Power and Sample Size in some Nonparametric Hypothesis Tests

Lucas Ayres

Federal University of Minas Gerais (Bachelor's thesis)

EMABG Welcome Course
August 17, 2020

Introduction

- The problem of induction
- Hypothesis testing: Neyman-Pearson's approach
- Advantages of nonparametric tests
- Sample size for nonparametric tests: scarcity of information to lay users

1. Chi-squared goodness-of-fit test (Pearson, 1900)
2. Wilcoxon signed-rank test (Wilcoxon, 1945)
3. Wilcoxon rank-sum test (Wilcoxon, 1945)
4. Kruskal-Wallis test (Kruskal and Wallis, 1952)

Chi-Squared Goodness-of-Fit Test

- Consider a multinomial dist. (n, p).
- Test $H_0 : \mathrm{p} = \mathrm{p}_0$ versus $H_1 : \mathrm{p} \neq \mathrm{p}_0$.
- Test statistic: $Q = \sum_{i=1}^{k} \frac{(O_i - np_{0i})^2}{np_{0i}}$.
- Under H_0, $Q_n \xrightarrow{d} \chi^2_{k-1}$.
  Under H_1, $Q_n \xrightarrow{d} \chi'^2_{k-1,\lambda}$, where $\lambda = n \sum_{i=1}^{k} \frac{(p_{1i} - p_{0i})^2}{p_{0i}}$.
- Critical region: $\left[\chi^2_{1-\alpha,k-1}, +\infty\right)$.
- Power: $P_{H_1}\left(Q \geq \chi^2_{1-\alpha,k-1}\right)$.
- Find λ such that $\chi^2_{1-\alpha,k-1} = \chi'^2_{\beta,k-1,\lambda}$.
  (Find root of eqn $\chi^2_{1-\alpha,k-1} - \chi'^2_{\beta,k-1,\lambda} = 0$.)
- Once we know λ, determine the sample size:
  $n = \lambda \left[\sum_{i=1}^{k} \frac{(p_{1i} - p_{0i})^2}{p_{0i}}\right]^{-1}$.

Chi-Squared Goodness-of-Fit Test

Example (dice manufacturer)

In order to assess whether there is a relevant departure from the hypothesis that their faces are equiprobable, how many dice should be rolled?
The null hypothesis will be tested using a chi-squared test with $\alpha = 0.01$ and $\beta = 0.05$. The effect size 0.1 represents the maximum amount of deviation tolerated by the consumer.

