4 ANÁLISE DE VARIÂNCIA

Um problema que se apresenta com maior freqüência do que qualquer outro na análise estatística é o de avaliar se duas ou mais amostras diferem significativamente com relação a alguma variável.

Este tipo de problema ocorre tão freqüentemente porque os pesquisadores muitas vezes propõem experimentos para comparar dois ou mais tratamentos (amostras) entre si. Por exemplo, uma nova técnica de aplicação de vermífugo em caprino é comparada com a técnica tradicional, diferentes tipos de adubos orgânicos são avaliados na cultura do tomate, diferentes variedades de milho forrageiro são avaliadas numa determinada região, etc..

Em função disso, é necessário um método estatístico para solucionar problemas dessa natureza. Um dos métodos mais utilizados para resolver tais problemas é conhecido como análise de variância.

4.1 Análise de Variância

A análise de variância foi introduzida por Fisher e é essencialmente um processo baseado na decomposição da variação total existente entre uma série de observações, em partes que podem ser atribuídas a causas conhecidas e numa parte devida a causas desconhecidas ou não suscetíveis de controle. Como exemplo das causas conhecidas, pode-se citar o efeito de diferentes inseticidas no controle do pulgão em batata (*Solanum tuberosum* L.) cv. RADOSA, e como exemplo das causas desconhecidas, as diferenças existentes entre as plantas (parcelas), condicionando um tipo diferente de resposta a um mesmo inseticida. Os efeitos dessas causas desconhecidas, ou não controláveis, contribuem para uma porção da variação total, que é isolada na análise de variância, recebendo a denominação de **Erro** ou **Resíduo**.

A variação que contribui para o erro experimental pode ser de dois tipos:

a) Inerente à própria variabilidade do material experimental;

b) Proveniente da falta de uniformidade do ambiente em que é conduzido o experimento.

Na análise de variância, quando a variação total é decomposta, as causas conhecidas e desconhecidas representam, respectivamente, a variação entre amostragens (tratamentos) e a variação dentro de amostragens (erro ou resíduo).

Como a variação total é medida em termos de variância, é calculada a soma de quadrados total, bem como o número de graus de liberdade, as quais representam, respectivamente, o numerador e o denominador de equação da variância. Através do desdobramento da soma de quadrados total de duas ou mais amostras de dados, obtém-se as suas respectivas somas de quadrados entre amostragens e dentro de amostragens.

Tais somas de quadrados divididas pelos seus respectivos graus de liberdade fornecem os quadrados médios (variâncias) entre amostragens e dentro de amostragens, respectivamente, os quais são confrontados através de um teste de hipótese (por exemplo, o teste F) para verificar se as amostras avaliadas diferem significativamente ou não com relação a alguma variável.

Os dados relativos às somas de quadrados e aos graus de liberdade, bem como os quadrados médios serão colocados numa tabela, chamada de Quadro de Análise de Variância. A composição desta tabela está explicitada na TABELA 4.1.

TABELA 4.1 – QUADRO DA ANÁLISE DE VARIÂNCIA SEGUNDO UM ÚNICO CRITÉRIO*

Causa de Variação	Graus de Liberdade (GL)	Soma de Quadrados (SQ)	Quadrados Médios (QM)	F Calculado
Entre Amostragens	$t - 1$	$SQ1$	$QM1 = \frac{SQ1}{t-1}$	$F = \frac{QM1}{QM2}$
Dentro de Amostragens	$t\,(r - 1)$	$SQ2 = SQ\,Total - SQ1$	$QM2 = \frac{SQ2}{t\,(r-1)}$	
Total	$t\,x\,r - 1$	$SQ\,Total$		

*: A análise de variância é denominada “segundo um único critério”, porque, no caso apresentado, foi levado em consideração apenas um critério, representado pelos efeitos das várias amostragens (tratamentos). Os experimentos planejados com base neste tipo de análise são denominados “experimentos inteiramente casualizados”.

As fórmulas matemáticas e o processo de análise de variância para cada tipo de experimento serão vistos em capítulos posteriores, quando for feita uma abordagem sobre cada delineamento estatístico.

