Análise da variação qualitativa em amostras pequenas

F. G. BRIEGER
da Cadeira de Genética da
Escola Superior de Agricultura
"Luiz de Queiroz"

ÍNDICE

I — Introdução 36
II — Amostras com a frequência esperada p menor do que 0,1 ... 41
III — Amostras com frequência esperada maior do que 0,1 e número total menor do que 50 43
VI — Testes da homogeneidade de distribuições internas 44
Resumo 46
Abstract 48
Literatura 50

Apresentado para publicação em 4-11-1947.

I — INTRODUÇAO

Discuti num trabalho recente (BRIEGER, 1947) os métodos estatísticos que devemos empregar para determinar qual o número mínimo necessário em amostras para garantir o aparecimento de determinados tipos. O objeto da presente publicação tem por fim explicar quais os processos de análise estatística que podem ser aplicados quando dispmos, por qualquer razão, apenas de amostras muito pequenas.

A análise da variação qualitativa pode ser executada por dois processos . o delta teste, baseado na distribuição de Gauss, e o X2-teste, baseado numa forma derivada das distribuições de Pearson (BRIEGER, 1945).

I. No primeiro caso formamos um desvio relativo, divid'ndo a diferença entre a frequência observada e a frequência esperada pelo seu êrro standard. Uma vez que êste último é calculado por uma fórmula geral, e não derivado dos valores observados, podemos atribuir-lhe um número infinito como gráu de liberdade. A fórmula é a seguinte :

delta (desvio relativo)

$$\text{em números} = \frac{f\ (\text{obs}) - p.N}{\sqrt{p(1-p)\ N}} \quad \frac{nf\ (1) = 1}{nf(2) = \text{inf.}} \quad \ldots \quad 1a$$

$$\text{em percentagem} = \frac{f\,(\text{obs.}\%) - p\%}{\sqrt{\dfrac{p\%\,(100 - p\%)}{N}}} \quad \frac{nf(1) = 1}{nf(2) = \text{inf.}} \quad \ldots \quad 1b$$

Êste teste pode evidentemente ser aplicado apenas na variação alternativa, onde p representa a frequência ideal do tipo ou da classe esperada, enquanto que (1-p) representa a frequência do seu não-aparecimento.

Na análise da variação qualitativa múltipla, com numerosos tipos ou classes, temos que calcular um êrro standard e um desvio relativo para cada classe.

Além disso o teste apenas pode ser aplicado quando p é um valor entre 0,1 e 0,9. Quando a frequência esperada for menor do que 0,1 a distribuição do binômio torna-se tão assimétrica que podemos substituí-la pela série de **Poisson**. Também neste caso podíamos calcular o desvio relativo, lembrando que agora a frequência esperada é representada pela média da série de Poisson m e que o êrro standard de uma série de Poisson é igual a raiz quadrada da média.

$$\text{desvio relativo} = \frac{f(obs) - m}{\sqrt{\bar{m}}} \quad . \; . \; . \quad 1(c)$$

Porém êste têrmo não tem utilidade na análise estatística, porque os seus limites de acaso são diferentes para cada série de Poisson, sendo pràticamente impossível dar táboas completas para tôdas as séries de Poisson.

II — O X^2-teste pode ser aplicado em qualquer caso da variação qualitativa tanto alternativa ou múltipla, calculando-se para cada classe o quociente do quadrado da diferença entre a frequência observada e esperada, dividido pela frequência esperada:

em números:

$$X^2 = \frac{\{f(obs.) - f(esp.\}^2}{f\ (esp.)} \quad . \; . \; . \quad 2a$$

em percentagem:

$$X^2 = \frac{\{f(obs.\ \%) - f(esp.\ \%)\}^2 . N}{f(esp.\ \%)\ 100} \quad . \; . \; . \quad 2b$$

Designando as probabilidades das diversas classes com as letras p1, p2, etc., devemos distinguir dois casos neste X^2-teste.

II — a) : Se o valor de p1 referente a uma das classes, for muito grande, isto é, não muito diferente de um, podemos desprezar o respectivo valor de X^2, por ser em geral muito pequeno. Analisamos individualmente os valores de X^2 de cada uma das classes restantes, cada uma com 1 grau de liberdade, podendo também somar os valores de X^2 individuais. Como limite da aproximação a um, podemos aceitar um valor de p maior do que 0,9, e como valor suficientemente pequeno para cada uma das demais classes podemos aceitar valores menores do que 0,1.

II b) : Se de outro lado os valores de p são maiores do que 0,1 e menores do que 0,9, não será mais justificada uma análise simples dos quocientes X^2 individuais, mas será necessário estudar o X^2 total, isto é, a soma de todos os valores individuais X^2 total, isto é, a soma de todos os valores individuais X^2, com grau de liberdade igual a seu número menos um.

Devemos ainda lembrar uma limitação do X^2-teste; êle pode ser aplicado sòmente quando a frequência esperada em números for igual a cinco pelo menos, ou maior.

Os dois testes, delta-teste e X^2-teste, não representam testes diferentes. O valor de X^2-teste para a variação alternativa e para valores de p entre 0,1 e 0,9, é igual ao quadrado de delta, calculado para os mesmos dados. Esta relação algébrica é fácil de demonstrar.

Supomos que temos duas classes com a e b indivíduos, sendo o total N igual a sua soma, e que as frequências ideais são iguais a pa e pb de modo que (pa + pb) é igual a um, e teremos :

$$X^2 = \frac{(a - Np_a)^2}{Np_a} + \frac{(b - Np_b)^2}{Np_b} \quad . \; . \; . \quad 3$$

Evidentemente os dois desvios têm um valor igual exceto o sinal que é oposto, como se vê pela dedução seguinte :

$$\begin{array}{lcl} a + b = N & \text{ou} & a = N - b \\ p_a + p_b = 1 & \text{ou} & p_a = 1 - p_b \\ \hline \end{array}$$

$$\begin{aligned} \text{temos : } a - Np_a &= (N-b) - N(1-p_b) \\ &= Np_b - b \\ &= -(b-Np_b) \end{aligned}$$

Agora podemos simplificar a equação (3) :

$$\begin{aligned} X^2 &= (a-Np_a)^2 . \left(\frac{1}{Np_a} + \frac{1}{N(1-p_a)} \right) \\ &= (a-Np_a)^2 . \frac{(1-p_a) + p_a}{p_a . (1-p_a) N} \\ &= \frac{(a-Np_a)^2}{p_a (1-p_a) N} = (\text{delta})^2 \quad . \; . \; . \quad 4 \end{aligned}$$

Assim é apenas uma simples questão de conveniência se usamos na variação alternativa o delta teste com os limites bilaterais da distribuição de Gauss ou o X^2-teste com os limites unilaterais da distribuição de Pearson com nf1=1; nf2 = infinito. Os limites dêstes dois testes são idênticos.

