 2K - 4Kq^K + 2Kq^{2K} + K^2 - 2K^2 q^K + K^2 q^{2K}] \right\}$$

Simplificando, tem-se que $\sigma_x^2 = \frac{N}{K} (K^2 q^K - K^2 q^{2K})$

$$\Rightarrow \sigma_x^2 = NK (q^K - q^{2K}). \quad (2)$$

Uma medida de economia no custo relativo ao número de testes é definida por Dorfman (1943) como:

$$\Rightarrow C = \frac{N(1 - q^K) + \frac{N}{K}}{N}$$

$$\Rightarrow C = (1 - q^K) + \frac{1}{K}$$

$$\Rightarrow C = \frac{K+1}{K} - q^K. \quad (3)$$

Observa-se que quanto menor é C, maior é a economia proporcionada pelo processo.

A economia alcançada pelo uso do método depende do tamanho do grupo e da probabilidade p de ocorrência de doença.

A Tabela 1 e a Figura 1 ilustram um exemplo de cálculo do número esperado de testes tendo-se fixado N, o tamanho da população e variando os valores de K em função da probabilidade de ocorrência p.

TABELA 1 - Número esperado de testes tendo-se fixado N , o tamanho da população em 10000, em que K é o número de indivíduos por grupos e p , a probabilidade de ocorrência.

K	$p = 0,01$	$p = 0,02$	$p = 0,05$	$p = 0,10$	$p = 0,15$
4	2894	3276	4355	5939	7280
16	2110	3387	6224	8772	9882
20	2321	3824	6915	9284	10112
25	2622	4365	7626	9682	10228
40	3560	5793	8965	10102	10235
50	4150	6558	9431	10148	10197
80	5650	8139	9960	10122	10125
100	6440	8774	10041	10100	10100

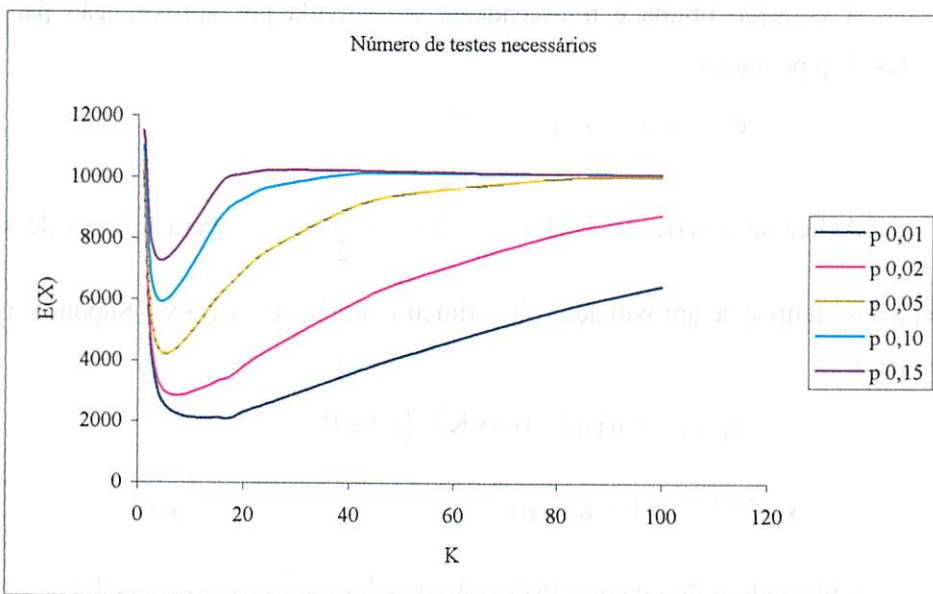


FIGURA 1 - Número esperado de testes em que K é o número de indivíduos por grupos e p , a probabilidade de ocorrência.

O valor de K que minimiza E(X) será denominado K ótimo e obtido da seguinte forma:

Na equação 1, supondo K variando continuamente, deriva-se E(X) em relação a K para se encontrar o mínimo.

$$\frac{dE(X)}{dK} = -Nq^K \ln q - \frac{N}{K^2}$$

Fazendo $\frac{dE(X)}{dK} = 0$, tem-se

$$-Nq^K \ln q - \frac{N}{K^2} = 0$$

$$q^K \ln q + \frac{1}{K^2} = 0. \quad (4)$$

A equação obtida é transcendente e resolvida por aproximação para valores de p pequenos.

$$e^{\ln(q)} = q \rightarrow q^x = e^{x \ln(q)}. \quad (5)$$

Aplica-se a série de Taylor $e^x = 1 + x + \frac{x^2}{2!} + \dots$. Para valores de x pequenos, tem-se a aproximação de primeira ordem $e^x \cong 1 + x$. Supondo p pequeno:

$$q \cong 1 \rightarrow \ln(q) \cong 0 \rightarrow K \ln(q) \cong 0$$

$$\rightarrow e^{K \ln(q)} \cong 1 + K \ln(q). \quad (6)$$

Utilizando o fato de que $\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \dots$ para $-1 \leq x \leq 1$

e que $q = 1 - p$, então $\ln(q) = \ln(1 - p) = -p - \frac{1}{2}(-p)^2 + \dots$

Desprezando os termos de ordem superior, tem-se que:

$$\ln(q) = -p. \quad (7)$$

Substituindo (5), (6) e (7) em (4), tem-se:

$$e^{K \ln(q)} + (-p) + \frac{1}{K^2} = 0 \quad (4)$$

$$[1 + K(-p)](-p) + \frac{1}{K^2} = 0$$

$$(1 - Kp)(-p) + \frac{1}{K^2} = 0$$

$$\Rightarrow \frac{1}{K^2} = p(1 - Kp)$$

$$\Rightarrow K^2 = \frac{1}{p - Kp^2}.$$

Como a probabilidade de ocorrência é pequena, tem-se que

$$\Rightarrow K^2 \cong \frac{1}{p}$$

$$\Rightarrow K \cong \frac{1}{\sqrt{p}}. \quad (8)$$

Este valor será então considerado como valor ótimo do tamanho dos grupos. Substituindo (8) na equação (1), tem-se:

$$E(X) = N \left(1 - q^{1/\sqrt{p}} \right) + \frac{N}{1/\sqrt{p}}. \quad (9)$$

Teorema Binomial (Mood et al., 1974)-se m é qualquer número real, então:

$$(1-x)^m = \sum_{i=0}^{\infty} \binom{m}{i} (-1)^i, \text{ para todo valor de } x, |x| < 1.$$

Do teorema, tem-se que:

$$(1-x)^m = 1 - mx + \frac{m(m-1)x^2}{2!} - \frac{m(m-1)(m-2)x^3}{3!} + \dots$$

$$\Rightarrow q^{1/\sqrt{p}} = (1-p)^{1/\sqrt{p}} = 1 - \frac{1}{\sqrt{p}}p + \frac{1}{2\sqrt{p}}\left(\frac{1}{\sqrt{p}} - 1\right)p^2 + \dots$$

Desprezando os termos de ordem superior, supondo p pequeno:

$$(1-p)^{1/\sqrt{p}} \cong 1 - \frac{1}{\sqrt{p}}p. \quad (10)$$

Substituindo (10) em (9), tem-se:

$$E(X) \cong N \left[1 - \left(1 - \frac{1}{\sqrt{p}}p \right) \right] + N\sqrt{p}$$

$$\Rightarrow E(X) \cong N\sqrt{p} + N\sqrt{p} \cong 2N\sqrt{p}, \quad (11)$$

sendo este então o valor mínimo esperado para o número total de testes. Este procedimento em dois estágios será denominado método de Watson. O custo mínimo dos testes, denominado por c , é dado por:

$$c = 2\sqrt{p}. \quad (12)$$

2.2. Uma hipótese simplificadora

Finucan (1965) retoma o trabalho de Dorfman (1943) adotando a seguinte abordagem heurística. Supondo p suficientemente pequeno e o tamanho K dos grupos não muito grande, é razoável supor que para cada grupo pode ocorrer no máximo um indivíduo efetivo. Com esta hipótese, a situação se torna

bem mais simples, pois, por exemplo, o número de grupos efetivos passa a ser exatamente igual ao número de indivíduos efetivos.

Pode-se considerar esta abordagem como uma aproximação de primeira ordem do problema, similar à adotada anteriormente de se tomarem apenas os termos lineares da expansão de Taylor. Com esta abordagem, possíveis flutuações no número de indivíduos infectados são iguais a flutuações no número de grupos efetivos. Pode-se então trabalhar com a esperança do número de indivíduos efetivos, isto é, se a probabilidade de um indivíduo ser efetivo é p , o número esperado de efetivos é Np , que é o mesmo que supor uma percentagem p da população seja efetiva. Tem-se então NpK grupos efetivos, e como para cada grupo efetivo todos os indivíduos devem ser testados individualmente, o número total de testes será $\frac{N}{K} + NpK$.

Neste caso, a economia é:

$$C = \frac{1}{K} + pK.$$

Seja:

$$r = \frac{1}{p} : \text{em média, um em } r \text{ indivíduos está infectado.}$$

$$\Rightarrow C = \frac{1}{K} + \frac{K}{r}.$$

Deriva-se C em relação a K para se encontrar o custo mínimo c :

$$\frac{dC}{dK} = -\frac{1}{K^2} + \frac{1}{r}$$

Fazendo-se $\frac{dC}{dK} = 0$

$$-\frac{1}{K^2} + \frac{1}{r} = 0$$

$$\Rightarrow \frac{1}{K^2} = \frac{1}{r}$$

$$\Rightarrow K = \sqrt{r}$$

Substituindo $K = \sqrt{r}$ em $C = \frac{1}{K} + \frac{K}{r}$, tem-se que $c = \frac{2}{\sqrt{r}} = 2\sqrt{p}$.

