

Advanced Topics in Biostatistics: Non-Parametric Methods

Prof. Dr. Annette Kopp-Schneider
Biostatistics – C060
kopp@dkfz.de

Outline

- Overview test procedures
- Wilcoxon signed rank test
- Wilcoxon rank sum test = Mann-Whitney U-Test
- Kruskal-Wallis-test
- Spearman rank correlation

Test procedures for quantitative variables:

Comparison of means

1 sample:

One-sample t-test

2 paired samples:

paired t-test

2 independent samples:

t-test (equal standard deviation in both groups)

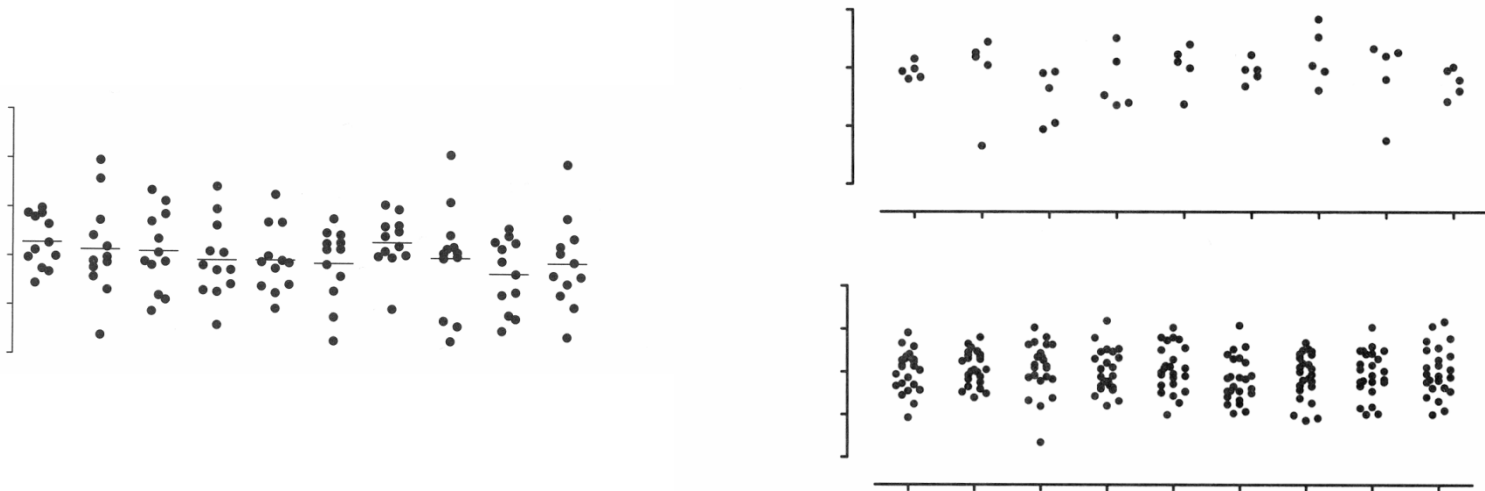
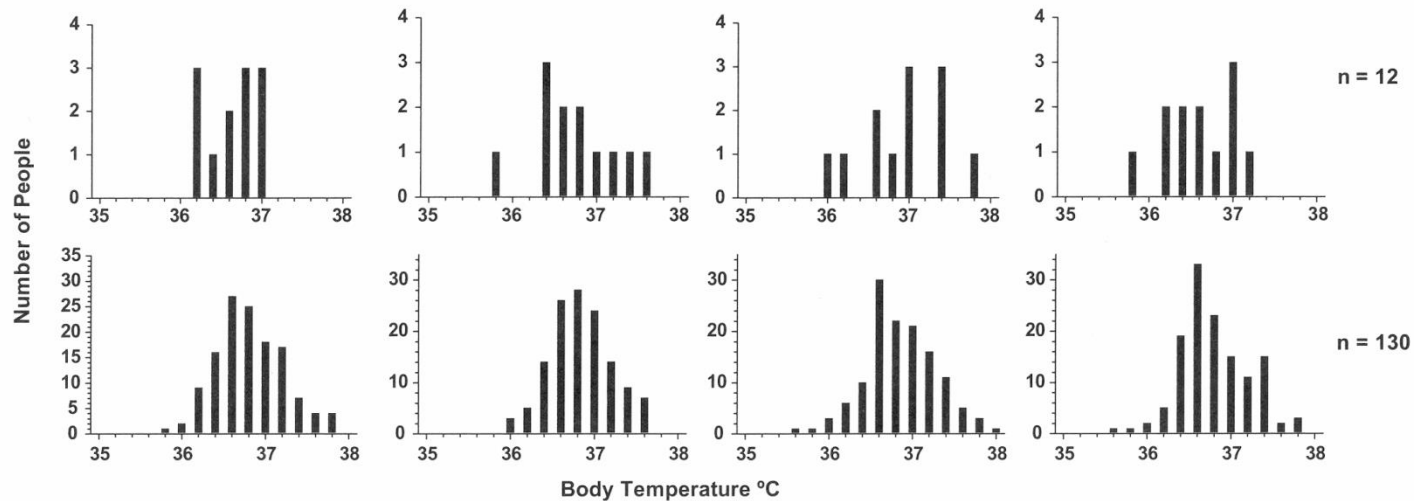
Welch-test (different standard deviation in both groups)

