BRAGANTIA
Boletim Científico do Instituto Agronômico do Estado de São Paulo

Vol. 24 Campinas, maio de 1965 N.º 25

CONTRIBUIÇÃO PARA O ESTUDO DE TABELAS DE CLASSIFICAÇÃO (¹)

ANDRÉ TOSELLO, *engenheiro-agrônomo, Centro Tropical de Pesquisas e Tecnologia de Alimentos, anexo ao Instituto Agronômico*

RESUMO

O presente trabalho estabelece um método geral para se construir tabela de classificação de grãos ou outros produtos por tipo, isto é, baseado no número de defeitos contidos numa determinada amostra de produto.

A teoria é formulada impondo-se a condição de que as amostras devem representar com igual precisão os tipos.

É apresentada uma fórmula matemática que permite construir uma escala de classificação por tipos, partindo-se de um único ponto da escala, isto é, de um tipo cujo número de defeitos é escolhido *a priori*. A fórmula permite construir escalas com a precisão que se queira.

Na aplicação prática, as escalas construídas matemàticamente de acôrdo com êste trabalho, devem satisfazer à condição de que os tipos estabelecidos devem ser fàcilmente diferenciáveis.

O trabalho dá um exemplo de aplicação no estudo das tabelas de classificação, por tipo, do café beneficiado.

1 — INTRODUÇÃO

Na classificação comercial de grãos, costuma-se correlacionar o tipo com a aparência, ou melhor, com a quantidade de defeitos (grãos defeituosos) contidos em determinada amostra. Uma escala de classificação por tipos nada mais é do que uma sucessão A_1 A_2 A_3 .... A_n, de tal modo que a cada têrmo A corresponda certo número de defeitos x contidos numa amostra de m grãos.

Para que essa escala de classificação seja boa, devem ser satisfeitas, pelo menos, duas condições: *a*) os tipos devem ser diferenciáveis a olho nu ou por um processo rápido e expedito; *b*) as amostras devem representar os tipos com igual precisão.

A primeira condição, que vamos chamar de ordem prática, é verificada por meio da experimentação direta e após a satisfação da segunda.

(¹) Recebido para publicação em 1.º de junho de 1965.

O presente trabalho visa resolver a segunda condição, que chamaremos condição matemática.

2 — TEORIA

Se A significa que, numa mesma amostra de *m* grãos, devem estar contidos *x* defeitos, é evidente que, quando se tomam *m* grãos, a freqüência mais provável dos defeitos é a correspondente à probabilidade $p = \frac{x}{m}$. Significa dizer que $p_1 = \frac{x_1}{m}$ $p_2 = \frac{x_2}{m}$ $p_3 = \frac{x_3}{m}$ .... $p_n = \frac{x_n}{m}$ (1) são as probabilidades parciais para eventos dos grãos defeituosos, na amostra *m*, correspondentes aos tipos A_1 A_2 A_3 .... A_n, respectivamente.

Admitindo-se que l_1 l_2 l_3 .... l_n sejam os desvios permitidos, para os eventos considerados, e para que na escala de classificação não haja superposição, deve-se ter:

$$\begin{aligned} l_1 + l_2 &= x_2 - x_1 \\ l_2 + l_3 &= x_3 - x_2 \\ &\vdots \\ l_{n-1} + l_n &= x_n - x_{n-1} \end{aligned} \qquad (2)$$

Sabe-se, para *n* grande, que se pode tomar, com boa aproximação, para valor da probabilidade do evento compreendido no desvio l:

$$p_{-l}^{+l} = \frac{2}{\sqrt{\pi}} \int_0^{\lambda} e^{-\varepsilon^2} d\varepsilon = \theta(\lambda) \qquad (*)$$

em que:

$$\lambda = \frac{l}{\sqrt{2mp\,(1-p)}}$$

ou

$$\frac{l^2}{p\,(1-p)} = 2\,m\,\lambda^2$$

(*) CASTELNUOVO. Calcolo delle probabilitá. 1.º vol. p. 106. Edição 1933.

lembrando-se que a condição matemática se verifica para

$$p_{-l_1}^{+l_1} = p_{-l_2}^{+l_2} = p_{-l_3}^{+l_3} = \ldots\ldots = p_{-l_n}^{+l_n} = \theta(\lambda)$$

e, portanto: ..

$$\frac{l_1^2}{p_1(1-p_1)} = \frac{l_2^2}{p_2(1-p_2)} = \frac{l_3^2}{p_3(1-p_3)} = \\ = \ldots\ldots = \frac{l_n^2}{p_n(1-p_n)} = 2m\lambda^2 \qquad (3)$$

ou ainda:

$$\frac{l_1^2}{x_1(m-x_1)} = \frac{l_2^2}{x_2(m-x_2)} = \frac{l_3^2}{x_3(m-x_3)} = \\ = \ldots\ldots = \frac{l_n^2}{x_n(m-x_n)} = \frac{2\lambda^2}{m} \qquad (4)$$

