

Um estudo sobre o tamanho e poder dos testes *t-Student* e *Wilcoxon*

Emílio Augusto Coelho Barros* e Josmar Mazucheli

Departamento de Estatística, Universidade Estadual de Maringá, Av. Colombo, 5790 – Bl. F-67, 87020-900, Maringá, Paraná, Brasil. *Autor para correspondência. e-mail: emilioaugusto@pop.com.br

RESUMO. Um dos problemas mais comuns em estatística consiste em testar a hipótese $H_0: \mu = \mu_0$ versus a alternativa $H_1: \mu \neq \mu_0$, em que μ_0 é algum valor específico do parâmetro μ . A partir de uma amostra aleatória, sob a suposição de que a mesma é proveniente de uma distribuição Normal com média μ e desvio padrão σ , ambos desconhecidos, pode-se aplicar o bem conhecido teste *t-Student*. Como alternativa, sob a suposição de simetria da distribuição dos dados, pode-se utilizar o teste não-paramétrico conhecido na literatura como teste de *Wilcoxon* (Conover, 1971). Neste artigo, é conduzido um estudo de simulação Monte-Carlo com o intuito de avaliar o tamanho e o poder dos testes *t-Student* e de *Wilcoxon* sob diferentes instâncias. Para o cálculo do tamanho de ambos os testes, foram realizadas $B = 100.000$ simulações Monte-Carlo com 10 diferentes tamanhos de amostras, $n = 10, 20, \dots, 90, 100$. Cada uma das $B = 100.000$ amostras foram geradas das distribuições Normal, Laplace, Uniforme, *t-Student* e Logística sob a hipótese nula, sem perda de generalizades, com $\mu = 0$. Para o cálculo do poder, novamente $B = 100.000$ amostras foram geradas sob a hipótese alternativa com $\mu_0 = -1.0, -0.9, \dots, 0.9, 1.0$.

Palavras-chave: tamanho do teste, poder do teste, simulação Monte-Carlo, teste *t-Student*, teste de *Wilcoxon*.

ABSTRACT. A study about the size and power of *t-Student* and *Wilcoxon* tests. A typical problem in statistics data analysis consists in testing a null hypothesis $H_0: \mu = \mu_0$ versus an alternative $H_1: \mu \neq \mu_0$, where μ_0 is some specific value of the true parameter μ . From a random sample with normal distribution with mean μ and variance σ^2 , both unknown, we might use the well known t-student test. As an alternative to the usual t-student test, under the symmetric supposition, we can use the nonparametric *Wilcoxon on* test (Conover, 1971). In this paper, a Monte-Carlo simulation study was conducted to calculate the empirical size and power of the t-student and *Wilcoxon* tests. In this study, several instances were considered. For the size of both tests, we considered sample sizes equal to $n = 10, 20, \dots, 90, 100$ simulated from Normal, Laplace, Uniform, *t-Student* and Logistic distributions, under the null hypothesis with $\mu = 0$. In the power study, from the same sample sizes and for all distributions, random samples were simulated from the alternative hypothesis considering $\mu_0 = -1.0, -0.9, \dots, 0.9, 1.0$.

Key words: empirical size, power of tests, Monte-Carlo simulation, *t-Student* test, *Wilcoxon* test.

Introdução

Seja (x_1, \dots, x_n) uma amostra aleatória proveniente de uma população com distribuição simétrica escrita na forma $f\left(\frac{x-\mu}{\sigma}\right)$ (Casella e Berger, 1990), com parâmetro de locação μ e escala σ , ambos desconhecidos. Suponha o interesse em testar a hipótese $H_0: \mu = \mu_0$ versus a alternativa $H_1: \mu \neq \mu_0$, em que μ_0 é algum valor especificado do parâmetro μ . Sob a suposição de que (x_1, \dots, x_n) é proveniente de uma distribuição Normal com média μ e desvio padrão

σ , $X \sim N(\mu, \sigma)$, é bem conhecida que a estatística definida na forma:

$$T = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s},$$

tem distribuição *t-Student* com $(n-1)$ graus de liberdade; em que $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ e $s = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2}$ são, respectivamente, a média e o desvio padrão amostral. Para hipóteses alternativas da forma $H_1: \mu \neq \mu_0$, rejeita-se $H_0: \mu = \mu_0$, em nível de significância α , ($0 < \alpha < 1$), se $|T| > t_{(n-1; \alpha/2)}$, em que $t_{(n-1; \alpha/2)}$ é o $100 \times \alpha/2$ percentil

da distribuição *t-Student* com $(n-1)$ graus de liberdade. Quando não é possível garantir a suposição de normalidade, na presença de amostras suficientemente grandes, o Teorema Central do Limite garante que se os dados são provenientes de uma população com média μ e variância finita σ^2 , então a variável aleatória T pode ser aproximada por uma distribuição normal para $n \rightarrow \infty$ (ver, por exemplo, Ferguson, 1967; Graybill, 1974; Casella e Berger, 1990). Na ausência da suposição de normalidade e assumindo que os dados são provenientes de uma distribuição simétrica, uma alternativa ao teste *t-Student* é o teste de *Wilcoxon* (Conover, 1971). O teste de *Wilcoxon* é baseado nos postos de (x_1, \dots, x_n) , cuja estatística é escrita na forma:

$$W = \sum_{i=1}^{n^+} R_i^+ - \frac{n_t(n_t+1)}{4},$$

em que R_i^+ é o posto de $|x_i - \mu_0|$ para $x_i \neq \mu_0$, n_t e n^+ são, respectivamente, o número de observações em que $x_i \neq \mu_0$ e $(x_i - \mu_0) > 0$. Na presença de empates, a média dos postos é utilizada. Por exemplo, se $x_1 = x_2$ e o posto de $|x_1 - \mu_0| = 1$ e o posto de $|x_2 - \mu_0| = 2$, então os postos dessas diferenças são $R_1^+ = 1.5$ e $R_2^+ = 1.5$, respectivamente. Para $n_t \leq 20$, a distribuição amostral de w é obtida a partir da enumeração de todas as possíveis amostras sob H_0 (Conover, 1971; Hollander e Wolfe, 1973). Para $n_t > 20$, a estatística:

$$W^* = \frac{W\sqrt{n_t-1}}{\sqrt{n_t V - W^2}},$$

tem aproximadamente distribuição *t-Student*, com $(n-1)$ graus de liberdade, em que

$$V = \frac{1}{24} n_t(n_t+1)(2n_t+1) - 0.5 \sum_{i=1}^{n^-} t_i(t_i+1)(t_i-1),$$

n^- é o número de grupos com empates e t_i é o número de empates no i -ésimo grupo (Iman, 1974).

