A DETERMINAÇÃO DOS NÚMEROS DE INDIVÍDUOS MÍNIMOS NECESSÁRIOS NA EXPERIMENTAÇÃO GENÉTICA

F. G. Brieger
Chefe da Seção Técnica de Genética
da Escola Superior de Agricultura "Luiz de Queiroz"
da Universidade de São Paulo

ÍNDICE

(*) Entregue para a publicação em 27-3-1947.

INTRODUÇÃO

Um problema de grande importância em experimentos agrícolas e em estudos experimentais de genética, consiste na determinação do tamanho dos experimentos. Quantas vezes deve ser repetido o experimento e quantos indivíduos serão necessários para combinar maior eficiência com a maior economia.

É evidente que se nós escolhermos números demasiadamente grandes vamos desnecessàriamente aumentar as despesas e o volume do trabalho. Se formos econômicos demais, reduzimos despezas e trabalhos ,corremos o risco de perder o experimento inteiro por não tirar conclusões, em consequência da falta de material.

Experimentei nos anos passados processos que permitem resolver o problema e que foram o assunto de uma conferência que realizei há um ano no Instituto Fitotécnico de Estanzuela, Uruguai (7). Mais recentemente o Dr. William J. Madow discutiu alguns aspectos teóricos do problema em nossa Escola (12). A presente publicação tem por finalidade apresentar tanto a sua base teórica como a aplicação dos processos, escolhendo para a discussão e melhor explicação os seguintes problemas :

A) Qual será o número mínimo de repetições em experimentosá, que permitirá a exclusão da possibilidade que um dos tratamentos, variedades, etc., estudados apareça sempre como uma das melhores ?

B) Qual será o número mínimo de indivíduos necessários, em experimentos genéticos, para se obter, de um tipo esperado com a frequência p no mínimo um determinado número de indivíduos que poderá ser um ou mais ?

C) Qual será o número total mínimo necessário para que se possa distinguir com precisão entre duas fórmulas mendelianas seguindo as frequências p1 ou p2 para uma classe de tipos ?

D) Qual será o número total mínimo de indivíduos necessário para que se possa distinguir entre uma frequência p1 e duas frequências p2 e p3, sendo uma delas maior e a outra menor do que p1 ?

E) Conhecendo a frequência p(esp) quais são os valores extremos de p(obs) que devemos tomar em consideração num total de n observações, ou tendo obtido um valor p(obs) em n observações quais os valores de p(esp) dos quais êste valor de p(obs) pode ser um desvio de acaso ?

Estas cinco perguntas servem muito bem para ilustrar e explicar os princípios fundamentais empregados.

Com referência ao primeiro problema (A), devemos definir em forma matemática, quando considerarmos um tratamento, uma variedade como sendo "entre as melhores". Às vezes queremos excluir apenas a possibilidade que um dos tratamentos fôsse acidentalmente o melhor. Mas em outros casos, se êles me parecem mais frequentes, temos que ser menos exigentes, ficando satisfeitos quando um dos tratamentos não seja acidentalmente, um dos dois ou três melhores.

Podemos também inverter a pergunta. Em vez de determinar o número de repetições que serão necessárias para executar com eficiência o experimento, perguntamos se os resultados obtidos podem ou não ser causados pelo acaso. Chamei a solução dêste segundo problema de "teste de sequência. (9,10)

Nos problemas B, C e D encontramos frequentemente uma dificuldade inicial na determinação dos valores das frequências p. Por exemplo : não podemos sempre em estudos genéticos usar as proporções ideais mendelianas de : 3:1, 9:7, etc., mas temos que tomar em consideração as complicações adicionais. Uma fonte de complicações é a diferença da viabilidade dos diferentes segregados mendelianos. Assim não é raro o caso que os recessivos homozigotos tenham viabilidade inferior aos dominantes. Supomos que num caso concreto de uma segregação monofatorial, a frequência mendeliana dos recessivos é p=0,25, sendo a sua viabilidade, porém, apenas 50%, e queremos obter no mínimo 5 indivíduos adultos dêste tipo. A solução certa será determinar o número mínimo para uma expectativa de 0,25x0,50=0,125, mas podemos também preferir manter o valor de p=0,25 e dobrar o número de indivíduos desejados para deixar assim u'a margem para a eliminação de uma parte dêles.

A situação torna-se ainda mais complicada quando encontramos não processos simples de eliminação, mas uma competição entre gametófitos, sejam entre tubos polínicos (4) ou entre megásporos (3).

Nota : No trabalho citado o têrmo "megasporo" foi substituido pelo têrmo "megaspório" contra minha vontade, pois a meu ver os têrmos arquespório, microspório, megaspório, indicam o tecido que forma os respectivos esporos, e a competição se dá naturalmente entre êstes últimos e não entre os tecidos formativos.

Pelos poucos exemplos citados fica evidente que mesmo para estudos da genética mendeliana não será suficiente preparar tábons dos números mínimos para alguns valores especiais de p apenas, mas que temos de achar fórmulas gerais que permitam o cálculo para qualquer frequência.

Deveremos ainda definir inicialmente o que usaremos como nível de precisão. Expliquei em várias publicações (1937, 1945, 1946) que não podemos determinar, de uma forma absoluta e final, o que é o limite de precisão. Propuz uma fórmula empírica que se mostrou de bastante utilidade, para definir o limite do provável e do improvável, sendo em N observações ou repetições o limite de probabilidade igual (1:5N) e o limite de improbabilidade igual a (1:1ON). Consideramos como improvável qualquer acontecimento esperado apenas com a probabilidade de P.lim igual ou inferior a 1:10n, e como provável qualquer outro acontecimento esperado com a probabilidade P.lim igual ou maior que1:5N. Com respeito aos acontecimentos esperados com uma frequência intermediária, não podemos fazer previsão segura, de modo que chamei êste intervalo entre os limites a **"região de dúvida"**.

Estas fórmulas empíricas porém, não podem sempre ser aplicadas e nos casos a serem resolvidos neste trabalho, o valor de N é justamente a quantidade desconhecida que pretendemos determinar. Assim empregaremos apenas os três limites convencionais de precisão : 5%(ou P.lim=0,05); 1%(P.lim=0,01 e 1%o(P.lim=0,001)

II — SOLUÇÃO DOS PROBLEMAS

1) Os princípios básicos dos métodos.

O princípio básico consiste em determinar a frequência com a qual podemos esperar o resultado desejado (p) e a frequência de todos os outros resultados não desejados (q), de modo que o total de todos os acontecimentos possíveis será $p+q=1$. As frequências de tôdas as combinações de resultados favoráveis e desfavoráveis em **n** repetições ou em **n** indivíduos são definidos pelos têrmos do binômio $(p+q)n$. Temos agora que determinar quais as combinações de resultados favoráveis ou desfavoráveis que não queremos obter para depois achar um valor do expoente **n** do binômio tal que a soma das frequências dos têrmos não desejados **ficará igual ou menor** o limite de precisão;

$$q^{n} + nq^{n-1} \cdot p + \frac{n(n-1)}{2!} \cdot q^{n-2} \cdot p^{2} + \frac{n(n-1)(n-2)}{3!} \, q^{n-3} \, p^{3} \gtrless P.lim. \quad \text{--(1a)}$$

Frequentemente usamos também a seguintes transformação que facilita o cálculo:

$$n!\,p^{n} \cdot \left\{ \frac{1}{n!} \left(\frac{q}{p}\right)^{n} + \frac{1}{(n-1)!} \left(\frac{q}{p}\right)^{n-1} + \frac{1}{2!(n-2)!} \left(\frac{q}{p}\right)^{n-2} + \frac{1}{3!(n-3)!} \left(\frac{q}{p}\right)^{n-3} \cdots\cdots \gtreqless P.lim. \quad \text{--(1b)}\right.$$

Quando precisamos calcular um ou apenas poucos têrmos do binômio, o trabalho é relativamente fácil, mas precisando-se de mais têrmos o trabalho de cálculo torna-se muito penoso e até impraticável. Assim devemos ver se não será possível substituir a fórmula que exige o cálculo dos têrmos do binômio por outra mais simples.

Agora é um fato bem conhecido que a série binominal $(p+q)^n$ aproxima-se a uma distribuição normal ou de Gauss com média $p\,n$ e com êrro standard $\sqrt{p(1-p)\,n}$ quando o valor do expoente **n** tornar-se bastante grande. Esta aproximação será bastante satisfatória quando **n** for maior do que 30. Abaixo do valor **r** igual a 10 devemos recorrer em geral ao próprio binômio.

Substituindo a série dos têrmos binominais pela aproximação á distribuição de **Gauss** (normal) temos que determinar um valor de **n** de tal modo que a área externa na extremidade da curva, cortada pela absissa $pn + \delta \sqrt{p(1-p)\,n}$ ou $pn = \delta \sqrt{p(1-p)\,n}$ será igual ou inferior ao valor do nível de precisão escolhido. O têrmo delta representa os valores na distribuição de **Gauss** correspondentes aos níveis de precisão.

Assim teremos que achar o valor de **n** que satisfaça uma das duas equações:

$$\left.\begin{aligned} pn + \delta \cdot \sqrt{p(1-p)n} &= P.lim. \\ pn - \delta \cdot \sqrt{p(1-p)n} &= P.lim. \end{aligned}\right\} \quad \text{--------------- (2)}$$

A possibilidade de substituição do binômio pela distribuição de Gauss depende não sòmente do valor do expoente **n**, mas é apenas justificada quando p e $q=(1-p)$ não são muito desiguais. Quando **p** é bem menor do que **q** o binômio torna-se tão assimétrico que a sua substituição por uma distribuição simétrica como aquela de **Gauss** não é mais admissível. Nestes casos podemos aplicar uma outra aproximação e substituir o binômio pela série de **Poisson.** Podemos aceitar como limite um valor de **p** aproximadamente igual ou menor do que 0,1 ou maior ou igual a 0,9 A aproximação de **Poisson** é em geral boa quando o expoente **n** for maior do que 30, e tolerável quando êle for entre 10 e 30 O cálculo dos têrmos das séries de **Poisson** é mais fácil do que aquêles do binômio, não precisando a determinação dos valores de têrmos fatoriais muito elevados. Teremos que calcular, de acôrdo com a definição bem conhecida da série de **Poisson,** um valor médio m=n.p que satisfaça a equação :

$$e^{-\bar{m}}\left(1+\bar{m}+\frac{\bar{m}^2}{2!}+\frac{\bar{m}^3}{3!}\cdots\cdots\cdots\right)\leqq \text{lim.} \cdots\cdots (3)$$

$$\text{onde:} \quad \bar{m} = n.p \quad \text{ou:} \quad n = \frac{\bar{m}}{p}$$

Assim o cálculo de **n** é feito em dois passos. Em primeiro lugar determinamos a média m da série de **Poisson,** que satisfaz à equação (3) para depois calcular **n** pela divisão desta média m pela frequência **p.**

O cálculo dos têrmos dos binômios será muito facilitado quando se usar táboas especiais, e **FISHER** e **YATEES** (11) deram por exemplo os valores dos têrmos e dos seus logarítmos desde 2! até 400!

A determinação dos têrmos da série de **Poisson** nem sempre será necessária pois já existem táboas próprias. Assim MOLINA (13) deu as frequências simples e acumuladas das séries com m=0,001 até m=100.

2) Solução do primeiro problema

Supomos que nós queremos comparar a produção de a variedades e que queremos excluir a possibilidade de que uma delas seja acidentalmente sempre uma das melhores. A probabilidade de qualquer variedade ser a melhor é (1:a) e a probabilidade dela ser a segunda, a terceira, etc., é também (1:a), sendo os acontecimentos mùtuamente exclusivos. A probabilidade de uma variedade ser **ou a melhor ou** a seg da, mas não a terceira, será então (2+a) para cada repetição. e a probabilidade dela ocupar o 1.º, 2.º, 3.º ...m.º lugar será então (m:a).

Finalmente a probabilidade de que uma variedade ocupe êste lugar em 1, 2... n repetições é (m:a1, (m:a)2 ou então (m÷a)n, segundo o teorema da multiplicação de probabilidades. Esta então é a frequência do acontecimento que nós não queremos obter, de modo que temos finalmente a equação.

$$\left(\frac{m}{a}\right)^n = Plim. \text{------------------} (4)$$

A mesma equação obteremos partindo do binômio $(p+q)n$ onde $p=m:a$ é a frequência dos acontecimentos desejados, e q igual a $(1—m:a)$ é a frequência dos acontecimentos não desejados.

$$(p+q)^n = \left(\frac{m}{a}+\frac{a-m}{a}\right)^n = \frac{m^n}{a} + \ldots\ldots$$

tomamos em consideração apenas o primeiro têrmo o qual deve ser no máximo igual ao limite de precisão:

Para o cálculo usamos a transformação logarítmica:

$$n = \frac{\log Plim.}{\log (m:a)} = \frac{\log Plim.}{\log m - \log a} \text{-----------} (5)$$

Explicaremos o emprêgo desta fórmula num caso concreto.