As distribuições de X^2 podem ser transformadas em distribuições de Pearson, dividindo o valor X^2 pelo grau de liberdade nf1, que é no nosso caso igual a um, e extraindo a raiz quadrada. A distribuição de Pearson com $n1=1$; $n2=$ infinito é igual à metade da distribuição de Gauss de modo que os limites unilaterais superiores desta distribuição de Pearson são iguais aos limites bilaterais superiores da distribuição de Gauss.

Esta derivação dá ainda a explicação matemática da razão pela qual os valores individuais de X^2 na variação alternativa e com valores de p entre 0,1 e 0,9 não têm interêsse, e que devemos dar maior importância ao X^2 total.

Os valores de X^2 individual são quocientes do mesmo têrmo, do quadrado do desvio, dividido ou por p.N ou por (1-p)N. Quando $p = (1-p) = 0,5$ os dois valores de X^2 são idênticos e quando p for diferente de (1-p) um dos valores de X^2 tem que ser menor do que o outro, tratando-se de uma simples consequência algébrica, e não de qualquer resultado estatístico. Na variação múltipla, com três ou mais classes, a situação é diferente e aqui tanto os valores dos X^2 individuais como o do X^2 total tem interêsse estatístico.

Pelo exposto ficou claro que as distribuições do X^2, das quais são calculados os limites de precisão na execução do teste, enumerados nas táboas em uso geral (veja táboa III), são contínuas, e temos que demonstrar quando êstes limites podem ser aplicados, pois as frequências na variação qualitativa não são contínuas, mas ao contrário valores descontínuos. Podemos sempre obter 1,2,3.. indivíduos de um determinado tipo, mas não meio indivíduo ou qualquer outra fração. Assim temos antes de mais nada resolver qual o tipo de distribuição descontínua seguida pelas frequências observadas e quando podemos aplicar uma aproximação a uma distribuição contínua.

Limitaremos a discussão aos casos da variação alternativa, sendo fácil de derivar depois a solução para os casos mais complexos da variação múltipla.

A variação da frequência p pode ser calculada pela expansão do binômio.

$$(p + q)^N \quad \text{onde } p + q = 1 \quad . \quad . \quad . \quad . \quad (5)$$

Êste têrmo binominal se aproxima a dois extremos:

Quando p é bem pequeno, de modo que q fica aproximadamente igual a 1, a série binominal se transforma numa série de Poisson, com frequência média $\overline{m}$ igual a (p. N) e êrro standard igual à raiz quadrada desta média $\overline{m}$.

Se de outro lado N torna-se muito grande, aproximando-se ao infinito, a série descontínua binominal se transforma na distribuição contínua de Gauss, devendo ainda o valor de N ser tanto maior quanto mais p e q forem desiguais.

Para melhor ilustrar estas duas aproximações, utilizaremos os dados contidos nos Quadros 1 a 3, nos quais encontramos as frequências de distribuições de Poisson na segunda coluna à esquerda e os valores de X^2-total, baseados na distribuição modificada de Gauss, na segunda coluna da direita.

Nas duas colunas centrais encontramos as frequências de duas distribuições binominais com $p=1/20$ ou 0,05 e com $p=1/10$ ou 0,10. Foram escolhidas frequências esperadas iguais a 5 (Quadro 1), 10 (Quadro 2) e 16 (Quadro 3). As linhas horizontais indicam a posição dos limites de 5%, 1% e 0,1% de precisão, isto é, os pontos que cortam uma área em ambas as extremidades das distribuições de tal modo que a soma das duas áreas excluidas em ambas as extremidades é igual a 5%, 1% e 0,1% da área total da distribuição (compare BRIEGER, 1945 : Fig. 4).

A posição dos limites do binômio $(1|20 + 19|20)N$ (terceira coluna) se aproxima muito aos limites da séries de Poisson (segunda coluna à esquerda) e aquela dos limites do binômio $(1|10 + 9|10)N$ se aproxima dos limites de X^2-total, para valores de $N = 50$, 100 e 150.

Fica assim demonstrado que é correta a conclusão de BRIEGER (1946), MOLINA (1943), e outros que os limites entre as séries de Poisson e as séries binominais podem ser fixados aproximadamente num valor de p igual a 0,1 e que a série binominal se torna razoàvelmente contínua para expoentes maiores do que 50 (BRIEGER, 1945 : Fig. 1)

As distribuições de Poisson são por si mesmas assimétricas, mas para que possamos aplicar como aproximação uma distribuição modificada de Gauss devemos exigir uma simetria da distribução além da continuidade.

O X^2-total do Quadro 1 mostra uma forte assimetria devida à pressão do limite absoluto zero, pois o valor menor possível da frequência observada é evidentemente zero, não podendo haver número de indivíduos com sinal negativo. Assim apesar do que a variação da frequência observada possa subir até qualquer valor, além de 0,1% limites de probabilidade, o valor mais extremo na outra extremidade é o zero, justamente localizado além do limite de 5% de probabilidade, e dentro do limite de 1%.

O X^2-total do Quadro 2, que se refere a uma frequência esperada de 10 já não acusa mais uma assimetria causada pela pressão do limite absoluto zero.

Para melhor ilustrar estas relações calculei os valores de X^2

individual que constam nos Quadros 4 a 6, para os valores da frequência esperada igual a 5,10 e 15.

Com f(esp) igual a 5, temos uma variação ainda bem assimétrica, podendo atingir apenas os desvios negativos do 1% limite de precisão, limite êste porém mutas vezes satisfatório. Com f(esp) igual a 10 atingimos justamente, no lado de desvios negativos, o 0,1% limite de precisão e com f(esp igual a 15 estamos pràticamente livres do efeito da pressão do limite absoluto zero e da assimetria por êste causada.

