Poder e Tamanho Amostral em alguns Testes de Hipótese Não Paramétricos

Lucas Lima Ayres

Orientador: Prof. Thiago Rezende dos Santos

Universidade Federal de Minas Gerais

24 de novembro de 2017

Introdução

- O problema da indução
- Testes de hipótese: a abordagem de Neyman-Pearson
- Vantagens dos testes não paramétricos
- Dimensionamento amostral não paramétrico: escassez de informações ao público

1. Teste de aderência do qui-quadrado (Pearson, 1900)
2. Teste dos postos sinalizados de Wilcoxon (Wilcoxon, 1945)
3. Teste da soma dos postos de Wilcoxon (Wilcoxon, 1945)
4. Teste de Kruskal-Wallis (Kruskal e Wallis, 1952)

Teste de Aderência do Qui-Quadrado

- Considere-se uma dist. multinomial $(n, \mathbf{p})$.
- Testa-se $H_0 : \mathbf{p} = \mathbf{p_0}$ contra $H_1 : \mathbf{p} \neq \mathbf{p_0}$.
- Estatística-teste: $Q = \sum_{i=1}^{k} \frac{(O_i - np_{0i})^2}{np_{0i}}$.
- Sob H_0, $Q_n \xrightarrow{d} \chi^2_{k-1}$.
  Sob H_1, $Q_n \xrightarrow{d} \chi'^2_{k-1,\lambda}$, onde $\lambda = n \sum_{i=1}^{k} \frac{(p_{1i} - p_{0i})^2}{p_{0i}}$.
- Região crítica: $\left[\chi^2_{1-\alpha,k-1}, +\infty\right)$.
- Poder: $P_{H_1}\left(Q \geq \chi^2_{1-\alpha,k-1}\right)$.
- Encontrar λ tal que $\chi^2_{1-\alpha,k-1} = \chi'^2_{\beta,k-1,\lambda}$.
  (Calcular a raiz da eq. $\chi^2_{1-\alpha,k-1} - \chi'^2_{\beta,k-1,\lambda} = 0$.)
- Conhecido o valor de λ, determina-se o tamanho amostral:
  $n = \lambda \left[\sum_{i=1}^{k} \frac{(p_{1i} - p_{0i})^2}{p_{0i}}\right]^{-1}$.

Teste de Aderência do Qui-Quadrado

Exemplo (fábrica de dados)

Para aferir se há algum desvio relevante da hipótese de equiprobabilidade das faces, quantos dados precisam ser lançados? A hipótese nula será testada através do teste do qui-quadrado para $\alpha = 0{,}01$ e $\beta = 0{,}05$. O tamanho do efeito igual a 0,1 representa o desvio máximo tolerável.

Teste de Aderência do Qui-Quadrado

No R:

```
> chisq.test.pss(effectsize = 0.1, df = 5,
sig.level = 0.01, power = 0.95)

 Sample Size for Pearson's Chi-Squared Test

  Sample size: 2577 ( 2576.206 )
  Significance level: 0.01
  Power: 0.95
  Effect size: 0.1
  Degrees of freedom: 5
```

Teste de Aderência do Qui-Quadrado

Com o programa PASS:

Chi-Square Tests

Numeric Results for Chi-Square Test

Power	N	W	Chi-Square	DF	Alpha	Beta
0,95008	2577	0,1000	25,7700	5	0,01000	0,04992

References

Cohen, Jacob. 1988. Statistical Power Analysis for the Behavioral Sciences, Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the size of the sample drawn from the population. To conserve resources, it should be small.
W is the effect size--a measure of the magnitude of the Chi-Square that is to be detected.
DF is the degrees of freedom of the Chi-Square distribution.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.

