



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA



Tamanho de Amostra e Poder para Três Testes Não-Paramétricos

Autor: Piel Antônio Castro e Souza
Orientador: Prof. Msc. Vanessa Bielefeldt Leotti Torman

Porto Alegre, 13 de Julho de 2012.
Universidade Federal do Rio Grande do Sul

Instituto de Matemática
Departamento de Estatística

Tamanho de Amostra e Poder para Três Testes Não-Paramétricos

Autor: Piel Antônio Castro e Souza

Monografia apresentada para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:
Prof. Msc. Vanessa Bielefeldt Leotti Torman
Prof. Dra. Patrícia Klarmann Ziegelmann

Porto Alegre, 13 de Julho de 2012.

*Dedico este trabalho à minha família e amigos,
que estiveram comigo durante todo este percurso.*

Resumo

Introdução: Testes não-paramétricos são muito flexíveis quanto a suas suposições e muitas vezes são a única opção disponível aos pesquisadores, mas, diferente dos teste paramétricos, formas de cálculo de tamanho de amostra e poder para estes testes não são bem conhecidas.

Objetivo: Reunir formas de calcular tamanho de amostra e poder para os testes Mann-Whitney, Wilcoxon e Kruskal-Wallis e apresentar programas que façam tais cálculos.

Metodologia: Revisão bibliográfica, com leitura de artigos e livros sobre o assunto, análise de programas envolvendo o tema e desenvolvimento de exemplos para explicar as funções dos programas.

Resultados e Conclusão: Programas que utilizam métodos semelhantes para os cálculos obtiveram resultados bem parecidos, como SAS 9.3, PASS 11 e GPower 3.1 para Mann-Whitney, e PASS 11 e GPower 3.11 para Wilcoxon. O programa PASS 11 pareceu ser o mais completo. Os programas livres apresentam soluções para os testes Mann-Whitney e Wilcoxon, mas não para o teste Kruskal-Wallis. Apenas o método do Banco de Dados Exemplo para o teste Mann-Whitney apresentou a vantagem de não necessitar escolher distribuição.

Palavras-Chave: Tamanho de amostra; Poder; Mann-Whitney; Wilcoxon, Kruskal-Wallis.

Abstract

Background: Non-parametric tests are very flexible about their assumptions and are often the only option available to researchers, but, unlike parametric tests, ways of calculating sample size and power for these tests are not well known.

Aim: Gather ways of calculating sample size and power for Mann-Whitney, Wilcoxon and Kruskal-Wallis tests, and programs that make such calculations for these tests.

Methodology: Review literature by reading articles and books on the subject, analysis of programs involving the theme and examples to explain the functionalities of the programs.

Results and conclusion: Programs that had similar methods for calculating had very similar results, as SAS 9.3, PASS 11 and GPower for Mann-Whitney, and PASS 11 and GPower 3.1 for Wilcoxon. The program PASS 11 appeared to be the most complete. The free software presented solutions to Mann-Whitney and Wilcoxon, but not for the Kruskal-Wallis test. Only the Exemplary Dataset Method to the Mann-Whitney test showed the advantage of not requiring choosing a distribution.

Keywords: Sample Size; Power; Mann-Whitney; Wilcoxon; Kruskal-Wallis.

Sumário

Introdução.....	8
Mann-Whitney	9
SAS 9.3.....	10
R 2.15.0.....	10
PASS 11.....	11
nQuery Advisor 7.....	11
GPower 3.1.....	12
Banco de Dados Exemplo.....	12
Exemplo 1.....	13
SAS 9.3.....	14
R 2.15.0.....	17
PASS 11.....	18
nQuery 7.....	18
GPower 3.1.....	18
Banco de Dados Exemplo.....	19
Wilcoxon.....	21
PASS 11.....	21
GPower 3.1.....	21
Exemplo 2.....	21
PASS 11.....	22
GPower 3.1.....	23
Kruskal-Wallis.....	23
Exemplo 3.....	24
Considerações finais.....	25
Referências.....	27

Este artigo será submetido à Revista do HCPA.

Introdução

Muitas vezes pesquisadores necessitam utilizar testes não-paramétricos, seja pelos dados não atenderem a alguma suposição dos testes paramétricos ou porque a amostra é pequena. Testes não-paramétricos não tem formas de cálculo de tamanho de amostra e poder tão bem difundidas como os testes paramétricos. A falta destes pode dificultar discussões mais aprofundadas sobre o poder de um estudo já realizado e o planejamento de novos estudos. Por exemplo, digamos que após um teste de hipóteses que tenha resultado em uma diferença não significativa, estamos interessados em saber qual o poder para detectar a diferença desejada alcançado com o tamanho de amostra utilizado. Com esta informação, caso o poder seja alto podemos ficar mais confortáveis com o resultado, ou, e se ele for pequeno, podemos planejar um novo estudo com uma amostra maior com poder suficiente para detectar a diferença mínima esperada.

Nosso objetivo foi reunir formas de cálculo de tamanho de amostra e poder para os testes Mann-Whitney, Wilcoxon e Kruskal-Wallis. Foram escolhidos estes testes por serem bastante utilizados na prática. O principal objetivo foi encontrar programas que realizem tais cálculos, descrever as informações necessárias para realizar os procedimentos nos devidos programas e apresentar exemplos. Os programas pagos que foram utilizados na pesquisa são o nQuery Advisor 7 (1), PASS 11 (2) e SAS 9.3 (3). Os programas livres utilizados foram GPower 3.1 (4) e o R 2.15.0 (5).

Agora serão apresentadas informações introdutórias sobre os testes abordados para melhor entendimento do leitor, para mais informações sugere-se, por exemplo, consultar Sheskin (6). O teste de Mann-Whitney (também chamado de teste de Wilcoxon-Mann-Whitney ou teste de soma de postos de Wilcoxon) é utilizado para testar hipóteses envolvendo amostras de dois grupos independentes, como no caso onde se quer comparar dois tratamentos aplicados a dois grupos de pacientes diferentes. O teste, se rejeitar a sua hipótese nula, indica alta possibilidade que exista uma diferença entre as medianas das duas populações de onde as duas amostras foram extraídas. O teste Mann-Whitney é alternativo ao teste t para duas amostras independentes, que é paramétrico. A estatística U do teste tem uma distribuição exata, comumente tabelada com os valores críticos de U em referência aos tamanhos de cada grupo. Para amostras grandes pode-se utilizar uma transformação na estatística U do teste que se aproxima da distribuição normal.

O teste de Wilcoxon é utilizado para situações com duas amostras dependentes (relacionadas, ou pareadas), como no caso de medições realizadas em pacientes antes e após um tratamento. Este teste é uma extensão do teste de Wilcoxon para uma amostra, pois é utilizada a diferença entre os valores dos pares de cada amostra para o cálculo da estatística. A hipótese nula testada é se a mediana da população de diferenças é zero. Se uma diferença significativa é encontrada a possibilidade é alta que as amostras foram extraídas de duas populações com medianas diferentes. A estatística calculada no teste é comparada com a distribuição exata da estatística, comumente tabelada. O teste Wilcoxon é alternativo ao teste t , paramétrico, para duas amostras dependentes. A distribuição da estatística do teste de Wilcoxon com uma transformação pode ser aproximada pela normal, para amostras suficientemente grandes.

