28

STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING
LEIDING: PROF. DR D. VAN DANTZIG
ADVISEUR VOOR STATISTISCHE CONSULTATIE: PROF. DR J. HEMELRIJK

On the combination of independent two sample tests of Wilcoxon

(corrected version)

by

Ph.van Elteren

1958

The Mathematical Centre at Amsterdam, founded the 11th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government through the Netherlands Organization for Pure Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

On the combination of independent two sample tests of Wilcoxon [1]

(corrected version)

by

Ph.van Elteren

1. Introduction

In statistical practice often the situation is met with, that one wants to draw conclusions from data, which have not all been gathered under the same conditions. When these conditions may affect the observed quantities, the data must be divided into groups that are homogeneous with respect to the conditions and the effects under consideration have to be tested within each group. The groups seperately often are too small to draw a conclusion. Then it is tried to draw an over-all-conclusion, applying a combination technique on the results of the individual tests. A well known technique is that of R.A. FISHER (1932) based on the probability integral transformation. The underlying idea of Fisher's technique is to use the logarithm of the product of the tailerrors (or probabilities of exceedance) of the individual tests as a test statistic. This statistic multiplied by -2 has a χ^2-distribution with 2k degrees of freedom, k being the number of the tests, provided the hypothesis tested is true for all combined tests. This simple technique can be applied for a large number of tests but has the following disadvantages:

1. It is only exact, if the statistics of the combined tests have continuous distributions (cf. W.A. WALLIS (1942)).
2. Attemps to change the weights of the individual tests make the techniques much more complicated.

For these reasons, the statistical department of the Mathematical Centre often used another easy combination method. It is based on a linear combination of the statistics of the individu-

---

1) Report SP 68 of the Statistical Department of the Mathematical Centre, Amsterdam. Head of the Department is Prof. Dr D. van Dantzig.

al tests. The obtained over-all statistic is in most cases approximately normally distributed under the hypothesis tested, as either the individual statistics have approximately normal distributions or the number of the combined tests is so large, that the Central Limit Theorem applies. The method can be used for many tests with symmetrically distributed statistics, and has a one-sided and a two-sided version. By an adequate choice of the combination coefficients the method can obtain special consistency or effeciency properties.

In the present paper the qualities of this combination method will be illustrated on Wilcoxon's two sample test. Wilcoxon himself has recommended the use of the sum of the statistics if a conclusion has to be drawn on k pairs of samples (cf. F. WILCOXON (1946)). Two linear combinations, in certain special cases equivalent with the sum, will be treated here. One of them yields a test, with a region of consistency that is independent of the proportion of the sample sizes and the other has in an important special case the largest efficiency.

2. Notation and definitions

Wilcoxon's two sample test can be applied on samples of two random variables $\underline{x}$ and $\underline{y}$ (cf. F. WILCOXON (1945), H.B. MANN and D.R. WHITNEY (1947)). In the present paper k pairs of random variables $\underline{x}_i$, $\underline{y}_i$ $(i=1,2,\ldots,k)$ are considered with distribution functions denoted by $F_i(x)$ and $G_i(x)$ respectively. Samples of independent observations of these variables are assumed to be available. The sample sizes will be denoted by m_i (for $\underline{x}_i$) and n_i (for $\underline{y}_i$).

The hypothesis H_o to be tested states that $F_i(x) \equiv G_i(x)$ for $i=1,2,..,k$.

Let $\underline{x}_{i,r}$ $(r=1,2,\ldots,m_i)$ and $\underline{y}_{i,s}$ $(s=1,2,\ldots,n_i)$ be the r^{th} observation of $\underline{x}_i$ and the s^{th} observation of $\underline{y}_i$ respectively. Let sgn(z) be defined by

$$(2.1) \qquad \mathrm{sgn}(z) \overset{\text{def}}{=} \begin{cases} -1 & \text{if} \quad z<0 \\ 0 & \text{if} \quad z=0 \\ +1 & \text{if} \quad z>0 \end{cases}$$

then Wilcoxon's statistic for the i^{th} pair of samples is a linear function of

$$(2.2) \qquad w_i \stackrel{\text{def}}{=} \sum_{r=1}^{m_i} \sum_{s=1}^{n_i} \operatorname{sgn}(\underline{x}_{ir} - \underline{y}_{is})$$

(cf. D. VAN DANTZIG and J. HEMELRIJK (1953)).
As mentioned in section 1, the statistics considered in this paper are of the type:

$$(2.3) \qquad \underline{W} \stackrel{\text{def}}{=} \sum_{i=1}^{k} c_i \underline{w}_i .$$

The numbers c_i are called the "weights". They have to be real and can depend on the sample sizes.