≥ 3 independent samples:

Analysis of Variance (ANOVA)

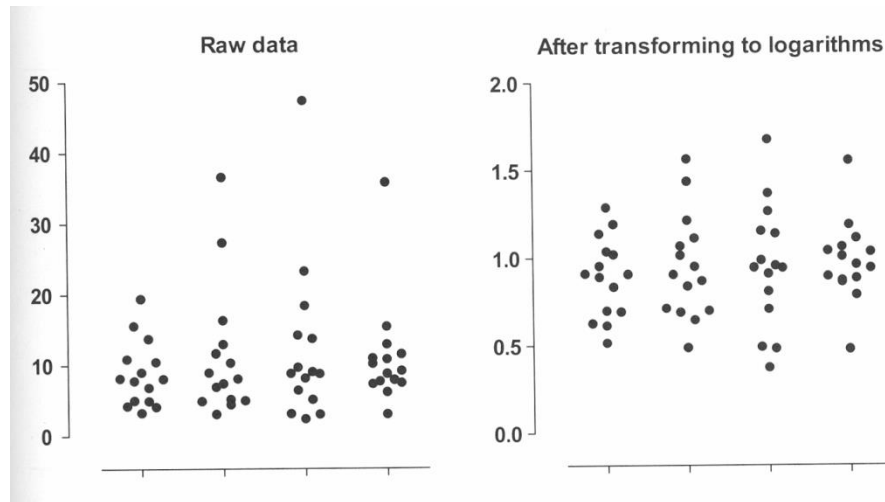
General assumption: Data come from Normal distribution

Normal distribution does not always look ,normal‘



From: H. Motulsky, *Intuitive Biostatistics*, 2010.

Transformation to normal distribution



Raw data from a log-normal distribution seem to have outliers.

Note: most outlier tests are inappropriate when values come from a non-Gaussian distribution

From: H. Motulsky, Intuitive Biostatistics, 2010.

Non-parametric procedures based on ranks

When to use...

- Data do not come from normal distribution
- No data transformation available to transform data to normal distribution
- Transformation of data to normal distribution is hard to interpret
- Presence of outlying observations

Test Procedures for Quantitative Variables: Comparison of Means

Parametric test (assuming normal distribution)	<i>Non-parametric test</i>
<u>1 sample:</u> One-sample t-test	
<u>2 paired samples:</u> paired t-test	<i>Sign test, Wilcoxon signed rank test</i>
<u>2 independent samples:</u> t-test Welch-test	<i>Wilcoxon rank sum test (= Mann-Whitney U-test)</i>
<u>≥ 3 independent samples:</u> Analysis of Variance (ANOVA)	<i>Kruskal-Wallis-test</i>