Supondo que o número de variedades a seja igual a 20 e que m seja igual a 2, e empregando ainda os três limites de precisão 5% e 1%) ou 3 vezes (limite de precisão 1%o).

$$n = \frac{\log 0{,}05}{\log (2:20)} = \frac{0{,}69897-2}{-1} = 1{,}3 < 2$$

$$n = \frac{\log 0{,}01}{\log (2:20)} = \frac{-2}{-1} = 2$$

$$n = \frac{\log 0{,}001}{\log (2:20)} = \frac{-3}{-1} = 3$$

Resultado: Para evitar que uma das 20 variedades seja acidentalmente a melhor ou a segunda em produtividade, temos que repetir o experimento no mínimo 2 vezes (limite de precisão 5% e 1%) ou 3 vezes (limite de precisão 1%).

Devemos lembrar ainda que o argumento usado acima, para determinar o número mínimo de repetições não é o único que devemos tomar em consideração no planejamento de experimentos. Não podemos, por exemplo, deixar de prestar atenção à possibilidade do campo experimental ser heterogêneo o que poderá induzir-nos a aumentar o número de repetições. Também não devemos esquecer que não somos apenas interessados se uma ou outra das variedades é melhor do que as demais, mas queremos saber quanto mais produz. O t-teste necessário para isso torna-se tanto mais eficiente quanto maior o número de repetições, pois o êrro standard das médias diminui proporcionalmente com a rais quadrada de n e os limites da distribuição de **Standard** decrescem com o aumento dêste número de **repetições.**

3) SOLUÇÃO DO SEGUNDO PROBLEMA

A) Cálculo pela série binominal.

Passemos agora para o segundo problema mencionado na introdução. Qual o número mínimo **n** de indivíduos necessário para ter no mínimo 1, 2... a indivíduos de um determinado fenótipo em experimentos genéticos.

Se **p** é a probabilidade de obter um determinado tipo e (1-p)=q a probabilidade de não obtê-lo, podemos calcular as frequências de obter 1, 2...a indivíduos dêste tipo, expandindo o binômio (p+q)n até o têrmo (a+1).

A soma dêstes têrmos (a+1) deve ser no máximo igual ao limite de precisão. Temos então, segundo a equação. (1 b) :

$$n!\,p^{n}.\left\{\frac{1}{n!}\left(\frac{q}{p}\right)^{n}+\frac{1}{(n-1)!}\left(\frac{q}{p}\right)^{n-1}\cdots\cdots+\frac{1}{m!(n-m)!}\left(\frac{q}{p}\right)^{n-m}\right\}\leqq P\text{lim}\ldots\ldots(6)$$

Darei como exemplo o cálculo dos valores de **n** para p=0,25 e a=0,1, 2 e 3, isto é, a resposta quantos indivíduos serão necessários numa segregação mendeliana monofatorial seguindo a proporção (3A— +1aa), para ter no mínimo 1, 2, 3 ou 4 indivíduos do tipo recessivo (aa). Calculamos em primeiro lugar para os três limites convencionais de precisão, as frequências dos quatro primeiros têrmos dos binômios com expoentes n=10 15, 20 25, 30, 35, 40, 45 e 50, conforme consta dos Quadro I e II. As frequências acumuladas dos têrmos sucessivos dos 9 binômios constam do Quadro III. Desenhamos as curvas que correspondem a cada linha horizontal dêste Quadro III, para determinar os pontos de interseção com as linhas que correspondem aos níveis de precisão. Podemos fazer um único gráfico, porém para a ilustração numa escala fácil de compreender foram executados 3 gráficos separados, um para cada nível de precisão (Fig. 1 a 3). O valor de **n** desejado é o número inteiro imediatamente superior ao ponto de interseção, como indicado nos gráficos por flechas.

Os resultados finais são os seguintes (Quadros V a VII):

Para se ter no mínimo um ou mais indivíduos do fenótipo esperado com a frequência p=0,25 devemos estudar o total de 11 indivíduos (Precisão 5%), 17 indivíduos (Precisão 1%) ou 24 indivíduos (Precisão 1%o).

Para se ter no mínimo dois ou mais indivíduos do fenótipo os números totais de indivíduos são respectivamente : 18, 24 e 29. Para se obter no mínimo três ou mais necessitamos do mesmo modo de 24, 31 e 40. Para se obter no mínimo quatro ou mais os números totais necessários são 29, 37 e 48.

O exemplo serve não sòmente para ilustrar o processo do cálculo, mas. também, para demonstrar que êle é muito laborioso. Uma vez que o menor número de m achado para 5% de precisão e para um ou mais indivíduos seja superior a 10, podemos também aplicar a aproximação do binômio à distribuição de Gauss e explicaremos mais tarde êsse processo.

Quando podemos limitar-nos ao primeiro têrmo do binômio querendo saber apenas o número mínimo total de indivíduos necessários para obter no mínimo um ou mais indivíduos do tipo esperado com a frequência p, o cálculo torna-se fácil, pois, temos então apenas a solucionar a equação que já conhecemos. (Fórmulas 4 e 5).

$$q^n = PLim$$

$$n = \frac{\log RLim}{\log q}$$

B) Cálculos pela aproximação à distribuição de Gauss

Passemos agora para a discussão dos processos baseados na aproximação do binômio à distribuição de Gauss :

Se nós esperamos um acontecimento com a frequência p, teremos em n indivíduos (p.n.) casos esperados. Mas devido às causas acidentais o número geralmente observado nem sempre é igual ao número esperado, pois, existirá uma certa variação em volta do valor esperado ou médio (pn) caracterizado pelo êrro standard $\pm\sqrt{p(1-p)n}$. Esta variação segue a distribuição "normal" ou de Gauss como já explicado quando as frequências p e (1-p) não sejam demasiadamente desiguais de (0,1 até 0,9), e quando n é um número razoàvelmente grande (maior do que 10). Indicando os limites da distribuição de Gauss nos diferentes níveis de precisão, com a letra grega delta podemos dizer que os valores extremos da variação serão:

$$np \pm \delta\sqrt{p(1-p)n}$$

Podemos agora resolver o nosso primeiro problema : qual seja o número mínimo de indivíduos para que um tipo esperado com a frequência p apareça no mínimo num número a de indivíduos. A resposta é dada pela equação :

$$\left.\begin{array}{l} pn - \delta.\sqrt{p(1-p)n} = a - 1 = m \\ b = \delta.\sqrt{\dfrac{1-p}{p}} \qquad c = \dfrac{m}{p} \\ n = \left\{\dfrac{b + \sqrt{b^2 + 4c}}{2}\right\}^2 \end{array}\right\} \text{-------------} (7)$$

Os valores de delta para os limites unilaterais da distribuição de Gauss, são : 1,64; 2,33; 3,09. É evidente que devemos aplicar apenas os limites unilaterais pois tomamos em consideração sòmente desvios no sentido negativo em relação ao valor ideal, pn querendo saber apenas qual o desvio negativo maior, isto é, qual o número mínimo que pode acontecer, sem interessar-nos pelos desvios positivos e os valores máximos.

Para compensar a aproximação devida ao emprêgo dos limites de distribuição de Gauss é indicado usar ainda uma compensação nos casos nos quais p tem valores entre 0,2 e 0,8, sendo desnecessária a compensação para os valores de 0,2 até 0,1, ou 0,8 até 0,9.

A compensação consiste no acréscimo do têrmo seguinte :

$$\left.\begin{array}{l} n' = \dfrac{1}{p} \\ n(cor.) = n + n' \end{array}\right\} \text{---------------} (8)$$

Para ilustrar a aplicação das fórmulas (7 e 8) calculamos os números totais mínimos necessários para obter 2 ou mais, 3 ou mais, 4 ou mais de indivíduos esperados com a frequência p=0,25, e para os três níveis convencionais de precisão. O cálculo consta do Quadro IV.

Quando quizermos ter apenas garantia de obter no mínimo um indivíduo do tipo esperado com a frequência p, o cálculo torna-se mais fácil ainda. A fórmula (7) transforma-se do modo seguinte :

$$\left.\begin{array}{l} c = \frac{m}{p} = \frac{a-1}{p} = 0 \quad \textit{para} \quad a = 1 \\ n = \left\{\frac{b+\sqrt{b^2}}{2}\right\}^2 = b^2 = \delta^2\,\frac{1-p}{p} \\ n' = \frac{1}{p} \\ n(\text{cor.}) = n + n' \end{array}\right\} \text{-----------} (9)$$

Para o exemplo escolhido de p=0,25 obtemos então os resultados **seguintes**:

$$n = 1{,}64^2\,\frac{0{,}75}{0{,}25} = 8{,}07 < 9 \qquad n' = 4 \qquad n(\text{cor.}) = 13$$

$$n = 2{,}33^2\,\frac{0{,}75}{0{,}25} = 16{,}29 < 17 \qquad n' = 4 \qquad n(\text{cor.}) = 21$$

$$n = 3{,}09^2\,\frac{0{,}75}{0{,}25} = 28{,}65 < 29 \qquad n' = 4 \qquad n(\text{cor.}) = 33$$

Os resultados finais do cálculo todo dos valores n e n (cor) constam no Quadro V. a VII.

C) Aproximação da série Poisson.

Expliquei na introdução a êste capítulo que podemos substituir a série binominal pela série de **Poisson,** quando p for menor do que 0,1. Teremos então que determinar os têrmos de séries de **Poisson,** seguindo a sua definição matemática bem conhecida :

$$e^{-\overline{m}}\left(1+\overline{m}+\frac{\overline{m}^2}{2!}+\frac{\overline{m}^3}{3!}\cdots\cdots\right) \text{-----------} (10)$$

e escolher o valor médio $\overline{m}$=n.p para que a frequência do primeiro têrmo, a soma das frequências dos dois primeiros têrmos, dos três têrmos, etc., fique igual ou inferior ao limite de precisão. Para isso devemos calcular, como fizemos para as séries binominais com diferentes expoentes, as frequências dos têrmos para valores escolhidos de $\overline{m}$=2,3,4, etc. e obter os valores desejados por interpolação gráfica. Mas as táboas de MOLINA (10) permitem dispensar êste processo laborioso. Podemos simplesmente constatar nestas táboas muito úteis por exemplo, que para m=3,0 a frequência do primeiro têrmo tem o valor de 0,0498 o que é justamente inferior a 0,05 limite de precisão, que para $\overline{m}$=4,7 a frequência do primeiro têrmo

0,009095 é justamente inferior a 0,01 limite de precisão e para $\overline{m}=7,0$ a sua frequência de 0,000912 é justamente inferior a 0,001 limite de precisão.

Obtemos assim os valores de $\overline{m}$ que constam na Táboa 1, e podemos com a sua ajuda calcular o valor de n pela fórmula :

$$n = \frac{\overline{m}}{p}$$

Usamos de novo um exemplo e queremos saber qual o número total mínimo de indivíduos necessário para se obter no mínimo 3 indivíduos de um tipo esperado com a frequência $p=0,07$. Achamos na Táboa 1 os seguintes valores de $\overline{m}$ para êste caso' 6,3 (Precisão 0,05); 8,5 (Precisão 0,01 e 11,3 (Precisão 0,001). Assim podemos calcular os valores de n :

$$\left.\begin{array}{lll} \textit{Precisão } 5\% & n = \frac{6,30}{0,07} = 90,0 < 91 \\ \textit{Precisão } 1\% & n = \frac{8,50}{0,07} = 121,4 < 122 \\ \textit{Precisão } 0,1\% & n = \frac{11,30}{0,07} = 161,4 < 162 \end{array}\right\}$$

D) Comparação dos três processos

Devemos agora comparar os três métodos de cálculos explicados nos capítulos anteriores. Consideramos sempre como o valor mais acertado aquêle calculado na base da série binominal, sendo os outros apenas aproximações, devendo-se verificar se estas aproximações são satisfatórias. Os valores aproximados não devem ser muito diferentes dos valores exatos, e devem ser sempre maiores do que os valores exatos de modo que a aproximação nunca reduz, mas sim, aumenta a precisão.

Comecemos com os valores do Quadro V que contém os números mínimos para p-0,5. Podemos constatar que os valores calculados com a aproximação à distribuição de Gauss, sem correção são todos menores que aquêles da série binominal, de modo, que a aproximação não pode ser considerada como satisfatória. Os valores corrigidos porém são iguais ou um pouco maiores do que os valores exatos da série binominal, e portanto satisfatório.

Os valores para $p=0,25$ (Quadro VI) mostram que nêste caso a correção dos valores calculados pela aproximação normal não é mais tão necessária. Os valores não compensados são apenas pequenos demais no limite de 0,05 da precisão, de modo, que a correção é realmente necessária apenas para êste limite. No limite de 1% os valores não corrigidos são iguais aos valores exatos e no limite de 1‰ êles são um pouco maiores. Assim nestes casos a correção não é mais necessária.