Observa-se que este valor é o mesmo obtido na equação 12.

Finucan (1965) também observa que o método das diferenças finitas é mais adequado que a diferenciação para encontrar o valor ótimo de K . A fórmula de interpolação de Newton (diferença à frente) propicia uma boa aproximação para $f(x+h)$ com poucos termos da expansão.

Seja:

$$C_K = \frac{1}{K} + (1 - q^K)$$

$$\Delta C_K = C_{K+1} - C_K$$

$$\Rightarrow \Delta C_K = \frac{1}{K+1} + 1 - q^{K+1} - \left(\frac{1}{K} + 1 - q^K \right)$$

$$\Rightarrow \Delta C_K = \frac{1}{K+1} - \frac{1}{K} + q^K - q^{K+1}$$

$$\Rightarrow \Delta C_K = \frac{K - K - 1}{K(K+1)} + q^K(1 - q)$$

$$\Rightarrow \Delta C_K = -\frac{1}{K(K+1)} + pq^K.$$

Esta diferença muda de sinal exatamente antes do C mínimo. Assim, o mínimo é exatamente depois da solução formal $\Delta C = 0$. Resolvendo $\Delta C = 0$, tem-se:

$$\frac{1}{K(K+1)} = pq^K$$

$$\Rightarrow \frac{1}{K(K+1)} = \frac{1}{r} \cdot \left(1 - \frac{1}{r}\right)^K$$

$$\Rightarrow K(K+1) = \frac{r}{\left(1 - \frac{1}{r}\right)^K}$$

$$\Rightarrow K(K+1) = r \left(1 - \frac{1}{r}\right)^{-K}$$

Fazendo $K = \sqrt{r} + z$, em que z é um valor desconhecido, e substituindo na expressão acima, tem-se:

$$\begin{aligned} (\sqrt{r} + z)(\sqrt{r} + z + 1) &= r \left(1 - \frac{1}{r}\right)^{-(\sqrt{r} + z)} \\ r + 2z\sqrt{r} + \sqrt{r} + z^2 + z &= r \left(1 - \frac{1}{r}\right)^{-(\sqrt{r} + z)} \\ \Rightarrow 1 + \frac{2z}{\sqrt{r}} + \frac{1}{\sqrt{r}} + \frac{z^2}{r} + \frac{z}{r} &= \left(1 - \frac{1}{r}\right)^{-(\sqrt{r} + z)} \end{aligned} \quad (13)$$

Utilizando o Teorema Binomial da seção 2.1, tem-se:

$$\begin{aligned} (1-x)^{-m} &= 1 + mx + \frac{m(m+1)x^2}{2!} + \frac{m(m+1)(m+2)x^3}{3!} + \dots \\ \Rightarrow \left(1 - \frac{1}{r}\right)^{-(\sqrt{r} + z)} &= 1 + \frac{(\sqrt{r} + z)}{r} + \frac{(\sqrt{r} + z)(\sqrt{r} + z + 1)}{2!r^2} + \dots = \\ &= 1 + \frac{1}{\sqrt{r}} + \frac{z}{r} + \frac{1}{2r} + \frac{2z+1}{2r\sqrt{r}} + \frac{z^2+z}{2r^2} + \dots \cong \\ &\cong 1 + \frac{1}{\sqrt{r}} + \frac{z}{r} + \frac{1}{2r} + \frac{2z+1}{2r\sqrt{r}} \end{aligned} \quad (14)$$

Para valores grandes de r , isto é, para pequenos valores de p , sabe-se que a partir de $\frac{z^2 + z}{2r^2} + \dots$, os termos tendem a zero. Logo, igualando (13) e (14),

tem-se:

$$\begin{aligned}
 1 + \frac{1}{\sqrt{r}} + \frac{z}{r} + \frac{1}{2r} + \frac{2z+1}{2r\sqrt{r}} &= 1 + \frac{2z}{\sqrt{r}} + \frac{1}{\sqrt{r}} + \frac{z^2}{r} + \frac{z}{r} \\
 \Rightarrow \frac{1}{2r} + \frac{2z+1}{2r\sqrt{r}} &= \frac{2z}{\sqrt{r}} + \frac{z^2}{r} \\
 \Rightarrow z &= \frac{1}{4\sqrt{r}}, \text{ aproximadamente.}
 \end{aligned}$$

Dessa forma, $\Delta C = 0$ quando $K \cong \sqrt{r} + \frac{1}{4\sqrt{r}}$, isto é, C é mínimo quando K é um inteiro mais próximo desse valor. A Tabela 2 está de acordo com este resultado e pode ser considerada como uma verificação do método de solução adaptado aqui, com exceção para $p = 0,07$, em que $C = 0,53$, para $K = 4$, enquanto, para $K = 5$, $C = 0,55$, denotando que o tamanho ótimo para $p = 0,07$ é $K = 4$ e não $K = 5$, como encontrado por Dorfman (1943).

Finucan (1965) conclui que o resultado $K \cong \sqrt{r}$, também encontrado em Watson (1961), é bom para valores pequenos de p .

TABELA 2 - Tamanho ótimo do grupo e custo relativo (em %) ao número de testes em função da probabilidade de ocorrência de infecção.

Probabilidade de ocorrência	Tamanho ótimo do grupo	Custo
0,01	11	20
0,02	8	27
0,03	6	33
0,04	6	38
0,05	5	43
0,06	5	47
0,07	5	50
0,08	4	53
0,09	4	56
0,10	4	59
0,12	4	65
0,13	3	67
0,15	3	72
0,20	3	82
0,25	3	91
0,30	2	99

FONTE: Dorfman (1943)

2.2.1 Procedimento em três estágios

Uma das vantagens do método de triagem por grupos é de que ele pode ser aplicado de forma iterativa. O procedimento anterior de separar em grupos e testar todos indivíduos dos grupos efetivos será denominado procedimento em dois estágios. Os grupos são testados, e de dois estágios. Por meio do procedimento em três estágios (Finucan, 1965), a população é dividida em grupos com K indivíduos de acordo com a hipótese simplificadora de no máximo um indivíduo efetivo por grupo, Np grupos são recusados. Tem-se então um total de NpK indivíduos para serem testados. Divide-se novamente estes indivíduos em subgrupos com K_1 elementos e novamente são rejeitados Np subgrupos com NpK_1 indivíduos, que serão testados individualmente. O número total de testes é:

$$\frac{N}{K} + \frac{NpK}{K_1} + NpK_1.$$

A economia no custo relativo do número de testes é:

$$C = \frac{1}{K} + \frac{pK}{K_1} + pK_1$$
$$\Rightarrow C = \frac{1}{K} + \frac{K}{rK_1} + \frac{K_1}{r}.$$

Deriva-se C em relação a K , em relação a K_1 :

$$\frac{dC}{dK} = -\frac{1}{K^2} + \frac{1}{rK_1} \quad \text{e} \quad \frac{dC}{dK_1} = -\frac{K}{rK_1^2} + \frac{1}{r}.$$

Fazendo-se $\frac{dC}{dK} = 0$ e $\frac{dC}{dK_1} = 0$:

$$\frac{1}{K^2} = \frac{1}{rK_1} \quad \text{e} \quad \frac{K}{rK_1^2} = \frac{1}{r}$$

$$\Rightarrow K^2 = rK_1 \text{ e } K = K_1^2$$

$$\Rightarrow K^2 = rK^{1/2}$$

$$\Rightarrow K = (\sqrt[3]{r})^2 \text{ e } K_1 = \sqrt[3]{r}.$$

Dessa forma, o custo mínimo é:

$$c = \frac{1}{(\sqrt[3]{r})^2} + \frac{(\sqrt[3]{r})^2}{r\sqrt[3]{r}} + \frac{\sqrt[3]{r}}{r}$$

$$\Rightarrow c = \frac{1}{(\sqrt[3]{r})^2} + \frac{(\sqrt[3]{r})^2}{(\sqrt[3]{r})^4} + \frac{\sqrt[3]{r}}{r}$$

$$\Rightarrow c = \frac{1}{(\sqrt[3]{r})^2} + \frac{1}{(\sqrt[3]{r})^2} + \frac{1}{(\sqrt[3]{r})^2}$$

$$\Rightarrow c = \frac{3}{(\sqrt[3]{r})^2}.$$

2.2.2 Procedimento em quatro estágios

De forma análoga ao procedimento em três estágios, são formados, seqüencialmente, grupos com K , K_1 e K_2 indivíduos (Finucan, 1965). O número total de testes é:

$$\frac{N}{K} + \frac{NpK}{K_1} + \frac{NpK_1}{K_2} + NpK_2$$

$$C = \frac{1}{K} + \frac{pK}{K_1} + \frac{pK_1}{K_2} + pK_2$$

$$\Rightarrow C = \frac{1}{K} + \frac{K}{rK_1} + \frac{K_1}{rK_2} + \frac{K_2}{r}$$

Deriva-se C em relação a K, K₁ e em relação a K₂.