Test Procedures for Quantitative Variables: Comparison of Means

Parametric test (assuming normal distribution)	<i>Non-parametric test</i>
<u>1 sample:</u> One-sample t-test	
<u>2 paired samples:</u> paired t-test	<i>Sign test,</i> <i>Wilcoxon signed rank test</i>
<u>2 independent samples:</u> t-test Welch-test	<i>Wilcoxon rank sum test</i> <i>(= Mann-Whitney U-test)</i>
<u>≥ 3 independent samples:</u> Analysis of Variance (ANOVA)	<i>Kruskal-Wallis-test</i>

Wilcoxon signed rank test: Example

A non-parametric alternative to the one-sample t-test or the paired t-test

Example:

Results of a placebo-controlled trial to test the effectiveness of a sleeping drug:
Hours of sleep

Patient	Drug	Placebo
1	6.1	5.2
2	6.0	7.9
3	8.2	3.9
4	7.6	4.7
5	6.5	5.3
6	5.4	7.4
7	6.9	4.2
8	6.7	6.1
9	7.4	3.8
10	5.8	7.3

Wilcoxon signed rank test: Example

A non-parametric alternative to the one-sample t-test or the paired t-test

Example:

Results of a placebo-controlled trial to test the effectiveness of a sleeping drug:
Hours of sleep

Patient	Drug	Placebo	Diff
1	6.1	5.2	0.9
2	6.0	7.9	-1.9
3	8.2	3.9	4.3
4	7.6	4.7	2.9
5	6.5	5.3	1.2
6	5.4	7.4	-2.0
7	6.9	4.2	2.7
8	6.7	6.1	0.6
9	7.4	3.8	3.6
10	5.8	7.3	-1.5

Wilcoxon signed rank test: Example

A non-parametric alternative to the one-sample t-test or the paired t-test

Example:

Results of a placebo-controlled trial to test the effectiveness of a sleeping drug:
Hours of sleep

Patient	Drug	Placebo	Diff	Diff
1	6.1	5.2	0.9	0.9
2	6.0	7.9	-1.9	1.9
3	8.2	3.9	4.3	4.3
4	7.6	4.7	2.9	2.9
5	6.5	5.3	1.2	1.2
6	5.4	7.4	-2.0	2.0
7	6.9	4.2	2.7	2.7
8	6.7	6.1	0.6	0.6
9	7.4	3.8	3.6	3.6
10	5.8	7.3	-1.5	1.5

Wilcoxon signed rank test: Example

A non-parametric alternative to the one-sample t-test or the paired t-test

Example:

Results of a placebo-controlled trial to test the effectiveness of a sleeping drug:
Hours of sleep

Patient	Drug	Placebo	Diff	Diff	Rank
1	6.1	5.2	0.9	0.9	
2	6.0	7.9	-1.9	1.9	
3	8.2	3.9	4.3	4.3	
4	7.6	4.7	2.9	2.9	
5	6.5	5.3	1.2	1.2	
6	5.4	7.4	-2.0	2.0	
7	6.9	4.2	2.7	2.7	
8	6.7	6.1	0.6	0.6	
9	7.4	3.8	3.6	3.6	
10	5.8	7.3	-1.5	1.5	

Wilcoxon signed rank test: Example

A non-parametric alternative to the one-sample t-test or the paired t-test

Example:

Results of a placebo-controlled trial to test the effectiveness of a sleeping drug:
Hours of sleep

Patient	Drug	Placebo	Diff	Diff	Rank
1	6.1	5.2	0.9	0.9	2
2	6.0	7.9	-1.9	1.9	5
3	8.2	3.9	4.3	4.3	10
4	7.6	4.7	2.9	2.9	8
5	6.5	5.3	1.2	1.2	3
6	5.4	7.4	-2.0	2.0	6
7	6.9	4.2	2.7	2.7	7
8	6.7	6.1	0.6	0.6	1
9	7.4	3.8	3.6	3.6	9
10	5.8	7.3	-1.5	1.5	4

Sum of ranks of positive differences: $T_+ = 40$

Sum of ranks of negative differences: $T_- = 15$

Wilcoxon signed rank test: Procedure

A non-parametric alternative to the one-sample t-test or the paired t-test

Situation:

Two paired samples

No assumption about distribution of differences

Hypotheses:

$$H_0 : \text{Median}(F_{\text{drug-placebo}}) = 0$$

$$H_1 : \text{Median}(F_{\text{drug-placebo}}) \neq 0$$

Test statistic:

Follow algorithm:

1. Exclude differences = 0, $N = \# \text{nonzero differences}$
2. Rank absolute differences
3. $T_+ = \text{rank sum of positive differences}$
 $T_- = \text{rank sum of negative differences}$
 $T = \min\{T_+, T_-\}$

Test decision for small sample size:

If $T < \text{critical value from Table} \rightarrow \text{reject } H_0$

Table A7 Critical values for the Wilcoxon matched pairs signed rank test.

Reproduced from Table 21 of White *et al.* (1979) with permission of the authors and publishers.

N = number of non-zero differences; T = smaller of T_+ and T_- ; Significant if T < critical value.

One-sided P -value					One-sided P -value				
	0.05	0.025	0.01	0.005		0.05	0.025	0.01	0.005
Two-sided P -value					Two-sided P -value				
N	0.1	0.05	0.02	0.01	N	0.1	0.05	0.02	0.01
5	1				30	152	137	120	109
6	2	1			31	163	148	130	118
7	4	2	0		32	175	159	141	128
8	6	4	2	0	33	188	171	151	138
9	8	6	3	2	34	201	183	162	149
10	11	8	5	3	35	214	195	174	160
11	14	11	7	5	36	228	208	186	171
12	17	14	10	7	37	242	222	198	183
13	21	17	13	10	38	256	235	211	195
14	26	21	16	13	39	271	250	224	208
15	30	25	20	16	40	287	264	238	221
16	36	30	24	19	41	303	279	252	234
17	41	35	28	23	42	319	295	267	248
18	47	40	33	28	43	336	311	281	262
19	54	46	38	32	44	353	327	297	277
20	60	52	43	37	45	371	344	313	292
21	68	59	49	43	46	389	361	329	307
22	75	66	56	49	47	408	397	345	323
23	83	73	62	55	48	427	397	362	339
24	92	81	69	61	49	446	415	380	356
25	101	90	77	68	50	466	434	398	373
26	110	98	85	76					
27	120	107	93	84					
28	130	117	102	92					
29	141	127	111	100					

Table A7 Critical values for the Wilcoxon matched pairs signed rank test.

Reproduced from Table 21 of White *et al.* (1979) with permission of the authors and publishers.

N = number of non-zero differences; T = smaller of T_+ and T_- ; Significant if $T < \text{critical value}$.

One-sided P -value					One-sided P -value				
	0.05	0.025	0.01	0.005		0.05	0.025	0.01	0.005
Two-sided P -value					Two-sided P -value				
N	0.1	0.05	0.02	0.01	N	0.1	0.05	0.02	0.01
5	1				30	152	137	120	109
6	2	1			31	163	148	130	118
7	4	2	0		32	175	159	141	128
8	6	4	2	0	33	188	171	151	138
9	8	6	3	2	34	201	183	162	149
10	11	8	5	3	35	214	195	174	160
11	14	11	7	5	36	228	208	186	171
12	17	14	10	7	37	242	222	198	183
13	21	17	13	10	38	256	235	211	195
14	26	21	16	13	39	271	250	224	208
15	30	25	20	16	40	287	264	238	221
16	36	30	24	19	41	303	279	252	234
17	41	35	28	23	42	319	295	267	248
18	47	40	33	28	43	336	311	281	262
19	54	46	38	32	44	353	327	297	277
20	60	52	43	37	45	371	344	313	292
21	68	59	49	43	46	389	361	329	307
22	75	66	56	49	47	408	397	345	323
23	83	73	62	55	48	427	397	362	339
24	92	81	69	61	49	446	415	380	356
25	101	90	77	68	50	466	434	398	373
26	110	98	85	76					
27	120	107	93	84					
28	130	117	102	92					
29	141	127	111	100					

A non-parametric alternative to the one-sample t-test or the paired t-test

Example:

$N = 10$

$T_+ = 40$

$T_- = 15$

$\rightarrow T = \min\{T_+, T_-\} = 15$

$15 > 8$

\rightarrow do not reject H_0 at 5% level

Wilcoxon signed rank test: Procedure (2)

A non-parametric alternative to the one-sample t-test or the paired t-test

Test decision for large sample size:

$$Z = \frac{T - \frac{1}{4}N(N+1)}{\sqrt{\frac{N(N+1)(2N+1)}{24}}} = -1.274$$

$T = \min\{T_+, T_-\}$, N : #non-zero differences

z has Standard Normal Distribution \rightarrow compare with quantile $z_{1-\alpha/2}$ from SND

If $|z| > 1.96 \rightarrow$ reject H_0 at 5% level

Remark:

Confidence interval for median of differences can be calculated by tedious method.

Wilcoxon signed rank test: What is a large sample size?

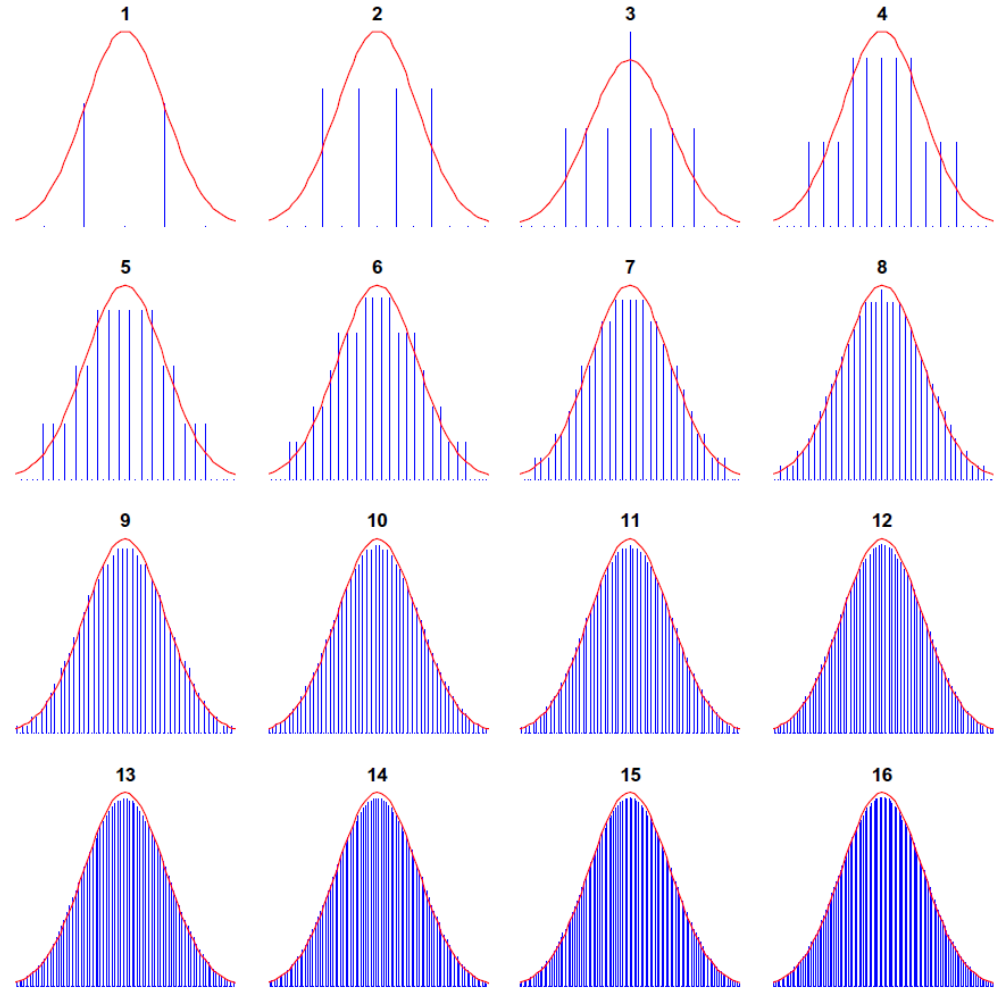


Figure 2. Exact distribution of the Wilcoxon signed-rank statistic for sample of sizes ($1 \leq n \leq 16$), and curve of the normal distribution with mean $n(n+1)/4$ and variance $n(n+1)(2n+1)/24$.