Em problemas reais, quando não é razoável assumir que os dados são provenientes de uma distribuição simétrica ou normal, pode-se recorrer ao teste do sinal como alternativa aos testes *Wilcoxon* e *t-Student*. Vale lembrar que na prática, em geral, é mais fácil garantir a simetria do que a normalidade, o que torna o teste de *Wilcoxon* uma boa alternativa não paramétrica ao teste *t-Student* (Conover, 1971).

Neste artigo, um estudo de simulação Monte-Carlo avalia o comportamento dos testes *t-Student* e *Wilcoxon* com relação ao tamanho e poder. Para o cálculo do tamanho empírico de ambos os testes, para várias distribuições, valores pseudo-aleatórios são gerados sob a hipótese nula com $\mu = 0$. Na avalia-

ção do poder, são gerados valores sob a hipótese alternativa atribuindo-se valores a μ , que variam entre -1 e 1, com passo igual a 0.1.

Na Seção “Estudo de Simulação”, são discutidas as diferentes instâncias avaliadas no estudo de simulação. A Seção “Tamanho dos Testes” avalia o comportamento de ambos os testes com relação ao tamanho dos mesmos. O poder é avaliado na Seção “Poder dos Testes”. A Seção “Conclusões” finaliza este artigo com algumas conclusões.

Estudo de simulação

O poder de um teste estatístico é definido como a probabilidade de rejeitar a hipótese nula, dado que a mesma é falsa. Na prática, é importante que se tenham testes com nível de significância próximos do nível de significância nominal e que o poder seja alto, mesmo em situações de amostras pequenas. Na suposição de amostras provenientes de uma distribuição Normal, hipóteses da forma $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$ podem ser avaliadas a partir da estatística *t-Student*, como descrito na Seção “Introdução”. Alternativamente, na suposição de simetria, pode-se utilizar o teste não-paramétrico de *Wilcoxon*. Para ambos, os testes, quando essas suposições estão satisfeitas, espera-se que o erro Tipo I seja próximo do nível de significância nominal e que o poder seja alto. Por outro lado, quando ocorre a quebra de alguma dessas suposições, é importante avaliar o comportamento da taxa de rejeição de H_0 quando a mesma é verdadeira. Essa taxa de rejeição é definida como o tamanho empírico do teste e pode ser calculada via simulação Monte-Carlo, gerando-se amostras sob a hipótese nula. Por outro lado, gerando-se valores sob a hipótese alternativa, tem-se que a proporção de vezes em que H_0 é rejeitada define o poder de um teste estatístico.

Para os testes *t-Student* e *Wilcoxon*, o tamanho e o poder dos mesmos foram calculados considerando-se a distribuição Normal, Laplace, Uniforme, *t-Student* e Logística (Tabela 1).

Para todas as instâncias descritas a seguir, foram adotados níveis de significância nominais $\alpha = 0.01$ e $\alpha = 0.05$. As simulações foram conduzidas utilizando macros desenvolvidas no software SAS, nos quais os valores das estatísticas *t-Student* e *Wilcoxon* foram calculadas utilizando o procedimento *univariate*.

Tamanho dos testes

No cálculo dos tamanhos de ambos os testes, *t-Student* e *Wilcoxon*, para cada uma das distribuições apresentadas na Tabela 1, foram geradas sob a hipótese nula, $H_0 : \mu = 0$, $B = 100.000$ amostras de

tamanhos $n = 10, 20, \dots, 90, 100$. Para cada n e cada uma das distribuições apresentadas na Tabela 1, a proporção de vezes, entre as $B = 100.000$ amostras, em que os testes levam a falsa rejeição de H_0 define o tamanho empírico dos testes. Os resultados do estudo de simulação estão apresentados nas Tabelas 2 a 6. A Figura 1 apresenta o comportamento dos tamanhos dos testes em relação aos tamanhos de amostra.

Uma vez que o número de vezes em que H_0 é rejeitada tem distribuição Binomial, a hipótese nula para avaliar se o tamanho empírico é estatisticamente igual ao tamanho nominal foi testada em nível de significância de 5%. Utilizou-se o teste Binomial exato, sendo que esses resultados também são apresentados nas Tabelas 2 a 6.