$$\frac{dC}{dK} = -\frac{1}{K^2} + \frac{1}{rK_1}$$

$$\frac{dC}{dK_1} = -\frac{K}{rK_1^2} + \frac{1}{rK_2}$$

$$\frac{dC}{dK_2} = -\frac{K_1}{rK_2^2} + \frac{1}{r}$$

Fazendo-se $\frac{dC}{dK} = 0$, $\frac{dC}{dK_1} = 0$ e $\frac{dC}{dK_2} = 0$, tem-se, respectivamente:

$$\frac{1}{K^2} = \frac{1}{rK_1}, \quad \frac{K}{rK_1^2} = \frac{1}{rK_2} \quad \text{e} \quad \frac{K_1}{rK_2^2} = \frac{1}{r}$$

$$\Rightarrow K^2 = rK_1, \quad \frac{K}{K_1^2} = \frac{1}{K_2} \quad \text{e} \quad K_1 = K_2^2$$

$$K^2 = rK_1$$

$$\Rightarrow K = r^{1/2}K_1^{1/2}$$

$$\frac{K}{K_1^2} = \frac{1}{K_2}$$

$$\Rightarrow \frac{K}{K_1^2} = \frac{1}{K_1^{1/2}}$$

$$\Rightarrow K = K_1^{3/2}$$

$$\begin{aligned}
K &= r^{1/2} K_1^{1/2} \\
\Rightarrow K_1^{3/2} &= r^{1/2} K_1^{1/2} \\
\Rightarrow K_1 &= r^{1/2} \\
\Rightarrow K &= (\sqrt[4]{r})^3, K_1 = \sqrt{r} \text{ e } K_2 = \sqrt[4]{r}.
\end{aligned}$$

Dessa forma, o custo mínimo é:

$$\begin{aligned}
c &= \frac{1}{(\sqrt[4]{r})^3} + \frac{(\sqrt[4]{r})^3}{r\sqrt{r}} + \frac{\sqrt{r}}{r\sqrt[4]{r}} + \frac{\sqrt[4]{r}}{r} \\
\Rightarrow c &= \frac{1}{(\sqrt[4]{r})^3} + \frac{1}{(\sqrt[4]{r})^3} + \frac{1}{(\sqrt[4]{r})^3} + \frac{1}{(\sqrt[4]{r})^3} \\
\Rightarrow c &= \frac{4}{(\sqrt[4]{r})^3}.
\end{aligned}$$

2.2.3 Procedimento em cinco estágios

O número total de testes é:

$$\begin{aligned}
\frac{N}{K} + \frac{NpK}{K_1} + \frac{NpK_1}{K_2} + \frac{NpK_2}{K_3} + NpK_3 \\
C = \frac{1}{K} + \frac{pK}{K_1} + \frac{pK_1}{K_2} + \frac{pK_2}{K_3} + pK_3 \\
\Rightarrow C = \frac{1}{K} + \frac{K}{rK_1} + \frac{K_1}{rK_2} + \frac{K_2}{rK_3} + \frac{K_3}{r}.
\end{aligned}$$

Deriva-se C em relação a K, K₁, K₂ e em relação a K₃.

$$\frac{dC}{dK} = -\frac{1}{K^2} + \frac{1}{rK_1}$$

$$\frac{dC}{dK_1} = -\frac{K}{rK_1^2} + \frac{1}{rK_2}$$

$$\frac{dC}{dK_2} = -\frac{K_1}{rK_2^2} + \frac{1}{rK_3}$$

$$\frac{dC}{dK_3} = -\frac{K_2}{rK_3^2} + \frac{1}{r}$$

Fazendo-se $\frac{dC}{dK} = 0$, $\frac{dC}{dK_1} = 0$, $\frac{dC}{dK_2} = 0$ e $\frac{dC}{dK_3} = 0$, tem-se,

respectivamente:

$$\frac{1}{K^2} = \frac{1}{rK_1} , \frac{K}{rK_1^2} = \frac{1}{rK_2} , \frac{K_1}{rK_2^2} = \frac{1}{rK_3} \text{ e } \frac{K_2}{rK_3^2} = \frac{1}{r}$$

$$\Rightarrow K^2 = rK_1 , \frac{K}{K_1^2} = \frac{1}{K_2} , \frac{K_1}{K_2^2} = \frac{1}{K_3} \text{ e } K_2 = K_3^2$$

$$\frac{K_1}{K_2^2} = \frac{1}{K_3}$$

$$\Rightarrow K_2 = K_1^{2/3}$$

$$\frac{K}{K_1^2} = \frac{1}{K_2}$$

$$\Rightarrow K_1 = K^{3/4}$$

$$K^2 = rK_1$$

$$\Rightarrow K = r^{4/5}.$$

$$\text{Logo, } K = (\sqrt[5]{r})^4, K_1 = (\sqrt[5]{r})^3, K_2 = (\sqrt[5]{r})^2 \text{ e } K_3 = \sqrt[5]{r}.$$

Dessa forma, o custo mínimo é:

$$c = \frac{1}{(\sqrt[5]{r})^4} + \frac{(\sqrt[5]{r})^4}{r(\sqrt[5]{r})^3} + \frac{(\sqrt[5]{r})^3}{r(\sqrt[5]{r})^2} + \frac{(\sqrt[5]{r})^2}{r\sqrt[5]{r}} + \frac{\sqrt[5]{r}}{r}$$

$$\Rightarrow c = \frac{1}{(\sqrt[5]{r})^4} + \frac{1}{(\sqrt[5]{r})^4} + \frac{1}{(\sqrt[5]{r})^4} + \frac{1}{(\sqrt[5]{r})^4} + \frac{1}{(\sqrt[5]{r})^4}$$

$$\Rightarrow c = \frac{5}{(\sqrt[5]{r})^4}.$$

2.2.4 Procedimento em vários estágios

Seqüencialmente são formados grupos com $K, K_1, \dots, K_{(s-1)}$ indivíduos.

O número total de testes é:

$$\frac{N}{K} + \frac{NpK}{K_1} + \dots + NpK_{s-1}$$

$$C = \frac{1}{K} + \frac{pK}{K_1} + \dots + pK_{s-1}$$

$$\Rightarrow C = \frac{1}{K} + \frac{K}{rK_1} + \dots + \frac{K_{s-1}}{r}.$$

Pode-se mostrar (Finucan, 1965) que a solução do sistema

$$\frac{dC}{dK_i} = 0 \quad i = 0, \dots, s-1 \text{ forma uma progressão geométrica de razão } \sqrt[5]{r}.$$

$$K = (\sqrt[s]{r})^{s-1}, K_1 = (\sqrt[s]{r})^{s-2}, \dots \text{ e } K_{s-1} = \sqrt[s]{r}.$$

Dessa forma, o correspondente custo mínimo é:

$$c = \frac{1}{(\sqrt[s]{r})^{s-1}} + \frac{(\sqrt[s]{r})^{s-1}}{r(\sqrt[s]{r})^{s-2}} + \dots + \frac{\sqrt[s]{r}}{r}$$

$$\Rightarrow c = \frac{s}{(\sqrt[s]{r})^{s-1}}$$

$$\Rightarrow c(s) = \frac{s\sqrt[s]{r}}{r}.$$

Aplicando logaritmo e denotando $\ln[c(s)]$ por $L(s)$, tem-se

que:

$$L(s) = \ln s + (\ln r)/s - \ln r.$$

Para se obter o número de estágios que é mínimo para o custo, deriva-se L :

$$\frac{dL}{ds} = 1/s - (\ln r)/s^2$$

$$\text{Fazendo-se } \frac{dL}{ds} = 0$$

$$\frac{1}{s} = (\ln r)/s^2$$

$$\Rightarrow s = \ln r.$$

Logo, aproximadamente, o número ótimo de estágios é dado pelo inteiro mais próximo de $\ln r$. Substituindo na expressão para o custo mínimo, tem-se:

$$c_{\min} = \frac{(\ln r)r^{1/\ln r}}{r}$$

$$\Rightarrow c_{\min} = \frac{e \ln r}{r}$$

A Tabela 3 expressa a comparação do método de Watson com o procedimento em s estágios descrito por Finucan (1965). Pode-se observar que quando se adota o procedimento em vários estágios, o custo diminui. Quando o número ótimo de estágios é dois, existem diferenças nos resultados para o custo devido às aproximações efetuadas no método de Watson.

TABELA 3 – Comparação entre o método de Watson e o procedimento em s estágios para determinar o tamanho ótimo do grupo e o custo relativo (em %) ao número de testes em função da probabilidade de ocorrência de infecção p.

Probabilidade de ocorrência	Tamanho ótimo do grupo	Custo	Número ótimo de estágios	Custo mínimo
0,01	11	20	5	13
0,02	8	28	4	21
0,03	6	35	4	29
0,04	6	40	3	35
0,05	5	45	3	41
0,06	5	49	3	45
0,07	5	53	3	51
0,08	4	57	3	55
0,09	4	60	2	59
0,10	4	63	2	63
0,12	4	69	2	69
0,13	3	72	2	72
0,15	3	77	2	77
0,20	3	89	2	87
0,25	3	100	2	94
0,30	2	109	2	98

2.3 Método Básico de Matrizes

Um outro método de triagem por grupos, descrito por Berger et al. (2000), consiste em indexar cada indivíduo com uma entrada em uma matriz retangular. Os grupos de observação são formados pelos elementos indexados pelas linhas L_i e as colunas C_j desta matriz. Observa-se que, neste caso, cada elemento pertence simultaneamente a dois grupos e pode ocorrer tamanho de grupos diferentes. Cada elemento, como está em dois grupos, é testado pelo menos duas vezes, o que aumenta a confiabilidade dos resultados. Este método será denominado método básico de matrizes (MBM), utilizado, por exemplo, em testes de DNA e AIDS.

