STATISTICAL TESTS BASED ON RANKS

Jana Jurečková

Charles University, Prague

1. Parametric and nonparametric models

Example 1. Model of measurement.
Let $\mathbf{X} = (X_1, \ldots, X_n)$ measurements of some physical entity μ. If we admit random fluctuations, then we consider the model

$$X_i = \mu + e_i, \quad i = 1, \ldots, n.$$

What can we assume about the vector of errors $e_1, \ldots, e_n$? We can assume that
(1) The distribution of vector $(e_1, \ldots, e_n)$ is independent of μ.
(2) *Moreover,* $e_1, \ldots, e_n$ are independent.
(3) *Moreover,* $e_1, \ldots, e_n$ are identically distributed.
(3) *Moreover,* the distribution of e_1 has a density, symmetric about 0.
(4) *Moreover,* the distribution of e_1 is normal $\mathcal{N}(0, \sigma^2)$ with unknown σ.
(5) *Moreover,* σ is even known.
If we assume 1–5, then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an

efficient estimate of μ. But often we are not sure of the normal distribution, and even the assumption 3 may be unrealistic, if e.g. μ is the length of the volume.

Example 2. Comparison of two treatments.
Let $X_1, \ldots, X_m$ be the blood pressure of m patients after an application of some medicament and $Y_1, \ldots, Y_n$ be the blood pressure of the control group, which received a placebo. Let F and G be the respective distribution functions of X and Y.
We wish to test the hypothesis $\mathbf{H}: \ F \equiv G$ (no effect). But the test depends on the alternative under consideration:

(1) F and G are absolutely continuous, otherwise unknown, and the medicament has reduces the blood pressure, i.e.

$$\mathbf{K}_1: \ G(z) \leq F(z) \ \forall z, \ G(z_0) \leq F(z_0)$$

(Y is stochastically larger than X).

(2) Moreover, $\mathbf{K}_2: \ G(z) = F(z - \Delta) \ \forall z$ with some $\Delta > 0$ (the alternative of shift in location).

(3) $F \sim \mathcal{N}(\mu_1, \sigma_1^2),\ G \sim \mathcal{N}(\mu_2, \sigma_2^2)$, where $\mu_1, \mu_2, \sigma_1, \sigma_2$ are unknown, $\mathbf{K}_3 :\ \mu_1 < \mu_2$, where generally $\sigma_1 \neq \sigma_2$.

(4) $F \sim \mathcal{N}(\mu_1, \sigma^2),\ G \sim \mathcal{N}(\mu_2, \sigma^2)$, where μ_1, μ_2, σ are unknown, $\mathbf{K}_4 :\ \mu_1 < \mu_2$.

We would use the t-test against $\mathbf{K}_4$; testing $\mathbf{H}$ against $\mathbf{K}_3$ is known as the Behrens-Fisher problem. We would use the rank tests for $\mathbf{H}$ against $\mathbf{K}_1$ and $\mathbf{K}_2$.

2. Practical problems which we can solve with the aid of rank tests or tests based on generalized ranks:

(1) *Two sample tests of equality of two treatments effects against alternatives of shift in location or scale.*

(Wilcoxon, van der Waerden, median rank tests; Siegel-Tukey and quartile rank tests). Some of them tests we shall describe later in detail.

(2) *Two sample tests of equality of two treatments effects against general one sided or two sided alternatives.*

(Kolmogorov-Smirnov tests).

(3) *Tests of equality of effects of several treatments* (Kruskal-Wallis rank test).

(4) *Tests of equality of effects of several treatments on observations organized in blocks* (Friedman rank test).

(5) *Tests of equality of effects of several treatments on observations categorized in contingency tables* (Kruskal-Wallis test with midranks).

(6) *Tests of equality of effects of two treatments based on paired observations* (signed-rank tests: one-sample Wilcoxon, van der Waerden, sign test).

(7) *Tests of independence in bivariate population* (Spearman rank correlation coefficient, Kendall's tau, quadrant test).

For most of these cases, there exists also a **permutation test**, based on the order statistics.

(8) *Tests of hypothesis* $\mathbf{H}: \ \boldsymbol{\beta} = \mathbf{0}$ *or more generally* $\mathbf{H}: \ \mathbf{A}\boldsymbol{\beta} = \mathbf{b}$ *in the linear regression*

model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$.

(9) *Tests of hypothesis on some components of* $\boldsymbol{\beta}$ *in the linear regression model, with the other components nuisance, without a neccessity to estimate the nuisance parameter* (tests based on so called **regression rank scores**).

(10) *Tests on the parameters of the linear autoregressive time series model.* The nuisance parameters are either estimated (alligned rank tests) or the tests are based on the **autoregression rank scores**. Especially, tests on the order of the autoregression.

(11) *Tests of independence of two autoregressive time series* (based on autoregression rank scores) - often desired in practice, but there were no reasonable tests untill recently.

3. Nonparametric hypotheses and tests

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be the vector of observations. The hypothesis and alternative $\mathbf{H}$ and $\mathbf{K}$ are two disjoint sets of probability distributions of $\mathbf{X}$. The hypothesis is usually the homogeneous, symmetric, independent, while the alternative means an inhomogeneity, asymmetry, dependence etc.

Every rule, which assigns just one of the decisions "to accept $\mathbf{H}$" or "to reject $\mathbf{H}$" to any point $\mathbf{x} = (x_1, \ldots, x_n)$, is called *the test* (non-randomized) of hypothesis $\mathbf{H}$ against alternative $\mathbf{K}$. 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 $\mathbf{H}$ if $\mathbf{x} \in A_K$ and accepts $\mathbf{H}$ if $\mathbf{x} \in A_H$.

To simplify the structure of the tests, we supplement the family of tests by the *randomized tests.* A randomized test rejects $\mathbf{H}$ 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.

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

(1) We reject $\mathbf{H}$ even if it is correct *(error of the first kind)*;

(2) we accept $\mathbf{H}$ even if it is incorrect *(error of the second kind).*

If $\mathbf{X}$ has distribution P, then the test Φ rejects $\mathbf{H}$ with the probability

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

The probability $\beta_\Phi(Q) = \mathbf{E}_Q(\Phi(\mathbf{X})),\ Q \in K$, is called *the power* of the test Φ against the alternative Q and 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 $\mathbf{K}$ and has the small probability (smaller than a prescribed α) of the error of the first kind for all distributions from the hypothesis $\mathbf{H}$.