Only the right-sided test will be considered here, where hypothesis H_o is rejected if the observed value of $\underline{W}$ is equal to W_α or larger. W_α is defined as the smallest value that can be attained by $\underline{W}$ for which:

$$(2.4) \qquad P\left[\underline{W} \geq W_\alpha \mid H_o\right] \leq \alpha ,$$

where α is the level of significance.

If the distribution of $\underline{W}$ under hypothesis H_o is symmetric with respect to 0, which is true if the distributions $F_i(x)$ and $G_i(x)$ $(i=1,2,\ldots,k)$ are continuous, then the corresponding left sided test will have a critical region: $W \leq -W_\alpha$ and the two-sided test a critical region $|W| \geq W_{\alpha/2}$, both at the level of significance α. The properties of these tests can easily be derived from the properties of the right sided test treated below.
In this paper each test based on a statistic of the type (2.3), with critical region defined by $\underline{W} \geq W_\alpha$ will be called a W-test.

3. General properties of the distribution of $\underline{W}$

In this and the next section the following assumptions are assumed to be valid.

A 3.1 The random variables $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k$, $\underline{y}_1, \ldots, \underline{y}_k$ are independent.

A 3.2 The distribution functions $F_i(x)$ and $G_i(x)$ are continuous $(i=1,2,\ldots,k)$.

Assumption A 3.2 is not necessary for all results to be treated, but it is introduced for convenience in order to avoid the complications due to ties.

Well known properties of Wilcoxon's statistic in the case of one pair of samples yield immediately the following results:

$$(3.1) \qquad \mu_o \overset{\text{def}}{=} \mathcal{E}(\underline{W} \mid H_o) = 0 \ ,$$

$$(3.2) \qquad \sigma_o^2 \overset{\text{def}}{=} \operatorname{var}\{\underline{W} \mid H_o\} = \frac{1}{3} \sum_{i=1}^{k} c_i^2 m_i n_i (m_i + n_i + 1)$$

(cf. H.B. MANN and D.R. WHITNEY (1947)) and under alternative hypotheses:

$$(3.3) \qquad \mu \overset{\text{def}}{=} \mathcal{E}\,\underline{W} = \sum_{i=1}^{k} c_i m_i n_i b_i$$

where

$$(3.4) \qquad b_i \overset{\text{def}}{=} 2P[\underline{x}_i > \underline{y}_i] - 1 = 2 \int_{-\infty}^{+\infty} G_i(x)\ dF_i(x) - 1 \ .$$

For the variance σ^2 of $\underline{W}$ a more complicated expression in the distribution functions $F_i(x)$ and $G_i(x)$ is found (cf. D. VAN DANTZIG (1951)).

If the sample sizes m_i and n_i $(i=1,2,\ldots,k)$ are large, the distribution of $\underline{W}$ will be approximately normal. This can be concluded from a limit theorem by E.L. LEHMANN (1951) for the case of large sample sizes and from the Central Limit Theorem for the case of large k. These theorems are valid under very general conditions, not only under hypothesis H_o, but also under alternative hypotheses if the $|b_i|$ are smaller than 1.

It follows that the critical value W_α, defined in section 2, is approximately equal to

$$(3.5) \qquad \hat{W}_\alpha = \frac{1}{3} u_\alpha \sqrt{\sum_{i=1}^{k} c_i^2 m_i n_i \ (m_i + n_i + 1)} \ ,$$

where u_α is given by

$$(3.6) \qquad \frac{1}{\sqrt{2\pi}} \int_{u_\alpha}^{\infty} e^{-\frac{1}{2}x^2}\ dx = \alpha \ .$$

The power of the W-test with respect to a given alternative (given set of distribution functions $F_i(x)$ and $G_i(x)$ will be approximately equal to

$$(3.7) \qquad \hat{P}_{\alpha}(F_i,G_i) = 1-\phi((u_{\alpha} - \frac{\mu}{\sigma_0})\frac{\sigma_0}{\sigma}) ,$$

where $\phi(x)$ denotes the distribution function of a N(0,1)-variable.

