

DISTRIBUIÇÃO DAS MEDIDAS DE INDURAÇÃO E ERITEMA NA PROVA TUBERCULÍNICA: APLICAÇÃO DE UM MÉTODO GRÁFICO DE DECOMPOSIÇÃO DE UMA DISTRIBUIÇÃO DE FREQUÊNCIAS EM COMPONENTES NORMAIS *

Odécio Sanches **

RSPU-B/243

SANCHES, O. — *Distribuição das medidas de induração e eritema na prova tuberculínica: aplicação de um método gráfico de decomposição de uma distribuição de frequências em componentes normais.* **Rev. Saúde públ., S. Paulo**, 9:15-24, 1974.

RESUMO: O teste tuberculínico, segundo a técnica de Mantoux, pretende classificar as pessoas de acordo com o tamanho da reação provocada por uma única e especificada dose de tuberculina. Dados empíricos mostram que as medidas de tais reações, obtidas em populações genéricas, representam a mistura de distintos grupos componentes. Com a finalidade de separar tais grupos, objetivando a solução posterior do problema de classificação, foi utilizado um método gráfico de decomposição de uma distribuição de frequências em componentes normais para decompor distribuições de frequências de medidas de induração e eritema. A decomposição obtida permite verificar que o número de componentes que descreve a distribuição da variável induração não é o mesmo que para a variável eritema, para os dados considerados.

UNITERMOS: *Teste tuberculínico. Técnica de Mantoux. Induração. Eritema. Mistura de distribuições normais. Distribuição normal truncada.*

1. INTRODUÇÃO

Ruffino Netto⁴ (1970) afirma existir evidências de que alguns aspectos mensuráveis na prova tuberculínica devem ser melhor elucidados. Neste sentido, tenta quantificar a relação induração-eritema pressupondo que este par tenha uma distribuição Normal bi-dimensional, o que implica uma distribuição Normal unidimensional para cada uma das marginais.

Os resultados do teste tuberculínico, segundo a técnica de Mantoux, como afirma Guld³ (1953), tem por finalidade classificar pessoas em distintos grupos conforme o tamanho da reação provocada por uma única e especificada dose de tuberculina. Assim sendo, os resultados do referido teste, obtidos em uma população genérica, devem representar uma mistura

* Extraído da tese: "Extensão do Método de Bhattacharya para decomposição de uma distribuição de frequências em componentes normais. Aplicação a um problema biológico", apresentada à Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, em 28-12-73, para obtenção do título de doutor em Ciências

** Do Departamento de Bioestatística e Estatística Vital da Escola de Enfermagem de Ribeirão Preto, USP — Campus de Ribeirão Preto, SP — Brasil

de distintos grupos sendo de se esperar que uma única distribuição de probabilidades dificilmente descreva, de maneira satisfatória, o conjunto de dados observados.

Logo, estudar as distribuições das medidas de induração ou de eritema, obtidas em uma população genérica, significa tentar buscar as distribuições de probabilidades que descrevem os grupos componentes e determinar as proporções em que ocorre tal mistura para posteriormente, se de interesse, estudar o problema de classificação.

O objetivo deste trabalho é, usando um método gráfico aproximado e os dados obtidos por Ruffino Netto⁴ (1970), verificar se a distribuição de freqüências para um conjunto de medidas de induração pode ser decomposta em três grupos distintos como evidencia o trabalho de Edwards et al.² (1959) e, ao mesmo tempo, verificar se a distribuição de freqüências da variável eritema se comporta de forma análoga já que, conforme citação em Ruffino Netto⁴ (1970), vários autores concordam que a variável eritema é tão boa quanto a variável induração para classificar indivíduos submetidos ao teste tuberculínico.

2. MÉTODO GRÁFICO PARA DECOMPOSIÇÃO DE UMA DISTRIBUIÇÃO DE FREQUÊNCIAS EM COMPONENTES NORMAIS COM OU SEM TRUNCAMENTO

Distribuições truncadas ocorrem comumente na prática, sendo o truncamento uma consequência da impossibilidade de se observar determinados valores o que é devido, em geral, à imprecisão do aparelho de medida ou à pequena sensibilidade do método de mensuração empregado.