O teste Kruskal-Wallis é uma extensão do teste de Mann-Whitney para envolver desenhos com duas ou mais amostras independentes (k amostras), e se $k=2$ o resultado do teste será equivalente ao teste de Mann-Whitney. É o teste alternativo à ANOVA de um fator, paramétrica. O resultado do teste Kruskal-Wallis quando significativo indica alta probabilidade da existência de diferença entre pelo menos duas medianas no grupo de k medianas populacionais, e nesse caso é necessário conduzir um teste de comparação múltipla para identificar quais pares de medianas são possivelmente diferentes. A distribuição Qui-Quadrado é utilizada para aproximar a distribuição da estatística de teste.

Para a construção deste trabalho foram realizadas revisões bibliográficas, como leitura de livros e artigos sobre o assunto. Também foram feitas análises sobre os programas com leitura dos manuais e utilização dos mesmos para resolver exemplos. Nas seções a seguir serão explicados os programas e métodos utilizados, cada uma referente a um teste e com um exemplo ao final. Primeiramente será apresentado o teste Mann-Whitney com os programas SAS 9.3, PASS 11, nQuery Advisor 7, GPower 3.1, e R 2.15 e o método do banco de dados exemplo proposto por Divine *et al* (7) e inserir a referência. Para o teste Wilcoxon serão apresentados os programas PASS 11 e GPower 3.1. Finalizando, explicaremos os procedimentos do programa PASS 11 para o teste Kruskal-Wallis.

Mann-Whitney

O teste Mann-Whitney é utilizado para testar hipóteses envolvendo amostras de dois grupos independentes e tem as seguintes suposições: cada amostra foi selecionada

aleatoriamente da população que representa; as amostras são independentes uma da outra; a variável aleatória observada é originalmente contínua; a distribuição das populações das amostras deve ter mesmo formato.

SAS 9.3

O programa SAS 9.3 (3) utiliza o método de O'Brien-Castelloe (8) para o cálculo de tamanho de amostra e poder do teste. Este realiza uma aproximação baseada no comportamento assintótico da transformação da estatística U da estatística à distribuição normal para tamanhos grandes de amostra. Expressa o tamanho do efeito pelo logaritmo da razão de chances do teste:

$$\log(\text{WMWodds}) = \log\left[\frac{p''}{1-p''}\right]$$

onde $p'' = P(X < Y)$, X e Y sendo amostras aleatórias das duas distribuições comparadas. A qualidade da aproximação do poder diminui para amostras pequenas por utilizar a propriedade assintótica da estatística do teste.

É necessário escolher uma distribuição para a variável sob a hipótese alternativa para que seja realizado o cálculo do tamanho da amostra ou do poder do teste. O SAS 9.3 possui as seguintes distribuições: categórica ordinal, Beta, Binomial, Exponencial, Gama, Laplace, Logística, Lognormal, Normal, Poisson e Uniforme.

Esta escolha de distribuição, que como veremos está presente na maioria dos métodos, é uma grande desvantagem, pois é incoerente com o fato de que justamente recorre-se aos testes não-paramétricos para não ter que exigir uma distribuição para a variável estudada. Entretanto, para se fazer essa escolha, pode-se recorrer à literatura ou proceder testes de aderência.

O procedimento **PROC POWER** com a opção **TWOSAMPLEWILCOXON** é o que calcula tanto o tamanho de amostra quanto o poder para o teste Mann-Whitney no software SAS 9.3. Os comandos necessários para cálculo estarão detalhados no exemplo. Os resultados são exibidos pelo software em forma de tabela e pode-se adicionar um gráfico dos valores do poder contra o tamanho de amostra. O programa pode trabalhar com tamanhos de amostras diferentes entre os grupos.

R 2.15.0

No pacote *samplesize* do R 2.15 (5) existe a função *n.wilcox.ord* para calcular o tamanho de amostra para o teste de Mann-Whitney para variáveis categóricas ordinais. O cálculo é feito a partir do método de Zhao-Rahardja-Qu (equação E1). Não foi encontrada

nenhuma função para cálculo de poder do teste Mann-Whitney no R. O uso da função *n.wilcox.ord* será mais detalhado no exemplo.

$$N = \frac{(Z_{\alpha} + Z_{\beta})^2 (1 - \sum_{c=1}^D ((1-t)p_c + tq_c)^3)}{12t(1-t) (\sum_{c=2}^D p_c \sum_{d=1}^{c-1} q_d + 0,5 \sum_{c=1}^D p_c q_c - 0,5)^2} \quad (E1)$$

Na Equação (E1) temos que D é número de possíveis respostas, p_c e q_c são as proporções hipotéticas no nível c das duas populações que estão sendo comparadas, m e n são os tamanhos de amostra para cada um dos grupos, $N = m + n$, $t = n/N$ é a proporção de observações no segundo grupo, e Z_{α} e Z_{β} são os valores da distribuição normal padrão associados ao erro do tipo I e II, respectivamente. Z_{α} é o valor positivo com área nas caudas igual a α e Z_{β} o ponto que possui área à direita no valor de β .

PASS 11

O método seguinte é o utilizado no software PASS 11 (2) que realiza uma aproximação do tamanho de amostra e o poder a partir dos valores obtidos para o teste t (9), o tamanho do ajuste depende da distribuição sob a hipótese alternativa. Os coeficientes que são multiplicados pelo tamanho de amostra necessário para o teste t são 1 para distribuição uniforme, 2/3 para Laplace, $9/\pi^2$ para distribuição logística e $\pi/3$ para distribuição normal.

Os parâmetros e opções que devem ser definidos são as médias e desvios padrões da variável em cada grupo comparado, nível de significância, poder ou erro tipo II (apenas para calcular tamanho de amostra), tamanho da amostra em cada grupo (apenas para o cálculo do poder), o tipo da hipótese alternativa e qual das distribuições sob a hipótese alternativa utilizar.

Utilizando uma lista de valores para o tamanho de amostra temos como resultado uma tabela com os valores de poder para cada um, e pode-se solicitar um gráfico para ilustrar esses resultados, como será visto no exemplo

nQuery Advisor 7

A quarta forma de cálculo de tamanho de amostra e poder é a do programa nQuery Advisor 7 (1). Para variáveis contínuas o programa utiliza a aproximação da estatística do teste à normal a partir do método de Noether (10) e para variáveis ordinais computa o poder aproximando a distribuição condicional da estatística do teste sob a hipótese alternativa (11).

Para calcular o poder do teste para variáveis contínuas é necessário informar o nível de significância, se o teste é bilateral ou unilateral, média da variável em cada grupo, tamanho da amostra e desvio padrão comum aos grupos, não é possível especificar desvios padrões

diferentes, apenas um desvio conjunto e também não podemos especificar grupos com tamanhos diferentes. Com estas informações o programa calcula o tamanho do efeito e p'' . A hipótese nula é $p''=0,5$, o que implica em medianas iguais supondo que a distribuição da variável tem o mesmo formato em cada grupo. O programa apresenta como resultado o valor solicitado na célula designada com o nome **Power (%)**, no caso do cálculo do poder, ou **N Total**, no caso de cálculo de tamanho de amostra.