2.3.1 Custo do MBM

Sejam

$N = mn$: população;

m : número de linhas;

n : número de colunas;

$m + n$: número de grupos.

Definem-se as variáveis aleatórias:

L_i (respectivamente C_j): igual a um se a i -ésima linha (respectivamente, j -ésima coluna) testada for positiva e zero se for negativa, $1 \leq i \leq m$ ($1 \leq j \leq n$).

$L = L_1 + L_2 + \dots + L_m$: variável aleatória número de linhas positivas.

$C = C_1 + C_2 + \dots + C_n$: variável aleatória número de colunas positivas.

Após os testes dos grupos, os elementos que podem ser efetivos são indexados pelas intersecções entre as linhas e colunas dos grupos efetivos.

O número destas intersecções é dado pelo produto do número de linhas pelo de colunas.

A esperança do número total de testes é $E(X) = m + n + E[LC]$.

Tem-se que L e C são estocasticamente independentes. Então:

$$E[LC] = E\left[\left(\sum_{i=1}^m L_i\right)\left(\sum_{j=1}^n C_j\right)\right]$$

$$\Rightarrow E[LC] = mnE[L_1C_1]$$

pode ser :

$$0.0 = 0, 0.1=0, 1.0=0, 1.1=1.$$

Logo, $L_1C_1 = 0$ a menos que $L_1 = C_1 = 1$. Dessa forma $E[L_1C_1] = p(L_1 = C_1 = 1)$. O evento $L_1 = C_1 = 1$ ocorre quando o indivíduo $(1,1)$ é positivo ou quando este indivíduo é negativo mas pelo menos um dos outros $n-1$ indivíduos na linha 1 é positivo e pelo menos um dos outros $m-1$ indivíduos na coluna 1 é positivo. Desse modo:

$$E[L_1C_1] = p(L_1=C_1=1)$$

$$\Rightarrow E[L_1C_1] = p+q(1-q^{n-1})(1-q^{m-1})$$

$$\Rightarrow E[L_1C_1] = p+q(1-q^{n-1}-q^{m-1}+q^{m+n-2})$$

$$\Rightarrow E[L_1C_1] = 1-q+q-q^m-q^n+q^{m+n-1}$$

$$\Rightarrow E[L_1C_1] = 1-q^m-q^n+q^{m+n-1}.$$

Portanto, $E[LC] = mn(1-q^m-q^n+q^{m+n-1})$, e assim,

$E(X) = m + n + mn(1-q^m-q^n+q^{m+n-1})$, sendo o custo dado por:

$$C = \frac{m+n+mn(1-q^m-q^n+q^{m+n-1})}{mn}$$

$$C = \frac{1}{n} + \frac{1}{m} + 1 - q^m - q^n + q^{m+n-1}.$$

Exemplo: Procedimento para o teste Elisa.

Têm-se 96 indivíduos. Amostras combinadas são então formadas com 12 indivíduos em cada linha e 8 indivíduos em cada coluna. O primeiro estágio consiste de $8+12=20$ testes e o segundo estágio consiste em testar todos os indivíduos localizados nas linhas e colunas cuja intersecção foi identificada como infectada no primeiro estágio.

$$E(X) = \left\{ 20 + 96 \left[1 - (1-p)^{12} - (1-p)^8 + (1-p)^{19} \right] \right\}$$

$$\Rightarrow E(X) = 116 - 96(1-p)^{12} - 96(1-p)^8 + 96(1-p)^{19}.$$

A Figura 2 expressa $E(X)$ para $0 \leq p \leq 1$. Pelo gráfico, nota-se que o método MBM somente é vantajoso para $p \leq 0,2$, pois acima desse valor gastariam-se mais testes do que se fossem feitos testes individuais.

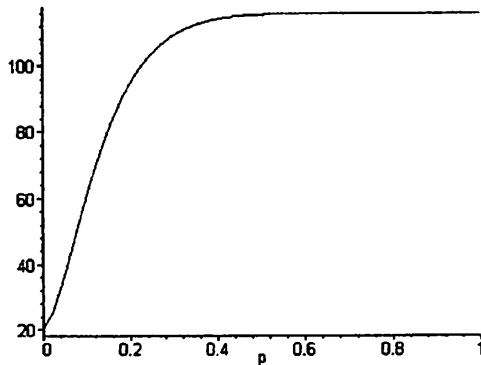


FIGURA 2 – Esperança do número total de testes para o MBM, sendo $m = 12$ e $n = 8$

O custo é dado por $C = 1 + 12^{-1} + 8^{-1} - q^{12} - q^8 + q^{19}$

$$\Rightarrow C = \frac{29}{24} - (1-p)^{12} - (1-p)^8 + (1-p)^{19}.$$

A Figura 3 apresenta o custo para $0 \leq p \leq 0,2$. Pode-se notar, pelo gráfico, que o custo é igual a 0,95 para $p = 0,2$, confirmando ser o MBM vantajoso apenas para $p \leq 0,2$.

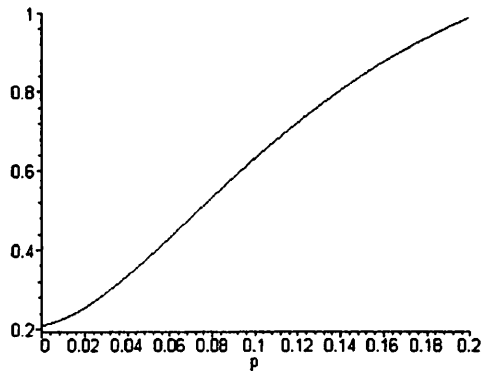


FIGURA 3 - Custo dos testes para o MBM, sendo $m = 12$ e $3 n = 8$

3 METODOLOGIA

Propõe-se um novo método de triagem por grupos. Comparações relativas ao custo foram feitas por meio analítico e processos gráficos, entre os métodos apresentados.

Com relação ao tamanho da população, foram considerados dois tamanhos que tiveram por objetivo refletir situações relativamente extremas, isto é, populações grandes e populações pequenas.

O pressuposto de que os elementos são independentes é a condição básica para que o problema admita uma solução explícita. Foi analisado o caso em que se sabe, a priori, que um valor fixado p de indivíduos na população seja efetivo. Neste caso, o número de efetivos nos grupos passa a ser agora variável aleatória dependente, e o tratamento teórico se torna extremamente complexo. Adotou-se, então, a estratégia de simulação computacional. Para efeito de comparação, foi desenvolvido um programa para o caso independente.

3.1 Simulação dos dados

Foram implementados quatro programas, utilizando o pacote estatístico SAS® (1990) para simular o número total de testes a serem realizados em uma população com uma probabilidade de ocorrência p e quando se tem um valor fixado de indivíduos efetivos na população, com o objetivo de comparar o procedimento MBM com o método de Watson. Várias comparações foram feitas considerando os casos independente e dependente. O procedimento SAS® (1990) utilizado para encontrar a média e o desvio padrão do número médio de testes necessários foi o PROC IML.

4 RESULTADOS E DISCUSSÃO

4.1 Procedimento divisão pela metade.

Um procedimento possível de ser implementado, graças aos avanços tecnológicos atuais que permitem amostras com apenas traços de substâncias químicas, que propicia uma importante economia no número de testes a serem realizados é o método que será denominado de divisão pela metade (procedimento semelhante ao descrito em Sobel & Elashoff (1975) para estimar taxas de infecção).

Seja:

N : população;

p : probabilidade de um indivíduo ser efetivo.

Formam-se dois grupos com $K = \frac{N}{2}$ indivíduos. Aqui é necessário que se mantenha um registro sobre em que grupo cada indivíduo se encontra. Os dois grupos são testados e toma-se um deles como sendo efetivo. Novamente, com os indivíduos deste grupo, formam-se dois grupos de tamanho $K_1 = \frac{N}{4}$. O processo é repetido até que um grupo efetivo com apenas um indivíduo seja obtido. Este é então retirado da população e todo o procedimento é repetido.

A Figura 4 expressa o procedimento de divisão pela metade, considerando uma população de 16 indivíduos, em que:

Divide-se a população em dois grupos e recusa-se um grupo;

Divide-se o grupo infectado em dois subgrupos e recusa-se um deles;

Proseguem-se as subdivisões nos subgrupos infectados até encontrar o indivíduo infectado.

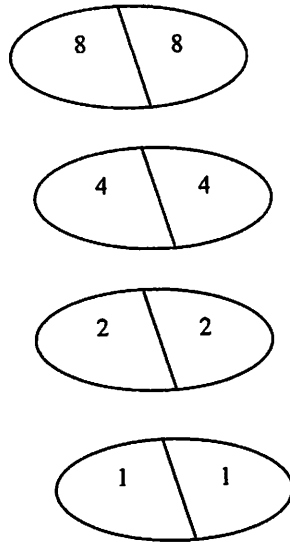


FIGURA 4 – Procedimento divisão pela metade em que um indivíduo seja efetivo, sendo $N = 16$, $K = 8$, $K_1 = 4$, $K_2 = 2$, $K_3 = 1$

No total, foram realizados quatro testes para se detectar um indivíduo doente. O número de testes para se detectar o primeiro indivíduo efetivo é $\log_2(N)$. Se N não é potência de 2, este número será o maior inteiro próximo de $\log_2(N)$.