4.2 Suposições da Análise de Variância

Além de aprender as regras para efetuar uma análise de variância, todo pesquisador deve buscar o domínio e a compreensão dos princípios inerentes à mesma, para não se defrontar com sérios problemas, como por exemplo, chegar a conclusões que não têm justificativas ou não alcançar conclusões importantes porque os dados não foram analisados adequadamente.

Desse modo, para que a análise de variância possa ter validade, o pesquisador deve atender às seguintes suposições:

a) **Os efeitos principais devem ser aditivos** – Nos experimentos, os vários efeitos devem ser aditivos, tanto é que para cada delineamento estatístico existe um modelo matemático denominado **modelo linear aditivo**. Para o delineamento inteiramente casualizado, este modelo é $X_{ij} = \hat{m} + t_i + e_{ij}$, onde expressa que o valor de qualquer unidade experimental é resultante de uma média geral, mais um efeito de tratamentos e mais um efeito do erro experimental. O modelo correspondente ao delineamento em blocos casualizados é: $X_{ij} = \hat{m} + t_i + b_j + e_{ij}$, onde o valor de qualquer unidade experimental é resultante de uma média geral, mais um efeito de tratamentos, mais um efeito de blocos e mais um efeito do erro experimental. Para o delineamento em quadrado latino, este modelo é: $X_{ijk} = \hat{m} + t_{(k)ij} + l_j + c_j + e_{ijk}$, onde o valor de qualquer unidade experimental é resultante de uma média geral, mais um efeito de tratamentos, mais um efeito de linhas, mais um efeito de colunas e mais um efeito do erro experimental. O aspecto importante, que deve notar-se nestes modelos, é que os efeitos se somam; daí o nome de modelo linear aditivo.

O modelo para o delineamento em blocos casualizados, por exemplo, implica que um efeito de tratamento é o mesmo para todos os blocos e que o efeito de bloco é o mesmo para todos os tratamentos. Em outras palavras, encontra-se que um tratamento aumenta a produção em certa quantidade acima da média geral, supomos que este tenha o mesmo efeito tanto nos blocos de alta produção como nos blocos de baixa produção.

Caso o que foi exposto acima não se verifique, é necessário transformar os dados experimentais para ajustá-los ao modelo aditivo.

b) **Os erros de observação devem ser independentes** – Cada observação possui um erro que deve ser independente dos demais. O princípio da casualização assegura a validade da estimativa do erro experimental, pois permite uma distribuição independente do mesmo. A casualização evita que todas as parcelas que recebem o mesmo tratamento ocupem posições adjacentes na área experimental, visto que as parcelas adjacentes, principalmente no campo, tendem a estar mais relacionadas entre si do que as parcelas distribuídas aleatoriamente.

c) **Os erros de observação devem ser normalmente distribuídos** – A única fonte de variação dentro de amostragens são os erros aleatórios. Estes devem ter distribuição normal (ou aproximadamente normal) com média igual a zero e variância igual a S^2. Felizmente, as variações da suposição de normalidade não afetam muito seriamente a validade da análise de variância.

A normalidade dos dados pode ser verificada por um teste de normalidade, como por exemplo, o teste do quiquadrado, desde que o número de amostras com as quais estão trabalhando seja definitivamente grande.

Quando se verifica que falta normalidade aos dados, usam-se as transformações para que os mesmos sejam normalmente distribuídos. De modo geral, dados médios de parcelas têm distribuição normal.

d) **As variâncias das diferentes amostras devem ser homogêneas** – Na análise de variância, o valor do Quadrado Médio do Resíduo, que corresponde à estimativa da variância do erro experimental, é utilizado nas fórmulas matemáticas dos testes de hipóteses. Tais testes são utilizados para verificar se existe ou não diferença significativa entre os tratamentos avaliados. Assim sendo, é importante que as estimativas das variâncias dos diferentes tratamentos (amostras) sejam homogêneas, ou seja, não deve haver uma variação muito grande entre suas estimativas, de modo que os resultados obtidos dos testes de hipóteses tenham validade. Vale ressaltar que, no delineamento inteiramente casualizado, o Quadrado Médio do Resíduo corresponde exatamente a média das estimativas das variâncias de cada tratamento. Já nos outros delineamentos estatísticos (delineamento em blocos casualizados e delineamento em quadrado latino) o Quadrado Médio do Resíduo é menor que a média das estimativas das variâncias de cada tratamento, em função das outras fontes de variação provocadas pela variação externa que são isoladas do resíduo através do uso do princípio do controle local (blocos no delineamento em blocos casualizados e linhas e colunas no delineamento em quadrado latino).