Tabela 1. Distribuições consideradas no estudo de simulação Monte-Carlo.

| Distribuição | Expressão da Densidade |
|--------------|--|
| Normal | $f(x \mu, \sigma = 1) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} (x - \mu)^2 \right]$ |
| Laplace | $f(x \mu, \sigma = 1) = \frac{1}{2\sigma} \exp \left(- x - \mu \right)$ |
| Uniforme | $f(x \mu) = \frac{1}{2+2\mu}, \quad -1 \leq x \leq 1 + 2\mu$ |
| t-Student | $f(x \mu) = \frac{\Gamma(\frac{2}{5})}{\sqrt{5}\mu \Gamma(2.5)} \left(1 + \frac{(x-\mu)^2}{5} \right)^{-3}$ |
| Logística | $f(x \mu, \sigma = 1) = \frac{\exp(x-\mu)}{[1 + \exp(x-\mu)]^2}$ |

Tabela 2. Taxas de Rejeição de H_0 para $X \sim N(\mu = 0, \sigma = 1)$

| n | $\alpha = 5\%$ | | $\alpha = 1\%$ | |
|-----|----------------|----------|----------------|---------------------|
| | t-Student | Wilcoxon | t-Student | Wilcoxon |
| 10 | 0.0509 | 0.0500 | 0.0108 | 0.0103 |
| 20 | 0.0506 | 0.0495 | 0.0101 | 0.0094 ⁺ |
| 30 | 0.0491 | 0.0495 | 0.0097 | 0.0111 |
| 40 | 0.0499 | 0.0499 | 0.0097 | 0.0104 |
| 50 | 0.0501 | 0.0513 | 0.0098 | 0.0104 |
| 60 | 0.0505 | 0.0507 | 0.0102 | 0.0104 |
| 70 | 0.0496 | 0.0511 | 0.0101 | 0.0107 |
| 80 | 0.0493 | 0.0490 | 0.0098 | 0.0102 |
| 90 | 0.0497 | 0.0499 | 0.0100 | 0.0104 |
| 100 | 0.0494 | 0.0493 | 0.0102 | 0.0101 |

⁺Significante ao nível de 5%.

Tabela 3. Taxas de Rejeição de H_0 para $X \sim Laplace(\mu = 0, \sigma = 1)$

| n | $\alpha = 5\%$ | | $\alpha = 1\%$ | |
|-----|---------------------|----------|---------------------|---------------------|
| | t-Student | Wilcoxon | t-Student | Wilcoxon |
| 10 | 0.0414 ⁺ | 0.0484 | 0.0057 ⁺ | 0.0097 |
| 20 | 0.0467 ⁺ | 0.0496 | 0.0072 ⁺ | 0.0094 |
| 30 | 0.0479 ⁺ | 0.0493 | 0.0084 ⁺ | 0.0120 ⁺ |
| 40 | 0.0487 | 0.0510 | 0.0085 ⁺ | 0.0102 |
| 50 | 0.0490 | 0.0506 | 0.0090 ⁺ | 0.0105 |
| 60 | 0.0479 ⁺ | 0.0502 | 0.0084 ⁺ | 0.0099 |
| 70 | 0.0488 | 0.0511 | 0.0092 | 0.0107 |
| 80 | 0.0491 | 0.0501 | 0.0094 | 0.0104 |
| 90 | 0.0492 | 0.0496 | 0.0090 ⁺ | 0.0103 |
| 100 | 0.0500 | 0.0508 | 0.0092 ⁺ | 0.0101 |

⁺Significante ao nível de 5%.

Tabela 4. Taxas de Rejeição de H_0 para $X \sim U(a = -1, b = 1)$.

| n | $\alpha = 5\%$ | | $\alpha = 1\%$ | |
|-----|---------------------|---------------------|---------------------|---------------------|
| | t-Student | Wilcoxon | t-Student | Wilcoxon |
| 10 | 0.0526 ⁺ | 0.0479 ⁺ | 0.0134 ⁺ | 0.0093 |
| 20 | 0.0527 ⁺ | 0.0495 | 0.0118 ⁺ | 0.0095 |
| 30 | 0.0513 | 0.0499 | 0.0112 ⁺ | 0.0114 ⁺ |
| 40 | 0.0507 | 0.0518 | 0.0108 | 0.0109 ⁺ |
| 50 | 0.0515 | 0.0513 | 0.0109 ⁺ | 0.0107 |
| 60 | 0.0499 | 0.0499 | 0.0100 | 0.0101 |
| 70 | 0.0497 | 0.0498 | 0.0104 | 0.0104 |
| 80 | 0.0505 | 0.0507 | 0.0108 | 0.0110 ⁺ |
| 90 | 0.0497 | 0.0497 | 0.0100 | 0.0100 |
| 100 | 0.0515 | 0.0517 | 0.0106 | 0.0107 |

⁺Significante ao nível de 5%.

Tabela 5. Taxas de rejeição de H_0 para $X \sim t_4$

| n | $\alpha = 5\%$ | | $\alpha = 1\%$ | |
|-----|---------------------|---------------------|---------------------|---------------------|
| | t-Student | Wilcoxon | t-Student | Wilcoxon |
| 10 | 0.0440 ⁺ | 0.0500 | 0.0075 ⁺ | 0.0102 |
| 20 | 0.0461 ⁺ | 0.0482 ⁺ | 0.0078 ⁺ | 0.0095 |
| 30 | 0.0456 ⁺ | 0.0486 | 0.0082 ⁺ | 0.0113 ⁺ |
| 40 | 0.0474 ⁺ | 0.0517 | 0.0081 ⁺ | 0.0108 |
| 50 | 0.0486 | 0.0513 | 0.0087 ⁺ | 0.0109 ⁺ |
| 60 | 0.0479 ⁺ | 0.0496 | 0.0085 ⁺ | 0.0109 ⁺ |
| 70 | 0.0471 ⁺ | 0.0498 | 0.0089 ⁺ | 0.0105 |
| 80 | 0.0477 ⁺ | 0.0505 | 0.0092 ⁺ | 0.0105 |
| 90 | 0.0487 | 0.0496 | 0.0086 ⁺ | 0.0097 |
| 100 | 0.0483 | 0.0509 | 0.0089 ⁺ | 0.0101 |

⁺Significante ao nível de 5%.

