

STATISTICAL TESTS BASED ON RANKS

Jana Jurečková
Charles University, Prague

Contents

1	Basic concepts of hypotheses testing in nonparametric setup	5
1.1	Introduction	5
1.2	Principle of invariance in hypotheses testing	6
2	Properties of ranks and of order statistics	9
3	Locally most powerful rank tests	11
4	Selected two-sample rank tests	15
4.1	Two-sample tests of location	15
4.2	Two-sample rank tests of scale	19
4.3	Rank tests of H_0 against general two-sample alternatives based on the empirical distribution functions.	20
4.4	Modification of tests in the presence of ties	22
5	Tests for comparison of the treatments based on paired observations	25
5.1	Rank tests of H_1	25
5.2	One-sample Wilcoxon test	27
5.3	Sign test	27
6	Tests of independence in bivariate population	29
6.1	Spearman test	29
6.2	Quadrant test	30
7	Rank test for comparison of several treatments	31
7.1	One-way classification	31
7.2	Kruskal-Wallis rank test	32
7.3	Two-way classification (random blocks)	33

Chapter 1

Basic concepts of hypotheses testing in nonparametric setup

1.1 Introduction

Let $\mathbf{X} = (X_1, \dots, X_n)$ be a random vector (vector of observations) and let H and K be two disjoint sets of probability distributions on $(\mathbb{R}^n, \mathcal{B}^n)$. We say that \mathbf{X} fulfills the hypothesis if the distribution of \mathbf{X} belongs to H and that \mathbf{X} fulfills the alternative if its distribution belongs to K . We shall use the same symbols H and K either to denote the hypotheses or the set. The hypothesis is usually the homogeneous, symmetric, independent of the statements while the alternative means inhomogeneity, asymmetry, dependence etc.

The problem is to decide between the hypothesis and alternative on the basis of observations X_1, \dots, X_n . Every rule, which assigns just one of the decisions "to accept H " or "to reject H " to any point $\mathbf{x} = (x_1, \dots, x_n)$, is called *the test* (nonrandomized) of hypothesis H against alternative K . Such test partitions the sample space \mathcal{X} into two complementary parts: the *critical region* (rejection region) A_K and *acceptance region* A_H . The test rejects H if $\mathbf{x} \in A_K$ and accepts H if $\mathbf{x} \in A_H$.

If we perform the test on the basis of observations \mathbf{x} , then either our decision is correct or we could make either of the following two kinds of errors:

- (1) We reject H even if it is correct (*error of the first kind*);
- (2) we accept H even if it is incorrect (*error of the second kind*).

It is desirable to use the test with the smallest possible probabilities of both errors. If the true distribution P of \mathbf{X} satisfies $P \in H$, then the probability of the error of the 1st kind equals $\mathbb{P}(\mathbf{X} \in A_K)$ and $\sup_{P \in H} P(\mathbf{X} \in A_K)$ is called *the size* of the test with the critical region A_K . If the true distribution Q of \mathbf{X} satisfies $Q \in K$, then the probability of the error of the second kind equals $Q(\mathbf{X} \in A_H) = 1 - Q(\mathbf{X} \in A_K)$. The probability $\beta(Q) = Q(\mathbf{X} \in A_K)$, $Q \in K$ is called *the power* of the test against the alternative Q . The function $\beta(Q) : K \mapsto [0, 1]$ is called *the power function* of the test. The desirable test maximizes the power function uniformly over the whole alternative and has the small probability of the error of the first kind for all distributions from the hypothesis.

The testing theory and searching for the optimum considerably simplifies when we supplement the family of tests by the *randomized tests*. A randomized test rejects H with the probability $\Phi(\mathbf{x})$ and accepts with probability $1 - \Phi(\mathbf{x})$ while observing \mathbf{x} , where $0 \leq \Phi(\mathbf{x}) \leq 1 \quad \forall \mathbf{x}$ is the *test function*. The set of randomized tests coincides with the set $\{\Phi(\mathbf{x}) : 0 \leq \Phi \leq 1\}$ and hence it is convex and weakly compact.

If X has distribution P , then the test Φ rejects H with the probability

$$\beta_{\Phi}(P) = \mathbb{E}_P(\Phi(\mathbf{X})) = \int_{\mathcal{X}} \Phi(\mathbf{x})dP(\mathbf{x}).$$

Intuitively, the best test should satisfy

$$\beta_{\Phi}(Q) = \mathbb{E}_Q(\Phi(\mathbf{X})) := \max \quad \forall Q \in K \quad (1.1)$$

and simultaneously

$$\beta_{\Phi}(P) = \mathbb{E}_P(\Phi(\mathbf{X})) := \min \quad \forall P \in H. \quad (1.2)$$

Because no test satisfies both conditions simultaneously, the optimal test is defined in the following way: Select a small number α , $0 < \alpha < 1$, called *the significance level*, and among all tests satisfying

$$\beta_{\Phi}(P) \leq \alpha \quad \forall P \in H \quad (1.3)$$

we look for the test satisfying (1.1). Such test, if it exists, is called *the uniformly most powerful test of size $\leq \alpha$* , briefly *the uniformly most powerful α -test of H against K* . The hypothesis [alternative] is called *simple* if H [K] is a one-point set; otherwise it is called *composite*. The test of a simple hypothesis against a simple alternative is given by the fundamental *Neyman-Pearson lemma*.

THEOREM 1.1.1 NEYMAN-PEARSON LEMMA. *Let P and Q be two probability distributions with densities p and q with respect to some measure μ (e.g., $\mu = P + Q$). Then, for testing the simple hypothesis $H : \{P\}$ against the simple alternative $K : \{Q\}$, there exists the test Φ and a constant k such that*

$$\mathbb{E}_P(\Phi(\mathbf{X})) = \alpha \quad (1.4)$$

and

$$\Phi(\mathbf{x}) = \begin{cases} 1 & \text{if } q(\mathbf{x}) > k \cdot p(\mathbf{x}) \\ 0 & \text{if } q(\mathbf{x}) < k \cdot p(\mathbf{x}). \end{cases} \quad (1.5)$$

This test is the most powerful α -test of H against K .

1.2 Principle of invariance in hypotheses testing

Let g be a 1:1 transformation $\mathcal{X} : \mathcal{X}$. We say that the problem of testing of H against K is *invariant* with respect to g if g retains both H and K , i.e.

$$\begin{aligned} \mathbf{X} \text{ satisfies } H & \text{ iff } g\mathbf{X} \text{ satisfies } H \\ \mathbf{X} \text{ satisfies } K & \text{ iff } g\mathbf{X} \text{ satisfies } K. \end{aligned}$$

If the problem of testing H against K is invariant with respect to the group \mathcal{G} of transformations of \mathcal{X} onto \mathcal{X} , then we could naturally consider only the tests which are equally invariant, i.e. the tests Φ satisfying

$$\Phi(g\mathbf{X}) = \Phi(\mathbf{X}) \quad \forall \mathbf{x} \in \mathcal{X}, \quad \forall g \in \mathcal{G}.$$

We shall then look for *the most powerful invariant α -test*. In some cases, there exists a statistic $T(\mathbf{X})$, called *maximal invariant*, such that every invariant test is a function of $T(\mathbf{X})$.

DEFINITION 1.2.1 *The statistic $T = T(\mathbf{X})$ is called maximal invariant with respect to the group \mathcal{G} of transformations, provided T is invariant, i.e.*

$$T(g\mathbf{x}) = T(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{X}, \quad \forall g \in \mathcal{G}$$

and

$$\text{if } T(\mathbf{x}_1) = T(\mathbf{x}_2) \text{ then there exists } g \in \mathcal{G} \text{ such that } \mathbf{x}_2 = g\mathbf{x}_1.$$

The structure of invariant tests is characterized in the following theorem.

THEOREM 1.2.1 *Let $T(\mathbf{X})$ be the maximal invariant with respect to the group of transformations \mathcal{X} . The test Φ is invariant with respect to \mathcal{G} if and only if there exists a function h such that*

$$\Phi(\mathbf{x}) = h(T(\mathbf{x})) \quad \forall \mathbf{x} \in \mathcal{X}.$$

PROOF. (i) If $\Phi(\mathbf{x}) = h(T(\mathbf{x})) \forall \mathbf{x}$, then $\Phi(g\mathbf{x}) = h(T(g\mathbf{x})) = h(T(\mathbf{x})) = \Phi(\mathbf{x}) \forall g \in \mathcal{G}$ and hence Φ is invariant.

(ii) Let Φ be invariant and let $T(\mathbf{x}_1) = T(\mathbf{x}_2)$. Then, by the definition, $\mathbf{x}_2 = g\mathbf{x}_1$ for some $f \in \mathcal{G}$ and hence $\Phi(\mathbf{x}_2) = \Phi(\mathbf{x}_1)$. ■

Examples of maximal invariants

1. Let $\mathbf{x} = (x_1, \dots, x_n)$ and let G be the group of translations

$$g\mathbf{x} = (x_1 + c, \dots, x_n + c), \quad c \in \mathbb{R}^1.$$

Then the maximal invariant is, e.g., $T(\mathbf{x}) = (x_2 - x_1, \dots, x_n - x_1)$.

2. Let G be the group of orthonormal transformations $\mathbb{R}^n \mapsto \mathbb{R}^n$. Then $T(\mathbf{x}) = \sum_{i=1}^n x_i^2$ is the maximal invariant.
3. Let G be the set of $n!$ permutations of x_1, \dots, x_n . Then the vector ordered components of \mathbf{x} (*vector of order statistics*)

$$T(\mathbf{x}) = (x_{n:1} \leq x_{n:2} \leq \dots \leq x_{n:n})$$

is the maximal invariant with respect to G .

4. Let G be the set of transformations $x'_i = f(x_i)$, $i = 1, \dots, n$ such that $f : \mathbb{R}^1 \mapsto \mathbb{R}^1$ is continuous and strictly increasing function. Consider only the points of the sample space \mathcal{X} with different components. Let R_i be the rank of x_i among x_1, \dots, x_n , i.e. $R_i = \sum_{j=1}^n I[x_j \leq x_i]$, $i = 1, \dots, n$. Then $T(\mathbf{x}) = (R_1, \dots, R_n)$ is the maximal invariant for G .

