

HERMANO VAZ DE ARRUDA

ENGENHEIRO - AGRÔNOMO

Chefe da Seção de Bioestatística do
Instituto Biológico - São Paulo

TRANSFORMAÇÃO ANGULAR DE DADOS
DE PORCENTAGENS, EM FACE DA
DISTRIBUIÇÃO BINOMIAL

Orientador: Prof. Frederico Pimentel Gomes

Tese apresentada à Escola Superior de
Agricultura "Luiz de Queiroz" da Universidade
de São Paulo, para obtenção do título de
Magister Scientiae (M. S.)

PIRACICABA

Estado de São Paulo

1971

Í N D I C E

	Página
1 - INTRODUÇÃO	1
2 - TEORIA DA TRANSFORMAÇÃO	2
3 - APRECIACÃO DA DISTRIBUIÇÃO DA VARIÁVEL Y , ATRAVÉS DA ANÁLISE DA VARIÂNCIA DOS DADOS TRANSFORMADOS	10
4 - RESULTADOS EXPERIMENTAIS DE DADOS DE PORCENTAGENS , P , ANALISADOS COM A TRANSFORMAÇÃO $\text{arc sen } \sqrt{P\% / 100}$ ENCONTRADOS NA LITERATURA ESPECIALIZADA	12
5 - MÉTODOS PARA DETERMINAR A HOMOGENEIDADE DAS VARIÂNCIAS, NUM DELINEAMENTO EM BLOCOS AO ACASO	13
5.1 - Teste da Correlação Múltipla	14
5.2 - Teste de F Máximo	15
6 - CONCLUSÕES	24
7 - LITERATURA	25
7.1 - Literatura Citada no Texto	25
7.2 - Literatura Citada no Quadro 9	26

A P R E S E N T A Ç Ã O

O presente trabalho foi feito com o objetivo de rever a teoria geral da transformação de dados experimentais, para a posterior análise de variância, com especial destaque para a transformação $\text{arc sen } \sqrt{P\%/100}$, para dados expressos em porcentagens. A análise de variância técnica, muito usada na análise e interpretação de resultados experimentais, requer para a sua correta aplicação que os dados satisfaçam certas exigências estatísticas, entre as quais está a homogeneidade da variância residual. Não deverá haver relação entre as médias e variâncias residuais dos diversos tratamentos e as transformações feitas, sugeridas pela teoria, procuram torná-las independentes.

As transformações mais comuns são a raiz quadrada, quando se trabalha com dados de contagens e a $\text{arc sen } \sqrt{P\%/100}$, quando com dados referentes a porcentagens. Será tratado com especial ênfase a segunda destas transformações, mostrando com dados colhidos em experimentos instalados no Instituto Biológico, que embora os dados se refiram a porcentagens, nem sempre se distribuem, segundo a distribuição binomial, na qual é baseada a transformação $\text{arc sen } \sqrt{P\%/100}$. Seria necessário fazer-se uma pesquisa sobre a homogeneidade das variâncias, em tais casos e são sugeridos três métodos diferentes, que poderão ser utilizados pelos interessados.

Para terminar, queremos deixar consignados os nossos agradecimentos aos Professores Frederico Pimentel Gomes e Izaias Rangel Nogueira, respectivamente, pela orientação e sugestões apresentadas

O a u t o r

1 - INTRODUÇÃO

É muito comum, na análise de variância de dados expressos em porcentagens, fazer-se a transformação destes dados para $\text{arc sen } \sqrt{P\%/100}$, antes de se proceder à análise. Esta transformação é feita admitindo-se que a variável P (porcentagem de plantas doentes, estacas enraizadas, etc.) tenha distribuição binomial. Neste caso, a transformação homogeneizará a variância experimental, o que é uma das exigências estatísticas para a validade dos testes de significância e dos intervalos de confiança para as médias dos tratamentos.

A variância de uma fração X/N , onde X corresponde ao número de indivíduos portadores do atributo analisado e N ao número total de indivíduos por unidade experimental, é dada por:

$$V\left(\frac{X}{N}\right) = \frac{p(1-p)}{N}$$

onde p corresponde à verdadeira proporção de indivíduos portadores do atributo na população.

A estimativa da variância é obtida através dos dados amostrais e pode ser expressa por:

$$\hat{V}\left(\frac{X}{N}\right) = \frac{\hat{p}(1-\hat{p})}{N}$$

No caso de um delineamento completamente casualizado, \hat{p} é dado pela média aritmética dos valores X/N obtidos em cada parcela do tratamento.

A variância de X/N (razão) é máxima para $p = 0,5$, diminuindo à medida que p se aproxima dos valores extremos, 0 e 1. A transformação é proposta justamente para corrigir essa heterogeneidade das variâncias, de sorte que as populações dos diversos tratamentos possam ter uma variância comum.

PEARCE (1965) afirma que, se a maioria dos dados estiverem num intervalo de 0,15 a 0,85, pouco se ganhará com a transformação, podendo-se em tais casos fazer a análise dos dados sem transformação.

Os dados de percentagens ($100 X/N$) devem ser estimados num número constante, N , de indivíduos por parcela, e sempre que possível este número deve ser igual ou superior a 20, conforme se verá no decorrer da teoria dada a seguir.