Finalmente os valores para p=0,1, contidos no Quadro VII, mostram que pràticamente podemos dispensar a correção por completo. Os valores calculados pela aproximação de Gauss são iguais ou apenas muito pouco menores do que os valores exatos da série binominal no limite 5% de precisão, e êles são um pouco maiores do que os valores exatos no limite 1% de precisão, e êles são bastante maiores para o limite 1‰ de precisão.

Podemos assim tirar a seguinte conclusão : **A aproximação à distribuição normal pode ser usada sem perda de precisão e sem compensação desde os valores de p=0,25 até p=0,10, mas, para valores de p=0,50 até p=0,25 deve ser acrescida a correção n'=1÷p.**

Explicamos que podemos usar para valores pequenos de p a aproximação da série de **Poisson**, sendo o limite de p=0,1. Os dados do Quadro VII justificam esta decisão. Os valores de **Poisson** nêste caso de p=0,1 são todos muito próximos e um pouco maiores do que os valores exatos da série binominal. Êles apresentam de fato já uma melhor aproximação do que os valores calculados com a aproximação à distribuição de Gauss.

4) SOLUÇÃO DO TERCEIRO PROBLEMA

A) Cálculo pela série binominal.

O terceiro problema mencionado na introdução : a distinção entre duas espectativas, p1 e p2, também pode ser resolvido empregando o método de calcular as frequências acumuladas dos têrmos binominais das séries (p1+q1)n e (p2+q2)n. Mas agora a solução algébrica é mais complicada ainda. Na solução do segundo problema, tratado nêste trabalho, sabemos quantos têrmos do binômio queríamos acumular, e a única incógnita é o expoente n. Agora porém, temos três valores desconhecidos, além do expoente n precisamos achar os números de têrmos m1 e m2 a serem acumulados em cada série binominal. Para poder determinar êstes três valores desconhecidos, precisamos estabelecer três equações independentes.

Suponhamos que a frequência p1 fôsse maior do que p2. O tipo esperado com estas duas frequências pode aparecer em :

1, 2, 3,... m2 ... n.p2... n indivíduos
1 2, 3... m1 ... n.p1... n indivíduos

Devemos escolher os dois valores m2 maior do que np2 e m1 menor do que np1 de tal modo que : a) as frequências acumuladas dos têrmos m2, m2+1, m2+2... n sejam no máximo iguais aos limites de precisão; b) que as frequências acumuladas dos têrmos 0, 1, 2, 3 ...m1 sejam também iguais ou inferiores ao mesmo limite de precisão, e c) que os valores, m1 e m2 sejam idênticos.

Assim teremos as seguintes equações :

$$\left.\begin{array}{l}
\textit{Para a serie } (p_1+q_1)^n \\
q_1^n + nq_1^{n-1} p_1 + \frac{n(n-1)}{2!} q_1^{n-2} p_1^2 \cdots + \frac{n(n-1)\ldots(n+1-m_1)}{1.2\ldots m_1} q_1^{n-m_1} p_1^{m_1} \lesseqgtr \text{Plim.} \\
\textit{Para a serie } (p_2+q_2)^n \\
q_2^n + nq_2^{n-1} p_2 + \frac{n(n-1)}{2!} q_2^{n-2} p_2^2 \cdots\cdots + \frac{n(n-1)\ldots(n+1-m_2)}{1.2\ldots m_2} q_2^{n-m_2} p_2^{m_2} \lesseqgtr \text{Plim.} \\
\textit{sendo, finalmente} \\
m_1 = m_2
\end{array}\right\} \quad (12)$$

Se nós acharmos uma solução que satisfaça estas três equações podemos esperar que as variações negativas do valor (p1n) e as variações positivas do valor (p2n) não coincidam dentro dos limites de precisão, de modo que podemos distinguir com segurança as expectativas de p1 e p2.

O trabalho do cálculo necessário para a solução destas equações consiste no seguinte : Para cada um dos dois binômios e para cada um dos três limites de precisão temos que calcular no mínimo quatro valores de m para construir curvas de m1 e m2, para achar por interpolação gráfica o valor de m para o ponto no qual estas curvas se cruzam. Assim para distribuições de 3 níveis de precisão precisamos de 2x3x4=24 valores de m.

Podemos dar ainda uma compreensão mais detalhada de trabalho de cálculo necessário se escolhermos valores concretos, por exemplo, p1=0,3 e p2=0,2. No limite 5% precisamos para cada um dos 8 valores de m o cálculo de cêrca de 40 têrmos de binômios com expoente de cêrca de 200 no limite de 1%, precisamos cêrca de 100 têrmos de binômios com expoente de cêrca de 400 e no limite de 1%o cêrca de 200 têrmos de binômios com expoente de mais ou menos 700, isto é, um total de 8(40+100+200)=2,720 têrmos binominais com expressões fatoriais muito elevadas Todo êste trabalho imenso serve apenas para resolver um único problema : a distinção entre as duas frequências de 0,2 e 0,3 O trabalho de cálculo é evidentemente excessivo e pràticamente inexequível.

B) Cálculo pela aproximação à distribuição de Gauss.

Pelo exposto acima, sabemos que a aproximação baseada na distribuição de Gauss é bastante satisfatória para o estudo de frequências entre 0,9 e 0,1, e quando os valores de n são superiores a 10.

Suponhamos que a frequência p1 seja maior do que p2. O desvio máximo negativo em relação a uma expectativa (n.p1) e o desvio máximo positivo em relação à expectativa (n.p2) são :

$$\left.\begin{array}{l}\text{Desvio mínimo em relação a } p_1\text{:} \\ np_1 - \delta.\sqrt{p_1(1-p_1)n} \\ \text{Desvio máximo em relação a } p_2\text{:} \\ np_2 + \delta.\sqrt{p_2(1-p_2)n}\end{array}\right\}$$

$$\left\{np_1 - \delta.\sqrt{p_1(1-p_1)n}\right\} - \left\{np_2 + \delta\sqrt{p_2(1-p_2)n}\right\} = 0$$

$$n(p_1 - p_2) - \delta.\left\{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}\right\}.\sqrt{n} = 0$$

quando delta significa os valores dos limites da distribuição de Gauss. A diferença entre êstes dois valores extremos deverá ser igual a zero ou maior ainda :

$$\left.\begin{array}{l} n = \left\{\delta\,\dfrac{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}{p_1 - p_2}\right\}^2 \\ n' = \dfrac{1}{p_1 - p_2} \\ n(\text{cor.}) = n + n' \end{array}\right\} \cdots\cdots (13)$$

Assim podemos resolver o problema com relativamente pouco cálculo, em contraste com a fórmula mais exata baseada no binômio (12) que exige um cálculo pràticamente inexequível.

Para melhor compreensão incluimos no Quadro VIII o cálculo dos números mínimos totais necessários para distinguir as expectativas $p1=0,2$ e $p2=0,3$. Os resultados são : Precisão 5%:209 indivíduos; Precisão 1%:412 indivíduos; Precisão 0,1%: 716 indivíduos.

Se não quizermos tornar o teste de distinção de duas frequências p1 e p2 mais rigoroso ainda, podemos exigir que a diferença entre o número máximo em relação a np2 e o número mínimo em relação a np1 seja maior do que zero e no mínimo igual a um número $a=m+1$. Assim a fórmula 8 se transformará na forma **seguinte**:

$$n(p_1-p_2) - \delta \cdot \left\{ \sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)} \right\} \cdot \sqrt{n} = m$$

$$\left.\begin{array}{l} b = \delta \cdot \dfrac{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}{p_1 - p_2} \qquad c = \dfrac{m}{p_1 - p_2} \\ n = \left\{ \dfrac{b + \sqrt{b^2 + 4c}}{2} \right\}^2 \\ n' = \dfrac{1}{p_1 - p_2} \\ n(cor) = n + n' \end{array}\right\} \text{-------------- (14)}$$

C) Cálculo pela aproximação de Poisson

Explicámos acima que podemos empregar a série de **Poisson** no lugar da série binominal quando p for menor do que 0,1. Assim podemos também empregar estas séries para distinguir duas frequências p1 e p2' ambas inferiores a 0,1. O raciocínio é o mesmo como antes (pg. 231). Devemos achar duas séries de **Poisson** com as médias m1 e m2 que satisfaçam as seguintes condições : o menor valor de m1 que não mais fôsse esperado num determinado nível de precisão coincide com o maior número em condições idênticas numa série m2. O processo de cálculo seria porém muito laborioso e preparei por isso uma tabela simples (Tâboa 2) com ajuda dos valores de MOLINA (13). Para o seu emprêgo precisamos saber um dêstes valores médios m1 ou m2 e a proporção m1 : m2 que deve ser igual a p1 : p2, para poder aplicar o processo.

Assim determinamos apenas os valores para 1% limite.

O emprêgo da tâboa explicamo-lo com a ajuda de um exemplo : Qual o número mínimo necessário para poder disdinguir entre $p1=0,08$ e $p2=0,04$?

Determinamos o quociente $p1 : p2=2,0$ e achamos o valor de $m2=31,3$ na táboa 2 para êste quociente.

Agora determinamos **n** pela equação :

$$n = \frac{\bar{m}_2}{p_2} \qquad (15)$$

$$n = \frac{31,30}{0,04} = 782,5 < \overline{783}$$

Seriam necessário no mínimo 783 indivíduos para distinguir entre as frequências de 0,08 e 0,04 com 1% de precisão.

D) Comparação dos três métodos.

Seria muito interessante comparar quantitativamente os resultados obtidos com os dois processos aproximados e com o processo exato. Mas, tive que desistir desta comparação em vista do trabalho excessivo que o processo dos têrmos binomiais exige. Porém, lembrando que já demonstramos que as aproximações são satisfatórias quando estudamos uma só frequência p e considerando que o raciocínio é o mesmo na solução dos problemas tratados, podemos concluir que a aproximação seria igualmente satisfatória na solução do segundo como do primeiro **problema.**

5) LIMITES BILATERAIS E A DISTRIBUIÇAO ENTRE TRÊS FREQUÊNCIAS

A) Distinção entre três frequências.

Depois do que já foi explicado nos capítulos anteriores o cálculo dos têrmos dos binômios torna-se impraticável quando se trata de apenas duas frequências a serem distinguidas de modo que não é mais necessário tomar êste processo em consideração no caso de três frequências. Devemos diretamente passar a aplicar a aproximação de **Gauss** sempre que as três frequências forem maiores do que 0,1 e aquela de **Poisson,** quando êles forem menores do que 0,1.

Supomos que temos p2 maior p1 maior p3' de modo que podemos formular as seguintes duas **equações** :

$$\left.\begin{aligned} n_{1.2}\cdot(p_1-p_2) - \delta.\left\{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}\right\}.\sqrt{n_{1.2}} &= P\,\text{lim.} \\ n_{1.3}\cdot(p_3-p_1) - \delta.\left\{\sqrt{p_3(1-p_3)} + \sqrt{p_1(1-p_1)}\right\}.\sqrt{n_{1.3}} &= P\,\text{lim.} \end{aligned}\right\}$$

$$\left.\begin{aligned} n_{1.2} &= \left\{\delta\,\frac{\sqrt{p_1(1-p_1)}+\sqrt{p_2(1-p_2)}}{p_1-p_2}\right\}^2 = P\,\text{lim.} \\ n_{1.3} &= \left\{\delta\,\frac{\sqrt{p_3(1-p_3)}+\sqrt{p_1(1-p_1)}}{p_3-p_1}\right\}^2 = P\,\text{lim.} \end{aligned}\right\} \text{---(16)}$$

O valor delta representa agora os limites bilaterais da distribuição de **Gauss**, pois tomamos em consideração simultaneamente variações positivas e negativas do valor n.p1).

Nós queremos porém achar um só valor n, em vez dos dois valores n1.2 e n1.3. Porém, não existe solução algébrica que possa satisfazer ao mesmo tempo duas equações independentes com uma só incógnita, de modo que teremos que solucionar ambas as equações separadamente e usar apenas o valor maior de n calculado.

Por exemplo: quantos indivíduos serão necessários para distinguir ao mesmo tempo as frequências 0,20, 0,25 e 0,30 com 1% precisão?

$$n_{1.2} = \left\{2{,}58\,\frac{\sqrt{0{,}25 \,.\, 0{,}75} + \sqrt{0{,}20 \,.\, 0{,}80}}{0{,}25 - 0{,}20}\right\}^2 = 1840$$

$$n_{1.3} = \left\{2{,}58\,\frac{\sqrt{0{,}30 \,.\, 0{,}70} + \sqrt{0{,}25 \,.\, 0{,}75}}{0{,}30 - 0{,}25}\right\}^2 = 2107$$

Será necessário um total de 2.107 indivíduos para, no máximo 1 vez em 100 apenas correr o risco de não poder decidir entre as duas frequências teóricas de $p=0{,}20$, $p=0{,}25$ e $p=0{,}30$.

