Comparação da robustez de alternativas do teste de igualdade de duas médias populacionais sob não normalidade por simulação Monte Carlo

Roberta Bessa Veloso Silva e Daniel Furtado Ferreira*

*Departamento de Ciências Exatas, Universidade Federal de Lavras, 37200-000, Lavras, Minas Gerais, Brasil. *Autor para correspondência. e-mail: danielff@ufla.br*

RESUMO. O presente trabalho teve por objetivos avaliar os riscos de se tomar decisões erradas (erro tipo I e erro tipo II) em populações não-normais, por meio de simulação computacional, e comparar três testes usualmente empregados. Foram comparados o teste t com a aproximação dos graus de liberdade proposto por Satterthwaite (1946), t com os graus de liberdade dado por $\upsilon = min(n_1 - 1,\ n_2 - 1)$ e método de bootstrap sob diferentes distribuições de probabilidade. Sob distribuições não-normais, t com ajuste de graus de liberdade por Satterthwaite e t com $\upsilon = min(n_1 - 1,\ n_2 - 1)$ não controlaram as probabilidades de erro tipo I; o critério de bootstrap controlou as probabilidades de erro tipo I e apresentou poder do teste equivalente, sendo considerado robusto com a violação da pressuposição de normalidade; o teste de t com graus de liberdade ajustado (Satterthwaite) em populações não-normais e amostras de tamanhos diferentes apresentou probabilidades de erro tipo I maiores que as nominais.

Palavras-chave: probabilidade do erro tipo I, poder do teste, bootstrap, não normalidade.

ABSTRACT. Comparison of the robustness of alternatives to the two sample test under non normality distributions through Monte Carlo simulation. This work aimed to evaluate the risks of committing type I and type II errors in non normal populations by means of computational simulation and to compare three tests usually applied. It was compared the t test with the approach of the degrees of freedom proposed by Satterthwaite (1946), t with the degrees of freedom given by $\upsilon = min(n_1 - 1,\ n_2 - 1)$ and bootstrap method under different distributions of probability. Under non normal distribution the t with Satterthwaite adjustment of degrees of freedom and with $\upsilon = min(n_1 - 1,\ n_2 - 1)$ degrees of freedom did not control type I error probabilities. The bootstrap criterion controlled the type I error rates and presented equivalent power being considered robust with the violation of the normality assumption. The t test under non normal distribution with Satterthwaite adjustment of degrees of freedom with samples of different sizes presented type I error rates greater than the nominal levels.

Key-words: type I error rates; power of the test; bootstrap, non normality.

Introdução

A comparação de duas médias populacionais é feita por meio de duas alternativas experimentais: (a) comparações pareadas, em que a amostra selecionada na população é avaliada antes e após a aplicação de um tratamento; (b) comparações independentes, em que as duas populações que se deseja comparar são amostradas de forma independente. O segundo caso é o mais freqüente na experimentação, sendo usado quando não existe nenhuma razão plausível para o pareamento (Snedecor e Cochran, 1980), embora o primeiro seja, em geral, mais eficiente.

Para essas comparações, e considerando principalmente o segundo caso das comparações independentes, a formulação do teste de t de Student é feita sob condições de normalidade das populações amostradas. Atendidas as outras pressuposições, independência e homocedasticidade, a validade do teste de t é garantida com a normalidade. Os pesquisadores das ciências biológicas não podem garantir que as amostras realizadas de forma aleatória são provenientes de populações normais e, ainda, que possuam variâncias homogêneas.

Muitos estudos, no entanto, têm sido realizados mostrando que o teste de t é robusto o suficiente para considerar substanciais desvios de seus pressupostos teóricos, principalmente quando as amostras são de mesmo tamanho e quando hipóteses bilaterais são consideradas (Box, 1953; Srivastava, 1958; Boneau, 1960; Poste *et al.*, 1982). Populações assimétricas afetam muito pouco os testes bilaterais, mas afetam severamente os testes unilaterais, principalmente com pequenas amostras. O poder do teste é menor do que o encontrado em situações ideais para populações platicúrticas e maior para as populações leptocúrticas, em especial para pequenas amostras (Glass *et al.*, 1972).