As transformações, de um modo geral, são indicadas quando houver heterogeneidade das variâncias, e a transformação adequada pode ser feita quando se conhecer a relação que existe entre as variâncias e respectivas médias dos tratamentos. Uma das finalidades de uma transformação é tornar as variâncias independentes das médias.

BARTLETT (1947) sugere substituir os valores extremos de X/N , 0 e 1, respectivamente, por

$$\frac{1}{2N} \quad e \quad 1 - \frac{1}{2N}$$

2 - TEORIA DA TRANSFORMAÇÃO

Suponha-se que uma variável casual y tenha média igual a μ e variância igual a $V(y)$. Admita-se que a variância possa ser expressa como uma função da média μ ,

$$V(y) = f(\mu).$$

O que se deseja é determinar uma função $\phi(y)$ que transforme y para uma nova variável casual.

$$Z = \phi (y) ,$$

de modo que a variância de Z seja independente de μ .

Pelo teorema de Taylor , numa vizinhança de $y = \mu$ ter-se-á aproximadamente

$$Z \approx \phi (\mu) + (y - \mu) \phi' (\mu)$$

Como a esperança de $(y - \mu)$ é igual a zero, conclui-se que

$$E (Z) \approx \phi (\mu) ,$$

e a variância de Z será:

$$V (Z) = E \left[Z - E (Z) \right]^2 \approx E \left[(y - \mu) \phi' (\mu) \right]^2 ,$$

$$V (Z) \approx \left[\phi' (\mu) \right]^2 V (y) ,$$

e sendo

$$V (y) = f (\mu) ,$$

$$V (Z) \approx \left[\phi' (\mu) \right]^2 f (\mu) , \tag{1}$$

logo

$$\left[\phi' (\mu) \right] \approx \frac{V (Z)}{f (\mu)} ,$$

$$\phi' (\mu) = \frac{\sqrt{V (Z)}}{\sqrt{f (\mu)}} ,$$

$$\phi' (\mu) d \mu = \frac{\sqrt{V (Z)}}{\sqrt{f (\mu)}} d \mu ,$$

Integrando-se e substituindo-se μ por \underline{y} , obtemos

$$\phi(y) = \sqrt{V(Z)} \int \frac{1}{\sqrt{f(y)}} dy \quad (2)$$

onde $\sqrt{V(Z)}$ deverá ser uma constante, C , independente de μ .

Será examinado a seguir o caso particular de \underline{Y} ser a variável X/N , que tem distribuição binomial.

$$V(y) = \frac{p(1-p)}{N}$$

$$\mu(y) = E(y) = p$$

A variância de \underline{Y} é, pois, uma função da média μ , dada por

$$V(y) = \frac{\mu(1-\mu)}{N}$$

De acôrdo com (2) ter-se-á:

$$\phi(y) = C \sqrt{N} \int \frac{1}{\sqrt{y(1-y)}} dy,$$

logo

$$\phi(y) = C_1 \cdot \arcsen \sqrt{y}$$

onde C_1 é constante e igual a $2 \sqrt{V(Z)N}$.

A integral proposta é facilmente calculada fazendo-se a transformação de variável

$$y = t^2, \quad dy = 2t dt,$$

$$\begin{aligned} \phi(y) &= C \sqrt{N} \int \frac{1}{\sqrt{y-y^2}} dy = C \sqrt{N} \int \frac{1}{\sqrt{t^2-t^4}} 2t dt \\ &= C \sqrt{N} \int \frac{2t dt}{\sqrt{1-t^2}} = C \sqrt{N} 2 \int \frac{dt}{\sqrt{1-t^2}} \end{aligned}$$

$$\begin{aligned}\phi(y) &= C \sqrt{N} \cdot 2 \arcsen t + C \\ &= C_1 \arcsen \sqrt{y} + C.\end{aligned}$$

A transformação adequada, ignorando-se as constantes, que não alteram a análise de variância, seria, pois,

$$\phi(y) = \arcsen \sqrt{y}$$

Voltando de novo à variância de Z , nova variável aleatória obtida através da transformação proposta, obtém-se:

$$Z = \phi(y) = \arcsen \sqrt{y} \qquad y = X/N$$

$$V(Z) \approx \left[\phi'(\mu) \right]^2 V(y)$$

$$\phi'(y) = \frac{1}{\sqrt{1-y}} \cdot \frac{1}{2\sqrt{y}}$$

$$\phi'(\mu) = \frac{1}{\sqrt{1-p}} \cdot \frac{1}{2\sqrt{y}}$$

$$\left[\phi'(\mu) \right]^2 = \frac{1}{(1-p)} \cdot \frac{1}{4p}$$

Então,

$$V(Z) \approx \left[\frac{1}{(1-p)} \cdot \frac{1}{4p} \right] p \frac{(1-p)}{N} = \frac{1}{4N}.$$

Vê-se, assim, que a variância de Z independe de μ , depende somente de N e se aproxima de

$$\frac{1}{4N},$$

quando Z é expresso em radianos.