Em vez de executar o cálculo podemos usar uma tábua que preparei há cêrca de 10 anos (BRIEGER 1, Tábua 12) e que nos dá imediatamente os valores mínimos de n para os limites de precisão 5% e 1%.

Quando os valores das frequências a serem empregados são menores do que 0,1 então temos que recorrer à distribuição de **Poisson**. Em analogia ao caso anterior, teríamos que fazer duas determinações e achar o número total mínimo tanto para distinção entre p1 e p2 como entre p1 e p3. A tábua I dêste trabalho, porém não pode ser usada para êste fim pois ela toma em consideração apenas limites unilaterais. Não calculei uma outra táboa para os limites bilaterais pois apenas muito raramente temos necessidade de aplicá-la.

B) Emprêgo dos limites bilaterais para fins informativos

Frequentemente encontramos as seguintes duas perguntas na experimentação que podemos agora responder com facilidade.

a) Esperando um certo tipo com a frequência p e usando um total de n indivíduos, quais os valores de variações extremas de p que podem ser encontrados ?

b) Tendo constatado em uma ou mais populações ou famílias o aparecimento de um determinado tipo com a frequência p(obs), qual poderá ser o valor ideal de p(esp), do qual o valor de p(obs) representará um desvio de acaso ?

A resposta naturalmente será diferente quando p(esp) for maior do que 0,1 ou quando êle for menor do que 0.1.

Para uma série de valores p(esp) os limites da variação de p(obs) em função do número de indivíduos estudados, e para os três níveis convencionais de precisão foram calculados (tábca 3) os valores de n para as diferenças de p(esp) e p(obs) igual a 0,05, 0,10, 0,15, etc., de acôrdo com as fórmulas :

$$\left.\begin{aligned} n &= \delta^2 \frac{\bar{p}(1-\bar{p})}{(\bar{p}-p)^2} \\ \bar{p} &= p\,(\textit{esperado}) \\ p &= p\,(\textit{observado}) \end{aligned}\right\} \text{-------------- (17)}$$

Supomos por exemplo que esperamos p=0,5 e que temos cêrca de 45 plantas em cada família estudada. Verificamos então na coluna encabeçada pelo valor 0,5 que passamos um valor perto a n=45 ns seguintes linhas horizontais :

0,25 e 0,75 (Precisão 1%).
0,30 e 0,70 (Precisão 1%)
0,35 e 0,65 Precisão 0,5%)

Isto quer dizer que em 1000 famílias de 45 indivíduos, uma família dará frequências mais extremas do que 0,25 ou 0,75, uma família em 100 dará valores de p(obs) mais extremos do que 0,3 e 0.7 e finalmente uma família em 20 dará valores de p(obs) maior de 0,35 e 0,65, sendo o centro de variação sempre o valor ideal p=0,5.

A resposta para a segunda pergunta formulada acima, pode ser obtida na táboa 3 do modo seguinte: obtivemos por exemplo numa família de 100 indivíduos um determinado tipo com a frequência p(obs)=0,30. Estudando os valores de n que constam da linha horizontal indicado por 0,30 da esquerda para a direita constatamos que o valor de n sobe passando o valor 100 no nível de precisão 1% um pouco antes da coluna que corresponde ao valor p(esp)=0,20, descendo em baixo de 100 de novo entre a sexta e sétima coluna, sendo os valores exatos p(esp)=0,40, n=160 e p(esp)=0,45, n=74. Assim podemos concluir que o valor observado de 0,3 poderá ser um desvio de qualquer frequência ireal entre cêrca de 0,20 e 0,43, com 1% de precisão.

Para os valores de p(esp) menores do que 0,1 preparei outra táboa com os limites bilaterais das distribuições de **Poisson** a qual deve ser usada da forma seguinte:

Supomos que temos uma frequência p(esp)=0,05 e famílias de 300 indivíduos, então o valor médio da série de **Poisson** será $\overline{m}=np=300 \times 0{,}05 = 15$. Procuramos então na **táboa** 4 a linha horizontal que corresponde a êste valor $\overline{m}=15$. Verificamos então que êste valor de m pode variar até 5, respectivamente 29 no nível de 0,01 precisão, até 7 e 26 no 1% nível, até 9 e 26 no 5% nível de precisão. Aplicando a fórmula p=m : n e substituindo n por 300 no nosso exemplo, podemos fàcilmente determinar os valores correspondentes de p.

95% das famílias variam entre 9:300=0,030 até 23:300 = 0,077
99 % das famílias variam entre 7:300=0,023 até 26:300 = 0,087
99,9% das famílias variam entre 5:300=0,017 até 29:300 = 0,097
sendo o valor ideal central p=0.050.

Também podemos resolver o seguinte problema. Suponhamos que foram achados 5 indivíduos de um determinado tipo num total de 100 indivíduos. Usando apenas o limite 1% de precisão, verificamos na **táboa** 4 que 5 indivíduos podem ser encontrados em todos os casos desde $\overline{m}=0$ até $\overline{m}=13$. Calculando p esp=m : n achamos assim os valores de 13 : 100=0,13 Assim o nosso valor de p(obs) igual a 0,05 pode ser um desvio de acaso de qualquer valor de p(esp) desde 0,0... até 0,13.

6) TESTES PARCIAIS PROGRESSIVOS

Pelo exposto acima, torna-se claro que precisamos às vezes números bem elevados de indivíduos para satisfazer as exigências estabelecidas. Mas nem sempre temos material bastante e em outros casos a execução de experimentos muitos extensos torna-se demais dispendiosa. Devemos então nos lembrar que o chamado "número total mínimo" represente tanto um mínimo como um máximo. Êle representa o mínimo necessário para obter o resultado desejado com uma boa margem de garantia, de acôrdo com o limite de precisão estabelecido, de modo que, não terá vantagem aumentar os números. Mas se nós aceitamos uma menor margem de precisão e queremos confiar mais na nossa sorte, podemos reduzir o número de indivíduos.

Suponhamos que queremos achar um indivíduo no mínimo de um tipo esperado com a frequência $p=0,25$. Pelas fórmulas dadas acima sabemos que estudando 16,3 indivíduos vamos ter no mínimo um indivíduo dêste tipo, falhando o nosso experimento apenas uma vez em 100 casos ou mais raramente ainda. De outro lado sabemos que a nossa definição de frequência esperada é igual a 0,25 ou 1 : 4, quer dizer que se não houvesse variação de acaso, um indivíduo em cada quatro seria do tipo esperado. Se dividirmos o nosso total de 16,3 em um conjunto de quatro amostras com quatro indivíduos cada um, sabemos que no mínimo em um dêles deverá aparecer um indivíduo do tipo desejado, mas não sabemos em qual deles. Podemos calcular a probabilidade de não achar êste indivíduo na primeira, segunda, etc. amostra do conjunto, considerando na fórmula (4) o valor de n como conhecido.

Os resultados podemos interpretar do modo seguinte: Uma vez em três casos não encontramos um indivíduo do tipo desejado na primeira amostra, uma vez em dez êle não aparece na primeira e segunda amostra, uma vez em trinta e dois êle não aparece em três amostras ou num total de 12 indivíduos. Ninguém aceitaria provàvelmente uma probabilidade de 1 : 3 apenas como satisfatória exceto quando o custo do experimento por indivíduos fôsse muito elevado. Mas a esperança de 1 em 10 já é as vezes aceitável. Às vezes será vantajoso começar o experimento com um menor número de indivíduos, apesar das "chances" reduzidas e, no caso de "azar" continuar o experimento até alcançar o resultado desejado. Assim procedendo progressivamente perdemos tempo, mas, limitamo-nos a produzir apenas o material absolutamente necessário. Citarei alguns exemplos dêste processo parcial ou progressivo dos nossos estudos genéticos em milho.

1º Exemplo — Nos estudos da genética do milho tunicata desejava-se saber, entre outras coisas, se as plantas homozigotas TuTu foram igualmente férteis, as plantas heterozigotas Tutu, usando sementes formadas nas flechas. A probabilidade de encontrar plantas TuTu é de 1 em 3 ou 0,33. Para identificar no mínimo um indivíduo TuTu precisa-se ento de acôrdo com as fórmulas dadas acima, no mínimo 9 famílias descendentes de indivíduos autofecundados (Precisão 5%) ou de 14 famílias (Precisão 1%) e preferia de ter no mínimo três ou quatro famílias de plantas TuTu, sendo então os números mínimos : 23 e 29 respectivamente (Precisão 5%) e 29 e 34 respectivamente (Precisão 1%). Porém, por falta de terreno, pude plantar apenas 12 famílias em três grupos sucessivos de 4 cada uma, estando disposto se necessário aumentar o número de famílias.

Os resultados obtidos no ano agricola de 1945-46 foram os seguintes :

	Observado :
Primeiro grupo de 4 famílias :	1 TuTu — 3 Tutu
Segundo grupo de 4 famílias :	1 TuTu — 3 Tutu
Terceiro grupo de 4 famílias :	1 TuTu — 3 Tutu
Total de 12 famílias	3 TuTu — 9 Tutu

Evidentemente tivemos "sorte". A probabilidade de obter uma familia TuTu em quatro podemos calcular. O segundo têrmo do binômio $\left(\frac{2}{3} + \frac{1}{3}\right)^4$ é igual a $4.\left(\frac{2}{3}\right)^3 . \frac{1}{3}$ ou 0,3950 e a probabilidade de obter êste resultado três vezes em seguida é igual a terceira presença de 0,3953 ou 0,06. Assim o resultado obtido podia ser esperado com a frequência de 1 em 17, isto é, bem frequentemente

2.º Exemplo : Em famílias de milho segregando na proporção de 9+3+3+1 para plantas roxas (B—Pl—), plantas roxas diluidas, plantas "sunred" e "sunred diluido", queremos isolar alguns indivíduos da constituição homozigota "roxo forte (BB Pl Pl). Êles são esperados dentro dos roxos fortes com a frequência p=1 : 9=0,11. O número mínimo necessário para achar um só indivíduo será 24 (Precisão 50) e 49 (Precisão 1%) e para ter no mínimo quatro indivíduos, os números serão : 7? (Precisão 5%) e 104 (Precisão 1%).

Não tivemos no ano passado bastante plantas à disposição de modo que, fomos forçados a proceder progressivamente (

iniciar o teste para homozigotia com o reduzido material à disposição. Os resultados obtidos até agora são os seguintes :

1.º teste : Em 23 famílias de plantas autofecundadas : Nenhum indivíduos de BB Pl Pl.

Apesar de ser a nossa expectativa média de encontrar indivíduos em cada 9 indivíduos estudados, tivemos o "azar" de não encontrar ainda nenhum indivíduo de constituição desejada nos primeiros 23 indivíduos estudados.

3.º **Exemplo** : Dos estudos sôbre a hereditariedade em milho indígena citamos também um exemplo. Num conjunto de 29 espigas do milho "Diamantino", cultivado pelos Bororós, podiam aparecer grãos brancos ou coloridos de acôrdo com 6 diferentes fórmulas genéticas, e era de interêsse saber se tôdas as seis diferentes proporções apareceram de fato. Mas, como mostra o Quadro 9, o número mínimo de indivíduos necessários para distinguir tôdas as diferentes frequências mend nas é em parte tão grande que é impossível encontrá-las em espigas individuais. Assim era a única esperança compensar a falta de números de grãos por espiga, aumentando o número de espigas até obter os resultados definitivos.

Os resultados obtidos constam no Quadro 10. Para a análise empregámos o método seguinte : As espigas foram organizadas em ordem crescente da porcentagem de grãos incolores e depois foram calculados os valores de X^2 para as diversas expectativas mendelianas. O Quadro 10 contém apenas os valores insignificantes, isto é, menores do que 6,66... (Precisão 1%).

E' evidente que um grande número de espigas segue a proporção 1:1, ficando apenas uma espiga duvidosa com valores de X^2 relativamente pequenos tanto para a razão 1:1 como 3:5. Com respeito às demais espigas a situção é mais complicada. Temos uma espiga que está de acôrdo apenas com a proporção 3:5, outra com a proporção 3:13 e duas com a proporção 1:7. Assim constatamos a existência de quatro das seis proporções mendelianas esperadas. Para os dois restantes os números ainda não são suficientes apesar de que dispomos de 11 espigas com um total de 1.333 grãos.

De um modo geral pode-se tirar a seguinte conclusão prática : Se o total de grãos nas espigas que estão de acôrdo com uma proporção mendeliana como demonstrado pelos valores de X^2 menores do que 6,66 (Precisão 1%) for igual ou maior do que o número mínimo total exigido para uma distinção podemos esperar que no mínimo uma destas espigas permitirá uma distinção clara entre as proporções estudadas. (Quadro 11)

Os três exemplos apresentados demonstram claramente que podemos ter a "sorte" de obter resultados decisivos mesmo quando o total de indivíduos for menor do que o número mínimo, necessário para ter relativa garantia dentro dos limites escolhidos de precisão.

