Statistics in Geophysics: Inferential Statistics III

Steffen Unkel

Department of Statistics
Ludwig-Maximilians-University Munich, Germany

Background

- Tests not requiring assumptions involving specific parametric distributions for the data or for the sampling distribution of the test statistics are called nonparametric.

- Nonparametric methods are appropriate if
  1. we know or suspect that the parametric assumption(s) required for a particular test are not met;
  2. a test statistic that is suggested or dictated by the problem at hand is a complicated function of the data, and its sampling distribution is unknown and/or cannot be derived analytically.

- Only a few nonparametric tests for location will be presented here.

One-sample Wilcoxon signed-rank test

- Let $X_1, \ldots, X_n$ be a random sample with continuous cdf $F_X(\cdot)$.
- Suppose that it is desired to test that the 0.5 quantile, x_{med}, of the population sampled from is a specific value, say δ_0.
- Consider the test problems:
  (a) $H_0 : x_{med} = \delta_0$ vs. $H_1 : x_{med} \neq \delta_0$
  (b) $H_0 : x_{med} \geq \delta_0$ vs. $H_1 : x_{med} < \delta_0$
  (c) $H_0 : x_{med} \leq \delta_0$ vs. $H_1 : x_{med} > \delta_0$.
- For $i = 1 \ldots, n$, let $D_i = X_i - \delta_0$ and define

$$Z_i = \begin{cases} 1 & \text{if} \quad D_i > 0 \\ 0 & \text{if} \quad D_i < 0 \end{cases} .$$

One-sample Wilcoxon signed-rank test

- Test statistic

$$W^+ = \sum_{i=1}^{n} R_i Z_i,$$

where R_i is the rank of $|D_i|$.

- Rejection region:
  (a) $W^+ > w^+_{1-\alpha/2}$ or $W^+ < w^+_{\alpha/2}$
  (b) $W^+ < w^+_{\alpha}$
  (c) $W^+ > w^+_{1-\alpha}$,

  where w^+_α denotes the α-quantile of the distribution of W^+.

One-sample Wilcoxon signed-rank test

- For sufficiently large samples: Approximation by $\mathcal{N}\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right)$.
- Test statistic:

$$Z = \frac{W^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \overset{a}{\sim} \mathcal{N}(0,1) \ .$$

- Rejection region:
  (a) $Z > z_{1-\alpha/2}$ or $Z < z_{\alpha/2}$
  (b) $Z < z_\alpha$
  (c) $Z > z_{1-\alpha}$,

  where z_α is the α-quantile of the standard normal distribution.

Wilcoxon signed-rank test for paired data

- We assume that the sampling situation is such that we observe paired data $(X_1, Y_1), \ldots, (X_n, Y_n)$.
- For $i = 1, \ldots, n$, the differences $D_i = X_i - Y_i$ arise from a continuous distribution and each pair (X_i, Y_i) is chosen randomly and independent.
- The null hypothesis is that the median difference, δ, between pairs of observations is zero.
- Consider the test problems:
  (a) $H_0 : \delta = 0$ vs. $H_1 : \delta \neq 0$
  (b) $H_0 : \delta \geq 0$ vs. $H_1 : \delta < 0$
  (c) $H_0 : \delta \leq 0$ vs. $H_1 : \delta > 0$.

Wilcoxon signed-rank test for paired data

- Define

$$Z_i = \begin{cases} 1 & \text{if} \quad D_i > 0 \\ 0 & \text{if} \quad D_i < 0 \end{cases}$$

- Test statistic:

$$W^+ = \sum_{i=1}^{n} R_i Z_i,$$

  where R_i is the rank of $|D_i|$.

- Rejection region:
  (a) $W^+ > w^+_{1-\alpha/2}$ or $W^+ < w^+_{\alpha/2}$
  (b) $W^+ < w^+_{\alpha}$
  (c) $W^+ > w^+_{1-\alpha}$,

  where w^+_α denotes the α-quantile of the distribution of W^+.

Wilcoxon rank-sum test

- Given two samples of independent data, the aim is to test for a possible difference in location.
- The null hypothesis is that the two data samples have been drawn from the same distribution.
- Under H_0 there are $n + m$ observations making up a single distribution, where n (m) denote the number of observations in sample 1 (sample 2).
- The test statistic is a function of the ranks of the data values within the $n + m$ observations that are pooled under H_0.

Wilcoxon rank-sum test

- Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be two random samples from populations with continuous cdfs $F_X(\cdot)$ and $F_Y(\cdot)$, respectively.

- Consider the test problems:
  (a) $H_0 : x_{med} = y_{med}$ vs. $H_1 : x_{med} \neq y_{med}$
  (b) $H_0 : x_{med} \geq y_{med}$ vs. $H_1 : x_{med} < y_{med}$
  (c) $H_0 : x_{med} \leq y_{med}$ vs. $H_1 : x_{med} > y_{med}$.

- Arrange the $n + m$ observations of the pooled sample $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ in ascending order.

- Define

$$V_i = \begin{cases} 1 & \text{if the } i\text{-th order statistic belongs to the } X \text{ sample} \\ 0 & \text{if the } i\text{-th order statistic belongs to the } Y \text{ sample} \end{cases}$$

Wilcoxon rank-sum test

- Test statistic:

$$W_{n,m} = \sum_{i=1}^{n+m} iV_i = \sum_{i=1}^{n} R(X_i) \ ,$$

where $R(X_i)$ is the rank of X_i in the pooled sample.

- Rejection region:
  (a) $W_{n,m} > w_{1-\alpha/2}(n,m)$ or $W_{n,m} < w_{\alpha/2}(n,m)$
  (b) $W_{n,m} < w_{\alpha}(n,m)$
  (c) $W_{n,m} > w_{1-\alpha}(n,m)$,
  where w_α denotes the α-quantile of the distribution of $W_{n,m}$.

- For sufficiently large samples: Approximation by $\mathcal{N}(n(n+m+1)/2, nm(n+m+1)/12)$.