Sanches⁵ (1973) estuda o método gráfico proposto por Bhattacharya¹ (1967), para decomposição de uma distribuição de freqüências em componentes Normais, a partir da pressuposição que as distribuições componentes sejam truncadas, com

duplo truncamento em pontos conhecidos. Esta hipótese adicional leva ao mesmo resultado obtido por Bhattacharya no que diz respeito à decomposição em si, havendo, no entanto, necessidade da correção para o truncamento para a obtenção das estimativas das proporções da mistura. Os casos de truncamento simples à direita ou à esquerda, e os de não truncamento, são obtidos facilmente como casos particulares.

O método proposto, ao mesmo tempo em que permite obter a separação das distribuições componentes, permite também obter estimativas dos parâmetros das mesmas, desde que algumas hipóteses, as quais serão indicadas a seguir, estejam satisfeitas, pelo menos aproximadamente.

Seja:

$$f(x) = \sum_{i=1}^k p_i f_{it}(x)$$

a função densidade de probabilidade de uma mistura de k distribuições Normais duplamente truncadas, onde:

a) os p_i são números reais, tais que

$$0 < p_i < 1 \text{ e } \sum_{i=1}^k p_i = 1.$$

b) $f_{it}(x)$ é a função densidade de probabilidade de uma Normal duplamente truncada nos pontos ξ_{1i} e ξ_{2i} , $\xi_{1i} < \xi_{2i}$, ambos conhecidos, e de parâmetros μ_i e σ_i^2 ; $i = 1, 2, \dots, k$.

Admitindo como hipóteses que:

a) as distribuições componentes são suficientemente separadas de tal modo a existir uma região, suficientemente ampla, para cada componente, onde os efeitos das demais componentes sejam desprezíveis;

b) a variável esteja agrupada em intervalos de classe de amplitude suficientemente pequena;

consideremos:

N_j a freqüência total da j^{ma} componente;

h amplitude constante de cada intervalo de classe;

$y(x)$: freqüência do intervalo tendo x como ponto médio.

Nestas condições, Sanches ⁵ (1973), vale o seguinte resultado:

$$\Delta \log y(x) \simeq -h \left(\rho_j^2 - \frac{h^2}{12} \right) \left(x - \mu_j + \frac{h}{2} \right) \sigma_j^{-4}$$

onde

$$\Delta \log y(x) = \log y(x+h) - \log y(x)$$

Este resultado mostra que o gráfico de $\Delta \log y(x)$ é, na região onde a j^{ma} componente está convenientemente separada das demais, uma reta com inclinação negativa, para h suficientemente pequeno.

Estimativas μ_j^* e σ_j^{*2} para μ_j e σ_j^2 são obtidas diretamente do gráfico e são dadas por:

$$a) \quad \sigma_j^{*2} = \frac{f_2 h}{f_1} \cotg \theta_j^* - \frac{h^2}{12}$$

onde θ_j^* é a medida do ângulo agudo feito pela reta, correspondente à j^{ma} componente, com o eixo dos x e medido diretamente no gráfico; f_1 e f_2 são dois fatores de escala, respectivamente, para os eixos x e $\Delta \log y$, já que estas variáveis, em geral, estão em escalas distintas.

Se, na prática, forem utilizados logaritmos na base 10 então o valor de f_2 deve ser corrigido multiplicando-o por $\log_{10} e$

$$b) \quad \mu_j^* = \delta_j^* + \frac{h}{2}$$

onde δ_j^* é, no gráfico de $\Delta \log y(x)$, o ponto onde a reta correspondente à j^{ma} componente corta o eixo dos x .

Obtidas as estimativas μ_j^* e σ_j^{*2} as proporções da mistura podem ser estimadas, por exemplo, por:

$$p_j^* = \frac{N_j^*}{\sum N_j^*}; \quad j = 1, 2, \dots, k$$

onde

$$N_j^* = \frac{\sum y}{\sum F_j^*(x)}$$

sendo as somas do numerador e denominador, nesta última expressão, restritas à região em consideração.