Actually, a continuous and increasing function does not change the ranks of the components of \mathbf{x} , i.e. T is invariant to G . On the other hand, let two different vectors \mathbf{x} and \mathbf{x}' have the same vector of ranks R_1, \dots, R_n . Put $f(x_i) = x'_i$, $i = 1, \dots, n$ and let f be linear on the intervals $[x_{n:1}, x_{n:2}], \dots, [x_{n:n-1}, x_{n:n}]$; define f in the rest of the real line so that it is strictly increasing. Such f always exists, hence T is the maximal invariant.

Chapter 2

Properties of ranks and of order statistics

Let $\mathbf{X} = (X_1, \dots, X_n)$ be the vector of observations; denote $X_{n:1} \leq X_{n:2} \leq \dots \leq X_{n:n}$ the components of \mathbf{X} ordered according to increasing magnitude. The vector $\mathbf{X}_{(\cdot)} = (X_{n:1}, \dots, X_{n:n})$ is called the *vector of order statistics* and $X_{n:i}$ is called the *i th order statistic*.

Assume that the components of \mathbf{X} are different and define the *rank* of X_i as $R_i = \sum_{j=1}^n I[X_j \leq X_i]$. Then the vector R of ranks of \mathbf{X} takes on the values in the set \mathcal{R} of $n!$ permutations (r_1, \dots, r_n) of $(1, \dots, n)$.

The first property of $\mathbf{X}_{(\cdot)}$ and R is described in the following proposition.

Proposition 1. The pair $(\mathbf{X}_{(\cdot)}, R)$ is a sufficient statistic for any family of absolute continuous probability distributions of \mathbf{X} .

PROOF. If $\mathbf{X}_{(\cdot)} = \mathbf{x}_{(\cdot)}$ and $R = r$ are prescribed, then

$$P(\mathbf{X} \in A | \mathbf{X}_{(\cdot)} = \mathbf{x}_{(\cdot)}, R = r) = P\{(X_{n:r_1}, \dots, X_{n:r_n}) \in A | \mathbf{X}_{(\cdot)} = \mathbf{x}_{(\cdot)}, R = r\} = 0 \text{ or } 1$$

depending on whether $(x_{n:r_1}, \dots, x_{n:r_n})$ is an element of A or not; this probability does not depend on the original distribution of \mathbf{X} and this is the property defining the sufficiency. ■

DEFINITION 2.0.1 We say that the random vector \mathbf{X} satisfies the hypothesis of randomness H_0 , if it has a probability distribution with density of the form

$$p(\mathbf{x}) = \prod_{i=1}^n f(x_i), \quad \mathbf{x} \in \mathbb{R}^n$$

where f is an arbitrary one-dimensional density. Otherwise speaking, X satisfies the hypothesis of randomness provided its components are independent identically distributed (i.i.d.) random variables with absolute continuous distribution.

The following theorem gives the general form of the distribution of $\mathbf{X}_{(\cdot)}$ and of R .

THEOREM 2.0.1 Let \mathbf{X} have the density $p_n(x_1, \dots, x_n)$.

(i) Then the vector $\mathbf{X}_{(\cdot)}$ of order statistics has the distribution with the density

$$\bar{p}(x_{n:1}, \dots, x_{n:n}) = \begin{cases} \sum_{r \in \mathcal{R}} p(x_{n:r_1}, \dots, x_{n:r_n}) & \dots x_{n:1} \leq \dots \leq x_{n:n} \\ 0 & \text{otherwise.} \end{cases} \quad (2.1)$$

(ii) The conditional distribution of R given $\mathbf{X}_{(\cdot)} = \mathbf{x}_{(\cdot)}$ has the form

$$\Pr(R = r | \mathbf{X}_{(\cdot)} = \mathbf{x}_{(\cdot)}) = \frac{p(x_{n:r_1}, \dots, x_{n:r_n})}{\bar{p}(x_{n:1}, \dots, x_{n:n})} \quad (2.2)$$

for any $r \in \mathcal{R}$ and any $x_{n:1} \leq \dots \leq x_{n:n}$.

The distributions of $X_{(\cdot)}$ and R considerably simplify under the hypothesis H_0 : This is described in the following theorem.

THEOREM 2.0.2 *If \mathbf{X} satisfies the hypothesis of randomness H_0 , then $\mathbf{X}_{(\cdot)}$ and R are independent, the vector of ranks R has the uniform discrete distribution*

$$\Pr(R = r) = \frac{1}{n!}, \quad r \in \mathcal{R} \quad (2.3)$$

and the distribution of $\mathbf{X}_{(\cdot)}$ has the density

$$\bar{p}(x_{n:1}, \dots, x_{n:n}) = \begin{cases} n!p(x_{n:1}, \dots, x_{n:n}) & \dots x_{n:1} \leq \dots \leq x_{n:n} \\ 0 & \dots \text{otherwise.} \end{cases} \quad (2.4)$$

Finally, the following theorem summarizes some properties of the marginal distributions of the random vectors R and $\mathbf{X}_{(\cdot)}$ under H_0 .

THEOREM 2.0.3 *Let \mathbf{X} satisfies the hypothesis H_0 . Then*

(i) $\Pr(R_i = j) = \frac{1}{n} \quad \forall i, j = 1, \dots, n.$

(ii) $\Pr(R_i = k, R_j = m) = \frac{1}{n(n-1)}$ for $1 \leq i, j, k, m \leq n, i \neq j, k \neq m.$

(iii) $\mathbb{E}R_i = \frac{n+1}{2}, \quad i = 1, \dots, n.$

(iv) $\text{var } R_i = \frac{n^2-1}{12}, \quad i = 1, \dots, n.$

(v) $\text{cov}(R_i, R_j) = -\frac{n+1}{12}, \quad 1 \leq i, j \leq n, i \neq j.$

(vi) *If \mathbf{X} has the density $p(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i)$, then $X_{n:k}$ has the distribution with the density*

$$f_{(n)}(x) = n \binom{n-1}{k-1} (F(x))^{k-1} (1-F(x))^{n-k} f(x), \quad x \in \mathbb{R}^1$$

where $F(x) = \int_{-\infty}^x f(y)dy, \quad k = 1, \dots, n.$

Remark. In the special case that \mathbf{X} is a sample from the uniform $R[0, 1]$ distribution, $X_{n:k}$ has beta distribution $B(k, n-k+1)$ with the expectation and variance

$$\mathbb{E}X_{n:k} = \frac{k}{n+1}, \quad \text{var } X_{n:k} = \frac{k(n-k+1)}{(n+1)^2(n+2)}.$$

Chapter 3

Locally most powerful rank tests

We want to test a hypothesis on the distribution of the random vector \mathbf{X} , say the hypothesis of randomness H_0 . The rank test depends on \mathbf{X} only through the ranks R_1, \dots, R_n of the components of \mathbf{X} and hence could be characterized by test function $\Phi(R)$. If we test the hypothesis of randomness H_0 against a simple alternative $K : \{Q\}$ [\mathbf{X} has the fixed distribution Q], then the most powerful rank α -test of H_0 against K follows directly from the Neyman-Pearson Lemma and from Theorem 2.1:

$$\Phi(r) = \begin{cases} 1 & \dots n!Q(R = r) > k_\alpha \\ 0 & \dots n!Q(R = r) < k_\alpha \\ \gamma & \dots n!Q(R = r) = k_\alpha, r \in \mathcal{R} \end{cases}$$

where k_α and γ are determined so that

$$\#\{r : n!Q(R = r) > k_\alpha\} + \gamma\#\{r : n!Q(R = r) = k_\alpha\} = n!\alpha, \quad 0 < \alpha < 1.$$

However, many composite alternatives of the practical interest are too rich and the uniformly most powerful rank tests against such alternatives do not exist. Then we may take excursion to the local tests and look for a rank test most powerful locally in a neighborhood of the hypothesis.

DEFINITION 3.0.1 *Let $d(Q)$ be a measure of distance of alternative $Q \in K$ from the hypothesis H . The α -test Φ_0 is called the locally most powerful in the class \mathcal{M} of α -tests of H against K if, given any other $\Phi \in \mathcal{M}$, there exists $\varepsilon > 0$ such that $\beta_{\Phi_0}(Q) \geq \beta_\Phi(Q) \forall Q$ satisfying $0 < d(Q) < \varepsilon$.*

We shall illustrate the structure of the locally most powerful rank tests of H_0 against a class of alternatives covering the shift and regression in the location and scale.

THEOREM 3.0.2 *Let A be a class of densities, $A = \{g(x, \theta) : \theta \in \mathcal{J}\}$ such that*

- (a) $\mathcal{J} \subset \mathbb{R}^1$ is an open interval, $\mathcal{J} \ni 0$.
- (b) $g(x, \theta)$ is absolutely continuous in θ for almost all x .
- (c) For almost all x , there exists the limit

$$\dot{g}(x, 0) = \lim_{\theta \rightarrow 0} \frac{1}{\theta} [g(x, \theta) - g(x, 0)]$$

and

$$\lim_{\theta \rightarrow 0} \int_{-\infty}^{\infty} |\dot{g}(x, \theta)| dx = \int_{-\infty}^{\infty} |\dot{g}(x, 0)| dx.$$

Consider the alternative $K = \{q_\Delta : \Delta > 0\}$, where

$$q_\Delta(x_1, \dots, x_n) = \prod_{i=1}^n g(x_i, \Delta c_i), \quad c_1, \dots, c_n \text{ given numbers.}$$

Then the test with the critical region

$$\sum_{i=1}^n c_i a_n(R_i, g) \geq k$$

is the locally most powerful rank test of H_0 against K with the significance level $\alpha = P(\sum_{i=1}^n c_i a_n(R_i, g) \geq k)$, where P is any distribution satisfying H_0 ,

$$a_n(i, g) = \mathbb{E} \left[\frac{\dot{g}(X_{n:i}, 0)}{g(X_{n:i}, 0)} \right], \quad i = 1, \dots, n$$

and $X_{n:1}, \dots, X_{n:n}$ are the order statistics corresponding to the random sample of size n from the population with the density $g(x, 0)$.