GPower 3.1

Como os demais programas, para o cálculo do poder e tamanho de amostra no GPower 3.1 (4) é necessário especificar uma distribuição sob a hipótese alternativa. As distribuições que o programa disponibiliza são Normal, Dupla Exponencial, Logística pelo método ARE. Entretanto, pode-se utilizar o método ARE mínimo e nesse caso não é necessário especificar distribuição, mas o poder será subestimado e o tamanho da amostra, superestimado. O método ARE (Eficiência Relativa Assintótica - *asymptotic relative efficiency* ou ARE) (12) define a razão entre o poder do teste Mann-Whitney e o poder do teste t para duas amostras independentes. Se os tamanhos de amostra em cada grupo $n = m$ é necessário para atingir o poder especificado para o teste Mann-Whitney e os tamanhos de amostra $n' = m'$ são necessárias para o teste t para atingir o mesmo poder, então a razão n'/n é chamada de eficiência do teste Mann-Whitney relativa ao teste t de duas amostras independentes. A eficiência limite para tamanhos de amostra $n = m$ crescendo para infinito é chamado de Eficiência Relativa Assintótica.

As informações necessárias para o cálculo do poder do teste são as médias e desvios-padrão de cada grupo, tipo do teste (unilateral ou bilateral), nível de significância, tamanho de amostra de cada grupo e a proporção entre o tamanho das amostras. Para calcular o tamanho de amostra são as mesmas opções exceto pela exclusão dos tamanhos de amostra e a adição do poder do teste desejado.

Para o cálculo do poder não podemos estipular diferentes tamanhos de amostra em cada grupo na sessão principal, apenas na janela de gráfico. Nesta janela podemos escolher uma amplitude de valores para o tamanho de amostra total e como resultado temos um gráfico do poder contra os tamanhos de amostra estipulados.

Banco de Dados Exemplo

A última forma, que chamaremos de método do banco de dados exemplo, foi proposta por Divine *et al* (7) a partir das formulas de Noether (10) e Zhao–Rahardja–Qu (13). É um método de cálculo de tamanho de amostra e poder que utiliza informações de um banco

de dados que reflita a hipótese alternativa a ser testada. Divine *et al* apontam que a fórmula proposta por Noether assumia que a variável observada era contínua e não havia empates, o que foi corrigido por Zhao *et al* na equação (E1).

Zhao–Rahardja–Qu afirmam que a fórmula é flexível o suficiente para lidar com dados com ou sem empates, dados que incluam uma mistura de empates e não empates ou observações contínuas. Divine *et al* derivam o método do banco de dados exemplo a partir da equação de Zhao *et al* ao isolar na equação (E1) a estatística Qui-Quadrado do teste Mann-Whitney. Se o tamanho de amostra observado foi N_{obs} (somando os tamanhos amostrais dos dois grupos) e denotamos a estatística Qui-Quadrado do teste Mann-Whitney como χ_{obs}^2 , temos

$$N = \frac{N_{obs}(Z_{\alpha}+Z_{\beta})^2}{\chi_{obs}^2} \quad (E2)$$

Isto é, se χ_{obs}^2 é realmente uma realização da estatística de teste utilizando dados que reflitam a hipótese alternativa e que tenham tamanho de amostra N_{obs} , segue que a equação (E2) representa o tamanho de amostra estimado pelo método do banco de dados exemplo.

Caso o objetivo seja estimar o poder do teste utilizando o método do banco de dados exemplo, podemos resolver a equação (E2) para Z_{β} . Serão apresentadas no exemplo duas funções que desenvolvemos no R para facilitar os cálculos.

Exemplo 1

O exemplo a seguir será utilizado para ilustrar todas as formas apresentadas de cálculo de tamanho de amostra e poder para o teste de Mann-Whitney. No exemplo de Puff City (14), pesquisadores querem testar o efeito de um tratamento de asma em 260 crianças afrodescendentes. Os pacientes foram selecionados aleatoriamente e alocados em dois grupos, controle e intervenção (tratamento para asma). O primeiro grupo tinha 129 crianças e o segundo, 131. A variável observada foi o número de visitas à emergência nos 12 meses após a alocação. A Figura 1 apresenta as distribuições do número de visitas para os dois grupos. A maioria das crianças selecionadas não tiveram visitas à emergência em ambos os grupos, entretanto, o grupo tratamento parece ter um deslocamento para zero. O teste de Mann-Whitney aplicado a esses dados não detecta diferença significativa ($U=7642$, $p=0,0656$). Informações mais detalhadas sobre os grupos encontram-se na Tabela 1.

	Grupo	
	Controle	Tratamento
Média	0,7519	0,4885
Mediana	0	0
Desvio Padrão	1,9326	1,9860
1º Quartil	0	0
3º Quartil	1	0

Tabela 1 - Análise descritivas dos grupos do exemplo de Puff City.

Um dos objetivos deste exemplo será calcular o tamanho de amostra necessário para considerar os dois grupos diferentes ao nível de significância de 5% e poder de 80%. O outro será verificar o poder que a amostra de tamanho 260 tem de detectar diferença a 5% de significância, dado a diferença encontrada nos dados.