The criterion of optimality for tests:
Select a small number α, $0 < \alpha < 1$, called *the significance level,* and among all tests satisfying

$$\beta_\Phi(P) \le \alpha \quad \forall P \in H$$

we look for the test satisfying

$$\beta_\Phi(Q) := \max \quad \forall Q \in \mathbf{K}.$$

Such test, if it exists, is called *the uniformly most powerful test of size* $\le \alpha$, briefly *the uniformly most powerful* α-test of $\mathbf{H}$ against $\mathbf{K}$.
Simple hypothesis [alternative] means that $\mathbf{H}$ [$\mathbf{K}$] is 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.

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 $\mathbf{H} : \{P\}$ against the simple alternative $\mathbf{K} : \{Q\}$, there exists the test Φ and a constant k such that

$$\mathbf{E}_P(\Phi(\mathbf{X})) = \alpha \tag{1}$$

and

$$\Phi(\mathbf{x}) = \begin{cases} 1 & \text{if } q(\mathbf{x}) > k.p(\mathbf{x}) \\ 0 & \text{if } q(\mathbf{x}) < k.p(\mathbf{x}). \end{cases} \qquad (2)$$

This test is the most powerful α-test of $\mathbf{H}$ against $\mathbf{K}$.

4. Invariant tests

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

$$\mathbf{X} \text{ satisfies } \mathbf{H} \textit{ iff } g\mathbf{X} \text{ satisfies } \mathbf{H}$$
$$\mathbf{X} \text{ satisfies } \mathbf{K} \textit{ iff } g\mathbf{X} \text{ satisfies } \mathbf{K}.$$

If the problem of testing $\mathbf{H}$ against $\mathbf{K}$ is invariant with respect to the group $\mathcal{G}$ of transformations of $\mathcal{X}$ onto $\mathcal{X}$, then we naturally consider only the invariant tests, which satisfy

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

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

The test Φ is invariant with respect to $\mathcal{G}$ if and only if it is a function of the maximal invariant.

Examples of maximal invariants

(1) Let G be the set of $n!$ permutations of $x_1, \ldots, x_n$. Then the vector ordered components of $\mathbf{x}$ *(vector of order statistics)*

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

is the maximal invariant with respect to G.

(2) Let G be the set of transformations $x_i' =$

$f(x_i),\ i = 1, \ldots, n)$ such that $f : \mathbf{R}^1 \mapsto \mathbf{R}^1$ 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 $x_1, \ldots, x_n$, i.e. $R_i = \sum_{j=1}^{n} I[x_j \leq x_i],\ i = 1, \ldots, n$. Then $T(\mathbf{x}) = (R_1, \ldots, R_n)$ 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 $R_1, \ldots, R_n$. Put $f(x_i) = x_i',\ i = 1, \ldots, n$ and let f be linear on the intervals $[x_{n:1}, x_{n:2}], \ldots, [x_{n:n-1}, x_{n:n}]$; 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.

5. Properties of ranks and of order statistics

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be the vector of observations; denote $X_{n:1} \leq X_{n:2} \ldots \leq X_{n:n}$ the components of $\mathbf{X}$ ordered according to increasing magnitude. The vector $\mathbf{X}_{(.)} = (X_{n:1}, \dots, X_{n:n})$ is called the *vector of order statistics* and $X_{n:i}$ is called the ith *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 $\mathbf{R}$ of ranks of $\mathbf{X}$ takes on the values in the set $\mathcal{R}$ of $n!$ permutations $(r_1, \dots, r_n)$ of (1,. . . , n).

The distribution of $\mathbf{X}_{(.)}$ *and of* R :

If $\mathbf{X}$ has density $p_n(x_1, \dots, x_n)$, then the vector $\mathbf{X}_{(.)}$ of order statistics has the distribution with the density

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

We say that the random vector $\mathbf{X}$ satisfies the **hypothesis of randomness** $\mathbf{H}_0$, if it has a probability distribution with density of the form

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

where f is an arbitrary one-dimensional density. Otherwise speaking, $\mathbf{X}$ satisfies the hypothesis of randomness provided its components are a random sample from an absolutely continuous distribution.

If $\mathbf{X}$ satisfies the hypothesis of randomness $\mathbf{H}_0$, then $\mathbf{X}_{(.)}$ and $\mathbf{R}$ are independent, the vector of ranks $\mathbf{R}$ has the uniform discrete distribution

$$\Pr(\mathbf{R}=r)=\frac{1}{n!},\ r\in\mathcal{R} \tag{3}$$

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

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

Marginal distributions of the random vectors $\mathbf{R}$ and $\mathbf{X}_{(.)}$ under $\mathbf{H}_0$:

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

(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) $\mathbf{E}R_i=\frac{n+1}{2},\ i=1,\ldots,n.$

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

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

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

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

where $F(x)$ is the distribution function of $X_1, \ldots, X_n$.

6. Locally most powerful rank tests

We want to test a hypothesis of randomness $\mathbf{H}_0$ on the distribution of $\mathbf{X}$. The rank test is characterized by test function $\Phi(\mathbf{R})$. The most powerful rank $\alpha-$test of $\mathbf{H}_0$ against a simple alternative $\mathbf{K} : \{Q\}$ [that $\mathbf{X}$ has the fixed distribution Q] follows directly from the Neyman-Pearson Lemma:

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

where k_α and γ are determined so that

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

If we want to test against a composite alternative and the uniformly most powerful rank tests do not exist, then we look for a rank test, *most powerful locally* in a neighborhood of the hypothesis.

Definition. Let $d(Q)$ be a measure of distance of alternative $Q \in K$ from the hypothesis $\mathbf{H}$. The $\alpha-$test Φ_0 is called the *locally most powerful* in the class $\mathcal{M}$ of $\alpha-$tests of $\mathbf{H}$ against $\mathbf{K}$ 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 $0 < d(Q) < \varepsilon$.

7. The structure of the locally most powerful rank tests of H_0 :

Let A be a class of densities, $A = \{g(x,\theta) : \theta \in \mathcal{J}\}$ such that
$\mathcal{J} \subset \mathbb{R}^1$ is an open interval, $\mathcal{J} \ni 0$.
$g(x,\theta)$ is absolutely continuous in θ for almost all x.
For almost all x, there exists the limit

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

and

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

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

$$q_\Delta(x_1,\dots,x_n) = \prod_{i=1}^{n} g(x_i, \Delta c_i),$$

$c_1,\dots,c_n$ 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 $\mathbf{H}_0$ against $\mathbf{K}$ 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 $\mathbf{H}_0$,

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

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 $g(x,0)$.