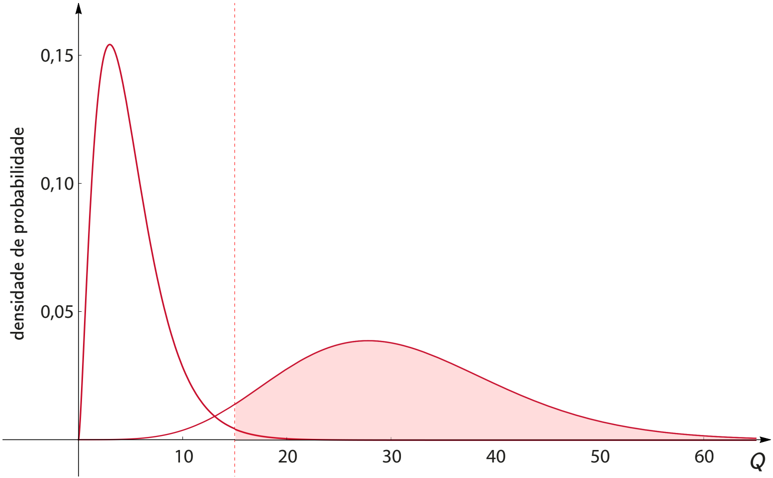
Summary Statements

A sample size of 2577 achieves 95% power to detect an effect size (W) of 0,1000 using a 5 degrees of freedom Chi-Square Test with a significance level (alpha) of 0,01000.

Design Tab

Solve For:	Sample Size
DF (Degrees of Freedom):	5
Power:	0,95
Alpha:	0,01
W (Effect Size):	0,1

Teste de Aderência do Qui-Quadrado



Teste dos Postos Sinalizados de Wilcoxon

- Seja X uma v.a. contínua cuja dist. é simétrica em torno de uma constante μ.
- Testa-se $\mu = 0$ contra $\mu > 0$.
- Estatística-teste: $T^+ = \sum_{i=1}^n R_i \psi_i$, onde

  R_i é o posto de $|X_i|$, $\psi_i = \begin{cases} 1, & \text{se } X_i > 0, \\ 0, & \text{c.c.} \end{cases}$
- Sob H_0, $\frac{T_n^+ - E_{H_0}(T_n^+)}{\sqrt{\mathrm{Var}_{H_0}(T_n^+)}} \xrightarrow{d} N(0,1)$.

  Sob H_1, $\frac{T_n^+ - E_{H_1}(T_n^+)}{\sqrt{\mathrm{Var}_{H_1}(T_n^+)}} \xrightarrow{d} N(0,1)$.
- Região crítica: $\left[\frac{n(n+1)}{4} + z_{1-\alpha}\sqrt{\frac{n(n+1)(2n+1)}{24}}, +\infty\right)$.
- Poder: $P_{H_1}\left(\frac{T^+ - E_{H_1}(T^+)}{\sqrt{\mathrm{Var}_{H_1}(T^+)}} \geq \frac{\frac{n(n+1)}{4} + z_{1-\alpha}\sqrt{\frac{n(n+1)(2n+1)}{24}} - E_{H_1}(T^+)}{\sqrt{\mathrm{Var}_{H_1}(T^+)}}\right)$.

Teste dos Postos Sinalizados de Wilcoxon

- Achar n tal que $\dfrac{\frac{n(n+1)}{4}+z_{1-\alpha}\sqrt{\frac{n(n+1)(2n+1)}{24}}-E_{H_1}(T^+)}{\sqrt{\mathrm{Var}_{H_1}(T^+)}}=z_\beta.$
- $E(T^+)=n\left[p_1+\frac{(n-1)}{2}p_2\right],$
  $\mathrm{Var}(T^+)=n\left\{p_1(1-p_1)+(n-1)\left[(p_1-p_2)^2+\frac{3}{2}\,p_2(1-p_2)+(n-2)\left(p_3-p_2^2\right)\right]\right\},$
  onde $p_1=P(X_i>0)$, $p_2=P\left(X_i+X_j>0\right)$, $p_3=P\left(X_i+X_j>0\cap X_i+X_k>0\right)$.
- Calcular a raiz da equação

$$\left\{\frac{n\,(n+1)}{4}+z_{1-\alpha}\sqrt{\frac{n\,(n+1)\,(2n+1)}{24}}-n\left[p_1+\frac{(n-1)}{2}p_2\right]\right.$$
$$\left.-z_\beta\sqrt{np_1\,(1-p_1)+\frac{n\,(n-1)}{2}\left[2\,(p_1-p_2)^2+3\,p_2\,(1-p_2)\right]+n\,(n-1)\,(n-2)\,\left(p_3-p_2^2\right)}\right\}=0.$$

Teste dos Postos Sinalizados de Wilcoxon

Exemplo (dist. uniforme)

Determinar o tamanho da amostra para testar $H_0 : \mu = 0$ contra $H_1 : \mu \neq 0$ com $\alpha = 0{,}1$, $\beta = 0{,}2$, quando o mecanismo gerador das observações é a v.a. uniforme $(-0{,}3; 0{,}7)$.

