

Universidade Aberta



Investigação Experimental: Potência estatística dos testes de aleatorização na comparação de dois grupos independentes

Fernando José dos Santos Branco

Tese orientada pela Professora Doutora Teresa Paula Costa Azinheira Oliveira

Doutoramento em Matemática
Especialidade Planeamento de Experiências

2010

Agradecimentos

Agradeço à Professora Doutora Teresa Paula Costa Azinheira Oliveira a orientação, o apoio e encorajamento com que me acompanhou ao longo da elaboração desta dissertação.

Agradeço também ao Professor Christoph Minder o interesse que manifestou pelo meu trabalho, a sua disponibilidade e paciência para ler em Português grande parte desta dissertação e as críticas e sugestões que a permitiram melhorar.

Resumo

O objectivo desta dissertação foi o de avaliar a potência dos testes de aleatorização em planeamentos de experiências completamente aleatorizadas, com dois grupos independentes. Realizámos três estudos experimentais, por simulação, tendo como variáveis independentes o tipo de distribuição, a dimensão e o equilíbrio das amostras e o tamanho do efeito. A potência do teste de aleatorização constituiu a variável dependente. No primeiro estudo, comparámos a potência dos testes de aleatorização no caso de populações gaussianas com a sua potência no caso de diversas distribuições não-normais. No segundo estudo, avaliámos o comportamento dos testes de aleatorização, em termos de potência, no caso de distribuições com assimetria e/ou curtose. Finalmente, no terceiro estudo, avaliámos a potência destes testes, no caso de distribuições normais, quando existe heterocedasticidade e desequilíbrio na dimensão das amostras. Como padrão de comparação, em todos os estudos calculámos também a potência do teste t de Student. Os resultados destes três estudos sugerem que, de um modo geral, quando os grupos são equilibrados, e os dados de cada grupo provêm de distribuições idênticas, ainda que não-normais ou com elevado grau de assimetria e/ou curtose, a potência dos teste de aleatorização não é muito afectada, verificando-se, por vezes, ganhos de potência. Contrariamente, quando se verifica a existência de heterocedasticidade e desequilíbrio na dimensão das amostras, mesmo com distribuições normais, a potência do teste de aleatorização é fortemente afectada. Em geral, as potências alcançadas pelos testes de aleatorização e t de Student são muito semelhantes, com uma ligeira vantagem para o primeiro, que tende a reduzir-se com o aumento da dimensão das amostras.

Abstract

The aim of this thesis was to evaluate the statistical power of randomization tests, in experiments with a completely randomized design, with two independent groups. We performed three experimental studies by simulation, using the type of distribution, the samples dimension and balance, and the effect size as independent variables. The statistical power was the dependent variable. In the first study, we compared the power of the randomization tests in the case of Gaussian distributions with their power in the case of non-normal distributions. In the second study, we evaluated the behaviour of the randomization tests, in terms of power, with data simulated from asymmetric and/or kurtotic distributions. Finally, in the third study, we evaluated the power of these tests, in the case of Gaussian distributions, when heteroscedasticity is present and the sample sizes are unequal. As a comparison standard, we also evaluated the power of the Student-t test. The results of these three studies suggest that, in general, when the group sizes are equal and the data from each sample come from identical distributions, even if they are non Gaussian or with a high degree of asymmetry and/or kurtosis, the power of the randomization tests is not highly affected. Moreover, in some cases, gains in power are observed. However, when heteroscedasticity is present and the sample sizes are unequal, even in the case of Gaussian distributions, the power of the randomization tests is highly affected. In general, the power achieved by randomization and Student-t tests are very similar, presenting the former a slight advantage, which tends to reduce with increased samples size.

Índice

INTRODUÇÃO	1
PARTE I.....	9
CAPÍTULO 1 - TESTES DE ALEATORIZAÇÃO	11
1. <i>Introdução</i>	13
2. <i>Definição do teste de aleatorização</i>	18
2.1. Descrição do procedimento.....	18
2.2. Teste de hipótese	20
2.3. Uma aplicação: um estudo experimental	22
2.4. Escolha do conjunto de referência.....	26
2.5. Escolha da estatística de teste.....	27
3. <i>Validade do procedimento do teste de aleatorização</i>	29
3.1 Teste de aleatorização exaustivo	29
3.2 Teste de aleatorização não-exaustivo.....	33
CAPÍTULO 2 - POTÊNCIA ESTATÍSTICA.....	35
1. <i>Definição de potência estatística</i>	37
2. <i>A função potência</i>	37
3. <i>Cálculo da potência estatística</i>	41
4. <i>Problemas da Análise de Potência</i>	44
4.1. Problemas teóricos.....	45
4.2. Problemas práticos.....	47
CAPÍTULO 3 - TESTES DE ALEATORIZAÇÃO E ANÁLISE DE POTÊNCIA	51
1. <i>Introdução</i>	53
2. <i>Diferentes abordagens para o cálculo da potência</i>	54
3. <i>Nível real e nível nominal</i>	56
PARTE II.....	61
1. <i>Introdução</i>	63
2. <i>Questões metodológicas comuns aos três estudos</i>	63
3. <i>O sistema R</i>	66
ESTUDO 1	69

1. <i>Introdução</i>	71
2. <i>Método</i>	77
3. <i>Apresentação e discussão dos resultados</i>	78
3.2. Comparação da potência dos testes de aleatorização e t de Student	80
3.3. Influência da não-normalidade na potência dos testes de aleatorização.....	82
3.3.1. Testes bicaudais.....	83
3.3.2. Testes unicaudais	87
4. <i>Conclusões</i>	92
ESTUDO 2	95
1. <i>Introdução</i>	97
2. <i>Método</i>	104
3. <i>Apresentação e discussão dos resultados</i>	106
3.1. Comparação da potência dos testes de aleatorização e t de Student	107
3.2. Influência da assimetria e curtose na potência dos testes de aleatorização.....	108
3.2.1. Testes bicaudais.....	110
3.2.2. Testes unicaudais	113
4. <i>Conclusões</i>	117
ESTUDO 3	119
1. <i>Introdução</i>	121
2. <i>Método</i>	121
3. <i>Apresentação e discussão dos resultados</i>	125
3.2. Potência dos testes de aleatorização para $n_1 = n_2$	127
3.3. Potência dos testes de aleatorização para $n_1 < n_2$	129
3.4. Potência dos testes de aleatorização para $n_1 > n_2$	131
4. <i>Conclusões</i>	132
CONCLUSÃO	135
REFERÊNCIAS	141

ANEXOS	159
ANEXO I.....	161
ANEXO II.....	163
ANEXO III	165
ANEXO IV	169
ANEXO V.....	171
ANEXO VI	173
ANEXO VII	217
ANEXO VIII.....	219
ANEXO IX.....	221
ANEXO X.....	283

Índice de Tabelas

[Capítulo 1]

Tabela 1.1 Vectores experimentais e estatísticas de teste 24

Tabela 1.2 Distribuição de frequências das estatísticas de teste
(ET) 25

[Capítulo 3]

Tabela3.1 Estatísticas descritivas (mínimo, máximo e média) da
diferença entre a potência do teste de aleatorização exaustivo
e não-exaustivo, para os cinco conjuntos de 20 000 amostras. 59

[Estudo 1]

Tabela E1.1 Percentagem de amostras consideradas com
distribuição normal pelos testes de Lillifors, Anderson-
Darling e Shapiro-Wilk para dados simuladas de uma
distribuição normal e de uma distribuição uniforme (para $\alpha =$
0.05). 73

Tabela E1.2 As 15 distribuições Marron-Wand 74

Tabela E1.3 Potência dos testes de aleatorização, bi e unicaudais,
para valores de tamanho do efeito (TE) de 0.2, 0.5 e 0.8, em
função do tamanho das amostras 80

Tabela E1.4 Diferença entre a potência dos testes de
aleatorização e t de Student para a distribuição Marron-Wand
05 ('Outlier'). 82

[Estudo 2]

Tabela E2.1 Médias e desvios-padrão das 20 distribuições g - e h
simuladas 105

Tabela E2.2 Diferença entre a potência dos testes de aleatorização e t de Student (testes bicaudais)..... 108

[Estudo 3]

Tabela E3.1 Dimensões e desvios-padrão das amostras dos 19.grupos de condições 124

Tabela E3.2 Estatísticas (mínimos e máximos) para a diferença entre a potência dos testes de aleatorização e t de Student 125

Índice de Figuras

[Capítulo 1]

Figura 1.1 Gráfico da distribuição de frequências das estatísticas de teste.....	25
----------------------------------------------------------------------------------	----

[Capítulo 2]

Figura 2.1 Funções potência para os testes t de Student unicaudal à esquerda, unicaudal à direita e bicaudal, com $\alpha = 0.05$ e $n_1 = n_2 = 8$, sob os pressupostos paramétricos clássicos	39
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----

Figura 2.2 Funções potência para o teste t de Student bicaudal, com $\alpha = 0.05$, para grupos equilibrados com 4, 8, 16, 32 e 64 elementos, sob os pressupostos paramétricos clássicos.....	40
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----

Figura 2.3 Funções potência para o teste t de Student bicaudal, com $n_1 = n_2 = 8$, para diversos valores de α , sob os pressupostos paramétricos clássicos	41
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----

[Estudo 1]

Figura E1.1 Densidades das 15 distribuições Marron-Wand	76
---------------------------------------------------------------	----

Figura E1.2 Potência dos testes de aleatorização bicaudais, para a distribuição normal, com $\alpha = 0.05$, em função do tamanho do efeito e do tamanho das amostras	79
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----

Figura E1.3 Potência dos testes de aleatorização unicaudais, para a distribuição normal, com $\alpha = 0.05$, em função do tamanho do efeito e do tamanho das amostras	79
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----

Figura E1.4 Diferença de potência entre os testes de aleatorização e o teste t de Student (bicaudais) para a distribuição M-W 05 ('Outlier'), em função do tamanho do efeito e do tamanho das amostras	81
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----

Figura E1.5 Diferença entre a potência dos testes de aleatorização bicaudais, com $n_1 = n_2 = 8$, para as distribuições Marron-Wand 02 a 05 e a sua potência para a distribuição normal, em função do tamanho do efeito	84
Figura E1.6 Diferença entre a potência dos testes de aleatorização bicaudais, com $n_1 = n_2 = 8$, para as distribuições Marron-Wand 06 a 15 e a sua potência para a distribuição normal, em função do tamanho do efeito	85
Figura E1.7 Diferença entre a potência dos testes de aleatorização unicaudais, com $n_1 = n_2 = 8$, para as distribuições Marron-Wand 02 a 05 e a sua potência para a distribuição normal, em função do tamanho do efeito	88
Figura E1.8 Diferença entre a potência dos testes de aleatorização unicaudais, com $n_1 = n_2 = 8$, para as distribuições Marron-Wand 06 a 15 e a sua potência para a distribuição normal, em função do tamanho do efeito	89
[Estudo 2]	
Figura E2.1 Gráficos da função densidade de probabilidade $f_g(y)$ para quatro valores de g (0.2, 0.4, 0.6 e 0.8).	99
Figura E2.2 Gráficos da função densidade de probabilidade $f_h(y)$ para três valores de h (0.1, 0.2 e 0.3).	100
Figura E2.3 Gráficos da função densidade de probabilidade $f_{g,h}(y)$ para as combinações de quatro valores de g (0.2, 0.4, 0.6 e 0.8) com três valores de h (0.1, 0.2 e 0.3)	100
Figura E2.4 Testes de aleatorização (bicaudais): diferença entre a sua potência para as distribuições assimétricas e a sua potência para a distribuição normal, com $n_1 = n_2 = 8$	111

Figura E2. 5 Testes de aleatorização (bicaudais): diferença entre a sua potência para as distribuições com curtose e a sua potência para a distribuição normal, com $n_1 = n_2 = 8$	112
Figura E2. 6 Testes de aleatorização unicaudais: diferença entre a sua potência para as distribuições assimétricas e a sua potência para a distribuição normal, com $n_1 = n_2 = 8$	114
Figura E2. 7 Testes de aleatorização unicaudais: diferença entre a sua potência para as distribuições com curtose e a sua potência para a distribuição normal, com $n_1 = n_2 = 8$	116
[Estudo 3]	
Figura E3.1 Potência do teste de aleatorização bicaudal para as condições com amostras equilibradas	127
Figura E3.2 Potência do teste de aleatorização bicaudal para as condições em que a maior variância está associada à amostra menor em comparação com a condição em que há homogeneidade de variâncias.....	129
Figura E3.3 Sobreposição dos dois gráficos da Figura E3.2	130
Figura E3.4 Potência do teste de aleatorização bicaudal para as condições em que a maior variância está associada à amostra maior em comparação com a condição em que há homogeneidade de variâncias.....	131
Figura E3.5 Sobreposição dos dois gráficos da Figura E3.4	132

INTRODUÇÃO

Os estudos realizados no âmbito desta dissertação tiveram como objectivo avaliar a potência dos testes de aleatorização em planeamentos de experiências completamente aleatorizadas, com dois grupos independentes.

Em geral, quando se usa este tipo de planeamentos, por exemplo, em investigações experimentais nas áreas da Biologia, da Medicina, da Psicologia, da Educação, o teste estatístico clássico a que os investigadores recorrem mais frequentemente é o teste t de Student, ou o teste F da Análise de Variância.

Raras são as experiências, porém, em que os experimentadores extraem amostras aleatórias de populações bem definidas. Um analista político, que realiza uma sondagem eleitoral, poderá enumerar toda a população e dela extrair uma amostra aleatória, mas um investigador nas áreas da Biologia, da Psicologia ou em qualquer outra área, não tem muitas vezes a possibilidade de seleccionar aleatoriamente amostras da população em que está interessado. E nos casos raros em que isso acontece, trata-se de populações tão específicas que o seu interesse é reduzido. Muitos autores, de que podemos citar, Cotton (1967), Keppel (1973), Kirk (1968) e Spence et al. (1976), concorrem na afirmação de que amostras não-aleatórias são prevalentes nos estudos experimentais.

Alguns autores, entre os quais Edgington e Onghena (2007), defendem que as amostras aleatórias não só são raras na investigação experimental típica, como, em muitos casos, não tem sentido utilizá-las. Estes autores afirmam que a população em que o investigador está interessado é, quase sempre, uma população que não permite a extracção de uma amostra aleatória. Suponhamos o caso de um biólogo que está empenhado em estabelecer uma lei biológica geral aplicável não só a indivíduos existentes num dado momento como a indivíduos que ainda não nasceram ou já morreram. Mesmo que fosse possível extrair uma amostra aleatória

da população de indivíduos existentes num dado momento, as leis biológicas resultantes não contemplariam a população mais vasta em que o biólogo está interessado, mas unicamente aquela da qual essa amostra foi retirada.

A validade da aplicação dos testes clássicos, como o teste t de Student ou o teste F da Análise de Variância, depende dos pressupostos de normalidade, homocedasticidade e independência. É questionável que, em muitas situações, esses pressupostos possam ser defendidos realisticamente, ainda que alguns estudos de simulação (por exemplo, Boneau, 1960; Glass, Peckham, & Sanders, 1972) tenham mostrado que essas técnicas são relativamente robustas face a violações dos seus pressupostos.

Há um pressuposto básico, no entanto, que é comum a todos os métodos da ‘teoria normal’, o da obtenção de amostras aleatórias das populações em estudo. Hays (1972), por exemplo, chama a atenção, no seu conhecido manual de Estatística, para o facto de “[...]todas as técnicas e teoria que iremos discutir se aplicam a amostras aleatórias, e não se mantêm [válidas] necessariamente para dados recolhidos de qualquer forma.”

Se as amostras não forem aleatórias, não parece justificável usar na análise estatística dos dados experimentais, como técnicas de primeira escolha, as que se baseiam no modelo da amostragem aleatória. Será mais apropriado usar as técnicas que se baseiam no modelo da distribuição aleatória, os testes de aleatorização, que com os recursos computacionais actualmente disponíveis são fáceis de usar.

Os testes de aleatorização são testes estatísticos cuja validade se baseia no modelo de distribuição aleatória. São testes com uma base teórica sólida formulada inicialmente por Pitman (1937a, 1937b, 1938) e Welch

(1938), na sequência de Fisher (1966/1935), e desenvolvida por muitos outros estatísticos, de que destacamos Kempthorne (1952, 1955) e Edgington (1964, 1966, 1969a, 1969b, 1995).

Quando, na análise de dados experimentais, se calcula uma estatística de teste como t ou F , o que distingue o teste de aleatorização do teste clássico é a forma de calcular a significância. No caso do teste de aleatorização a significância é calculada por um procedimento em que repetidamente se permutam os dados, e a significância calculada dessa maneira é exacta.

Com este procedimento, o investigador pode avaliar a significância de qualquer teste estatístico, mesmo que ainda não tenha sido derivada a distribuição amostral da estatística do teste. Na análise dos seus dados, o investigador poderá escolher o teste que julgue mais conveniente e adequado às hipóteses em estudo, sendo possível construir uma ‘tabela’ que lhe indica a significância dos resultados experimentais.

As vantagens dos testes de aleatorização são evidentes: os dados não necessitam de constituir uma amostra aleatória; os dados podem provir de uma qualquer população; e são extremamente versáteis, podendo ser usados para calcular a significância de qualquer estatística de teste.

Em diversas obras sobre a aplicação destes testes, como as de Edgington & Onghena (2007), Maritz (1995), Manly (1997), referem-se utilizações de testes de aleatorização em áreas tão diversas como Biologia, Farmacologia, Medicina, Psicologia e Química.

Mas apesar da sua crescente utilização, muitos problemas interessantes necessitam de investigação, nomeadamente os relacionados com a sua potência.

Existe uma extensa literatura sobre a potência estatística dos testes paramétricos, face a violações dos seus pressupostos, como os estudos clássicos de Pearson (1931), Box (1953) e Boneau (1960). Nestes estudos a preocupação fundamental era com o erro de Tipo I, mas estudos posteriores começaram a examinar o desempenho dos testes relativamente à probabilidade de cometer erros de Tipo II e o problema relacionado da potência dos testes.

A partir dos anos 70, surgem muitos estudos de simulação que comparam o comportamento de testes paramétricos e não-paramétricos (Lehman, 1975; Randles, 1980; Blair & Higgins, 1985).

Não existem, porém, muitos estudos sobre a potência dos testes de aleatorização, embora exista a suposição mais ou menos generalizada que estes testes são robustos relativamente a não-normalidade, heterocedasticidade e problemas associados com amostras de pequena dimensão.

Esta suposição é baseada nalguns estudos, como o de Kempthorne e Doerfler (1969), que mostrou que, em determinadas circunstâncias, os testes de aleatorização são mais potentes que os testes não-paramétricos clássicos. Também Romano (1989) indica terem os testes de aleatorização, para grandes amostras, potência semelhantes aos testes clássicos quando se verificam os seus pressupostos. Em Bradbury (1987) e Routledge (1997) os testes de aleatorização surgem como mais robustos para pequenas amostras.

Mas, em sentido contrário, Lindman (1974) e Chen e Chen (1998) defendem que os testes clássicos, face a pequenas violações dos seus pressupostos, são mais potentes do que os de aleatorização.

O facto de com um teste de aleatorização não existirem pressupostos sobre uma distribuição populacional – o teste é realizado com os dados recolhidos – poderá explicar, pelo menos parcialmente, porque não há muita investigação sobre a potência deste tipo de teste.

Na verdade, para realizar uma análise a priori da potência de teste será necessário invocar alguns pressupostos sobre as respostas que podem ser esperadas, o que pode parecer contraditório com uma das vantagens dos testes de aleatorização face aos testes paramétricos clássicos: a de não ser necessário pressupor distribuições específicas.

No entanto, a análise da potência de teste é a única ferramenta estatística para decidir, ao planear uma experiência, quantas unidades experimentais deveremos afectar a cada tratamento. Pode tornar-se portanto conveniente fazer pressupostos adicionais, sendo evidente que a validade da análise de potência vai depender da validade dos pressupostos invocados.

Assim, parece relevante procurar investigar e desenvolver a temática da potência dos testes de aleatorização. Nesta dissertação, investigamos o caso particular do planeamento de experiências completamente aleatorizadas, com dois grupos independentes. Neste sentido, procurámos construir diversos cenários, manipulando o tipo de distribuições, o tamanho do efeito, a variabilidade nas respostas e a dimensão das amostras, esperando cobrir um leque alargado das situações com que um investigador poderá deparar na sua prática de investigação.

Esta dissertação consta de duas partes. Na Parte I, começamos por apresentar, no Capítulo 1, o conceito de teste de aleatorização; no Capítulo 2, o conceito de potência estatística; e, finalmente, no Capítulo 3, são relacionados esses conceitos, com a análise de potência dos testes de aleatorização.

Na Parte II, após uma introdução em que se apresentam algumas questões metodológicas comuns aos estudos realizados, apresentamos um primeiro estudo, em que comparamos a potência dos testes de aleatorização no caso de populações gaussianas com a sua potência no caso de diversas distribuições, misturas de normais. Os resultados obtidos neste estudo, levaram-nos a realizar um segundo estudo, em que avaliamos o comportamento dos testes de aleatorização, em termos de potência, no caso de distribuições com grau variado de assimetria e/ou curtose. Num terceiro, avaliamos a potência destes testes quando existe heterocedasticidade e desequilíbrio no tamanho dos grupos. Finalmente, apresentamos uma conclusão geral, a que seguem as referências bibliográficas e os anexos.

PARTE I

CAPÍTULO 1 - TESTES DE ALEATORIZAÇÃO

1. Introdução

Um teste de aleatorização é um teste estatístico cuja validade tem por base a distribuição aleatória das unidades experimentais pelos tratamentos.

Assim, o teste de aleatorização tem por base o modelo de distribuição aleatória, enquanto os testes estatísticos clássicos, como o teste t de Student ou o teste F da Análise de Variância, têm por base o modelo de amostragem aleatória. Também é neste mesmo modelo da amostragem aleatória, que se baseiam os testes de permutação.¹

A ideia básica subjacente aos testes de aleatorização e de permutação é a de calcular uma estatística de teste para cada permutação dos dados, de modo a construir o conjunto de referência a partir do qual se determina a probabilidade associada à estatística de teste observada. Mas, baseados em diferentes modelos, estes testes diferem em termos das hipóteses testadas e no alcance das inferências.

Começamos por apresentar a origem e as principais contribuições para a teoria destes testes, deixando para o ponto seguinte a definição operacional de um teste de aleatorização, com a descrição do procedimento, regra de decisão e critério de validade.

A ideia deste tipo de testes surge com Fisher (1966) que na sua obra ‘The Design of Experiments’, originalmente publicada em 1935, reanalisa, no Capítulo III, uma parte dos dados de uma experiência realizada por Darwin sobre o mérito relativo da auto-fertilização e da fertilização cruzada no crescimento de plantas. Fisher analisa os dados

¹ Usamos a terminologia proposta Kempthorne e Doerfler (1969) que distinguem ‘randomization tests’ (testes de aleatorização), baseados no modelo de distribuição aleatória, de ‘permutation tests’ (testes de permutação), baseados no modelo de amostragem aleatória.

referentes a *Zea Mays* (com 15 observações emparelhadas), após criticar a análise feita por Galton, a pedido de Darwin. Fisher começa por indicar qual o objectivo da experiência, testar se os dois grupos de medidas provêm de uma mesma população e afirma que a ‘teoria dos erros’ permite testar uma hipótese mais limitada, a de que os dois grupos de medidas são amostras extraídas da mesma população com distribuição normal. Para testar essa hipótese, para a qual ‘está disponível um critério estatístico exacto’, Fisher usa o teste *t* de Student para grupos emparelhados. Seguidamente, após referir que a única falha na experiência de Darwin foi a de não ter havido distribuição aleatória dos elementos de cada par, vai testar uma hipótese mais ampla, a de que as observações de cada grupo são amostras da mesma população, sem especificar que esta tem uma distribuição normal. Usa as quinze diferenças entre os pares de plantas com fertilização cruzada e autofertilizadas, diferenças que sob a hipótese nula têm igual probabilidade de ter um sinal positivo ou negativo². Escolhe como estatística de teste a soma das diferenças e vai verificar quantas das 2^{15} estatísticas que se podem calcular (atribuindo a cada diferença alternadamente um sinal positivo ou negativo), excedem a soma das diferenças observadas.

Como conclusão, Fisher afirma que como, sob a hipótese nula, cada uma dessas 2^{15} combinações tem uma igual probabilidade de ocorrência, o número de estatísticas iguais ou superiores à estatística observada permite calcular a sua significância. Essa probabilidade é exacta e, tendo a probabilidade associada à estatística do teste *t* de Student (0.04970) um valor muito aproximado ao valor calculado com este método (0.05267), vem justificar o uso do teste *t* de Student.

Importa chamar a atenção para o facto de Fisher usar, como base dos testes que efectuou, o modelo de amostragem aleatória: no caso do teste *t*

² Nos dados de Darwin, 13 das diferenças tinham um sinal positivo (indicando superioridade da fertilização cruzada) e 2 sinal negativo.

de Student, a hipótese testada é a de que as observações de cada grupo são amostras da mesma população normal; no caso do teste de permutação, a hipótese é a de que as observações de cada grupo são amostras da mesma população.

Fisher refere especificamente a necessidade de distribuição aleatória ('randomization'), no planeamento da experiência, mas com um objectivo de controlo, de modo a garantir a validade interna da experiência.

Dadas as dificuldades de computação, na época, os testes de permutação têm, para Fisher, um interesse sobretudo teórico, justificando o uso de testes mais expeditos, como o teste t de Student. Assim, Fisher (1936), após apresentar um exemplo de uma comparação de dois grupos independentes, com um teste de permutação, afirma que “na realidade, o estatístico não utiliza este processo muito enfadonho, mas as suas conclusões não têm outra justificação para além do facto de que poderiam ter sido alcançadas por este método muito elementar”.

A ideia de um teste ‘muito elementar’ tendo por base o modelo da distribuição aleatória, ou seja, a ideia de um teste de aleatorização, deve-se a Pitman que publica uma série de três artigos sobre ‘testes de significância que podem ser aplicados a amostras de quaisquer populações’ (Pitman, 1937a, 1937b, 1938).

Pitman (1937a) após demonstrar a validade dos testes para amostras extraídas de populações infinitas, conclui que a validade se mantém “quando as amostras são extraídas com reposição de uma população finita desde que as amostras sejam simultâneas - isto é, desde que ambas sejam extraídas antes que qualquer delas seja repostas” e “em particular [...] quando as amostras em conjunto constituem toda a população”.

Este autor apresentou originalmente um quadro teórico para os testes de aleatorização, dando em cada um dos seus três artigos (Pitman, 1937a, 1937b, 1938) exemplos de testes que se baseiam unicamente no modelo da distribuição aleatória.

Muitos investigadores apresentaram desenvolvimentos importantes para a teoria dos testes de aleatorização, dos quais evidenciamos, Welch (1937), Kempthorne (1952, 1955), Dwass (1957), Chung e Fraser (1958) e Hope (1968). Edgington, para além de contribuições em diversos artigos (Edgington, 1964, 1966, 1969a, 1969b), publica em 1980 'Randomization Tests', obra de referência nesta área, com sucessivas edições em 1987, 1995 e 2007, esta última tendo como co-autor Patrick Onghena. Importa ainda referir Pesarin (2001), com relevantes contributos na área dos testes multivariados.

O procedimento usual para realizar um teste de aleatorização é o da repetida permutação dos dados. Esta abordagem é por isso chamada Abordagem de Permutação dos Dados.

Kempthorne e Doerfler (1969) apresentaram uma nova conceptualização dos testes de aleatorização, com a Abordagem da Aleatorização de Referência. Nessa abordagem, o conjunto de referência não é construído a partir da permutação dos dados, mas por um procedimento com duas fases. Numa primeira fase gera-se o conjunto de todas as possíveis distribuições das unidades experimentais pelas condições, sendo uma escolhida aleatoriamente para a experiência. Numa segunda fase, os valores experimentais observados são sobrepostos às distribuições geradas na primeira fase, permitindo construir o conjunto de referência. Embora, para a realização de um teste, se use a Abordagem da Permutação dos Dados, a abordagem proposta por Kempthorne e seus colaboradores é útil para compreender a lógica dos testes de aleatorização e demonstrar a sua validade. Mas, em termos da decisão

final e de validade, as duas abordagens, a da Permutação dos Dados e a da Aleatorização de Referência são equivalentes (Edgington & Onghena, 2007).

Com o advento do computador, possibilitando a realização fácil e rápida da computação dos dados, os testes de aleatorização passaram a ter uma utilização prática, evidenciada na análise de dados experimentais.

Em muitas situações, no entanto, a quantidade de cálculo necessária excede a capacidade de um computador. Por exemplo, o número de maneiras de distribuir 32 unidades experimentais por 4 condições, com 8 unidades experimentais por condição, é de 9.95611×10^{16} . Este problema foi resolvido por Dwass (1957), que sugeriu o uso de “testes de aleatorização modificados”. Quando não é possível usar todas as permutações dos dados, é extraída uma amostra aleatória dessas permutações para construir o conjunto de referência. Este procedimento conduz a um teste de aleatorização válido, pois quando a hipótese nula é verdadeira a probabilidade de a rejeitar, para qualquer nível de significância α , não é maior que α . Quando a hipótese nula é falsa, porém, o teste será menos potente que o teste baseado na totalidade das permutações.

Também Chung e Fraser (1958) propuseram o uso de um subconjunto do conjunto de referência para determinar a significância de um teste, mas esse subconjunto é construído de forma sistemática e não aleatoriamente, como o proposto Dwass. O interesse desta contribuição é fundamentalmente teórico, com a introdução do conceito de grupo de permutação.

Como se referiu anteriormente, os testes de aleatorização tinham inicialmente um interesse meramente teórico, servindo para validar os métodos clássicos, nomeadamente no caso de se verificarem violações

dos seus pressupostos. Os testes de aleatorização estiveram, também, na base dos testes baseados em números de ordem ('ranks') que foram propostos nos anos 40, bem como na construção das suas tabelas de valores críticos.

Com a facilidade e rapidez de cálculo possibilitada pelos avanços nos computadores, os testes de aleatorização passaram a ser considerados por muitos estatísticos como técnicas de primeira escolha (por exemplo, Brillinger, Jones & Tukey, 1978).

2. Definição do teste de aleatorização

Definimos um teste de aleatorização como sendo um teste estatístico cuja validade se baseia na distribuição aleatória de unidades pelos tratamentos. Esta formulação abstracta será operacionalizada nos pontos seguintes.

2.1. Descrição do procedimento

Descrevemos o procedimento do teste de aleatorização de acordo com Onghena (1994), que usa a Abordagem da Aleatorização de Referência proposta por Kempthorne e colaboradores (ver, por exemplo, Gabriel & Hall, 1983; Kempthorne & Doerfler, 1969). Edgington e Onghena (2007), expõem o procedimento do teste de aleatorização usando a Abordagem da Permutação dos Dados, mostrando, como referimos já, que as duas abordagens são equivalentes em termos da decisão final e de validade. Para a descrição do procedimento usamos, sem perda de generalidade, um planeamento completamente aleatorizado com dois tratamentos. O

procedimento, porém, pode ser facilmente generalizado para outros planeamentos experimentais.

Começaremos por introduzir a seguinte notação:

Sejam

$\mathbf{u}' = (u_1, u_2, \dots, u_i, \dots, u_n)$ um vector de etiquetas para as N unidades experimentais;

$\mathbf{e}_j = (e_{j1}, e_{j2}, \dots, e_{ji}, \dots, e_{jn})$ um vector experimental de N parâmetros de um planeamento de experiências e_{ji} associados à experiência j , representando a afectação das unidades aos tratamentos; e

$\mathbf{y}' = (y_1, y_2, \dots, y_i, \dots, y_N)$ um vector resposta de N observações y_i .

Relativamente ao vector \mathbf{u}' , num planeamento completamente aleatorizado com dois tratamentos, n_1 unidades são aleatoriamente afectas ao Tratamento A e $n_2 = N - n_1$ unidades são afectas ao tratamento B.

Quanto ao vector \mathbf{e}' , num planeamento completamente aleatorizado com dois tratamentos, e_{ji} é igual a 1 se a experiência j atribui a unidade i ao tratamento A e e_{ji} é igual a 0 se a experiência j atribui a unidade i ao tratamento B.

Assim, a selecção aleatória de um vector \mathbf{e}_j , do conjunto de todos os possíveis vectores experimentais $\mathcal{E} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_j, \dots, \mathbf{e}_k)$, implica a aleatorização na distribuição das unidades experimentais. Designamos por vector experimental observado \mathbf{e}_{obs} o vector experimental \mathbf{e}_j seleccionado aleatoriamente e por conjunto de referência o conjunto dos vectores experimentais \mathcal{E} .

A cardinalidade desse conjunto de referência, num planeamento completamente aleatorizado com dois tratamentos, é igual a

$$k = C_{n_1}^{n_1+n_2} \quad (1.1)$$

Seja, $s_j = (\mathbf{e}_j, \mathbf{y})$ a estatística de teste, função do vector experimental e do vector resposta e representemos por $s_{\text{obs}} = (\mathbf{e}_{\text{obs}}, \mathbf{y})$ o valor observado da estatística de teste.

Considera-se habitualmente $s_{\text{obs}} = s_1$ e designam-se os outros $k-1$ valores da estatística do teste por estatísticas de referência adicionais. Designa-se por distribuição de referência a distribuição de frequência das estatísticas de referência (constituídas pela estatística observada e pelas estatísticas de referência adicionais).

2.2. Teste de hipótese

A hipótese nula genérica de um teste de aleatorização é a de que as respostas são independentes do tratamento em que são observadas. Assim, sob a hipótese nula, as respostas (\mathbf{e} , portanto, a estatística de teste observada) seriam as mesmas se se tivesse escolhido outro vector experimental. Formalizando:

$$H_0 : s_{\text{obs}} = s_j, \forall \mathbf{e}_j \in \mathcal{F} \quad (1.2)$$

Suponhamos que a estatística de teste foi definida de modo a que valores elevados sejam inconsistentes com a hipótese nula.

No procedimento do valor crítico, a distribuição de referência é constituída ordenando as estatísticas de referência s_j :

$$s_{(1)} \leq s_{(2)} \leq \dots \leq s_{(j)} \leq \dots \leq s_{(k)} \quad (1.3)$$

e determina-se o valor crítico da estatística de teste, s_{crit} :

$$s_{\text{crit}} = s_{(k - \text{trunc}(k\alpha))} \quad (1.4)$$

em que $\text{trunc}(r)$ é a função de truncagem, que devolve o maior inteiro que é menor ou igual a r .

Se $k\alpha$ for um inteiro, então s_{crit} é $s_{(k(1-\alpha))}$; se $k\alpha$ não for um inteiro, então s_{crit} é a estatística de referência cujo número de ordem é o menor inteiro superior a $k(1-\alpha)$.

A regra de decisão do teste é a de rejeitar a hipótese nula se a estatística observada for maior que a estatística crítica:

$$\text{Rejeitar } H_0 \text{ caso se verifique a condição: } s_{\text{obs}} > s_{\text{crit}} \quad (1.5)$$

Se se preferir usar o procedimento do valor-p, então o valor-p (valor de probabilidade associada à estatística do teste) é a proporção de estatísticas de referência que são iguais ou superiores à estatística observada:

$$p = k^{-1} \sum_{j=1}^k I(s_j \geq s_{\text{obs}}) \quad (1.6)$$

com $I(t)$, a função indicadora, sendo igual a 1 se a proposição t for verdadeira e igual a 0 no caso contrário. O número de estatísticas de referência iguais ou superiores à estatística observada, η ,

$$\text{com } \eta = kp = \sum_{j=1}^k I(s_j \geq s_{obs}) = 1 + \sum_{j=2}^k I(s_j \geq s_{obs}) \quad (1.7)$$

é, pelo menos, igual a 1, visto que $s_{obs} = s_1$ é também uma estatística de referência.

A regra de decisão do teste, segundo este procedimento, consiste em rejeitar a hipótese nula sempre que o valor-p seja igual ou inferior ao nível de significância α :

$$\text{Rejeitar } H_0 \text{ quando se verificar: } p \leq \alpha \quad (1.8)$$

Em termos de decisão do teste, os dois procedimentos, o do valor crítico e o do valor-p, são equivalentes. Tem-se:

$$\begin{aligned} \text{Rejeitar } H_0 &\Leftrightarrow p \leq \alpha \\ &\Leftrightarrow \eta \leq k\alpha \\ &\Leftrightarrow \eta \leq \text{trunc}(k\alpha) \\ &\Leftrightarrow \sum_{j=1}^k I(s_j \geq s_{obs}) \leq \sum_{j=1}^k I(s_j > s_{crit}) \\ &\Leftrightarrow s_{obs} > s_{crit} \end{aligned} \quad (1.9)$$

2.3. Uma aplicação: um estudo experimental

Suponhamos que um investigador deseje efectuar uma experiência para testar o efeito de dois tratamentos, A e B, para combater uma dada doença. Admita que é usada na experiência uma medida em que um valor mais elevado significa um efeito mais positivo. O investigador supõe que o efeito do tratamento A é superior ao do tratamento B, pelo que opta

por um teste unilateral à direita, estabelecendo um nível de significância de 0.05.

O investigador decide realizar previamente um estudo-piloto, com 8 doentes: $\mathbf{u}' = (a, b, c, d, e, f, g, h)$. Quatro dos doentes são escolhidos aleatoriamente para receberem o tratamento A, recebendo os outros quatro o tratamento B. Isto significa que um dos $k = C_4^8 = 70$ vectores experimentais do conjunto de referência \mathcal{E} foi extraído aleatoriamente. Esses vectores estão listados na Tabela 1.1 (em que também se indicam as estatísticas de teste).

Suponhamos que os primeiros quatro doentes são afectos ao tratamento A e os restantes quatro ao tratamento B, o que significa que $\mathbf{e}_{\text{obs}} = \mathbf{e}_1$. Admitamos, também, que o investigador decidiu usar a diferença entre as médias dos dois tratamentos como estatística de teste: $s_j = s(\mathbf{e}_j, \mathbf{y}) = n_1^{-1} \mathbf{e}_j' \mathbf{y} - n_2^{-1} (\mathbf{1}' - \mathbf{e}_j)' \mathbf{y}$.

A hipótese nula é a de que não há efeito diferencial do tratamento, ou seja, para cada doente, o valor observado seria o mesmo independentemente do tratamento recebido.

São obtidos os seguintes resultados: $\mathbf{y}' = (12, 15, 16, 17, 10, 11, 13, 14)$, sendo a estatística de teste $s_1 = s(\mathbf{e}_1, \mathbf{y}) = n_1^{-1} \mathbf{e}_1' \mathbf{y} - n_2^{-1} (\mathbf{1}' - \mathbf{e}_1)' \mathbf{y} = 3.0$.

Tabela 1.1 Vetores experimentais e estatísticas de teste

Vector experimental	Estatística de teste	Vector experimental	Estatística de teste
$e'_1 = (1, 1, 1, 1, 0, 0, 0, 0)$	3.0	$e'_{36} = (0, 1, 1, 1, 1, 0, 0, 0)$	2.0
$e'_2 = (1, 1, 1, 0, 1, 0, 0, 0)$	-0.5	$e'_{37} = (0, 1, 1, 1, 0, 1, 0, 0)$	2.5
$e'_3 = (1, 1, 1, 0, 0, 1, 0, 0)$	0.0	$e'_{38} = (0, 1, 1, 1, 0, 0, 1, 0)$	3.5
$e'_4 = (1, 1, 1, 0, 0, 0, 1, 0)$	1.0	$e'_{39} = (0, 1, 1, 1, 0, 0, 0, 1)$	4.0
$e'_5 = (1, 1, 1, 0, 0, 0, 0, 1)$	1.5	$e'_{40} = (0, 1, 1, 0, 1, 1, 0, 0)$	-1.0
$e'_6 = (1, 1, 0, 1, 1, 0, 0, 0)$	0.0	$e'_{41} = (0, 1, 1, 0, 1, 0, 1, 0)$	0.0
$e'_7 = (1, 1, 0, 1, 0, 1, 0, 0)$	0.5	$e'_{42} = (0, 1, 1, 0, 1, 0, 0, 1)$	0.5
$e'_8 = (1, 1, 0, 1, 0, 0, 1, 0)$	1.5	$e'_{43} = (0, 1, 1, 0, 0, 1, 1, 0)$	0.5
$e'_9 = (1, 1, 0, 1, 0, 0, 0, 1)$	2.0	$e'_{44} = (0, 1, 1, 0, 0, 1, 0, 1)$	1.0
$e'_{10} = (1, 1, 0, 0, 1, 1, 0, 0)$	-3.0	$e'_{45} = (0, 1, 1, 0, 0, 0, 1, 1)$	2.0
$e'_{11} = (1, 1, 0, 0, 1, 0, 1, 0)$	-2.0	$e'_{46} = (0, 1, 0, 1, 1, 1, 0, 0)$	-0.5
$e'_{12} = (1, 1, 0, 0, 1, 0, 0, 1)$	-1.5	$e'_{47} = (0, 1, 0, 1, 1, 0, 1, 0)$	0.5
$e'_{13} = (1, 1, 0, 0, 0, 1, 1, 0)$	-1.5	$e'_{48} = (0, 1, 0, 1, 1, 0, 0, 1)$	1.0
$e'_{14} = (1, 1, 0, 0, 0, 1, 0, 1)$	-1.0	$e'_{49} = (0, 1, 0, 1, 0, 1, 1, 0)$	1.0
$e'_{15} = (1, 1, 0, 0, 0, 0, 1, 1)$	0.0	$e'_{50} = (0, 1, 0, 1, 0, 1, 0, 1)$	1.5
$e'_{16} = (1, 0, 1, 1, 1, 0, 0, 0)$	0.5	$e'_{51} = (0, 1, 0, 1, 0, 0, 1, 1)$	2.5
$e'_{17} = (1, 0, 1, 1, 0, 1, 0, 0)$	1.0	$e'_{52} = (0, 1, 0, 0, 1, 1, 0, 1)$	-2.5
$e'_{18} = (1, 0, 1, 1, 0, 0, 1, 0)$	2.0	$e'_{53} = (0, 1, 0, 0, 1, 0, 1, 1)$	-2.0
$e'_{19} = (1, 0, 1, 1, 0, 0, 0, 1)$	2.5	$e'_{54} = (0, 1, 0, 0, 0, 1, 1, 1)$	-1.0
$e'_{20} = (1, 0, 1, 0, 1, 1, 0, 0)$	-2.5	$e'_{55} = (0, 0, 1, 1, 1, 1, 0, 0)$	-0.5
$e'_{21} = (1, 0, 1, 0, 1, 0, 1, 0)$	-1.5	$e'_{56} = (0, 0, 1, 1, 1, 0, 1, 0)$	0.0
$e'_{22} = (1, 0, 1, 0, 1, 0, 0, 1)$	-1.0	$e'_{57} = (0, 0, 1, 1, 1, 0, 0, 1)$	1.0
$e'_{23} = (1, 0, 1, 0, 0, 1, 1, 0)$	-1.0	$e'_{58} = (0, 0, 1, 1, 0, 1, 1, 0)$	1.5
$e'_{24} = (1, 0, 1, 0, 0, 1, 0, 1)$	-0.5	$e'_{59} = (0, 0, 1, 1, 0, 1, 0, 1)$	1.5
$e'_{25} = (1, 0, 1, 0, 0, 0, 1, 1)$	0.5	$e'_{60} = (0, 0, 1, 1, 0, 0, 1, 1)$	2.0
$e'_{26} = (1, 0, 0, 1, 1, 1, 0, 0)$	-2.0	$e'_{61} = (0, 0, 1, 0, 1, 1, 1, 0)$	3.0
$e'_{27} = (1, 0, 0, 1, 1, 0, 1, 0)$	-1.0	$e'_{62} = (0, 0, 1, 0, 1, 1, 0, 1)$	-2.0
$e'_{28} = (1, 0, 0, 1, 1, 0, 0, 1)$	-0.5	$e'_{63} = (0, 0, 1, 0, 1, 0, 1, 1)$	-1.5
$e'_{29} = (1, 0, 0, 1, 0, 1, 1, 0)$	-0.5	$e'_{64} = (0, 0, 1, 0, 0, 1, 1, 1)$	-0.5
$e'_{30} = (1, 0, 0, 1, 0, 1, 0, 1)$	0.0	$e'_{65} = (0, 0, 0, 1, 1, 1, 1, 0)$	0.0
$e'_{31} = (1, 0, 0, 1, 0, 0, 1, 1)$	1.0	$e'_{66} = (0, 0, 0, 1, 1, 1, 0, 1)$	-1.5
$e'_{32} = (1, 0, 0, 0, 1, 1, 1, 0)$	-4.0	$e'_{67} = (0, 0, 0, 1, 1, 0, 1, 1)$	-1.0
$e'_{33} = (1, 0, 0, 0, 1, 1, 0, 1)$	-3.5	$e'_{68} = (0, 0, 0, 1, 0, 1, 1, 1)$	0.0
$e'_{34} = (1, 0, 0, 0, 1, 0, 1, 1)$	-2.5	$e'_{69} = (0, 0, 0, 0, 1, 1, 1, 1)$	0.5
$e'_{35} = (1, 0, 0, 0, 0, 1, 1, 1)$	-2.0	$e'_{70} = (0, 0, 0, 0, 1, 1, 1, 1)$	-3.0

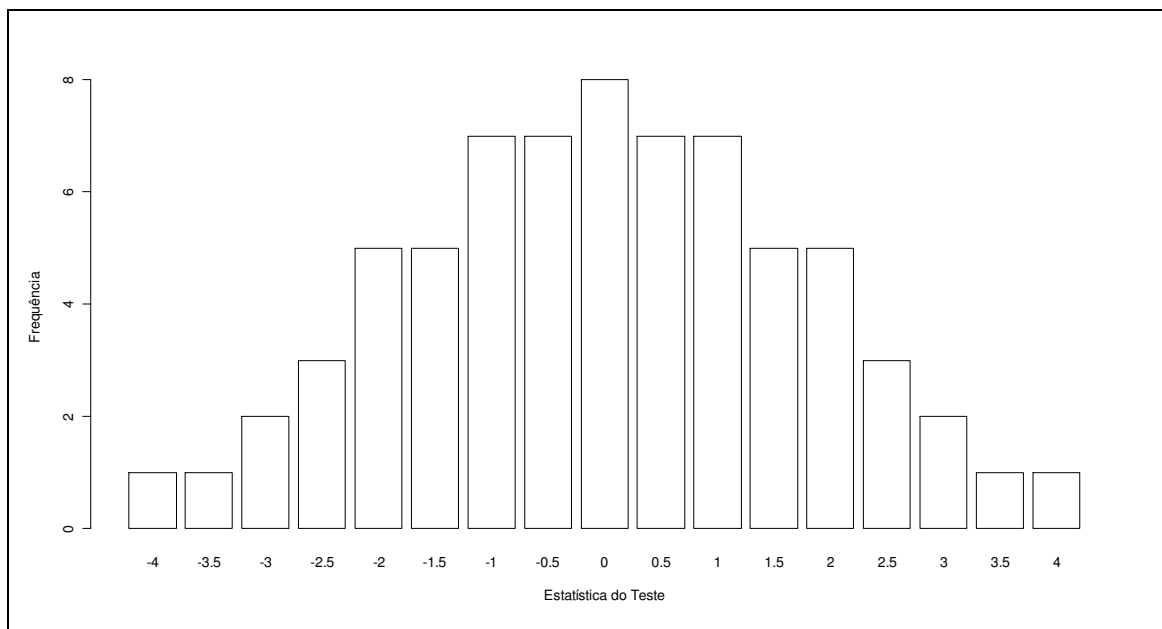
Para construir o conjunto de referência, a fim de determinar a probabilidade associada à estatística de teste, o investigador calcula os resultados que poderiam ter ocorrido, na condição de a hipótese nula ser verdadeira, para as outras possíveis afectações dos doentes às condições (já apresentados na Tabela 1.1).

A Tabela 1.2 mostra a distribuição de frequências para as 70 estatísticas de teste, que se representa graficamente na Figura 1.1.

Tabela 1.2 Distribuição de frequências das estatísticas de teste (ET)

ET	-4	-3.5	-3	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	3	3.5	4
Freq.	1	1	2	3	5	5	7	7	8	7	7	5	5	3	2	1	1

Figura 1.1 Gráfico da distribuição de frequências das estatísticas de teste



Como $\text{trunc}(k\alpha) = \text{trunc}(3.5) = 3$, s_{crit} é igual a $s_{(67)} = 3$. Sendo o valor da estatística observada não superior ao valor crítico, não há evidência

estatística para rejeitar a hipótese nula a favor da superioridade do Tratamento A.

Se o investigador tivesse usado o procedimento do valor-p, teria chegado à mesma conclusão. Existem outros três vectores que resultam em estatísticas iguais ou superiores à estatística observada: $s_{(38)}$ com uma estatística de 3.5; $s_{(39)}$ com uma estatística de 4.0; e $s_{(61)}$ com uma estatística de 3.0. Assim, o valor-p é igual $4/70 = 0.057$, e como este valor é superior ao valor de $\alpha = 0.05$, o investigador não tem evidência estatística conducente à rejeição da hipótese nula.

2.4. Escolha do conjunto de referência

Em geral, a escolha do conjunto de referência corresponde à escolha do planeamento da experiência. Considerações práticas podem levar, porém, a optar por um conjunto de referência não-exaustivo.

Na aplicação descrita no ponto anterior, foi possível enumerar todas as (70) combinações, dado o pequeno número de unidades experimentais em cada grupo. No entanto, com o aumento do número de unidades experimentais, o número de combinações aumenta exponencialmente (por exemplo, com 16 observações por grupo, o número de combinações seria superior a 600 milhões), não sendo possível realizar um teste de aleatorização num tempo razoável. Nestes casos pode e deve recorrer-se ao princípio de Monte Carlo, usando um subconjunto das k combinações. Para o fazer, extrai-se uma amostra aleatória, com reposição, de k^*-1 estatísticas de referência do conjunto das k estatísticas de referência. Assim, o tempo de cálculo é controlado pelo investigador, que determina a cardinalidade do conjunto de referência \mathcal{F}^* . (Usamos o sobrescrito *

para distinguir um teste de aleatorização não-exaustivo do teste de aleatorização exaustivo.)

Nesta dissertação só nos ocupamos de planeamentos de experiências completamente aleatorizadas com dois grupos. Mas os testes de aleatorização podem ser usados em outros tipos de planeamentos, como se constata se se consultarem, por exemplo, as obras de Edgington e Onghena (2007) ou de Manly (2007). Nessas obras são apresentados diversos planeamentos de experiências, incluindo planeamentos com medidas repetidas, com blocos aleatorizados, factoriais e multivariados.

2.5. Escolha da estatística de teste

No que respeita à estatística de teste, é escolhida uma estatística que reflecta o efeito antecipado pelo investigador. Nesse sentido, tendo em conta as suas expectativas, o experimentador, na aplicação acima descrita, escolheu como estatística de teste a diferença entre as médias dos dois tratamentos, $\bar{A} - \bar{B}$. Noutra experiência, outro investigador usaria uma estatística que reflectisse variabilidade ou qualquer outra característica dos dados que tivesse interesse em estudar.

É de notar que a estatística de teste, $\bar{A} - \bar{B}$, é equivalente à estatística *t* de Student.

Duas estatísticas de teste são equivalentes sob aleatorização se e unicamente se estiverem relacionadas de forma monótona estrita no conjunto de todos os vectores experimentais. Assim, duas estatísticas de teste equivalentes originam distribuições de referência com igual forma e, em consequência, o mesmo valor de probabilidade associado à estatística observada.

A estatística *t* de Student é:

$$t = \frac{\bar{A} - \bar{B}}{s \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}, \quad (1.10)$$

em que \bar{A} e \bar{B} são as médias dos dois grupos, n_A e n_B o número de observações em cada grupo e s é um estimador da variância, calculada a partir das variâncias dos dois grupos, s_A^2 e s_B^2 :

$$s = \sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}}. \quad (1.11)$$

No caso da estatística t de Student, o denominador é uma medida de escala, necessária para o teste paramétrico, mas sendo uma constante em todos os vectores experimentais, a sua remoção não tem qualquer efeito na ordenação das estatísticas de teste e, conseqüentemente, na probabilidade associada à estatística observada (valor- p).

A vantagem de usar estatísticas de teste equivalentes é a de simplificar o cálculo. No caso deste exemplo, o investigador poderia ter simplificado ainda mais o cálculo se usasse, como estatística de teste:

$$\sum_{i=1}^{n_A} A_i, \quad (1.12)$$

representativa da soma das observações do Tratamento A, tendo em consideração que, em cada vector experimental, a soma total e o número de observações são constantes. A estatística 1.12 continua a ser uma estatística equivalente à do teste t de Student.

Segundo Edgington e Onghena (2007), a estatística de teste de um teste de aleatorização é, em geral, mais simples que a de um teste paramétrico, pois apenas necessita de reflectir o tipo de efeito que é antecipado. No caso do teste paramétrico, a estatística deve incluir também uma estimativa da variabilidade do efeito antecipado.

Importa ainda referir que, quando se usa um teste paramétrico clássico, é necessário que tenha sido derivada a distribuição amostral da estatística do teste. No caso dos testes de aleatorização, como para cada conjunto de dados é construída a distribuição de referência, o investigador pode usar a estatística que julgue mais apropriada para detectar o efeito em que está interessado.

3. Validade do procedimento do teste de aleatorização

Um teste estatístico é válido se a probabilidade de rejeitar a hipótese nula, quando ela é verdadeira, é menor ou igual a α . É exactamente válido, quando essa probabilidade é igual a α .

3.1 Teste de aleatorização exaustivo

Demonstramos a validade dos testes de aleatorização exaustivos seguindo a apresentação de Onghena (1994), baseada nos argumentos de Hoeffding (1952) e Box e Anderson (1955), que provaram que os testes de aleatorização são válidos.

Para demonstrar a validade do teste de aleatorização exaustivo, é necessário considerar se $k\alpha$ é um inteiro (ou não) e se a estatística crítica

é igual (ou não) a outras estatísticas de referência que têm um número de ordem superior na ordenação apresentada em 1.3.

1) No caso de $k\alpha$ ser um inteiro e $s_{crit} = s_{(k(1-\alpha))} \neq s_{(k(1-\alpha)+1)}$, então

$$\begin{aligned} \sum_{j=1}^k I(s_j \leq s_{crit}) &= \sum_{j=1}^k I(s_j \leq s_{(k(1-\alpha))}) \\ &= k(1-\alpha) \end{aligned} \quad (1.13)$$

e o complemento é

$$\begin{aligned} \sum_{j=1}^k I(s_j > s_{crit}) &= k - (k(1-\alpha)) \\ &= k\alpha \end{aligned} \quad (1.14)$$

Como, sob a hipótese nula, a distribuição aleatória das unidades pelos tratamentos corresponde à extração aleatória de uma estatística de referência do conjunto de todas as possíveis estatísticas de referência, tem-se,

$$\begin{aligned} P(s_j \leq s_{crit}) &= k^{-1} \sum_{j=1}^k I(s_j \leq s_{crit}) \\ &= 1 - \alpha \end{aligned} \quad (1.15)$$

sendo o seu complemento

$$\begin{aligned} P(s_j > s_{crit}) &= k^{-1} \sum_{j=1}^k I(s_j > s_{crit}) \\ &= \alpha \end{aligned} \quad (1.16)$$

2) No caso de $k\alpha$ ser um inteiro e $s_{crit} = s_{(k(1-\alpha))} = s_{(k(1-\alpha)+1)} = \dots = s_{(k(1-\alpha)+\lambda)}$, em que λ , o número de estatísticas de referência iguais à estatística crítica, é igual ou superior a 1, então

$$\begin{aligned} \sum_{j=1}^k I(s_j > s_{crit}) &= k - (k(1-\alpha)) - \lambda \\ &= k\alpha - \lambda \end{aligned} \quad (1.17)$$

e, sob a hipótese nula,

$$\begin{aligned} P(s_j > s_{crit}) &= k^{-1} \sum_{j=1}^k I(s_j > s_{crit}) \\ &= \alpha - \frac{\lambda}{k} \\ &< \alpha \end{aligned} \quad (1.18)$$

3) No caso de $k\alpha$ não ser um inteiro e $s_{crit} = s_{(k-trunc(k\alpha))} \neq s_{(k-trunc(k\alpha)+1)}$, então

$$\begin{aligned} \sum_{j=1}^k I(s_j \leq s_{crit}) &= \sum_{j=1}^k I(s_j \leq s_{(k-trunc(k\alpha))}) \\ &= k - trunc(k\alpha) \end{aligned} \quad (1.19)$$

e o seu complemento é

$$\begin{aligned} \sum_{j=1}^k I(s_j > s_{crit}) &= k - (k - trunc(k\alpha)) \\ &= trunc(k\alpha) \end{aligned} \quad (1.20)$$

Sob a hipótese nula,

$$\begin{aligned}
P(s_j \leq s_{crit}) &= k^{-1} \sum_{j=1}^k I(s_j \leq s_{crit}) & (1.21) \\
&= 1 - \frac{trunc(k\alpha)}{k}
\end{aligned}$$

e o complemento é

$$\begin{aligned}
P(s_j > s_{crit}) &= k^{-1} \sum_{j=1}^k I(s_j > s_{crit}) & (1.22) \\
&= \frac{trunc(k\alpha)}{k} \\
&< \alpha
\end{aligned}$$

4) No caso de $k\alpha$ não ser um inteiro e $s_{crit} = S_{(k-trunc(k\alpha))} = S_{(k-trunc(k\alpha)+1)} = \dots = S_{(k-trunc(k\alpha)+\lambda)}$, então

$$\begin{aligned}
\sum_{j=1}^k I(s_j > s_{crit}) &= k - (k - trunc(k\alpha)) - \lambda & (1.23) \\
&= trunc(k\alpha) - \lambda
\end{aligned}$$

e, sob a hipótese nula

$$\begin{aligned}
P(s_j > s_{crit}) &= k^{-1} \sum_{j=1}^k I(s_j > s_{crit}) & (1.24) \\
&= \frac{trunc(k\alpha) - \lambda}{k} \\
&< \alpha
\end{aligned}$$

Assim, podemos concluir que o teste de aleatorização é exactamente válido no caso de $k\alpha$ ser um inteiro e a estatística crítica não ser igual a uma ou mais estatísticas de referência com um número de ordem superior na ordenação 1.3: a probabilidade de cometer um erro de Tipo I é igual a

α (Equação 1.16). Nos outros casos, o teste é válido, sendo a probabilidade de cometer um erro de Tipo I inferior a α (Equações 1.18, 1.22 e 1.24).

3.2 Teste de aleatorização não-exaustivo

No que concerne a validade dos testes de aleatorização não-exaustivos, Dwass (1957), invocando alguns pressupostos, e Noreen (1986) mostraram que os testes de aleatorização não-exaustivos são válidos e Foutz (1980) provou que podem ser construídos testes Monte Carlo exactamente válidos, empregando uma variável aleatória auxiliar para separar valores repetidos.

Note-se que $k^* \alpha$ é habitualmente um inteiro, pois o número k^* de estatísticas de referência é definido pelo investigador. O problema surge com valores repetidos nas estatísticas de referência. No entanto, o emprego da variável aleatória auxiliar permite distinguir esses valores e, sob a hipótese nula, a probabilidade associada à estatística do teste é igual a α . Se não se usar essa variável aleatória auxiliar, o teste continua a ser válido, pois, sob a hipótese nula, a probabilidade associada à estatística do teste é igual ou inferior a α . Note-se que, excepto se $s_{\text{obs}} = s_j$, para algum j , o teste será exactamente válido.

CAPÍTULO 2 - POTÊNCIA ESTATÍSTICA

1. Definição de potência estatística

A potência de um teste estatístico é a probabilidade de rejeitar uma hipótese nula falsa.

Quando se realiza um teste estatístico, tomam-se decisões correctas quando se rejeita uma hipótese nula falsa ou não se rejeita uma hipótese nula verdadeira. Mas podem-se tomar decisões incorrectas quando se rejeita uma hipótese nula verdadeira (erro de Tipo I) ou não se rejeita uma hipótese nula falsa (erro de Tipo II).

A probabilidade de cometer um erro de Tipo I é designada por α e a probabilidade de cometer um erro de Tipo II é designada por β . Assim, a potência de um teste é o complemento de β .

Os três determinantes da potência de um teste são: o nível de significância α ; o tamanho da amostra n ; e o tamanho do efeito δ .

A potência será tanto mais elevada quanto maiores forem os valores de α , de n e de δ .

2. A função potência

A função potência de um teste estatístico é a potência estatística em função do tamanho do efeito, para um dado nível de significância e tamanho da amostra.

No contexto de estudos de potência, a potência de um teste é definida independentemente da verdade da hipótese nula. Assim, a potência é o complemento de β , quando a hipótese nula é falsa; mas se a hipótese nula

é verdadeira, a potência é a probabilidade de cometer um erro de Tipo I, não devendo ser maior que o nível de significância α .

Para ilustrar as relações entre potência e nível de significância, tamanho da amostra e o tamanho do efeito, apresentamos graficamente a potência do teste t de Student para várias combinações desses parâmetros.

Na Figura 2.1 apresentam-se as funções potência para os testes t de Student unicaudal à esquerda, unicaudal à direita e bicaudal para grupos independentes, para um nível de significância de 0.05 e amostras com 8 observações cada, sob os pressupostos de normalidade, homocedasticidade e independência. O tamanho do efeito é a diferença entre as médias populacionais em unidades de desvio-padrão:

$$\delta = \frac{\mu_1 - \mu_2}{\sigma} \quad (2.1)$$

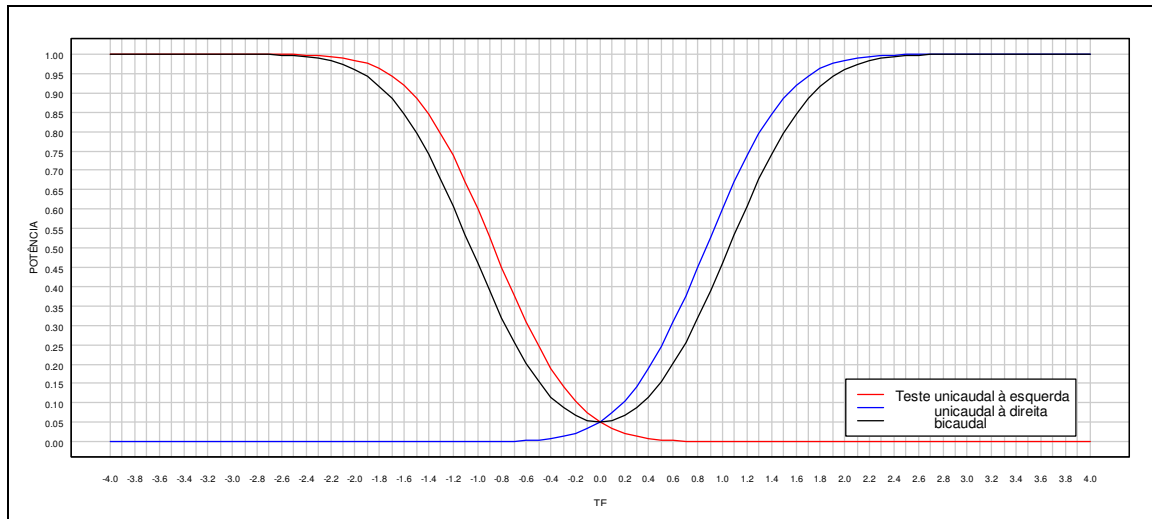
em que:

μ_1 = média da primeira população;

μ_2 = média da segunda população; e

σ = desvio-padrão de qualquer das populações.

Figura 2.1 Funções potência para os testes t de Student unicaudal à esquerda, unicaudal à direita e bicaudal, com $\alpha = 0.05$ e $n_1 = n_2 = 8$, sob os pressupostos paramétricos clássicos

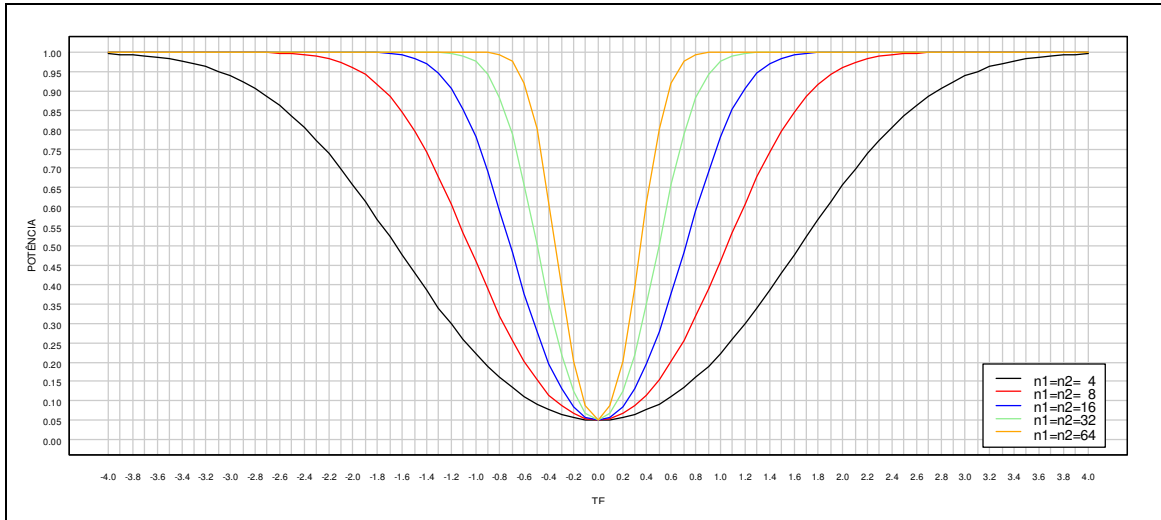


Podemos verificar que, para o teste t de Student unicaudal à esquerda, a potência aumenta em função inversa do tamanho do efeito; para o teste t de Student unicaudal à direita, a potência aumenta em função do tamanho do efeito; e para o teste t de Student bicaudal, a potência aumenta em função do valor absoluto do tamanho do efeito

Quando o tamanho do efeito é igual a zero, a potência é igual ao nível de significância.

Para valores negativos do tamanho do efeito, o teste unicaudal à esquerda é mais potente que o teste bicaudal; para valores positivos, o teste unicaudal à direita é mais potente que o teste bicaudal.

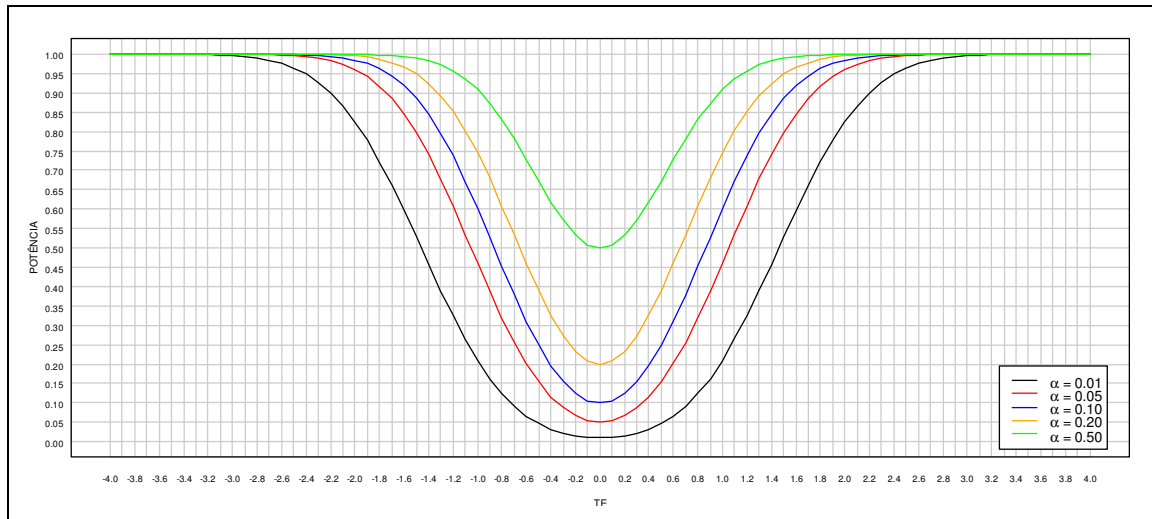
Figura 2.2 Funções potência para o teste t de Student bicaudal, com $\alpha = 0.05$, para grupos equilibrados com 4, 8, 16, 32 e 64 elementos, sob os pressupostos paramétricos clássicos



A Figura 2.2 mostra a potência do teste t de Student bicaudal em função do tamanho do efeito e do tamanho das amostras (com $n_1 = n_2$), para o nível de significância de 0.05, sob os pressupostos paramétricos clássicos.

Pode-se verificar que a potência, para um determinado valor do efeito, é uma função do tamanho da amostra: quanto maior a amostra, maior a potência. A potência mínima é de 0.05, para todos os tamanhos das amostras, que corresponde ao valor do nível de significância.

Figura 2.3 Funções potência para o teste t de Student bicaudal, com $n_1 = n_2 = 8$, para diversos valores de α , sob os pressupostos paramétricos clássicos



A Figura 2.3 mostra a potência do teste t bicaudal em função de α , para $n_1 = n_2 = 8$, sob os pressupostos paramétricos clássicos. Podemos ver que, para cada tamanho do efeito, a potência é tanto maior quanto mais elevado o nível de significância. As probabilidades dos erros de Tipo I e de Tipo II estão inversamente relacionadas, o que implica que não se podem minimizar ambos os erros simultaneamente.

3. Cálculo da potência estatística

Para determinar a potência de um teste, podem ser usadas as tabelas e gráficos de obras como, por exemplo, as de Kraemer e Thiemann (1987), Cohen (1988), Lipsey (1990) e Murphy e Myors (1998).

Kraemer e Thiemann (1987) derivaram um modelo para a análise da potência estatística baseado no coeficiente de correlação intraclasse. Desenvolveram métodos para expressar muitos testes estatísticos em

termos desse coeficiente, o que lhes possibilitou apresentar uma tabela geral baseada no coeficiente de correlação intraclasse.

Na obra mais completa sobre análise de potência, Cohen (1988) apresentou tabelas para um grande número de estatísticas, usando diferentes medidas de tamanho do efeito e cálculos de potência para cada tipo de estatísticas.

Já Lipsey (1990), para calcular a potência de vários testes estatísticos, usou como base o teste t de Student, e Murphy e Myers (1998) a estatística F (ou uma transformação dessa estatística).

Este era o procedimento corrente há alguns anos atrás, embora, por vezes, não se encontrassem os valores requeridos ou fosse necessário interpolar para os obter.

Hoje, para calcular a potência da maioria dos testes estatísticos existem diversos programas informáticos específicos, como, por exemplo, o G*Power 3 de Faul et al. (2007).

Mas muitos programas informáticos estatísticos de uso geral possibilitam o cálculo da potência, incluindo funções para distribuições não-centrais. É o caso do programa R (R Development Core Team, 2009), que usamos para calcular as funções potência que foram apresentadas graficamente nas Figuras 2.1, 2.2 e 2.3.

A distribuição t não-central depende, para além dos usuais graus de liberdade, do chamado parâmetro não-central. Quando o valor deste parâmetro é igual a zero, obtém-se a distribuição t usual.

A potência de um teste t de Student pode ser calculada a partir da distribuição t não-central (Johnson, Kotz & Balakrishnan, 1995).

A distribuição t não-central com ν graus de liberdade e parâmetro não-central λ define-se como a distribuição de

$$t_{\nu} = \frac{U + \lambda}{\sqrt{\frac{X}{\nu}}} \quad (2.2)$$

em que U e X são variáveis aleatórias independentes, $U \sim N(0,1)$ e $X \sim \chi^2(\nu)$, e λ é uma constante.

A função densidade de probabilidade é dada por

$$f(x) = \frac{\nu^{\nu/2} e^{-\lambda^2/2}}{\sqrt{\pi} \Gamma(\nu/2) (\nu + x^2)^{(\nu+1)/2}} \times \sum_{j=0}^{\infty} \Gamma\left(\frac{\nu + j + 1}{2}\right) \left(\frac{\lambda^j}{j!}\right) \left(\frac{2x^2}{\nu + x^2}\right)^{j/2} \quad (2.3)$$

com $x, \lambda \in R$, $\nu \in N$ e em que $\Gamma(x)$ é a função gama.

A função densidade de probabilidade é enviesada para a direita. O parâmetro da forma é ν e o parâmetro de não-centralidade é λ . Para um valor fixo de ν , a cauda torna-se mais pesada quando λ aumenta. Para valores elevados de ν , a função densidade de probabilidade é aproximadamente simétrica.

Os dois primeiros momentos são:

$$\mu_1 = \frac{\Gamma((\nu-1)/2) \sqrt{\nu/2}}{\Gamma(\nu/2)} \lambda \quad (2.4.1)$$

$$\mu_2 = \frac{\nu}{\nu-2} (1 + \lambda^2) - \left(\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}\right)^2 (\nu/2) \lambda^2 \quad (2.4.2)$$

Sob a hipótese nula ($\mu_1 = \mu_2$), a estatística t tem uma distribuição t com $n_1 + n_2 - 2$ graus de liberdade:

$$t \sim t'_{n_1+n_2-2} (\lambda = 0) \quad (2.5)$$

Sob uma hipótese alternativa, a estatística t tem uma distribuição t não-central com $n_1 + n_2 - 2$ graus de liberdade:

$$t \sim t'_{n_1+n_2-2} \left(\lambda = \frac{\mu_1 - \mu_2}{\sigma} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \right) \quad (2.6)$$

Comparando a equação 2.6 com a equação 2.1, que define o tamanho do efeito de Cohen, podemos ver que o parâmetro não-central é igual ao tamanho do efeito vezes a raiz quadrada de metade da média harmônica dos tamanhos das amostras:

$$\lambda = \delta \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \quad (2.7)$$

Para o cálculo da potência do teste t de Student usamos, como dissemos, o programa R. É fácil calcular a potência com este programa, pois tem uma função, 'power.t.test', que calcula a potência do teste ou determina os parâmetros para obter uma dada potência. No Anexo I apresentamos o programa que foi usado para calcular as funções potência representadas na Figura 2.1.

4. Problemas da Análise de Potência

A análise de potência levanta alguns problemas, tanto ao nível da teoria, como na sua utilização prática.

4.1. Problemas teóricos

Nem todas as escolas de inferência estatística aceitam o conceito de potência estatística, tal como foi definido nos pontos anteriores. Relatamos, de forma resumida, a introdução desse conceito por Neyman e Egon Pearson.

Foi Fisher que fundou a estatística inferencial moderna, tornando-a acessível aos investigadores em várias áreas do conhecimento. Na sequência do trabalho de Gossett (1908), Fisher dá uma fundamentação matemática às conjecturas de Gossett e amplia os seus resultados, que servem de base a uma série de testes de hipótese associados a diferentes modelos estatísticos, que apresenta num conjunto de artigos que culminam na sua influente obra ‘Statistical methods for research workers’ (Fisher, 1925).

Para Fisher, os principais elementos de um teste são os seguintes:

1. uma hipótese nula;
2. uma estatística de teste;
3. a distribuição dessa estatística, dada a hipótese nula; e
4. a probabilidade associada à estatística do teste (o valor p).

O valor p é visto como uma indicação da plausibilidade da hipótese nula, tendo em conta os dados observados. A rejeição ou aceitação da hipótese nula é uma outra questão, é uma questão separada.

O programa de investigação desenvolvido por Neyman e Pearson (1928a, 1928b, 1933), tem como motivação aquilo que consideravam a falta de base lógica, na formulação de Fisher, para a escolha de estatísticas de teste e o uso do valor p como medida de credibilidade da hipótese nula.

A solução que propuseram foi a de encarar um teste de hipótese como uma escolha entre hipóteses rivais, introduzindo, para além da noção de hipótese nula, a noção de uma hipótese alternativa. Assim, a finalidade de um teste de hipótese deixava de ser o de calcular uma medida de credibilidade que os dados forneciam à hipótese nula, para passar a ser o de decidir rejeitar ou não rejeitar a hipótese nula com base nos dados observados. O objectivo de um teste é o de formular uma regra de decisão, com base numa partição do espaço amostral em dois conjuntos complementares, designados como região de aceitação e região de rejeição.

Para Neyman e Pearson, os principais elementos de um teste são:

1. uma hipótese nula e uma hipótese alternativa;
2. uma estatística de teste;
3. a distribuição dessa estatística, dada a hipótese nula;
4. o nível de significância, α ;
5. a região de rejeição

A decisão de rejeitar ou não rejeitar a hipótese nula implica a possibilidade de cometer dois tipos de erros. Comete-se um erro de Tipo I quando se rejeita uma hipótese nula verdadeira; um erro de Tipo II, quando não se rejeita uma hipótese falsa. Esses erros são interpretados por Neyman e Pearson em termos da repetição da experiência um grande número de vezes.

Como existe uma relação inversa entre a probabilidade de cometer os dois tipos de erros, não sendo possível minimizá-los simultaneamente, é necessário encontrar um compromisso.

A solução encontrada por estes autores foi a de considerarem a hipótese nula como mais importante que a alternativa, o que significa estabelecer

uma probabilidade baixa para o erro de Tipo I e escolher um teste que minimize a probabilidade do erro de Tipo II (ou, por outras palavras, que maximize a sua potência).

Neyman e Pearson demonstraram, com o seu famoso lema, que o teste a escolher é o teste uniformemente mais potente, para um dado nível de significância α , podendo ser usado o critério da razão de verosimilhanças como método geral para a construção de testes.

Estes autores pensavam que a sua teoria era um desenvolvimento das ideias expressas por Fisher em 1925, mas Fisher sempre se opôs fortemente ao conceito de dois tipos de erros e aos cálculos de potência.

Gigerenzer (1987, 1993), Gigerenzer et al. (1989), Cowles (1989) e Oakes (1990) dão conta da polémica entre Fisher e Neyman e das implicações para a investigação em Psicologia e outras ciências sociais.

A posição teórica que está subjacente à definição de potência estatística que apresentámos nos pontos anteriores é a proposta por Neyman e Pearson (1933), com a interpretação de probabilidade frequentista. É a corrente principal na estatística matemática (por exemplo, Lehmann & Romano, 2005; Mood, Graybill, & Boes, 1974), bem como nas ciências sociais e do comportamento (por exemplo, Hays, 1994; Kirk, 1999; Winer, Brown & Michels, 1991).

4.2. Problemas práticos

Quando se realiza uma experiência surgem problemas práticos quando se pretende efectuar o cálculo de potência *a priori*. O investigador, desejando saber qual deverá ser o tamanho da sua amostra, deverá

especificar um tamanho de efeito e definir os custos dos erros de Tipo I e II.

Em muitos casos não poderá especificar uma hipótese alternativa simples, mas poderá indicar um leque de valores do tamanho do efeito dignos de consideração. Um investigador não terá base estatística para escolher entre uma amostra com 8 elementos ou uma com 80, se não puder especificar o tamanho do efeito em que está interessado.

Lipsey (1990) sugeriu três estratégias que um investigador pode seguir, ao planear um estudo, para fazer uma estimativa do mínimo tamanho de efeito que o estudo deveria detectar.

Uma primeira estratégia é a de escolher o tamanho do efeito com base nos valores de tamanho de efeito detectados em investigações anteriormente publicadas que usem as mesmas medidas e tratamentos.

Outra estratégia que o investigador pode seguir é a de escolher um tamanho de efeito baseado em diferenças de médias que considere relevantes, sendo necessário fazer uma estimativa do desvio-padrão, que poderá ser feita a partir de estudos anteriores ou com um estudo-piloto.

Finalmente, o tamanho do efeito poderá ser baseado na comparação de dois grupos não-experimentais que difiram em termos práticos na variável em que o investigador está interessado. Importa avaliar a relevância prática de efeito do tratamento pois que, como Prentice e Miller (1992) mostraram, tamanhos de efeito pequenos podem ter consequências práticas importantes, enquanto grandes tamanhos de efeito podem resultar em efeitos que, do ponto de vista prático, sejam negligenciáveis.

Cohen (1988) indica alguns valores convencionais para tamanhos de efeito pequenos, médios e grandes nas ciências do comportamento. Para o teste t de Student para grupos independentes, esses valores são, respectivamente, de 0.20, 0.50 e 0.80.

Lipsey (1990) forneceu suporte empírico a esses valores convencionais. No seu estudo, que engloba os resultados de 102 meta-análises, que incluem os resultados de 6 700 estudos individuais na área das ciências do comportamento, principalmente em educação, Lipsey ordenou os tamanhos dos efeitos positivos e dividiu-os em três grupos: o primeiro grupo compreendia 33% dos valores mais baixos, o segundo 34% dos valores médios e o terceiro grupo 33% dos valores mais altos. As medianas dos valores do tamanho do efeito nesses três grupos foram, respectivamente, de 0.15, 0.45 e 0.90, que são semelhantes aos valores propostos por Cohen.

Mais difícil é o problema de quantificar os custos associados aos erros de Tipo I e de Tipo II. Hays (1994) afirma que os investigadores nas ciências sociais e do comportamento raramente podem quantificar as perdas associadas aos dois tipos de erros. Contudo, tendo em conta as convenções respeitantes ao controlo do erro de Tipo I, não há razões para não considerar as consequências dos erros de Tipo II. Poderá ser possível considerar os custos relativos dos erros de Tipo I e de Tipo II, mesmo quando é difícil considerar os custos absolutos de cada um dos erros isoladamente.

Nas ciências sociais e humanas, e nomeadamente em Psicologia, a noção de potência estatística foi introduzida por Cohen (1962). Defendeu a necessidade de os investigadores tomarem em consideração o erro de Tipo II e, com a publicação da sua manual sobre análise de potência (Cohen, 1969), revista em 1977 e actualizada em 1988, permitiu que a análise de potência fosse realizada de forma prática.

Este autor chamou a atenção para o facto de que negligenciar a análise de potência na fase de planeamento de um estudo pode implicar desperdiçar esforço, tempo e dinheiro, pois pode suceder que com o plano, tratamentos, medidas e tamanhos de amostras escolhidas seja virtualmente impossível obter um resultado estatisticamente significativo para tamanhos de efeito razoáveis. Aliás, em áreas específicas de investigação têm sido documentados os efeitos prejudiciais da fraca potência estatística (Cohen, 1962; Rosenthal, 1990; Rosnow & Rosenthal, 1989; Sedlmeier & Gigerenzer, 1989).

Importa ainda considerar o problema dos pressupostos paramétricos para a realização do cálculo de potência. Para os testes paramétricos, o problema não se põe: os testes paramétricos são unicamente válidos se se verificarem os seus pressupostos e o cálculo de potência não requer pressupostos adicionais (sendo apenas necessário especificar o tamanho do efeito). Contudo, para os testes não-paramétricos pode existir um problema, pois que, para o cálculo de potência, para além dos pressupostos associados com a hipótese nula, podem ser necessários outros pressupostos associados com a hipótese alternativa. A análise de potência dependerá, também, da validade desses outros pressupostos.

CAPÍTULO 3 - TESTES DE ALEATORIZAÇÃO E ANÁLISE DE POTÊNCIA

1. Introdução

A potência estatística dos testes de aleatorização não tem sido muito estudada, talvez porque o conceito de teste de aleatorização tenha a sua origem em Fisher e o conceito de potência estatística pertença à escola de Neyman-Pearson.

Se se consultarem os manuais sobre testes de aleatorização de Eggington e Onghena (2007), Lunneborg (2000), Manly (2007), Noreen (1989), não se encontram indicações sobre o número de participantes a incluir em cada tratamento com base em considerações de potência; e as obras sobre análise de potência referidas no capítulo 2, como a de Cohen (1988), não incluem indicações sobre a maneira de calcular a potência dos testes de aleatorização.

Possivelmente, para além das origens diferentes destes conceitos, uma outra razão poderá explicar a falta de interesse em estudar a potência dos testes de aleatorização: quando se realiza um teste de aleatorização, não são feitos pressupostos distribucionais, mas é impossível realizar uma análise de potência *a priori* sem pressupostos sobre as respostas que se podem esperar.

Em todo o caso, a análise de potência é a única ferramenta estatística que pode ser usada para determinar o número de participantes a incluir numa experiência e, por isso, pode valer a pena fazer pressupostos adicionais, ainda que a validade do cálculo de potência dependa da validade desses pressupostos adicionais.

2. Diferentes abordagens para o cálculo da potência

Foram propostas várias abordagens para calcular a potência dos testes não-paramétricos, no quadro do modelo de distribuição aleatória e do modelo de amostragem aleatória.

No quadro do modelo de distribuição aleatória, assumem-se valores básicos específicos e uma função do efeito para calcular a potência (Gabriel & Hall, 1983; Kempthorne & Doerfler, 1969). Embora interessante do ponto de vista teórico, esta abordagem é demasiado pesada em termos de computação e os pressupostos referentes aos valores básicos difíceis de defender.

No quadro da distribuição aleatória, pode usar-se a técnica *bootstrap* se não for especificada uma distribuição populacional (Hamilton & Collings, 1991) ou a técnica Monte Carlo, se for especificada uma distribuição populacional (Lehmann, 1975; Milton, 1970).

Para usar a técnica *bootstrap*, no quadro da análise de potência, é necessário realizar previamente um estudo-piloto, a partir do qual se deriva uma função de distribuição empírica que serve para estimar a distribuição populacional e a partir da qual se extraem amostras, com reposição, para estimar a potência.

Não usaremos esta técnica, nesta dissertação, pois ela implica a realização, para cada caso, de um estudo-piloto. Vamos utilizar a técnica Monte Carlo, assumindo diversas distribuições populacionais.

Embora para testes não-paramétricos baseados em números de ordem (*ranks*) seja possível usar uma abordagem algébrica (ver, por exemplo, Lehmann, 1975), para dados não transformados em números de ordem tal abordagem não é possível.

Também a abordagem assintótica tem sido sugerida na literatura (Kraemer & Thiemann, 1987; Siegel & Castellan, 1988) para calcular a potência de um teste não paramétrico baseado em números de ordem. Consiste em calcular a potência do teste paramétrico correspondente e ajustar esse valor pela eficiência relativa assintótica do teste de não-paramétrico baseado em números de ordem em comparação com o teste paramétrico.

Hoeffding (1951,1952) demonstrou que os testes de aleatorização são assintoticamente tão potentes como os testes paramétricos convencionais, quando os pressupostos destes testes se verificam. A eficiência relativa assintótica do teste de aleatorização é igual a 1 no caso de duas amostras extraídas de populações com distribuição normal, tanto em termos de eficiência Pitman (Lehmann, 1986), como em termos de eficiência Bahadur (Bahadur & Raghavachari, 1970). Em consequência, poderíamos estimar a potência de um teste de aleatorização calculando a potência do teste t de Student. No entanto, este resultado é obtido assumindo uma amostra infinitamente grande e essa estimativa pode ser pouco precisa no caso de pequenas amostras.

Assim, vamos simular dados de distribuições populacionais específicas e a proporção de rejeições da hipótese nula servirá para estimar a potência do teste de aleatorização. Este estimador é designado por Hammersley e Handscomb (1964) 'crude Monte Carlo estimator', sendo um estimador não-enviesado da potência do teste.

3. Nível real e nível nominal

Importa ainda referir um problema que pode surgir ao analisar a função potência do teste de aleatorização. Distinguem-se o nível real do teste (α_0) e o seu nível nominal (α). A probabilidade de cometer um erro de Tipo I, ou seja, a potência do teste quando o tamanho do efeito é igual a zero, é o nível real do teste, por oposição ao seu nível nominal ou nível de significância.

No que concerne um teste paramétrico, quando os seus pressupostos se verificam, o nível real do teste é igual ao nível nominal. Mas o mesmo não se passa com um teste não-paramétrico, com uma distribuição de referência discreta: neste caso, o nível real é muitas vezes inferior ao nominal.

Como referimos, em capítulo anterior, o nível real de um teste de aleatorização é inferior ao seu nível nominal em duas situações. Em primeiro lugar, quando $k\alpha$ não for um inteiro; em segundo lugar, quando a estatística crítica for igual a outras estatísticas do conjunto de referência que tem um número de ordem superior na ordenação 1.3, ou seja $s_{\text{crit}} = s_{(k(1-\alpha))} = s_{(k(1-\alpha)+1)} \dots = s_{(k(1-\alpha)+\lambda)}$.

Nos estudos por simulação que vamos efectuar, na parte prática desta dissertação, a primeira situação não se verifica, tendo em conta os valores de k e α que iremos escolher, de modo a que $k\alpha$ seja um inteiro.

Relativamente à segunda situação, como as nossas ‘medidas’, valores simulados de distribuições contínuas, são muito ‘precisas’, seria improvável a repetição de valores das estatísticas no conjunto de referência, se usássemos conjuntos de referência exaustivos.

No entanto, tendo em conta os tamanhos das amostras que vamos simular, não é possível usar conjuntos de referência exaustivos. Vamos avaliar a potência do teste de aleatorização na comparação de dois grupos independentes com 8, 16 e 32 elementos por grupo. Com 8 elementos por grupo, o número de combinações é de 12 870; com 16 elementos por grupo, o número de combinações é já de 601 080 390; e com 32 elementos por grupo, o número de combinações é de 1.83262×10^{18} . Usaremos, por isso, uma amostra aleatória de 1 000 dessas combinações³.

Assim, com conjuntos de referência não-exaustivos, existe a possibilidade de repetições dos valores das estatísticas no conjunto de referência, visto ser possível extrair o mesmo vector mais de uma vez. Isso implica que a potência estimada será inferior a que seria obtida com conjuntos de referência exaustivos.

O número de vectores distintos na amostra (de combinações) é uma variável aleatória com a seguinte distribuição de probabilidade (Barton & David, 1959; Johnson, Kemp & Kotz, 2005):

$$P(V_{m,n} = j) = \binom{m}{j} \sum_{k=0}^j (-1)^k \binom{j}{k} \left(\frac{j-k}{m}\right)^n, \quad j \in \{1, 2, \dots, \min(m, n)\}, \quad (3.1)$$

em que m é o tamanho da população e n o tamanho da amostra.

O valor esperado e a variância dessa distribuição são, respectivamente:

$$E(V_{m,n}) = m \left(1 - \left(1 - \frac{1}{m}\right)^n\right) \quad (3.2.1)$$

³ Na Parte II, na introdução metodológica, justificamos a escolha do número de amostras simuladas em cada condição (20 000) e do número de reamostragens efectuadas em cada amostra (999) para obter um conjunto de referência com 1 000 estatísticas.

$$\text{var}(V_{m,n}) = m(m-1) \left(1 - \frac{2}{m}\right)^n + m \left(1 - \frac{1}{m}\right)^n - m^2 \left(1 - \frac{1}{m}\right)^{2n} \quad (3.2.2)$$

No caso das simulações com 16 e 32 elementos por grupo, a probabilidade de vectores repetidos é praticamente nula. O mesmo não acontece no caso das simulações com 8 elementos por grupo, pois a fracção amostra/população é elevada: 1000/12870 (cerca de 8%).

Aqui a média de valores distintos é de 962.2, com um desvio-padrão de 5.84.⁴

Para verificar qual a perda de potência em que podíamos incorrer, nesta situação, realizámos um estudo por simulação em que comparámos a potência do teste de aleatorização calculada a partir do conjunto de referência exaustivo e a potência estimada a partir de uma amostra de 1000 combinações.

De distribuições normais, gerámos 5 conjuntos de 20 000 amostras de 16 elementos, 8 por grupo. A média do primeiro grupo (primeiros 8 elementos), variava em termos do tamanho do efeito, com os seguintes quinze valores: -4.00, -3.00, -2.00, -1.00, -0.75, -0.50, -0.25, 0.00, 0.25, 0.50, 0.75, 1.00, 2.00, 3.00 e 4.00. A média do segundo grupo era igual a 0, sendo o desvio-padrão de ambos os grupos igual a 1.

Em cada amostra testámos a diferença entre grupos com uma estatística equivalente à diferença das médias, com $\alpha = 0.05$, para testes unicaudais

⁴ Para calcular, com o programa R, a média e a variância dos valores distintos, escrevemos as seguintes funções:

```
med.nvd<-function(m,n){m * (1-(1-(1/m))^n)}
```

```
var.nvd<-function(m,n){m * (m-1) * (1-(2/m))^n + m * (1-(1/m))^n  
- m^2 * (1-(1/m))^(2*n)}
```

à esquerda, testes unicaudais à direita e testes bicaudais, usando quer a totalidade das combinações (conjunto de referência exaustivo), quer uma amostra aleatória de 1 000 dessas combinações (conjunto de referência não-exaustivo).

Para cada grupo de 20 000 amostras, e para cada tamanho do efeito, a potência foi calculada dividindo o número de amostras em que se rejeitou a hipótese nula pelo número total de amostras.

No Anexo II apresentamos o programa em R que usámos para comparar a potência e no Anexo III as matrizes com os resultados das simulações.

Na Tabela 3.1 são apresentadas estatísticas descritivas (Mínimo, Máximo e Média) da diferença entre a potência estimada com o conjunto de referência exaustivo e a potência estimada com o conjunto de referência não-exaustivo, para os cinco conjuntos de 20 000 amostras.

Tabela3.1 Estatísticas descritivas (mínimo, máximo e média) da diferença entre a potência do teste de aleatorização exaustivo e não-exaustivo, para os cinco conjuntos de 20 000 amostras.

Estatísticas	Conjunto	Teste		
		unicaudal à esquerda	unicaudal à direita	bicaudal
Mínimo	1	-0.0014	-0.0001	-0.0008
	2	-0.0005	-0.0007	-0.0011
	3	-0.0009	-0.0007	-0.0005
	4	-0.0009	-0.0002	0.0000
	5	-0.0001	0.0000	0.0000
Máximo	1	0.0035	0.0033	0.0025
	2	0.0016	0.0011	0.0013
	3	0.0025	0.0026	0.0021
	4	0.0038	0.0025	0.0022
	5	0.0030	0.0022	0.0022
Média	1	0.0005	0.0010	0.0006
	2	0.0002	0.0004	0.0000
	3	0.0007	0.0004	0.0004
	4	0.0005	0.0005	0.0008
	5	0.0008	0.0005	0.0006

Como se pode verificar, nos cinco conjuntos de 20 000 amostras, as diferenças de potência são pequenas, com as perdas máximas a variarem entre 0.0011 e 0.0038. Se observarmos, em cada conjunto, os valores das médias das diferenças de potência para os quinze valores de tamanho do efeito, podemos verificar que as perdas de potência não são relevantes.

Estes resultados sugerem, assim, não ser um problema o facto de, para o caso de amostras com 8 elementos por grupo, a fracção amostra/população ser relativamente elevada, levando a valores repetidos no conjunto de referência.

PARTE II

1. Introdução

Nos três estudos que apresentamos seguidamente avaliamos o comportamento dos testes de aleatorização, em termos de potência, no caso de planejamentos de experiências completamente aleatorizadas com dois grupos. Para efeitos de comparação, em todos os estudos avaliámos também a potência alcançada pelo teste t de Student.

No Estudo 1, comparámos a potência alcançada pelos testes de aleatorização no caso da distribuição normal com a alcançada no caso de distribuições não-normais.

Os resultados deste primeiro estudo, sugeriram a realização do Estudo 2, em que avaliamos a influência da assimetria e/ou curtose sobre a potência dos testes de aleatorização.

Finalmente, no Estudo 3, estudámos, em distribuições normais, a influência da heterocedasticidade e do desequilíbrio na dimensão das amostras sobre a potência dos testes de aleatorização.

2. Questões metodológicas comuns aos três estudos

Em todos os estudos, por simulação, os dados foram gerados com programas escritos em R (R Development Core Team, 2008), recorrendo-se, quando necessário e como se indica em cada caso, a pacotes específicos do sistema R.

A potência estatística, que constitui a variável dependente, foi avaliada para testes bicaudais, unicaudais à esquerda e unicaudais à direita, para o nível de significância de 0.05.

As variáveis independentes que manipulámos foram o tipo de distribuição, o tamanho do efeito, a dimensão das amostras e, no caso do Estudo 3, também o equilíbrio das amostras e a heterocedasticidade.

Importa referir que, em todos os estudos, a estatística de teste que usámos, no caso dos testes de aleatorização, foi a indicada no capítulo 1, em 1.12:

$$\sum_{i=1}^{n_A} A_i,$$

que representa a soma das observações do primeiro grupo, e é uma estatística equivalente à do teste t de Student.

Justificamos aqui, porque são comuns aos três estudos, os valores escolhidos relativamente à dimensão dos grupos e ao tamanho do efeito, bem como as opções que fizemos no que respeita ao número de simulações (Nsim) e número de reamostragens (Nream).

Dimensão dos grupos

Usámos grupos com 8, 16 e 32 elementos por estes valores corresponderem, respectivamente, a amostras consideradas pequenas, médias e grandes em estudos experimentais nas Ciências Sociais e Humanas, nomeadamente em Psicologia.

Tamanho do efeito (TE)

Utilizámos 81 valores para o tamanho do efeito: de -4 a 4, em intervalos de 0.1. Estes valores foram escolhidos de modo a, por um lado, não ter muitos valores de TE a resultar em potência com valor de 1 (ou aproximado) e, por outro lado, a permitir haver suficiente diferenciação entre eles.

Entre os valores de TE, chamamos a atenção para os valores de 0.2, 0.5 e 0.8 que correspondem, segundo Cohen (1962, 1988) a efeitos “pequenos”, “médios” e “grandes”, na investigação em ciências sociais, nomeadamente em Psicologia. Estes valores convencionais receberam suporte empírico com um estudo de Lipsey (1990) englobando os resultados de 102 meta-análises, que sumariam os resultados de 6 700 estudos individuais na área das ciências do comportamento.

Número de simulações (Nsim) e número de reamostragens (Nream)

Para simular dados das várias distribuições, escrevemos programas em R. Para cada distribuição simulámos 20 000 amostras e, para cada uma destas amostras, gerámos 999 reamostragens para estimar a significância dos testes de aleatorização.

Os valores referentes ao número de amostras e ao número de reamostragens foram escolhidos tendo em atenção as recomendações de Oden (1991), Westfall & Young (1993) e Boos & Zhang (2000).

Westfall & Young (1993) afirmam que, para analisar um conjunto de dados reais, o número de reamostragens (Nream) deve ser tão grande quanto possível (10 000 ou mais). Assim, sendo os dados analisados por diferentes investigadores haverá pouca variação nos resultados. Contudo, no caso de estudos por simulação, como os que apresentamos nesta dissertação, estes autores, seguindo Hope (1968), afirmam ser muito mais importante que o número de simulações (Nsim) seja elevado.

Tendo em conta o tempo de computação, estabelecemos um valor de 20 000 000 para $N_{sim} \times N_{ream}$ e escolhemos uma combinação de valores para N_{sim} e N_{ream} de modo a minimizar o erro de estimativa da potência do teste de aleatorização.

A estratégia seguida por Oden (1991) foi a de minimizar o erro médio quadrático máximo (cfr. Oden, 1991, eq. 3) para um leque de valores de α (0.0, 1.0) e obteve $N_{ream} = 2\sqrt{N_{sim}}$ sob a hipótese nula, sugerindo a seguinte regra: para um valor fixo de $N_{sim} \times N_{ream}$, deve ser escolhido um valor de N_{ream} no intervalo $N_{ream} = 2\sqrt{N_{sim}}$ a $N_{ream} = 4\sqrt{N_{sim}}$.

Contudo, Boos & Zhang (2000), seguindo a mesma estratégia, consideraram ter mais interesse o leque de valores para α (0.00, 0.10) e poder não ser relevante minimizar o erro quadrático médio sob a hipótese nula para estimar a potência de hipótese alternativas. Os resultados das simulações que efectuaram levaram-nos a concluir que, para $\alpha = 0.05$ (e $\alpha = 0.10$), o valor que minimiza o erro quadrático médio é de cerca de 7 a 8 para a razão $N_{ream}/\sqrt{N_{sim}}$.

Assim, seguindo as indicações destes autores, para o valor fixo de 20 000 000 para $N_{sim} \times N_{ream}$, escolhemos os valores de 20 000 para N_{sim} e de 999 para N_{ream} (as 999 combinações reamostradas, mais a observada, constituem uma amostra de 1 000 combinações), com uma razão $N_{ream}/\sqrt{N_{sim}}$ de 7.1.

3. O sistema R

R é um ‘ambiente de programação para análise de dados e gráficos’. A linguagem R foi escrita originalmente por Ross Ihaka e Robert Gentleman, na Universidade de Auckland, Nova Zelândia (Ihaka & Gentleman, 1996), tendo sido desenvolvida, posteriormente por uma equipa de cerca de uma dúzia de pessoas, estatísticos e informáticos (‘R Development Core Team’).

Na perspectiva de um utilizador, a linguagem R é muito semelhante à linguagem S desenvolvida por John Chambers e outros nos Laboratórios Bell da AT&T (Becker, Chambers, & Wilks, 1998) e contém alguma da linguagem para modelar funções descrita em Chambers e Hastie (1992). Mas a crença que a linguagem S poderia ser desenvolvida de uma maneira diferente e mais vantajosa deu origem à linguagem R, que se baseou em ideias da linguagem Scheme.

R é um ‘software’ de fonte aberta, distribuído gratuitamente, que está continuamente a ser ampliado com a incorporação de novos métodos e que conta com o apoio de uma forte comunidade de utilizadores. Pode ser obtido do R-project em <http://www.r-project.org>. Como o R é também uma linguagem de programação, o utilizador não está limitado aos procedimentos pré-programados de pacotes estatísticos como o SAS ou o SPSS, sendo relativamente fácil programar novos métodos. É cada vez mais usado pelos investigadores em Estatística, como se constata se se compulsarem os jornais da especialidade.

Nesta dissertação, todas as análises estatísticas e gráficos foram realizados com o sistema R. Foram escritos alguns programas, que incorporaram, sempre que possível, as funções da linguagem ou rotinas disponíveis.

Em particular, na simulação de dados, usou-se o gerador de números pseudo-aleatórios ‘Mersenne Twister’ (Matsumoto & Nishimura, 1998), que tem um período extremamente longo e passou a maioria dos testes estatísticos para medir a qualidade de um conjunto de números aleatórios, nomeadamente a bateria de testes Diehard (Marsaglia, 1996). No caso da distribuição normal, foi usado o método ‘Inversion’, baseado no algoritmo AS 241 de Wichura (1988), que fornece resultados precisos até cerca de 16 dígitos.

ESTUDO 1

1. Introdução

O objectivo deste estudo é avaliar a potência dos testes de aleatorização, na comparação de dois grupos independentes equilibrados, no caso de distribuições não-normais. Avaliamos também a potência do teste t de Student, que comparamos com a alcançada pelo teste de aleatorização.

Quando, em investigações experimentais nas áreas das ciências humanas e sociais se comparam dois grupos, é frequente usar-se o teste t de Student, pressupondo-se a normalidade das distribuições.

A crença na tendência das populações para terem uma distribuição normal tem-se desvanecido ao longo do tempo, como relata Stigler (1973).

