Example 1.: Postsurgical treatment

The new postsurgical treatment is compared with the standard one. Data in the table are the recovery times in days: X_i for the patients with the new treatment, Y_j those for control. Used test: *Kolmogorov-Smirnov*.

	i	Z_i	R_i	V_i	$\frac{i}{2} - V_1 - \ldots - V_i$
X_i	1	19	1	0	0.5
	2	22	4	1	0.0
	3	25	6	1	-0.5
	4	26	7	0	0.0
	5	28	8	1	-0.5
	6	29	9	0	0.0
	7	34	12	0	0.5
	8	37	14	0	1.0
	9	38	15	0	1.5
Y_j	10	20	2	1	1.0
	11	21	3	1	0.5
	12	24	5	0	1.0
	13	30	10	1	0.5
	14	32	11	0	1.0
	15	36	13	0	1.5
	16	40	16	1	1.0
	17	48	17	1	0.5
	18	54	18	1	0.0

Example 1: Continuation

$D_{mn}^{+} = \frac{N}{mn} \cdot \max_{1 \le i \le N} \left(\frac{i}{2} - V_1 - \ldots - V_i \right) = \frac{1}{3}$

$nD_{mn}^{+} = 3$

By Table 29: $P_{H_0}(nD_{mn}^{+} \le 4) = 0.9271$

$P_{H_0}(nD_{mn}^{+} \le 5) = 0.9832$

$\Longrightarrow$ $P_{H_0}(nD_{mn}^{+} \ge 5) = 0.0629$

$P_{H_0}(nD_{mn}^{+} \ge 6) = 0.0168.$

Notation: $N = m + n$

We do not reject the hypothesis $\mathbf{H}_0$.

Example 2. Psychological counseling

In a test of the effect of psychological counseling, 80 boys are divided at random into a control group of 40 to whom only the normal counseling facilities are available, and a treatment group of 40 who receive a special counseling. At the end of the study, an assessment is made of each boy who is then classified as having a good (1), fairly good (2), fairly poor (3), or poor (4) adjustment, with the following results:

	(1)	(2)	(3)	(4)	Total
Treatment	5	7	16	12	40
Control	7	9	15	9	40
Sum	12	16	31	21	80

$\mathbf{H_0}$: no significance effect of special counseling
$\mathbf{H_1}$: results are better after special counseling

Used test: *Wilcoxon test with midranks*

d_i	ranks	midrank
5+7=12	1,. . . ,12	6.5
7+9=16	13,. . . ,28	20.5
16+15=31	29,. . . ,59	44
12+9=21	60,. . . ,80	70

X_i's		midrank
5	$X_1, \ldots, X_5$	6.5
7	$X_6, \ldots, X_{12}$	20.5
16	$X_{13}, \ldots, X_{28}$	44
12	$X_{29}, \ldots, X_{40}$	70

Y_i's		midrank
7	$Y_1, \ldots, Y_7$	6.5
9	$Y_8, \ldots, Y_{16}$	20.5
15	$Y_{17}, \ldots, Y_{31}$	44
9	$Y_{32}, \ldots, Y_{40}$	70

$m = n = 40, \; N = m + n = 80$

$$W_N^* = \sum_{i=1}^{40} [\text{midrank of } \; X_i] = 1720$$

$$\mathbf{E} W_N^* = \frac{m(N+1)}{2} = 1620$$

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

$$= 10800 - 945.2 = 9854.8$$

$$(\text{var } W_N^*)^{1/2} = 99.27$$

$$\frac{W_N^* - \mathbf{E}W_N^*}{(\text{var } W_N^*)^{1/2}} = \frac{100}{99.27} = 1.007$$

$$< \Phi^{-1}(0.95) = 1.64$$

Hypothesis is not rejected. The test indicates no significant effect of the special counseling on the prevention of the juvenile delinquency.

Example 3: Effectiveness of a new medicament against headaches

15 patients got two bottles with pills, denoted A and B. They should always one pill when they suffered from headaches, alternating A and B (only the doctor knew which bottle contained the new medicament and which contained the standard one). 10 among 15 patients reported in favor of the new medicament. Is this result significant, can we claim that the new medicament is significantly better?