8. Special cases:

I. *Alternative of the shift in location*:
$\mathbf{K}_1 : \{q_\Delta : \Delta > 0\}$ where

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

where f is a fixed absolute continuous density such that $\int_{-\infty}^{\infty} |f'(x)|dx < \infty$. Then the locally most powerful rank $\alpha-$test of $\mathbf{H}_0$ against $\mathbf{K}$ has the critical region

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

where k satisfies the condition
$P(\sum_{i=m+1}^{N} a_N(R_i,f) \geq k) = \alpha,\ P \in \mathbf{H}_0$ and

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

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

$$a_N(i,f) = \mathbf{E}\varphi(U_{N:i}, f),\ i = 1,\ldots,N$$

where $\varphi(u,f) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))},\ 0 < u < 1$

and $U_{N:1},\ldots,U_{N:N}$ are the order statistics corresponding to the sample of size N from the uniform $R(0,1)$ distribution. The scores can be also expressed in the form

$$a_N(i,f) =$$
$$N\begin{pmatrix} N-1 \\ i-1 \end{pmatrix} \int_{-\infty}^{\infty} f'(x)F^{i-1}(x)(1-F(x))^{N-i}dx.$$

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

$$a_N(i,f) = \varphi\left(\frac{i}{N+1}\right) = \varphi(\mathbf{E}U_{N:i}, f),$$

$i = 1,\ldots,N$. The asymptotic critical values coincide for both types of scores.

II. *Alternative of two samples differing by scales:* $\mathbf{K}_2 : \{q_\Delta : \Delta > 0\}$ where

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

where density f satisfies $\int_{-\infty}^{\infty} |xf'(x)|dx < \infty$ and μ is the nuisance parameter. Then the locally most powerful test has the critical region

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

where k is determined by the condition $P(\sum_{i=m+1}^{N} a_{1N}(R_i, f) \geq k) = \alpha,\ P \in \mathbf{H}_0$ and the scores have the form

$$\begin{aligned} a_{1N}(i, g) &= \mathbf{E}\left\{-1 - X_{N:i}\frac{f'(X_{N:i})}{f(X_{N:i})}\right\} \\ &= \mathbf{E}\varphi_1(U_{N:i}, f),\ i = 1, \ldots, N, \end{aligned}$$

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 can replace the scores by the approximate scores $a_{1N}(i, f) = \varphi_1\left(\frac{i}{N+1}, f\right), i = 1, \ldots, N.$

III. *Alternative of simple regression:*
$\mathbf{K}_3 = \{q_\Delta : \Delta > 0\}$ where
$q_\Delta(x_1, \ldots, x_N) = \prod_{i=1}^N f(x_i - \Delta c_i)$ with a fixed absolutely continuous density f and with given constants $c_1, \ldots, c_N$, $\sum_{i=1}^N c_i^2 > 0$. Then the locally most powerful test has the critical region $\sum_{i=1}^N c_i a_N(R_i, f) \geq k$ with the the same scores as in I. and with k determined by the condition $P(\sum_{i=1}^N c_i a_N(R_i, f) \geq k) = \alpha$.

9. Selected two-sample rank tests

Denote $(X_1, \ldots, X_m, Y_1, \ldots, Y_n) = (Z_1, \ldots, Z_N)$ with $N = m+n$, where $(X_1, \ldots, X_m)$ has distribution function F and $(Y_1, \ldots, Y_n)$ has distribution function G.
Consider testing $\mathbf{H}_0 : \ F \equiv G$ against the alternative
$\mathbf{K}_1 : \ G(x) \leq F(x) \ \forall x \in \mathbf{R}^1, \ G(x) \neq F(x)$ at least for one x.
$\mathbf{K}_1$ is a one-sided alternative stating that the random variable Y is *stochastically larger* than X.

The problem of testing $\mathbf{H}_0$ against $\mathbf{K}_1$ is invariant to the group $\mathcal{G}$ of transformations $z_i' = g(z_i),\ i = 1, \ldots, N$ where g is any continuous strictly increasing function, with the vector of ranks $R_1, \ldots, R_N$ of $Z_1, \ldots, Z_N$ as the maximal invariant. The class of invariant tests thus coincides with that of rank tests.

Because both $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$ are random samples, the distribution of the vector of ranks $(R_1, \ldots, R_m, R_{m+1}, \ldots, R_{m+n})$ is symmetric in the first m and the last n arguments. Hence, the vectors of ordered ranks

$R_1' < \ldots < R_m'$ and $R_{m+1}' < ... < R_{m+n}'$

are sufficient. Because either of these vectors determines the other, the family of invariant tests of $\mathbf{H}_0$ against $\mathbf{K}_1$ reduces to the tests dependent only on the ordered ranks of one of the samples, e.g. on the ordered ranks of $Y_1, \ldots, Y_n$.

Vector $R_{m+1}', ..., R_N'$ runs over $\binom{N}{n}$ combinations. All these combinations are equally probable under $\mathbf{H}_0$ and hence the critical region of each rank test of the size $\alpha = k/\binom{N}{n}$ consists of just k points $s_1, \ldots, s_n$, $1 \leq s_1 < \ldots < s_n \leq N$. The rank tests mutually differ in the points included in the critical regions.

The above alternative $\mathbf{K}_1$ is still to rich and hence there does not exist the uniformly most powerful rank test of $\mathbf{H}_0$ against $\mathbf{K}_1$. However, we are able to find rank tests locally most powerful for $\mathbf{H}_0$ against some important subsets of $\mathbf{K}_1$.

11. Two-sample tests of location

Consider the special alternative of $\mathbf{K}_1$, namely that G differ from F by a shift in location, i.e.,

$$\mathbf{K}_2: \ G(x) = F(x - \Delta), \ \Delta > 0.$$

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. 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 $a_N(i) = \mathbb{E}\varphi(U_{N:i})$ (or approximate $a_N(i) = \varphi\left(\frac{i}{N+1}\right)$), $i = 1, \ldots, N$, are generated by an appropriate *score function* $\varphi : (0,1) \mapsto \mathbf{R}^1$.