Êstes Quadros dão assim a razão pela qual alguns autores admitem o emprêgo do X^2-teste para valores mínimos da frequência esperada de pelo menos 5 (BRIEGER) ou de pelo menos 10 (SNEDECOR e outros).

Em resumo podemos formular a seguinte conclusão :

O X^2-teste, e do mesmo modo o delta-teste na análise da variação qualitativa, se baseiam na hipótese que as frequências observadas variam em volta do valor da frequência esperada e de acôrdo com uma distribuição modificada de Gauss. Uma aproximação à distribuição de Gauss, pode ser aplicada a priori apenas a) quando a frequência esperada em fração é maior do que 0,1. b) quando a frequência esperada em números é maior do que 5. c) quando o número total que corresponde ao expoente do têrmo binominal, é maior do que 50.

Êstes três valores críticos podemos reunir numa equação :

$$\begin{aligned} f(obs) &= N - p \\ 5 &= 50 - 0{,}1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 6 \end{aligned}$$

Assim surgem dois problemas que discutiremos nos capítulos seguintes : a) como analisar amostras onde a frequência esperada, em frações, for menor do que 0,1.

b) como analisar amostras com p maior do que 0,1, mas onde o número total for menor do que 50, de modo que a frequência esperada em números é menor do que 5.

II — Amostras com a frequência esperada p menor do que 0,1

A relação entre o valor X^2 individual e a distribuição de Poisson é fàcilmente demonstrada.

Lembrando-se que a frequência esperada numa série de Poisson é igual à média da série, m, e que o êrro standard é igual à raiz quadrada desta média, podemos formular o desvio relativo :

$$\text{desvio relativo (Poisson)} = \frac{f(\text{obs.}) - f(\text{esp})}{\text{erro}} = \frac{f(\text{obs}) - \overline{m}}{\sqrt{\overline{m}}}$$

$$(\text{desvio relativo})^2 = \frac{(f(\text{obs}) - \overline{m})^2}{\overline{m}} = \frac{\{f(\text{obs}) - f(\text{esp})\}^2}{f(\text{esp})} = X^2 \quad .\ .\ 7$$

Assim torna-se evidente que podemos aplicar o têrmo X^2, mas devíamos empregar no teste de significância limites de precisão calculados de uma série de Poisson, em vez da distribuição modificada de Gauss.

A diferença entre os limites para a frequência observada calculada nas duas **hipóteses consta no quadro 7.**

Se nós compararmos a diferença entre o limite de 0,1% precisão da distribuição de Poisson e do X^2-teste, para diferentes valores de f(esp) notamos que as diferenças em números variam pouco. Mas não devemos esquecer que uma divergência de dois indivíduos (última linha) representa uma inexatidão de 4% apenas em relação a frequência esperada de 50, quando uma divergência de 3 indivíduos (na primeira linha) representa uma inexatidão de 300% em relação à frequência esperada de 1.

Como inexatidão ainda tolerável podemos aceitar os valores contidos na linha que corresponde a uma frequência esperada de pelo menos 5 ou melhor ainda de pelo menos 10.

A explicação teórica da possibilidade de empregar na análise duma série de Poisson os limites da distribuição de Gauss consiste no seguinte : É um fato matemático bem conhecido que mesmo as distribuições de Poisson, como qualquer distribuição binominal assimétrica, tendem para a continuidade e simetria quando o expoente N torna-se bastante grande. No nosso caso podemos calcular fàcilmente êste valor N pela equação

$$f(\text{obs.}) = N.p$$

e substituindo por f(obs) o valor 10 e para p o valor 0,1, o qual nós aceitamos como limites entre as distribuições de Poisson e do Binômio :

$$10 = N \cdot 0{,}1$$
$$N = 100$$

Se o valor de p for menor ainda, podemos até aceitar o valor 5 como limite inferior para a frequência esperada em números : com precisão de 5% ou 0,05 :

$$5 = N.0,05$$

$$N = 100$$

Os valores exatos dos limites de f(obs), calculados na base da distribuição de Poisson, estão contidos na táboa 1. Incluimos nesta táboa os limites para as frequências esperadas de p igual a 0,001 até 15.

A aplicação da táboa é fàcilmente explicada :

Esperamos por exemplo um acontecimento com uma frequência de p = 0,005 e estudámos 60 indivíduos ao todo. Assim teremos uma frequência esperada de 0,005 vezes 60 ou 0,3 indivíduos. Da sétima linha da táboa podemos deduzir que em 95 casos entre 100 obteremos qualquer valor entre 0 e 2 inclusive, isto é : 0 ou 1 ou 2 indivíduos do tipo esperado. Apenas 5 vezes em 100 repetições (limite de 5% precisão) podemos obter 3 ou mais indivíduos, e uma vez em 1000 repetições (limite de 0,1% de precisão) obteremos 4 indivíduos ou mais do tipo.

Invertendo a pergunta, e obtendo num experimento 4 ou mais indivíduos, podemos concluir que êste desvio do valor esperado 0,3 é causado por agentes especiais ou que a nossa hipótese básica era errada e que o valor esperado na realidade não era 0,005, mas que era outro e maior. Num outro trabalho recente (BRIEGER, 1947) já foi dada a solução dêste último problema, isto é, como num caso dêste achar o valor da frequência esperada mais provável.

A táboa I vai apenas até valores da frequência esperada igual a 15, pois como já foi explicado para valores maiores podemos sempre aplicar o X^2-teste na sua forma comum.

III — Amostras com frequência esperada maior do que 0,1 e número total menor do que 50.

Já explicàmos acima na equação (4) que o X^2-teste para duas classes é idêntico ao delta-teste, mas que não é justificável a priori, substituir nos cálculos dos limites de precisão uma distribuição de Gauss para binômios com expoentes relativamente pequenos.

A dimensão das divergências entre os limites verdadeiros de distribuições binominais e os limites de X^2 calculados com aproximação de Gauss ficam ilustradas pelos dados contidos nos Quadros 8 e 9.