Essas violações do teste, em geral, afetam os riscos de se cometer o erro tipo I, ou seja, de rejeitar uma hipótese verdadeira, e o do tipo II, de aceitar uma hipótese falsa. Com a alteração das probabilidades de erros, o pesquisador tem grande chance de tomar decisões erradas.

Duas aproximações são citadas na literatura para se aplicar o teste t. A primeira é a aproximação de Satterthwaite (1946), na qual modifica-se o cálculo do número de graus de liberdade para se empregar a estatística. A segunda alternativa é considerar os graus de liberdade como sendo $\upsilon = min\,(\,n_1 - 1,\; n_2 - 1)$. Embora essa alternativa seja raramente usada, nenhum estudo foi realizado comparando sua performance com a primeira em situações de não normalidade.

Uma terceira opção que vem sendo usada na literatura, além de outras abordagens, é a dos testes que fazem uso de computação intensiva. Esses testes são referenciados como testes de bootstrap e aleatorização (Manly, 1998). Esses dois procedimentos diferem apenas, na maioria das circunstâncias, se a reposição é ou não considerada no processo de reamostragem. Segundo Manly (1998), são substituídas, em cada amostra as observações pelos resíduos obtidos da subtração da observação amostral pela média amostral correspondente. Os resíduos de ambas as amostras são combinadas em uma única amostra de tamanho n_1+n_2, sendo então retiradas amostras com reposição de tamanhos n_1 e n_2, respectivamente, e calculada a estatística t do teste convencional. Esses valores são obtidos milhares de vezes e o valor do teste obtido com a amostra original é confrontado com valores dessa distribuição gerada (Manly, 1998). Por não terem sido encontrados trabalhos comparando essa abordagem com as anteriormente mencionadas sob situações de não normalidade das populações estudadas, o presente trabalho foi realizado, tendo por objetivos avaliar os riscos de se tomar decisões erradas (erro tipo I e erro tipo II) em populações não-normais, através de simulação computacional, utilizando-se o teste t com a aproximação dos graus de liberdade proposto por Satterthwaite (1946), com os graus de liberdade dado por $\upsilon = min\,(\,n_1 - 1,\; n_2 - 1)$ e usando a abordagem de bootstrap e comparar os três critérios propostos sob diferentes distribuições de probabilidade das duas populações consideradas.

Material e Métodos

Para se avaliarem as probabilidades de se cometer o erro tipo I e o erro tipo II, duas populações foram geradas através de simulação computacional. Para a geração dessas populações, foram considerados os modelos probabilísticos normal, com média igual a zero e variância igual a 1, em ambas as populações (situação ideal), exponencial, lognormal e Weibull. Para cada uma dessas situações, as populações foram geradas com médias iguais e com médias diferentes. Para este último caso, foram consideradas as diferenças padronizadas entre as médias populacionais de k=1, 2, 3 e 4 erros-padrão, definidos em função dos momentos de primeira ordem (média) e de segunda ordem (variância) de cada modelo probabilístico e dos valores dos parâmetros desses modelos especificados para cada população. Os parâmetros desse modelo foram especificados de tal maneira a atender um coeficiente de variação (CV) dado pela razão do desvio padrão da distribuição adotada pela média geral do modelo considerado. O valor considerado de CV foi 10%.

Para cada uma das diferentes situações, foram retiradas 2.000 amostras independentes de cada população, sendo aplicado o teste ***t*** para a hipótese da igualdade entre as médias populacionais, H_0: $\mu_1=\mu_2$. Diferentes tamanhos de amostras foram utilizados de ambas as populações. O teste ***t*** foi realizado através das aproximações de Satterthwaite (1946) ou ***t*** de Student "clássico" e $\upsilon = min\,(\,n_1 - 1,\; n_2 - 1)$ para os graus de liberdade e da abordagem de bootstrap proposta por Manly (1998). O número de erros cometidos nas 2.000 repetições foi computado para os níveis de significância de 5% e de 1%.