A preocupação com a violação do pressuposto de normalidade não é recente. Já em 1930 Egon Pearson alertava

“[...] não é suficiente saber que a amostra *pode* provir de uma população normal; deve ser igualmente claro, que é improvável que ela possa provir de uma população que se diferencie tanto da normal ao ponto de invalidar o uso de testes baseados na ‘teoria normal’ no tratamento subsequente do material.” (Pearson, 1930, p. 239)

A mesma preocupação leva Pitman (1937a, 1937b, 1938) a propor os testes de aleatorização que, segundo o título dos seus artigos, são “testes de significância que podem ser aplicados a amostras de qualquer população”.

Estudos realizados por Micceri (1989) e Stigler (1977) mostram que muitos dos conjuntos de dados recolhidos em investigações empíricas apresentam, usualmente, não-normalidade. No mesmo sentido, concorrem muitos outros autores, dos quais podemos citar Mosteller e Tukey (1968), Bradley (1977) e Wilcox (1995a, 1995b).

Mas os dados analisados por Micceri e Stigler sugerem diferentes tipos de não-normalidade: enquanto os dados de Stigler são aproximadamente normais, com caudas pesadas e alguns valores extremos ('outliers'), os dados descritos por Micceri apresentam uma não-normalidade mais extrema, com distribuições não só assimétricas mas multimodais.

Dada esta situação, que deve fazer um investigador quando deseja analisar um conjunto de dados? Habitualmente, é aconselhado a realizar primeiro um teste de ajustamento à distribuição normal e decidir, em função do resultado, que modelo estatístico usar.

Thode (2002) apresenta cerca de 40 procedimentos propostos para testar especificamente a normalidade, para além de métodos gráficos e outros testes de ajustamento de uso geral que podem também ser usados para testar a normalidade de um conjunto de dados.

Contudo, com conjuntos de dados de pequena dimensão, como tantas vezes se encontram em estudos experimentais, os resultados dos testes prévios à normalidade das distribuições não são uma resposta ao problema.

A título de ilustração, simulámos amostras de 8, 16 e 32 elementos, com distribuição normal, e amostras, com os mesmos tamanhos, com distribuição uniforme (o número de amostras simuladas foi, em cada caso de 100 000). Usámos três testes: (1) o teste de Lillifors, uma adaptação do teste de Kolmogorov-Smirnov, o teste mais conhecido baseado na

função de distribuição empírica; (2) o teste de Anderson-Darling, também baseado na função de distribuição empírica e considerado superior em termos de potência (Stephens,1986) e o teste de Shapiro-Wilks, baseado em estatísticas de ordem. Estabelecemos uma probabilidade de cometer um erro de Tipo I de 0.05.

Tabela E1.1 Percentagem de amostras consideradas com distribuição normal pelos testes de Lilliefors, Anderson-Darling e Shapiro-Wilk para dados simuladas de uma distribuição normal e de uma distribuição uniforme (para $\alpha = 0.05$).

TESTE Distribuição simulada	LILLIEFORS		ANDERSON- DARLING		SHAPIRO-WILK	
	Normal	Uniforme	Normal	Uniforme	Normal	Uniforme
n = 8	95.1	94.1	95.0	93.1	94.9	92.6
n = 16	95.1	91.9	95.0	87.3	95.0	86.0
n = 32	95.2	84.8	95.1	67.5	95.0	57.7

Como se pode ver, os três testes têm resultados semelhantes no que se refere aos dados simulados com distribuição normal, rejeitando apenas erradamente 5% dos casos, como seria de esperar. Mas todos erram, no que respeita aos dados simulados com distribuição uniforme: o erro vai diminuindo com o aumento do tamanho das amostras, mas mesmo no caso de amostras com 32 elementos, o melhor teste erra em 57.7 % dos casos.

Não se pode, evidentemente, generalizar os resultados obtidos com a distribuição uniforme para outras distribuições e é de supor que a eficiência destes testes melhorará com o aumento do tamanho das amostras. Mas também se sabe que, com grandes amostras, pequenos desvios à normalidade levam à rejeição da hipótese testada.

Os dados recolhidos em investigações empíricas podem provir das mais variadas distribuições e não é possível cobrir todas as possibilidades. Neste estudo usamos o conjunto de 15 distribuições usadas por Marron e Wand (1992) no seu estudo de simulação e que foram usadas

posteriormente em muitos outros, e que são misturas de normais. Essas 15 distribuições incluem algumas distribuições do tipo descrito por Stigler, mas na maioria apresentam não-normalidade extrema, como as analisadas por Micceri.

As 15 distribuições Marron-Wand podem ser escritas como misturas de normais, como se pode ver na Tabela E1.1, na qual se indicam, para cada distribuição, os parâmetros das normais que a constituem, juntamente com a respectiva ponderação.

Tabela E1.2 As 15 distribuições Marron-Wand

Distribuição	Média	Variância	Ponderação
1. Gaussian	0	1	1
2. Skewed	-0.3	1.4400000	0.2
	0.3	0.6400000	0.2
	1.0	0.4444444	0.6
3. Strong Skewed	0.000000	1.00000000	0.125
	-1.000000	0.44444444	0.125
	-1.666667	0.19753086	0.125
	-2.111111	0.08779149	0.125
	-2.407407	0.03901844	0.125
	-2.604938	0.01734153	0.125
	-2.736626	0.00770734	0.125
-2.824417	0.00342548	0.125	
4. Kurtotic	0	1.00	0.6666667
	0	0.01	0.3333333
5. Outlier	0	1.00	0.1
	0	0.01	0.9
6. Bimodal	-1	0.4444444	0.5
	1	0.4444444	0.5
7. Separated Bimodal	-1.5	0.25	0.5
	1.5	0.25	0.5
8. Asymmetric Bimodal	0.0	1.0000000	0.75
	1.5	0.1111111	0.25
9. Trimodal	-1.2	0.3600	0.45
	1.2	0.3600	0.45
	0.0	0.0625	0.10
10. Claw	0.0	1.00	0.5
	-1.0	0.01	0.1
	-0.5	0.01	0.1
	0.0	0.01	0.1
	0.5	0.01	0.1
	1.0	0.01	0.1

[Continua]

Tabela E1.2 [Continuação] As 15 distribuições Marron-Wand

11. Double Claw	-1.0	0.4444444	0.490000000
	1.0	0.4444444	0.490000000
	-1.5	0.0001000	0.002857143
	-1.0	0.0001000	0.002857143
	-0.5	0.0001000	0.002857143
	0.0	0.0001000	0.002857143
	0.5	0.0001000	0.002857143
	1.0	0.0001000	0.002857143
	1.5	0.0001000	0.002857143
12. Asymmetric Claw	0.0	1.000000	0.50000000
	-1.5	0.160000	0.25806452
	-0.5	0.040000	0.12903226
	0.5	0.010000	0.06451613
	1.5	0.002500	0.03225806
2.5	0.000625	0.01612903	
13. Asymmetric Double Claw	-1.0	0.4444444	0.460000000
	1.0	0.4444444	0.460000000
	-1.5	0.0001000	0.003333333
	-1.0	0.0001000	0.003333333
	-0.5	0.0001000	0.003333333
	0.5	0.0049000	0.023333333
	1.0	0.0049000	0.023333333
	1.5	0.0049000	0.023333333
14. Smooth Comb	-1.4761905	0.2579994961	0.50793651
	0.8095238	0.0644998740	0.25396825
	1.9523810	0.0161249685	0.12698413
	2.5238095	0.0040312421	0.06349206
	2.8095238	0.0010078105	0.03174603
	2.9523810	0.0002519526	0.01587302
15. Discrete Comb	-2.1428571	0.081632653	0.28571429
	-0.4285714	0.081632653	0.28571429
	1.2857143	0.081632653	0.28571429
	2.2857143	0.002267574	0.04761905
	2.5714286	0.002267574	0.04761905
	2.8571429	0.002267574	0.04761905

Para se poder visualizar a forma destas distribuições, apresentamos na Figura E1.1 gráficos das suas densidades.

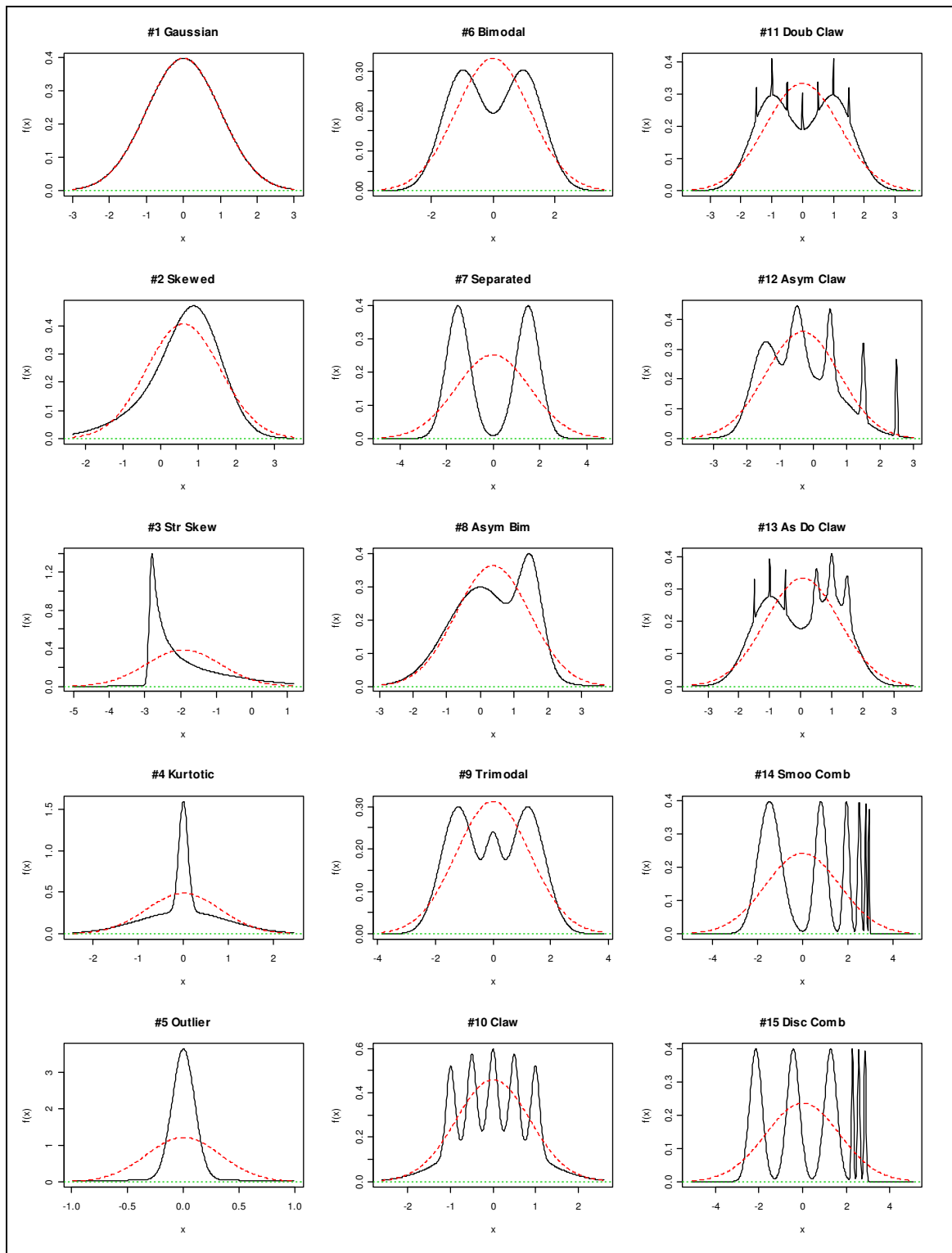


Figura E1.1 Densidades das 15 distribuições Marron-Wand

2. Método

Avaliámos a potência dos testes de aleatorização e a dos testes t de Student, na comparação de dois grupos independentes equilibrados, manipulando as seguintes três variáveis:

Tamanho das amostras, com três níveis: $n_1 = n_2 = 8$, $n_1 = n_2 = 16$ e $n_1 = n_2 = 32$;

Tamanho do efeito, com 81 níveis: de -4 a 4, em intervalos de 0.1; e

Distribuição populacional, com 15 níveis: as distribuições Marron-Wand, misturas de normais, descritas na Tabela E1.1.

A potência foi avaliada para testes bicaudais, unicaudais à esquerda e unicaudais à direita, para $\alpha = 0.05$.

Para simular dados destas distribuições, escrevemos programas em R (R Development Core Team, 2008) (Anexo IV). Para gerar dados das distribuições Marron-Wand, usámos o package 'norlmix' (Mächler, 2007). Para cada distribuição simulámos 20 000 amostras e, para cada uma destas amostras, gerámos 999 combinações aleatórias para estimar a significância dos testes de aleatorização.

Calculámos a potência dos testes como a proporção de amostras em que o valor de probabilidade associado à estatística do teste era igual ou inferior a 0.05.

No referente aos testes de aleatorização, com amostras de $n_1 = n_2 = 8$, o número total de combinações é de 12 870; este número sobe para 601 080 390, no caso de amostras de $n_1 = n_2 = 16$, e para 1.83262E+18, no caso de amostras de $n_1 = n_2 = 32$. Como usámos, para estimar a significância do teste de aleatorização, um conjunto de referência de 1 000 combinações (1 observada, mais 999 geradas aleatoriamente),

construímos intervalos de confiança a 99% para cada valor de potência. Quando compararmos a potência dos testes de aleatorização com a potência do teste t de Student, usaremos a informação fornecida por estes intervalos de confiança.

3. Apresentação e discussão dos resultados

As matrizes com os resultados das simulações efectuadas podem ser consultadas no CD anexo, ficheiro ‘MatrizesDados_Estudo1’.

Começamos por apresentar a potência dos testes de aleatorização no caso da distribuição normal, seguidamente comparamos a potência dos testes de aleatorização com a potência dos testes t de Student e, finalmente, para os testes de aleatorização, comparamos a sua potência no caso das distribuições Marron-Wand 02 a 14 relativamente à sua potência no caso da distribuição normal.

3.1. Potência dos testes de aleatorização (distribuição normal)

A potência dos testes de aleatorização, no caso da distribuição normal, apresenta-se graficamente nas Figura E1.2 (testes bicaudais) e E2.3 (testes unicaudais). Como seria de esperar, a potência dos testes aumenta em função do tamanho das amostras.

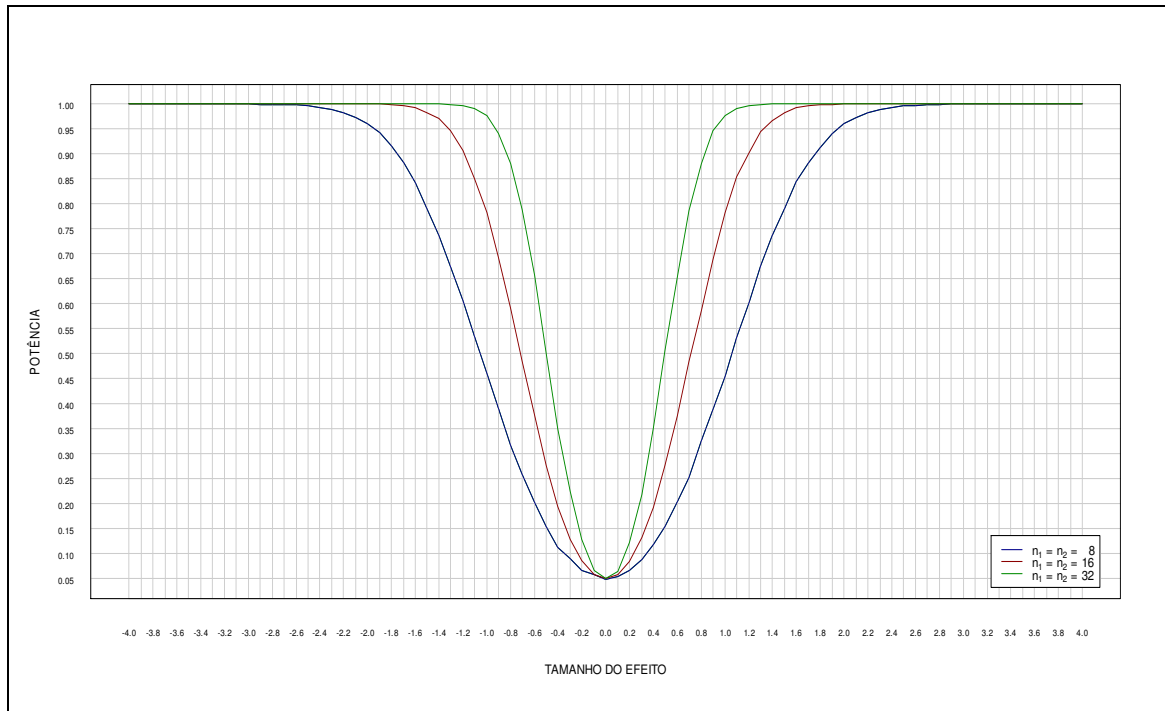


Figura E1.2 Potência dos testes de aleatorização bicaudais, para a distribuição normal, com $\alpha = 0.05$, em função do tamanho do efeito e do tamanho das amostras

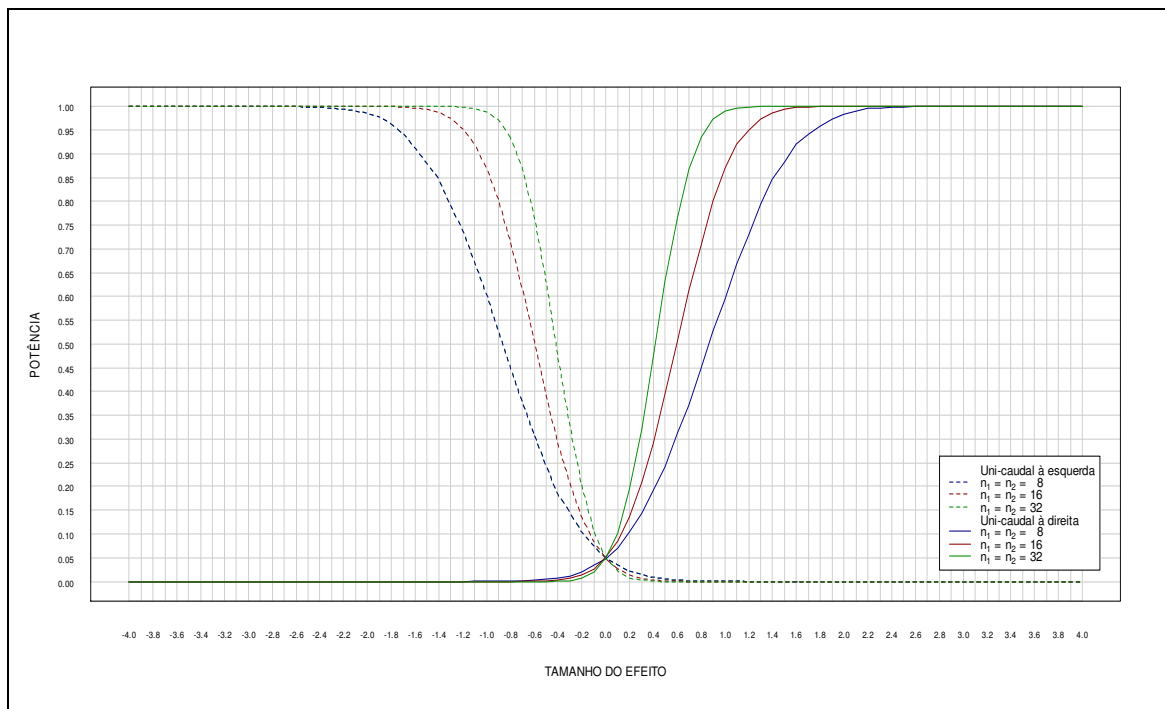


Figura E1.3 Potência dos testes de aleatorização unicaudais, para a distribuição normal, com $\alpha = 0.05$, em função do tamanho do efeito e do tamanho das amostras

No caso dos testes bicaudais, para atingir uma potência de 0.80, valor considerado convencionalmente como mínimo aceitável, é necessário que os valores do tamanho do efeito, para amostras equilibradas com 8, 16 e 32 sujeitos por grupo, sejam, respectivamente, de aproximadamente 1.5, 1.0 e 0.7. No caso dos testes unicaudais, para alcançar a mesma potência, os valores do tamanho do efeito passam para 1.3, 0.9 e 0.6.

Na tabela E1.3, apresenta-se a potência dos testes de aleatorização para os valores de TE de 0.2, 0.5 e 0.8 que correspondem, segundo as indicações de Cohen, a efeitos “pequenos”, “médios” e “grandes”.

Tabela E1.3 Potência dos testes de aleatorização, bi e unicaudais, para valores de tamanho do efeito (TE) de 0.2, 0.5 e 0.8, em função do tamanho das amostras

TE	$n_1 = n_2 = 8$		$n_1 = n_2 = 16$		$n_1 = n_2 = 32$	
	Unicaudal	Bicaudal	Unicaudal	Bicaudal	Unicaudal	Bicaudal
0.2	0.10	0.07	0.14	0.08	0.20	0.12
0.5	0.24	0.15	0.39	0.28	0.63	0.50
0.8	0.45	0.32	0.71	0.59	0.94	0.88

Como se pode verificar, apenas para efeitos considerados “grandes” e para amostras com 32 sujeitos por grupo a potência alcançada supera o valor de 0.80.

3.2. Comparação da potência dos testes de aleatorização e t de Student

No Anexo V apresentam-se estatísticas descritivas da diferença de potência entre os testes de aleatorização e t de Student. A observação dessas estatísticas permite constatar que as diferenças são, de um modo geral, muito pequenas e tendem a reduzir-se com o aumento do tamanho

das amostras. Se não considerarmos a distribuição M-W 05 ('Outlier'), as diferenças de potência variam entre -0,013 e 0.015.

Relativamente à distribuição M-W 05 ('Outlier'), e para testes bicaudais, verificam-se diferenças de potência a favor dos testes de aleatorização, como se pode observar na Figura E1.4.

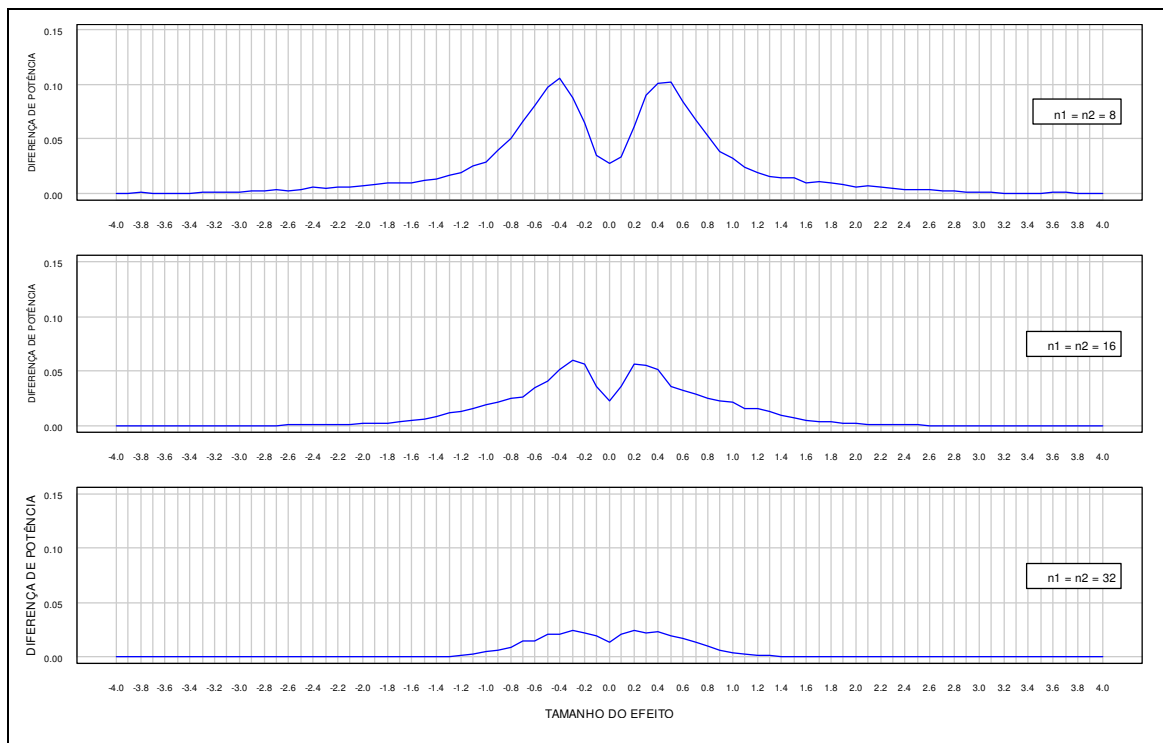


Figura E1.4 Diferença de potência entre os testes de aleatorização e o teste t de Student (bicaudais) para a distribuição M-W 05 ('Outlier'), em função do tamanho do efeito e do tamanho das amostras

Contudo, importa dizer que só se observam diferenças significativas, para testes bicaudais, no caso de amostras com 8 e 16 elementos por grupo, e para testes unicaudais apenas para amostras com 8 elementos por grupo, como se mostra na Tabela E1.4. Com o aumento do tamanho das amostras não só o tamanho das diferenças diminui, como o leque de valores de TE em que se observam diferenças diminui.

Tabela E1.4 Diferença entre a potência dos testes de aleatorização e t de Student para a distribuição Marron-Wand 05 ('Outlier').

TE	$n_1 = n_2 = 8$			$n_1 = n_2 = 16$			$n_1 = n_2 = 32$		
	Esq	Dir	Bi	Esq	Dir	Bi	Esq	Dir	Bi
-1.0	0.008	0.000	0.029	0.006	0.000	0.019	0.001	0.000	0.005
-0.9	0.010	0.000	0.040	0.005	0.000	0.022	0.000	0.000	0.006
-0.8	0.017	0.000	0.050	0.005	0.000	0.025	0.001	0.000	0.009
-0.7	0.023	0.000	0.066	0.005	0.000	0.027	0.000	0.000	0.015
-0.6	0.032	0.000	0.081	0.006	0.000	0.034	0.001	0.000	0.015
-0.5	0.046	0.000	0.098	0.009	0.000	0.040	0.002	0.000	0.021
-0.4	0.061	0.000	0.106	0.014	0.000	0.051	0.001	0.000	0.021
-0.3	0.066	0.000	0.089	0.022	0.000	0.060	0.002	0.000	0.024
-0.2	0.057	0.001	0.065	0.028	0.000	0.056	0.004	0.000	0.022
-0.1	0.036	0.005	0.035	0.021	0.002	0.036	0.005	0.001	0.019
0.0	0.018	0.018	0.028	0.009	0.010	0.023	0.003	0.002	0.013
0.1	0.006	0.035	0.034	0.002	0.021	0.035	0.000	0.006	0.021
0.2	0.001	0.055	0.061	0.000	0.027	0.057	0.000	0.005	0.024
0.3	0.000	0.066	0.091	0.000	0.021	0.055	0.000	0.001	0.022
0.4	0.000	0.061	0.101	0.000	0.016	0.051	0.000	0.001	0.023
0.5	0.000	0.049	0.102	0.000	0.012	0.036	0.000	0.002	0.020
0.6	0.000	0.034	0.085	0.000	0.008	0.032	0.000	0.002	0.017
0.7	0.000	0.024	0.068	0.000	0.007	0.028	0.000	0.001	0.014
0.8	0.000	0.016	0.053	0.000	0.005	0.025	0.000	0.001	0.010
0.9	0.000	0.012	0.038	0.000	0.006	0.022	0.000	-0.001	0.007
1.0	0.000	0.008	0.032	0.000	0.005	0.021	0.000	0.000	0.004

Nota: Esc: testes unicaudal à esquerda; Dir: teste unicaudal à direita; Bi: teste bicaudal. A negrito indicam-se as diferenças significativas.

Para esta distribuição, Marron-Wand 05 ('Outlier'), o teste t de Student é demasiado conservador, pois para $TE = 0.0$ a potência alcançada, respectivamente para os testes unicaudal à esquerda, unicaudal à direita e bicaudal, é de 0.033, 0.032 e 0.021, enquanto que os valores correspondentes para o teste de aleatorização são de 0.050, 0.050 e 0.049.

3.3. Influência da não-normalidade na potência dos testes de aleatorização

No Anexo VI, apresentam-se, graficamente, os resultados relativos à potência alcançada pelos testes de aleatorização no caso das

distribuições Marron-Wand 02 a 15 em comparação com a sua potência no caso da distribuição Marron-Wand 01 (Normal).

Descrevemos, primeiro, a influência da não-normalidade para o caso de testes bicaudais, seguidamente para o caso de testes unicaudais. Na apresentação dos resultados, distinguiremos dois grupos de distribuições Marron-Wand: o primeiro grupo inclui as distribuições 02 a 05, semelhantes às analisadas por Stigler, o segundo grupo compreende as distribuições 06 a 15, do tipo das descritas por Micceri.

3.3.1. Testes bicaudais

Começamos por apresentar, na Figura E1.5, para amostras de 8 sujeitos por grupo, a diferença entre a potência dos testes bicaudais para as distribuições Marron-Wand 02 a 05 e a sua potência para a distribuição normal. Estas quatro distribuições são todas unimodais, ao contrário das restantes distribuições que apresentam duas ou mais modas.

As diferenças de potência não são uniformes para o conjunto dos 81 tamanhos de efeito. Pode-se observar que essas diferenças apresentam aproximadamente um mesmo padrão, simétrico: a partir do valor de TE de 0.0, começam a verificar-se ganhos crescentes de potência até aproximadamente aos valores de TE de |0.8|, a partir destes valores os ganhos começam a diminuir até cerca dos valores de TE de |1.6|. Então, inverte-se a tendência, observando-se perdas crescentes até aproximadamente aos valores de TE de |2.0|, valores a partir dos quais as perdas vão diminuindo.

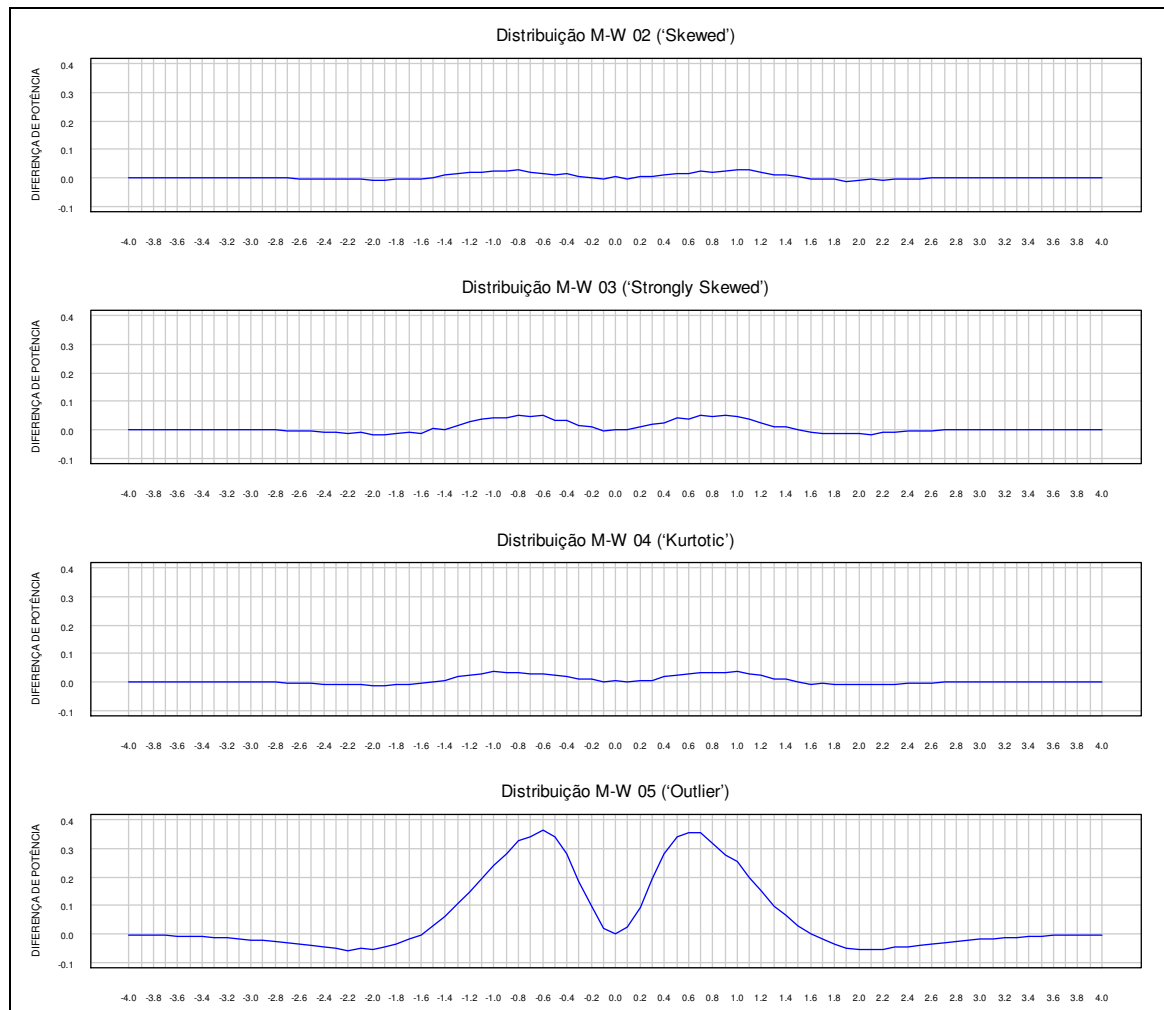


Figura E1.5 Diferença entre a potência dos testes de aleatorização bicaudais, com $n_1 = n_2 = 8$, para as distribuições Marron-Wand 02 a 05 e a sua potência para a distribuição normal, em função do tamanho do efeito

Se exceptuarmos a distribuição Marron-Wand 05 ('Outlier'), as diferenças de potência são relativamente pequenas: as perdas de potência são no máximo de 0.019 e os ganhos máximos de 0.052 (distribuição Marron-Wand 03, 'Strongly Skewed'). No que respeita à distribuição Marron-Wand 05, a perda máxima é de 0.058, mas o ganho máximo apresenta um valor elevado: 0.362.

A diferença entre a potência dos testes bicaudais, com 8 sujeitos por grupo, para as distribuições Marron-Wand 06 a 15 e a sua potência para a

distribuição normal, é apresentada na Figura E1.6. Para estas distribuições o padrão das diferenças de potência é inverso ao acima descrito, observando-se, a partir do valor de TE de 0.0, primeiro perdas e depois ganhos de potência. Mas essas diferenças são relativamente pequenas: as perdas são no máximo de 0.046 e os ganhos de 0.017 (distribuição Marron-Wand 07, ‘Separated Bimodal’).

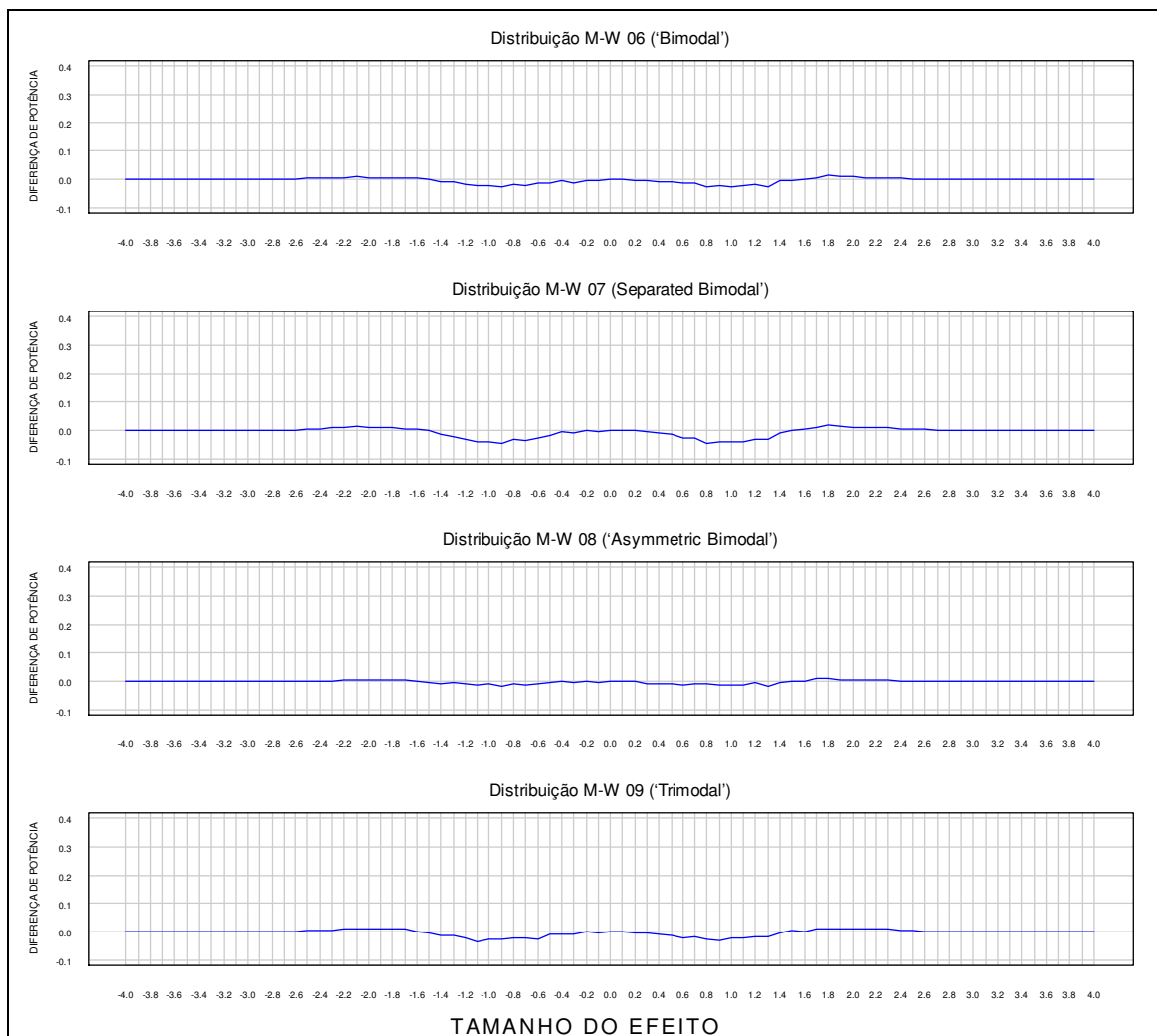


Figura E1.6 Diferença entre a potência dos testes de aleatorização bicaudais, com $n_1 = n_2 = 8$, para as distribuições Marron-Wand 06 a 15 e a sua potência para a distribuição normal, em função do tamanho do efeito [Continua]

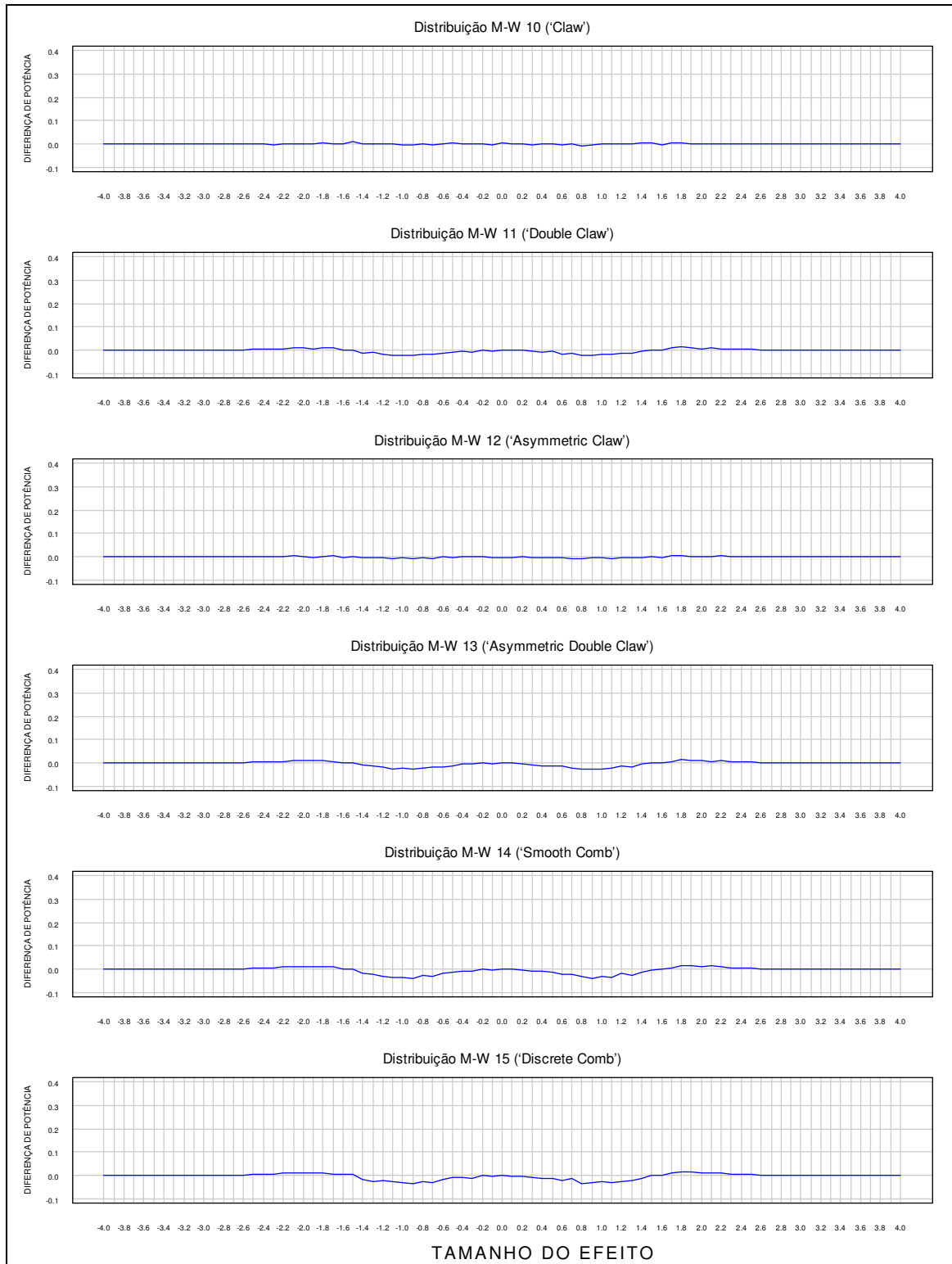


Figura E1.6 [Continuação] Diferença entre a potência dos testes de aleatorização bicaudais, com $n_1 = n_2 = 8$, para as distribuições Marron-Wand 06 a 15 e a sua potência para a distribuição normal, em função do tamanho do efeito

Os efeitos que descrevemos dizem respeito a amostras com $n_1 = n_2 = 8$. Verifica-se uma diminuição destes efeitos no caso das amostras maiores: embora com padrão semelhante, não só o tamanho das diferenças diminui, como o leque de valores de TE em que se verificam diferenças de potência é mais pequeno.

Assim, para a distribuição Marron-Wand 05 ('Outlier'), no caso de $n_1 = n_2 = 16$, verificam-se ganhos de potência a partir do valor de TE de 0 até 11.01, com um máximo de 0.219, para um valor de TE de cerca de 10.51. No caso de $n_1 = n_2 = 32$, os ganhos de potência observam-se a partir do valor de TE de 0 até 10.71, com um máximo de 0.115, para um valor de TE de cerca de 10.41. No que respeita às perdas de potência, também se observam diminuições: se a perda máxima era de 0.058, para amostras com $n_1 = n_2 = 8$, esse valor passa para 0.037, no caso de $n_1 = n_2 = 16$, e para 0.027 para $n_1 = n_2 = 32$.

Quanto à distribuição Marron-Wand 07 ('Separated Bimodal'), em que as perdas eram no máximo de 0.046 e os ganhos de 0.017 no caso de $n_1 = n_2 = 8$, verifica-se que as perdas máximas passam para cerca de 0.022 e de 0.011, respectivamente para amostras com $n_1 = n_2 = 16$ e com $n_1 = n_2 = 32$; quanto aos ganhos máximos, que já eram muito pequenos, também eles diminuem com o aumento do tamanho das amostras.

3.3.2. Testes unicaudais

Na Figura E1.7 apresentamos, para amostras de 8 sujeitos por grupo, a diferença entre a potência dos testes unicaudais para as distribuições Marron-Wand 2 a 05 (as distribuições unimodais) e a sua potência para a distribuição normal.

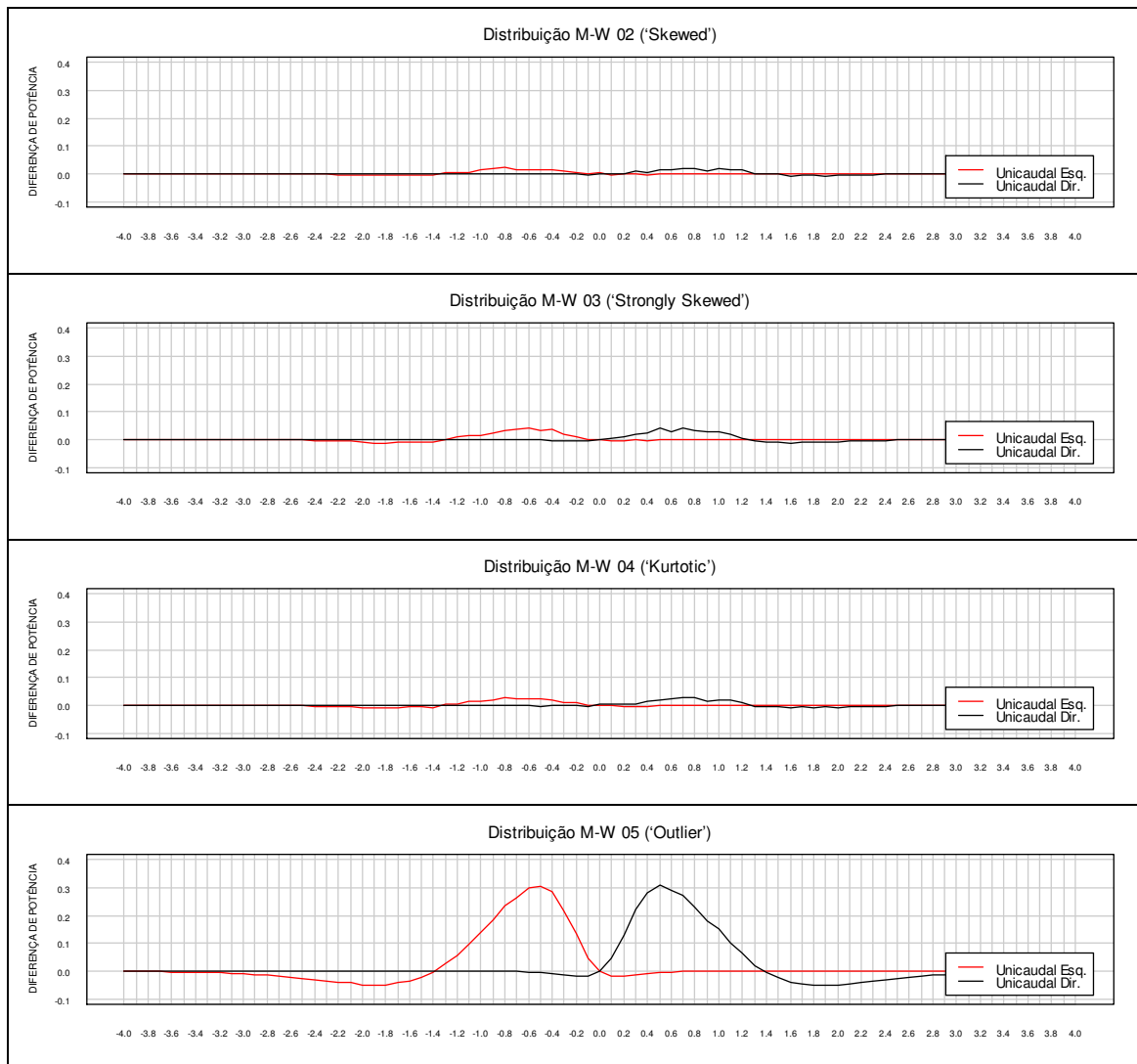


Figura E1. 7 Diferença entre a potência dos testes de aleatorização unicaudais, com $n_1 = n_2 = 8$, para as distribuições Marron-Wand 02 a 05 e a sua potência para a distribuição normal, em função do tamanho do efeito

As diferenças de potência para os testes unicaudais à esquerda e à direita são aproximadamente iguais, pelo que passamos a descrever as diferenças apenas para estes últimos testes.

O padrão das diferenças é semelhante ao anteriormente observado no caso dos testes bicaudais, mas observa-se uma diminuição dos seus tamanhos.

Não considerando a distribuição Marron-Wand 05, as perdas de potência são no máximo de 0.013 e os ganhos máximos de 0.041 (distribuição Marron-Wand 03, ‘Strongly Skewed’). No que respeita à distribuição Marron-Wand 05, ‘Outlier’, a perda máxima é de 0.051 e o ganho máximo é de 0.308.

Apresenta-se na Figura E1.8 a diferença entre a potência dos testes unicaudais, com 8 sujeitos por grupo, para as distribuições Marron-Wand 06 a 15 e a sua potência para a distribuição normal.

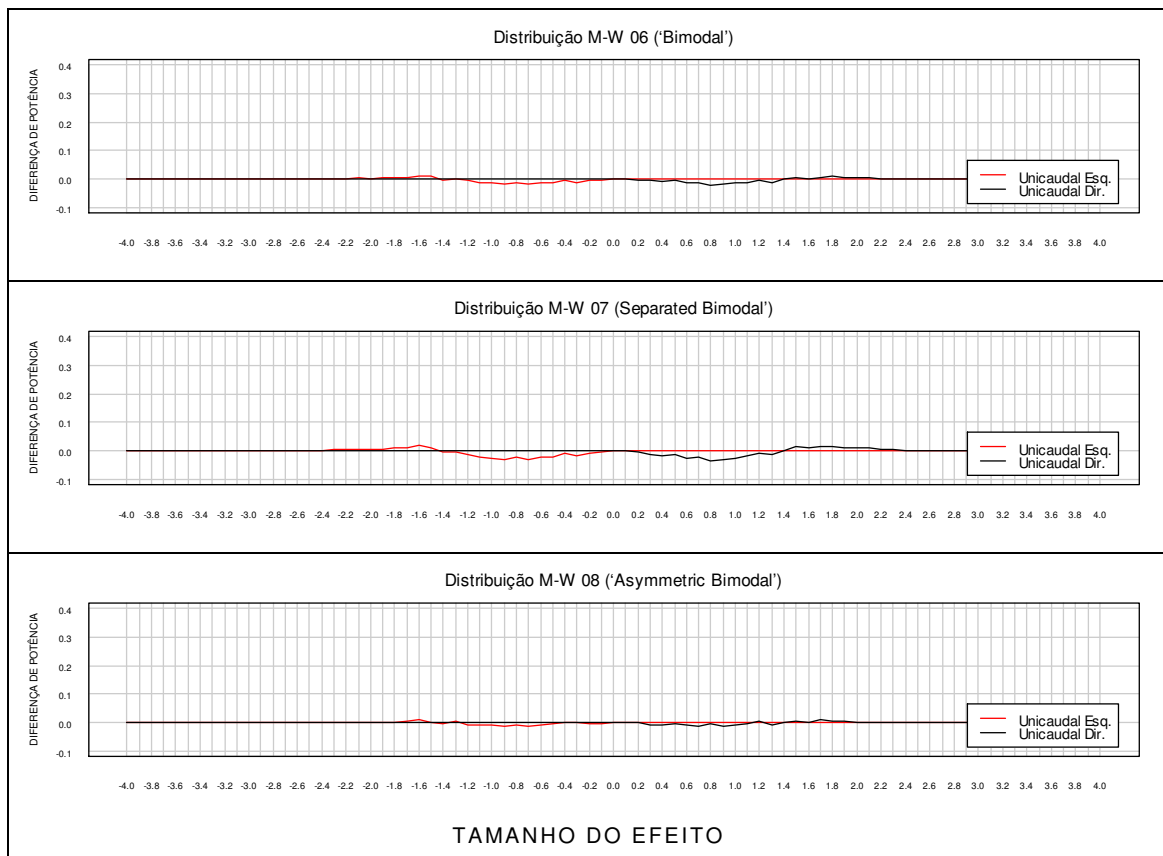


Figura E1.8 Diferença entre a potência dos testes de aleatorização unicaudais, com $n_1 = n_2 = 8$, para as distribuições Marron-Wand 06 a 15 e a sua potência para a distribuição normal, em função do tamanho do efeito [Continua]

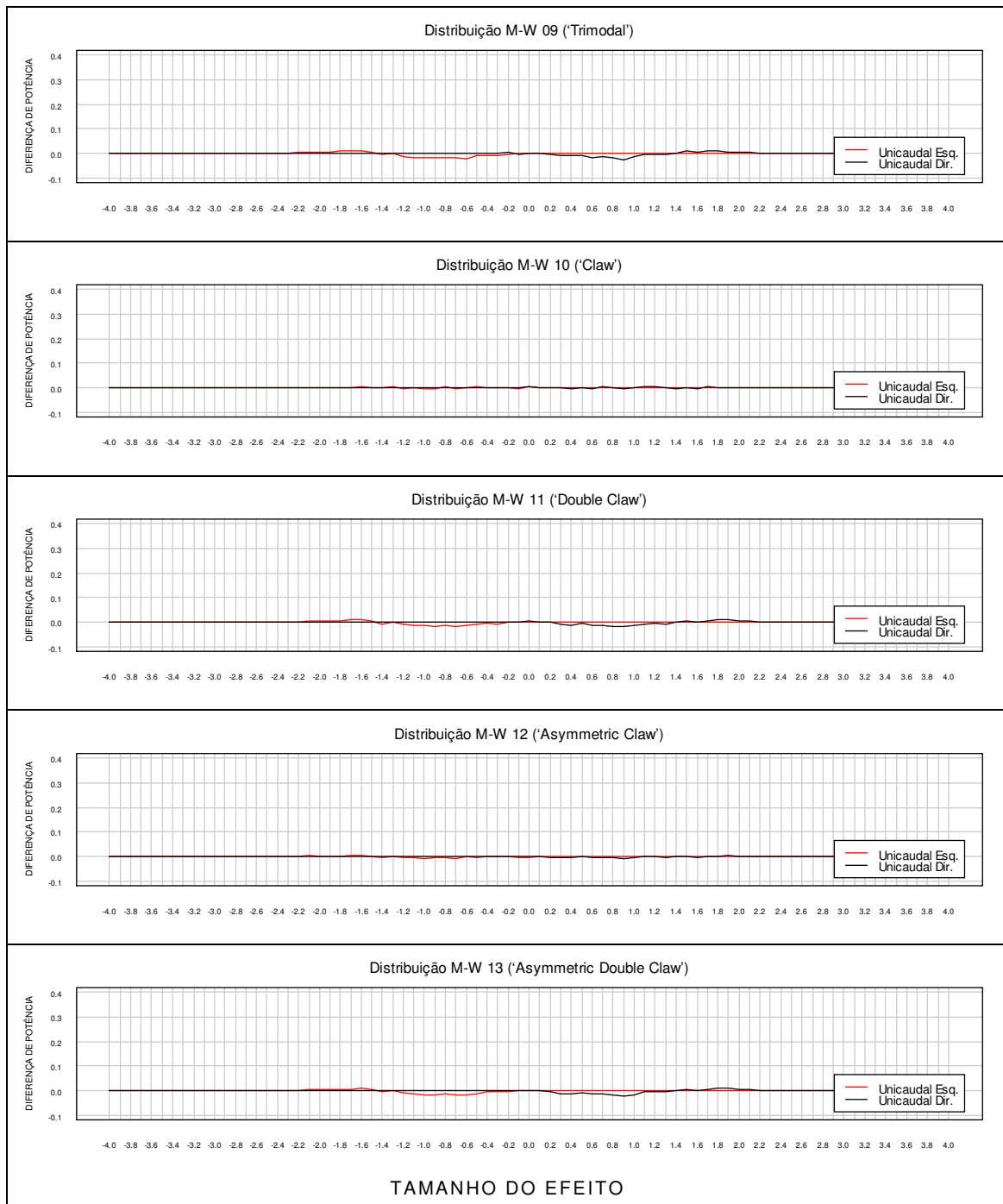


Figura E1.8 [Continuação] Diferença entre a potência dos testes de aleatorização uncaudais, com $n_1 = n_2 = 8$, para as distribuições Marron-Wand 06 a 15 e a sua potência para a distribuição normal, em função do tamanho do efeito [Continua]

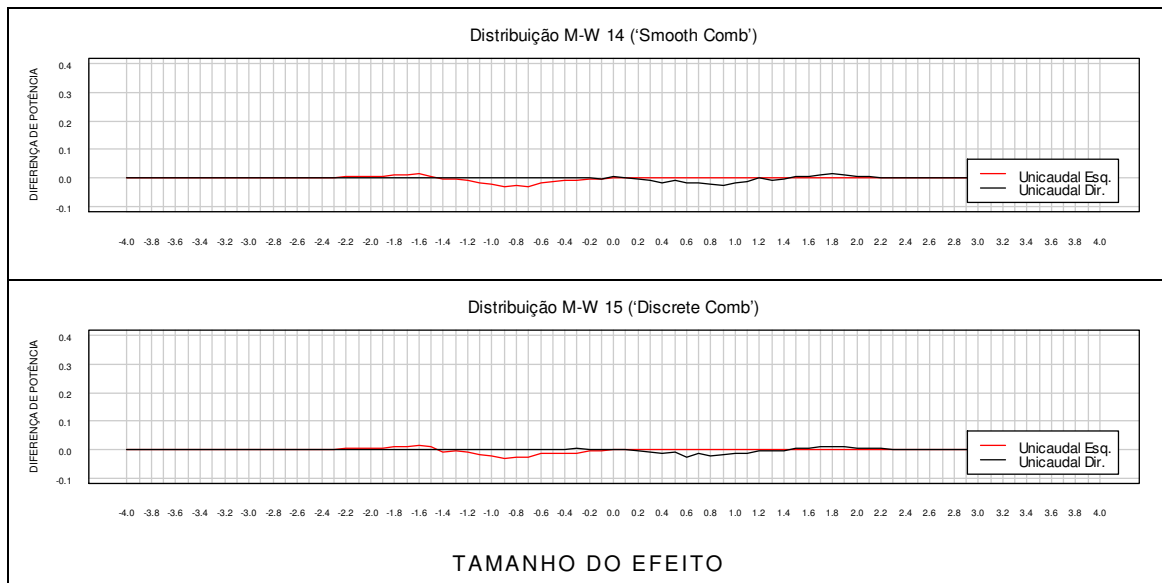


Figura E1.8 [Continuação] Diferença entre a potência dos testes de aleatorização unicaudais, com $n_1 = n_2 = 8$, para as distribuições Marron-Wand 06 a 15 e a sua potência para a distribuição normal, em função do tamanho do efeito

Também para estas distribuições o padrão das diferenças de potência é semelhante ao observado no caso dos testes bicaudais, verificando-se, a partir do valor de TE de 0.0, primeiro perdas e depois ganhos de potência. Mas as diferenças são menores: as perdas são no máximo de 0.036 e os ganhos de 0.014 (distribuição Marron-Wand 07, 'Separated Bimodal').

Os resultados apresentados respeitam a amostras com $n_1 = n_2 = 8$. Tal como para os testes bicaudais, verifica-se uma diminuição destes efeitos no caso das amostras maiores. Embora com padrão semelhante, o tamanho das diferenças diminui e é mais pequeno o leque de valores de TE em que se verificam diferenças de potência.

4. Conclusões

Neste estudo foi avaliada a potência, dos testes de aleatorização, na comparação de duas amostras independentes e equilibradas, para distribuições não-normais. O efeito na potência destes testes é diferente para os dois tipos de não-normalidade considerados.

Para as distribuições Marron-Wand 02 a 05, semelhantes às descritas por Stigler (normais com caudas pesadas e alguns valores extremos), todas unimodais, observa-se, para valores mais baixos de TE (em que se incluem os valores referidos por Cohen), um aumento de potência; para valores mais elevados de TE, constata-se perdas. Para o conjunto dos 81 tamanhos de efeito considerados, os ganhos e as perdas são aproximadamente iguais e de pequena dimensão, salvo para a última distribuição, 'Outlier', em que os ganhos são relevantes e superam as perdas.

No que respeita às distribuições Marron-Wand 06 a 15, semelhantes às analisadas por Micceri, com não-normalidade mais extrema, com múltiplas modas, observa-se um padrão inverso: para valores mais baixos de TE, verificam-se perdas, para valores mais elevados, ganhos de potência. Mas aqui as perdas superam os ganhos, embora tanto aquelas como estes sejam de pequena dimensão.

As diferenças de potência, relativamente à distribuição normal, verificam-se tanto no caso dos testes bicaudais como no caso dos unicaudais, sendo tanto maiores quanto menor a dimensão dos grupos.

No que concerne à comparação do teste de aleatorização com o teste t de Student, em termos de potência, verificou-se que, em geral, apresentam uma potência semelhante, com alguma vantagem para o primeiro. Essa vantagem só é significativa no caso de uma das distribuições simuladas,

a distribuição Marron-Wand 05, distribuição com forte curtose e valores extremos ('outliers'), e diminui com o aumento do tamanho dos grupos. Estes resultados sugerem que se um investigador, ao planejar uma experiência, escolher o tamanho dos seus grupos em função de um determinado tamanho de efeito populacional e de uma dada potência, assumindo uma distribuição normal, a potência do seu teste de aleatorização não será diminuída em mais de 0.05, se os seus dados provierem de distribuições não-normais.

No entanto, é importante referir que os resultados deste estudo foram obtidos com grupos equilibrados e provenientes da mesma distribuição. Não podem, por isso, ser generalizados para situações em os grupos não sejam equilibrados, possuam diferentes distribuições e sejam heterocedásticos.

ESTUDO 2

1. Introdução

O objectivo deste segundo estudo é avaliar o comportamento, em termos de potência, dos testes de aleatorização, na comparação de dois grupos independentes equilibrados, em função da forma das distribuições. A potência dos testes *t* de Student, para as mesmas distribuições, servirá de padrão de comparação.

A forma das distribuições será manipulada em termos de assimetria e curtose e, com essa finalidade, simularemos dados de distribuições *g-e-h*.

As distribuições *g-e-h* foram introduzidas por Tukey (1977), e o estudo das suas propriedades foi ampliado por Hoaglin (1983, 1985), Martinez e Iglewicz (1984), Badrinath e Chatterjee (1988, 1991), Mills (1995), e Dutta e Babbel (2002), Field e Genton (2006) e Headrick, Kowalchuk e Sheng (2008).

Tukey introduziu a família de distribuições *g-e-h* pela seguinte transformação de uma variável normal padronizada *Z*

$$Y_{g,h}(Z) = \left(\frac{e^{gZ} - 1}{g} \right) e^{hZ^2/2} \quad (\text{E2.1})$$

em que *g* e *h* são quaisquer números reais.

O parâmetro *g* é responsável pela assimetria ou enviesamento da distribuição *g-e-h*. Para *g* = 0, a distribuição é simétrica. À medida que o valor absoluto de *g* aumenta, a quantidade de assimetria aumenta. O sinal de *g* controla a direcção da assimetria (mas não a quantidade). Valores positivos de *g* enviesam a distribuição para a cauda direita, enquanto valores negativos de *g* enviesam a distribuição para a cauda esquerda. O valor de *h* controla o alongamento no centro ou o peso das caudas da distribuição. Para *h* = 0, o alongamento é equivalente ao de uma

distribuição normal. Para $h = 1$, o alongamento é aproximadamente equivalente ao de uma distribuição de Cauchy. Valores de h situam-se tipicamente no intervalo $(0,1)$.

É de notar que é possível introduzir mais dois parâmetros, de localização e de escala, nas distribuições g - e - h . Seja X uma variável aleatória criada por uma transformação linear de $Y_{g,h}(Z)$:

$$X = A + BY_{g,h}(Z) = A + B \left(\frac{e^{gZ} - 1}{g} \right) e^{hZ^2/2}, \quad (\text{E2.2})$$

em que A é o parâmetro de localização e B o de escala. Hoaglin (1985) indica que o parâmetro A é a mediana da variável aleatória X .

No presente estudo, tendo em conta os nossos objectivos, estes parâmetros têm pouca importância, pelo que restringimos a nossa atenção às distribuições g - e - h com $A = 0$ e $B = 1$, ou seja, usamos a equação E2.1.

É ainda de referir que, em certas situações, para aproximar uma distribuição, pode ser necessário trabalhar com os parâmetros g e h , não considerados como constantes, mas como polinómios de ordem mais elevada de Z (ver, por exemplo, Hoaglin (1985) e Dutta e Perry (2006)).

Quando $h = 0$, a distribuição g - e - h reduz-se a

$$Y_g(Z) = \frac{(e^{gZ} - 1)}{g}, \quad (\text{E2.3})$$

que também é conhecida como distribuição g .

Quando $g = 0$, a distribuição g - e - h reduz-se a

$$Y_h(Z) = Ze^{hZ^2/2}, \quad (\text{E2.4})$$

que também é conhecida como distribuição h .

Para ilustrar como o valor de g afecta o enviesamento, apresentamos na Figura E2.1 gráficos da função de densidade de probabilidade $f_g(y)$ para os valores de g que usámos neste estudo.

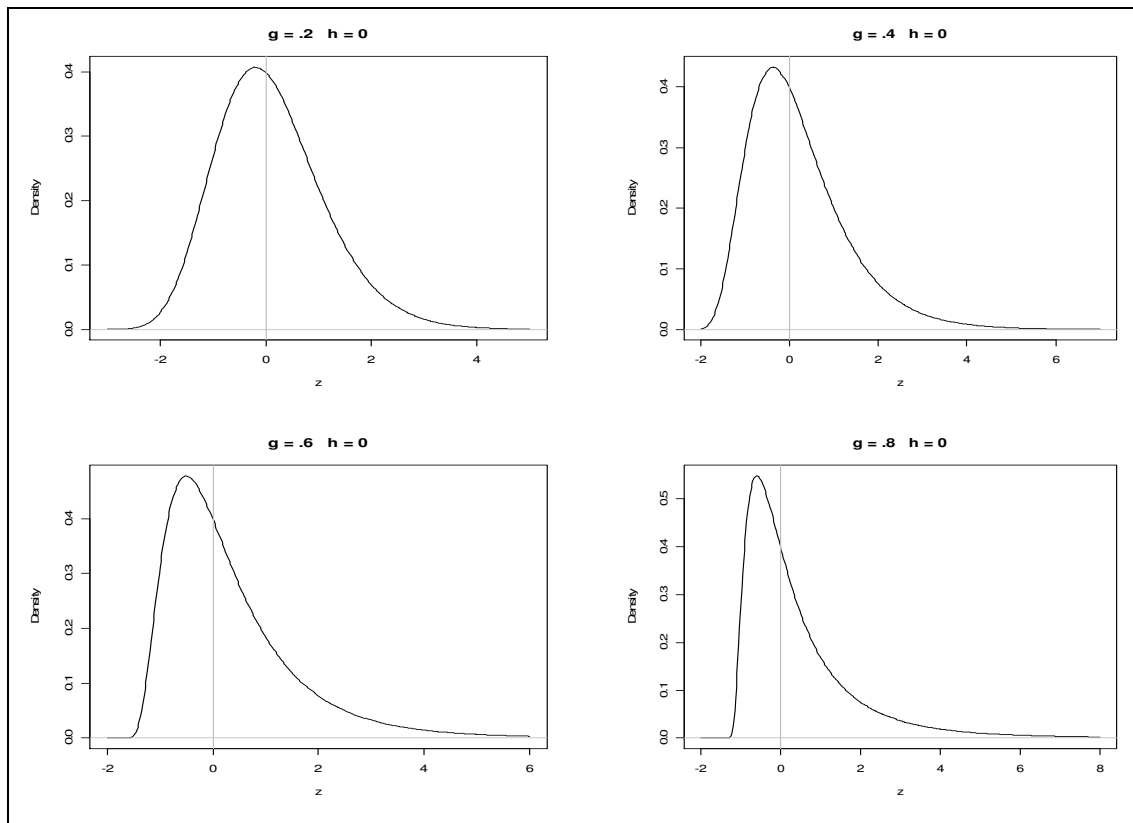


Figura E2.1 Gráficos da função densidade de probabilidade $f_g(y)$ para quatro valores de g (0.2, 0.4, 0.6 e 0.8).

Na Figura E2.2, mostramos como o valor de h influencia o alongamento, apresentando gráficos da função de densidade de probabilidade $f_h(y)$ para os valores de h usados neste estudo.

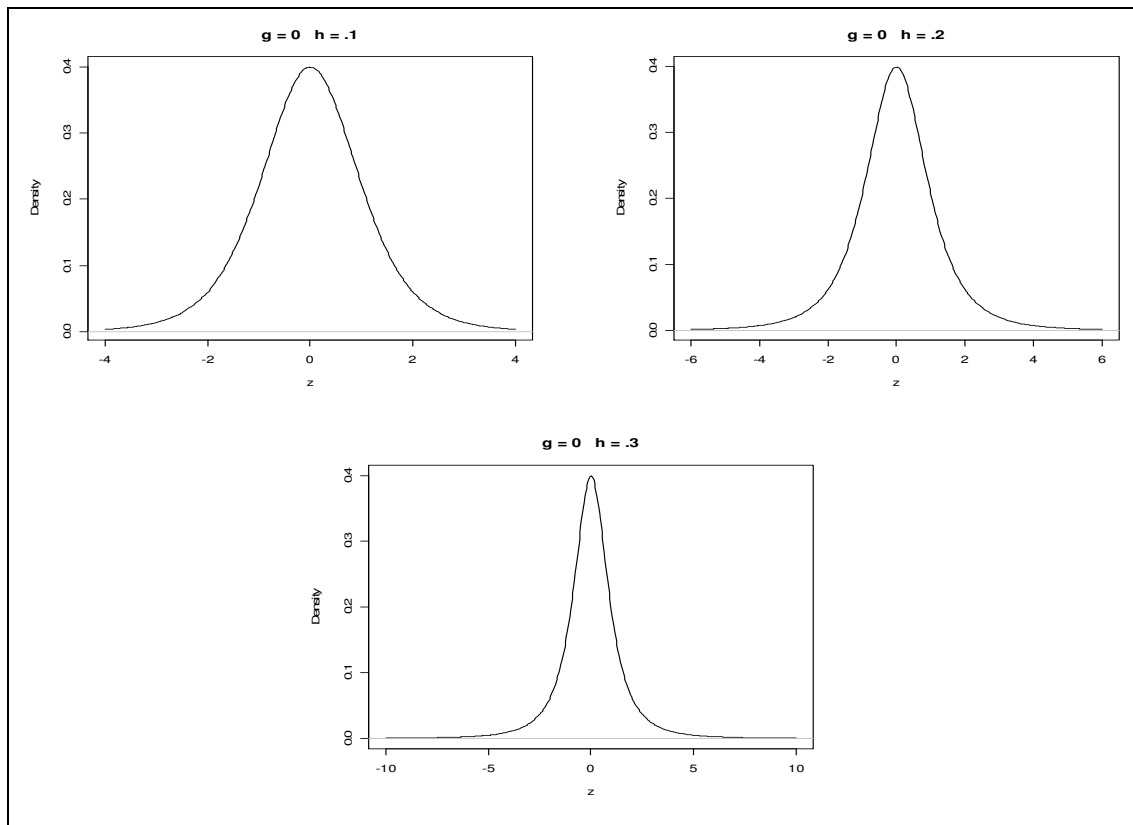


Figura E2.2 Gráficos da função densidade de probabilidade $f_h(y)$ para três valores de h (0.1, 0.2 e 0.3).

Para mostrar o efeito conjunto dos parâmetros g e h , são apresentados na Figura E2.3 gráficos de densidade de probabilidade $f_{g,h}(y)$ para os valores desses parâmetros que utilizámos no presente estudo.

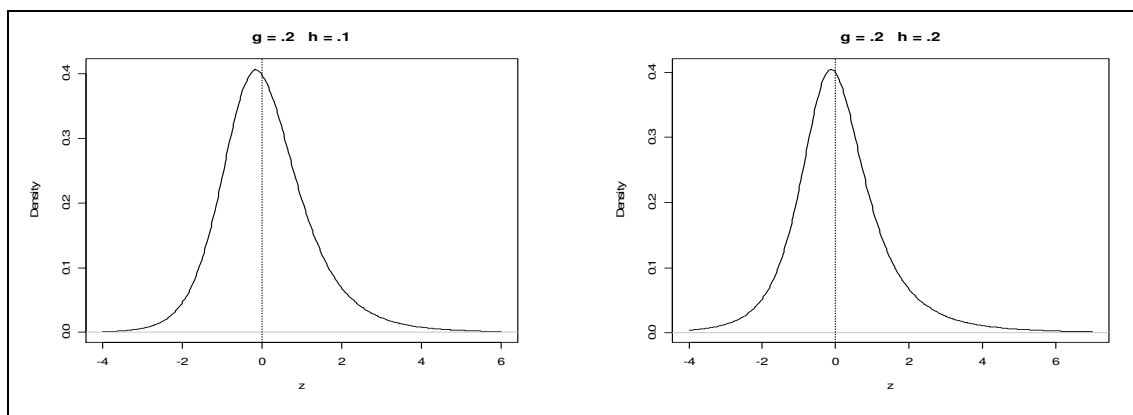


Figura E2.3 Gráficos da função densidade de probabilidade $f_{g,h}(y)$ para as combinações de quatro valores de g (0.2, 0.4, 0.6 e 0.8) com três valores de h (0.1, 0.2 e 0.3). [continua]

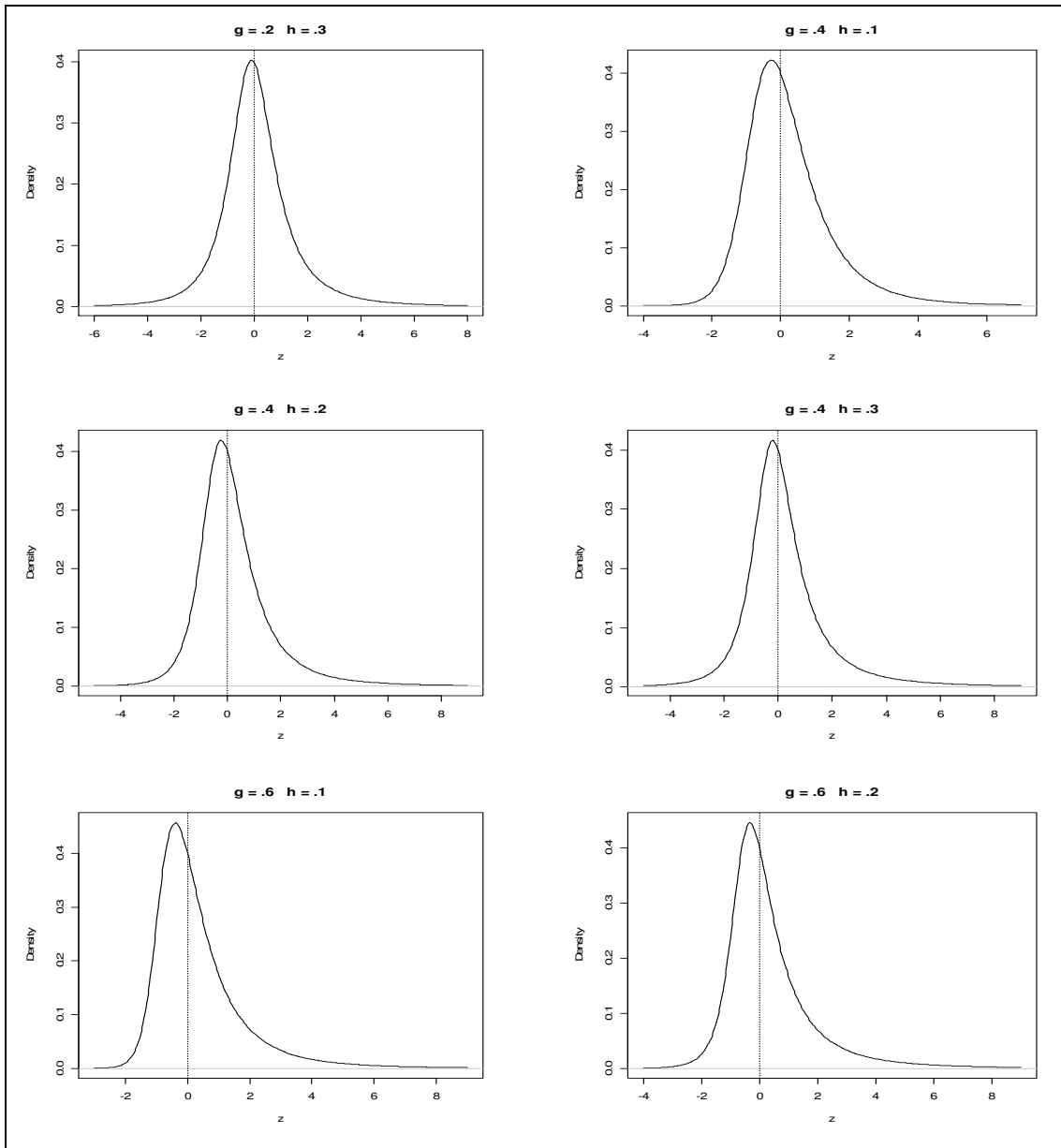


Figura E2.3 [Continuação] Gráficos da função densidade de probabilidade $f_{g,h}(y)$ para as combinações de quatro valores de g (0.2, 0.4, 0.6 e 0.8) com três valores de h (0.1, 0.2 e 0.3). [continua]

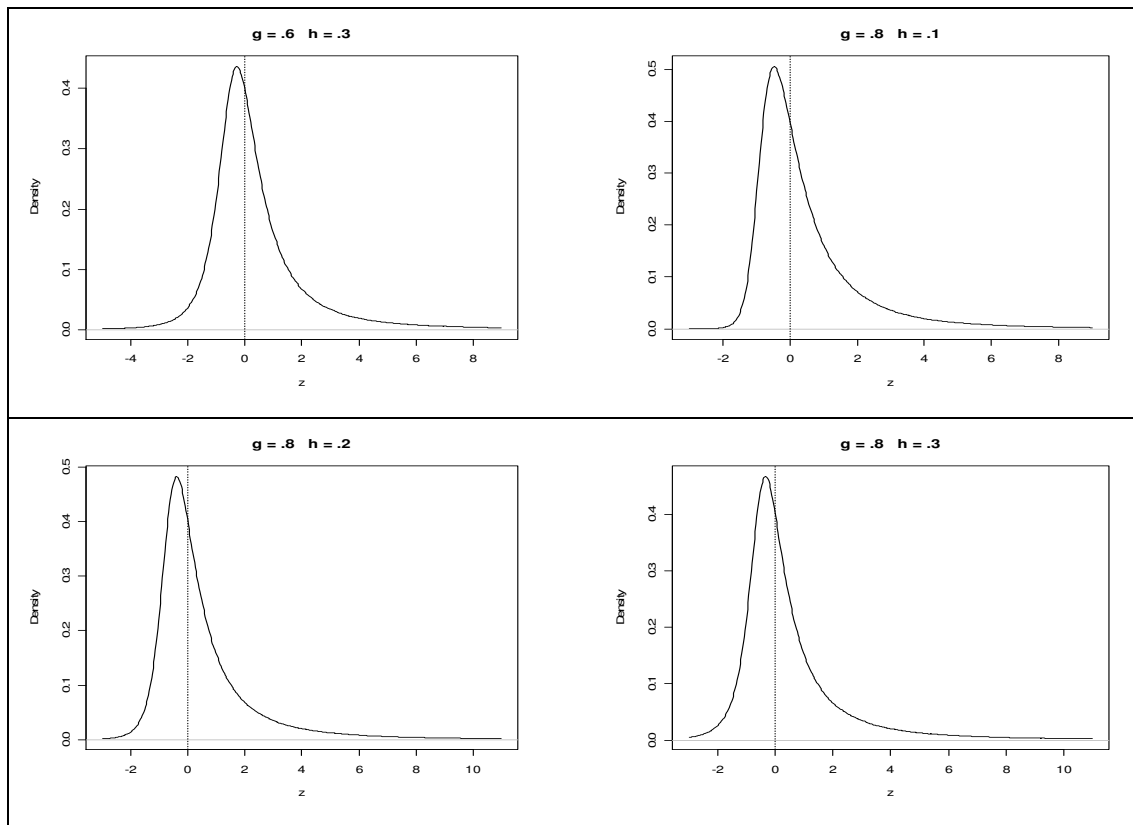


Figura E2.3 [Continuação] Gráficos da função densidade de probabilidade $f_{g,h}(y)$ para as combinações de quatro valores de g (0.2, 0.4, 0.6 e 0.8) com três valores de h (0.1, 0.2 e 0.3).

Vamos referir, seguidamente, os momentos destas distribuições, limitando-nos a apresentar os que nos interessam directamente no contexto desta investigação: os dois primeiros momentos. Começaremos por apresentar os casos mais simples da distribuição g e da distribuição h , depois o caso mais geral em que os valores de g e h são ambos diferentes de zero.

Distribuição g

Nesta sub-família, $Y(Z) = (e^{gZ} - 1)/g$. Os dois primeiros momentos são:

$$E(Y) = (e^{g^2/2} - 1)/g, \text{ e} \quad (\text{E2.5.1})$$

$$\text{Var}(Y) = e^{g^2} (e^{g^2} - 1)/g^2. \quad (\text{E2.5.2})$$

Distribuição h

Nesta sub-família, $Y(Z) = Ze^{hZ^2/2}$; a simetria das distribuições relativamente a zero implica que os momentos de ordem ímpar de uma distribuição h são zero, quando finitos. O segundo momento é:

$$\text{Var}(Y) = \frac{1}{(1-2h)^{3/2}}, \quad \text{com } 0 \leq h < \frac{1}{2}. \quad (\text{E2.6})$$

Verifica-se, assim, que a variância de Y só é finita para $h < \frac{1}{2}$.

Distribuição g - e - h

Para esta distribuição, $Y = [(e^{gZ} - 1)/g] e^{hZ^2/2}$.

Os dois primeiros momentos são:

$$E(Y) = \frac{1}{g\sqrt{1-h}} e^{g^2/2(1-h)} - 1, \quad \text{com } 0 \leq h < 1, \text{ e} \quad (\text{E2.7.1})$$

$$\text{Var}(Y) = \left[\frac{1}{g^2\sqrt{1-2h}} e^{2g^2/(1-2h)} - 2e^{g^2/2(1-2h)} + 1 \right] - \frac{1}{g^2(1-h)} \left[e^{g^2/2(1-h)} - 1 \right]^2,$$

$$\text{com } 0 \leq h < \frac{1}{2}. \quad (\text{E2.7.2})$$

Outros momentos de ordem superior podem ser calculados usando os resultados de Martinez (1981). Este autor derivou o momento de ordem n relativamente à origem, quando $g \neq 0$ e $0 \leq h < 1/n$:

$$E(Y^n) = \frac{1}{g^n \sqrt{1-nh}} \sum_{i=0}^n (-1)^i \binom{n}{i} e^{[(n-1)g]^2/2(1-nh)} \quad (\text{E2.8})$$

A vantagem de utilizar a distribuição g - e - h para modelar distribuições empíricas reside na sua maior flexibilidade em dar conta de uma grande variedade de propriedades das distribuições que quaisquer outras distribuições paramétricas.

Martinez e Iglewicz (1984) referem que a distribuição g - e - h engloba ou aproxima uma grande variedade de distribuições dependendo dos valores de g e h . Por exemplo, torna-se a distribuição normal padrão quando $g = 0$ e $h = 0$, a distribuição lognormal para $g = 1$ e $h = 0$ e a distribuição de Cauchy para $g = 0$ e $h = 0.97$. Um valor positivo de h implica caudas mais pesadas que as da distribuição normal padrão, enquanto que um valor negativo implica caudas menos pesadas (Badrinath e Chatterjee, 1991). No caso de $h = 0$, numa distribuição simétrica (ou seja, na distribuição normal padrão), existe alongamento neutro; mas essa definição de alongamento neutro também se aplica quando $g \neq 0$, isto é, para distribuições lognormais (Hoaglin, 1985).

Estes autores referem ainda que, para determinados valores de g e h , a distribuição g - e - h pode aproximar bem uma grande variedade de distribuições como as distribuições uniforme, beta, exponencial, Weibull, t de Student, logística e dupla exponencial.

2. Método

Avaliámos a potência dos testes de aleatorização e a dos testes t de Student, na comparação de dois grupos independentes equilibrados, manipulando as seguintes três variáveis:

Tamanho das amostras, com três níveis: $n_1 = n_2 = 8$, $n_1 = n_2 = 16$ e $n_1 = n_2 = 32$;

Tamanho do efeito, com 81 níveis: de -4 a 4, em intervalos de 0.1; e

Distribuição populacional, com 20 níveis: distribuições g - e - h , com valores g de 0, 0.2, 0.4, 0.6 e 0.8 cruzados com os valores de h de 0, 0.1, 0.2 e 0.3.

A potência foi avaliada para testes bicaudais, unicaudais à esquerda e unicaudais à direita, para $\alpha = 0.05$.

Tabela E2.1 Médias e desvios-padrão das 20 distribuições g - e - h simuladas

Distribuição		g	h	μ	σ
1	Gaussiana	0	0	0	1
2	Com	0.2	0	0.101	1.030
3	assimetria	0.4	0	0.208	1.128
4		0.6	0	0.329	1.314
5		0.8	0	0.471	1.630
6	Com	0	0.1	0	1.182
7	curtose	0	0.2	0	1.467
8		0	0.3	0	1.988
9	Com	0.2	0.1	0.118	1.229
10	assimetria	0.2	0.2	0.142	1.548
11	e	0.2	0.3	0.173	2.161
12	curtose	0.4	0.1	0.245	1.381
13		0.4	0.2	0.294	1.816
14		0.4	0.3	0.362	2.758
15		0.6	0.1	0.389	1.678
16		0.6	0.2	0.470	2.364
17		0.6	0.3	0.584	4.104
18		0.8	0.1	0.563	2.207
19		0.8	0.2	0.687	3.420
20		0.8	0.3	0.866	7.165

Na Tabela E2.1 apresentamos as 20 distribuições g - e - h que usámos neste estudo, indicando as suas médias e desvios-padrão. Escolhemos combinações de valores g e h que permitissem um leque variado de distribuições, desde distribuições relativamente aproximadas da Normal, até distribuições com forte assimetria e/ou alongamento.

Para simular dados destas distribuições, escrevemos programas em R (R Development Core Team, 2008), versão 2.7.1. (Anexo VII). Para cada condição simulámos 20 000 amostras e, para cada uma destas amostras, gerámos 999 combinações aleatórias para estimar a significância dos testes de aleatorização.

Tendo estabelecido o nível de significância de 0.05, calculámos a potência dos testes como a proporção de amostras em que a significância era igual ou inferior àquele valor.

No referente aos testes de aleatorização, com amostras de $n_1 = n_2 = 8$, o número total de combinações é de 12 870; este número sobe para 601 080 390, no caso de amostras de $n_1 = n_2 = 16$, e para $1.83262E+18$, no caso de amostras de $n_1 = n_2 = 32$. Como usámos, para estimar a significância do teste de aleatorização, um conjunto de referência de 1 000 combinações (1 ‘observada’, mais 999 geradas aleatoriamente), construímos intervalos de confiança a 99% para cada valor de potência. Quando comparámos a potência dos testes de aleatorização com a potência do teste t de Student, usámos a informação fornecida por estes intervalos de confiança.

3. Apresentação e discussão dos resultados

As matrizes com os resultados das simulações efectuadas podem ser consultadas no CD anexo, ficheiro ‘MatrizesDados_Estudo2’.

Não descrevemos os resultados da avaliação da potência dos testes de aleatorização, no caso da distribuição normal, pois essa descrição já foi feita no Estudo 1, para amostras com a mesma dimensão e iguais valores do tamanho do efeito. Assim, começamos por comparar a potência dos

testes de aleatorização com a potência dos testes t de Student e, seguidamente, para os testes de aleatorização, comparamos a sua potência no caso das distribuições com assimetria e/ou curtose com a sua potência no caso da distribuição normal.

3.1. Comparação da potência dos testes de aleatorização e t de Student

No Anexo VIII apresentam-se estatísticas descritivas da diferença de potência entre os testes de aleatorização e t de Student. A observação dessas estatísticas permite constatar que as diferenças são, de um modo geral, muito pequenas, variando entre -0,004 e 0.070. O tamanho das diferenças aumenta com o grau de assimetria e/ou curtose e diminui com o tamanho das amostras.

Contudo, importa dizer que só se observam diferenças significativas para valores de TE entre -0.7 e 0.7 e unicamente no caso de testes bicaudais (com uma exceção, abaixo assinalada). Na Tabela E2.2, apresentam-se, as diferenças significativas entre os dois testes, para o caso de testes bicaudais.

Tabela E2.2 Diferença entre a potência dos testes de aleatorização e t de Student (testes bicaudais)

$n_1 = n_2 = 8$						
	D 14	D 16	D 17	D 18	D 19	D 20
TE	(g = 0.4, h = 0.3)	(g = 0.6, h = 0.2)	(g = 0.6, h = 0.3)	(g = 0.8, h = 0.1)	(g = 0.8, h = 0.2)	(g = 0.8, h = 0.3)
-0.7	0.040	0.043	0.033	0.044	0.043	0.024
-0.6	0.040	0.042	0.039	0.048	0.048	0.028
-0.5	0.043	0.044	0.047	0.050	0.053	0.036
-0.4	0.039	0.039	0.054	0.044	0.057	0.045
-0.3	0.035	0.033	0.055	0.035	0.055	0.059
-0.2	0.025	0.025	0.045	0.027	0.045	0.070
-0.1	0.018	0.017	0.028	0.019	0.027	0.050
0	0.013	0.013	0.017	0.015	0.018	0.020
0.1	0.017	0.017	0.026	0.020	0.025	0.047
0.2	0.025	0.023	0.043	0.026	0.044	0.070
0.3	0.033	0.034	0.051	0.037	0.054	0.058
0.4	0.041	0.039	0.058	0.044	0.059	0.047
0.5	0.043	0.044	0.047	0.049	0.056	0.036
0.6	0.038	0.041	0.038	0.045	0.046	0.026
0.7	0.032	0.035	0.028	0.040	0.039	0.021

$n_1 = n_2 = 16$			$n_1 = n_2 = 32$
	D 17	D 19	D 20
TE	(g = 0.6, h = 0.3)	(g = 0.8, h = 0.2)	(g = 0.8, h = 0.3)
-0.7	0.021	0.027	0.016
-0.6	0.026	0.031	0.020
-0.5	0.032	0.038	0.025
-0.4	0.039	0.042	0.033
-0.3	0.044	0.045	0.043
-0.2	0.042	0.040	0.060
-0.1	0.024	0.024	0.046
0	0.015	0.016	0.019
0.1	0.024	0.024	0.047
0.2	0.039	0.038	0.057
0.3	0.044	0.046	0.047
0.4	0.040	0.046	0.036
0.5	0.034	0.038	0.027
0.6	0.023	0.032	0.018
0.7	0.021	0.024	0.016

Nota: A negrito indicam-se as diferenças significativas

Para amostras com 8 sujeitos por grupo, as diferenças observam-se nas distribuições D14 ($g = 0.4, h = 0.3$), D16 ($g = 0.6, h = 0.2$), D17($g = 0.6, h = 0.3$), D18 ($g = 0.8, h = 0.1$), D19 ($g = 0.8, h = 0.2$) e D20 ($g = 0.8, h = 0.3$). Para amostras com 16 sujeitos por grupo, as diferenças só se observam nas distribuições D17($g = 0.6, h = 0.3$), D19 ($g = 0.8, h = 0.2$) e D20 ($g = 0.8, h = 0.3$) e, finalmente, nas amostras com 32 sujeitos por grupo, unicamente na D20 ($g = 0.8, h = 0.3$).

Não existem, assim, diferenças de potência entre os testes de aleatorização e os testes *t* de Student para as distribuições simuladas que só apresentam assimetria (D2 a D5) ou só apresentam curtose (D6 a D8). É necessário a combinação de assimetria e curtose, e valores elevados de assimetria e/ou de curtose, para que se possam observar essas diferenças. Mas, com o aumento do número dos sujeitos por grupo, as diferenças vão diminuindo.

No que respeita a testes unicaudais, não se observaram diferenças significativas, com uma exceção: no caso de 8 sujeitos por grupo, para a D20 ($g = 0.8, h = 0.3$), a distribuição com mais assimetria e curtose, existem diferenças significativas para valores de TE de -0.2, -0.1, 0.1 e 0.2. Essas diferenças foram, respectivamente, de 0.046, 0.039, 0.038 e 0.044.

3.2. Influência da assimetria e curtose na potência dos testes de aleatorização

No Anexo IX, apresentam-se, graficamente, os resultados relativos à potência alcançada pelos testes de aleatorização no caso das distribuições *g-e-h* em comparação com a sua potência no caso da distribuição normal. Descrevemos, primeiro, a influência da assimetria,

da curtose e da assimetria e curtose para o caso de testes bicaudais, seguidamente para o caso de testes unicaudais.

3.2.1. Testes bicaudais

Na Figura E2.4 apresenta-se, para amostras de 8 sujeitos por grupo, a diferença entre a potência dos testes bicaudais para distribuições com assimetria e a sua potência para a distribuição normal.

Como se pode constatar, quanto mais forte a assimetria, maiores as diferenças de potência. Essa diferença apresenta um mesmo padrão, simétrico, para as quatro distribuições g simuladas (com assimetria positiva).

No caso da distribuição com mais forte assimetria, podemos observar que a partir do valor de TE de 0 há ganhos crescentes de potência até TE = |0.8|, em que o ganho é de 0.19, seguidamente os ganhos vão decrescendo até TE = |1.7|. A partir deste valor, observam-se pequenas perdas de potência, com um máximo de 0.03, para TE = |2.2|.

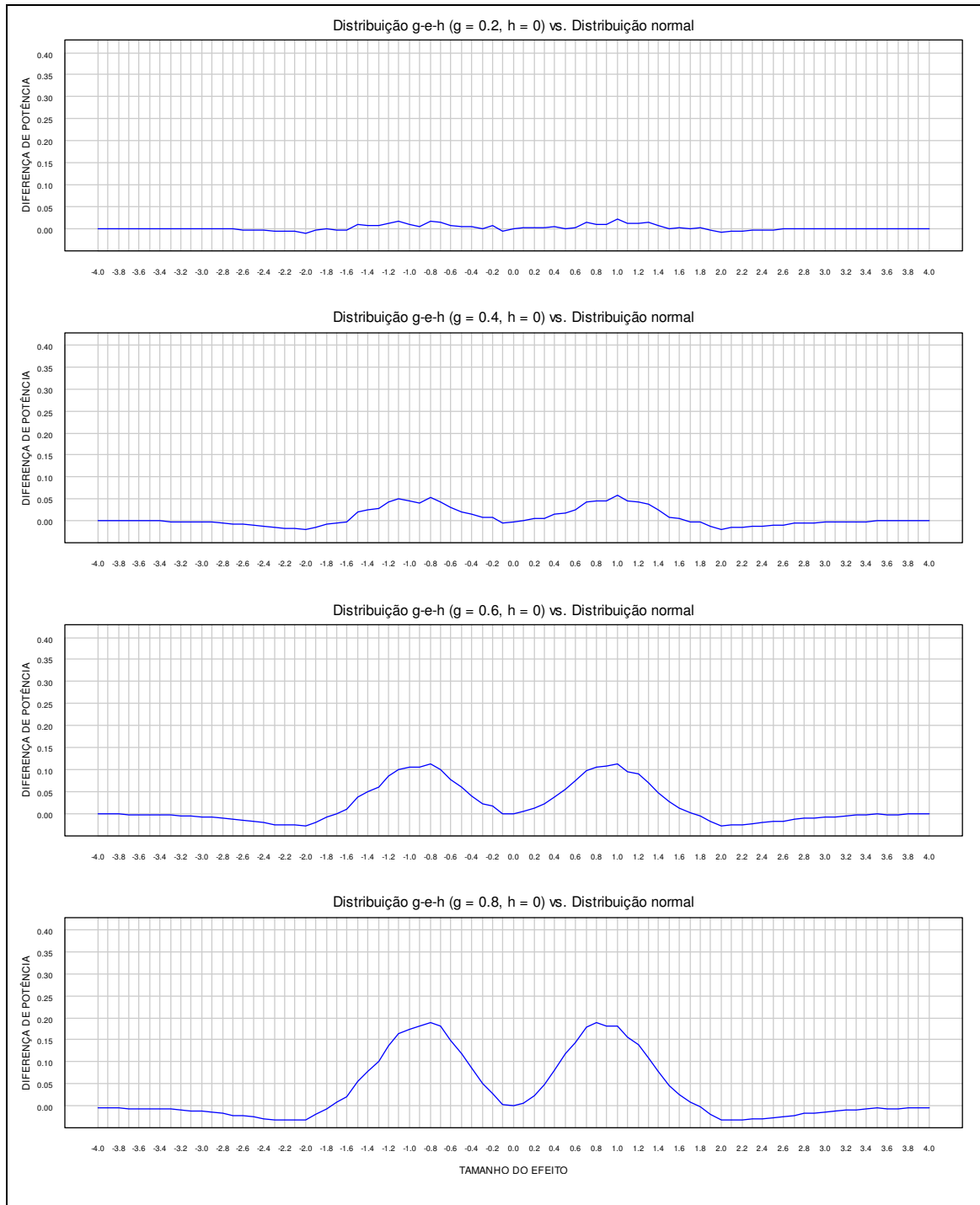


Figura E2.4 Testes de aleatorização (bicaudais): diferença entre a sua potência para as distribuições assimétricas e a sua potência para a distribuição normal, com $n_1 = n_2 = 8$.

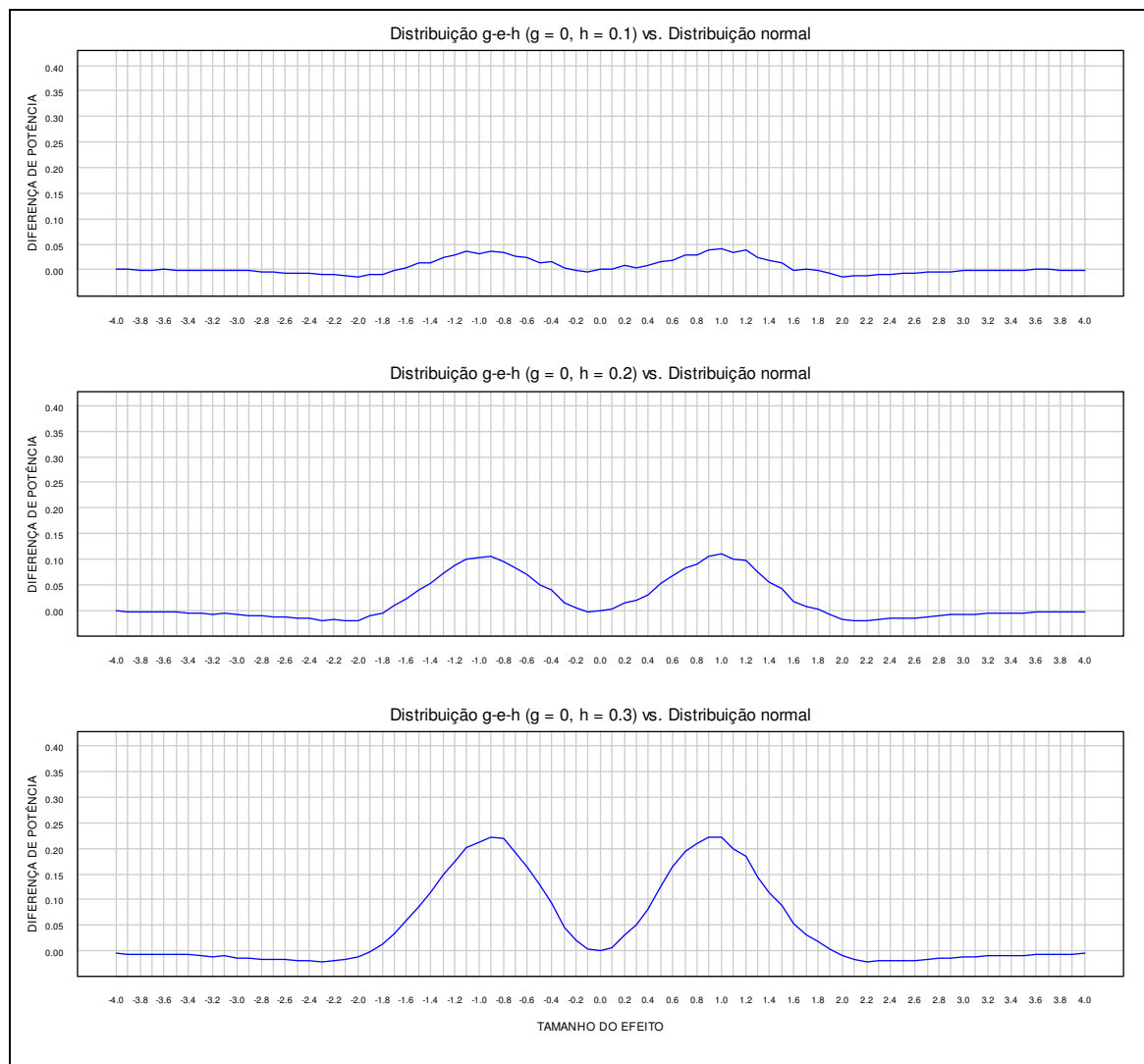


Figura E2.5 Testes de aleatorização (bicaudais): diferença entre a sua potência para as distribuições com curtose e a sua potência para a distribuição normal, com $n_1 = n_2 = 8$.

Na Figura E2.5 apresenta-se a diferença entre a potência dos testes bicaudais para distribuições com curtose e a potência dos mesmos testes para a distribuição normal.

Pode-se verificar que, quanto maior a curtose, maiores as diferenças de potência. O padrão das diferenças é semelhante ao anteriormente observado nas distribuições com assimetria.

No caso da distribuição que apresenta maior curtose, pode-se ver que a partir do valor de TE de 0 há ganhos crescentes de potência até TE = |1.0|, em que o ganho é de 0.22, seguidamente os ganhos vão decrescendo até TE = |1.9|. A partir deste valor, observam-se pequenas perdas de potência, com um máximo de 0.02, para TE = |2.2|.

Para as outras 12 distribuições simuladas (ver Anexo IX), com assimetria e curtose, observa-se o mesmo padrão das diferenças de potência. Quanto mais assimetria e curtose, maior o efeito, os quais não são apenas aditivos. Assim, por exemplo, para a distribuição com maior assimetria e curtose ($g = 0.8$ e $h = 0.3$), verificam-se ganhos de potência a partir do valor de TE de 0 até |2.2|, com um máximo de 0.65, para um valor de TE de cerca de |0.5|. A perda de potência, a partir do valor de TE de |2.2| é muito pequena, sendo, no máximo de 0.01 para um valor de TE de cerca de |2.8|.