Three basic tests of this type the most often used in practice:

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

$$W = \sum_{i=m+1}^{N} R_i \geq k_\alpha \tag{4}$$

i.e., the test function

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

where k_α is determined so that
$P_{H_0}(W > k_\alpha) + \gamma P_{H_0}(W = k_\alpha) = \alpha$,
($\alpha = 0.05,\ \alpha = 0.01$). This test is the locally most powerful against $\mathbf{K}_2$ with F logistic with the density

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

For small m and n, the critical value k_α can be directly determined: For each combination $s_1 < \ldots < s_n$ of the numbers $1, \ldots, N$ 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 $M_N = \alpha \begin{pmatrix} N \\ n \end{pmatrix}$; if there is no integer M_N satisfying this condition, we find the largest integer M_N less than $\alpha \begin{pmatrix} N \\ n \end{pmatrix}$ 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 can 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 can alternatively use the *dual form* of the Wilcoxon statistic: Let $Z_1 < \ldots < Z_{N:N}$ be the order statistics and define $V_1, \ldots, V_N$ 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 $m, n \to \infty$, then, under $\mathbf{H}_0$, W_N has asymptotically normal distribution in the following sense:

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

where Φ is the standard normal distribution function.

To be able to use the normal approximation, we must know the expectation and variance of W_N under $\mathbf{H}_0$. 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. Let the random vector $R_1, \ldots, R_N$ have the discrete uniform distribution on the set R of all permutations of numbers $1, \ldots, N$, i.e. $\Pr(R = r) = \frac{1}{N!}$, $r \in \mathcal{R}$; let $c_1, \ldots, c_N$ and $a_1 = a(1), \ldots, a_N = a(N)$ are arbitrary constants. Then the expectation and variance of the linear statistic $S_N = \sum_{i=1}^{N} c_i a(R_i)$ are

$$\mathbf{E} S_N = \frac{1}{N} \sum_{i=1}^{N} c_i \sum_{j=1}^{N} a_j$$

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

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

Parameters of the Wilcoxon statistic under $\mathbf{H}_0$:

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

The distribution of W_N under $\mathbf{H}_0$ is symmetric around $\mathbf{E} W_N$. If we test the $\mathbf{H}_0$ against the left-sided alternative ($\Delta < 0$, the second sample shifted to the left with respect the first one), we reject $\mathbf{H}_0$ if $W_N < 2\mathbb{E}W_N - k_\alpha$.

(ii) van der Waerden test. Consider the approximate scores 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 $\mathbf{H}_0$ against $\mathbf{K}_1$ if the distribution function F has approximately normal tails. In fact, the test is asymptotically optimal for $\mathbf{H}_0$ 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 can be recommended; for large m, n, if we do not have the tables at disposal, we can 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, the distribution of S_N under $\mathbf{H}_0$ is symmetric around 0.

(iii) Median test. The median test uses the scores generated by the score function

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

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 $\mathbf{H}_0$, S_N has the hypergeometric probability distribution:

$$\Pr(S_N = k|H_0) =$$

$$= \begin{cases} \dfrac{\begin{pmatrix} M \\ k \end{pmatrix}\begin{pmatrix} M \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} \\ \ldots \max(0, n-M) \leq k \leq \min(M, n) \\ 0 \ldots \quad \text{otherwise.} \end{cases}$$

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

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

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

12. Two-sample rank tests of scale

Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ 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. $\mathbf{H}_0 : F \equiv G$, against the two-sample alternative of scale

$$\mathbf{K}_4 : \; G(x-\mu) = F\left(\frac{x-\mu}{\sigma}\right) \; \forall x \in \mathbf{R}^1, \; \sigma > 1.$$

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. The score function φ_1 for the scale alternatives is $U-$shaped and the test statistics are of the form

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

(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 $\mathbf{H}_0$ is the same as that of the Wilcoxon statistic. Let $Z_{N:1} < Z_{N:2} < \ldots < Z_{N:N}$ be the order statistics corresponding to the pooled sample of $N = m + n$ variables. Re-order this vector in the following way:

$$Z_{N:1}, Z_{N:N}, Z_{N:N-1}, Z_{N:2}, Z_{N:3},$$
$$Z_{N:N-2}, Z_{N:N-3}, Z_{N:4}, Z_{N:5}, \ldots$$

and denote $\tilde{R}_i$ the new rank of Z_i with respect to the new order $i = 1, \ldots, N$. 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 $k'\alpha$ 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 $\mathbf{H}_0$ coincides with the distribution of the Wilcoxon statistic, hence we can use the tables of the Wilcoxon test.

(ii) Quartile test is based on the scoree function

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

and we get the test statistic

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

and reject $\mathbf{H}_0$ 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 $\mathbf{H}_0$, analogously as the median test.

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

Again, $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are two samples with the respective distribution functions

F and G. We wish to test the hypotheses of randomness $\mathbf{H}_0: \; F \equiv G$ either against the one-sided alternative

$$\mathbf{K}_5^+: \; G(x) \le F(x) \; \forall x, \; F \ne G$$

or against the general alternative

$$\mathbf{K}_5: \; F \ne G.$$

Testing against $\mathbf{K}_5$ 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. We shall describe the Kolmogorov -Smirnov tests; another known test of this type is the Cramér - von Mises test.

The empirical distribution function $\widehat{F}_m$ corresponding to the sample $X_1, \ldots, X_m$ is defined as

$$\widehat{F}_m(x) = \frac{1}{m} \sum_{i=1}^{m} I[X_i \le x], \; x \in \mathbf{R}^1;$$

analogously is defined the empirical d.f. $\widehat{G}_n$ for the sample $Y_1, \ldots, Y_n$. Denote

$$D^+_{mn} = \max_{x \in \mathbf{R}^1} [\widehat{F}_m(x) - \widehat{G}_n(x)]$$

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

The Kolmogorov-Smirnov test against $\mathbf{K}_5$ has the test function $\Phi(\mathbf{X}, \mathbf{Y})$:

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

The statistic D_{mn} is the rank statistic, though not linear. To see this, consider the order statistics $Z_{N:1} < \ldots < Z_{N:N}$ of the pooled sample and establish the indicators $V_1, \ldots, V_N$ where $V_j = 0$ if $Z_{N:j}$ comes from the $X-$sample and $Z_{N:j} = 1$ otherwise.
Because $\widehat{F}_m$ and $\widehat{G}_n$ are nondecreasing step functions, the maximum can be attained only in either of the points $Z_{N:1}, \ldots, Z_{N:N}$; moreover

$$\widehat{F}_m(Z_{N:j}) - \widehat{G}_n(Z_{N:j})$$

$$= \frac{m+n}{mn} \left[j \frac{mn}{m+n} - V_1 - \ldots - V_j \right], \; j = 1, \ldots, N$$

what gives the value of the test criterion

$$D_{mn}$$
$$= \frac{m+n}{mn} . \max_{1 \le j \le N} \left| j \frac{mn}{m+n} - V_1 - \ldots - V_j \right| .$$

Notice that this expression depends only on $V_1, \ldots, V_N$; on the other hand, $V_i = 1 \Longleftrightarrow$ one of the ranks $R_{m+1}, \ldots, R_N$ is equal to i, while $V_i = 0 \Longleftrightarrow$ one of the ranks $R_1, \ldots, R_m$ is equal to i. Thus $V_1, \ldots, V_N$ are dependent only on the ranks, and so is also D_{mn}. This implies that the distribution of D_mn under $\mathbf{H}_0$ is the same for all F. This expression is also used for the calculation of D_{mn}. Analogous consideration holds for the one-sided Kolmogorov-Smirnov criterion D^+_{mn} which can be expressed in the form

$$D^+_{mn} =$$
$$= \frac{m+n}{mn} . \max_{1 \le j \le N} \left[j \frac{mn}{m+n} - V_1 - \ldots - V_j \right] .$$

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

$$\lim_{m,n \to \infty} P_{H_0} \left\{ \left(\frac{mn}{m+n} \right)^{1/2} D^+_{mn} \le x \right\} =$$
$$= 1 - \exp\{-2x^2\}, \ x > 0.$$

14. 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 can 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 and of the test criterion 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 $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ as the data rounded from the continuous data
$X_1^*, \ldots, X_m^*, Y_1^*, \ldots, Y_n^*$. Then the possible values of $\widehat{F}_m(x) - \widehat{G}_n(x),\ x \in \mathbf{R}^1$ form a subset of possible values of $\widehat{F}_m^*(x) - \widehat{G}_n^*(x),\ x \in \mathbf{R}^1$ where $\widehat{F}_m^*$ and $\widehat{G}_n^*$ are the empirical distribution functions of X_i^*'s and Y_j^*'s, respectively; hence

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

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.*

15. Randomization

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

$$(Z_i, U_i) < (Z_j, U_j) \Longleftrightarrow$$

$$\begin{cases} \text{either} & Z_i < Z_j \\ \text{or} & Z_i = Z_j \text{ and } U_i < U_j. \end{cases}$$

Denote $R_1^*, \ldots, R_n^*$ the ranks of the pairs $(Z_1, U_1), \ldots, (Z_N, U_N)$. We shall say that $Z_1, \ldots, Z_N$ 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 $R_1^*, \ldots, R_n^*$ is uniformly distributed over the set $\mathcal{R}$ of permutations of $1, \ldots, N$.

16. 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

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

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

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

Because the distribution of $(R'_1, \ldots, R'_N)$ 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 can 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 $d_1, \ldots, d_e$ and hence the whole test is conditional. We have

$$\mathbf{E}(W_N^* | d_1, \ldots, d_e) = n\frac{N+1}{2} = \mathbf{E}W_N$$

The variance of W_N^* is equal

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

The first term 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.

17. Comparison of two treatments based on paired observations

To exclude the effects due to the inhomogeneity of the data, we can divide the experimental units in n homogeneous pairs, and apply the new treatment to one unit of the pair while the other unit serves for the control. We can also apply both treatments successively to the same unit.

Let $Y_1, \ldots, Y_N$ be the measurements of the effects of the new treatment and $X_1, \ldots, X_N$ be the control measurements. Then $(X_1, Y_1), \ldots, (X_N, Y_N)$ is a random sample from a bivariate distribution with the distribution function $F(x, y)$; it is generally unknown and assumed being continuous.

The hypothesis $\mathbf{H}_1$ 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.

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

Under the alternative of a positive effect of the new treatment, the distribution of the random vector (X, Y) is shifted toward the positive halfplane $y > x$.

Rank tests of $\mathbf{H}_1$

Transform $(X_i, Y_i),\ i = 1, \ldots, n$ in the following way:

$$Z_i = Y_i - X_i,\ W_i = X_i + Y_i,\ i = 1, \ldots, n.$$

Under $\mathbf{H}_1$, the distribution of the vector $(Z_1, W_1), \ldots, (Z_N, W_N)$ is symmetric around the $w-$axis, while under the alternative it is shifted in the direction of the positive half-axis z. The problem is invariant with respect to the transformations $z_i' = z_i,\ w_i' = g(w_i),\ i = 1, \ldots, n$, where g is a $1:1$ function with finite number of discontinuities. The invariant tests depend only on $(Z_1, \ldots, Z_N)$, because it is the maximal invariant. It is a sample from some one-dimensional distribution with a continuous distribution function D. The problem of testing $\mathbf{H}_1$ is then equivalent to stating that the distribution D is symmetric around 0,

$$H_1' : \ D(z) + D(-z) = 1 \quad z \in \mathbb{R}^1$$

against the alternative that the distribution is shifted in the direction of the positive z,

$$K_1' : \quad D(z+\Delta)+D(-z+\Delta) = 1\ \forall z \in \mathbb{R}^1,\ \Delta > 0$$

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

$$\mathbf{H}_1'' : p = 1/2,\ F_2 = F_1,\ \mathbf{K}_1'' : p < 1/2,\ F_2 \leq F_1.$$

This problem is invariant with respect to the transformations $G: \quad z_i' = g(z_i),\ i = 1, \ldots, n$, where g is continuous, odd and increasing function. The maximal invariant is
$(S_1, \ldots, S_m, R_1, \ldots, R_n)$, where $S_1, \ldots, S_m$ are the ranks of the absolute values of negative Z's among $|Z_1|, \ldots, |Z_N|$ and $R_1, \ldots, n$ are the ranks of positive Z's among $|Z_1|, \ldots, |Z_N|$. Moreover, the vectors $S_1' < \ldots < S_m'$ and $R_1' < \ldots < R_n'$ of ordered ranks are sufficient for $(S_1, \ldots, S_m, R_1, \ldots, R_n)$ and, further, one of them uniguely determines the other; hence it is finally consider only, e.g., $R_1' < \ldots < R_n'$ and the invariant tests of $\mathbf{H}_1$ [or of $\mathbf{H}_1'$] depends only on $R_1' < \ldots < R_n'$.