Os limites são pràticamente idênticos para o binômio $\left(\frac{1}{2} + \frac{1}{2}\right)^N$ exceto algumas pequenas divergências sem muita importância. Assim podemos aplicar o teste na forma comum e sem respeitar o limite de pelo menos 5 indivíduos para a frequência esperada. Apenas não devemos esquecer que em amostras menores, ficamos com uma variação assimétrica e pràticamente unilateral. Devido à pressão do limite zero, não podemos obter de forma alguma desvios negativos significantes quando o número esperado é menor do que 4.

A situação é sòmente pouco diferente para uma distribuição levemente assimétrica como o binômio $\left(\frac{1}{4} + \frac{3}{4}\right)^N$ (Quadro 9).

Mesmo quando descemos até frequências esperadas iguais a 1,5 ou 2 ou a números totais de indivíduos iguais a 6 ou 8, encontramos apenas uma diferença de no máximo um indivíduo nos limites, calculados pelo dois processos.

Apesar de que podemos assim pràticamente aplicar o X^2-teste na variação alternativa com duas classes, mesmo para as menores amostras que podem acontecer em experimentos, calculei para êstes dois binômios os limites exatos de precisão e que constam na táboa II. A sua aplicação não precisa de nenhuma explicação detalhada.

IV — Testes da homogeneidade de distribuições internas

Além dos casos da variação qualitativa, o X^2-teste é amplamente aplicado na análise estatística quando queremos comparar as frequências em distribuições observadas com aquelas em distribuições teóricas adequadas, num teste que chamamos de teste da distribuição inteira.

Como exemplo escolhi a análise das médias do diâmetro de frutas em 217 árvores de laranja Baianinha, obtidas pelo enxêrto de borbulhas da mesma árvore matriz em numerosos cavalos de laranja azêda (BRIEGER, MOREIRA, 1944). Podemos esperar que a variação dêstes diâmetros acompanhasse a distribuição de Gauss, e damos no Quadro 10 as frequências para intervalos tanto em frações do êrro standard como em milímetros.

Na terceira coluna constam as frequências esperadas de acôrdo com a distribuição de Gauss enquanto na quarta constam as frequências observadas experimentalmente.

Aplicando agora o X^2-teste encontramos uma situação classificada na introdução como o caso IIb. Para as oito classes centrais com frequências esperadas maiores podemos calcular o X^2 sem hesitação. Para as demais classes devemos antes verificar se os valores de frequência esperada são maiores do que 5. Verificamos que êste não é o caso para cinco classes em cada extremidade, de modo que temos que acumular as suas frequências das extremidades para dentro até atingir o número mínimo necessário. Somando as cinco classes chegámos a um valor da frequência esperada igual a 4,92 ou quase igual a 5. Agora podemos executar o X^2-teste. Comparando o total dos 10 valores de X^2, com 9 graus de liberdade, com os limites que constam na tábua III, verificamos que êle é altamente significante e comparando os valores de X^2 individuais, com um grau de liberdade cada um, constatamos que apenas os valores do X^2 referentes às classes extremas não significantes. Assim sabemos que a anormalidade da distribuição observada consiste na alta frequência de valores nas classes com desvios mais extremos, sejam das maiores do que + 2,00 vezes o êrro sandard, ou menores do que — 2,00 vezes o êrro standard.

O emprêgo da tábua I nos permite ir além dêste resultado do X^2-teste. Na terceira classe, correspondendo ao intervalo de —3,5 até —4,5 esperamos 0,04 valores e temos nenhum, na classe seguinte esperamos 0,24 e temos um, sendo ambas as frequências observadas dentro dos limites de precisão dado na Tábua I.

Na classe seguinte de —2,5 até —3,0 esperamos 1,03 valores e os limites que nos indicam quais as frequências que não devem ser mais encontradas são, de acôrdo com a Tábua I.

5% precisão : 4

1% precisão : 5

0,1% precisão : 7

Assim o valor da frequência observada 7 é fora do 0,1% de precisão e portanto altamente significante.

Pelo mesmo processo podemos também provar que a anomalia da outra extremidade de distribuição é devido às frequências no intervalo de +2,5 até +3,0 vezes o êrro standard.

Assim empregando, os limites contidos na nova Tábua I, depois da aplicação do X^2-teste que é um teste muito sumário para as extremidades da distribuição, podemos obter resultados mais exatos e detalhados.

RESUMO

Na aplicação do X^2-teste devemos distinguir dois casos :

A) Quando as classes de variáveis são caracterizadas por frequências esperadas entre p = 0,1 e p = 0,9, podemos aplicar o X^2-teste pràticamente sem restrição. É talvez aconselhável, mas não absolutamente necessário limitar o teste aos casos nos quais a frequência esperada é pelo menos igual a 5. e porisso incluimos na Táboa II os limites da variação de dois binômios $\left(\frac{1}{2}+\frac{1}{2}\right)^N$ e $\left(\frac{1}{4}+\frac{3}{4}\right)^N$ para valores pequenos de N e nos três limites convencionais de precisão : 5%, 1% e 0,1%.

Neste caso, os valores dos X^2-individuais têm apenas valor limitado e devemos sempre tomar em consideração principalmente o X^2 total.

O valor para cada X^2 individual pode ser calculado por qualquer das expressõe seguintes :

$$X^2 = \frac{(f\ obs - f\ esp)^2}{f\ esp.}$$

$$= \frac{(f\ obs - pN)^2}{pN}$$

$$= \frac{(f\ obs\% - p\%)^2 N}{p\%.\ 100}$$

O delta-teste dá o mesmo resultado estatístico como o X^2-teste com duas classes, sendo o valor do X2-total algébricamente igual ao quadrado do valor de delta. Assim pode ser mais fácil às vezes calcular o X^2 total como quadrado do desvio relativo da variação alternativa :

$$X^2 = \frac{(f\ obs - pN)^2}{p.(1-p).N}$$

$$= \frac{(f\ obs\% - p\%)^2.\ N}{p\%.\ (100 - p\%)}$$

B) Quando há classes com frequência esperada menor do que $p = 0,1$, podemos analisar os seus valores individuais de X^2, e desprezar o valor X^2 para as classes com p maior do que 0,9. O X^2-teste, todavia, pode agora ser aplicado apenas, quando a frequência esperada for pelo menos igual ou maior do que 5 ou melhor ainda, igual ou maior do que 10.

Quando a frequência esperada for menor do que 5, a variação das frequências observadas segue uma distribuição de Poisson, não sendo possível a sua substituição pela aproximação Gausseana.