Let us apply the theorem to find the locally most powerful rank tests of H_0 against some standard alternatives.

I. We shall start with the *alternative of the shift in location* and test H_0 on the random vector (X_1, \dots, X_N) against the alternative $K_1 : \{q_\Delta : \Delta > 0\}$ where

$$q_\Delta(x_1, \dots, x_n) = \prod_{i=1}^m f(x_i) \prod_{i=m+1}^N f(x_i - \Delta), \quad (3.1)$$

where f is a fixed absolute continuous density such that

$$\int_{-\infty}^{\infty} |f'(x)| dx < \infty. \quad (3.2)$$

Then the family of densities A with $g(x, \theta) = f(x - \theta)$ and $\mathcal{J} = \mathbb{R}^1$ fulfills the conditions (a) - (c) of Theorem 3.2. Then the locally most powerful rank α -test of H_0 against K has the critical region

$$\sum_{i=m+1}^N a_N(R_i, f) \geq k \quad (3.3)$$

where k satisfies the condition $P(\sum_{i=m+1}^N a_N(R_i, f) \geq k) = \alpha$, $P \in H_0$ and

$$a_N(i, f) = \mathbb{E} \left[-\frac{f'(X_{N:i})}{f(X_{N:i})} \right], \quad i = 1, \dots, N \quad (3.4)$$

and $X_{N:1} < \dots < X_{N:N}$ are the order statistics corresponding to the sample of size N from the distribution with the density f . The scores (3.4) may be also written as

$$a_N(i, f) = \mathbb{E} \varphi(U_{N:i}, f), \quad i = 1, \dots, N \quad (3.5)$$

where $\varphi(u, f) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}$, $0 < u < 1$ and $U_{N:1}, \dots, U_{N:N}$ are the order statistics corresponding to the sample of size N from the uniform $R(0, 1)$ distribution. The scores (3.4) could be also expressed in the form

$$a_N(i, f) = N \binom{N-1}{i-1} \int_{-\infty}^{\infty} f'(x) F^{i-1}(x) (1 - F(x))^{N-i} dx. \quad (3.6)$$

Remark. The computation of the scores (3.4) (see also (3.6) is difficult for some densities; if there are no tables of the scores at disposal, they are often replaced by the *approximate scores*

$$a_N(i, f) = \varphi\left(\frac{i}{N+1}\right) = \varphi(\mathbb{E}U_{N:i}, f), \quad i = 1, \dots, N. \quad (3.7)$$

The asymptotic critical values coincide for both types of scores.

II. H_0 against the alternative of two samples differing by scales. Consider H_0 against the alternative $K_2 : \{q_\Delta : \Delta > 0\}$ where

$$q_\Delta(x_1, \dots, x_N) = \prod_{i=1}^m f(x_i - \mu) \prod_{i=m+1}^N e^{-\Delta} f\left(\frac{x_i - \mu}{e^\Delta}\right), \quad \Delta > 0 \quad (3.8)$$

where f is an absolutely continuous density satisfying $\int_{-\infty}^{\infty} |xf'(x)|dx < \infty$ and μ is the nuisance parameter. Then the family of densities A with $g(x, \theta) = e^{-\theta}f((x - \mu)e^{-\theta})$, $\mathcal{J} = \mathbb{R}^1$, fulfills the conditions (a) - (c) of Theorem 3.2 and the locally most powerful test has the critical region

$$\sum_{i=m+1}^N a_{1N}(R_i, f) \geq k, \quad (3.9)$$

where k is determined by the condition $P(\sum_{i=m+1}^N a_{1N}(R_i, f) \geq k) = \alpha$, $P \in H_0$ and the scores have the form

$$a_{1N}(i, g) = \mathbb{E}\left\{-1 - X_{N:i} \frac{f'(X_{N:i})}{f(X_{N:i})}\right\} = \mathbb{E}\varphi_1(U_{N:i}, f), \quad (3.10)$$

$i = 1, \dots, N$, where $\varphi_1(u, f) = -1 - F^{-1}(u) \frac{f'(F^{-1}(u))}{f(F^{-1}(u))}$, $0 < u < 1$. In this case, too, we could replace the scores (3.10) by the approximate scores $a_{1N}(i, f) = \varphi_1\left(\frac{i}{N+1}, f\right)$, $i = 1, \dots, N$.

III. *Tests of H_0 against the alternative of simple regression.* Consider H_0 against the alternative $K = \{q_\Delta : \Delta > 0\}$ where $K_3 : q_\Delta(x_1, \dots, x_N) = \prod_{i=1}^N f(x_i - \Delta c_i)$ with a fixed absolutely continuous density f satisfying (3.2) and with given constants c_1, \dots, c_N , $\sum_{i=1}^N c_i^2 > 0$. Then the locally most powerful test has the critical region

$$\sum_{i=1}^N c_i a_N(R_i, f) \geq k \quad (3.11)$$

with the scores (3.5) and with k determined by the condition $P(\sum_{i=1}^N c_i a_N(R_i, f) \geq k) = \alpha$.

Chapter 4

Selected two-sample rank tests

Consider two random samples (X_1, \dots, X_m) and (Y_1, \dots, Y_n) with the respective distribution functions F and G . For the sake of brevity, we shall also denote $(X_1, \dots, X_m, Y_1, \dots, Y_n) = (Z_1, \dots, Z_N)$ with $N = m + n$. The hypothesis of randomness for the vector (Z_1, \dots, Z_N) in this special case could be reformulated as $H_0 : F \equiv G$. Consider first testing H_0 against the alternative $K_1 : G(x) \leq F(x) \forall x \in \mathbb{R}^1, G(x) \neq F(x)$ at least for one x .

K_1 is a one-sided alternative stating that the random variable Y is *stochastically larger* than X .

The problem of testing H_0 against K_1 is invariant to the group \mathcal{G} of transformations $z'_i = g(z_i)$, $i = 1, \dots, N$ where g is any continuous strictly increasing function. As we have seen before, the vector of ranks R_1, \dots, R_N of Z_1, \dots, Z_N is the maximal invariant with respect to \mathcal{G} . Then, by Theorem 1.2, the class of invariant tests coincides with that of rank tests.

Hence, we shall restrict our considerations to the rank tests. However, we could still reduce the class of tests due to the following considerations. Because both (X_1, \dots, X_m) and (Y_1, \dots, Y_n) are random samples, the distribution of the vector of ranks $(R_1, \dots, R_m, R_{m+1}, \dots, R_{m+n})$ is symmetric in the first m and the last n arguments under all pairs of distributions F and G . Hence, the sufficient statistic for the vector $(R_1, \dots, R_m, R_{m+1}, \dots, R_{m+n})$ are two vectors of ordered ranks

$$R'_1 < \dots < R'_m \quad \text{and} \quad R'_{m+1} < \dots < R'_{m+n} \quad (4.1)$$

of random variables X_1, \dots, X_m and Y_1, \dots, Y_n , ordered according to the increasing magnitude. Because either of the vectors in (4.1) determines the other, the family of invariant tests of H_0 against K_1 in turns reduces to the tests dependent only on the ordered ranks of one of the samples, e.g. on the ordered ranks of Y_1, \dots, Y_n .

Vector R'_{m+1}, \dots, R'_N could be equal to any of $\binom{N}{n}$ combinations. All these combinations are equally probable under H_0 and hence the critical region of each rank test of the size $\alpha = k / \binom{N}{n}$ consists of just k points s_1, \dots, s_n , $1 \leq s_1 < \dots < s_n \leq N$. The rank test differ in the points which they include in the critical regions.

The above alternative K_1 is still too rich and hence there does not exist the uniformly most powerful rank test of H_0 against K . However, we are able to find rank tests locally most powerful for H_0 against some important subsets of K .

4.1 Two-sample tests of location

Consider the special alternative of K_1 , namely that G differ from F by a shift in location, i.e.,

$$K_2 : G(x) = F(x - \Delta), \quad \Delta > 0.$$

If we know that F is normal, we use the two-sample t -test. Generally, the test statistic of any rank test is a function of the ordered ranks of the second sample. Theorem 3.2 and the following example I. show that the locally most powerful test generally has the critical region of the form

$$\sum_{i=m+1}^N a_N(R_i) \geq k;$$

hence the test criterion really depends only on the ordered ranks of Y_i 's. The scores $a_N(i) = \mathbb{E}\varphi(U_{:,i})$ (which could be approximated by $a_N(i) = \varphi\left(\frac{i}{N+1}\right)$), $i = 1, \dots, N$, are generated by an appropriate *score function* $\varphi : (0, 1) \mapsto \mathbb{R}^1$. We shall now describe three basic tests of this type which are the most often used in practice. Every one is locally most powerful for some special F , but the probabilities of the error of the first kind are the same for all $F \in H_0$.

(i) **Wilcoxon / Mann-Whitney test.** The Wilcoxon test has the critical region

$$W = \sum_{i=m+1}^N R_i \geq k_\alpha \quad (4.2)$$

i.e., the test function

$$\Phi(x) = \begin{cases} 1 & \dots W > k_\alpha \\ 0 & \dots W < k_\alpha \\ \gamma & \dots W = k_\alpha \end{cases}$$

where k_α is determined so that $P_{H_0}(W > k_\alpha) + \gamma P_{H_0}(W = k_\alpha) = \alpha$, $0 < \alpha < 1$ ($\alpha = 0.05$, $\alpha = 0.01$). This test is the locally most powerful against K_2 with F logistic with the density

$$f(x) = \frac{e^{-x}}{(1 + e^{-x})^2}, \quad x \in \mathbb{R}.$$

For small m and n , the critical value k_α could be directly determined: For each combination $s_1 < \dots < s_n$ of the numbers $1, \dots, N$ we calculate $\sum_{i=1}^n s_i$ and order these values in the increasing magnitude. The critical region is formed of the M_N largest sums where $M_N = \alpha \binom{N}{n}$; if there is no integer M_N satisfying this condition, we find the largest integer M_N less than $\alpha \binom{N}{n}$ and randomize the combination which leads to the $(M_N + 1)$ -st largest value. However, this systematic way, though precise, becomes difficult for large N , where we should use the tables of critical values.