Let ν be the number of positive components of $(Z_1, \ldots, Z_N)$. Then ν is a binomial random variable $B(N, \pi)$; $\pi = 1/2$ under $\mathbf{H}_1$ and, for any fixed n,

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

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

However, there generally is no uniformly most powerful test for $\mathbf{H}''_1$ against $\mathbf{K}''_1$. We usually consider the alternative of *shift in location* that $(Z_1, \ldots, Z_N)$ has the density $q_\Delta, \Delta > 0$:

$$q_\Delta(z_1, \ldots, z_N) = \prod_{i=1}^{N} f(z_i - \Delta) : \Delta > 0 \qquad (5)$$

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

$$\sum_{i=1}^{N} a_N^+(R_i^+, f) \text{sign } Z_i \geq k_\alpha \qquad (6)$$

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

$$a_N^+(i, f) = \mathbb{E}\varphi^+(U_{(i)}, f), \ i = 1, \ldots, N$$

$$\varphi^+(u, f) = \varphi(\frac{u+1}{2}, f), \ 0 < u < 1$$

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.

One-sample Wilcoxon test

The one-sample Wilcoxon test is based on the criterion

$$W_N^+ = \sum_{i=1}^{n} \text{sign } Z_i . R_i^+ \qquad (7)$$

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

$$W_N^{++} = \sum_{i=1}^{\nu} R_i \qquad (8)$$

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

We reject $\mathbf{H}_1$ 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:

$$P_{H_1}\left\{\frac{W_N^+ - \mathbf{E}W_N^+}{\text{var } W_N^+} \le x\right\} \to \infty \quad \text{as} \quad N \to \infty \tag{9}$$

where

$$\mathbf{E}W_N^+ = 0, \qquad \text{var } W_N^+ = \frac{1}{6}N(N+1)(2N+1) \tag{10}$$

The parameters follow from the following proposition:

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

The one-sample Wilcoxon test is convenient for the densities of logistic type.

Sign test

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

We want to test the hypothesis of symmetry of all distributions around 0, against the alternative that all distributions are shifted toward the positive values:

$$H_1^*: \; D_i(z) + D_i(-z) = 1, \quad z \in \mathbf{R}^1, \; i = 1, \ldots, N$$

The problem is invariant with respect to all transformations $z_i' = f_i(z_i), \; i = 1, \ldots, N$, where f_i's are continuous, increasing and odd functions. The maximal invariant is the number n of positive components. The invariant tests depend only on n, and the uniformly most powerful among them has the form

$$\Phi(n) = \begin{cases} 1 & \ldots n > C_\alpha \\ \gamma & \ldots n = C_\alpha \\ 0 & \ldots n < C_\alpha \end{cases} \tag{11}$$

where C_α and γ are determined by the equation

$$\sum_{n > C_\alpha} \binom{N}{n} \left(\frac{1}{2}\right)^N + \gamma \binom{N}{C_\alpha} \left(\frac{1}{2}\right)^N = \alpha. \tag{12}$$

The criterion of the *sign test* is simply the number of positive components among $Z_1, \ldots, Z_N$

and its distribution under $\mathbf{H}_1$ is binomial $b(N, 1/2)$. For large N we can again use the normal approximation.

If all distribution functions $D_1, \ldots, D_N$ coincide, the sign test is the locally most powerful rank test of $\mathbf{H}_1$ for double-exponential D with density $d(z) = \frac{1}{2}\mathrm{e}^{-|z-\Delta|}, \; z \in \mathbb{R}^1$. For using the rank test we need not to know the exact values $X_i, Y_i, \; i = 1, \ldots, N$; it is sufficient to know the signs of the differences $Y_i - X_i$. This is a very convenient property: we can 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.

18. Tests of independence in bivariate population

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

$$H_2: \; F(x, y) = F_1(x)F_2(y) \qquad (13)$$

where F_1 and F_2 are arbitrary distribution functions. The most natural alternative for $\mathbf{H}_2$ is the positive [or negative] dependence, but it is too wide and there is no uniformly most powerful test. We rather consider the alternative

$$\begin{array}{l} X_i = X_i^0 + \Delta Z_i \\ Y_i = Y_i^0 + \Delta Z_i \end{array} \quad \Delta > 0, \ i = 1, \dots, n, \qquad (14)$$

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 $\mathbf{H}_2$ against the alternative $\mathbf{K}_2$ in which X_i^0 has the density f_1 and Y_i^0 the density f_2, has the critical region

$$\sum_{i=1}^{n} a_n(R_i, f_1) a_n(R_i, f_2) > C_\alpha \qquad (15)$$

where the scores $a_n(i, f)$ are usually replaced be approximate scores.

Two the most well-known rank tests of independence.

Spearman test

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

$$r_S = \frac{\frac{1}{n}\sum_{i=1}^n R_i S_i - \bar{R}\bar{S}}{[\frac{1}{n}\sum_{i=1}^n (R_i - \bar{R})^2[\frac{1}{n}\sum_{i=1}^n (S_i - \bar{S})^2]^{1/2}}$$

where

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

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

Then we can express the criterion in a simpler form

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

The test rejects $\mathbf{H}_2$ 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

$$\mathcal{S}' = \sum_{i=1}^n (R_i - S_i)^2 \qquad (16)$$

for which $r_S = 1 - \frac{6}{n^3-n}\mathcal{S}'$. The test based on $\mathcal{S}'$ rejects $\mathbf{H}_2$ if $\mathcal{S}' < C'_\alpha$.