Entre os vários testes estatísticos utilizados para verificar a homogeneidade de variâncias, têm o teste F-máximo, proposto por Hartley.

O teste F-máximo é simples e rápido, porém apresenta menor precisão quando as amostras têm graus de liberdade diferentes.

A fórmula do referido teste é a seguinte:

$$\text{F-máximo} = \frac{s^2 \, máxima}{s^2 \, mínima}$$

onde:
s^2 máxima = maior valor das estimativas das variâncias entre as amostras;
s^2 mínima = menor valor das estimativas das variâncias entre as amostras.

O valor calculado de F-máximo é confrontado com o valor de F-máximo tabelado, com K = número de estimativas das variâncias das diferentes amostras e (N – 1) graus de liberdade associados a cada estimativa de variância, sendo N = número de observação de cada amostra (TABELA A.1).

Logo tem-se:

F-máximo calculado > F-máximo tabelado (1%) - ** (as estimativas das variâncias são estatisticamente diferentes no nível de 1% de probabilidade, isto é, não há homogeneidade de variâncias);

F-máximo calculado $\leq$ F-máximo tabelado (1%) - recorre-se no nível de 5% de probabilidade;

F-máximo calculado > F-máximo tabelado (5%) - * (as estimativas das variâncias são estatisticamente diferentes no nível de 5% de probabilidade, isto é, não há homogeneidade de variâncias);

F-máximo calculado $\leq$ F-máximo tabelado (5%) - ns (as estimativas das variâncias não diferem estatisticamente entre si no nível de 5% de probabilidade, isto é, as variâncias são homogêneas).

Quando os graus de liberdade para cada amostra são diferentes, toma-se a média aritmética dos mesmos para usar a TABELA A.1.

Exemplo 1: Verificar se as variâncias são homogêneas pelo teste F-máximo a partir dos dados da TABELA 4.2.

TABELA 4.2 – PESOS DE 20 CAPULHOS, EM GRAMAS, DE VARIEDADES DE ALGODÃO HERBÁCEO NO MUNICÍPIO DE VIÇOSA-AL

Variedades	I	II	III	IV	V	VI	Totais de Variedades
1 – ALLEN - 333/57	78,0	90,0	90,0	75,0	70,0	88,0	491,0
2 – AFC - 65/5236	100,0	65,0	78,0	92,0	85,0	90,0	510,0
3 – IAC - 13.1	102,0	95,0	102,0	85,0	80,0	98,0	562,0
4 – IPEANE – SU – 01	98,0	70,0	85,0	85,0	88,0	80,0	506,0

FONTE: FERREIRA (1977).

As variâncias de cada variedade são:

$$s_1^2 = \frac{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}{N-1}$$

$$= \frac{(78{,}0)^2 + (90{,}0)^2 + (90{,}0)^2 + (75{,}0)^2 + (70{,}0)^2 + (88{,}0)^2 - \frac{(491{,}0)^2}{6}}{6-1}$$

$$= \frac{6.084{,}0 + 8.100{,}0 + 8.100{,}0 + 5.625{,}0 + 4.900{,}0 + 7.744{,}0 - \frac{241.081{,}0}{6}}{5}$$

$$= \frac{40.553{,}0 - 40.180{,}1667}{5}$$

$$= \frac{372{,}8333}{5} \cong \mathbf{74{,}5667}$$

$$s_2^2 = \frac{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}{N-1}$$

$$= \frac{(100{,}0)^2 + (65{,}0)^2 + (78{,}0)^2 + (92{,}0)^2 + (85{,}0)^2 + (90{,}0)^2 - \frac{(510{,}0)^2}{6}}{6-1}$$