4. Consistency; designfree W-test

A test is consistent against an alternative hypothesis H and with respect to a parameter N if the power of the test against hypothesis H tends to 1 for $N \to \infty$. In this section the classes of alternatives, against which the W-test is consistent, are investigated. The numbers k, m_i and n_i are supposed to be non-decreasing functions of a natural number N that tends to infinity. The dependence on N is denoted, if necessary, by writing k(N), $m_i(N)$, $n_i(N)$, $\mu(N)$ etc.

The following special cases will be considered:

Case I: $k(N)=k$ for each N, and $m_i(N)$ and $n_i(N)$ tend to infinity for $N \to \infty$ but $m_i(N)/N$ and $n_i(N)/N$ remain bounded and larger than a positive number.

Case II: k(N) tends to infinity for $N \to \infty$ and $m_i(N)$ and $n_i(N)$ remain bounded. For convenience it is assumed that $k(N)=N$ and that $m_i(N)$ and $n_i(N)$ are constants (denoted by m_i and n_i) for $i \leq N$.

An argument similar to that given by D. VAN DANTZIG (1951) shows that the power of the W-test tends to 1 if $\mu(N)/\sigma_0(N) \to \infty$. This yields the following theorem:

Theorem 4.1

The W-test is consistent with respect to N against all alternative hypotheses for which

$$N^{-\frac{3}{2}}\left\{\sum_{i=1}^{k} c_i^{\,2}(N)\right\}^{-\frac{1}{2}} \sum_{i=1}^{k} c_i(N)m_i(N)n_i(N)b_i \longrightarrow +\infty \quad \text{if} \quad N \longrightarrow \infty$$

in case I and

$$\left\{\sum_{i=1}^{N} c_i^{\,2}(N)\right\}^{-\frac{1}{2}} \sum_{i=1}^{N} c_i(N)m_i n_i b_i \longrightarrow +\infty \qquad \text{if} \quad N \longrightarrow \infty$$

in case II.

In both cases the test is for sufficiently small α not consistent against other alternatives.

According to theorem 4.1 the consistency conditions depend on the sample sizes. Given the b_i and the weights $c_i(N)$ the test can be made consistent by an appropriate choice of the sample sizes. This can be avoided by a special choice of the weights:

$$c_i(N) = \frac{c}{m_i(N)\ n_i(N)}, \tag{4.1}$$

where c is an arbitrary constant not equal to 0. If this constant is positive the test will be consistent against all alternatives for which:

$$N^{+\frac{1}{2}} \sum_{i=1}^{k} b_i \to \infty \quad \text{for } N \to \infty \qquad \text{in case I}$$

and

$$N^{-\frac{1}{2}} \sum_{i=1}^{N} b_i \to \infty \quad \text{for } N \to \infty \qquad \text{in case II}$$

and, for sufficiently small α, against no other alternatives.

As the consistency conditions of the W-test based on (4.1) do not depend on the sample design (i.e. on the numbers m_i and n_i), it will be called the "designfree W-test". The use of designfree tests has been recommended by C. VAN EEDEN and J. HEMELRIJK (1955).

The statistic of the designfree W-test is equal to the sum of the individual statistics w_i if $m_1=m_2=\ldots=m_k$ and $n_1=n_2=\ldots=n_k$.

In practice the set of pairs of distribution functions $F_i(x)$ and $G_i(x)$ $(i=1,2,\ldots,k)$ often can be considered as a random sample from a population of such pairs. Alternative hypotheses with this property will be called "randomized alternatives". If a randomized alternative is true the b_i are observations of a random variable $\underline{b}$. It can be shown that the W-test with positive weights is in case II consistent against all randomized alternatives for which $\mathcal{E}\underline{b}$ is positive. In case I the situation is more complicated.

5. Locally best W-test

In this section not only the numbers of observations but also the distribution functions $F_i(x)$ and $G_i(x)$ and thus the quantities b_i are supposed to depend on N. This will be denoted by writing $F_i(x;N)$, $G_i(x;N)$ and $b_i(N)$.

Further the following assumption is made:

A 5.1 For sufficiently large N

(5.1) $$\Delta_i(x;N) \overset{\text{def}}{=\!=} \sqrt{N}\left\{G_i(x;N) - F_i(x;N)\right\}$$

is bounded for $N \to \infty$.