Para se detectar o segundo indivíduo efetivo, o número de testes necessários é $\log_2(N-1)$ (ou o inteiro mais próximo). Para o terceiro, $\log_2(N-2)$, e assim sucessivamente. O número esperado de efetivos é Np ; portanto, o número esperado de testes é:

$$E(X) = \log_2(N) + \log_2(N-1) + \dots + \log_2(N - Np + 1)$$

$\cong Np \log_2 N$ (se p é suficientemente pequeno).

Conseqüentemente, o custo é $C = \frac{Np \log_2(N)}{N} = p \log_2(N)$.

4.2 Comparação de custos

Foram feitas comparações algébricas e comparações exatas utilizando o software de computação algébrica Maple V (1998) para comparar os custos do procedimento MBM com os métodos de divisão pela metade e o método de Watson.

4.2.1 MBM versus método de Watson

Quando se considera o caso em que o tamanho da população é grande, tem-se, para o MBM:

$$E(X) = m+n+mn(1-q^m - q^n + q^{m+n-1}).$$

Sabe-se que $q^m \cong 1 - mp$ e $q^n \cong 1 - np$, então $q^{m+n-1} \cong 1 - (m+n-1)p$. Dessa forma, substituindo e simplificando, tem-se que:

$$E(X) = m + n + mnp.$$

Tem-se também que se m e n são grandes e próximos, então $m \cong n$ e: $E(X) \cong 2m + m^2p$.

No método de Watson (seção 2.1), tem-se que:

$$E(X) = 2N\sqrt{p}$$

$$\Rightarrow E(X) \cong 2m^2\sqrt{p}.$$

Ao se compararem os dois métodos, tem-se que $2m + m^2p < 2m^2\sqrt{p}$, pois $2 + mp < 2m\sqrt{p}$ e $mp \ll m\sqrt{p}$.

Logo, pode-se concluir que para valores grandes de N (no exemplo $N = 10000$, ou seja, $m = 100$), o MBM gera um menor número de testes, como ilustra a Figura 5, sendo $E(X) \cong 2m^2\sqrt{p}$ para o método de Watson e $E(X) \cong 2m + m^2p$ para o procedimento MBM.

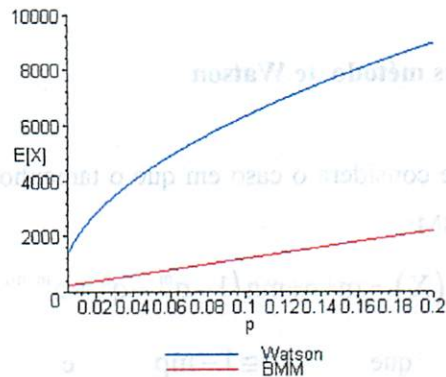


FIGURA 5 - Número esperado de testes (MBM e método de Watson) para $N = 10000$ e $0,005 < p < 0,2$

Quando se considera o caso em que o tamanho da população é pequeno, por exemplo, $N = 96$, $m = 12$, $n = 8$ e $0,005 < p < 0,2$, tem-se que o procedimento MBM gera um menor número de testes apenas no intervalo $0,014 < p < 0,1$, como ilustra a Figura 6, sendo $E(X) \cong 2mn\sqrt{p}$ para o

método de Watson e $E(X) = m + n + mn(1 - q^m - q^n + q^{m+n-1})$ para o procedimento MBM.

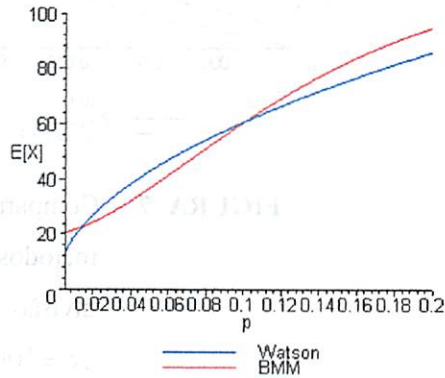


FIGURA 6 - Número esperado de testes (MBM e método de Watson) para $N = 96$ e $0,005 < p < 0,2$

4.2.2 MBM versus método de Watson versus método divisão pela metade

As Figuras de 7 a 10 ilustram a comparação do custo dos testes para os métodos MBM, Watson e o da divisão pela metade, para $N = 10000$ e $N = 100$. Verificou-se que para $N = 10000$, sendo $C = 2\sqrt{p}$ (Watson), $C = p \log_2 10000$ (divisão pela metade) e $C = 0,02 + p$ (MBM), o custo no método de divisão pela metade é menor que no método de Watson para $0 \leq p \leq 0,0227$ (Figura 7), enquanto, para $0,0227 \leq p \leq 0,2$ (Figura 8), esta relação se inverte.

O procedimento MBM gerou um custo menor do que os demais para qualquer valor de p , como indicam as Figuras 7 e 8.

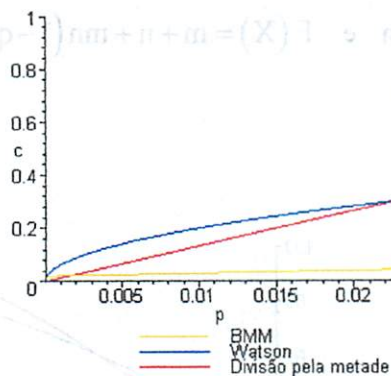


FIGURA 7 - Comparação de custos entre os métodos MBM, Watson e o da divisão pela metade para $N = 10000$ e $0 \leq p \leq 0,0227$

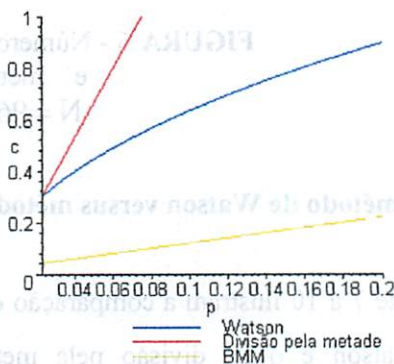


FIGURA 8 - Comparação de custos entre os métodos MBM, Watson e o da divisão pela metade para $N = 10000$ e $0,0227 \leq p \leq 0,2$

Quando $m = n = 10$, ou seja, $N = 100$, sendo $C = 2\sqrt{p}$ (Watson),

$$C = p \log_2 100 \quad (\text{divisão pela metade}) \quad \text{e} \quad C = \frac{1}{n} + \frac{1}{m} + 1 - q^m - q^n + q^{m+n-1}$$

(MBM), verificou-se que o custo no método de divisão pela metade é menor que

os outros dois métodos considerando $0,01 \leq p \leq 0,09$ (Figura 9), enquanto que para $0,09 \leq p \leq 0,2$ (Figura 10), o custo no método descrito em Watson é menor que o dos outros dois métodos, sendo o custo no procedimento MBM menor que no método de divisão pela metade.

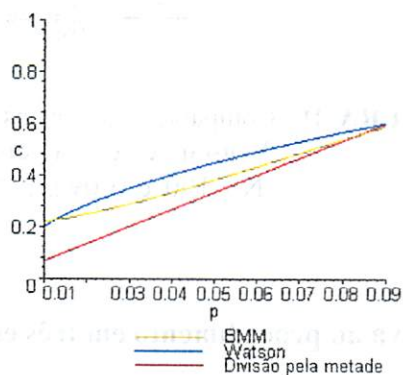


FIGURA 9 - Comparação de custos entre os métodos MBM, Watson e o da divisão pela metade para $N = 100$ e $0,01 \leq p \leq 0,09$.

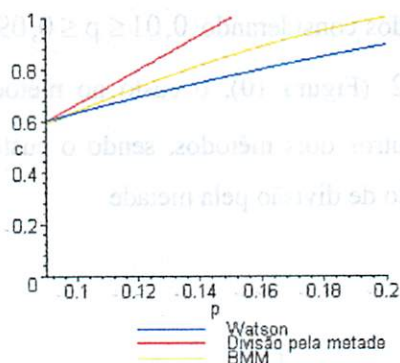


FIGURA 10. Comparação de custos entre os métodos MBM, Watson e o da divisão pela metade para $N = 100$ e $0,09 \leq p \leq 0,2$

4.3 Uma alternativa ao procedimento em três estágios

O procedimento em três estágios, descrito na seção 2.2.1, consiste no fato de que todos os elementos dos grupos efetivos formam novamente uma população, que será novamente dividida em grupos. Estes novos grupos não guardam nenhuma relação com os grupos anteriores.

No estudo a seguir, um grupo efetivo é dividido em subgrupos. Este procedimento será denominado de subdivisão dos grupos efetivos. A Figura 11 ilustra este procedimento.

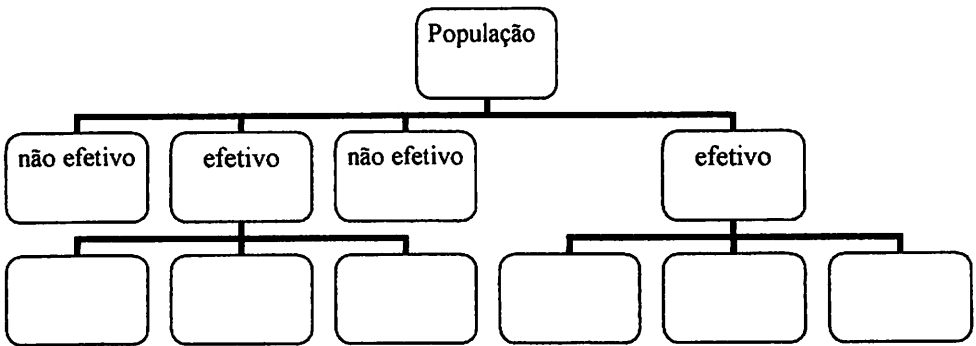


FIGURA 11 – Procedimento de subdivisão dos grupos efetivos.