Verifica-se que

$$V(Z) \approx \frac{821}{N}$$

quando Z é expresso em graus, desde que Z' (em graus) é dado por

$$Z' = \frac{180}{\pi} Z$$

$$V(Z)' = \frac{(180)^2}{\pi^2} \cdot V(Z)$$

$$V(Z') = \frac{(180)^2}{\pi^2} \cdot \frac{1}{4N} \approx \frac{821}{N}$$

FISHER (1950) afirma que, com a transformação, em amostras grandes, a variância de Z depende somente do número de observações nas quais as porcentagens (razões) são baseadas.

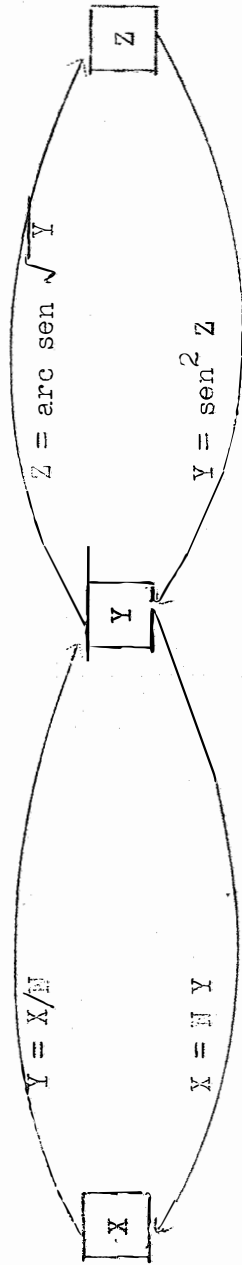
Com o fito de apreciar o efeito do tamanho da amostra, ou seja de N , sobre a estabilização das variâncias, foram calculadas as variâncias de Z , para $p = 0,20$ e $0,50$ (sendo p a probabilidade de ocorrência de um indivíduo portador do atributo na população), após a transformação para $\text{arc sen } \sqrt{X/N}$, para diferentes valores de $N: 5, 10, 20$ e 30 . Os resultados estão reunidos no quadro 1.

QUADRO 1 - Variâncias teóricas da variável $Z = \text{arc sen } \sqrt{X/N}$, em função de dois valores de p (0,2 e 0,5) e de quatro valores de N 5, 10, 20, 30, em confronto com os valores aproximados das variâncias, dados por

$$\frac{821}{N}$$

N	Variâncias Teóricas		Variâncias aproximada = 821 / N
	p = 0,2	p = 0,5	
5	282,76	253,39	164,20
10	125,80	93,85	82,10
20	47,25	43,41	41,05
30	29,70	28,16	27,37

QUADRO 2 - Esquema de transformação de variável binomial X, para X/N e $Z = \arcsen \sqrt{X/N}$



X	P (x)	Y	P (Y)	Z em graus	P (Z)
0	$(1-p)^N$	0		0	
1	$N(1-p)^{N-1}p$	1/N			
2	$\binom{N}{2}(1-p)^{N-2}p^2$	2/N			
...					
X	$\binom{N}{X}(1-p)^{N-X}p^X$	X/N	$\binom{N}{NY}p^{NY}(1-p)^{N(1-Y)}$	Z	$\binom{N}{N \text{ sen}^2 Z}p^{N \text{ sen}^2 Z}(1-p)^{N \cos^2 Z}$
...					
N	p^N	1		90°	

$$P(X=x) = P(Y = X/N) = P(Z)$$

$$\mu_Z = \sum Z P(Z)$$

$$\sigma^2(Z) = \sum (Z - \mu)^2 P(Z)$$

QUADRO 3 - Cálculo da média e variância para a variável

$$Z = \text{arc sen } \sqrt{X/N} \text{ , corresponde à } p = 0,2 \text{ e } N = 5$$

X	100 X/N	Z	P (z)	Z - μ	(Z - μ) ²
0	0	0	1 / 3125	- 68,05	4.630,80
1	20	26,56	20 / 3125	- 41,49	1.721,42
2	40	39,23	160 / 3125	- 28,82	830,59
3	60	50,77	640 / 3125	- 17,28	298,60
4	80	63,44	1280 / 3125	- 4,61	21,25
5	100	90,00	1024 / 3125	+ 21,95	481,80
			3125 / 3125		

$$\Sigma (Z - \mu) P (Z) = 0$$

$$\mu = E (Z) = \Sigma Z P (Z) = \frac{212.664,00}{3.125} = 68,05$$

$$\sigma^2 (Z) = \Sigma (Z - \mu)^2 P (Z) = \frac{88.360,80}{3.125} = 282,76$$

$$\sigma_{\text{aprox.}}^2 = \frac{821}{5} = 164,20$$

Sendo $Y = X/N$ uma variável discreta o termo $P (Y)$ corresponde a $P (X)$ conforme o esquema anexo.

Calculou-se inicialmente a nova variável Z de acordo com as tabelas de FISHER e YATES (1948) e a seguir $\mu(Z)$

$$\mu(Z) = \Sigma Z P (Z) \text{ ,}$$

e depois a variância

$$\Sigma \left[Z - E(Z) \right]^2 P(Z)$$

Verifica-se, pelos dados do quadro 1, que as variâncias de Z, passam a ser praticamente independentes dos valores de P (média populacional dos valores X/N), quando N assume valores iguais ou superiores a 20.