Para x pequeno em relação a m, tem-se

$$\frac{l_1^2}{x_1} = \frac{l_2^2}{x_2} = \frac{l_3^2}{x_3} = \ldots\ldots = \frac{l_n^2}{x_n} = 2\lambda^2 \qquad (4a)$$

ou também

$$l_1 = \lambda\sqrt{2x_1} \qquad l_2 = \lambda\sqrt{2x_2} \qquad l_3 = \lambda\sqrt{2x_3} \\ \ldots\ldots \; l_n = \lambda\sqrt{2x_n} \qquad (4b)$$

Combinando-se as expressões (2) e (4), tem-se

$$\left(1 + \frac{2\lambda^2}{m}\right)x_n^2 - 2(x_{n-1} + l_{n-1} + \lambda^2)x_n + \\ + (x_{n-1} + l_{n-1})^2 = 0$$

Considerando-se que

$$1 + \frac{2\lambda^2}{m} \simeq 1$$

e resolvendo-se, tem-se

$$x_n = (x_{n-1} + l_{n-1} + \lambda^2) + \lambda\sqrt{\lambda^2 + 2x_{n-1} + 2l_{n-1}} \qquad (5)$$

expressão que permite calcular o ponto x_n, quando se conhece o anterior x_{n-1}.

A precisão dos cálculos pode ser verificada observando-se que

$$x_n - x_1 = (l_1 + l_n) + 2(l_2 + l_3 + \ldots + l_{n-1}) \qquad (6)$$

A simples verificação da expressão (5) nos dá uma falsa impressão de que o valor de m não entra no cálculo do valor de x, ou, em outras palavras, que o tamanho da amostra não influi na construção da tabela. Na verdade, essa influência existe, e pode ser perfeitamente determinada. É suficiente verificar que a expressão (5) é uma fórmula de recorrência; portanto, todos os x vão depender de x_1, que, por sua vez, é determinado em função me m :

$$\frac{l_1^2}{x_1} = 2\lambda^2 \left(1 - \frac{x_1}{m}\right) \qquad (7)$$

Se, em vez de m, tomarmos $a.m$, e, portanto, em vez de x_1, ax, tem-se:

$$l_{a1} = l_1 \sqrt{a} \qquad (7a)$$

3 — APLICAÇÃO

Vamos tomar, como exemplo de aplicação, o estudo de uma escala para classificação, por tipos, de café beneficiado. A escala atual, utilizada pelo Instituto Brasileiro do Café, é uma sucessão assim constituída:

A_1 = Tipo 2 = (4 defeitos contidos numa amostra de 300 g)
A_2 = Tipo 3 = (12 defeitos contidos numa amostra de 300 g)
A_3 = Tipo 4 = (26 defeitos contidos numa amostra de 300 g)
A_4 = Tipo 5 = (46 defeitos contidos numa amostra de 300 g) (8)
A_5 = Tipo 6 = (86 defeitos contidos numa amostra de 300 g)
A_6 = Tipo 7 = (160 defeitos contidos numa amostra de 300 g)
A_7 = Tipo 8 = (360 defeitos contidos numa amostra de 300 g)

Admitem-se as hipóteses de que cada defeito é representado por um grão defeituoso e que 300 gramas contêm 2.300 grãos.

Pode-se demonstrar, fàcilmente, que essa escala não satisfaz à condição matemática, aplicando-se a expressão (4b) aos valores de x para os casos em que são pequenos em relação a m.

O problema consiste em construir algumas escalas matemáticas e, em seguida, verificar a sua aplicação prática.

Se quisermos que a amostra de 300 gramas, ou de 2.300 grãos, represente, para todos os tipos, o tipo com uma precisão de 95%, deve-se ter:

$$\theta(\lambda) = 0{,}95 \qquad \lambda = 1{,}38 \quad \lambda^2 = 1{,}92 \qquad (1)$$

Tomando-se : $x_1 = 4$ $$\frac{l_1^2}{x_1(m - x_1)} = \frac{2\lambda^2}{m} \quad (*)$$

tem-se: $l_1 = 3{,}63$

Aplicando-se a fórmula de recorrência (5), calculam-se todos os valores de *x* e de *l* restantes, e tem-se:

$$\begin{array}{lll} x_1 = 4 & l_1 = 3{,}63 & \\ x_2 = 13{,}69 & l_2 = 6{,}06 & \\ x_3 = 31{,}67 & l_3 = 11{,}92 & \\ x_4 = 58{,}5 & l_4 = 15{,}91 & (9) \\ x_5 = 93{,}23 & l_5 = 19{,}82 & \\ x_6 = 135{,}77 & l_6 = 22{,}72 & \\ x_7 = 185{,}01 & l_7 = 26{,}52 & \end{array}$$

$$\begin{array}{lll} x_1 = 2 & l_1 = 2{,}76 & \\ x_2 = 11{,}33 & l_2 = 6{,}57 & \\ x_3 = 28{,}28 & l_3 = 10{,}38 & \\ x_4 = 52{,}72 & l_4 = 14{,}06 & (10) \\ x_5 = 84{,}70 & l_5 = 17{,}92 & \\ x_6 = 124{,}04 & l_6 = 21{,}42 & \\ x_7 = 170{,}98 & l_7 = 25{,}52 & \end{array}$$

Procedendo-se do mesmo modo, porém admitindo-se um grau de precisão maior, isto é, $\theta(\lambda) = 0{,}99$, o que corresponde a $\lambda = 1{,}83$, e partindo-se de $x_1 = 2$, tem-se:

(*) CASTELNUOVO. Idem, tabela II.