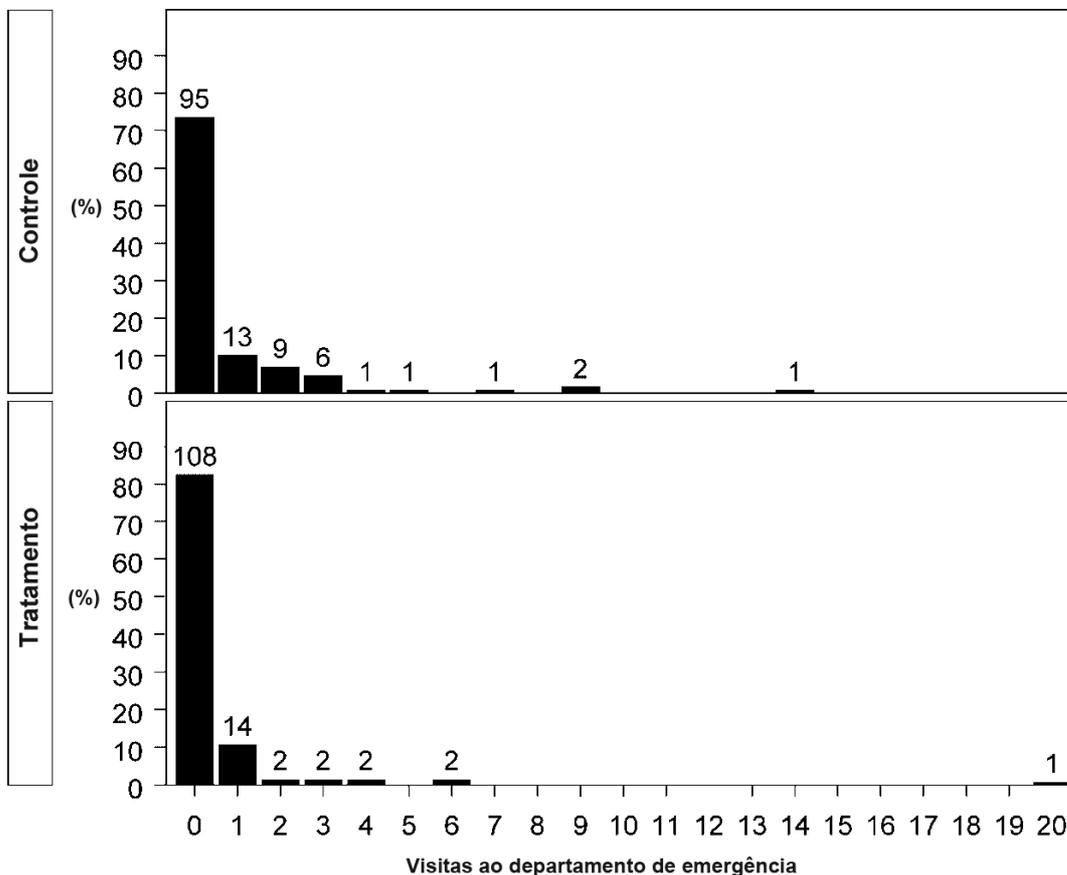


Figura 1 - Distribuição do número crianças por número de visitas à emergência do exemplo Puff City para o grupo controle e o grupo tratamento.

SAS 9.3

Vamos ilustrar como calcular o tamanho de amostra no SAS primeiramente informando toda a distribuição ordinal da variável, e após se assumirá que a mesma tem

distribuição Normal. É importante ressaltar que a escolha da distribuição normal não é adequada para os dados deste exemplo, que são discretos e assimétricos, mas será utilizada aqui para fins de comparação dos métodos.

A seguir temos as linhas de códigos para o cálculo do tamanho de amostra no modo ordinal.

```
1. proc power;  
   a. twosamplewilcoxon  
   b. vardist("Controle") = ordinal ((0 1 2 3 4 5 6 7 8 9 10 11 12  
      13 14 15 16 17 18 19 20) : (.735 .1 .07 .047 .008 .008 0 .008  
      0 .016 0 0 0 0 .008 0 0 0 0 0 0))  
   c. vardist("Tratamento") = ordinal ((0 1 2 3 4 5 6 7 8 9 10 11  
      12 13 14 15 16 17 18 19 20) : (.8245 .1075 .015 .015 .015 0  
      .015 0 0 0 0 0 0 0 0 0 0 0 0 .008))  
   d. variables = "Controle" | "Tratamento"  
   e. nbins = 10000  
   f. ntotal = .  
   g. alpha = 0.05  
   h. power = 0.1 0.8 0.9;  
2. plot step=0.1;  
3. run;
```

Abaixo temos a descrição das linhas de comando anteriores.

1. Procedimento para cálculo de tamanho de amostra e poder do programa SAS.
 - a. **twosamplewilcoxon**, opção que define o tipo de teste, com visto na introdução o teste Mann-Whitney pode ser chamado também de Wilcoxon-Mann-Whitney.
 - b. **vardist**, define a distribuição da variável do grupo 1 (Controle). Foram necessários arredondamentos nos valores das proporções para que a soma resultasse um.
 - c. **vardist**, define a distribuição da variável do grupo 2 (Tratamento). Foram necessários arredondamentos nos valores das proporções para que a soma resultasse um.
 - d. **variables**, nomeia as variáveis a serem testadas.
 - e. **nbins**, número de classes que será dividida a distribuição de cada variável, utilizado na computação interna, uma maior quantidade de classes levará a um resultado com maior precisão. Foi deixado o valor que vem por definição que é 10 mil divisões.

- f. **ntotal**, tamanho da amostra total. Quando se deseja calcular o tamanho de amostra, basta apontá-lo como *missing* (`ntotal = .`) Ao invés deste, pode-se colocar **npergroup**, quando se deseja o tamanho da amostra por grupo.
 - g. **alpha**, nível de significância escolhido.
 - h. **power**, poder do teste. Pode ser uma lista de valores quando for calcular o tamanho de amostra. Para o cálculo do poder do teste deve ser informado como *missing* (`power = .`).
2. **plot**, opção para gerar um gráfico de poder contra tamanho de amostra quando se define mais de um valor de poder ou tamanho de amostra. **step** define o espaçamento entre os valores.
 3. **run**, Linha de comando final para sinalizar fim do processo.

As linhas de código para o exemplo considerando distribuição normal para as variáveis, apresentadas abaixo, tem pouca diferença em relação ao código apresentado anteriormente, mudando apenas as linhas 1.b e 1.c onde se define a distribuição da variável em cada grupo e a falta do comando **plot**.

```

1. proc power;
   a. twosamplewilcoxon
   b. vardist("Controle") = normal (0.7519 , 1.93258 )
   c. vardist("Tratamento") = normal (0.4885 , 1.98598 )
   d. variables = "Controle" | "Tratamento"
   e. nbins = 10000
   f. ntotal = .
   g. alpha = 0.05
   h. power = 0.8;
2. run;

```

Para os dados do exemplo o PROC POWER do SAS 9.3 apresentou uma estimativa de tamanho de amostra total de 576 (288 por grupo) utilizando a distribuição ordinal dos valores. Supondo uma distribuição normal o programa encontrou uma amostra total de 1822, muito maior que utilizando a distribuição categórica ordinal.

Analisando o tamanho de amostra original do estudo ,260 indivíduos, pelo programa foi encontrado um poder de 47,10% como podemos ver na Figura 2, o gráfico de tamanho de amostra contra poder, considerando a distribuição ordinal da variável. No gráfico também podemos encontrar o tamanho de amostra necessário para atingir os 80% desejados, assim como o ganho de poder começa a diminuir para amostras maiores de 600.