Hypothesis $\mathbf{H}_1$: There is no difference in the effects of two medicaments.
Used test: the **sign test**
S_{15} = the number of positive records among 15.
Under $\mathbf{H}_1$: S_{15} has the binomial distribution $b(15, \frac{1}{2})$.

$$P_{H_1}(S_{15} \geq 10) = \sum_{i=10}^{15} \binom{15}{i} \left(\frac{1}{2}\right)^{15} = 0.1509$$

$$P_{H_1}(S_{15} \geq 11) = 0.0592$$

$$P_{H_1}(S_{15} \geq 12) = 0.0176.$$

10 positive records is not yet significant for rejecting the hypothesis. 11 positive records would be on the border of significance on the level 0.05.

Example 4: Effect of thiamin on learning

74 children were divided in 37 matched pairs. One child in each pair was receiving B_1. The table below gives the gain in IQ during the 6 weeks of experiment for 12 of the pairs.

Used test: One-sample Wilcoxon test
(with midranks)

$$W_N^* = \sum_{i:Z_i>0} R_i^+ \qquad W_N^+ = \sum_{i=1}^{N} \text{sign } Z_i . R_i^+$$

$$W_N^* = \frac{1}{2} W_N^+ + \frac{1}{4} N(N+1)$$

$R_1^+, \ldots, R_N^+$ are the ranks of $|Z_1|, \ldots, |Z_N|$. Under $\mathbf{H}_1$ (symmetry), the sign Z_i and R_i^+ are independent. This enables to calculate $\mathbf{E}W_N^*$, $\mathbf{E}W_N^+$ and var W_N^* :

$$\mathbf{E}W_N^+ = 0 \Longrightarrow \mathbf{E}W_N^* = \frac{1}{4} N(N+1)$$

$$\text{var } W_N^* = \frac{1}{24} N(N+1)(2N+1).$$

Modification of Wilcoxon in the presence of nulls and ties:

Assume that among $|Z_1|, \dots, |Z_N|$ are d_0 nulls and from the remaining are e values different:

d_1	equal to the	smallest
d_2	equal to the	2nd smallest
...		
d_e	equal to the	largest,

$d_0 + d_1 + \dots + d_e = N$.
We omit zeros and calculate the midranks $\tilde{R}_i$ for remaining $|Z_i|$.
$\tilde{W}_N = \sum_{i:Z_i>0} \tilde{R}_i$ is the modified Wilcoxon statistic.

Corrected parameters of $\tilde{W}_N$ with respect to nulls and ties:

$$\mathbf{E}\tilde{W}_N = \frac{1}{4}N(N+1) - \frac{1}{4}d_0(d_0+1)$$

$$\text{var } \tilde{W}_N = \frac{1}{24}N(N+1)(2N+1)$$

$$-\frac{1}{24}d_0(d_0+1)(2d_0+1) - \frac{1}{48}\sum_{i=1}^{e} d_i(d_i^2-1)$$

$$\mathbf{P}_{H_1}\left\{\frac{\tilde{W}_N - \mathbf{E}\tilde{W}_N}{(\text{var } \tilde{W}_N)^{1/2}} \le x\right\} \longrightarrow \Phi(x)$$

as $N \to \infty$, where Φ is the standard normal distribution function.

i	treated Y_i	control X_i	$Z_i = Y_i - X_i$	sign $Z_i \cdot \tilde{R}_i$
1	14	8	6	8.0
2	18	26	-8	-10.0
3	2	-7	9	11.0
4	4	-1	5	6.5
5	-5	2	-7	-9.0
6	14	9	5	6.5
7	-3	0	-3	-4.0
8	-1	-4	3	4.0
9	1	13	-12	-12.0
10	6	3	3	4.0
11	3	3	0	0.0
12	3	4	-1	-2.0

$$\tilde{W}_N = \sum_{i:Z_i>0} \tilde{R}_i = 40$$

$$d_0 = 1,\ d_1 = 3,\ d_2 = 2$$

$$\mathbf{E}\tilde{W}_N = 39 - \frac{1}{2} = 38.5$$

$$\text{var } \tilde{W}_N = 161.75$$

$$\frac{\tilde{W}_N - \mathbf{E}\tilde{W}_N}{(\text{var } \tilde{W}_N)^{1/2}}$$

$$= \frac{40 - 38.5}{12.72} = 0.12 < 1.64 = \Phi^{-1}(0.95)$$

Hypothesis $\mathbf{H}_1$ is not rejected. The data do not confirm the effect of thiamin on learning.

Example 5: IQ scores at four universities

The table gives the IQ scores of 100 stage 1 students at each of four New Zealand universities: Auckland, Wellington, Canterbury and Otago.