Teste dos Postos Sinalizados de Wilcoxon

$$X \sim \text{uniforme}(-0{,}3; 0{,}7)$$

$$p_1 = P\left(X > 0\right) = \int_0^{0,7} dx = 0{,}7$$

$$p_2 = P\left(X_1 + X_2 > 0\right) = \int_{-0,3}^{0,3}\int_{-x_1}^{0,7} dx_2 dx_1 + \int_{0,3}^{0,7}\int_{-0,3}^{0,7} dx_2 dx_1 = 0{,}82$$

$$\begin{aligned} p_3 &= P\left(X_1 + X_2 > 0 \cap X_1 + X_3 > 0\right) \\ &= \int_{0,3}^{0,7}\int_{-0,3}^{0,7}\int_{-0,3}^{0,7} dx_3 dx_2 dx_1 + \int_{-0,3}^{0,3}\int_{-x_1}^{0,7}\int_{-x_1}^{0,7} dx_3 dx_2 dx_1 = 0{,}712 \end{aligned}$$

Teste dos Postos Sinalizados de Wilcoxon

No R:

```
> wilcox.test.pss(p1 = 0.7, p2 = 0.82, p3 = 0.712,
sig.level = 0.1, alternative = "two.sided", power = 0.8)

  Sample Size for Wilcoxon's Signed Rank Test

   Sample size: 18 ( 17.38723 )
   Significance level: 0.1
   Power: 0.8
   p1 = 0.7
   p2 = 0.82
   p3 = 0.712
```

Teste dos Postos Sinalizados de Wilcoxon

Com o programa PASS:

Tests for One Mean (Simulation)

Numeric Results for Testing One Mean = Mean0. Hypotheses: H0: Mean1 = Mean0; H1: Mean1 ≠ Mean0
H0 Distribution: UniformMS(0 0,28867513459482)
H1 Distribution: UniformMS(0,2 0,28867513459482)
Test Statistic: Wilcoxon Signed-Rank Test

Power	N	H0 Mean0	H1 Mean1	Target Alpha	Actual Alpha	Beta
0,81948	18	0,0	0,2	0,10000	0,09879	0,18052

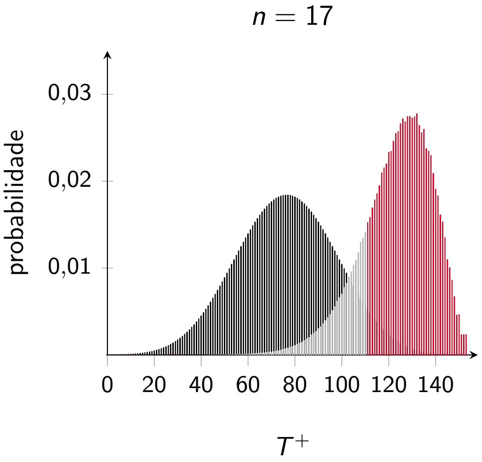
Notes
Simulations: 1000000. Run Time: 35,85 minutes.

References

Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
Devroye, Luc. 1986. Non-Uniform Random Variate Generation. Springer-Verlag. New York.
Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, MA.

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N is the size of the sample drawn from the population.
Mean0 is the value of the mean assuming the null hypothesis. This is the value being tested.
Mean1 is the actual value of the mean. The procedure tests whether Mean0 = Mean1.
Target Alpha is the probability of rejecting a true null hypothesis. It is set by the user.
Actual Alpha is the alpha level that was actually achieved by the experiment.
Beta is the probability of accepting a false null hypothesis.

Teste dos Postos Sinalizados de Wilcoxon



$1 - \beta = 0{,}812474852925232$

Teste da Soma dos Postos de Wilcoxon

- Sejam X e Y v.as contínuas e independentes tais que $Y \stackrel{d}{=} X + \Delta$.
- Testa-se $H_0 : \Delta = 0$ contra $H_1 : \Delta > 0$.
- Estatística-teste: $T = \sum_{i=1}^{n} R_i$, onde R_i é o posto de y_i.
- Sob H_0, $\frac{T_{\min(m,n)} - E_{H_0}(T_{\min(m,n)})}{\sqrt{\mathrm{Var}_{H_0}(T_{\min(m,n)})}} \xrightarrow{d} N(0,1)$.