Wilcoxon signed rank test in SigmaPlot

The screenshot shows the SigmaPlot interface. The 'Statistics' menu is open, and the path 'Before and After' > 'Signed Rank Test...' is highlighted. In the background, a data table is visible with columns 1, 2, 3, and 4. Column 1 is labeled 'Patient' and contains values from 1 to 11. Column 2 is labeled 'Drug' and contains values from 1.0000 to 10.0000. Column 3 is labeled 'Placebo' and contains values from 6.1000 to 7.3000. Column 4 is empty.

Signed Rank Test 2 Report

Wilcoxon Signed Rank Test Saturday, December 10, 2011, 1:06:40 PM

Data source: Data 1 in Notebook1

Normality Test (Shapiro-Wilk) Passed (P = 0.344)

Group	N	Missing	Median	25%	75%
Col 2	11	1	6.600	5.950	7.450
Col 3	11	1	5.250	4.125	7.325

W = -25.000 T+ = 15.000 T- = -40.000
Z-Statistic (based on positive ranks) = -1.274
P(est.) = 0.221 P(exact) = 0.232

The change that occurred with the treatment is not great enough to exclude the possibility that it is due to chance (P = 0.232).

Wilcoxon signed rank test in Web tool

Treatment 1	Treatment 2	Sign	Abs	R	Sign R
6.1	5.2	1	0.9	2	2
6.0	7.9	-1	1.9	5	-5
8.2	3.9	1	4.3	10	10
7.6	4.7	1	2.9	8	8
6.5	5.3	1	1.2	3	3
5.4	7.4	-1	2	6	-6
6.9	4.2	1	2.7	7	7
6.7	6.1	1	0.6	1	1
7.4	3.8	1	3.6	9	9
5.8	7.3	-1	1.5	4	-4

Significance Level:

☐ .01

☒ .05

1 or 2-tailed hypothesis?:

☐ One-tailed

☒ Two-tailed

Result Details

W-value: 15
Mean Difference: -1.24
Sum of pos. ranks: 40
Sum of neg. ranks: 15

Z-value: -1.2741
Mean (W): 27.5
Standard Deviation (W): 9.81

Sample Size (N): 10

Result 1 - Z-value

The value of z is -1.2741. The p-value is .20408.

The result is not significant at $p < .05$.

Result 2 - W-value

The value of W is 15. The critical value for W at N = 10 ($p < .05$) is 8.

The result is not significant at $p < .05$.

<https://www.socscistatistics.com/tests/signedranks/default2.aspx>

Test Procedures for Quantitative Variables: Comparison of Means

Parametric test (assuming normal distribution)	<i>Non-parametric test</i>
<u>1 sample:</u> One-sample t-test	
<u>2 paired samples:</u> paired t-test	<i>Sign test, Wilcoxon signed rank test</i>
<u>2 independent samples:</u> t-test Welch-test	<i>Wilcoxon rank sum test (= Mann-Whitney U-test)</i>
<u>≥ 3 independent samples:</u> Analysis of Variance (ANOVA)	<i>Kruskal-Wallis-test</i>

Wilcoxon rank sum test: Example

Example:

Comparison of birth weights (in kg) of children born to 14 heavy smoking mothers with those of children born to 15 non-smoking mothers

Heavy smokers (group 1) weight	Non-smokers (group 0) weight
2.34	2.71
2.38	3.31
2.74	3.36
2.86	3.41
2.90	3.51
3.18	3.54
3.23	3.60
3.27	3.61
3.42	3.70
3.53	3.73
3.60	3.83
3.65	3.89
3.65	3.99
3.69	4.08
	4.13

Wilcoxon rank sum test: Example

Example:

Comparison of birth weights (in kg) of children born to 14 heavy smoking mothers with those of children born to 15 non-smoking mothers