A táboa I dá os limites da variação da série de Poisson para frequências esperadas (em números) desde 0,001 até 15.

A vantagem do emprêgo da nova táboa I para a comparação, classe por classe, entre distribuições esperadas e observadas é explicada num exemplo concreto. Por meio desta táboa obtemos informações muito mais detalhadas do que pelo X^2-teste devido ao fato que neste último temos que reunir as classes nas extremidades das distribuições até que a frequência esperada atinja pelo menos o valor 5.

Incluimos como complemento uma táboa dos limites X^2, para 1 até 30 graus de liberdade, tirada de um outro trabalho recente (BRIEGER, 1946). Para valores maiores de graus de liberdade, podemos calcular os limites por dois processos :

Podemos usar uma solução dada por Fischer :

$$\sqrt{2\,X^2} - \sqrt{2\,nf} = \text{delta}$$

Devem ser aplicados os limites unilaterais da distribuição de Gauss : 5%:1, 64; 1%:2,32; 0,1%:3,09 :

Uma outra solução podemos obter segundo BRIEGER (1946) calculando o valor :

$$\sqrt{\frac{X^2}{nf}} = \text{teta}$$

e procurando os limites nas táboas para limites unilaterais de distribuições de Fischer, com $n1 = nf(X^2)$; $n2 = inf$; (BRIEGER, 1946).

ABSTRACT

The main object of the paper is a revision of the methods for the statistical analysis of qualitative variation in small samples. In general, this analysis is carried out by means of two tests, which are mathematically different, but give identical statistical results: the analysis of relative deviates, i.e. of te quotients between the deviate of the observed frequency, with regerds to the expected frequency divided by its standard error, or an analysis using the X^2-test.

We must distinguish, in our discussion, two cases, which are quite different. The basic distribuition of all cases of qualitative variation is the binomial. If we expect any qualitative result to occur with probability p ,its non-occurence having probability q equal to (1-p), the expectancy to have, 0, 1, 2... m cases or individuals of the expected type in N trials may be calculated by expanding the binomial.

$$(p + q)^N$$

Such a binomial may be substituited by other distribuitions in two special cases:

a) If p is very small and thus q is aproaching the value one, we may substitute the binomial by a Poisson series.

b) If the exponent N becomes very large, the binomial is approaching a continous distribution, i.e. the normal or Gaussean distribution.

Only in the second case, the application of the X^2-test is really justified, and in all other cases, we are dealing only with approximations. Indvidual values of X^2- should follow a modified distribution of Pearson, with $n1 = 1$; $n2 = \text{inf.}$ Since this distributon corresponds exactly to one half of the Gaussean dstribution, it follows that the bilateral limits of the latter are equal to the unilateral limits of the former. These points have been explained fully elsewhere (BRIEGER, 1945, 1946).

We have now to decide which value of p may be accepted as a satisfactory limit between a Poisson and a binomial series. Quadros 1-3, show that the conventional limit of $p=0,1$ is fully justified, from a practical point of view. In these tables we find in the second column from the left the frequencies of a Poisson series, and in the second column from the night the values of X^2 based, as explained, on a modified Gaussean

distribution. The two columns in the centre correspond to two binomials and it is evident that the first with $p=0,05$ has its limits of precision almost at the same level as the Poisson series, while the other with $p=0,1$ agrees fairly well with the limits of the X^2 series. Thus is seems justified to treat separately the cases with expected frequencies of p equal or smaller that 0,1 and those with p larger than 0,1.

A) When the different classes, wich may be two (alternative variability) or more (multiple variability, have all expected frequences of p between 0,1 and 0,9, we may use practically the X^2 test with out any restriction. Quadros 8 and 9 show that the limits calculated for two binomials are practically identical with those of the X^2 total.

Nevertheless a special table is given (table 11) for the limits of binomials with p equal 0,5 and 0,25 and expected class frequencies of less than 10.

One must not forget that in these cases the individual values of X^2 for each class are of less importance than their sum, the X^2 total.

The value of X^2 for each class may be calculated either with the general formula, using actual numbers or with a modified formula using percentages :

$$X^2 = \frac{(\text{f obs} - \text{f esp})^2}{\text{f esp}} = \frac{(\text{f esp} - Np)^2}{Np}$$

$$= \frac{(\text{f obs\%} - p\%)^2 . N}{p\% . 100}$$

In the case of alternative variability, we may calculate directly the value of the X^2 total, by squaring the relative deviate :

$$X^2 \text{ total} = \frac{(\text{f obs} - \text{f esp})^2}{p.(1-p) . N} = \frac{(\text{f obs} - Np)^2}{p. (1-p).N}$$

$$= \frac{(\text{f obs\%} - p\%)^2 .N}{p\% . (100-p\%)}$$

B) If we have one or more classes with expected frequencies equal os smaller than 0,1 we have to deal with a Poisson series. As shown in Quadro 7 the agreement between the limits of the Poisson series and the X^2-test for one classe (simple X^2) is only satisfactory when the expected frequency is larger than 10 and tolerable when it is between 5 and 10. If the expected number should be smaller still, we cannot use anymore the X^2-test, but should use the values given in table I, calculated for Poisson series witr expected frequencies (in numbers) from 1 to 15.

Very frequently the X^2-test is used for comparing in detail observed and expected distributions, a test called sometimes "homogeneity test". Since generally the frequencies in the marginal classes are less than five, we have to accumulate by summing the frequencies from the more extreme classes towards the center, untill all accumulated and remaining values are at least equal to five. The statistical information, lost in this accumulating process, may be recovered when comparing the individual class frequencies with the limiting values in table 1. As ilustration, a concrete exemple is discussed. (Quadro 10).

The formulas and tables of this paper have been tried out first during sometime and, having been found of considerable value in the execution of statistical analysis, are now published. In order to permit a more general use, a table of ordinary limits for the X^2-test is included, taken from a recent paper (BRIEGER, 1946).

LITERATURA

BRIEGER, F. G. — 1945 — As distribuições do Acaso — Anais da Escola, Vol: II, pp. 321-391.

BRIEGER, F. G. — 1946 — Limites Unilaterais e Bilaterais na amostra estatística — Bragantia — Vol : VI, pp. 479-545.