  Sob H_1, $\frac{T_{\min(m,n)} - E_{H_1}(T_{\min(m,n)})}{\sqrt{\mathrm{Var}_{H_1}(T_{\min(m,n)})}} \xrightarrow{d} N(0,1)$.
- Região crítica: $\left[\frac{n(m+n+1)}{2} + z_{1-\alpha}\sqrt{\frac{mn(m+n+1)}{12}}, +\infty\right)$.
- Poder: $P_{H_1}\left(\frac{T - E_{H_1}(T)}{\sqrt{\mathrm{Var}(T)}} \geq \frac{\frac{n(m+n+1)}{2} + z_{1-\alpha}\sqrt{\frac{mn(m+n+1)}{12}} - E_{H_1}(T)}{\sqrt{\mathrm{Var}_{H_1}(T)}}\right)$.

Teste da Soma dos Postos de Wilcoxon

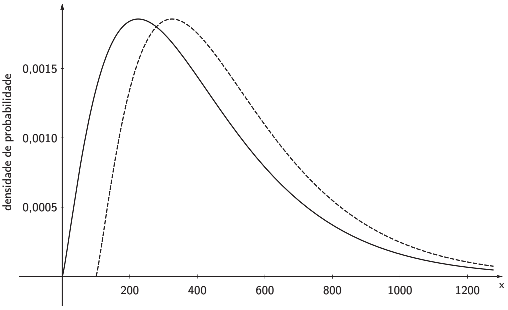
- Achar n tal que $\dfrac{\frac{n(m+n+1)}{2}+z_{1-\alpha}\sqrt{\frac{mn(m+n+1)}{12}}-E_{H_1}(T)}{\sqrt{\mathrm{Var}_{H_1}(T)}}=z_\beta$.
- $E(T)=mnp_1+\frac{n(n+1)}{2}$,
  $\mathrm{Var}(T)=mn\left[p_1(1-p_1)+(n-1)\left(p_2-p_1^2\right)+(n-1)\left(p_3-p_1^2\right)\right]$,
  onde $p_1=P\left(X_i<Y_j\right)$, $p_2=P\left(X_i<Y_j\cap X_i<Y_k\right)$, $p_3=P\left(X_i<Y_j\cap X_k<Y_j\right)$.
- Calcular a raiz da equação

$$\left\{\frac{n\left(an+n+1\right)}{2}+z_{1-\alpha}\sqrt{\frac{an^2\left(an+n+1\right)}{12}}-an^2p_1-\frac{n\left(n+1\right)}{2}\right.$$
$$\left.-z_\beta\sqrt{an^2\left[p_1\left(1-p_1\right)+\left(n-1\right)\left(p_2-p_1^2\right)+\left(an-1\right)\left(p_3-p_1^2\right)\right]}\right\}=0.$$

Teste da Soma dos Postos de Wilcoxon

Exemplo (dist. gama)

Considerem-se $X \sim \Gamma(9, 25; 180)$ e $Y \stackrel{d}{=} X + 100$. Determinar o tamanho da amostra para testar $H_0 : \Delta = 0$ contra $H_1 : \Delta > 0$ com $\alpha = 0{,}05$ e $\beta = 0{,}1$.



Teste da Soma dos Postos de Wilcoxon

$$X \sim \Gamma(2{,}25; 180)$$
$$Y \overset{d}{=} X + 100$$

$$p_1 = P\left(X < Y\right) = \int_{100}^{\infty}\int_{0}^{y} f_{XY}(x, y)dxdy \approx 0{,}623$$

$$W = \min\left(Y_1, Y_2\right)$$

$$p_2 = P\left(X_1 < Y_1 \cap X_1 < Y_2\right) = P\left(X < W\right) = \int_{100}^{\infty}\int_{0}^{w} f_{XW}(x, w)dxdw$$
$$\approx 0{,}485$$

$$Z = \max\left(X_1, X_2\right)$$

$$p_3 = P\left(X_i < Y_j \cap X_k < Y_j\right) = P\left(Z < Y\right) = \int_{100}^{\infty}\int_{0}^{y} f_{ZY}(z, y)dzdy$$
$$\approx 0{,}447$$

Teste da Soma dos Postos de Wilcoxon

No R:

```
> wilcox2.test.pss(p1 = 0.623, p2 = 0.485, p3 = 0.447,
sig.level = 0.05, alternative = "greater", power = 0.9,
a = 1)

 Sample Size for Wilcoxon's Rank-Sum Test

  Sample size (group 1): 93 ( 92.10933 )
  Sample size (group 2): 93 ( 92.10933 )
  Significance level: 0.05
  Power: 0.9
  p1 = 0.623
  p2 = 0.485
  p3 = 0.447
```

Teste da Soma dos Postos de Wilcoxon

Com o programa PASS:

Mann-Whitney-Wilcoxon Tests (Simulation)

Numeric Results for Testing Mean Difference = Diff0.   Hypotheses: H0: Diff1 = Diff0; H1: Diff1 < Diff0
H0 Dist's: GammaMS(405 270) & GammaMS(405 270)
H1 Dist's: GammaMS(405 270) & GammaMS(405 270) + 100
Test Statistic: Mann-Whitney-Wilcoxon Test

Power	N1/N2	H0 Diff0	H1 Diff1	Target Alpha	Actual Alpha	Beta
0,901	92/92	0,0	-100,0	0,050	0,050	0,099

Notes
Pool Size: 2000000. Simulations: 1000000. Run Time: 1,64 hours.

References

Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
Devroye, Luc. 1986. Non-Uniform Random Variate Generation. Springer-Verlag. New York.
Matsumoto, M. and Nishimura,T. 1998. 'Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator.' ACM Trans. On Modeling and Computer Simulations.
Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N1 is the size of the sample drawn from population 1.
N2 is the size of the sample drawn from population 2.
Diff0 is the mean difference between (Grp1 - Grp2) assuming the null hypothesis, H0.
Diff1 is the mean difference between (Grp1 - Grp2) assuming the alternative hypothesis, H1.
Target Alpha is the probability of rejecting a true null hypothesis. It is set by the user.

Teste de Kruskal-Wallis

- Seja X uma v.a. contínua e $X_i \stackrel{d}{=} X + \Delta_i$ uma translação dela.
- Testa-se $H_0 : \Delta_1 = \Delta_2 = \cdots = \Delta_k$ contra $H_1 : \neg H_0$.
- Estatística-teste: $H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{\left(\sum_{j=1}^{n_i} R_{ij}\right)^2}{n_i} - 3\,(n+1)$, onde R_{ij} é o posto da observação x_{ij}.
- Sob H_0, $H_{\min(n_1,\ldots,n_k)} \xrightarrow{d} \chi^2_{k-1}$.

  Sob H_1, $H_{\min(n_1,\ldots,n_k)} \xrightarrow{d} \chi'^2_{k-1,\lambda}$, onde $\lambda = 12n\left[\int_{-\infty}^{\infty} f_X(x)^2 dx\right]^2 \sum_{i=1}^{k} \frac{n_i}{n}\left(\Delta_i - \bar{\Delta}\right)^2$.
- Região crítica: $\left[\chi^2_{1-\alpha,k-1}, +\infty\right)$.
- Poder: $P_{H_1}\left(H \geq \chi^2_{1-\alpha,k-1}\right)$.
- Encontrar λ tal que $\chi^2_{1-\alpha,k-1} = \chi'^2_{\beta,k-1,\lambda}$.

  (Calcular a raiz da eq. $\chi^2_{1-\alpha,k-1} - \chi'^2_{\beta,k-1,\lambda} = 0$.)
- Conhecido o valor de λ, determina-se o tamanho amostral: $n = \lambda\left[12\int_{-\infty}^{\infty} f_X(x)^2 dx \sum_{i=1}^{k} a_i \left(\Delta_i - \bar{\Delta}\right)^2\right]^{-1}$.

Teste de Kruskal-Wallis

Exemplo (conscrição lotérica)

A tabela a seguir mostra o resultado de um sorteio realizado para definir a ordem de convocação de homens para uma guerra. Aplicando-se o teste de Kruskal-Wallis aos meses do ano, a hipótese nula é rejeitada ao nível de significância de 0,01. Fazer o gráfico do poder do teste como função do parâmetro λ.