Tabela 6. Taxas de Rejeição de H_0 para $X \sim Logística(\mu = 0, \sigma = 1)$

| n | $\alpha = 5\%$ | | $\alpha = 1\%$ | |
|-----|---------------------|---------------------|---------------------|---------------------|
| | t-Student | Wilcoxon | t-Student | Wilcoxon |
| 10 | 0.0455 ⁺ | 0.0479 ⁺ | 0.0087 ⁺ | 0.0093 |
| 20 | 0.0489 | 0.0495 | 0.0095 ⁺ | 0.0095 |
| 30 | 0.0490 | 0.0499 | 0.0095 | 0.0114 ⁺ |
| 40 | 0.0496 | 0.0518 | 0.0095 | 0.0109 ⁺ |
| 50 | 0.0508 | 0.0513 | 0.0098 | 0.0107 |
| 60 | 0.0488 | 0.0499 | 0.0089 ⁺ | 0.0101 |
| 70 | 0.0496 | 0.0498 | 0.0096 | 0.0104 |
| 80 | 0.0497 | 0.0507 | 0.0100 | 0.0110 ⁺ |
| 90 | 0.0495 | 0.0497 | 0.0096 | 0.0100 |
| 100 | 0.0509 | 0.0517 | 0.0103 | 0.0107 |

⁺Significante ao nível de 5%.

Poderdo teste

No cálculo do poder, para ambos os testes e para cada uma das distribuições apresentadas na Tabela 1, também foram geradas $B = 100.000$ amostras de tamanhos $n = 10, 20, \dots, 90, 100$, sob a hipótese alternativa com $\mu = -1.0, -0.9, \dots, 0.9, 1.0$. Em cada uma das $T = 5 \times 10 \times 100.000$ simulações, foram calculados os *p-valores* associados a ambos os testes. A proporção de vezes em que H_0 foi corretamente rejeitada definiu o poder dos testes. Os resultados são apresentados nas Figuras 2 a 6. As curvas internas às figuras, representadas por (-o-), mostram o poder exato para o teste *t-Student*. Para cada uma das distribuições, os níveis de significância adotados foram os próprios tamanhos dos testes apresentados nas Tabelas 2 a 6.

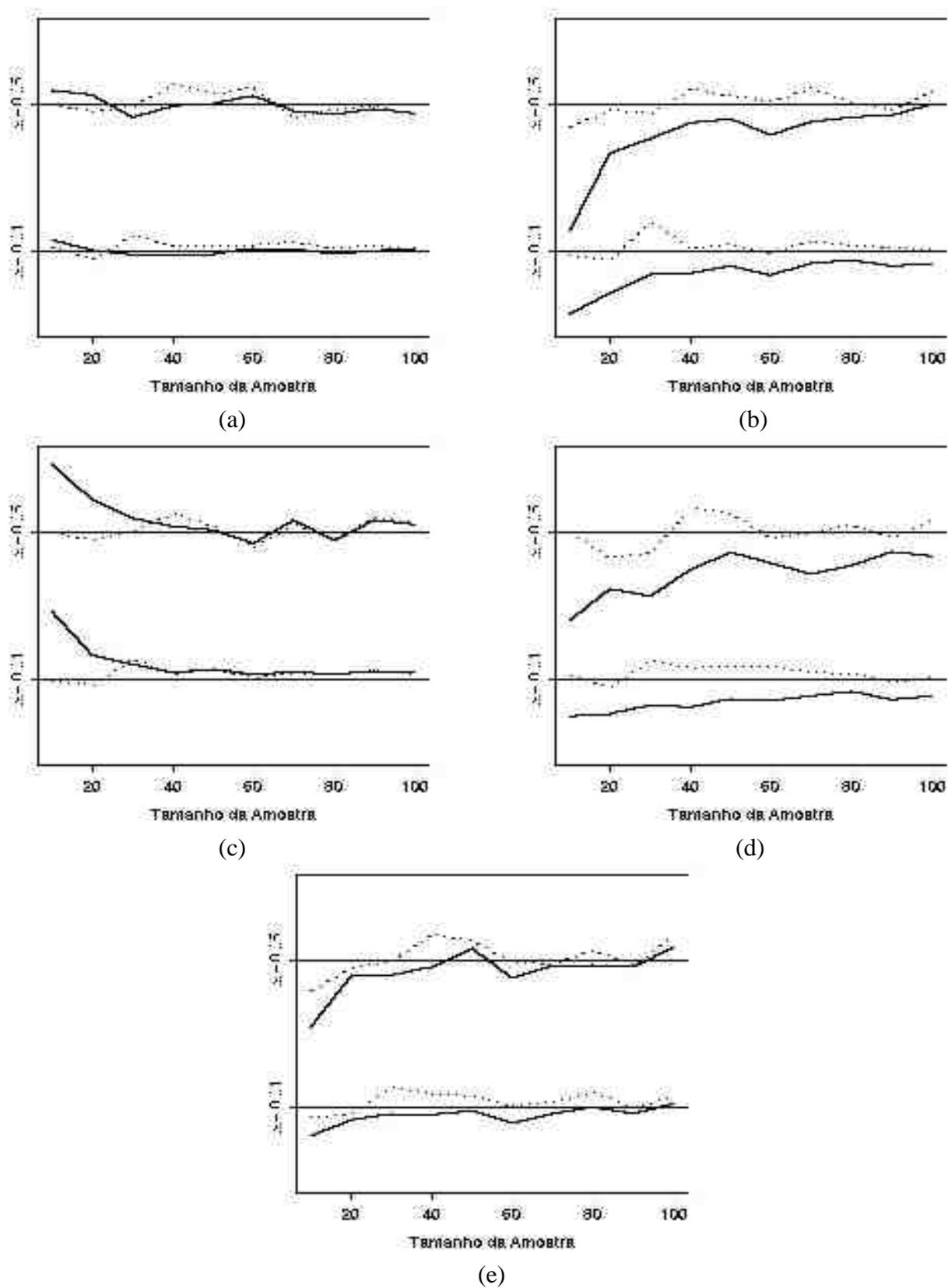


Figura 1. Tamanho dos testes; (—): Teste *t-Student*, (.....): Teste de *Wilcoxon*; (a): Distribuição Normal; (b): Distribuição de Laplace; (c): Distribuição Uniforme; (d): Distribuição *t-Student*; (e): Distribuição Logística.