IQ	1	2	3	4	5	6	7	8	9	10	11	
A	1	2	9	13	16	16	14	13	9	5	2	100
W	1	2	9	9	12	15	18	14	9	6	5	100
C	3	5	7	13	15	14	12	12	9	6	4	100
O	2	3	7	13	17	15	11	14	8	5	5	100
d_i	7	12	32	48	60	60	55	53	35	22	16	400

Hypothesis H_0: There is not significance difference between the universities

Used test: Kruskal=Wallis test with midranks

$N = 400$, $e = 11$ different values, $n_1 = n_2 = n_3 = n_4 = 100$.

i	d_i	IQ	midrank
1	7	90-94	4.0
2	12	95-99	13.5
3	32	100-104	35.5
4	48	105-109	75.5
5	60	110-114	129.5
6	60	115-119	189.5
7	55	120-124	247.0
8	53	125-129	301.0
9	35	130-134	345.0
10	22	135-139	373.5
11	16	140-	392.5

$$K^* = \frac{\frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \left(R_{i.}^* - \frac{N+1}{2} \right)^2}{1 - \frac{\sum_{i=1}^{e}(d_i^3 - d_i)}{N(N^2-1)}}$$

$R_{1.}^* = 195.645$ $\quad R_{2.}^* = 213.95$

$R_{3.}^* = 193.7$ $\quad R_{4.}^* = 198.435$

$K^* = 1.909$

$\mathbf{P}(\chi_3^2 \geq 7.8) = 0.0503$

Because $K^* = 1.909 < 7.8$, we do not reject the hypothesis. The data did not indicate a significant difference between the universities.

Example 6: Effectiveness of hypnosis

In a study of hypnosis, the emotions of fear (1), happiness (2), depression (3) and calmness (4) were requested (in random order) from each of eight subjects during the hypnosis. The following table gives the resulting measurements of skin potentials in millivolts.

Hypothesis H_2 : No significant difference between the four emotions.

Used test: Friedman test for observations divided in blocks

1	2	3	4	5	6	7	8	
23.1	57.6	10.5	23.6	11.9	54.6	21.0	20.3	
22.7	53.2	9.7	19.6	13.8	47.1	21.0	20.3	
22.5	53.7	10.8	21.1	13.7	39.2	13.7	16.3	
22.6	53.1	8.3	21.6	13.3	37.0	14.8	14.8	

R_{1j}	R_{2j}	R_{3j}	R_{4j}	R_{5j}	R_{6j}	R_{7j}	R_{8j}	$R_{\cdot j}$
4	4	3	4	1	4	4	3	27/8
3	2	2	1	4	3	3	4	22/8
1	3	4	2	3	2	1	2	18/8
2	1	1	3	2	1	2	1	13/8

$p = 4$ treatments, $N = 8$ blocks (8 patients)
$R_{\cdot j} = \frac{1}{N} \sum_{i=1}^{N} R_{ij}$

$$Q_N = \frac{12N}{p(p+1)} \sum_{j=1}^{p} \left[R_{\cdot j} - \frac{1}{2}(p+1) \right]^2$$

$$= \frac{96}{20} \left[\left(\frac{27}{8} - \frac{5}{2} \right)^2 + \left(\frac{22}{8} - \frac{5}{2} \right)^2 + \left(\frac{18}{8} - \frac{5}{2} \right)^2 + \left(\frac{13}{8} - \frac{5}{2} \right)^2 \right] = 7.72$$

Table M: $\mathbf{P}_{H_2}(Q_N \geq 7.5) = 0.052$
$\mathbf{P}_{H_2}(Q_N \geq 7.65) = 0.049$

Because $Q_N = 7.72 > 7.65$, we reject the hypothesis. The data indicate a significant difference in the skin potentials.

Example 7: Comparison of effects of four medicaments

The following table presents the total number of coughs per days of seven patients, under three medicaments:
heroin, 5 mg (1)
dextromethorphan, 10 mg (2)
codeine, 10 mg (3)
placebo (4).
The subscripts are the ranks for each patient Determine whether there is a significant difference between the four treatments.

	1	2	3	4	5	6	7	$R_{\cdot j}$
(1)	251	126	49	45	233	291	1385	2.43
(2)	207	180	123	85	232	208	1204	2.29
(3)	167	104	63	147	233	158	1611	2.0
(4)	301	120	186	100	250	183	1913	3.29

$N = 7, \; p = 4$
$Q_N = 3.86 < 7.63 =$ the 95% critical value. We do not reject hypothesis of no significant difference between the treatments.