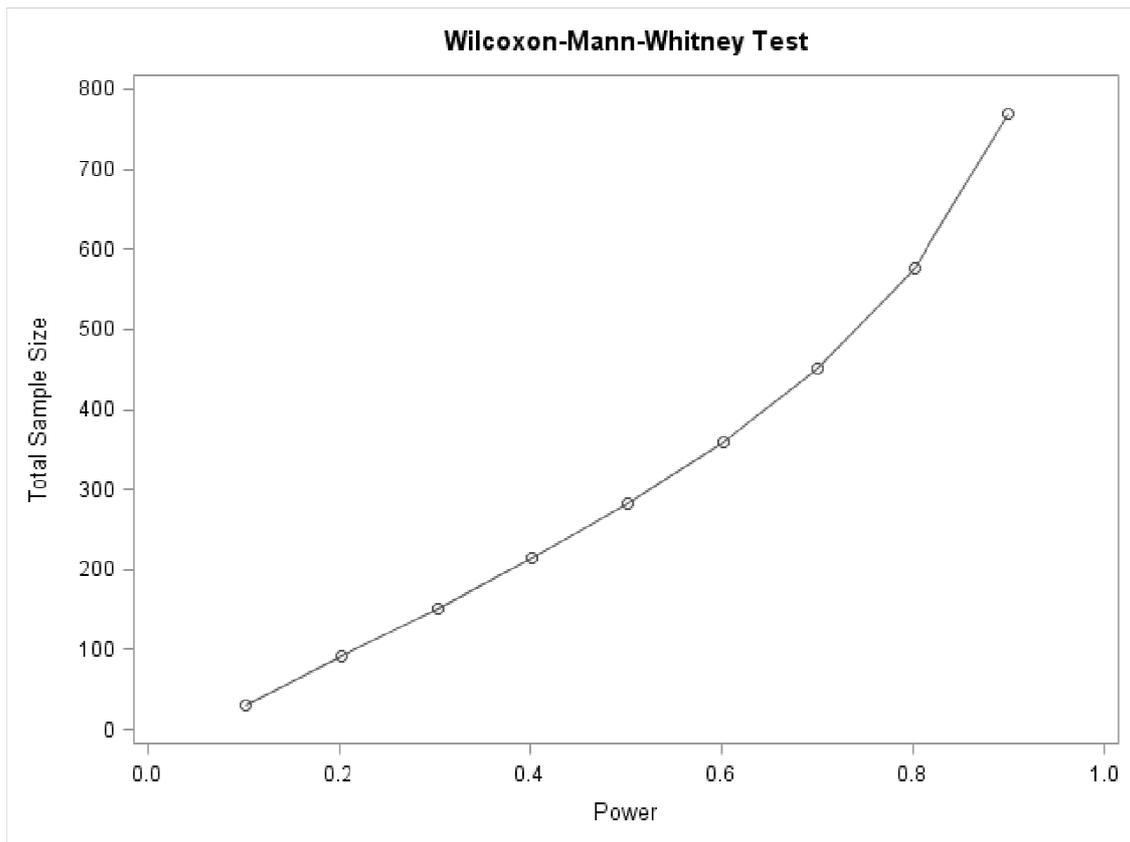


Figura 2 - Poder contra Tamanho de Amostra pelo programa SAS 9.2 considerando distribuição ordinal.

R 2.15.0

Para aplicar a fórmula de Zhao *et al* (E1), implementada na função *n.wilcox.ord* do R, o primeiro passo é calcular p_c , que é a proporção de crianças no grupo 1 para cada classe c , e q_c , que é a mesma proporção mas para o grupo 2. Por exemplo, para calcular p_0 , a proporção de crianças do grupo 1 que não fizeram visitas à emergência, toma-se o número de crianças do grupo 1 que não tiveram visitas (95) dividido pelo número total de crianças no grupo 1 (129).

No software R, o tamanho de amostra pode ser calculado com os seguintes comandos.

1. `grupo1<- c(0.736434109, 0.100775194, 0.069767442, 0.046511628, 0.007751938, 0.007751938, 0, 0.007751938, 0.015503876, 0.007751938, 0)`
2. `grupo2<- c(0.824427481, 0.106870229, 0.015267176, 0.015267176, 0.015267176, 0, 0.015267176, 0, 0, 0, 0.007633588)`
3. `library(samplesize)`
4. `n.wilcox.ord(beta=0.2, alpha=0.05, t=0.5, p=grupo1, q=grupo2)`

A seguir temos a descrição das linhas de comandos e dos argumentos necessários para a função *n.wilcox.ord*.

1. Criação do vetor de proporções das categorias para o número de visitas à emergência no grupo 1, para cada categoria é número de crianças da categoria dividido pelo total de crianças. O vetor deve somar 1.
2. Criação do vetor de proporções das categorias para o número de visitas à emergência no grupo 2, para cada categoria é número de crianças da categoria dividido pelo total de crianças. O vetor deve somar 1.
3. Função para carregar o pacote *samplesize*, o qual tem a função *n.wilcox.ord*.
4. Função que calcula o tamanho de amostra para o teste de Mann-Whitney. Necessita dos argumentos.
 - a. **beta**, nível de erro tipo II.
 - b. **alpha**, nível de erro tipo I
 - c. **t**, quociente do tamanho de amostra n/N , onde n é o tamanho de amostra do grupo B e N é o tamanho total da amostra, no caso, estamos considerando grupos com tamanho de amostra iguais.
 - d. **p**, vetor de proporções das categorias do grupo 1.
 - e. **q**, vetor de proporções das categorias do grupo 2.

Foi obtido o valor de 600 (599,7698 exatamente) para o tamanho de amostra total, ou seja, 300 por grupo.

PASS 11

Por não ter solução para dados categóricos, assumiu-se distribuição Normal para resolver o exemplo no programa PASS 11, resultando num tamanho de amostra total de 1822, mesmo resultado encontrado pelos programas SAS e G*Power. O programa encontrou um poder de 18,39% para o teste com a amostra original de 260 indivíduos.

nQuery 7

No programa nQuery Advisor 7, não pode-se utilizar o modo de dados categóricos no exemplo Puff City, pois aceita apenas 8 níveis de resposta para variáveis ordinais e no exemplo temos 11 níveis. Utilizando o modo de aproximação pela normal o programa resultou em 904 amostras por grupo, um total de 1808. Analisando a amostra original de 260 indivíduos o programa encontrou um poder de 18%.

GPower 3.1

No software GPower, utilizando o método ARE mínimo foi encontrado que será necessário uma amostra de 2014 indivíduos, 1007 por grupo. Realizando o cálculo escolhendo

a distribuição normal foi encontrado o tamanho de amostra de 1822, assim como no software SAS. Na Figura 3 temos o gráfico de tamanho de amostra contra poder pelo método ARE mínimo, podemos ver nele o crescimento do poder do teste com o incremento do tamanho da amostra. Utilizando o programa para analisar a amostra original de 260 indivíduos foi encontrado um poder de 17,08%.

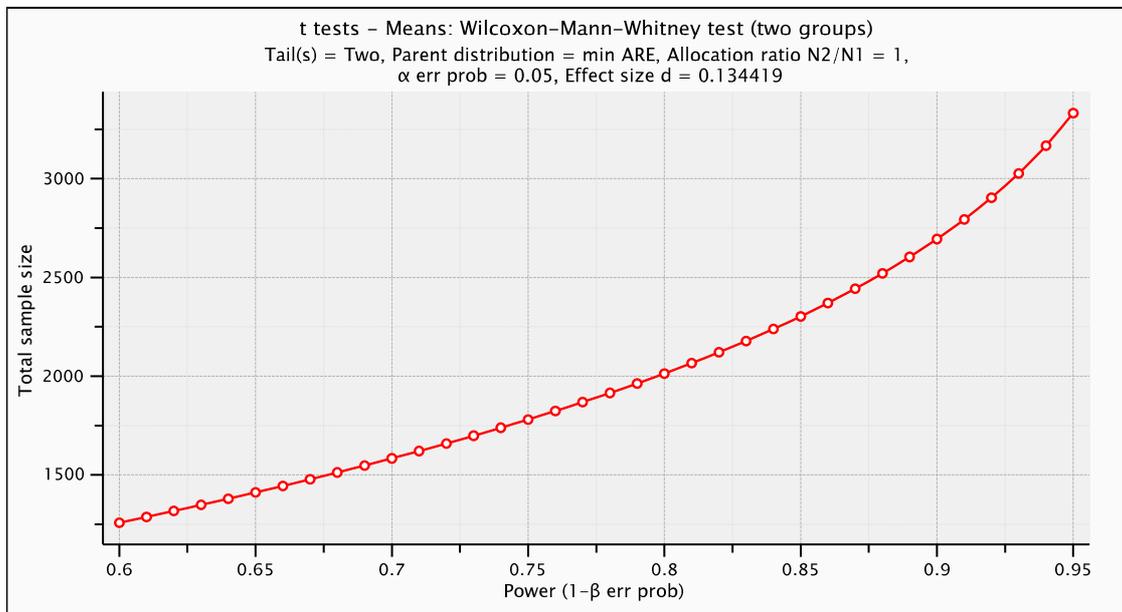


Figura 3 - Poder contra Tamanho de Amostra No Programa G*Power 3.1 pelo método ARE mínimo para o exemplo de Puff City.

