MS-C1620 Statistical inference

3 Non-parametric tests

Pekka Pere

Department of Mathematics and Systems Analysis
School of Science
Aalto University

Academic year 2023–2024
Period III–IV

Non-parametric statistics

Statistical models and methods can be divided roughly into two classes, parametric and non-parametric.

A model is parametric if it is fully defined by a set of parameters $\theta_1, \ldots, \theta_K$ (we assume a certain "family" of distributions, e.g. "some normal" or "some exponential").

A model or method is non-parametric or distribution-free (roughly) if it does not assume a specific family of distributions.

Stricter assumptions $\leftrightsquigarrow$ *more powerful results* $\leftrightsquigarrow$ *more narrow area of application.*

How to understand a statistical test

or: How it works, why it works, when it works

Rule 1: Suppose the hypothesis is true. Try to understand how the test statistic is then distributed. What kinds of values do you expect?

Rule 2: Suppose the hypothesis is false. Try to understand how the test statistic is then distributed.

Contents

1 Sign tests

2 Rank tests

One-sample sign test

One-sample sign test is applied in similar testing problems as the one-sample t-test. However, it makes much milder distributional assumptions.

One-sample sign test, assumptions

Let $x_1, x_2, \ldots, x_n$ be an i.i.d. sample from a continuous distribution with the median m.

One-sample sign test, hypotheses

$$H_0 : m = m_0 \quad H_1 : m \neq m_0.$$

One-sample sign test

One-sample sign test, test statistic

- The test statistic S equals the number of cases with $x_i > m_0$ (alternatively, the number of the cases with $x_i < m_0$) and follows binomial distribution with the parameters n and $1/2$ under H_0.
- Under H_0, the expected value of the test statistic is $\frac{1}{2}n$ (its variance is $\frac{1}{4}n$) and both **large** and **small** values of the test statistic suggest that the null hypothesis H_0 is false.

The distribution of the test statistic S is tabulated and statistical software gives exact p-values of the test.

One-sample sign test, asymptotic distribution

- If the sample size is large, then the standardized test statistic,

$$Z = \frac{S - n/2}{\sqrt{n/4}},$$

follows approximately the standard normal distribution under H_0.

- For large n, the p-value of the test can thus be retrieved from a normal table.
- The approximation is usually good enough if $n > 20$. For smaller samples, it is better to rely on the exact distribution of the test statistic S (which is the binomial distribution).

One sample sign test and discrete distributions

We assumed above that the observations come from a continuous distribution. However, the sign test can be used for discrete variables as well.

In that case it is possible that for some of the differences, $x_i - m_0 = 0$, are neither positive nor negative.

If the number of zeros is small compared to the sample size, these observations can be deleted and the sample size can be modified accordingly.

If the number of zeros is large, then the zeroes should be dealt with such that they are against rejecting the null hypothesis (conservative approach).

- For example: Consider the case of two-tailed alternative hypothesis, 3 negative differences, 15 positive differences and 6 zero differences. Now the test should be conducted as there were 9 negative differences and 15 positive ones.

Paired sign test

Paired sign test is a non-parametric version of the paired t-test.

Paired sign test, assumptions

Observations consist of an i.i.d. sample of pairs $(x_{i1}, x_{i2}), i = 1, 2, \ldots, n$ (the values **within** a pair need not be independent). The distribution of the differences $d_i = x_{i2} - x_{i1}$ is continuous (denote the median of this distribution by m_d).

Paired sign test, hypotheses

$$H_0 : m_d = 0 \quad H_1 : m_d \neq 0.$$

The test is conducted by applying the one-sample sign test to the differences d_i.

Contents

1 Sign tests

2 Rank tests

One-sample signed rank test/Wilcoxon test

More sophisticated versions of the sign tests are given by signed rank tests, which consider not only the signs of the observations but also their relative order (and thus use more information).

One-sample signed rank test, assumptions

Let $x_1, x_2, \ldots, x_n$ be an i.i.d. sample from continuous, symmetric distribution with the median m.

One-sample signed rank test, hypotheses

$$H_0 : m = m_0 \quad H_1 : m \neq m_0.$$

One-sample signed rank test

- Calculate the absolute values of the differences $|d_i| = |x_i - m_0|, i = 1, 2, \ldots, n$, and order the absolute values from the smallest to the largest.
- Define signed ranks r_i such that r_i is the rank of the absolute value $|d_i| = |x_i - m_0|$ multiplied with the sign of the difference $(x_i - m_0)$.

One-sample signed rank test, test statistic

- The test statistic,

$$W = \sum_{r_i > 0} r_i$$

  is the sum of the positive ranks (alternatively, the sum of the negative ranks) and follows a certain distribution under H_0.
- Under H_0, the expected value of the test statistic is $\frac{n(n+1)}{4}$ (its variance is $\frac{n(n+1)(2n+1)}{24}$) and both **large** and **small** values of the test statistic suggest that the null hypothesis H_0 is false.