O tamanho destes subgrupos foi obtido aplicando-se a teoria para dois estágios, isto é, o tamanho ótimo quando se considera como população o grupo.

Sejam:

N : população;

p : probabilidade de ocorrência;

$K \cong \frac{1}{\sqrt{p}}$: tamanho (ótimo) dos grupos;

$m = \frac{N}{K}$: número de grupos;

$1 - q^k$: probabilidade de um grupo ser efetivo;

B : variável aleatória número de grupos efetivos.

B é uma variável binomial com parâmetros m e $1 - q^k$, portanto, sua esperança é:

$$E(B) = m(1 - q^K) = \frac{N}{K}(1 - q^K).$$

Para se dividir um grupo efetivo em subgrupos de tamanho ótimo, é necessário utilizar a teoria das probabilidades condicionais (Casella & Berger, 1990), isto é, a probabilidade de ocorrência p_1 , dado que o grupo é efetivo. Tem-se que:

$$p_1 = \frac{p}{1 - q^K}.$$

Sejam:

$K_1 \cong \frac{1}{\sqrt{p_1}}$: tamanho (ótimo) dos subgrupos dos grupos efetivos;

$n = \frac{K}{K_1}$: número de subgrupos para cada grupo efetivo;

B_j : variável aleatória número de subgrupos efetivos de um grupo efetivo.

Novamente B_j é uma variável binomial com parâmetros n e $1 - q_1^{K_1}$ ($q_1 = 1 - p_1$).

A esperança de B_j é dada por $n(1 - q_1^{K_1}) = \frac{K}{K_1}(1 - q_1^{K_1})$. Portanto a esperança do número de testes para cada grupo efetivo é

$$\frac{K}{K_1} + K_1 \left[\frac{K}{K_1} (1 - q_1^{K_1}) \right].$$

Se X é a variável aleatória, número total de testes, então:

$$E(X) = \frac{N}{K} + E(B) \left[\frac{K}{K_1} + K(1 - q_1^{K_1}) \right]$$

$$= \frac{N}{K} + \frac{N}{K}(1-q^K) \left[\frac{K}{K_1} + K(1-q_1^{K_1}) \right].$$

Do Teorema Binomial da seção 2.1 e, sendo p e p_1 pequenos, têm-se as aproximações:

$$(1-q^K) = 1 - (1-p)^K \cong 1 - (1-Kp) \cong Kp$$

$$(1-q_1^{K_1}) = 1 - (1-p_1)^{K_1} \cong 1 - (1-K_1p_1) \cong K_1p_1;$$

em que:

$$E(X) \cong \frac{N}{K} + \frac{N}{K} p K \left(\frac{K}{K_1} + K p_1 K_1 \right);$$

$$E(X) \cong \frac{N}{K} + Np \frac{K}{K_1} + Np K p_1 K_1.$$

Como $pp_1 \cong 0$, então $E(X) \cong \frac{N}{K} + Np \frac{K}{K_1}$.

Tem-se que $p_1 = \frac{p}{1-q^K} = \frac{p}{1-(1-Kp)} \cong \frac{p}{Kp} \cong \frac{1}{K}$,

em que $K_1 = \frac{1}{\sqrt{p_1}} = \frac{1}{\sqrt{1/K}} = \sqrt{K}$ e:

$$E(X) \cong \frac{N}{K} + Np \frac{K}{\sqrt{K}}$$

$$= \frac{N}{K} + Np\sqrt{K} = \frac{N}{1/\sqrt{p}} + Np\sqrt{1/\sqrt{p}} = N\sqrt{p} + \frac{Np}{\sqrt[4]{p^3}} \cong N\sqrt{p}.$$

O custo deste método é $C = \frac{N\sqrt{p}}{N} = \sqrt{p} = \frac{1}{\sqrt{r}}$.

Comparando com o custo mínimo para o procedimento em três estágios

exposto em Finucan (1965), tem-se que o custo no método de subdivisão dos grupos efetivos (Figuras 12 e 13) é maior para $0 \leq p \leq 0,0014$ e menor para $0,0014 \leq p < 0,2$.

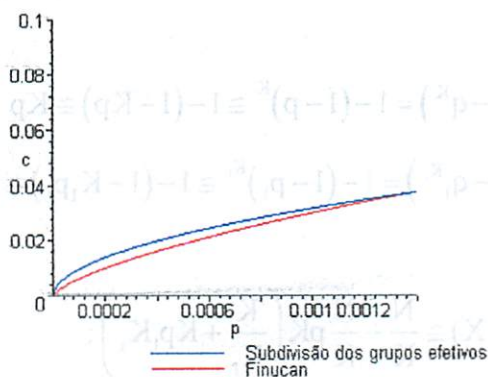


FIGURA 12 - Comparação de custos de subdivisão dos grupos efetivos versus método descrito em Finucan $0 \leq p \leq 0,0014$

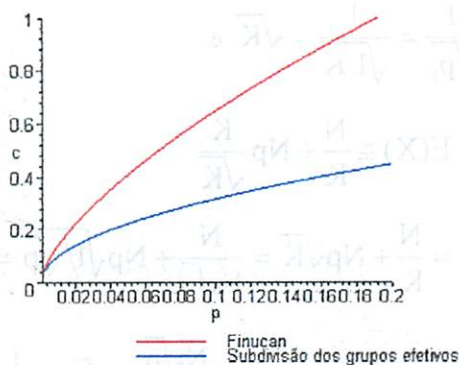


FIGURA 13 - Comparação de custos de subdivisão dos grupos efetivos versus método descrito em Finucan $0,0014 \leq p \leq 0,2$

4.4 Comparação via simulação computacional dos casos dependente e independente

Como já foi observado no caso independente, os indivíduos são estocasticamente independentes, tendo uma probabilidade p de serem efetivos, enquanto, no caso dependente, existe um valor fixado p de indivíduos efetivos na população. Os casos são de natureza semelhante, porém o segundo não admite solução explícita ou esta é muito complexa. Para compará-los, foram realizadas 1000 simulações com uma população de $N = 10000$ e vários valores de K para o tamanho dos grupos. O número de estágios é dois, isto é, ao se testar um grupo efetivo, todos os elementos deste grupo são testados individualmente.

As Tabelas 4 e 5 expressam os valores médios do número de testes realizados com seus respectivos desvios padrão. Os resultados para a esperança do número total de testes são próximos. Tal fato era esperado, pois sendo N grande e o valor fixado p de indivíduos efetivos pequeno, a dependência fica diluída e se aproxima de independência. Com relação ao desvio padrão, tem-se que o caso dependente apresenta resultados expressivamente menores, como, por exemplo, quando se considera $p = 0,0004$ (Tabelas 4 e 5), em que o tamanho ótimo do grupo é 50 e o desvio padrão no caso independente é igual a 99,76 (Tabela 4), enquanto, no caso dependente, é igual a 9,39 (Tabela 5), implicando numa menor variabilidade dos dados em relação ao caso independente. Pode-se observar que quando o valor de p aumenta o tamanho do grupo diminui consideravelmente como por exemplo para $p = 0,2$ em que, o tamanho ótimo é próximo de 2. Dessa forma a economia com os testes diminui consideravelmente, Isto confirma o fato de que o método de Watson somente é viável para valores de p pequenos conforme os resultados descritos na Tabela 2

K	$p = 0,0004$	$p = 0,004$	$p = 0,004$	$p = 0,04$	$p = 0,0625$	$p = 0,2$
2	5008 (0)	5078 (0,50)	5784 (5,36)	6211 (8,00)	8602 (21,35)	
4	2516 (0,22)	2655 (1,96)	4008 (16,99)	4777 (25,00)	8406 (56,70)	
5	2020 (0,27)	2193 (2,78)	3844 (24,96)	4760 (35,00)	8725 (60,92)	
16	689 (1,33)	1230 (16,65)	5419 (103,32)	7073 (125,00)	9929 (61,31)	
25	500 (3,14)	1332 (29,56)	6797 (150,50)	8403 (146,00)	10362 (29,77)	
50	399 (9,39)	1975 (81,06)	8896 (198,65)	9809 (131,00)	10200 (0)	
80	441 (17,11)	2811 (150,66)	9745 (155,17)	10066 (67,00)	10125 (0)	

TABELA 5 - Número esperado de testes com respectivos desvios padrão entre parênteses, em que $N = 10000$, variando-se K, o tamanho dos grupos, tendo-se um valor fixado p de indivíduos efetivos na população para o caso dependente

K	$p = 0,0004$	$p = 0,004$	$p = 0,004$	$p = 0,04$	$p = 0,0625$	$p = 0,2$
2	5006 (4,15)	5078 (12,00)	5784 (39,00)	6212 (46,00)	8601 (69,00)	
4	2517 (8,28)	2655 (25,00)	4008 (71,00)	4771 (83,00)	8402 (100,00)	
5	2020 (10,10)	1371 (24,00)	3843 (87,00)	4755 (102,00)	8720 (104,00)	
16	689 (31,06)	1230 (97,00)	5429 (203,00)	7863 (190,00)	9928 (64,05)	
25	499 (48,86)	1331 (147,00)	6794 (235,00)	8411 (199,00)	10364 (30,00)	
50	401 (99,76)	1979 (268,00)	8886 (250,00)	9804 (141,00)	10200 (4,00)	
80	451 (156,37)	2806 (406,00)	9748 (176,00)	10064 (71,00)	10125 (0)	

TABELA 4 - Número esperado de testes com respectivos desvios padrão entre parênteses, em que $N = 10000$, variando-se K, o tamanho dos grupos, em função da probabilidade de ocorrência p para o caso independente

As Tabelas 6 e 7 expressam os valores médios do número de testes realizados com seus respectivos desvios padrão. Foram realizadas 1000 simulações com uma população de $N = 100$ e vários valores de K para os grupos. Não se observou diferença considerável no número de testes. Com relação ao desvio padrão, tem-se que o caso dependente apresenta resultados expressivamente menores, implicando numa menor variabilidade dos dados em relação ao caso independente.