Banco de Dados Exemplo

Para utilizar o método do banco de dados exemplo, é necessário ter a estatística do teste Qui-Quadrado de Mann-Whitney para a comparação dos dois grupos, que no caso foi de 3,393, para o tamanho de amostra total de 260. A equação (E2) foi implementada na função do R conforme linhas de comando a seguir.

1. `source(file="https://sites.google.com/site/monografiapiel/home/tamanho-de-amostra/tamanho_de_amostra_bde.txt")`
2. `tam.bde(nobs=260, xobs=3.393, alfa=0.05, beta=0.20, bilateral=T)`
3. 601.4467

Abaixo as descrições do código da função para cálculo de tamanho de amostra baseada no método do banco de dados exemplo.

1. Comando que busca na internet a função para o tamanho de amostra e a carrega no programa R.
2. A função `tam.bde` necessita das seguintes informações.
 - i. **nobs**, tamanho de amostra do banco de dados exemplo.

- ii. **xobs**, estatística Qui-Quadrado do banco de dados exemplo.
- iii. **alfa**, nível de significância desejado, por definição é 5%.
- iv. **beta**, nível de erro tipo II desejado, por definição é 20% (poder de 80%).
- v. **bilateral**, se o teste será bilateral ou unilateral.

3. Resultado da função.

Assim, obtém-se o tamanho de amostra total de 602 arredondando, o que equivale a 301 por grupo. É importante ressaltar que o método do banco de dados exemplo só trabalha com grupos de tamanhos iguais.

Pelo método do banco de dados exemplo o poder pode ser encontrado ao se isolar Z_β na equação (E2), como citado anteriormente. Uma vez isolado basta substituir os valores de $N_{obs} = 260$, $\chi_{obs} = 3,393$ e N , este último foi escolhido como 260 também para encontrarmos o poder para a amostra atual. Encontrado o valor de $Z_\beta = -0,1180$, verificamos na distribuição normal padrão o valor sob a curva à direita de Z_β , 54,69% foi o poder encontrado. Este processo foi implementado na função do R a seguir.

1. `source(file="https://sites.google.com/site/monografiapiel/home/tamanho-de-amostra/poder_bde.txt")`
2. `pod.bde(n=260, nobs=260, xobs=3.393, alfa=0.05, bilateral=T)`
3. 0.5469

A seguir as descrições do código da função acima para cálculo de tamanho de amostra baseada no método do banco de dados exemplo.

1. Função que busca na internet a função para o poder e a carrega no programa R.
2. A função *pod.bde* necessita das seguintes informações.
 - i. **n**, tamanho de amostra do estudo futuro.
 - ii. **nobs**, tamanho de amostra do banco de dados exemplo.
 - iii. **xobs**, estatística Qui-Quadrado do banco de dados exemplo.
 - iv. **alfa**, nível de significância desejado, por definição é 5%.
 - v. **bilateral**, se o teste será bilateral ou unilateral.
3. Resultado da função.

Wilcoxon

O teste de Wilcoxon é utilizado para situações com duas amostras dependentes e tem as seguintes suposições: cada amostra foi selecionada aleatoriamente da população que representa; a variável aleatória observada é originalmente contínua; os valores das diferenças são distribuídos simetricamente em torno da mediana da população de diferenças.

PASS 11

Assim como o para o teste de Mann-Whitney, o programa PASS aproxima o tamanho de amostra e o poder a partir da relação com o poder do teste t (9), o tamanho do ajuste depende da distribuição sob a hipótese alternativa. As distribuições disponíveis são as mesmas que para o teste Mann-Whitney.

Os parâmetros e opções que devem ser definidos são a média da diferença entre os pares, desvio padrão da diferença entre os pares, nível de significância, poder ou nível de erro tipo II (para calcular tamanho de amostra), tamanho da amostra (para o cálculo do poder), tipo da hipótese alternativa e qual das distribuições sob a hipótese alternativa segue a diferença entre os pares. Possui as mesmas opções de resultados que para o teste Mann-Whitney.

GPower 3.1

Como o programa GPower 3.1 é um software livre e está constantemente em desenvolvimento as informações sobre o método utilizado não estavam completas no endereço dos desenvolvedores (15), mas por se tratar de um teste semelhante ao Mann-Whitney supomos que tenha uma construção parecida.

Os dados necessários para a realização dos cálculos de tamanho de amostra são nível de significância, poder (para cálculo de tamanho de amostra), se o teste é unilateral ou bilateral, a distribuição considerada e as informações sobre os parâmetros que podem ser definidas separadamente por grupo (média e desvios da variável em cada grupo e correlação entre os grupos) ou utilizar a diferença entre os valores (média e desvio padrão das diferenças).

Exemplo 2

Como exemplo extraído do manual do PASS 11 (16), digamos que pesquisadores queiram estudar o impacto de um programa de exercícios na massa corporal de um indivíduo.

Para tanto eles selecionam N indivíduos, medem a massa corporal, colocam o indivíduo no programa de exercícios e medem novamente a massa. A variável de interesse não é a massa em si, mas sim o quanto houve de mudança do estado inicial para o final.

Para estudar o impacto do programa de exercícios, os pesquisadores escolhem conduzir o teste pareado de Wilcoxon, pois acreditam que a distribuição das diferenças entre o estado inicial e final não deve se adequar as suposições do teste t para amostras pareadas. Experimentos passados deste tipo usam desvio padrão de 10 quilogramas. Os pesquisadores desejam encontrar uma diferença média de 5 quilogramas ou mais. O nível de significância será 5%, poder de 80%, num teste bilateral.

PASS 11

O tamanho de amostra calculado pelo programa PASS 11 para o exemplo foi de 36 indivíduos. Para o cálculo foi escolhido que a distribuição sob a hipótese alternativa era a distribuição normal. Podemos ver na Figura 4 um gráfico com o estudo do poder contra tamanho de amostra, quanto maior o tamanho de amostra maior o poder para identificar uma diferença significativa e o incremento de poder é pouco para amostras maiores do que 70 indivíduos.