Para simular dados de cada população (X_{ij}, i=1, 2 e j=1, 2, ..., n_i) com média μ e variância σ^2, estipulados conforme a situação, foi usado um algoritmo em Pascal para inversão da função de distribuição dos modelos normal, exponencial, weibull e lognormal, conforme procedimentos relatado por Dachs (1988), baseado no teorema da probabilidade integral. O modelo linear geral adotado é:

$$X_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, 2 \;\; e \;\; j = 1, 2, ..., n_i$$

em que X_{ij} é o valor da i-ésima população na j-ésima unidade amostral; μ é uma constante geral arbitrária estipulada como 100, sem perda de generalidade; τ_i é o efeito da i-ésima população; $\mu_i = \mu + \tau_i$ é a média geral da i-ésima população e ε_{ij} é o efeito do erro experimental associado à observação X_{ij}, com distribuição dada de acordo com os modelos probabilísticos mencionados anteriormente.

Baseado no teorema da probabilidade integral, foi gerada uma variável U ~ U (0,1) (uniforme 0-1) e, sendo F a função de distribuição qualquer, então o valor do efeito do erro experimental foi obtido por:

$\varepsilon_{ij} = F^{-1}$ (U)

o qual tem função de densidade de probabilidade f(ε), em que f(ε) é a densidade do modelo de probabilidade adotado, primeira derivada de F(ε) (Morgan, 1995).

A função de distribuição normal dos dados da população i é adotada por:

$$F(\varepsilon) = \int_{-\infty}^{\varepsilon} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(t-\mu_i)^2}{2\sigma_i^2}} dt .$$

A rotina para inversão dessa função de distribuição, igualando F a um número uniforme u utilizado, foi aquela apresentada em Dachs (1988). Os parâmetros de ambas as populações foram considerados iguais entre si, ou seja, $\mu_1 = \mu_2 = 0$ e $\sigma_1^2 = \sigma_2^2 = 1$ (situação ideal) para o estudo das probabilidades de erro tipo I e $\sigma_1^2 = \sigma_2^2 = 1$ com $\mu_1 = 0$ e $\mu_2 = k\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ (situação ideal) para o estudo do poder dos testes, sendo k=1, 2, 3 e 4.

Para a distribuição exponencial, foi usado o mesmo teorema. Nesse caso, para se gerar um valor $\varepsilon_{ij,}$ foi usada diretamente a seguinte relação:

$$\varepsilon_{ij} = -\frac{1}{\alpha}\ln(1 - u_{ij})$$

em que α é o parâmetro da distribuição exponencial, u_{ij} é o valor aleatório da distribuição uniforme para a i-ésima população e j-ésima a unidade amostral. A distribuição exponencial tem média e variância dadas por:

$$\mu(\varepsilon) = \frac{1}{\alpha} \text{ e } \sigma_\varepsilon^2 = \frac{1}{\alpha^2} .$$

Para simular os dados de ambas as populações, foi considerado inicialmente o valor de α=1/10 com a finalidade de se avaliar a probabilidade de erro tipo I sob essa distribuição.

Em uma segunda etapa, foi considerado $\mu_1 = \frac{1}{\alpha} + \mu$ e $\mu_2 = \mu_1 + k\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$, $\sigma_1^2 = \sigma_2^2 = \frac{1}{\alpha^2}$, tomando-se valores de k=1, 2, 3 e 4 erro(s)-padrão entre as duas médias populacionais. Nessa situação, foram simulados casos semelhantes àqueles sob normalidade, considerada situação ideal, para o estudo do poder do teste.