There exist various tables of the Wilcoxon test, organized in various ways. Many tables provide the critical values of the Mann-Whitney's statistic

$$U_N = \sum_{i=m+1}^N \sum_{j=1}^m I[Z_i \geq Z_j];$$

we could easily see that U_N and W_N are in one-to-one relation $W_N = U_N + \frac{n(n+1)}{2}$.

For an application of the Wilcoxon test, we could alternatively use the *dual form* of the Wilcoxon statistic: Let $Z_1 < \dots < Z_{N:N}$ be the order statistics and define V_1, \dots, V_N in the following way:

$V_i = 0$ if $Z_{N:i}$ belongs to the 1st sample and $V_i = 1$ if $Z_{N:i}$ belongs to the second sample.

Then $W_N = \sum_{i=1}^N iV_i$.

For large m and n , where there are no tables, we use the *normal approximation* of W_N : If $m, n \rightarrow \infty$, then, under H_0 , W_N has asymptotically normal distribution in the following sense:

$$\lim_{m, n \rightarrow \infty} P_{H_0} \left\{ \frac{W_N - \mathbb{E}W_N}{\sqrt{\text{var } W_N}} < x \right\} = \Phi(x), \quad x \in \mathbb{R}^1, \quad (4.3)$$

where Φ is the standard normal distribution function.

To be able to use the normal approximation (4.3), we must know the expectation and variance of W_N under H_0 . The following theorem gives the expectation and the variance of a more general linear rank statistic, covering the Wilcoxon as well other rank tests.

THEOREM 4.1.1 *Let the random vector R_1, \dots, R_N have the discrete uniform distribution on the set \mathcal{R} of all permutations of numbers $1, \dots, N$, i.e. $\Pr(R = r) = \frac{1}{N!}$, $r \in \mathcal{R}$; let c_1, \dots, c_N and $a_1 = a(1), \dots, a_N = a(N)$ are arbitrary constants. Then the expectation and variance of the linear statistic*

$$S_N = \sum_{i=1}^N c_i a(R_i) \quad (4.4)$$

are

$$\mathbb{E}S_N = \frac{1}{N} \sum_{i=1}^N c_i \sum_{j=1}^N a_j \quad (4.5)$$

$$\text{var } S_N = \frac{1}{N-1} \sum_{i=1}^N (c_i - \bar{c})^2 \sum_{j=1}^N (a_j - \bar{a})^2, \quad (4.6)$$

where $\bar{c} = \frac{1}{N} \sum_{i=1}^N c_i$ and $\bar{a} = \frac{1}{N} \sum_{i=1}^N a_i$.

PROOF. Actually,

$$\begin{aligned} \mathbb{E}S_N &= \sum_{i=1}^N c_i \cdot \mathbb{E}a(R_i) = \sum_{i=1}^N c_i \frac{1}{N} \sum_{j=1}^N a_j, \\ \text{var } S_N &= \sum_{i=1}^N c_i^2 \cdot \text{var } a(R_i) + \sum_{i \neq j} c_i c_j \text{cov}(a(R_i), a(R_j)) \\ &= \text{var } a(R_1) \sum_{i=1}^N c_i^2 + \text{cov}(a(R_1), a(R_2)) \sum_{i \neq j} c_i c_j \\ &= N^2 \bar{c}^2 \cdot \text{cov}(a(R_1), a(R_2)) + \sum_{i=1}^N c_i^2. \end{aligned}$$

Theorem 2.3 further gives

$$\begin{aligned} \text{var } a(R_1) &= \frac{1}{N} \sum_{i=1}^N (a_i - \bar{a})^2, \\ \text{cov}(a(R_1), a(R_2)) &= -\frac{1}{N(N-1)} \sum_{i=1}^N (a_i - \bar{a})^2, \end{aligned}$$

hence

$$\text{var } S_N = -\frac{N}{N-1}\bar{c}^2 \sum_{i=1}^N (a_i - \bar{a})^2 + \frac{1}{N-1} \sum_{i=1}^N (a_i - \bar{a})^2 \cdot \sum_{j=1}^N c_j^2.$$

■

As a special case, we get the parameters of the Wilcoxon statistic under H_0 :

$$\mathbb{E}W_N = \frac{n(N+1)}{2}, \quad \text{var } W_N = \frac{mn(N+1)}{12}. \quad (4.7)$$

The tables of critical values profit from the fact that the distribution of W_N under H_0 is symmetric around $\mathbb{E}W_N$. If we test the H_0 against the left-sided alternative ($\Delta < 0$, the second sample shifted to the left with respect the first one), we reject H_0 if $W_N < 2\mathbb{E}W_N - k_\alpha$.

A sufficient condition for the symmetry of the linear rank statistic, which cover the Wilcoxon, follows from the following theorem:

THEOREM 4.1.2 *Let (R_1, \dots, R_N) be a random vector with discrete uniform distribution on the set \mathcal{R} of permutations of $1, \dots, N$. Let c_1, \dots, c_N and $a_1 = a(1), \dots, a_N = a(N)$ be constants such that either*

$$a_i + a_{N-i+1} = K = \text{const}, \quad i = 1, \dots, N \quad (4.8)$$

or

$$c_i + c_{N-i+1} = K = \text{const}, \quad i = 1, \dots, N. \quad (4.9)$$

Then the distribution of the statistic $S_N = \sum_{i=1}^N c_i a(R_i)$ is symmetric around $\mathbb{E}S_N$, i.e. $S_N - \mathbb{E}S_N$ and $-(S_N - \mathbb{E}S_N)$ have the same distributions.

PROOF. Under (4.8),

$2N\bar{a} = \sum_{i=1}^N a_i + \sum_{i=1}^N a_{N-i+1} = NK$, hence $a_i + a_{N-i+1} = 2\bar{a}$, $i = 1, \dots, N$. Because $(N - R_1 + 1, \dots, N - R_N + 1)$ and (R_1, \dots, R_N) have the same distributions, $S'_N = \sum_{i=1}^N c_i a(N - R_i + 1)$ has the same distribution as S_N and

$$S'_N = 2\bar{a} \sum_{i=1}^N c_i - S_N = 2\mathbb{E}S_N - S_N \implies S'_N - \mathbb{E}S'_N = \mathbb{E}S_N - S_N$$

$$\implies \Pr(S_N - \mathbb{E}S_N = s) = \Pr(S'_N - \mathbb{E}S'_N = s) = \Pr(\mathbb{E}S_N - S_N = s)$$

holds for any s .

Analogously, under (4.9), $c_i + c_{N-i+1} = 2\bar{c}$, $i = 1, \dots, N$ and (R_N, \dots, R_1) has the same distribution as (R_1, \dots, R_N) . Hence, $S''_N = \sum_{i=1}^N c_{N-i+1} a(R_i) = \sum_{i=1}^N c_i a(R_{N-i+1})$ has the same distribution as S_N and

$$S''_N = 2\bar{c} \sum_{i=1}^N a_i - S_N = 2\mathbb{E}S_N - S_N \quad S''_N - \mathbb{E}S_N = \mathbb{E}S_N - S_N.$$

The rest of the proof follows the steps of the first part. ■

(ii) **van der Waerden test.** Consider the test criterion (3.3) with the approximate scores (3.7) corresponding to the score function $\varphi(u) = \Phi^{-1}(u)$, $0 < u < 1$, where Φ is the standard normal distribution function. The van der Waerden test is convenient for testing H_0 against K_1 if the distribution function F has approximately normal tails. In fact, the test is asymptotically optimal for H_0 against the normal alternatives and its *relative asymptotic efficiency (Pitman efficiency)* with respect to the t -test is equal to 1 under normal F and ≥ 1 under all nonnormal F . For

these good properties the test could be recommended; for large m, n , if we do not have the tables at disposal, we could use the critical values of the test based on the normal approximation $N(\mathbb{E}S_N, \text{var } S_N)$ where in the van der Waerden case, by Theorem 4.1,

$$\mathbb{E}S_N = 0, \quad \text{var } S_N = \frac{mn}{N(N-1)} \sum_{i=1}^N \left[\Phi^{-1} \left(\frac{i}{N+1} \right) \right]^2.$$

Moreover, by Theorem 4.2, the distribution of S_N under H_0 is symmetric around 0.

. **(iii) Median test.** The median test is based on the criterion 3.3) with the scores (3.7) generated by the score function

$$\varphi(u) = \begin{cases} 0 & \dots 0 < u < \frac{1}{2} \\ \frac{1}{2} & \dots u = \frac{1}{2} \\ 1 & \dots \frac{1}{2} < u < 1. \end{cases} \quad (4.10)$$

The test statistic has a simple interpretation: Let μ be the median of the pooled sample $X_1, \dots, X_n, Y_1, \dots, Y_n$. The test statistic is equal to the number of Y -observations situated above μ , increased by $\frac{1}{2}$ for odd N .

If N is even, $M = N/2$ then, under H_0 , S_N has the hypergeometric probability distribution:

$$\Pr(S_N = k | H_0) = \begin{cases} \frac{\binom{M}{k} \binom{M}{n-k}}{\binom{N}{n}} & \dots \max(0, n-M) \leq k \leq \min(M, n) \\ 0 & \dots \text{otherwise.} \end{cases}$$

Actually, the vector (R'_{m+1}, \dots, R'_N) of ordered ranks of Y_1, \dots, Y_n in the pooled sample could be equal to any of $\binom{N}{n}$ combinations, each with the same probability. If $S_N = k$, then k elements in the combinations are greater than μ and $n-k$ elements are less than μ ; the number of such combinations is $\binom{M}{k} \binom{M}{n-k}$.

Hence, we could use the critical values from the tables of the hypergeometric distribution. For large number of observations we use the normal approximation with the parameters

$$\mathbb{E}S_N = n/2, \quad \text{var } S_N = \frac{mn}{4(N-1)}.$$

The median test is the most convenient for the heavy tailed F with the density f such that while $\lim_{x \rightarrow \pm\infty} f(x) = 0$, this convergence is much slower than in the case of the normal or logistic distributions (e.g., for the Cauchy distribution).