III — CONCLUSAO

O processo para calcular os números mínimos que devem ser considerados como os mais exatos, consiste no seguinte : a) determinação da frequência p dos acontecimentos desejados e da frequência $q=1-p$ dos acontecimentos não desejados; b) determinação da frequência total de tôdas as combinações de acontecimentos desejados e não desejados as quais queremos evitar em n repetições, sendo necssário para isso calcular a soma acumulada dos primeiros **m** têrmos do binômio $(p+q)n$; c) estabelecer o limite de precisão que queremos aplicar; d) achar o valor do expoente n do binômio de tal modo que o valor da soma acumulada das frequências mencionadas no ponto b, seja no máximo igual ao limite de precisão escolhido.

$$q^{n} + nq^{n-1} \cdot p + \frac{n(n-1)}{2!} q^{n-2} \cdot p^{2} + \frac{n(n-1)(n-2)}{3!} q^{n-3} p^{3} \leqq P.lim. - - (1)$$

Com referência à escolha do limite de precisão não podemos esquecer o fato. que expliquei antes (1937, 1945, 1946) que não existe um limite absoluto que possa ser aplicado de um modo geral. Além dos fatores subjetivos do julgamento individual, depende o limite de precisão do número de observações e repetições a serem feitas. Recomendei como valor indicado para o limite mínimo de probabilidade de um acontecimento o valor $P.lim=1 \div 5n$ e como limite máximo da improbabilidade o valor $P.lim=1 \div 10n$, ficando entre ambos o que chamei a região da dúvida. Uma vez que nos casos a serem tratados nesta publicação o valor de n é justamente a desconhecida a ser determinada, teremos que recorrer ao emprêgo dos limites convencionais de precisão : $P.lim=0{,}05(5\%)$, $P.lim=0{,}01(1\%)$ e $P.lim=0{,}001(1‰)$.

A aplicação do teorema do binômio, seguindo a fórmula básica (1) torna-se em geral inexequível pelo trabalho do cálculo excessivo, de modo que, temos de achar fórmulas aproximadas. Foi demonstrado que podemos usar sem perda de precisão, as seguintes duas aproximações :

1) A distribuição normal de **Gauss** com média n.p e com êrro standard igual a $\sqrt{p(1-p)\,n}$, quando n fôr um número maior do que 10 e p tiver valores entre 0,1 e 0,9.

2) As distribuições de **Poisson** com média $\bar{m}=n.p$ quando p fôr entre 0,0 e 0,1.

Podemos agora dar as soluções para os cinco problemas enumerados na introdução.

A) O número mínimo de repetições necessário para que um determinado tratamento, variedade entre a tratamentos estudados, não ocupe em **n** repetições acidentalmente sempre o primeiro, segundo... o m.º lugar, determina-se pela **fórmula**:

$$\left(\frac{m}{a}\right)^{n} \gtrless P.lim. \quad \text{------------------} (4)$$

$$n = \frac{\log P.lim.}{\log (m:a)} = \frac{\log P.lim.}{\log.m - \log.a} \quad \text{-----------} (5)$$

B) O número total mínimo necessário para ter um determinado número $a=m+1$ de indivíduos de um certo tipo esperado com a frequência p, determina-se do modo seguinte:

1) Quando p tem qualquer valor entre 0,1 e 0,9 e **n** for maior do que 10.

$$pn - \delta.\sqrt{p(1-p)n} = a - 1 = m$$

$$\left.\begin{array}{l} b = \delta.\sqrt{\dfrac{1-p}{p}} \qquad c = \dfrac{m}{p} \\[2ex] n = \left\{\dfrac{b+\sqrt{b^2+4c}}{2}\right\}^2 \end{array}\right\} \text{-------------} (7)$$

$$\left.\begin{array}{l} n' = \dfrac{1}{p} \\[2ex] n(cor.) = n + n' \end{array}\right\} \text{----------------} (8)$$

A correção n' é apenas necessária quando p tem valores entre 0,2 e 0,8.

Se o valor a=1 a fórmula se simplifica :

$$
\left.
\begin{array}{l}
c = \frac{m}{p} = \frac{a-1}{p} = 0 \quad \textit{para} \quad a = 1 \\
n = \left\{ \frac{b+\sqrt{b^2}}{2} \right\}^2 = b^2 = \delta^2 \frac{1-p}{p} \\
n' = \frac{1}{p} \\
n(cor.) = n + n'
\end{array}
\right\} \text{-----------} (9)
$$

Os limites unilaterais de Gauss são δ =1,96 (P.lim=0,05), δ =2,33 (P.lim=0,01), δ =3,09 P.lim = 0,001).

2) Quando p for menor do que 0,1, escolhe-se na táboa 1 o valor de m da série de **Poisson**, e calcula-se :

$$n = \overline{m} : p$$

C) O número mínimo de indivíduos necessário para poder distinguir entre duas frequências p1 e p2 determina-se pelos processos seguintes :

1) Quando p1 e p2 são valores entre 0,1 e 0,9 :

$$
\left.
\begin{array}{l}
n = \left\{ \delta \frac{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}{p_1 - p_2} \right\}^2 \\
n' = \frac{1}{p_1 - p_2} \\
n(cor.) = n + n'
\end{array}
\right\} \text{--------} (13)
$$

Os valores dos limites unilaterais da distribuição de **Gauss** são os mesmos citados acima (1,64-2,33-3,09).

Uma fórmula mais complicada para casos especiais onde queremos mais rigor está dada na **fórmula** :

$$n(p_1-p_2) - \delta \cdot \left\{ \sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)} \right\} \cdot \sqrt{n} = m$$

$$b = \delta \cdot \frac{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}{p_1 - p_2} \qquad c = \frac{m}{p_1 - p_2}$$

$$
\left.
\begin{array}{l}
n = \left\{ \frac{b+\sqrt{b^2+4c}}{2} \right\}^2 \\
n' = \frac{1}{p_1 - p_2} \\
n(cor) = n + n'
\end{array}
\right\} \text{--------------} (14)
$$

2) Quando p1 e p2 são menores do que 0,1 empregamos a táboa 2, calculando o quociente de p1 dividido por p2' e procurando na táboa o valor correspondente a $\bar{m}2$ da série de **Poisson.**

Calcula-se depois :

$$n = \frac{\bar{m}_2}{p_2} \text{ ---------------------- (15)}$$

) Os processos necessários para determinar o número total mínimo necessário na distinção de três frequências, são os seguintes :

1) Quando as três frequências, p2 maior do que p1 maior que p3, têm valores entre 0,1 e 0,9, resolvem-se ambas as equações seguintes, usando depois o valor maior de $\bar{n}$ achado :

$$\left.\begin{array}{l} n_{1,2} = \left\{ \delta \dfrac{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}{p_1 - p_2} \right\}^2 = \text{P.lim.} \\ n_{1,3} = \left\{ \delta \dfrac{\sqrt{p_3(1-p_3)} + \sqrt{p_1(1-p_1)}}{p_3 - p_1} \right\}^2 = \text{P.lim.} \end{array}\right\} \text{-- (16)}$$

Os valores dos limites bilaterais da distribuição de Gauss que teremos que usar aqui, são : δ =10,96 (P.lim=0,05); δ = **2,58** (P.lim=0,1) e δ =3,29 (P.lim=0,001).

Para evitar o cálculo, podemos também usar uma táboa publicada por BRIEGER (1937, táboa 12).

2) Não foi dada uma táboa para os limites bilaterais das respectivas séries de **Poisson** que deveríamos usar quando os valores de p foram, todos os três, menores do que 0,1. Pois êstes casos so raros e os números de indivíduos necessários em geral excessivamente grandes, de modo que não vale a pena calcular uma táboa especial. Com aproximação pode-se usar a táboa 2, apesar de serem empregados nelas os limites unilaterais e não bilaterais da série de **Poisson.**

E) Finalmente podemos resolver a pergunta informativa : tendo achado em n indivíduos uma frequência p, queremos saber quais os valores extremos de p(esp) dos quais o valor f(obs) pode ser um desvio do acaso.

1) Quando p(obs) e também a frequência p(esp) têm valores entre os extremos 0,1 e 0,9 empregamos a táboa 3, verificando na **linha** horizontal que corresponde aos valores f(obs) em que coluna ou entre quais colunas o valor n da táboa cor-

responde ao número do experimento, achando assim os valores extremos de p(esp).

A táboa pode também ser usada da forma inversa. Sabendo num experimento qual o valor p(esp), podemos determinar os valores extremos de p(obs) que podem ser encontrados para qualquer valor de n. Começamos então com as **colunas** que correspondem a p(esp) e verificamos em que linha ou entre quais linhas horizontais encontra-se o respectivo valor de n.

2) Quando os valores forem menores do que 0,1 temos que usar a transformação de **Poisson** com $\overline{m} = n.p$, empregando a táboa 4. Tendo achado um valor qualquer de p(obs) em n indivíduos calculamos o número correspondente de $\overline{m}$(obs). Comparando êste valor com os limites dados na táboa 4, achamos fàcilmente os valores de m(esp) dos quais m(obs) pode ainda ser um desvio de acaso. Pela relação $p = \overline{m} + n$ achamos então os valores correspondentes de p(esp).

Como no caso anterior e apenas invertendo o processo podemos na mesma táboa determinar os limites de variação de uma frequência $p(esp) = \overline{m} + n$.

TESTES PARCIAIS PROGRESSIVOS

Não muito raramente torna-se impossível ou dispendioso demais a obtenção de um número tão elevado como o número mínimo necessário para ter resultados garantidos, dentre os limites de precisão escolhidos. Podemos então confiar em nossa "sorte" e iniciar o experimento com um número bem menor, aumentando o número até alcançar o resultado desejado, e frequentemente não será mesmo necessário continuar até atingir o número mínimo calculado. Trabalhando assim progressivamente podemos economizar material, perdendo porém em compensação, tempo. Exemplos concretos do processo foram discutidos.

IV) — ABSTRACT

The main object of the present paper consists in giving formulas and methods which enable us to determine the minimum number of repetitions or of individuals necessary to garantee some extent the success of an experiment. The theoretical basis of all processes consists essentially in the following. Knowing the frequency of the desired **p** and of the non desired events q we may calculate the frequency of all possi-

ble combinations, to be expected in n repetitions, by expanding the binomium $(p+q)n$.

Determining which of these combinations we want to avoid we calculate their total frequency, selecting the value of the exponent **n** of the binomium in such a way that this total frequency is equal or smaller than the accepted limit of precision.

$$n!p^n \cdot \left\{ \frac{1}{n!}\left(\frac{q}{p}\right)^n + \frac{1}{(n-1)!}\left(\frac{q}{p}\right)^{n-1} + \frac{1}{2!(n-2)!}\left(\frac{q}{p}\right)^{n-2} + \frac{1}{3!(n-3)!}\left(\frac{q}{p}\right)^{n-3} \cdots\cdots \overline{\lesssim} \text{P.lim.} \quad \text{--(1b)} \right.$$

There does not exist an absolute limit of precision since its value depends not only upon psychological factors in our judgement, but is at the same sime a function of the number of repetitions For this reasen y have proposed (1,56) two relative values, one equal to $1 \div 5n$ as the lowest value of probability and the other equal to $1 \div 10n$ as the highest value of improbability, leaving between them what may be called the "region of doubt However these formulas cannot be applied in our case since this number n is just the unknown quantity. Thus we have to use, instead of the more exact values of these two formulas, the conventional limits of P.lim equal to 0,05 (Precision 5%), equal to 0,01 (Precision 1%, and to 0,001 (Precision P, 1%).

The binominal formula as explained above (cf. formula 1, pg. 85), however is of rather limited applicability owing to the excessive calculus necessary, and we have thus to procure approximations as substitutes. We may use, without loss of precision, the following approximations: a) The normal or **Gaussean** distribution when the expected frequency p has any value between 0,1 and 0,9, and when n is at least superior to ten.

b) The **Poisson** distribution when the expected frequecy p is smaller than 0,1.

Tables V to VII show for some special cases that these approximations are very satisfactory.

The pratical solution of the following problems, stated in the introduction can now be given:

A) What is the minimum number of repititions necessary in order to avoid that any one of **a** treatments, varieties etc. may be accidentally always the best, on the best and second best, or the first, second, and third best or finally one of the n best treatments, varieties etc. Using the first term of the binomium, we have the following equation for n:

$$n = \frac{\log Rlim.}{\log (m:a)} = \frac{\log Rlim.}{\log m - \log a} \text{ .... } (6)$$

B) What is the minimun number of individuals necessary in order that a ceratin type, expected with the frequency p, may appaer at least in one, two, three or a=m+1 individuals.