Para se gerarem dados de uma distribuição lognormal, foi considerada uma variável Y com distribuição normal com média α e variância β^2. Em seguida, foi implementada a transformação $\varepsilon = e^y$, então ε tem distribuição lognormal dada por:

$$f(\varepsilon) = \frac{1}{\varepsilon\beta\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(\varepsilon)-\alpha}{\beta}\right)^2} \quad , \quad \varepsilon > 0$$

cuja média e variância são dadas por:

$$\mu(\varepsilon) = e^{\alpha + \frac{\beta^2}{2}} \text{ e } \sigma_\varepsilon^2 = \mu^2(\varepsilon)\left(e^{\beta^2} - 1\right).$$

Foram fixados os valores de α e β^2, para se ter variâncias de acordo com os coeficientes de variação estipulados anteriormente. As duas situações, de igualdade das médias populacionais e de não-igualdade considerando diferenças consecutivas de k=1 ou 4 erro(s)-padrão, foram consideradas.

Finalmente, o modelo Weibull de um parâmetro foi adotado, cuja densidade é dada por:

$$f(\varepsilon) = \beta\varepsilon^{\beta-1}e^{-\varepsilon^\beta} \quad , \quad \varepsilon \geq 0 \text{ e } \beta > 0 .$$

A geração de números aleatórios dessa distribuição foi feita da seguinte forma:

$$\varepsilon_{ij} = \sqrt[\beta]{-\ln(1 - u_{ij})}$$

em que β foi fixado em 2 e consideradas as situações de médias iguais e de médias diferentes por uma quantidade dada por k=1, 2, 3 e 4 erro(s)-padrão, sabendo que a média e a variância dessa população são dadas por:

$$\mu_\varepsilon = \Gamma\left(\frac{\beta+1}{\beta}\right) \text{ e } \sigma_\varepsilon^2 = \Gamma\left(\frac{\beta+2}{\beta}\right) - \Gamma^2\left(\frac{\beta+1}{\beta}\right).$$

Foram considerados tamanhos amostrais combinando valores de n_1 extraídos da população 1 com valores de n_2 da população 2. Os valores considerados foram: n_1=5, n_2=5; n_1=5, n_2=20;

$n_1=5$, $n_2=60$; $n_1=20$, $n_2=20$; $n_1=60$, $n_2=60$. Portanto foram simuladas 6 combinações de tamanhos amostrais e consideradas 4 distribuições de probabilidade, totalizando 24 situações. Considerando as condições de médias iguais e diferentes e as 2.000 repetições de cada simulação, bem como as análises de bootstrap que foram repetidas 2.000 vezes, o número total de simulações atingiu a ordem de bilhões de análises.

Resultados e discussão

Um teste ideal é aquele que apresenta probabilidades de erro tipo I semelhantes ao valor nominal adotado e o poder superior ou igual ao dos testes concorrentes. Embora a expectativa de um experimentador fosse que esse teste apresentasse probabilidades de erro tipo I nulas e poder do teste de 100%. Um teste como esse não existe, devendo as comparações entre os testes avaliados se restringirem à descrição de um teste ideal, apresentada anteriormente.

Na Tabela 1, estão apresentadas as probabilidades de erro tipo I (k=0) para as três opções do teste, para as quatro distribuições e para os dois valores nominais de significância adotados em função de diferentes combinações dos tamanhos das amostras (n_1 e n_2). Sob distribuição normal, as probabilidades de erro tipo I obtidas por simulação Monte Carlo para todas as alternativas dos testes podem ser consideradas iguais aos valores nominais de significância, conforme preconizado pela teoria. Essa conclusão é obtida pela comparação entre os valores simulados e o intervalo de confiança para proporções (Leemis e Trivedi, 1996), cujos intervalos de 99% de confiança para o valor nominal de 5% é [3,82%; 6,39%] e para 1% é [0,52%; 1,73%]. O fato de as alternativas apresentarem valores de probabilidades de erro tipo I concordantes com o valor nominal indica que elas seriam equivalentes quanto ao controle dessa probabilidade de erro sob situação ideal. As demais distribuições são consideradas individualmente, a seguir.