$$= \frac{10.000{,}0 + 4.225{,}0 + 6.084{,}0 + 8.464{,}0 + 7.225{,}0 + 8.100{,}0 - \frac{260.100{,}0}{6}}{5}$$

$$= \frac{44.098{,}0 - 43.350{,}0}{5}$$

$$= \frac{748{,}0}{5} = \mathbf{149{,}6000}$$

$$s_3^2 = \frac{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}{N-1}$$

$$= \frac{(102{,}0)^2 + (95{,}0)^2 + (102{,}0)^2 + (85{,}0)^2 + (80{,}0)^2 + (98{,}0)^2 - \frac{(562{,}0)^2}{6}}{6-1}$$

$$= \frac{10.404{,}0 + 9.025{,}0 + 10.404{,}0 + 7.225{,}0 + 6.400{,}0 + 9.604{,}0 - \frac{315.844{,}0}{6}}{5}$$

$$= \frac{53.062{,}0 - 52.640{,}6667}{5}$$

$$= \frac{421{,}3333}{5} \cong \mathbf{84{,}2667}$$

$$s_4^2 = \frac{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}{N-1}$$

$$= \frac{(98{,}0)^2 + (70{,}0)^2 + (85{,}0)^2 + (85{,}0)^2 + (88{,}0)^2 + (80{,}0)^2 - \frac{(506{,}0)^2}{6}}{6-1}$$

$$= \frac{9.604{,}0 + 4.900{,}0 + 7.225{,}0 + 7.225{,}0 + 7.744{,}0 + 6.400{,}0 - \frac{256.036{,}0}{6}}{5}$$

$$= \frac{43.098{,}0 - 42.672{,}6667}{5}$$

$$= \frac{452{,}3333}{5} \cong \mathbf{85{,}0667}$$

$$\text{F-máximo} = \frac{s^2 máxima}{s^2 mínima}$$

$$= \frac{149{,}6000}{74{,}5667} \cong \mathbf{2{,}01}$$

F-máximo tabelado (K = 4; N – 1 = 5): 1% = 28,0; 5% = 13,7.

Logo, F-máximo = 2,01 ns. Assim, chega-se à conclusão de que as estimativas das variâncias do peso de 20 capulhos de variedades de algodão herbáceo são homogêneas.

Uma regra prática e rápida para verificar a homogeneidade de variâncias é que a relação entre a maior e a menor delas não pode ser superior a mais de quatro vezes para que elas sejam homogêneas.

Quando às variâncias das diferentes amostras não são homogêneas, têm-se diversos cursos a seguir. Primeiro, pode-se separar as amostras em grupos, de modo que as variâncias dentro de cada grupo sejam homogêneas. Assim, a análise de variância poderá ser efetuada para cada grupo. Segundo, pode-se utilizar um método descrito em textos mais avançados de estatística, o qual contempla um procedimento bastante complicado para ponderar médias de acordo com suas variâncias. Terceiro, pode-se transformar os dados de tal forma que eles fiquem homogêneos. Este método é o mais utilizado na prática.

4.3 Transformações de Dados

Conforme foi visto, na análise de variância, algumas condições são exigidas para que os testes de hipóteses tenham validade. Contudo, como tais condições raramente são verificadas na prática, vários procedimentos são utilizados com o fim de reparar (pelo menos aproximadamente) a falta de verificação dessas condições. Dentre os procedimentos, geralmente utilizam-se transformações de dados.

Uma transformação é qualquer alteração sistemática num conjunto de dados onde certas características são mudadas e outras permanecem inalteradas.