Os efeitos descritos respeitam a amostras com $n_1 = n_2 = 8$. Verifica-se uma diminuição destes efeitos no caso das amostras maiores: embora o padrão seja semelhante, não só o tamanho das diferenças diminui, como o leque de valores de TE em que se verificam ganhos de potência é mais pequeno. Assim, para a distribuição com maior assimetria e curtose, no caso de $n_1 = n_2 = 16$, verificam-se ganhos de potência a partir do valor de TE de 0 até |1.5|, com um máximo de 0.62, para um valor de TE de cerca de |0.4|. No caso de $n_1 = n_2 = 32$, os ganhos de potência observam-se a partir do valor de TE de 0 até |1.0|, com um máximo de 0.57, para um valor de TE de cerca de |0.3|.

3.2.2. Testes unicaudais

Na Figura E2.6 apresenta-se, para amostras com $n_1 = n_2 = 8$, a diferença entre a potência dos testes unicaudais para distribuições com assimetria e para a distribuição normal.

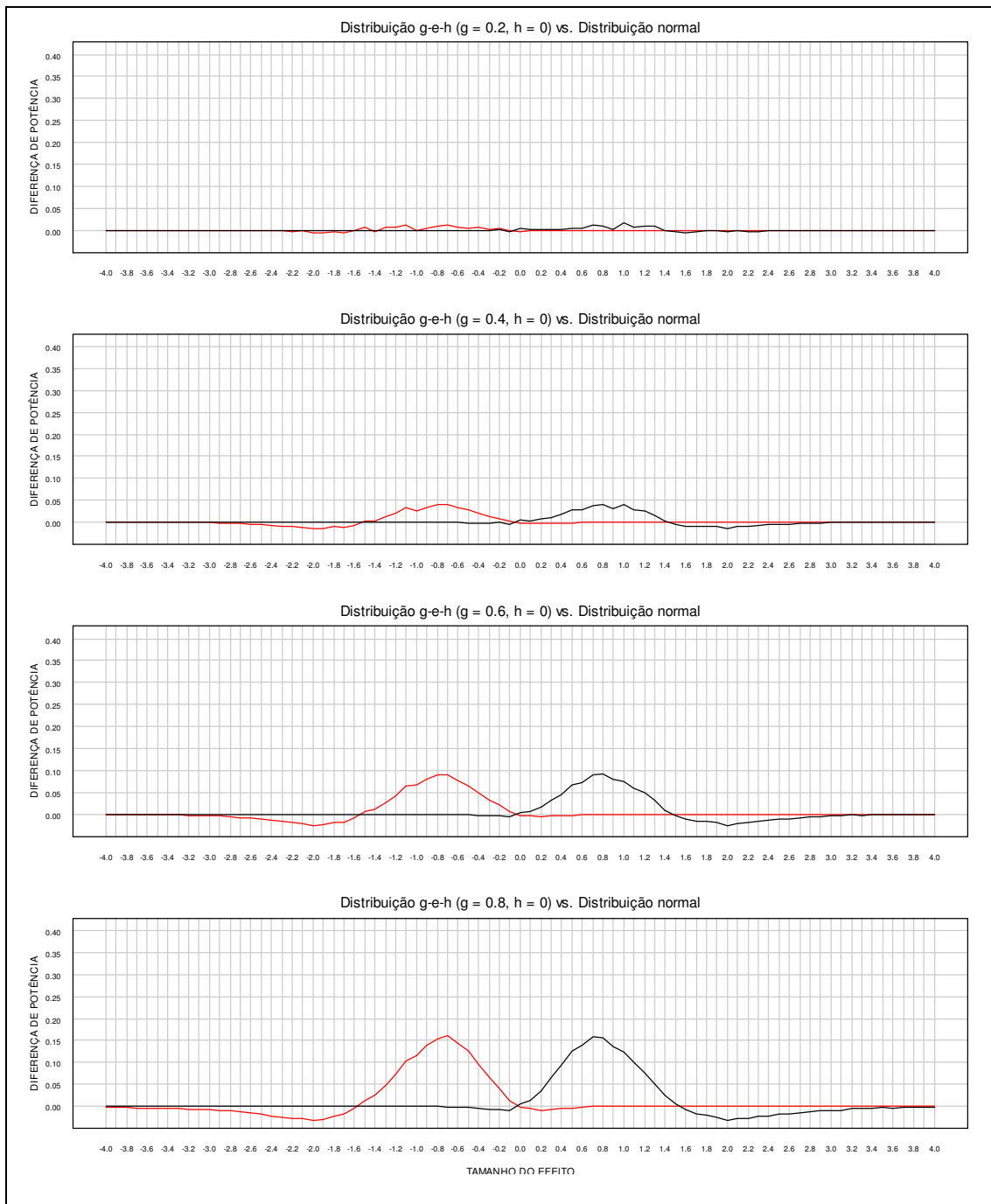


Figura E2. 6 Testes de aleatorização unicaudais: diferença entre a sua potência para as distribuições assimétricas e a sua potência para a distribuição normal, com $n_1 = n_2 = 8$ [A vermelho: teste unicaudal à esquerda; a negro: teste unicaudal à direita]

Também no caso dos testes unicaudais, quanto mais forte a assimetria, maiores as diferenças de potência. Como o padrão das diferenças é

aproximadamente igual para os testes à esquerda e à direita, descrevemos apenas as diferenças para estes últimos.

Para a primeira distribuição (com $g = 0.2$), as diferenças são muito pequenas. O padrão das diferenças começa a tornar-se perceptível nas distribuições seguintes. No caso da distribuição com mais forte assimetria, a partir de $TE = 0$ podemos verificar ganhos de potência até $TE = 0.7$, em que o ganho é de 0.16, decrescendo os ganhos até $TE = 1.5$. A partir desse valor, observam-se pequenas perdas de potência, com um máximo de 0.03 para $TE = 2.0$. A partir desse valor de TE , as diferenças vão diminuindo.

Na Figura E2.7 apresenta-se a diferença entre a potência dos testes unicaudais para distribuições com curtose e a potência dos mesmos testes para a distribuição normal.

Pode-se verificar que, quanto maior a curtose, maior as diferenças de potência. O padrão das diferenças é semelhante ao que anteriormente observámos nas distribuições com assimetria.

No caso da distribuição que apresenta maior curtose, a partir de $TE = 0$ podemos verificar ganhos de potência até $TE = 0.7$, com um ganho de 0.19, decrescendo os ganhos até $TE = 1.5$. A partir desse valor, observam-se pequenas perdas de potência, com um máximo de 0.02 para $TE = 2.2$. A partir desse valor de TE , as diferenças vão diminuindo.

Para as outras 12 distribuições simuladas (ver Anexo IX), com assimetria e curtose, observa-se o mesmo padrão de diferenças de potência. Quanto maior a assimetria e curtose, maiores os efeitos, que não são meramente aditivos. Assim, por exemplo, para a distribuição com maior assimetria e curtose ($g = 0.8$ e $h = 0.3$), verificam-se ganhos de potência para valores de TE de 0 a 2.0, com um máximo de 0.60, para $TE = 0.5$. A perda de

potência, a partir do valor de TE de 2.0 é muito pequena, sendo, no máximo de 0.01, para um valor de TE de 2.5.

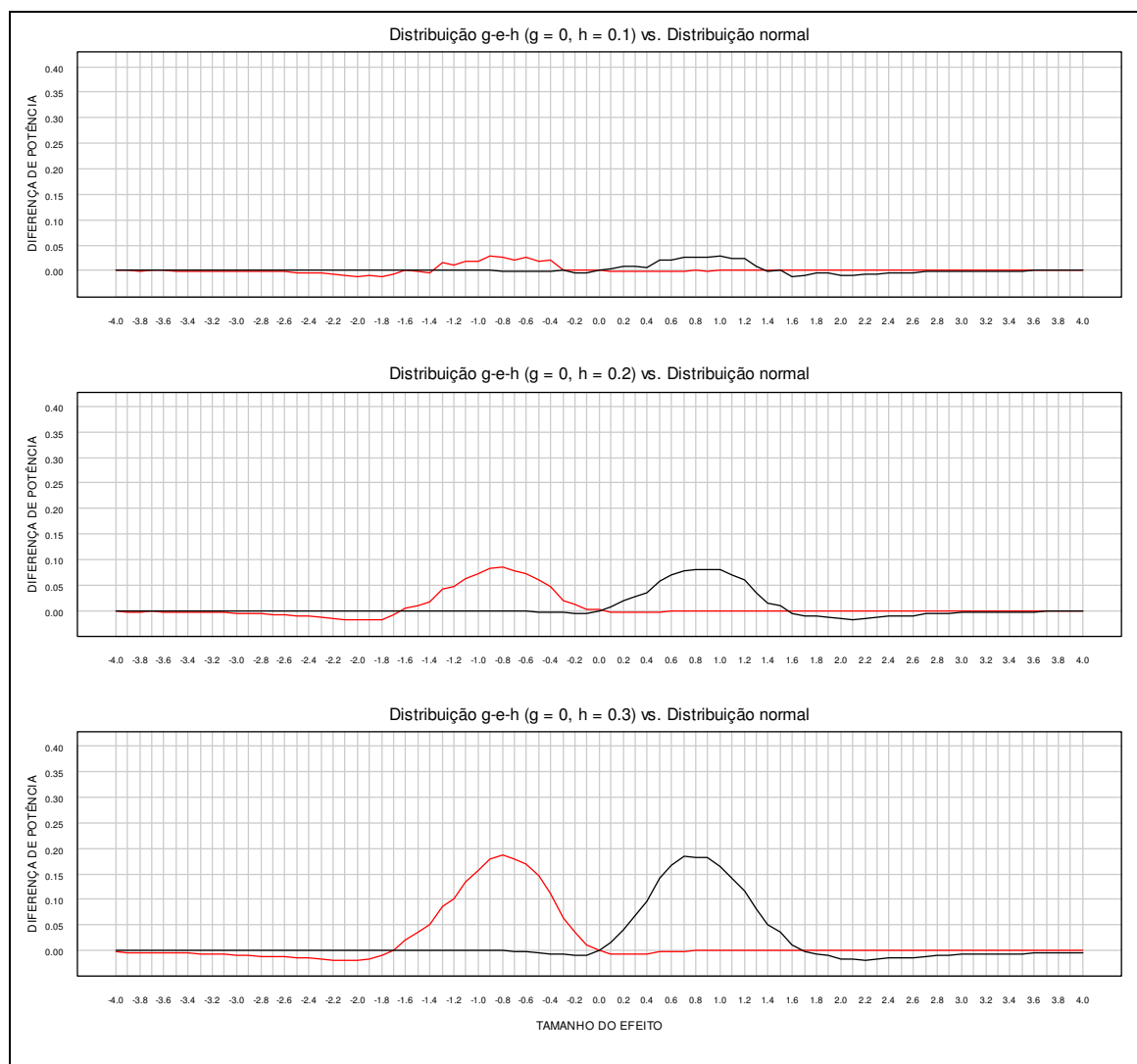


Figura E2. 7 Testes de aleatorização unicaudais: diferença entre a sua potência para as distribuições com curtose e a sua potência para a distribuição normal, com $n_1 = n_2 = 8$. [A vermelho: teste unicaudal à esquerda; a negro: teste unicaudal à direita]

Os efeitos que acabámos de descrever foram obtidos com amostras com $n_1 = n_2 = 8$. Verifica-se uma diminuição destes efeitos no caso das amostras maiores: o padrão é semelhante, mas não só o tamanho das diferenças diminui, como o leque de valores de TE em que se verificam ganhos de potência é mais pequeno.

Assim, para a distribuição com maior assimetria e curtose, no caso de $n_1 = n_2 = 16$, verificam-se ganhos de potência a partir do valor de TE de 0 até 1.3, com um máximo de 0.55, para um valor de TE de 0.4. No caso de $n_1 = n_2 = 32$, os ganhos de potência observam-se a partir do valor de TE de 0 até 0.9, com um máximo de 0.51, para um valor de TE de cerca de 0.3.

4. Conclusões

Neste estudo foi avaliada a potência, dos testes de aleatorização, na comparação de duas amostras independentes e equilibradas, em função da forma das distribuições.

Pudemos constatar que a potência destes testes para as distribuições simuladas, com diferentes graus de assimetria, de curtose e de assimetria e curtose, foi superior, em geral, à sua potência no caso da distribuição normal. Os ganhos de potência foram tanto maiores quanto maior a assimetria, quanto maior a curtose e quanto maior a interação da assimetria e curtose.

Mas esta constatação tem de ser qualificada. Em primeiro lugar, os ganhos de potência observaram-se apenas para um intervalo relativamente pequeno de valores do tamanho do efeito, embora nesse intervalo estejam contidos os valores que, segundo Cohen, são os mais comuns nos estudos experimentais nas ciências do comportamento. Em segundo lugar, esses ganhos estão inversamente relacionados com o tamanho das amostras, sendo de supor que com amostras maiores do que as usadas, os ganhos de potência se tornem irrelevantes. Em terceiro lugar, importa referir que fora desse intervalo de valores do tamanho do efeito em que se observaram os ganhos de potência, se verificaram

perdas de potência, ainda que muito pequenas quando comparadas com os ganhos.

No que respeita à comparação do teste de aleatorização com o teste t de Student, em termos de potência, verificou-se que, em geral, apresentam uma potência semelhante, com alguma vantagem para o primeiro. Essa vantagem só se observa quando as amostras são pequenas ou quando as distribuições têm um grau elevado de assimetria e curtose.

Estes resultados sugerem que se um investigador, ao planejar uma experiência, escolher o tamanho das suas amostras em função de um determinado tamanho do efeito populacional e de uma dada potência, assumindo uma distribuição normal, a potência do teste não será diminuída se as suas amostras provierem de distribuições assimétricas e com curtose.

Importa, ainda, salientar que os resultados deste estudo não podem ser extrapolados para situações em que as amostras não sejam equilibradas e provenientes de distribuições não contínuas e heterocedásticas.

ESTUDO 3

1. Introdução

Os testes clássicos para a comparação de dois grupos independentes, o teste t de Student e o teste F da Análise de Variância, são considerados robustos relativamente à violação do pressuposto de heterocedasticidade, desde que as amostras sejam equilibradas. Mas, se as amostras tiverem diferentes tamanhos, variâncias desiguais afectam criticamente as probabilidades dos erros de Tipo I e de Tipo II (Boneau, 1960; Games & Howell, 1976; Kohr & Games, 1977; Rogan & Keselman, 1977; Scheffe, 1959; Zimmerman, 2006).

Quando a maior amostra tem a variância mais elevada, verifica-se uma diminuição acentuada da probabilidade de rejeitar a hipótese nula para todas as diferenças entre médias, incluindo uma diferença de zero. Ou seja, nesta situação há uma diminuição da potência do teste, e a probabilidade de cometer um erro de Tipo I fica abaixo da nominal. Quando a menor amostra tem a variância mais elevada, constata-se um efeito inverso: há uma elevação acentuada da probabilidade de rejeitar a hipótese nula incluindo a probabilidade de cometer um erro de Tipo I.

O objectivo deste estudo é avaliar o comportamento, em termos de potência, do teste de aleatorização, na comparação de dois grupos independentes, para distribuições normais, em condições em que existe heterocedasticidade e as amostras não são equilibradas. A potência do teste t de Student, para as mesmas condições, servirá de padrão de comparação.

2. Método

Avaliámos a potência do teste de aleatorização e a do teste t de Student, na comparação de dois grupos independentes, com distribuição normal,

variando o tamanho das amostras, o seu equilíbrio, a homogeneidade das variâncias e o tamanho do efeito.

Relativamente ao tamanho e equilíbrio das amostras, escolhemos amostras de 8, 16 e 32 elementos, com amostras equilibradas de $n_1 = n_2 = 8$, $n_1 = n_2 = 16$ e $n_1 = n_2 = 32$, e amostras não-equilibradas com $n_1 = 8$ e $n_2 = 16$, $n_1 = 8$ e $n_2 = 32$, $n_1 = 16$ e $n_2 = 8$, e $n_1 = 32$ e $n_2 = 8$.

No que respeita à variabilidade das amostras, simulámos dados de distribuições com $\sigma = 1$, para a primeira amostra, e com $\sigma = 1, 3$ ou 10 para a segunda amostra.

Quanto ao Tamanho do Efeito (TE), usámos 81 valores, de -4 a 4 , com intervalos de 0.1 .

Quando se comparam amostras de populações com igual variância, é usual definir o TE como a diferença entre as médias populacionais em unidades de desvio-padrão populacional. É a medida ‘d’ de Cohen (1988).

Neste estudo, porém, simulámos dados de populações com diferentes variâncias, o que levanta a questão de qual desvio-padrão usar para definir o tamanho do efeito. Decidimos usar o desvio-padrão da primeira amostra, que em todas as condições é igual a 1:

$$TE = \frac{\mu_1 - \mu_2}{\sigma_1} \quad (\text{E3.1})$$

Esta fórmula corresponde à medida ‘ Δ ’, proposta por Glass (1976, 1977):

$$\Delta = \frac{\bar{x}_1 - \bar{x}_2}{s_2}, \quad (\text{E3.2})$$

que tem sido usada em estudos meta-analíticos, em que a primeira amostra corresponde ao grupo de tratamento e a segunda ao grupo de controlo, com a justificação de que o desvio-padrão da população que não recebeu o tratamento será provavelmente mais comparável nos diversos estudos meta-analisados (Becker, 1988; Carlson & Schmidt, 1999; Rosenthal, 1994).

Embora Hedges e Olkin (1985) tenham defendido o uso de um desvio-padrão baseado na combinação dos desvios-padrão das duas amostras em vez do desvio-padrão do grupo de controlo, para obter menor enviesamento e variância da estimativa do TE, usamos neste estudo o TE definido pela equação E3.1, pois o nosso objectivo não é o de estimação, mas o de definir uma métrica para as diferenças de médias entre as populações, que, por planeamento, são heterocedásticas.

A potência foi avaliada para testes bicaudais, unicaudais à esquerda e unicaudais à direita, para $\alpha = 0.05$.

Na Tabela E3.1 apresentamos os 19 grupos de condições, definidos em termos da dimensão, equilíbrio e homogeneidade das variâncias das amostras, sendo que para cada um desses grupos de condições avaliámos a potência dos testes para 81 tamanhos de efeito. O número total de condições deste estudo é, assim, de 1539 (19x81).

Para simular os dados deste estudo, escrevemos programas em R (R Development Core Team, 2008), versão 2.7.1. (Anexo X). Para cada condição simulámos 20 000 amostras e, para cada uma destas amostras, gerámos 999 combinações aleatórias para estimar a significância dos testes de aleatorização.

Calculámos a potência dos testes como a proporção de amostras em que a significância era igual ou inferior a 0.05, o valor do nível de significância.

Tabela E3.1 Dimensões e desvios-padrão das amostras dos 19 grupos de condições

Grupos de condições	n_1	n_2	σ_1	σ_2
1	8	8	1	1
2	8	8	1	3
3	8	8	1	10
4	16	16	1	1
5	16	16	1	3
6	16	16	1	10
7	32	32	1	1
8	32	32	1	3
9	32	32	1	10
10	8	16	1	1
11	8	16	1	3
12	8	16	1	10
13	8	32	1	1
14	8	32	1	3
15	8	32	1	10
16	16	8	1	3
17	16	8	1	10
18	32	8	1	3
19	32	8	1	10

Relativamente ao teste de aleatorização, como o número de combinações é muito elevado (é de 1.83262E+18, no caso de amostras de $n_1 = n_2 = 32$), usámos um conjunto de referência de 1 000 combinações (1 ‘observada’, mais 999 geradas aleatoriamente) para estimar a significância do teste. Calculámos intervalos de confiança a 99% para os valores estimados de potência do teste de aleatorização, que utilizámos quando comparámos a potência deste teste com a potência do teste t de Student.

3. Apresentação e discussão dos resultados

As matrizes com os resultados das simulações efectuadas podem ser consultadas no CD anexo, ficheiro ‘MatrizesDados_Estudo3’.

3.1. Comparação da potência dos testes de aleatorização e dos t de Student

Na Tabela E3.2 apresentamos estatísticas (mínimo e máximo) da diferença entre a potência do teste de aleatorização e a potência do teste t de Student, por grupo de condições.

Tabela E3.2 Estatísticas (mínimos e máximos) para a diferença entre a potência dos testes de aleatorização e t de Student

Grupo de condições	n1	n2	σ_1	σ_2		Esq	Dir	Bi
1	8	8	1	1	Mínimo	-0.003	-0.004	-0.004
					Máximo	0.001	0.001	0.002
2	8	8	1	3	Mínimo	-0.003	-0.003	0.000
					Máximo	0.002	0.002	0.008
3	8	8	1	10	Mínimo	-0.001	-0.002	0.003
					Máximo	0.002	0.002	0.008
4	16	16	1	1	Mínimo	-0.004	-0.002	-0.004
					Máximo	0.001	0.001	0.001
5	16	16	1	3	Mínimo	-0.003	-0.003	-0.001
					Máximo	0.000	0.000	0.005
6	16	16	1	10	Mínimo	-0.002	-0.003	0.001
					Máximo	0.001	0.001	0.006
7	32	32	1	1	Mínimo	-0.003	-0.002	-0.003
					Máximo	0.001	0.000	0.001
8	32	32	1	3	Mínimo	-0.002	-0.002	-0.002
					Máximo	0.001	0.001	0.003
9	32	32	1	10	Mínimo	-0.003	-0.004	-0.001
					Máximo	0.001	0.001	0.004

[Continua]

Tabela E3.2 [Continuação] Estatísticas (mínimos e máximos) para diferença entre a potência dos testes de aleatorização e t de Student

Grupo de condições	n1	n2	σ_1	σ_2		Esq	Dir	Bi
10	8	16	1	1	Mínimo	-0.003	-0.003	-0.003
					Máximo	0.001	0.001	0.001
11	8	16	1	3	Mínimo	0.000	0.000	-0.002
					Máximo	0.014	0.014	0.004
12	8	16	1	10	Mínimo	0.000	0.000	-0.001
					Máximo	0.014	0.015	0.003
13	8	32	1	1	Mínimo	-0.003	-0.005	-0.003
					Máximo	0.001	0.001	0.001
14	8	32	1	3	Mínimo	0.000	0.000	-0.004
					Máximo	0.015	0.014	0.002
15	8	32	1	10	Mínimo	0.000	0.000	-0.001
					Máximo	0.011	0.013	0.002
16	16	8	1	3	Mínimo	-0.020	-0.019	0.001
					Máximo	0.000	0.000	0.010
17	16	8	1	10	Mínimo	-0.024	-0.023	0.010
					Máximo	-0.001	-0.001	0.017
18	32	8	1	3	Mínimo	-0.023	-0.023	-0.003
					Máximo	0.000	0.000	0.001
19	32	8	1	10	Mínimo	-0.031	-0.032	-0.003
					Máximo	-0.004	-0.003	0.001

Como se pode verificar, os testes de aleatorização e t de Student são semelhantes em termos de potência. Para as condições, em que as amostras são equilibradas, as diferenças são todas inferiores a 0.010, quer para testes unicaudais, quer para testes bicaudais. Para as condições em que a maior variância está associada à amostra maior, a maior diferença é de 0.015, para testes unicaudais, sendo inferior a 0.010 para testes bicaudais. Relativamente às condições em que a maior variância está associada à amostra menor, a maior diferença é de -0.032, para testes unicaudais, e de 0.017 para testes bicaudais.

Os intervalos de confiança calculados para cada valor de potência do teste de aleatorização, permitem afirmar que nenhuma destas diferenças é significativa, do ponto de vista estatístico (para $\alpha = 0.01$).

3.2. Potência dos testes de aleatorização para $n_1 = n_2$

Na Figura E3.1 apresentamos a potência do teste de aleatorização para amostras equilibradas (teste bicaudal).

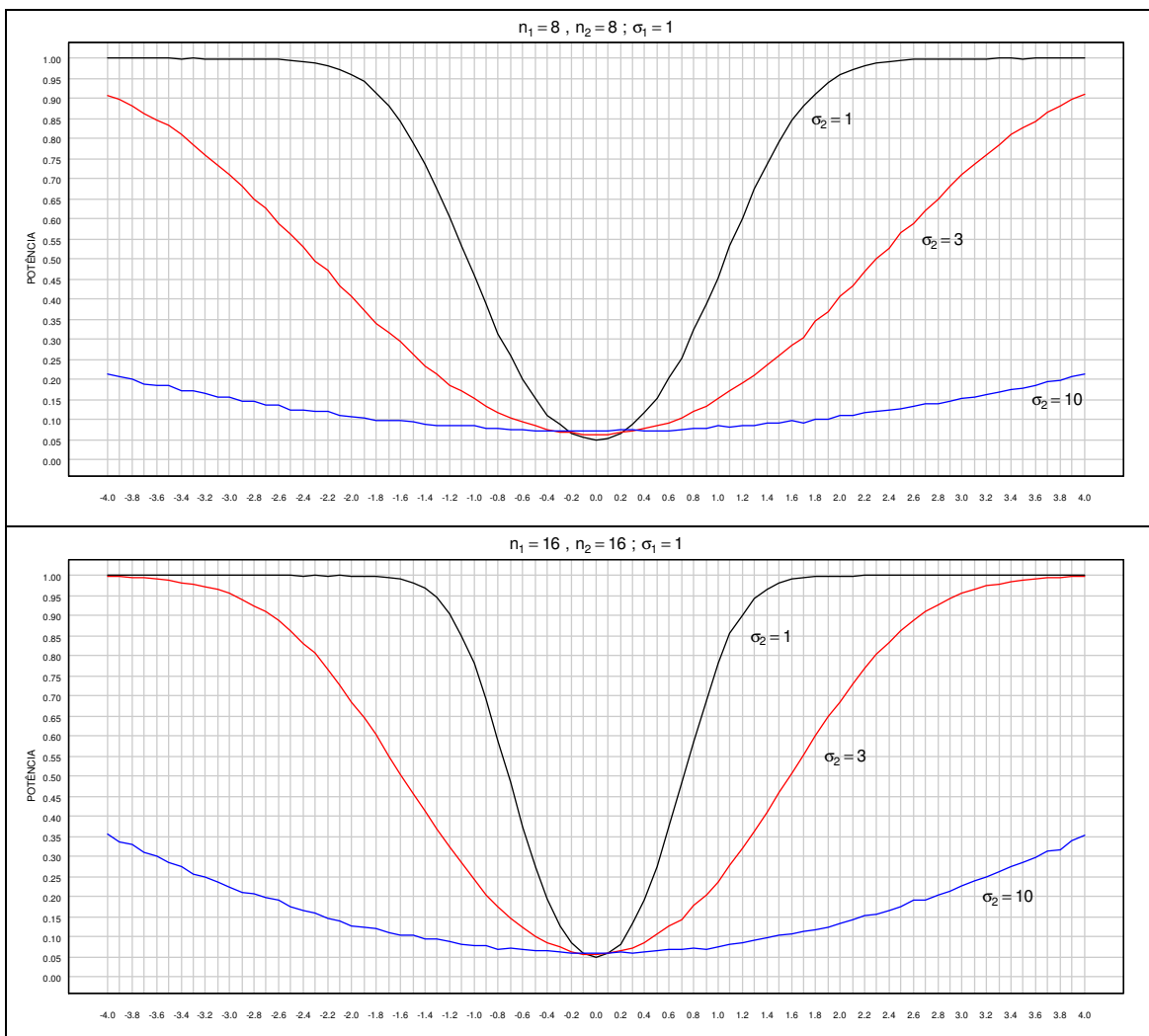


Figura E3.1 Potência do teste de aleatorização bicaudal para as condições com amostras equilibradas [Continua]

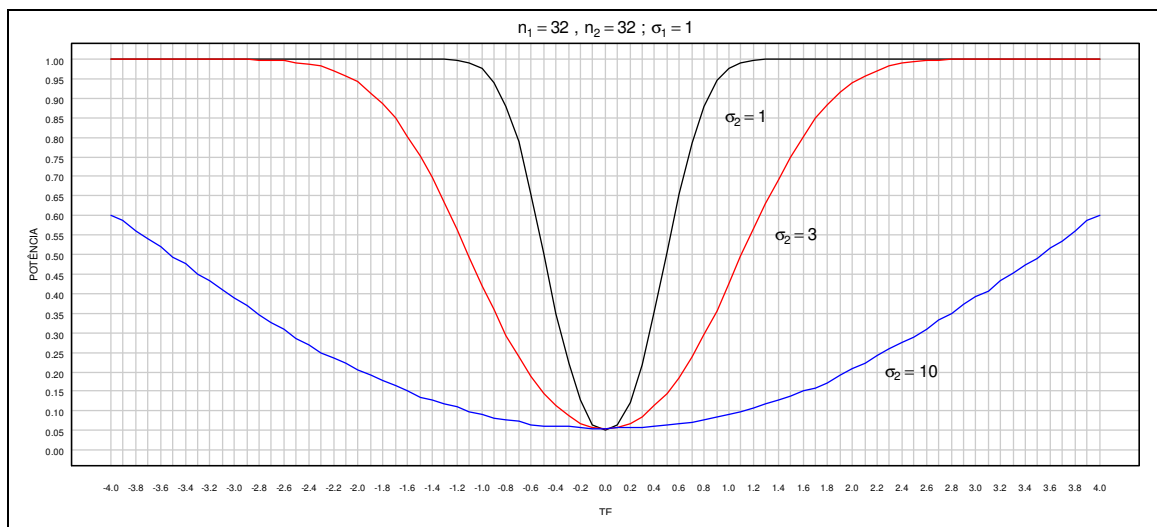


Figura E3.1 [Continuação] Potência do teste de aleatorização bicaudal para as condições com amostras equilibradas

Pode-se constatar que a potência do teste é fortemente influenciada pela heterogeneidade de variâncias: para todos os três níveis de tamanho das amostras, quanto maior a heterogeneidade, menor a potência.

Para as duas situações com heterogeneidade há aumentos de potência com o aumento do tamanho das amostras. Mas, esses aumentos são menores que o verificado na situação de homogeneidade.

As perdas relativas de potência, das situações com heterogeneidade em comparação com a situação de homogeneidade, são maiores com o aumento das amostras, para alguns valores de TE: quando $\sigma_2 = 3$, no intervalo de -1.5 a 1.5; e quando $\sigma_2 = 10$, no intervalo de -2.0 a 2.0.

Para um TE de 0.0, quando há heterogeneidade de variâncias, a potência excede o valor 0.05, mas, com o aumento do tamanho das amostras, tende a coincidir com o valor nominal.

Os resultados dos testes unicaudais são semelhantes, com a sua potência a diminuir em função do grau de heterogeneidade de variâncias.

3.3. Potência dos testes de aleatorização para $n_1 < n_2$

Apresentamos na Figura E3.2 os resultados da potência do teste de aleatorização bicaudal para as condições em que a maior variância está associada à amostra maior, em comparação com a condição em que as variâncias são iguais.

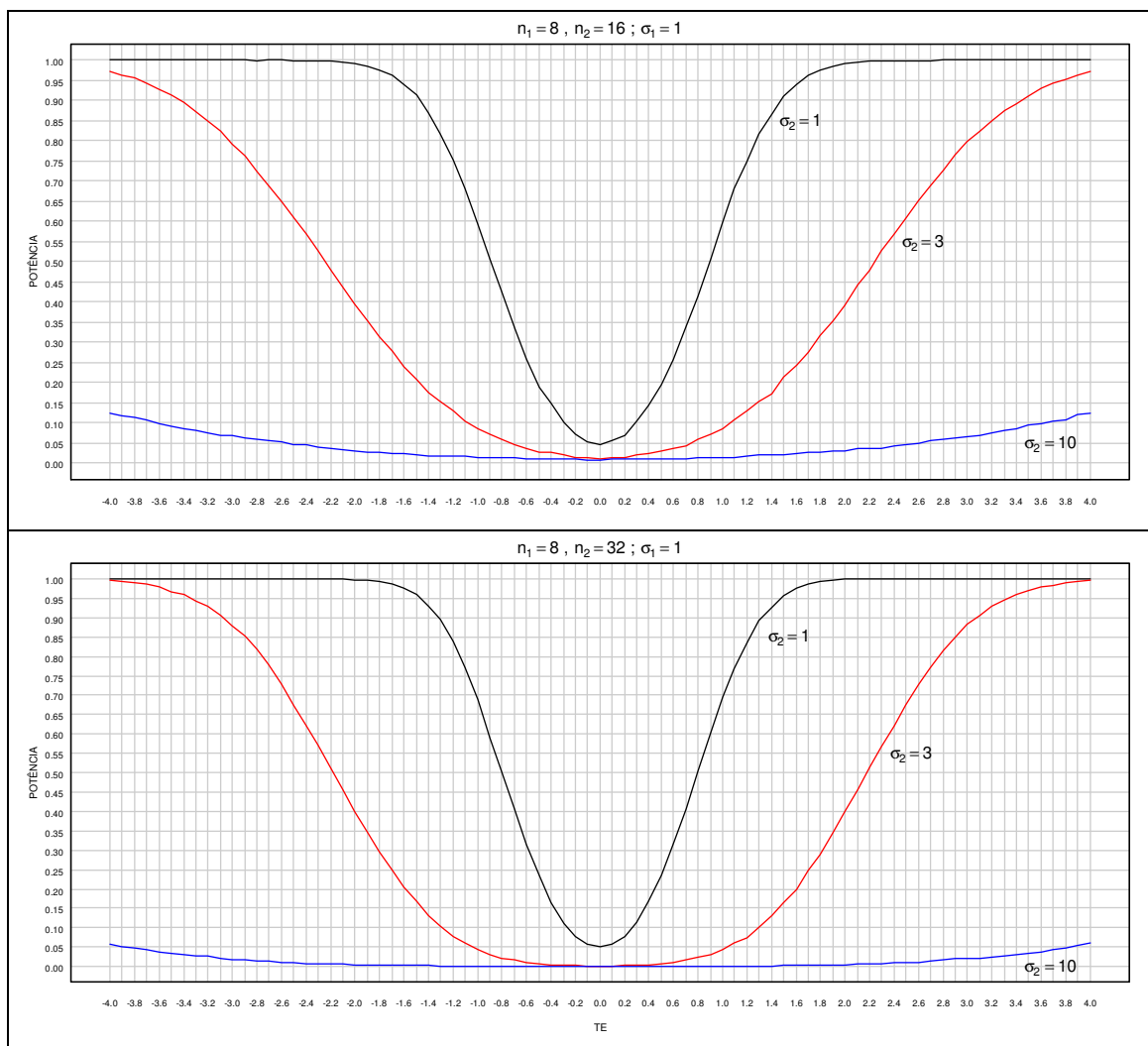


Figura E3.2 Potência do teste de aleatorização bicaudal para as condições em que a maior variância está associada à amostra menor em comparação com a condição em que há homogeneidade de variâncias

Como se pode observar na Figura E3.2, quanto maior o grau de heterogeneidade das variâncias maior é a perda de potência relativamente à situação em que há homogeneidade.

Quando existe homogeneidade ($\sigma_1 = 1, \sigma_2 = 1$), a potência é maior, para todos os valores de TE, quando $n_2 = 32$ do que quando $n_2 = 16$, ou seja, quando as amostras são mais desequilibradas (mas, no total, há mais elementos).

Quando não existe homogeneidade, porém, o efeito do desequilíbrio das amostras varia segundo o nível de heterogeneidade das variâncias. Assim, no caso de $\sigma_2 = 3$, a potência é maior quando há menos desequilíbrio ($n_1 = 8, n_2 = 16$) para valores de TE de 0.0 a |1.9|; a partir de TE = |2.0|, a potência é maior quando há mais desequilíbrio ($n_1 = 8, n_2 = 32$). No caso de $\sigma_2 = 10$, a potência é maior, para todos os valores de TE, quando há menos desequilíbrio.

Para melhor visualizar estes efeitos, apresentamos na Figura E3.3, a sobreposição dos gráficos mostrados na Figura E3.2.

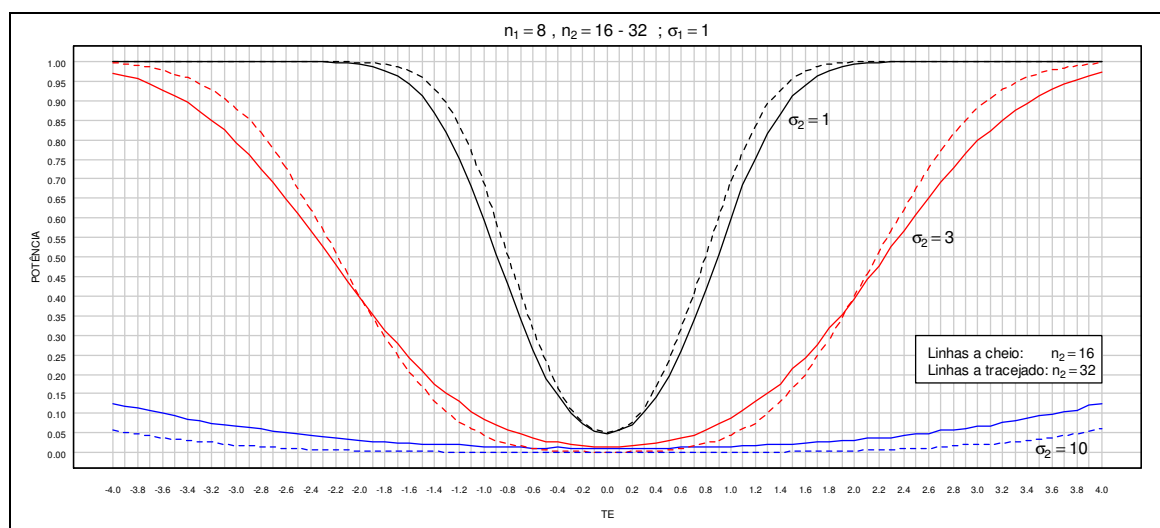


Figura E3.3 Sobreposição dos dois gráficos da Figura E3.2

Quanto aos testes unicaudais, a sua potência, para as várias combinações de tamanhos das amostras e razões de variâncias, segue o padrão descrito para os testes bicaudais.

3.4. Potência dos testes de aleatorização para $n_1 > n_2$

Os resultados da potência do teste de aleatorização bicaudal para as condições em que a maior variância está associada à amostra menor, em comparação com a condição em que as variâncias são iguais, são apresentados na Figura E3.4.

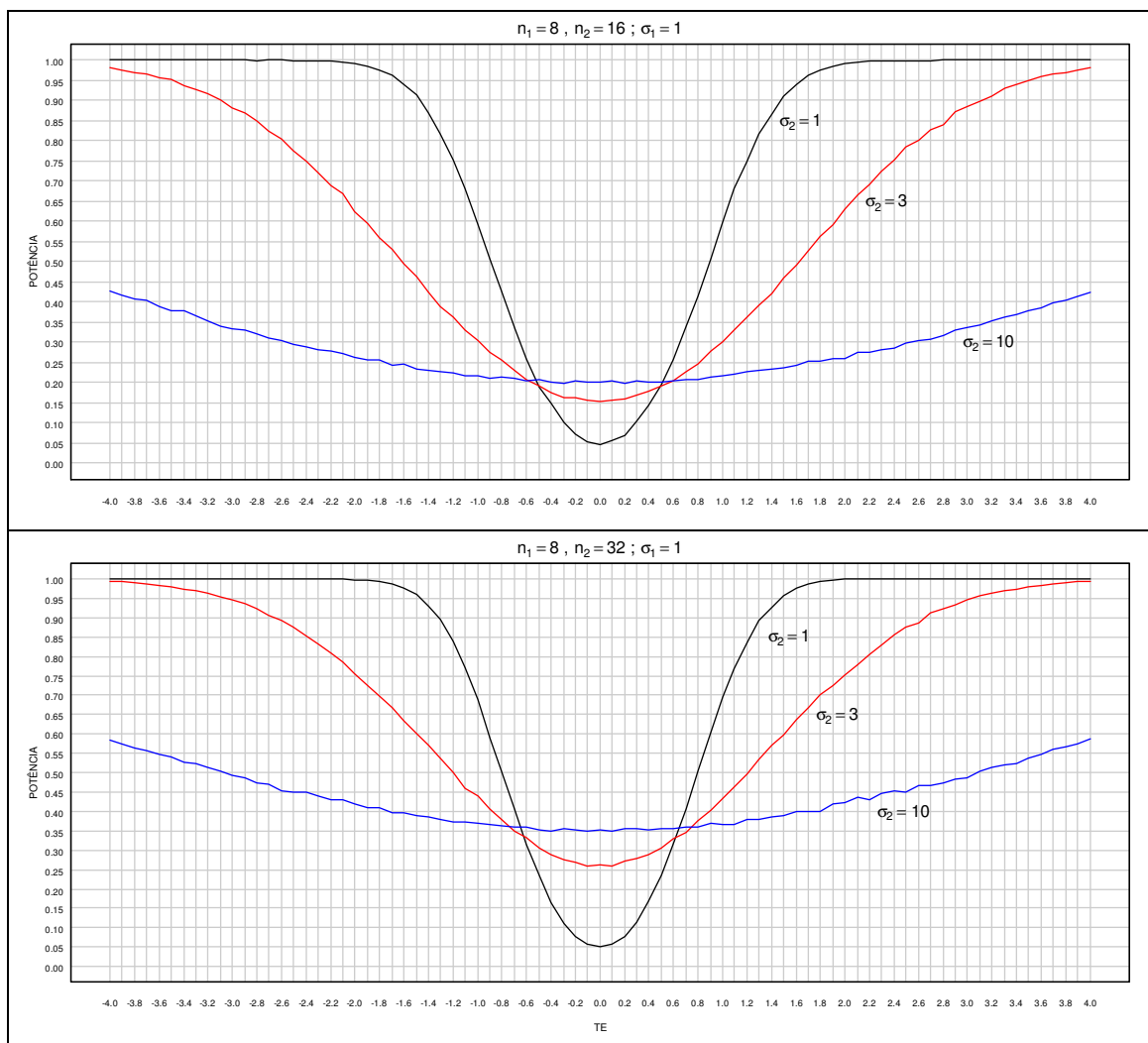


Figura E3.4 Potência do teste de aleatorização bicaudal para as condições em que a maior variância está associada à amostra maior em comparação com a condição em que há homogeneidade de variâncias

Como se pode constatar, nas condições em que a maior variância está associada à amostra maior, a potência do teste de aleatorização é fortemente afectada: em comparação com a condição em que existe

homogeneidade de variâncias, verifica-se um aumento de potência para valores de TE num intervalo aproximadamente de -0.5 a de 0.5; a partir destes valores observam-se perdas de potência, sendo estas tanto maiores quanto mais elevados os valores de TE.

Importa referir que, para os dois níveis de heterogeneidade de variâncias, a potência é sempre superior no caso das amostras mais desequilibradas, como se pode ver na Figura E3.5, em que se sobrepõem os dois gráficos da Figura E3.4.

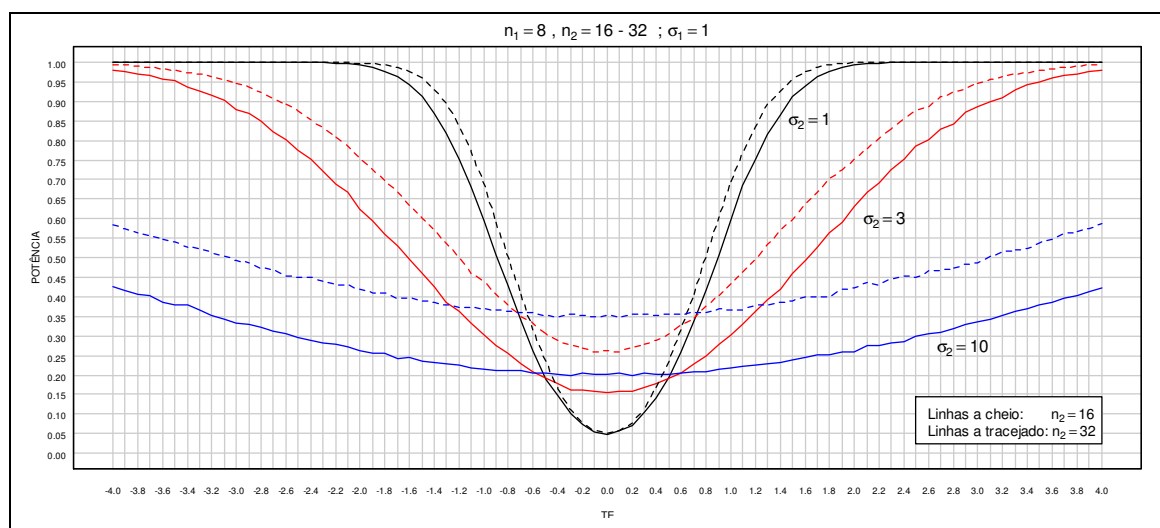


Figura E3.5 Sobreposição dos dois gráficos da Figura E3.4

A potência dos testes unicaudais, para as várias combinações de tamanhos das amostras e razões de variâncias, segue o padrão descrito para os testes bicaudais.

4. Conclusões

Neste estudo foi avaliada a potência, dos testes de aleatorização, na comparação de dois grupos independentes, para distribuições normais, em condições em que existe heterocedasticidade e as amostras não são equilibradas.

Verificamos, em primeiro lugar, que a potência deste teste é tão fortemente afectada pela heterocedasticidade e desequilíbrio das amostras quanto o teste t de Student.

Amostras equilibradas

Mesmo nas condições com amostras equilibradas, a potência do teste é influenciada pela heterogeneidade de variâncias: quanto maior a heterogeneidade, menor a potência, tanto para testes bicaudais, como para os unicaudais.

Assim, estes resultados não permitem afirmar que ‘a heterocedasticidade não é um problema, quando as amostras são equilibradas’, para amostras dentro dos limites dos tamanhos e graus de heterocedasticidade usados neste estudo.

Amostras desequilibradas, com a maior variância associada à amostra maior

Quando a maior variância está associada à amostra maior, observou-se uma diminuição da probabilidade de rejeitar a hipótese nula para todas as diferenças entre médias, incluindo uma diferença de zero, tal como está amplamente descrito na literatura para o teste t de Student e para a Análise de Variância.

Com o aumento da heterogeneidade de variâncias, verificou-se uma diminuição da potência do teste de aleatorização. Os resultados sugerem que o efeito do desequilíbrio das amostras na potência, no sentido de a diminuir, tende a ser maior quanto maior for a heterogeneidade de variâncias.

Amostras desequilibradas, com a maior variância associada à amostra menor

Quando a maior variância está associada à amostra menor, a potência do teste de aleatorização é fortemente afectada: em comparação com a condição em que existe homogeneidade de variâncias, observa-se elevação da potência, mas só para valores de TE num reduzido intervalo: de cerca de -0.5 a cerca de 0.5; fora desse intervalo, pelo contrário, a potência é reduzida. Quanto ao efeito do desequilíbrio das amostras, os resultados mostram que a potência é sempre mais elevada na situação de mais desequilíbrio (mas com número total de sujeitos mais elevado).

Em resumo, dentro dos limites dos valores simulados neste estudo no que respeita à dimensão e equilíbrio das amostras e heterogeneidade de variâncias, a capacidade do teste de aleatorização para detectar diferenças de localização foi fortemente afectada pela heterocedasticidade.

CONCLUSÃO

Os três estudos realizados tinham como finalidade avaliar a potência dos testes de aleatorização, em planejamentos de experiências completamente aleatorizadas, com dois grupos. Com esse objectivo simulámos dados de diferentes distribuições e variámos a dimensão das amostras, o seu equilíbrio, a homogeneidade das variâncias e o tamanho do efeito. A potência foi avaliada para testes bicaudais, unicaudais à esquerda e unicaudais à direita, para $\alpha = 0.05$.

No primeiro estudo, os dados foram simulados de distribuições não-normais. Tendo como referência a potência alcançada pelos testes de aleatorização no caso de distribuições normais, pudemos constatar que, de um modo geral, a influência da não-normalidade sobre a potência foi reduzida. Para o conjunto de distribuições aproximadamente normais (normais com caudas pesadas e alguns valores extremos), todas unimodais, observaram-se, para valores mais baixos do tamanho do efeito, aumentos de potência; para valores mais elevados do tamanho do efeito, perdas de potência. Para o conjunto das distribuições com não-normalidade mais extrema, todas com múltiplas modas, observa-se um padrão inverso: para valores mais baixos do tamanho do efeito, verificam-se perdas, para valores mais elevados, ganhos de potência.

As diferenças de potência, relativamente à distribuição normal, verificaram-se tanto no caso dos testes bicaudais como no caso dos unicaudais, sendo tanto menores quanto maior a dimensão dos grupos. Mas as perdas de potência foram no máximo de 0.05 e, para o conjunto dos 81 tamanhos do efeito, os ganhos superaram as perdas.

No que respeita à comparação do teste de aleatorização com o teste t de Student, em termos de potência, verificou-se que, em geral, apresentam uma potência semelhante, com alguma vantagem para o primeiro.

Os resultados deste estudo, principalmente os ganhos elevados de potência verificados com uma distribuição simétrica, com forte curtose, e, secundariamente, os ganhos, ainda que menos elevados, verificados com uma distribuição assimétrica, levaram-nos a realizar o segundo estudo, em que manipulámos a assimetria e a curtose das distribuições.

No segundo estudo pudemos constatar que a potência dos testes de aleatorização para as distribuições simuladas, com diferentes graus de assimetria, de curtose e de assimetria e curtose, foi superior, em geral, à sua potência no caso da distribuição normal. Os ganhos de potência foram tanto maiores quanto maior a assimetria, quanto maior a curtose e quanto maior a interacção da assimetria e curtose.

Mas os ganhos de potência observaram-se apenas para um intervalo relativamente pequeno de valores do tamanho do efeito, embora nesse intervalo estejam contidos os valores que, segundo Cohen, são os mais comuns nos estudos experimentais nas ciências do comportamento. Importa ainda referir que esses ganhos estão inversamente relacionados com o tamanho das amostras.

Também neste caso, ao comparar o teste de aleatorização com o teste t de Student, em termos de potência, verificámos que, em geral, apresentam uma potência semelhante, com alguma vantagem para o primeiro, nomeadamente quando as amostras são pequenas ou quando as distribuições têm um grau elevado de assimetria e curtose.

Nos dois primeiros estudos, os dados simulados provinham de distribuições idênticas e os grupos eram equilibrados. No terceiro estudo foi avaliada a potência dos testes de aleatorização, no caso de distribuições normais, em condições em que existia heterocedasticidade e as amostras não eram equilibradas.

Verificámos, em primeiro lugar, que a potência do teste de aleatorização é tão fortemente afectada pela heterocedasticidade e desequilíbrio das amostras quanto o teste t de Student.

Mesmo nas condições com amostras equilibradas, a potência do teste é influenciada pela heterogeneidade de variâncias: quanto maior a heterogeneidade, menor a potência, tanto para testes bicaudais, como para os unicaudais. Assim, os resultados que obtivemos não permitem afirmar que ‘a heterocedasticidade não é um problema, quando as amostras são equilibradas’.

Quando as amostras são desequilibradas, com a maior variância associada à amostra maior, observou-se uma diminuição da probabilidade de rejeitar a hipótese nula para todas as diferenças entre médias, incluindo uma diferença de zero, tal como está amplamente descrito na literatura para o teste t de Student e para a Análise de Variância.

Quando a maior variância está associada à amostra menor, a potência do teste de aleatorização é fortemente afectada: em comparação com a condição em que existe homogeneidade de variâncias, observa-se elevação da potência, mas só para valores de TE num reduzido intervalo: de cerca de -0.5 a cerca de 0.5; fora desse intervalo, pelo contrário, a potência é reduzida.

Em conclusão, os resultados destes três estudos sugerem que, de um modo geral, quando os grupos são equilibrados, e os dados de cada grupo provêm de distribuições idênticas, ainda que não-normais ou com elevado grau de assimetria e/ou curtose, a potência dos teste de aleatorização não é muito afectada. Contrariamente, quando se verifica a existência de heterocedasticidade e desequilíbrio na dimensão das

amostras, mesmo com distribuições normais, a potência do teste de aleatorização é fortemente afectada.

Para finalizar, gostaríamos de dizer que temos consciência que muito ficou por investigar sobre a potência dos testes de aleatorização, mesmo no âmbito limitado do planeamento de experiências completamente aleatorizadas com dois grupos independentes.

Em futuras investigações, seria importante estudar o comportamento dos testes de aleatorização, em termos de potência, no caso de outras distribuições e dimensão das amostras, nomeadamente com amostras de menor dimensão.

Em particular, gostaríamos de investigar como se comportam os testes de aleatorização na presença de valores extremos ('outliers'), em comparação com outros testes.

Seria importante, também, estudar a potência estatística dos testes de aleatorização no caso de dados medidos com menor precisão: no caso das simulações que efectuámos, os valores gerados são muito 'precisos', não se observando valores repetidos nas nossas amostras. Tal não acontece com os dados recolhidos em investigações nas áreas da Medicina, da Biologia ou da Psicologia, por exemplo, em que a precisão é limitada, havendo repetições nos dados.

REFERÊNCIAS

- Badrianth, S. G., & Chatterjee, S. (1988). On measuring skewness and elongation in common stock return distributions: The case of the market index. *Journal of Business*, *61*, 4, 451-72.
- Badrianth, S. G., & Chatterjee, S. (1991). A data-analytic look at skewness and elongation in common-stock return distributions. *Journal of Business and Economic Statistics*, *9*, 9, 223-33.
- Bahadur, R. R., & Raghavachari, M. (1970). *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability*. Berkeley: University of California Press.
- Barton, D. E., & David, F. N. (1959). Contagious occupancy. *Journal of the Royal Statistical Society, B*, *21*, 120-133.
- Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*, 257-278.
- Becker, R. A., Chambers, J. M., & Wilks, A. R. (1988). *The new S language: A programming environment for data analysis and graphics*. New York: Chapman & Hall. (Monterey: Wadsworth & Brooks/Cole.)
- Blair, R. C., & Higgins, J. J. (1985). Comparison of the power of the paired samples t test to that of the Wilcoxon' signed-rank test under various population shapes. *Psychological Bulletin*, *97*, 119-128.
- Boneau, C. A. (1960). The effects of violation of assumptions underlying the t-test. *Psychological Review*, *69*, 246-256.
- Boos, D. D., & Zhang, J. (2000). Monte Carlo evaluation of resampling-based hypothesis tests. *Journal of the American Statistical Association*, *95*, 486-492.

- Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika*, 40, 318-335.
- Box, G. E. P., & Anderson, S. L. (1955). Permutation theory in the derivation and the study of departures from assumption. *Journal of the Royal Statistical Society, B*, 17, 1-34.
- Bradbury, I. (1987). Analysis of variance versus randomization tests: a comparison. *British Journal of Mathematics and Statistical Psychology*, 40, 177-187.
- Bradley, J. V. (1977). A common situation conducive to bizarre distribution shapes. *The American Statistician*, 31, 147-150.
- Brillinger, D. R., Jones, L. V., & Tukey, J. W. (1978). *The management of weather resources. Volume II. The role of statistics in weather resources management*. Report of the Statistical Task Force to the Weather Modification Board, Department of Commerce, Washington, DC.
- Carlson, K. D., & Schmidt, F. L. (1999). Impact of experimental design on effect size: Findings from the research literature on training. *Journal of Applied Psychology*, 84, 851-862.
- Chambers, J. M. & Hastie, T. J. (Eds) (1992). *Statistical models in S*. New York: Chapman & Hall. (Monterey: Wadsworth and Brooks/ Cole.).
- Chen, S. Y., & Chen, H. J. (1998). Single-stage analysis of variance under heterocedasticity. *Communications in Statistics - Simulation and Computation*, 27, 641-666.

- Chung, J. H. & Fraser, D. A. S. (1958). Randomization tests for a multivariate two sample problem. *Journal of the American Statistical Association*, 53, 729-735.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, 65, 145-153.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cotton, J. W. (1967). *Elementary statistical theory for behavior scientists*. Reading, MA: Addison-Wesley.
- Cowles, M. (1989). *Statistics in psychology: An historical perspective*. Hillsdale, NJ: Lawrence Erlbaum.
- Dutta, K., & Perry, J. (2006). *A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital*. Federal Reserve Bank of Boston, Working Paper No 06-13.
- Dutta, K. K., & Babbel, D. F. (2002). *On measuring skewness and kurtosis in short rate distributions: The case of the Us Dollar London Inter Bank Offer Rates*. Wharton Financial Institutions Center Working Paper.

- Dwass, M. (1957). Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics*, 28, 181-187.
- Edgington, E. S. (1964). Randomization tests. *Journal of Psychology*, 57, 445-449.
- Edgington, E. S. (1966). Statistical inference and nonrandom samples. *Psychological Bulletin*, 66, 485-487.
- Edgington, E. S. (1969a). Approximate randomization tests. *Journal of Psychology*, 72, 143-149.
- Edgington, E. S. (1969b). *Statistical inference: The distribution-free approach*. New York: McGraw-Hill.
- Edgington, E. S. (1980). *Randomization tests*. New York: Marcel Dekker.
- Edgington, E. S. (1987). *Randomization tests* (2nd ed.). New York: Marcel Dekker.
- Edgington, E. S. (1995). *Randomization tests* (3rd ed.). New York: Marcel Dekker.
- Edgington, E. S. & Onghena, P. (2007). *Randomization tests* (4th ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Field, C. A., & Genton, M. G. (2006). The multivariate *g*-and-*h* distribution, *Technometrics*, 48, 104-111.

- Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.
- Fisher, R. A. (1936). "The coefficient of racial likeness" and the future of craniometry. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland*, 66, 57-63.
- Fisher, R. A. (1966). *The design of experiments* (8th ed.). Edinburgh: Oliver & Boyd. [Primeira edição: 1935]
- Foutz, R. V. (1980). A method for constructing exact tests from tests statistics that have unknown null distributions. *Journal of Statistics and Computer Simulation*, 10, 187-193.
- Gabriel, K. R. & Hall, W. J. (1983). Rerandomization inference on regression and shift effects: Computationally feasible methods. *Journal of the American Statistical Association*, 78, 384, 827-836.
- Games, J., & Howell, J. F. (1976). Pairwise multiple comparisons procedures with unequal n's and/or variances: a Monte Carlo study. *Journal of Educational Statistics*, 1, 113-125.
- Gigerenzer, G. (1987). Probabilistic thinking and the fight against subjectivity. In L. Krüger, G. Gigerenzer & M. Morgan (Eds.), *The probabilistic revolution*, vol. 2: *Ideas in the sciences* (pp. 11-33). Cambridge, MA: MIT Press.
- Gigerenzer, G. (1993). The Superego, the Ego and the Id in statistical reasoning. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 311-339). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance: How probability changed science and everyday life*. Cambridge: University Press.
- Glass, G. V (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3-8.
- Glass, G.V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5, 351-379
- Glass, G., Peckham, P., & Sanders, J. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, 42, 237-288.
- Gossett, W. (1908). The probable error of the mean. *Biometrika*, 6, 1-25.
- Hamilton, M. A., & Collings, B. J. (1991). Determining the appropriate sample size for nonparametric tests for location shift. *Technometrics*, 3, 327-337.
- Hammersley, J. M., & Handscomb, D. C. (1964). *Monte Carlo methods*. London: Chapman & Hall.
- Hays, W. L. (1994). *Statistics* (5th ed.). Belmont, CA: Wadsworth.
- Headrick, T. C., Kowalchuk, R. K., & Sheng, Y. (2008). Parametric probability densities and distribution functions for Tukey *g*-and-*h* transformations and their use for fitting data. *Applied Mathematical Sciences*, 2, 9, 449-462.

- Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Hoaglin, D. C. (1983). *g*-and-*h* distributions. In S. Kotz and N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences, Vol. 3* (pp. 298-301). New York: Wiley.
- Hoaglin, D. C. (1985). Summarizing shape numerically: The *g*-and-*h* distributions. In D. C. Hoaglin, F. Mosteller & J. W. Tukey (eds.), *Exploring data tables, trend, and shapes* (pp. 417-513). New York: Wiley.
- Hoeffding, W. (1951). A combinatorial central limit theorem. *Annals of Mathematical Statistics*, 22, 4, 558-566.
- Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. *Annals of Mathematical Statistics*, 23, 2, 169-192.
- Hope, A. C. A. (1968). A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistic Society, B*, 30, 582-598.
- Ihaka, R. & Gentleman, R. (1996) R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5, 299–314.
- Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate discrete distributions* (3th ed.). Hoboken, NJ: Wiley.
- Johnson, N. L., Kotz S., & Balakrishnan, N. (1995). *Continuous univariate distributions*, vol. 2 (2nd ed.). New York: Wiley.
- Kempthorne, O. (1952). *The design and analysis of experiments*. New York: Wiley.

- Kempthorne, O. (1955). The randomization theory of experimental inference. *Journal of the American Statistical Association*, 50, 964-967.
- Kempthorne, O., & Doerfler, T. E. (1969). The behavior of some significance tests under experimental randomization. *Biometrika*, 56, 231-247.
- Keppel, G. (1973). *Design and analysis: A researcher handbook*. Englewood Cliffs, NJ: Prentice-Hall.
- Kirk, R. E. (1968). *Introductory statistics*. Belmont, CA: Brooks/Cole.
- Kirk, R. E. (1999). *Statistics: An introduction* (4th ed.). Fort Worth: Harcourt Brace.
- Kohr, R. L., & Games, P. A. (1977). Testing complex a priori contrasts in means for independent samples. *Journal of Educational Statistics*, 1, 207-216.
- Kraemer, H. C., & Thiemann, S. (1987). *How many subjects?* Newbury Park, CA: Sage.
- Lehmann, E. L. (1975). *Nonparametrics, statistical methods based on ranks*. San Francisco, CA: Holden-Day.
- Lehmann, E. L. (1986). *Testing statistical hypothesis* (2nd ed.). New York: Wiley.
- Lehmann, E. L., & Romano, J. P. (2005). *Testing statistical hypotheses* (3rd ed.). New York: Springer.
- Lindman, H. R. (1974). *Analysis of variance in complex experimental designs*. San Francisco: W. H. Freeman.

- Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.
- Lunneborg, C. E. (2000). *Data analysis by resampling: Concepts and applications*. Pacific Grove, CA: Duxbury.
- Mächler, M. (2007). *nor1mix: Normal (1-d) Mixture Models (S3 Classes and Methods)*. R package, version 1.0-7.
- Manly, B. F. J. (2007). *Randomization, bootstrap and Monte Carlo methods in Biology* (3rd ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Maritz, J. S. (1995). *Distribution-free statistical methods* (2nd. ed.). London: Chapman & Hall.
- Marron, J. S., & Wand, M. P. (1992). Exact mean integrated squared error. *Annals of Statistics*, 20, 712–736.
- Marsaglia, G. (1996). DIEHARD: A battery of tests of randomness. <http://stat.fsu.edu/~geo/diehard.html>.
- Martinez, J. (1981). *Some properties of robust scale estimators*. Ph.D. dissertation, Temple University, Philadelphia, PA.
- Martinez, J., & Iglewicz, B. (1984). Some properties of the Tukey g and h family of distributions. *Communications in Statistics – Theory and Methods*, 13, 3, 353-69.

- Matsumoto, M., & Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation*, 8, 1, 3-30.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105, 156-166.
- Mills, T. C. (1995). Modelling skewness and kurtosis in the London Stock Exchange FT-SE Index Return Distributions. *The Statistician*, 44, 3, 323-32.
- Milton, R. C. (1970). *Rank order probabilities: Two-sample normal shift alternatives*. New York: Wiley.
- Mood A.M., Graybill F.A., & Boes D.C. (1974). *Introduction to the theory of statistics* (3rd ed.). New-York: McGraw-Hill.
- Mosteller, F., & Tukey, J.W. (1968). Data analysis, including statistics. In G. Lindzey & E. Aronson (Eds.), *The handbook of Social Psychology: Research methods* (pp. 80-203). Reading, MA: Addison-Wesley.
- Murphy, K. R., & Myers, B. (1998). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests*. Mahwah, NJ: Lawrence Erlbaum.
- Neyman, J., & Pearson, E. (1928a). On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. *Biometrika*, 20A, 175-240.
- Neyman, J., & Pearson, E. (1928b). On the use and interpretation of certain test criteria for purposes of statistical inference: Part II. *Biometrika*, 20A, 263-294.

- Neyman, J., & Pearson, E. (1933). On the problem of the most efficient tests of statistical hypothesis. *Philosophical Transactions of the Royal Society of London, Series A*, 231, 289-337.
- Noreen, E. (1986). *On the validity of randomization tests*. Working paper, School of Business, University of Washington.
- Noreen, E. W. (1989). *Computer-intensive methods for testing hypotheses: An introduction*. New York: Wiley.
- Oakes, M. (1990). *Statistical Inference*. Chestnut Hill, MA: Epidemiology Resources Inc.
- Oden, N. L. (1991). Allocations of effort in Monte Carlo simulation for power of permutations tests. *Journal of the American Statistical Association*, 86, 1074-1076.
- Onghena, P. (1994). *The power of randomization tests for single-case designs*. Unpublished Ph.D. thesis, Faculteit der Psychologie en Pedagogische Wetenschappen, Katholieke Universiteit Leuven.
- Pearson, E. S. (1930). A further development of tests of normality. *Biometrika*, 22, 239-249.
- Pearson, E. S. (1931). The analysis of variance in cases of non-normal variation. *Biometrika*, 23, 114-133.
- Pesarin, F. (2001). *Multivariate permutation tests with applications in Biostatistics*. New York: Wiley.

- Pitman, E. J. G. (1937a). Significance tests which may be applied to samples from any population. *Journal of the Royal Statistical Society, B, 4*, 119-130.
- Pitman, E. J. G. (1937b). Significance tests which may be applied to samples from any population. II. The correlation coefficient. *Journal of the Royal Statistical Society, B, 4*, 225-232.
- Pitman, E. J. G. (1938). Significance tests which may be applied to samples from any population. III. The analysis of variance test. *Biometrika, 29*, 322-335.
- Prentice, D. A. & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin, 112*, 160-164.
- R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Randles, R. H. (1980). Nonparametric statistical tests of hypotheses. In R. V. Hogg (Ed.), *Modern statistics: Methods and applications* (pp. 31-40). Providence, RI: American Mathematical Society.
- Rogan, J. C., & Keselman, H. J. (1977). Is the ANOVA F-test robust to variance heterogeneity when sample sizes are equal?: An investigation via a coefficient of variation. *American Educational Research Journal, 14*, 493-498.

- Romano, J. P. (1989). Bootstrap and randomization tests of some nonparametric hypothesis. *Annals of Statistics*, *17*, 141-159.
- Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist*, *45*, 775-777.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-260). New York: Russell Sage Foundation.
- Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*, 1276-1284.
- Routledge, R. D. (1997). P values from permutation and F tests. *Computational Statistics and Data Analysis*, *24*, 379-386.
- Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, *105*, 309-316.
- Sheffe, H. (1959). *The analysis of variance*. New York: Wiley.
- Siegel, S., & Castellan, N. J., Jr. (1988). *Nonparametric statistics for the behavioural sciences* (2nd ed.). New York: McGraw-Hill.
- Spence, J. T., Cotton, J. W., Underwood, B. J., & Duncan, C. P. (1976). *Elementary statistics* (3th ed.). New York: Appleton-Century-Crofts.
- Stephens, M. A. (1986). Tests based on EDF statistics. In: R.B. D'Agostino & M. A. Stephens (Eds.), *Goodness-of-Fit Techniques*. New York: Marcel Dekker.

- Stigler, S. M. (1973). Simon Newcomb, Percy Daniell, and the history of robust estimation 1885-1920. *Journal of the American Statistical Association*, 68, 872-879.
- Stigler, S. M. (1977). Do robust estimators work with real data? *The Annals of Statistics*, 5, 1055-1098.
- Thode, H.C. (2002). *Testing for Normality*. New York: Marcel Dekker.
- Tukey, J. W. (1977). *Modern techniques in data analysis*. NSF-sponsored regional research conference at Southeastern Massachusetts University, North Dartmouth, MA.
- Welch, B. L. (1937). On the z-test in randomized blocks and Latin squares. *Biometrika*, 29, 21-52.
- Welch, B. L. (1938). The significance of the differences between two means when the population variations are unequal. *Biometrika*, 29, 350-362.
- Westfall, P. H., & Young, S. S. (1993). *Resampling-based multiple testing*. New York: Wiley.
- Wichura, M. J. (1988). Algorithm AS 241: The Percentage Points of the Normal Distribution. *Applied Statistics*, 37, 477-484.
- Wilcox, R. R. (1995a). ANOVA: A paradigm for low power and misleading measures of effect size? *Review of Educational Research*, 65, 51-77.
- Wilcox, R. R. (1995b). ANOVA: The practical importance of heteroscedastic methods, using trimmed means versus means, and designing simulation studies. *British Journal of Mathematical and Statistical Psychology*, 48, 99-114.

- Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.
- Zimmerman, D. W. (2006). Two separate effects of variance heterogeneity on the validity and power of significance tests of location. *Statistical Methodology*, 3, 4, 351-374.

ANEXOS

Anexo I

Programas em R para realizar a Figura 2.1

1. Cálculo da função potência

2. Representação gráfica

1. Programa para calcular a função potência para o teste t de Student de diferença de médias (unicaudal à esquerda, unicaudal à direita e bicaudal), para $n_1 = n_2 = 8$, $\alpha = 0.05$ e TE de -4 a 4, com intervalos de 0.1.

```
te<-matrix(seq(-4,4,by=0.1),81,1)
pot<-matrix(0,81,3)
for(i in 1:81){a<-power.t.test(n = 8, delta =
te[i],sd = 1,
  sig.level = 0.05,power = NULL,type =
"two.sample",alternative = "one.sided")
  pot[82-i,1]<-a[[5]]
  a<-power.t.test(n = 8, delta = te[i], sd = 1,
  sig.level = 0.05,power = NULL,type =
"two.sample",alternative = "one.sided")
  pot[i,2]<-a[[5]]
  a<-power.t.test(n = 8, delta = te[i], sd = 1,
  sig.level = 0.05,power = NULL,type =
"two.sample",alternative = "two.sided", strict =T)
  pot[i,3]<-a[[5]]}
fp<-cbind(te,pot)
```

2. Programa para realizar a representação gráfica:

```
te<-matrix(seq(-4,4,by=.1),81,1)
win.metafile("Rplot%02d.wmf", width = 9, height =
4,pointsize = 8)
vf<-matrix(
c(rep(0,times=40),rep(1,times=41)),81,1)
par(mar=c(3,3,2,1), mgp=c(2, 0.5,0))
plot(te,vf,type="l", col="white", axes=F ,
xlab="TE",ylab="POTÊNCIA",
cex.axis=.5,xpd=NA,cex.lab=.8)
axis(side=1,labels = T, at = seq(-4,4,.1),tck = 1,
lty=1, las=1, col="gray80", cex.axis=.6)
axis(yaxp=c(0, 1, 20),side=2,labels = T, tck = 1,
lty=1, las=1, col="gray80", cex.axis=.6)
box()
```

```
lines(res[,1],res[,2],col="red")
lines(res[,1],res[,3],col="blue")
lines (res[,1],res[,4],col="black")
box()
legend(x= "bottomright",c("Teste unicaudal à
esquerda",
"          unicaudal à direita", "
bicaudal"), inset=.05, cex=1,col =
c("red", "blue", "black"),lty=c(1,1,1),bty="o",bg="wh
ite",y.intersp=.8)
dev.off()
```

Nota: As Figuras 2.2 e 2.3 foram realizadas com programas semelhantes a este, com as alterações necessárias nos parâmetros.

Anexo II

Programa em R para comparar a potência do teste de aleatorização: Conjunto de referência exaustivo vs. conjunto não-exaustivo

```

set.seed(12345)
#Função: Teste de Aleatorização (total de
combinações)
comb<-function(x,y){
  xx<-c(x,y)
  etb1<-NA;etb2<-NA
  ncomb<-nCm(length(xx),length(x))
  etb1<-combn(xx,length(x),fun=sum)
  etb2<-(etb1^2/length(x))+(sum(xx)-
etb1)^2/length(y)
  sigesq<-(length(etb1[etb1<=etb1[1]]))/ncomb
  sigdir<-(length(etb1[etb1>=etb1[1]]))/ncomb
  sigbi<-(length(etb2[etb2>=etb2[1]]))/ncomb
  sigta<-c(sigesq,sigdir,sigbi)
  sigta}
#Função: Teste de Aleatorização (amostra de
combinações)
permg<-function(x,y,nream=999){
xx<- c(x,y)
etobs1<-sum(x) ; etobs2 <-
(sum(x)^2/length(x))+(sum(y)^2/length(y))
etb1 <- NA; etb2 <- NA
for(i in 1:nream){
  etb1[i]<- sum(sample(xx, size = length(x),
replace = F))
  etb2[i]<- etb1[i]^2/length(x)+(sum(xx)-
etb1[i])^2/length(y)}
  vpesq <- (1+length(etb1[etb1 <=
etobs1]))/(nream+1)
  vpdire <- (1+length(etb1[etb1 >=
etobs1]))/(nream+1)
  vpbire <- (1+length(etb2[etb2 >=
etobs2]))/(nream+1)
  sigtasample<-c(vpesq,vpdire,vpbire)}
nsim<-20000
ngr1<-8;ngr2<-8
alfa<-0.05
te<-1
dadgr1<-matrix(rnorm(nsim*ngr1,te,1),nsim,ngr1)
dadgr2<-matrix(rnorm(nsim*ngr2,0,1),nsim,ngr2)
sigta<-matrix(0,nsim,3)

```

```
sigtasample<-matrix(0,nsim,3)
pta<-matrix(0,1,3); ptas<-matrix(0,1,3)
for(i in 1:nsim) {
  sigta[i,]<-comb(dadgr1[i,],dadgr2[i,])
  sigtasample[i,]<-permg(dadgr1[i,],dadgr2[i,])
  if(sigta[i,1] <=alfa) pta[1,1]= pta[1,1] +1
  if(sigta[i,2] <=alfa) pta[1,2]= pta[1,2] +1
  if(sigta[i,3] <=alfa) pta[1,3]= pta[1,3] +1
  if(sigtasample[i,1] <=alfa) ptas[1,1]=
ptas[1,1] +1
  if(sigtasample[i,2] <=alfa) ptas[1,2]=
ptas[1,2] +1
  if(sigtasample[i,3] <=alfa) ptas[1,3]=
ptas[1,3] +1}
pta<-pta/nsim;ptas<-ptas/nsim
matpot<-cbind(pta,ptas)
colnames(matpot)<-c("TA_esq", "TA_dir", "TA_bi",
"TA_s_esq", "TA_s_dir", "TA_s_bi")
matpot
```

Anexo III**Matrizes de resultados da comparação da potência do teste de aleatorização:
Conjunto de referência exaustivo vs. conjunto de referência não-exaustivo.**

Abreviaturas usadas:

TE	Tamanho do efeito
Uni_esq	Teste de aleatorização unicaudal à esquerda
Uni_dir	Teste de aleatorização unicaudal à direita
Bi	Teste de aleatorização bicaudal

Cinco conjuntos de 20 000 amostras:

TE	Conjunto de referência						Diferença (Exaustivo - Não-exaustivo)		
	Exaustivo			Não-exaustivo			Uni_esq	Uni_dir	Bi
	Uni_esq	Uni_dir	Bi	Uni_esq	Uni_dir	Bi			
-4.00	1.0000	0.0000	1.0000	1.0000	0.0000	1.0000	0.0000	0.0000	0.0000
-3.00	1.0000	0.0000	0.9998	0.9999	0.0000	0.9998	0.0000	0.0000	0.0000
-2.00	0.9848	0.0000	0.9610	0.9843	0.0000	0.9601	0.0005	0.0000	0.0009
-1.00	0.6011	0.0000	0.4601	0.5976	0.0000	0.4595	0.0035	0.0000	0.0006
-0.75	0.4131	0.0008	0.2869	0.4134	0.0009	0.2864	-0.0003	-0.0001	0.0005
-0.50	0.2453	0.0041	0.1548	0.2437	0.0042	0.1539	0.0016	0.0000	0.0009
-0.25	0.1214	0.0179	0.0765	0.1228	0.0175	0.0773	-0.0014	0.0004	-0.0008
0.00	0.0524	0.0501	0.0513	0.0524	0.0496	0.0514	0.0001	0.0005	-0.0001
0.25	0.0180	0.1222	0.0772	0.0177	0.1214	0.0764	0.0004	0.0009	0.0009
0.50	0.0051	0.2426	0.1550	0.0053	0.2405	0.1528	-0.0002	0.0022	0.0022
0.75	0.0011	0.4129	0.2891	0.0011	0.4112	0.2881	0.0000	0.0017	0.0010
1.00	0.0004	0.5995	0.4616	0.0004	0.5962	0.4591	0.0001	0.0033	0.0025
2.00	0.0000	0.9831	0.9580	0.0000	0.9832	0.9575	0.0000	-0.0001	0.0005
3.00	0.0000	1.0000	0.9998	0.0000	1.0000	0.9998	0.0000	0.0000	0.0000
4.00	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000	0.0000	0.0000

TE	Conjunto de referência						Diferença (Exaustivo - Não-exaustivo)		
	Exaustivo			Não-exaustivo			Uni_esq	Uni_dir	Bi
	Uni_esq	Uni_dir	Bi	Uni_esq	Uni_dir	Bi			
-4.00	1.0000	0.0000	1.0000	1.0000	0.0000	1.0000	0.0000	0.0000	0.0000
-3.00	1.0000	0.0000	1.0000	1.0000	0.0000	1.0000	0.0000	0.0000	0.0000
-2.00	0.9845	0.0000	0.9619	0.9845	0.0000	0.9617	0.0000	0.0000	0.0001
-1.00	0.6023	0.0003	0.4608	0.6027	0.0003	0.4611	-0.0005	0.0000	-0.0003
-0.75	0.4165	0.0011	0.2880	0.4149	0.0010	0.2878	0.0016	0.0001	0.0002
-0.50	0.2452	0.0041	0.1542	0.2452	0.0043	0.1539	0.0000	-0.0003	0.0003
-0.25	0.1220	0.0165	0.0727	0.1220	0.0166	0.0731	0.0000	-0.0001	-0.0004
0.00	0.0488	0.0486	0.0480	0.0484	0.0493	0.0482	0.0004	-0.0007	-0.0001
0.25	0.0164	0.1183	0.0737	0.0166	0.1172	0.0738	-0.0003	0.0011	-0.0002
0.50	0.0047	0.2367	0.1503	0.0050	0.2362	0.1499	-0.0003	0.0005	0.0004
0.75	0.0010	0.4048	0.2781	0.0010	0.4038	0.2792	0.0000	0.0010	-0.0011
1.00	0.0001	0.5956	0.4519	0.0002	0.5951	0.4507	-0.0001	0.0005	0.0013
2.00	0.0000	0.9841	0.9606	0.0000	0.9835	0.9601	0.0000	0.0006	0.0005
3.00	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000	0.0000	0.0000
4.00	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000	0.0000	0.0000

TE	Conjunto de referência						Diferença (Exaustivo - Não-exaustivo)		
	Exaustivo			Não-exaustivo			Uni_esq	Uni_dir	Bi
	Uni_esq	Uni_dir	Bi	Uni_esq	Uni_dir	Bi			
-4.00	1.0000	0.0000	1.0000	1.0000	0.0000	1.0000	0.0000	0.0000	0.0000
-3.00	1.0000	0.0000	0.9998	1.0000	0.0000	0.9999	0.0000	0.0000	0.0000
-2.00	0.9840	0.0000	0.9587	0.9835	0.0000	0.9581	0.0005	0.0000	0.0006
-1.00	0.5972	0.0004	0.4614	0.5946	0.0003	0.4593	0.0025	0.0001	0.0021
-0.75	0.4127	0.0012	0.2884	0.4110	0.0011	0.2874	0.0018	0.0000	0.0010
-0.50	0.2425	0.0048	0.1538	0.2414	0.0053	0.1540	0.0011	-0.0005	-0.0003
-0.25	0.1216	0.0182	0.0733	0.1225	0.0183	0.0730	-0.0009	0.0000	0.0003
0.00	0.0482	0.0512	0.0503	0.0478	0.0519	0.0506	0.0004	-0.0007	-0.0003
0.25	0.0154	0.1211	0.0746	0.0157	0.1203	0.0738	-0.0003	0.0008	0.0009
0.50	0.0039	0.2433	0.1505	0.0039	0.2429	0.1510	0.0000	0.0004	-0.0005
0.75	0.0008	0.4135	0.2867	0.0007	0.4132	0.2864	0.0001	0.0003	0.0004
1.00	0.0001	0.6021	0.4604	0.0001	0.5995	0.4596	0.0000	0.0026	0.0008
2.00	0.0000	0.9845	0.9609	0.0000	0.9845	0.9599	0.0000	0.0000	0.0010
3.00	0.0000	1.0000	0.9999	0.0000	1.0000	1.0000	0.0000	0.0000	0.0000
4.00	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000	0.0000	0.0000

TE	Conjunto de referência						Diferença (Exaustivo - Não-exaustivo)		
	Exaustivo			Não-exaustivo			Uni_esq	Uni_dir	Bi
	Uni_esq	Uni_dir	Bi	Uni_esq	Uni_dir	Bi			
-4.00	1.0000	0.0000	1.0000	1.0000	0.0000	1.0000	0.0000	0.0000	0.0000
-3.00	1.0000	0.0000	0.9999	1.0000	0.0000	0.9998	0.0000	0.0000	0.0001
-2.00	0.9839	0.0000	0.9581	0.9828	0.0000	0.9563	0.0011	0.0000	0.0018
-1.00	0.6023	0.0000	0.4603	0.5985	0.0000	0.4589	0.0038	0.0000	0.0014
-0.75	0.4097	0.0012	0.2892	0.4095	0.0012	0.2883	0.0002	0.0000	0.0010
-0.50	0.2451	0.0052	0.1550	0.2460	0.0052	0.1533	-0.0009	0.0000	0.0017
-0.25	0.1215	0.0168	0.0783	0.1218	0.0168	0.0773	-0.0003	0.0000	0.0010
0.00	0.0521	0.0507	0.0508	0.0521	0.0508	0.0507	-0.0001	-0.0001	0.0001
0.25	0.0186	0.1245	0.0783	0.0182	0.1247	0.0783	0.0004	-0.0002	0.0000
0.50	0.0048	0.2480	0.1578	0.0050	0.2476	0.1572	-0.0002	0.0004	0.0005
0.75	0.0014	0.4161	0.2915	0.0014	0.4153	0.2896	0.0000	0.0009	0.0019
1.00	0.0004	0.6007	0.4639	0.0003	0.5982	0.4617	0.0001	0.0025	0.0022
2.00	0.0000	0.9826	0.9572	0.0000	0.9822	0.9562	0.0000	0.0004	0.0010
3.00	0.0000	1.0000	0.9997	0.0000	1.0000	0.9997	0.0000	0.0000	0.0000
4.00	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000	0.0000	0.0000

TE	Conjunto de referência						Diferença (Exaustivo - Não-exaustivo)		
	Exaustivo			Não-exaustivo			Uni_esq	Uni_dir	Bi
	Uni_esq	Uni_dir	Bi	Uni_esq	Uni_dir	Bi			
-4.00	1.0000	0.0000	1.0000	1.0000	0.0000	1.0000	0.0000	0.0000	0.0000
-3.00	1.0000	0.0000	0.9999	1.0000	0.0000	0.9999	0.0000	0.0000	0.0000
-2.00	0.9849	0.0000	0.9611	0.9848	0.0000	0.9603	0.0001	0.0000	0.0008
-1.00	0.6066	0.0003	0.4627	0.6036	0.0003	0.4623	0.0030	0.0000	0.0005
-0.75	0.4155	0.0013	0.2864	0.4140	0.0013	0.2856	0.0015	0.0000	0.0007
-0.50	0.2443	0.0051	0.1566	0.2432	0.0049	0.1559	0.0011	0.0003	0.0006
-0.25	0.1194	0.0155	0.0754	0.1187	0.0154	0.0751	0.0008	0.0001	0.0003
0.00	0.0496	0.0480	0.0481	0.0496	0.0479	0.0475	-0.0001	0.0001	0.0006
0.25	0.0176	0.1202	0.0746	0.0176	0.1200	0.0745	-0.0001	0.0002	0.0001
0.50	0.0049	0.2355	0.1505	0.0053	0.2352	0.1501	-0.0004	0.0003	0.0004
0.75	0.0008	0.4104	0.2806	0.0007	0.4098	0.2784	0.0001	0.0006	0.0022
1.00	0.0001	0.5989	0.4593	0.0001	0.5967	0.4573	0.0000	0.0022	0.0020
2.00	0.0000	0.9831	0.9609	0.0000	0.9822	0.9596	0.0000	0.0009	0.0012
3.00	0.0000	1.0000	0.9998	0.0000	1.0000	0.9998	0.0000	0.0000	0.0000
4.00	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000	0.0000	0.0000

Anexo IV

Programas em R para o cálculo de potência do Estudo 1

```

# MW.nm15
#Definições
nsim <- 20000
ngr1 <- 32; ngr2 <- 32
alfa<-0.05
te<- seq(-4,4,by=.10)
set.seed(12345)
#Função: Teste t de Student
teste.t <- function(x1, x2) {
  n1 <- length(x1); n2 <- length(x2)
  mx1 <- mean(x1); mx2 <- mean(x2)
  s1 <- var(x1); s2 <- var(x2)
  s <- ((n1-1)*s1 + (n2-1)*s2)/(n1+n2-2)
  et <- (mx1 - mx2)/sqrt(s*(1/n1 + 1/n2))
  et
}
#Função: Teste de Aleatorização (sample)
permg<-function(x,y,nream=999){
xx<- c(x,y)
etobs1<-sum(x) ; etobs2 <-
(sum(x)^2/length(x))+(sum(y)^2/length(y))
etb1 <- NA
etb2 <- NA
for(i in 1:nream){
etb1[i]<- sum(sample(xx, size = length(x), replace = F))
etb2[i]<- etb1[i]^2/length(x)+(sum(xx)-
etb1[i])^2/length(y)
}
vpesq <- (1+length(etb1[etb1 <= etobs1]))/(nream+1)
vpdir <- (1+length(etb1[etb1 >= etobs1]))/(nream+1)
vpbi <- (1+length(etb2[etb2 >= etobs2]))/(nream+1)
res<-matrix(c(vpesq,vpdir,vpbi),1,3)
}
res.tt<-matrix(0,length(te),3); res.ta<-matrix(0,
length(te),3)
matsig <- matrix(0,nsim,3)
for(k in 1 : length(te)){
dadgr1 <- matrix(rnorMix(nsim*ngr1,
MW.nm15)+(te[k]*1.6909791),nsim,ngr1)
dadgr2 <- matrix(rnorMix(nsim*ngr1, MW.nm15),nsim,ngr2)
ncesq <- 0; ncdir <- 0; nbi <- 0
et<- matrix(0,nsim,1)
for(i in 1:nsim) {et[i,1] <- teste.t(dadgr1[i,],
dadgr2[i,])
}
}

```

```

        if(et[i,1]<= qt(alfa, (ngr1+ngr2-2)))
ncesq=ncesq+1
        if(et[i,1]>= qt(1-alfa, (ngr1+ngr2-2)))
ncdir=ncdir+1
        if(et[i,1]<= qt(alfa/2, (ngr1+ngr2-2))|et[i,1]>=
qt(1-alfa/2, (ngr1+ngr2-2))) nbi=nbi+1
        matsig[i,]<-permg(dadgr1[i,],dadgr2[i,])
    }
res.tt[k,1]<-ncesq/nsim; res.tt[k,2]<-ncdir/nsim;
res.tt[k,3]<-nbi/nsim
vecsigesq<-matsig[,1]
vecsigdir<-matsig[,2]
vecsigbi<-matsig[,3]
res.ta[k,1]<-length(vecsigesq[vecsigesq<=alfa])/nsim
res.ta[k,2]<-length(vecsigdir[vecsigdir<=alfa])/nsim
res.ta[k,3]<-length(vecsigbi[vecsigbi<=alfa])/nsim
}
sink("ResMW15_32+32_20000_999", append=F)
cat(res.tt)
cat(res.ta)
sink()

```

Nota: Este programa foi usado para a Distribuição Marron-Wand 15, para amostras com 32 elementos por grupo. Os programas usados nas outras condições foram semelhantes a este, com as alterações necessárias nos parâmetros (tipo de distribuição, desvio-padrão e tamanhos das amostras).

Anexo V**Estatísticas Descritivas (Mínimo, Máximo e Média) da Diferença de Potência entre Testes de Aleatorização e Testes t de Student**

Distribuição	Estatísticas	n ₁ = n ₂ = 8			n ₁ = n ₂ = 16			n ₁ = n ₂ = 32		
		Esq	Dir	Bi	Esq	Dir	Bi	Esq	Dir	Bi
M-W 1	Mínimo	-0.003	-0.004	-0.004	-0.004	-0.003	-0.004	-0.003	-0.002	-0.003
	Máximo	0.001	0.001	0.002	0.001	0.001	0.002	0.001	0.000	0.001
	Média	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000	0.000
M-W 2	Mínimo	-0.002	-0.003	-0.003	-0.002	-0.002	-0.003	-0.003	-0.003	-0.002
	Máximo	0.002	0.002	0.005	0.002	0.002	0.005	0.001	0.001	0.001
	Média	0.000	0.000	0.001	0.000	0.001	0.000	0.000	0.000	0.000
M-W 3	Mínimo	-0.003	-0.002	-0.003	-0.002	-0.003	-0.002	-0.004	-0.003	-0.001
	Máximo	0.006	0.005	0.015	0.001	0.005	0.015	0.001	0.000	0.002
	Média	0.000	0.000	0.002	0.000	0.002	0.000	0.000	0.000	0.000
M-W 4	Mínimo	-0.002	-0.003	-0.003	-0.004	-0.002	-0.003	-0.002	-0.003	-0.001
	Máximo	0.001	0.003	0.008	0.000	0.003	0.008	0.000	0.000	0.002
	Média	0.000	0.000	0.002	0.000	0.002	0.000	0.000	0.000	0.000
M-W 5	Mínimo	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000
	Máximo	0.066	0.066	0.106	0.028	0.066	0.106	0.005	0.006	0.024
	Média	0.005	0.005	0.021	0.002	0.021	0.002	0.000	0.000	0.004
M-W 6	Mínimo	-0.004	-0.006	-0.008	-0.003	-0.004	-0.006	-0.002	-0.003	-0.003
	Máximo	0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.001
	Média	0.000	0.000	-0.002	0.000	-0.002	0.000	0.000	0.000	0.000
M-W 7	Mínimo	-0.005	-0.005	-0.013	-0.002	-0.005	-0.005	-0.003	-0.002	-0.004
	Máximo	0.001	0.003	0.002	0.001	0.003	0.002	0.001	0.001	0.000
	Média	-0.001	-0.001	-0.003	0.000	-0.003	0.000	0.000	0.000	0.000
M-W 8	Mínimo	-0.004	-0.004	-0.006	-0.003	-0.004	-0.004	-0.003	-0.002	-0.003
	Máximo	0.001	0.001	0.000	0.001	0.001	0.000	0.000	0.000	0.001
	Média	0.000	0.000	-0.002	0.000	-0.002	0.000	0.000	0.000	0.000
M-W 9	Mínimo	-0.004	-0.005	-0.009	-0.002	-0.004	-0.005	-0.003	-0.002	-0.004
	Máximo	0.001	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.001
	Média	0.000	0.000	-0.002	0.000	-0.002	0.000	0.000	0.000	0.000
M-W 10	Mínimo	-0.003	-0.003	-0.004	-0.003	-0.003	-0.003	-0.004	-0.003	-0.002
	Máximo	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	Média	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000	0.000
M-W 11	Mínimo	-0.004	-0.004	-0.008	-0.003	-0.004	-0.004	-0.003	-0.003	-0.004
	Máximo	0.001	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000
	Média	0.000	0.000	-0.002	0.000	-0.002	0.000	0.000	0.000	0.000
M-W 12	Mínimo	-0.004	-0.003	-0.006	-0.004	-0.004	-0.003	-0.003	-0.002	-0.002
	Máximo	0.000	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.000
	Média	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000	0.000

[continua]

Distribuição	Estatísticas	$n_1 = n_2 = 8$			$n_1 = n_2 = 16$			$n_1 = n_2 = 32$		
		Esq	Dir	Bi	Esq	Dir	Bi	Esq	Dir	Bi
M-W 13	Mínimo	-0.004	-0.006	-0.009	-0.004	-0.004	-0.006	-0.002	-0.003	-0.004
	Máximo	0.001	0.001	0.000	0.001	0.001	0.000	0.000	0.001	0.001
	Média	0.000	0.000	-0.002	0.000	-0.002	0.000	0.000	0.000	0.000
M-W 14	Mínimo	-0.005	-0.005	-0.012	-0.003	-0.005	-0.005	-0.003	-0.002	-0.004
	Máximo	0.001	0.000	0.000	0.001	0.000	0.000	0.002	0.001	0.000
	Média	-0.001	-0.001	-0.003	0.000	-0.003	0.000	0.000	0.000	0.000
M-W 15	Mínimo	-0.006	-0.005	-0.011	-0.002	-0.006	-0.005	-0.003	-0.003	-0.004
	Máximo	0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.001
	Média	0.000	0.000	-0.003	0.000	-0.003	0.000	0.000	0.000	0.000

Nota: Esc: testes unicaudal à esquerda; Dir: teste unicaudal à direita; Bi: teste bicaudal.