Chi-Squared Goodness-of-Fit Test

In 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
```

Chi-Squared Goodness-of-Fit Test

Using 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.

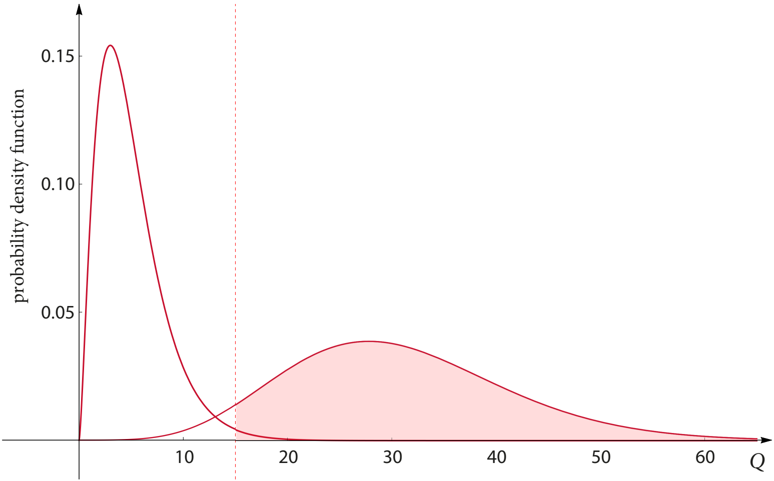
Summary Statements

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

Chi-Squared Goodness-of-Fit Test



Wilcoxon Signed-Rank Test

- Let X be a continuous r.v. whose dist. is symmetric around a constant μ.
- Test $\mu = 0$ versus $\mu > 0$.
- Test statistic: $T^+ = \sum_{i=1}^n R_i \psi_i$, where

  R_i is the rank of $|X_i|$, $\psi_i = \begin{cases} 1, & \text{if } X_i > 0, \\ 0, & \text{otherwise.} \end{cases}$
- Under H_0, $\frac{T_n^+ - E_{H_0}(T_n^+)}{\sqrt{\mathrm{Var}_{H_0}(T_n^+)}} \xrightarrow{d} N(0,1)$.

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

Wilcoxon Signed-Rank Test

- Find n such that $\dfrac{\frac{n(n+1)}{4}+z_{1-\alpha}\sqrt{\frac{n(n+1)(2n+1)}{24}}-E_{H_1}(T^+)}{\sqrt{\mathrm{Var}_{H_1}(T^+)}}=z_\beta$.
- $E(T^+)=n\left[p_1+\frac{(n-1)}{2}p_2\right]$,
  $\mathrm{Var}(T^+)=n\left\{p_1(1-p_1)+(n-1)\left[(p_1-p_2)^2+\frac{3}{2}\,p_2(1-p_2)+(n-2)\left(p_3-p_2^2\right)\right]\right\}$,
  where $p_1=P(X_i>0)$, $p_2=P\left(X_i+X_j>0\right)$, $p_3=P\left(X_i+X_j>0\cap X_i+X_k>0\right)$.
- Find the root of the equation

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

Wilcoxon Signed-Rank Test

Example (uniform dist.)

Determine the sample size needed to test $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ with $\alpha = 0.1$, $\beta = 0.2$, when the data-generating mechanism is the uniform $(-0.3, 0.7)$ r.v.

Wilcoxon Signed-Rank Test

$$X \sim \text{uniform}(-0.3, 0.7)$$

$$p_1 = P\left(X > 0\right) = \int_0^{0.7} dx = 0.7$$

$$p_2 = P\left(X_1 + X_2 > 0\right) = \int_{-0.3}^{0.3} \int_{-x_1}^{0.7} dx_2 dx_1 + \int_{0.3}^{0.7} \int_{-0.3}^{0.7} dx_2 dx_1 = 0.82$$

$$\begin{aligned} p_3 &= P\left(X_1 + X_2 > 0 \cap X_1 + X_3 > 0\right) \\ &= \int_{0.3}^{0.7} \int_{-0.3}^{0.7} \int_{-0.3}^{0.7} dx_3 dx_2 dx_1 + \int_{-0.3}^{0.3} \int_{-x_1}^{0.7} \int_{-x_1}^{0.7} dx_3 dx_2 dx_1 = 0.712 \end{aligned}$$

Wilcoxon Signed-Rank Test

In 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
```

Wilcoxon Signed-Rank Test

Using 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.

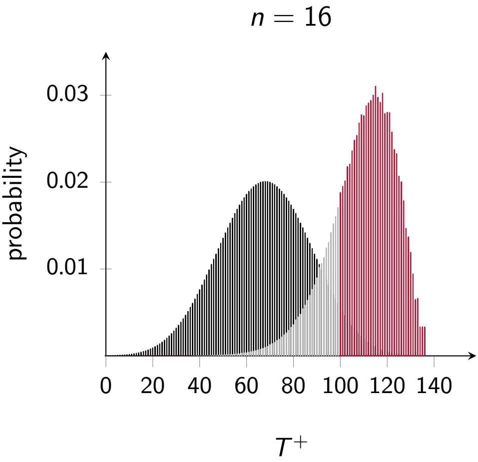
References

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.