However, let us still consider the difference between the codeine and placebo. In this case the Friedman test reduces to the sign test: 6 among 7 differences between the values for codeine and placebo are positive. For the binomial random variable $B = b(7, \frac{1}{2})$ we have

$$P(B = 7) = 0.0078, \; P(B = 6) = 0.0547,$$
$$P(B \geq 6) = 0.0625$$
$$P(B = 7) + \gamma P(B = 6) = 0.05$$
$$\text{for} \quad \gamma = 0.77 \approx 0.8.$$

Because our value 6 is on the border of significance 0.05, we should make the randomization, which leads to the conclusion that we should reject the hypothesis, that there is no difference between the codeine and placebo, with probability 0.8, and do not reject with probability 0.2.

Example 8: Test of independence of performance in language and arithmetics

From a group of 98 students enrolled for a statistics course, 9 are selected at random and given a simple arithmetic tests and an artificial language test. Using the Spearman test, we should test for independence of both performances.
The following table gives the scores L_i in language and A_i in arithmetics of the tested students. The R_i are the ranks with respect to the language scores and S_i the ranks with respect to the arithmetics scores.

i	L_i	R_i	A_i	S_i	$(S_i - R_i)^2$
1	50	6	38	8	4
2	23	2	28	6	16
3	28	3	14	1	4
4	34	4	26	4	0
5	14	1	18	2	1
6	54	9	40	9	0
7	46	5	23	3	4
8	52	7	30	7	0
9	53	8	27	5	9
					$\sum = 38$

$$\mathcal{S}^* = \sum_{i=1}^{n} (S_i - R_i)^2 = 38$$

Table 12: $n = 9$, $\alpha = 0.05$

$$\mathbf{P}_{H_3}(\mathcal{S}^* \leq 50) \leq 0.05$$

$$\mathbf{P}_{H_3}(\mathcal{S}^* \leq 51) > 0.05$$

$$\alpha = 0.01$$

$$\mathbf{P}_{H_3}(\mathcal{S}^* \leq 28) \leq 0.01$$

$$\mathbf{P}_{H_3}(\mathcal{S}^* \leq 29) > 0.01.$$

Because $\mathcal{S}^* < 50$ and the small values of $\mathcal{S}^*$ are significant, we reject the hypothesis of independence in favour of the alternative of positive dependence between performances in languages and arithmetics at the level $\alpha = 0.05$. However, we do not reject on the level $\alpha = 0.01$, because $\mathcal{S}^* > 29$.

Example 9: Crying babies and their IQ

Sperman correlation coefficient with midranks:

We have observations

$$\begin{array}{lll} X_1, & \ldots, & X_N \\ Y_1, & \ldots, & Y_N \end{array}$$

and assume there are e_1 different values among the X_i, among them
d_1 equal to the smallest, d_2 equal to the 2nd smallest, etc., d_{e_1} equal to the largest.
Similarly, there are e_2 different values among the Y_i, among them
f_1 equal to the smallest, f_2 equal to the 2nd smallest, etc., f_{e_2} equal to the largest.
Let $R_1^*, \ldots, R_N^*$ be the midranks of the X_i
and $S_1^*, \ldots, S_N^*$ be the midranks of the Y_i,
$i = 1, \ldots, N$.
The modified Spearman statistic is

$$\mathcal{S}^{**} = \sum_{i=1}^{N} (S_i^* - R_i^*)^2.$$

For testing we use the normal approximation with the parameters

$$\mathbf{E}_{H_3}(\mathcal{S}^{**}) = \frac{N^3 - N}{6} - \sum_{i=1}^{e_1} \frac{d_i^3 - d_i}{12} - \sum_{j=1}^{e_2} \frac{f_j^3 - f_j}{12}$$

$$\text{var}_{H_3}\ \mathcal{S}^{**} = \frac{(N-1)N^2(N+1)^2}{36}$$

$$\cdot \left\{ \left[1 - \sum_{i=1}^{e_1} \frac{d_i^3 - d_i}{N^3 - N} \right] \cdot \left[1 - \sum_{j=1}^{e_2} \frac{f_j^3 - f_j}{N^3 - N} \right] \right\}^{-1} .$$