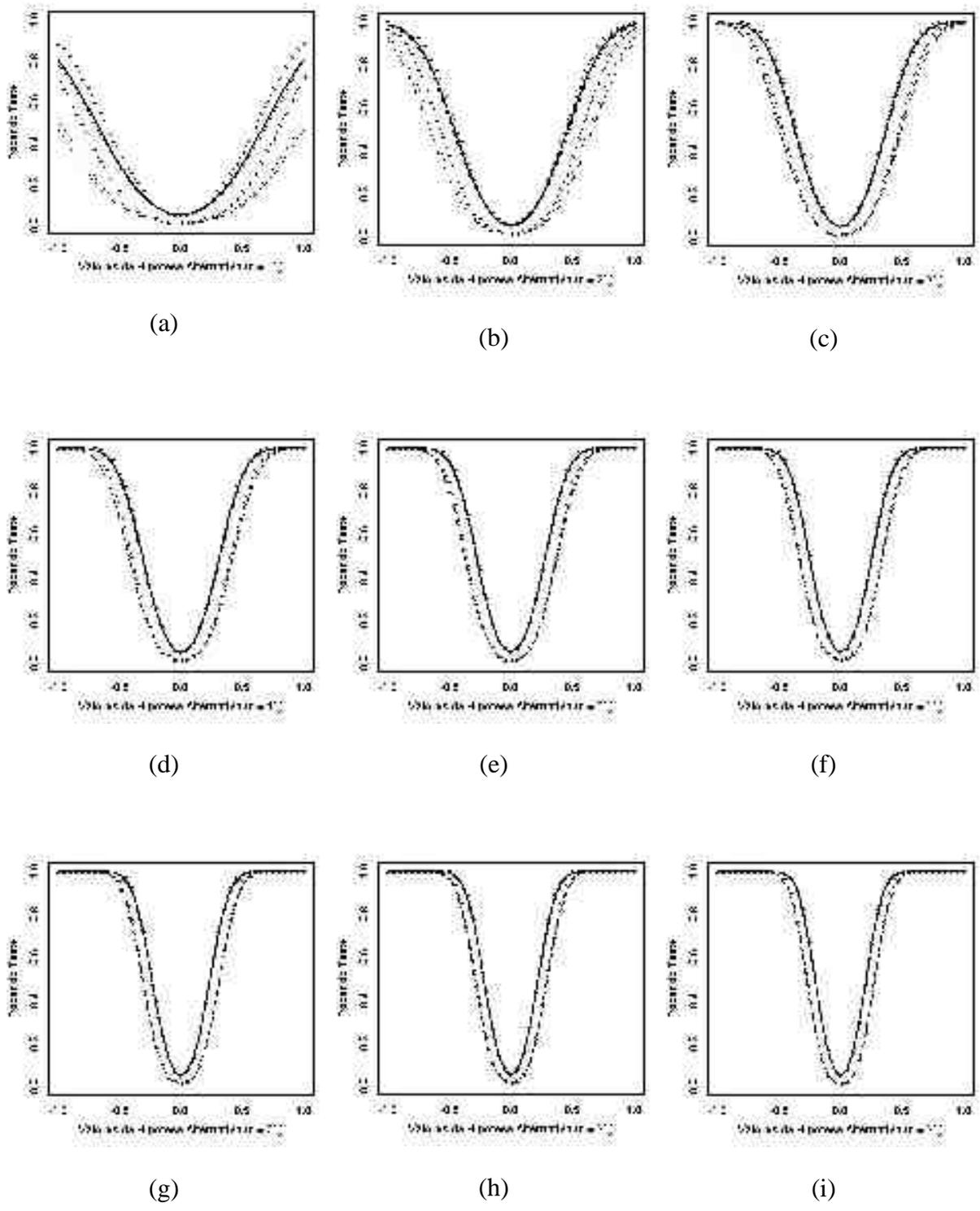


Figura 2. Valores da hipótese alternativa versus o poder dos testes para a distribuição Normal; $\alpha = 0.05$: (—) *t-Student*, (---) *Wilcoxon*; $\alpha = 0.01$: (· ·) *t-Student*, (---) *Wilcoxon*.