1) For p between 0,1 and 0,9 and using the Gaussean approximation we **have**:

$$pn - \delta\sqrt{p(1-p)n} = a - 1 = m$$

$$\left.\begin{array}{l} b = \delta\sqrt{\frac{1-p}{p}} \quad c = \frac{m}{p} \\ n = \left\{\frac{b + \sqrt{b^2 + 4c}}{2}\right\}^2 \end{array}\right\} \text{ .... } (7)$$

$$\left.\begin{array}{l} n' = \frac{1}{p} \\ n(cor.) = n + n' \end{array}\right\} \text{ .... } (8)$$

We have to use the correction n' when p has a value between 0,25 and 0,75. The greek letters delta represents in the present case the unilateral limits of the Gaussean distribution for the three conventional limits of precision: 1,64; 2,33; and 3,09 respectively.

If we are only interested in having at least one individual, and m becomes equal to zero, the formula reduces to:

$$c = \frac{m}{p} = \frac{a-1}{p} = 0 \quad \textit{para} \quad a = 1$$

$$\left.\begin{array}{l} n = \left\{\frac{b + \sqrt{b^2}}{2}\right\}^2 = b^2 = \delta^2 \frac{1-p}{p} \\ n' = \frac{1}{p} \\ n(cor.) = n + n' \end{array}\right\} \text{ .... } (9)$$

2) If p is smaller than 0,1 we may use table 1 in order to find the mean m of a **Poisson distribution** and determine.

$$n = \overline{m} : p$$

C) Which is the minimun number of individuals necessary for distinguishing two frequencies p1 and p2 ?

1) When p1 and p2 are values between 0,1 and 0,9 we have:

$$\left.\begin{array}{l} n = \left\{ \delta \dfrac{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}{p_1 - p_2} \right\}^2 \\ n' = \dfrac{1}{p_1 - p_2} \\ n(cor.) = n + n' \end{array}\right\} \text{- - - - - - - (13)}$$

We have again to use the unilateral limits of the **Gaussean** distribution. The correction n' should be used if at least one of the valors p1 or p2 has a value between 0,25 and 0,75.

A more complicated formula may be used in cases where whe want to increase the precision :

$$n(p_1 - p_2) - \delta \cdot \left\{ \sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)} \right\} \cdot \sqrt{n} = m$$

$$b = \delta \cdot \frac{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}{p_1 - p_2} \qquad c = \frac{m}{p_1 - p_2}$$

$$\left.\begin{array}{l} n = \left\{ \dfrac{b + \sqrt{b^2 + 4c}}{2} \right\}^2 \\ n' = \dfrac{1}{p_1 - p_2} \\ n(cor.) = n + n' \end{array}\right\} \text{- - - - - - - - - - - - - (14)}$$

2) When both p1 and p2 are smaller than 0,1 we determine the quocient (p1÷p2) and procure the corresponding number m2 of a **Poisson** distribution in table 2. The value n is found by the equation :

$$n = \frac{m_2}{p_2} \text{ - (15)}$$

D) What is the minimun number necessary for distinguishing three or more frequencies, p2 p1 p3.

1. If the frequecies p1 p2 p3 are values between 0,1 e 0,9 we have to solve the individual equations and sue the higest value of n thus determined :

$$\left.\begin{array}{l} n_{1,2} = \left\{ \delta \dfrac{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}{p_1 - p_2} \right\}^2 = F.lim \\ n_{1,3} = \left\{ \delta \dfrac{\sqrt{p_3(1-p_3)} + \sqrt{p_1(1-p_1)}}{p_3 - p_1} \right\}^2 = P.lim \end{array}\right\} \text{- - (16)}$$

Delta represents now the bilateral limits of the : **Gaussean** distribution : 1,96-2,58-3,29.

2) No table was prepared for the relatively rare cases of a comparison of three or more frequencies below 0,1 and in such cases extremely high numbers would be required.

E) A process is given which serves to solve two problems of informatory nature : a) if a special type appears in **n** individuals with a frequency p(obs), what may be the corresponding ideal value of p(esp), or; b) if we study samples of n in diviuals and expect a certain type with a frequency p(esp) what may be the extreme limits of **p(obs) in individual** fa... lies ?

1) If we are dealing with values between 0,1 and 0,9 we may use table 3. To solve the first question we select the respective horizontal line for p(obs) and determine which column corresponds to our value of n and find the respective value of p(esp) by interpolating between columns.

In order to solve the second problem we start with the respective column for p(esp) and find the horizontal line for the given value of n either diretly or by approximation and by interpolation.

2) For frequencies smaller than 0,1 we have to use table 4 and transform the fractions p(esp) and p(obs) in numbers of **Poisson** series by multiplication with n.

In order to solve the first broblem, we verify in which line the lower **Poisson** limit is equal to m(obs) and transform the corresponding value of m into frequecy p(esp) by dividing through n. The observed frequency may thus be a chance deviate of any value between 0,0... and the values given by dividing the value of m in the table by n.

In the second case we transform first the expectation p(esp) into a value of m and procure in the horizontal line. corresponding to m(esp) the extreme values om m which than must be transformed, by dividing through n into values of p(obs).

F) **Partial and progressive tests** may be recomended in all cases where there is lack of material or where the loss of time is less importent than the cost of large scale experiments since in many cases the minimun number necessary to garantee the results within the limits of precision is rather large.

One should not forget that the minimun number really represents at the same time a maximun number, necessary only if one takes into consideration essentially the disfavorable variations, but smaller numbers **may frequently already** satisfactory results.

For instance, by definition, we know that a frequecy of p means that we expect one individual in every total o(f1÷p). If there were no chance variations, this number (1÷p) will be suficient, and if there were favorable variations a smaller number still may yield one individual of the desired type.

Thus trusting to luck one may start the experiment with numbers, smaller than the minimun calculated according to the formulas given above, and increase the total untill the desired result is obtained and this may well b ebefore the "minimum number" is reached.

Some concrete examples of this partial or progressive procedure are given from our genetical experiments with maize.

LITERATURA CITADA

1 — BRIEGER, F. G. — 1937 — Táboas e Fórmulas para Estatística. **Comp. Melhoramentos S. Paulo.**

2 — BRIEGER, F. G. — 1942 — Coeficiente de Variação e Índice de Variança. **Bragantia, 2 :** 315-332.

3 — BRIEGER, F. G. — 1945 — Competição entre megásporos em milho. **Anais da E. S. A. "Luiz de Queiroz", 2:** 239-267.

4 — BRIEGER, F. G. — 1945 — A ação dos gens gametofiticos com referência ao milho. **Anais da E. S. A. "Luiz de Queiroz". 2 :** 269-297.

5 — BRIEGER, F. G. — 1945 — As distribuições do Acaso. **Anais da E. S A. "Luiz de Queiroz", 2 :** 321-391.

6 — BRIEGER, F. G. — 1946 — Limites Bilaterais e Unilaterais. **Bragantia, 6 :** (em impressão).

7 — BRIEGER, F. G. — 1946 — Números mínimos na análise mendeliana. **Anais do Inst. Fitotécnico "La Estanzuela".** (em impressão).

8 — BRIEGER, F. G. — 1946 — Princípios e métodos de amostragem. **Anais do Inst. Fitotécnico de "La Estanzuela".** (em impressão).

9 — BRIEGER, F. G. SÍLVIO MOREIRA e Z. LEME — 1941 — Estado sôbre o melhoramento da laranja "Baía" III. **Bragantia, 1:** 567-610

10 — BRIEGER, F. G. e SILVIO MOREIRA — 1945 — Experiências de cavalos para Citrus II. **Bragantia, 5:** 597-658.

11 — FISCHER, R. A. and E. YATES — 1943 — Statistical Tables. **Oliver and Boyd.** Londres, 2.a **Ed.**

12 — MADOW, W. J. — 1946 — Resumo de conferências sôbre amostragem. **Piracicaba** (mimeografado).

13 — MOLINA, E. C. — 1943 — Poisson's Exponential Binomial Limite. **Van Nostrand Co. New York.**

Figura 1

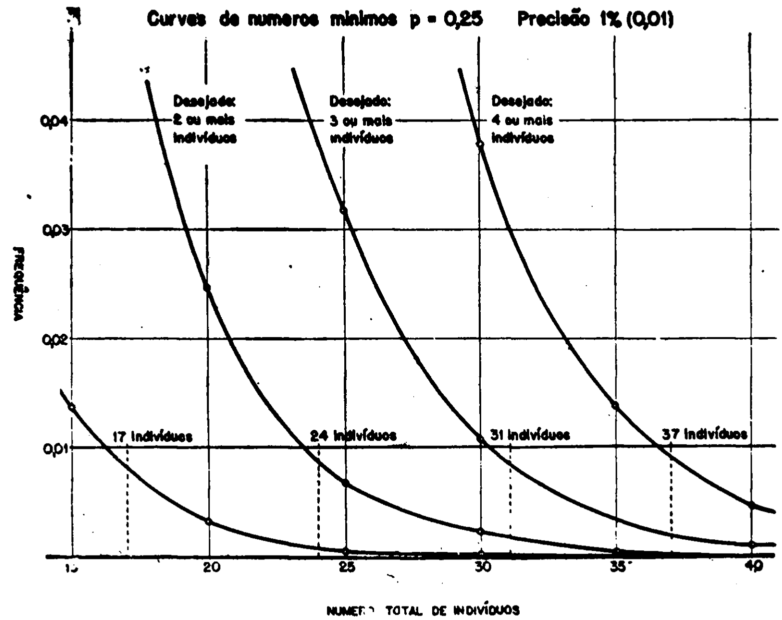


Figura 2

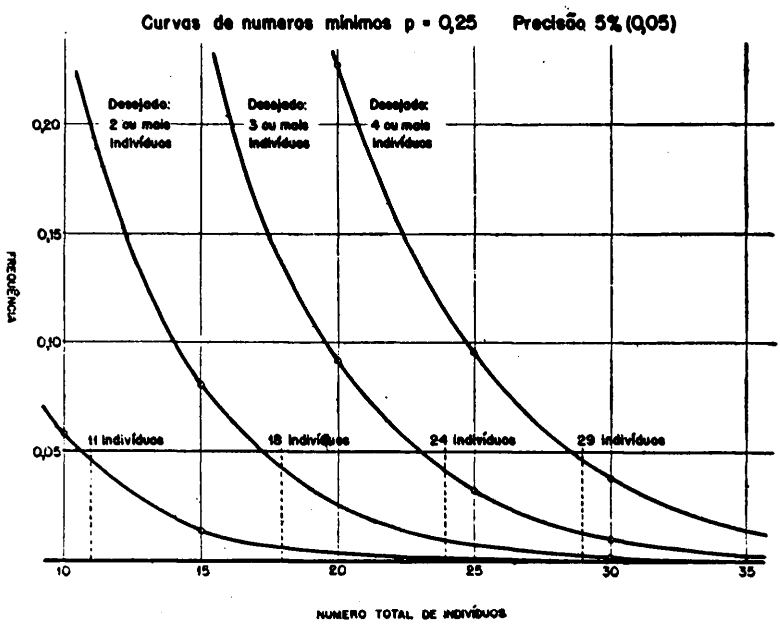
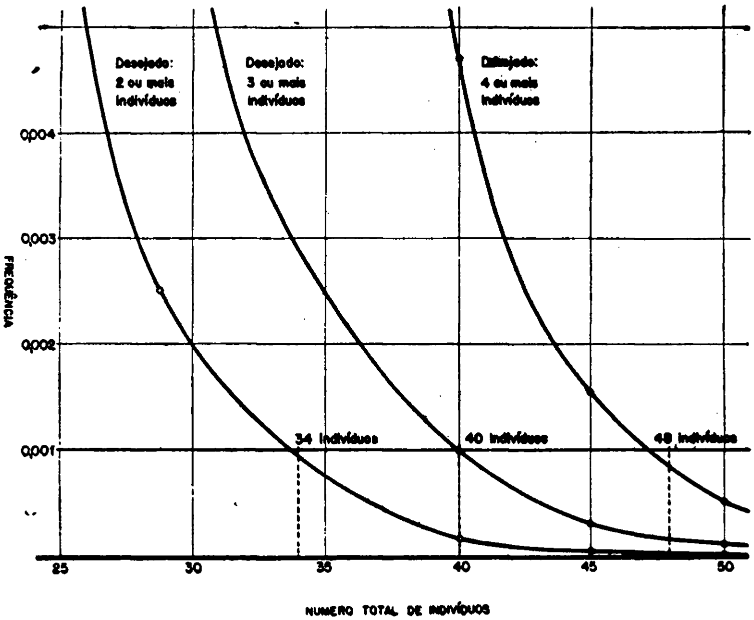


Figura 3

Curvas de numeros minimos p = 0,25 Precisão 1‰ (0,001)



TÁBOA 1

N.º de indivíduos do tipo desejado	Medias de distribuições de Poisson $\bar{m}$:		
	5%	1%	1‰
1 e mais	3,0	4,7	7,0
2 e mais	4,8	6,7	9,5
3 e mais	6,3	8,5	11,3
4 e mais	7,8	10,1	13,1

