EFEITO DE HETEROGENEIDADE DE VARIÂNCIA E DISTRIBUIÇÃO DE PROBABILIDADE DOS DADOS SOBRE O PODER E TAMANHO DO TESTE F[1]

FRANCISCO JOSÉ PFEILSTICKER ZIMMERMANN[2]

RESUMO - Estudou-se o efeito da heterogeneidade de variância dos tratamentos e da distribuição de probabilidade dos dados sobre o tamanho e o poder do teste F, em dois delineamentos experimentais: blocos ao acaso (com 3, 5, 7, 9 e 11 tratamentos) e quadrado latino (tamanhos de 3 x 3 até 7 x 7). Os dados foram gerados a partir de sete distribuições probabilísticas (normal, uniforme, logística, Laplace, Weibull, exponencial e Cauchy). Para cada caso, geraram-se 10.000 experimentos e contou-se o número de vezes que o teste F foi significativo, tanto nos casos de hipótese nula falsa, quanto verdadeira. Com base nesta contagem, concluiu-se que a não-normalidade dos dados não influenciou o tamanho e o poder do teste F, que só foram afetados quando houve heterogeneidade de variância. Recomenda-se, então, a transformação dos dados somente em casos de heterocedasticidade, principalmente quando associada à distribuição normal, pois, aí, verificou-se que o teste F foi totalmente prejudicado no aspecto de tamanho do teste.

Termos para indexação: estatística matemática, heterocedasticidade, simulação, transformação de dados.

EFFECTS OF THE HETEROGENEITY OF VARIANCE AND PROBABILITY DISTRIBUTION OF THE DATA ON THE POWER AND SIZE OF THE F TEST

ABSTRACT - The effect of the heterocedasticity of treatments and the distribution of probability over the size and power of the F test was studied, in two experimental designs: complete randomized blocks (with 3, 5, 7, 9 e 11 treatments) and latin square (from sizes 3 x 3 to 7 x 7). Data were generated from seven probability distributions, to wit, normal, uniform, logistic, Laplace, Weibull, exponential and Cauchy. In each case 10,000 experiments were generated and it was counted the number of times the F test was significant, for the null hipotheses either true or false. Based on this counting, it could be concluded that non-normality does not affect the size and the power of the F test, which were affected only when heterocedasticity of the treatments occurred. Transformation of the data is recommended only when there is heterogeneity of variance mainly when associated to normal distribution, in which case the F test was not efficient and had a very bad test-size.

Index-terms: mathematical statistics of treatments, heterocedasticity simulation, data transformation.

INTRODUÇÃO

A análise de variância é usada fundamentalmente na busca de solução de problemas de dois tipos (Eisenhart 1947):

1. detecção e estimação de relações físicas entre médias; e

2. detecção e estimação de componentes de variância associados a uma população composta.

O procedimento computacional e a mecânica do teste estatístico de significância são os mesmos em ambos os casos; o mesmo teste (F) é avaliado e referido aos mesmos níveis de significância, tanto num caso como no outro.

Para tornar válidos os procedimentos já tão conhecidos de análise de variância, apesar de algumas diferenças básicas de ordem inferencial, com respeito aos dois problemas antes referidos, são necessárias as seguintes pressuposições (Eisenhart 1947, Cochran 1947):

a) os números a serem analisados são valores observados de variáveis aleatórias, distribuídas ao redor de suas médias verdadeiras, que são constantes fixas;

b) as relações entre os parâmetros componentes do modelo matemático, associados aos valores obtidos, são relações aditivas;

c) as variáveis aleatórias são de igual variância e mutuamente não correlacionadas; e

d) as variâncias aleatórias são conjuntamente distribuídas segundo uma distribuição multivariada

[1] Aceito para publicação em 22 de dezembro de 1986.

[2] Eng. - Agr., Ph.D., EMBRAPA/Centro Nacional de Pesquisa de Arroz e Feijão (CNPAF), Caixa Postal 179, CEP 74000 Goiânia, GO.

normal.

Quando todas estas pressuposições são satisfeitas, os procedimentos usuais de análise de variância para estimação e teste de relações lineares fixas entre as médias ou de componentes de variância tornam-se estritamente válidos (Scheffé 1959).

Cochran (1947) afirmou que, em adição aos efeitos sobre a validade dos testes de significância é possível que a não-normalidade seja acompanhada de menor eficiência na estimação dos efeitos e haja perda correspondente de poder no teste F. Se for efetuada a análise de variância quando a variância verdadeira do erro diferir de uma observação para outra, será quase certo que ocorrerá perda de eficiência na estimação dos efeitos de tratamentos e haverá, também, uma perda de sensibilidade nos testes de significância. Quanto maiores forem as diferenças na variância, maiores serão estas perdas.