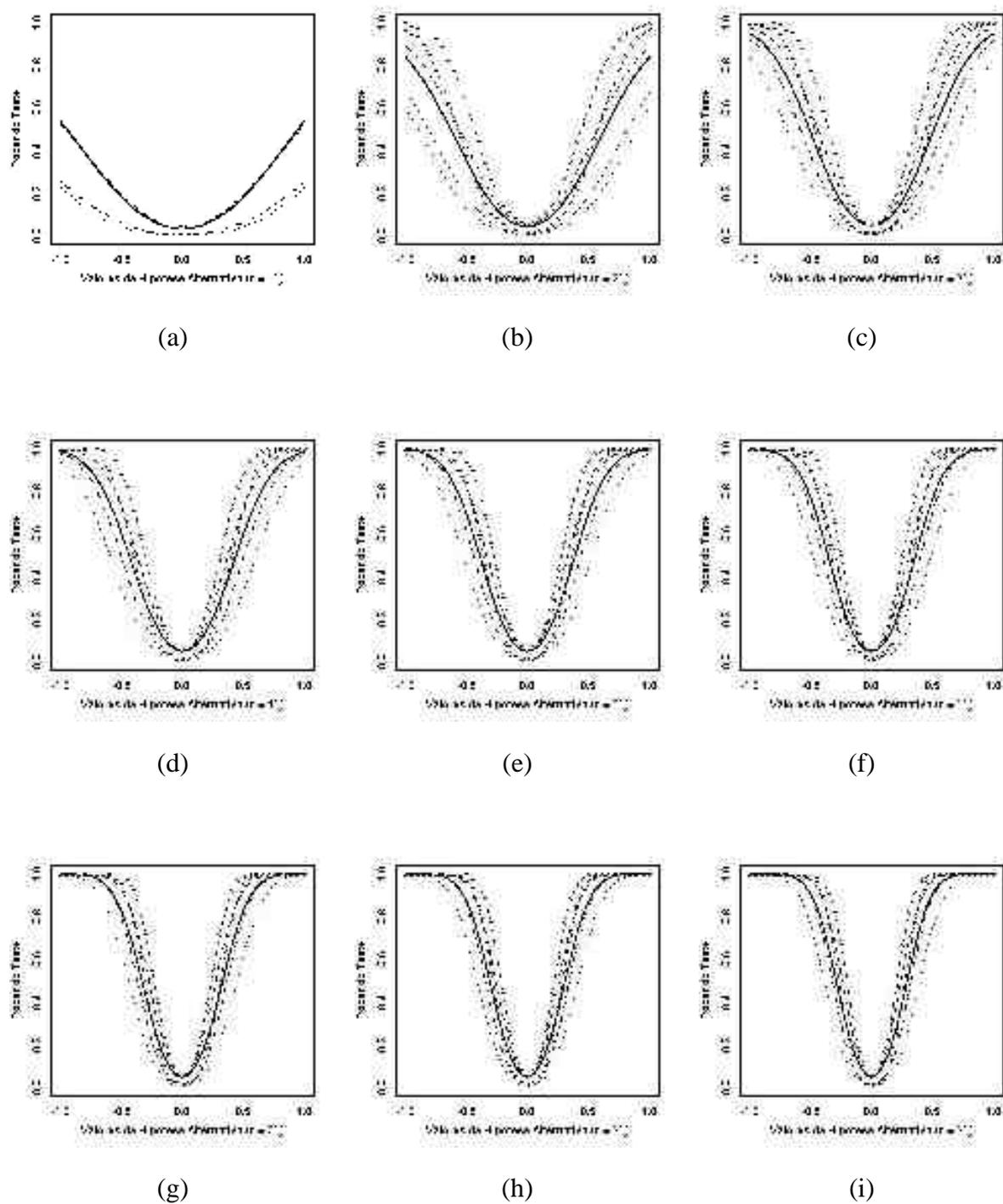


Figura 3. Valores da hipótese alternativa versus o poder dos testes para a distribuição Laplace; $\alpha = 0.05$: (—) *t-Student*, (---) *Wilcoxon*; $\alpha = 0.01$: (· ·) *t-Student*, (· - ·) *Wilcoxon*.

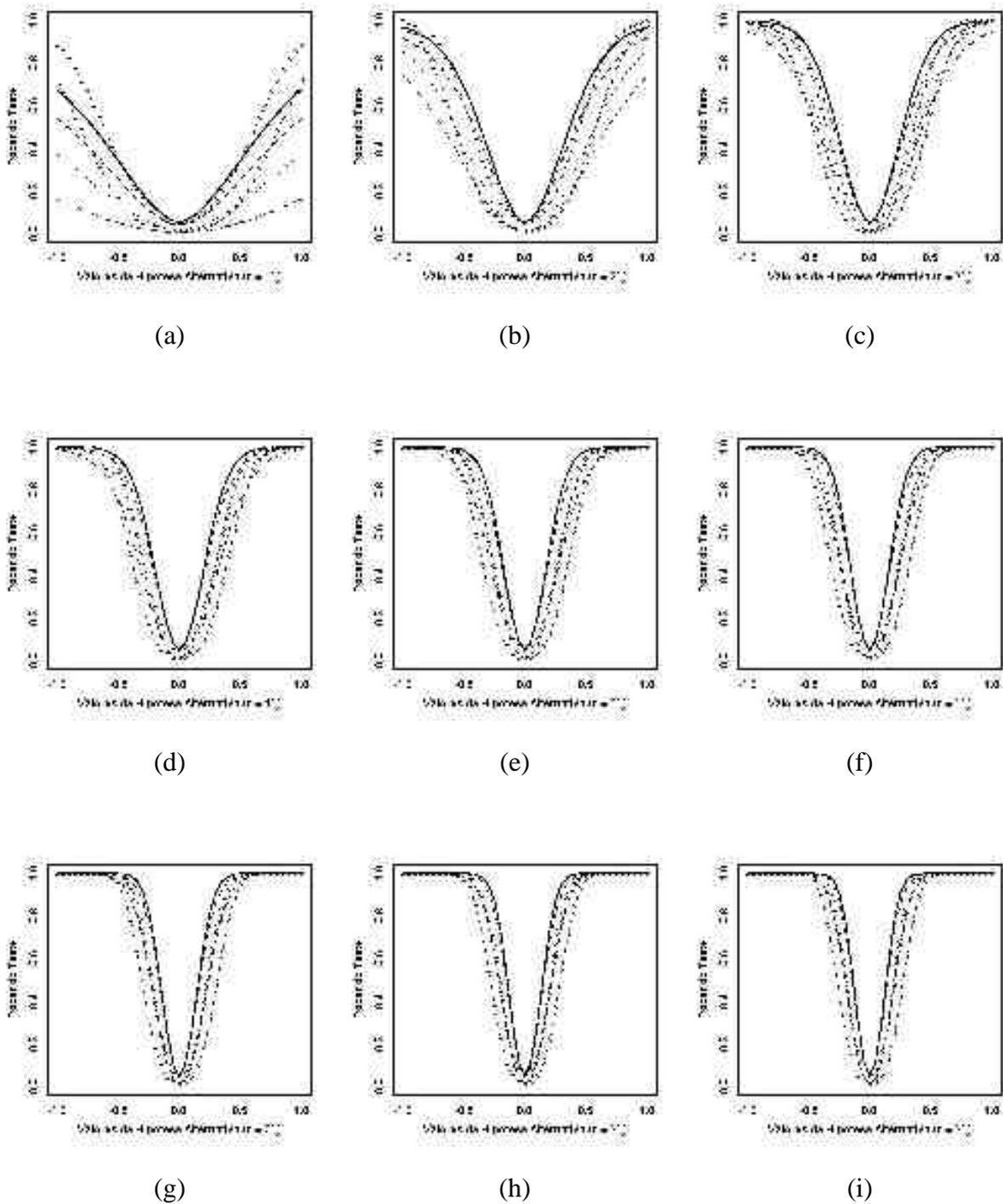


Figura 4. Valores da hipótese alternativa versus o poder dos testes para a distribuição Uniforme; $\alpha = 0.05$: (—) *t-Student*, (---) *Wilcoxon*; $\alpha = 0.01$: (· ·) *t-Student*, (-·-) *Wilcoxon*.