4.2 Two-sample rank tests of scale

Let X_1, \dots, X_m and Y_1, \dots, Y_n be two samples with the respective distribution functions $F(x - \mu)$ and $G(y - \mu)$, where μ is an unknown nuisance shift parameter. We wish to test the hypothesis of randomness, i.e. $H_0 : F \equiv G$, against the two-sample alternative of scale

$$K_4 : G(x - \mu) = F \left(\frac{x - \mu}{\sigma} \right) \quad \forall x \in \mathbb{R}^1, \quad \sigma > 1.$$

The locally most powerful rank test against K_4 is given by (3.9) and (3.10). However, instead of the tests optimal against some special shapes of F with complicated form of the scores, we shall rather describe tests with simple scores which are really used in the practice. Notice, by (3.10), that the score function φ_1 for the scale alternatives is not more monotone but U -shaped and the test statistics are of the form

$$S_N = \sum_{i=m+1}^N \varphi_1 \left(\frac{R_i}{N+1} \right). \quad (4.11)$$

(i) The Siegel-Tukey test. This test is based on reordering the observations, leading to new ranks, and to the test statistics whose distribution under H_0 is the same as that of the Wilcoxon statistic. Let $Z_{N:1} < Z_{N:2} < \dots < Z_{N:N}$ be the order statistics corresponding to the pooled sample of $N = m + n$ variables. Reorder this vector in the following way:

$$Z_{N:1}, Z_{N:N}, Z_{N:N-1}, Z_{N:2}, Z_{N:3}, Z_{N:N-2}, Z_{N:N-3}, Z_{N:4}, Z_{N:5}, \dots \quad (4.12)$$

and denote \tilde{R}_i the new rank of Z_i with respect to the order (4.12), $i = 1, \dots, N$. The critical region of the Siegel-Tukey test has the form

$$S'_N = \sum_{i=m+1}^N \tilde{R}_i \leq k'_\alpha$$

where k'_α is determined so that $P_{H_0}(S'_N < k'_\alpha) + \gamma P_{H_0}(S'_N = k'_\alpha) = \alpha$. The distribution of S'_N under H_0 coincides with the distribution of the Wilcoxon statistic, hence we could use the tables of the Wilcoxon test. However, unlike in the case of the Wilcoxon test, the Pitman efficiency of the Siegel-Tukey test to the F -test is rather low under normal F , namely $\frac{6}{\pi^2} \doteq 0.608$. Anyway, we should not use the two sample F -test of scale unless we are sure of the normality; namely this test is very sensitive to the deviations from the normal distribution.

(ii) Quartile test. Put in (4.11)

$$\varphi_1(u) = \begin{cases} 0 & \dots 0.25 < u < 0.75 \\ 0.5 & \dots u = 0.25, u = 0.75 \\ 1 & \dots 0 < u < 0.25 \text{ and } 0.75 < u < 1 \end{cases}$$

and we get the test statistic

$$S_N = \frac{1}{2} \sum_{i=m+1}^N \left[\text{sign} \left(\left| \frac{R_i}{N+1} - \frac{1}{2} \right| - \frac{1}{4} \right) + 1 \right] \quad (4.13)$$

and reject H_0 for large values of S_N . The value of S_N is, unless $N+1$ is divisible by 4, the number of observations of the Y -sample which belong either to the first or to the fourth quartile of the pooled sample.

If N is divisible by 4, then S_N has the hypergeometric distribution under H_0 , analogously as the median test.

4.3 Rank tests of H_0 against general two-sample alternatives based on the empirical distribution functions.

Again, X_1, \dots, X_m and Y_1, \dots, Y_n are two samples with the respective distribution functions F and G . We wish to test the hypotheses of randomness $H_0 : F \equiv G$ either against the one-sided alternative

$$K_5^+ : G(x) \leq F(x) \quad \forall x, \quad F \neq G$$

or against the general alternative

$$K_5 : F \neq G.$$

This case is not covered by Theorem 3.2; moreover, testing against K_5 is invariant to all continuous functions and there is no reasonable maximal invariant under this setup. In this case we usually use the tests based on the empirical distribution functions, which are the maximal likelihood estimators of the theoretical distribution functions in such nonparametric setup. Among these tests, we shall describe the Kolmogorov -Smirnov tests; another known test of this type is the Cramér - von Mises test.

The empirical distribution function \hat{F}_m corresponding to the sample X_1, \dots, X_m is defined as

$$\hat{F}_m(x) = \frac{1}{m} \sum_{i=1}^m I[X_i \leq x], \quad x \in \mathbb{R}^1; \quad (4.14)$$

analogously is defined the empirical d.f. \hat{G}_n for the sample Y_1, \dots, Y_n . Denote

$$D_{mn}^+ = \max_{x \in \mathbb{R}^1} [\hat{F}_m(x) - \hat{G}_n(x)] \quad (4.15)$$

$$D_{mn} = \max_{x \in \mathbb{R}^1} |\hat{F}_m(x) - \hat{G}_n(x)|.$$

The Kolmogorov-Smirnov test against K_5 has the test function $\Phi(\mathbf{X}, \mathbf{Y})$:

$$\Phi(\mathbf{X}, \mathbf{Y}) = \begin{cases} 1 & \dots D_{mn} > C_\alpha \\ \gamma & \dots D_{mn} = C_\alpha \\ 0 & \dots D_{mn} < C_\alpha \end{cases}$$

The statistic D_{mn} is the rank statistic, though not linear. To see this, consider the order statistics $Z_{N:1} < \dots < Z_{N:N}$ of the pooled sample and establish the indicators V_1, \dots, V_N where $V_j = 0$ if $Z_{N:j}$ comes from the X -sample and $Z_{N:j} = 1$ otherwise.

Because \hat{F}_m and \hat{G}_n are nondecreasing step functions, the maximum in (4.15) could be attained only in either of the points $Z_{N:1}, \dots, Z_{N:N}$; moreover

$$\hat{F}_m(Z_{N:j}) - \hat{G}_n(Z_{N:j}) = \frac{m+n}{mn} \left[j \frac{mn}{m+n} - V_1 - \dots - V_j \right], \quad j = 1, \dots, N$$

what gives the value of the test criterion

$$D_{mn} = \frac{m+n}{mn} \cdot \max_{1 \leq j \leq N} \left| j \frac{mn}{m+n} - V_1 - \dots - V_j \right|. \quad (4.16)$$

Notice that this expression depends only on V_1, \dots, V_N ; on the other hand, $V_i = 1 \iff$ one of the ranks R_{m+1}, \dots, R_N is equal to i , while $V_i = 0 \iff$ one of the ranks R_1, \dots, R_m is equal to i . Thus V_1, \dots, V_N are dependent only on the ranks, and so is also D_{mn} . This implies that the distribution of D_{mn} under H_0 is the same for all F . (4.16) is also used for the calculation of D_{mn} . Analogous consideration could be made for the one-sided Kolmogorov-Smirnov criterion D_{mn}^+ which could be expressed in the form

$$D_{mn}^+ = \frac{m+n}{mn} \cdot \max_{1 \leq j \leq N} \left[j \frac{mn}{m+n} - V_1 - \dots - V_j \right].$$

For large values m, n , we could use the limit critical values of the tests, but the asymptotic distributions of the criteria are not normal. More precisely, it holds

$$\lim_{m, n \rightarrow \infty} P_{H_0} \left\{ \left(\frac{mn}{m+n} \right)^{1/2} D_{mn}^+ \leq x \right\} = 1 - \exp\{-2x^2\}, \quad x > 0. \quad (4.17)$$

4.4 Modification of tests in the presence of ties

If both distribution functions F and G are continuous, then all observations are different with probability 1 and the ranks are well defined.

However, we round the observations to a finite number of decimal places and thus, in fact, we express all measurement on a countable network. In such case, the possibility of ties cannot be ignored and we should consider the possible modifications of rank tests for such situation. Let us first make several general remarks:

- If the tied observations belong to the same sample, then their mutual ordering does not affect the value of the test criterion. Hence, we should mainly consider the ties of observations from different samples.
- A small number of tied observations could be eventually omitted but this is paid by a loss of information.
- Some test statistics are well defined even in the presence of ties; the ties may only change the probabilities of errors of the 1st and 2nd kinds. Let us mention the Kolmogorov - Smirnov test as an example: The definitions of the empirical distribution function (4.14) and of the test criterion (4.16) make sense even in the presence of ties. However, if we use the tabulated critical values of the Kolmogorov - Smirnov test in this situation, the size of the critical region will be less than the prescribed significance level. Actually, we may then consider our observations $X_1, \dots, X_m, Y_1, \dots, Y_n$ as the data rounded from the continuous data $X_1^*, \dots, X_m^*, Y_1^*, \dots, Y_n^*$. Then the possible values of $\hat{F}_m(x) - \hat{G}_n(x)$, $x \in \mathbb{R}^1$ form a subset of possible values of $\hat{F}_m^*(x) - \hat{G}_n^*(x)$, $x \in \mathbb{R}^1$ where \hat{F}_m^* and \hat{G}_n^* are the empirical distribution functions of X_i^* 's and Y_j^* 's, respectively; hence

$$\max_{x \in \mathbb{R}^1} [\hat{F}_m(x) - \hat{G}_n(x)] \leq \max_{x \in \mathbb{R}^1} [\hat{F}_m^*(x) - \hat{G}_n^*(x)]$$

and similarly for the maxima of absolute values.

We shall describe two possible modifications of the rank tests in the presence of ties: *randomization* and *method of midranks*.

Randomization

Let Z_1, \dots, Z_N be the pooled sample. Take independent random variables U_1, \dots, U_N , uniformly $R(0, 1)$ distributed and independent of Z_1, \dots, Z_N . Order the pairs $(Z_1, U_1), \dots, (Z_N, U_N)$ in the following way:

$$(Z_i, U_i) < (Z_j, U_j) \iff \begin{cases} \text{either } Z_i < Z_j \\ \text{or } Z_i = Z_j \text{ and } U_i < U_j. \end{cases} \quad (4.18)$$

Denote R_1^*, \dots, R_n^* the ranks of the pairs $(Z_1, U_1), \dots, (Z_N, U_N)$. We shall say that Z_1, \dots, Z_N satisfy the hypothesis \bar{H} if they are independent and identically distributed (not necessarily with an absolutely continuous distribution). Then, under \bar{H} , the vector R_1^*, \dots, R_n^* is uniformly distributed over the set \mathcal{R} of permutations of $1, \dots, N$. We shall demonstrate it on an important special case when Z_1, \dots, Z_N could take on the equidistant values, e.g. when the data are rounded on k decimal places.