Dependendo da existência ou não de truncamento os $F_j^*(x)$ são dados por:

$$F_j^*(x) = \begin{cases} K_{jt}^* \Delta \phi^*(x); & \text{se o truncamento é duplo} \\ K_{je}^* \Delta \phi^*(x); & \text{para truncamento só à esquerda} \\ K_{jd}^* \Delta \phi^*(x); & \text{para truncamento só à direita} \\ \Delta \phi^*(x); & \text{se não há truncamento} \end{cases}$$

onde:

$$\Delta \phi^*(x) = \phi \left(\frac{x + \frac{h}{2} - \mu_j^*}{\sigma_j^*} \right) - \phi \left(\frac{x - \frac{h}{2} - \mu_j^*}{\sigma_j^*} \right)$$

$\phi(\cdot)$: função distribuição acumulada da $N(0;1)$

$$K_{jt}^* = \left\{ \phi\left(\frac{\xi_{2j} - \mu_j^*}{\sigma_j^*}\right) - \phi\left(\frac{\xi_{1j} - \mu_j^*}{\sigma_j^*}\right) \right\}^{-1}$$

$$K_{je}^* = \left\{ 1 - \phi\left(\frac{\xi_{1j} - \mu_j^*}{\sigma_j^*}\right) \right\}^{-1}$$

$$K_{jd}^* = \left\{ \phi\left(\frac{\xi_{2j} - \mu_j^*}{\sigma_j^*}\right) \right\}^{-1}$$

Convém ressaltar que em alguns casos especiais onde a hipótese de conveniente separabilidade entre as componentes não está satisfeita para uma ou algumas das componentes da mistura e, nos casos mais gerais, quando não se sabe se as condições básicas do método estão satisfeitas, a validade do mesmo pode ser estabelecida numa primeira etapa, a partir da construção do gráfico de $\Delta \log y(x)$ para os dados que se tem em mãos. Nestes casos, inicialmente, determinam-se aquelas componentes que são evidenciadas pelo gráfico. As freqüências esperadas para estas componentes são então subtraídas das freqüências observadas obtendo-se o que denominamos freqüências residuais. Se tais freqüências residuais não são desprezíveis o processo é então repetido e novas componentes possivelmente determinadas. Caso ocorra a existência de freqüências residuais não desprezíveis e a inexistência de mais componentes, isto significa que o método não se aplica ao conjunto de dados observados.

3. ESTUDO DAS DISTRIBUIÇÕES DAS VARIÁVEIS INDURAÇÃO E ERITEMA

3.1. Induração

Os dados por nós utilizados, Tabela 1, são uma adaptação daqueles obtidos por

Ruffino Netto⁴ (1970) nos anos de 1967 e 1968 no Complexo Hidroelétrico de Urubupungá. Dos dados originais excluimos as medidas 0 (zero) por significarem, obviamente, ausência de reação. Além disto, por razões de sensibilidade do método de mensuração empregado, truncamos as observações, à esquerda, no ponto correspondente à medida de 2 mm. A tuberculina assim como a dose aplicada e as técnicas de mensuração empregadas encontram-se descritas em Ruffino Netto⁴ (1970).

A Tabela 1 apresenta, além das freqüências consideradas, uma coluna correspondente às diferenças dos logaritmos, calculados na base 10, das freqüências dos intervalos de classe adjacentes.

TABELA 1

Medidas de induração na prova tuberculínica* e diferenças dos logaritmos das freqüências

Induração (mm)	Freqüências (y)	$\Delta \log_{10} y(x_i)$
2 — 4	253	-0,3697
4 — 6	108	-0,4652
6 — 8	37	0,0229
8 — 10	39	-0,1140
10 — 12	30	-0,0969
12 — 14	24	-0,1761
14 — 16	16	-0,3010
16 — 18	8	-0,6021
18 — 20	2	
Total	512	

* Dados adaptados de Ruffino Netto⁴ (1970)

A Fig. 1 mostra o gráfico de $\Delta \log_{10} y(x_i)$, onde x_i é o ponto médio do i^{mo} intervalo; $i = 1, 2, \dots, 8$.