No entanto, Silva e Ferreira (2002) relatam que a opção t de Student com graus de liberdade $\upsilon = min(n_1 - 1,\ n_2 - 1)$ apresenta deficiências no controle da probabilidade de erro tipo I sob heterocedastia. As situações de amostras de tamanhos diferentes ($n_1 \neq n_2$) não afetam os resultados para as probabilidades de erro tipo I dos três testes, sob homocedastia e diferentes modelos probabilísticos para o resíduo.

Para a distribuição exponencial, o teste de bootstrap foi o único a apresentar probabilidades de erros tipo I semelhantes às nominais em todas as situações (Tabela 1), tanto com $\alpha=5\%$ como para $\alpha=1\%$. Os testes com graus de liberdade ajustados por Satterthwaite ou dados por $\upsilon = min(n_1 - 1,\ n_2 - 1)$ não controlam o erro tipo I em várias circunstâncias. A opção de Satterthwaite mostrou controle do erro tipo I para amostras de tamanhos iguais e superiores a 20 com $\alpha=5\%$. Para $\alpha=1\%$ esse teste apresentou problemas somente com amostras de tamanhos diferentes ($n_1=5$ e $n_2=20$). A opção do teste t com $\upsilon = min(n_1 - 1,\ n_2 - 1)$ para $\alpha=5\%$ controlou as probabilidades de erro tipo I com grandes amostras e iguais $n_1=60$ e $n_2=60$ e para amostras de tamanhos diferentes ($n_1=5$ e $n_2=20$). Para $\alpha=1\%$, o problema de não controle do erro tipo I só se manifestou em amostras de tamanhos iguais e pequenos ($n_1=10$ e $n_2=10$), embora com $n_1=5$ e $n_2=5$ tenha apresentado probabilidades similares a nominal.

Tabela 1. Probabilidades de erro tipo I (k=0) para as três opções do teste, para as quatro distribuições e para os dois valores nominais de significância adotados em função de diferentes combinações dos tamanhos amostrais (n_1 e n_2)

	5%			1%		
	t (Satt)	t (Glmin)	t (Boot)	t (Satt)	t (Glmin)	t (Boot)
	$n_1=5$ e $n_2=20$					
Normal	5,2500	4,8000	5,1000	1,2000	0,8500	1,1000
Exponencial	7,7000	4,5500	5,4000	2,9000	0,6000	1,0500
LogNormal	7,0000	1,7000	3,1500	0,9000	0,0500	0,2500
Weibull	5,3000	4,9500	5,0000	1,4500	0,7000	0,9000
	$n_1=5$ e $n_2=5$					
Normal	4,9000	4,9000	5,2500	0,9000	0,9000	0,9500
Exponencial	3,7000	3,6500	5,4000	0,6500	0,6500	0,8000
LogNormal	0,8500	0,8500	2,4500	0,0500	0,0500	0,2500
Weibull	4,5000	4,4500	4,9500	1,0000	1,0000	0,9500
	$n_1=10$ e $n_2=10$					
Normal	5,4000	5,4000	5,4500	1,2000	1,1500	1,2000
Exponencial	3,6000	3,1500	4,4500	0,7500	0,5000	0,9500
LogNormal	1,6000	1,5000	3,5000	0,1000	0,1000	0,3500
Weibull	4,8500	4,8000	5,0500	0,8000	0,8000	0,9500
	$n_1=20$ e $n_2=20$					
Normal	5,4000	5,4000	5,3500	1,2000	1,2000	1,1500
Exponencial	3,9500	3,7000	4,3500	0,6500	0,6500	0,8000
LogNormal	2,1000	1,9500	3,4500	0,1500	0,1500	0,5000
Weibull	4,1500	4,1000	4,2000	0,9000	0,9000	1,0000
	$n_1=60$ e $n_2=60$					
Normal	5,4500	5,4500	5,5500	1,1000	1,1000	1,1000
Exponencial	4,2000	4,1000	4,3000	0,7500	0,7500	0,9000
LogNormal	3,0500	3,0000	3,9000	0,2000	0,2000	0,6000
Weibull	5,7000	5,7000	5,6000	1,0500	1,0000	1,1000