As principais transformações são:

a) **Raiz quadrada –** Própria para certos tipos de dados em que a média é aproximadamente igual à variância, ou seja, para dados oriundos de uma distribuição de Poisson (tipo de distribuição em que os dados apresentam uma probabilidade muito baixa de ocorrência em qualquer indivíduo – os fenômenos naturais são os exemplos mais óbvios desse tipo de ocorrência). Tais tipos de dados ocorrem quando as variáveis são oriundas de contagem como: número de sementes por planta ou por parcela, número de dias para enraizamento de bulbos por parcela, número de insetos por planta ou por parcela, número de plantas atacadas por um determinado patógeno por parcela, número de carrapatos por animal ou por parcela, número de animais doentes por parcela, etc.. Os dados provenientes de uma escala de notas também devem ser transformados através da raiz quadrada. Também os dados de porcentagens, referentes às contagens, quando variam de 0 a 20% ou de 80 a 100%, podem ser transformados através da raiz quadrada. Neste caso, as porcentagens entre 80 e 100% devem ser, de preferência, subtraídas de 100, antes de se fazer à transformação. A transformação da raiz quadrada é, ainda, indicada no caso de porcentagens, fora dos limites acima considerados, quando as observações estão claramente numa escala contínua.

Neste caso tem-se: $\sqrt{x}$.

Quando nesse tipo de transformação os dados variam de 0 a 10, trabalha-se com $\sqrt{x + 0{,}5}$ ou $\sqrt{x + 1}$, em lugar de $\sqrt{x}$, pois evita-se o problema dos valores de X iguais a zero.

b) **Logarítmica –** É usada sempre que se têm dados em que os desvios padrões das amostras são aproximadamente proporcionais às médias, ou seja, todas as amostras apresentam o mesmo coeficiente de variação. Também quando os efeitos principais são multiplicativos, em vez de aditivos, os dados devem ser transformados através desse tipo de transformação. Essa transformação é satisfatória quando os dados se referem à contagem de bactérias, de esporos, de grãos de pólen, de ovos de insetos, de ácaros, etc.. Dados provenientes de adição de vitaminas em animais também devem ser transformados através da transformação logarítmica. É utilizada, ainda, quando os dados são apresentados por porcentagens que abrangem uma grande amplitude de variação. Esse tipo de transformação resolve tanto o problema de heterogeneidade de variâncias como a falta de aditividade no modelo.

Nesse caso tem-se: log x.

Na transformação logarítmica, quando a amostra possui dados iguais a zero ou muito próximos de zero, trabalha-se com log (x + 1), pois se evita que se usem números negativos na análise, além de resolver o problema de valores de X iguais a zero.

Essa transformação deve ser usada quando as variâncias de cada amostra possuem, no mínimo, 12 observações.

c) **Arcoseno ou angular –** Própria para dados em que a média é proporcional à variância, ou seja, para dados oriundos de uma distribuição binomial (tipo de distribuição em que os dados apresentam uma probabilidade calculável de ocorrência ou não em qualquer indivíduo). Tais tipos de dados ocorrem quando as variáveis são oriundas de proporção como: porcentagem de germinação de sementes, porcentagem de mortalidade de plantas infectadas com vírus, porcentagem de sobrevivência de bezerros da raça Nelore, etc..

Nesse caso tem-se: arco seno $\sqrt{x(\%)}$.

Na transformação arco seno, quando todos os dados estão entre 30 e 70% não precisa usar a transformação. Se os dados extrapolam esta amplitude, usa-se então a transformação.

Quando o número de observações for menor que 50 (N < 50), a proporção 0% deve ser substituída por $\frac{1}{4}$N e a proporção 100% para 100 – $\frac{1}{4}$ N, antes de transformar os dados em arco seno $\sqrt{x(\%)}$.

Existe uma tabela própria para esta transformação (TABELA A.2).

4.3.1 Escolha da melhor transformação

Em alguns casos fica-se sem saber qual seria a transformação mais adequada. Quando se defrontar com tais situações, têm-se várias maneiras para escolher a melhor transformação. Entre as várias maneiras, uma das mais simples é por meio de gráficos, onde se coloca no eixo dos **x** e **y** as médias e variâncias respectivas de cada amostra para cada transformação e seleciona-se a que apresentar menor dispersão.