Heavy smokers (group 1)		Non-smokers (group 0)	
weight	rank	weight	rank
2.34	1	2.71	3
2.38	2	3.31	10
2.74	4	3.36	11
2.86	5	3.41	12
2.90	6	3.51	14
3.18	7	3.54	16
3.23	8	3.60	
3.27	9	3.61	
3.42	13	3.70	
3.53	15	3.73	
3.60		3.83	
3.65		3.89	
3.65		3.99	
3.69		4.08	
		4.13	

Wilcoxon rank sum test: Example

Example:

Comparison of birth weights (in kg) of children born to 14 heavy smoking mothers with those of children born to 15 non-smoking mothers

Heavy smokers (group 1)		Non-smokers (group 0)	
weight	rank	weight	rank
2.34	1	2.71	3
2.38	2	3.31	10
2.74	4	3.36	11
2.86	5	3.41	12
2.90	6	3.51	14
3.18	7	3.54	16
3.23	8	3.60	17.5
3.27	9	3.61	
3.42	13	3.70	
3.53	15	3.73	
3.60	17.5	3.83	
3.65		3.89	
3.65		3.99	
3.69		4.08	
		4.13	

Wilcoxon rank sum test: Example

Example:

Comparison of birth weights (in kg) of children born to 14 heavy smoking mothers with those of children born to 15 non-smoking mothers

Heavy smokers (group 1)		Non-smokers (group 0)	
weight	rank	weight	rank
2.34	1	2.71	3
2.38	2	3.31	10
2.74	4	3.36	11
2.86	5	3.41	12
2.90	6	3.51	14
3.18	7	3.54	16
3.23	8	3.60	17.5
3.27	9	3.61	19
3.42	13	3.70	23
3.53	15	3.73	24
3.60	17.5	3.83	25
3.65	20.5	3.89	26
3.65	20.5	3.99	27
3.69	22	4.08	28
		4.13	29

Wilcoxon rank sum test: Example

Example:

Comparison of birth weights (in kg) of children born to 14 heavy smoking mothers with those of children born to 15 non-smoking mothers

Heavy smokers (group 1)		Non-smokers (group 0)	
weight	rank	weight	rank
2.34	1	2.71	3
2.38	2	3.31	10
2.74	4	3.36	11
2.86	5	3.41	12
2.90	6	3.51	14
3.18	7	3.54	16
3.23	8	3.60	17.5
3.27	9	3.61	19
3.42	13	3.70	23
3.53	15	3.73	24
3.60	17.5	3.83	25
3.65	20.5	3.89	26
3.65	20.5	3.99	27
3.69	22	4.08	28
Rank sum: $R_1 = 150.5$		4.13	29
		Rank sum: $R_0 = 284.5$	

Situation:

Two independent samples
No assumptions about distribution

Hypotheses:

$$H_0 : F_{\text{smokers}} = F_{\text{non-smokers}}$$
$$H_1 : F_{\text{smokers}} \neq F_{\text{non-smokers}}$$

Test statistic:

Follow algorithm:

1. Rank values from both groups together
2. T = rank sum in group with smaller sample size

Test decision for small sample size:

Use Table for critical values

If $T \leq W_{\alpha/2}$ or $T \geq W_{1-\alpha/2} \rightarrow$ reject H_0 at level α