3 - APRECIÇÃO DA DISTRIBUIÇÃO DA VARIÁVEL Y, ATRAVÉS DA ANÁLISE DA VARIÂNCIA DOS DADOS TRANSFORMADOS

QUENOUILLE (1953) afirma que quando cada porcentagem (100 X/N) fôr o resultado de um grande número, N, de observações independentes, o valor teórico do quadrado médio residual, obtido após a transformação proposta se aproximará do valor

$$\frac{821}{N}$$

Num delineamento completamente ao acaso este fato fica logo evidenciado uma vez que a variância residual é a variação dentro dos grupos e esta é uma estimativa da variância teórica para diversos valores de $\mu = p$ e iguais todos a

$$\frac{821}{N}$$

O quadrado médio residual dos dados transformados estará em torno de

$$\frac{821}{N}$$

se a variável $Y = X/N$ tiver distribuição binomial, o que implica em se ter constante o valor de p para cada tentativa, no caso, parcela experimental de um mesmo tratamento. É claro que, na prática, em experimentos de campo, esta admissão nem sempre se verifica: a probabilidade p de ocorrer um indivíduo doente numa parcela experimental poderá depender da quantidade de inóculo e assim variar de uma parcela para outra, em função do número de plantas doentes já existentes ou mesmo da produção (número de frutos) das parcelas adjacentes.

Em alguns experimentos (como os referentes à determinação de porcentagens de pegamento dos enxertos, em função dos vários tipos de enxertia, ou porcentagens de enraizamento de estacas em função de diversos tipos de hormônios vegetais) parece ser mais razoável a admissão da constância dos valores de p .

A verificação da concordância do quadrado médio residual dos dados transformados com o teórico

$$\left(\frac{821}{N}\right),$$

admitindo-se a distribuição binomial é feita pelo teste χ^2 :

$$\chi^2 = \frac{f(Q. M.) N}{821}$$

Com f graus de liberdade, onde f é o número de graus de liberdade do resíduo ; Q. M. é o quadrado médio do resíduo ; N é o número de indivíduos por unidade experimental.

Sempre que for significativo, o valor de χ^2 observado, abandona-se a hipótese teórica, isto é, de que a distribuição seja binomial, e passa-se a usar o quadrado médio do resíduo para os testes de significância.

Assim, MATHER (1946) admite que a distribuição poderá não ser completamente casual (binomial) quando o χ^2 for significativo.

COCHRAN (1943) introduz a existência de fontes estranhas de variação associadas à binomial assim que o quadrado médio do resíduo fôr superior ao teórico.

Tudo indica que se estas fontes estranhas tiverem efeitos casuais o quadrado médio do resíduo continuará sendo casual e daí o seu uso para os testes de significância.

Um fato porém fica em dúvida, que é o referente à eficiência da transformação na homogeneização das variâncias. Seria interessante, antes de usar o quadrado médio residual para os testes de significância, verificar a homogeneidade das variâncias.

4 - RESULTADOS EXPERIMENTAIS DE DADOS DE PORCENTAGENS , p , ANALISADOS COM A TRANSFORMAÇÃO $\text{arc sen } \sqrt{P\%/100}$ ENCONTRADOS NA LITERATURA ESPECIALIZADA

Aqui será analisados somente os dados de porcentagens obtidos com a mostras constituídas de 20 ou mais indivíduos, quando a transformação proposta terá maior possibilidade de homogeneizar as variâncias, em se tratando de u ma distribuição binomial.

Êstes dados estão reunidos no quadro 2 , no qual aparece o material estudado e a referência bibliográfica, além daqueles de interesse estatístico.

5 - MÉTODOS PARA DETERMINAR A HOMOGENEIDADE DAS VARIÂNCIAS, NUM DELINEAMENTO EM BLOCOS AO ACASO

Como já foi dito, quando o valor de χ^2 fôr significativo, indicando que os dados não têm uma distribuição binomial característica, é interessante pesquisar-se a homogeneidade da variância residual. Um dos métodos consiste em decompor-se a variância residual na variância dos diversos contrastes de tratamentos de interesse na análise conforme COCHRAN (1947). Esta decomposição é viável num delineamento em blocos ao acaso não o sendo para o quadrado latino. Quando o número de graus de liberdade das variâncias residuais for o mesmo, pode-se usar o teste de Cochran, citado por BLISS (1967).

Recentemente, HAN (1969) propôs um teste para a homogeneidade de variâncias numa tabela de dupla entrada, na qual se pode enquadrar o delineamento em blocos ao acaso. O autor considera o modelo:

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \dots 1.1$$

$$i = 1, 2, \dots, t$$

$$j = 1, 2, \dots, n$$

$$n > t$$

no qual:

μ é a média geral

α_i é o efeito de i -ésima coluna

β_j é o efeito de j -ésima linha

ϵ_{ij} é o termo do erro que é $N(0; \sigma_i^2)$

A correlação entre ϵ_{ij} e ϵ_{km} é igual a zero se $j \neq m$ e igual a ρ se $j = m$. Assim, os termos são igualmente correlacionados dentro das linhas, porém, são independentes entre linhas. Quando $\rho = 0$, o modelo re-

duz-se ao usual de blocos ao acaso.