Para a distribuição lognormal com $\alpha=5\%$, somente o t de bootstrap com grandes amostras de tamanhos iguais ($n_1=n_2=60$) é que apresentou controle do erro tipo I. De forma geral, houve uma tendência dos testes serem rigorosos, ou seja, apresentaram probabilidades de erro tipo I inferiores ao valor nominal de 5%. Para $\alpha=1\%$, resultados semelhantes aos de $\alpha=5\%$ foram observados, excetuando-se a opção de Satterthwaite com $n_1=5$ e $n_2=20$.

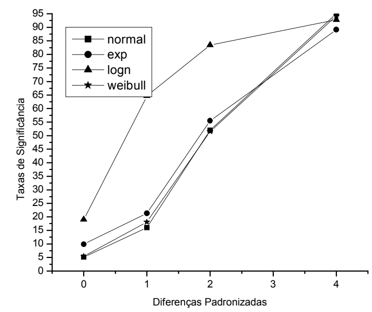
Para a distribuição Weibull, observou-se que as probabilidades de erro tipo I dos três testes são semelhantes ao valor nominal tanto para 5% como para 1%. Esse fato talvez possa ser explicado pela escolha do parâmetro β igual a 2. A distribuição obtida com essa escolha, apesar de ser assimétrica à direita, está bem próxima da simetria, aproximando-se do modelo normal.

Na Figura 1, encontram-se as comparações entre as distribuições para cada teste em função das diferenças entre as médias para $\alpha=5\%$. Quando o k for igual a zero, o valor da ordenada refere-se às probabilidades de erro tipo I e para todos os valores de k, ao poder do teste. Para k=0, o teste de Satterthwaite não controlou as probabilidades de erro tipo I para as distribuições exponencial e lognormal, apresentando probabilidades de erro superiores ao valor nominal de 5%. O mesmo resultado foi observado para o t com $\upsilon = min\,(\,n_1-1,\ n_2-1)$. O teste t de bootstrap só apresentou problemas com a distribuição lognormal e, ao contrário de seus concorrentes, subestimou o valor nominal de significância.

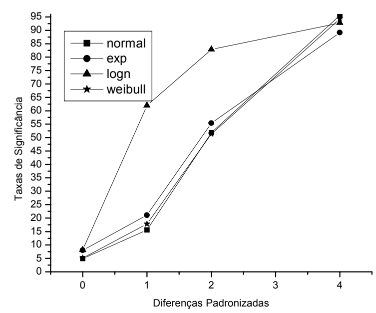
Na Figura 2, destacou-se a comparação dos testes para duas distribuições selecionadas: normal e lognormal (Figura 2, a e b) em função do valor de k. Para k=0 e sob normalidade (Figura 2a), os três testes apresentaram controle do erro tipo I. Para k=0 e distribuição lognormal (Figura 2b), nenhum dos três testes apresentou controle das probabilidades de erro tipo I, sendo que os t Satterthwaite e $\upsilon = min\,(\,n_1-1,\ n_2-1)$ superestimaram o valor nominal, e a opção de bootstrap o subestimou.

Na Tabela 2, está apresentado o poder dos testes para k=1 e k=4, em relação a diferentes tamanhos amostrais e para todas as distribuições estudadas. A distribuição lognormal foi, em geral, mais poderosa que as demais distribuições, principalmente com k=1. Para k=4, as distribuições tenderam a se igualarem em poder dos testes, tanto para $\alpha=1\%$ quanto para $\alpha=5\%$.