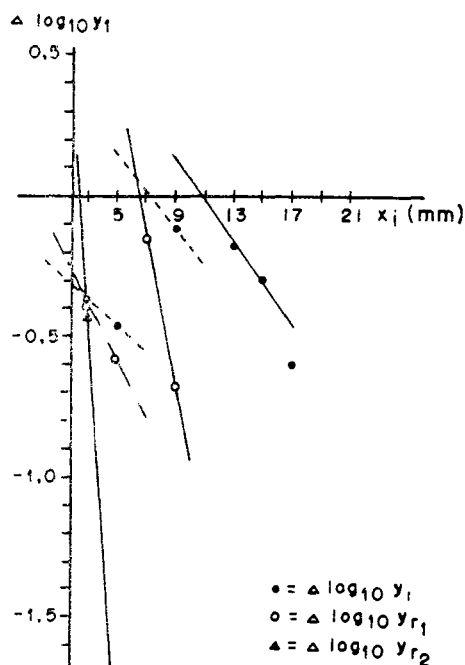


Fig 1 — Medidas de diâmetros de induração, na prova tuberculínica. Dados da Tabela 1.

Um exame deste gráfico mostra 3 pontos razoavelmente bem alinhados: os pontos correspondentes a $x = 11$ mm, $x = 13$ mm e $x = 15$ mm, o que indica, de acordo com o método exposto, a existência de uma componente Normal. Além disto, observamos a possibilidade de outras duas componentes correspondentes à retas passando por apenas 2 pontos, o que, possivelmente, está vinculado a uma acentuada intersecção entre os grupos componentes. Por este motivo a aplicação do método de decomposição ao presente conjunto de dados, com a finalidade de obtenção das estimativas dos parâmetros das diferentes componentes, deve ser feito por etapas.

Inicialmente, utilizando-nos do gráfico da Fig. 1 obtemos as estimativas dos parâmetros correspondentes à reta por 3 pontos a qual, como evidencia a Figura, deve ser a terceira componente.

Da Fig. 1 obtemos:

$$\delta_3^* = 10,80 \text{ mm}; \quad \theta_3^* = 55$$

$$f_1 = 10,00; \quad f_2 = 83,83$$

e desde que $h = 2$ mm. obtemos:

$$\mu_3^* = 11,80 \text{ mm e } \sigma_3^* = 3,43 \text{ mm.}$$

Considerando apenas as freqüências dos intervalos de classe na região que contém os pontos que determinam a reta, obtemos:

Com tais resultados foram calculadas as freqüências esperadas (aproximadas para o inteiro mais próximo) para esta terceira componente. Subtraindo-se tais freqüências das freqüências observadas (Tabela 1) obtivemos as freqüências residuais que denominados y_{r1} .

No mesmo gráfico da Fig. 1 são também representados os pontos correspondentes a $\Delta \log_{10} y_{r1}(x_i)$, conforme legenda.

Observe-se que este procedimento de agir por etapas permite corrigir as posições dos pontos no gráfico, eliminando a influência da intersecção, pelo menos, com uma das componentes.

Em virtude de tal fato calculamos as estimativas dos parâmetros apenas para a segunda componente usando a região da componente correspondente à reta que passa pelos pontos $x = 7$ mm e $x = 9$ mm os quais já estão livres, da influência da terceira componente. É claro que se aí existir uma influência da primeira componente ela será bem menor que na região correspondente aos pontos que estão à esquerda de $x = 7$ mm.

A partir da Fig. 1, considerando a reta corrigida para a segunda componente, obtemos:

$$\delta_2^* = 6,60 \text{ mm}; \quad \theta_2^* = 79^\circ 30'$$

Os quais permitem obter:

$$\mu_2^* = 7,60 \text{ mm}; \quad \delta_2^* = 1,79 \text{ mm},$$

e finalmente:

$$N_2^* = 65$$

A partir destes dados foram calculadas as freqüências esperadas para esta segunda componente as quais foram subtraídas de y_{r1} obtendo-se novas freqüências residuais, que denominamos y_{r2} . Como tais freqüências não foram desprezíveis repetiu-se o método para as mesmas e representou-se no gráfico da Fig. 1, conforme legenda lá indicada, as diferenças logarítmicas $\Delta \log_{10} y_{r2}$ obtendo-se a reta corrigida correspondente à primeira componen-

te, a qual, por pressuposição inicial, foi tomada como truncada, à esquerda, no ponto $\xi_{11} = 2 \text{ mm}$.