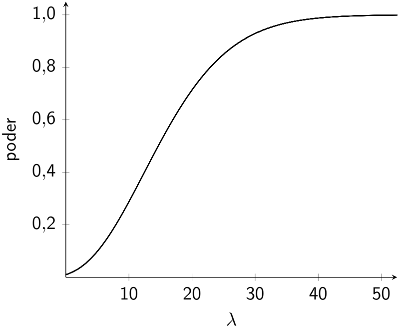
Teste de Kruskal-Wallis

dia	jan.	fev.	mar.	abr.	maio	jun.	jul.	ago.	set.	out.	nov.	dez.
1	305	086	108	032	330	249	093	111	225	359	019	129
2	159	144	029	271	298	228	350	045	161	125	034	328
3	251	297	267	083	040	301	115	261	049	244	348	157
4	215	210	275	081	276	020	279	145	232	202	266	165
5	101	214	293	269	364	028	188	054	082	024	310	056
6	224	347	139	253	155	110	327	114	006	087	076	010
7	306	091	122	147	035	085	050	168	008	234	051	012
8	199	181	213	312	321	366	013	048	184	283	097	105
9	194	338	317	219	197	335	277	106	263	342	080	043
10	325	216	323	218	065	206	284	021	071	220	282	041
11	329	150	136	014	037	134	248	324	158	237	046	039
12	221	068	300	346	133	272	015	142	242	072	066	314
13	318	152	259	124	295	069	042	307	175	138	126	163
14	238	004	354	231	178	356	331	198	001	294	127	026
15	017	089	169	273	130	180	322	102	113	171	131	320
16	121	212	166	148	055	274	120	044	207	254	107	096
17	235	189	033	260	112	073	098	154	255	288	143	304
18	140	292	332	090	278	341	190	141	246	005	146	128
19	058	025	200	336	075	104	227	311	177	241	203	240
20	280	302	239	345	183	360	187	344	063	192	185	135
21	186	363	334	062	250	060	027	291	204	243	156	070
22	337	290	265	316	326	247	153	339	160	117	009	053
23	118	007	256	252	319	109	172	116	119	201	182	162
24	059	236	258	002	031	358	023	036	195	196	230	095
25	052	179	343	351	361	137	067	286	149	176	132	084
26	092	365	170	340	357	022	303	245	018	007	309	173
27	355	205	268	074	296	064	289	352	233	264	047	078
28	077	299	223	262	308	222	088	167	257	094	281	123
29	349	285	362	191	226	353	270	061	151	229	099	016
30	164		217	208	103	209	287	333	315	038	174	003
31	211		030		313		193	011		079		100

Teste de Kruskal-Wallis

No R:

```
> lambda <- seq(0, 50, length = 5000)
> poder <- pchisq(qchisq(1 - 0.01, df = 11), df = 11,
ncp = lambda, lower.tail = FALSE)
> plot(lambda, poder, cex = 0.01)
```



Conclusão

- Os métodos apresentados são computacionalmente menos exigentes que os de simulação.
- Alguns elementos do *input* podem ser difíceis para o utilizador.
- Em um dos exemplos, a função `wilcox.test.pss` mostrou-se adequada mesmo para uma amostra pequena.
- A fórmula de Noether (1987) produziu resultados diferentes nos exemplos relacionados aos testes de Wilcoxon.
- O método da eficiência relativa assintótica subestimou o tamanho da amostra em um dos exemplos.
- Sugestões de investigação:
  - testes de Friedman, de Cochran, de Jonckheere-Terpstra;
  - admitir a possibilidade de empates;
  - tamanho do efeito para o Teste de Kruskal-Wallis.

Referências

AL-SUNDUQCHI, M.S. *Determining the appropriate sample size for inferences based on the Wilcoxon statistics*. Thesis (Ph.D.) – University of Wyoming, Laramie, 1990.

CHOW, S. *et al. Sample size calculations in clinical research*. Boca Raton: Taylor & Francis, 2017.

COHEN, J. *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum, 1988.

FAN, C.; ZHANG, D.; ZHANG, C. On sample size of the Kruskal–Wallis test with application to a mouse peritoneal cavity study. *Biometrics*, v. 67, p. 213–224, 2011.

FIENBERG, S.E. Randomization and social affairs: the 1970 draft lottery. *Science*, v. 171, p. 255–261, 1971.

HETTMANSPERGER, T.P. *Statistical inference based on ranks*. Nova York: Wiley, 1984.

KRUSKAL, W.H.; WALLIS, W.A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.*, v. 47, p. 583–621, 1952.

Referências

NOETHER, G.E. Sample size determination for some common nonparametric tests. *J. Am. Stat. Assoc.*, v. 82, p. 645–647, 1987.

PASS 15 Power Analysis and Sample Size Software. Kaysville: NCSS, LLC, 2017. <www.ncss.com/software/pass/>.

PEARSON, K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag. Ser. 5*, v. 50, p. 157–175, 1900.

R CORE TEAM. R: A language and environment for statistical computing. Viena: R Foundation for Statistical Computing, 2017. <https://www.r-project.org/>.

VAN DE WIEL, M.A. Exact non-null distributions of rank statistics. *Commun. Stat. Simulat. Comput.*, v. 30, p. 1011–1029, 2001.

WILCOXON, F. Individual comparisons by ranking methods. *Biometrics Bull.*, v. 1, p. 80–83, 1945.