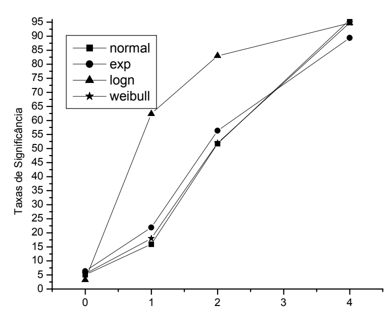
Para melhor avaliar a comparação dos testes, buscou-se apresentá-los em um único gráfico para o caso da normal e da lognormal (Figura 2, a e b). Para o caso normal, situação ideal teórica do t, os três testes não diferiram entre si com relação à probabilidade de erro tipo I (k=0) e poder do teste (k>0). Esse resultado reforça a idéia da não necessidade da recomendação do bootstrap, uma vez que o esforço computacional é intenso. Para a lognormal (Figura 2, b), o erro tipo I (k=0) não foi controlado para o teste t critério de Satterthwaite. Para o poder do teste (k>0), verifica-se que esse critério apresentou valores maiores com k<2 e tendeu a se igualar aos demais para k>2. Esse resultado não pode servir de base para sua recomendação, uma vez que não controlou a probabilidade de erro tipo I. Então, os critérios de t com graus de liberdade mínimos e bootstrap poderiam ser considerados robustos, pois controlaram a probabilidade de erro tipo I e apresentaram poder do teste praticamente igual sob distribuição não-normal.



(a) Satterthwaite

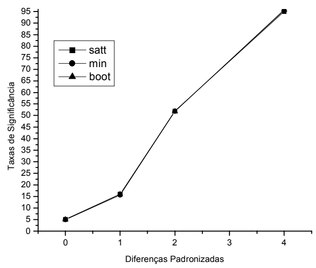


(b) GL mínimo

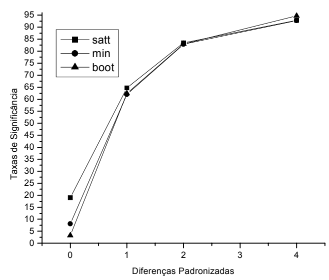


(c) bootstrap

Figura 1. Proporções de significância para o valor nominal de 5% dos testes de t com correção de graus de liberdade corrigidos por (a) Satterthwaite, (b) $\nu = \min\ (\,n_1-1,\ n_2-1)$ e (c) teste t de bootstrap em função das distribuições e das diferenças padronizadas entre as médias ($n_1=5$ e $n_2=60$)



(a) normal



(b) lognormal

Figura 2. Proporções de significância para o valor nominal de 5% para as distribuições (a) normal, (b) lognormal em função das diferenças padronizadas entre as médias populacionais e dos diferentes testes (n_1=5 e n_2=60)

Tabela 2. Poder dos testes com k=1 e k=4 das três opções do teste em relação às quatro distribuições e para os dois valores nominais de significância adotados em função de diferentes combinações dos tamanhos amostrais (n_1 e n_2)

