Nonparametric statistics

Markus Neuhäuser

Department of Mathematics and Technology,

RheinAhrCampus,

Koblenz University of Applied Sciences,

Remagen, Germany

neuhaeuser@rheinahrcampus.de

HOCHSCHULE KOBLENZ
UNIVERSITY OF APPLIED SCIENCES
RheinAhrCampus

Bonn, 18. Sept 2018

Non-normal data are very common, normality is the exception rather than the norm in applied research.

Check for normality not possible in case of small samples.

Transformations

Search for a suitable transformation by *trial and error* (p-hacking)

→ not acceptable (especially in case of small samples).

Transformations have to be predetermined.

The hypotheses before and after the transformation may differ.

Student's *t* test and other parametric methods are regarded as *robust*,

but:

- parametric tests are less robust
  - in case of small sample sizes
  - in case of very small significance levels ($\alpha << 0.05$)

    (Beasley et al. 2004, *Applied Statistics* 53, 95-108),

- nonparametric tests are more powerful

  in case of non-normal distributions.

→ **Apply nonparametric tests**

e.g. Wilcoxon rank sum test (Mann-Whitney *U* test),

in case of small samples and/or ties:

exact permutation tests instead of asymptotic or approximate tests

Tests for the general alternative (any difference between samples):

- Kolmogorov-Smirnov test
- Baumgartner-Weiß-Schindler (BWS) test

  (Baumgartner et al. 1998, *Biometrics* 54, 1129-1135)
- modified BWS test

  (Murakami 2006, *Journal of the Japanese Society of Computational Statistics* 19, 1-13)
- test proposed by Zhang (2006)

  (*Technometrics* 48, 95-103)

Power comparison: new tests better than Kolmogorov-Smirnov test

(Neuhäuser et al. 2017, *Communications in Statistics – Simulation and Computation* 46, 903-909).

Tests for a location shift

- Wilcoxon rank sum test (Mann-Whitney *U* test)
- Fisher-Pitman permutation test (nonparametric *t* test)
- Maximum to combine t test and Wilcoxon rank sum test

  (Neuhäuser 2015, *Journal of Applied Statistics* 42, 2769-2775)

- Tests based on (a modification of) Chebyshev's inequality, can yield small p-values even with very small sample sizes

  (Beasley et al. 2004, *Applied Statistics* 53, 95-108)

Tests for the Behrens-Fisher problem

(location-shift but difference in variability possible):

- Welch *t* test based on ranks (e.g. Cribbie et al. 2007, *J Mod Appl Stat Methods* 6, 117–132)
- Brunner-Munzel test (*Biometrical Journal* 2000; 42, 17-25)
- Permutation test with Brunner-Munzel statistic

  (Neubert & Brunner, 2007, *Computational Statistics and Data Analysis* 51, 5192-5204)

- Combination: Maximum test with different test statistics

  (Welz et al., *Journal of Statistical Computation and Simulation*, published online Jan. 2018)

Most <u>nonparametric tests</u> can be carried out as permutation or bootstrap tests, useful especially for small sample sizes and/or in the presence of ties.

→ no need for a restriction to continuous data.

discrete data → ties possible

Usually, mean ranks are used in practice (e.g. twice 3.5 instead of 3 and 4)

> Mean ranks give a more efficient test in comparison to randomly broken ties (Putter 1955, *Ann. Math. Statist.* 1955; 26, 368-386).

Be careful when a variable is computed,

in many software systems small numerical differences occur between identical results:

R (version 3.1.2):

```
> a <- log(3)-log(1)
> b <- log(6)-log(2)          # e.g. log-transformation and difference to baseline
> v <- c(a,b)
> v
[1] 1.098612  1.098612
> a-b
[1] 2.220446e-16
> rank(v, ties.method="average")
[1] 2 1
```

→ Rounding necessary before computing ranks

because pseudo-precision can reduce efficiency

(Neuhäuser et al. 2007, *Methods of Information in Medicine* 46, 538-541).

In many situations, exact permutation tests can be carried out nowadays,
not only for small samples,
e.g. an exact Pearson χ^2 test can easily be performed for large samples.

It is an irony, that asymptotic methods are carried out and justified as "good approximations, with no mention of the fact that the gold standard analysis [i.e. the exact test] they are trying to approximate are themselves readily available" (Berger et al., 2008, *Statistical Methods in Medical Research* 17, 231-242).

In a Socratic dialogue, Socrates asked: "If you can observe the exact p-value, then why would you go on to attempt to approximate it?"
(Berger, 2009, *Journal of Modern Applied Statistical Methods* 8, 316-321)

Thank you very much for your attention!