BRIEGER, F. G. — 1947 — A determinação dos números de indivíduos mínimos necessários na Experimentação Genética — Anais da Escola. Vol : IV (em impressão).

MOLINA, E. C. — 1943 — Poisson's Exponential Binominal Limits — Van Nostrand Co. — New York.

QUADRO 1

f. (obs)	Frequência: Poisson $\bar{m} = 5$	Frequência: Binômio $\left(\frac{1}{20}+\frac{19}{20}\right)^{100}$ N = 100, N.p = 5	Frequência: Binômio $\left(\frac{1}{10}+\frac{9}{10}\right)^{50}$ N = 50, N.p = 5	X^2— total N=50 Np=5	Lim. de Prec.
0	0,00 674	0,00 592	0,00 516	5,36	5%
1	0,03 369	0,03 115	0,02 862	3,56	
—	—	—	—	—	
—	—	—	—	—	
—	—	—	—	—	
9	0,03 627	0,03 490	0,03 322	3,65	5%
10	0,01 813	0,01 672	0,01 518	5,56	1%
11	0,00 824	0,00 720	0,00 613	8,00	
12	0,00 343	0,00 281	0,00 221	10,89	0,1%
13	0,00 132	0,00 105	0,00 072	14,22	
14	0,00 047	0,00 033	0,00 021	—	
15	0,00 016	0,00 010	0,00 006	—	
16	0,00 005	0,00 003	0,00 001	—	
17	0,00 002	0,00 001	—	—	

QUADRO 2

f. (obs)	Frequência: Poisson $\bar{m} = 10$	Frequência: Binômio $\left(\frac{1}{20}+\frac{19}{20}\right)^{200}$ N = 200, Np = 10	Frequência: Binômio $\left(\frac{1}{10}+\frac{9}{10}\right)^{100}$ N = 100, Np = 10	X^2— total N = 100, Np = 10	Lim. de Prec.
0	0,00 005	0,00 004	0,00 003	11,11	0,1%
1	0,00 045	0,00 037	0,00 030	9,00	
2	0,00 227	0,00 193	0,00 162	7,10	1%
3	0,00 757	0,00 671	0,00 589	5,44	
4	0,00 892	0,01 740	0,01 587	4,00	5%
5	0,03 783	0,03 590	0,03 387	2,78	
—	—	—	—	—	
—	—	—	—	—	
—	—	—	—	—	
15	0,03 472	0,03 378	0,03 268	2,78	5%
16	0,02 170	0,02 056	0,01 929	4,00	
17	0,01 276	0,01 171	0,01 059	5,44	1%
18	0,00 709	0,00 627	0,00 543	7,11	
19	0,00 373	0,00 316	0,00 260	9,00	0,1%
20	0,00 187	0,00 150	0,00 117	11,11	
21	0,00 089	0,00 068	0,00 050	—	
22	0,00 040	0,00 029	0,00 020	—	
23	0,00 018	0,00 012	0,00 007	—	
24	0,00 007	0,00 005	0,00 001	—	
25	0,00 003	0,00 002	—	—	
26	0,00 001	0,00 001	—	—	

QUADRO 3

f. (obs)	Frequência: Poisson $\bar{m} = 15$	Frequência: Binômio $\left(\frac{1}{20}+\frac{19}{20}\right)^{300}$ N = 300 Np = 15	Frequência: Binômio $\left(\frac{1}{10}+\frac{9}{10}\right)^{150}$ N = 150 Np = 15	X^2— total N = 150 Np = 15	Lim. de Prec.
0	—	—	—	—	
1	—	—	—	—	
2	0,00 003	0,00 003	0,00 002	12,52	0,1%
3	0,00 017	0,00 013	0,00 010	10,67	
4	0,00 054	0,00 053	0,00 041	8,96	
5	0,00 194	0,00 164	0,00 137	7,34	1%
6	0,00 484	0,00 425	0,00 368	5,33	
7	0,01 037	0,00 939	0,00 842	4,74	5%
8	0,01 944	0,01 810	0,01 672	3,84	
—	—	—	—	—	
—	—	—	—	—	
—	—	—	—	—	
22	0,02 036	0,01 940	0,01 832	3,84	5%
23	0,01 328	0,01 234	0,01 133	4,74	
24	0,00 830	0,00 750	0,00 666	5,33	1%
25	0,00 498	0,00 436	0,00 373	7,34	
26	0,00 287	0,00 242	0,00 199	8,96	
27	0,00 160	0,00 129	0,00 101	10,67	0,1%
28	0,00 086	0,00 066	0,00 050	12,52	
29	0,00 044	0,00 033	0,00 023	—	
30	0,00 022	0,00 016	0,00 010	—	
31	0,00 011	0,00 007	0,00 004	—	
32	0,00 005	0,00 003	0,00 002	—	
33	0,00 002	0,00 001	0,00 001	—	
34	0,00 001	—	—	—	

QUADRO 4

Limites do X^2 individual (uma classe)

	f.esp) 5				
	f.(obs)		desvio	X^2	
	4	6	1	0,20	
	3	7	2	0,80	
	2	8	3	1,80	
5%	1	9	4	3,20	5%
1%	0	10	5	5,00	1%
	—	11	6	7,20	
0,1%	—	12	7	9,80	0,1%
	—	13	8	12,80	
	—	14	9	16,20	

QUADRO 5

Limites do X^2 individual (uma classe)

	f.(esp) 10				
	f.(obs)		desvio	X^2	
	9	11	1	0,10	
	8	12	2	0,40	
	7	13	3	0,90	
	6	14	4	1,60	
	5	15	5	2,50	
5%	4	16	6	3,60	5%
	3	17	7	4,90	
1%	2	18	8	6,40	1%
	1	19	9	8,10	
0,1%	0	20	10	10,00	0,1%
	—	21	11	12,10	
	—	22	12	14,40	
	—	23	13	16,90	

QUADRO 6

Limites do X^2 individual (uma classe)

	f.(esp) 15				
	f.(obs)		desvio	X^2	
	14	16	1	0,07	
	13	17	2	0,27	
	12	18	3	0,59	
	11	19	4	0,71	
	10	20	5	1,67	
	9	21	6	2,40	
5%	8	22	7	3,27	5%
	7	23	8	4,27	
	6	24	9	5,40	
1%	5	25	10	6,67	1%
	4	26	11	8,07	
0,1%	3	27	12	9,60	0,1%
	2	28	13	11,27	
	1	29	14	13,07	
	0	30	15	15,00	