For large n we use the normal approximation with

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

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

Quadrant test

This test is based on the criterion

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

and rejects $\mathbf{H}_2$ 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 $\mathbf{H}_2$, the *hypergeometric distribution*

$$\Pr(Q = q) = \frac{\binom{m}{q}\binom{m}{m-q}}{\binom{n}{m}} \qquad (17)$$

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

$$\mathbb{E}Q = n/4, \qquad \text{var } Q = \frac{n^2}{16(n-1)}. \qquad (18)$$

19. Rank tests for comparison of several treatments
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 $x_{i1}, \ldots, x_{in_i}$, $i = 1, \ldots, p$, $\sum_{i=1}^{p} n_i = n$. Then $x_{i1}, \ldots, x_{in_i}$ is a random sample from a distribution with a distribution function F_i, $i = 1, \ldots, p$. The hypothesis of no difference between the treatments can be then expressed as the hypothesis of equality of p distribution functions, namely

$$\mathbf{H}_2: \ F_1 \equiv F_2 \equiv \ldots \equiv F_p \tag{19}$$

and we can consider this hypothesis either against the general alternative

$$\mathbf{K}_2: \ F_i(x) \neq F_j(x) \tag{20}$$

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

$$\mathbf{K}_2': \ F_i(x) = F(x - \Delta_i), \ i = 1, \ldots, p \tag{21}$$

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

The alternative 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 under the normality, $F_i \sim \mathcal{N}(\mu+\alpha_i, \sigma^2)$, $i = 1, \ldots, p$. We obtain the usual model of variance analysis

$$X_{ij} = \mu + \alpha_i + e_{ij}, \ j = 1, \ldots, n_i; \ i = 1, \ldots, p \tag{22}$$

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

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

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

$$\mathcal{F} = \frac{n-p}{p-1} \frac{\sum_{i=1}^{p} n_i (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2}{\sum_{i=1}^{p} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i\cdot})^2} \geq C_\alpha \quad (23)$$

where

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

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

Kruskal-Wallis rank test

Consider the vector of all observations and their ranks

$$R_{11}, \ldots, R_{1n_1}; R_{21}, \ldots, R_{2n_2}; \ldots; R_{p1}, \ldots, R_{pn_p}.$$

Let $R'_{i1} < \ldots < R'_{in_i}$ be the ordered ranks of the i-th sample, $i = 1, \ldots, p$. Then, under $\mathbf{H}$, it holds for any permutation $\{r_{11}, \ldots, r_{pn_p}\}$ of

$1, \ldots, N$ such that $r_{i1} < \ldots < r_{in_i}$, $i = 1, \ldots, p$, that

$$\mathbf{P}\left(R'_{11} = r_{11}, \ldots, R'_{pn_p} = r_{pn_p}\right) = \frac{n_1! \ldots n_p!}{N!}.$$

The Kruskal-Wallis rank test rejects $\mathbf{H}$ provided

$$K_N = \frac{12}{N(N+1)} \sum_{i=1}^{p} n_i (R_{i\cdot} - \frac{N+1}{2})^2$$
$$= \frac{12}{N(N+1)} \sum_{i=1}^{p} n_i R_{i\cdot}^2 - 3(N+1) > \mathcal{K}_\alpha$$

where

$$R_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} R_{ij}, \ i = 1, \ldots, p.$$

It can be formally obtained from the F-test, if we insert $R_{i\cdot}$ for $\bar{X}_{i\cdot}$ and $R_{\cdot\cdot} = \frac{N+1}{2}$ for $\bar{X}_{\cdot\cdot}$

If $n_i \to \infty$, $i = 1, \ldots, p$ and $p > 3$, then K_N has asymptotically $\chi^2(p-1)$ distribution under $\mathbf{H}_2$. In practice we can use the χ^2 approximation for $p > 3$ and $n_i > 5, i = 1, \ldots, p$.

In the special case $p = 2$, the Kruskal-Wallis reduces to the two-sided (two-sample) Wilcoxon test.

Modification in presence of tied observations:

Assume that there are e different values among the components $X_{11}, \ldots, X_{pn_p}$, and d_1 are equal to the smallest, ..., d_e are equal to the largest one. Let $(R^*_{11}, \ldots, R^*_{pn_p})$ be the midranks of $X_{11}, \ldots, X_{pn_p}$. Then the modified Kruskal-Wallis statistic has the form

$$K^*_N = \frac{\frac{12}{N(N+1)} \sum_{i=1}^{p} n_i \left(R^*_{i\cdot} - \frac{N+1}{2} \right)^2}{1 - \frac{1}{N^3 - N} \sum_{k=1}^{e} (d_k^3 - d_k)}.$$

The distribution of K^*_N conditioned by given $d_1, \ldots, d_e$ is approximately $\chi^2(p-1)$ under $\mathbf{H}$ for large $n_1 \ldots, n_p$. In the special case $p = 2$, the Kruskal-Wallis reduces to the two-sided (two-sample) Wilcoxon test.

Two-way classification (random blocks)

We want to compare p treatments and simultaneously to reduce the effect of non-homoheneity of the sample units. 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. The simplest model has n independent blocks, each containing p elements, and each treatment is applied just once in each block.

The observations can be formally described by the following table:

Treatm. *Block*	1	2	3	...	p
1	x_{11}	x_{12}	x_{13}	$\cdots$	x_{1p}
2	x_{21}	x_{22}	x_{23}	$\cdots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
n	x_{n1}	x_{n2}	x_{n3}	$\cdots$	x_{np}

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

X_{ij} are independent, X_{ij} has a continuous distribution function F_{ij}, $j = 1, \ldots, n$; $i = 1, \ldots, p$. We test the hypothesis that there is no significant difference among the treaments, hence

$$\mathbf{H}_3 : F_{i1} \equiv F_{i2} \equiv \ldots \equiv F_{ip} \; \forall i = 1, \ldots, n \quad (24)$$

against the alternative

$$\mathbf{K}_3 : F_{ij} \neq F_{ik} \quad (25)$$

at least for one i and at least for one pair j, k, or against a more special alternative

$$\mathbf{K}_3' : \qquad F_{ij}(x) = F_i(x - \Delta_j),$$
$$j = 1, \ldots, n; \; i = 1, \ldots, p$$
$$\Delta_j \neq \Delta_k \text{ at least for one pair } j, k.$$

The classical test of $\mathbf{H}_3$ is the F-test in the model

$$X_{ij} = \mu + \alpha_i + \beta_j + E_{ij}, \; j = 1, \ldots, n; \; i = 1, \ldots, p, \tag{26}$$

where E_{ij} are independent 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, $j = 1, \ldots, n; \; i = 1, \ldots, p$. The hypothesis $\mathbf{H}_3$ then reduces to the form

$$\beta_1 = \beta_2 = \ldots = \beta_p.$$

critical regionThe F-test of $\mathbf{H}_3$ has the critical region

$$\mathcal{F} = \frac{(n-1)\sum_{j=1}^{p}(\bar{X}_{\cdot j} - \bar{X}_{\cdot\cdot})^2}{\sum_{j=1}^{p}\sum_{i=1}^{n}(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}_{\cdot\cdot})^2} > C_\alpha \tag{27}$$