Then it can be proved, that $\sigma_o(N)/\sigma(N) \to 1$ and $\mu(N)/\sigma_o(N)$ is bounded for $N \to \infty$. Thus the power of the W-test can be approximated by

(5.2) $$1 - \phi\left(u_\alpha - \frac{\mu(N)}{\sigma_o(N)}\right) .$$

The quantity $\mu(N)/\sigma_o(N)$ is given by

(5.3) $$\frac{\mu(N)}{\sigma_o(N)} = \frac{\sum_{i=1}^{k(N)} c_i(N)m_i(N)n_i(N)b_i(N)}{\sqrt{\frac{1}{3}\sum_{i=1}^{k(N)} c_i^2(N)m_i(N)n_i(N)\left(m_i(N)+n_i(N)+1\right)}}$$

It is easily seen that the right hand member of (5.3) and thus asymptotically for $N \to \infty$ the power of the W-test, attains its largest value if

(5.4) $$c_i(N) = \frac{cb_i(N)}{m_i(N)+n_i(N)+1} \quad (i=1,2,\dots,k(N))$$

provided $k(N)$, $m_i(N)$, $n_i(N)$ and $b_i(N)$ are given positive functions of N (c is an arbitrary positive constant).

Consequently the W-test with

(5.5) $$c_i(N) = \frac{c}{m_i(N)+n_i(N)+1}$$

has for $N \to \infty$ asymptotically the largest power against all alternatives for which all $b_i(N)$ are positive and $b_i(N)/b_j(N) \to 1$ for $N \to \infty$ and each pair (i,j) of natural numbers $(i \leq k(N),\ j \leq k(N))$.

Because of its conditional optimality the W-test based on (5.5) will be called the "locally best W-test".

It can be shown that in case II the locally best W-test has asymptotically for $N \to \infty$ the largest power against all "randomized" alternatives" (cf. section 4) with positive $\mathcal{E}\,\underline{b}(N)$ fulfilling the condition that $\mathcal{E}\,\underline{a}(N)$ and $\operatorname{var}\{\underline{b}(N)\}$ are of order $N^{-\frac{1}{2}}$ for $N \to \infty$, if $\underline{a}(N)$ is the random variable, of which the numbers:

$$a_i(N) \overset{\text{def}}{=} \int \left| G_i(x;N) - F_i(x;N) \right| \, d(\max_x \{F_i(x;N) \, , \, G_i(x;N)\})$$

$$(i=1,2,\ldots,N)$$

are assumed to be observations.

6. "Designfree" versus "locally best" W-test

The designfree and the locally best W-test are equivalent if and only if

$$K_i \overset{\text{def}}{=} \frac{m_i \, n_i}{m_i+n_i+1} = K \tag{6.1}$$

independent of i for $i=1,2,\ldots,k$. This condition is fulfilled if for $i=1,2,\ldots,k$: $m_i=m$ and $n_i=n$ or $m_i=n$ and $n_i=m$ independent of i.

There are more possibilities e.g. for:

$K = 0{,}8$:	$m_i=1$	$n_i=8$	or	$m_i=2$	$n_i=2$
$1{,}5$:	2	9		3	4 etc.

In order to compare the asymptotic efficiencies of both tests assumption A 5.1 and the following assumptions are made:
For $i=1,2,\ldots,k(N)$, each real value of x and $N \to \infty$:

A 6.1 $\int \Delta_i(x;N) \, dF_i(x;N)$ tends to a finite limit independent of i.

A 6.2 In case I : $N^{-1}K_i(N)$ has a limit for each i.
In case II: let $N(a)$ be the number of the values of i for which $K_i=a$, then for each possible value of a, $N(a)/N$ tends to a finite limit.

The following notation is used:

Case I : $\varkappa_1, \varkappa_2, \ldots, \varkappa_\nu$ are the values assumed by the limits of $N^{-1}K_i(N)$,
η_j denotes k^{-1} multiplied by the number of values of i for which $N^{-1}K_i(N) \to \varkappa_j$ $(i=1,2,\ldots,k \; ; \; j=1,2,\ldots,\nu)$.