THEOREM 4.4.1 *Let Z_1, \dots, Z_N be random variables satisfying the hypothesis \bar{H} which take on the values from the set*

$$a + kd; \quad k = 0, \pm 1, \pm 2, \dots, \quad a \in \mathbb{R}^1, \quad d > 0.$$

Then the vector $R^ = (R_1^*, \dots, R_N^*)$ has the probability distribution*

$$\Pr(R^* = r) = \frac{1}{N!} \quad r \in \mathcal{R}.$$

PROOF. We may assume, without loss of generality, that Z_1, \dots, Z_N take on the integer values. Then the random variable $T_i = Z_i + U_i$ is equivalent to the pair (Z_i, U_i) , because $Z_i = [T_i]$ and $U_i = T_i - [T_i]$ with probability 1, $i = 1, \dots, N$. Because $\Pr(T_i = t) = 0 \quad \forall t \in \mathbb{R}^1$, the distribution function of T_i is continuous. Moreover,

$$(Z_i, U_i) < (Z_j, U_j) \iff T_i < T_j,$$

hence R_1^*, \dots, R_N^* are the ranks of the continuous random variables T_1, \dots, T_N satisfying the hypothesis of randomness H_0 ; this implies their distribution. ■

Method of midranks

The idea behind this method is that the equal observations should have equal ranks; the joint value of their rank is then taken as an average of all ranks of the group. We shall mainly describe this method on the Wilcoxon test, but it is applicable also to other tests.

Assume that there are e different values among N observations; among them, d_1 observations equal to the smallest value, d_2 observations equal to the second smallest value, etc., d_e observations equal to the largest value, $\sum_{i=1}^e d_i = N$. The average ranks of the individual groups are

$$\begin{aligned} v_1 = \dots = v_{d_1} &= \frac{1}{2}(d_1 + 1) \\ v_{d_1+1} = \dots = v_{d_1+d_2} &= d_1 + \frac{1}{2}(d_2 + 1) \\ v_{d_1+d_2+1} = \dots = v_{d_1+d_2+d_3} &= d_1 + d_2 + \frac{1}{2}(d_3 + 1) \\ &\dots\dots\dots \\ v_{d_1+d_2+\dots+d_{e-1}+1} = \dots &= d_1 + d_2 + \dots + d_{e-1} + \frac{1}{2}(d_e + 1). \end{aligned} \tag{4.19}$$

Let R'_1, \dots, R'_N denote the midranks of the observations Z_1, \dots, Z_N . We have the modified Wilcoxon statistic

$$W_N^* = \sum_{i=1}^N R'_i. \tag{4.20}$$

Because the distribution of (R'_1, \dots, R'_N) under \bar{H} is not more uniform on \mathcal{R} , (and the values may not be integer), we cannot use the standard tables of Wilcoxon critical values. If the numbers of equal observations are small comparing with N then we could use the normal approximation for sufficiently large m, n . To use this approximation, we must know the expectation and the variance of W_N^* under \bar{H} . These characteristics are conditional given the values d_1, \dots, d_e and hence the whole test is conditional. We have

$$\mathbb{E}(W_N^* | d_1, \dots, d_e) = n \frac{N + 1}{2} = \mathbb{E}W_N \tag{4.21}$$

and hence the expected value of the midrank statistic coincides with that of ordinary Wilcoxon statistic. Actually,

$$\begin{aligned} \mathbb{E}(W_N^*|d_1, \dots, d_e) &= \sum_{i=m+1}^N \mathbb{E}(R'_i|d_1, \dots, d_e) \\ &= \sum_{i=m+1}^N \frac{1}{N} \left[d_1 \frac{1}{2}(d_1 + 1) + d_2 \left(d_1 + \frac{1}{2}(d_2 + 1) \right) + \dots + d_e \left(d_1 + \dots + d_{e-1} + \frac{1}{2}d_e \right) \right] \\ &= \frac{1}{N} \sum_{i=m+1}^N (1 + \dots + N) = n \frac{N+1}{2}. \end{aligned}$$

The variance of W_N^* is equal

$$\text{var } W_n^* = \frac{mn(N+1)}{12} - \frac{mn \sum_{i=1}^e (d_i^3 - d_i)}{12N(N-1)}. \quad (4.22)$$

The first term in (4.22) is the variance of the standard Wilcoxon statistic, while the second term is a correction for the ties which vanishes if there are no ties among the observations.

Chapter 5

Tests for comparison of the treatments based on paired observations

When we want to compare the new treatment with the standard one, we wish to exclude the effects due to the inhomogeneity of the data. One possibility how to do that is to divide the experimental units in n pairs, as much homogeneous as possible, and apply the new treatment to one unit of the pair while the other unit serves for the control. We could also apply both treatments successively to the same unit.

Let Y_1, \dots, Y_N be the measurements of the effects of the new treatment and X_1, \dots, X_N be the control measurements. Then $(X_1, Y_1), \dots, (X_N, Y_N)$ could be considered as a random sample from a bivariate distribution with the distribution function $F(x, y)$; it is generally unknown and we only assume that it is continuous.

The hypothesis H_1 of no effect of the new treatment is equivalent to the statement that the distribution function $F(x, y)$ is symmetric around the straight line $y = x$, i.e.

$$H_1 : F(x, y) = F(y, x) \quad \forall x, y \in \mathbb{R}^1. \quad (5.1)$$

The alternative of the positive effect of the new treatment generally means that the distribution of the random vector (X, Y) is shifted toward the positive halfplane $y > x$. In this section we shall consider the rank tests of H_1 against various alternatives.

5.1 Rank tests of H_1

Apply the following transformation to (X_i, Y_i) , $i = 1, \dots, n$:

$$Z_i = Y_i - X_i, \quad W_i = X_i + Y_i, \quad i = 1, \dots, n \quad (5.2)$$

Under H_1 , the distribution of the vector $(Z_1, W_1), \dots, (Z_N, W_N)$ is symmetric around the w -axis, while under the alternative it is shifted in the direction of the positive half-axis z . The problem of testing H_1 against such alternative is invariant with respect to the transformations $z'_i = z_i$, $w'_i = g(w_i)$, $i = 1, \dots, n$, where g is a $1 : 1$ function with finite number of discontinuities. The vector (Z_1, \dots, Z_N) is the maximal invariant with respect to such transformations, hence the invariant tests will be only the functions of (Z_1, \dots, Z_N) , which forms a random sample from some one-dimensional distribution with a continuous distribution function D . The problem of testing H_1 is then equivalent to

$$H'_1 : D(z) + D(-z) = 1 \quad z \in \mathbb{R}^1 \quad (5.3)$$

stating that the distribution D is symmetric around 0, against the alternative

$$K'_1 : D(z + \Delta) + D(-z + \Delta) = 1 \quad \forall z \in \mathbb{R}^1, \quad \Delta > 0 \quad (5.4)$$

what means that the distribution is shifted in the direction of the positive z .

The distribution D is uniquely determined by the triple (p, F_1, F_2) with $p = \Pr(Z < 0)$, $F_1(z) = \Pr(|Z| < z | Z < 0)$ and $F_2(z) = \Pr(Z < z | Z > 0)$. Equivalent expressions for H'_1 and K'_1 are

$$H''_1 : p = 1/2, \quad F_2 = F_1, \quad p < 1/2, \quad F_2 \leq F_1.$$

This problem is invariant with respect to the group of transformations $G : z'_i = g(z_i)$, $i = 1, \dots, n$, where g is continuous, odd and increasing function. We could easily see that the maximal invariant with respect to G is $(S_1, \dots, S_m, R_1, \dots, R_n)$, where S_1, \dots, S_m are the ranks of the absolute values of negative Z 's among $|Z_1|, \dots, |Z_N|$ and R_1, \dots, R_n are the ranks of positive Z 's among $|Z_1|, \dots, |Z_N|$. Moreover, the vectors $S'_1 < \dots < S'_m$ and $R'_1 < \dots < R'_n$ of ordered ranks are sufficient for $(S_1, \dots, S_m, R_1, \dots, R_n)$ and, further, one of them uniquely determines the other; hence it is finally consider only, e.g., $R'_1 < \dots < R'_n$ and the invariant tests of H_1 [or of H'_1] depends only on $R'_1 < \dots < R'_n$.

Let ν be the number of positive components of (Z_1, \dots, Z_N) . Then ν is a binomial random variable $B(N, \pi)$; $\pi = 1/2$ under H_1 and, for any fixed n ,

$$P_{H_1}(R'_1 = r_1, \dots, R'_\nu = r_\nu, \nu = n) = \quad (5.5)$$

$$P_{H_1}(R'_1 = r_1, \dots, R'_\nu = r_\nu | \nu = n) P_{H_1}(\nu = n) = \frac{1}{\binom{N}{n}} \binom{N}{n} \left(\frac{1}{2}\right)^N = \left(\frac{1}{2}\right)^N$$

for any n -tuple (r_1, \dots, r_n) , $1 \leq r_1 < \dots < r_n \leq N$. The number of such tuples is $\sum_{n=0}^N \binom{N}{n} = \left(\frac{1}{2}\right)^N$. The critical region of any rank test of the size $\alpha = \frac{1}{2^N}$ contains just k such points (r_1, \dots, r_n) .

However, among such critical regions, there generally does not exist the uniformly most powerful one for H'_1 against K''_1 . We usually consider H''_1 against the alternative of *shift in location* under which (Z_1, \dots, Z_N) has the density $q_\Delta, \Delta > 0$:

$$q_\Delta(z_1, \dots, z_N) = \prod_{i=1}^N f(z_i - \Delta) : \Delta > 0 \quad (5.6)$$

where f is a one-dimensional symmetric density, $f(-x) = f(x)$, $x \in \mathbb{R}^1$. $\Delta = 0$ under H_1 [or H''_1]. The locally most powerful rank test of H_1 against (5.6) has the critical region

$$\sum_{i=1}^N a_N^+(R_i^+, f) \text{sign } Z_i \geq k_\alpha \quad (5.7)$$

where R_i^+ is the rank of $|Z_i|$ among $|Z_1|, \dots, |Z_N|$ and the scores $a_N^+(i, f)$ have the form

$$a_N^+(i, f) = \mathbb{E}\varphi^+(U_{(i)}, f), \quad i = 1, \dots, N \quad \varphi^+(u, f) = \varphi\left(\frac{u+1}{2}, f\right), \quad 0 < u < 1. \quad (5.8)$$

and where $\varphi(u, f) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}$, $0 < u < 1$. We shall describe two main tests of this type: the one sample Wilcoxon test and the sign test.