		n_1=5 e n_2=5					
		5%			1%		
		t (Satt)	t (Glmin)	t (Boot)	t (Satt)	t (Glmin)	t (Boot)
	Normal	14,7500	14,6000	15,1000	4,2500	4,2500	4,0500
K=1	Exponencial	17,3500	17,1000	19,9500	4,5000	4,4000	5,3000
	LogNormal	61,5000	61,1500	66,3500	42,8000	41,9000	49,0000
	Weibull	13,1500	13,1000	13,8000	3,7500	3,7500	3,7000
	Normal	93,7000	93,6500	93,6000	15,6000	15,3500	15,0500
K=4	Exponencial	93,1500	93,0500	93,3000	70,7000	70,3000	69,6500
	LogNormal	94,0500	94,0000	95,9000	87,5500	87,2500	90,7500
	Weibull	93,1000	92,9000	93,2500	71,4500	71,0500	70,3500
		n_1=20 e n_2=20					
	Normal	17,3000	17,2000	17,0500	5,4000	5,3500	5,3000
K=1	Exponencial	16,2500	16,0000	16,9000	5,0500	4,6500	6,4000
	LogNormal	47,9000	47,5000	51,6500	32,8500	32,1500	38,2000
	Weibull	17,4000	17,3000	17,5000	5,6000	5,4500	5,7500
	Normal	97,5000	97,4500	97,4500	88,1500	88,1500	88,8000
K=4	Exponencial	95,6500	95,4500	95,9500	87,5500	87,2000	88,1500
	LogNormal	95,4000	95,4000	96,2500	91,6000	91,3500	93,6500
	Weibull	96,3500	96,3500	96,4000	88,6500	88,4500	88,7000
		n_1=5 e n_2=20					
	Normal	15,5000	14,6500	14,6500	5,0500	4,4500	5,1000
K=1	Exponencial	22,4000	21,7000	22,8000	11,2500	5,9000	10,3000
	LogNormal	64,8500	60,5500	65,9500	51,7500	40,5000	52,6000
	Weibull	17,7500	17,6000	17,6000	5,3000	4,3500	5,1500
	Normal	96,4500	96,4000	96,9000	86,1500	85,9000	86,5000
K=4	Exponencial	88,4500	88,4000	89,8500	82,2500	82,0000	82,5500
	LogNormal	93,2500	92,8500	93,9500	87,9000	85,6500	91,2500
	Weibull	94,9500	94,8000	95,5000	85,7500	85,6500	86,0000

Conclusão

Sob distribuições não-normais, o t com ajuste de graus de liberdade por Satterthwaite e t com $\upsilon = min\,(n_1 - 1,\ n_2 - 1)$ não controlaram as probabilidades de erro tipo I; o critério de bootstrap controlou as probabilidades de erro tipo I e apresentou poder do teste equivalente, sendo considerado robusto com a violação da pressuposição de normalidade; o teste de t com graus de liberdade ajustado (Satterthwaite) em populações não-normais e amostras de tamanhos diferentes apresentou probabilidades de erro tipo I maiores que as nominais.

Referências

BONEAU, C.A. The effects of violations of assumptions underlying t test. *Psychol. Bull.*, Washington, DC., v.57, n.1, p.49-64, 1960.

BOX, G.E.P. Non-normality and tests on variances, *Biometrika*, London, v.40, n.3-4, p.318-335, 1953.

DACHS, J.N.W. *Estatística computacional.* Uma introdução ao Turbo Pascal. Rio de Janeiro: Livros Técnicos e Científicos, 1988, 236p.

GLASS, G.V. *et al.* Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Rev. Educ. Res.*, Washington, DC., v.42, n.3, p.237-288, 1972.

LEEMIS, L.M.; TRIVEDI, K.S. A comparison of approximate interval estimators for the Bernoulli parameter. *Am. Stat. Assoc.*, Alexandria, v.50, n.1, p.63-68, 1996.

MANLY, B.F.J. *Randomization, Bootstrap and Monte Carlo methods in Biology*, 2. ed. London: Chapman & Hall, 1998.

MORGAN, B.J.T. *Elements of simulation*, London: Chapman & Hall, 1995.

POSTEN, H.O. et al. Robustness of the two sample t-test under violations of the homogeneity of variance assumption.*Communic. Statistic -Theor. Meth.*, New York, p.109-126, 1982.

SATTERTHWAITE, F.E. An approximate distribution of estimates of variance components. *Biometric Bull.*, London, v. 2, n.6, p.110-114, 1946.

SILVA, R.B.V.; FERREIRA, D.F. Avaliação de uma alternativa para o teste t com variâncias heterogêneas por simulação. *Cienc. Agrotecnol.*, Lavras, 2002 (prelo).

SNEDECOR, G.W.; COCHRAN, W.G. *Statistical methods*, 7. ed., Ames: The Iowa State University Press, 1980.

SRIVASTAVA, A.B.L. Effect of non-normality on the power function of the t-test. *Biometrika*, London, v.45, n.3-4, p.421-429, 1958.

Received on June 14, 2002.
Accepted on November 20, 2002.