Anexo VI**Comparação da Potência dos Testes de Aleatorização:
Distribuições Marron-Wand (2 a 15) vs. Distribuição Normal**

1. Amostras com $n_1 = n_2 = 8$
 - 1.1. Testes bi-caudais
 - 1.2. Testes uni-caudais

2. Amostras com $n_1 = n_2 = 16$
 - 2.1. Testes bi-caudais
 - 2.2. Testes uni-caudais

3. Amostras com $n_1 = n_2 = 32$
 - 3.1. Testes bi-caudais
 - 3.2. Testes uni-caudais

1. Amostras com $n_1 = n_2 = 8$

1.1. Testes bi-caudais

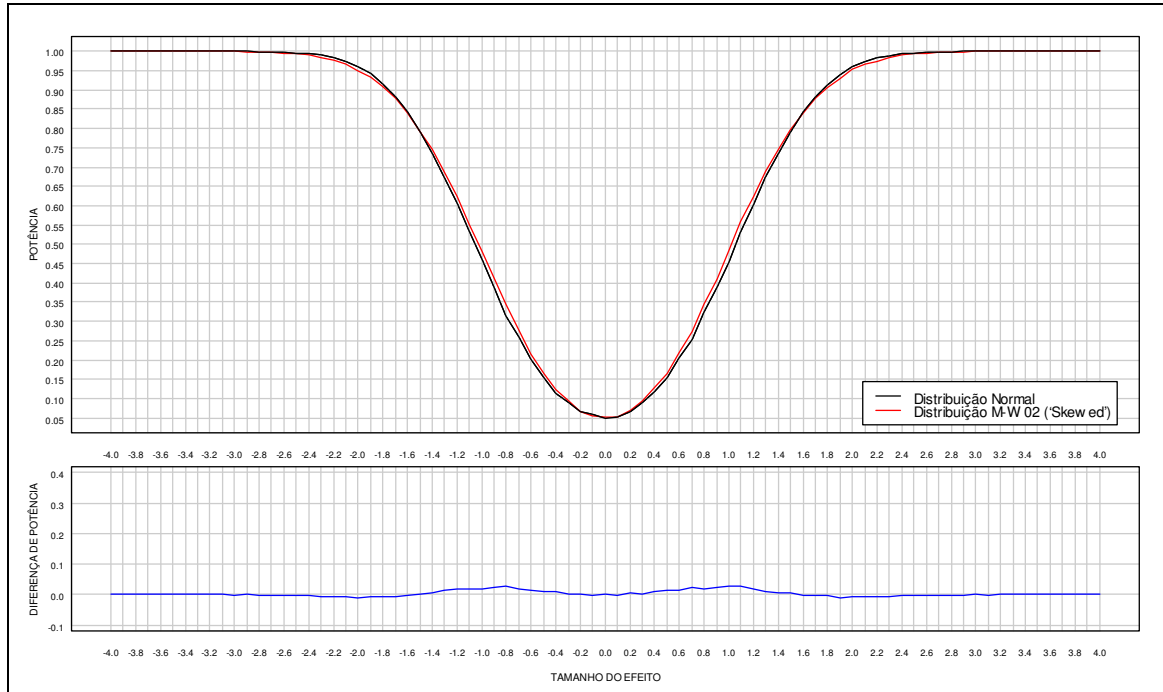


Figura Anexo VI. 1 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 02 ('Skewed') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

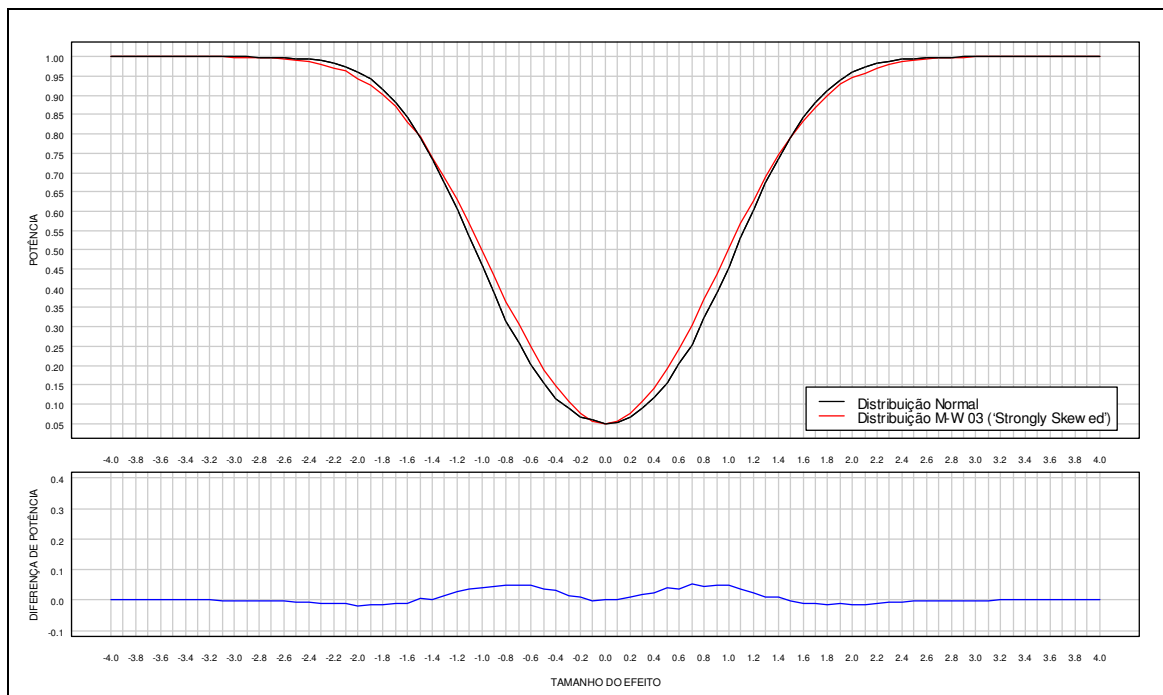


Figura Anexo VI. 2 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 03 ('Strongly Skewed') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

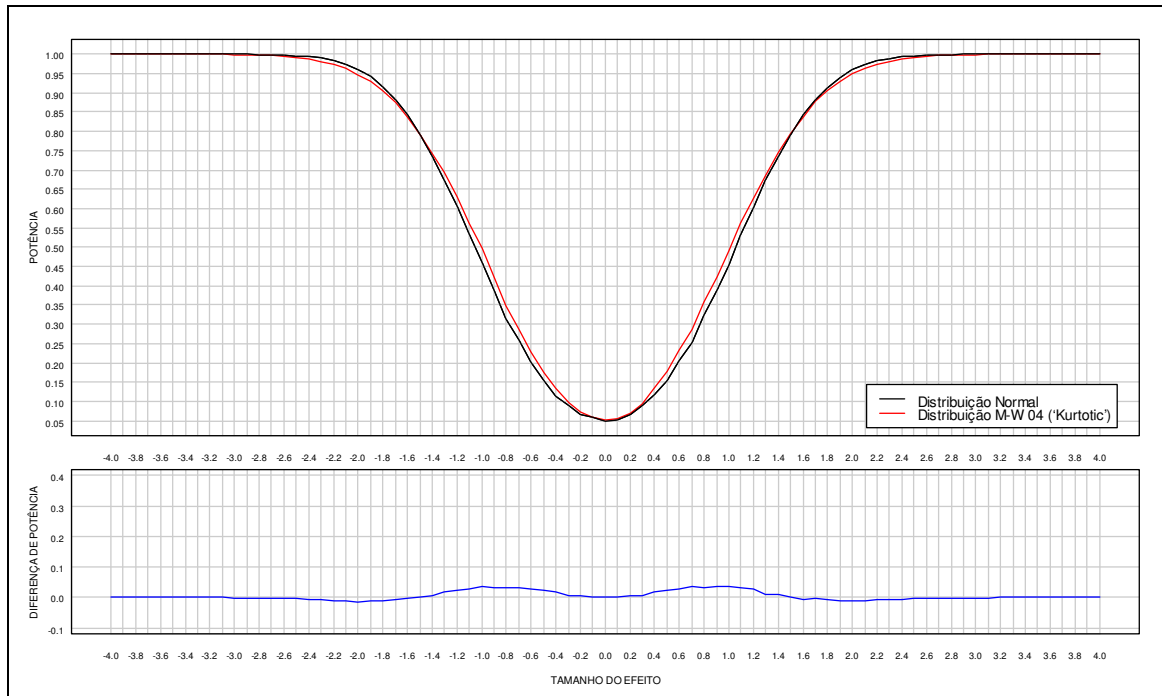


Figura Anexo VI. 3 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 04 ('Kurtotic') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

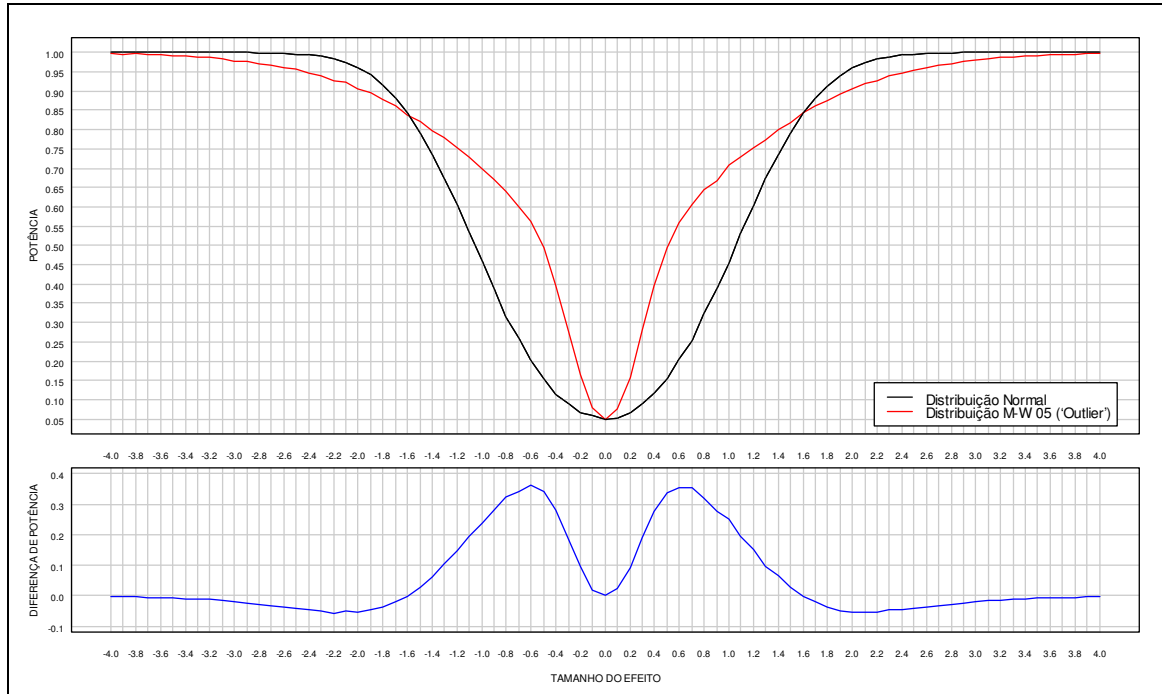


Figura Anexo VI. 4 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 05 ('Outlier') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

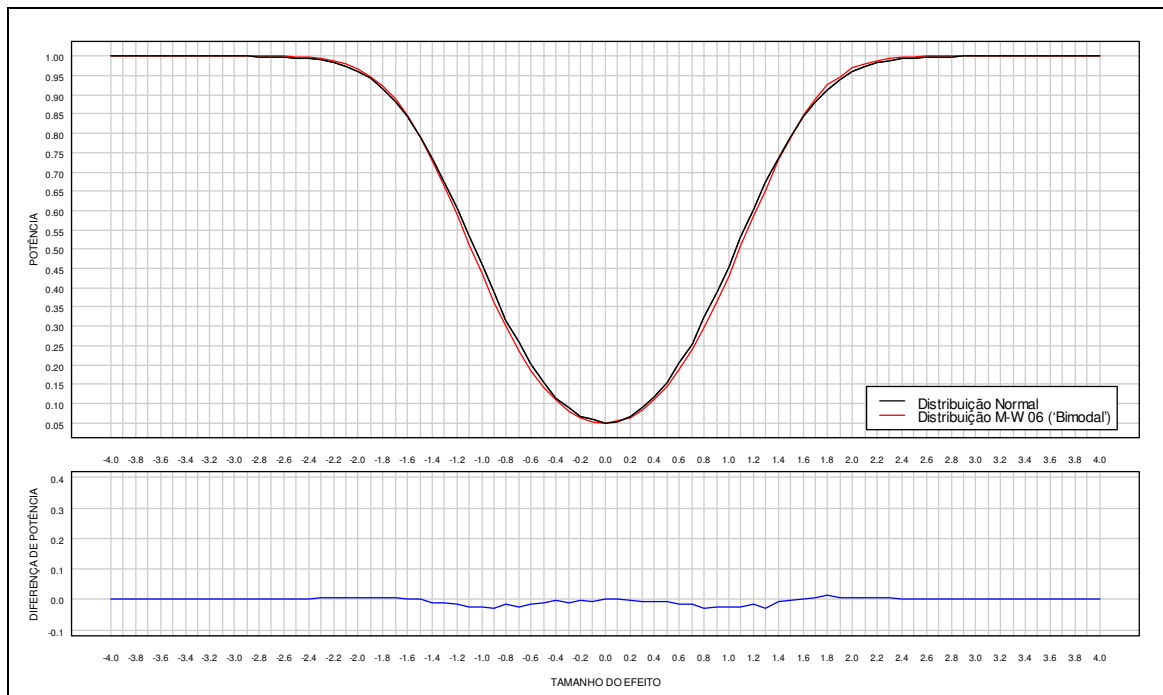


Figura Anexo VI. 5 [Painel superior] Potência do teste de aleatorização bi-causal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 06 ('Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

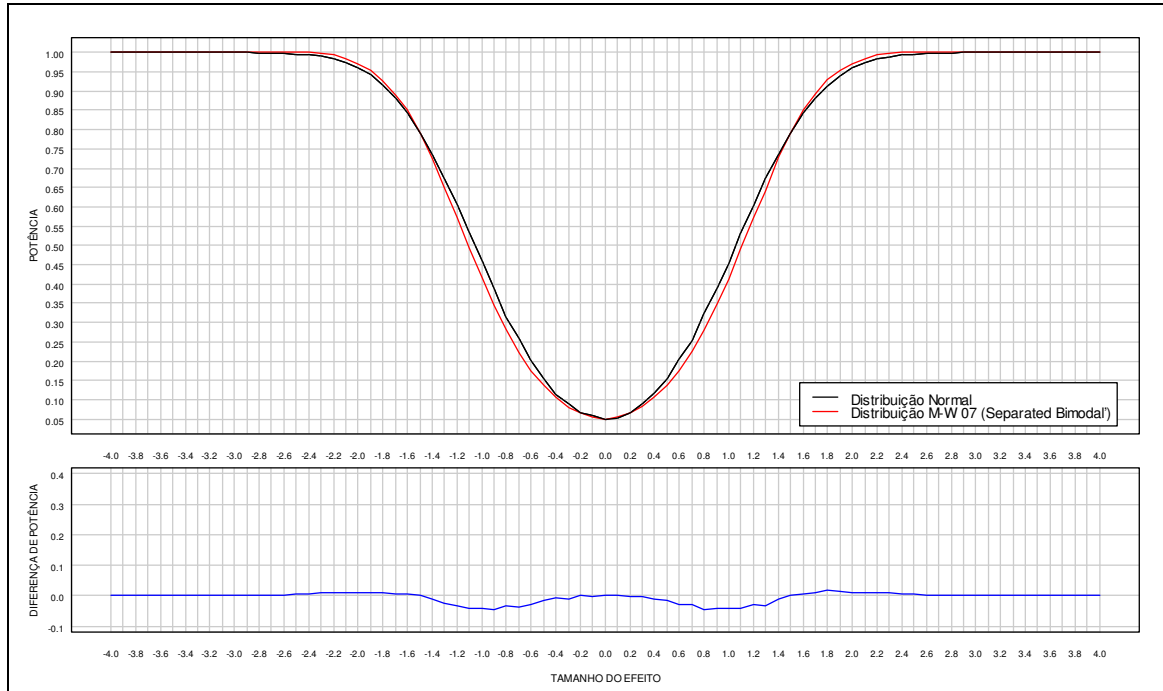


Figura Anexo VI. 6 [Painel superior] Potência do teste de aleatorização bi-causal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 07 ('Separated Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

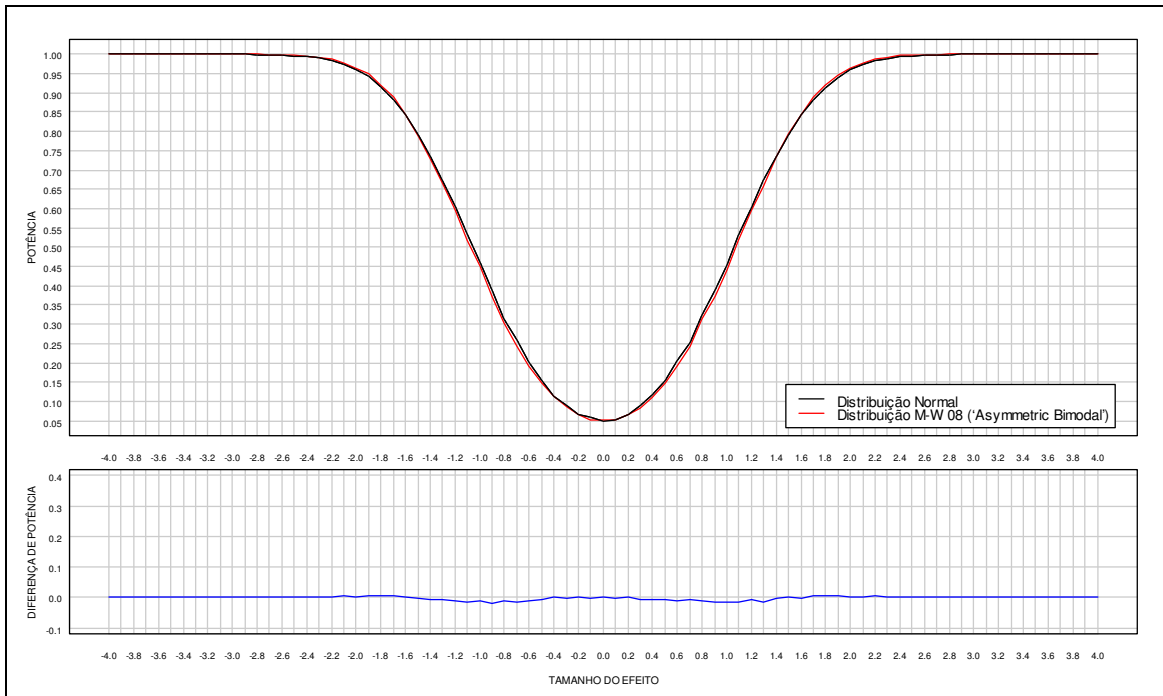


Figura Anexo VI. 7 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 08 ('Asymmetric Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

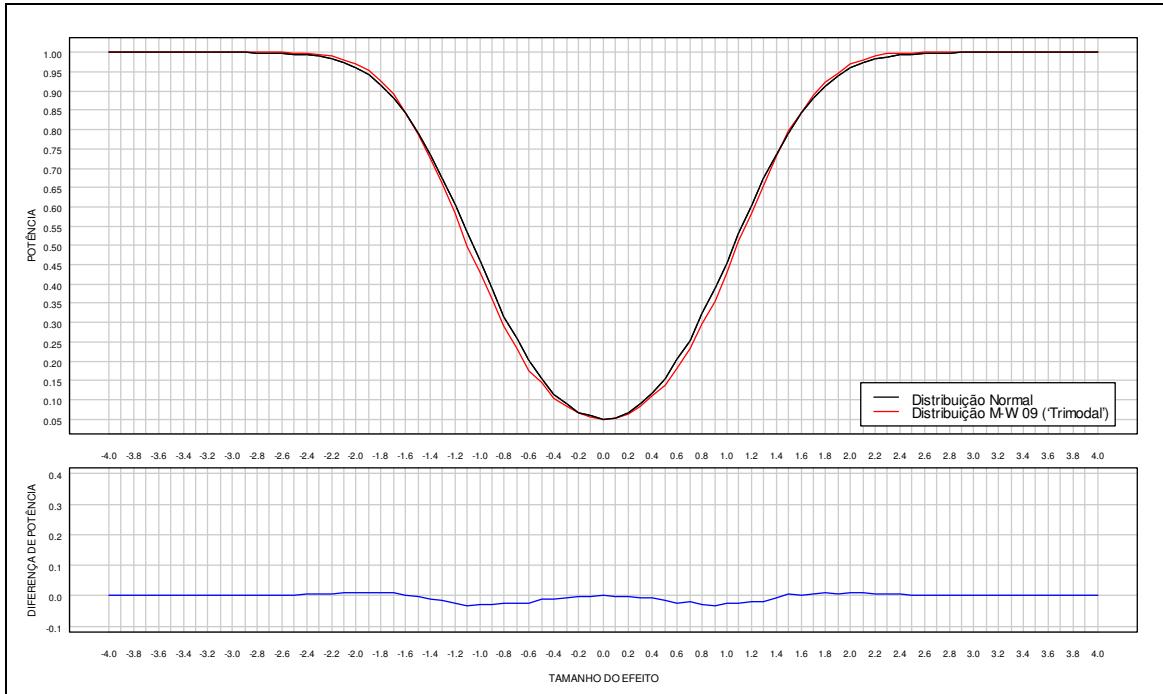


Figura Anexo VI. 8 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 09 ('Trimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

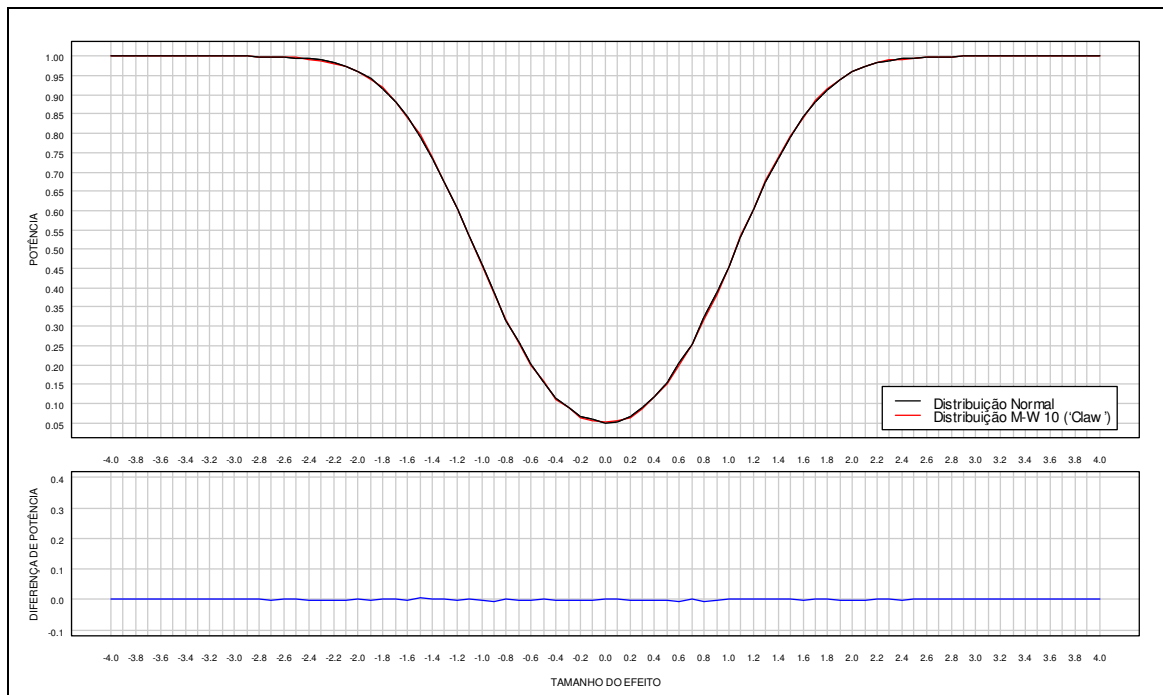


Figura Anexo VI. 9 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 10 ('Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

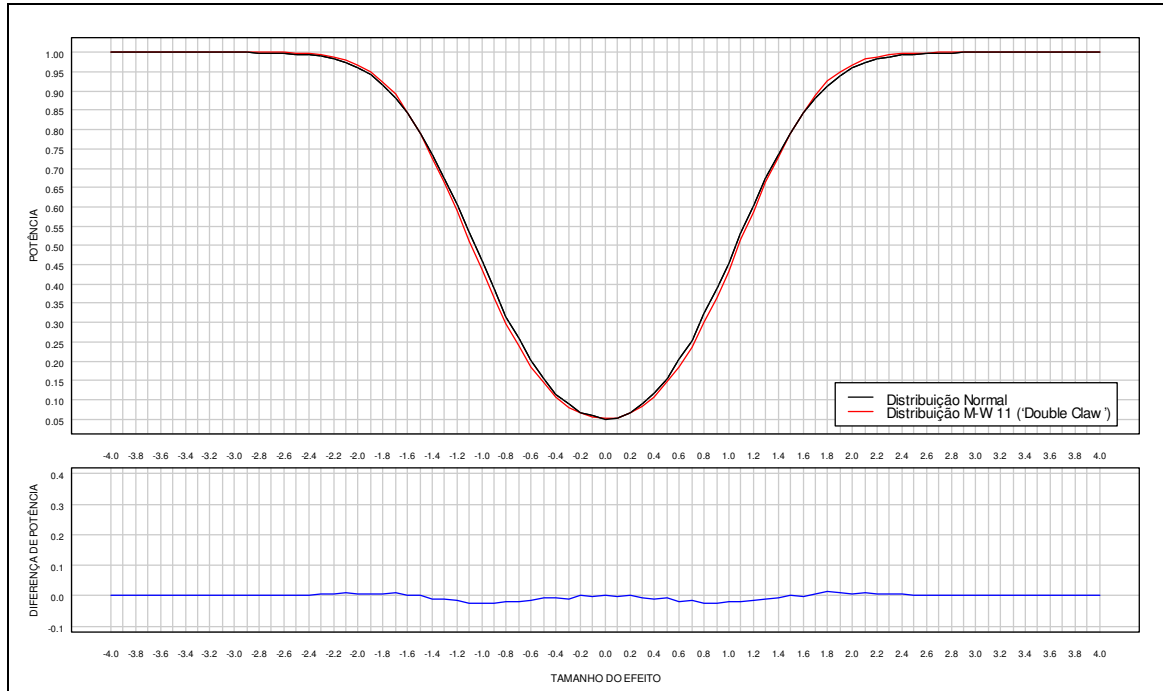


Figura Anexo VI. 10 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 11 ('Double Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

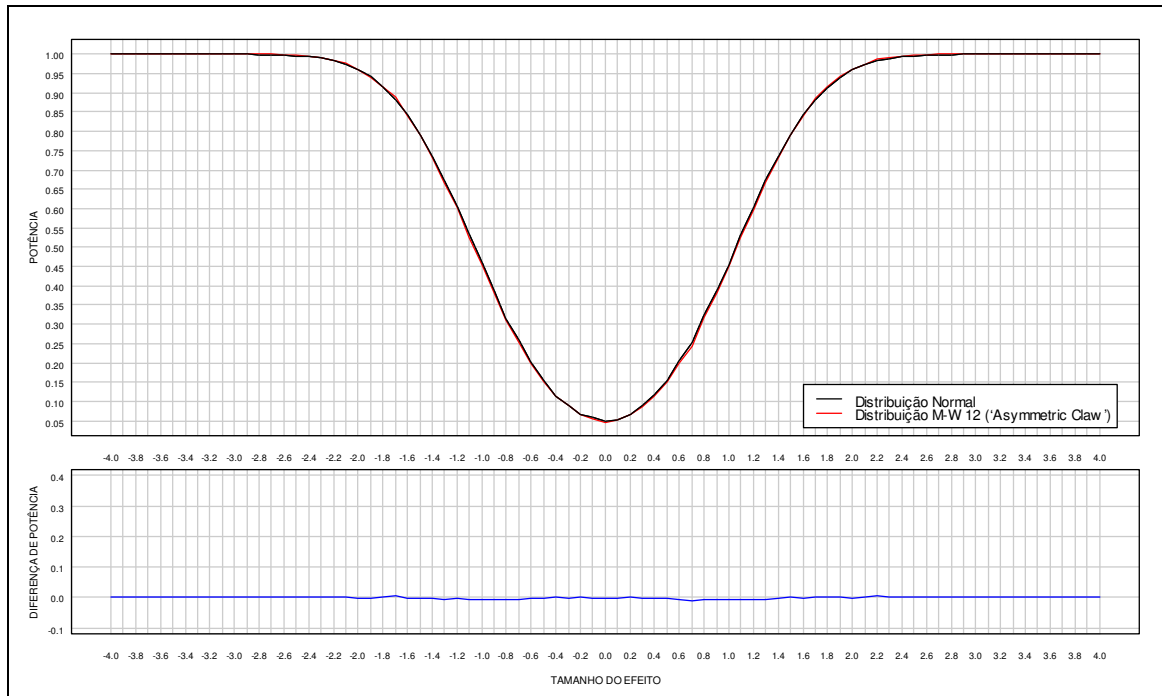


Figura Anexo VI. 11 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 12 ('Asymmetric Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

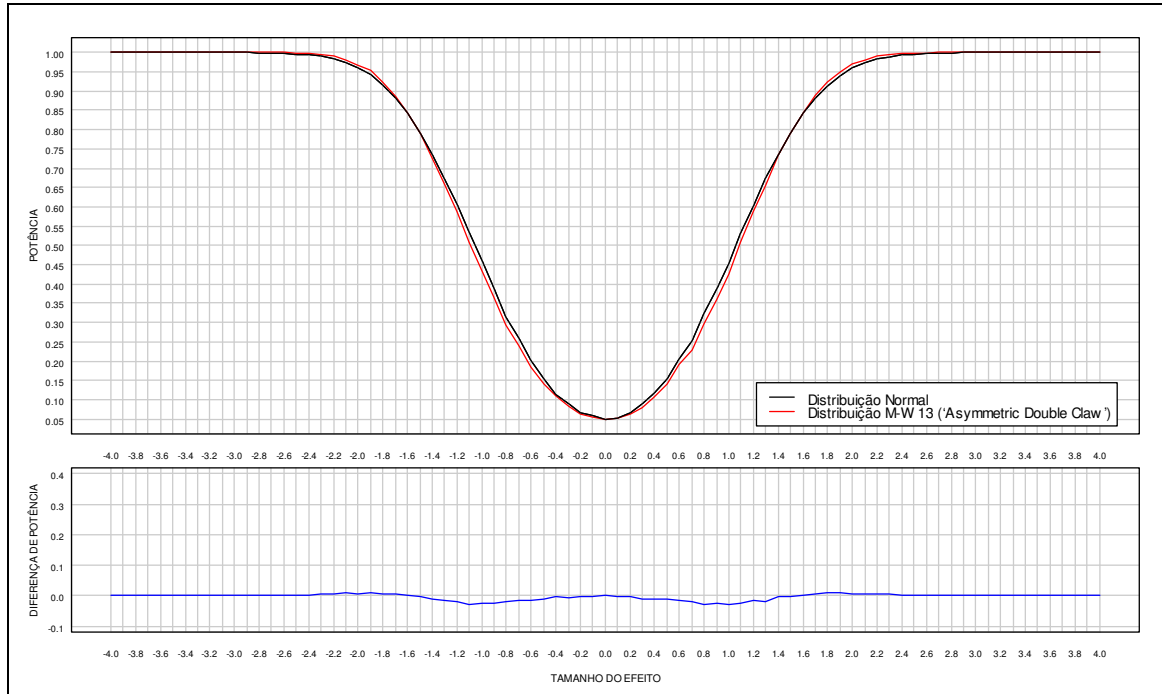


Figura Anexo VI. 12 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 13 ('Asymmetric Double Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

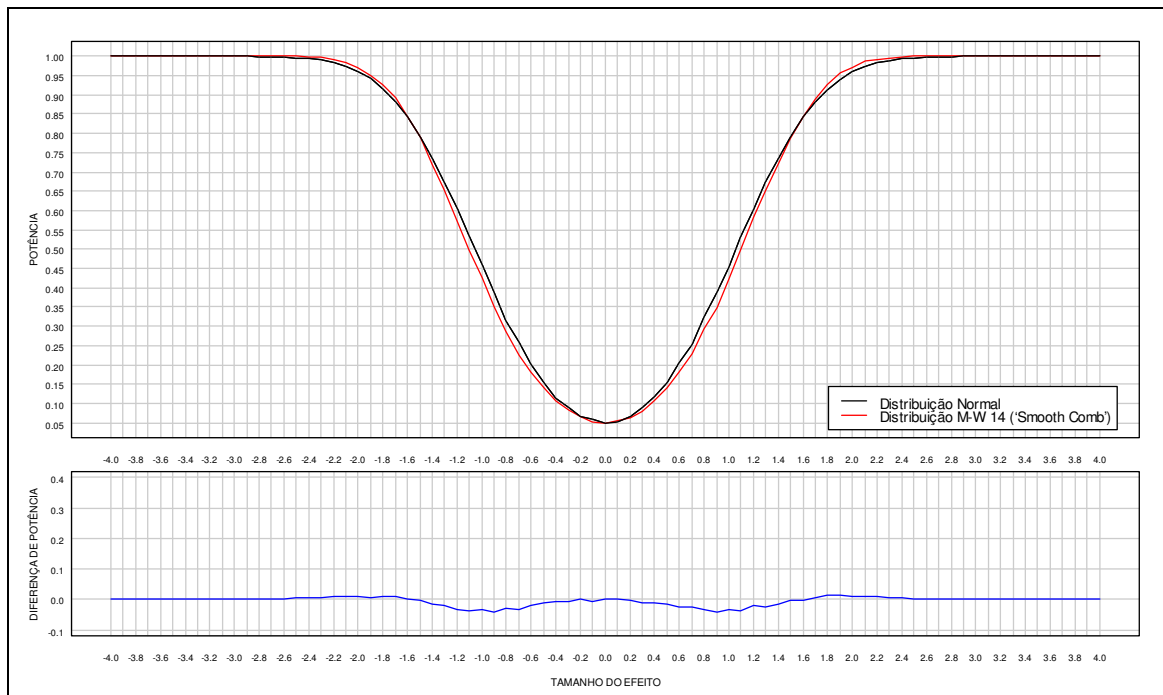


Figura Anexo VI. 13 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 14 ('Smooth Comb') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

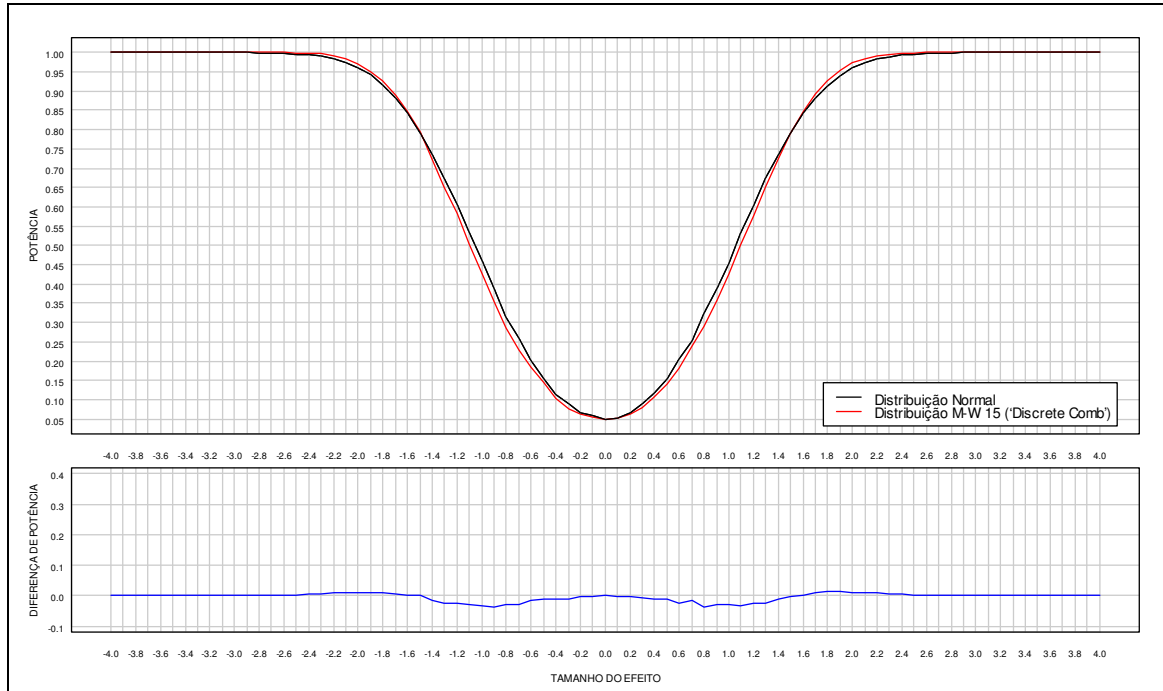


Figura Anexo VI. 14 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 15 ('Discrete Comb') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

1.2. Testes uni-caudais

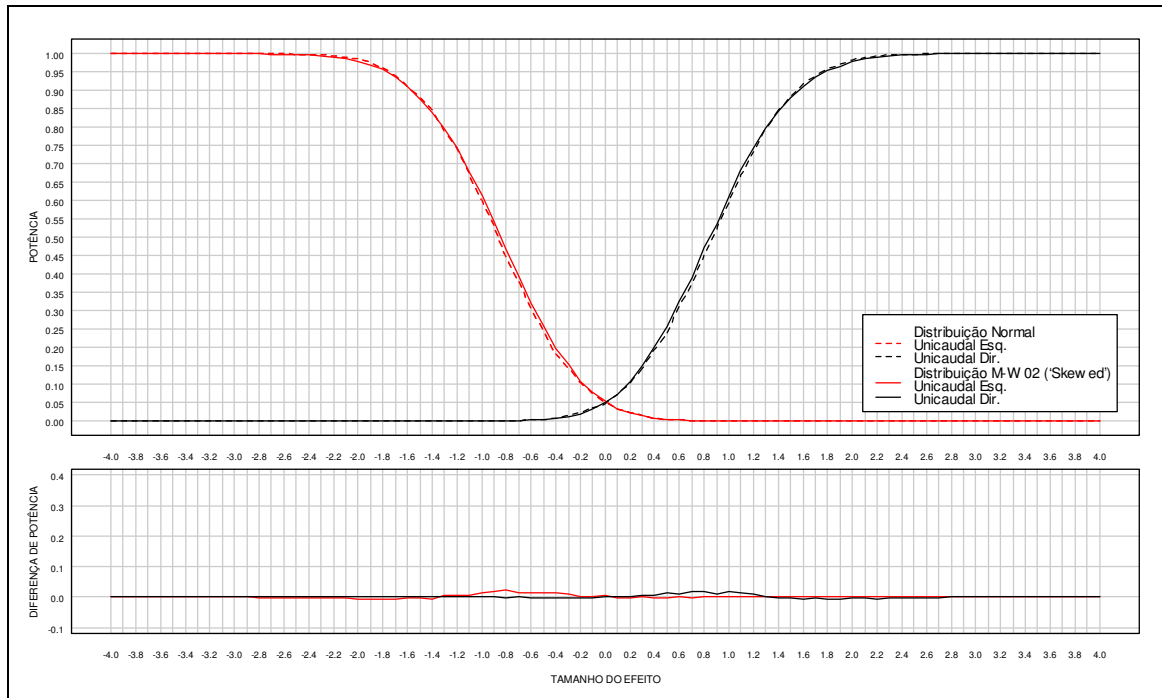


Figura Anexo VI. 15 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 02 ('Skewed') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

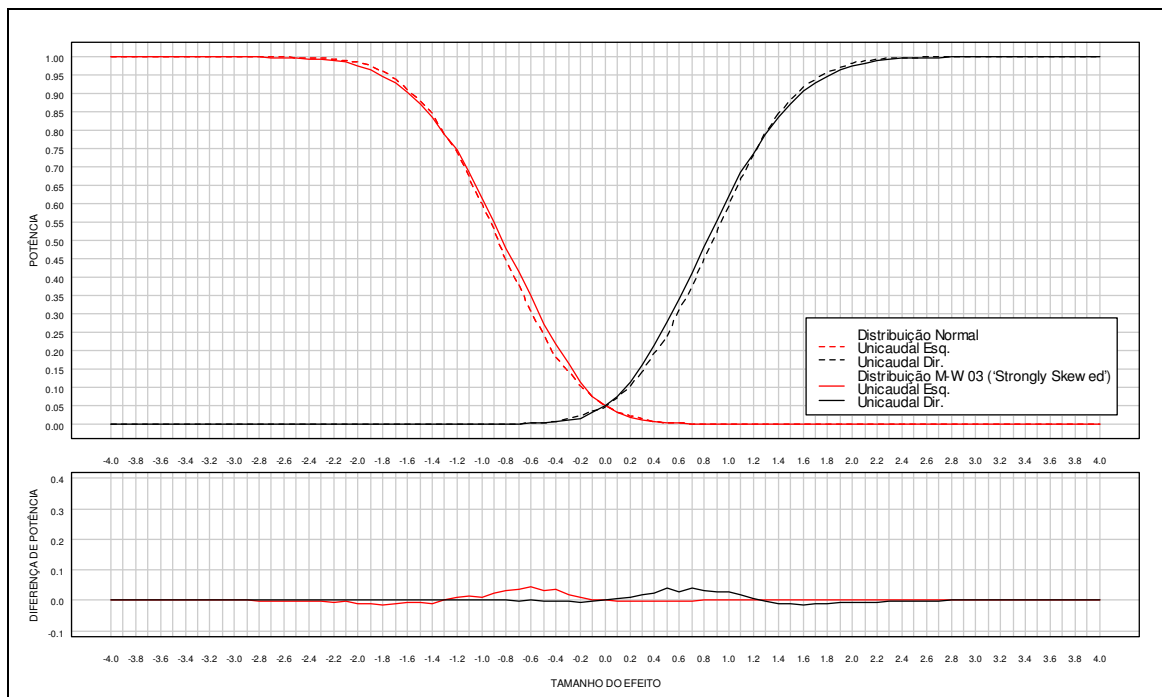


Figura Anexo VI. 16 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 03 ('Strongly Skewed') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

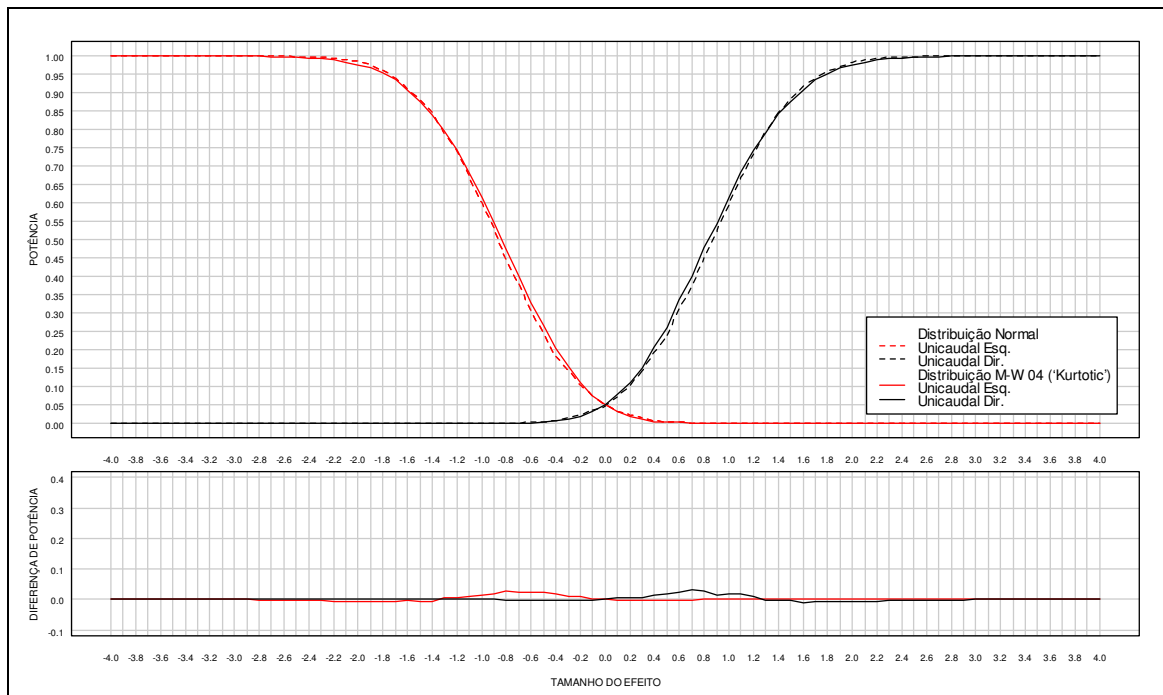


Figura Anexo VI. 17 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 04 ('Kurtotic') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

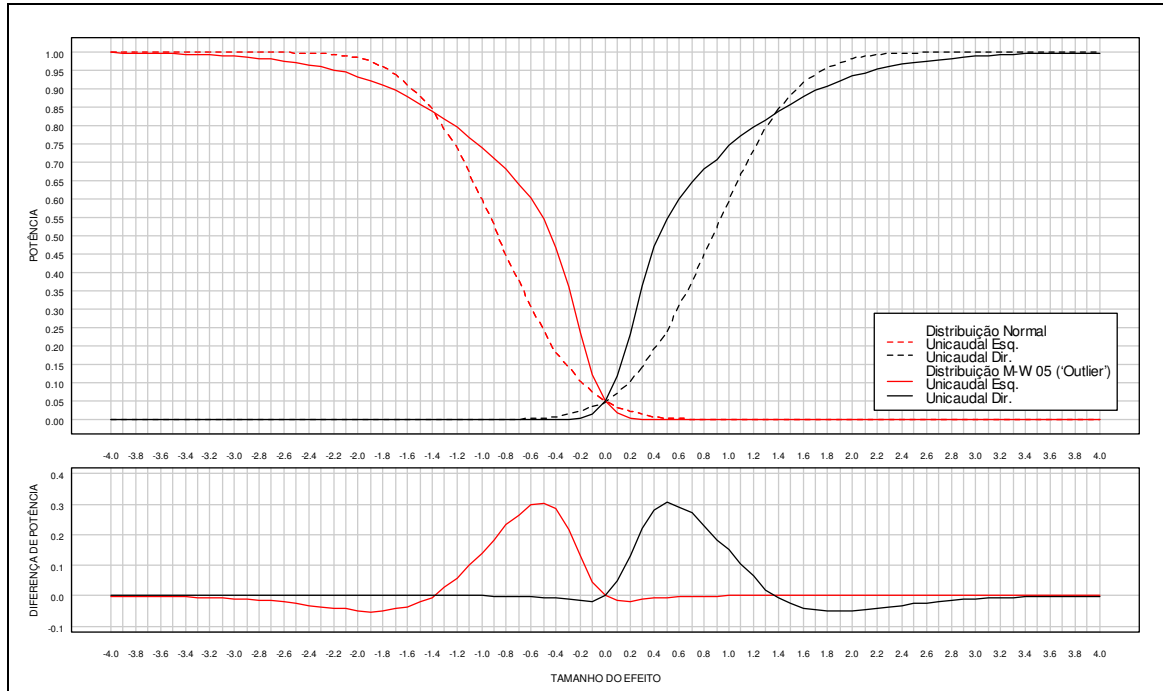


Figura Anexo VI. 18 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 05 ('Outlier') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

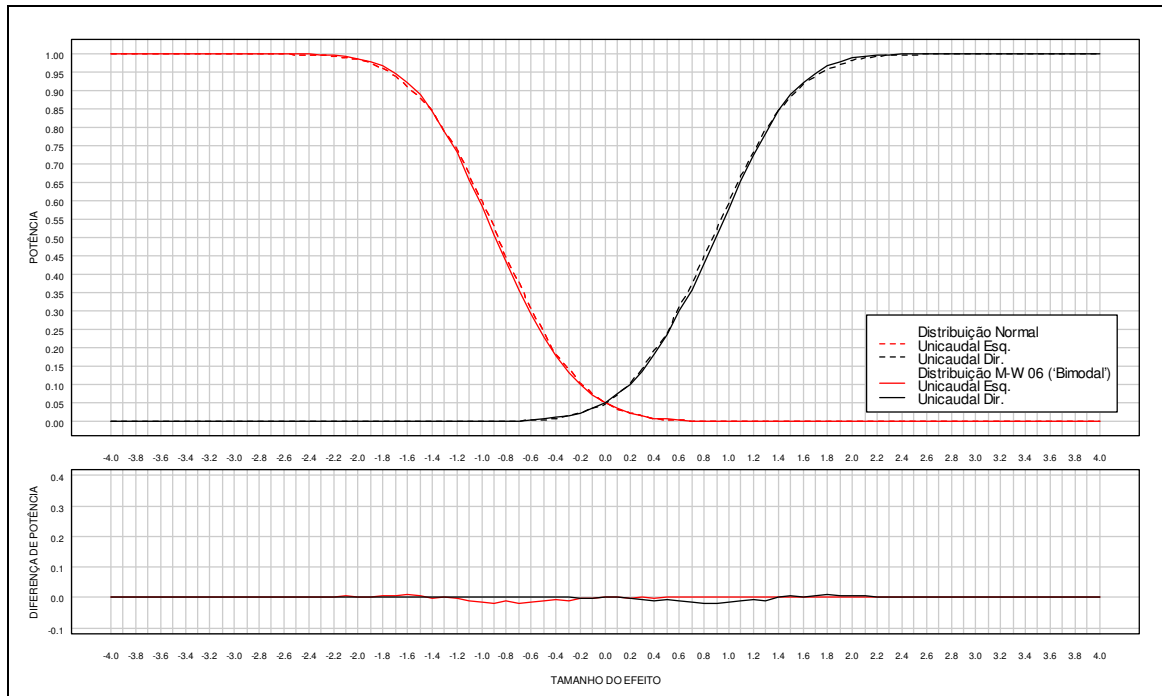


Figura Anexo VI. 19 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 06 ('Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

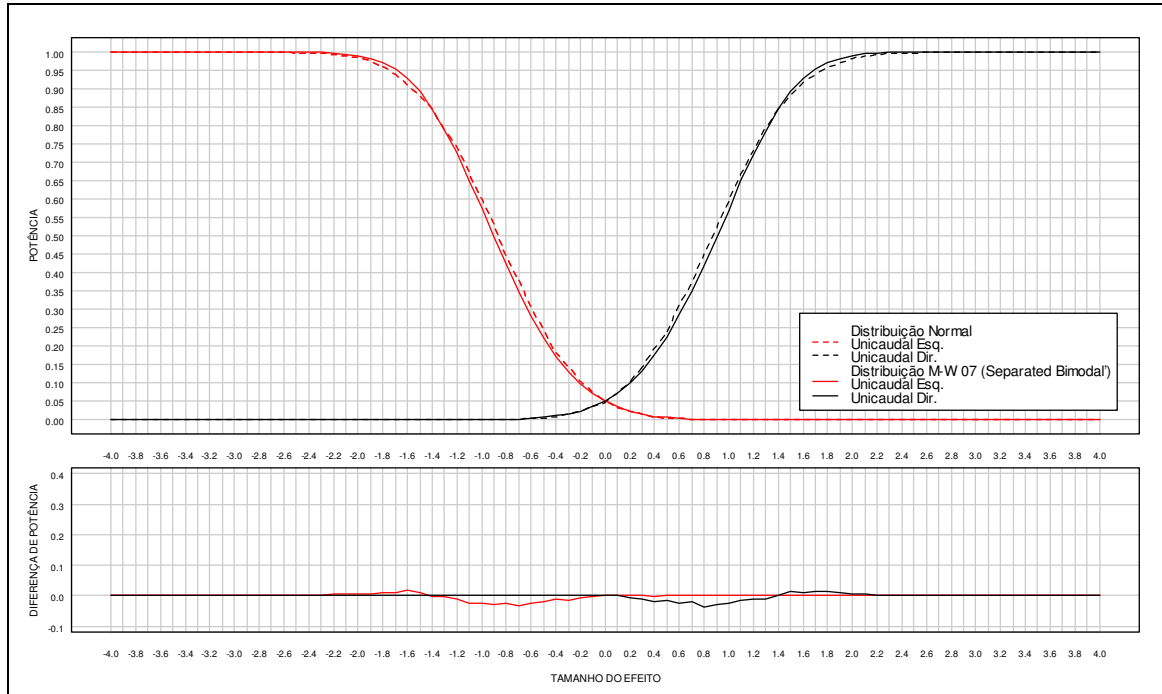


Figura Anexo VI. 20 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 07 ('Separated Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

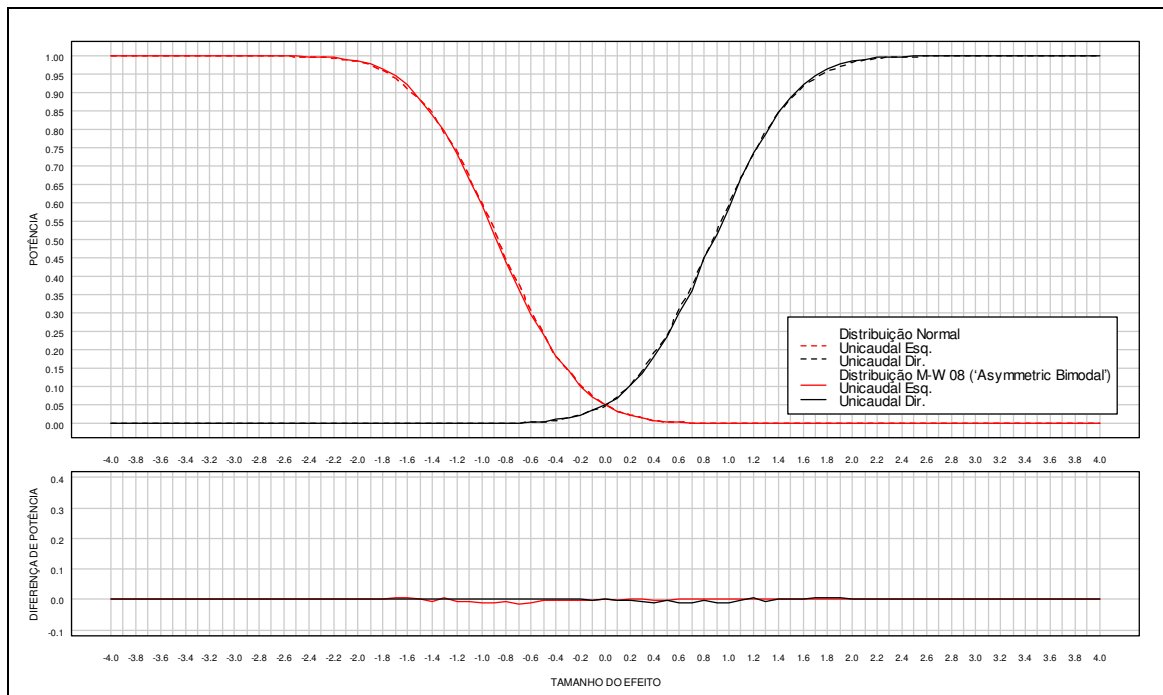


Figura Anexo VI. 21 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 08 ('Asymmetric Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

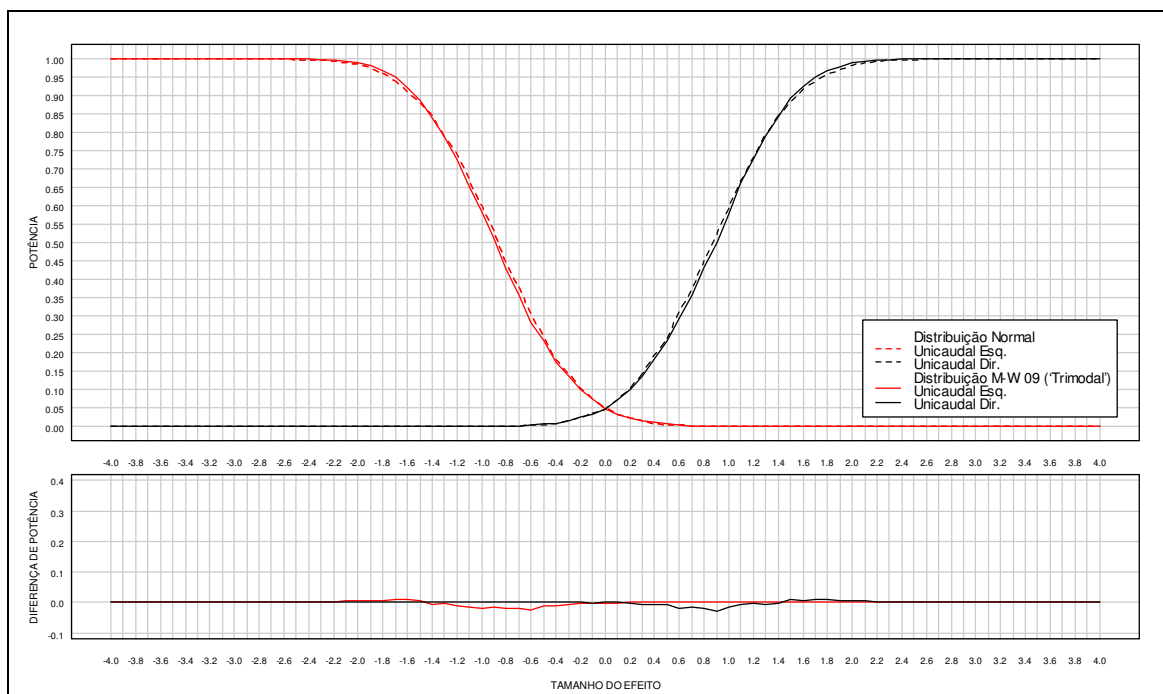


Figura Anexo VI. 22 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 09 ('Trimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

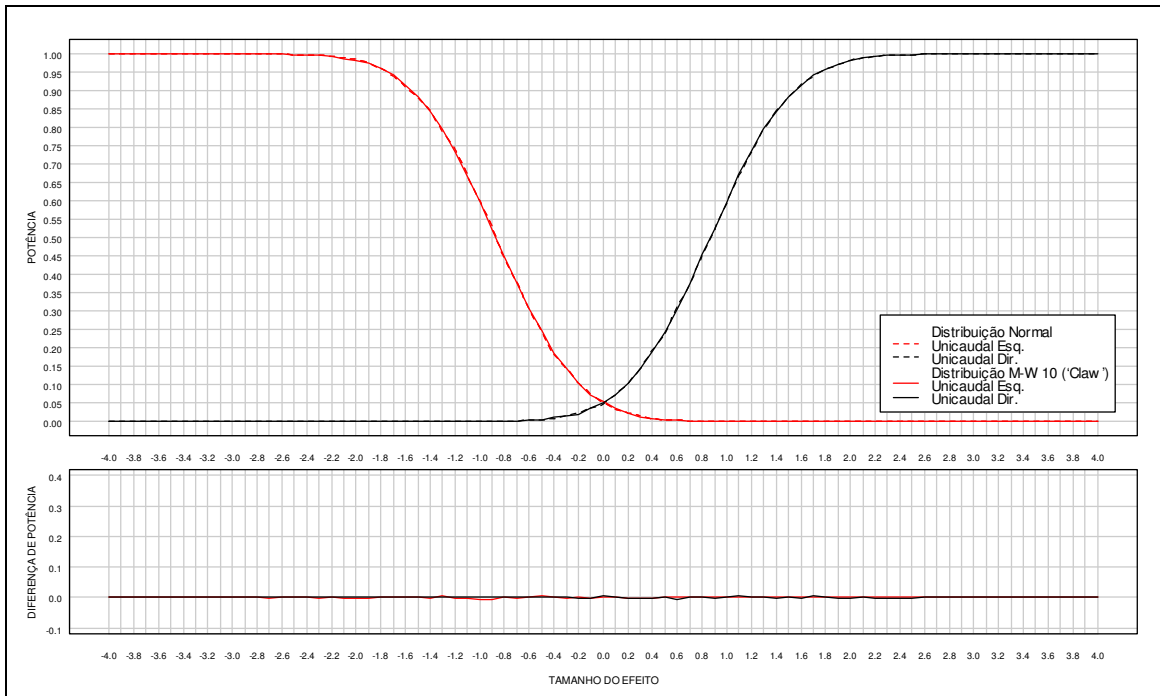


Figura Anexo VI. 23 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 10 ('Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

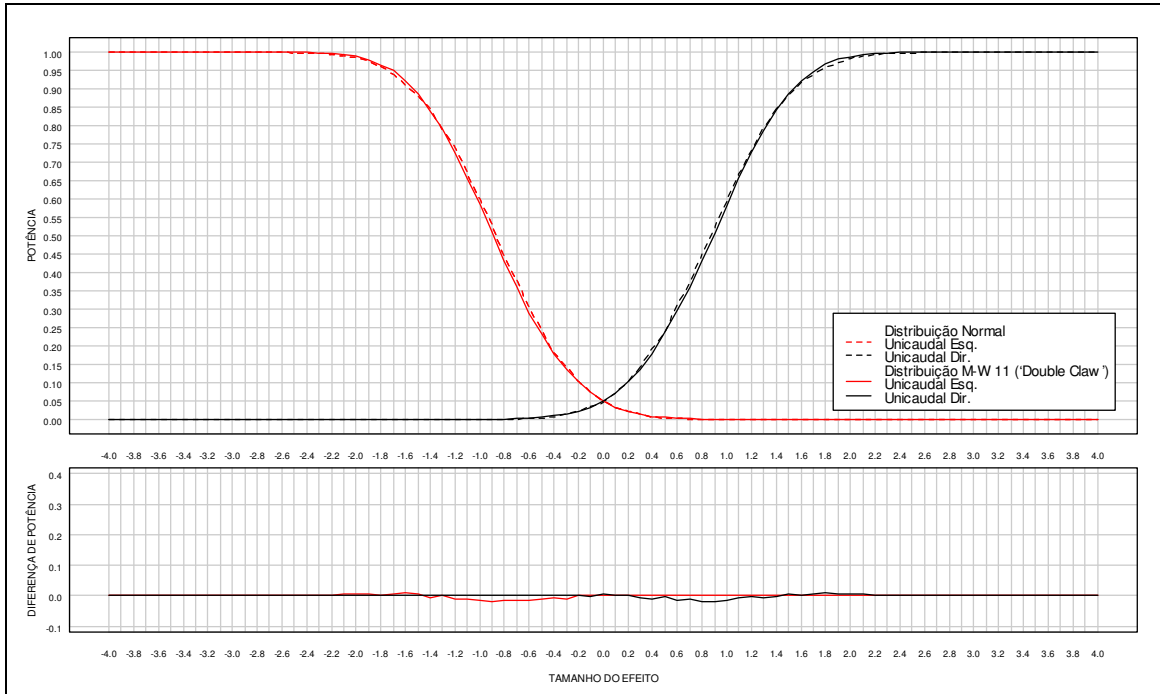


Figura Anexo VI. 24 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 11 ('Double Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

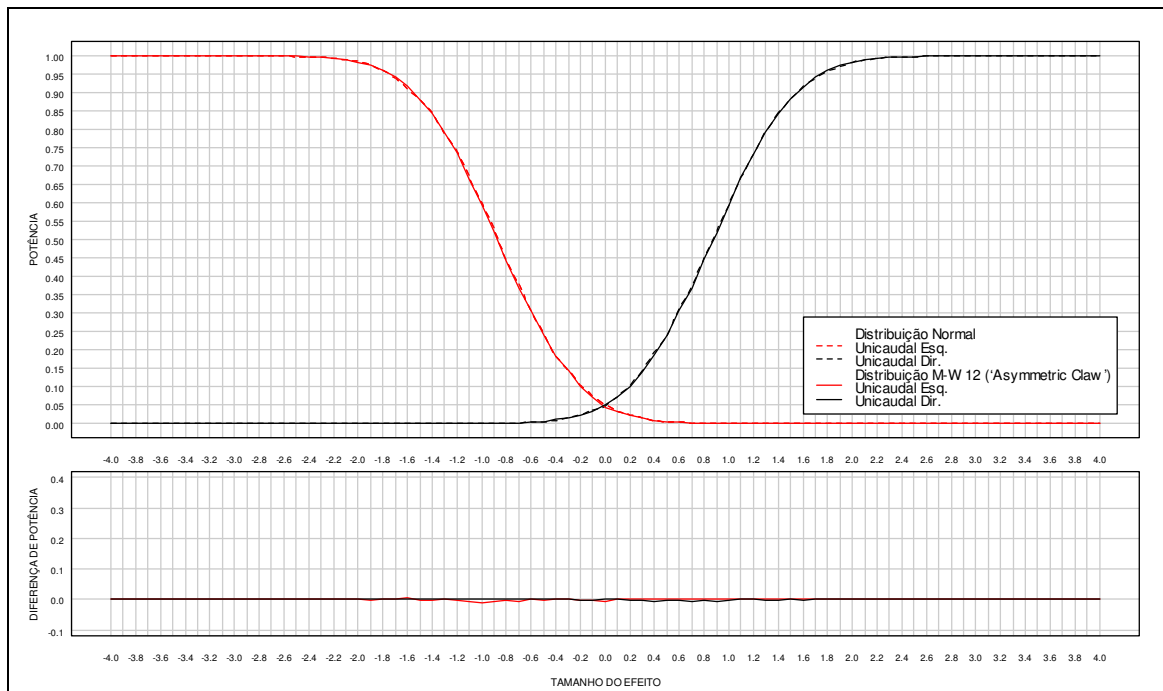


Figura Anexo VI. 25 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 03 ('Asymmetric Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

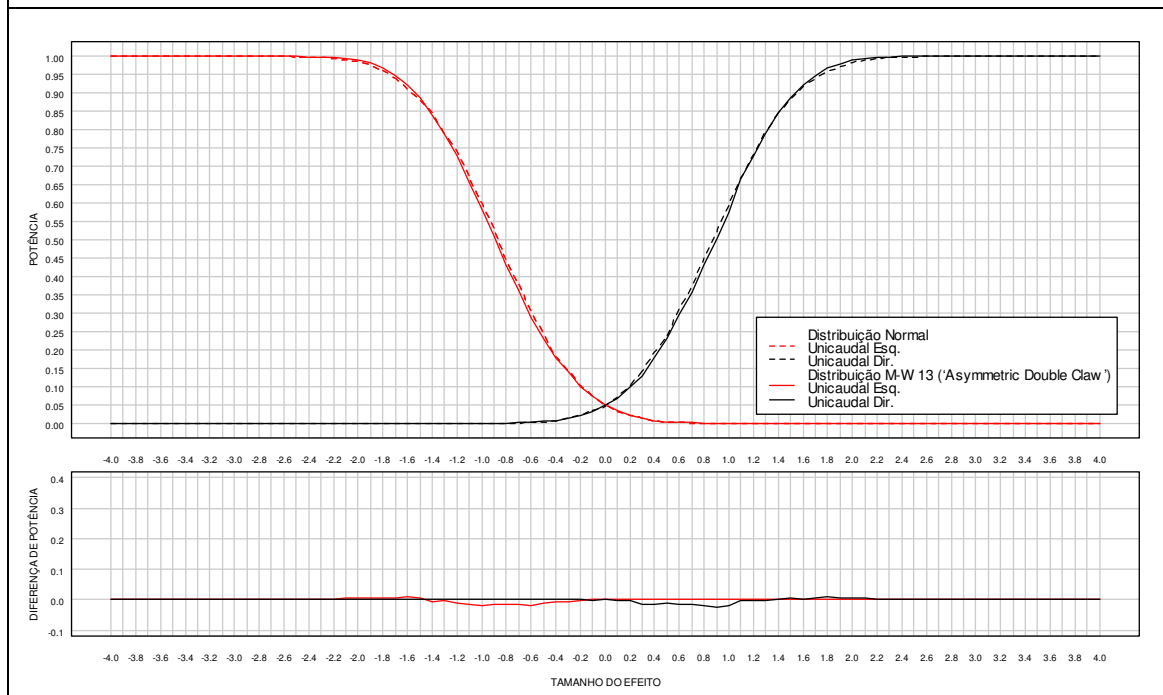


Figura Anexo VI. 26 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 13 ('Asymmetric Double Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

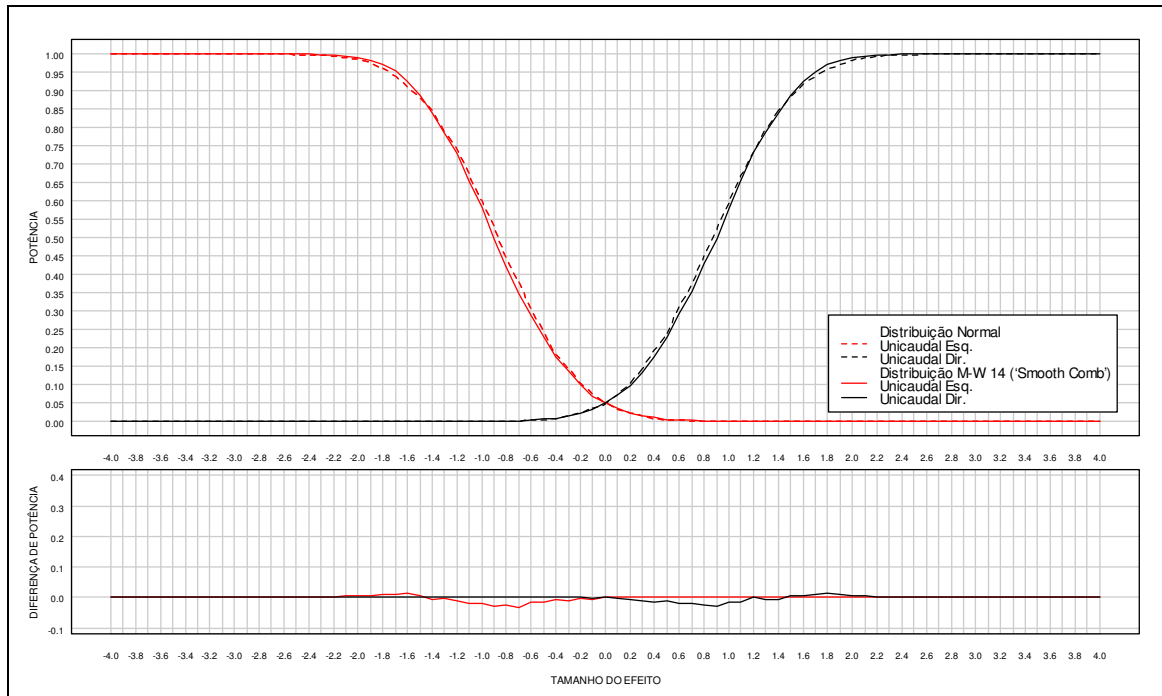


Figura Anexo VI. 27 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 14 ('Smooth Comb') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

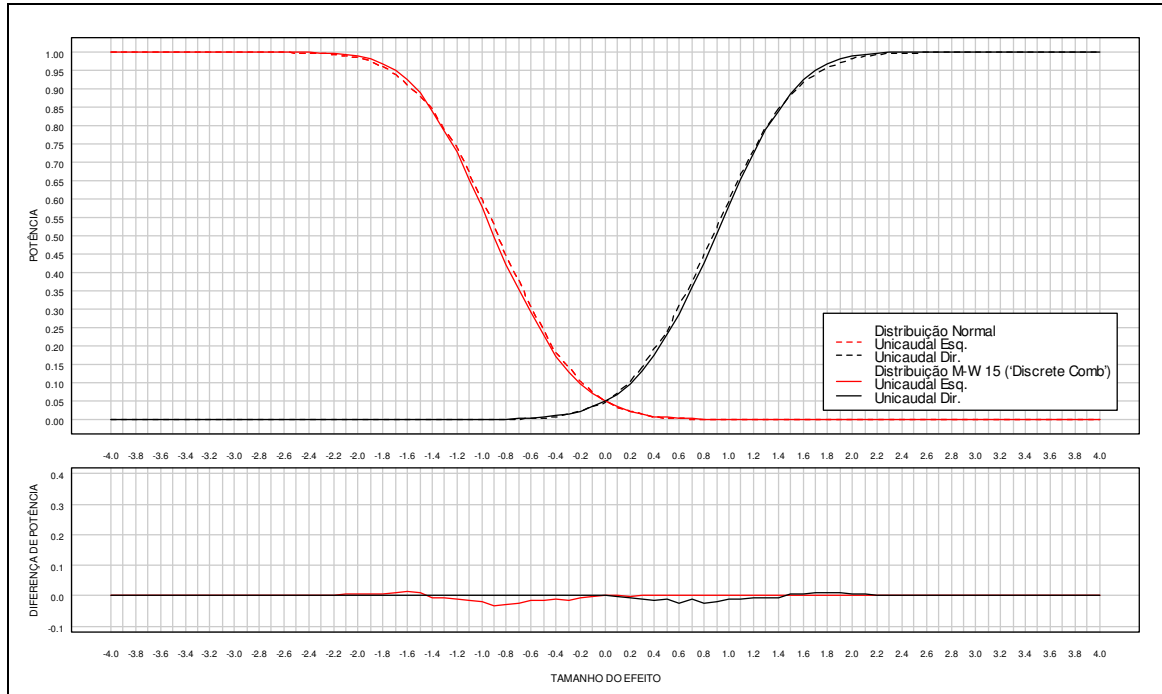


Figura Anexo VI. 28 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição Marron-Wand 15 ('Discrete Comb') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

2. Amostras com $n_1 = n_2 = 16$

2.1. Testes bi-caudais

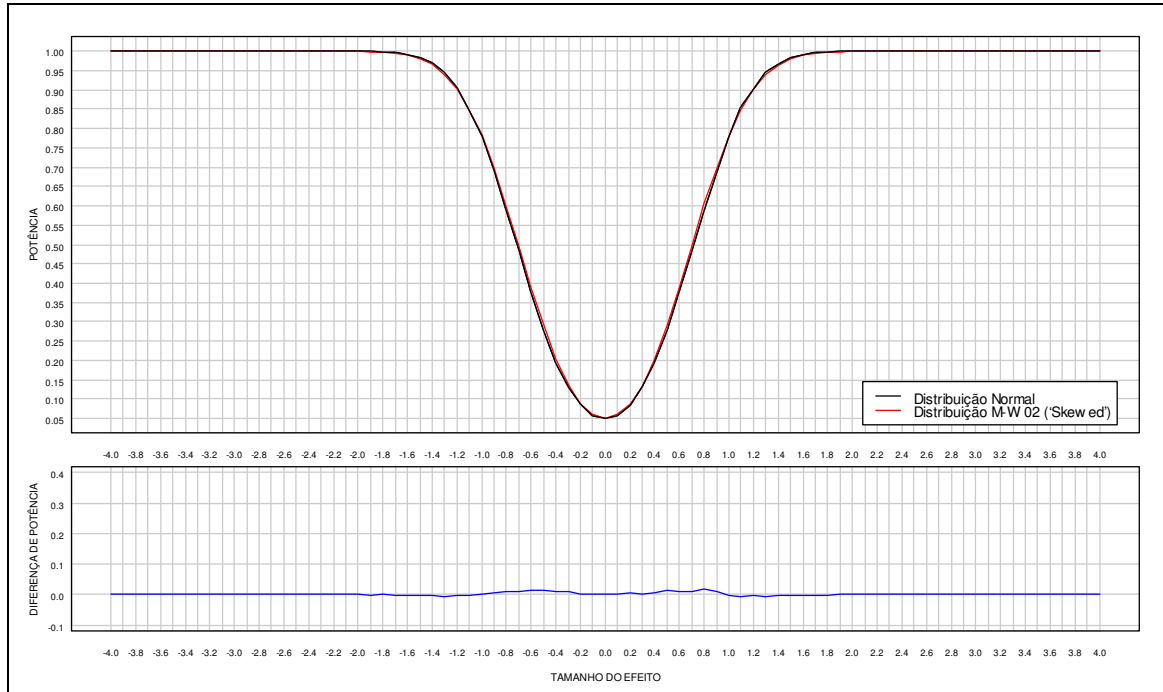


Figura Anexo VI. 29 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 02 ('Skewed') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

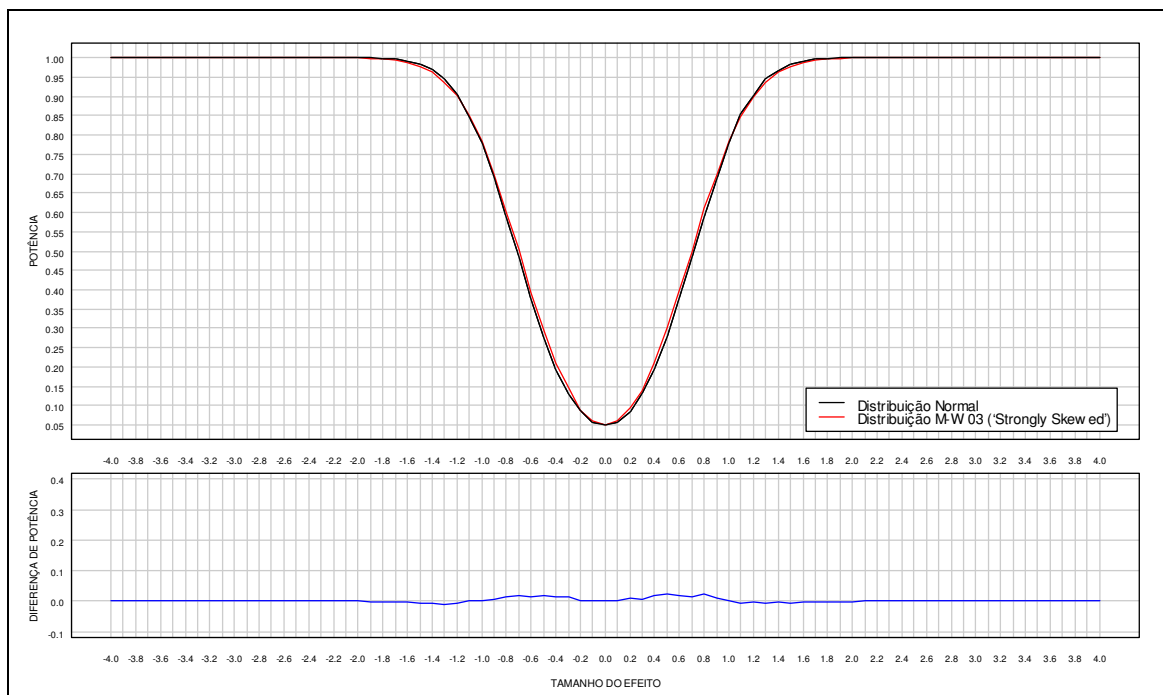


Figura Anexo VI. 30 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 03 ('Strongly Skewed') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

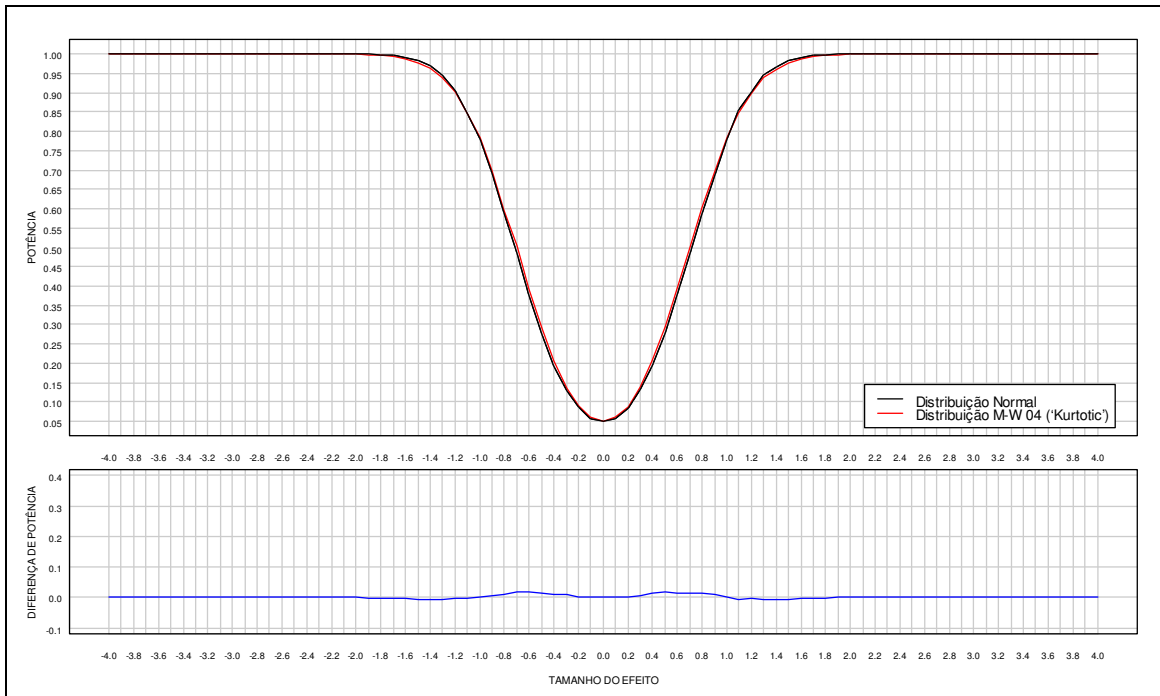


Figura Anexo VI. 31 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 04 ('Kurtotic') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

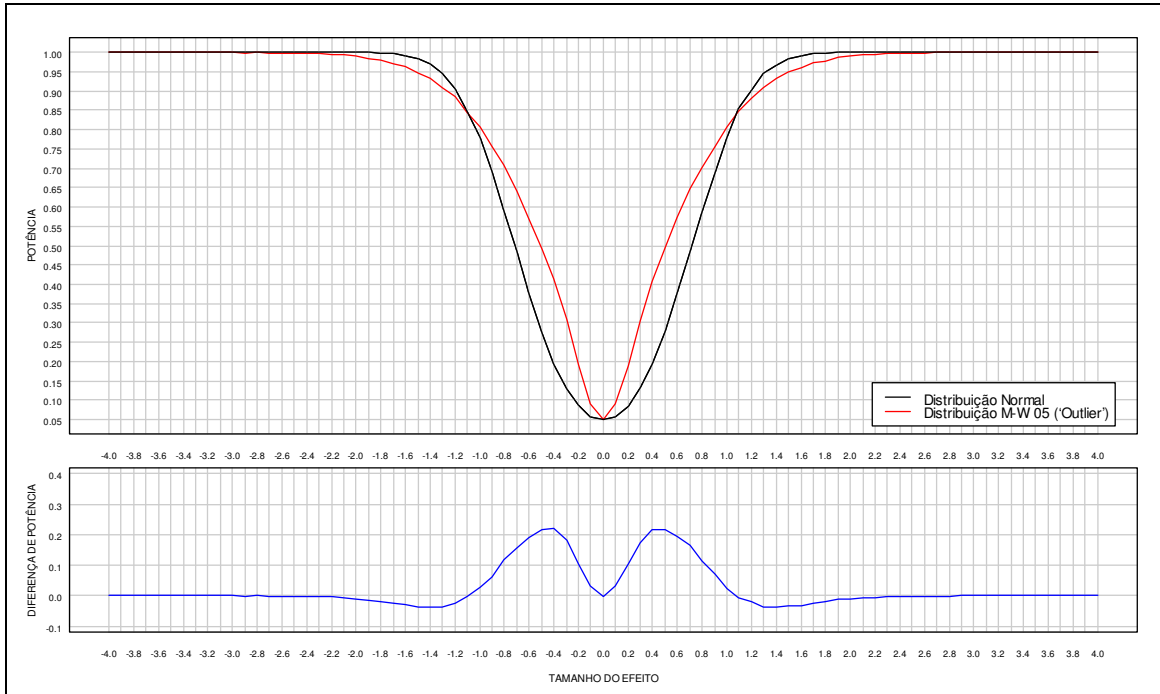


Figura Anexo VI. 32 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 05 ('Outlier') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

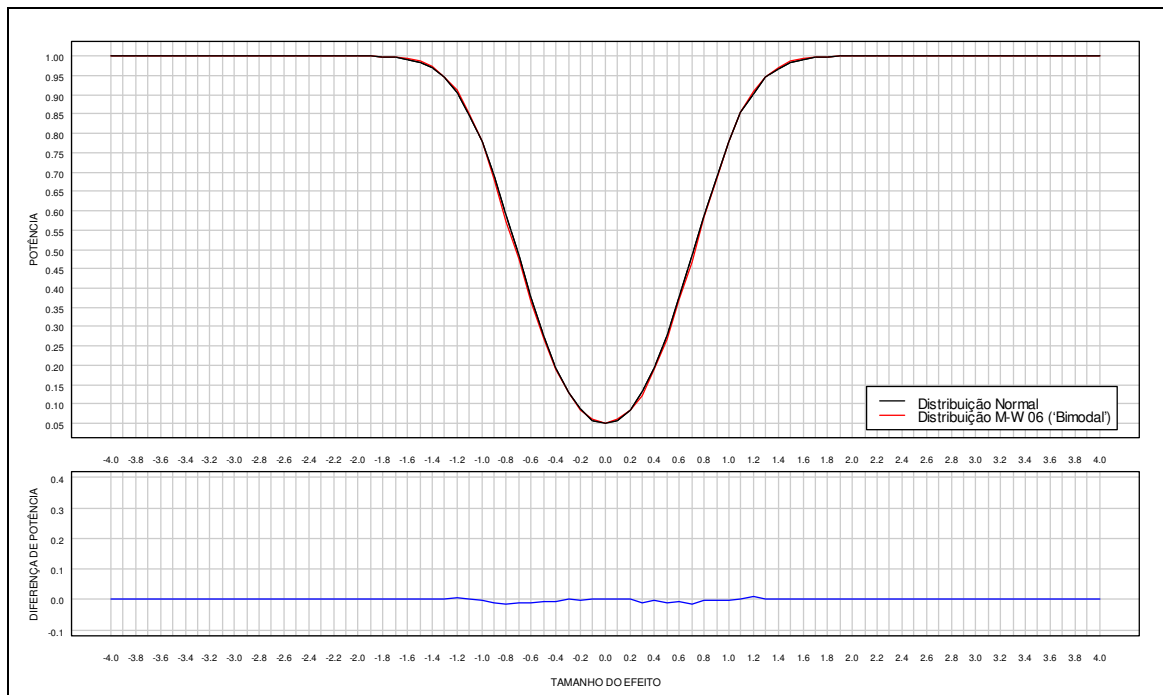


Figura Anexo VI. 33 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 06 ('Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

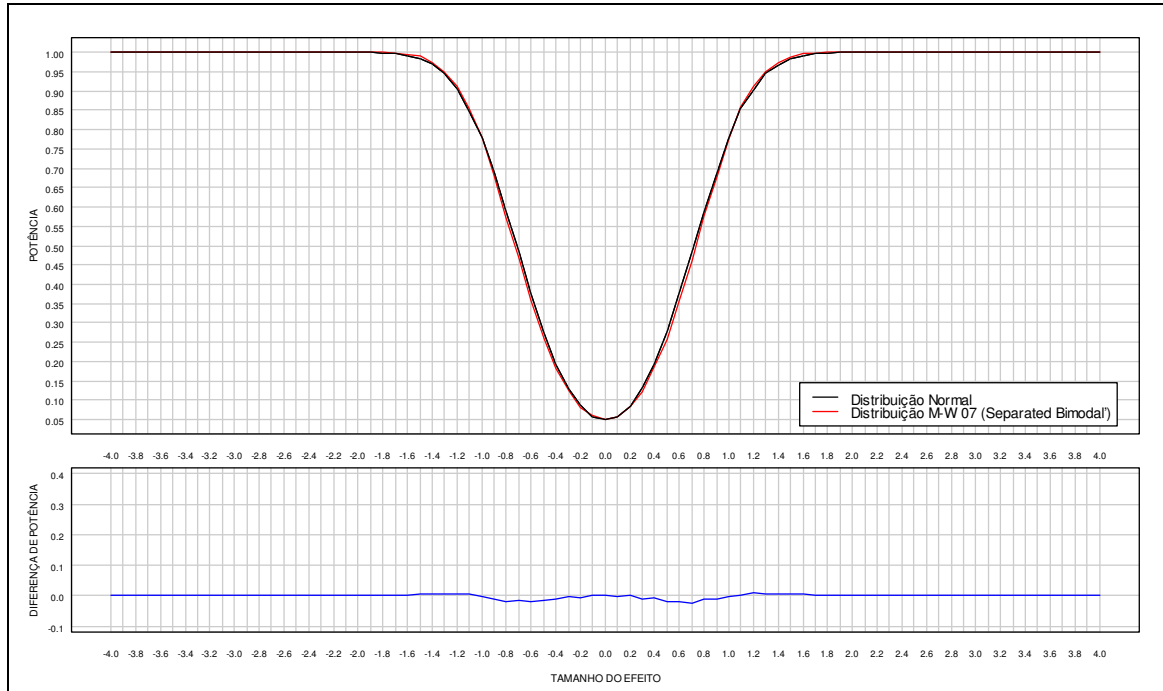


Figura Anexo VI. 34 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 07 ('Separated Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

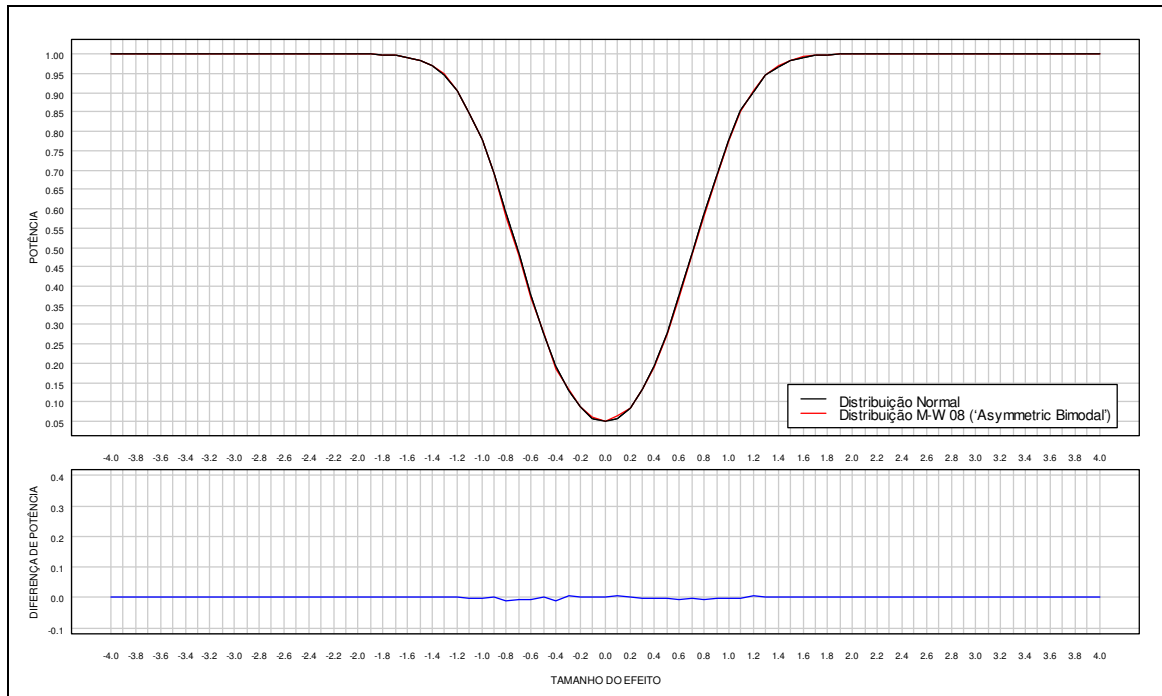


Figura Anexo VI. 35 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 08 ('Asymmetric Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

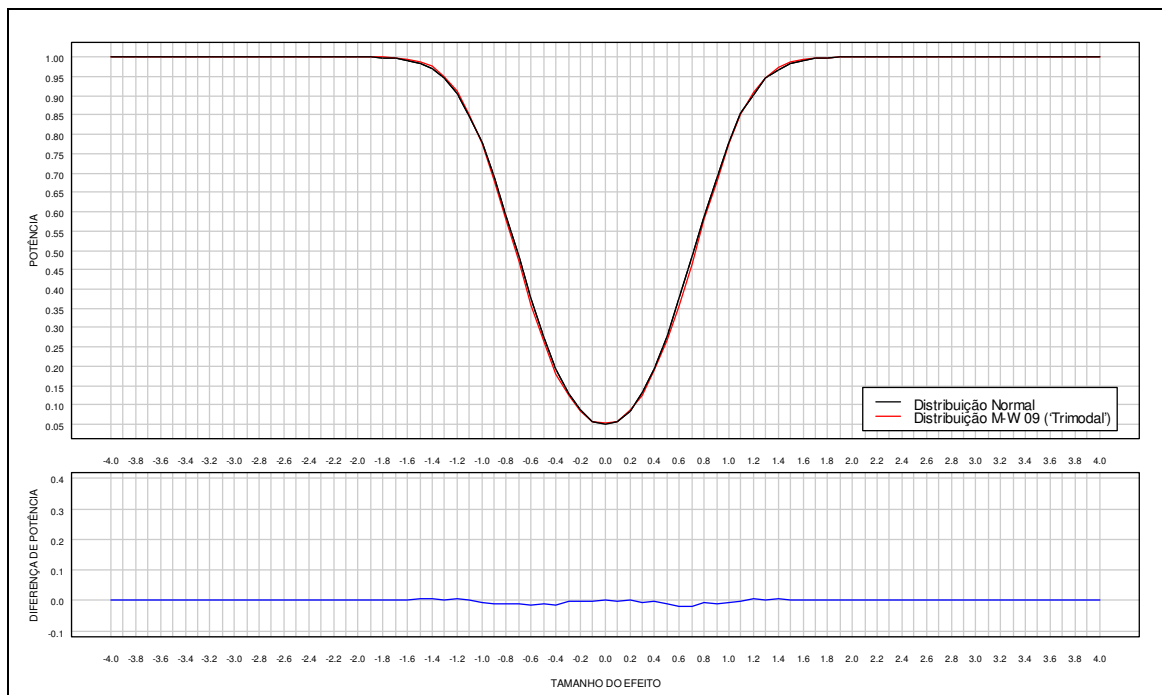


Figura Anexo VI. 36 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 09 ('Trimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

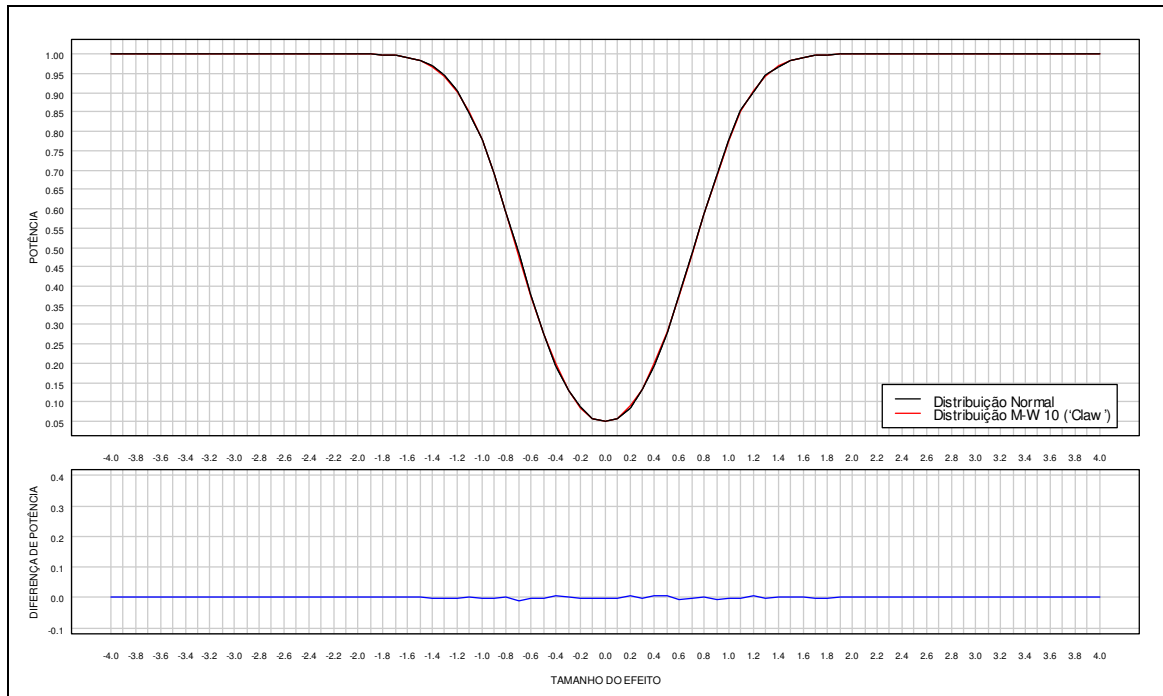


Figura Anexo VI. 37 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 10 ('Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

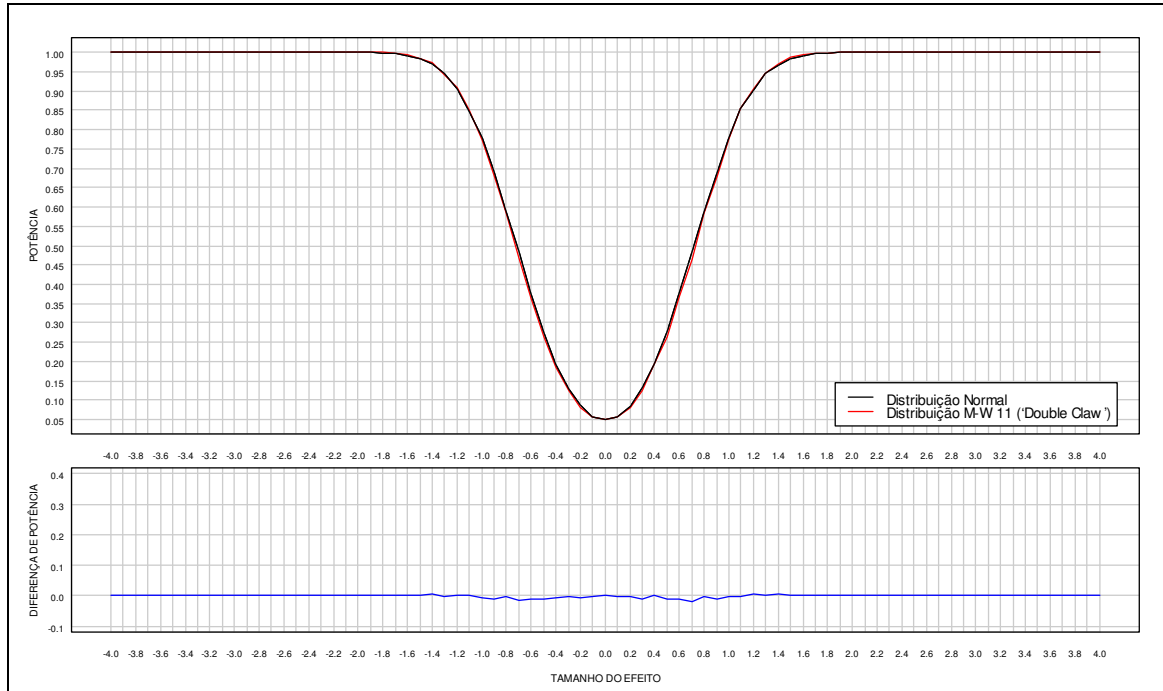


Figura Anexo VI. 38 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 11 ('Double Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

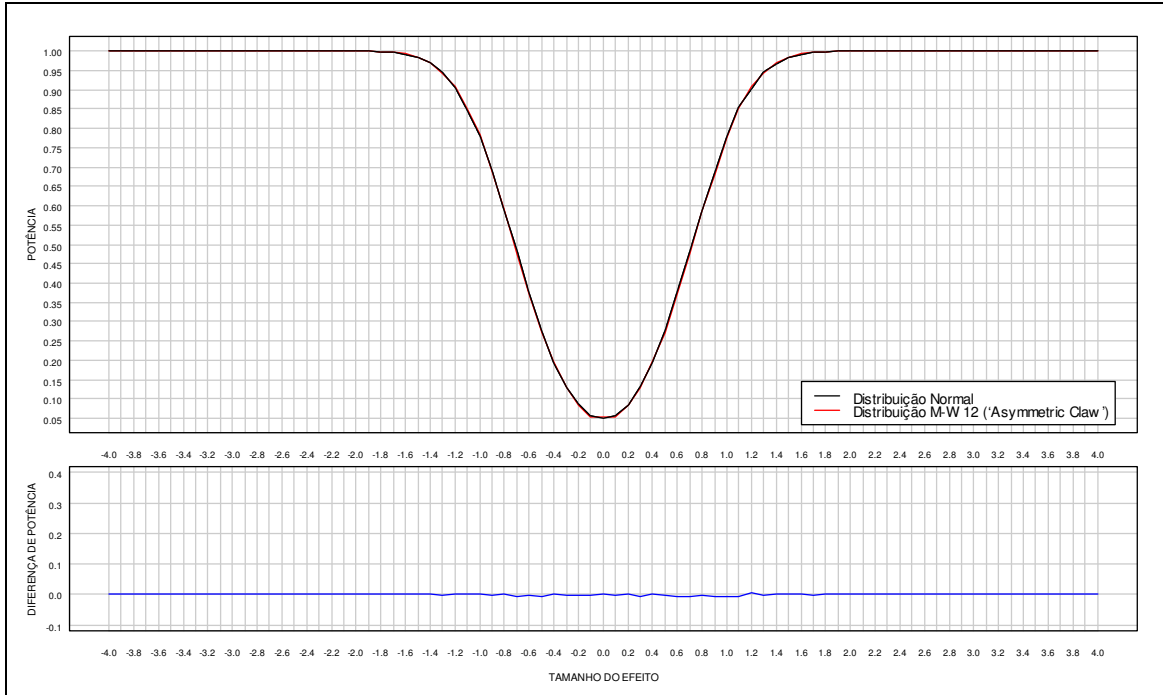


Figura Anexo VI. 39 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 12 ('Asymmetric Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

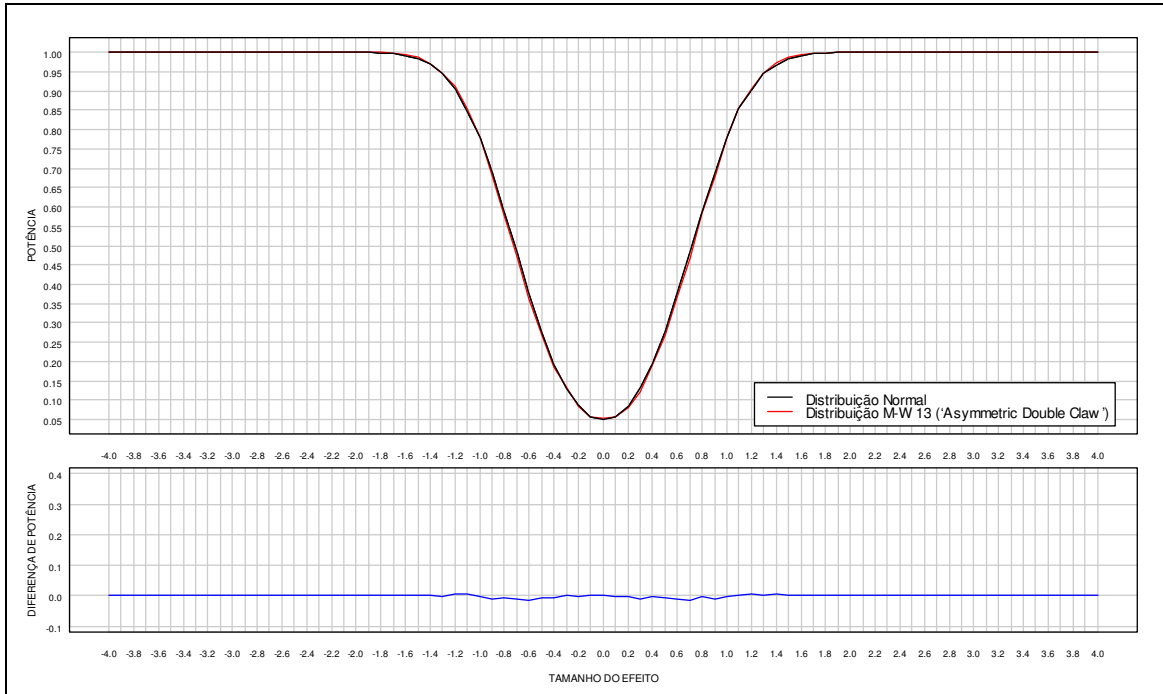


Figura Anexo VI. 40 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 13 ('Asymmetric Double Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

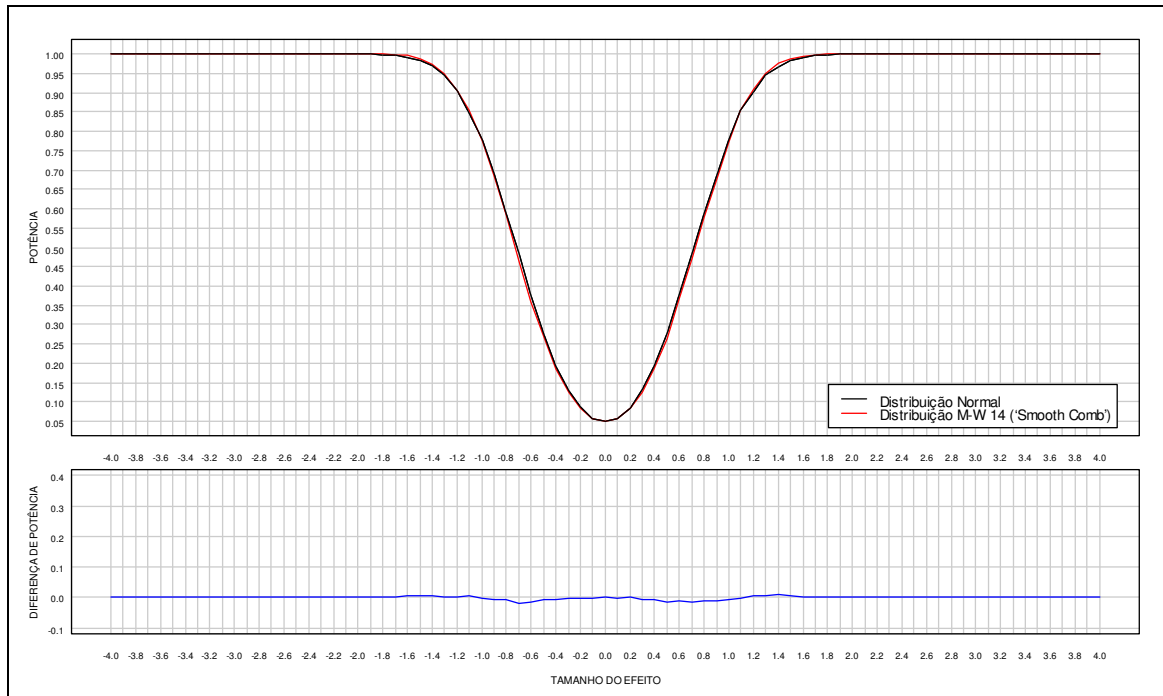


Figura Anexo VI. 41 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 14 ('Smooth Comb') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

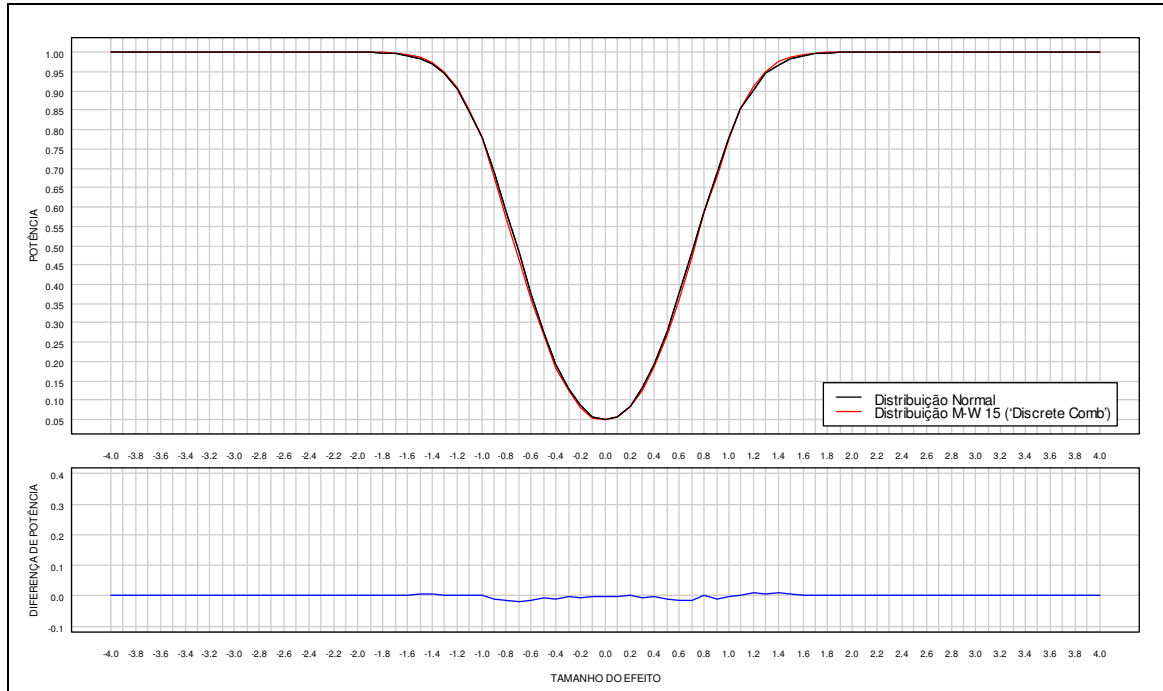


Figura Anexo VI. 42 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 15 ('Discrete Comb') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

2.2. Testes uni-caudais

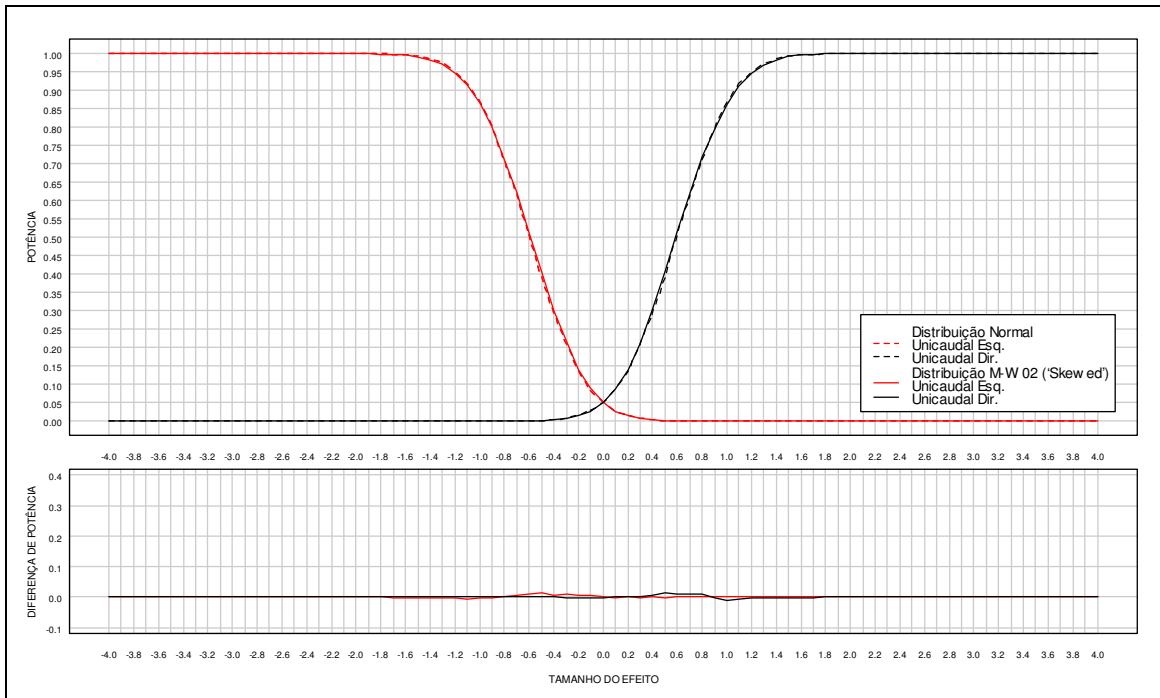


Figura Anexo VI. 43 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 02 ('Skewed') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