Para esta reta obtivemos:

$$\delta_1^* = 2,40 \text{ mm}; \quad \theta_1^* = 86^\circ$$

o que nos permite obter:

$$\mu_1^* = 3,40 \text{ mm}; \quad \delta_1^* = 0,94 \text{ mm}$$

$$N_1^* = 345$$

Finalmente foram estimadas as proporções da mistura:

$$p_1^* = 65,34\%; \quad p_2^* = 12,11\%$$

$$\text{e } p_3^* = 22,54\%$$

A Tabela 2 é um quadro comparativo onde apresentamos as freqüências esperadas para as três componentes determinadas e as freqüências observadas.

TABELA 2

Distribuição de medidas de induração: freqüências esperadas para três grupos componentes e freqüências observadas

Induração (mm)	Freqüências esperadas				Freqüências observadas
	1.ª comp.	2.ª comp.	3.ª comp.	Total	
2 — 4	251	1	2	254	253
4 — 6	93	11	4	108	108
6 — 8	1	26	10	37	37
8 — 10		21	20	41	39
10 — 12		5	27	32	30
12 — 14			26	26	24
14 — 16			18	18	16
16 — 18			9	9	8
18 — 20			3	3	2
Total	345	64	119	528	517

SANCHES, O. — Distribuição das medidas de induração e eritema na prova tuberculínica: aplicação de um método gráfico de decomposição de uma distribuição de freqüências em componentes normais. *Rev. Saúde públ.*, S. Paulo, 9:15-24, 1975.

A Fig. 2 mostra as três distribuições Normais esperadas em confronto com o histograma observado.

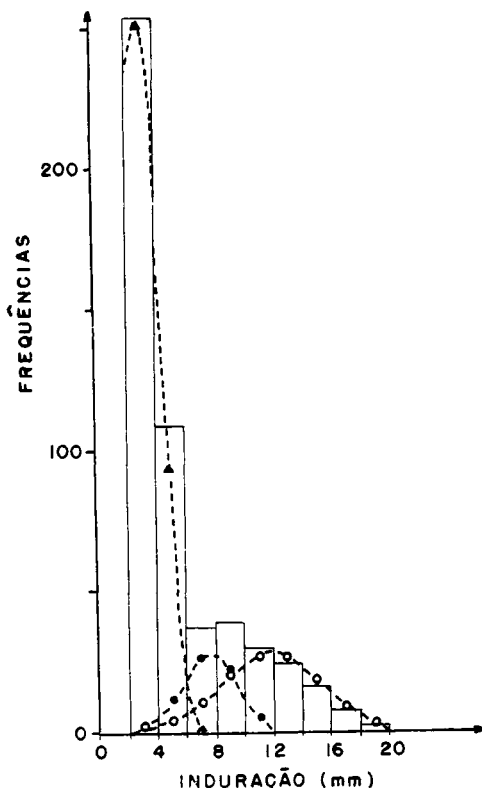


Fig. 2 — Histograma observado e decomposição esperada, em componentes normais, para a distribuição de medidas de diâmetros de induração.

3.2. Eritema

Os dados da Tabela 3 referem-se a medidas de diâmetros de eritema na prova tuberculínica, segundo a técnica de Mantoux, e são também uma adaptação de dados de Ruffino Netto⁴ (1970). A distribuição desta variável é estudada impondo-se um truncamento, à esquerda, no ponto correspondente à medida 1 mm.