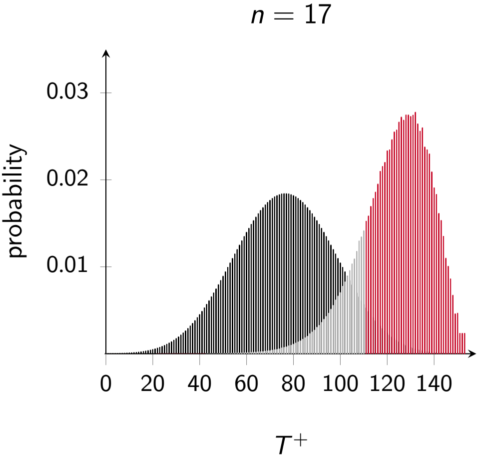
Summary Statements

Wilcoxon Signed-Rank Test



$1 - \beta = 0.7835736630247174$

Wilcoxon Signed-Rank Test



$1 - \beta = 0.812474852925232$

Wilcoxon Rank-Sum Test

- Let X and Y be continuous and independent r.vs such that $Y \overset{d}{=} X + \Delta$.
- Test $H_0 : \Delta = 0$ versus $H_1 : \Delta > 0$.
- Test statistic: $T = \sum_{i=1}^{n} R_i$, where R_i is the rank of y_i.
- Under H_0, $\frac{T_{\min(m,n)} - E_{H_0}(T_{\min(m,n)})}{\sqrt{\mathrm{Var}_{H_0}(T_{\min(m,n)})}} \xrightarrow{d} N(0,1)$.

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

Wilcoxon Rank-Sum Test

- Let X and Y be continuous and independent r.vs such that $Y \overset{d}{=} X + \Delta$.
- Test $H_0 : \Delta = 0$ versus $H_1 : \Delta > 0$.
- Test statistic: $T = \sum_{i=1}^{n} R_i$, where R_i is the rank of y_i.
- Under H_0, $\frac{T_{\min(m,n)} - E_{H_0}(T_{\min(m,n)})}{\sqrt{\mathrm{Var}_{H_0}(T_{\min(m,n)})}} \xrightarrow{d} N(0,1)$.

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

Wilcoxon Rank-Sum Test

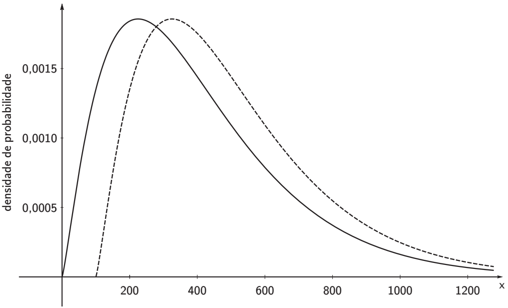
- Find n such that $\dfrac{\frac{n(m+n+1)}{2}+z_{1-\alpha}\sqrt{\frac{mn(m+n+1)}{12}}-E_{H_1}(T)}{\sqrt{\mathrm{Var}_{H_1}(T)}}=z_\beta$.
- $E(T)=mnp_1+\frac{n(n+1)}{2}$,
  $\mathrm{Var}(T)=mn\left[p_1(1-p_1)+(n-1)\left(p_2-p_1^2\right)+(n-1)\left(p_3-p_1^2\right)\right]$,
  where $p_1=P\left(X_i<Y_j\right)$, $p_2=P\left(X_i<Y_j\cap X_i<Y_k\right)$, $p_3=P\left(X_i<Y_j\cap X_k<Y_j\right)$.
- Find the root of the equation

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

Wilcoxon Rank-Sum Test

Example (gamma dist.)

Consider $X \sim \Gamma(2.25; 180)$ and $Y \overset{d}{=} X + 100$. Determine the sample size needed to test $H_0 : \Delta = 0$ against $H_1 : \Delta > 0$ with $\alpha = 0.05$ and $\beta = 0.1$.