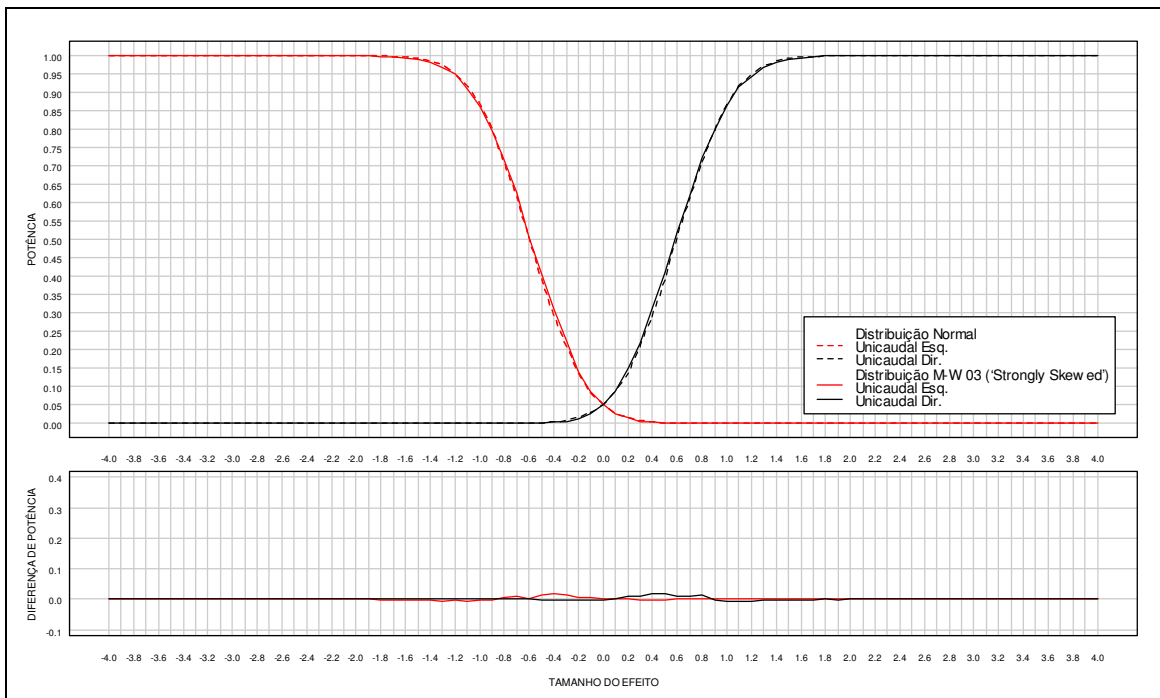


Figura Anexo VI. 44 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 03 ('Strongly Skewed') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

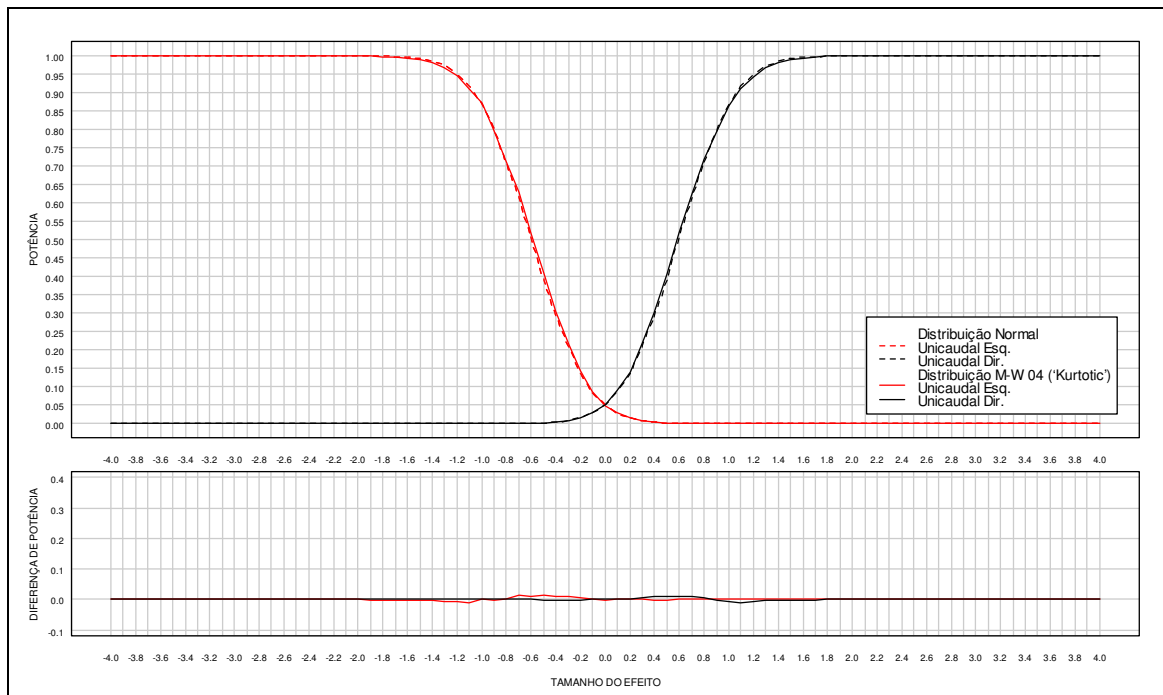


Figura Anexo VI. 45 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 04 ('Kurtotic') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

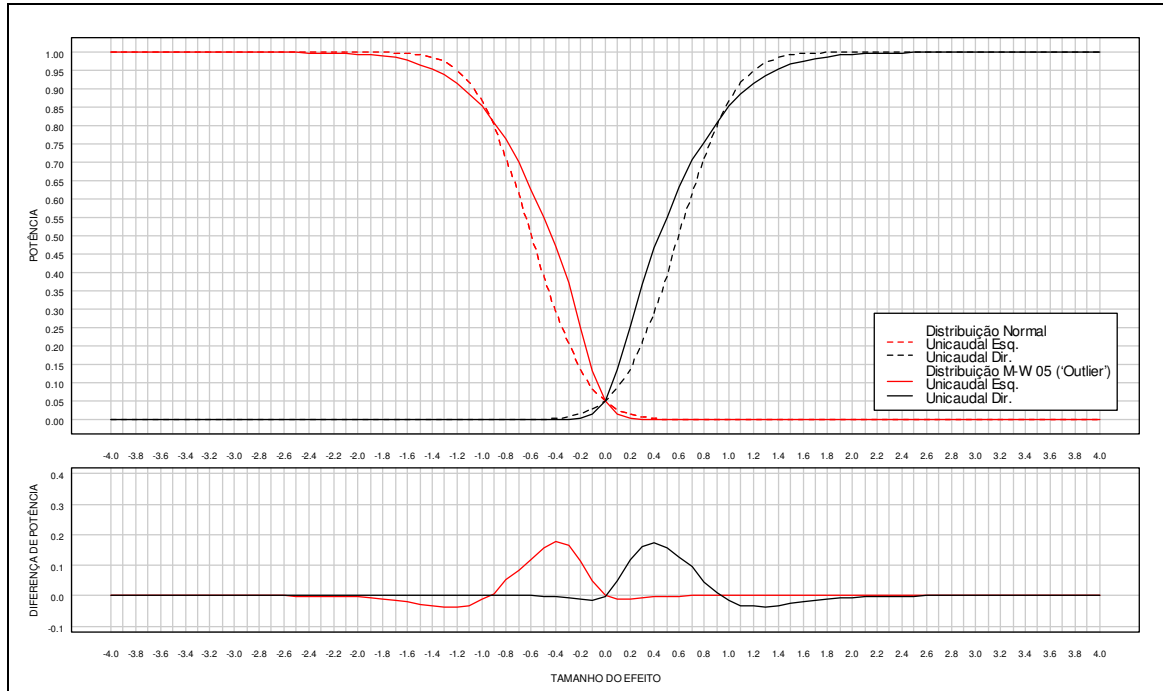


Figura Anexo VI. 46 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 05 ('Outlier') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

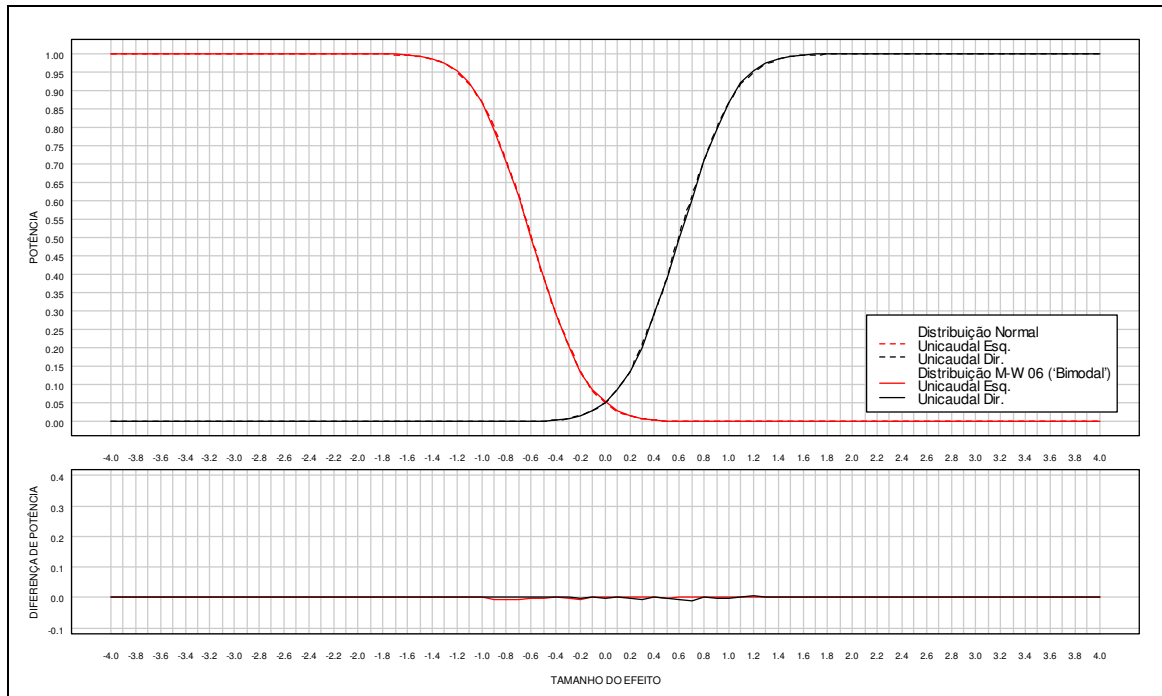


Figura Anexo VI. 47 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 06 ('Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

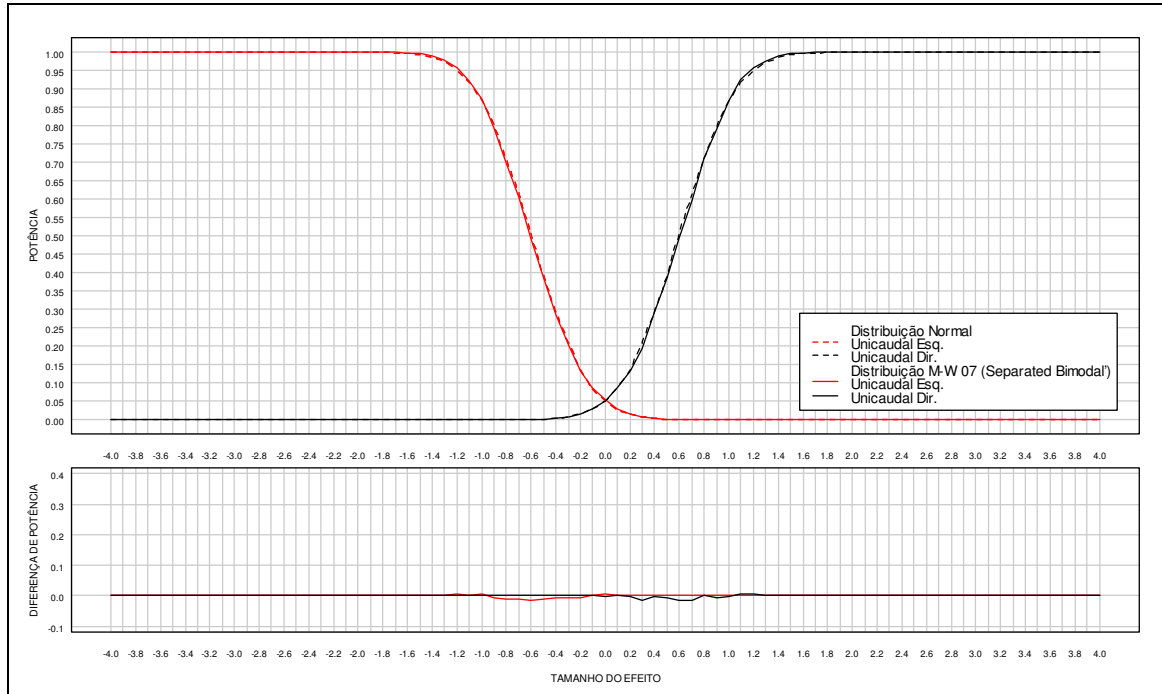


Figura Anexo VI. 48 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 07 ('Separated Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

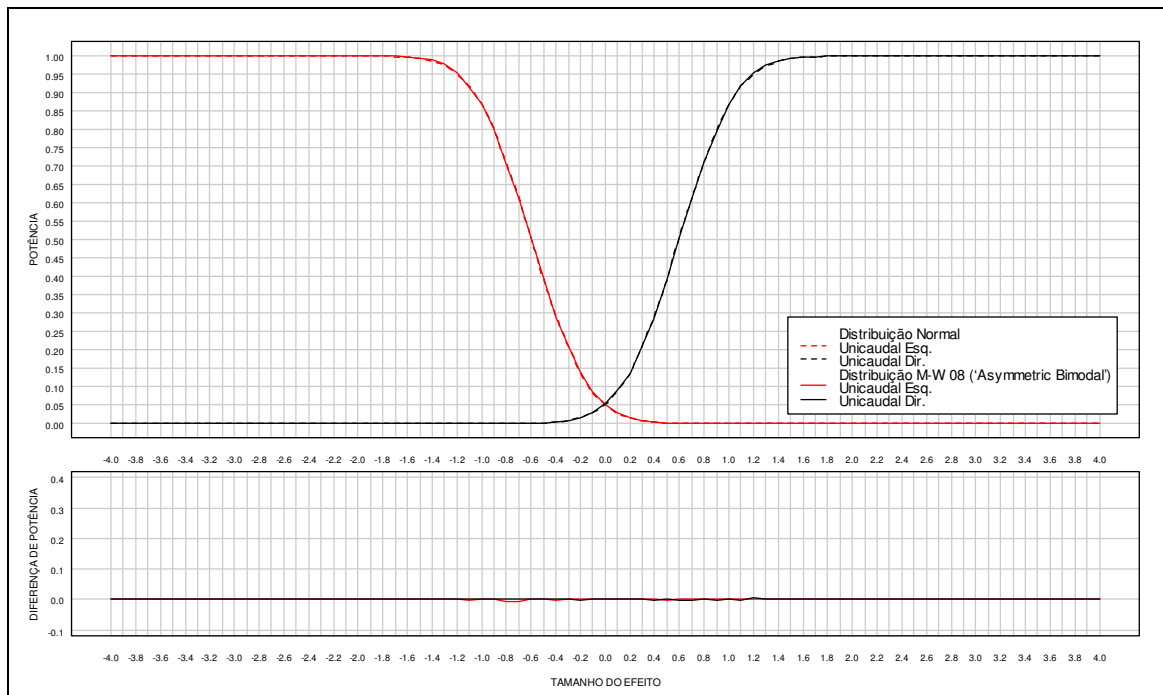


Figura Anexo VI. 49 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 08 ('Asymmetric Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

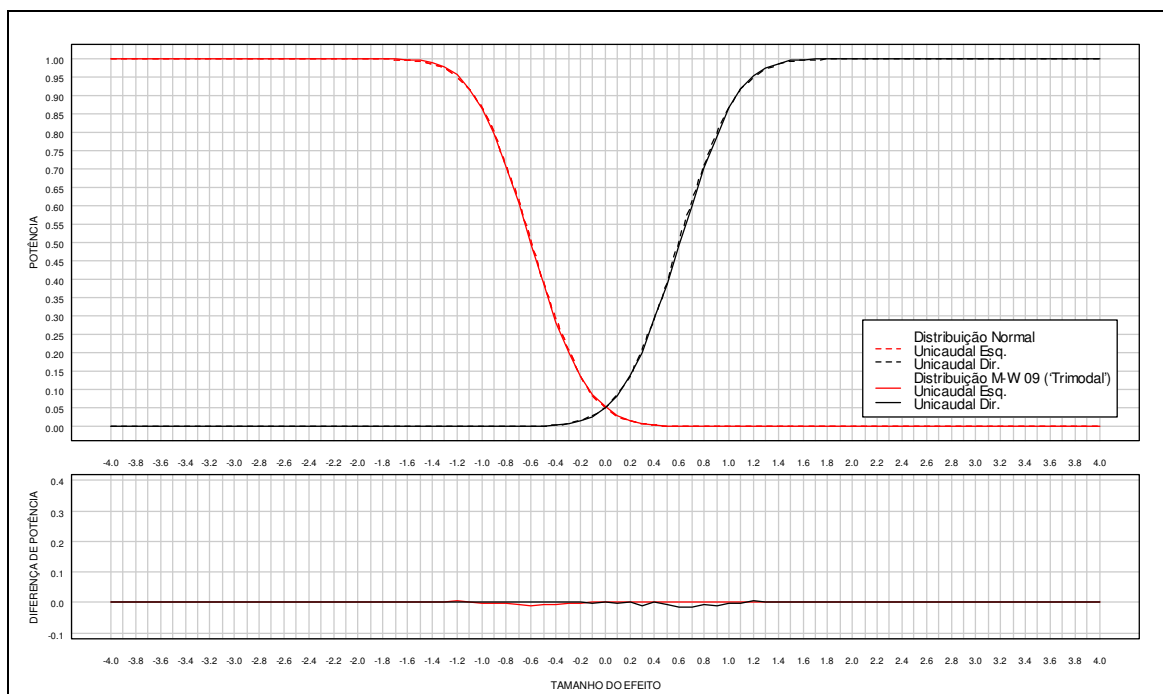


Figura Anexo VI. 50 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 09 ('Trimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

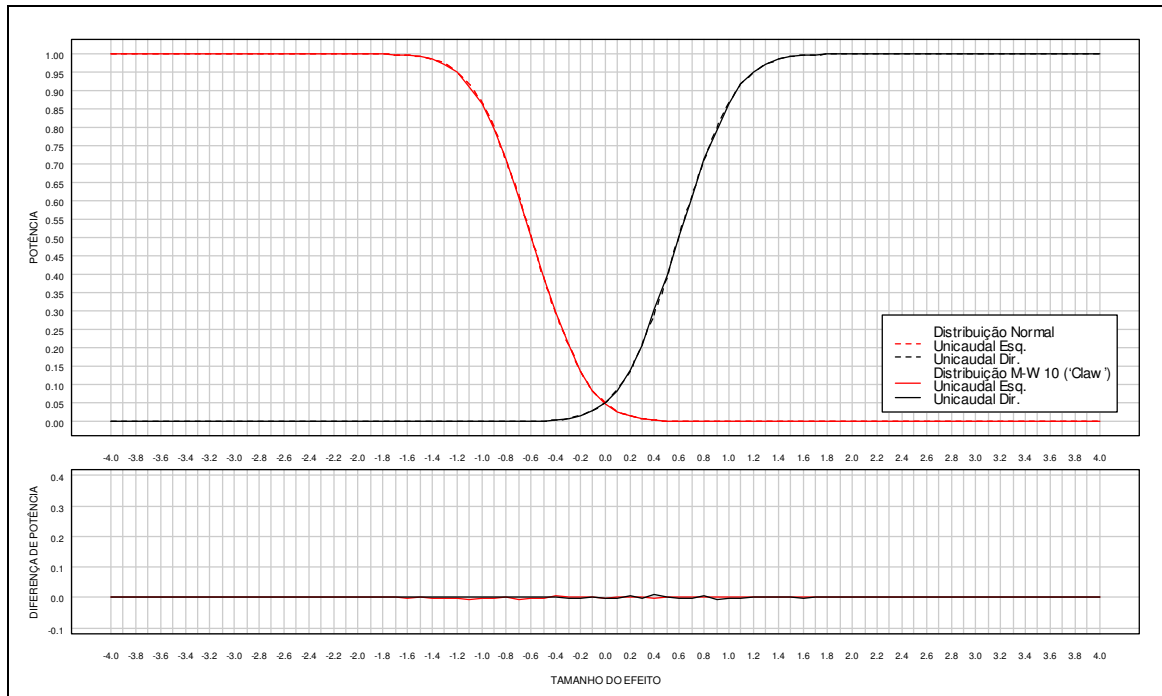


Figura Anexo VI. 51 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 10 ('Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

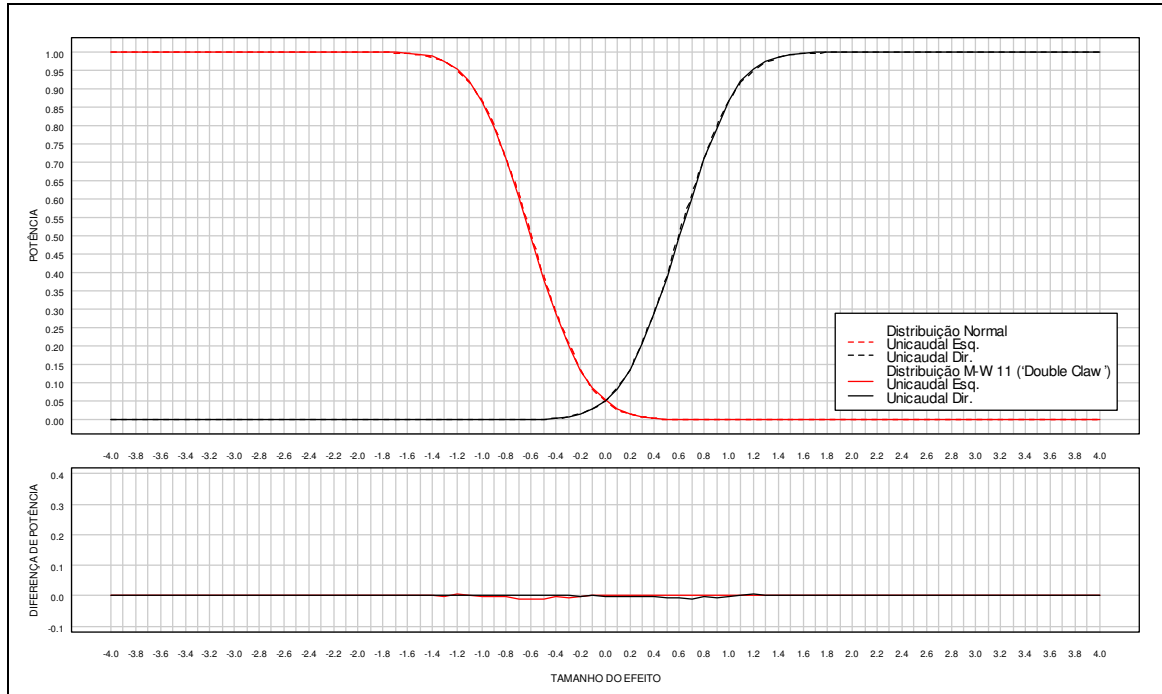


Figura Anexo VI. 52 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 11 ('Double Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

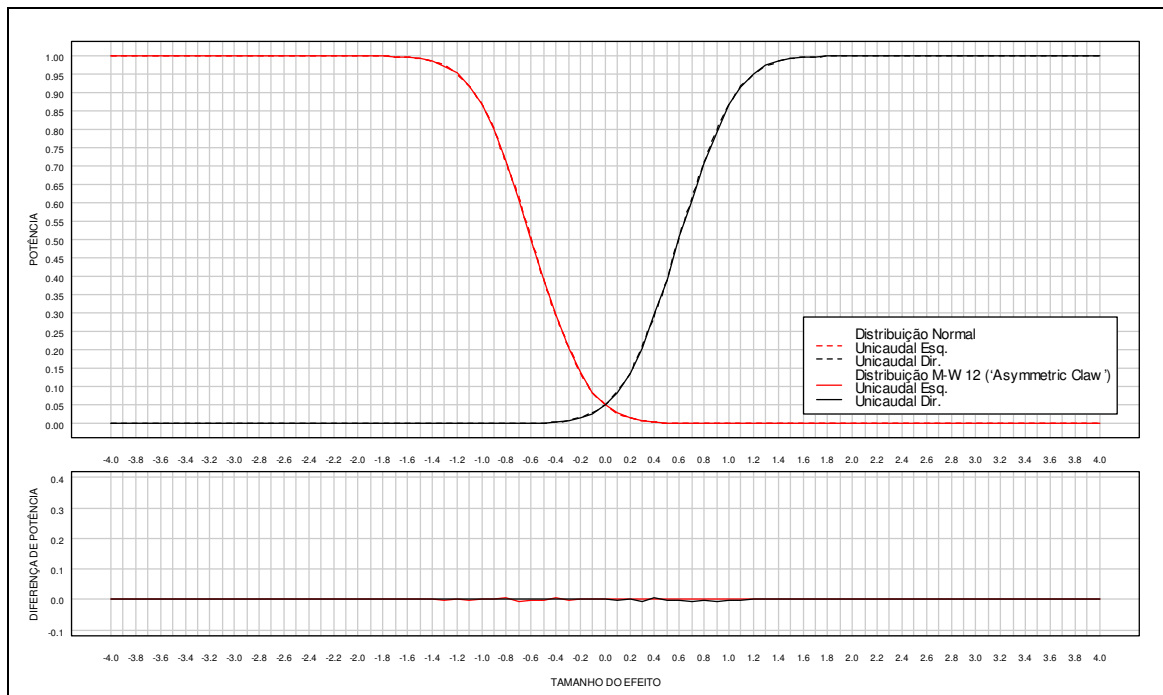


Figura Anexo VI. 53 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 12 ('Asymmetric Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

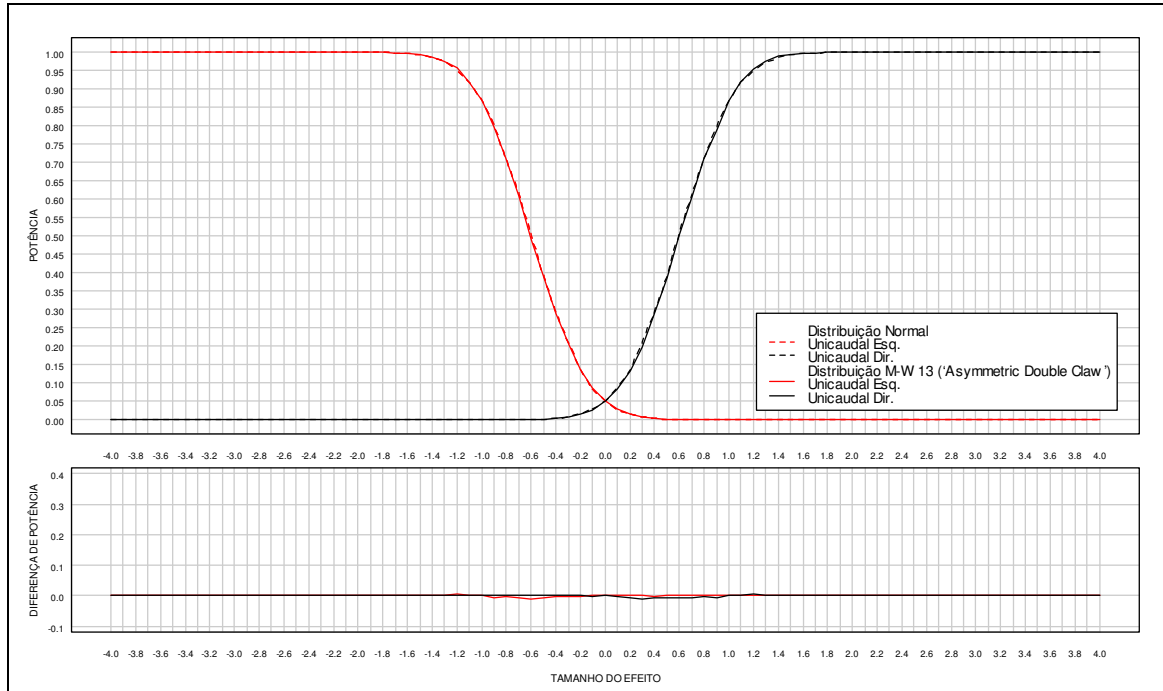


Figura Anexo VI. 54 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 13 ('Asymmetric Double Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

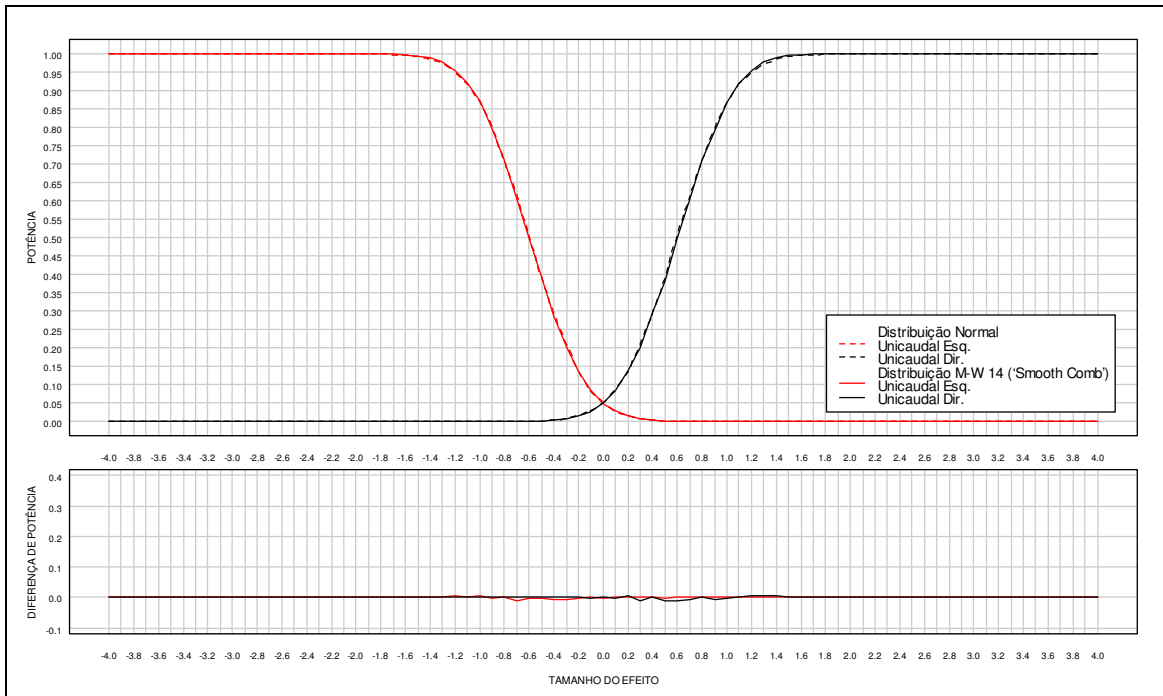


Figura Anexo VI. 55 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 14 ('Smooth Comb') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

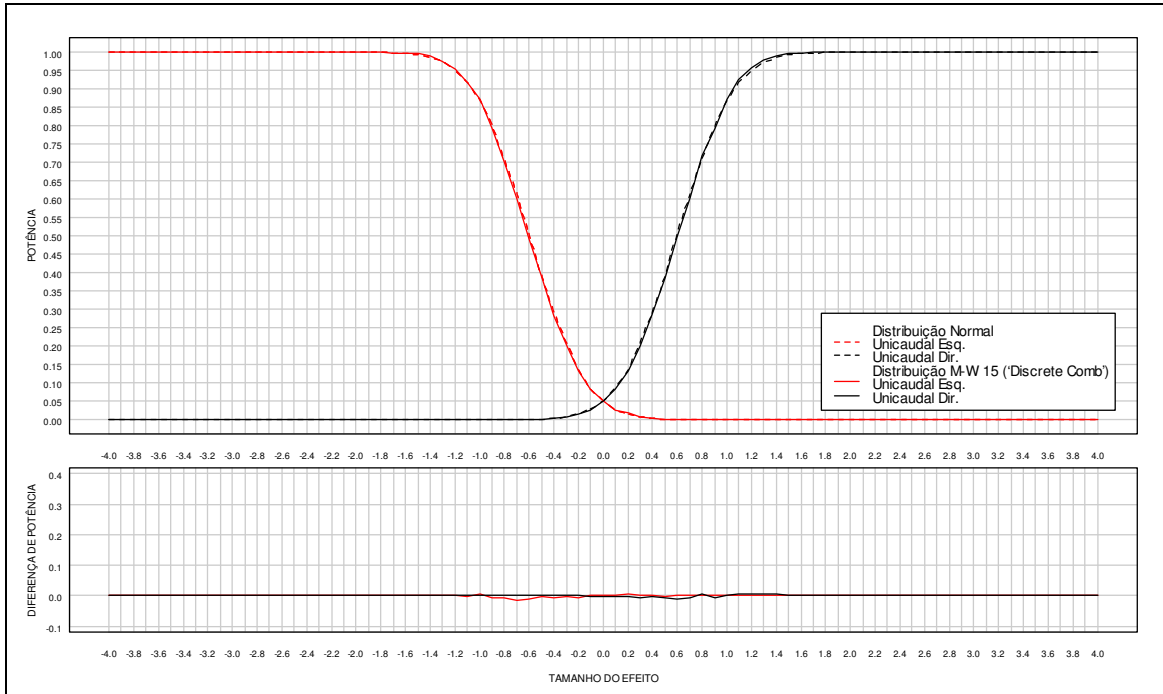


Figura Anexo VI. 56 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição Marron-Wand 15 ('Discrete Comb') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

3. Amostras com $n_1 = n_2 = 32$

3.1. Testes bi-caudais

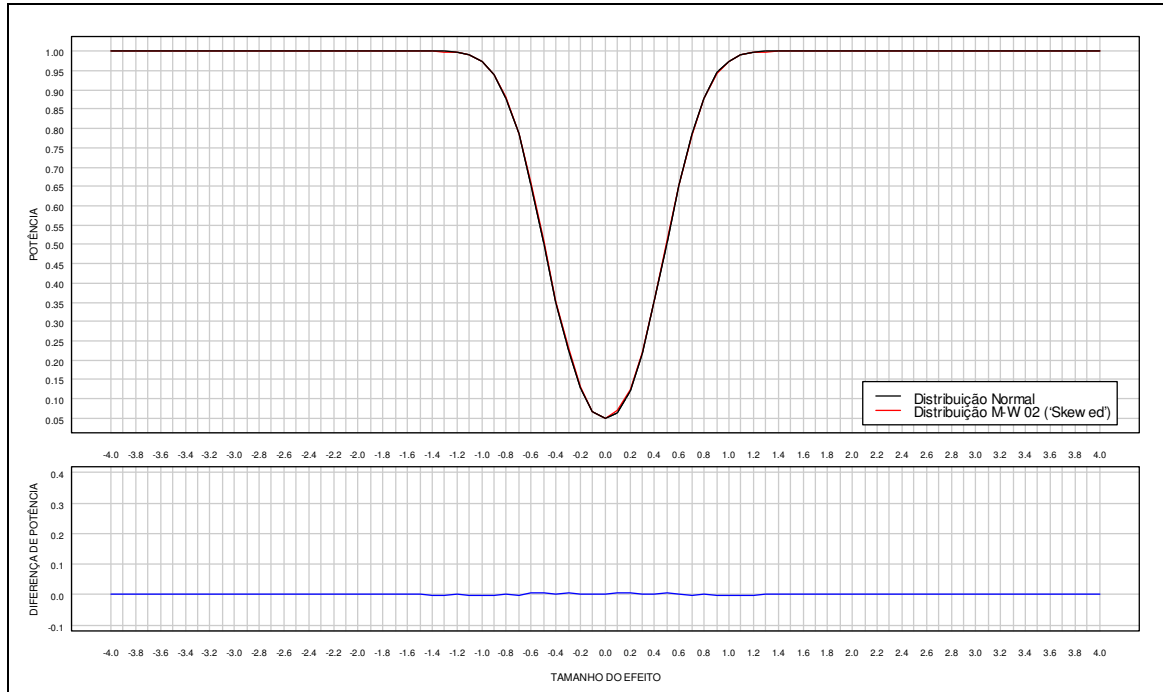


Figura Anexo VI. 57 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 02 ('Skewed') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

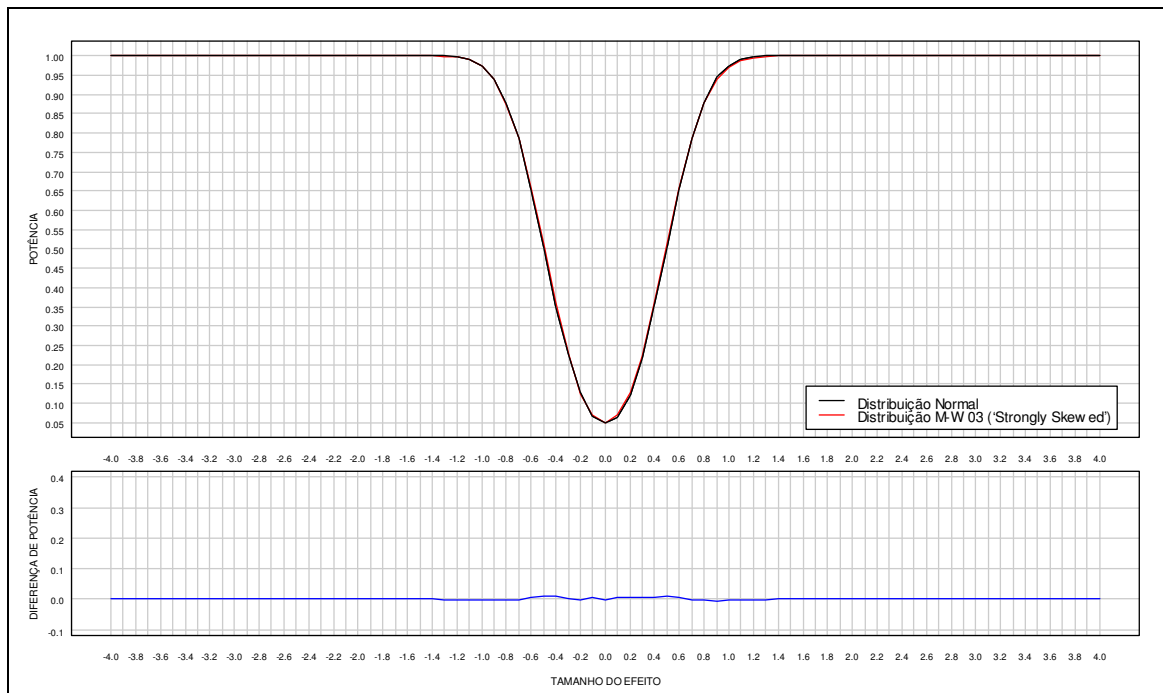


Figura Anexo VI. 58 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 03 ('Strongly Skewed') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

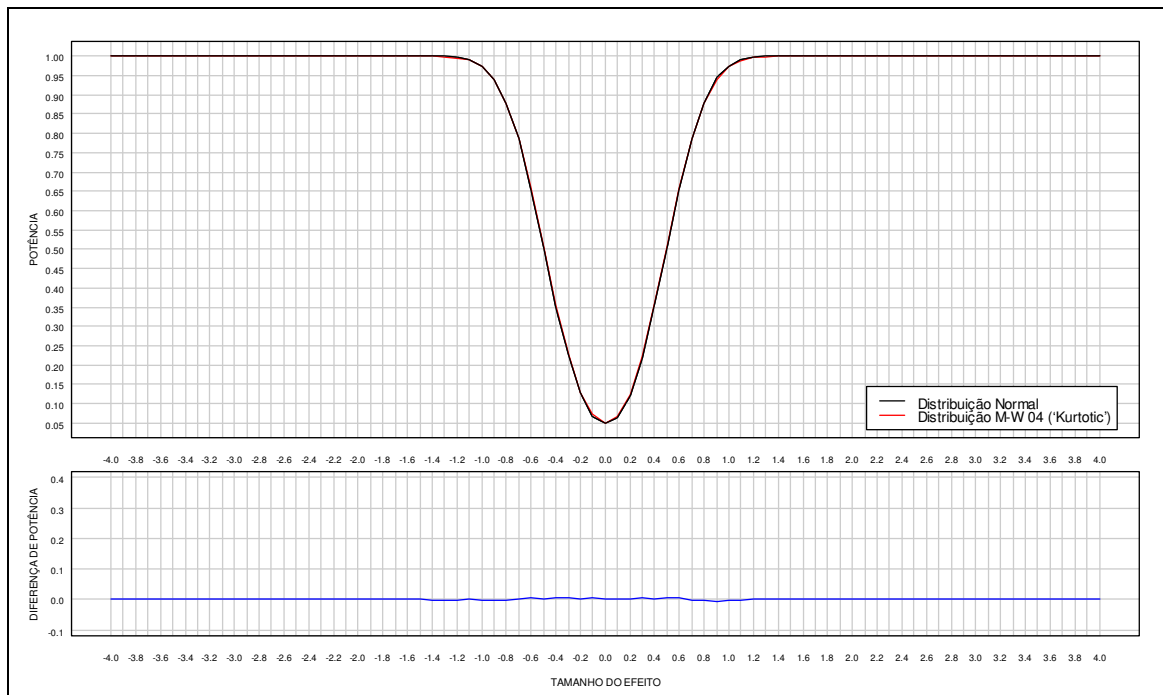


Figura Anexo VI. 59 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 04 ('Kurtotic') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

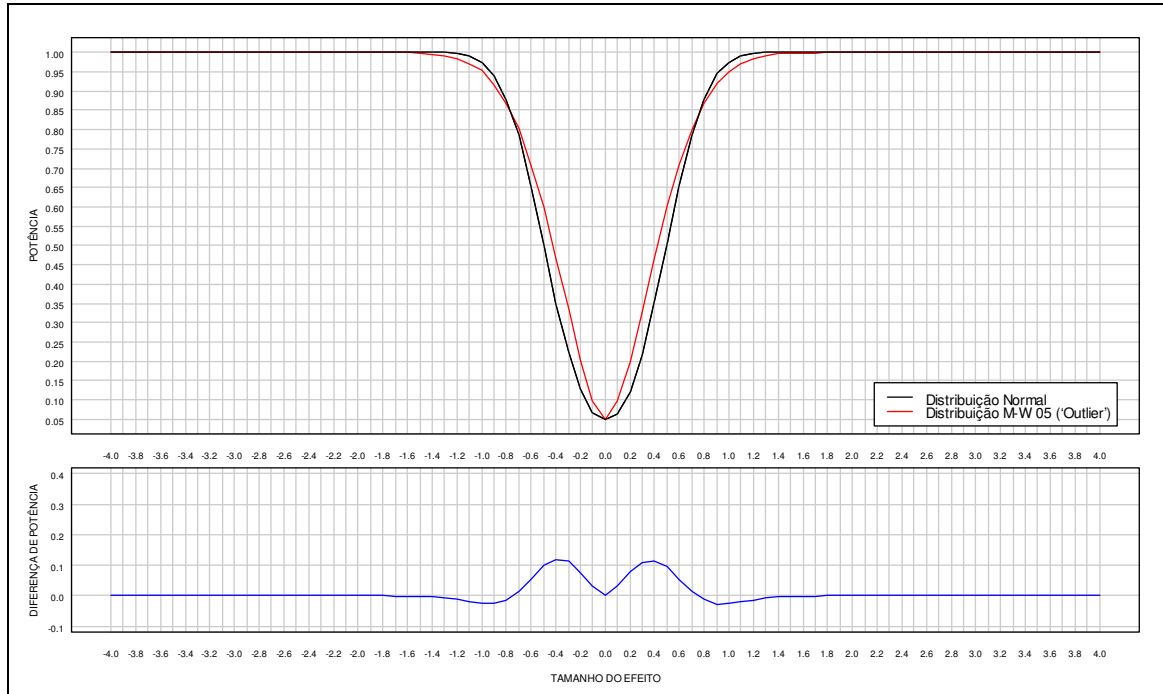


Figura Anexo VI. 60 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 05 ('Outlier') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

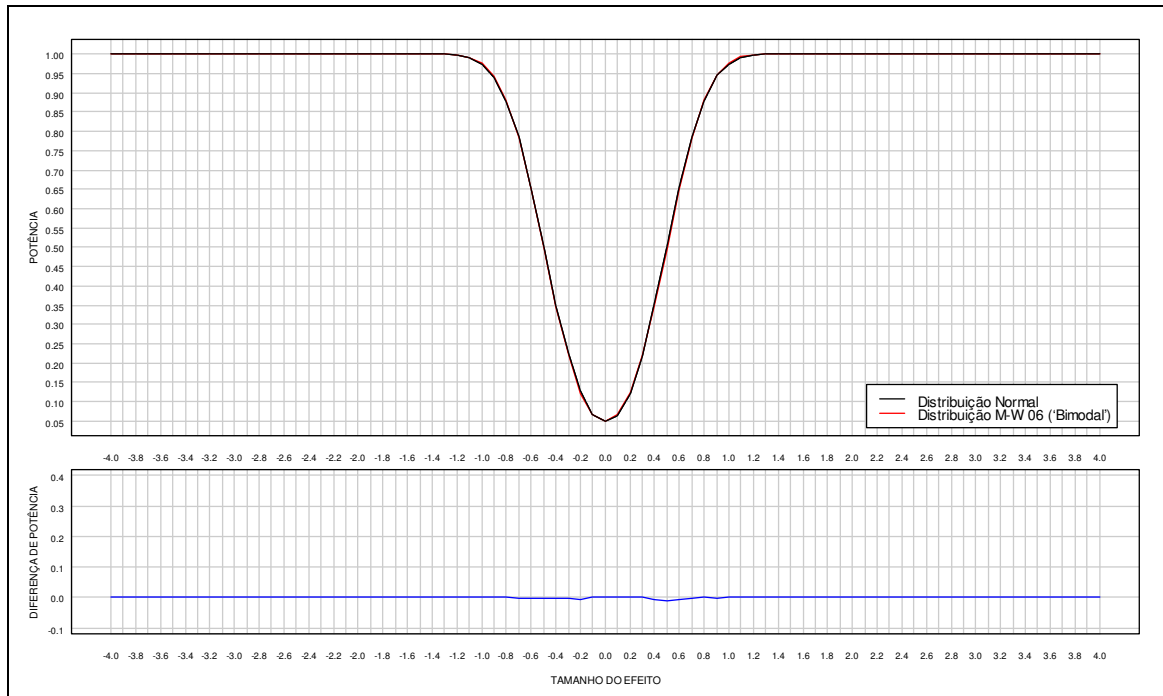


Figura Anexo VI. 61 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 06 ('Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

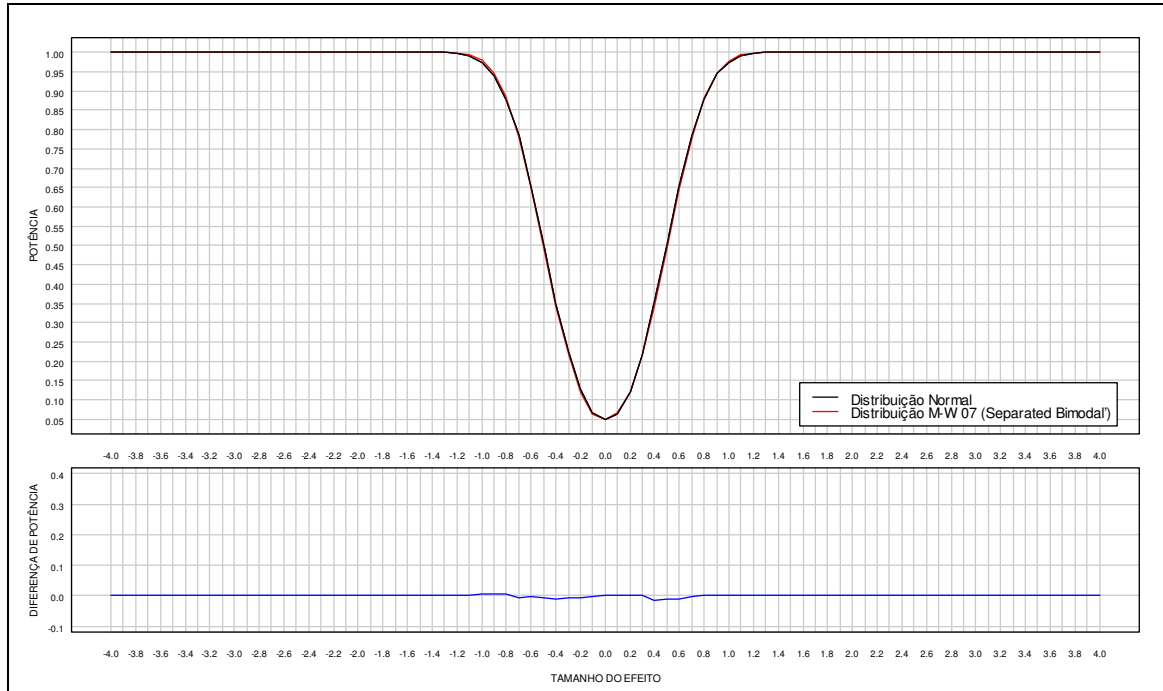


Figura Anexo VI. 62 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 07 ('Separated Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

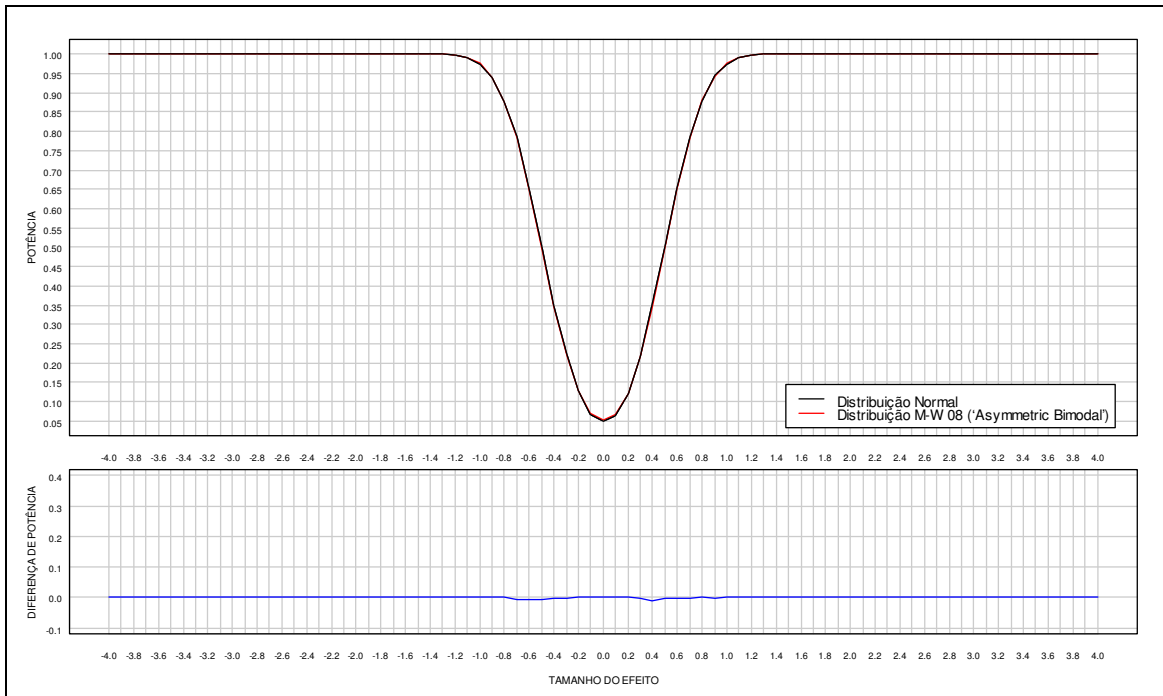


Figura Anexo VI. 63 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 08 ('Asymmetric Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

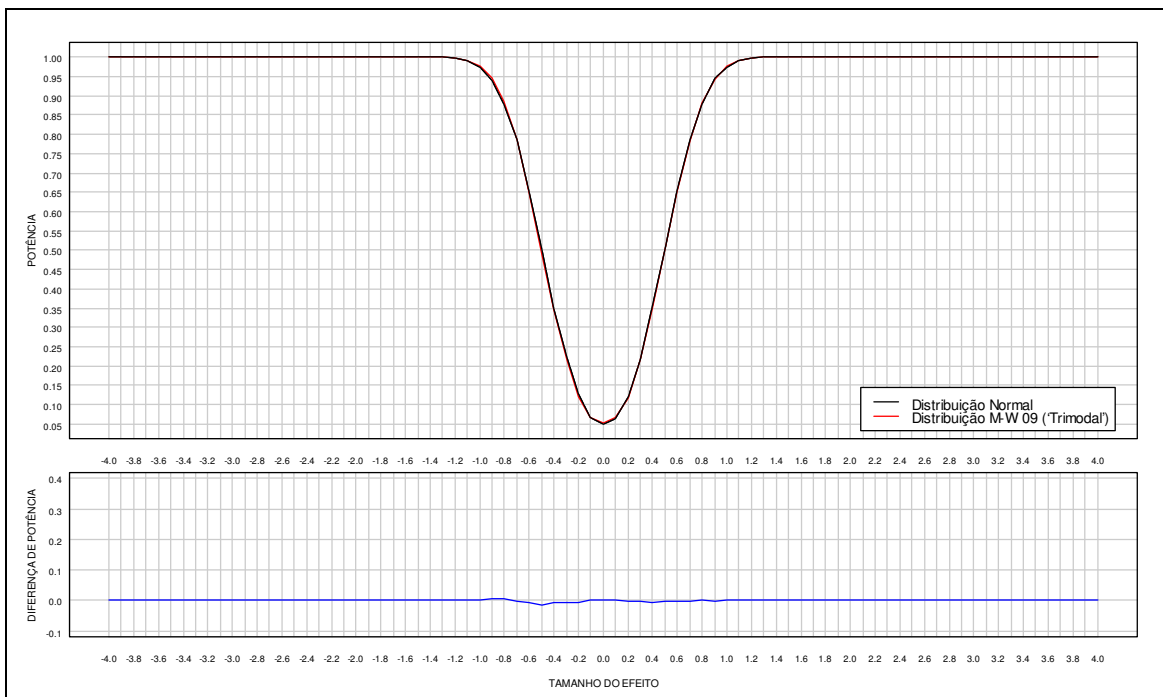


Figura Anexo VI. 64 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 09 ('Trimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

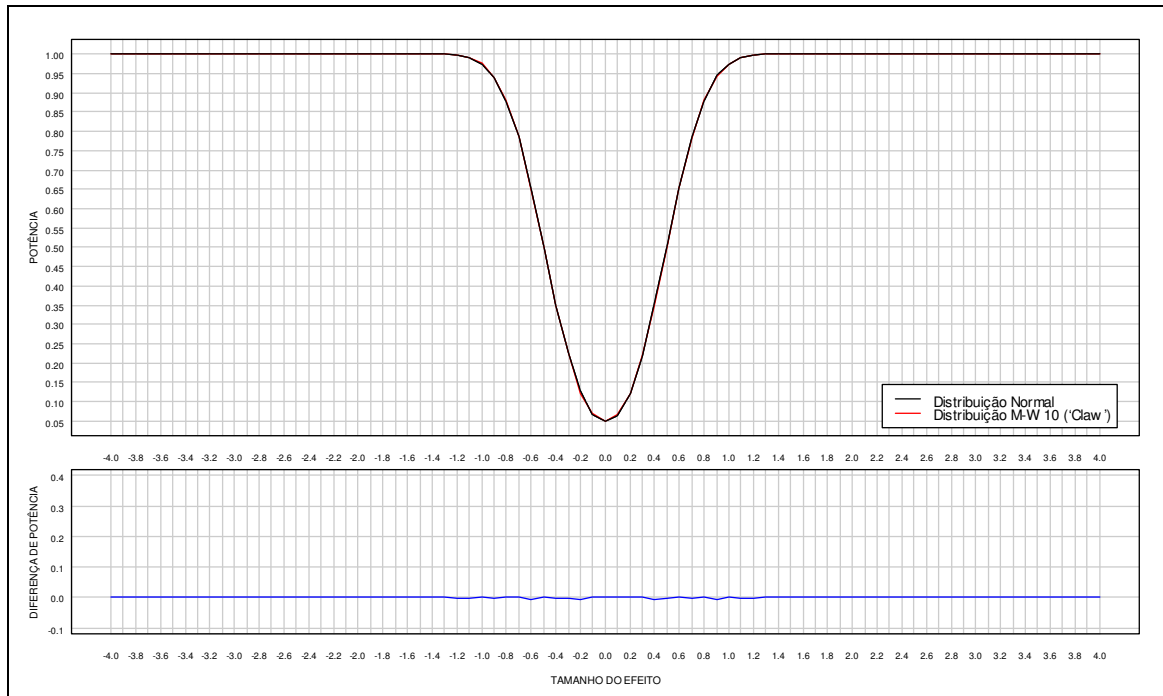


Figura Anexo VI. 65 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 10 ('Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

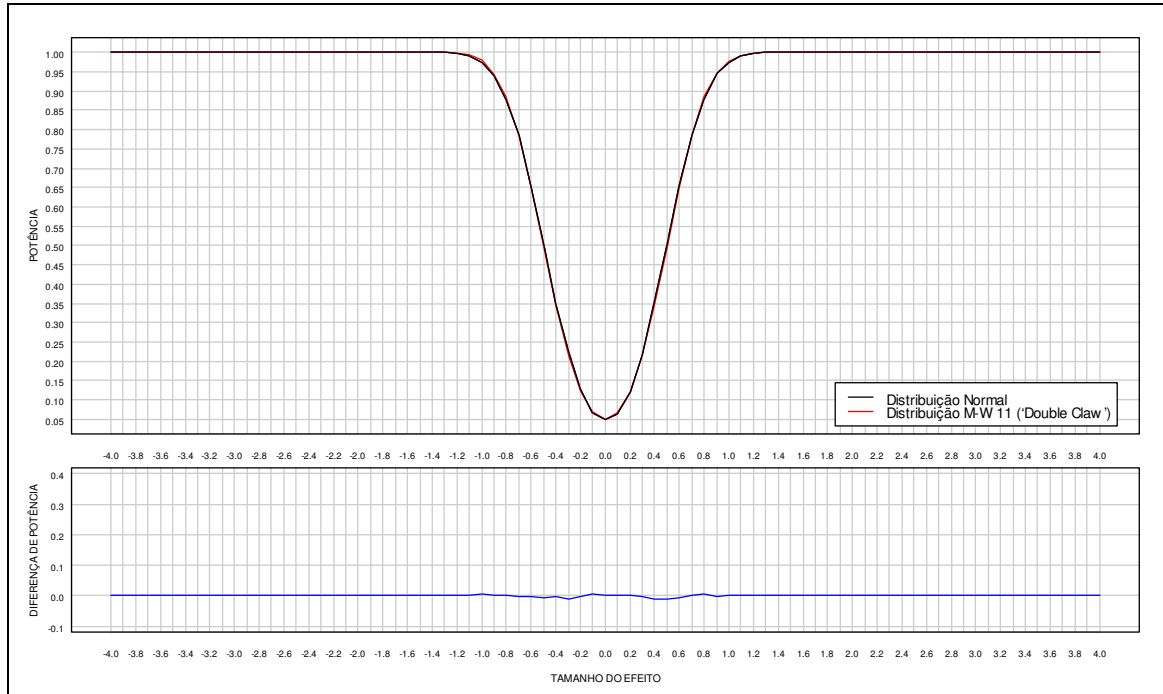


Figura Anexo VI. 66 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 11 ('Double Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

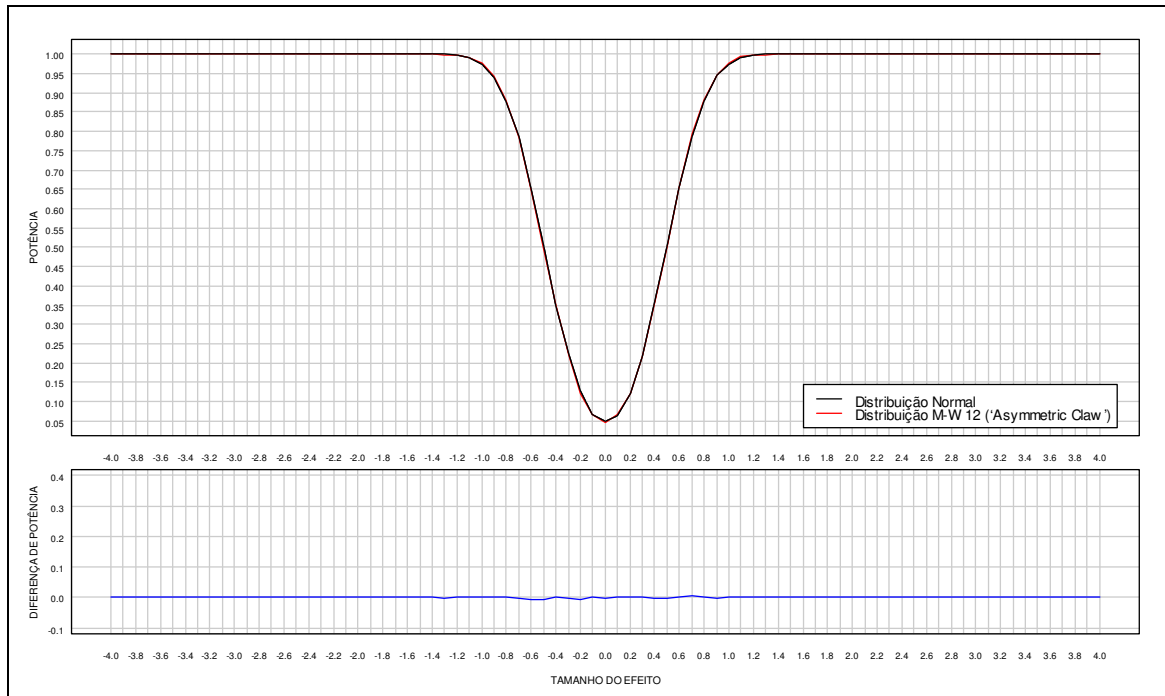


Figura Anexo VI. 67 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 12 ('Asymmetric Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

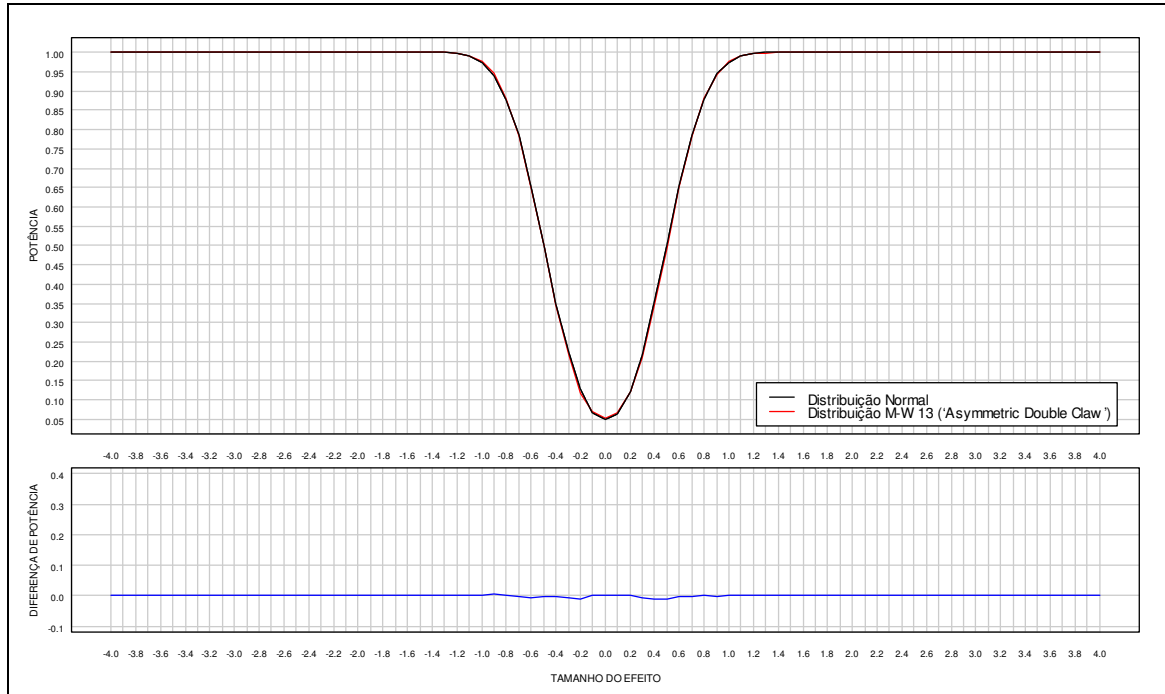


Figura Anexo VI. 68 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 13 ('Asymmetric Double Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

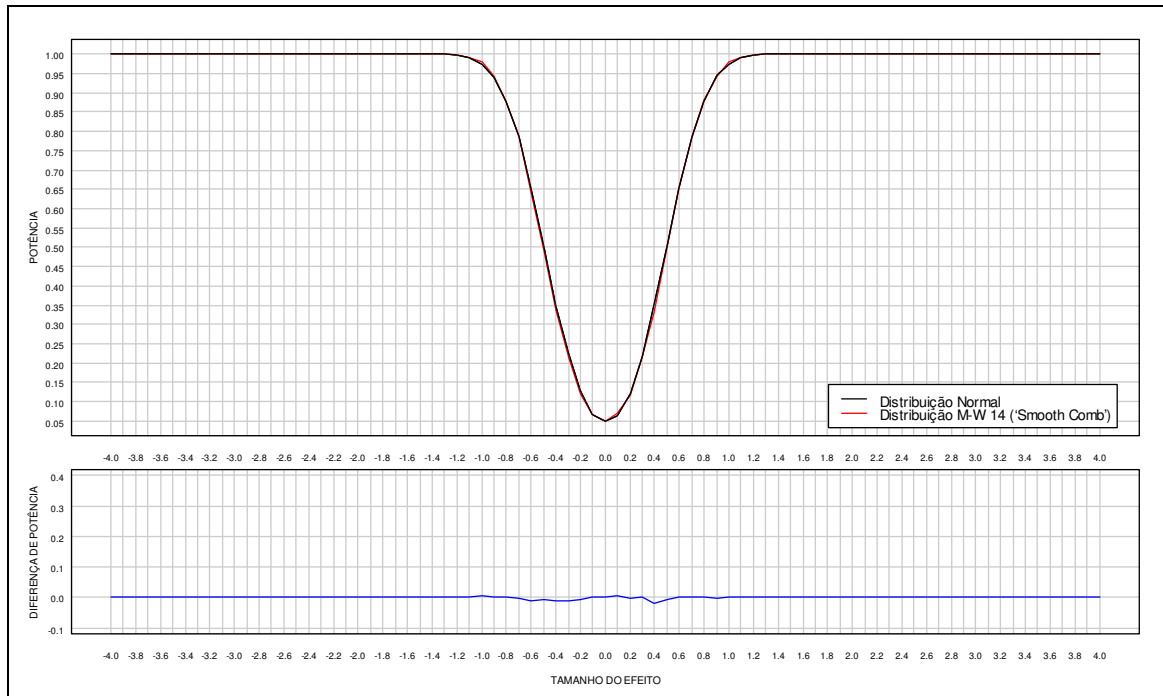


Figura Anexo VI. 69 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 14 ('Smooth Comb') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

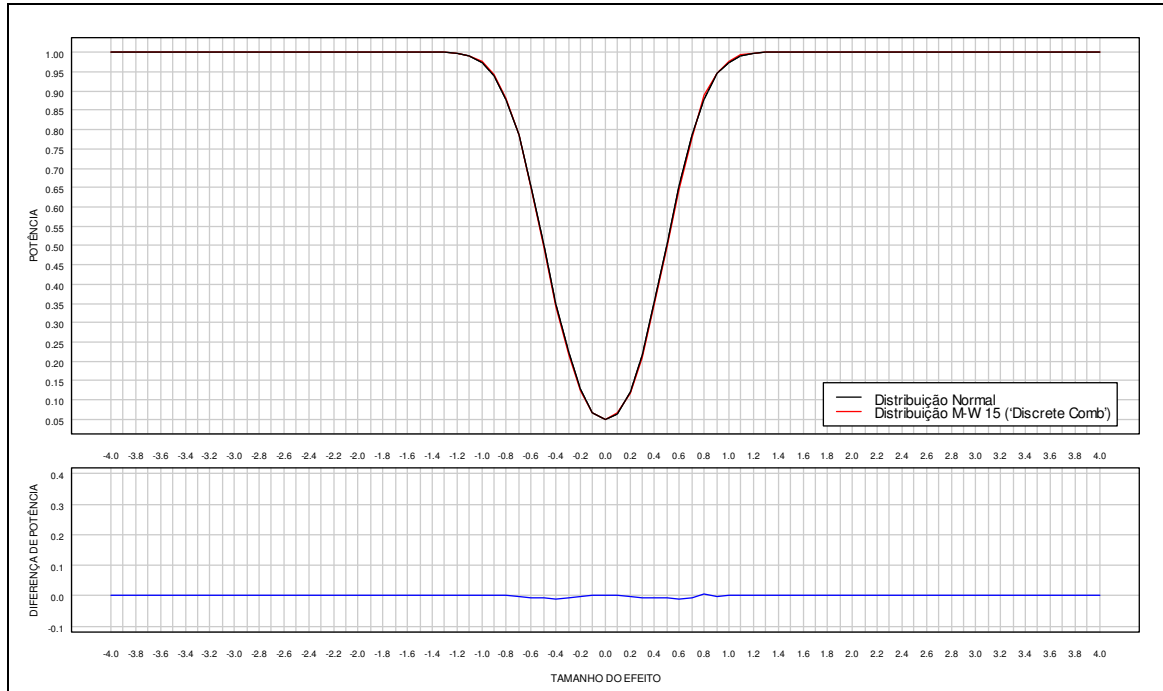


Figura Anexo VI. 70 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 15 ('Discrete Comb') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

3.2. Testes uni-caudais

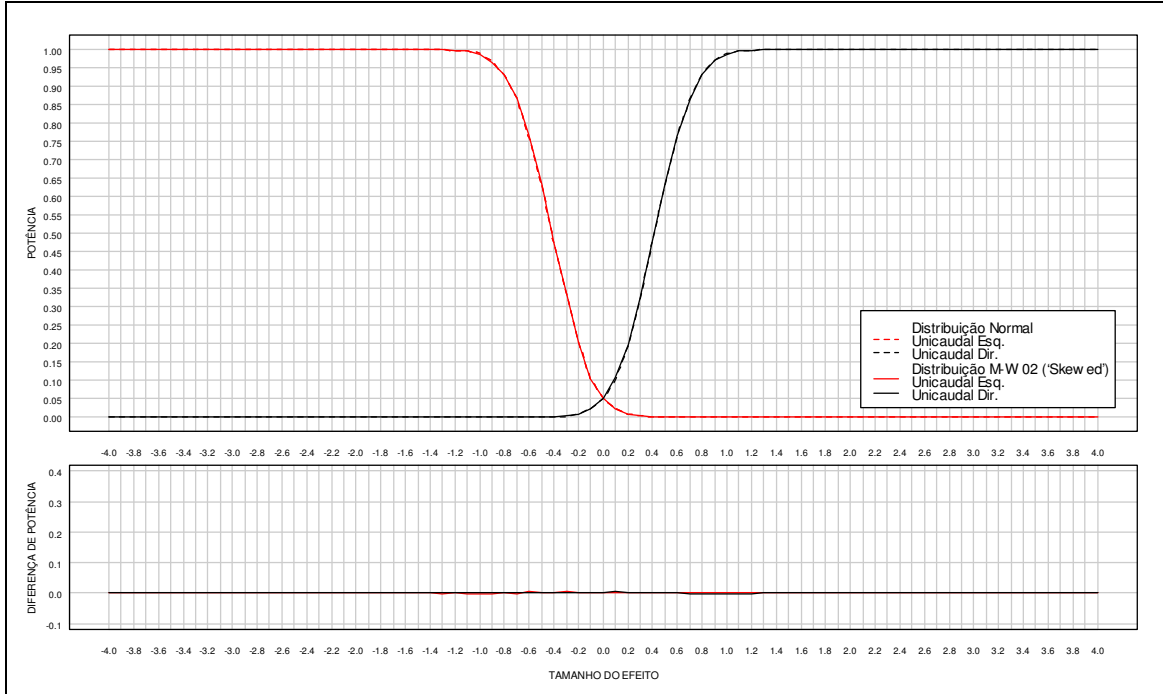


Figura Anexo VI. 71 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 02 ('Skewed') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

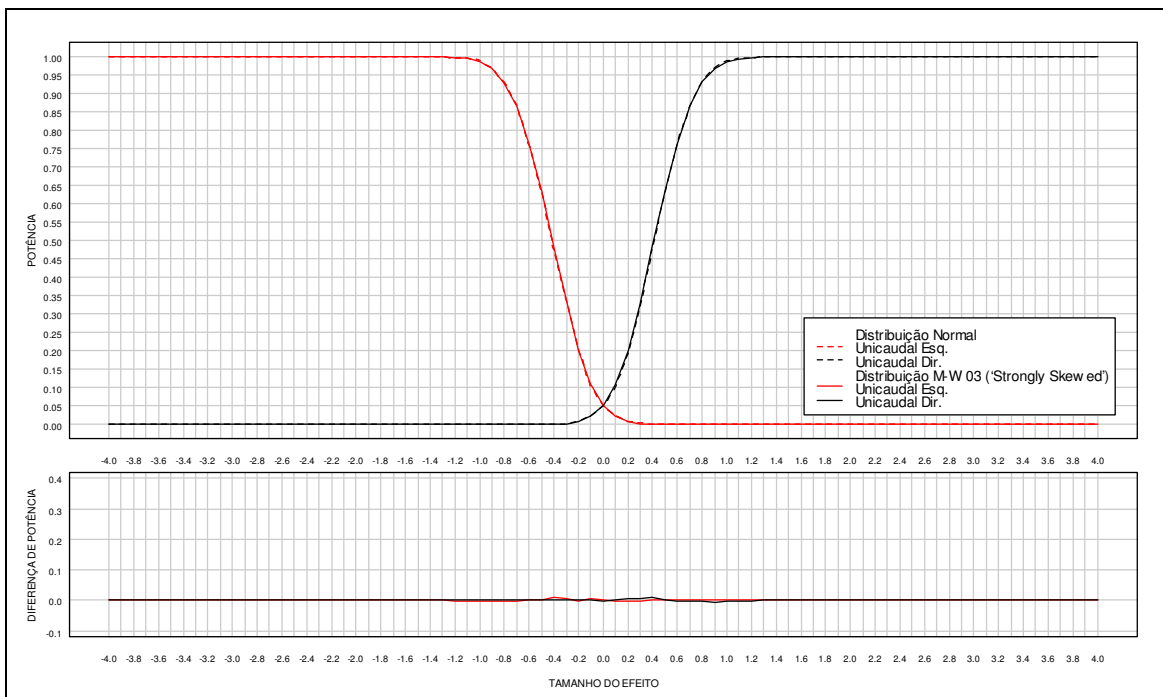


Figura Anexo VI. 72 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 03 ('Strongly Skewed') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

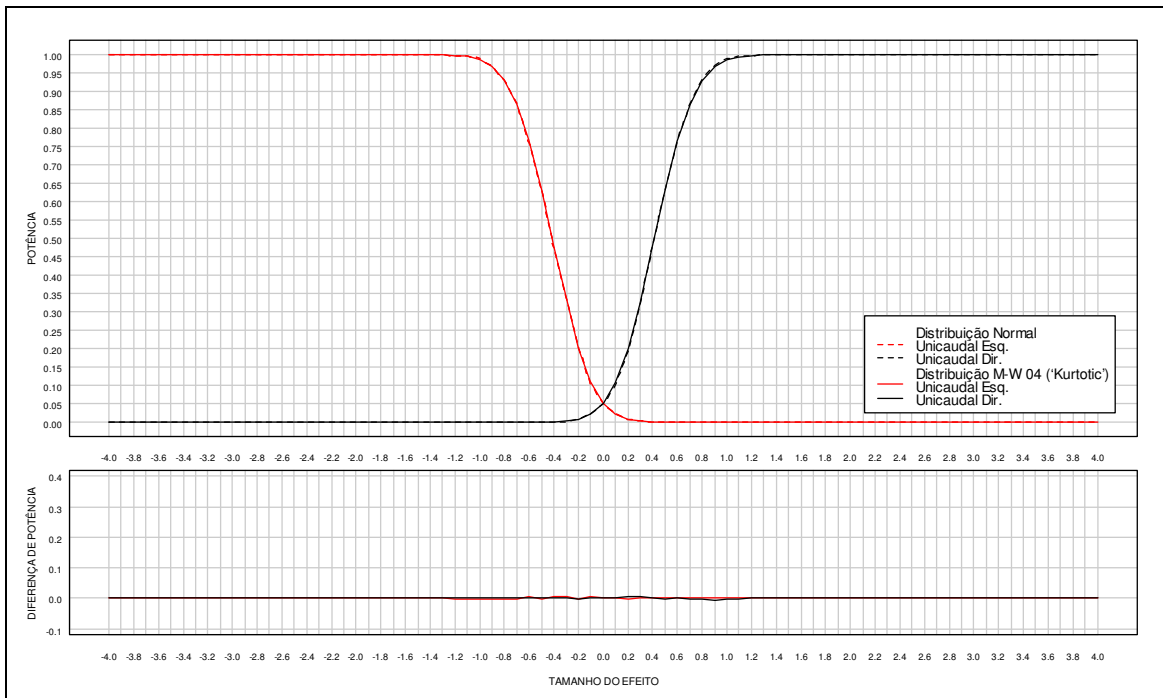


Figura Anexo VI. 73 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 04 ('Kurtotic') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

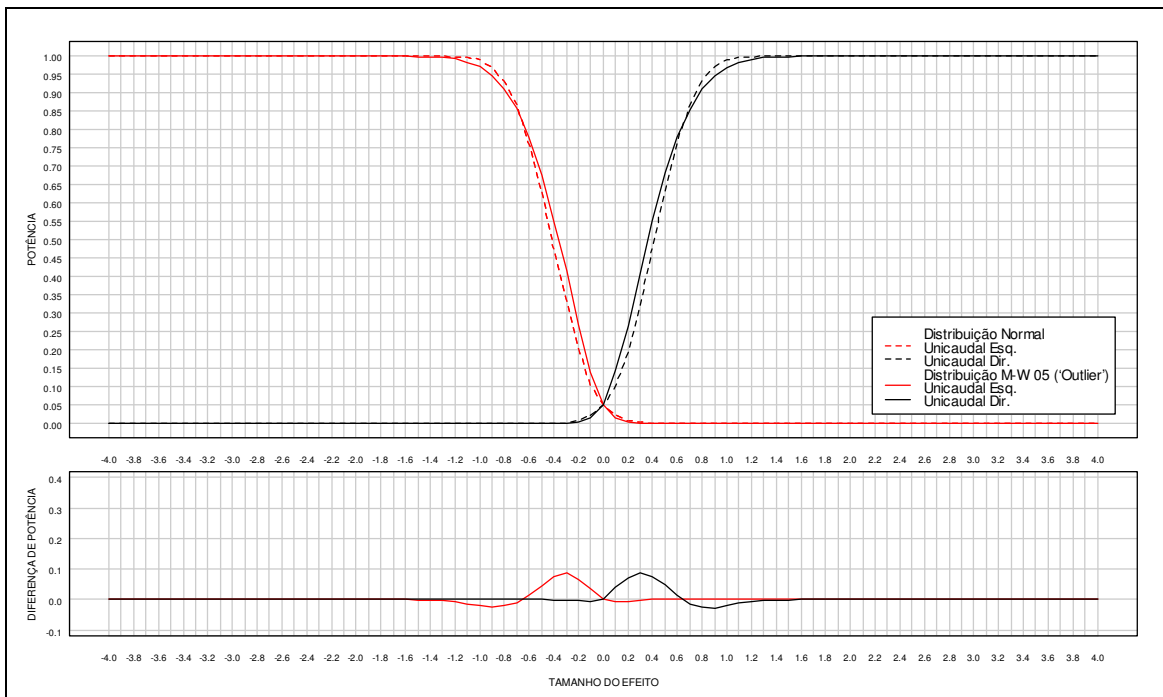


Figura Anexo VI. 74 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 05 ('Outlier') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

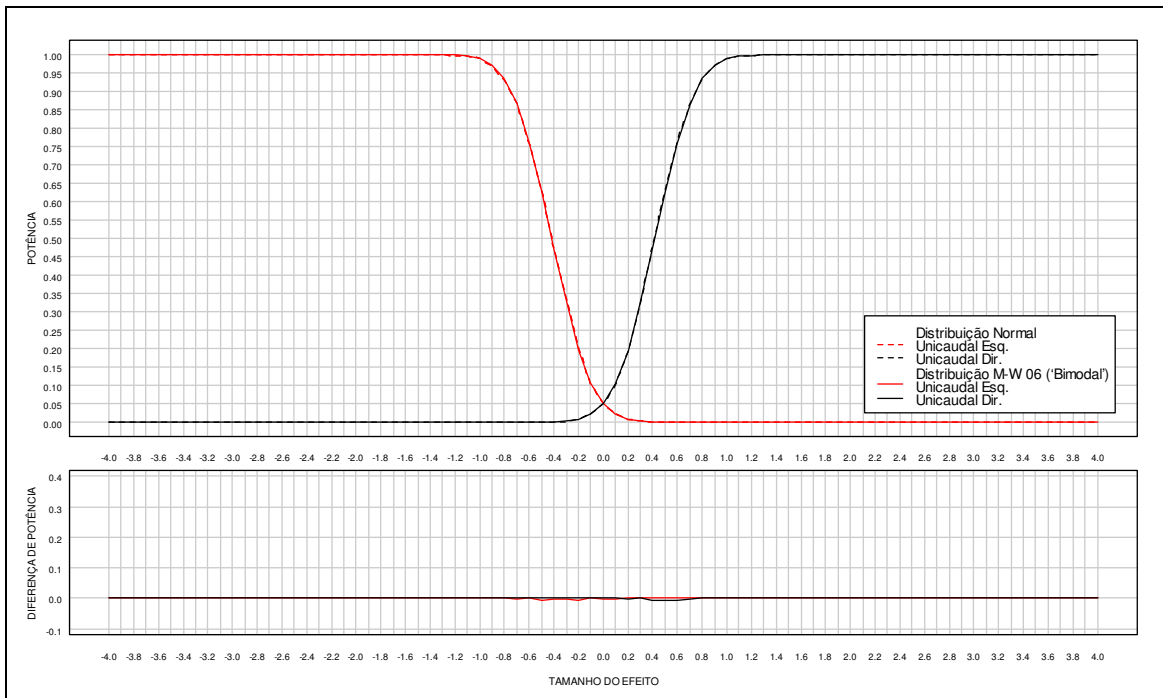


Figura Anexo VI. 75 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 06 ('Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

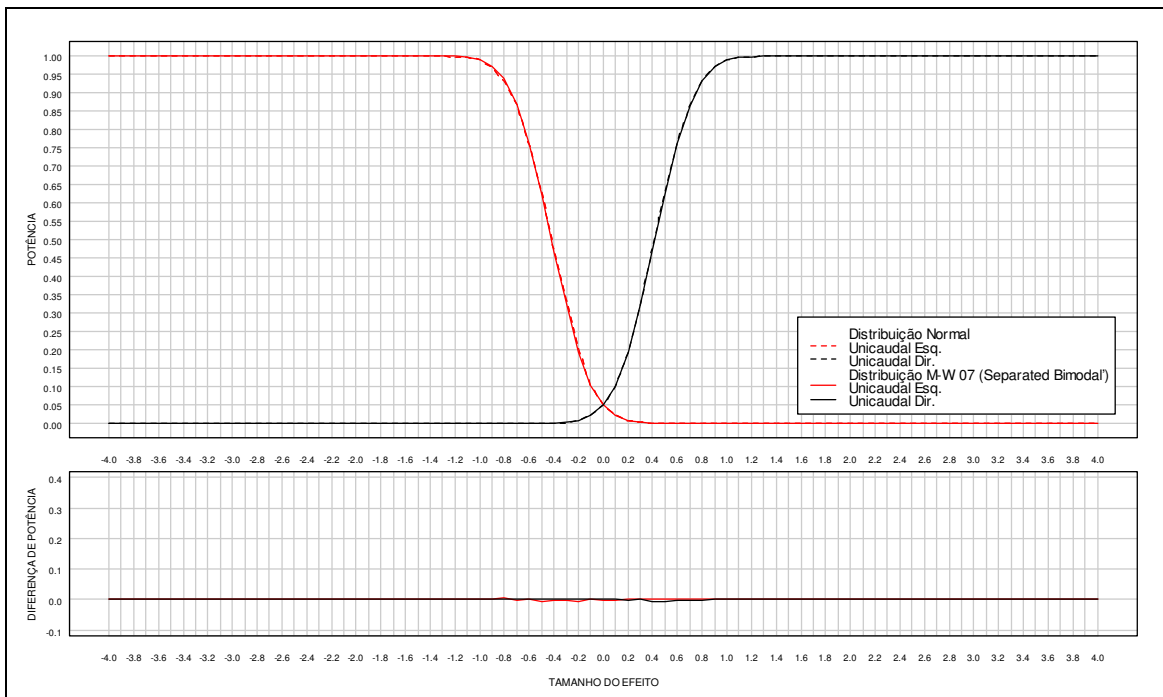


Figura Anexo VI. 76 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 07 ('Separated Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

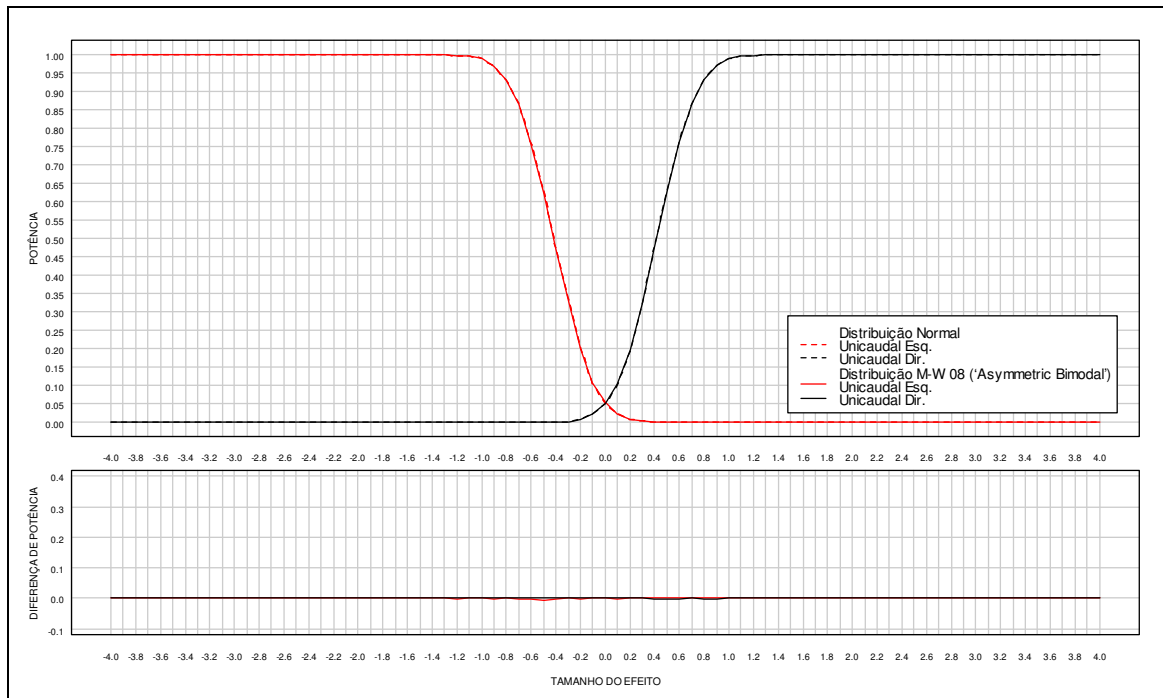


Figura Anexo VI. 77 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 08 ('Asymmetric Bimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

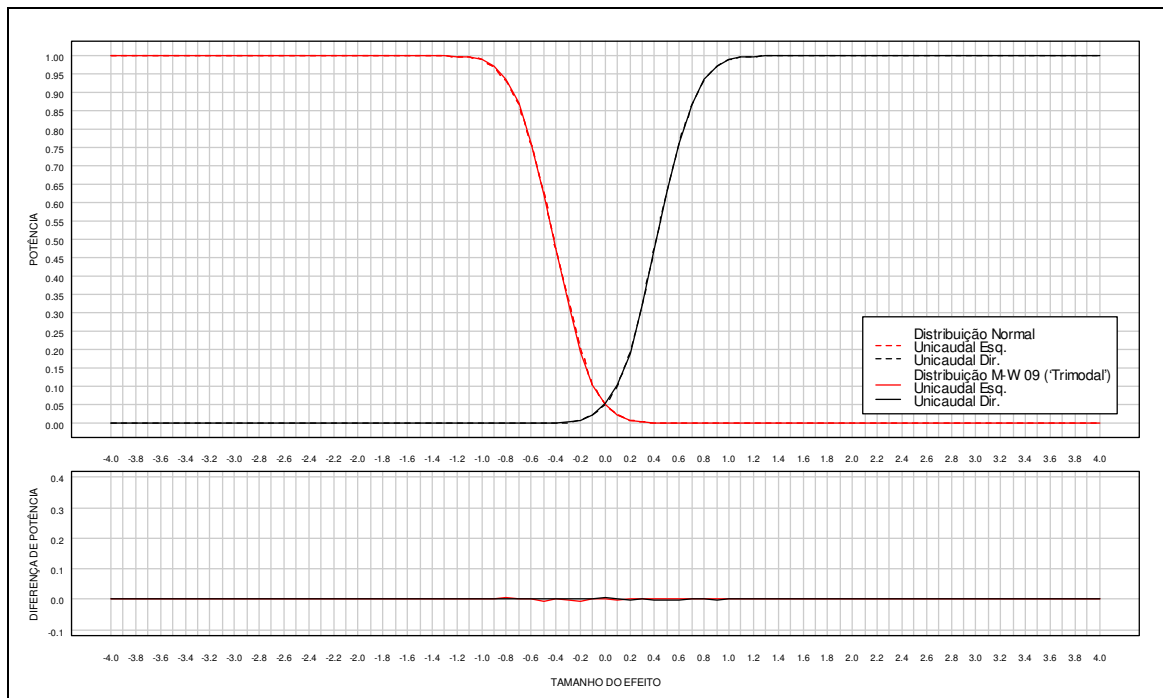


Figura Anexo VI. 78 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 09 ('Trimodal') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

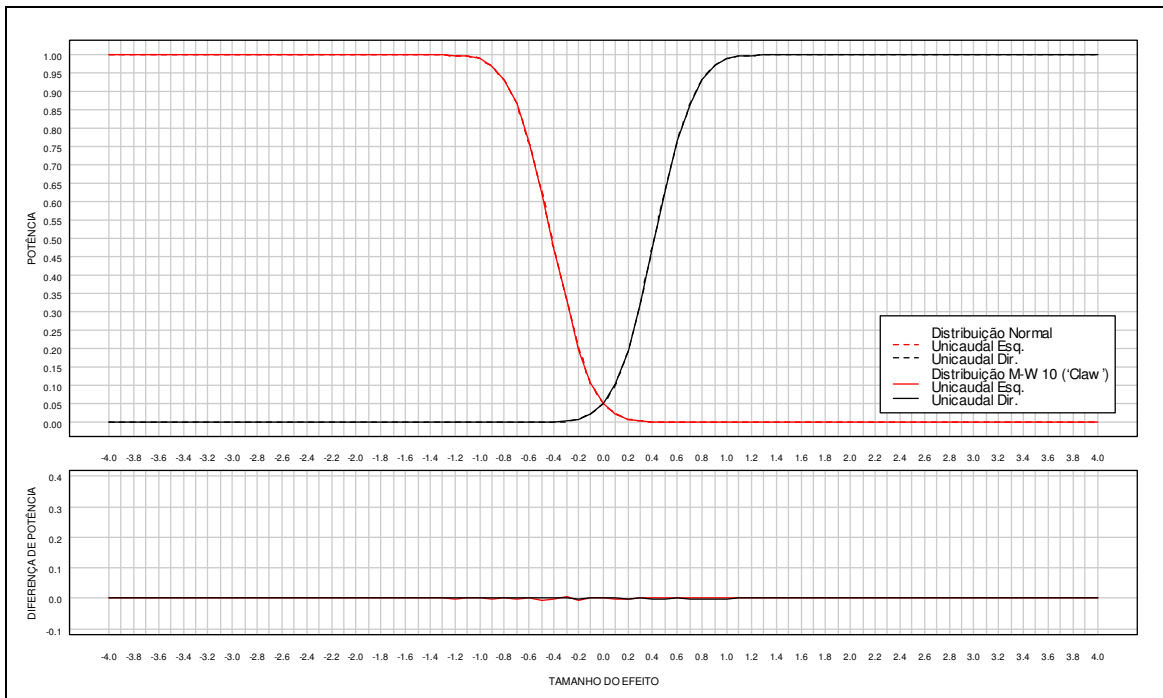


Figura Anexo VI. 79 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 10 ('Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

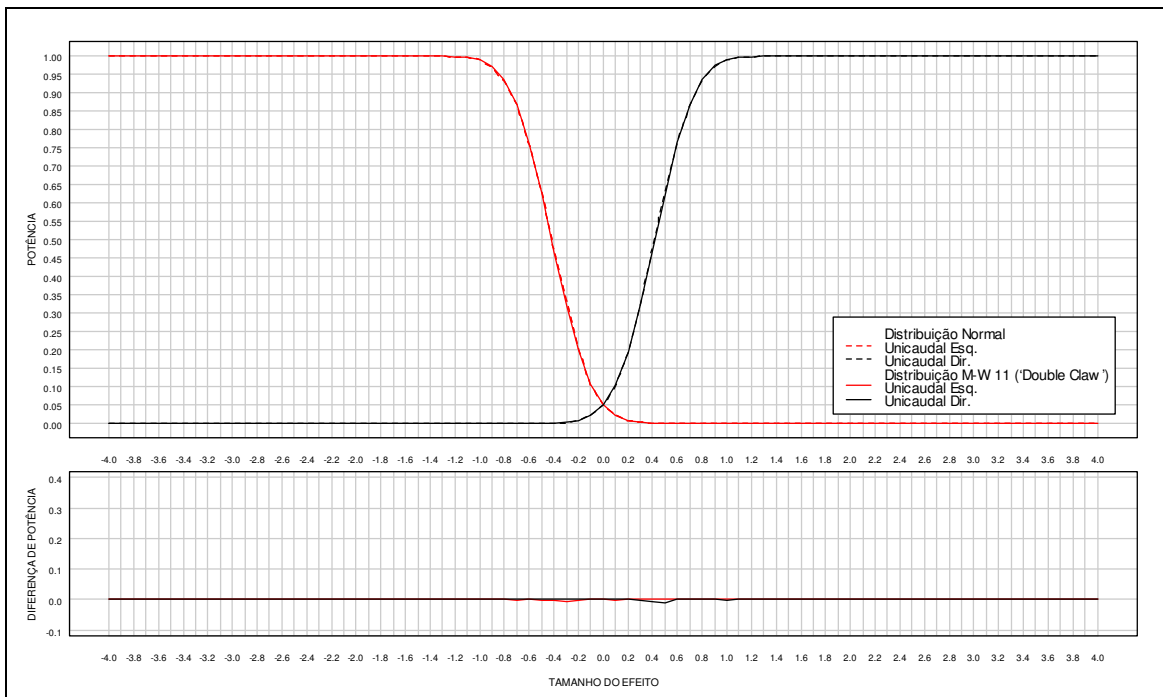


Figura Anexo VI. 80 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 11 ('Double Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

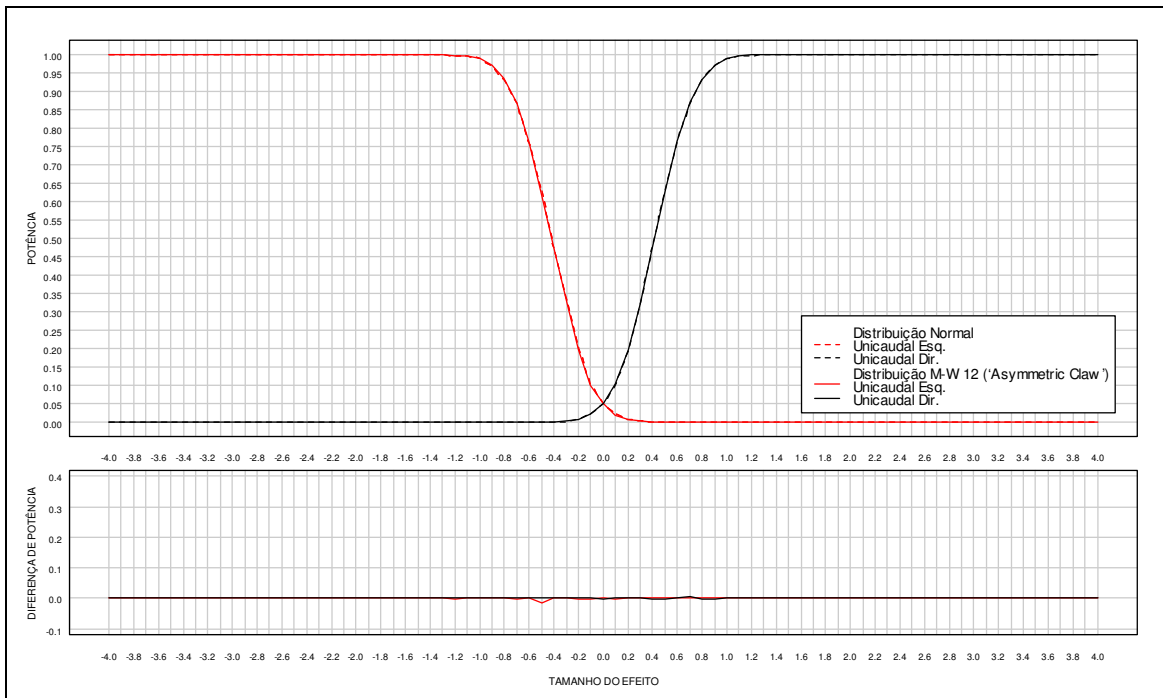


Figura Anexo VI. 81 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 12 ('Asymmetric Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

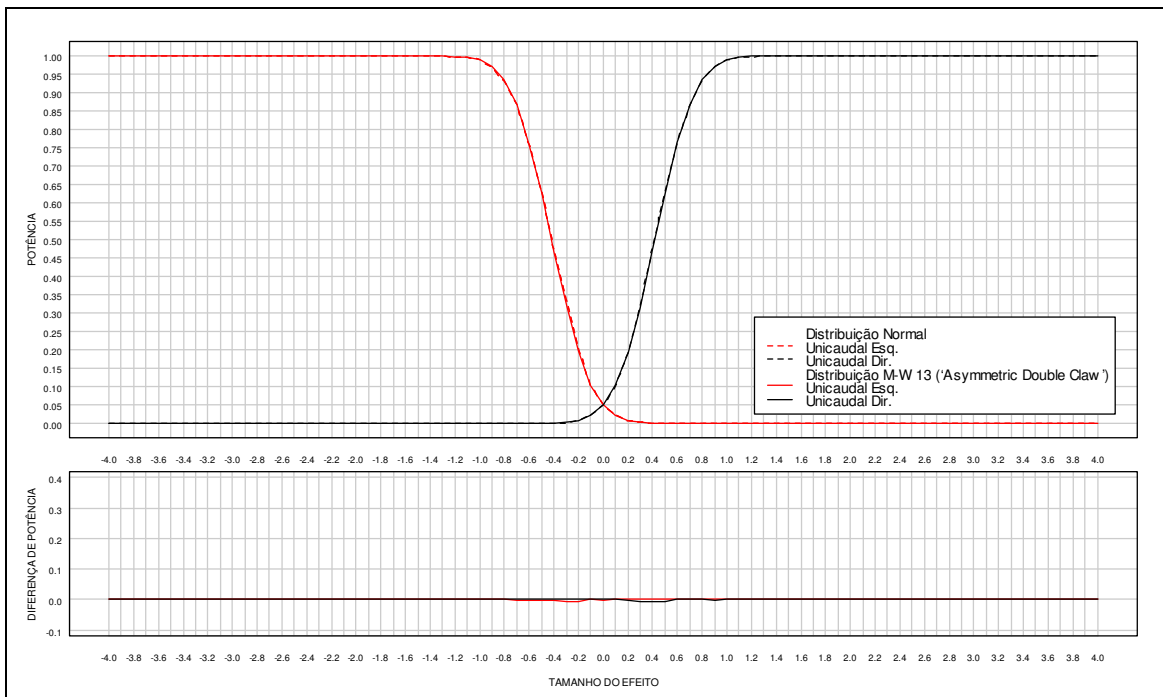


Figura Anexo VI. 82 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 13 ('Asymmetric Double Claw') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

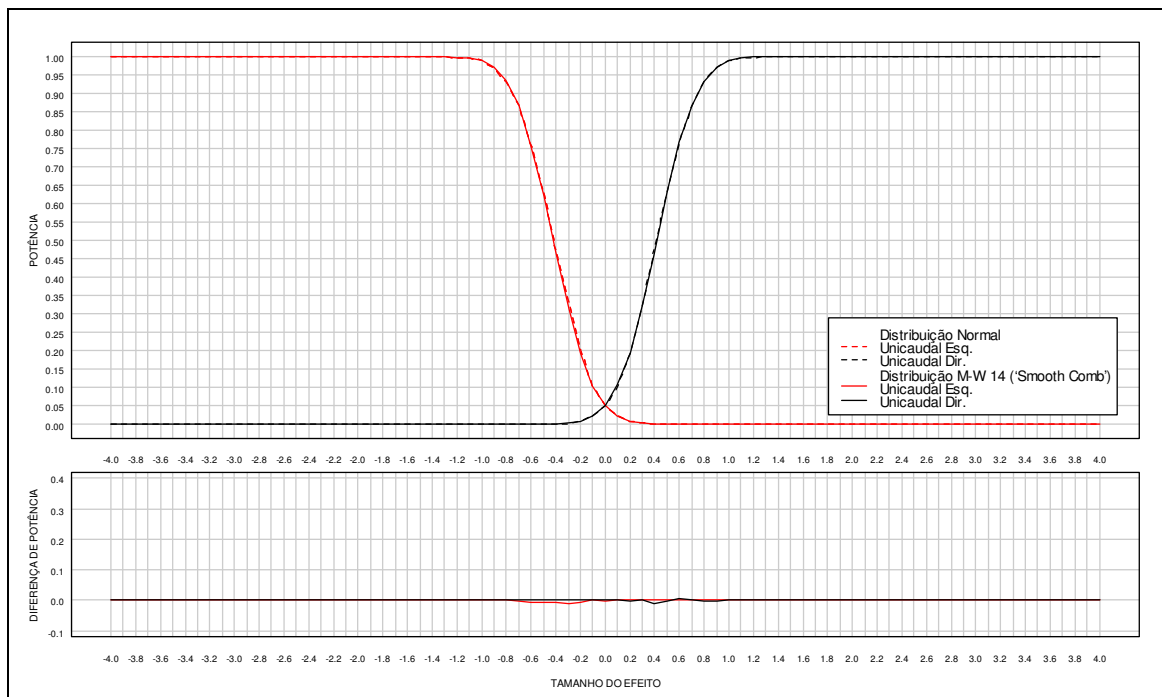


Figura Anexo VI. 83 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 14 ('Smooth Comb') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

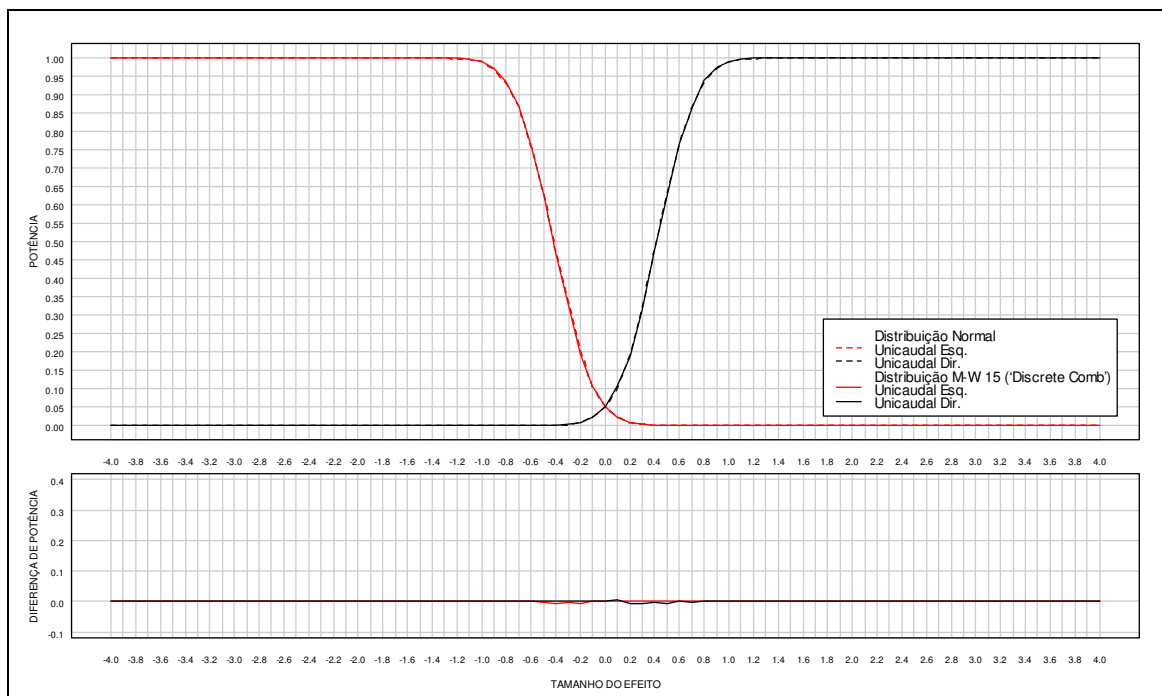


Figura Anexo VI. 84 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição Marron-Wand 15 ('Discrete Comb') em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

Anexo VII

Programas em R para o cálculo de potência do Estudo 2

```

# Dist. g-e-h (g=.8 ;h=.3)
#Def
nsim <- 20000
ngr1 <- 8; ngr2 <- 8
alfa<-0.05
te<- seq(-4,4,by=.10)
set.seed(23451)
g<-.8 ;h<-.3
#Função: Teste t de Student
teste.t <- function(x1, x2) {
n1 <- length(x1); n2<-length(x2)
mx1 <- mean(x1);mx2<-mean(x2)
s1 <- var(x1);s2<-var(x2)
s <- ((n1-1)*s1+(n2-1)*s2)/(n1+n2-2)
et <- (mx1 - mx2)/sqrt(s*(1/n1 + 1/n2))
et
}
#Função: Teste de Aleatorização (sample)
permg<-function(x,y,nream=999){
xx<- c(x,y)
etobs1<-sum(x) ; etobs2 <-
(sum(x)^2/length(x))+(sum(y)^2/length(y))
etb1 <- NA
etb2 <- NA
for(i in 1:nream){
etb1[i]<- sum(sample(xx, size = length(x), replace = F))
etb2[i]<- etb1[i]^2/length(x)+(sum(xx)-
etb1[i])^2/length(y)
}
vpesq <- (1+length(etb1[etb1 <= etobs1]))/(nream+1)
vmdir <- (1+length(etb1[etb1 >= etobs1]))/(nream+1)
vpbi <- (1+length(etb2[etb2 >= etobs2]))/(nream+1)
res<-matrix(c(vpesq,vmdir,vpbi),1,3)
}
#Função: Distribuição g-e-h
func.gh<-function(z,g,h){((exp(g*z)-
1)/g)*(exp((h*z^2)/2))}
res.tt<-matrix(0,length(te),3); res.ta<-matrix(0,
length(te),3)
matsig <- matrix(0,nsim,3)
for(k in 1 : length(te)){
dadgr1 <-
matrix(func.gh(z=rnorm(nsim*ng1),g=g,h=h)+(te[k]*

```

```

7.164677),nsim,ngr1)
dadgr2 <-
matrix(func.gh(z=rnorm(nsim*ngr1),g=g,h=h),nsim,ngr2)
ncesq <- 0; ncdir <- 0; nbi <- 0
et<- matrix(0,nsim,1)
for(i in 1:nsim) {et[i,1] <- teste.t(dadgr1[i,],
dadgr2[i,])
if(et[i,1]<= qt(alfa,(ngr1+ngr2-2)))
ncesq=ncesq+1
if(et[i,1]>= qt(1-alfa,(ngr1+ngr2-2)))
ncdir=ncdir+1
if(et[i,1]<= qt(alfa/2,(ngr1+ngr2-2))|et[i,1]>=
qt(1-
alfa/2,(ngr1+ngr2-2))) nbi=nbi+1
matsig[i,]<-permg(dadgr1[i,],dadgr2[i,])
}
res.tt[k,1]<-ncesq/nsim; res.tt[k,2]<-ncdir/nsim;
res.tt[k,3]<-nbi/nsim
vecsigesq<-matsig[,1]
vecsigdir<-matsig[,2]
vecsigbi<-matsig[,3]
res.ta[k,1]<-length(vecsigesq[vecsigesq<=alfa])/nsim
res.ta[k,2]<-length(vecsigdir[vecsigdir<=alfa])/nsim
res.ta[k,3]<-length(vecsigbi[vecsigbi<=alfa])/nsim
}
sink("Res_gh(g<-.8 ;h<-.3)_8+8_20000_999",append=F)
cat(res.tt,res.ta)
sink()

```

Nota: Este programa foi usado para simular dados da distribuição 20 (distribuição g - e - h , com $g = 0.8$ e $h = 0.3$), para $n_1 = n_2 = 8$. Os outros programas são semelhantes, usando, em vez da função `func.gh` as funções:

```

func.g<-function(z,g){(exp(g*z)-1)/g} e
func.h<-function(z,h){z*exp(h*z^2/2)},

```

respectivamente, para as distribuições g (D2 a D5) e para as distribuições h (D6 a D8), e substituindo os valores dos tamanhos dos grupos, os valores de g e/ou de h e o valor do desvio-padrão da distribuição.