TABELA 6 - Número esperado de testes com respectivos desvios padrão entre parênteses, em que $N = 100$, variando-se K , o tamanho dos grupos, em função da probabilidade de ocorrência p para o caso independente

K	$p = 0,02$	$p = 0,04$	$p = 0,1$	$p = 0,2$	$p = 0,3$
2	54 (2,73)	58 (3,83)	69 (5,51)	86 (6,66)	100 (7,05)
4	33 (5,54)	40(7,22)	59 (9,27)	84 (9,63)	101 (8,40)
5	30 (6,64)	39 (8,65)	61 (11,00)	87 (10,80)	102 (8,37)
10	28 (11,98)	44 (14,69)	75 (15,26)	100 (9,74)	107 (5,20)
20	39 (21,43)	61 (22,13)	93 (14,14)	104 (4,67)	105 (1,26)
50	67 (34,30)	89 (23,83)	101 (4,98)	102 (0)	102 (0)

TABELA 7 - Número esperado de testes com respectivos desvios padrão entre parênteses, em que $N = 100$, variando-se K , o tamanho dos grupos, tendo-se um valor fixado p de indivíduos efetivos na população para o caso dependente

K	$p = 0,02$	$p = 0,04$	$p = 0,1$	$p = 0,2$	$p = 0,3$
2	54 (0,22)	58 (0,50)	69 (1,18)	86 (2,24)	101 (2,96)
4	33 (0,73)	40 (1,63)	60 (3,78)	85 (5,58)	102 (5,57)
5	30 (1,05)	39 (2,36)	62 (4,75)	88 (6,73)	104 (6,47)
10	29 (2,90)	45 (6,02)	77 (9,57)	101 (8,40)	107 (4,38)
20	41 (8,28)	63 (13,15)	96 (11,90)	104 (3,77)	105 (1,09)
50	77 (25,01)	96 (15,40)	102 (0)	102 (0)	102 (0)

4.4.1 Procedimento MBM versus método de Watson.

As Tabelas 8, 9, 10 e 11 expressam os valores médios do número de testes realizados com seus respectivos desvios padrão. Foram realizadas 1000 simulações com uma população de $N = 96$ (procedimento para o teste Elisa) e vários valores de K para os grupos.

O procedimento MBM é viável somente para valores de m e n próximos de uma matriz retangular e valores pequenos de p como descrito em Berger et al.(2000), como pode-se verificar nas Tabelas 8 e 10.

Para $p = 0,05$, em que $m = 8$ e $n = 12$ (Tabelas 8 e 10), o procedimento MBM gera um menor número de testes tanto para o caso independente quanto para o caso dependente, se comparado com o método de Watson (Tabelas 9 e 11). Para $p = 0,2$, em que $m = 8$ e $n = 12$, o procedimento MBM gera um maior

<i>K</i>	<i>m</i>	$p = 0,01$	$p = 0,05$	$p = 0,1$	$p = 0,2$
8	12	20 (7,48)	45 (13,16)	68 (13,67)	91 (10,74)
6	16	22 (5,67)	41 (10,81)	62 (11,66)	86 (10,94)
4	24	28 (3,96)	43 (7,75)	59 (9,95)	80 (9,50)
2	48	50 (1,91)	58 (4,25)	67 (5,51)	82 (6,59)

TABELA 9 - Número esperado de testes com respectivos desvios padrão entre parênteses, em que $N = 96$, variando-se K , o tamanho dos grupos, em função da probabilidade de ocorrência p , para o método de Watson considerando o caso independente

<i>m</i>	<i>n</i>	$p = 0,01$	$p = 0,05$	$p = 0,1$	$p = 0,2$
8	12	22 (2,40)	37 (11,22)	63 (17,04)	94 (13,48)
6	16	24 (2,19)	39 (11,13)	62 (14,12)	90 (12,35)
4	24	30 (2,41)	43 (8,84)	60 (10,61)	84 (10,25)
2	48	51 (1,71)	60 (4,75)	69 (5,96)	84 (6,71)

TABELA 8 - Número esperado de testes com respectivos desvios padrão entre parênteses, em que $N = 96$, variando-se K , o tamanho dos grupos, em função da probabilidade de ocorrência p , para o MBM considerando o caso independente

número de testes nos dois casos considerados. Estes resultados estão em padrão, tem-se que o caso dependente apresenta resultados expressivamente menores ao se considerarem os dois procedimentos.

<i>K</i>	<i>m</i>	$d = 0,01$	$d = 0,05$	$d = 0,1$	$d = 0,2$
2	48	50 (0)	58 (0,64)	67 (1,23)	83 (2,22)
4	24	28 (0)	43 (1,99)	59 (3,69)	81 (5,14)
6	16	22 (0)	43 (3,50)	63 (6,12)	87 (7,30)
8	12	20 (0)	46 (5,48)	70 (8,07)	93 (7,73)

TABELA 11 - Número esperado de testes com respectivos desvios padrão entre parênteses, em que $N = 96$, variando-se K , o tamanho dos grupos, tendo-se um valor fixado p de indivíduos efetivos na população para o método de Watson considerando o caso dependente

<i>m</i>	<i>n</i>	$d = 0,01$	$d = 0,05$	$d = 0,1$	$d = 0,2$
2	48	51 (0)	57 (1,30)	69 (1,24)	84 (2,24)
4	24	29 (0)	39 (2,60)	61 (4,75)	85 (5,36)
6	16	23 (0)	34 (2,73)	62 (7,60)	92 (8,17)
8	12	21 (0)	32 (2,84)	63 (8,70)	96 (9,55)

TABELA 10 - Número esperado de testes com respectivos desvios padrão entre parênteses, em que $N = 96$, variando-se K , o tamanho dos grupos, tendo-se um valor fixado p de indivíduos efetivos na população para o MBM considerando o caso dependente

5 CONCLUSÕES

1. O método de subdivisão em grupos apresenta ganhos de economia em relação ao método em três estágios proposto em Finucan (1965), para o intervalo $0,0014 \leq p \leq 0,2$.
2. O procedimento de divisão pela metade apresenta a esperança do número de testes comparável aos outros procedimentos descritos na literatura.
3. O procedimento MBM é mais eficiente no caso em que $m \cong n$ e N grande, apresentando um custo menor que os procedimentos Watson e divisão pela metade. Para o caso que N é pequeno, seu custo é maior que o procedimento Watson, porém este é um teste mais confiável, uma vez que cada indivíduo é testado duas vezes.
4. Por meio de simulação computacional, foi constatado que o número esperado de testes é semelhante para os casos em que se considera dependência ou independência. No entanto, o caso independente apresenta desvio padrão expressivamente maior que o caso dependente.

6 REFERÊNCIAS BIBLIOGRÁFICAS

BERGER, T.; MANDELL, J. W.; SUBRAHMANYA, P. Maximally Efficient Two-Stage Screening. **Biometrics**, Washington, v. 56, n. 3, p. 833-840, Sept. 2000.

CASELLA, G.; BERGER, R. L. **Statistical Inference**. Belmont, CA: Duxbury Press, 1990. 650 p.

CHEN, C. L.; SWALLOW, W. H. Using Group Testing to Estimate a proportion, and to test the Binomial Model. **Biometrics**, Washington, v. 46, n. 4, p. 1035-1046, Dec. 1990.

DORFMAN, R. The Detection of Defective Members of Large Populations. **Annals of Mathematical Statistics**, Washington, v.14, p. 436-440, 1943.

FELLER, W. **An Introduction to Probability Theory and its Applications**. New York: John & Wiley, 1968.v. 1, 460 p.

FINUCAN, H. M. The Blood Testing Problem. **Journal of the Royal Statistical Society, series C-Applied Statistics**, London, v. 13, n. 1, p. 43-50, 1965.

HEAL, K. M.; HANSEN, M. L.; RICKARD, K. M. **Maple V; Learning Guide**. Waterloo: Waterloo Maple, 1998. 284 p. (Release, 5).

HEPWORTH, G. Exact Confidence Intervals for Proportions Estimated by Group Testing. **Biometrics**, Washington, v. 52, n. 3. p. 1134-1146, Sept. 1996.

HUNG, M. C.; SWALLOW, W. H. Use of Binomial Group Testing in Tests of Hypotheses for Classification or Quantitative Covariables. **Biometrics**, Washington, v. 56, n. 1, p. 204-212, Mar. 2000.

IVANOVA, T.; MALONE, L.; MOLLAGHASEMI, M. Comparison of a Two-Stage Group-Screening Design to a standard 2^{K-p} . Design for a Whole-line semiconductor manufacturing simulation model. In: Proceedings of the 1999 Winter Simulation conference. 1999, p. 640-646. Disponível em: <<http://www.informs-cs.org/wsc99papers/091>>. Acesso em: 10 de janeiro de 2003.