Asymptotic one-sample Wilcoxon signed rank test

- The distribution of the test statistic W is tabulated and statistical software gives exact p-values of the test.
- If the sample size is large the standardized test statistic,

$$Z = \frac{W - E(W)}{\sqrt{var(W)}},$$

  where $E(W) = \frac{n(n+1)}{4}$ and $var(W) = \frac{n(n+1)(2n+1)}{24}$ follows approximately the standard normal distribution under H_0.
- For large n, the p-value of the test can thus be retrieved from a normal table.
- The approximation is usually good enough if $n > 20$. For smaller samples, it is better to rely on the exact distribution of the test statistic W.

One-sample signed rank test

We assumed above that the observations come from a continuous distribution but the Wilcoxon signed rank test can be applied for discrete observations as well.

However, it is then possible that some points share the same rank.

In that case, all tied points are assigned rank equal to the median of the corresponding ranks.

- If two observations have the same rank corresponding to ranks 7 and 8, then both points are assigned to have rank 7.5.
- If three sample points have the same rank corresponding to ranks 3, 4, and 5, then each is assigned to have rank 4.

Paired signed rank test

- Also the one-sample signed rank test can be used for paired data, in a way analogous to paired t-test and paired sign test.

Paired signed rank test, assumptions

Observations consist of an i.i.d. sample of pairs $(x_{i1}, x_{i2}), i = 1, 2, \ldots, n$ (the values **within** a pair need not be independent). The difference $d_i = x_{i1} - x_{i2}$ has a symmetric distribution, the median of which is denoted by m_d.

Paired signed rank test, hypotheses

$$H_0 : m_d = 0 \quad H_1 : m_d \neq 0.$$

The test is conducted by applying the one-sample signed rank test to the differences d_i.

When to use sign tests and signed rank tests

Both families of tests are non-parametric counterparts of the t-test and worthy alternatives when normality cannot be assumed.

Both types of tests are appropriate in similar problems:

- one sample — comparison of the location to a constant,
- paired samples — comparison of the locations.

The assumptions of the two types of tests differ:

- sign test — continuous distribution,
- signed rank test — continuous symmetric distribution.

If normality can be assumed, use t-tests. If symmetry (but no normality) can be assumed, use signed rank tests. Otherwise, use sign tests.

Two-sample rank test/Wilcoxon rank-sum test

The two-sample rank sum test (a.k.a. Mann-Whitney U-test) is as the two sample t-test, but requires milder assumptions.

Two-sample rank test, assumptions

- Let $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$ be mutually independent i.i.d. samples from continuous distributions.
- Assume that the distributions of the samples are equal up to a location shift and denote their medians by m_x and m_y, respectively.

Two-sample rank test, hypotheses

$$H_0 : m_x = m_y \quad H_1 : m_x \neq m_y$$

Two-sample rank test

- The two-sample rank test is based on analyzing the order of all the observations.
- Assume without loss of generality that $n \leq m$ and combine $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ into one sample $z_1, ..., z_{n+m}$. Order then the observations z_i from smallest to largest and let r_i be the rank of z_i in the combined sample $z_1, z_2, \ldots, z_{n+m}$.

Two-sample rank test, test statistic

- The test statistic
$$W = \sum_{i=1}^{n} r_i$$
is the sum of the ranks of the smaller sample and it follows a certain distribution under H_0.
- Under H_0, the expected value of the test statistic is $n(n+m+1)/2$ (its variance is $nm(n+m+1)/12$) and both **large** and **small** values of the test statistic suggest that the null hypothesis H_0 is false.

Two-sample rank test, asymptotic distribution

- The distribution of the test statistic W is tabulated and statistical software gives the exact p-values.
- If the sample size is large the standardized test statistic,

$$Z = \frac{W - E(W)}{\sqrt{var(W)}},$$

  where $E(W) = n(n+m+1)/2$ and $var(W) = nm(n+m+1)/12$, follows approximately the standard normal distribution, under H_0.
- For large n, the p-value of the test can thus be retrieved from a normal table.
- The approximation is usually good enough if $n, m > 10$. For smaller samples, the exact distribution of the test statistic W should be relied upon.

Two-sample rank test

- The two-sample rank test can be used also when the observations are discrete, however then it is possible that some of the sample points have the same rank. The ties are handled as in one-sample signed rank test.
- The two sample rank test is a non-parametric counterpart of the two-sample t-test.
- If normality can be assumed, use the two-sample t-test. Otherwise, use the two-sample rank test.

Final note on rank tests

Note that ranks can be used even when a variable cannot be measured numerically but the observations can be ranked (for example, one could rank bands, or qualities of apartments, without measuring them numerically).