Anexo VIII

Estatísticas descritivas (mínimo, máximo e média) da diferença entre a potência do teste de aleatorização e a do teste t de Student

Distribuição		Estatísticas	$n_1 = n_2 = 8$			$n_1 = n_2 = 16$			$n_1 = n_2 = 32$		
			Esq	Dir	Bi	Esq	Dir	Bi	Esq	Dir	Bi
D1	N(0,1)	Mínimo	-0.003	-0.004	-0.004	-0.004	-0.002	-0.003	-0.003	-0.001	-0.003
		Máximo	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.000	0.001
		Média	0.000	0.000	-0.001	0.000	0.000	0.000	0.000	0.000	0.000
D2	g = 0.2	Mínimo	-0.003	-0.003	-0.002	-0.003	-0.002	-0.002	-0.002	-0.002	-0.002
		Máximo	0.001	0.002	0.002	0.001	0.000	0.002	0.001	0.001	0.001
		Média	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
D3	g = 0.4	Mínimo	-0.002	-0.001	-0.002	-0.002	-0.001	-0.001	-0.002	-0.002	-0.001
		Máximo	0.004	0.004	0.010	0.001	0.002	0.005	0.001	0.001	0.003
		Média	0.000	0.000	0.002	0.000	0.000	0.001	0.000	0.000	0.000
D4	g = 0.6	Mínimo	-0.001	-0.001	-0.001	0.000	-0.001	0.000	-0.003	-0.001	0.000
		Máximo	0.009	0.011	0.023	0.004	0.004	0.011	0.001	0.001	0.007
		Média	0.001	0.001	0.005	0.000	0.000	0.003	0.000	0.000	0.001
D5	g = 0.8	Mínimo	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Máximo	0.019	0.019	0.035	0.007	0.007	0.021	0.002	0.003	0.012
		Média	0.002	0.002	0.009	0.001	0.001	0.005	0.000	0.000	0.002
D6	h = 0.1	Mínimo	-0.001	-0.002	-0.001	-0.002	-0.002	0.000	-0.002	-0.003	-0.002
		Máximo	0.002	0.002	0.007	0.001	0.001	0.005	0.001	0.001	0.003
		Média	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000
D7	h = 0.2	Mínimo	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	-0.001	0.000
		Máximo	0.009	0.008	0.017	0.003	0.003	0.011	0.002	0.002	0.008
		Média	0.001	0.001	0.005	0.000	0.000	0.002	0.000	0.000	0.001
D8	h = 0.3	Mínimo	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Máximo	0.016	0.015	0.031	0.008	0.008	0.022	0.005	0.006	0.017
		Média	0.002	0.002	0.009	0.001	0.001	0.005	0.001	0.001	0.003
D9	g = 0.2, h = 0.1	Mínimo	-0.001	-0.002	-0.001	-0.001	-0.002	0.000	-0.002	-0.003	0.000
		Máximo	0.005	0.005	0.009	0.001	0.001	0.006	0.000	0.001	0.004
		Média	0.000	0.000	0.002	0.000	0.000	0.001	0.000	0.000	0.001
D10	g = 0.2, h = 0.2	Mínimo	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000
		Máximo	0.009	0.009	0.022	0.004	0.004	0.015	0.003	0.002	0.010
		Média	0.001	0.001	0.006	0.000	0.000	0.003	0.000	0.000	0.002
D11	g = 0.2, h = 0.3	Mínimo	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
		Máximo	0.019	0.018	0.034	0.010	0.010	0.027	0.007	0.007	0.020
		Média	0.002	0.002	0.009	0.001	0.001	0.006	0.001	0.001	0.004
D12	g = 0.4, h = 0.1	Mínimo	-0.001	0.000	0.000	0.000	0.000	0.000	-0.001	-0.002	0.000
		Máximo	0.010	0.009	0.020	0.003	0.003	0.013	0.002	0.001	0.007
		Média	0.001	0.001	0.005	0.000	0.000	0.003	0.000	0.000	0.001
D13	g = 0.4, h = 0.2	Mínimo	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Máximo	0.015	0.017	0.033	0.009	0.008	0.021	0.006	0.005	0.017
		Média	0.002	0.002	0.009	0.001	0.001	0.006	0.001	0.001	0.003
D14	g = 0.4, h = 0.3	Mínimo	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000
		Máximo	0.025	0.024	0.043	0.016	0.018	0.033	0.012	0.011	0.030
		Média	0.003	0.003	0.011	0.002	0.002	0.008	0.001	0.001	0.005

Distribuição		Estatísticas	$n_1 = n_2 = 8$			$n_1 = n_2 = 16$			$n_1 = n_2 = 32$		
			Esq	Dir	Bi	Esq	Dir	Bi	Esq	Dir	Bi
D15	g = 0.6, h = 0.1	Mínimo	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Máximo	0.017	0.018	0.034	0.009	0.008	0.023	0.004	0.005	0.015
		Média	0.002	0.002	0.009	0.001	0.001	0.005	0.000	0.000	0.003
D16	g = 0.6, h = 0.2	Mínimo	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
		Máximo	0.025	0.026	0.044	0.017	0.017	0.034	0.012	0.011	0.027
		Média	0.003	0.003	0.012	0.002	0.002	0.008	0.001	0.001	0.005
D17	g = 0.6, h = 0.3	Mínimo	0.000	0.000	0.002	0.000	0.000	0.001	0.000	0.000	0.000
		Máximo	0.035	0.035	0.058	0.027	0.025	0.044	0.018	0.019	0.038
		Média	0.003	0.003	0.012	0.002	0.002	0.009	0.002	0.002	0.007
D18	g = 0.8, h = 0.1	Mínimo	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
		Máximo	0.028	0.027	0.050	0.016	0.017	0.035	0.008	0.009	0.025
		Média	0.003	0.003	0.013	0.002	0.002	0.008	0.001	0.001	0.005
D19	g = 0.8, h = 0.2	Mínimo	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000
		Máximo	0.037	0.037	0.059	0.027	0.024	0.046	0.017	0.018	0.038
		Média	0.004	0.004	0.014	0.002	0.002	0.010	0.002	0.002	0.007
D20	g = 0.8, h = 0.3	Mínimo	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000
		Máximo	0.046	0.044	0.070	0.032	0.032	0.060	0.031	0.028	0.048
		Média	0.003	0.003	0.012	0.002	0.002	0.009	0.002	0.002	0.007

Nota: Esq - Teste unicaudal à esquerda; Dir - Teste unicaudal à direita; Bi - Teste bicaudal

ANEXO IX**Comparação da potência dos testes de aleatorização:
Distribuições *g-e-h* vs. distribuição normal**1. Amostras com $n_1 = n_2 = 8$

1.1. Testes bi-caudais

1.2. Testes uni-caudais

2. Amostras com $n_1 = n_2 = 16$

2.1. Testes bi-caudais

2.2. Testes uni-caudais

3. Amostras com $n_1 = n_2 = 32$

3.1. Testes bi-caudais

3.2. Testes uni-caudais

1. Amostras com $n_1 = n_2 = 8$
 1.1. Testes bi-caudais

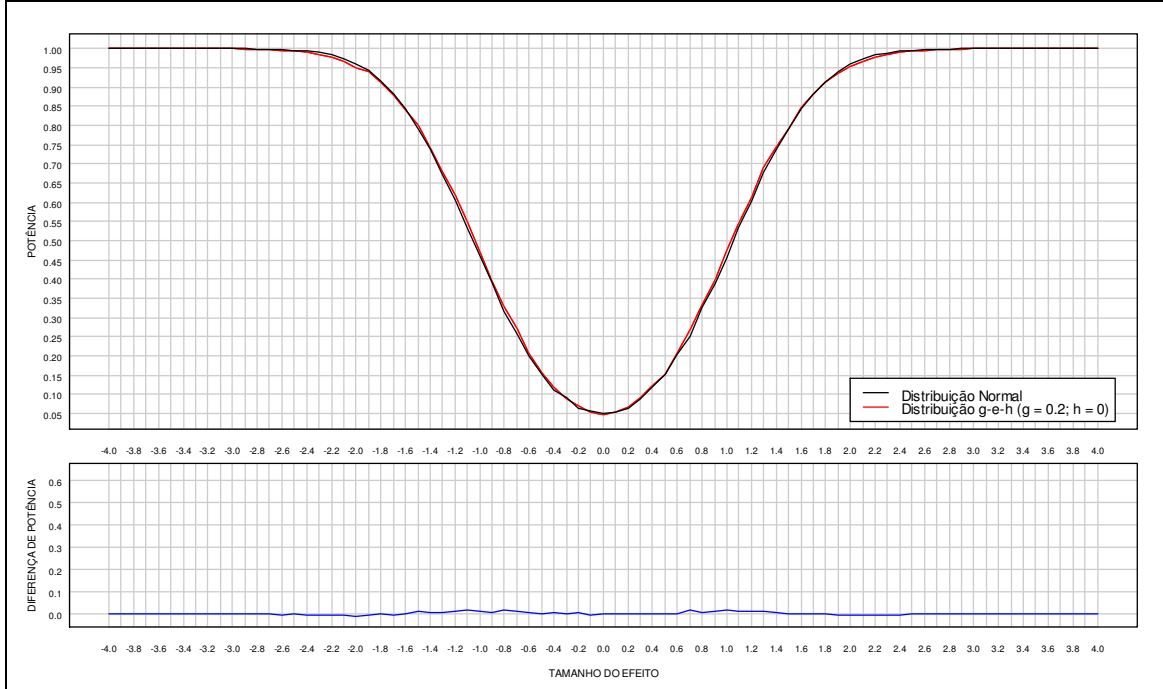


Figura Anexo IX. 1 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0$, em comparação com sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

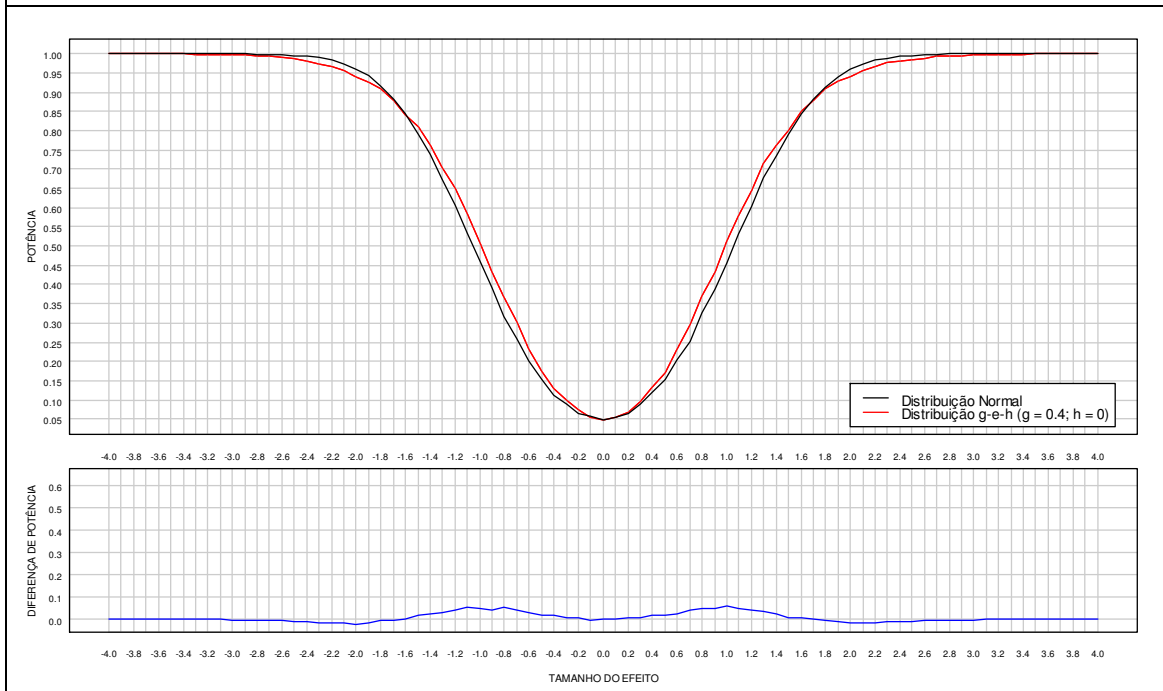


Figura Anexo IX.2 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0$, em comparação com sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

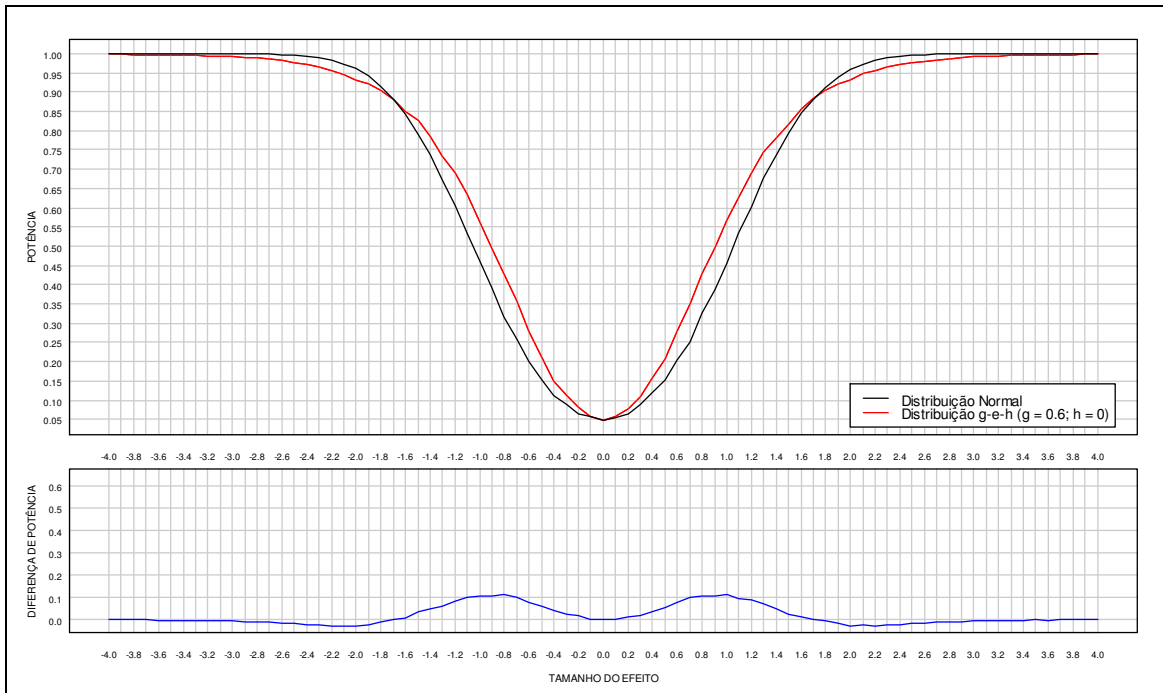


Figura Anexo IX.3 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0$, em comparação com sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

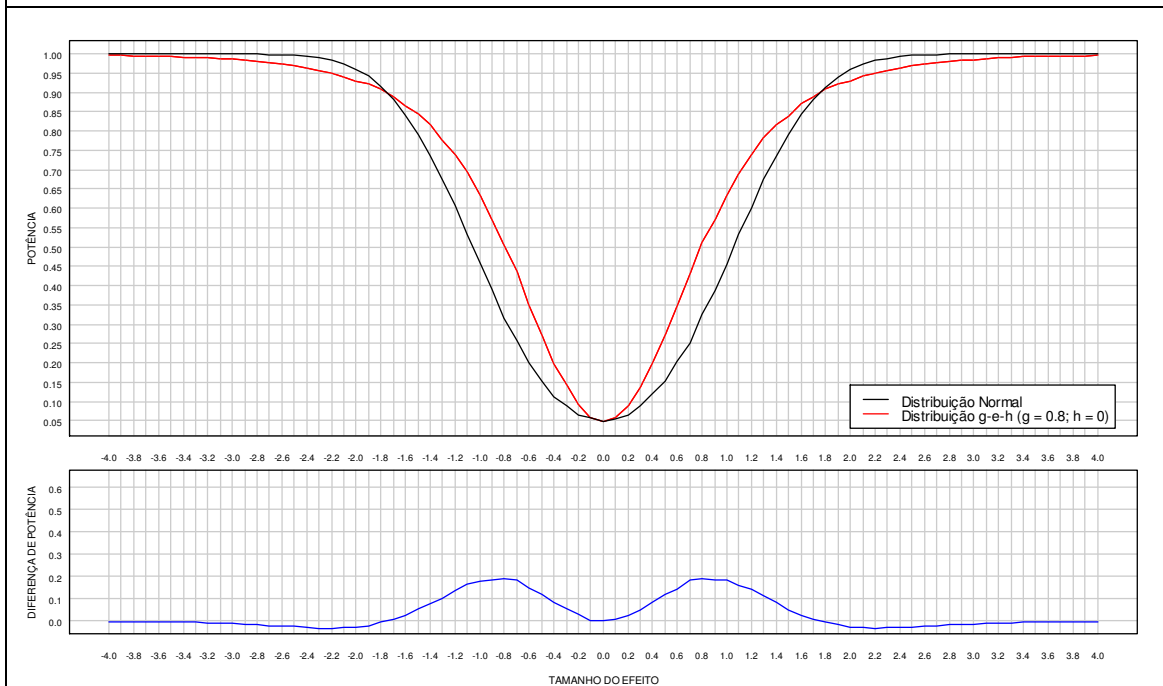


Figura Anexo IX.4 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0$, em comparação com sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

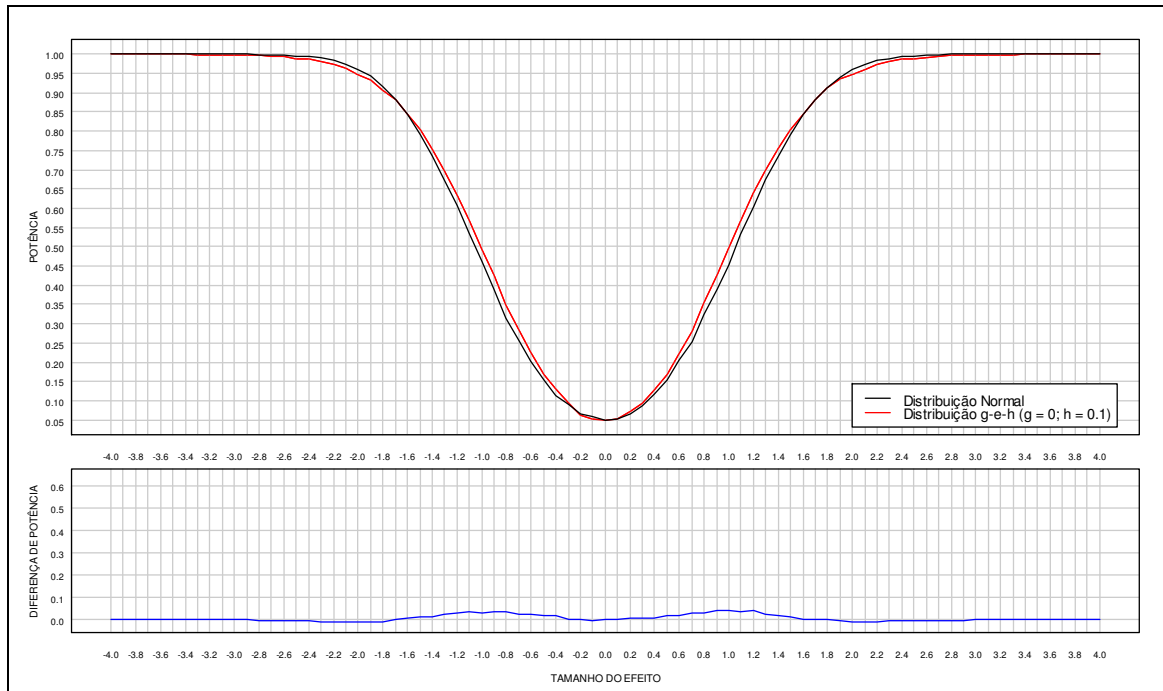


Figura Anexo IX.5 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0$ e $h = 0.1$, em comparação com sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

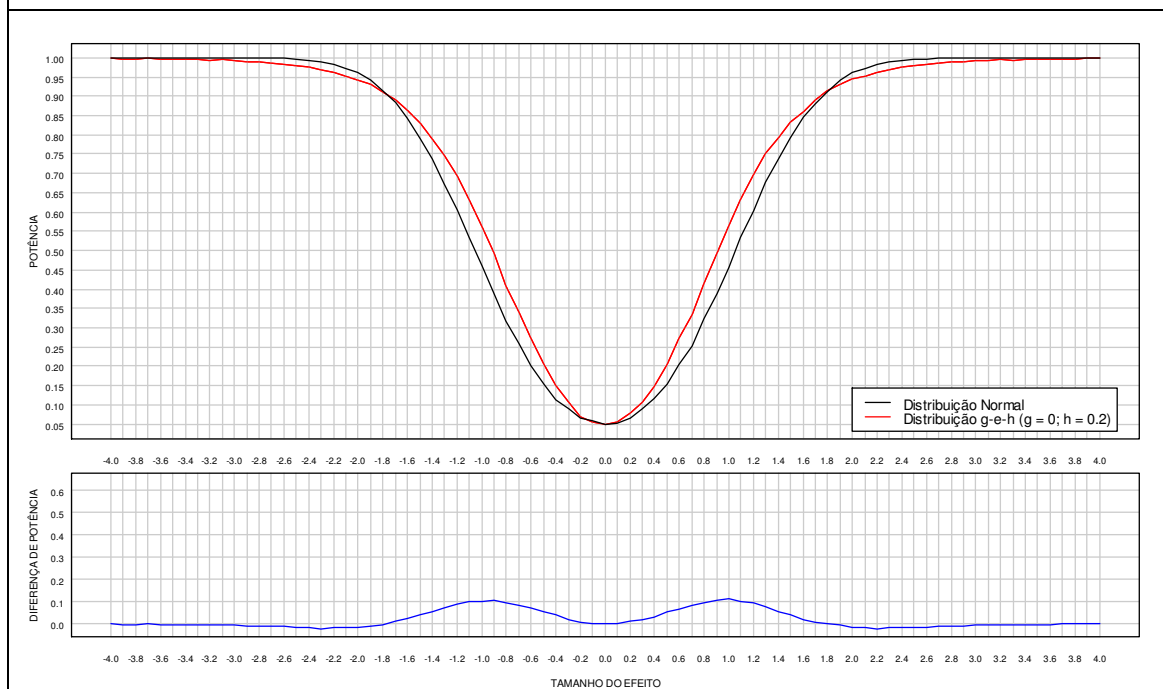


Figura Anexo IX.6 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0$ e $h = 0.2$, em comparação com sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

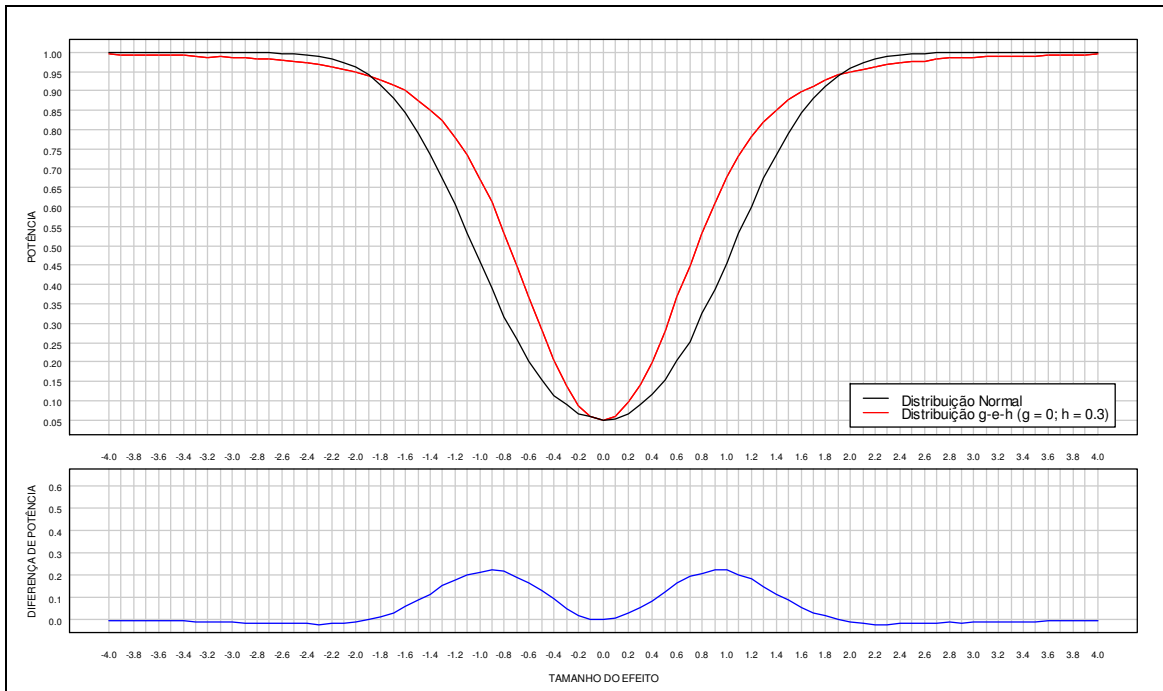


Figura Anexo IX.7 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

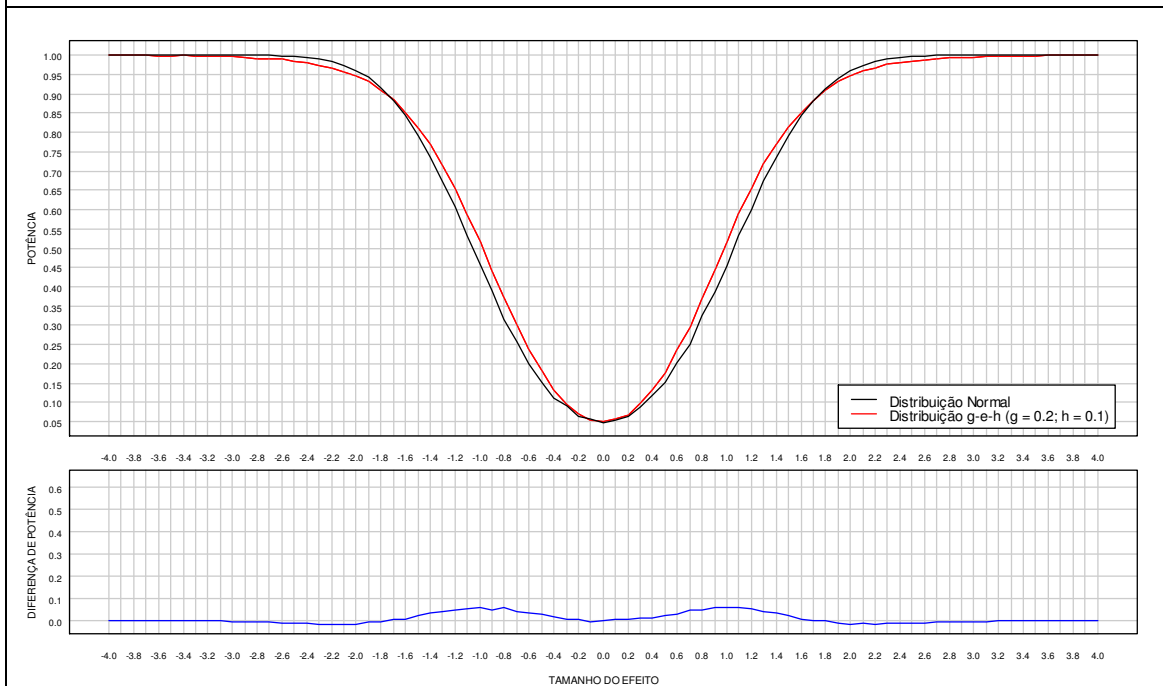


Figura Anexo IX.8 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

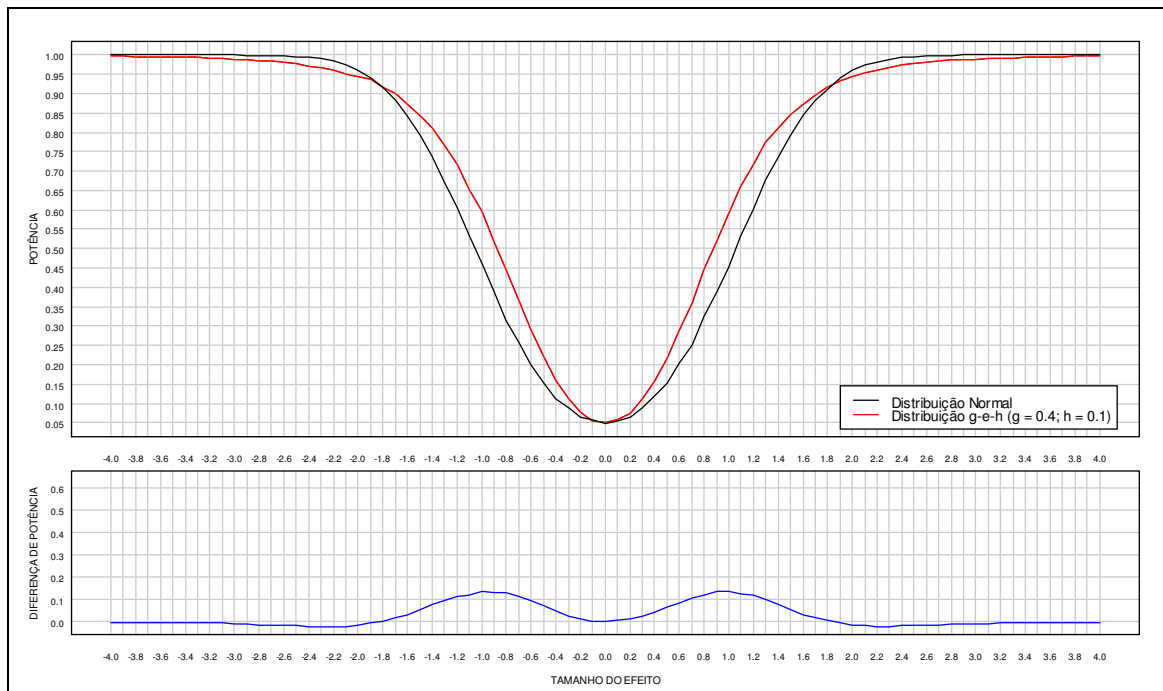


Figura Anexo IX.9 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

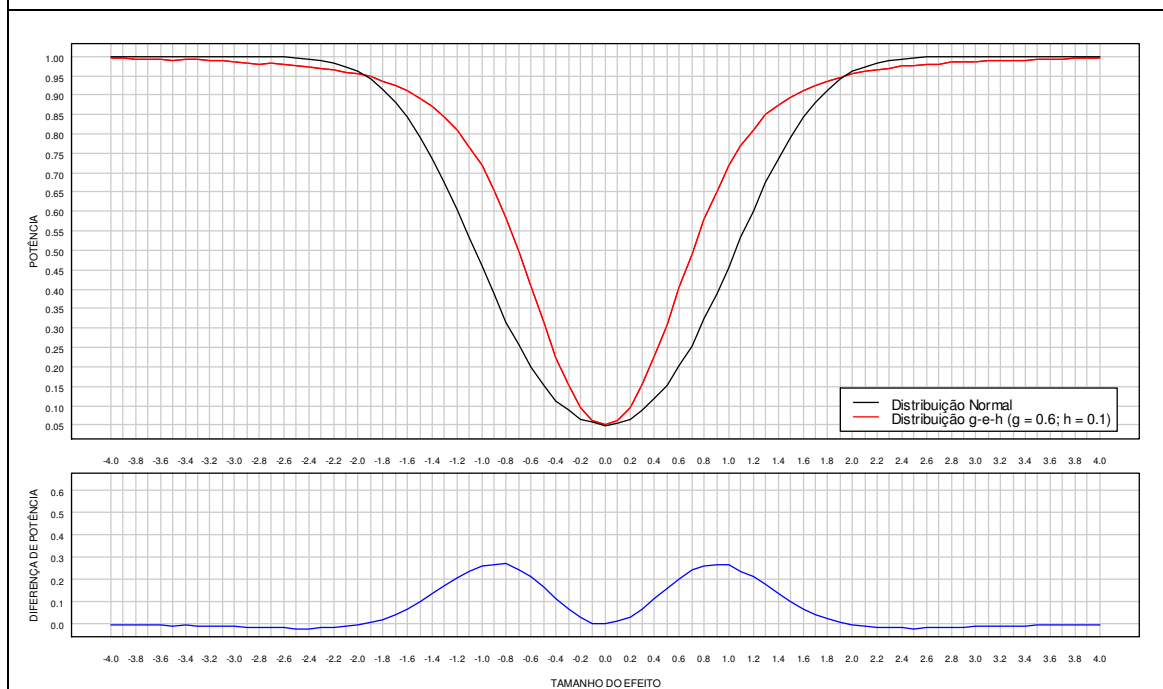


Figura Anexo IX.10 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

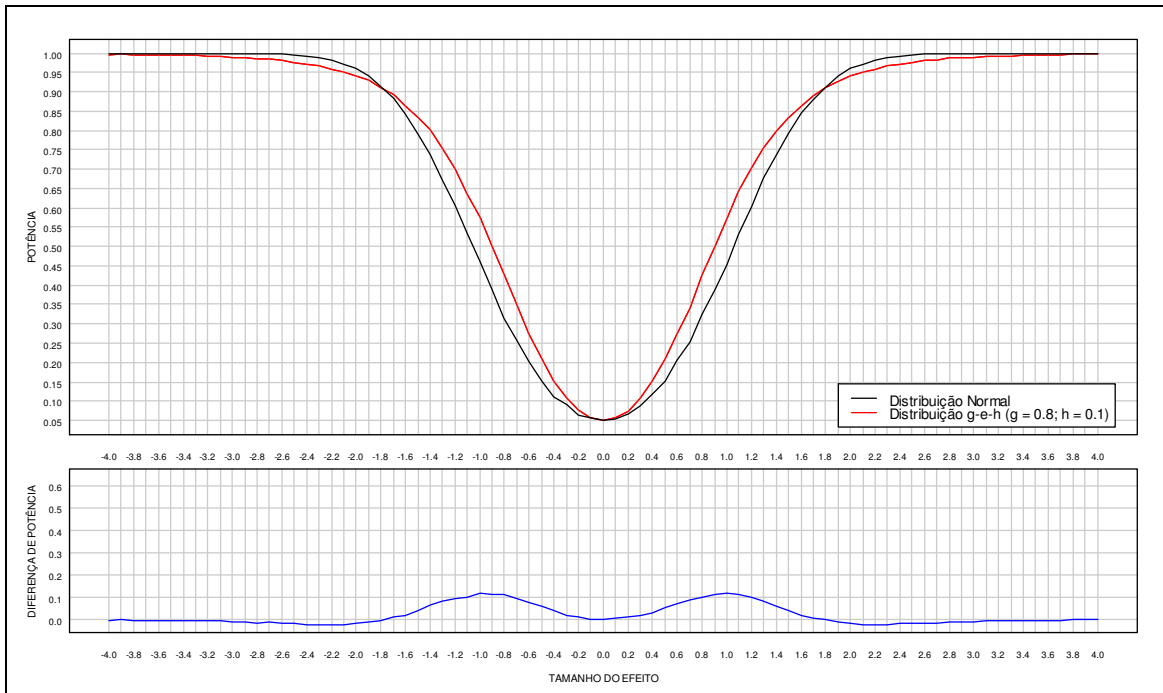


Figura Anexo IX.11 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

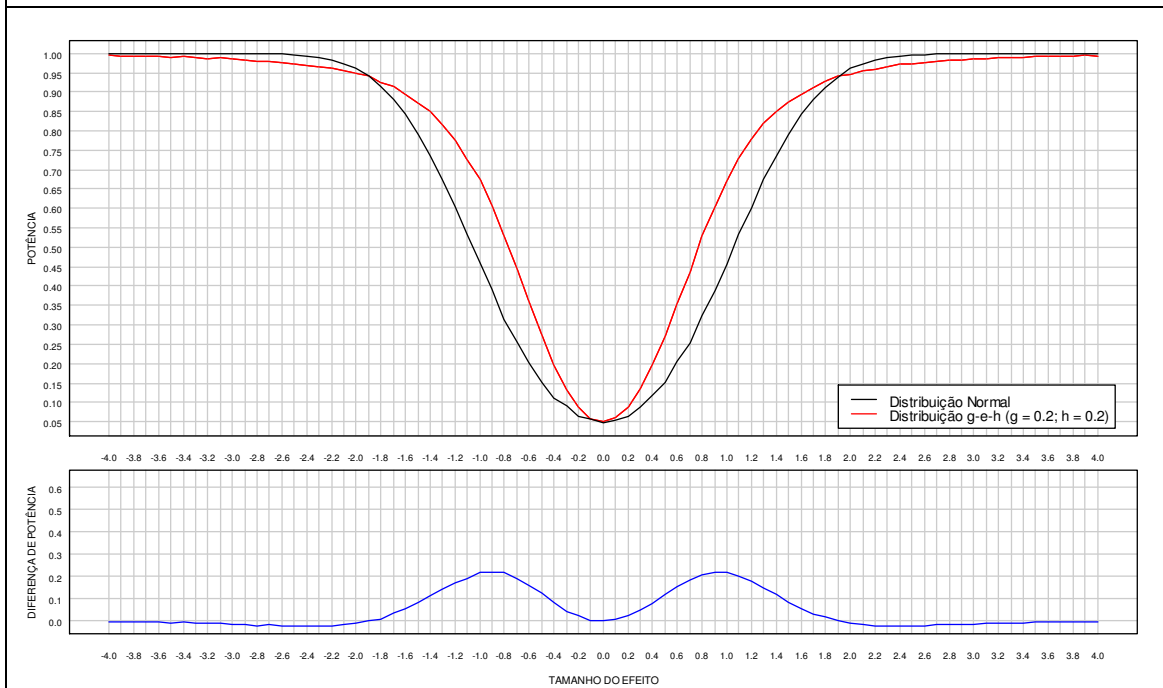


Figura Anexo IX.12 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

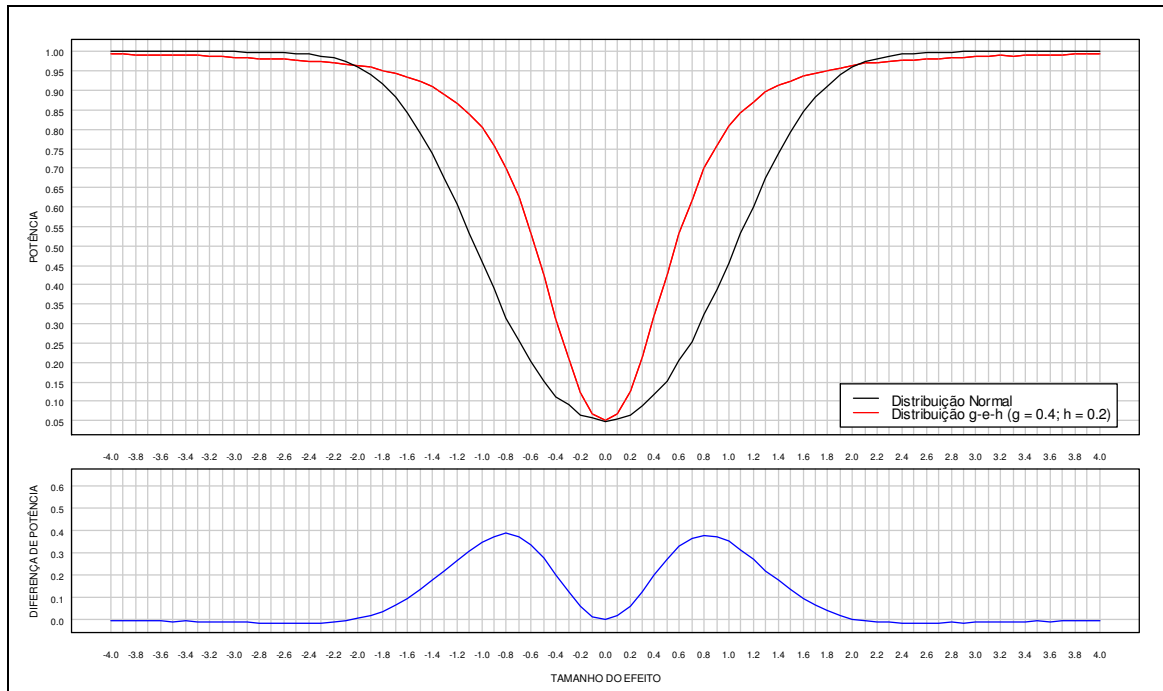


Figura Anexo IX.13 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

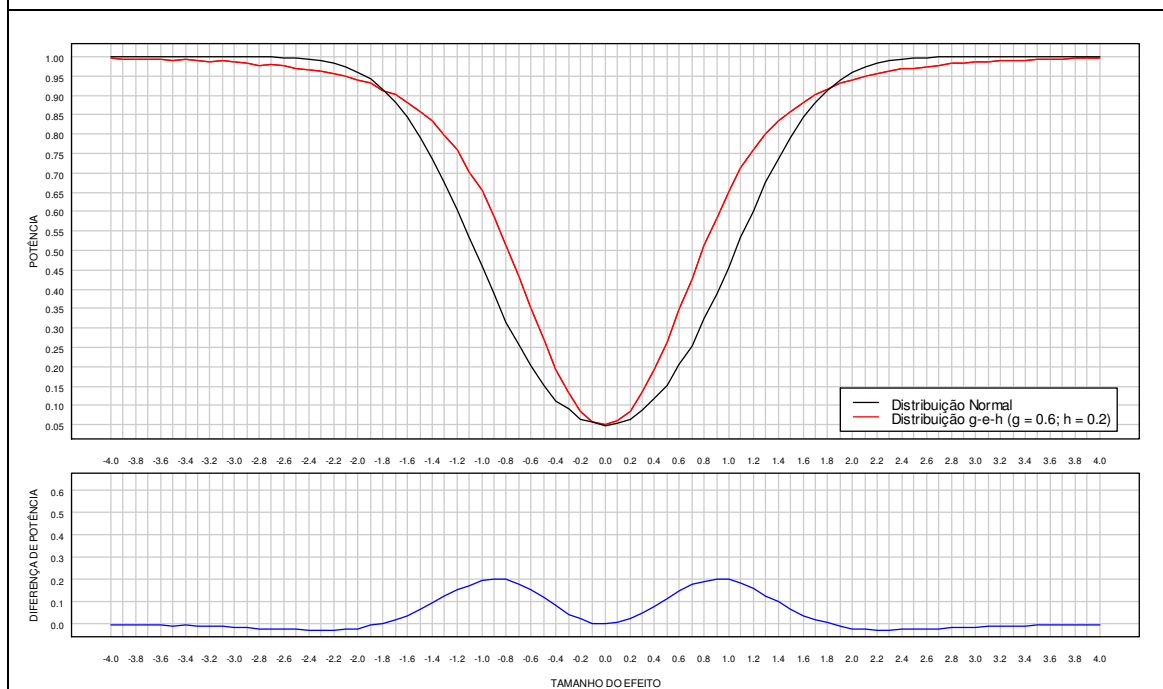


Figura Anexo IX.14 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

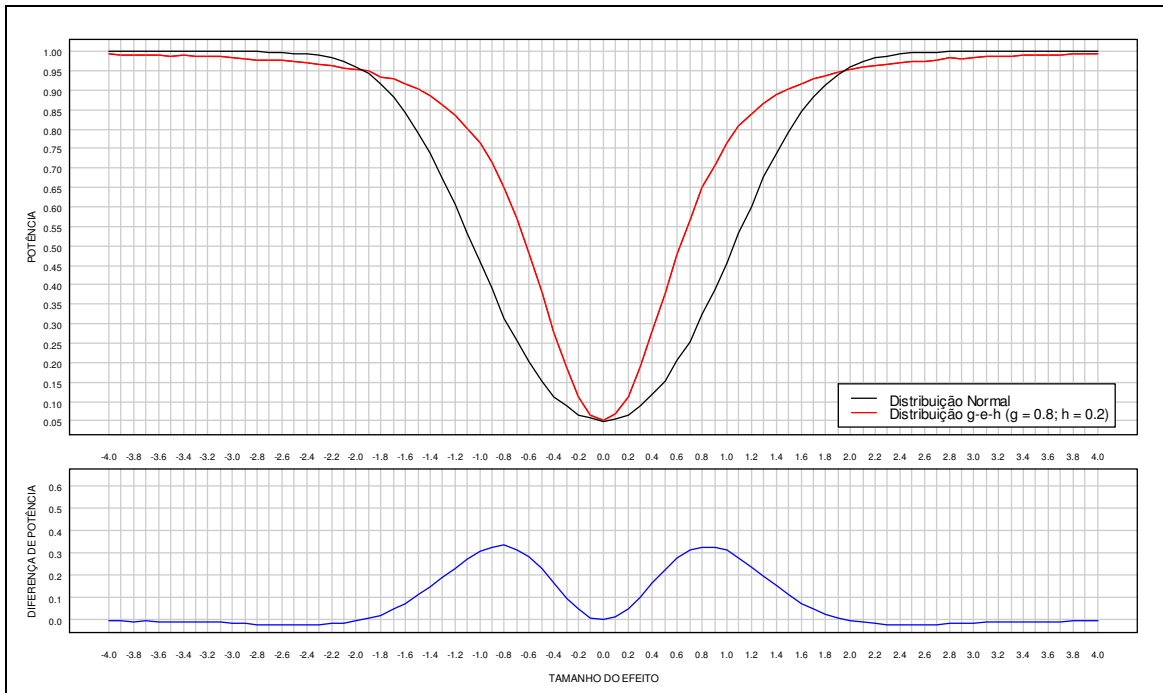


Figura Anexo IX.15 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

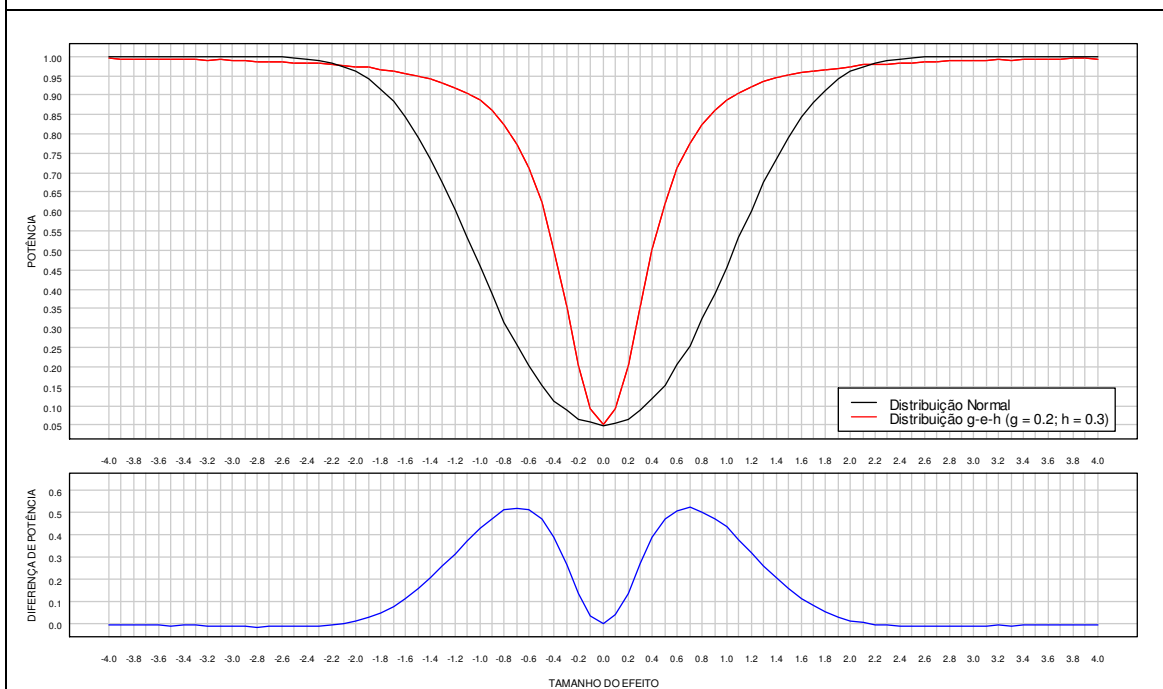


Figura Anexo IX.16 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

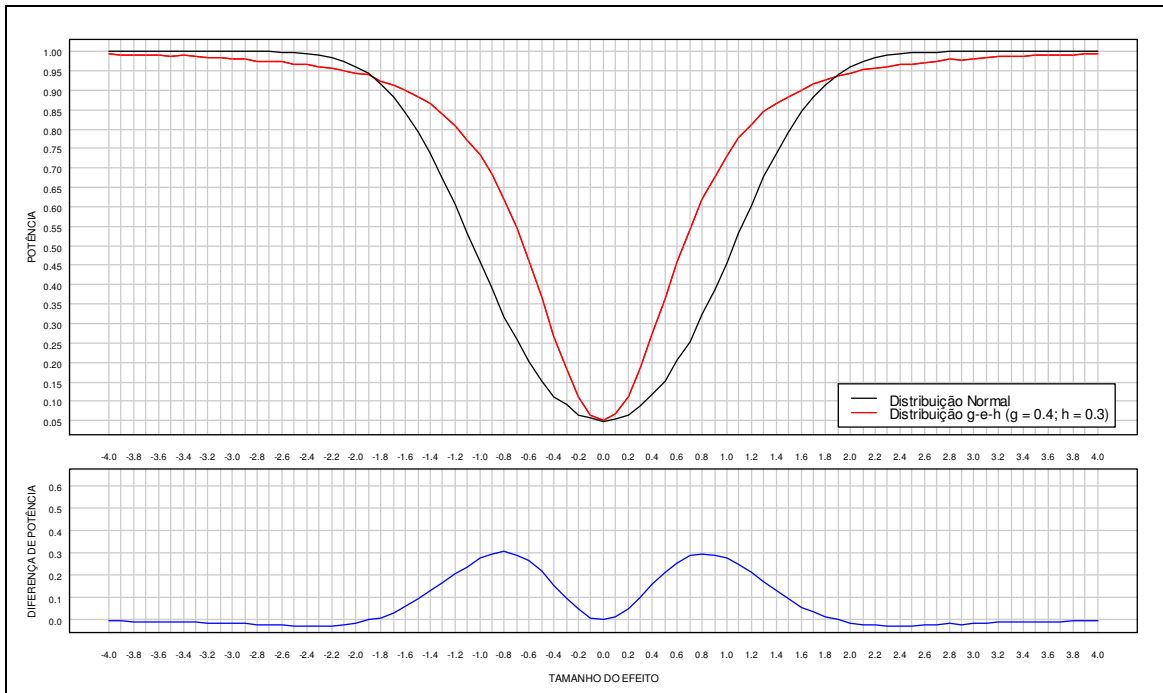


Figura Anexo IX.17 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

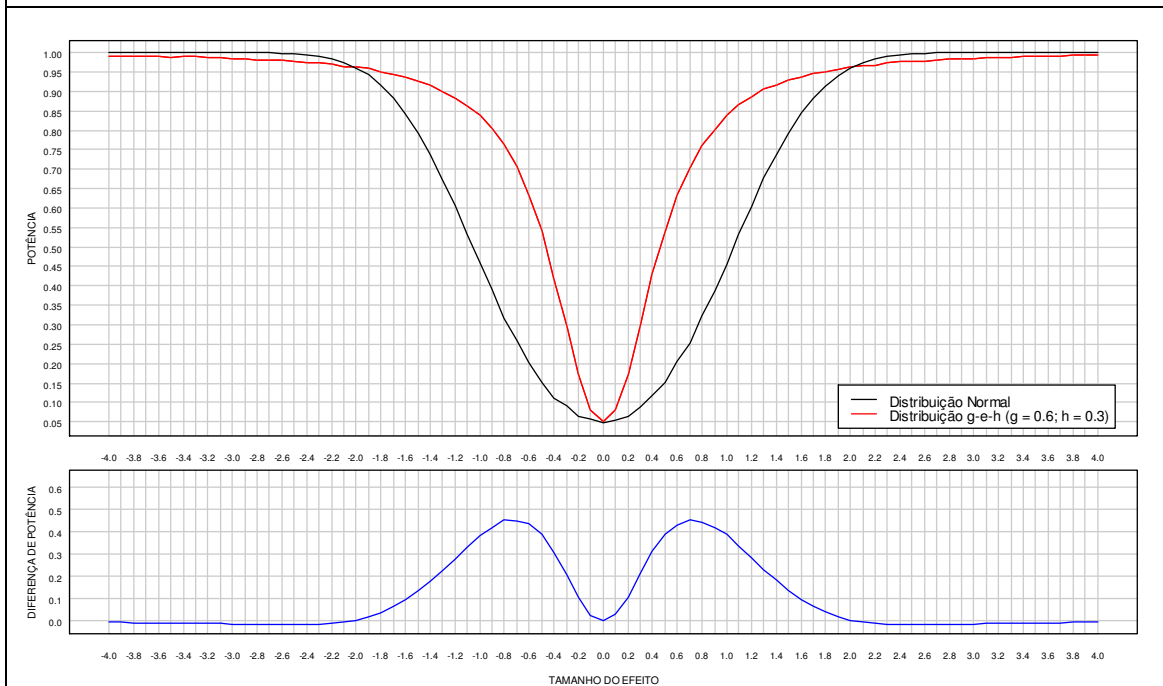


Figura Anexo IX.18 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

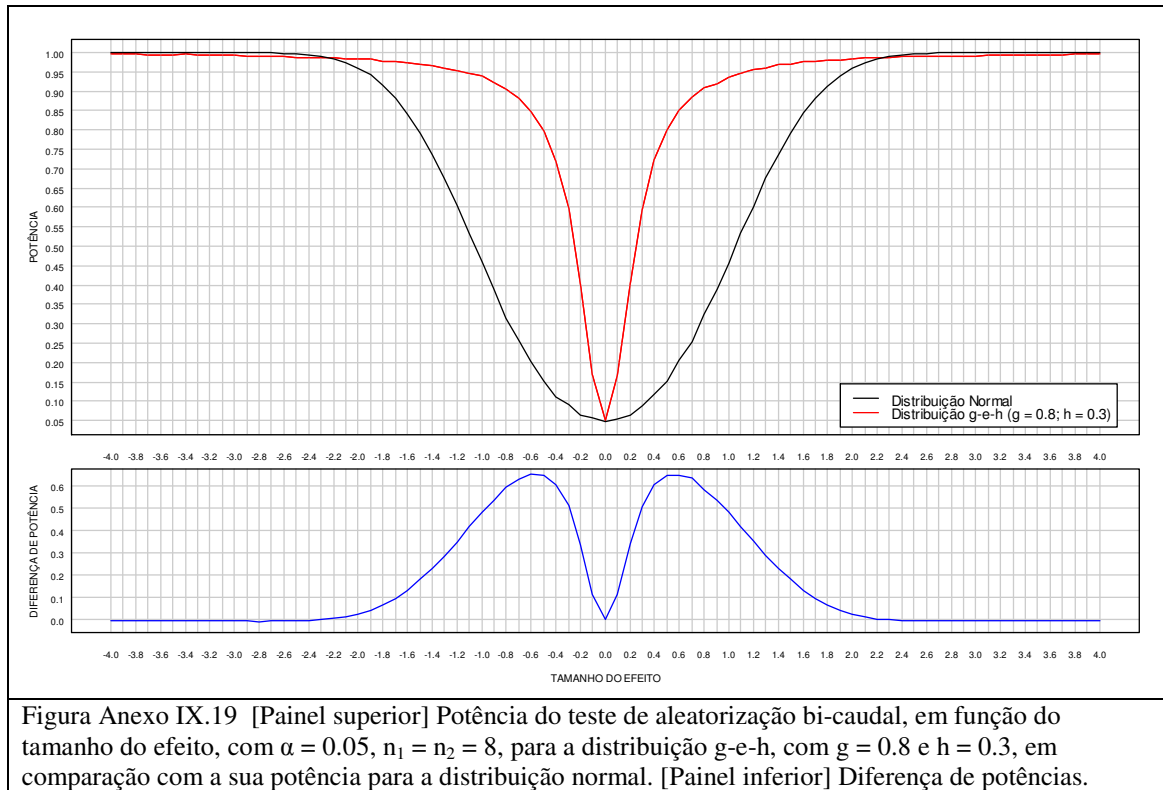


Figura Anexo IX.19 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

1.2. Testes uni-caudais

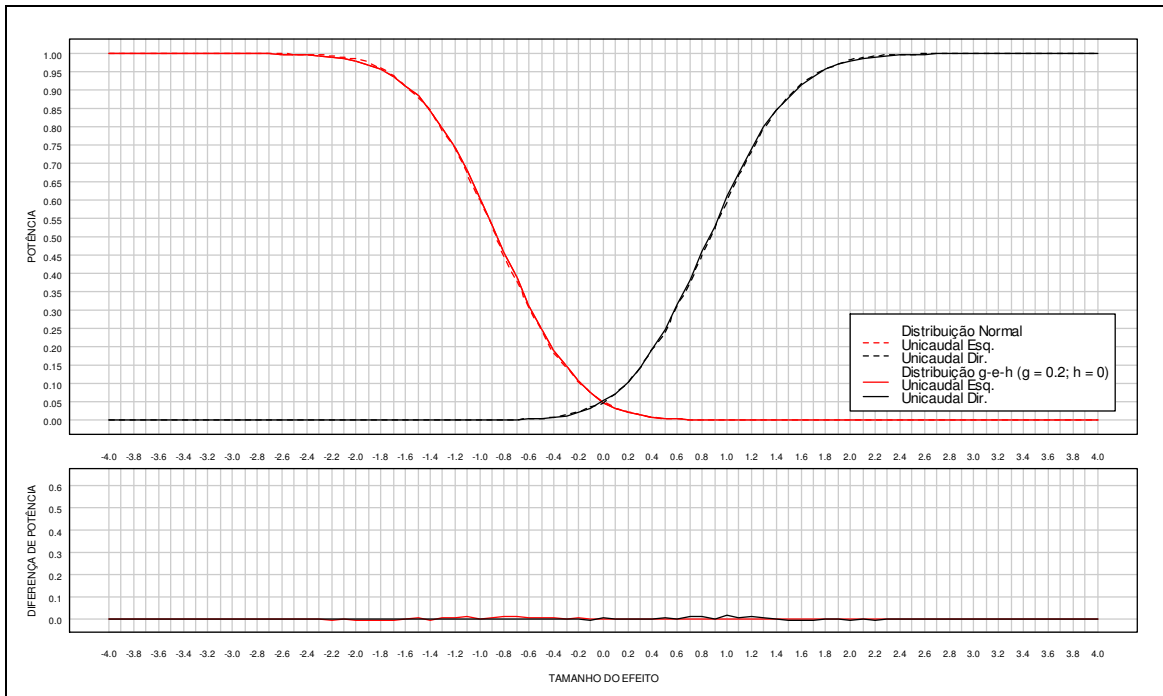


Figura Anexo IX.20 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

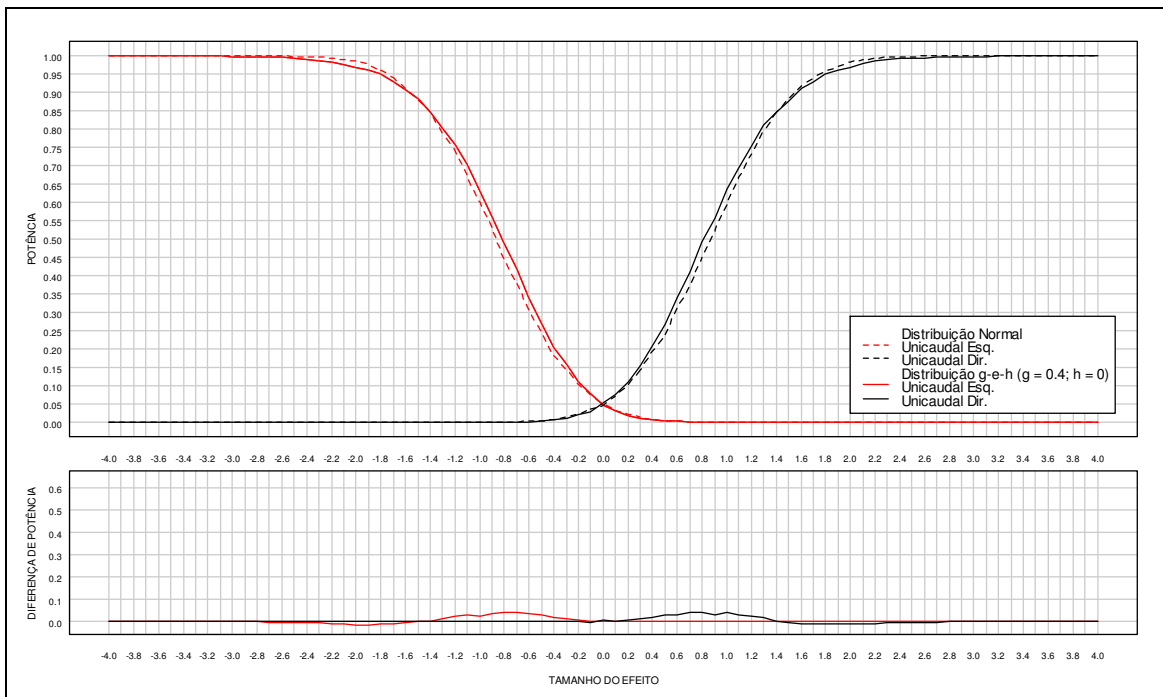


Figura Anexo IX.21 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

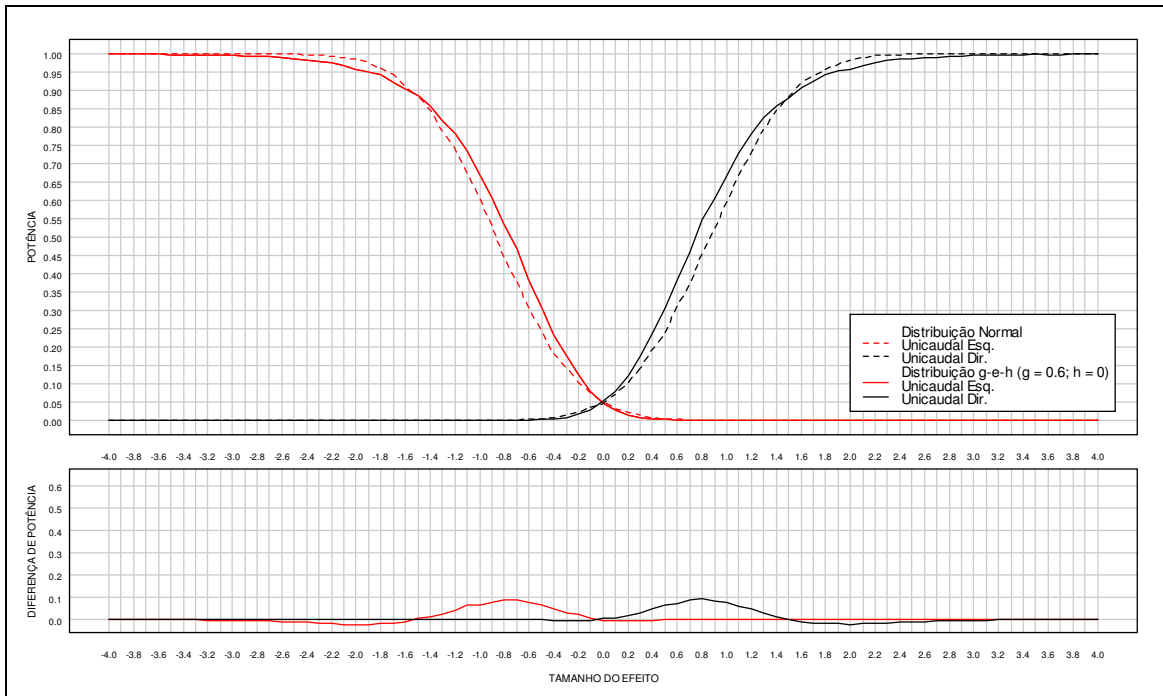


Figura Anexo IX.22 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

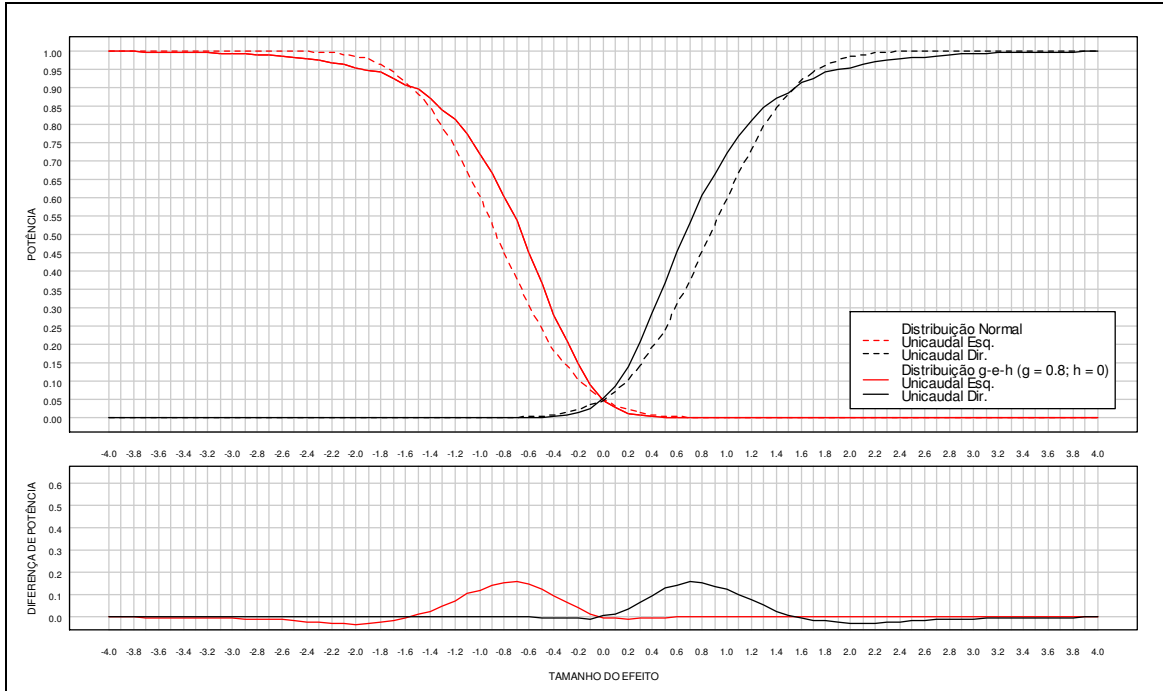


Figura Anexo IX.23 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

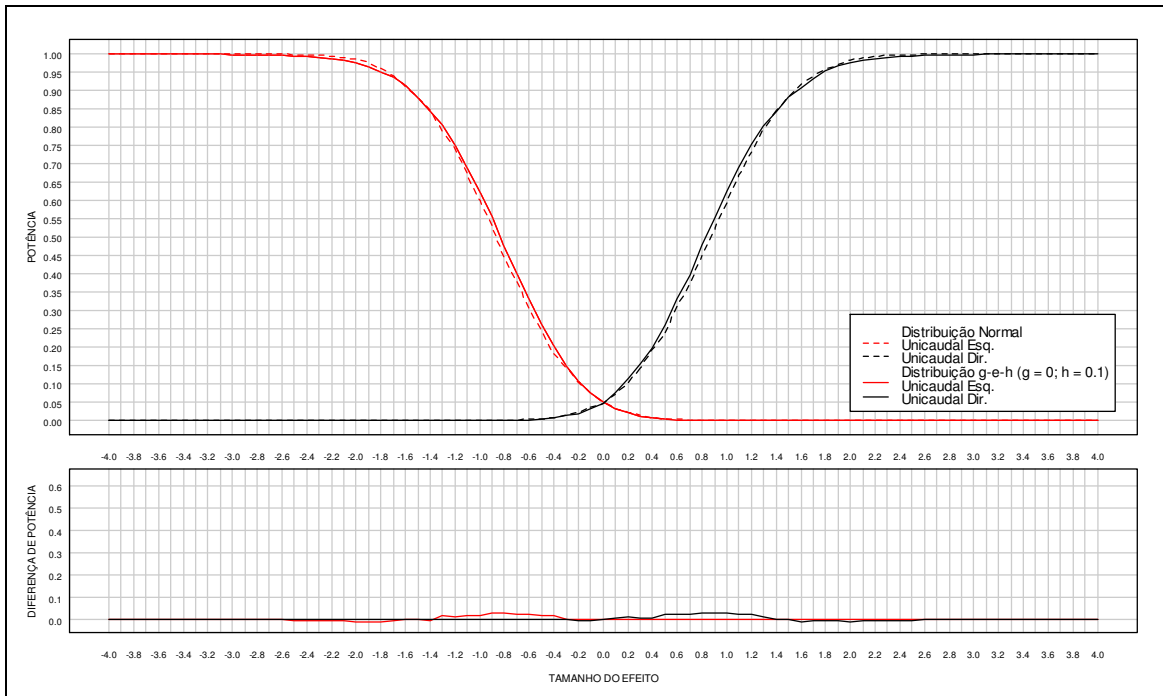


Figura Anexo IX.24 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

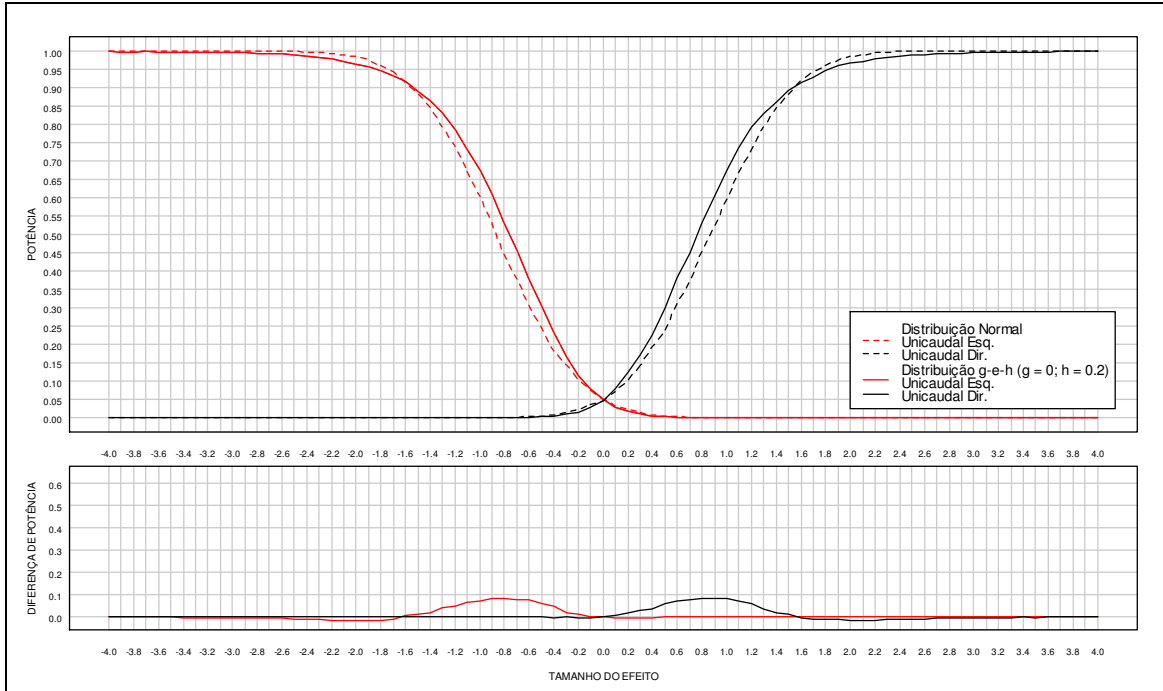


Figura Anexo IX.25 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

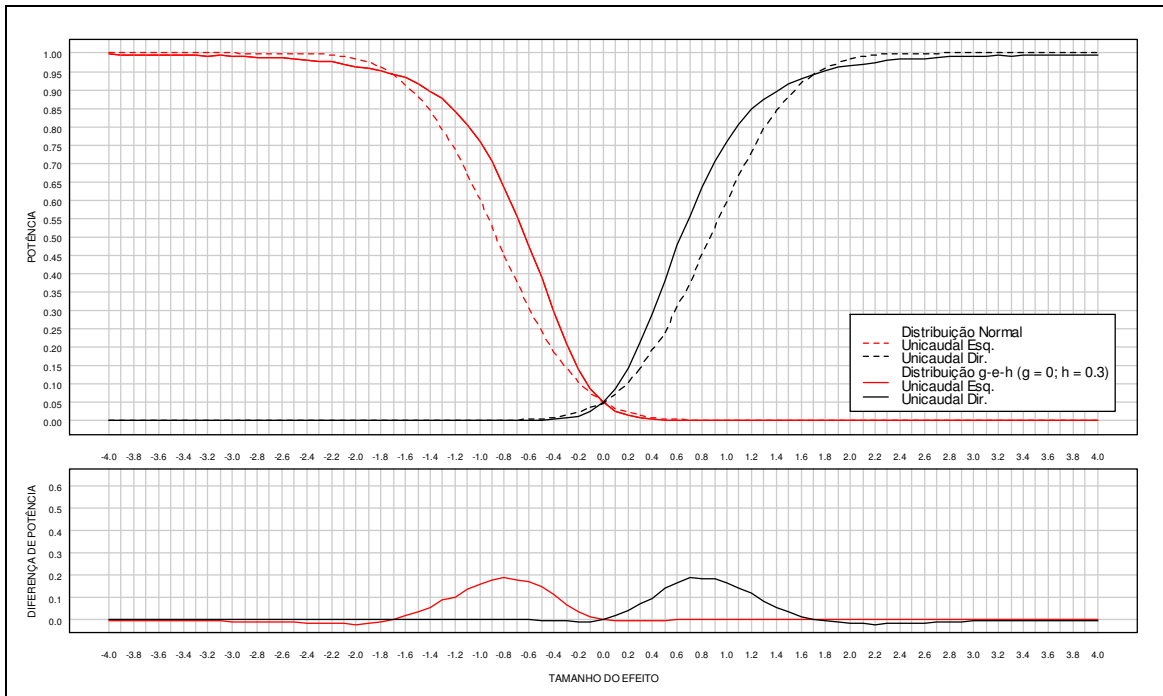


Figura Anexo IX.26 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

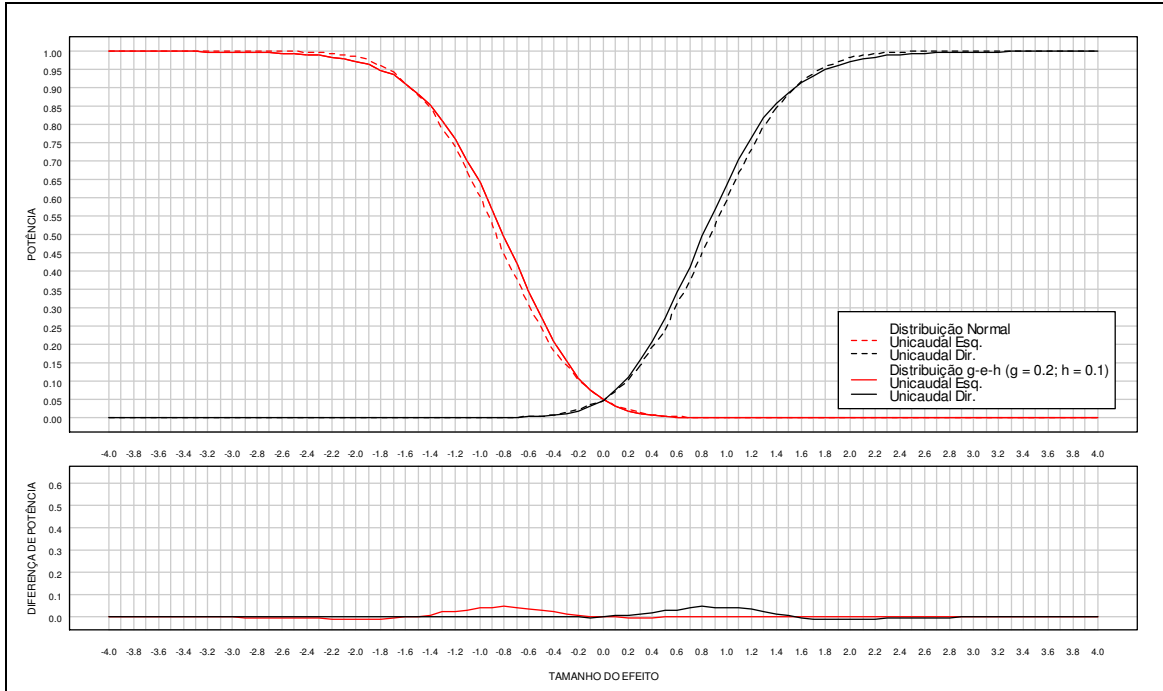


Figura Anexo IX.27 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

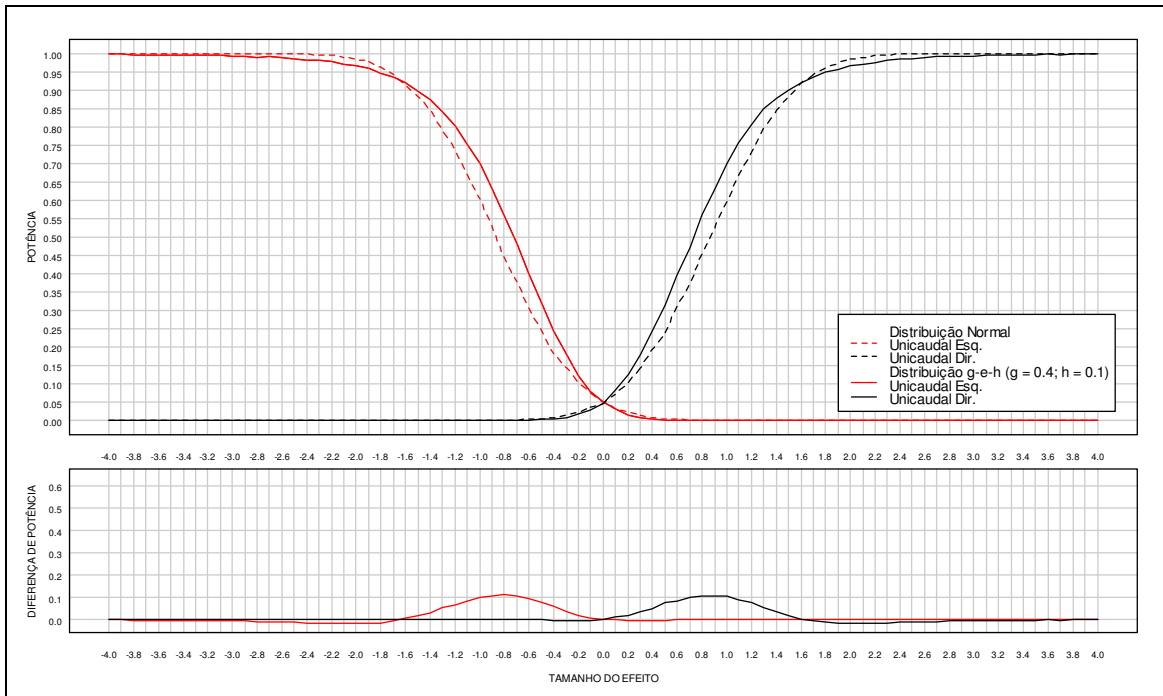


Figura Anexo IX.28 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

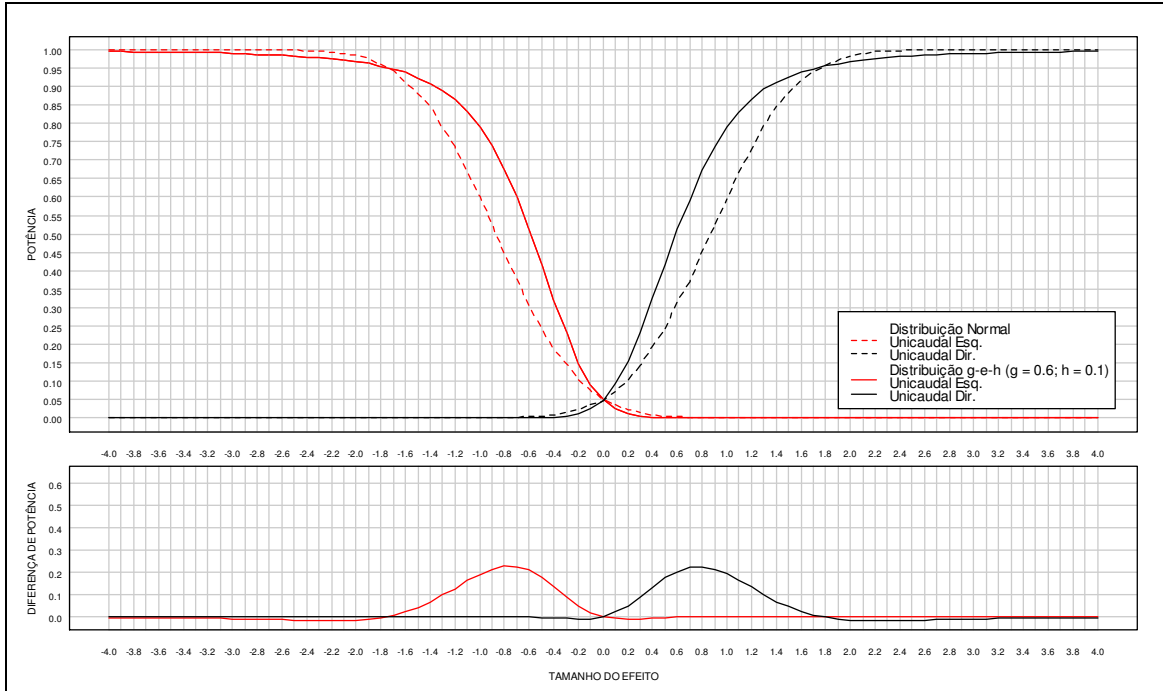


Figura Anexo IX.29 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

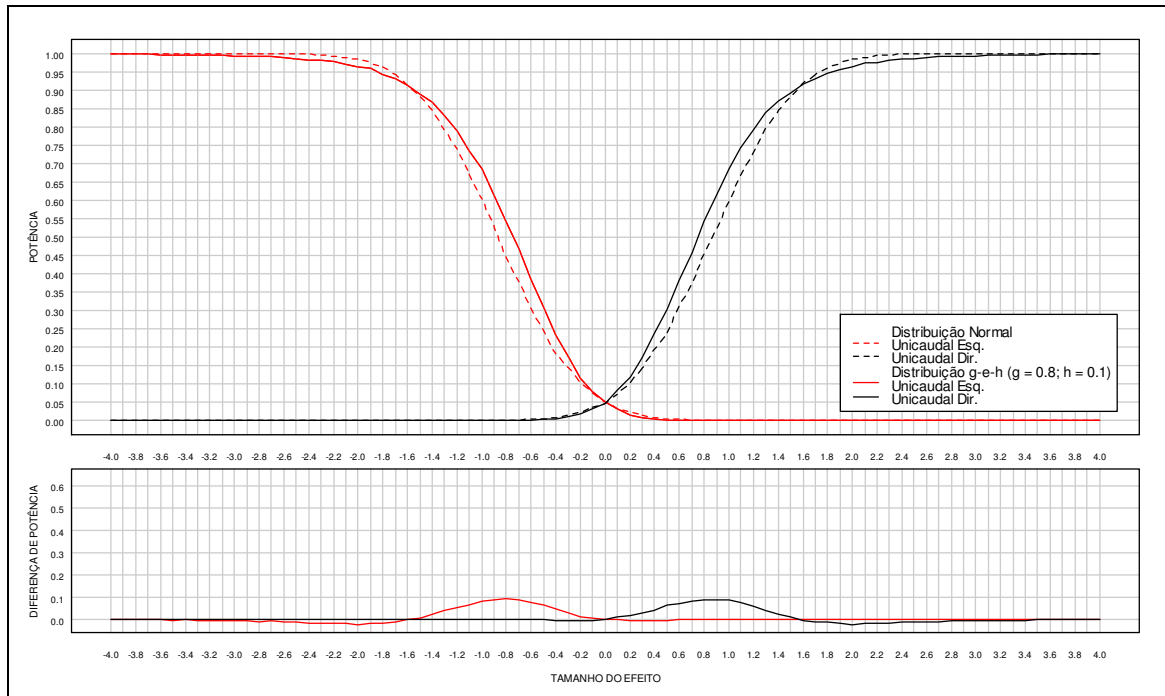


Figura Anexo IX.30 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

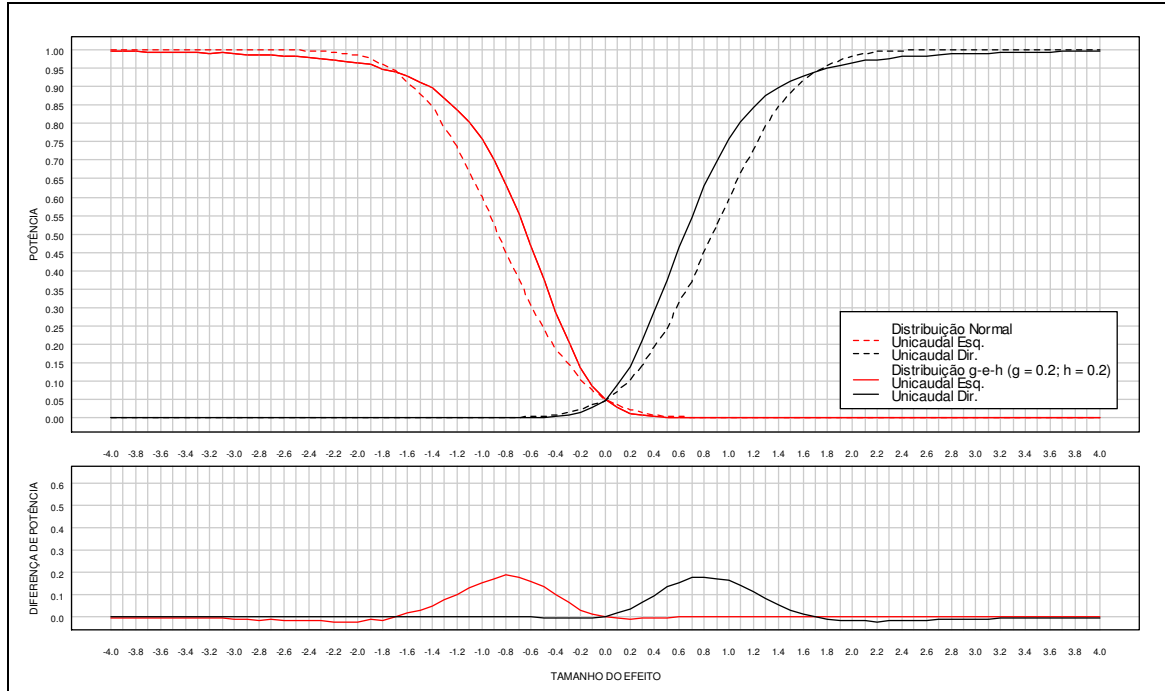


Figura Anexo IX.31 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

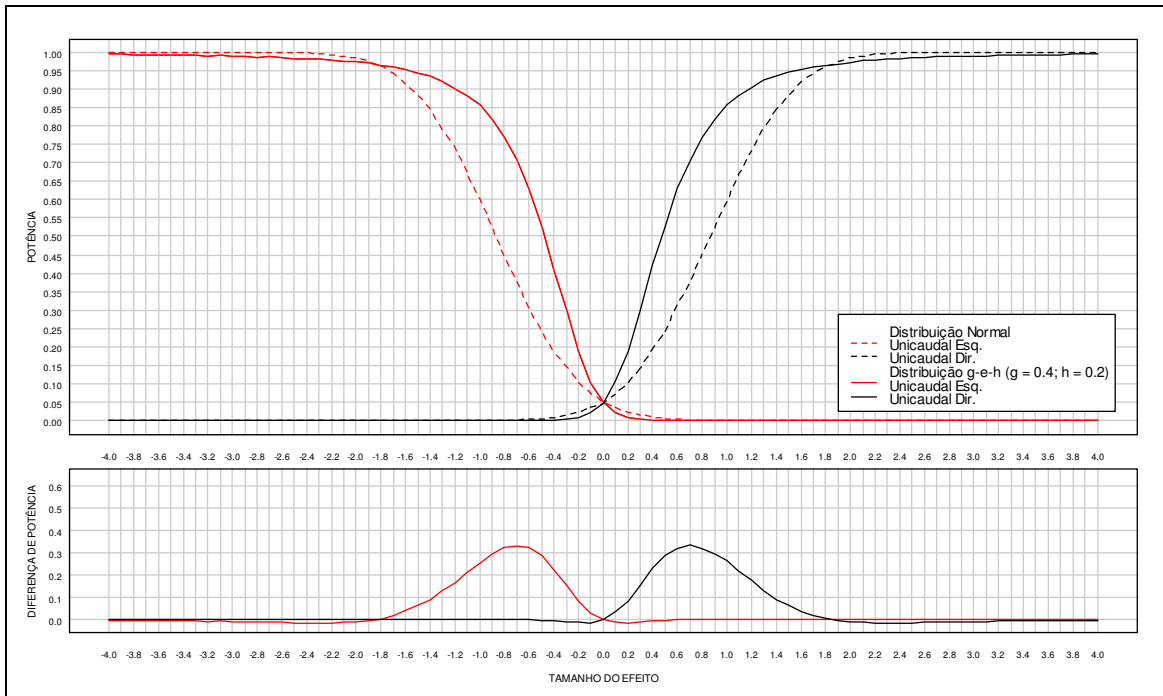


Figura Anexo IX.32 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

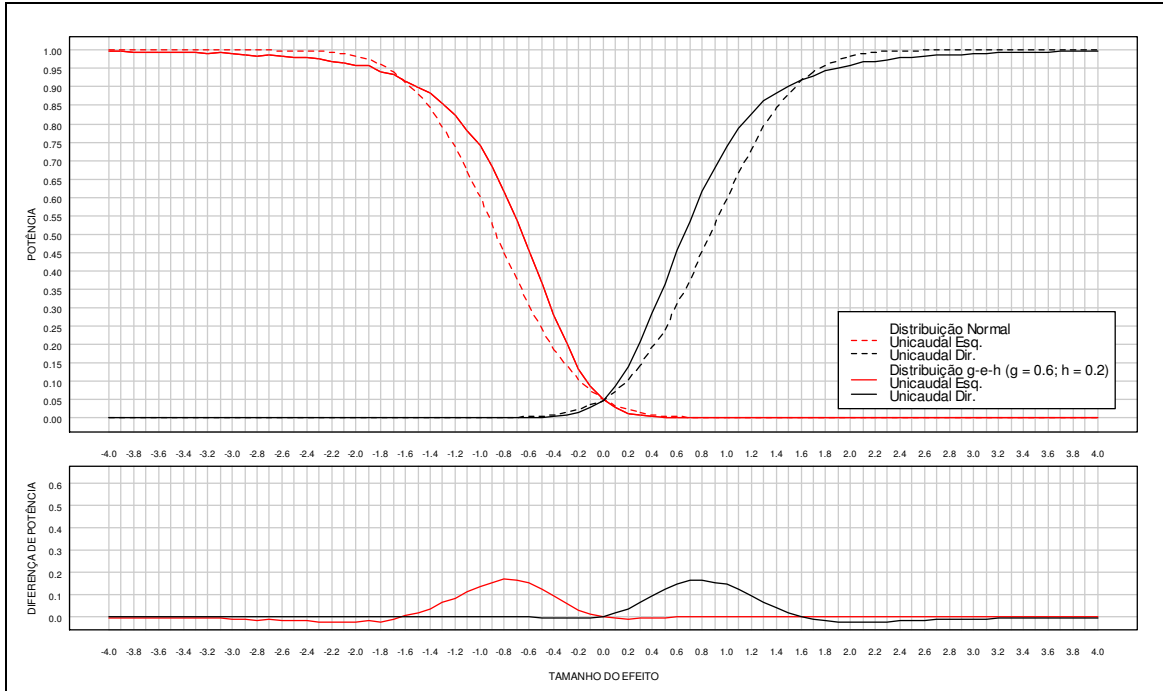


Figura Anexo IX.33 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

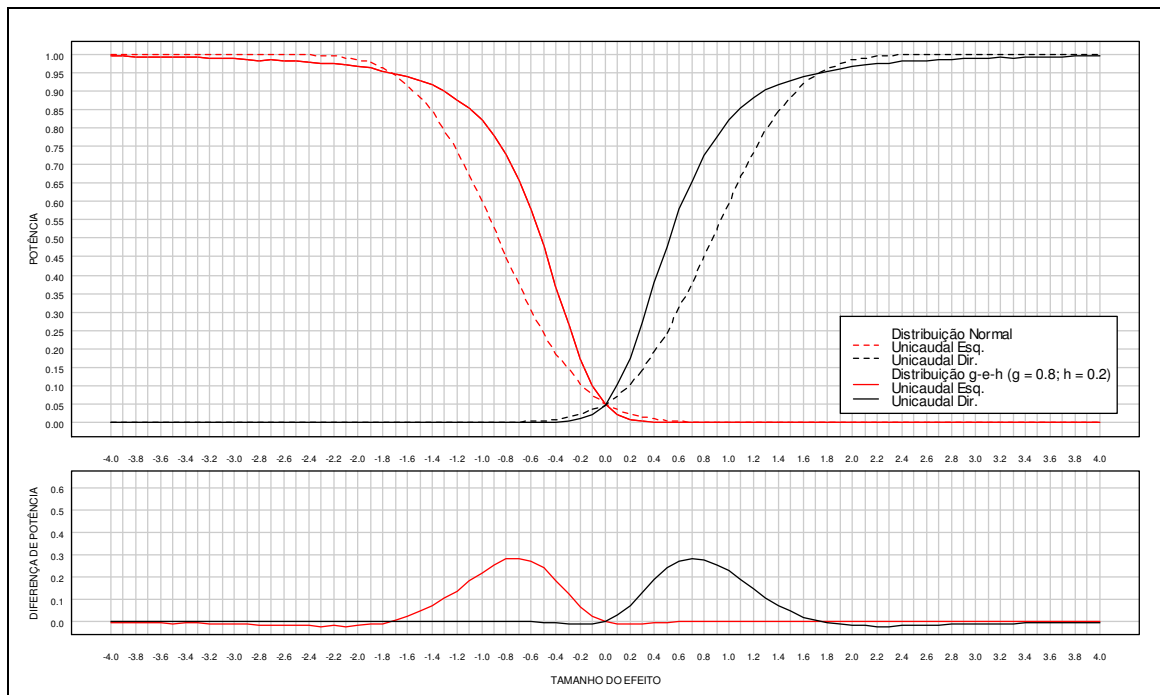


Figura Anexo IX.34 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

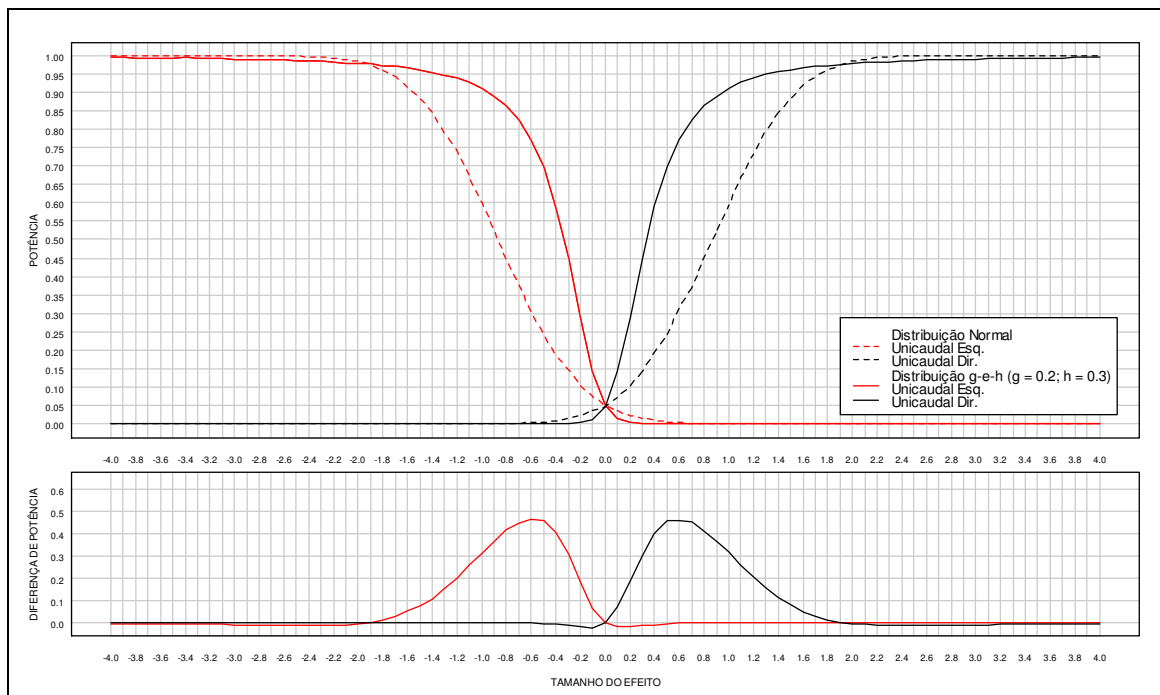


Figura Anexo IX.35 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

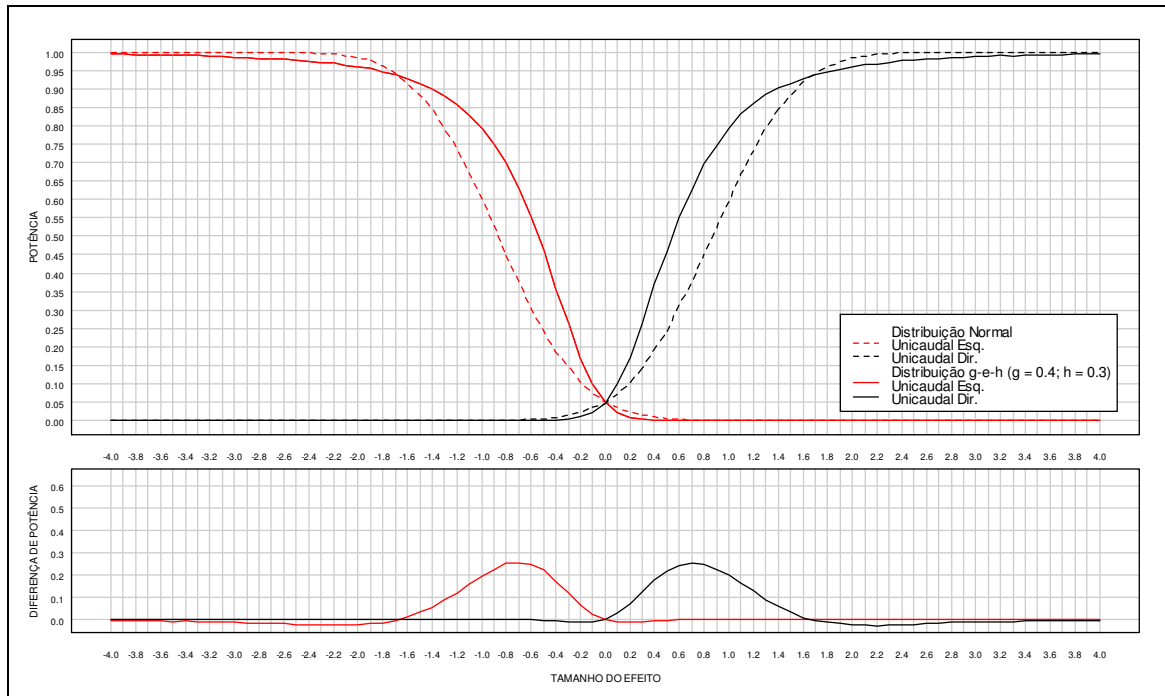


Figura Anexo IX.36 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

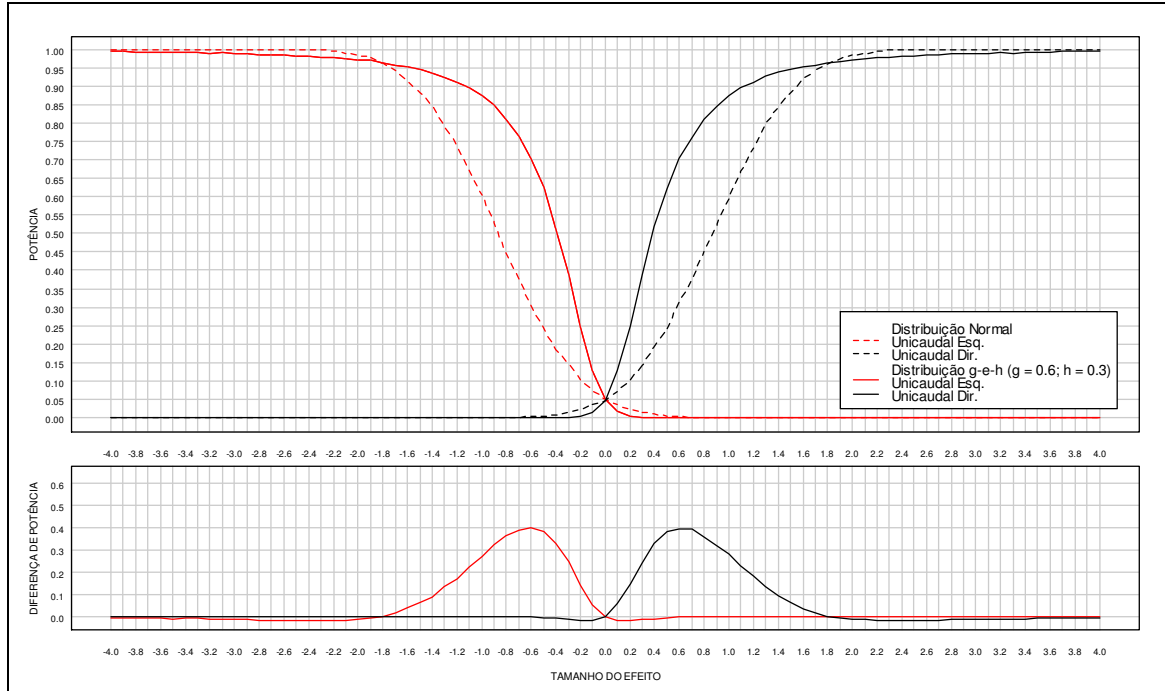


Figura Anexo IX.37 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

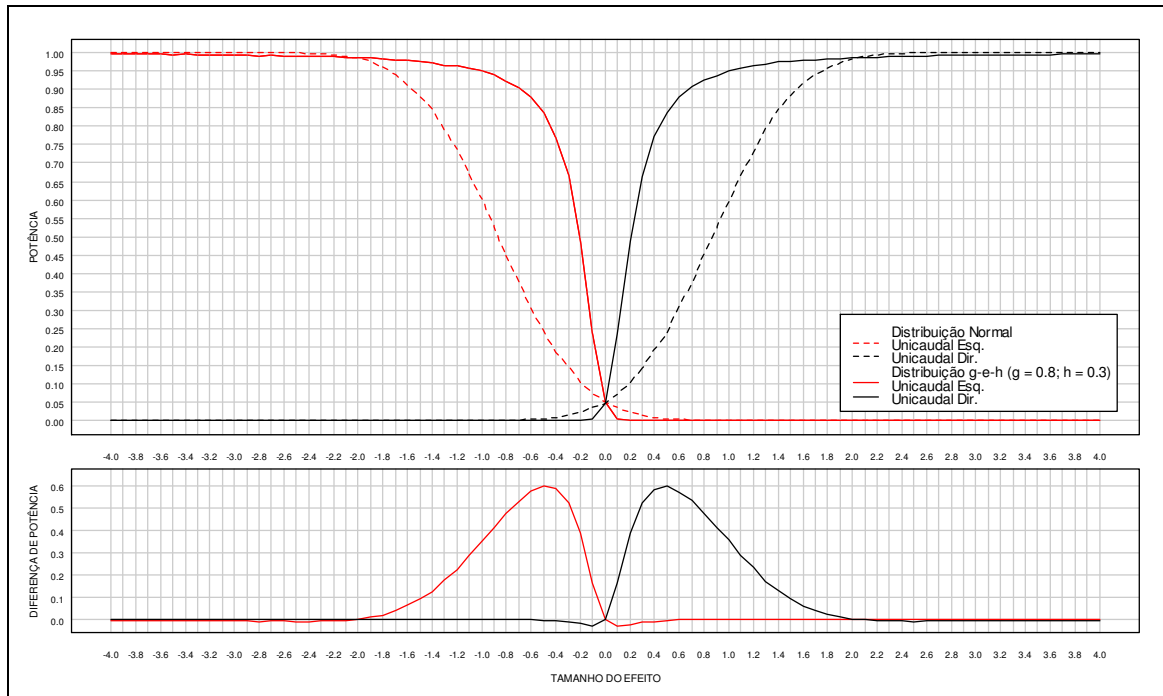


Figura Anexo IX.38 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 8$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