Case II: $\varkappa_1, \varkappa_2, \ldots, \varkappa_\nu$ are the values assumed by the K_i,

$$\eta_j \overset{\text{def}}{=} \lim N(\varkappa_j)/N \qquad (i=1,2,\ldots,N \; ; \; j=1,2,\ldots,\nu) \; .$$

Then the asymptotic efficiency $e_{D,L}$, as defined by E.J.G. PITMAN (1947) (cf. G.E. NOETHER (1955)), of the designfree W-test relative to the locally best W-test is given by

$$(6.2)\qquad e_{D,L}=\left\{\sum_{j=1}^{\nu}\eta_j x_j\sum_{l=1}^{\nu}\eta_l x_l^{-1}\right\}^{-1}=\left\{1+\tfrac{1}{2}\sum_{j=1}^{\nu}\sum_{l=1}^{\nu}\eta_j\eta_l\frac{(x_j-x_l)^2}{x_j x_l}\right\}^{-1}.$$

Thus $e_{D,L}\leq 1$ which agrees with the optimality of the locally best test.

If the x_j are given and $x_1<x_2<\ldots<x_l$, $e_{D,L}$ will be minimum if $\eta_1=\eta_l=\frac{1}{2}$ and $\eta_2=\eta_3=\ldots=\eta_{\nu-1}=0$. Then:

$$e_{D,L}=\left\{1+\frac{1}{4}\frac{(x_\nu-x_1)^2}{x_\nu x_1}\right\}^{-1}$$

Consequently $e_{D,L}\longrightarrow 0$ if $x_\nu\rightarrow\infty$ and x_1 is kept constant. Two examples are given below.

Example 1:

$$\left.\begin{array}{lll} \text{Case I : } m_1(N)=2N & n_1(N)=2N & \multirow{2}{*}{$\left.\right\}\ x_1=1$} \\ \text{\quad " II : } m_1=2 & n_1=3 & \\ \text{Case I : } m_2(N)=3N & n_2(N)=6N & \multirow{2}{*}{$\left.\right\}\ x_2=2$} \\ \text{\quad " II : } m_1=4 & n_1=5 & \end{array}\right\}\quad e_{D,L}=\frac{8}{9}=0{,}89 \text{ if } \eta_1=\eta_2=\tfrac{1}{2}.$$

Example 2:

$$\left.\begin{array}{lll} x_1=1 \text{ (like in example 1)} & & \\ \text{Case I : } m_2(N)=10N & n_2(N)=10N & \multirow{2}{*}{$\left.\right\}\ x_2=5$} \\ \text{\quad " II : } m_2=10 & n_2=11 & \end{array}\right\}\quad e_{D,L}=\frac{5}{9}=0{,}56 \text{ if } \eta_1=\eta_2=\tfrac{1}{2}.$$

The tests can also be compared with respect to the asymptotic values of their power against special classes of alternatives fulfilling A 5.1 and A 6.1. Such a class is e.g. defined by

$$(6.3)\qquad \begin{cases} F_i(x;N)=H(x+\gamma_i) \\ G_i(x;N)=H(x+\gamma_i+\theta N^{-\frac{1}{2}}) \end{cases},$$

where $H(x)$ is a continuous distribution function, γ_i an arbitrary real number and θ a finite positive constant independent of i.

Another example is the class defined by

$$(6.4) \qquad G_i(x;N) = \left\{F_i(x;N)\right\}^{1-\Theta N^{-\frac{1}{2}}}$$

where Θ has the same properties as mentioned at (6.3).

The first class consists of shift alternatives as considered by PITMAN, the second of so called distributionfree alternatives of LEHMANN (1953).

If $h(x)$ is the derivative of $H(x)$,

$$A \stackrel{\text{def}}{=} \int h^2(x)\, dx,$$

$$B \stackrel{\text{def}}{=} \begin{cases} \sqrt{3k\left\{\sum_{j=1}^{\nu} \eta_j x_j^{-1}\right\}^{-1}} & \text{in case I} \\ \sqrt{3\left\{\sum_{j=1}^{\nu} \eta_j x_j^{-1}\right\}^{-1}} & \text{in case II,} \end{cases}$$

$$C \stackrel{\text{def}}{=} \begin{cases} \sqrt{3k \sum_{j=1}^{\nu} \eta_j x_j} & \text{in case I} \\ \sqrt{3 \sum_{j=1}^{\nu} \eta_j x_j} & \text{in case II,} \end{cases}$$

then the power of each of the tests tends for $N \longrightarrow \infty$ to the values given below.