5.2 One-sample Wilcoxon test

The one-sample Wilcoxon test is based on the criterion

$$W_N^+ = \sum_{i=1}^n \text{sign } Z_i \cdot R_i^+ \quad (5.9)$$

where R_i^+ is the rank of $|Z_i|$ among $|Z_1|, \dots, |Z_N|$, or in the equivalent form

$$W_N^{++} = \sum_{i=1}^{\nu} R_i \quad (5.10)$$

where R_i is the rank of $Z_i > 0$ among $|Z_1|, \dots, |Z_N|$, ν is the number of positive components. Obviously $W_N^+ = 2W_N^{++} - \frac{1}{2}N(N+1)$.

We reject H_1 if $W_N^+ > C_\alpha$, i.e. if the test criterion exceeds the critical value. For large N , when the tables of critical values are not available, we may use the normal approximation:

$$P_{H_1} \left\{ \frac{W_N^+ - \mathbb{E}W_N^+}{\text{var } W_N^+} \leq x \right\} \rightarrow \infty \quad \text{as } N \rightarrow \infty \quad (5.11)$$

where

$$\mathbb{E}W_N^+ = 0, \quad \text{var } W_N^+ = \frac{1}{6}N(N+1)(2N+1) \quad (5.12)$$

The parameters (5.12) follow from the following proposition:

Proposition 5.2.1 *Let Z be a random variable with continuous distribution function symmetric around 0, i.e. $F(z) + F(-z) = 1$, $z \in \mathbb{R}^1$. Then Z and $\text{sign } Z$ are independent.*

PROOF. Obviously $P(\text{sign } Z = 1) = P(\text{sign } Z = -1) = 1/2$. Then $P(\text{sign } Z = 1, |Z| < z) = P(0 < Z < z) = P(-z < Z < 0) = P(\text{sign } Z = -1, |Z| < z) = \frac{1}{2}P(|Z| < z)$. ■ Similarly as the two-sample Wilcoxon test, the one-sample Wilcoxon test is convenient for the densities of logistic type.

5.3 Sign test

Consider a more general situation that Z_1, \dots, Z_N are independent random variables, Z_i distributed according to the distribution function D_i , and not necessarily all D_1, \dots, D_N are equal. This situation occurs when we compare two treatments under different experimental conditions or using different methods.

Under this situation we want to verify the hypothesis

$$H_1^* : D_i(z) + D_i(-z) = 1, \quad z \in \mathbb{R}^1, \quad i = 1, \dots, N \quad (5.13)$$

of symmetry of all distributions around 0, against the alternative that all distributions are shifted toward the positive values.

Such problem is invariant with respect to all transformations of the type $z'_i = f_i(z_i)$, $i = 1, \dots, N$, where f_i 's are continuous, increasing and odd functions. The maximal invariant with respect to such transformations is the number n of positive components. The uniformly most powerful invariant test (most powerful among the tests dependent only on n) has the form

$$\Phi(n) = \begin{cases} 1 & \dots n > C_\alpha \\ \gamma & \dots n = C_\alpha \\ 0 & \dots n < C_\alpha \end{cases} \quad (5.14)$$

where C_α and γ are determined by the equation

$$\sum_{n > C_\alpha} \binom{N}{n} \left(\frac{1}{2}\right)^N + \gamma \binom{N}{C_\alpha} \left(\frac{1}{2}\right)^N = \alpha. \quad (5.15)$$

The criterion of the *sign test* is simply the number of positive components among Z_1, \dots, Z_N and its distribution under H_1 is binomial $b(N, 1/2)$. For large N we could again use the normal approximation.

If all distribution functions D_1, \dots, D_N coincide, the sign test is the locally most powerful rank test of H_1 for D of the double-exponential type with the density $d(z) = \frac{1}{2}e^{-|z-\Delta|}$, $z \in \mathbb{R}^1$.

When we want to use the rank test, we need not to know the exact values X_i, Y_i , $i = 1, \dots, N$; it is sufficient to know the signs of the differences $Y_i - X_i$. This is a very convenient property of the sign test: we could use this test even for the qualitative observations of the type: "drogue A gives a better pain relief than drogue B". As a matter of fact, we do not have any better test under such conditions.

Chapter 6

Tests of independence in bivariate population

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a random sample from a bivariate distribution with a continuous distribution function $F(x, y)$. We want to test the hypothesis of independence

$$H_2 : F(x, y) = F_1(x)F_2(y) \quad (6.1)$$

where F_1 and F_2 are arbitrary distribution functions. The most natural alternative for H_2 is the positive [or negative] dependence, but it is too wide and we could hardly expect to find a uniformly most powerful test against such alternative. Instead of it we consider the alternative

$$\begin{aligned} X_i &= X_i^0 + \Delta Z_i \\ Y_i &= Y_i^0 + \Delta Z_i \end{aligned} \quad \Delta > 0, \quad i = 1, \dots, n, \quad (6.2)$$

where $X_i^0, Y_i^0, Z_i, i = 1, \dots, n$ are independent and their distributions are independent of i . The independence then means that $\Delta = 0$.

Let R_1, \dots, R_n be the ranks of X_1, \dots, X_n and let S_1, \dots, S_n be the ranks of Y_1, \dots, Y_n , respectively. Under the hypothesis of independence, the vectors (R_1, \dots, R_n) and (S_1, \dots, S_n) are independent and both have the uniform distribution on the set \mathcal{R} of permutations of $1, \dots, n$.

The locally most rank powerful test of H_2 against the alternative K_2 in which X_i^0 has the density f_1 and Y_i^0 the density f_2 , respectively, both densities continuously differentiable, has the critical region

$$\sum_{i=1}^n a_n(R_i, f_1) a_n(R_i, f_2) > C_\alpha \quad (6.3)$$

with the scores $a_n(i, f)$ given in (3.4), which are usually replaced by approximate scores (3.7).

We shall briefly describe two the most well-known rank tests of independence.

6.1 Spearman test

The Spearman test is based on the correlation coefficient of (R_1, \dots, R_n) and (S_1, \dots, S_n) :

$$r_{S} = \frac{\frac{1}{n} \sum_{i=1}^n R_i S_i - \bar{R} \bar{S}}{\left[\frac{1}{n} \sum_{i=1}^n (R_i - \bar{R})^2 \frac{1}{n} \sum_{i=1}^n (S_i - \bar{S})^2 \right]^{1/2}} \quad (6.4)$$

where

$$\bar{R} = \bar{S} = \frac{n+1}{2},$$

$$\frac{1}{n} \sum_{i=1}^n (R_i - \bar{R})^2 = \frac{1}{n} \sum_{i=1}^n (S_i - \bar{S})^2 = \frac{1}{n} \sum_{i=1}^n i^2 - \left(\frac{n+1}{2} \right)^2 = \frac{n^2 - 1}{12}. \quad (6.5)$$

Then we could express (6.4) in the simpler form

$$r_S = \frac{12}{n(n^2 - 1)} \sum_{i=1}^n R_i S_i - \frac{3(n+1)}{n-1}. \quad (6.6)$$

The Spearman test rejects H_2 if $r_S > C_\alpha$, or, equivalently, if $\mathcal{S} = \sum_{i=1}^n R_i S_i > C_\alpha^*$. In some tables we find the critical values for the statistic

$$\mathcal{S}' = \sum_{i=1}^n (R_i - S_i)^2 \quad (6.7)$$

for which $r_S = 1 - \frac{6}{n^3 - n} \mathcal{S}'$. The test based on \mathcal{S}' rejects H_2 if $\mathcal{S}' < C'_\alpha$.

For large n we use the normal approximation with

$$\mathbb{E}\mathcal{S} = \frac{n(n+1)^2}{4}, \quad \text{var } \mathcal{S} = \frac{n^2(n+1)^2(n-1)}{144}.$$

The Spearman test is the locally most powerful against the alternatives of the logistic type.

6.2 Quadrant test

This test is based on the criterion

$$Q = \frac{1}{4} \sum_{i=1}^n [\text{sign}(R_i - \frac{n+1}{2}) + 1][\text{sign}(S_i - \frac{n+1}{2}) + 1] \quad (6.8)$$

and rejects H_2 for large values of Q . For even n is Q equal to the number of pairs (X_i, Y_i) , for which X_i lies above the X -median and Y_i lies above the Y -median. Statistic Q then has, under the hypothesis H_2 , the *hypergeometric distribution*

$$\Pr(Q = q) = \frac{\binom{m}{q} \binom{m}{m-q}}{\binom{n}{m}} \quad (6.9)$$

for $q = 0, 1, \dots, m$, $m = n/2$. For large n we use the normal approximation with the parameters

$$\mathbb{E}Q = n/4, \quad \text{var } Q = \frac{n^2}{16(n-1)}. \quad (6.10)$$

Chapter 7

Rank test for comparison of several treatments

7.1 One-way classification

We want to compare the effects of p treatments; the experiment is organized in such a way that the i -th treatment is applied on n_i subjects with the results x_{i1}, \dots, x_{in_i} , $i = 1, \dots, p$, $\sum_{i=1}^p n_i = n$. Then x_{i1}, \dots, x_{in_i} is a random sample from a distribution with a distribution function F_i , $i = 1, \dots, p$. The hypothesis of no difference between the treatments could be then expressed as the hypothesis of equality of p distribution functions, namely

$$\mathbf{H}_2 : F_1 \equiv F_2 \equiv \dots \equiv F_p \quad (7.1)$$

and we could consider this hypothesis either against the general alternative

$$\mathbf{K}_2 : F_i(x) \neq F_j(x) \quad (7.2)$$

at least for one pair i, j at least for some $x = x_0$,
or against a more special alternative

$$\mathbf{K}'_2 : F_i(x) = F(x - \Delta_i), \quad i = 1, \dots, p \quad (7.3)$$

and $\Delta_i \neq \Delta_j$ at least for one pair i, j .