Neste trabalho procurou-se medir o efeito da heterocedasticidade das variâncias e/ou a não-normalidade dos dados sobre o tamanho e o poder do teste F, utilizando-se dois delineamentos experimentais dos mais conhecidos, blocos completos casualizados e quadrado latino, cada um com diversos números de tratamentos.

MATERIAL E MÉTODOS

Com o emprego do computador IBM 4341, 10.000 experimentos foram gerados para cada uma das combinações dos seguintes fatores:

a) Delineamento, número de tratamentos, repetições.
   - a.1. blocos completos casualizados, com 3, 5, 7, 9 e 11 tratamentos e, respectivamente, 7, 4, 3, 3 e 3 repetições.
   - a.2. quadrado latino, com 3, 4, 5, 6 e 7 tratamentos.

b) Médias de tratamentos.
   - b.1. todos os tratamentos com média zero;
   - b.2. tratamento i, com média igual ao inteiro de i dividido por 3;
   - b.3. tratamento i, com média i.

c) Variância dos tratamentos.
   - c.1. todos os tratamentos com variância igual a 1;
   - c.2. tratamento com variâncias diferentes e razão entre variância máxima e mínima igual a 4;
   - c.3. tratamentos com variâncias diferentes e razão entre variância máxima e mínima igual a 16.

d. Distribuição de probabilidade dos dados.

d.1. uniforme, $f(x) = \dfrac{x}{b-a}$

d.2. normal, $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

d.3. logística, $f(x) = \dfrac{e^{-(x-\alpha)/\beta}}{\beta[1+e^{-(x-\alpha)/\beta}]^2}$

d.4. exponencial, $f(x) = \lambda\, e^{\lambda(x-\alpha)}$

d.5. Weibull, $f(x) = abx^{b-1}\, e^{-a(x+\alpha)^b}$

d.6. Laplace, $f(x) = \dfrac{1}{2\beta}\, e^{-(x-\alpha)/\beta}$

d.7. Cauchy, $f(x) = \dfrac{1}{\pi\beta\left\{1+[(x-\alpha)/\beta]^2\right\}}$

No caso da distribuição de Cauchy, que não possui momentos finitos, utilizou-se, para a geração, o parâmetro de localização α, como média, e o parâmetro de escala β, como variância.

As amostras foram geradas a partir da distribuição uniforme (0,1), usando a sub-rotina IBM RANDOM, modificada por Garber (Zimmermann 1983). Foi, então, aplicada a transformação de probabilidade integral sugerida por Schreider (1966), para todas as distribuições, exceto a normal, para a qual se empregou o método de Box & Muller (1958).

RESULTADOS E DISCUSSÃO

Define-se tamanho de um teste ou nível de significância como a probabilidade de rejeitar-se a hipótese nula (Ho), quando ela é verdadeira, e poder de um teste como a probabilidade de rejeitar a hipótese nula (Ho), quando falsa. Assim, pode-se definir o tamanho empírico como a proporção de rejeição da hipótese nula, sendo Ho verdadeira, em M amostras de tamanho N, e usando uma região crítica para rejeição da hipótese nula para um dado nível α; e o poder empírico, como a proporção de rejeições de Ho, sendo esta falsa, para o mesmo nível α de significância.

A não-existência das pressuposições básicas de um teste deve, teoricamente, fazer com que os valores de significância nominal α e o tamanho empírico se tornem discrepantes. Assim, a ocorrência de uma boa aproximação dos dois valores indica ou o preenchimento das pressuposições, ou a robustez do teste, quando elas não ocorrem.

As Tabelas 1 e 2 mostram o tamanho empírico

TABELA 1. Percentagem de significância (x 10) do Teste F, no caso de tratamentos com médias iguais (Tamanho do teste). Delineamento: blocos ao acaso.

Distribuição de probabilidade	Variância														
	Homogênea					Heterogênea (Razão 4)					Heterogênea (Razão 16)				
	Nº de tratamentos					Nº de tratamentos					Nº de tratamentos				
	3	5	7	9	11	3	5	7	9	11	3	5	7	9	11
Uniforme	54	55	56	53	58	57	67	70	66	68	64	83	111	115	104
Normal	51	50	50	49	54	260	270	180	166	203	699	614	661	699	721
Logística	47	48	51	47	53	82	83	69	71	84	56	70	90	87	91
Exponencial	41	40	44	50	53	53	56	52	55	61	77	77	91	94	91
Weibull	50	49	48	54	54	54	62	59	62	66	63	77	92	100	94
Laplace	48	42	44	40	45	49	49	49	45	53	56	60	73	74	79
Cauchy	14	17	17	19	20	16	18	19	20	21	19	23	28	24	28

TABELA 2. Percentagem de significância (x 10) do Teste F, no caso de tratamentos com médias iguais (Tamanho do teste). Delineamento: quadrado latino.