(6.5)

	Class (6.3)	Class (6.4)
Design free	$1-\phi(u_\alpha-2\Theta AB)$	$1-\phi(u_\alpha-\frac{1}{2}\Theta B)$
Locally best	$1-\phi(u_\alpha-2\Theta AC)$	$1-\phi(u_\alpha-\frac{1}{2}\Theta C)$

Table (6.6) gives numerical results obtained by the substitutions

$$h(x) = (2\pi)^{-\frac{1}{2}}\, e^{-\frac{1}{2}x^2}, \quad \Theta = \begin{cases} \sqrt{\pi/k} & \text{for class (6.3), case I} \\ \sqrt{\pi} & \text{" \quad " \quad (6.3), case II} \\ 2/\sqrt{k} & \text{for class (6.4), case I} \\ 2 & \text{" \quad " \quad (6.4), case II,} \end{cases}$$

x_j and η_j according to the examples given above and $\alpha = 0{,}025$.

(6.6)

	Example 1	Example 2
Design free	0,516	0,609
Locally best	0,564	0,851

The following conclusions may be drawn from these results

1. If large differences between the quantities b_i are possible the designfree W-test should be used, as in that case the locally best test is not optimal and its consistency conditions strongly depend on the sample sizes.
2. If it is reasonable to assume that the b_i have values close to 0 and if the sample design shows large differences between the numbers K_i (defined by (6.1)) the locally best test may be preferred because of its larger efficiency.
3. If the numbers K_i are (approximately) equal both tests are (nearly) equivalent.

Résumé

Sur la combinaison de tests indépendents pour deux échantillons de Wilcoxon

Dans cet article l'auteur analyse une classe de tests, dont les valeurs typiques sont des combinaisons linéaires $\sum_{i=1}^{k} c_i w_i$ des valeurs typiques $w_1, \ldots, w_k$ de k tests indépendents pour deux échantillons de Wilcoxon. Deux combinaisons spéciales sont examinées en particulier. Les coefficients de ces combinaisons sont définis par:

(1) $c_i = \frac{c}{m_i\ n_i}$, où c est un nombre réel et m_i et n_i sont les effectifs des échantillons du i-ième test.

(2) $c_i = \frac{c}{m_i+n_i+1}$.

Le test (1) a une région de consistance indépendante des effectifs. Pour une classe importante d'alternatives le test (2) est asymptotiquement le plus puissant pour $k \to \infty$ ou $m_i \to \infty$, $n_i \to \infty$ et m_i/n_i possède une borne supérieure.

Les efficacités des deux méthodes ont été comparées pour des cas spéciaux.

References

VAN DANTZIG, D (1951), On the consistency and the power of Wilcoxon's two sample test,
Proc.Kon.Ned.Akad.v.Wet. A 54 } (1951) 1-8.
Indagationes Mathematicae 13 }

VAN DANTZIG, D. and J. HEMELRIJK (1953), Statistical methods based on few assumptions,
Bull. of the I.S.I. 24 (1954) 239-267.

VAN EEDEN, C. and J. HEMELRIJK (1955), A test for the equality of probabilities against a class of specified alternative hypotheses including trend,
Proc. Kon.Ned.Akad.v.Wet. 58 } (1955) 191-198, 301-308.
Indagationes Mathematicae 17 }

FISHER, R.A. (1932), Statistical methods for research workers,
Oliver and Boyd, Edinburgh and London, 4th ed.

LEHMANN, E.L. (1951), Consistency and unbiasedness of certain non-parametric tests,
Ann. of Math. Statistics 22 (1951) 165-179.

LEHMANN, E.L. (1953), The power of rank tests,
Ann. of Math. Statistics 24 (1953) 23-43.

MANN, H.B. and D.R. WHITNEY (1947), On a test whether one of two random variables is stochastically larger than the other,
Ann. of Math. Statistics 18 (1947) 50-60.

NOETHER, G.E. (1955), On a theorem of Pitman,
Ann. of Math. Statistics 26 (1955) 64-68.

PITMAN, E.J.G. (1949), Lecture notes on non-parametric statistical inference,
Columbia University, 1949.

WALLIS, W.A. (1942), Compounding probabilities from independent significance tests,
Econometrica 10 (1942) 229-248.

WILCOXON, F. (1945), Individual comparisons by ranking methods,
Biometrics 1 (1945) 80-82.

WILCOXON, F. (1946), Individual comparisons of grouped data by ranking methods,
Journal of Entomology 39 (1946) 269-270.