O numero mínimo total de indivíduos para qualquer valor de p entre 0,0... e 0,1 determina-se pela fórmula :

$$m = \bar{m} : p$$

Exemplo : — p = 1 em 16 e queremos ter três ou mais indivíduos com 1% precisão. Achamos então na táboa o valor de $\bar{m}$ = 8,5, e obtemos :

$$n = 8,5 : \left(\frac{1}{16}\right) = 136 \text{ indivíduos}$$

Example : — Which is the minimum number necessary in order to obtain at least thee or more individuals of a type expected with a frequency of p equal to 1:16. We find in the third line of the table the values for the means of Poisson distributions, and determine n by the formula :

$$n = \bar{m} : p$$

5% precision :	n = 6,3 : (1:16) = 100,8	101
1% precision :	n = 8,5 : (1:16) = 136,0	137
1‰ precision :	n = 11,3 : (1:16) = 180,8	181

TABOA 3

O uso da táboa é explicado na pg. 235

The use of the table is explained on pg. 248

P	P esperado															P
obs.	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,85	obs.
0,10	552 196 340	174 62 107	91 33 56	57 21 36	15 25 40	11 18 29	8 14 22	7 11 17	5 9 14	— — —	— — —	— — —	— — —	— — —	— — —	0,10
0,15	— — —	246 427 693	73 125 203	36 63 102	22 38 62	15 26 42	11 19 30	8 14 23	6 11 17	— — —	— — —	— — —	— — —	— — —	— — —	0,15
0,20	196 340 552	— — —	289 500 812	81 140 228	39 68 110	24 41 65	16 27 43	11 19 31	8 14 22	6 10 17	— — —	— — —	— — —	— — —	— — —	0,20
0,25	50 85 139	246 427 693	— — —	323 560 909	88 152 247	42 72 116	24 42 68	16 27 44	11 19 30	8 13 22	6 16 16	— — —	— — —	— — —	— — —	0,25
0,30	22 38 62	62 107 174	289 500 812	— — —	350 607 985	93 160 260	43 74 120	25 42 68	16 27 43	11 18 29	8 13 21	6 9 15	— — —	— — —	— — —	0,30
0,35	13 22 35	28 48 77	73 125 203	323 560 909	— — —	369 640 1039	96 165 268	43 75 121	24 42 63	15 26 42	10 17 28	7 12 19	— — —	— — —	— — —	0,35
0,40	8 14 23	16 23 44	33 56 91	81 140 228	350 607 985	— — —	381 660 1072	97 167 271	43 74 120	24 41 65	15 25 40	10 16 26	6 11 17	— — —	— — —	0,40
0,45	6 10 16	10 18 28	19 32 51	36 63 102	88 152 247	369 640 1039	— — —	385 667 1083	96 165 268	42 72 116	22 38 62	13 23 37	9 14 23	6 9 15	— — —	0,45
0,50	— — —	7 12 20	12 21 33	21 36 57	39 68 110	93 160 260	381 660 1072	— — —	381 660 1072	93 160 260	39 68 110	21 36 57	12 21 33	7 12 20	— — —	0,50
0,55	— — —	6 9 15	9 14 23	13 23 37	22 38 62	42 72 116	96 165 268	385 667 1083	— — —	369 640 1039	88 152 247	36 63 102	19 32 51	10 18 28	6 10 16	0,55
0,60	— — —	— — —	6 10 17	10 16 26	15 25 40	24 41 65	43 74 120	97 167 271	381 660 1072	— — —	350 607 985	81 140 228	33 56 91	16 27 24	8 14 23	0,60
0,65	— — —	— — —	— — —	7 12 19	10 17 28	15 26 42	24 42 68	43 75 121	96 165 268	369 640 1039	— — —	323 560 909	73 125 203	28 48 77	13 22 35	0,65
0,70	— — —	— — —	— — —	6 9 15	8 13 21	11 18 29	16 27 43	25 42 68	43 74 120	93 160 260	350 607 985	— — —	289 500 812	63 107 174	22 28 62	0,70
0,75	— — —	— — —	— — —	— — —	6 10 16	8 14 22	11 19 30	16 27 44	24 42 68	42 72 116	88 152 247	323 560 909	— — —	246 427 693	50 85 139	0,75
0,80	— — —	— — —	— — —	— — —	— — —	6 10 17	8 14 22	11 19 31	16 27 43	24 41 65	39 368 110	81 140 228	298 500 872	— — —	196 340 552	0,80
0,85	— — —	— — —	— — —	— — —	— — —	— — —	6 11 17	8 14 22	11 19 30	15 26 42	22 38 62	36 63 102	73 125 203	246 427 693	— — —	0,85
0,90	— — —	— — —	— — —	— — —	— — —	— — —	5 9 14	6 11 17	8 14 22	11 18 29	15 25 40	21 36 57	33 587 91	62 107 174	196 340 552	0,90
P	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,85	P
obs.	P esperado															obs.

TABOA 2

Número mínimo de indivíduos necessários para distinguir duas probabilidades de p1 e p2, sendo ambas menores do que 0,1. Limites de Precisão : 0,01 (1%).

P1 : P2	$\overline{m}2$	P1 : P2	$\overline{m}2$
1,70	58,1	3,0 — 3,5	8,1
1,75	52,0	3,5 — 4,0	6,0
1,80	46,0	4,0 — 4,5	4,7
1,85	40,8	4,5 — 5,0	4,1
1,90	37,5	5,0 — 5,5	3,5
1,95	34,2	5,5 — 6,0	2,9
2,00	31,3	6,0 — 6,5	2,3
2,1	26,8	6,5 — 8,0	1,7
2,2	23,5	8,0 — 10,0	1,2
2,3	21,3		
2,4	18,8		
2,5	16,5		
2,6	14,8		
2,7	13,3		
2,8	11,8		
2,9	9,6		
3,0	8,1		

Interpolação de $\overline{m}2$:
Para valores de p1 ÷ p2 entre 1,70 e 3,00, pode-se usar uma interpolação proporcional :
Exemplo : p1:p2 = 2,36

	2,30	21,3
	2,40	18,8
Dif.	0,10	-2,5
	0,06	-2,5 X 0,06 = -1,50
	2,36	21,3 — 1,5 = 19,8

Para qualquer par de valores p1 e p2 determina-se o seu quociente e procura-se na táboa o valor correspondente de $\overline{m}$. O valor n do número mínimo total necessrio para distinguir p1 e p2 determina-se pela fórmula :

$$n = \overline{m}2 : p2$$

Exemplo : p1 = 0,04; p2 = 0,02

$$\frac{p1}{p2} = \frac{0,04}{0,02} = 2,0$$

Acha-se na táboa para 2,0 o valor de $\overline{m}$ = 31,3

$$n = \overline{m} : p2 = 31,3 : 0,02 = 1.565 \text{ indivíduos}$$

Example : — Wich is the minimum number necessary in order to distinguish the two expected frequencies of p equal to 0,04 and 0,02 ?

$$\frac{p1}{p2} = 2,0$$

We find for 2,0 in the first column of the table, the value of m equal to 31,3 in the second column. Thus we may determine the minimum number by the equation :

$$n = m : p2 = 31,3 : 0,02 = 1.565$$

TABOA 4

Números extremos das séries de Poisson

Limites de Precisão			$\bar{m}$	Limites de Precisão		
1‰	1%	5%		5%	1%	1‰
0	0	0	1	3	5	6
0	0	0	2	5	6	8
0	0	0	3	7	8	10
0	0	1	4	8	10	12
0	1	2	5	10	12	14
0	1	2	6	11	13	16
1	2	3	7	12	14	17
2	3	4	8	14	16	19
2	3	5	9	15	18	20
3	4	5	10	17	19	22
3	5	6	11	18	20	23
3	5	7	12	19	22	25
4	5	7	13	21	23	26
5	6	8	14	22	25	28
5	7	9	15	23	26	29
6	8	10	16	24	27	31
6	8	10	17	26	28	32
7	9	11	18	27	30	33
7	10	12	19	28	31	35
8	11	13	20	29	32	36

Exemplo : — Quais são os **valores** extremos da variação do acaso no limite de 1% precisão para $p = 0,05$ e $n = 200$? $m = 0,05 . 200 = 10$. Nas colunas de 1% de precisão achamos na linha de $m = 10$ os dois valores de 4 e 19. Temos então :

$$p(\text{max.}) = \frac{m(\text{max})}{n} = \frac{19}{200} = 0,095$$

$$p(\text{min.}) = \frac{m(\text{min})}{n} = \frac{4}{200} = 0{,}020$$

Exemplo : — Quais os valores de p(esp) dos quais um valor de p(obs) = 0,02, achado num total de 300 indivíduos, pode representar um desvio de acaso

$$m(\text{obs}) = 0{,}02.300 = 6$$

Usando apenas a coluna dos limites de 1% e descendo de cima para baixo, encontramos o valor de 6 como limite inferior na linha de $\overline{m} = 14$ e como limite superior de $\overline{m} = 2$. Assim temos :

p(esp) max = 14:300 = 0,0477
p(esp) min = 2:300 = 0,0067

Example : — 1) What are the extréme deviates, at the 1% limite of precision, for p(esp) = 0,05 e n = 200 ?

We have $\overline{m}$ = p.n = 0,05 x 200 = 10, and find in the 1% column of the table, in the horizontal line for $\overline{m}$ equal 10, the two values 4 and 19. Thus we gest :

$$p(\text{max}) = \frac{\overline{m}(\text{max})}{n} = \frac{19}{200} = 0{,}095$$

$$p(\text{min}) = \frac{\overline{m}(\text{min})}{n} = \frac{4}{200} = 0{,}002$$

2) Which are the possible values of p (esp), corresponding to a value of p (obs) = 0,02 found in a total of 300 individuals (using the 1% limite) ?

$$\overline{m}(\text{obs}) = 0{,}02 \times 300 = 6$$

Using the second column, we encounter the value 6 in the row with m equal to 14 and in the sixth column for $\overline{m} = 2$. Thus we have :

p (esp) max = 14:300 = 0,0477
p (esp) min = 2:300 = 0,0067

QUADRO I

Cálculo das frequências dos têrmos do binômio (p = 0,25)

Têrmos binomiais	n = 10	n = 15	n = 20	n = 25	n = 30
n!	6,55 976	12,11 650	18,38 612	25,19 065	32,42 366
pn	0,97 940 - 7	0,96 910 - 10	0,95 880 - 13	0,94 850 - 16	0,93 820 - 19
n!pn	0,53 916	3,08 560	6,34 492	10,13 915	14,36 186
n!pn	0,53 916	3,08 560	6,34 492	10,13 915	14,36 186
(q:p)n	4,77 120	7,15 680	9,54 240	11,92 800	14,31 360
n!pn(p:q)n	5,31 036	10,24 240	15,88 732	22,06 715	28,67 546
n!	6,55 976	12,11 650	18,38 612	25,19 065	32,42 366
Dif.	0,75 060 - 2	0,12 590 - 2	0,50 120 - 3	0,87 650 - 4	0,25 180 - 4
1.o têrmo	0,05631	0,01336	0,00317	0,00073	0,0001[illegible]
n!pn	0,53 916	3,08 560	6,34 492	10,13 915	14,36 186
(q:p)n-1	4,29 408	6,67 968	9,06 528	11,45 088	13,83 648
n!pn(q:p)n-1	4,83 324	9,76 528	15,41 020	21,59 003	28,19 834
(n-1)!	5,55 976	10,94 041	17,08 509	23,79 271	30,94 654
Dif.	0,27 348 - 1	0,82 487 - 2	0,32 511 - 2	0,79 732 - 3	0,25 180 - [illegible]
2.o têrmo	0,18771	0,06681	0,02114	0,00627	0,00179
n!pn		3,08 560	6,34 492	10,13 915	14,36 186
(q:p)n-2		6,20 256	8,58 816	10,97 376	13,35 936
n!pn(q:p)n-2		9,28 816	14,93 308	21,11 291	27,72 122
2!		0,30 103	0,30 103	0,30 103	0,30 103
(n-2!)!		9,79 428	15,80 634	22,41 249	29,48 414
2!(n-2)!		10,09 531	16,10 737	22,71 352	29,78 517
Dif.		0,19 285 - 1	0,82 571 - 2	0,39 939 - 2	0,93 605 - [illegible]
3.o têrmo		0,15590	0,06694	0,02508	0,0086[illegible]
n!pn			6,34 492	10,13 915	14,36 186
(q:p)n-3			8,11 104	10,49 664	12,88 224
n!pn(q:p)n-3			14,45 596	20,63 579	27,24 410
3!			0,77 718	0,77 718	0,77 718
(n-3!)!			14,55 107	21,05 077	28,03 698
3!(n-3)!			15,32 922	21,82 892	28,81 513
Dif.			0,12 674 - 1	0,80 687 - 2	0,42 897 - 2
4.o têrmo			0,13389	0,06410	0,02685

Os números neste Quadro representam logarítmos decádicos, exceto os números nas linhas para o 1.º termo, o 2.º termo, etc.