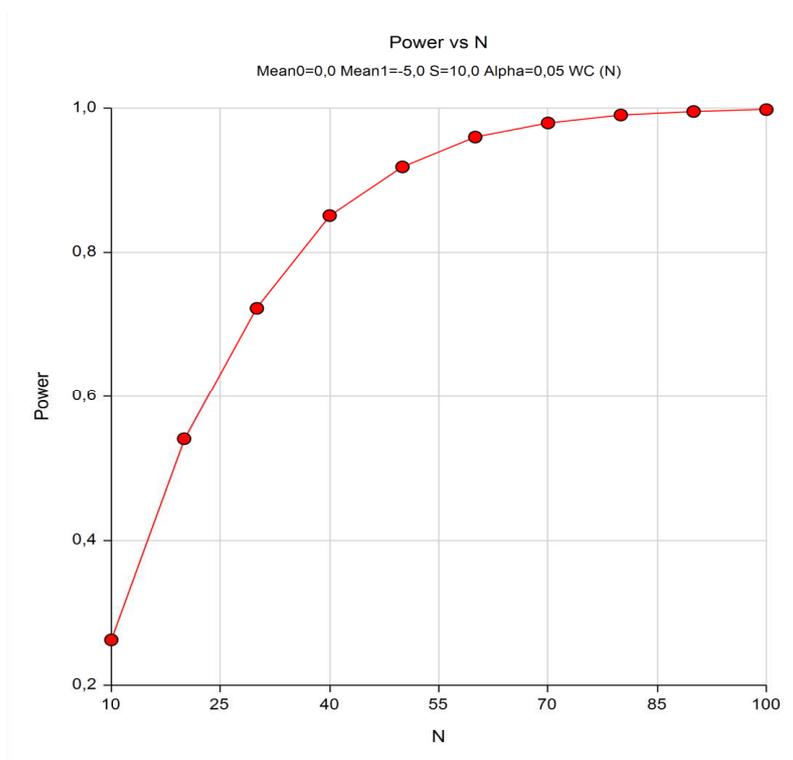


Figura 4 - Poder contra Tamanho de Amostra no Programa PASS 11 para o teste Wilcoxon.

GPower 3.1

O programa GPower 3.1 calculou um tamanho de amostra de 39 indivíduos para o exemplo utilizado o método ARE mínimo e 35 ao escolher a distribuição normal. Podemos ver na Figura 5 um estudo do poder versus o tamanho de amostra pelo método ARE mínimo, semelhante ao gerado pelo PASS, mas com os eixos invertidos e com outra escala.

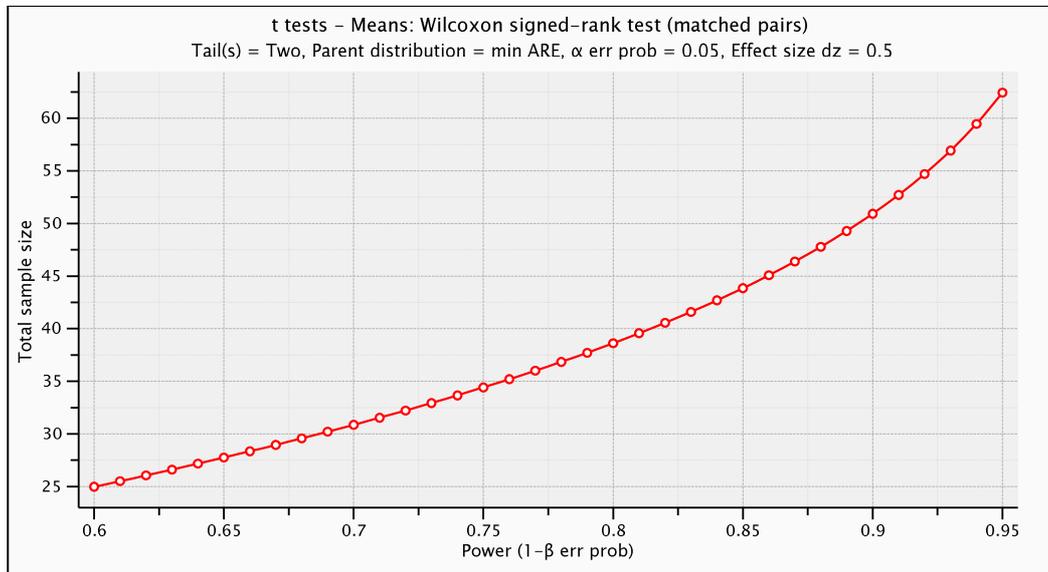


Figura 5 - Poder contra Tamanho de Amostra no Programa G*Power 3.1 pelo método ARE para o teste Wilcoxon.

Kruskal-Wallis

O teste Kruskal-Wallis envolve desenhos com duas ou mais amostras independentes e tem as seguintes suposições: cada amostra foi selecionada aleatoriamente da população que representa; as k amostras são independentes; a variável aleatória observada originalmente é contínua; a distribuição das populações das amostras deve ter mesmo formato.

O programa PASS possui uma rotina de simulação para o cálculo do tamanho de amostra e poder para o teste Kruskal-Wallis, que é a mesma utilizada para a ANOVA de um fator. Para cada cenário especificado são realizadas duas simulações, uma que estima o nível de erro do tipo I e a outra, o poder do teste.

Simplificando, o software necessita de quatro passos para realizar a simulação:

1) Especifica-se como o teste é realizado. Isto inclui se será considerado o teste F da ANOVA paramétrica ou o teste Kruskal-Wallis e o nível de significância especificado. Também deve-se escolher distribuições para a variável em cada grupo sob as hipóteses nula e

alternativa (as opções são Beta, Binomial, Cauchy, Constante, Exponencial, F, Gamma, Multinomial, Normal, Poisson, t de Student, Lambda de Tukey, Uniforme e Weibull). Define-se também o poder (para calcular tamanho de amostra), tamanho de amostra base para cada grupo (para o cálculo do poder), padrão do tamanho de amostra por grupo (para cálculo do poder) e número de grupos.

2) Então, amostras aleatórias são geradas da distribuição especificada pela hipótese alternativa. Calcula-se a estatística do teste a partir dos dados simulados e determina-se se a hipótese nula foi rejeitada ou não. Tabula-se o número de rejeições e utiliza-se este valor para calcular o poder do teste.

3) Após, gera-se amostras aleatórias da distribuição especificada pela hipótese nula. Calcula-se a estatística do teste a partir dos dados simulados e determina-se se a hipótese nula foi rejeitada ou não. Tabula-se o número de rejeições e utiliza este valor para calcular o nível de erro do tipo I do teste.

4) Repete-se os passos 2 e 3 um número grande de vezes, por definição no programa está especificado 2000 repetições.

O PASS tem dois métodos para gerar amostras aleatórias. O primeiro gera as amostras diretamente, uma a uma. O segundo método gera uma quantidade grande de valores aleatórios (mais de 10 mil) conforme a distribuição desejada e a partir destes valores armazenados são extraídos as amostras. Este segundo método pode cortar o tempo de simulação em até 70%.

Exemplo 3

Como exemplo extraído do manual do PASS 11 (16), digamos que pesquisadores estão planejando um experimento para comparação de medianas de quatro grupos utilizando o teste Kruskal-Wallis ao nível de significância de 5%. Estudos anteriores mostraram que o desvio padrão comum dos grupos é de 18. Acredita-se que as médias de cada grupo serão de 40, 10, 10 e 10, respectivamente. As distribuições associadas aos grupos sob a hipótese nula e a alternativa foram consideradas normais. Os pesquisadores estão interessados em estudar o poder do teste para tamanhos de amostra em cada grupo de 4, 8 e 12.