Distribuição de probabilidade	Variância														
	Homogênea					Heterogênea (Razão 4)					Heterogênea (Razão 16)				
	Nº de tratamentos					Nº de tratamentos					Nº de tratamentos				
	3	4	5	6	7	3	4	5	6	7	3	4	5	6	7
Uniforme	51	58	53	54	50	53	65	65	65	66	62	70	78	76	84
Normal	50	55	52	50	50	134	264	408	503	646	186	418	702	810	942
Logística	48	43	47	48	48	47	51	56	58	60	46	57	68	69	73
Exponencial	40	37	42	45	46	47	48	53	56	57	58	59	75	72	78
Weibull	45	52	50	51	50	52	57	61	63	61	54	65	78	75	77
Laplace	45	44	47	43	45	41	46	53	54	54	39	51	63	64	68
Cauchy	27	22	6	16	16	24	19	21	19	20	24	20	24	22	23

do teste F, nos dois delineamentos estudados. Verifica-se que, com variância homogênea todas as distribuições que possuem momentos finitos apresentam tamanho em torno de 5%, enquanto que, com dados provenientes da distribuição de Cauchy, o teste F mostrou-se conservador, com tamanho inferior a 2,7%.

Quando se trabalhou com variâncias heterogêneas, o teste continuou sendo conservador para dados de Cauchy, (sempre inferior a 3%), enquanto que, para as outras distribuições, exceto a normal, o tamanho foi muito pouco afetado, quando as variâncias não discreparam de muito (razão 4), e um pouco afetado, quando a razão de variância era maior. Entretanto, para dados oriundos da distribuição normal, observou-se uma perda total da eficiência do teste, com valores de tamanho de até 94%, para um esperado de 5%.

Estes dados, concordam em parte com a afirmação de Cochran (1947), de que nenhum erro sério é introduzido na significância do teste F por não-normalidade e adicionalmente realçam o fato de que o tamanho do teste é afetado pela heterogeneidade de variância, principalmente quando associada a dados provenientes da distribuição normal.

Com relação aos efeitos da não-normalidade associada, ou não, à heterogeneidade de variância no poder do teste F, pode-se verificar nas Tabela 3 (para blocos ao acaso) e Tabela 4 (para quadrado latino) que, a exemplo do que ocorreu com o tamanho do teste, com dados provenientes da distribuição Cauchy, o teste F não tem sensibilidade suficiente para aceitar ou recusar a hipótese nula, seja ela falsa ou verdadeira.

Com relação às outras distribuições, verificou-se que, à medida que se aumentou o número de tratamentos, aumentou o poder do teste, e que isto

TABELA 3. Percentagem de significância (x 10) do teste F, no caso de tratamentos com médias desiguais (Poder do teste). Delineamento: blocos ao acaso.

Distribuição de probabilidade	Valores das médias de tratamento	Variância: Homogênea — Nº de tratamentos 3	5	7	9	11	Heterogênea (Razão 4) — Nº de tratamentos 3	5	7	9	11	Heterogênea (Razão 16) — Nº de tratamentos 3	5	7	9	11
Uniforme	i	851	998	1.000	1.000	1.000	519	847	985	1.000	1.000	244	222	294	582	797
	Inteiro (i/3)	372	257	473	813	963	206	146	237	387	515	125	97	134	160	163
Normal	i	861	996	1.000	1.000	1.000	956	962	994	1.000	1.000	986	964	918	970	999
	inteiro (i/3)	393	281	528	837	970	696	524	503	741	914	899	771	759	802	886
Logística	i	846	988	999	1.000	1.000	819	994	999	1.000	1.000	282	244	328	637	817
	inteiro (i/3)	398	285	524	809	954	434	405	731	876	926	134	84	118	146	164
Exponencial	i	844	975	996	1.000	1.000	592	849	953	997	1.000	214	170	329	681	835
	inteiro (i/3)	440	321	561	821	935	205	141	291	454	584	89	68	99	155	163
Weibull	i	845	992	1.000	1.000	1.000	546	847	972	1.000	1.000	231	200	288	626	815
	inteiro (i/3)	387	270	498	815	956	198	138	249	405	535	106	86	110	148	164
Leplace	i	842	981	1.000	1.000	1.000	576	846	958	997	1.000	305	265	353	656	816
	inteiro (i/3)	418	300	539	819	942	243	159	283	440	567	140	78	106	139	156
Cauchy	i	94	149	212	289	379	55	81	115	154	195	36	32	41	51	63
	Inteiro (i/3)	40	30	42	53	64	30	23	29	34	36	26	24	31	27	31

TABELA 4. Percentagem de significância (x 10) do teste F, no caso de tratamentos com médias desiguais (Poder do teste). Delineamento: quadrado latino.