Outro procedimento é aplicar cada transformação para o maior e o menor dado de cada amostra. A amplitude dentro de cada amostra é determinada e a razão entre a maior e a menor amplitude é calculada. A transformação que produz a menor razão é a selecionada.

Exemplo 2: Escolher a melhor transformação a partir de dados da TABELA 4.3.

TABELA 4.3 – PERÍODO DE ENRAIZAMENTO (EM DIAS) DE CULTIVARES DE CEBOLA (*Allium cepa* L.) DE DIAS CURTOS. PIRACICABA – SP

Cultivares	I	II	Totais de Cultivares
01 – BAIA PERFORME	48,0	33,4	81,4
02 – BAIA DO CEDO SMP-V	18,4	10,2	28,6
03 – BAIS TRIUNFO SMJ-II	46,6	42,8	89,4
04 – BARREIRO SMJ-II	14,0	32,0	46,0
05 – COJUMATLAN L. 2691	10,6	2,4	13,0
06 – CREOLA CATARINENSE	64,0	44,7	108,7
07 – EXCEL BEMUDAS 986	31,0	14,8	45,8
08 – IPA – 2	17,0	10,8	27,8
09 – PIRA OURO A/R	16,8	26,8	43,6
10 – PIRA TROPICAL A/C	15,2	9,8	25,0
11 – TEXAS GRANO	11,4	2,5	13,9
12 – WHITE CREOLE	26,0	18,4	44,4
13 – BAIA DO CEDO SMJ-III	24,2	8,4	32,6
14 – BAIA SETE VOLTAS	19,4	18,2	37,6
15 – BARREIRO ROXA SMP-IV	8,0	14,2	22,2
16 – BARREIRO SMP-III	22,0	36,2	58,2
17 – CIGANINHA	4,6	6,2	10,8
18 – CREOLA	19,8	28,4	48,2
19 – PIRA COUTO	16,2	22,2	38,4
20 – PIRA GRANA	32,6	21,4	54,0
21 – PIRA LOPES A/R	25,8	5,0	30,8
22 – PIRA PERA A/C	19,4	16,0	35,4
23 – PIRA LOPES A/C	18,6	8,0	26,6
24 – ROXA CHATA SMP – IV	13,0	5,4	18,4
25 – TUBARÃO	19,2	13,2	32,4

FONTE: FERREIRA (1982).

Os resultados estão contidos no quadro a seguir:

Cultivares	Raiz Quadrada			Logarítmica		
	Maior	Menor	Amplitude	Maior	Menor	Amplitude
1	6,9282	5,7793	1,1489	1,6812	1,5237	0,1575
2	4,2895	3,1937	1,0958	1,2648	1,0086	0,2562
3	6,8264	6,5422	0,2842	1,6684	1,6314	0,0370
4	5,6569	3,7417	1,9152	1,5052	1,1461	0,3591
5	3,2558	1,5492	1,7066	1,0253	0,3802	0,6451
6	8,0000	6,6858	1,3142	1,8062	1,6503	0,1559
7	5,5678	3,8471	1,7207	1,4914	1,1703	0,3211
8	4,1231	3,2863	0,8368	1,2304	1,0334	0,1970
9	5,1769	4,0988	1,0781	1,4281	1,2253	0,2028
10	3,8987	3,1305	0,7682	1,1818	0,9912	0,1906
11	3,3764	1,5811	1,7953	1,0569	0,3979	0,6590
12	5,0990	4,2895	0,8095	1,4150	1,2648	0,1502
13	4,9193	2,8983	2,0210	1,3838	0,9243	0,4595
14	4,4045	4,2661	0,1384	1,2878	1,2601	0,0277
15	3,7683	2,8284	0,9399	1,1523	0,9031	0,2492
16	6,0166	4,6904	1,3262	1,5587	1,3424	0,2163
17	2,4900	2,1448	0,3452	0,7924	0,6628	0,1296
18	5,3292	4,4497	0,8795	1,4533	1,2967	0,1566
19	4,7117	4,0249	0,6868	1,3464	1,2095	0,1369
20	5,7096	4,6260	1,0836	1,5132	1,3304	0,1828
21	5,0794	2,2361	2,8433	1,4116	0,6990	0,7126
22	4,4045	4,0000	0,4045	1,2878	1,2041	0,0837
23	4,3128	2,8284	1,4844	1,2695	0,9031	0,3664
24	3,6056	2,3238	1,2818	1,1139	0,7324	0,3815
25	4,3818	3,6332	0,7486	1,2833	1,1206	0,1627
$Razão = \frac{Amp.\,máx.}{Amp.\,mín.}$	$\frac{2{,}8433}{0{,}1384} \cong 20{,}54$			$\frac{0{,}7126}{0{,}0277} \cong 25{,}73$		