Na Tabela 2 temos o poder para os três tamanhos de amostra especificados, o tamanho de amostra por grupo, o tamanho de amostra total, o nível de significância alvo que é o nível de significância especificado, o nível de significância erro tipo I e o nível de erro tipo II, os dois últimos foram calculados na simulação.

Tabela 2 - Resultado da simulação do programa PASS 11 para o exemplo do teste Kruskal-Wallis.

Poder	Tamanho do Grupo	Tamanho Total da Amostra	Nível de Significância Alvo	Nível de Erro Tipo I	Nível de Erro Tipo II
0,392	4	16	0,05	0,035	0,608
0,850	8	32	0,05	0,041	0,150
0,980	12	48	0,05	0,052	0,020

Para um tamanho de amostra de 8 por grupo, 32 no total, já alcançamos um poder superior à 80% e com um tamanho de amostra por grupo de 12 o poder é superior à 98%. Foram realizadas 2000 simulações.

Considerações finais

Para o teste Mann-Whitney obtivemos uma quantidade maior de programas, pagos e livres. Também temos o método do banco de dados exemplo, um método bem simples que não está implementado em nenhum dos programas abordados e por isto foi feita uma implementação no R. Para este teste, as informações solicitadas são semelhantes entre os programas.

O SAS 9.3 possui soluções para cálculo de tamanho de amostra e poder para o teste Mann-Whitney e uma variedade grande de distribuições para ajuste, mas sua interface não é muito intuitiva e pode afastar usuários. No programa R foi encontrado apenas cálculo para tamanho de amostra no caso de variáveis categóricas ordinais, apesar dessas restrições utiliza um dos métodos que leva em consideração empates. Os programas PASS 11 e GPower 3.1 são muito semelhantes quanto ao método e só trabalham com variáveis contínuas.

O método do banco de dados exemplo é o único presente no trabalho em que não se precisa especificar uma distribuição para a variável, seja para a hipótese alternativa ou a nula. Como foi dito anteriormente a escolha de distribuição sob hipótese alternativa ou nula, que está presente na maioria dos métodos, é uma grande desvantagem e deve ser feita com muito

cuidado. Por outro lado, uma desvantagem do banco de dados exemplo é que nem sempre o valor da estatística qui-quadrado é divulgado em artigos, que às vezes mostram apenas o valor-p.

O programa GPower 3.1 tem a opção de utilizar o método ARE mínimo, tanto para Mann-Whitney quanto para Wilcoxon, que tem a vantagem de ser válido independente da distribuição da variável, mas é conservador.

O teste Kruskal-Wallis foi encontrado apenas no programa PASS 11 e os cálculos de tamanho de amostra e poder são realizados por simulação. Este método exige muito conhecimento prévio das distribuições envolvidas tanto sob a hipótese nula quanto a alternativa para cada grupo do estudo.

Durante a revisão bibliográfica foram encontrados dois artigos que propunham métodos de cálculo de tamanho de amostra e poder para o teste Kruskal-Wallis. O primeiro propunha um método que se baseado em estudo piloto, podendo este ser não-paramétrico ou semi-paramétrico, e sem a necessidade de escolher um distribuição para a população (17). O segundo artigo propunha um método utilizando técnicas de *bootstrapping* para produzir uma estimativa do poder baseada na distribuição cumulativa empírica dos dados amostrais (18). Esses métodos têm como vantagem não precisar descrever a distribuição dos dados sob a hipótese nula nem sobre a alternativa.

Referências

1. Sample Size Software | Power Analysis Software | Effect Size Calculation. [Online] [Citado em: 13 de Junho de 2012.] <http://www.statistical-solutions-software.com/products-page/nquery-advisor-sample-size-software/>.
2. PASS: Power Analysis and Sample Size Software. [Online] [Citado em: 13 de Junho de 2012.] <http://www.ncss.com/pass.html>.
3. SAS 9.3 Software | SAS. [Online] [Citado em: 13 de Junho de 2012.] <http://www.sas.com/software/sas9/>.
4. Heinrich-Heine-Universität - Institut für experimentelle Psychologie. [Online] [Citado em: 13 de Junho de 2012.] <http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3>.
5. The R Project for Statistical Computing. [Online] [Citado em: 13 de Junho de 2012.] <http://www.r-project.org/>.
6. **DJ, Sheskin.** *Handbook of Parametric and Nonparametric Statistical Procedures*. Boca Raton : CHAPMAN & HALUCRC, 2004.
7. **Divine G, Kapke A, Havstad S, Joseph CL.** Exemplary data set sample size calculation for Wilcoxon-Mann-Whitney. *Statistics in Medicine*. 1, 2010, Vol. 29.
8. **O'Brien RG, Casteloe JM.** Exploiting the link between the Wilcoxon–Mann–Whitney test and a simple odds statistic. *Proceedings of the Thirty-first Annual SAS Users Group International Conference*. SAS Institute Inc. : Cary, NC, 2006.
9. **Al-Sunduqchi, Mahdi S.** Determining the Appropriate Sample Size for Inferences Based on the Wilcoxon Statistics. *Dissertação de Ph.D sob orientação de William C. Guenther*. Dept. de Estatística, University of Wyoming, Laramie, Wyoming, 1990.
10. **GE., Noether.** Sample size determination for some common nonparametric tests. *Journal of the American Statistical Association*. 1987, Vol. 82.
11. **JE, Kolassa.** A comparison of size and power calculations for the Wilcoxon statistic for ordered categorical data. *Statistics in Medicine*. 1995, Vol. 14.
12. **Lehmann, EL.** *Nonparametrics: Statistical methods based on ranks*. San Francisco, CA : Holden-Day, 1975.
13. **Zhao YD, Rahardja D, Qu Y.** Sample size calculation for the Wilcoxon–Mann–Whitney test adjusting for ties. *Statistics in Medicine*. 2008, Vol. 27.
14. **Joseph CLM, Peterson E, Havstad S, Johnson CC, Hoerauf S, Stringer S, Gibson-Scipio W, Ownby DR, Elston-Lafata J, Pallonen U, Strecher V.** A web-based, tailored asthma management program for urban African–American high school students. *American Journal of Respiratory Critical Care Medicine*. 2007, Vol. 175.

15. Manual GPower - Means: Wilcoxon signed-rank test (matched pairs). [Online] [Citado em: 13 de Junho de 2012.] http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/user-guide-by-distribution/t/means_wilcoxon_signed_rank_test_matched_pairs.
16. **JL, Hintze.** *User's Guide II, Power Analysis and Sample Size System*. Kaysville, Utah : NCSS, 2011.
17. **Fan C, Zhang D, Zhang CH.** On Sample Size of the Kruskal–Wallis Test with Application to a Mouse Peritoneal Cavity Study. *Biometrics*. 2011, Vol. 67.
18. **Mahoney M, Magel R.** Estimation of the Power of the Kruskal-Wallis Test. *Biometrical Journal*. 1996, Vol. 38, 5.