2. Amostras com $n_1 = n_2 = 16$

2.1. Testes bi-caudais

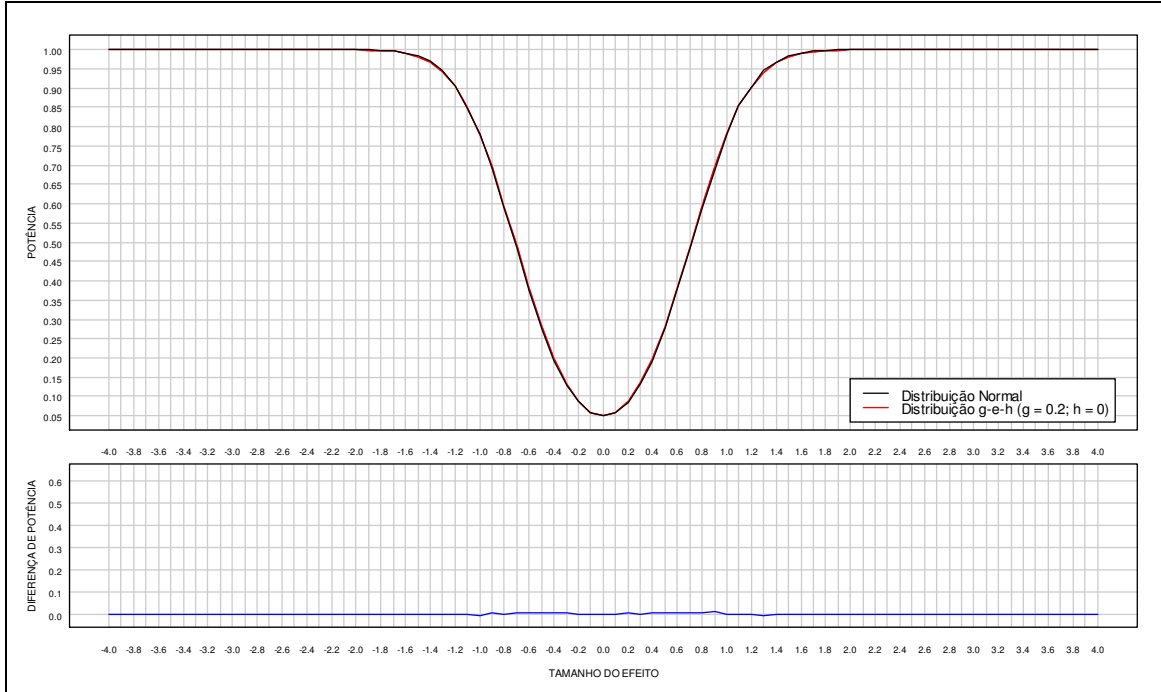


Figura Anexo IX.39 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

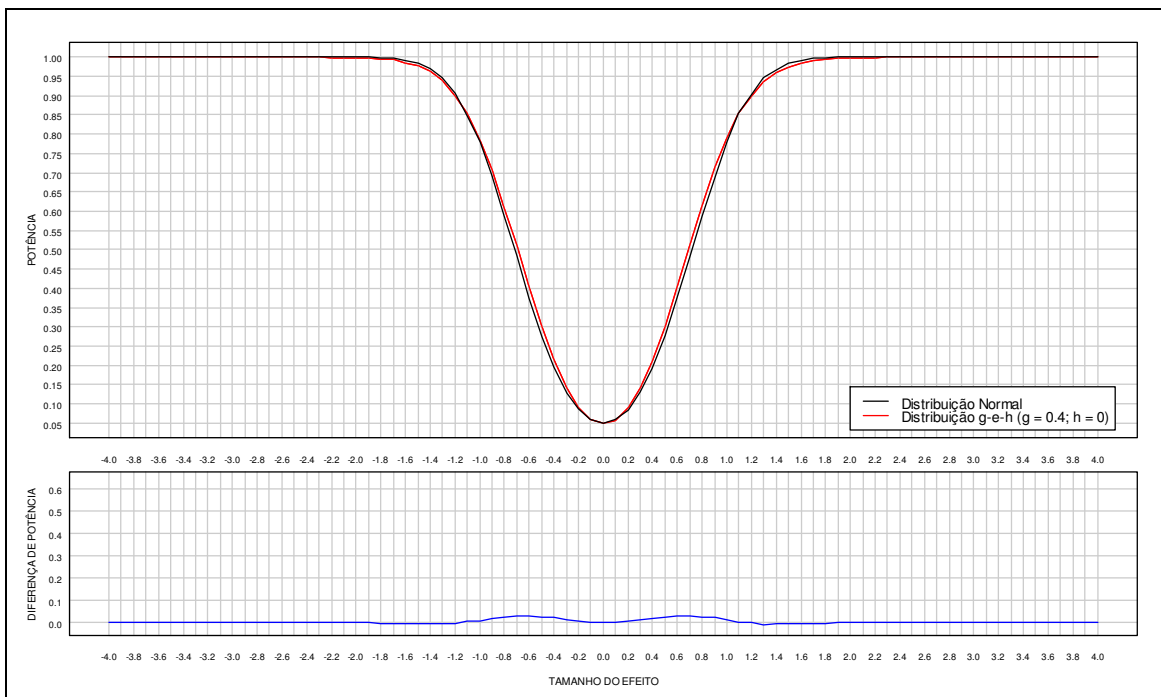


Figura Anexo IX.40 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

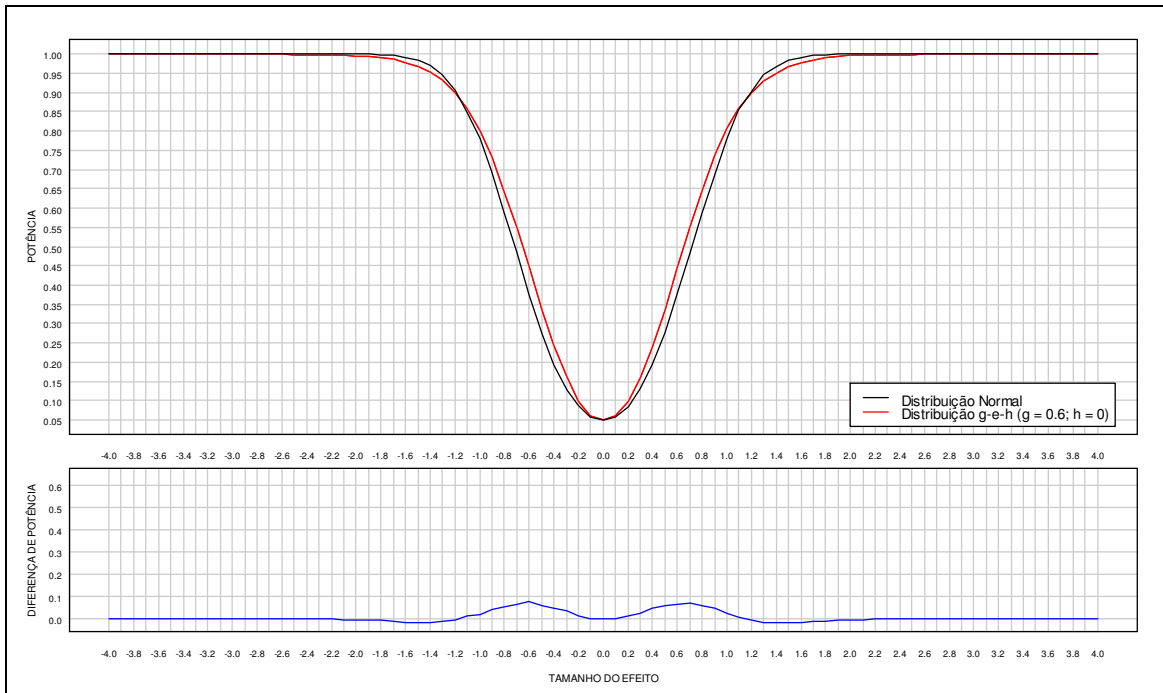


Figura Anexo IX.41 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

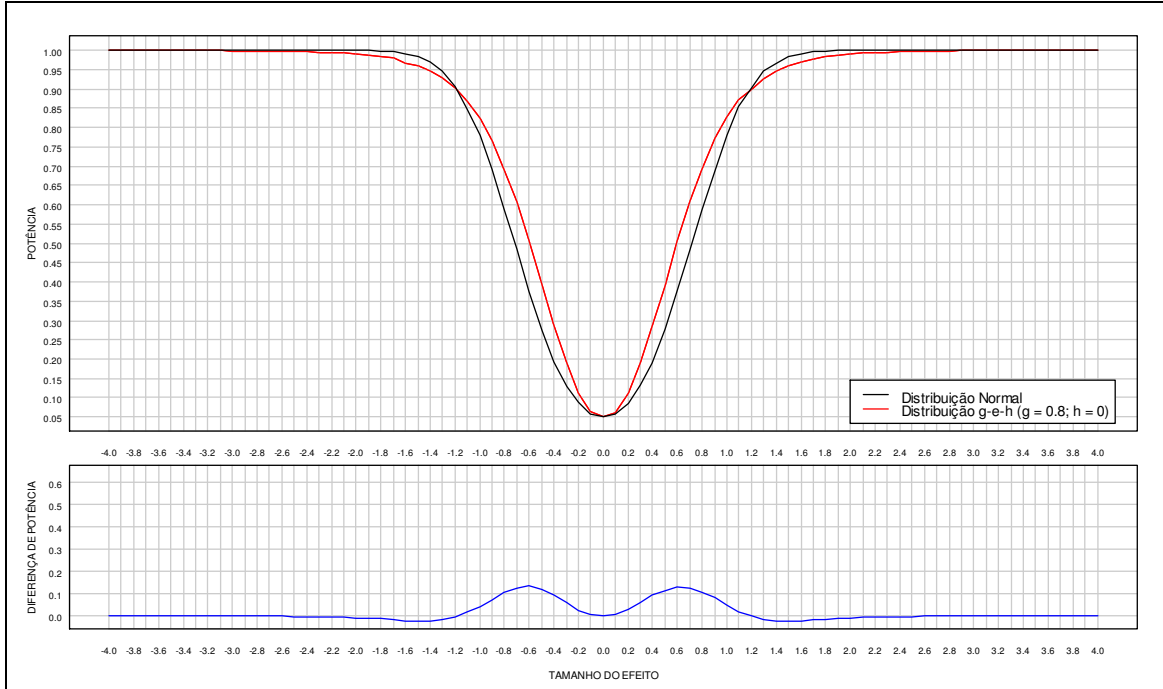


Figura Anexo IX.42 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

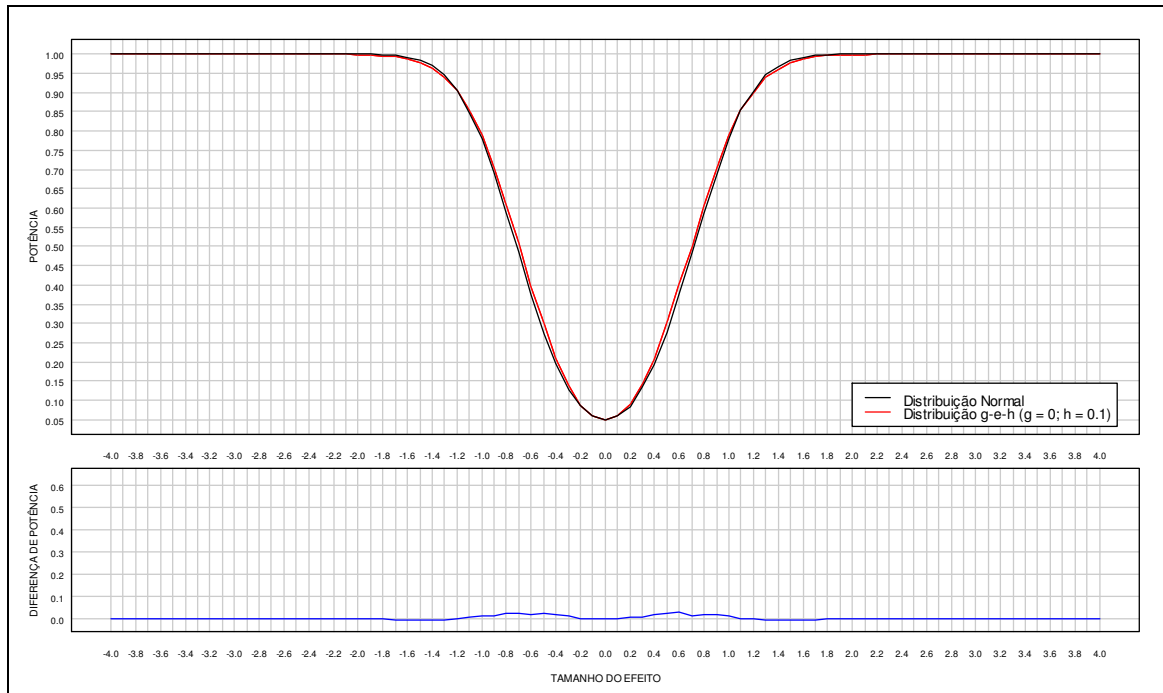


Figura Anexo IX.43 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

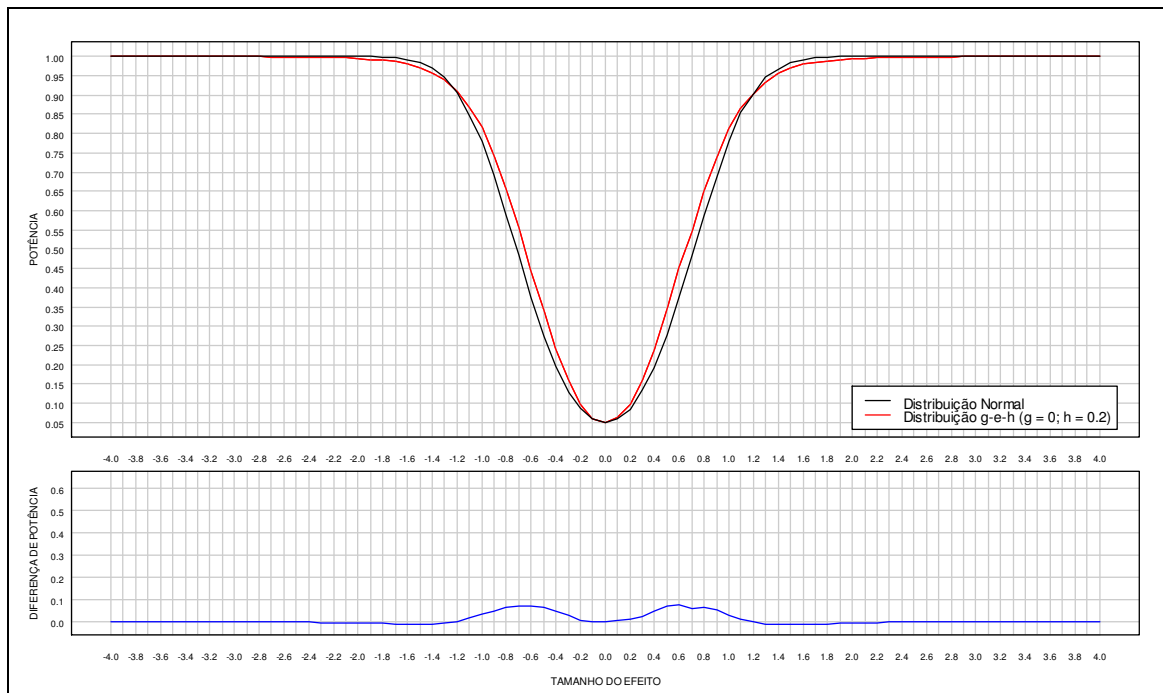


Figura Anexo IX.44 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

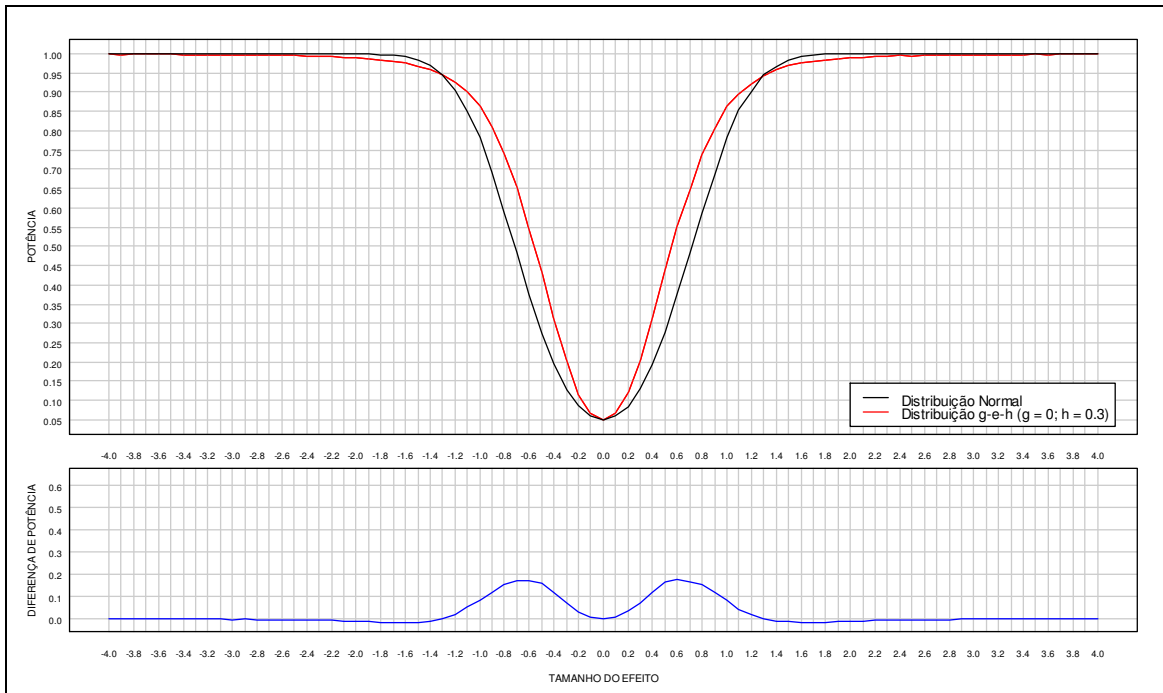


Figura Anexo IX.45 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

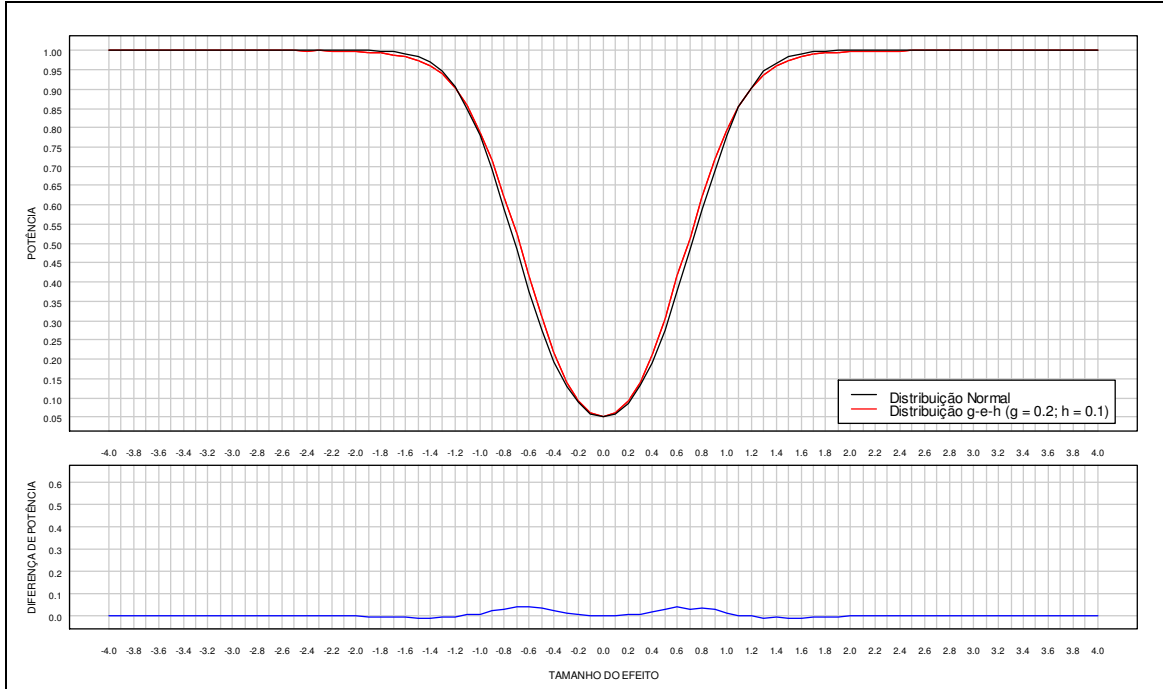


Figura Anexo IX.46 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

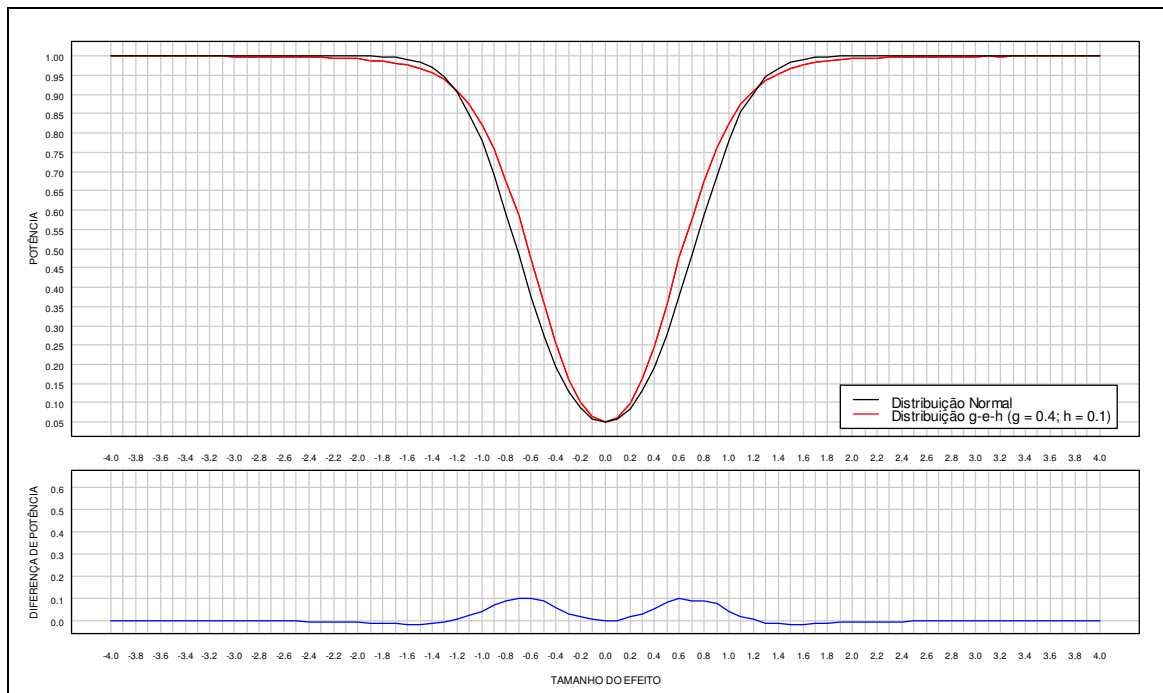


Figura Anexo IX.47 [Painel superior] Potência do teste de aleatorização bi-causal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

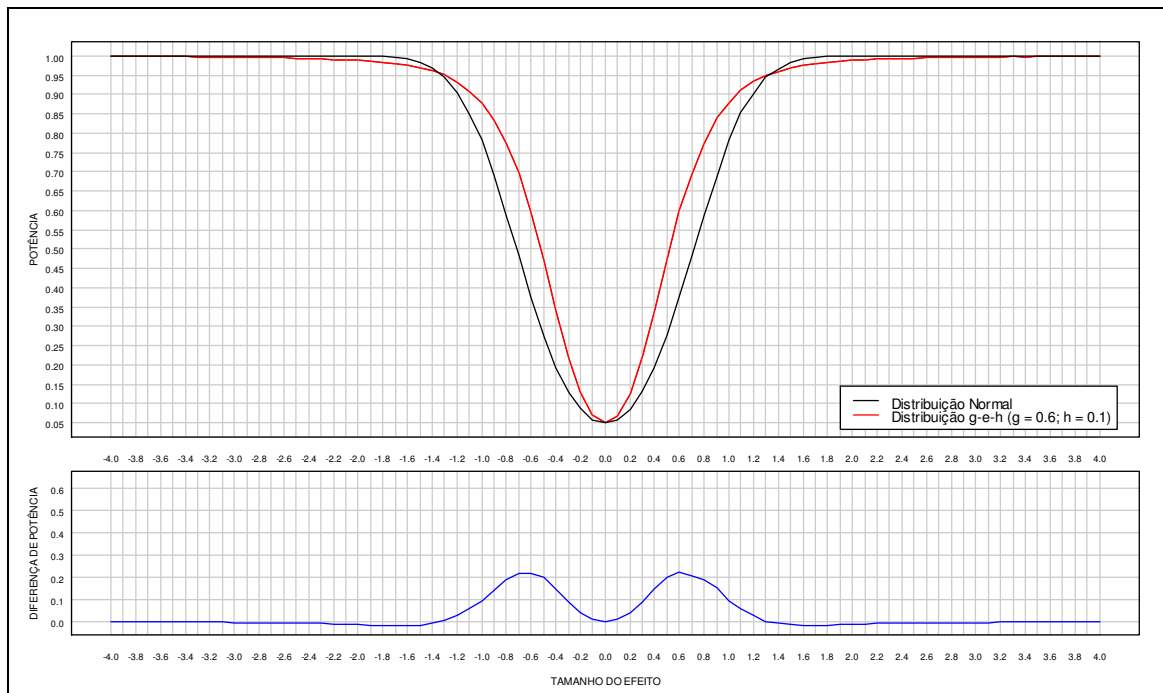


Figura Anexo IX.48 [Painel superior] Potência do teste de aleatorização bi-causal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

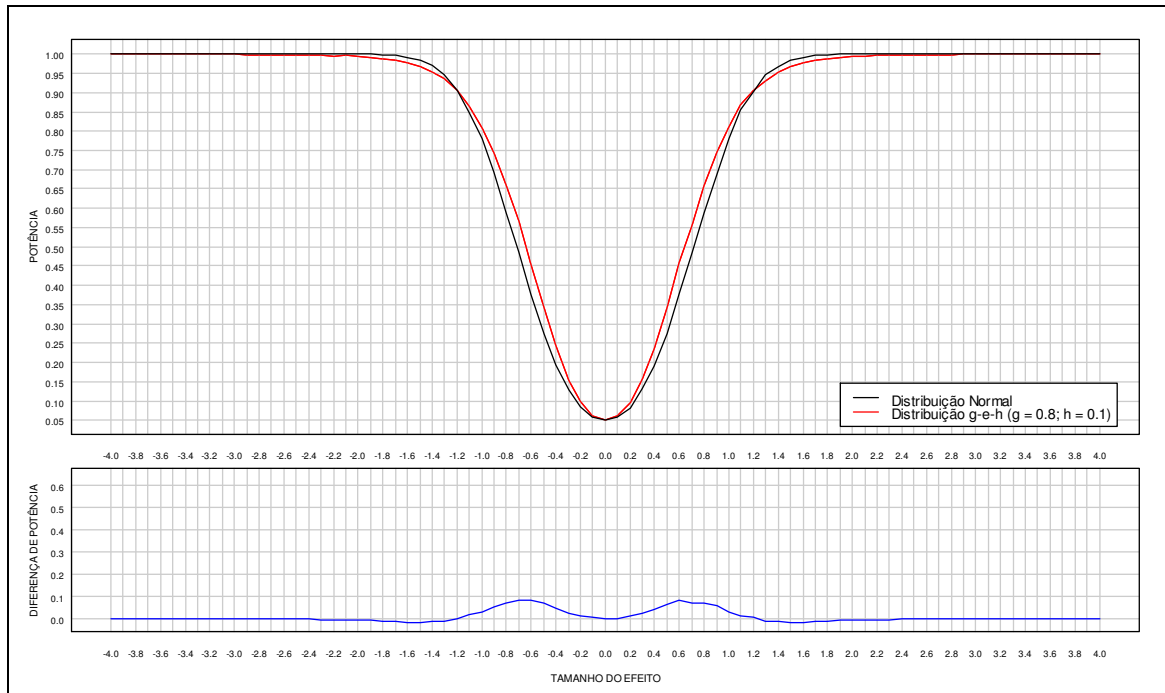


Figura Anexo IX.49 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

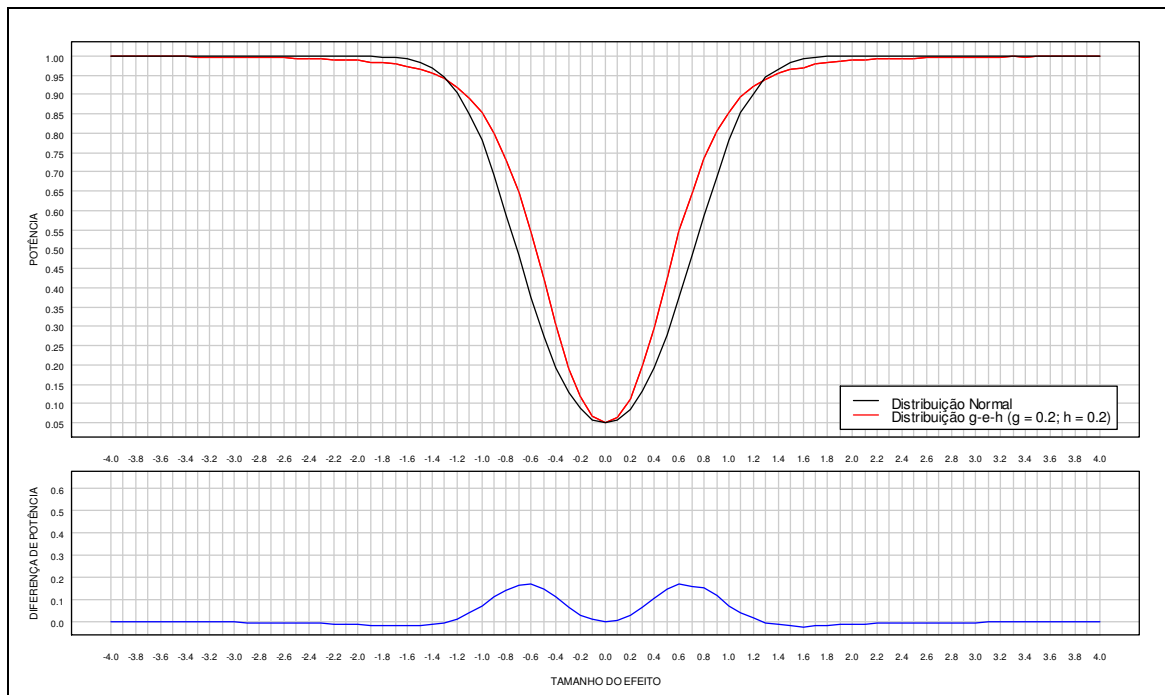


Figura Anexo IX.50 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

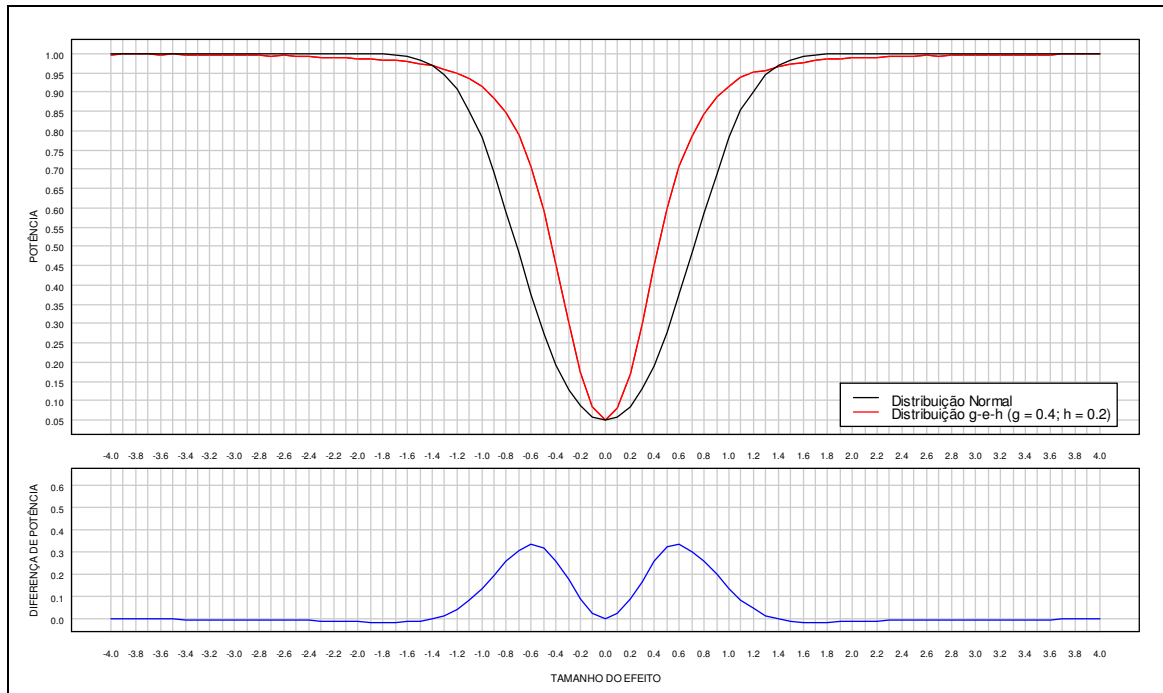


Figura Anexo IX.51 [Painel superior] Potência do teste de aleatorização bi-causal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

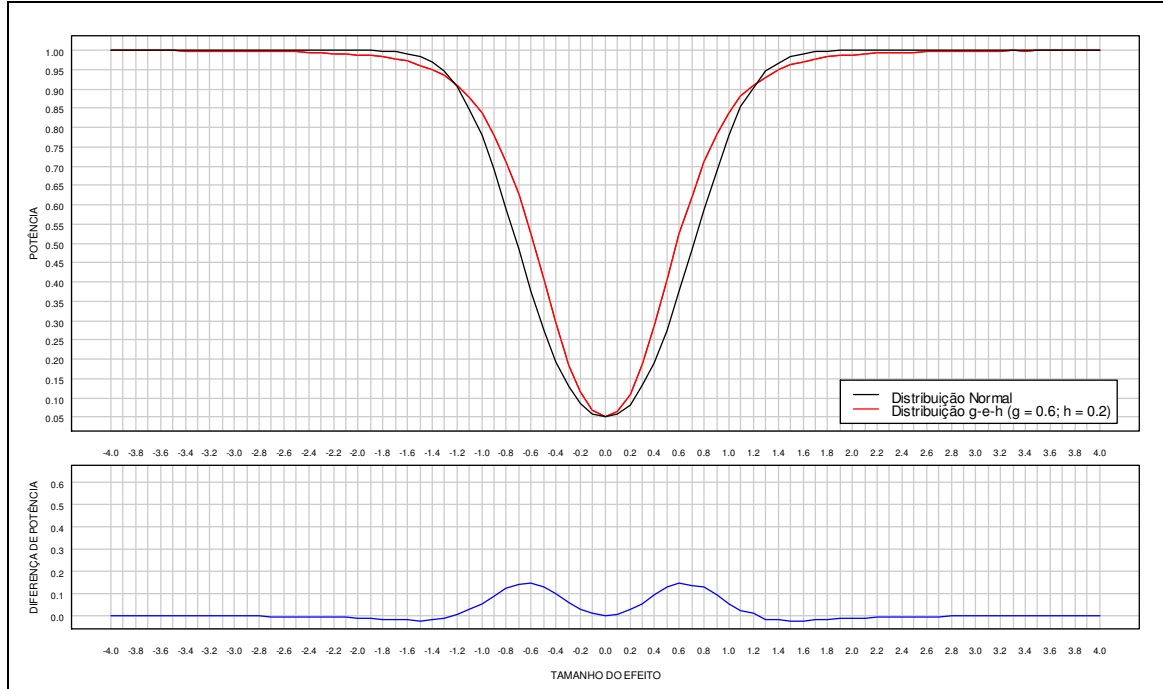


Figura Anexo IX.52 [Painel superior] Potência do teste de aleatorização bi-causal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

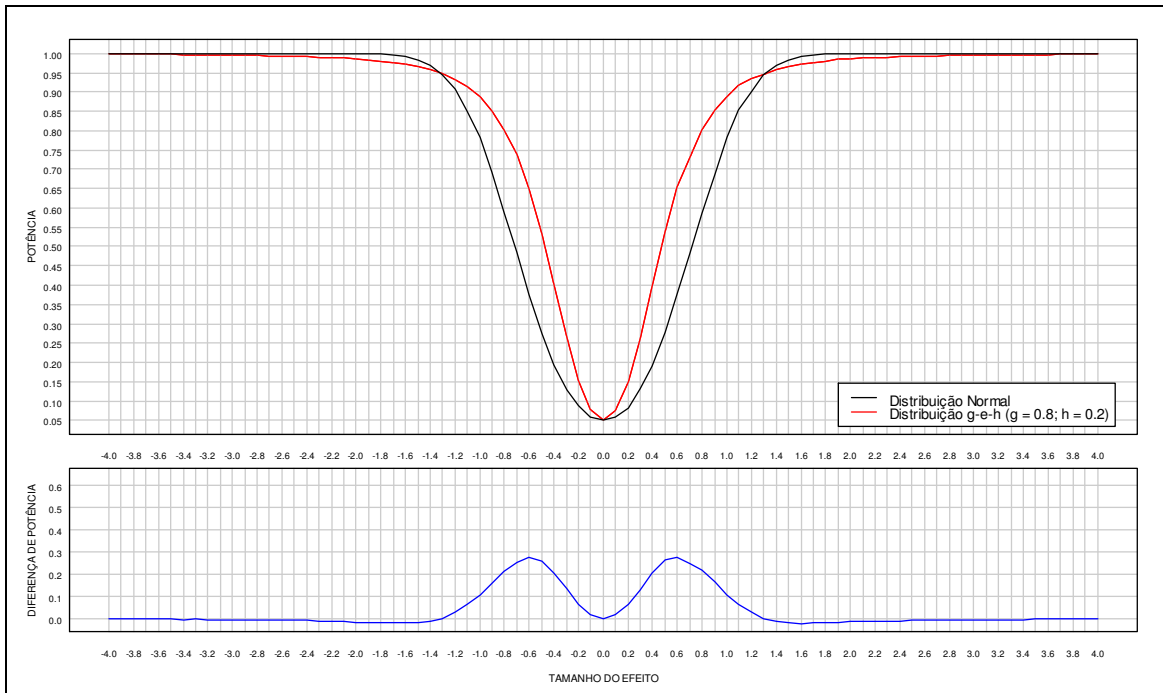


Figura Anexo IX.53 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

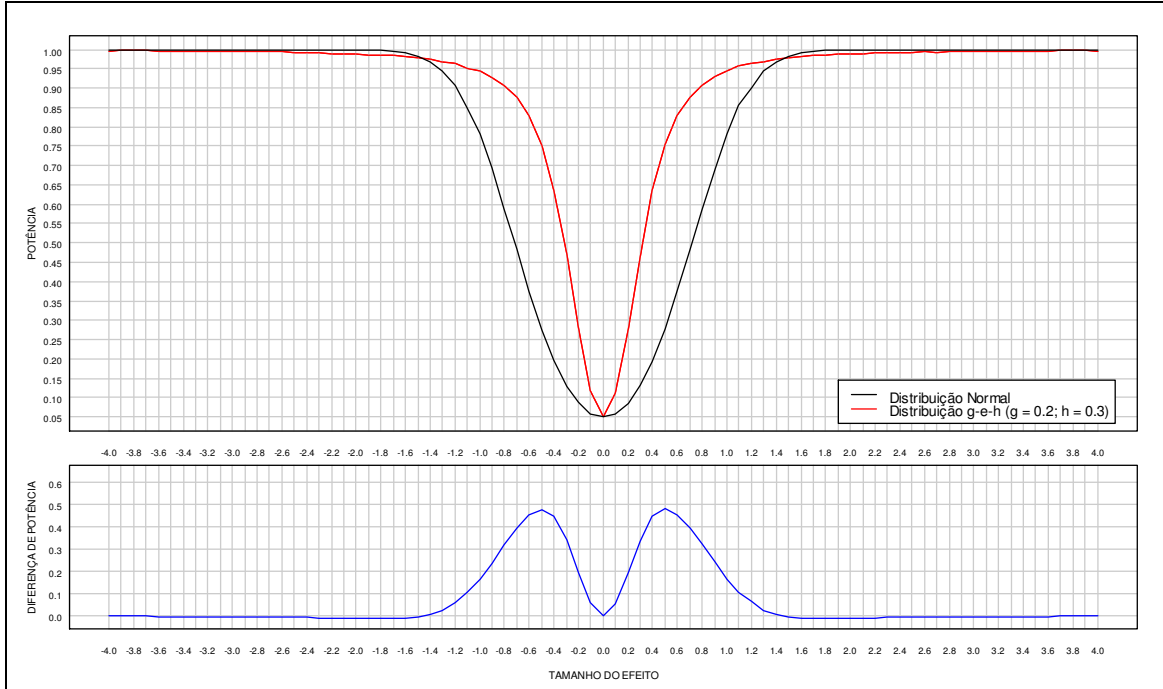


Figura Anexo IX.54 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

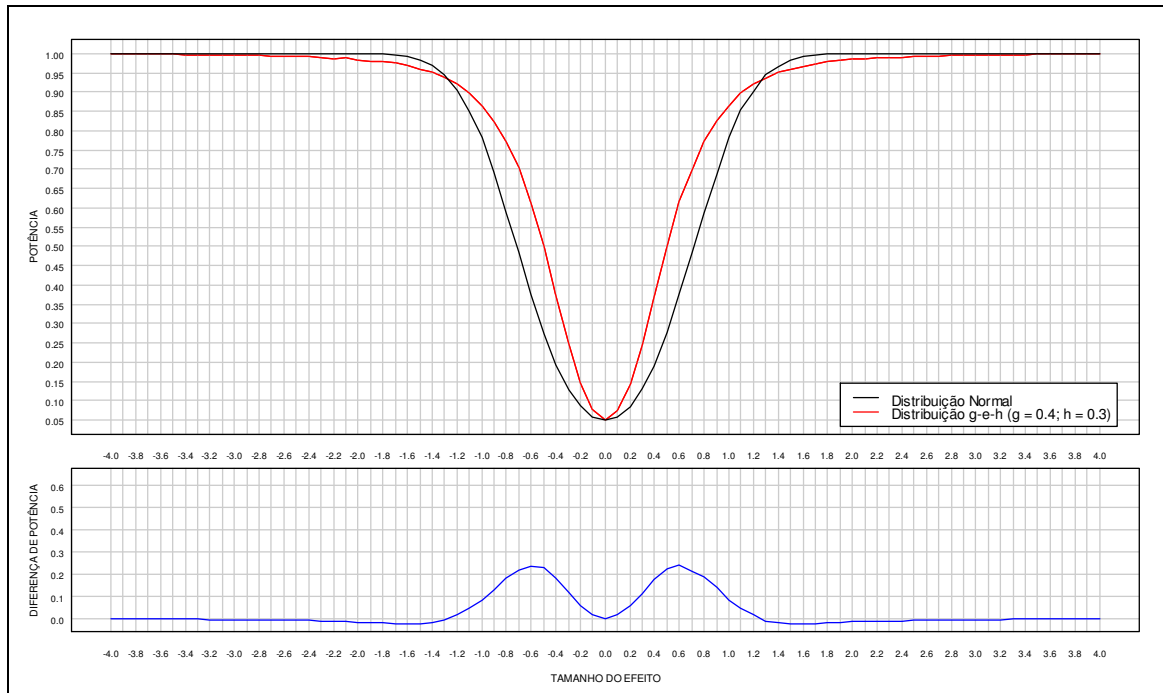


Figura Anexo IX.55 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

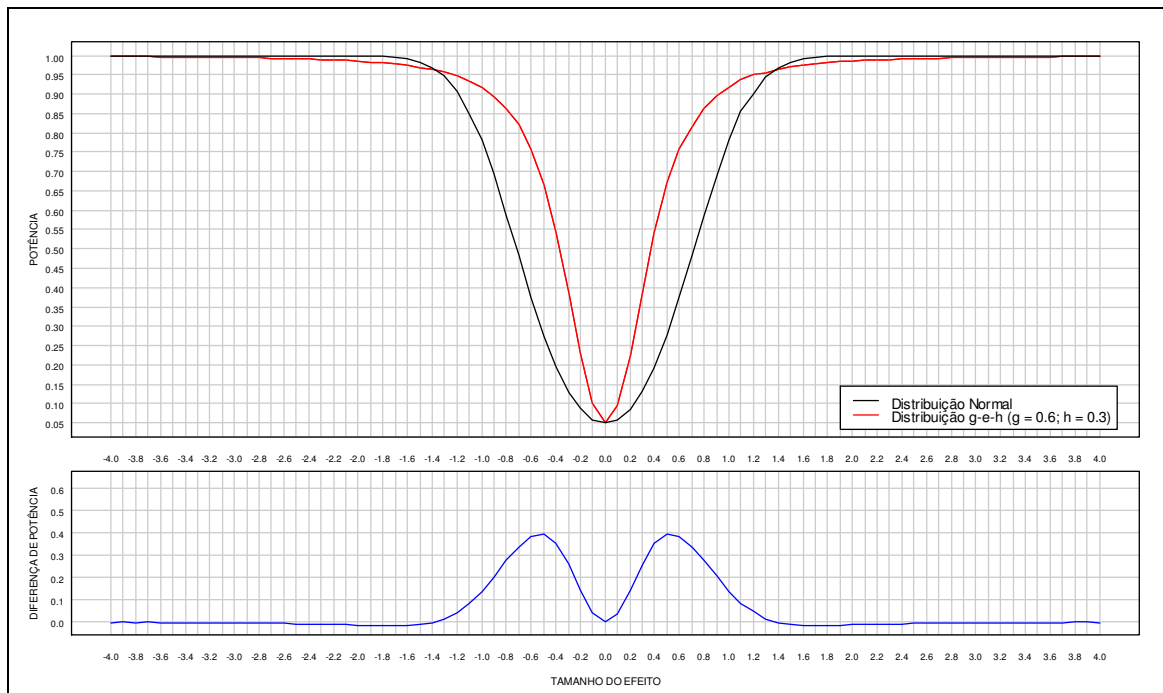


Figura Anexo IX.56 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

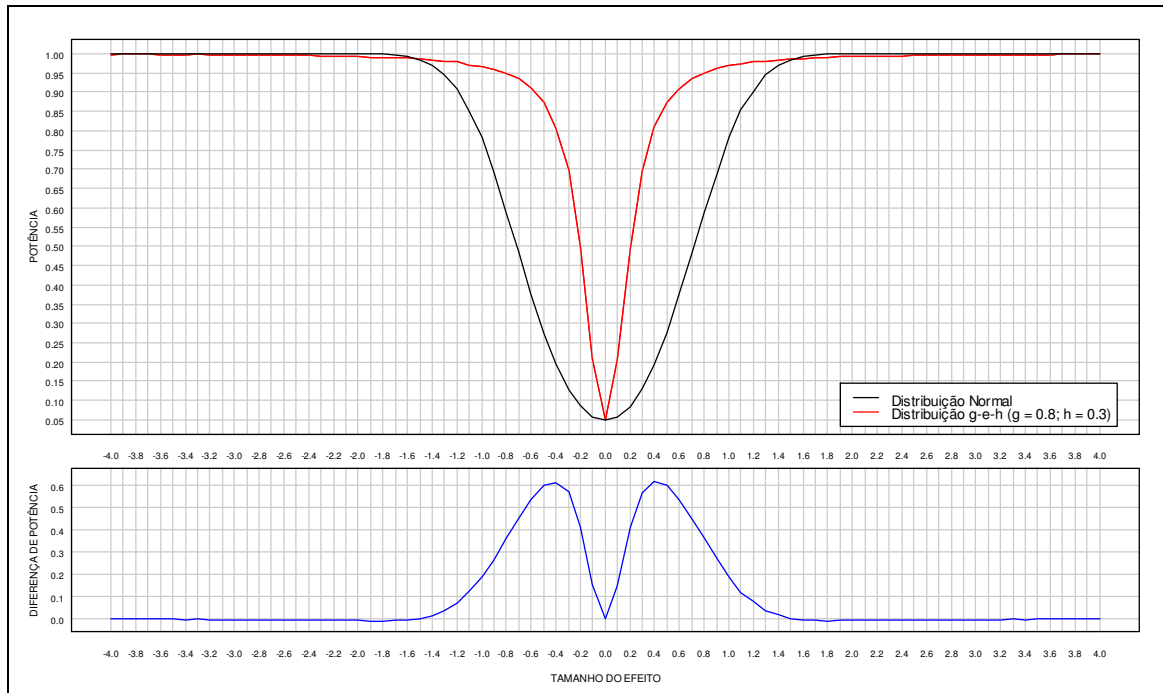


Figura Anexo IX.57 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

2.2. Testes uni-caudais

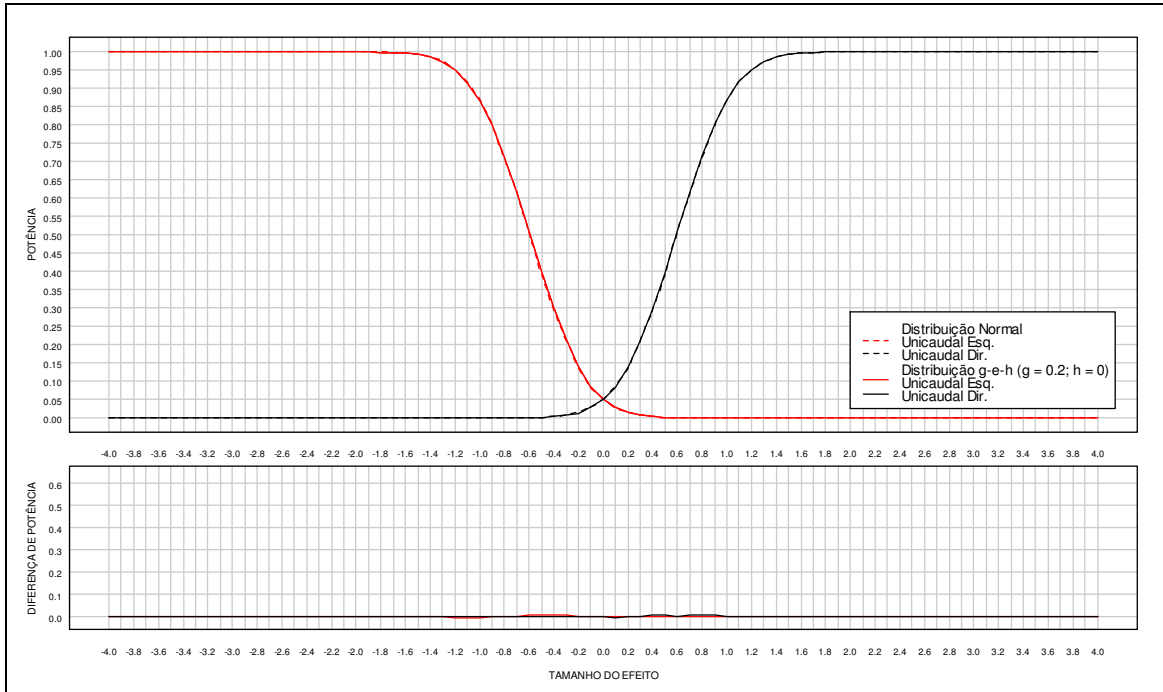


Figura Anexo IX.58 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

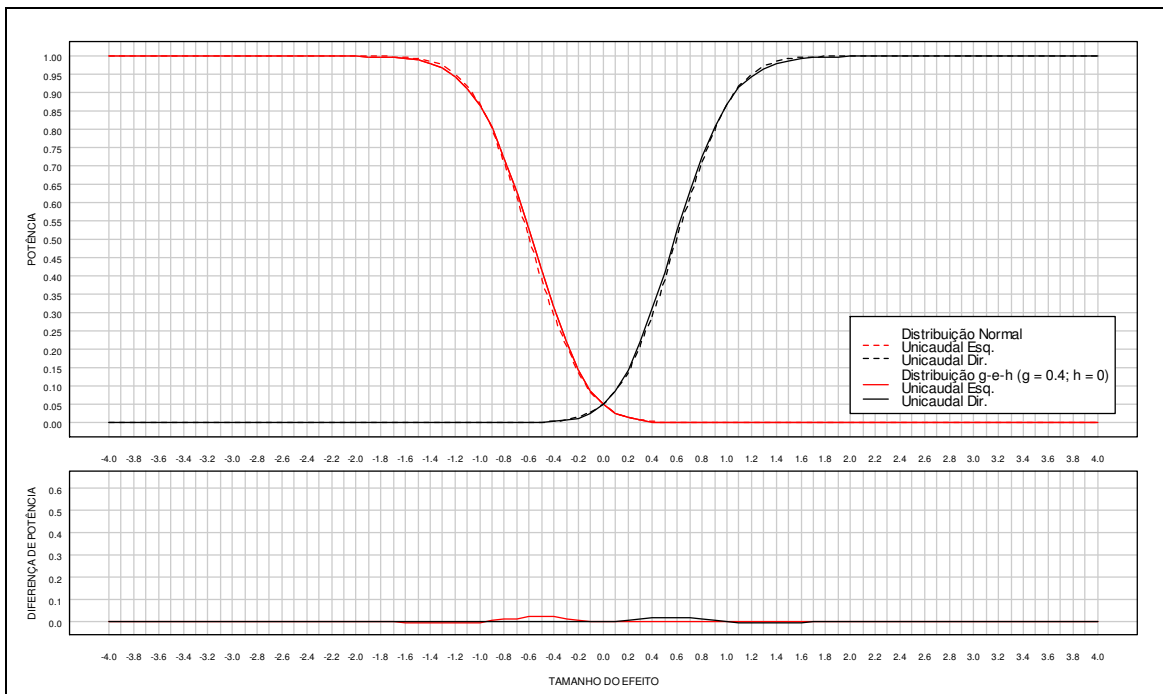


Figura Anexo IX.59 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

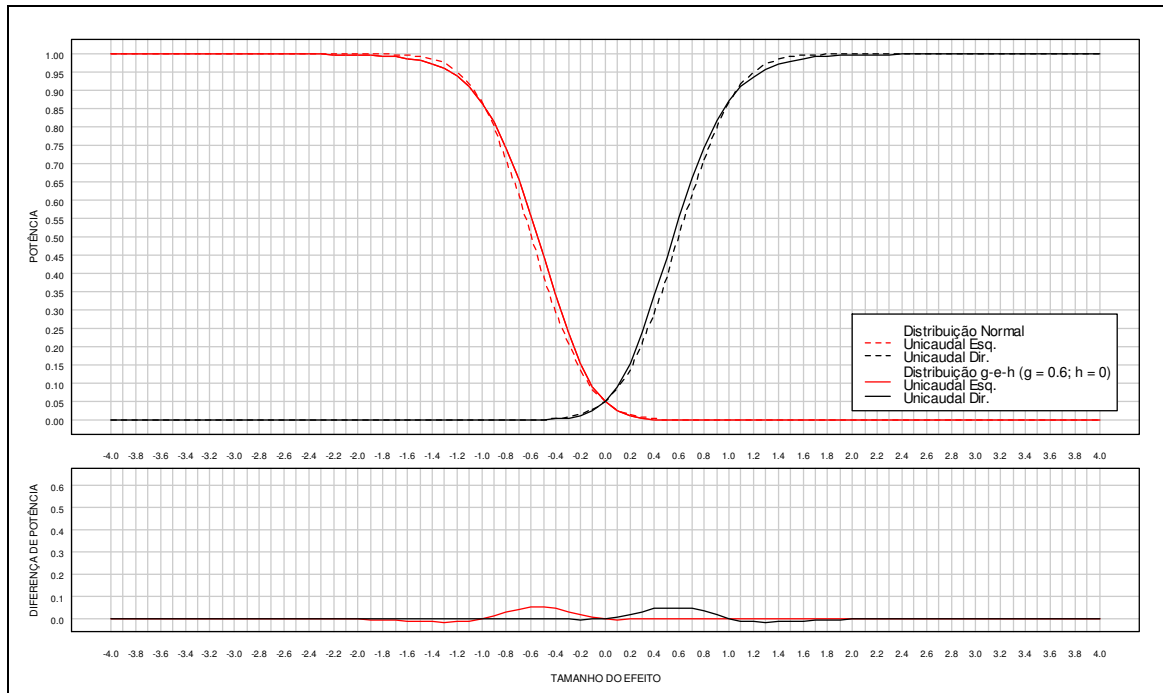


Figura Anexo IX.60 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

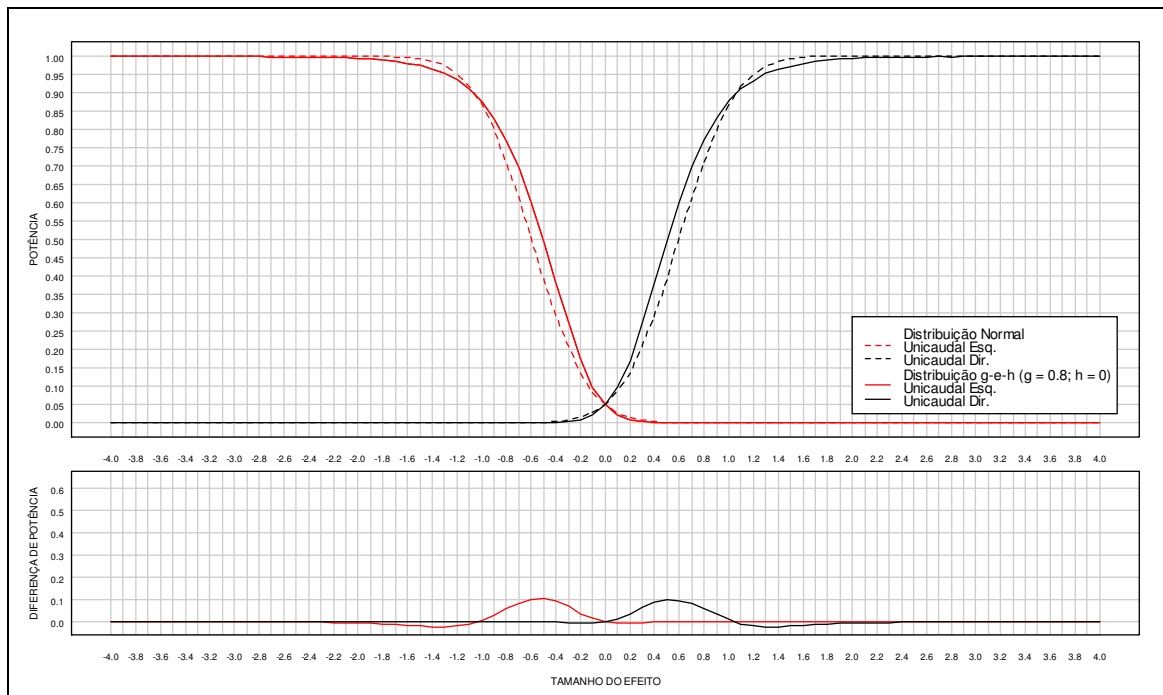


Figura Anexo IX.61 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

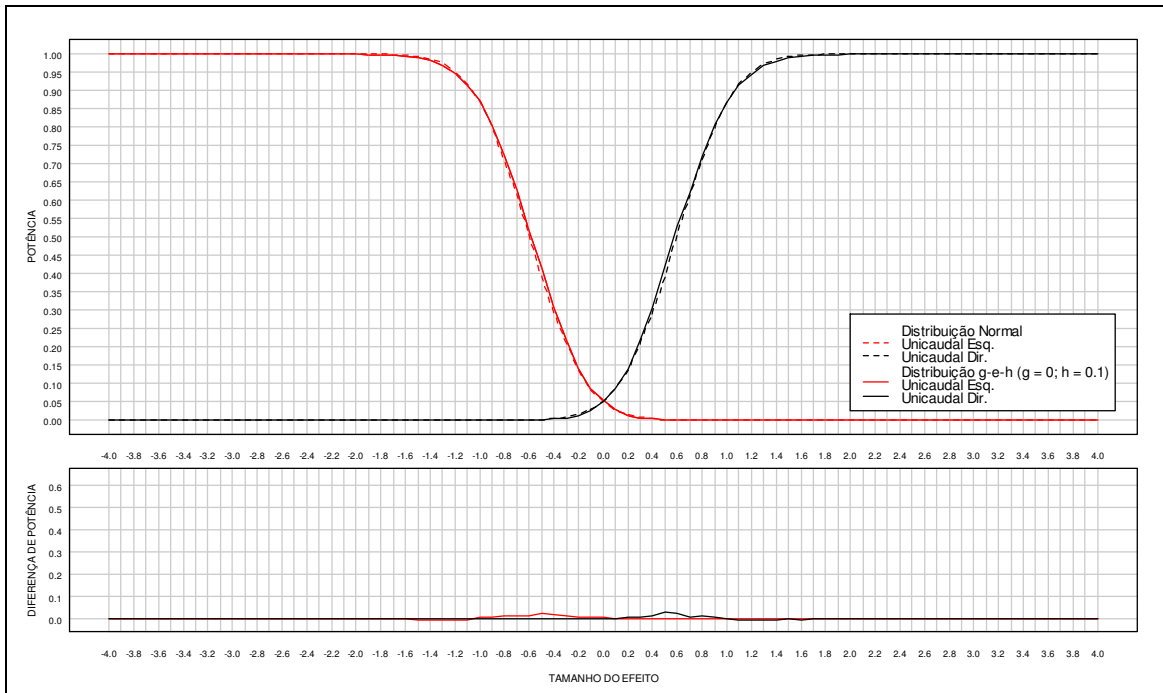


Figura Anexo IX.62 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

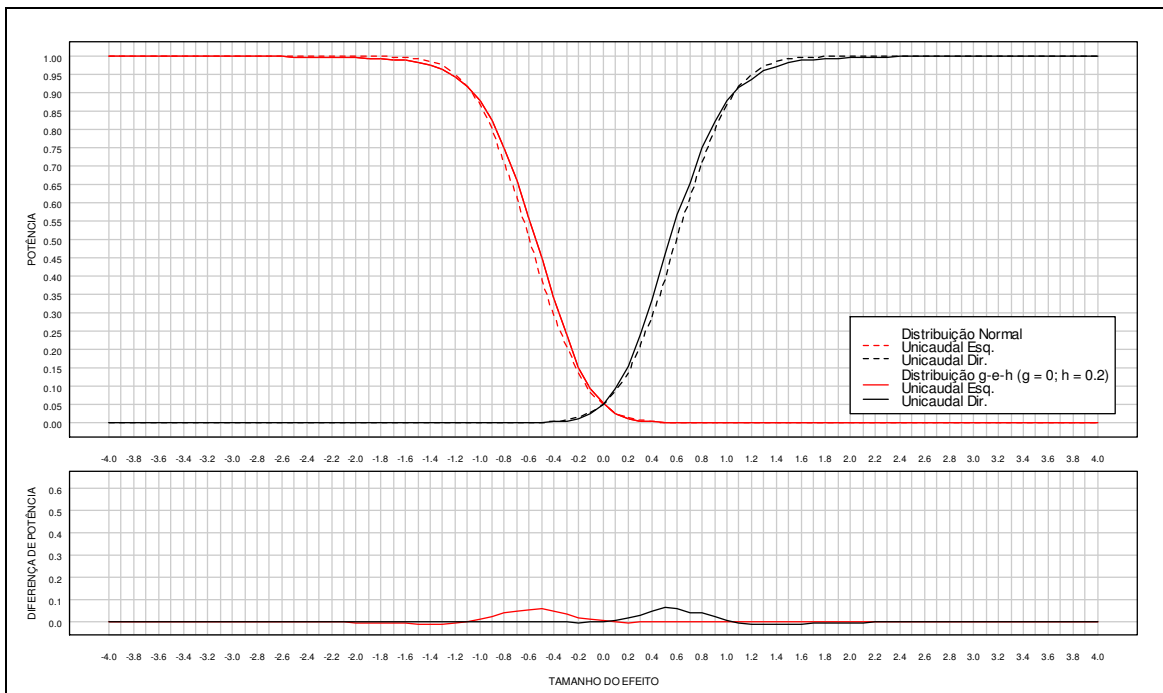


Figura Anexo IX.63 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.



Figura Anexo IX.64 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

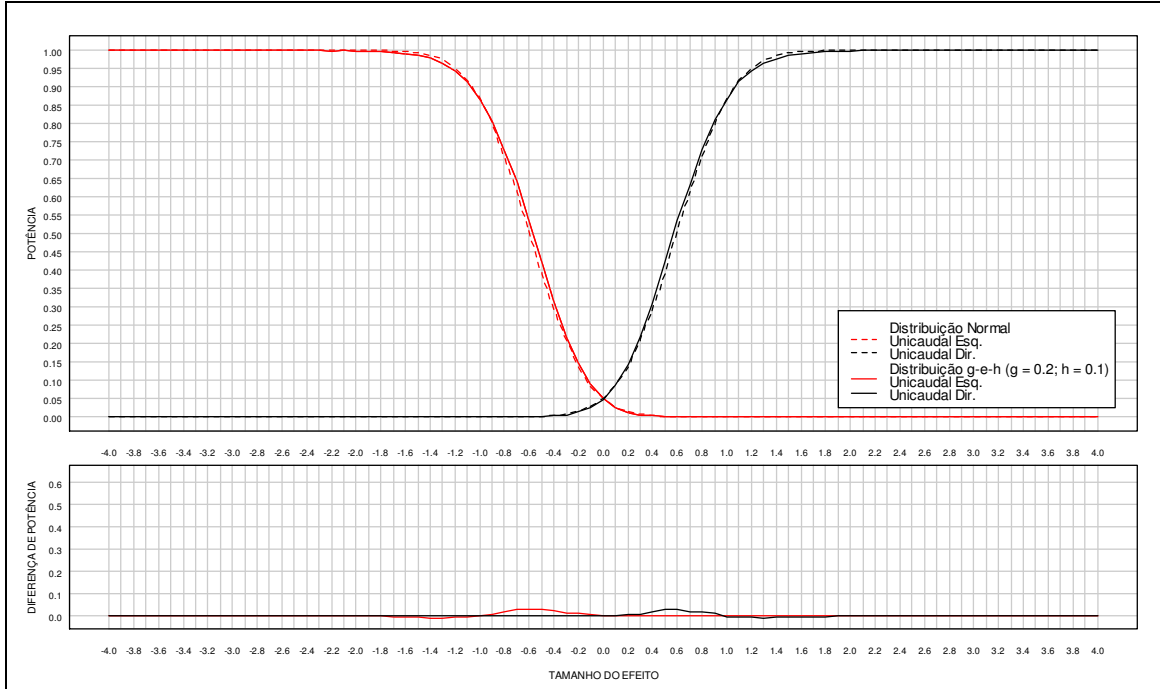


Figura Anexo IX.65 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

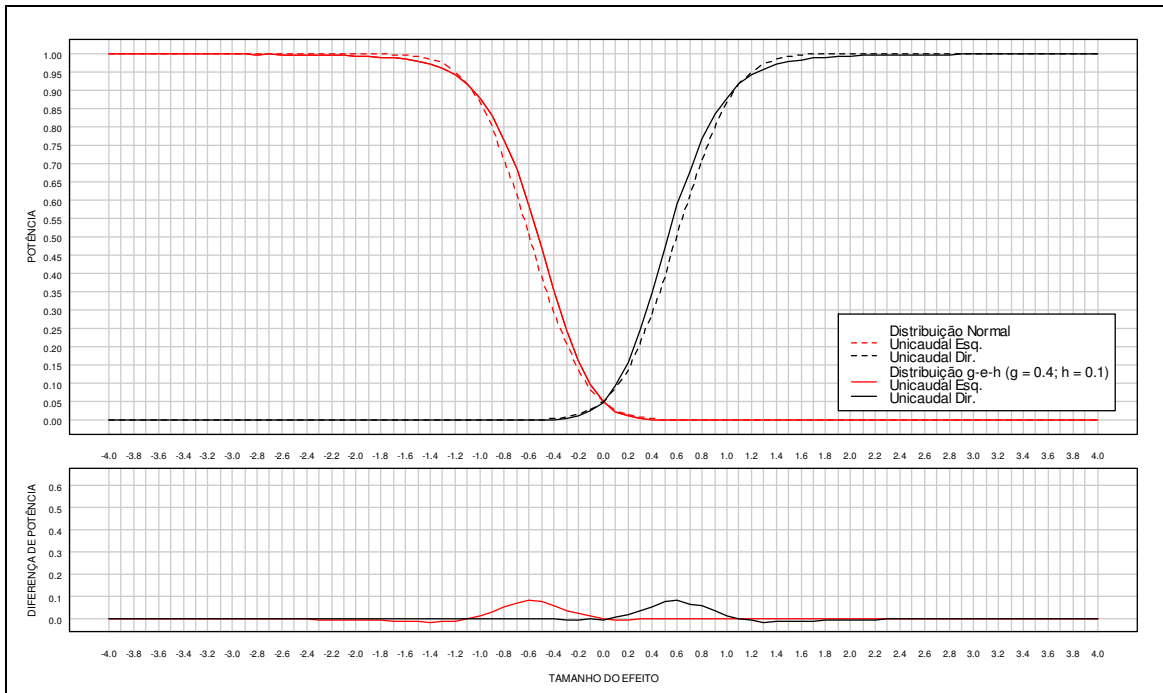


Figura Anexo IX.66 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

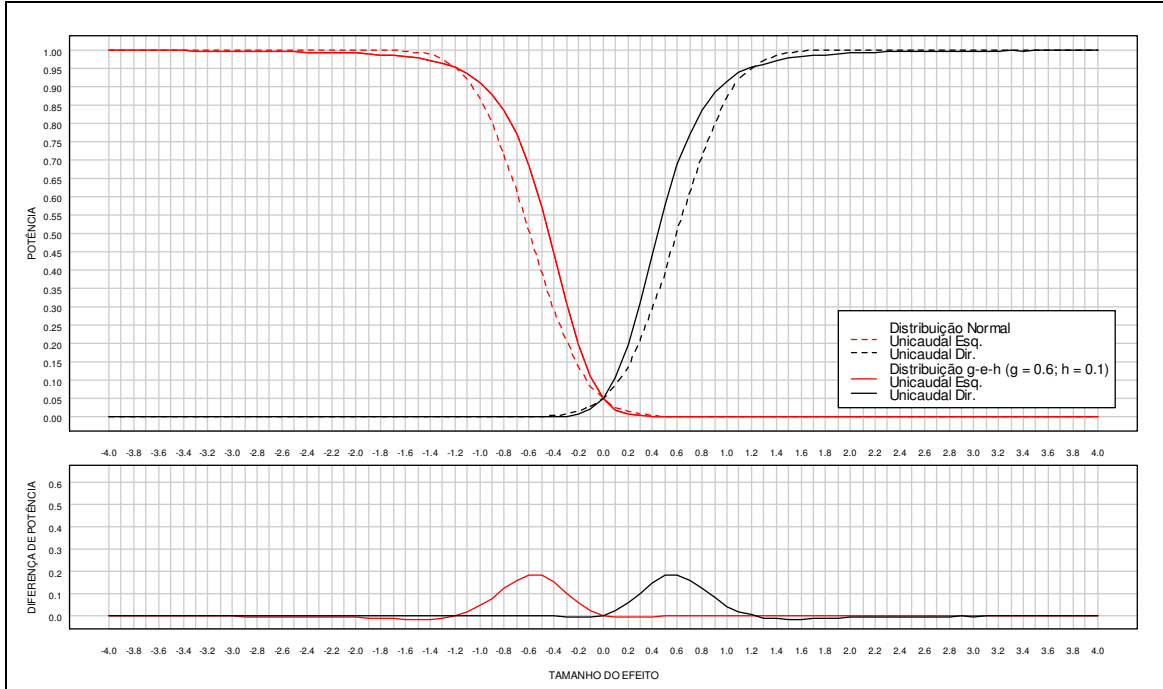


Figura Anexo IX.67 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

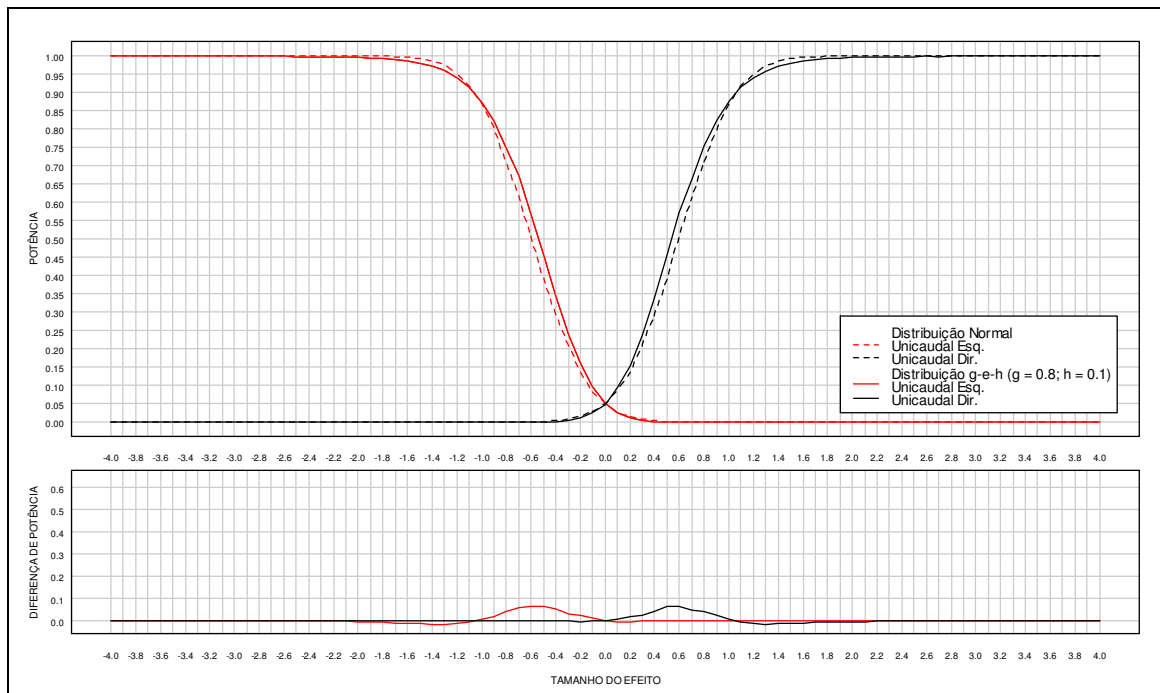


Figura Anexo IX.68 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

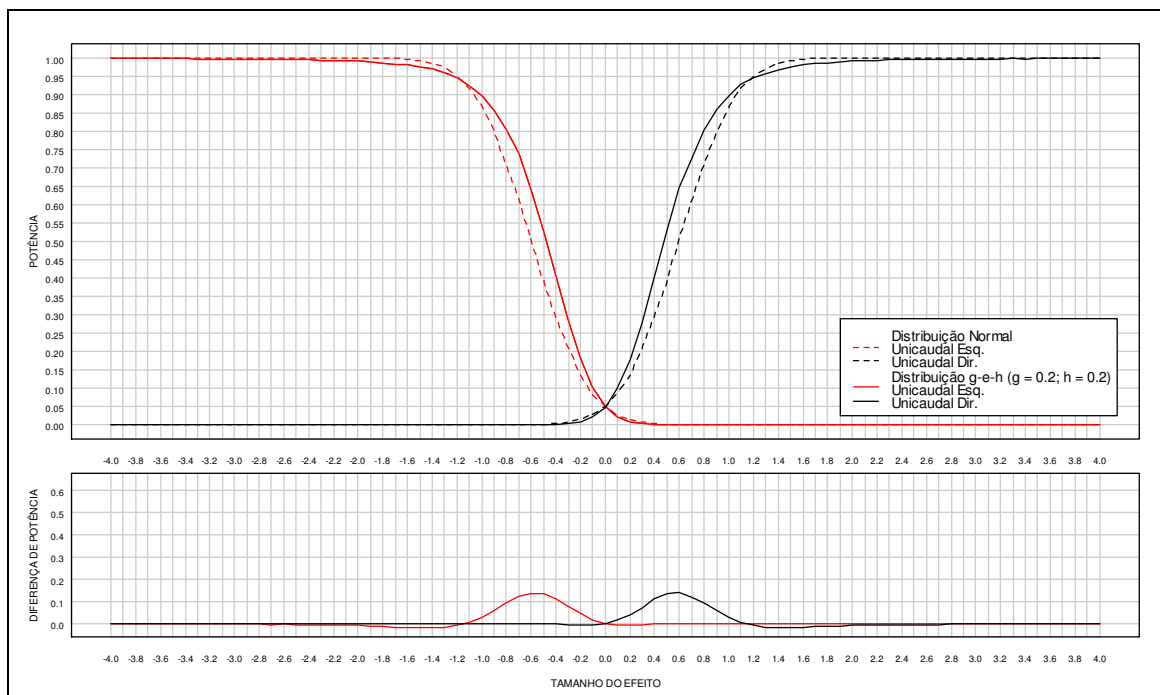


Figura Anexo IX.69 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

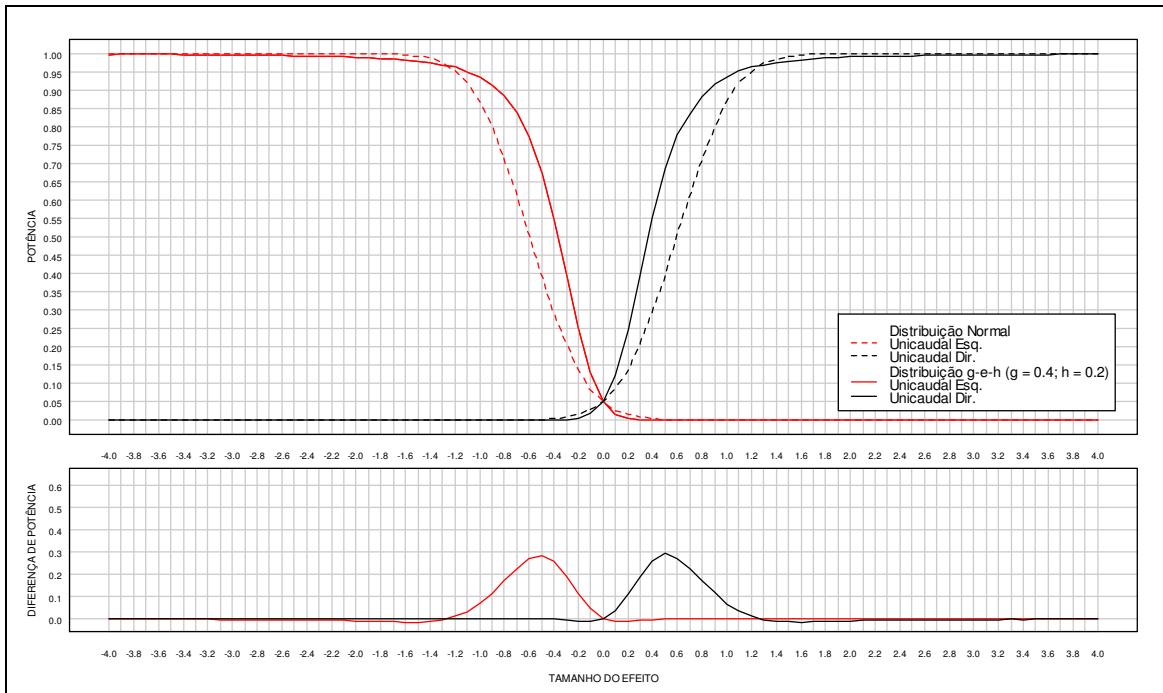


Figura Anexo IX.70 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

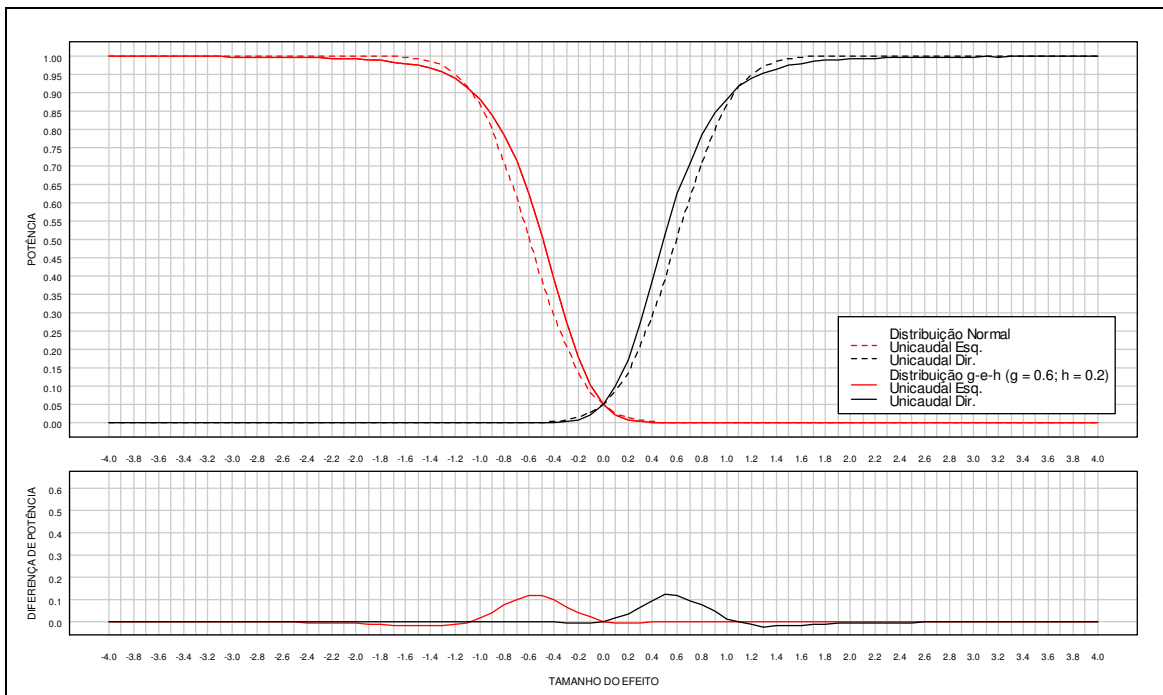


Figura Anexo IX.71 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

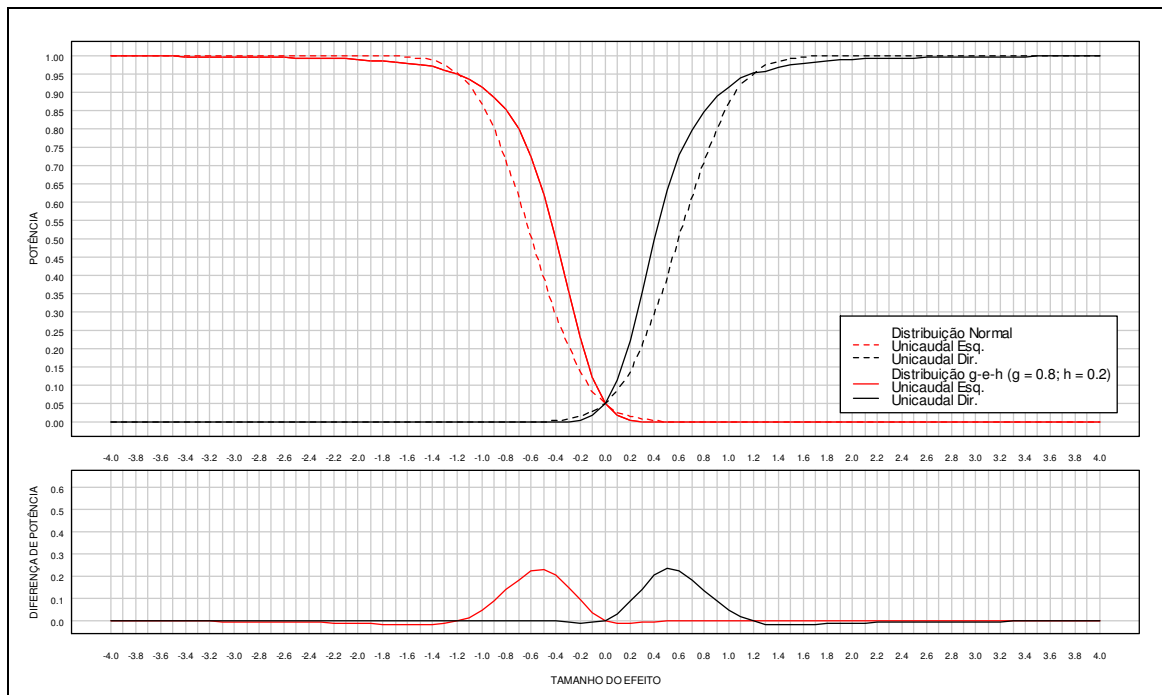


Figura Anexo IX.72 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

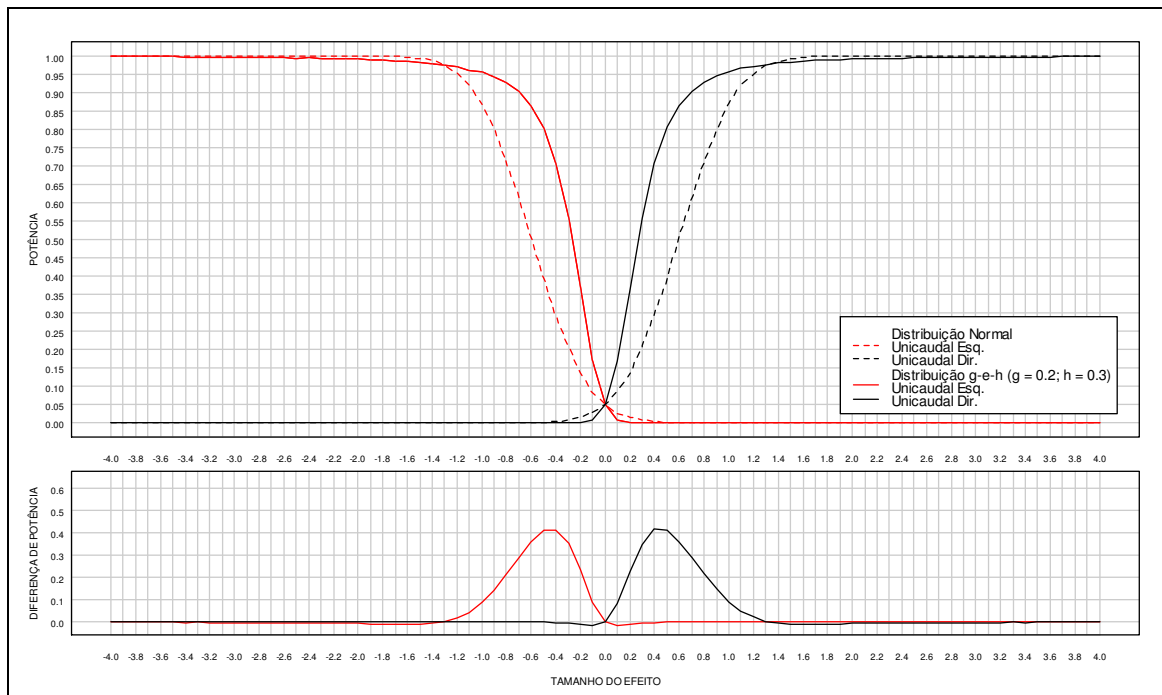


Figura Anexo IX.73 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

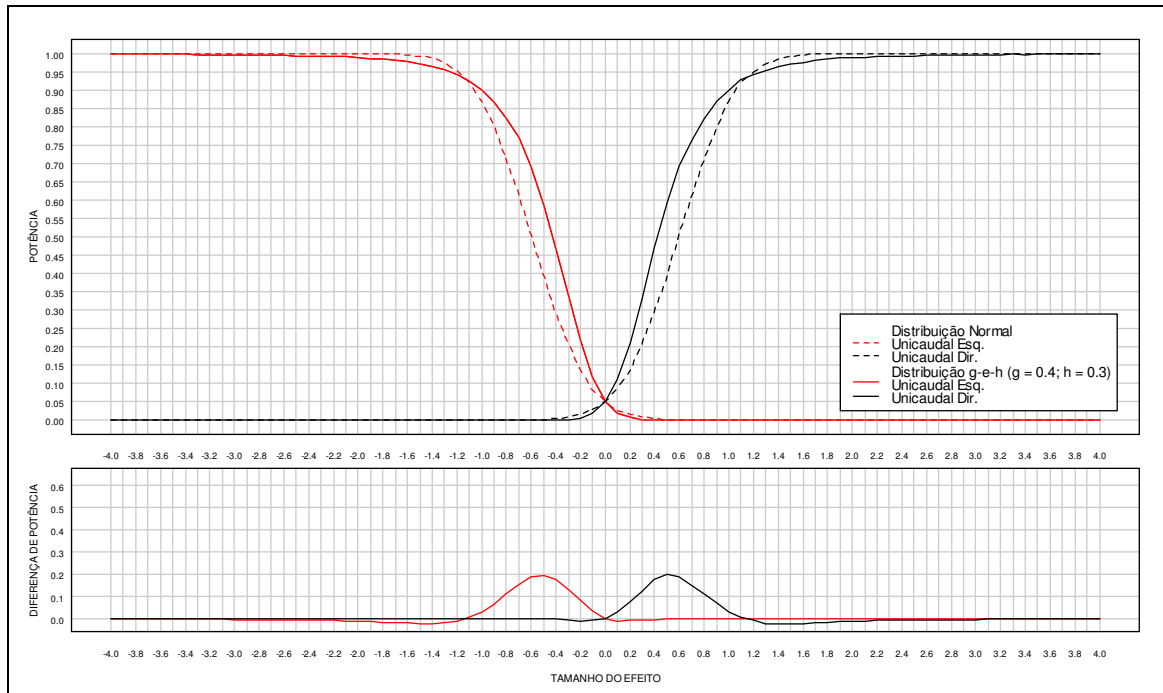


Figura Anexo IX.74 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

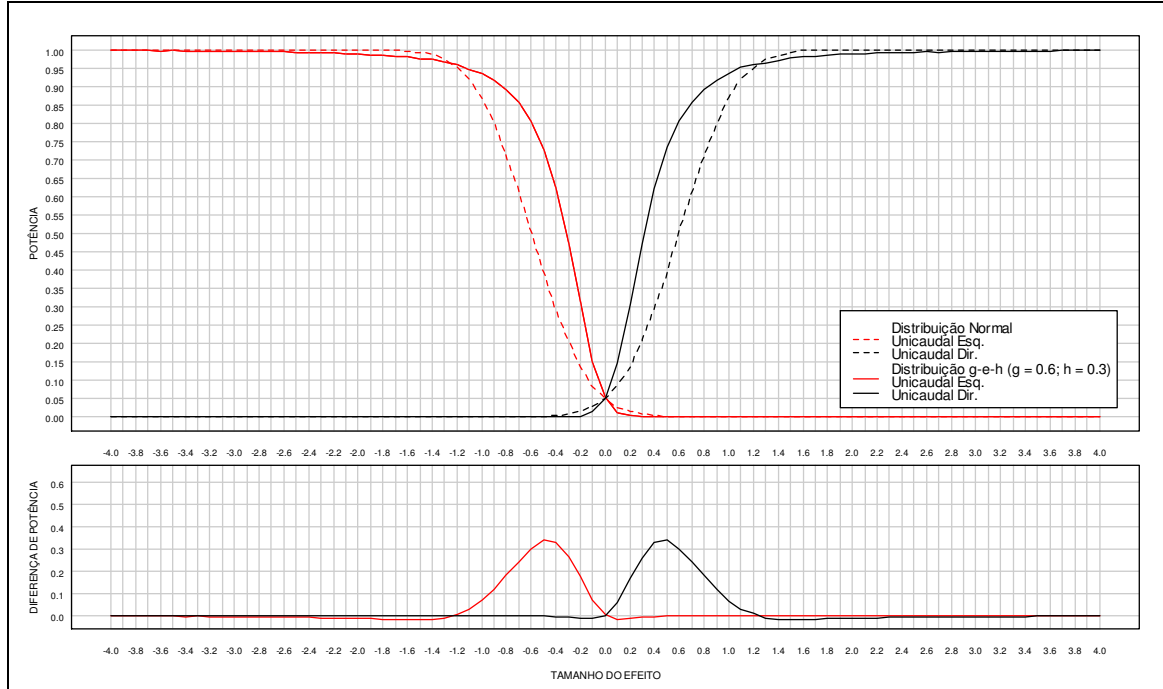


Figura Anexo IX.75 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

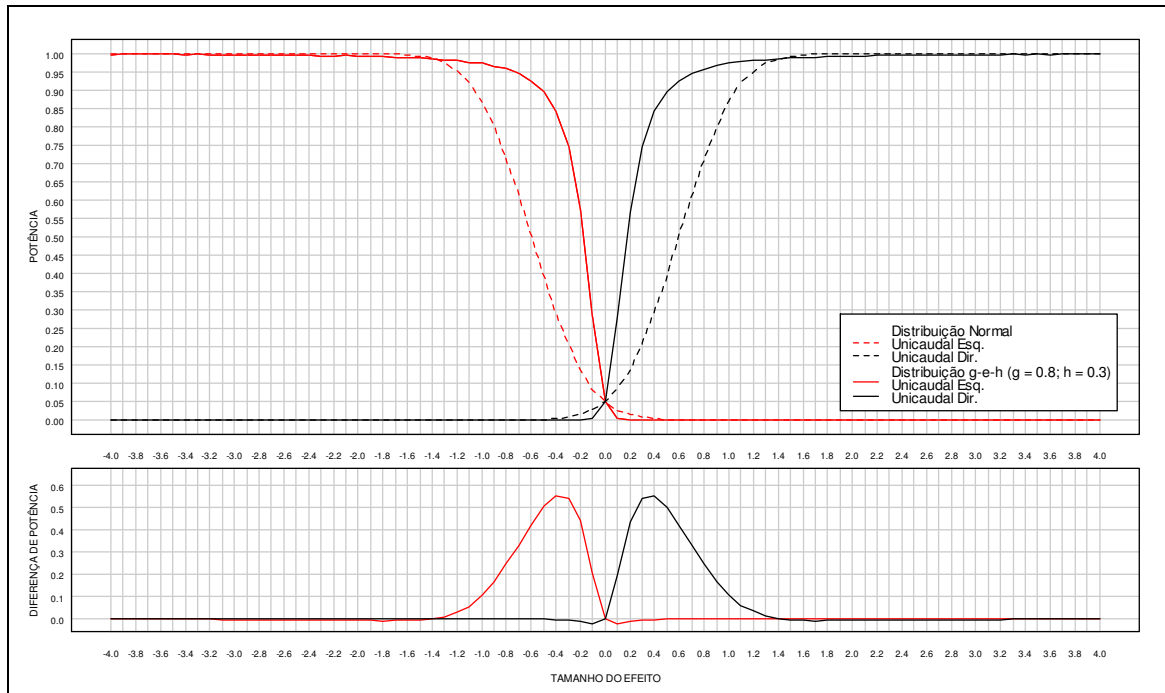


Figura Anexo IX.76 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 16$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