Pelos resultados apresentados acima, verifica-se que a transformação mais adequada é a raiz quadrada, pois a mesma apresentou o menor coeficiente entre as amplitudes (20,54).

4.3.2 Coeficiente de variação como indicativo para o uso de transformações

De um modo geral, uma indicação razoável do efeito favorável das transformações de dados é o coeficiente de variação (CV). Quando o valor do CV dos dados transformados for menor que o valor do CV dos dados originais ou não transformados, indica que a transformação foi válida. Em caso contrário, não se justifica o seu uso.

Considerando os dados do Exemplo 2, tem-se:

Dados originais CV = 38,26%

Dados transformados em $\sqrt{x}$ CV = 21,35%

Dados transformados em log x CV = 32,49%

Realmente, as transformações de dados foram válidas, pois houve uma redução muito significativa nos coeficientes de variação em relação aos dados originais, indicando que os dados experimentais foram ajustados de acordo com as exigências da análise de variância. Contudo, a transformação da raiz quadrada foi novamente confirmada como sendo a melhor transformação para tais dados.

4.3.3 Algumas considerações

Quando é utilizada uma transformação de dados, todas as comparações entre médias de tratamentos são feitas na escala transformada. Quando se achar preferível não apresentar os resultados na escala transformada, os dados finais devem ser transformados novamente para a escala original. Isto é feito elevando-se ao quadrado, no caso de $\sqrt{x}$; achando o antilogarítmo, no caso de log x; e procurando o valor correspondente na tabela de arco seno $\sqrt{x(\%)}$, no caso de transformação angular.

Em certos casos, não existe nenhuma transformação que possibilite o uso da análise de variância. Isto ocorre quando:

a) As médias são aproximadamente iguais e as variâncias heterogêneas;

b) As variâncias são homogêneas, porém os níveis dos tratamentos são heterogêneos em forma;

c) As médias variam independentemente das variâncias.

Se alguns destes casos ocorrem, a análise dos dados é feita através de métodos não-paramétricos.

4.4 Exercício

a) Considerando-se os dados da TABELA 4.4, pede-se:

a.1) Verifique se as variâncias dos tratamentos são homogêneas pelo teste F – máximo;

a.2) Transforme os dados originais através da transformação raiz quadrada;

a.3) Transforme os dados originais em porcentagem e, em seguida, na transformação arco seno;

a.4) Verifique qual é a transformação mais adequada para estes dados;

a.5) Comprove através do coeficiente de variação a necessidade do uso de transformação nestes dados.

TABELA 4.4 – DADOS DE GERMINAÇÃO DE 20 SEMENTES, APÓS 21 DIAS, REFERENTE AO USO DE DIFERENTES TRATAMENTOS PARA QUEBRA DE DORMÊNCIA EM TAMBORIL (*Enterolobium confortisiliquum*)

Tratamentos	I	II	III	IV	Totais de Tratamentos
1 – Ácido Sulfúrico Concentrado	8	8	7	6	29
2 – Ácido/água em 3:1	3	3	4	4	14
3 – Ácido/água em 2:1	5	5	6	4	20
4 – Ácido/água em 1:1	8	9	6	9	32
5 – Ácido/água em 1:2	10	10	8	10	38
6 – Ácido/água em 1:3	2	3	2	3	10
7 – Água Quente	5	5	7	5	22
8 – Testemunha	1	2	1	1	5

FONTE: Adaptado de SILVA e SILVA (1982).