Distribuição de probabilidade	Valores das médias de tratamento	Variância: Homogênea — Nº de tratamentos 3	4	5	6	7	Heterogênea (Razão 4) — Nº de tratamentos 3	4	5	6	7	Heterogênea (Razão 16) — Nº de tratamentos 3	4	5	6	7
Uniforme	i	175	170	1.000	1.000	1.000	160	680	999	1.000	1.000	122	461	976	1.000	1.000
	Inteiro (i/3)	93	205	318	806	973	89	183	291	701	953	81	143	213	480	804
Normal	i	186	795	1.000	1.000	1.000	402	984	1.000	1.000	1.000	403	976	1.000	1.000	1.060
	Inteiro (i/3)	100	216	348	835	978	257	685	936	999	1.000	297	754	972	1.000	1.000
Logística	i	190	785	997	1.000	1.000	184	720	993	1.000	1.000	137	538	946	999	1.000
	inteiro (i/3)	98	222	349	799	964	98	198	334	717	942	80	149	245	520	809
Exponencial	i	219	794	989	1.000	1.000	166	778	998	1.000	1.000	92	560	989	1.000	1.000
	inteiro (i/3)	106	254	386	812	949	71	175	311	761	965	45	103	198	526	887
Weibull	i	181	781	998	1.000	1.000	158	720	999	1.000	1.000	109	483	981	1.000	1.000
	inteiro (i/3)	93	209	349	810	964	80	172	296	722	959	65	126	209	488	827
Leplace	i	199	780	994	1.000	1.000	193	723	983	1.000	1.000	144	548	934	998	1.000
	inteiro (i/3)	96	227	360	809	954	95	210	344	727	931	77	155	259	543	808
Cauchy	i	42	88	162	212	271	47	84	161	196	271	42	62	130	153	229
	inteiro (i/3)	34	30	27	44	50	34	30	36	45	52	31	27	34	40	47

ocorreu também quando se aumentou a discrepância entre as médias dos tratamentos, passando do valor igual ao inteiro de (i/3) para valor igual a i.

As Tabelas 3 e 4 mostram, também, que houve diminuição no poder do teste F, quando havia heterocedasticidade dos tratamentos e que esta queda era maior quanto mais discrepante fossem as variâncias, exceto no caso da distribuição normal, onde se verificou um aumento no poder do teste. De maneira geral, verificou-se que, com médias bem discrepantes, no caso o tratamento i, média igual a i; com cinco ou mais tratamentos, já se tem poder do teste ao redor de 100%, ao passo que, para médias não muito diferentes, há a necessidade de maior número de tratamentos para rejeitar-se a hipótese nula. Estes dados concordam com a afirmação de Scheffée (1959), da robustez do teste F, com relação à não-normalidade dos dados, mas não quanto à sua robustez com relação à heterogeneidade de variância associada à distribuição normal.

CONCLUSÕES

1. O teste F é conservador e tem baixo poder de rejeição da hipótese nula, quando falsa, para dados oriundos da distribuição sem momentos finitos, isto é, Cauchy.

2. O tamanho e o poder do teste F não são afetados quando os dados têm variância homogênea, independentemente da distribuição de probabilidade.

3. O tamanho do teste é afetado pela heterogeneidade de variância, principalmente quando associada à distribuição normal.

4. O poder do teste, a exceção da distribuição normal, é reduzido sensivelmente, quando a razão das variâncias dos tratamentos é elevada.

5. Com base nas conclusões 2, 3 e 4, é recomendada a transformação dos dados somente nos casos de heterogeneidade de variância entre os tratamentos.

REFERÊNCIAS

BOX, G.E.P. & MULLER, M.E. A note on the generation of random normal deviates. **Ann. Math. Stat.**, 29: 610-1, 1958.

COCHRAN, W.G. Some consequences when the assumptions for the analysis of variance are not satisfied. **Biometrics**, 3:22-38, 1947.

EISENHART, C. The assumptions underlying the analysis of variance. **Biometrics**, 3:1-21, 1947.

SCHEFFÉ, H. **The analysis of variance.** New York, J. Wiley, 1959. 478p.

SCHREIDER, Y.A. **The Monte Carlo method**; the method of statistical trials. s.l., Pergamon, 1966. 275p.

ZIMMERMANN, F.J.P. **Distribution-free tests in the Latin square design.** Riverside, University of California, 1983. 44p. Tese Ph.D.