A fig. 3 mostra o gráfico de $\Delta \log_{10} y(x_1)$ o qual evidencia a existência de, pelo menos, 4 componentes. A mesma téc-

TABELA 3

Medidas de eritema na prova tuberculínica* e diferenças dos logaritmos das freqüências

Eritema (mm)	Freqüências (y)	$\Delta \log_{10} y(x_1)$
1 — 3	99	0,3585
3 — 5	226	-0,1898
5 — 7	146	-0,4400
7 — 9	53	0,0239
9 — 11	56	-0,2711
11 — 13	30	-0,1984
13 — 15	19	0,2895
15 — 17	37	-0,3921
17 — 19	15	0,0000
19 — 21	15	-0,3310
21 — 23	7	-0,3680
23 — 25	3	
706		

* Dados adaptados de Ruffino Netto⁴ (1970)

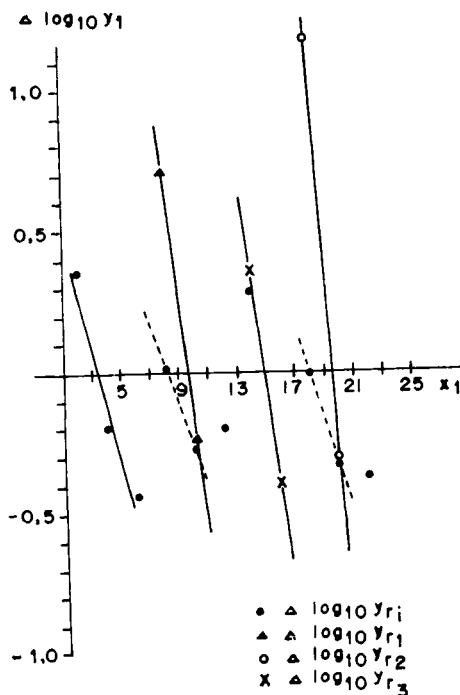


Fig. 3 — Medidas de diâmetros de eritema, na prova tuberculínica. Dados da Tabela 3.

SANCHES, O. — Distribuição das medidas de induração e eritema na prova tuberculínica: aplicação de um método gráfico de decomposição de uma distribuição de freqüências em componentes normais. *Rev. Saúde públ.*, S. Paulo, 9:15-24, 1975.

nica utilizada para a variável induração é aqui repetida no que diz respeito à obtenção das estimativas dos parâmetros das componentes.

rativo entre as freqüências esperadas, para as quatro componentes determinadas, e as freqüências observadas, enquanto que a Tabela 5 mostra as estimativas obtidas para os parâmetros das componentes da mistura.

A Tabela 4 mostra um quadro compa-

TABELA 4

Distribuição de diâmetros de eritema: freqüências esperadas para quatro grupos componentes e freqüências observadas

Eritema (mm)	Freqüências esperadas					Freqüências observadas
	1. ^a comp.	2. ^a comp.	3. ^a comp.	4. ^a comp.	Total	
1 — 3	113				113	99
3 — 5	207				207	226
5 — 7	151				151	146
7 — 9	43	10			53	53
9 — 11	5	51			56	56
11 — 13		30	2		32	30
13 — 15		2	18		20	19
15 — 17			35	1	36	37
17 — 19			16	7	23	15
19 — 21			2	12	14	15
21 — 23				5	5	7
23 — 25				1	1	3
Total	519	93	73	26	711	706

TABELA 5

Estimativas dos parâmetros das quatro componentes referidas na Tabela 4

Componentes	Parâmetros estimados		
	μ (mm)	σ (mm)	p (%)
Primeira (truncada em $\xi_{11} = 1$)	4,52	1,79	73,27
Segunda	10,50	1,22	13,15
Terceira	15,90	1,55	10,32
Quarta	19,60	1,40	3,25

Finalmente, a Fig. 4 estabelece um confronto entre o histograma observado e as quatro componentes esperadas.