A hipótese de nulidade a ser testada é:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_t^2$$

e a alternativa é a de que pelo menos uma das variâncias seja diferente.

O autor considera dois testes para pôr em prova a hipótese H_0 , o da correlação múltipla que é exato, e o teste de F máximo, que é mais simples, porém, só aplicável para $t \geq 3$.

5.1 - Teste da Correlação Múltipla

Partindo do modelo 1.1, HAN (1969) define $\bar{y}_{.j}$, a média da linha j , e

$$Z_{ij} = y_{ij} - \bar{y}_{.j}$$

$$H_0 (\sigma_1^2 = \sigma_2^2 = \dots = \sigma_t^2)$$

é verdadeira se e somente se $\bar{y}_{.j}$ não for correlacionada com Z_{ij} , para $i = 1, 2, \dots, t$.

Desde que

$$\sum_i^t Z_{ij} = 0,$$

tem-se que Z_{1j} é uma combinação linear de $Z_{2j}, Z_{3j}, \dots, Z_{tj}$. Agora, se $\bar{y}_{.j}$ não é correlacionada com Z_{kj} , $k = 2, 3, \dots, t$, também não será correlacionada com Z_{1j} . Assim, é suficiente calcular a regressão múltipla de $\bar{y}_{.j}$ com $Z_{2j}, Z_{3j}, \dots, Z_{tj}$. O teste da hipótese da homogeneidade das variâncias corresponde ao teste de que o coeficiente de correlação múltipla, R^2 , de $\bar{y}_{.j}$ para $Z_{2j}, Z_{3j}, \dots, Z_{tj}$ seja nulo. Assim, o teste da igualdade das variâncias, no modelo 1.1 é equivalente ao teste de $R^2 = 0$

O teste é feito calculando-se inicialmente o coeficiente de correlação R , usando-se as n linhas como indivíduos e depois o valor de F ,

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - t}{t - 1},$$

que tem distribuição de F com $(t - 1)$ e $(n - t)$ graus de liberdade.

5.2 - Teste de F Máximo

Para este teste devem-se calcular as variâncias dos valores Z_{ij} para cada coluna e fazer-se o quociente da máxima para a mínima variância

$$F_{\text{máx.}} = \frac{S_{\text{máx.}}^2}{S_{\text{mín.}}^2}$$

O valor encontrado será confrontado com o valor crítico da distribuição de $F_{\text{máx.}}$. Este valor é dado aproximadamente por:

$$F_{\text{máx.}}(\alpha) = e^{w_t(\alpha)} \sqrt{\frac{2}{n-1} \left[1 - (t-1)^{-2} \right]}$$

onde $w_t(\alpha)$ corresponde ao valor de q da distribuição de Tukey, para t tratamentos e infinitos graus de liberdade, ao nível de $\alpha = 0,05\%$ de probabilidade. BLISS (1969) apresenta uma tabela que dá os valores críticos de $F_{\text{máx.}}$ em função de k , número de estimativas de variâncias e ϕ , no caso $(n - 1)$ graus de liberdade, com as quais foram elas calculadas, valores estes que se aproximam bem dos calculados pela fórmula supra.

Afim de se aplicar os métodos sugeridos escolheu-se um experimento ainda não publicado da Secção de Inseticidas, do Instituto Biológico de São Paulo.

Os dados apresentados a seguir correspondem aos ângulos Z , obtidos através da transformação

$$Z = \text{arc sen } \sqrt{P\% / 100} .$$

QUADRO 4 - Dados de porcentagens, P , (porcentagens de frutos — laranjas — atacados pela mosca dos frutos), transformados para $\text{arc sen } \sqrt{P\% / 100}$, correspondentes à um delineamento em blocos ao acaso, com cinco tratamentos e quatro repetições

Tratamentos	R e p e t i ç õ e s				Totais dos Tratamentos
Testemunha	83,71	81,28	82,08	77,34	324,41
Gardona	56,35	52,48	58,76	39,82	207,41
Naled	45,63	54,15	58,76	38,82	190,18
Antracol	12,66	18,05	23,50	14,65	68,86
Cyanox	13,31	15,34	14,89	9,63	53,17
Totais de repetições	211,66	221,30	225,26	185,61	844,03

A análise de variância foi a seguinte, incluída no quadro 5 .

QUADRO 5 - Análise de variância dos dados apresentados no quadro 4

Causa de Variação	G. L.	S. Q.	Q. M.	F
Blocos	3	188,8800		
Tratamentos	4	12.380,1520	3.095,03	191,88 **
Erros	12	193,5319	16,13	
Total	19	12.762,5639		

O coeficiente de variação foi de 9,47% .