3. Amostras com $n_1 = n_2 = 32$

3.1. Testes bi-caudais

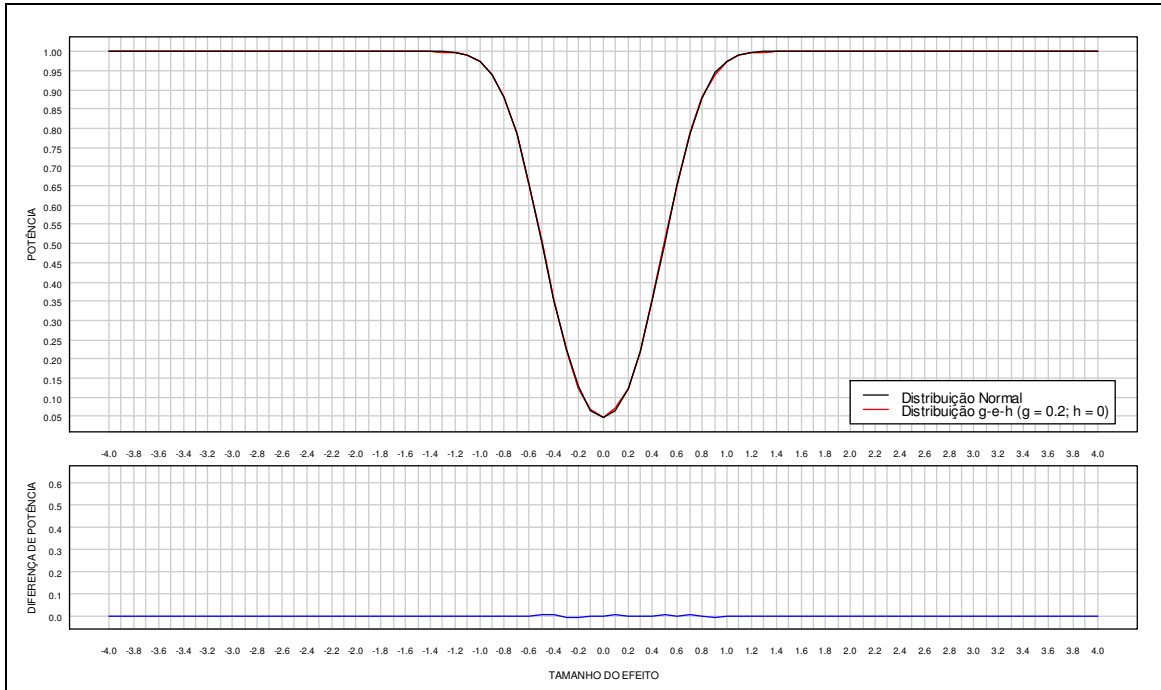


Figura Anexo IX.77 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

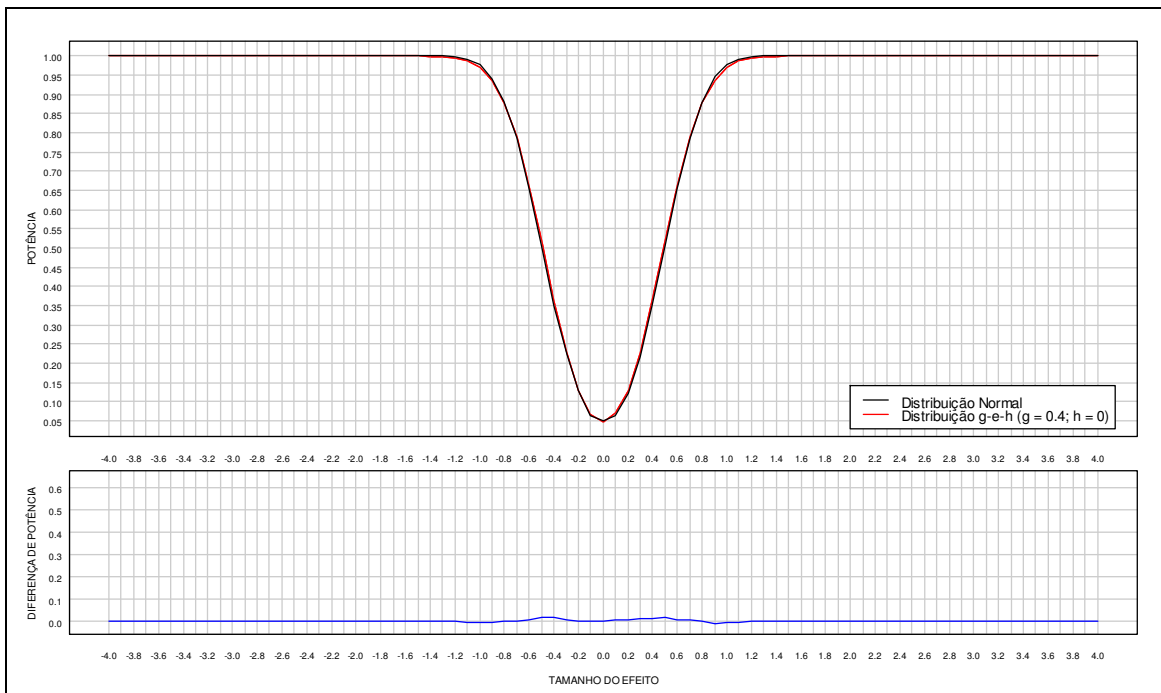


Figura Anexo IX.78 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

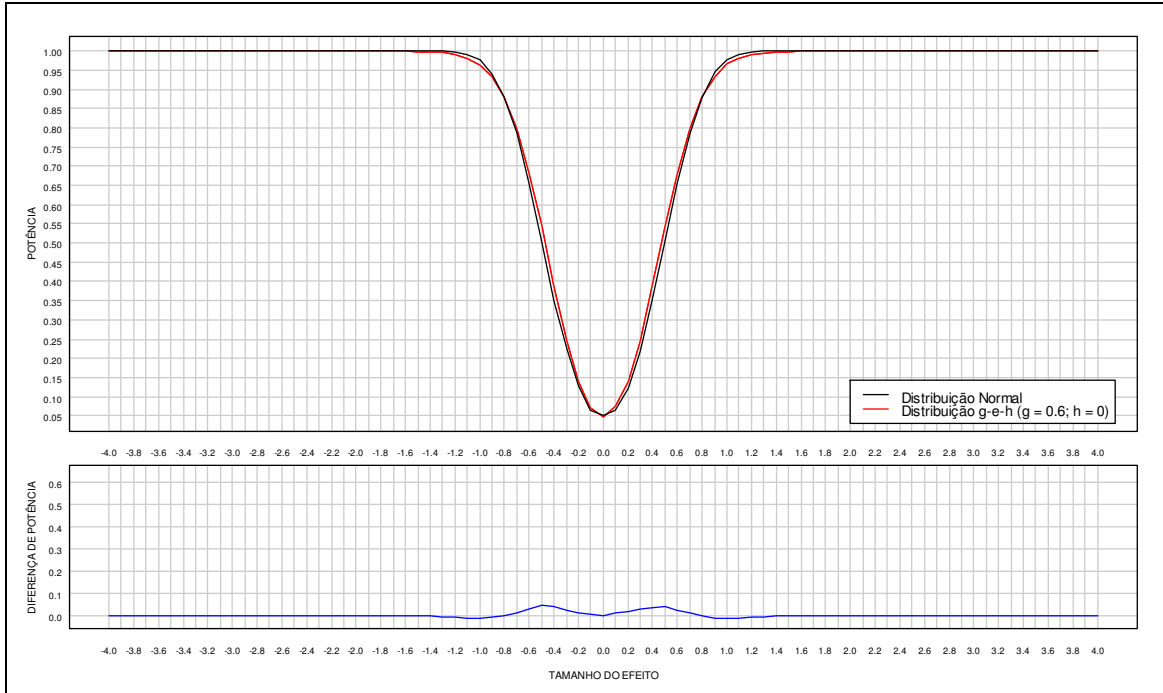


Figura Anexo IX.79 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

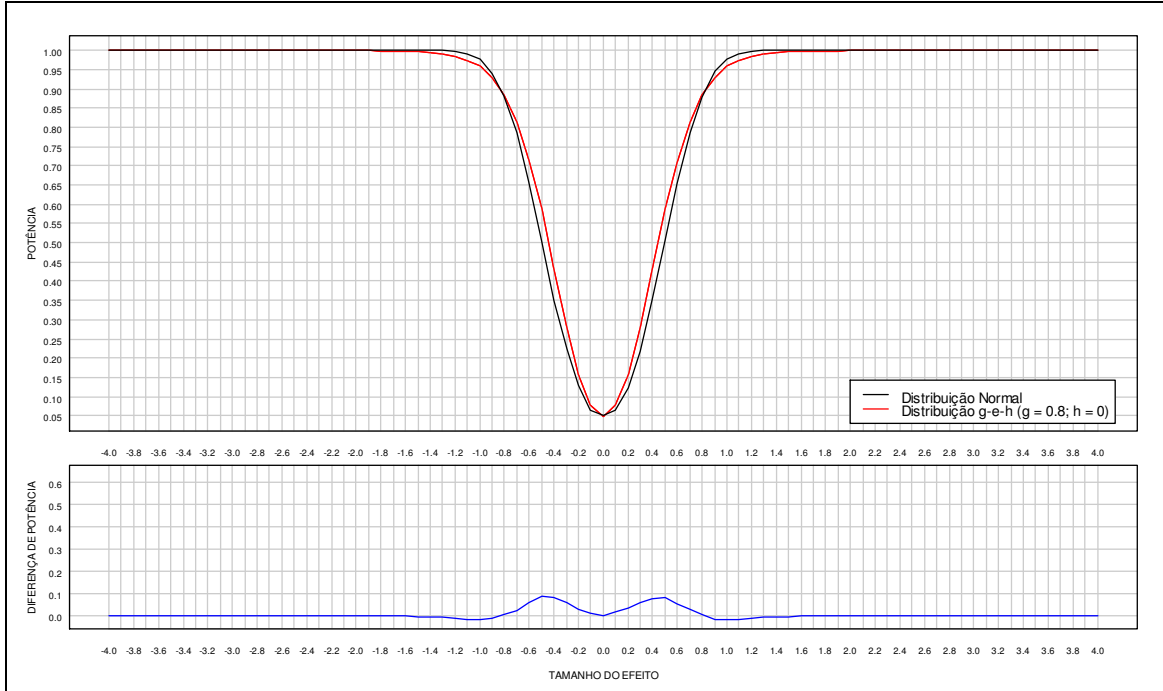


Figura Anexo IX.80 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

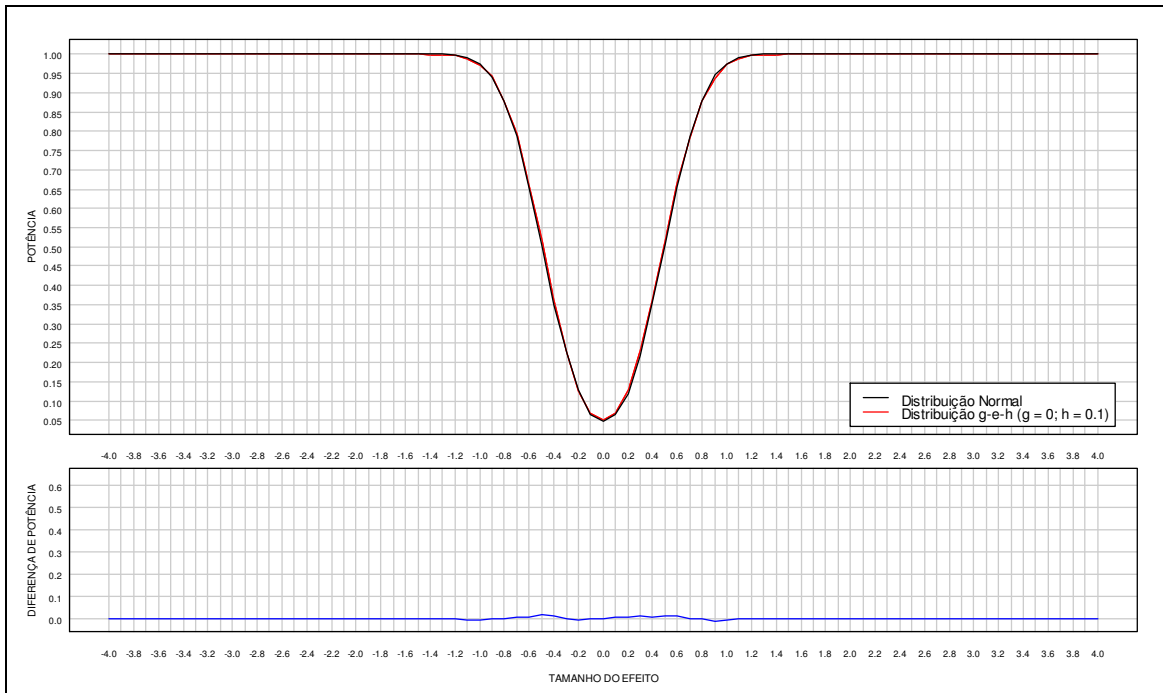


Figura Anexo IX.81 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

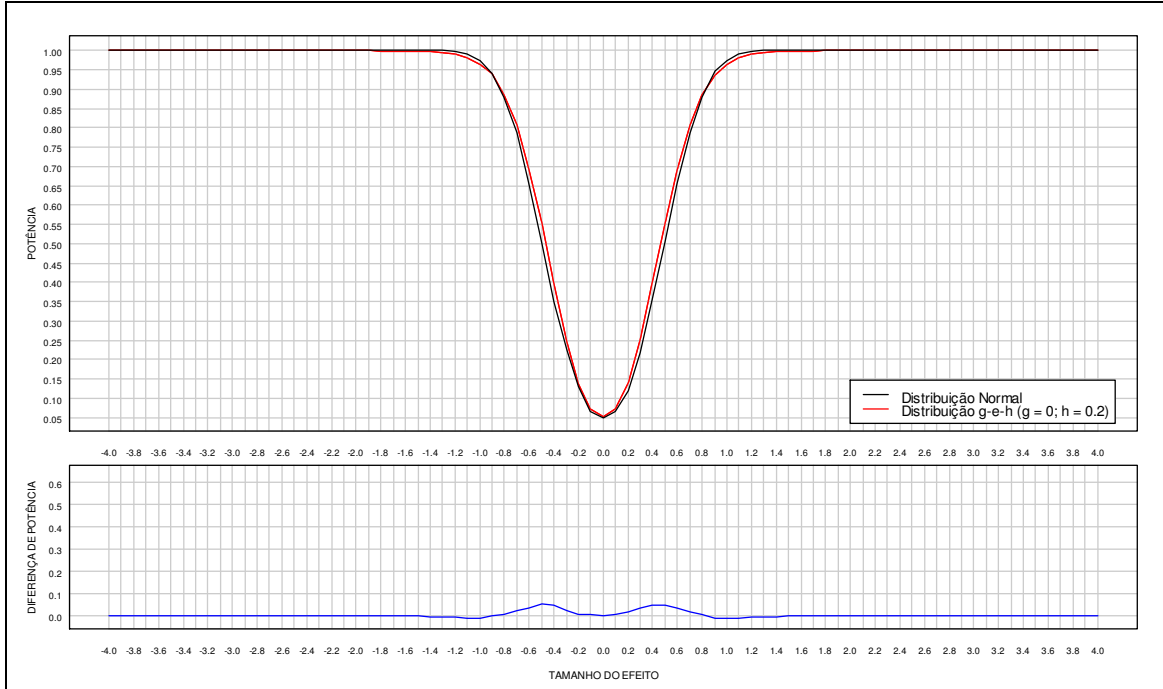


Figura Anexo IX.82 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

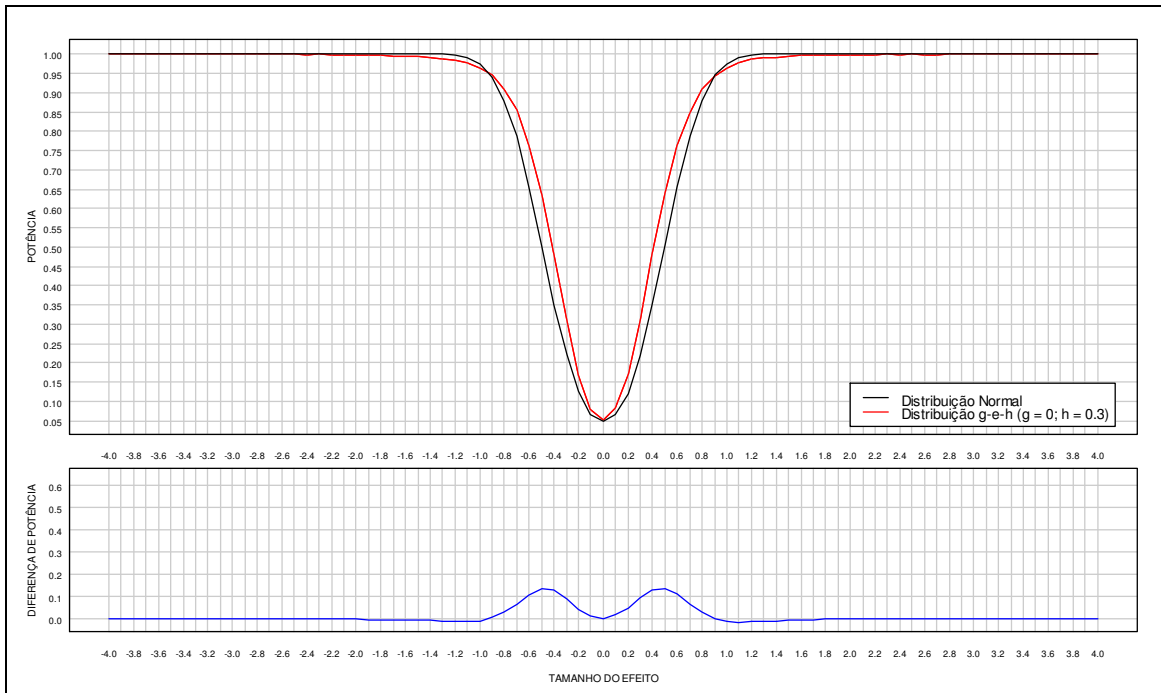


Figura Anexo IX.83 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

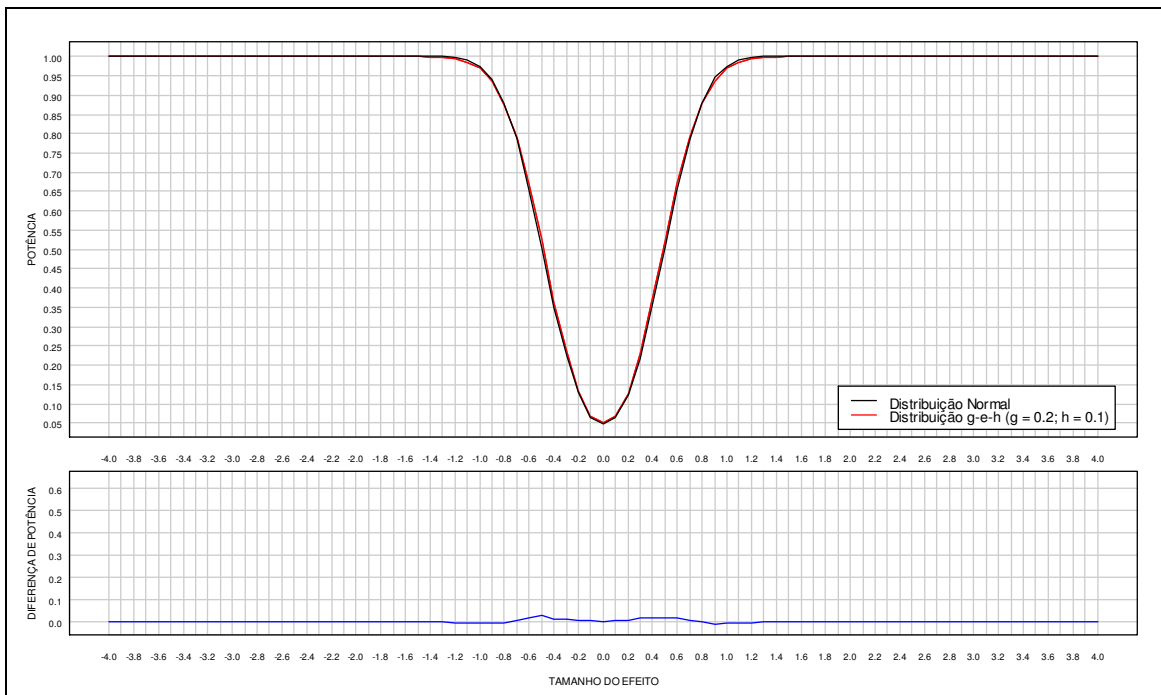


Figura Anexo IX.84 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

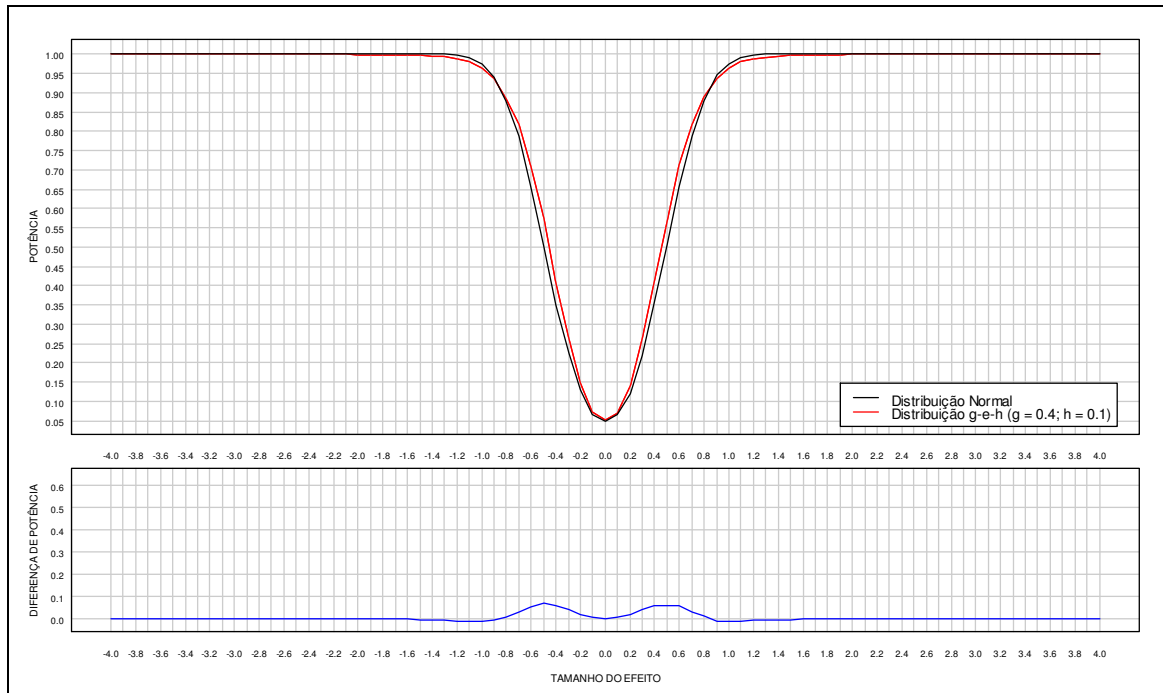


Figura Anexo IX.85 [Painel superior] Potência do teste de aleatorização bi-causal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

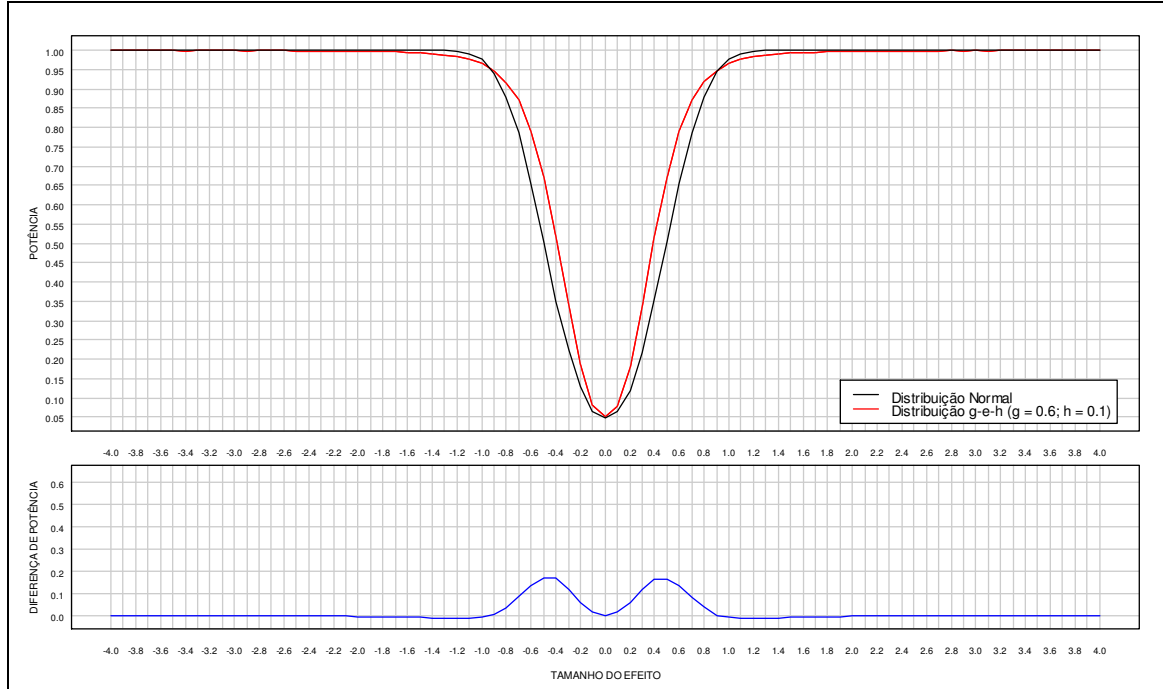


Figura Anexo IX.86 [Painel superior] Potência do teste de aleatorização bi-causal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

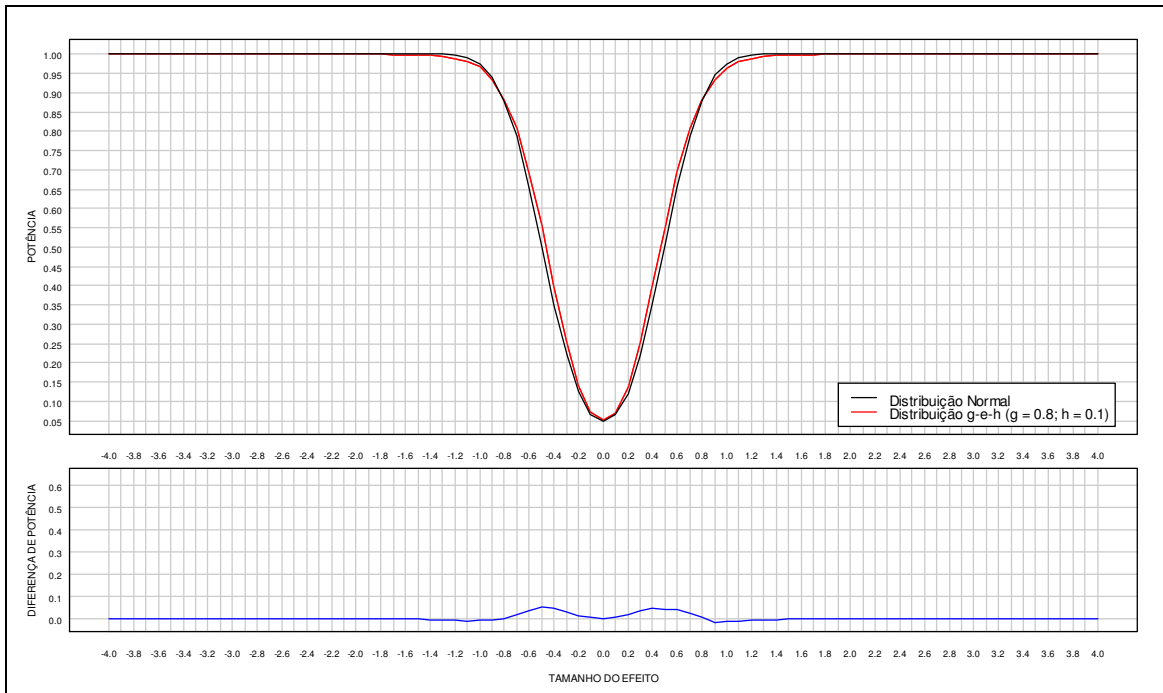


Figura Anexo IX.87 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

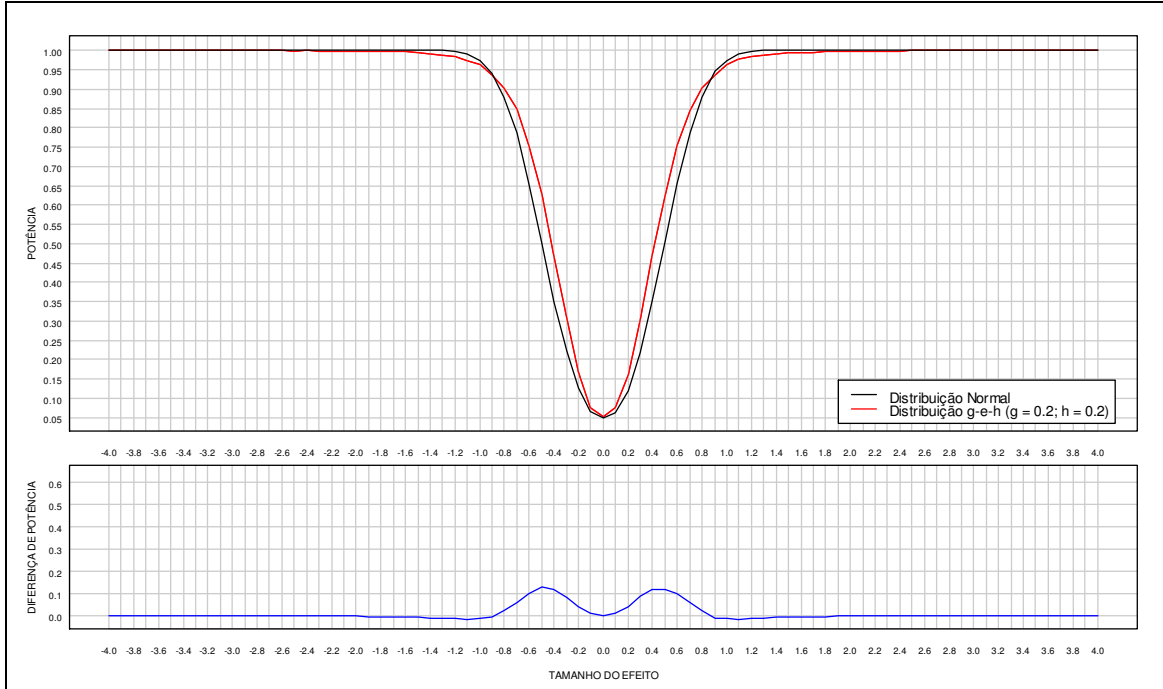


Figura Anexo IX.88 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

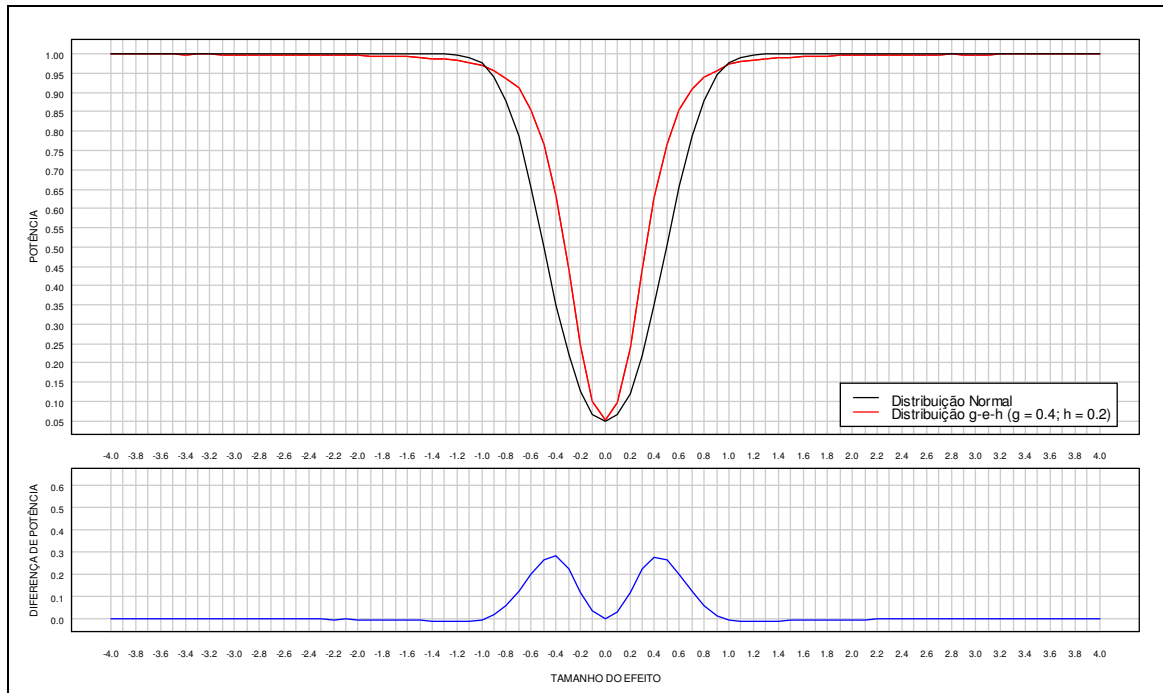


Figura Anexo IX.89 [Painel superior] Potência do teste de aleatorização bi-causal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

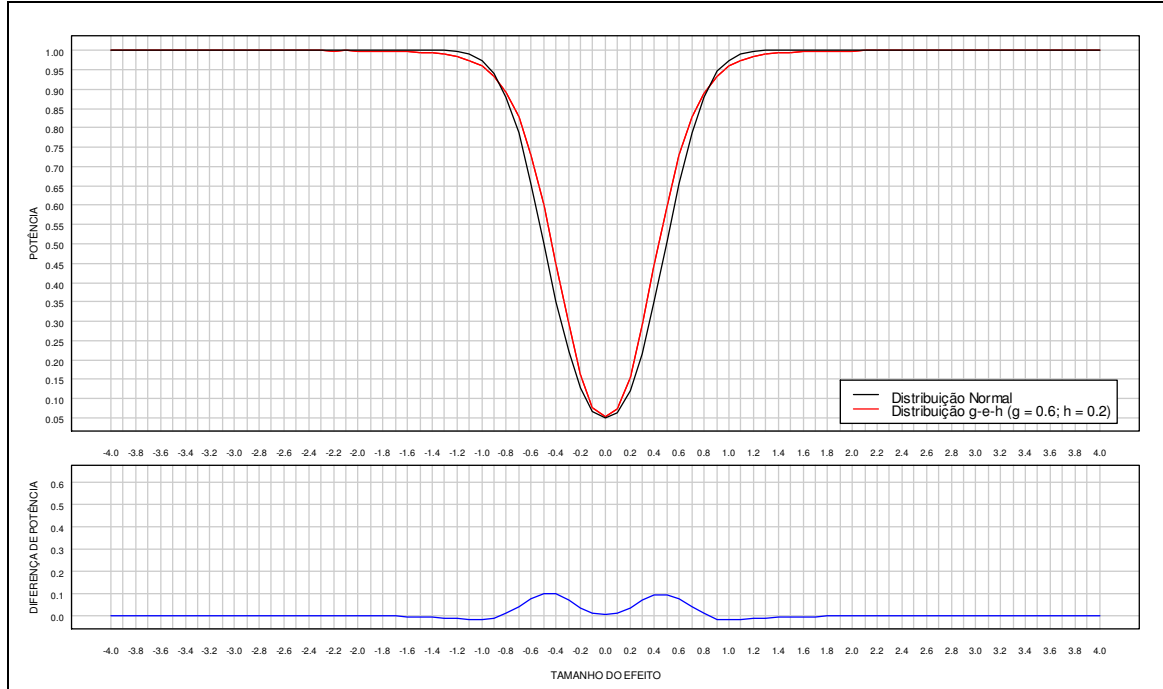


Figura Anexo IX.90 [Painel superior] Potência do teste de aleatorização bi-causal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

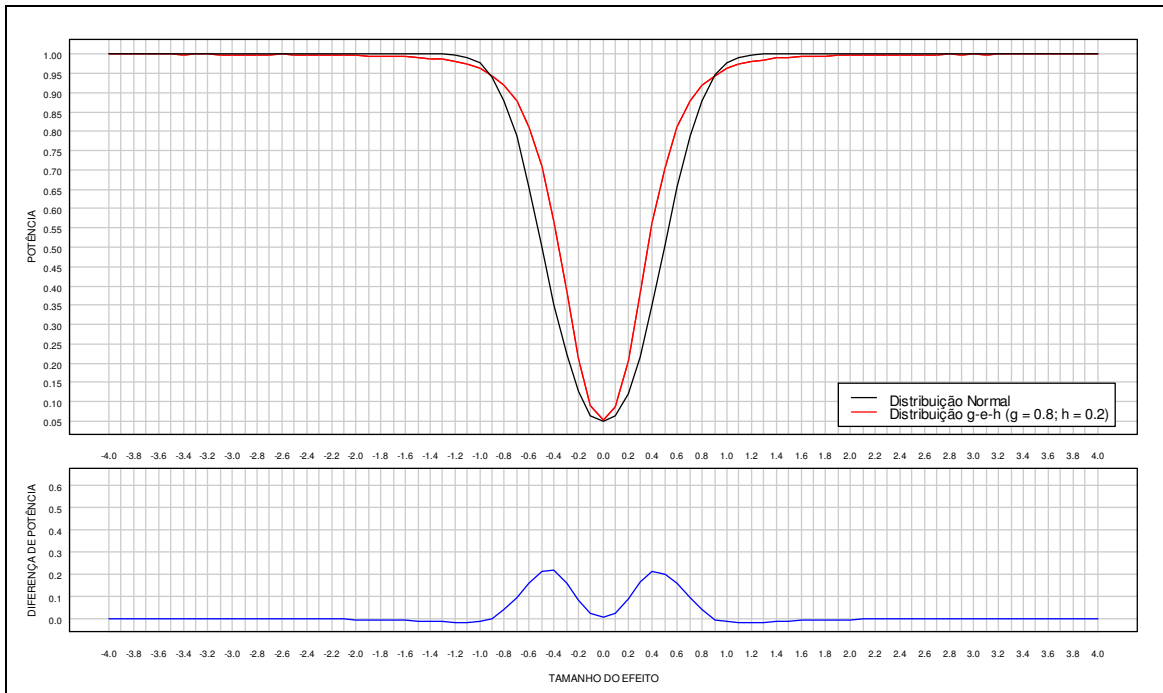


Figura Anexo IX.91 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

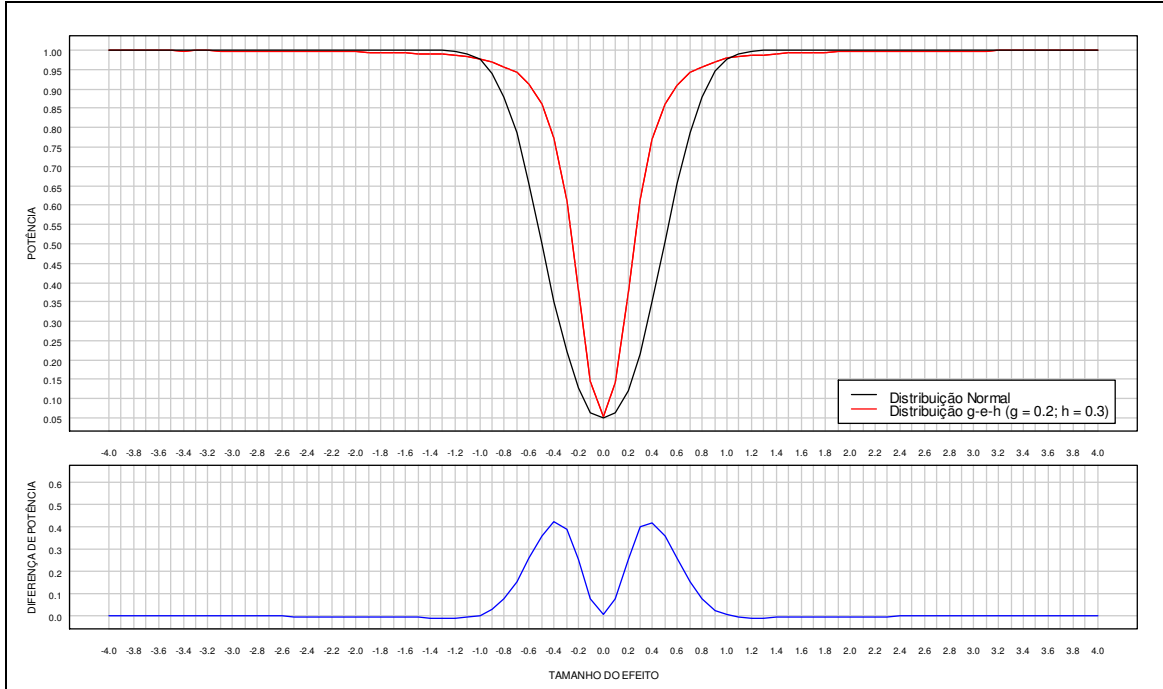


Figura Anexo IX.92 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

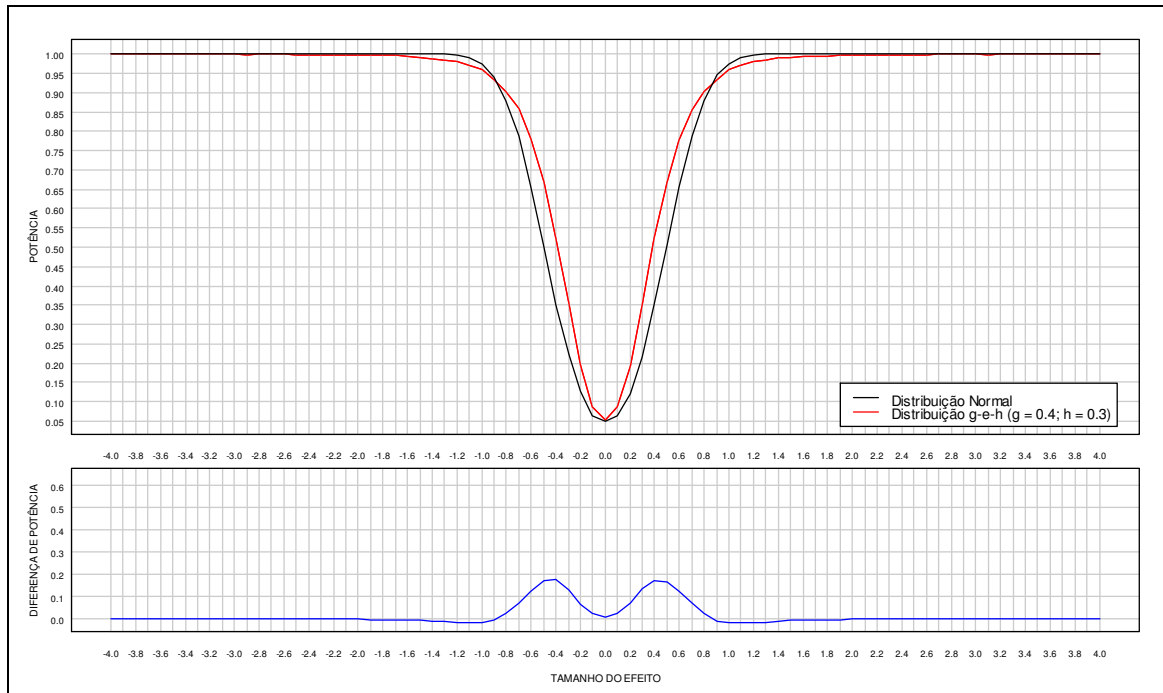


Figura Anexo IX.93 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

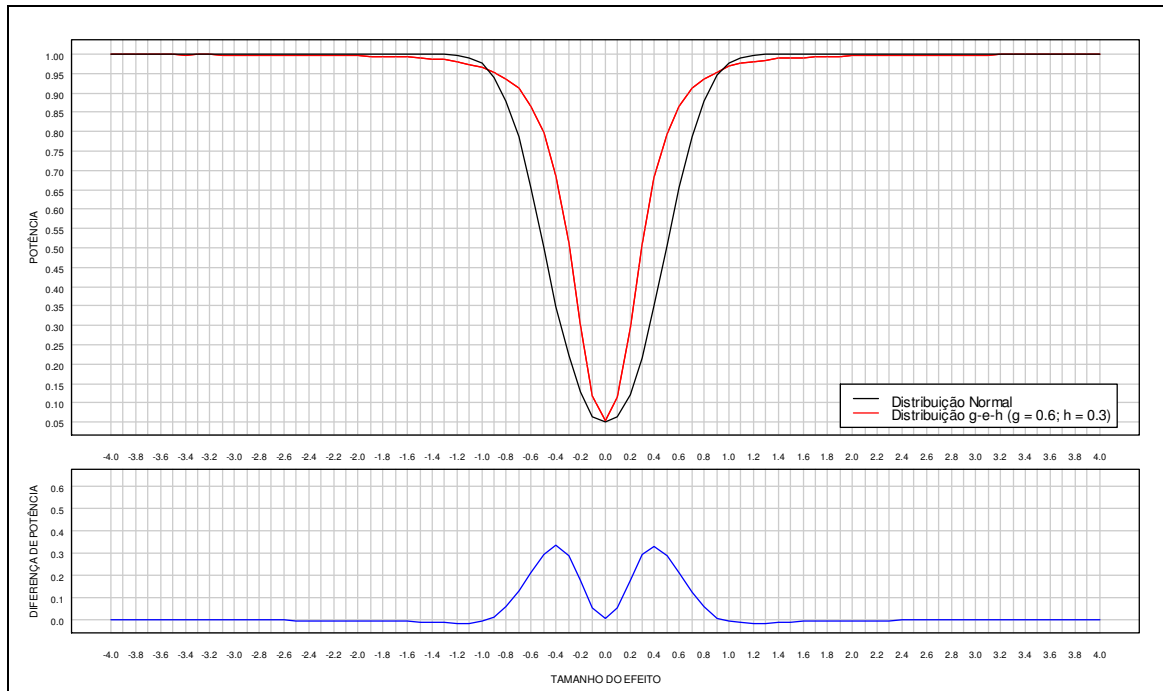


Figura Anexo IX.94 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

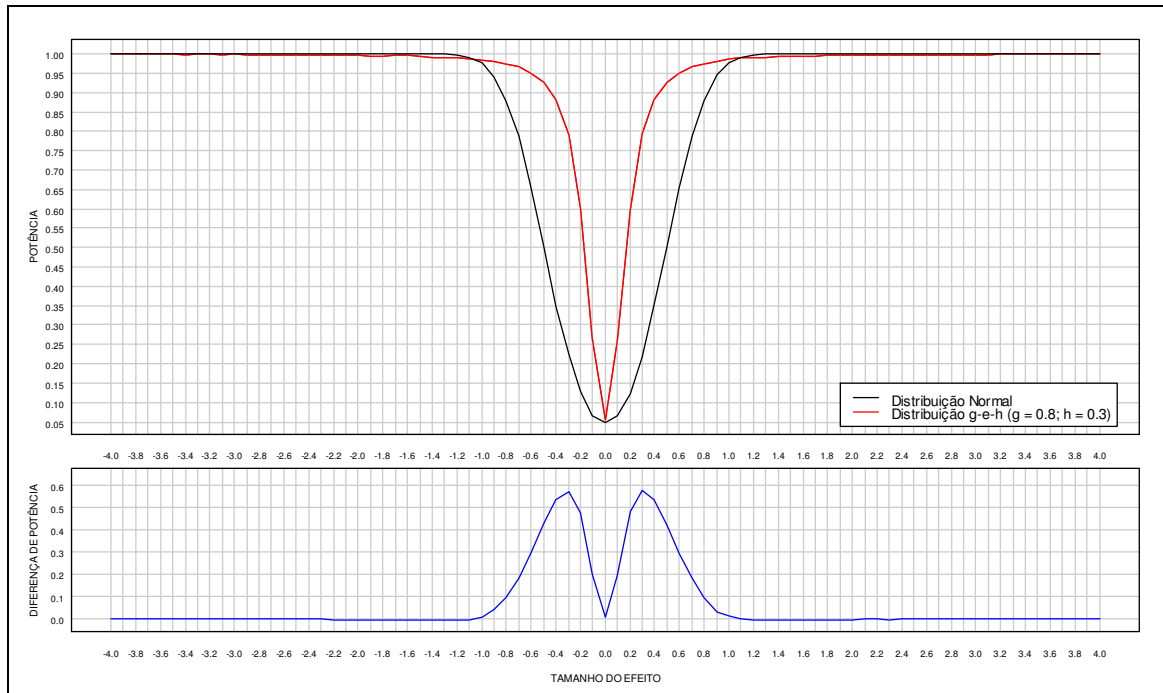


Figura Anexo IX.95 [Painel superior] Potência do teste de aleatorização bi-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

3.2. Testes uni-caudais

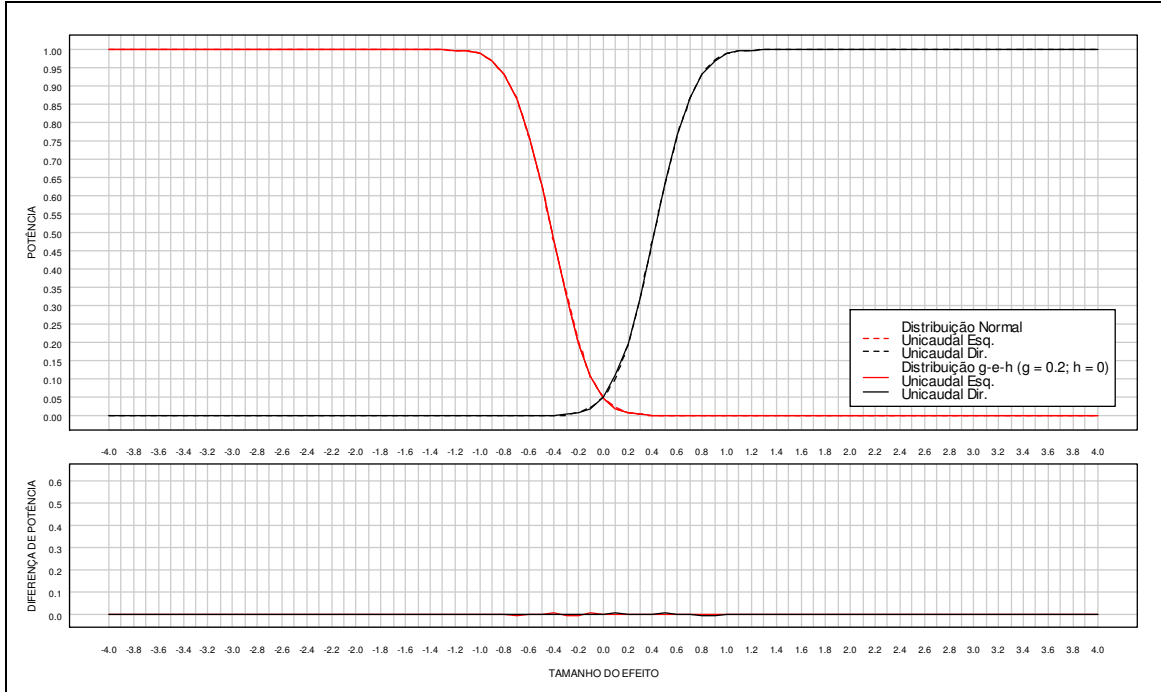


Figura Anexo IX.96 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

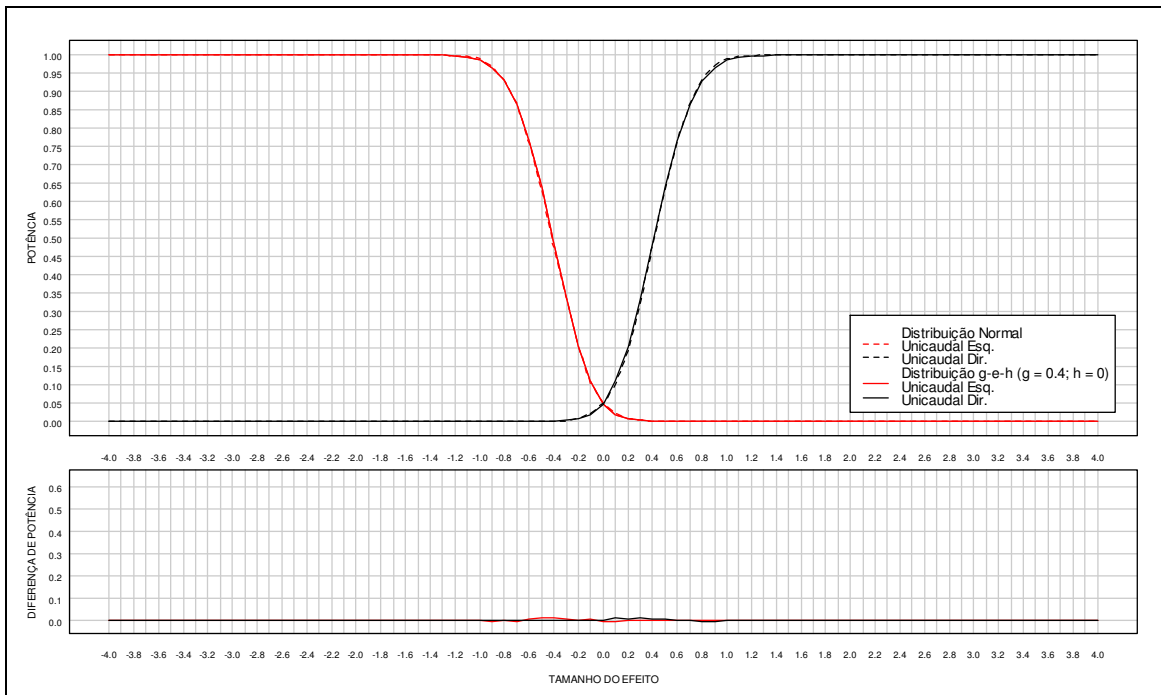


Figura Anexo IX.97 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

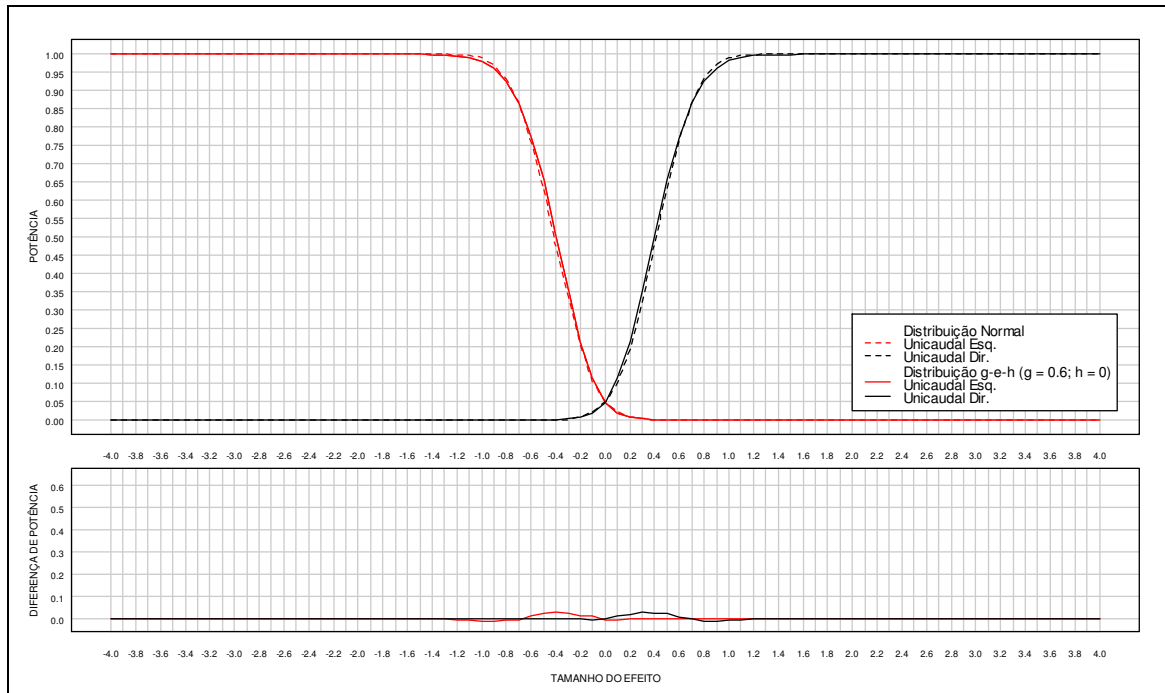


Figura Anexo IX.98 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

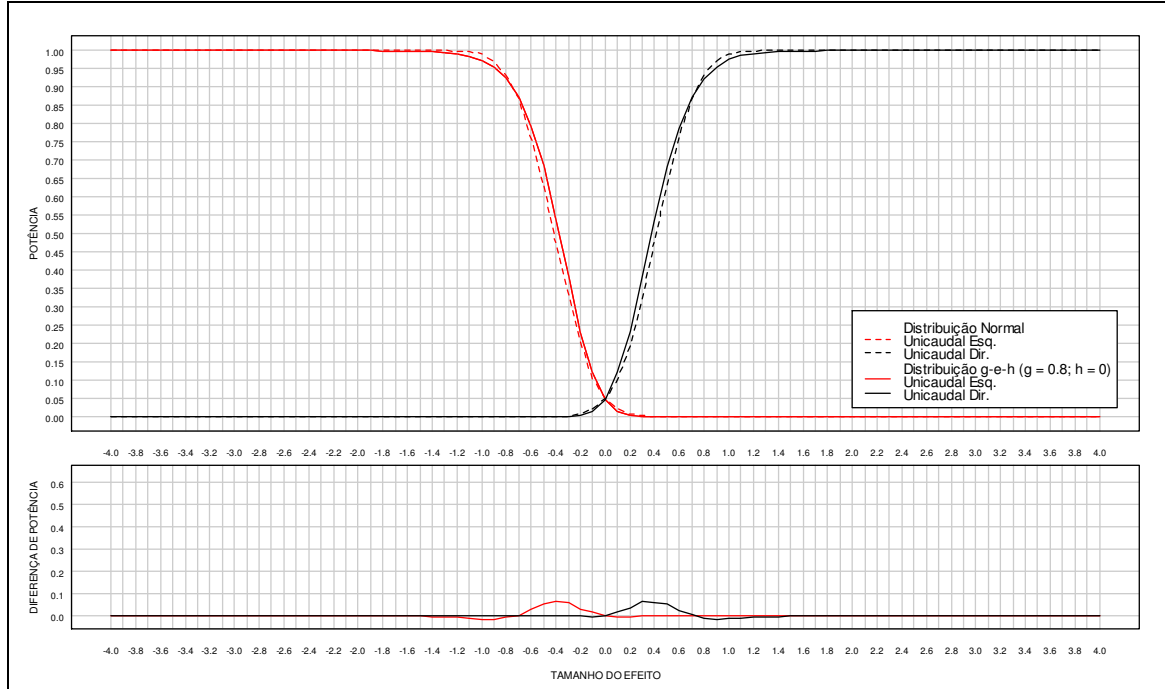


Figura Anexo IX.99 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

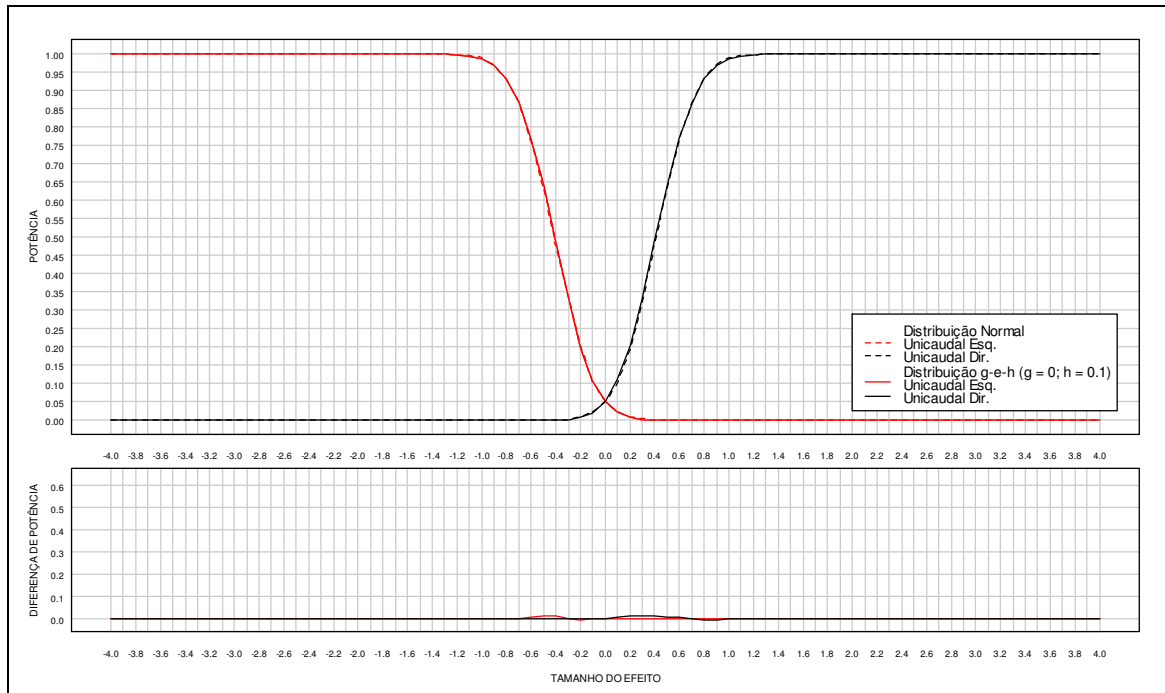


Figura Anexo IX.100 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

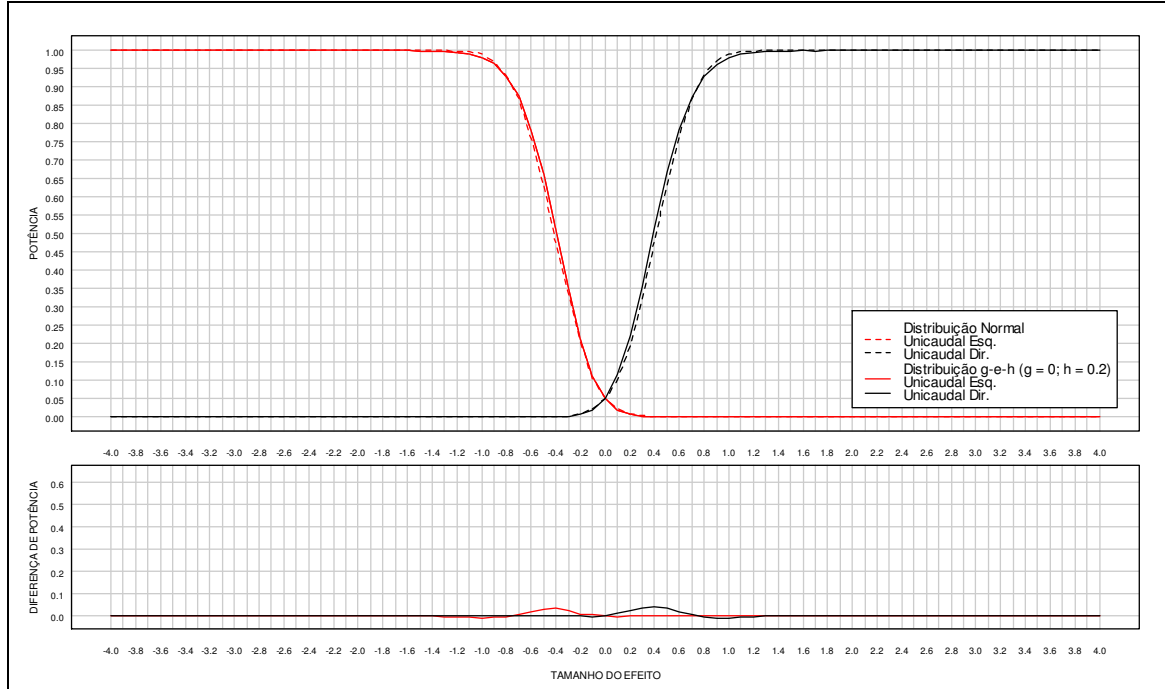


Figura Anexo IX.101 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

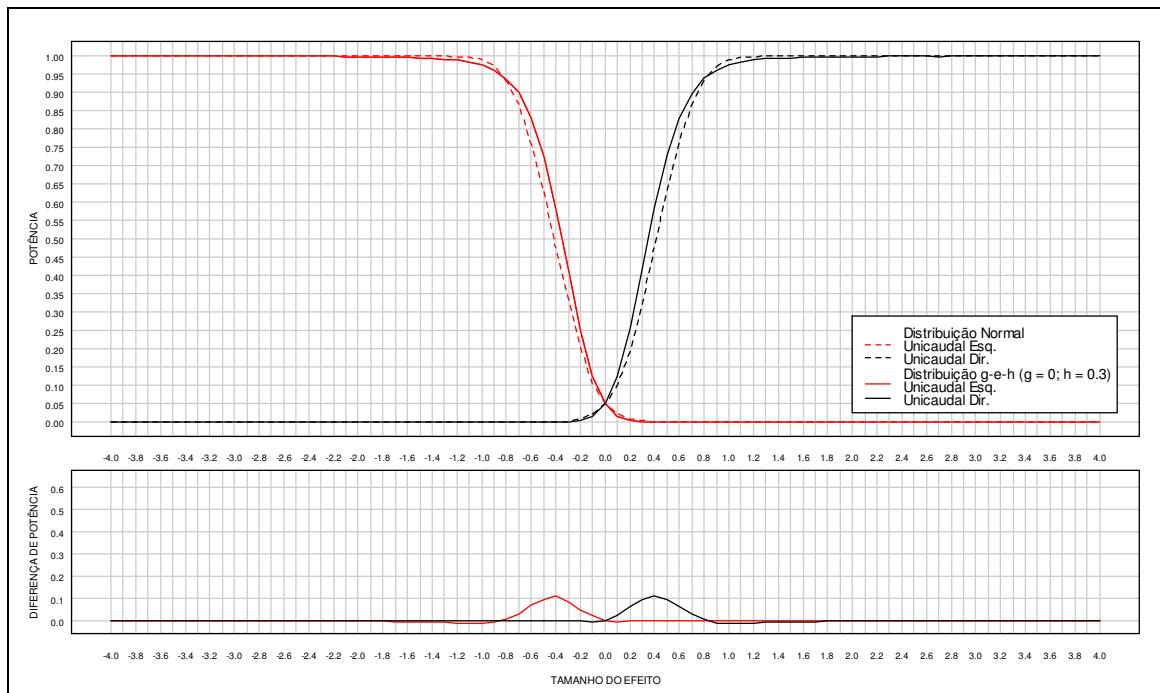


Figura Anexo IX.102 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

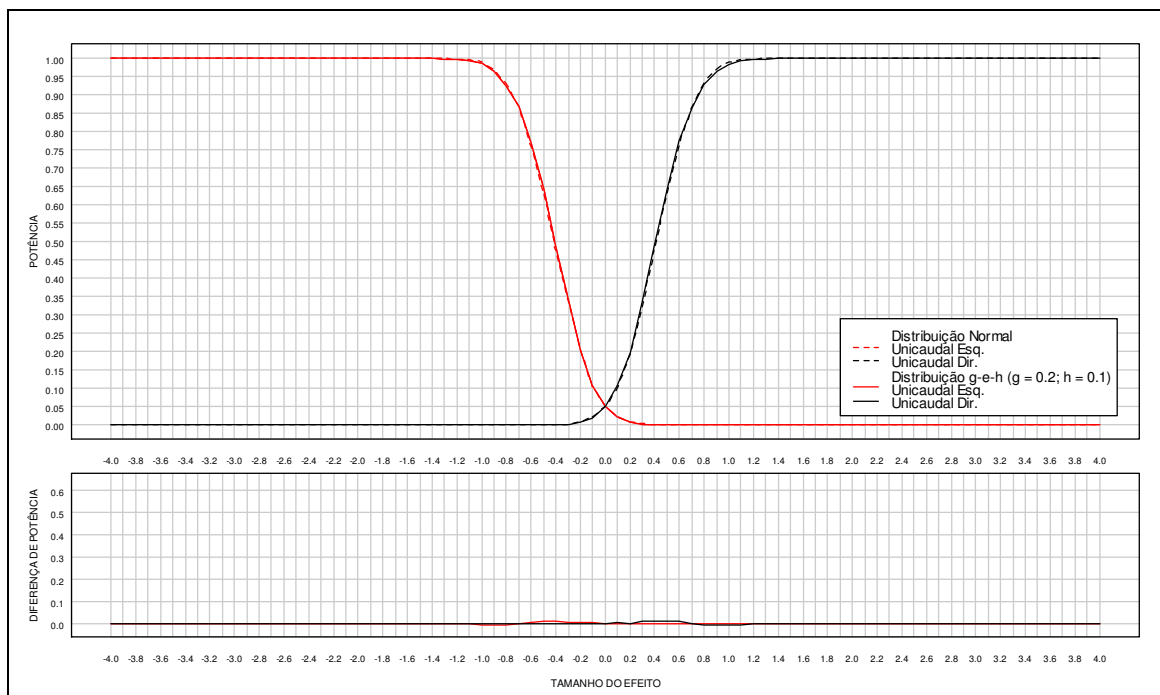


Figura Anexo IX.103 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

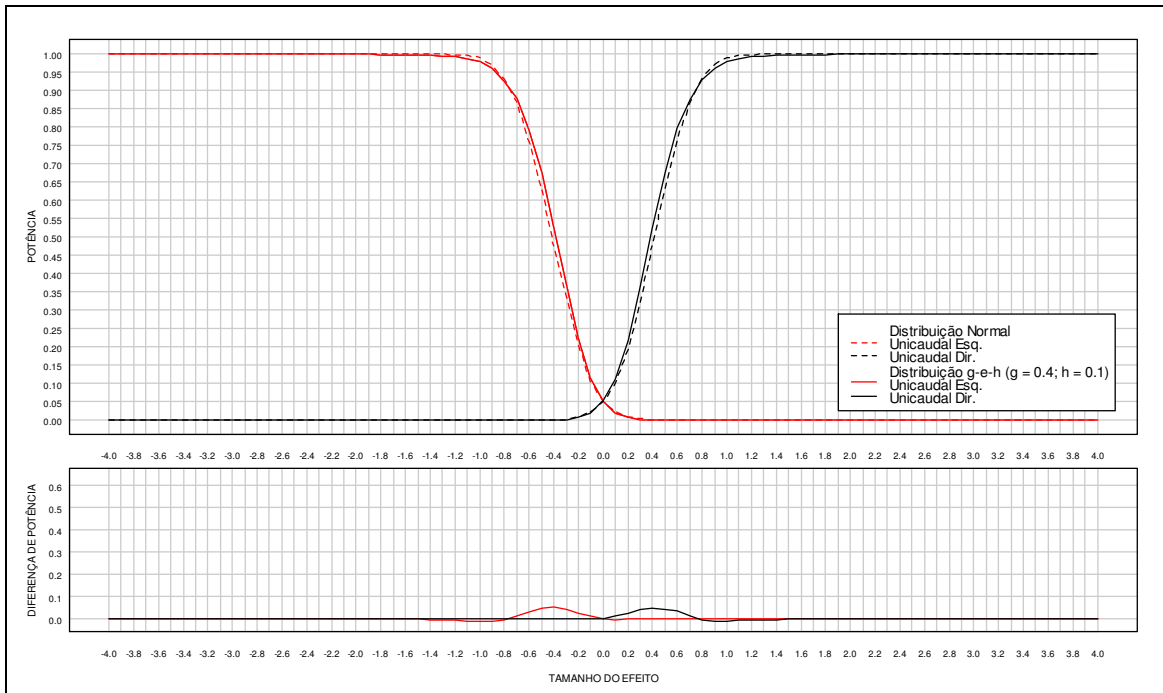


Figura Anexo IX.104 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

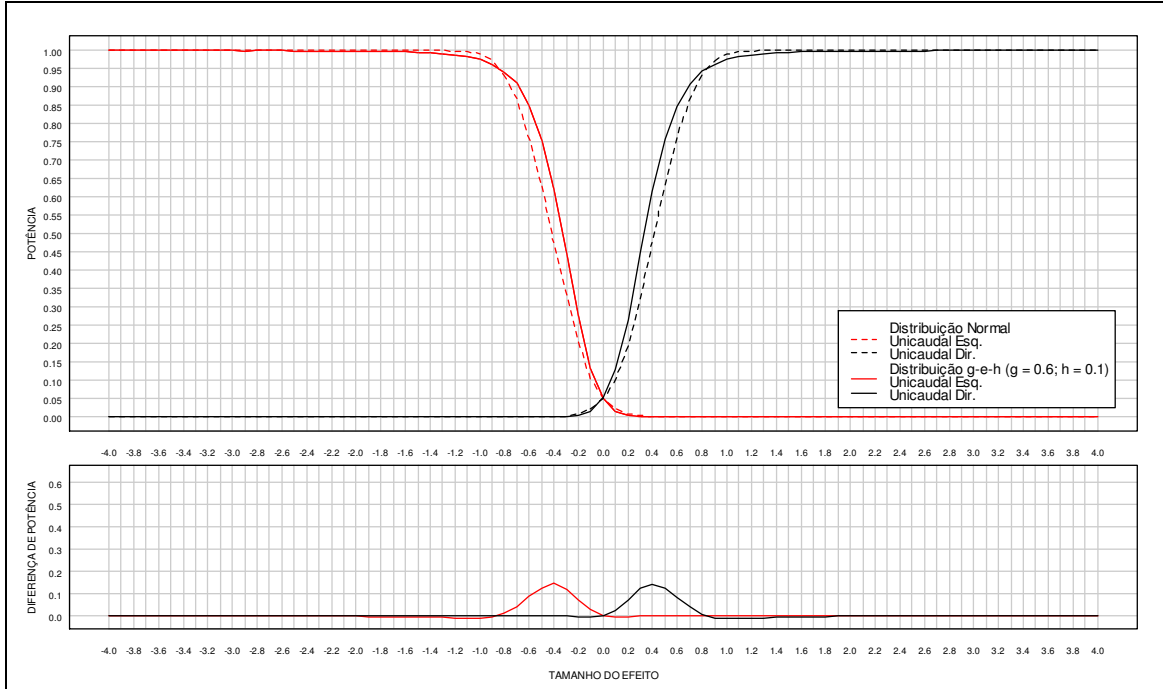


Figura Anexo IX.105 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

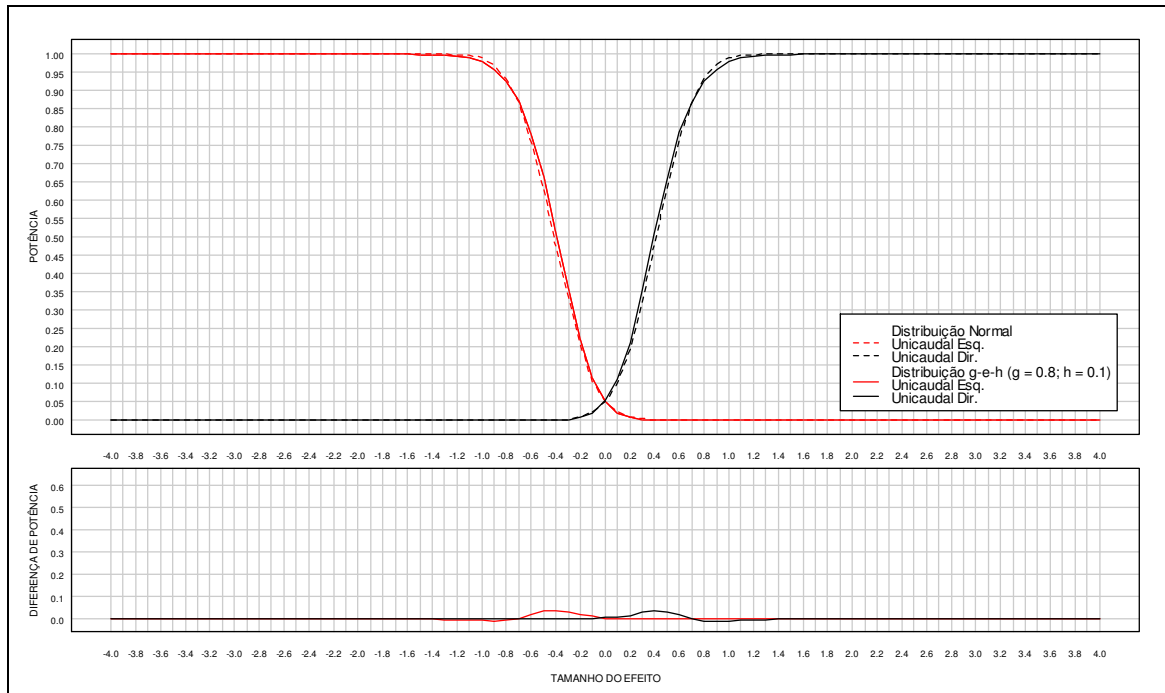


Figura Anexo IX.106 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.1$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

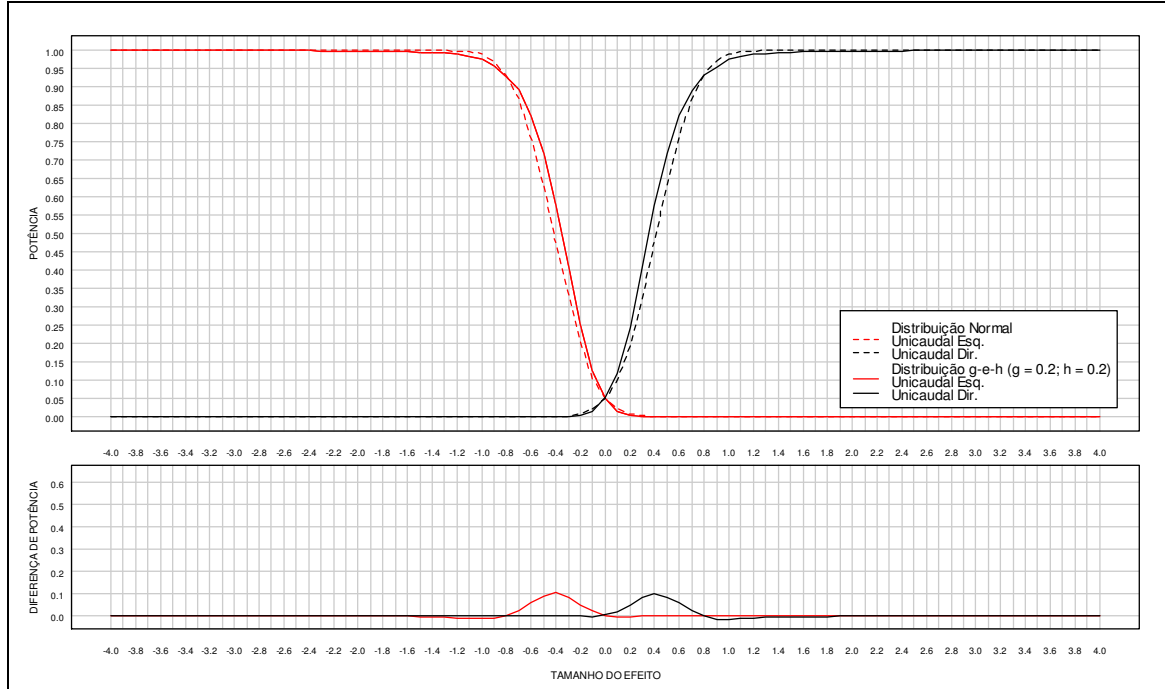


Figura Anexo IX.107 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

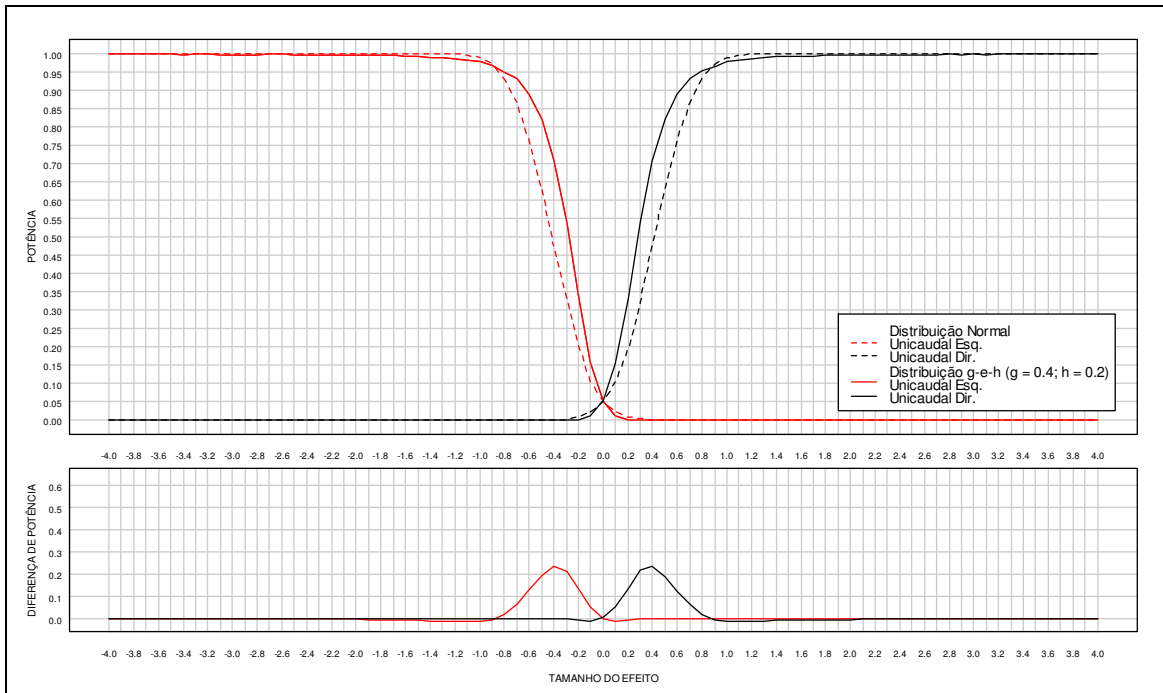


Figura Anexo IX.108 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

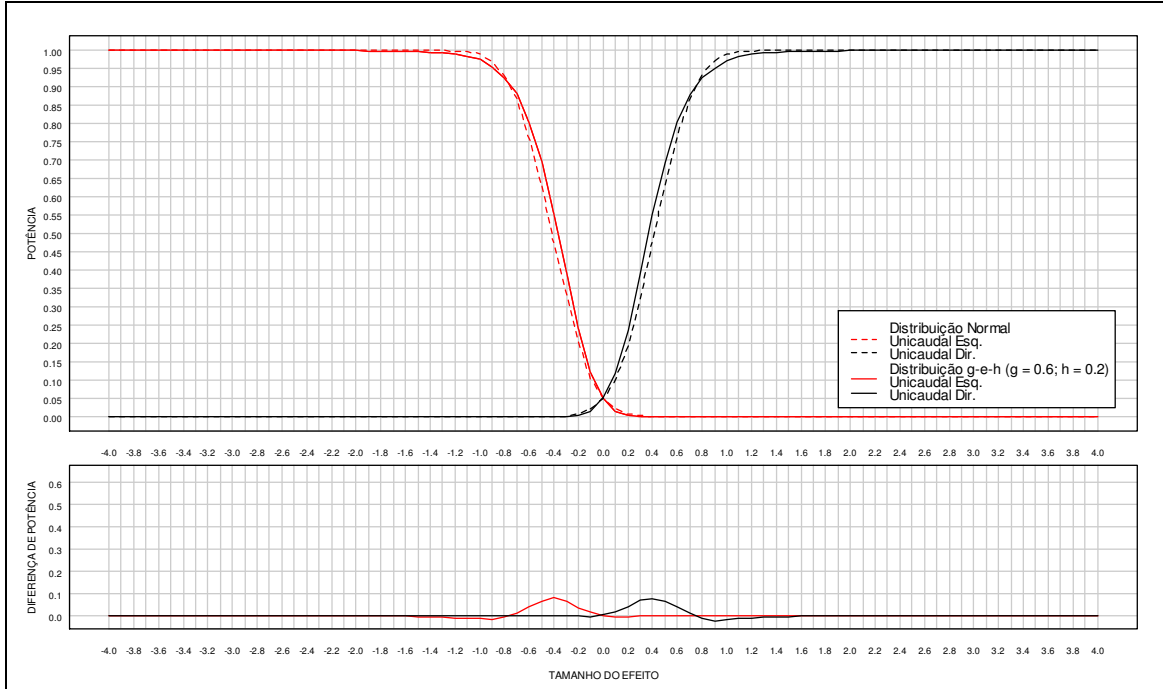


Figura Anexo IX.109 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

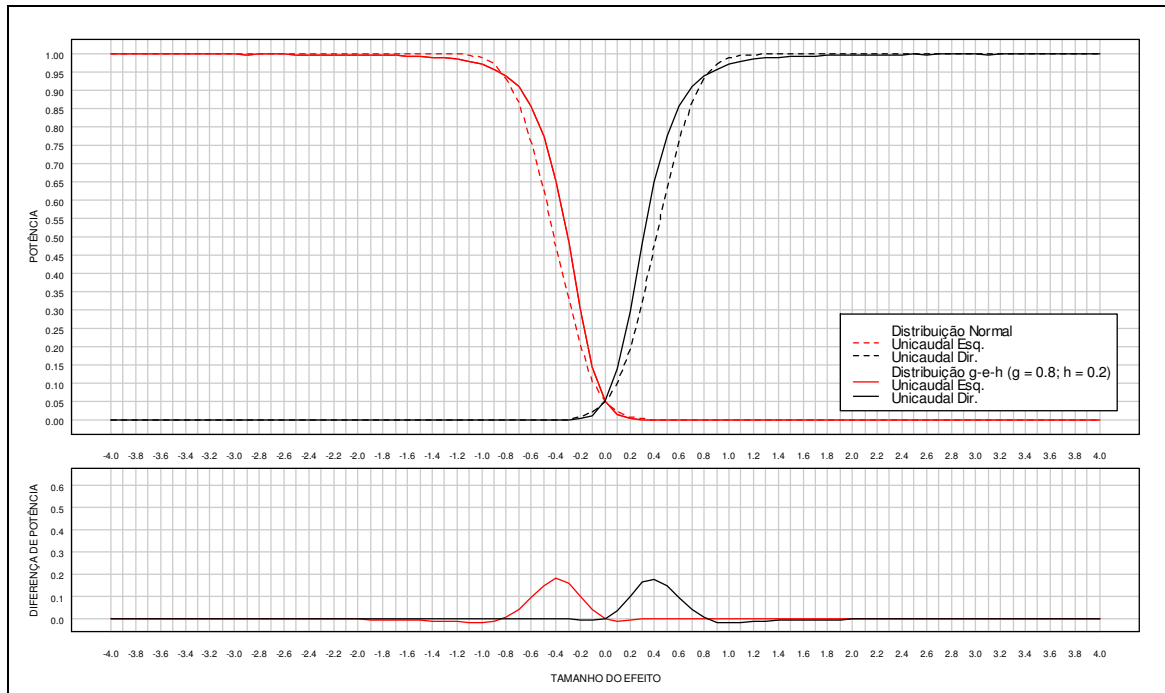


Figura Anexo IX.110 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.2$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

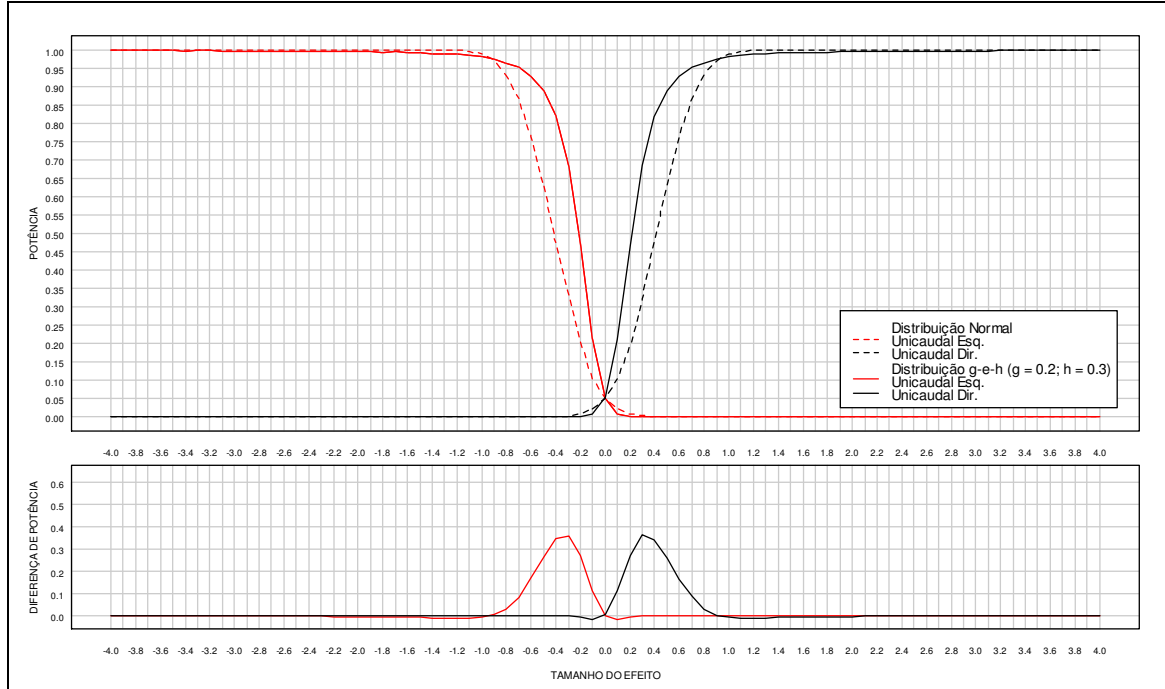


Figura Anexo IX.111 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.2$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

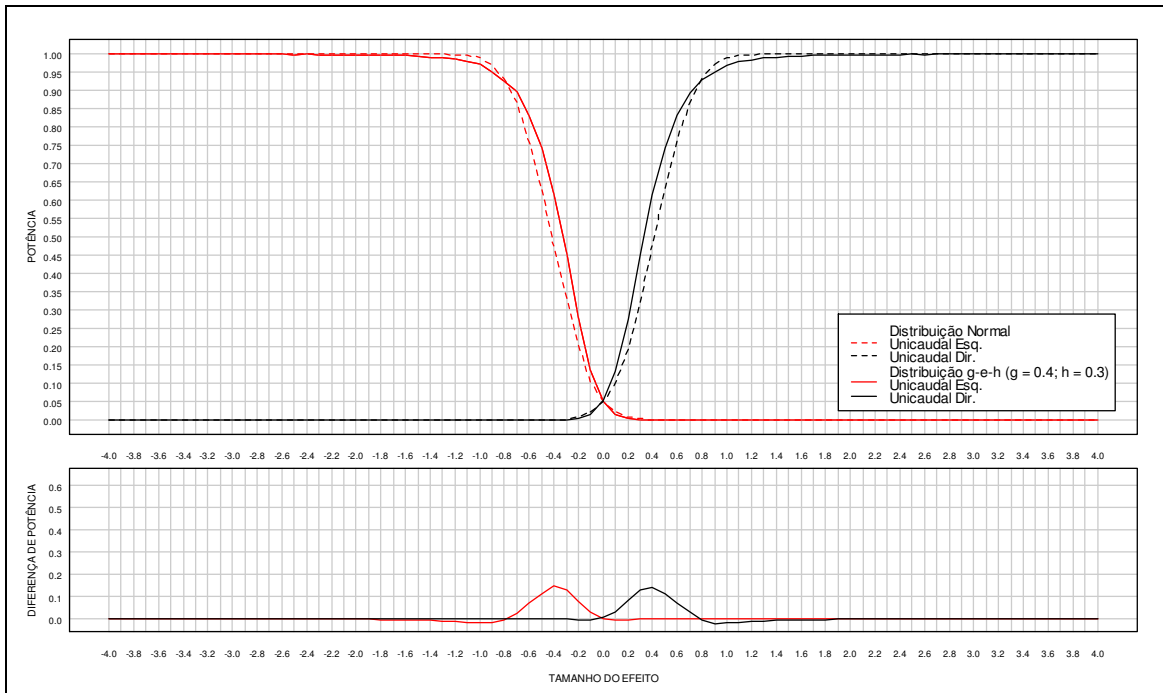


Figura Anexo IX.112 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.4$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

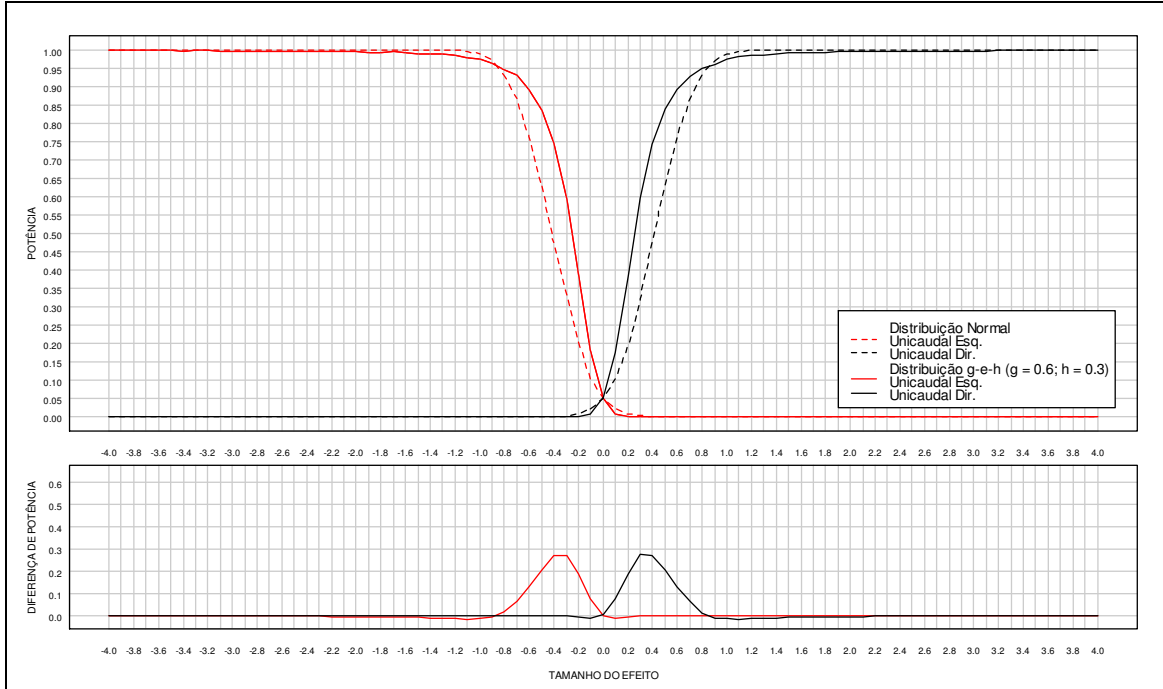


Figura Anexo IX.113 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.6$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

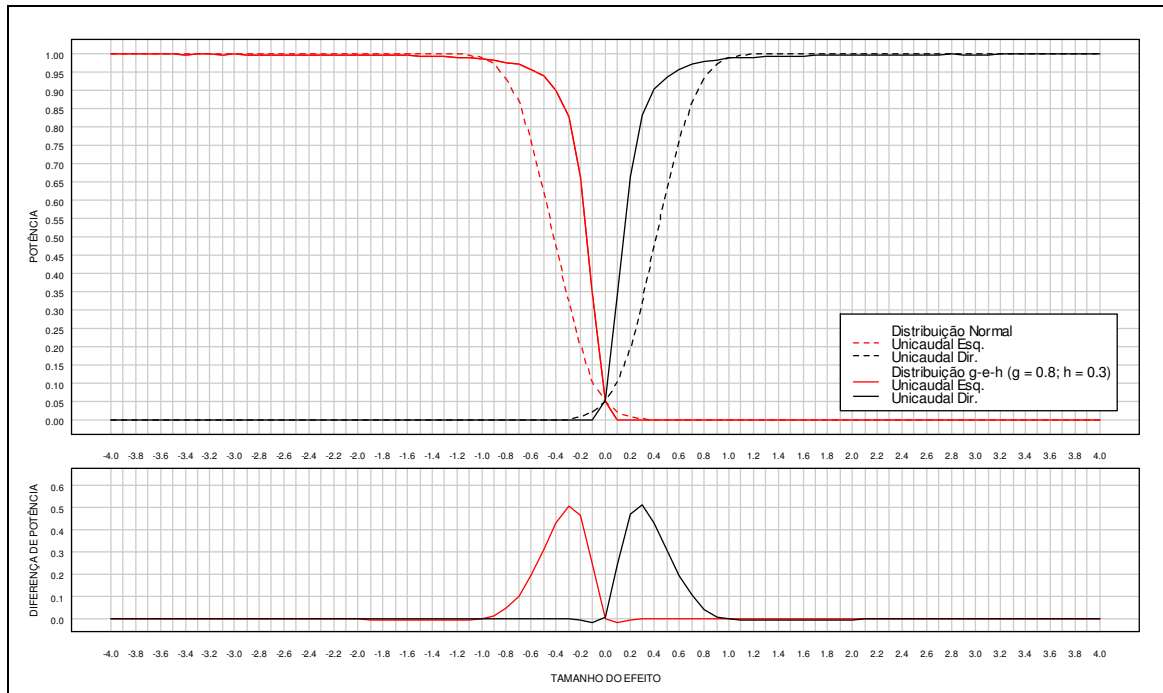


Figura Anexo IX.114 [Painel superior] Potência do teste de aleatorização uni-caudal, em função do tamanho do efeito, com $\alpha = 0.05$, $n_1 = n_2 = 32$, para a distribuição g-e-h, com $g = 0.8$ e $h = 0.3$, em comparação com a sua potência para a distribuição normal. [Painel inferior] Diferença de potências.

Anexo X

Programas em R para o cálculo de potência do Estudo 3

```

#E3C11
#Definições
nsim <- 20000
ngr1 <- 32; ngr2 <- 8
alfa<-0.05
te<- seq(-4,4,by=.10)
set.seed(12345)
#Função: Teste t de Student
teste.t <- function(x1, x2) {
  n1 <- length(x1); n2 <- length(x2)
  mx1 <- mean(x1); mx2 <- mean(x2)
  s1 <- var(x1); s2 <- var(x2)
  s <- ((n1-1)*s1 + (n2-1)*s2)/(n1+n2-2)
  et <- (mx1 - mx2)/sqrt(s*(1/n1 + 1/n2))
  et
}
#Função: Teste de Aleatorização (sample)
permg<-function(x,y,nream=999){
xx<- c(x,y)
etobs1<-sum(x) ; etobs2 <-
(sum(x)^2/length(x))+(sum(y)^2/length(y))
etb1 <- NA
etb2 <- NA
for(i in 1:nream){
etb1[i]<- sum(sample(xx, size = length(x), replace = F))
etb2[i]<- etb1[i]^2/length(x)+(sum(xx)-
etb1[i])^2/length(y)
}
vpesq <- (1+length(etb1[etb1 <= etobs1]))/(nream+1)
vpdir <- (1+length(etb1[etb1 >= etobs1]))/(nream+1)
vpbi <- (1+length(etb2[etb2 >= etobs2]))/(nream+1)
res<-matrix(c(vpesq,vpdir,vpbi),1,3)
}
res.tt<-matrix(0,length(te),3); res.ta<-matrix(0,
length(te),3)
matsig <- matrix(0,nsim,3)
for(k in 1 : length(te)){
dadgr1 <-
matrix(rnorm(n=nsim*ngr1,mean=0,sd=1),nsim,ngr1)
dadgr2 <- matrix(rnorm(n=nsim*ngr2,mean=te[k], sd =
10),nsim,ngr2)
ncesq <- 0; ncdire <- 0; nbi <- 0

```

```

    et<- matrix(0,nsim,1)
    for(i in 1:nsim) {et[i,1] <- teste.t(dadgr1[i,],
dadgr2[i,])
        if(et[i,1]<= qt(alfa,(ngr1+ngr2-2)))
ncesq=ncesq+1
        if(et[i,1]>= qt(1-alfa,(ngr1+ngr2-2)))
ncdir=ncdir+1
        if(et[i,1]<= qt(alfa/2,(ngr1+ngr2-2))|et[i,1]>=
qt(1-alfa/2,(ngr1+ngr2-2))) nbi=nbi+1
        matsig[i,]<-permg(dadgr1[i,],dadgr2[i,])
    }
res.tt[k,1]<-ncesq/nsim; res.tt[k,2]<-ncdir/nsim;
res.tt[k,3]<-nbi/nsim
vecsigesq<-matsig[,1]
vecsigdir<-matsig[,2]
vecsigbi<-matsig[,3]
res.ta[k,1]<-length(vecsigesq[vecsigesq<=alfa])/nsim
res.ta[k,2]<-length(vecsigdir[vecsigdir<=alfa])/nsim
res.ta[k,3]<-length(vecsigbi[vecsigbi<=alfa])/nsim
}
sink("ResE3C12",append=F)
cat(res.tt, "\n")

cat(res.ta)
sink()

```

Nota: Este programa foi usado para o último grupo de condições, com amostras desequilibradas, com $n_1 = 32$ e $n_2 = 8$, e heterocedasticidade, com $\sigma_1 = 1$ e $\sigma_2 = 10$. Os programas usados nas outras condições foram semelhantes a este, com as alterações necessárias nos parâmetros (desvio-padrão e dimensão das amostras).