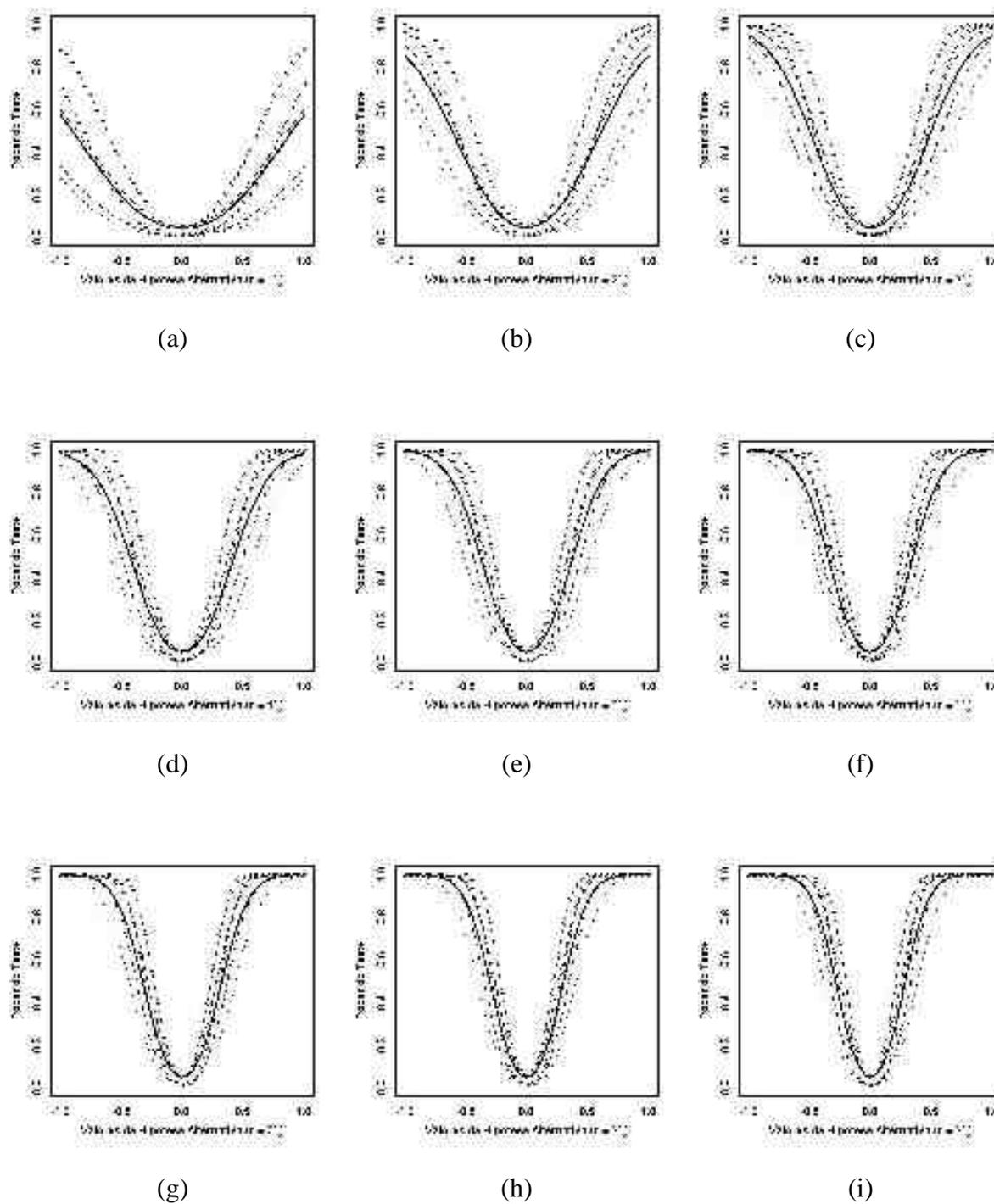


Figura 5. Valores da hipótese alternativa versus o poder dos testes para a distribuição *t-Student*; $\alpha = 0.05$: (—) *t-Student*, (---) *Wilcoxon*; $\alpha = 0.01$: (· ·) *t-Student*, (- -) *Wilcoxon*.