Usando o primeiro método para teste de homogeneidade da variância residual, calcularam-se as variâncias dos seguintes contrastes de tratamentos:

C_1 : Tratados versus Testemunha,

C_2 : Entre Tratados,

C_3 : Dentro dos Tratados,

que aparecem a seguir:

QUADRO 6 - Análise de Variância

Causa de Variação	G. L.	S. Q.	Q. M.
Tratados versus Testemunha	3	18,5156	6,17
Entre G_1 e G_2	3	39,1143	13,04
Dentro de G_1	3	113,1222	37,71
Dentro de G_2	3	22,7775	7,59
Erro	12	193,5296	

Agora, ter-se-á que testar a homogeneidade das variâncias dos contrastes. Neste caso, como todas as estimativas da variância residual têm o mesmo número de graus de liberdade, usou-se o teste de Cochran, C :

$$C = \frac{\text{Maior variância}}{\text{Soma das variâncias}} = \frac{37,71}{64,51} = 0,58$$

Este valor de C, 0,58, não é significativo, confrontando com o valor crítico de tabela, ao nível $\alpha = 0,05$ de probabilidade. Assim, não rejeitamos a hipótese de nulidade e podemos admitir a homogeneidade das variâncias dos contrastes.

Pelos métodos propostos por HAN (1969) , é necessário calcular os valores de $Z_{ij} = y_{ij} - \bar{y}_{.j}$:

QUADRO 7

Z_{1j}	Z_{2j}	Z_{3j}	Z_{4j}	$\bar{y}_{.j}$
2,61	0,18	0,98	- 3,76	81,10
4,50	0,63	6,91	- 12,03	51,85
- 1,91	6,61	- 1,51	- 3,17	47,54
- 4,55	0,84	6,29	- 2,56	17,21
0,02	2,05	1,60	- 3,66	13,29
n = 5			t = 4	

As variâncias dos Z_{ij} em cada coluna são:

$$S_1^2 = 12,83$$

$$S_2^2 = 6,94$$

$$S_3^2 = 13,09$$

$$S_4^2 = 15,51$$

$$F_{\max.} = \frac{15,51}{6,94} = 2,23$$

O valor de $F_{\max.}$ igual a 2,23 é menor que o valor crítico ao nível de $\alpha = 0,05$ de probabilidade, correspondente a 11,25 . Assim, como pelos testes anteriores não rejeitamos a hipótese da homogeneidade das variâncias residuais.

O método exato da regressão múltipla é o mais trabalhoso deles, obtido através do cálculo da regressão múltipla entre as variáveis Z_{2j} , Z_{3j} e Z_{4j} com $\bar{y}_{.j}$. Estas variáveis foram substituídas por X_1 , X_2 , X_3 e Y , res

pectivamente para simplificar a notação.

$$\Sigma (X_1 - \bar{X}_1)^2 = \Sigma x_1^2 = 27,7703$$

$$\Sigma (X_2 - \bar{X}_2)^2 = \Sigma x_2^2 = 52,3861$$

$$\Sigma (X_3 - \bar{X}_3)^2 = \Sigma x_3^2 = 62,0501$$

$$\Sigma (X_1 - \bar{X}_1)(X_2 - \bar{X}_2) = \Sigma x_1 x_2 = - 26,3125$$

$$\Sigma (X_1 - \bar{X}_1)(X_3 - \bar{X}_3) = \Sigma x_1 x_3 = 13,0584$$

$$\Sigma (X_2 - \bar{X}_2)(X_3 - \bar{X}_3) = \Sigma x_2 x_3 = - 32,1201$$

$$\Sigma (X_1 - \bar{X}_1) y = \Sigma x_1 y = - 31,8575$$

$$\Sigma (X_2 - \bar{X}_2) y = \Sigma x_2 y = - 106,6744$$

$$\Sigma (X_3 - \bar{X}_3) y = \Sigma x_3 y = - 109,5467$$

Admitindo-se o modelo de regressão múltipla:

$$Y_i = \alpha + b_1 (X_{1i} - \bar{X}_1) + b_2 (X_{2i} - \bar{X}_2) + b_3 (X_{3i} - \bar{X}_3) + \epsilon_i$$

conclui-se que a estimativa de b_1 , b_2 e b_3 , obtidas pelo método dos quadrados mínimos, são dadas matricialmente por

$$\hat{\beta} = S^{-1} X' Y ,$$

onde

$$S = \begin{bmatrix} 27,7703 & - 26,3125 & 13,0584 \\ - 26,3125 & 52,3861 & - 32,1201 \\ - 13,0584 & - 32,1201 & 62,0501 \end{bmatrix}$$

$$S^{-1} = \begin{bmatrix} 0,06976 & 0,03813 & 0,00506 \\ 0,03813 & 0,04481 & 0,01724 \\ 0,00506 & 0,01724 & 0,02398 \end{bmatrix}$$

$$X' Y = \begin{bmatrix} - 31,8575 \\ - 106,6744 \\ - 109,5467 \end{bmatrix}$$

e daí

$$\hat{b}_1 = - 6,8442$$

$$\hat{b}_2 = - 8,3101$$

$$\hat{b}_3 = - 4,6272$$

$$a = \bar{y} = 42,198$$

Com os valores de \hat{a} , \hat{b}_1 , \hat{b}_2 e \hat{b}_3 obtém-se a equação de regressão:

$$\hat{Y}_i = 56,6597 - 6,8442 X_{1i} - 8,3101 X_{2i} - 4,6272 X_{3i}$$

Os valores esperados \hat{Y}_i e os valores observados Y_i são dados a seguir:

QUADRO 8 - Valores esperados \hat{Y}_i e valores observados Y_i

\hat{Y}_i	Y_i
64,68	81,10
50,59	51,85
38,64	47,54
10,49	17,21
46,27	13,29

O coeficiente de correlação múltipla R , calculado através do coeficiente de correlação linear entre os pares $(\hat{Y}_i ; Y_i)$ foi de 0,7215 .