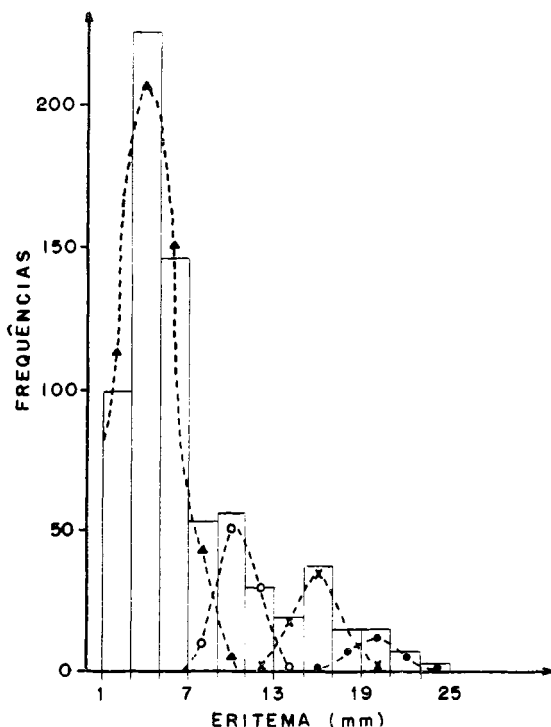


Fig. 4 — Histograma observado e decomposição esperada, em componentes normais; distribuição de medidas de diâmetros de eritema.

4. COMENTARIOS E CONCLUSÕES

Embora as medidas de diâmetros de induração e eritema, por nós utilizadas, proviessem de um mesmo grupo de pessoas, a maneira como as utilizamos levou a amostras de tamanhos distintos o que, para os objetivos que tínhamos fixado, não apresenta inconveniente algum.

O método proposto, embora aproximado, mostrou-se útil e permitiu, para os dados que tínhamos em mãos, obter a separação das distribuições de freqüências observadas em componentes Normais, apesar da relativa intersecção entre os grupos componentes.

Dos resultados obtidos, parece existir evidência de que as variáveis induração e eritema, sob o ponto de vista em que foram estudadas, não se comportam de maneira análoga.

Finalmente, obtida a separação da distribuição de freqüências observada em distintos grupos componentes, descritos por uma bem especificada distribuição de probabilidades, é possível, agora, estabelecer regiões de classificação dos elementos aos distintos grupos, assim como estimar as probabilidades de classificação errada. Para os dados aqui estudados isto será apresentado em outro artigo, no que se refere à classificação segundo as medidas de induração.

RSPU-B/243

SANCHES, O. — [The erythema and induration measurements in the tuberculin test: decomposition of the frequency distribution into normal components by a graphic method.] *Rev. Saúde públ.*, S. Paulo, 9:15-24, 1975.

SUMMARY: The tuberculin test by the Mantoux technique is used in order to classify people according to the size of the reaction caused by a single standard shot of tuberculin. Empirical data seem to demonstrate that such reactions obtained in general populations really represent a mixture of at least two separate groups. Aiming at the separation of these groups, in order to solve the classification problem later on, the Bhattacharya method was used. A decomposition of the frequency distribution of the erythema and induration measurements into normal components was thus obtained. The results of this decomposition showed that the number of components belonging to the distribution of the erythema variable was not the same as that for the induration one.

UNITERMS: Tuberculin test by Mantoux technique. Induration. Erythema. Mixture of normal distributions. Truncated normal distribution.

SANCHES, O. — Distribuição das medidas de induração e eritema na prova tuberculínica: aplicação de um método gráfico de decomposição de uma distribuição de frequências em componentes normais. *Rev. Saúde públ.*, S. Paulo, 9:15-24, 1975.

REFERÊNCIAS BIBLIOGRÁFICAS

1. BHATTACHARYA, C.G. — A simple method of resolution of a distribution into Gaussian components. *Biometrics*, 23:115-29, 1967.
2. EDWARDS, L.B. et al. — Sources of tuberculin sensitivity in human populations: A summing up of recent epidemiologic research. *Acta tuberc. scand.*, 47(Suppl.):77-97, 1959.
3. GULD, J. — Quantitative aspects of the intradermal tuberculin tests in humans. *Acta tuberc. scand.*, 28:222-34, 1953.
4. RUFFINO NETTO, A. — *Epidemiologia da Tuberculose: estudo de alguns aspectos mensuráveis na prova tuberculínica*. Ribeirão Preto, 1970. [Tese de Doutorado — Faculdade de Medicina de Ribeirão Preto da USP]
5. SANCHES, O. — *Extensão do método de Bhattacharya para decomposição de uma distribuição de frequências em componentes normais: aplicação a um problema biológico*. Ribeirão Preto, 1973. [Tese de Doutorado — Faculdade de Filosofia Ciências e Letras]

Recebido para publicação em 26-11-74
Aprovado para publicação em 13-01-75