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 $R_{i1}, \ldots, R_{ip}$; $i = 1, \ldots, n$. The ranks we arrange in the following table:

Treatm. *Block*	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}	$\frac{p+1}{2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
n	R_{n1}	R_{n2}	R_{n3}	...	R_{np}	$\frac{p+1}{2}$
Column average	$R_{\cdot 1}$	$R_{\cdot 2}$	$R_{\cdot 3}$	...	$\mathbf{R}_{\cdot p}$	Average $R_{\cdot\cdot} = \frac{p+1}{2}$

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

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

$$Q_n = \frac{12n}{p(p+1)} \sum_{j=1}^{p} (R_{.j} - \frac{p+1}{n})^2$$
$$= \frac{12n}{p(p+1)} \sum_{j=1}^{p} R_{.j}^{p} - 3n(p+1)$$

and the large value of the criterion are significant. As $n \to \infty$, 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 even in the situation that we observe only the ranks rather than exact values of the treatment effects.

PERMUTATION TESTS

Permutation tests are conditional tests under given vector of order stagtistics. We shall illustrate them on the test of hypothesis of randomness against a two-sample alternative.
Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two independent samples with the distribution functions F and G. We want to test

$\mathbf{H}_0: \ F \equiv G$ against $\mathbf{K}: \ G(x) = F(x-\Delta), \ \Delta > 0.$

The distribution function F is unknown, but we expect that it is normal. On the other hand, we wish to have a test good under the normality, but at least unbiased for all F with a continuous density. Such are the permutation tests.

For simplicity, denote
$(X_1, \ldots, X_m, Y_1, \ldots, Y_n) = (Z_1, \ldots, Z_N)$,
$N = m + n$ and let
$Z_{(1)} \leq Z_{(2)} \leq \ldots \leq Z_{(N)}$ be the corresponding order statistics. The permutation test is based only on $Z_{(1)}, Z_{(2)}, \ldots, Z_{(N)}$ and it should satisfy

$$\frac{1}{N!} \sum_{\mathbf{r} \in \mathcal{R}} \Phi(Z_{(r_1)}, \ldots, Z_{(r_N)})) = \alpha, \tag{28}$$

where $\mathcal{R}$ is the set of $N!$ permutations of $1, 2, \ldots, N$. The test is conditional, under given vector of order statistics $Z_{(1)}, Z_{(2)}, \ldots, Z_{(N)}$, variable are only the permutations $(r_1, \ldots, r_N)$.

Generally, the test rejecting the alternative that $(Z_1, \ldots, Z_N)$ has density $q(z_1, \ldots, z_N)$ has the form

$$\Phi(z_{r_1}, \ldots, z_{r_N}) = \Phi\left(r_1, \ldots, r_N | Z_{(\cdot)}\right)$$

$$= \begin{cases} 1 & \ldots \quad q(z_{r_1}, \ldots, z_{r_N}) > C(z_{(\cdot)}) \\ \gamma & \ldots \quad q(z_{r_1}, \ldots, z_{r_N}) = C(z_{(\cdot)}) \\ 0 & \ldots \quad q(z_{r_1}, \ldots, z_{r_N}) < C(z_{(\cdot)}) \end{cases}$$

where $C(z_{(\cdot)})$ is determined so that (28) is satisfied.

It means that we reject $\mathbf{H}_0$ for k permutations $r_1, \ldots, r_N$ of $z_{(1)}, \ldots, z_{(N)}$ leading to the largest values of $q(z_{(r_1)}, \ldots, z_{(r_N)})$, where $k + \gamma = \alpha N!$.

Special case: Two normal samples differing by a shift in location

Here

$$q(z_1, \ldots, z_N) = f(x_1) \ldots f(x_m) f(y_1 - \Delta) \ldots f(y_n - \Delta)$$

where f is the density of $\mathcal{N}(\mu, \sigma^2)$, i.e.

$$q(z_1, \ldots, z_N)$$

$$\frac{1}{(\sigma\sqrt{2\pi})^N} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}(x_i - \mu)^2 + \sum_{j=1}^{n}(y_j - \mu - \Delta)^2\right]\right\}$$

and this is large iff $\sum_{j=1}^{n} y_j$ is large.

Hence, the test rejects if

$$\sum_{j=1}^{n} y_j = \sum_{i=m+1}^{N} z_i > C_1(z_{(\cdot)}).$$

The vectors $z_{m+1}, \ldots, z_N$ run over $\left(\begin{array}{c} N \\ n \end{array} \right)$ combinations of $z_1, \ldots, z_N$. We reject the hypothesis for k largest values of $\sum_{i=m+1}^{N} z_i$, where

$$k + \gamma = \alpha \left(\begin{array}{c} N \\ n \end{array} \right). \qquad (29)$$

Practical procedure:

(i) Observe $(x_1, \ldots, x_m, y_1, \ldots, y_n) = (z_1, \ldots, z_N)$.

(ii) Determine integer k and fraction $\gamma \geq 0$ satisfying (29).

(iii) Calculate the values $\sum_{j=1}^{n} z_{\ell_j}$ for all combinations $\ell_1, \ldots, \ell_n$. Find the $\left[\left(\begin{array}{c} N \\ n \end{array} \right) - k + 1 \right]$-st largest sum, say a^*.

(iv) Reject $\mathbf{H}_0$, if $\sum_{j=1}^{n} y_j > a^*$ and reject with probability γ if $\sum_{j=1}^{n} y_j = a^*$.

References:

H. Bünning and G. Trenkler (1978). *Nichparametrische statistische Methoden (mit Aufgaben und Lösungen und einem Tabellenanhang).* W. de Gruyter, Berlin.

J.Hájek and Z.Šidák (1967). *Theory of Rank Tests.* Academia, Prague.

J. Jurečková (1981). *Rank Tests* (in Czech). Lecture Notes, Charles University in Prague.

J. P. Lecoutre and P. Tassi (1987). *Statistique non parametrique et robustesse*. Economica, Paris.

E. L. Lehmann (1975). *Nonparametrics: Statistical Methods Based on Ranks.* Holden-Day, San Francisco.

E. L. Lehmann (1986). *Testing Statistical Hypotheses.* Second Edition. J. Wiley.