A significância do valor de R é dada por:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - t}{t - 1} = \frac{0,4801}{0,5199} \cdot \frac{5 - 4}{4 - 1}$$

$$= 0,3078 \quad ,$$

valor este que, confrontado com o nível crítico da distribuição para 3 e 1 graus de liberdade, não é significativo. Outra vez, como ocorreu com os testes anteriores podemos admitir a homogeneidade da variância residual.

QUADRO 9 - Experimentos nos quais os dados de porcentagens, \underline{P} , foram transformados para $\text{arc sen } \sqrt{P\%/100}$, com valores de N, número de indivíduos por unidade experimental, iguais ou superiores a 20.

Material Estudado	Valor de N	Q. M. Residual	Variância Teórica 821/N	Coefficiente de Variação	χ^2	G. L.	Referência Bibliográfica
Vira-cabeça (tomateiro)	30	46,95	27,36	10,0	46,33 *	27	3
Estaquia (amoreira)	20	115,06	41,05	32,1			16
Estaquia (amoreira)	20	165,68	41,05	33,9			16
Estaquia (amoreira)	20	271,57	41,05	55,9			16
Murcha (quiabeiro)	20	110,29	41,05	25,5			N. P.
Heliobhis (milho)	25	34,44	32,84	20,9	25,17	30	8
Heliobhis (milho)	25	45,12	32,84	13,2	32,97	30	8
Heliobhis (milho)	25	25,21	32,84	10,8	18,42	30	8
Laphygn (milho)	50	57,24	16,42	17,1			6
Laphygn (milho)	50	40,50	16,42	13,9			6
Laphygn (milho)	50	33,62	16,42	13,6			6
Heliobhis (milho)	20	105,12	41,05	24,4	76,74 **	30	10
Heliobhis (milho)	20	54,08	41,05	24,2	39,48	30	10
Heliobhis (milho)	20	39,60	41,05	27,9	28,90	30	10
Heliobhis (milho)	20	54,08	41,05	29,0	39,49	30	10
Heliobhis (milho)	20	137,78	41,05	39,0	107,47 **	30	10
"Resinose" (abacaxi)	30	49,04	27,36	34,2	53,45	40	11
"Resinose" (abacaxi)	30	44,45	27,36	38,8	64,89 **	40	11

(continua ...)

N. P. = Dados não publicados

QUADRO 9 - (... continuação)

Material Estudado	Valor de N	Q. M. Residual	Variância Teórica 821/N	Coefficiente de Variação	χ^2	G. L.	Referência Bibliográfica
"Resinose" (abacaxi)	30	57,24	27,36	45,8			11
Moscas frutas (goiaba)	270	17,88	3,04	10,0			N. P.
Gorgulho (goiaba)	200	64,71	4,10	25,2			N. P.
Moscas frutas (caqui)	50	72,96	16,42	45,8			15
Moscas frutas (laranjeira)	900	22,12	0,91	20,5			14
Broca (caféero)	50	28,46	16,42	18,3			5
Broca (caféero)	50	36,05	16,42	18,6			5
Mariposa oriental (marmeleiro)	700	42,30	1,17	11,2			2
Ferrugem (laranja)	400	63,50	2,05	16,7			13
Ferrugem (laranja)	400	110,55	2,05	27,3			13
Cochonilha (laranjeira)	400	79,13	2,05	30,7			13
Cochonilha (laranjeira)	400	134,13	2,05	34,4			13
Bicho mineiro (caféero)	300	14,04	2,73	10,4			12
Bicho mineiro (caféero)	700	22,48	1,17	20,5			1
"Resinose" (abacaxi)	37	18,08	22,19	10,8	18,44	28	4
Heliothis , Diatrae (milho)	90	78,34	9,12	18,1			7
Heliothis , Diatrae (milho)	90	57,12	11,73	15,0			7

N. P. = Dados não publicados

6 - CONCLUSÕES

Analisando-se os dados apresentados no quadro 9 e a matéria apresentada nos itens anteriores, podem-se tirar as seguintes conclusões:

1 - A variância da variável transformada $z = \text{arc sen } \sqrt{X/N}$, onde X tem distribuição binomial, é praticamente independente de N, quando este valor for igual ou superior a 20.

2 - Se X tiver distribuição binomial, a variância residual na análise dos valores de Z se aproximará do valor

$$\frac{821}{N}$$

sendo N o número de indivíduos da unidade experimental.

3 - Em uma série de 35 experimentos analisados, referentes a dados de porcentagens, verificou-se após o uso da citada transformação, que em apenas 9 deles as variâncias residuais concordaram com a teórica,

$$\frac{821}{N}$$

esta admitindo ser casual e binomial a distribuição da variável X.

4 - Todos os casos em que houve a concordância com a distribuição binomial ocorreram em experimentos com as pragas Heliothis obsoleta, de milho, e Thecla brasilides (broca), de abacaxi.