The alternative (7.3) claims that the effects of treatments on the values of observations are linear and that at least two treatments differ in their effects.

The classical test for this situation is the F -test of the variance analysis; this test works well if we could assume that $F_i \sim \mathcal{N}(\mu + \alpha_i, \sigma^2)$, $i = 1, \dots, p$. We obtain the usual model of variance analysis

$$X_{ij} = \mu + \alpha_i + e_{ij}, \quad j = 1, \dots, n_i; \quad i = 1, \dots, p \quad (7.4)$$

where e_{ij} are independent random variables with the normal distribution $\mathcal{N}(0, \sigma^2)$. The hypothesis \mathbf{H}_2 could be then reformulated as

$$\mathbf{H}'_2 : \alpha_1 = \alpha_2 = \dots = \alpha_p = 0.$$

The F -test rejects the hypothesis \mathbf{H}'_2 provided

$$\mathcal{F} = \frac{n-p}{p-1} \frac{\sum_{i=1}^p n_i (\bar{X}_i - \bar{X}_{..})^2}{\sum_{i=1}^p \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2} \geq C_\alpha \quad (7.5)$$

where

$$\bar{X}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \text{ and } \bar{X}_{\cdot\cdot} = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} X_{ij},$$

$i = 1, \dots, p$ and where the critical value C_α is found in the tables of F -distribution with $(p-1, n-p)$ degrees of freedom.

7.2 Kruskal-Wallis rank test

Let us order all observations

$$x_{11}, \dots, x_{1n_1}, x_{21}, \dots, x_{2n_2}, \dots, x_{p1}, \dots, x_{pn_p}$$

according to the increasing magnitude. Let R_{i1}, \dots, R_{in_i} be the ranks of the observations x_{i1}, \dots, x_{in_i} . Let $R_{i1}^* < \dots < R_{in_i}^*$ be the same observations ordered in increasing magnitude. Then, under the hypothesis \mathbf{H}_2 ,

$$P_{\mathbf{H}_2} \left(R_{11}^* = r_{11}, \dots, R_{1n_1}^* = r_{1n_1}, \dots, R_{p1}^* = r_{p1}, \dots, R_{pn_p}^* = r_{pn_p} \right) = \frac{n_1! \dots n_p!}{n!} \quad (7.6)$$

for any permutation $(r_{11}, \dots, r_{1n_1}, \dots, r_{p1}, \dots, r_{pn_p})$ of numbers $1, \dots, n$ such that $r_{i1} < \dots < r_{in_i}$ for all $i = 1, \dots, p$.

Denote

$$\begin{aligned} R_{i\cdot} &= \frac{1}{n_i} \sum_{j=1}^{n_i} R_{ij}, \quad i = 1, \dots, p \\ R_{\cdot\cdot} &= \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} R_{ij} = \frac{n+1}{2}. \end{aligned} \quad (7.7)$$

If we replace $\bar{X}_{i\cdot}$ and $\bar{X}_{\cdot\cdot}$ in (7.5) by $R_{i\cdot}$ and $R_{\cdot\cdot}$, respectively, $i = 1, \dots, p$, we obtain

$$\frac{n-p}{p-1} \frac{\sum_{i=1}^p n_i (R_{i\cdot} - R_{\cdot\cdot})^2}{\sum_{i=1}^p \sum_{j=1}^{n_i} (R_{ij} - R_{i\cdot})^2}$$

and this is proportional to the criterion of the *Kruskal-Wallis test*,

$$\begin{aligned} \mathcal{K} &= \frac{12}{n(n+1)} \sum_{i=1}^p n_i \left(R_{i\cdot} - \frac{n+1}{2} \right)^2 \\ &= \frac{12}{n(n+1)} \sum_{i=1}^p n_i R_{i\cdot}^2 - 3(n+1). \end{aligned} \quad (7.8)$$

In the special case $p = 2$, the Kruskal-Wallis reduces to the two-sided (two-sample) Wilcoxon test. We reject the hypothesis \mathbf{H}_2 , provided $\mathcal{K} \geq C_\alpha$, where the critical value C_α is either obtained from special tables, or, if $p > 3$ and $n_i \geq 5$, $i = 1, \dots, p$, we use the asymptotic critical values:

It could be shown that, under \mathbf{H}_2 and for large n_1, \dots, n_p , the criterion \mathcal{K} has asymptotically χ^2 distribution with $p-1$ degrees of freedom.

Remark. In case of ties between the observations we replace the ranks by the midranks, similarly as in the case of the Wilcoxon test.

7.3 Two-way classification (random blocks)

We want to compare p treatments, but simultaneously we want to reduce the influence of non-homogeneity of the sample units. Then we could organize the experiment in such a way that we divide the subjects in n homogeneous groups, so called *blocks*, and compare the effects of treatments within each block separately. The subjects in the block are usually assigned the treatments in a random way.

Let us consider the simplest of these models with n blocks, each containing p elements, and each treatment is applied just once in each block. We assume that the blocks are independent of each other.

The observations could be formally described by the following table:

<i>Block</i>	<i>Treatment</i>	1	2	3	...	p
1		x_{11}	x_{12}	x_{13}	...	x_{1p}
2		x_{21}	x_{22}	x_{23}	...	x_{2p}
⋮		⋮	⋮	⋮		⋮
n		x_{n1}	x_{n2}	x_{n3}	...	x_{np}

The observation x_{ij} is the measured effect of the j -th treatment applied in the i -th block.

We assume that X_{ij} are independent random variables and X_{ij} has a continuous distribution function F_{ij} , $j = 1, \dots, n$; $i = 1, \dots, p$. We wish to verify the hypothesis that there is no significant difference among the treatments, hence

$$\mathbf{H}_3 : F_{i1} \equiv F_{i2} \equiv \dots \equiv F_{ip} \quad \forall i = 1, \dots, n \quad (7.9)$$

against the alternative

$$\mathbf{K}_3 : F_{ij} \neq F_{ik} \quad (7.10)$$

at least for one i and at least for one pair j, k ,

or against a more special alternative

$$\mathbf{K}'_3 : F_{ij}(x) = F_i(x - \Delta_j), \quad j = 1, \dots, n; \quad i = 1, \dots, p \quad (7.11)$$

$$\Delta_j \neq \Delta_k \quad \text{at least for one pair } j, k. \quad (7.12)$$

The classical test of \mathbf{H}_3 is the F -test, corresponding to the model

$$X_{ij} = \mu + \alpha_i + \beta_j + E_{ij}, \quad j = 1, \dots, n; \quad i = 1, \dots, p, \quad (7.13)$$

where E_{ij} are independent random variables with the normal distribution $\mathcal{N}(0, \sigma^2)$, μ is the main additive effect, α_i is the effect if the i -th block and β_j is the effect of the j -th treatment, $j = 1, \dots, n$; $i = 1, \dots, p$. The hypothesis \mathbf{H}_3 then reduces to the form

$$\beta_1 = \beta_2 = \dots = \beta_p.$$

The critical region of the F -test of \mathbf{H}_3 has the form

$$\mathcal{F} = \frac{(n-1) \sum_{j=1}^p (\bar{X}_{.j} - \bar{X}_{..})^2}{\sum_{j=1}^p \sum_{i=1}^n (X_{ij} - \bar{X}_{i.})^2} > C_\alpha \quad (7.14)$$

where C_α is the critical value of F -distribution with $p-1$ and $(p-1)(n-1)$ degrees of freedom.

Friedman rank test

Order the observations within each block and denote the corresponding ranks R_{i1}, \dots, R_{ip} ; $i = 1, \dots, n$. The ranks we arrange in the following table:

<i>Block</i>	<i>Treatment</i>	1	2	3	...	p	Row average
1		R_{11}	R_{12}	R_{13}	...	R_{1p}	$\frac{p+1}{2}$
2		R_{21}	R_{22}	R_{23}	...	R_{2p}	$\frac{p+1}{2}$
⋮		⋮	⋮	⋮		⋮	⋮
n		R_{n1}	R_{n2}	R_{n3}	...	R_{np}	$\frac{p+1}{2}$
Column average		$R_{.1}$	$R_{.2}$	$R_{.3}$...	$\mathbb{R}_{.p}$	Overall average $R_{..} = \frac{p+1}{2}$

where $R_{.j} = \frac{1}{n} \sum_{i=1}^n R_{ij}$ and $R_{..} = \frac{1}{np} \sum_{i=1}^n \sum_{j=1}^p R_{ij}$.

The *Friedman test* is based on the following criterion:

$$\begin{aligned}
 Q_n &= \frac{12n}{p(p+1)} \sum_{j=1}^p \left(R_{.j} - \frac{p+1}{n} \right)^2 \\
 &= \frac{12n}{p(p+1)} \sum_{j=1}^p R_{.j}^2 - 3n(p+1)
 \end{aligned} \tag{7.15}$$

while the large value of the criterion are significant.

As $n \rightarrow \infty$, then the distribution of Q_n is approximately χ^2 with $p-1$ degrees of freedom.

In case $p = 2$, the Friedman test is reduced to the two-sided sign test.

The Friedman test is applicable for the comparison of p treatments even in the situation that we observe only the ranks rather than exact values of the treatment effects.

Bibliography

- [1] H. Bünning and G. Trenkler (1978). *Nichparametrische statistische Methoden (mit Aufgaben und Lösungen und einem Tabellenanhang)*. W. de Gruyter, Berlin.
- [2] J.Hájek and Z.Šidák (1967). *Theory of Rank Tests*. Academia, Prague.
- [3] J. Jurečková (1981). *Rank Tests* (in Czech). Lecture Notes, Charles University in Prague.
- [4] J. P. Lecoutre and P. Tassi (1987). *Statistique non paramétrique et robustesse*. Economica, Paris.
- [5] E. L. Lehmann (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco.
- [6] E. L. Lehmann (1986). *Testing Statistical Hypotheses*. Second Edition. J. Wiley.