To test whether children who cry more actively s babies later tend to have higher IQ, a cry account X_i was taken for 22 children aged 5 days and the IQ scores Y_i at the age of 3 years. R_i^* and S_i^* in the following table are the midranks of X_i and Y_i, respectively, $i = 1, \ldots, N$, $N = 22$.

$$e_1 = 19,\ d_4 = d_7 = d_8 = 2,\quad \text{other}\quad d_i = 1$$

$$e_2 = 18,\ f_4 = f_6 = f_9 = f_{10} = 2,$$

$$f_3 = f_5 = f_7 = 3$$

$$\mathcal{S}^{**} = 1601.5$$

$$\mathbf{E}_{H_3}\mathcal{S}^{**} = 1761.5, \quad (\text{var}_{H_3}\ \mathcal{S}^{**})^{1/2} = 384.4$$

i	X_i	Y_i	R_i^*	S_i^*	$(S_i^* - R_i^*)^2$
1	20	90	16.0	1.0	225
2	17	94	9.0	2.0	1
3	15	100	4.0	3.0	1
4	19	103	14.0	4.5	90.25
5	23	103	19.5	4.5	225
6	14	106	2.0	6.0	16
7	27	108	22.0	7.0	225
8	17	109	9.0	8.5	0.25
9	18	109	11.5	8.5	9
10	15	112	4.0	10.5	42.25
11	15	112	4.0	10.5	42.25
12	23	113	19.5	12.0	56.25
13	21	114	17.5	13.0	20.25
14	16	118	6.5	14.0	56.25
15	12	119	1.0	15.0	196
16	19	120	14.0	16.0	4
17	18	124	11.5	17.0	30.25
18	19	132	14.0	18.0	16
19	16	133	6.5	19.0	156.25
20	17	141	9.0	20.0	121
21	26	155	21.0	21.0	0
22	21	157	17.5	22.0	20.25

$$\frac{1601.5 - 1761.5}{384.4} = -0.42 > -1.64 = -\Phi^{-1}(0.95).$$

We do not reject the hypothesis of independence.

Example 10: Pollution of Lake Michigan

The data give the number of "odor periods" observed in each year of the period 1950-64.

Hypothesis $\mathbb{H}$: No change of pollution with the time

Alternative: There was an upward trend in the pollution with the time

In the table, X_i is the year with the rank $R_i = i$, Y_i is the pollution, and S_i^* is the midrank of Y_i, $i = 1, \ldots, 15$.

i	X_i	Y_i	S_i^*	$(S_i^* - i)^2$
1	50	10	1	0
2	51	20	10	64
3	52	17	6.5	12.25
4	53	16	5	1
5	54	12	2	9
6	55	15	4	4
7	56	13	3	16
8	57	18	8	0
9	58	17	6.5	6.25
10	59	19	9	1
11	60	21	11	0
12	61	23	12.5	0.25
13	62	23	12.5	0.25
14	63	28	14.5	0.25
15	64	28	14.5	0.25

Test criterion:

$$\mathcal{S}^* = \sum_{i=1}^{N} (S_i^* - i)^2$$

which has the same distribution as the Spearman criterion under independence. Hence, if there are no ties, then under $\mathbf{H}_3$,

$$\mathbf{E}\mathcal{S}^* = \frac{N^3 - N}{6}$$

$$\text{var } \mathcal{S}^* = \frac{N^2(N+1)^2(N-1)}{36}$$

and we use the normal approximation or the tables of Spearman. The small values of $\mathcal{S}^*$ are significant.

Because the ties can appear only in the second row, with the midranks we have the parameters

$$\mathbf{E}\mathcal{S}^* = \frac{N^3 - N}{6} - \frac{1}{12} \sum_{i=1}^{e} (d_i^3 - d_i)$$

$$\text{var } \mathcal{S}^* = \frac{N^2(N+1)^2(N-1)}{36} \left[1 - \sum_{i=1}^{e} \frac{d_i^3 - d_i}{N^3 - N}\right].$$

For the lake Michigan,

$$\mathcal{S}^* = 114.5$$

$$\mathbf{E}\mathcal{S}^* = 558.5, \quad (\text{var } \mathcal{S}^*)^{1/2} = 149.26$$

$$\frac{114.5 - 558.5}{149.26} = -2.97 < -1.64 = \Phi^{-1}(0.05)$$

We reject the hypothesis and accept the alternative of the upward trend.