Wilcoxon Rank-Sum Test

$$X \sim \Gamma(2.25; 180)$$
$$Y \overset{d}{=} X + 100$$

$$p_1 = P\left(X < Y\right) = \int_{100}^{\infty} \int_{0}^{y} f_{XY}(x, y) dxdy \approx 0.623$$
$$W = \min\left(Y_1, Y_2\right)$$
$$p_2 = P\left(X_1 < Y_1 \cap X_1 < Y_2\right) = P\left(X < W\right) = \int_{100}^{\infty} \int_{0}^{w} f_{XW}(x, w) dxdw$$
$$\approx 0.485$$
$$Z = \max\left(X_1, X_2\right)$$
$$p_3 = P\left(X_i < Y_j \cap X_k < Y_j\right) = P\left(Z < Y\right) = \int_{100}^{\infty} \int_{0}^{y} f_{ZY}(z, y) dzdy$$
$$\approx 0.447$$

Wilcoxon Rank-Sum Test

In 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
```

Wilcoxon Rank-Sum Test

Using 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.

Kruskal-Wallis Test

- Let X be a continuous r.v. and $X_i \overset{d}{=} X + \Delta_i$ a shift.
- Test $H_0 : \Delta_1 = \Delta_2 = \cdots = \Delta_k$ versus $H_1 : \neg H_0$.
- Test statistic: $H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{\left(\sum_{j=1}^{n_i} R_{ij}\right)^2}{n_i} - 3\,(n+1)$, where R_{ij} is the rank of observation x_{ij}.
- Under H_0, $H_{\min(n_1,\dots,n_k)} \xrightarrow{d} \chi^2_{k-1}$.
  Under H_1, $H_{\min(n_1,\dots,n_k)} \xrightarrow{d} \chi'^2_{k-1,\lambda}$, where
  $\lambda = 12n\left[\int_{-\infty}^{\infty} f_X(x)^2 dx\right]^2 \sum_{i=1}^{k} \frac{n_i}{n}\left(\Delta_i - \bar{\Delta}\right)^2$.
- Critical region: $\left[\chi^2_{1-\alpha,k-1}, +\infty\right)$.
- Power: $P_{H_1}\left(H \geq \chi^2_{1-\alpha,k-1}\right)$.
- Find λ such that $\chi^2_{1-\alpha,k-1} = \chi'^2_{\beta,k-1,\lambda}$.
  (Find root of eqn $\chi^2_{1-\alpha,k-1} - \chi'^2_{\beta,k-1,\lambda} = 0$.)
- Once we know λ, determine the sample size:
  $n = \lambda\left[12\int_{-\infty}^{\infty} f_X(x)^2 dx \sum_{i=1}^{k} a_i \left(\Delta_i - \bar{\Delta}\right)^2\right]^{-1}$.

Kruskal-Wallis Test

Example (draft lottery)

The following table shows the result of of a lottery to decide the order of call to military service for Americans in the Vietnam war. By performing the Kruskal-Wallis test on months of the year, the null hypothesis is rejected at a significance level of 0.01.
Plot the test's power as a function of the parameter λ.

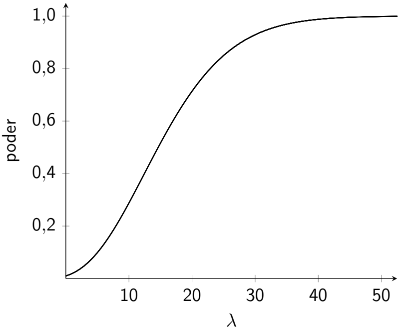
Kruskal-Wallis Test

day	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
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

Kruskal-Wallis Test

In R:

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



Conclusion

- The methods we presented are computationally less demanding than simulation ones.
- Some input elements may be difficult for users.
- In one example, the function `wilcox.test.pss` performed well even for a small sample.
- Noether's (1987) formula produced different results in the examples related to the Wilcoxon tests.
- The method of asymptotic relative efficiency underestimated sample size in one example.
- Future work suggestions:
  - tests of Friedman, Cochran, and Jonckheere-Terpstra;
  - admit the possibility of ties;
  - a measure of effect size for the Kruskal-Wallis test.

References

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*. New 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.

References

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. Wien: 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.