5 - Os coeficientes de variação foram normais, 9 deles acima de 30%; 9 de 20% a 30%; e os demais, ou seja, 17, com coeficientes abaixo de 20%.

6 - Nos casos em que a distribuição não possa ser aceita como binomial, seria necessário verificar se a transformação homogeneizou a variância residual.

7 - LITERATURA

7.1 - Literatura Citada no Texto

- BARTLETT, M. S. - 1947 - The Use of Transformations. Biometrics 3: 93-52 .
- BLISS, C. I. - 1967 - Statistics in Biology . McGraw-Hill Book Company.
New York.
- COCHRAN, W. G. - 1943 - Analysis of Variance for Percentages Based on Inequal Numbers. J. Am. Stat. Assoc. 38: 287-301 .
- - 1949 - Some Consequences when the Assumptions for the Analysis of Variance are not Satisfied. Biometrics 3: 22-38 .
- FISHER, R. A. - 1950 - Contributions to Mathematical Statistics. John Wiley and Sons. New York .
- and YATES, F. - 1948 - Tables for Biological. Agricultural and Medical Research. Hafner. New York.
- HAN , Chien-Pai - 1969 - Testing the Homogeneity of Variances in Two-way Classification. Biometrics 25: 153-158 .
- MATHER, K. - 1946 - Statistical Analysis in Biology. Interscience Publishers Inc. New York .
- PEARCE, K. - 1965 - Biological Statistics: an Introduction . McGraw-Hill Book Company . New York.
- QUENOUILLE, M. H. - 1953 - The Design and Analysis of Experiment. Hafner Publishing Company. New York .

7.2 - Literatura Citada no Quadro 9

- CALCAGNOLO, G. e LEIDERMAN, L. - 1957 - As Pulverizações Concentradas no Combate ao "Bicho - mineiro" do Cafeeiro. O Biológico XXIII: 203-219 .
- FALANGHE, O. , DIAS NETTO, N. e MOLINARI, J. B. A. - 1959 - Combate à Mariposa Oriental das Frutas, em Marmeleiros. Arquivos do Instituto Biológico 26: 89-101 .
- FRAGA, Jr. G. G. e COSTA, A. S. - 1950 - Análise de um Experimento para Combate do "Vira-cabeça" do Tomateiro. Bragantia 10: 305-316 .
- HEINRICH, W. O. - 1958 - Experiências para Combate à Thecla brasilides , Broca do Abacaxi. Arquivos do Instituto Biológico 25: 109-119 .
- - 1960 - Experiências de Campo para Comparação do Efeito de Inseticidas Modernos no Combate à Broca do Café. Arquivos do Instituto Biológico 27: 16-29 .
- LEIDERMAN, L. e SAUER, H. F. - 1954 - Ação de Alguns Inseticidas Orgânicos Sobre a Laphygma frugiperda , Atacando Milho. Arquivos do Instituto Biológico 21: 111-119 .
- - 1954 - Observações Sobre a Susceptibilidade de Cinco Variedades de Milho do Ataque de Heliothis obsoleta e Diatraea sp. O Biológico 73-77 .
- e SAUER, H. F. - 1954 - Resultados Preliminares da Ação de Inseticidas Orgânicos no Combate à Heliothis obsoleta em Espigas de Milho. Arquivos do Instituto Biológico 21: 101-110 .
- - 1954 - Novos Resultados com Inseticidas no Combate à Lagarta das Espigas de Milho , Heliothis obsoleta. O Biológico XX: 131-137 .

- LEIDERMAN, L. - 1955 - Combate das Lagartas de Heliothis obsoletas em Espigas de Milho com Modernos Inseticidas Orgânicos. Arquivos do Instituto Biológico 22: 14-22 .
- e VASCONCELLOS, F. T. C. - 1955 - Combate à "Resinosê" do Abacaxi, com Modernos Inseticidas Orgânicos. O Biológico XXI: 97-103 .
- e CALCAGNOLO, G. - 1958 - O Combate ao "Bicho - mineiro" do Cafeeiro por Meio de Pulvilhamentos com Inseticidas Orgânicos. O Biológico XXIV: 115-129 .
- - 1960 - Ação de Modernos Acaricidas Orgânicos no Combate ao "Ácaro da Ferrugem do Citrus" e a Cochonilha "Cabeça-de-prego" , em Laranjeiras. Arquivos do Instituto Biológico 27: 16-29 .
- PUZZI, D. e ORLANDO, A. - 1957 - Ensaio de Combate às "Moscas das Frutas" por Meio de Pulverizações e Iscas Envenenadas. O Biológico XXIII: 21-25 .
- , SAMPAIO, A. S. , ORLANDO, A. et alii - 1963 - Combate às "Moscas das Frutas" em Caqui , com Pulverizações de Diversos Inseticidas sob a Forma de Cobertura. O Biológico XXIX: 263-265 .
- RÚBIA, A. C. , INFORZATO, R. e PIMENTEL GOMES, F. - 1963 - Estacas de Amoreira Tratadas com Hormônios Vegetais em Dois Sistemas de Plantio e Diferentes Coberturas do Estufim. Anais da E. S. A. "Luiz de Queiroz" 22: Universidade de São Paulo.