$x_1 = 2$	$l_1 = 3{,}66$		
$x_2 = 16{,}15$	$l_2 = 10{,}49$		
$x_3 = 43{,}79$	$l_3 = 17{,}15$		
$x_4 = 85{,}79$	$l_4 = 24{,}45$	(11)	
$x_5 = 140{,}99$	$l_5 = 30{,}75$		
$x_6 = 209{,}09$	$l_6 = 38{,}35$		
$x_7 = 291{,}79$	$l_7 = 54{,}35$		

Pode-se, assim procedendo, determinar certo número de tabelas, tôdas satisfazendo à condição matemática. Parece-nos, porém, que sendo a atual tabela do I.B.C. bastante prática, qualquer nova tabela não deverá distanciar-se muito dela. Verifica-se, também, que as tabelas matemáticas tendem a aumentar os valores de x até x_4 ou x_5 e diminuir os valores de x_6 x_7 à medida que o grau de precisão exigido aumenta. Significa dizer que, ao se adotar uma tabela matemática com grau de precisão superior a 95%, os tipos médios serão beneficiados e os tipos baixos serão prejudicados. Veja-se o exemplo de n.º 9.

Outro problema a ser considerado é o do tamanho da amostra, isto é, o valor de m. Verifica-se, pela expressão (7a), que a redução do valor de m modifica sensìvelmente a construção da tabela. Pode-se tentar a redução de m e construir novas tabelas, a fim de serem experimentadas na prática. Fizemos alguns exemplos, que não nos pareceram interessantes de serem aplicados.

Analisando-se os exemplos (9), (10) e (11), que nos parecem mais interessantes de serem aplicados na prática, e aproximando-se os valores de x para números inteiros, vamos ter:

(9)

Precisão de 95%

A_1	$x_1 = 4$	(variável entre 1 e 7)
A_2	$x_2 = 13$	(variável entre 8 e 19)
A_3	$x_3 = 31$	(variável entre 20 e 43)
A_4	$x_4 = 58$	(variável entre 47 e 74)
A_5	$x_5 = 93$	(variável entre 75 e 113)
A_6	$x_6 = 135$	(variável entre 114 e 158)
A_7	$x_7 = 185$	(variável entre 159 e 211)

(10)

Precisão de 95%

A_1	$x_1 =$	2 (variável entre 0 e 4)
A_2	$x_2 =$	11 (variável entre 5 e 17)
A_3	$x_3 =$	28 (variável entre 18 e 38)
A_4	$x_4 =$	52 (variável entre 39 e 66)
A_5	$x_5 =$	84 (variável entre 67 e 102)
A_6	$x_6 =$	124 (variável entre 103 e 145)
A_7	$x_7 =$	171 (variácel entre 146 e 196)

(11)

Precisão de 99%

A_1	$x_1 =$	2 (variável entre 0 e 5)
A_2	$x_2 =$	16 (variável entre 6 e 26)
A_3	$x_3 =$	43 (variável entre 27 e 60)
A_4	$x_4 =$	85 (variável entre 61 e 110)
A_5	$x_5 =$	141 (variável entre 111 e 171)
A_6	$x_6 =$	209 (variável entre 172 e 247)
A_7	$x_7 =$	291 (variável entre 248 e 346)

A escala (9), em relação à atual, beneficiaria os tipos A_1 (tipo 2), A_2 (tipo 3), A_3 (tipo 4), A_4 (tipo 5) e A_5 (tipo 6) e prejudicaria os tipos 7 e 8.

A escala (10), em relação à atual, beneficiaria apenas os tipos 4 e 5 e prejudicaria os outros tipos.

A escala (11) beneficiaria todos, à exceção dos tipos 2 e 8.

O problema se resume em verificar quais as conseqüências práticas da adoção de uma dessas tabelas.

CONTRIBUTION TO A STUDY OF CLASSIFICATION TABLES

SUMMARY

In the commercial classification of grains it is the use to correlate the type with the aspect, or rather with the number of defects contained in a determined sample (defective grains). A scale of classification by types is nothing else than a succession A_1 A_2 A_3 .... A_n in such a way that to every A term there corresponds a certain number of defects x contained in a sample of m grains.

In order that this scale of classification be adequate, at least two conditions should be fulfilled: *a*) the types are to be subject to differentiation by naked eye or by a procedure so much as rapid as handy; *b*) the samples should represent the types with one and the same precision.

The first mentioned condition which is herein nominated as of practical order is determined by means of direct experimentation and after fulfilling the second one.

This paper has the purpose to solve the second condition which we may nominate mathematic condition.