QUADRO II

Cálculo das frequências dos têrmos do binômio (p = 0,25)

(Continuação)

Têrmos binomiais	n = 35	n = 40	n = 45	n = 50
n!	40,01 423	47,91 165	56,07 781	64,48 307
pn	0,92 790 - 22	0,91 760 - 25	0,90 730 - 28	0,89 700 - 31
n!pn	18,94 213	23,82 925	28,98 511	34,38 007
n!pn	18,94 213	23,82 925	28,98 511	34,38 007
(q:p)n	16,69 920	19,08 480	21,47 040	23,85 600
n!pn(p:q)n	35,64 133	42,91 405	50,45 551	58,23 607
n!	40,01 423	47,91 165	56,07 781	64.48 307
Dif.	0,62 710 - 5	0,00 240 - 5	0,37 240 - 6	0.75 300 - 7
1.o têrmo	**0,00004**	**0,00001**	**0,00000**	**0,00000**
n!pn	18,94 213	23,82 925	28,98 511	34,38 007
(q:p)n-1	16,22 208	18,60 768	20,99 328	23,37 888
n!pn(q:p)n-1	35,16 421	42,43 693	49.97 839	57,75 895
(n-1)!	38,47 016	46,30 959	54,42 460	62,78 410
Dif.	0,69 405 - 4	0,12 734 - 4	0,55 379 - 5	0,97 485 - 6
2.o têrmo	**0,00049**	**0,00013**	**0,00003**	**0,00001**
n!pn	18,94 213	23,82 925	28,98-511	34,38 007
(q:p) n-2	15,74 496	18,13 056	20,51 116	22,90 176
n!pn(q:p)n-2	34,68 709	41,95 981	49,49 627	57,28 183
2!	0,30 103	0,30 103	0,30 103	0,30 103
(n-2)!	36,93 869	44,71 852	52.78 115	61,09 391
2!(n-2)!	37,23 972	45,01 955	53.08 218	61,39 494
Dif.	0,44 737 - 3	0,94 026 - 4	0,41 409 - 4	0,88 689 - 5
3.o têrmo	**0,00280**	**0,00087**	**0,00026**	**0,00008**
n!pn	18,94 213	23,82 925	28,98 511	34,38 007
(q:p)n-3	15,26 784	17,65 344	20,03 904	22,42 464
n!pn(q:p)n-3	34,20 997	41,48 269	49,02 415	56,80 471
3!	0,77 815	0,77 815	0,77 815	0,77 815
(n-3)!	35,42 017	43,13 874	51,14 768	59,41 267
3!(n-3)!	36,19 832	43,91 689	51,92 583	60,19 082
Dif.	0,01 165 - 2	0,56 580 - 3	0,09 832 - 3	0,61 389 - 4
4.o têrmo	**0,01037**	**0,00368**	**0,00125**	**0,00041**

Os números neste Quadro representam logarítmos decádicos, exceto os números nas linhas para o 1.º termo, o 2.º termo, etc.

Quadro III

Têrmos do Binômio Acumulados	N. de indivíduos do tipo desejado (p = 0,25)	Número total de indivíduos: 10	15	20	25	30	35	40	45	50
		Frequência dos têrmos do Binômio								
1.o têrmo	1 e mais	0,05631	0,01336	0,00317	0,00013	0,00018	0,00004	0,00001	0,00000	0,00000
1.o-2.o	2 e mais	0,24402	0,08017	0,02431	0,00640	0,00197	0,00053	0,00014	0,00003	0,00001
1.o-2.o-3.o	3 e mais	—	0,23607	0,09125	0,03148	0,01060	0,00333	0,00001	0,00029	0,00009
1.o-2.o-3.o-4.o	4 e mais	—	—	0,22514	0,09558	0,03745	0,01371	0,00469	0,00159	0,00050

QUADRO IV
Cálculo do número mínimo pela aproximação da distribuição de Gauss

p = 0,25	2 ou mais indivíduos			3 ou mais indivíduos			4 ou mais indivíduos		
Precisão :	0,05	0,01	0,001	0,05	0,01	0,001	0,05	0,01	0,001
δ	1,64	2,33	3,09	1,64	2,33	3,09	1,64	2,33	3,09
δ^2	2,69	5,43	9,55	2 69	5,43	9,55	2,69	5,43	9,55
b^2	8,07	16,29	28,65	8,07	16,29	28,65	8,07	16,29	28,65
$4c$	16,00	16,00	16,00	32,00	32,00	32,00	48,00	48,00	48,00
$b^2 + 4c$	24,07	32,29	44,65	40,07	48,29	60,65	56,07	64,29	76,65
$\sqrt{b^2 + 4c}$	4,91	5,68	6,68	6,33	6,95	7,79	7,49	8,02	8,75
b	2,84	4,03	5,35	2.84	4.03	5,35	2,84	4,03	5,35
$b + \sqrt{b^2 + 4c}$	7,75	9,71	12,03	9.17	10,98	13,14	10,33	12,05	14,10
$\sqrt{n}$	3,88	4,86	6,02	4.58	5.49	6 57	5,16	6,02	7,05
n'	15,1	23,6	36,2	21,0	30.1	43,2	26,6	36,2	49,07
n'	4	4	4	4	4	4	4	4	4
n (cor.)	20	28	41	25	35	48	31	41	54

QUADRO V

p = 0,5 / Método de cálculo	Número total mínimo necessário para a precisão 5%	1%	1%
No mínimo 1 ou mais indivíduos			
Aproximação Normal Cor.	3	6	10
Aproximação Normal	5	8	12
Binômio	4	8	12
No mínimo 2 ou mais indivíduos			
Aproximação Normal Cor.	7	9	14
Aproximação Normal	9	11	16
Binômio	7	11	15
No mínimo 3 ou mais indivíduos			
Aproximação Normal Cor.	9	13	17
Aproximação Normal	11	15	19
Binômio	11	14	19
No mínimo 4 ou mais indivíduos			
Aproximação Normal Cor.	12	16	20
Aproximação Normal	14	18	22
Binômio	15	17	23

QUADRO VI

p = 0,25 / Método de cálculo	Número total mínimo necessário para a precisão 5%	1%	1%
No mínimo 1 ou mais indivíduos			
Aproximação Normal	13	21	33
Aproximação Normal Cor.	9	17	29
Binômio	11	17	24
No mínimo 2 ou mais indivíduos			
Aproximação Normal	20	28	41
Aproximação Normal Cor.	16	24	37
Binômio	18	24	34
No mínimo 3 ou mais indivíduos			
Aproximação Normal	25	35	48
Aproximação Normal Cor.	21	31	44
Binômio	24	31	40
No mínimo 4 ou mais indivíduos			
Aproximação Normal	31	41	54
Aproximação Normal Cor.	27	37	50
Binômio	29	37	48

QUADRO VII

p = 0,1	Número total mínimo necessário para a precisão		
Método de cálculo	5%	1%	1%
No mínimo 1 ou mais indivíduos			
Aproximação Normal	27	45	96
Aproximação Poisson	30	47	75
Binômio	27	44	66
No mínimo 2 ou mais indivíduos			
Aproximação Normal	42	68	105
Aproximação Poisson	48	67	95
Binômio	45	64	88
No mínimo 3 ou mais indivíduos			
Aproximação Normal	58	85	123
Aproximação Poisson	63	85	113
Binômio	61	81	108
No mínimo 4 ou mais indivíduos			
Aproximação Normal	72	100	140
Aproximação Poisson	78	101	131
Binômio	75	97	124

QUADRO VIII

Número mínimo para distinguir entre p1 = 0,2 e p2 = 0,3
Cálculo pela aproximação da distribuição de Gauss

Precisão	5%	1%	1%
δ	1,64	2,33	3,09
$\sqrt{p1(1-p1)}$	0,40	0,40	0,40
$\sqrt{p2(1-p2)}$	0,46	0,46	0,46
Soma das raizes	0,86	0,86	0,86
δ x soma	1,0104	2,0038	2,6574
(p1 — p2)	0,10	0,10	0,10
$\sqrt{n}$	14,10	20,04	26,57
n	198,81	401,60	705,96
n"	10,00	10,00	10,00
n (cor)	208,81	411,60	715,96
Valor final	209	412	716

QUADRO IX

Cruzamento "Incolor" x "Colorido" em Milho

Pares de Fatores	Proporção Mendeliana	% Incolor	No mínimo 1% limite	No mínimo 5% limite
um	1 para 1	50,00		
			344	175
Dois	3 para 5	62,50		
			549	277
Três	9 para 23	71,88		
			2964	1484
Dois	1 para 3	75,00		
			958	473
Três	3 para 13	81,25		
			728	374
Três	1 para 7	87,50		
			46	27
Puro	0: todos	100,00		

QUADRO XI

Proporção	N.os estudados Espigas	N.os estudados Grãos	Múmero mínimo necessário (Precisão de 1°/o)	Espigas seguramente classificadas	
1 para 1	10	1715		9	
			344		
3 para 5	8	788		1	
			549		
9 para 23	12	1505		—	
			2964		1
1 para 3	12	1551		—	
			958		
3 para 13	12	2372		1	
			728		
1 para 7	6	1382		2	

QUADRO X

Milho diamantino "incolor" x "colorido"

Número 1941	Total	Incolor n	Incolor o/o	X2 teste 1:1 (25)	3:5 (62,50)	9:23 (71,88)	1:3 (75)	3:13 (81,50)	1:7 (87,50)	RESULTADOS
16-118 x 4-119	76	31	40,79	2,58	—					1 : 1
10-117 x 17-116	39	18	46,15	0,23	—					
4-119 x 24-118	315	152	48,25	0,39	—					
3-117 x 20-116	138	68	49,27	0,03	—					
3-119 x 11-118	296	146	49,32	0,05	—					
5-119 x 15-118	331	169	51,06	0,15	—					
10-121 x 2-120	73	39	53,42	0,34	2,57					
2-119 x 21-118	52	28	53,84	0,31	1,66					
26-118 x 3-119	258	142	55,03	2,61	6,14					
Tôdas 9 famílias	1.578	793	20,25	6,69						
		X2 [illegible] (1:1) [illegible]								
23-118 x 9-119	137	83	60,58	6,13	0,22	8,64	—	—	—	3:5 — — — —
6-117 x 26-116	172	112	65,12	—	0,50	3.88	8,95	—	—	3:5 ou 9:23 — — —
2-121 x 7-120	89	58	65,16	—	0,27	1,98	4,60	—	—	3:5 ou 9:23 ou 1:3 — —
9-121 x 6-120	210	146	69,52	—	4,54	3.57	3,36	—	—	3:5 ou 9:23 ou 1:3 — —
7-117 x 5-111	[illegible]	130	63,69	—	4,33	0.52	2,59	—	—	3:5 ou 9:23 ou 1:3 — —
10-123 x 2-122	197	141	71.57	—	6,91	0,01	1,24	12,12	—	— 9:23 ou 1:3 — —
2-115 x 6-114	68	50	73,53	—	3,53	0.09	0,08	2,66	—	— 9:23 ou 1:3 ou 3:13 —
4-121 x 11-120	159	118	74,20	—	9,29	0.43	0,05	5,19	—	— 9:23 ou 1:3 ou 3:13 —
10-120 x 9-121	147	110	74,83	—	—	0.64	0,00	3,98	—	— 9:23 ou 1:3 ou 3:13 —
14-120 x 1-121	107	85	77,44	—	—	1.54	0,34	1,01	—	— 9:23 ou 1:3 ou 3:13 —
3-120 x 7-121	242	189	78,09	—	—	4.63	1,23	1,59	—	— 9:23 ou 1:3 ou 3:13 —
2-117 x 3-116	136	107	78.68	—	—	3,12	0,98	0,59	—	— 9:23 ou 1:3 ou 3:13 —
7-120 x 3-121	253	199	78,85	—	—	6.10	2,00	0,96	—	— 9:23 ou 1:3 ou 3:13 —
10-115 x 16-114	221	182	82,35	—	—	12.01	6.37	0.18	54,57	— — — 3:13 —
3-121 x 5-120	242	203	83,89	—	—	—	10.20	1,11	2,88	— — — 3:13 ou 1:7
7-121 x 11-120	414	356	86,00	—	—	—	—	6,13	0,85	— — — 3:13 ou 1:7
8-121 x 5-120	224	193	86,16	—	—	—	—	3,54	0,37	— — — 3:13 ou 1:7
6-121 x 9-120	159	139	87,43	—	—	—	—	3,99	0,00	— — — 3:13 ou 1:7
9-120 x 6-121	142	127	89,44	—	—	—	—	6,25	0,49	— — — — 1:7
1-121 x 1-120	201	180	89,55	—	—	—	—	—	0,71	— — — — 1:7