JOHNSON, W. O.; GASTWIRTH, J. L. Dual Group Screening. **Journal of Statistical Planning and Inference**, Amsterdam, v. 83, n. 2, p. 449-473, Feb. 2000.

MAURO, C. A. On the Performance of Two-Stage Group Screening Experiments. **Technometrics**, London, v. 26, n. 3, p. 371-388, Aug. 1984.

MAURO, C. A.; SMITH D. E. The Performance of 2-stages Group Testing in factor screening Experiments. **Technometrics**, London, v. 24, n. 4, p. 325-330, 1982.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to theory of statistics**. 3. ed. New York: Wiley & Sons, 1974. 842 p.

PATEL, M. S. Group screening with more than 2-stages. **Technometrics**, London, v. 4, n.2, p. 209-&, 1962.

SAS INSTITUTE / STAT. **SAS/ IML Software: Image and reference**, version 6. Cary, 1990. 501 p.

SOBEL, M.; ELASHOFF, R. M. Group testing with a new goal, estimation. **Biometrika**, Cambridge, v. 62, n. 1, p. 181-193, 1975.

WATSON, G. S. A study of the Group Screening Method. **Technometrics**, London, v. 3, p. 371-388, Aug. 1961.

WILKS, S. S. **Mathematical statistics**. New York: J. Wiley, 1962. 644 p.

ANEXOS

ANEXO

Página

- SIMULAÇÃO 1** Método de Watson – caso independente. Programa de simulação para determinar a distribuição do número de exames a serem realizados em uma população com probabilidade de sucesso p para cada indivíduo, sendo p essa probabilidade, N o tamanho da população, K o número de exames por grupo, m o número de grupos e X o número total de exames. 55
- SIMULAÇÃO 2** Método de Watson – caso dependente Programa de simulação para determinar a distribuição do número de exames a serem realizados em uma população com probabilidade de sucesso p para cada indivíduo, sendo p essa probabilidade, N o tamanho da população, K o número de exames por grupo, m o número de grupos e X o número total de exames. Nesse caso, o número de pessoas doentes na população será N_1 ($p = N_1/N$). Os indivíduos doentes serão sempre os N_1 primeiros: 1, 2, ..., N_1 . 56

SIMULAÇÃO 3 Procedimento MBM – caso independente Programa de simulação para determinar a distribuição do número de exames a serem realizados em uma população com probabilidade de sucesso p para cada indivíduo, sendo p essa probabilidade, mn o tamanho da população, m o número de linhas, n o número de colunas e X o número total de exames. 58

SIMULAÇÃO 4 Procedimento MBM- caso dependente Programa de simulação para determinar a distribuição do número de exames a serem realizados em uma população com probabilidade de sucesso p para cada indivíduo, sendo p essa probabilidade, mn o tamanho da população, m o número de linhas, n o número de colunas e X o número total de exames . Nesse caso, o número de pessoas doentes na população será N_1 ($p = N_1/N$). Os indivíduos doentes serão sempre os N_1 primeiros: 1, 2, ..., N_1 (população com doentes fixos). 60

SIMULAÇÃO 1

Método de Watson - caso independente. Programa de simulação para determinar a distribuição do número de exames a serem realizados em uma população com probabilidade de sucesso p para cada indivíduo, sendo p essa probabilidade, N o tamanho da população, K o número de exames por grupo, m o número de grupos e X o número total de exames.

```
Options ps=500 ls=80 nodate nonumber;
Proc Iml;
  p=0.0002;
  N=1024;k=32;m=32;/* restrito a m*K=N */
  nsim=1000;
  Xv=j(nsim,1,0);
  do i=1 to nsim;
    X=0;
    do ii=1 to m;
      achou=0;
      do iii=1 to K;
        u=ranuni(0);
        if u<=p then achou=1;
      end;
      if achou=1 then X=X+K+1;
      else X=X+1;
    end;
    Xv[i]=X;
  end;
  *print Xv;
  create elcio from Xv;
  append from Xv;
Quit;
Proc Univariate data=elcio normal;
  Var coll;
  Histogram coll/normal;
Run;Quit;
```

SIMULAÇÃO 2

Método de Watson - caso dependente. Programa de simulação para determinar a distribuição do número de exames a serem realizados em uma população com probabilidade de sucesso p para cada indivíduo, sendo p essa probabilidade, N o tamanho da população, K o número de exames por grupo, m o número de grupos e X o número total de exames. Nesse caso, o número de pessoas doentes na população será $N1$ ($p = N1/N$). Os indivíduos doentes serão sempre os $N1$ primeiros: 1, 2, ..., $N1$.

```
Options ps=5000 ls=80 nodate nonumber;
```

```
Proc Iml;
  N=1024;N1=21;p=N1/N;
  print N N1 p;
  Pop=j(N,1,0);
  do i=1 to N;
    pop[i]=i;
  end;
  k=16;m=64;/* restrito a m*K=N */
  nsim=1000;
  Xv=j(nsim,1,0);
  amostra=j(K,1,0);
  do i=1 to nsim;
    X=0;NN=N;PopAux=Pop;
    do ii=1 to m;
      do iii=1 to K;
        u=int(ranuni(0)*NN)+1;
        amostra[iii]=popaux[u];
        popaux[u]=popaux[nn];
        NN=NN-1;
      end;
      *print amostra;
      achou=0;
    do iii=1 to k;
      u=amostra[iii];
      if u<=N1 then achou=1;
    end;
  end;
```

```
    if achou=1 then X=X+K +1;
      else X=X+1;
        *print achou X;
          end;
    Xv[i]=X;
      end;
*print Xv;
create elcio from Xv;
append from Xv;
Quit;
Proc Univariate data=elcio normal;
  Var col1;
  Histogram col1/normal;
Run;Quit;
```

SIMULAÇÃO 3

Procedimento MBM - caso independente. Programa de simulação para determinar a distribuição do número de exames a serem realizados em uma população com probabilidade de sucesso p para cada indivíduo, sendo p essa probabilidade, mn o tamanho da população, m o número de linhas, n o número de colunas e X o número total de exames.

```
Proc lml;
p=0.20;
m=3;n=5;
nsim=10;
Xv=j(nsim,1,0);
do i=1 to nsim;
  pop=j(m,n,0);
  linha=j(m,1,0);
  col=j(n,1,0);
  X=m+n;
  do ii=1 to m;
    do iii=1 to n;
      u=ranuni(0);
      if u<=p then
        pop[ii,iii]=1;
      if pop[ii,iii]=1 then
        do;
          linha[ii]=1;
          col[iii]=1;
        end;
      end;
    end;
  end;
  do ii=1 to m;
    do iii=1 to n;
      if (linha[ii]=1 & col[iii]=1) then x=x+1;
    end;
  end;
  *print x pop linha col;
  Xv[i]=X;
end;
*print Xv;
create elcio from Xv;
append from Xv;
```

```
Quit;  
Proc Univariate data=elcio normal;  
  Var coll;  
  Histogram coll/normal;  
Run;Quit;
```

SIMULAÇÃO 4

Procedimento MBM- caso dependente. Programa de simulação para determinar a distribuição do número de exames a serem realizados em uma população com probabilidade de sucesso p para cada indivíduo, sendo p essa probabilidade, Sendo p essa probabilidade, mn o tamanho da população, m o número de linhas, n o número de colunas e X o número total de exames. Nesse caso, o número de pessoas doentes na população será $N1$ ($p = N1/N$). Os indivíduos doentes serão sempre os $N1$ primeiros: 1, 2, ..., $N1$ (população com doentes fixos).

```
Options ps=500 ls=80 nodate nonumber;
```

```
Proc Iml;  
p=0.20;  
m=8;n=12;  
ndoente=int(p*m*n);  
nsim=1000;  
Xv=j(nsim,1,0);  
amostra=j(m*n,1,0);  
NN=n*m;  
preal=ndoente/NN;  
print p preal ndoente NN;  
do i=1 to nsim;  
  pop=j(m,n,0);  
  linha=j(m,1,0);  
  col=j(n,1,0);  
  X=m+n;  
  NN=m*n;  
  do iv=1 to NN;  
    amostra[iv]=iv;  
  end;  
  *print amostra;  
do iv=1 to ndoente;  
  u=int(ranuni(0)*NN)+1;  
  amostra[NN]=amostra[u];  
  amostra[u]=NN;  
  NN=NN-1;
```

```

end;
  *print amostra;
  NN=m*n;
  do iv=NN-ndocente+1 to NN;
  u=amostra[iv];
  aux=mod(u,n);
  if aux>0 then li=int(u/n)+1;
  else li=int(u/n);
  co=u-(li-1)*n;
  pop[li,co]=1;
  end;
  *print pop;
  do ii=1 to m;
  do iii=1 to n;
  if pop[ii,iii]=1 then
  do;
  linha[ii]=1;
  col[iii]=1;
  end;
  end;
  end;
  do ii=1 to m;
  do iii=1 to n;
  if (linha[ii]=1 & col[iii]=1) then x=x+1;
  end;
  end;
  *print x pop linha col;
  Xv[i]=X;
end;
*print Xv;
create elcio from Xv;
append from Xv;
Quit;
Proc Univariate data=elcio normal;
  Var coll;
  Histogram coll/normal;
Run;Quit;

```