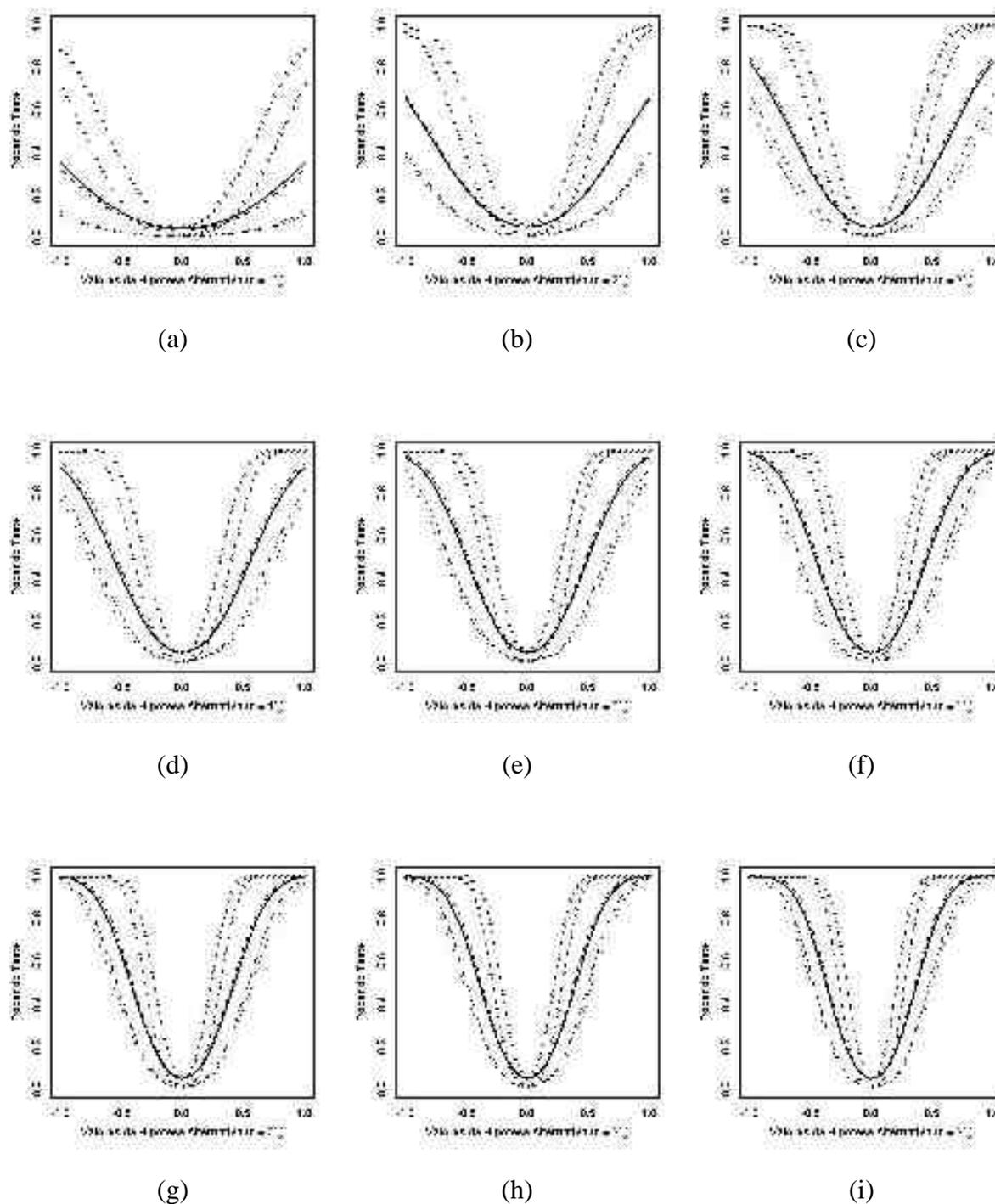


Figura 6. Valores da hipótese alternativa versus o poder dos testes para a distribuição Logística; $\alpha = 0.05$: (—) *t-Student*, (---) *Wilcoxon*; $\alpha = 0.01$: (· ·) *t-Student*, (- -) *Wilcoxon*.

Conclusão

A partir dos resultados apresentados nas Tabelas 2 a 6, observa-se para a distribuição Normal que os valores das taxas de erro tipo I dos testes *t-Student* e *Wilcoxon* se comportam de maneira similar e estão sempre próximos dos níveis de significância nominais, ($\alpha=5\%$ e $\alpha=1\%$), independente do tamanho da amostra. Para a distribuição Logística, em ambos os testes, as taxas de erro tipo I se aproximam do nível de significância à medida que o tamanho da amostra aumenta. Para $n \leq 30$, ambos os testes são conservativos; entretanto seus comportamentos são similares para amostras geradas da distribuição Normal. O comportamento das taxas de erro tipo I do teste de *Wilcoxon* para a distribuição de Laplace é bem diferente do teste *t-Student* principalmente para $n \leq 60$ em que a hipótese de que o tamanho empírico é igual ao tamanho nominal sempre é rejeitada em nível de significância de 5% (Tabela 3). Um comportamento similar é observado quando os dados são gerados das distribuições *t-Student* e Uniforme.

Como observação, as taxas de erro tipo I do teste de *Wilcoxon* estão na maioria das instâncias próximas do nível de significância nominal para qualquer distribuição, independentemente do tamanho da amostra. Esse fato está relacionado com o fato de que o teste de *Wilcoxon* é baseado na suposição de simetria.

Em relação ao poder, a partir das Figuras 2 a 6, observa-se que a medida que a hipótese alternativa se

afasta de $\mu = 0$, e o tamanho da amostra aumenta, o poder dos testes se aproximam de 1. Ambos os testes se comportam similarmente quando a distribuição é Normal e Logística, porém, para as demais distribuições, o teste de *Wilcoxon* apresenta um poder maior do que o teste *t-Student*.

Referências

- CASELLA, G.; BERGER, R.L. Statistical inference. *The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA*, 1990.
- CONOVER, W.J. *Practical Nonparametric Statistics*. New York: John Wiley and Sons, 1971.
- FERGUNSON, T.S. *Mathematical statistics*. New York: Academic Press, 1967.
- IMAN, R.L. Use of a t-statistic as an approximation to the exact distribution of the *Wilcoxon* signed ranks test statistic. *Communications in Statistic*, v. 3, p. 795-806, 1974.
- GRAYBILL, A M. et al. *Introduction to the Theory of Statistics*. São Paulo: McGraw-Hill, 1974.
- HOLLANDER, M.; WOLFE, D.A. W. *Nonparametric statistical inference*. New York: John Wiley & Sons, 1973.

Received on January 10, 2005.

Accepted on June 15, 2005.