QUADRO 7
Comparação dos limites de f(obs) para a distribuição de Poisson e o X^2-teste com uma classe

f.esp	Fórmula	Precisão 5%		Precisão 1%		Prec. 0,1%	
		Limites de f(esp)					
		Inf.	Sup.	Inf.	Sup.	Inf.	Sup.
1,0	Poisson	—	4	—	5	—	7
	X^2 simples	—	3	—	4	—	4
1,5	Poisson	—	5	—	6	—	8
	X^2 simples	—	4	—	5	—	6
2,0	Poisson	—	6	—	7	—	9
	X^2 simples	—	5	—	6	—	7
2,5	Poisson	—	7	—	8	—	10
	X^2 simples	—	6	—	7	—	8
3,0	Poisson	—	8	—	9	—	11
	X^2 simples	—	7	—	8	—	9
3,5	Poisson	—	9	—	10	—	12
	X^2 simples	—	8	—	9	—	10
4,0	Poisson	0	9	—	11	—	13
	X^2 simples	0	8	—	10	—	11
4,5	Poisson	0	10	—	12	—	14
	X^2 simples	0	9	—	10	—	12
5	Poisson	0	11	—	13	—	15
	X^2 simples	0	10	—	11	—	13
6	Poisson	1	12	0	14	—	17
	X^2 simples	1	11	—	13	—	15
7	Poisson	1	14	0	16	—	18
	X^2 simples	1	13	0	14	—	16
8	Poisson	2	15	1	17	0	20
	X^2 simples	2	14	0	16	—	18
9	Poisson	3	16	1	19	0	21
	X^2 simples	3	15	1	17	—	19
10	Poisson	3	18	2	20	1	23
	X^2 simples	3	17	1	19	—	21
11	Poisson	4	19	3	21	1	24
	X^2 simples	4	18	2	20	0	21
12	Poisson	5	20	3	23	1	26
	X^2 simples	5	19	2	22	0	24
15	Poisson	7	24	5	27	3	30
	X^2 simples	7	23	5	25	2	28
20	Poisson	11	30	9	33	6	37
	X^2 simples	11	29	8	32	5	35
30	Poisson	19	42	16	46	13	51
	X^2 simples	19	41	15	45	12	48
40	Poisson	27	54	24	58	20	63
	X^2 simples	27	53	23	57	19	61
50	Poisson	36	66	32	70	28	76
	X^2 simples	36	64	31	69	26	74

QUADRO 8
Comparação dos limites de f(obs) para a distribuição binominal $\left(\frac{1}{2}+\frac{1}{2}\right)^N$ e o X^2-teste com duas classes

N.	f.esp.	Fórmula	Precisão 5%		Precisão 1%		Prec. 0,1%	
			Limites de f(esp.)					
			Inf.	Sup.	Inf.	Sup.	Inf.	Sup.
5	2,5	Binômio	—	—	—	—	—	—
		X^2-Total	—	—	—	—	—	—
6	3,0	Binômio	—	6	—	—	—	—
		X^2-Total	0	6	—	—	—	—
7	3,5	Binômio	0	7	—	—	—	—
		X^2-Total	0	7	—	—	—	—
8	4,0	Binômio	0	8	0	8	—	—
		X^2-Total	1	7	0	8	—	—
9	4,5	Binômio	1	8	0	9	—	—
		X^2-Total	1	8	0	9	—	—
10	5,0	Binômio	1	9	0	10	—	—
		X^2-Total	1	9	0	10	—	—
11	5,5	Binômio	1	10	0	11	0	11
		X^2-Total	2	9	1	10	0	11
12	6,0	Binômio	2	10	1	11	0	12
		X^2-Total	2	10	1	11	0	12
14	7,0	Binômio	2	12	1	13	0	14
		X^2-Total	3	11	2	12	0	14
16	8,0	Binômio	3	13	2	14	1	15
		X^2-Total	4	12	2	14	1	15
18	9,0	Binômio	4	14	3	15	1	17
		X^2-Total	4	14	3	15	2	16
20	10,0	Binômio	5	15	3	17	2	18
		X^2-Total	5	15	4	16	2	18

QUADRO 9

Comparação dos limites de f(obs) para a distribuição binominal $\left(\frac{3}{4}+\frac{1}{4}\right)^N$ e o X^2-Teste com duas classes

N.	f.esp.	Fórmula	Precisão 5%		Precisão 1%		Prec. 0,1%	
			Limites de f(esp.)					
			Inf.	Sup.	Inf.	Sup.	Inf.	Sup.
4	1,0	Binômio	—	3	—	4	—	—
		X^2-Total	—	4	—	4	—	—
6	1,5	Binômio	—	5	—	5	—	6
		X^2-Total	—	4	—	5	—	5
8	2,0	Binômio	—	6	—	6	—	7
		X^2-Total	—	5	—	6	—	7
10	2,5	Binômio	—	6	—	7	—	8
		X^2-Total	—	6	—	7	—	8
12	3,0	Binômio	—	7	—	8	—	9
		X^2-Total	0	6	—	7	—	8
14	3,5	Binômio	0	8	—	9	—	10
		X^2-Total	0	7	—	8	—	9
16	4,0	Binômio	0	9	—	10	—	11
		X^2-Total	0	8	—	9	—	10
18	4,5	Binômio	0	9	—	11	—	12
		X^2-Total	0	9	—	10	—	11
20	5,0	Binômio	1	10	0	11	—	13
		X^2-Total	1	9	0	10	—	12
24	6	Binômio	1	11	0	13	—	15
		X^2-Total	1	11	0	12	—	13
28	7	Binômio	2	13	1	14	0	16
		X^2-Total	2	12	1	13	1	15
32	8	Binômio	3	14	1	16	0	18
		X^2-Total	3	13	1	15	0	16
36	9	Binômio	3	15	2	17	1	19
		X^2-Total	3	15	2	16	0	18
40	10	Binômio	4	17	3	18	1	21
		X^2-Total	4	16	2	18	0	20

QUADRO 10

Intervalos 0,5 σ	= 1,38 m/m	Frequências Esperadas		Frequências Observadas		X^2
— 4,0 até — 4,5	63,30 até 64,68	0,00		—		
— 3,5 até — 4,0	64,68 até 66,06	0,04		—		
— 3,0 até — 3,5	66,06 até 67,44	0,24	4,92	1	12	10,19
— 2,5 até — 3,0	67,44 até 68,82	1,06		7		
— 2,0 até — 2,5	68,82 até 70,20	3,58		4		
— 1,5 até — 2,0	70,20 até 71,58	8,55		13		2,32
— 1,0 até — 1,5	71,58 até 72,96	19,92		17		0,43
— 0,5 até — 1,0	72,96 até 74,34	32,53		29		0,38
0,0 até — 0,5	74,34 até 75,72	41,58		33		1,77
0,0	75,72					
0,0 até + 0,5	75,72 até 77,10	41,58		35		1,04
+ 0,5 até + 1,0	77,10 até 78,48	32,53		33		0,01
+ 1,0 até + 1,5	78,48 até 79,86	19,92		25		1,30
+ 1,5 até + 2,0	79,86 até 81,24	8,55		9		0,02
+ 2,0 até + 2,5	81,24 até 82,62	3,58		4		
+ 2,5 até + 3,0	82,62 até 84,00	1,06		7		
+ 3,0 até + 3,5	84,00 até 85,38	0,24	4,92	—	11	7,51
+ 3,5 até + 4,0	85,38 até 86,76	0,04		—		
+ 4,0 até + 4,5	86,76 até 88,14	0,00		—		
Total		217,00		217		24,97 nf (X^2) = 9

Limites de X^2: 5% 3, — 1% 6, — 0,1% 10, —
individual

TÁBOA I

Valores extremos da frequência observada na distribuição de Poisson que não devem mais ser encontrados

f.esp.	Precisão 5%		Precisão 1%		Precisão 0,1%	
	Limites de f(obs)					
	Inf.	Sup.	Inf.	Sup.	Inf.	Sup.
0,001	—	1	—	1	—	2
0,005	—	1	—	1	—	2
0,01	—	1	—	2	—	2
0,05	—	2	—	2	—	3
0,1	—	1	—	2	—	3
0,2	—	2	—	3	—	4
0,3	—	3	—	3	—	4
0,4	—	3	—	4	—	5
0,5	—	3	—	4	—	5
0,6	—	3	—	4	—	5
0,7	—	4	—	5	—	6
0,8	—	4	—	5	—	6
0,9	—	4	—	5	—	6
1,0	—	4	—	5	—	7
1,5	—	5	—	6	—	8
2,0	—	6	—	7	—	9
2,5	—	7	—	8	—	10
3,0	—	8	—	9	—	11
3,5	—	9	—	10	—	12
4,0	0	9	—	11	—	13
4,5	0	10	—	12	—	14
5,0	0	11	—	13	—	15
6,0	1	12	0	14	—	17
7,0	1	14	0	16	—	18
8,0	2	15	1	17	0	20
9,0	3	16	1	19	0	21
10,0	3	18	2	20	1	23
11,0	4	19	3	21	1	24
12,0	5	20	3	23	1	26
13,0	6	22	4	24	2	27
14,0	6	23	4	26	3	29
15,0	7	24	5	27	3	30

TÁBOA II

Valores extremos da frequência observada na distribuição binominal que não devem mais ser encontrados

Tot.	f.esp.	Precisão 5%		Precisão 1%		Precisão 0,1%	
		Limites de f(obs)					
		Inf.	Sup.	Inf.	Sup.	Inf.	Sup.
		Binômio $\left(\frac{1}{2}+\frac{1}{2}\right)^n$					
5	2,5	—	—	—	—	—	—
6	3,0	—	6	—	—	—	—
7	3,5	0	7	—	—	—	—
8	4,0	0	8	0	8	—	—
9	4,5	1	8	0	9	—	—
10	5,0	1	9	0	10	—	—
11	5,5	1	10	0	11	0	11
12	6,0	2	10	1	11	0	12
14	7,0	2	12	1	13	0	14
16	8,0	3	13	2	14	1	15
18	9,0	4	14	3	15	1	17
20	10,0	5	15	3	17	2	18
		Binômio $\left(\frac{3}{4}+\frac{1}{4}\right)^n$					
2	0,5	—	—	—	—	—	—
4	1,0	—	3	—	4	—	—
6	1,5	—	5	—	5	—	6
8	2,0	—	6	—	6	—	7
10	2,5	—	6	—	7	—	8
12	3,0	—	7	—	8	—	9
14	3,5	0	8	—	9	—	10
16	4,0	0	9	—	10	—	11
18	4,5	0	9	—	11	—	12
20	5,0	1	10	0	11	—	13
24	6,0	1	11	0	13	—	15
28	7,0	2	13	1	14	0	16
32	8,0	3	14	1	16	0	18
36	9,0	3	15	2	17	1	19
40	10,0	4	17	3	18	1	21

TABOA III

Limites de X^2

$n(X^2)$	0,1%	1%	5%	$n(X^2)$
1	10,83	6,66	3,84	1
2	13,82	9,21	5,99	2
3	16,27	11,35	7,82	3
4	18,47	13,28	9,49	4
5	20,52	15,09	11,07	5
6	22,46	16,81	12,59	6
7	24,32	18,48	14,07	7
8	26,13	20,09	15,51	8
9	27,88	21,67	16,92	9
10	29,59	23,21	18,31	10
11	31,26	24,73	19,68	11
12	32,91	26,22	21,03	12
13	34,53	27,69	22,36	13
14	36,12	29,14	23,69	14
15	37,70	30,58	25,00	15
16	39,25	32,00	26,30	16
17	40,79	33,41	27,59	17
18	42,31	34,81	28,87	18
19	43,82	36,19	30,14	19
20	45,32	37,57	31,41	20
21	46,80	38,93	32,67	21
22	48,27	40,29	33,92	22
23	49,73	41,64	35,17	23
24	51,18	42,98	36,42	24
25	52,62	44,31	37,65	25
26	54,05	45,64	38,89	26
27	55,48	46,96	40,11	27
28	56,89	48,28	41,34	28
29	58,30	49,59	42,56	29
30	59,70	50,89	43,77	30