Quantiles for Wilcoxon rank sum test

$$W_{\alpha/2}, W_{1-\alpha/2}$$

n_0, n_1	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.001$	n_0, n_1	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.001$
2, 8	3, 19			4, 13	18, 54	14, 58	10, 62
2, 9	3, 21			4, 14	19, 57	14, 62	10, 66
2, 10	3, 23			4, 15	20, 60	15, 65	10, 70
2, 11	4, 24			4, 16	21, 63	15, 69	11, 73
2, 12	4, 26			4, 17	21, 67	16, 72	11, 77
2, 13	4, 28			4, 18	22, 70	16, 76	11, 81
2, 14	4, 30			4, 19	23, 73	17, 79	12, 84
2, 15	4, 32			4, 20	24, 76	18, 82	12, 88
2, 16	4, 34			4, 21	25, 79	18, 86	12, 92
2, 17	5, 35			4, 22	26, 82	19, 89	13, 95
2, 18	5, 37			4, 23	27, 85	19, 93	13, 99
2, 19	5, 39	3, 41		4, 24	28, 88	20, 96	13, 103
2, 20	5, 41	3, 43		4, 25	28, 92	20, 100	14, 106
2, 21	6, 42	3, 45					
2, 22	6, 44	3, 47		5, 5	17, 38	15, 40	
2, 23	6, 46	3, 49		5, 6	18, 42	16, 44	
2, 24	6, 48	3, 51		5, 7	20, 45	17, 48	
2, 25	6, 50	3, 53		5, 8	21, 49	17, 53	
3, 5	6, 21			5, 9	22, 53	18, 57	15, 60
3, 6	7, 23			5, 10	23, 57	19, 61	15, 65
3, 7	7, 26			5, 11	24, 61	20, 65	16, 69
3, 8	8, 28			5, 12	26, 64	21, 69	16, 74
3, 9	8, 31	6, 33		5, 13	27, 68	22, 73	17, 78
3, 10	9, 33	6, 36		5, 14	28, 72	22, 78	17, 83
3, 11	9, 36	6, 39		5, 15	29, 76	23, 82	18, 87
3, 12	10, 38	7, 41		5, 16	31, 79	24, 86	18, 92
3, 13	10, 41	7, 44		5, 17	32, 83	25, 90	19, 96
3, 14	11, 43	7, 47		5, 18	33, 87	26, 94	19, 101
3, 15	11, 46	8, 49		5, 19	34, 91	27, 98	20, 105
3, 16	12, 48	8, 52		5, 20	35, 95	28, 102	20, 110
3, 17	12, 51	8, 55		5, 21	37, 98	29, 106	21, 114
3, 18	13, 53	8, 58		5, 22	38, 102	29, 111	21, 119
3, 19	13, 56	9, 60		5, 23	39, 106	30, 115	22, 123
3, 20	14, 58	9, 63		5, 24	40, 110	31, 119	23, 127
3, 21	14, 61	9, 66	6, 69	5, 25	42, 113	32, 123	23, 132
3, 22	15, 63	10, 68	6, 72				
3, 23	15, 66	10, 71	6, 75	6, 6	26, 52	23, 55	
3, 24	16, 68	10, 74	6, 78	6, 7	27, 57	24, 60	
3, 25	19, 71	11, 76	6, 81	6, 8	29, 61	25, 65	21, 69
				6, 9	31, 65	26, 70	22, 74
4, 4	10, 26			6, 10	32, 70	27, 75	23, 79
4, 5	11, 29			6, 11	34, 74	28, 80	23, 85
4, 6	12, 32	10, 34		6, 12	35, 79	30, 84	24, 90
4, 7	13, 35	10, 38		6, 13	37, 83	31, 89	25, 95
4, 8	14, 38	11, 41		6, 14	38, 88	32, 94	26, 100
4, 9	15, 41	11, 45		6, 15	40, 92	33, 99	26, 106
4, 10	15, 45	12, 48		6, 16	42, 96	34, 104	27, 111
4, 11	16, 48	12, 52		6, 17	43, 101	36, 108	28, 116
4, 12	17, 51	13, 55		6, 18	45, 105	37, 113	29, 121

Quantiles for Wilcoxon rank sum test

$$W_{\alpha/2}, W_{1-\alpha/2}$$

n_0, n_1	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.001$	n_0, n_1	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.001$
6, 19	46, 110	38, 118	29, 127	9, 15	79, 146	70, 155	60, 165
6, 20	48, 114	39, 123	30, 132	9, 16	82, 152	72, 162	61, 173
6, 21	50, 118	40, 128	31, 137	9, 17	84, 159	74, 169	63, 180
6, 22	51, 123	42, 132	32, 142	9, 18	87, 165	76, 176	65, 187
6, 23	53, 127	43, 137	33, 147	9, 19	90, 171	78, 183	66, 195
6, 24	55, 131	44, 142	34, 152	9, 20	93, 177	81, 189	68, 202
				9, 21	95, 184	83, 196	70, 209
7, 7	36, 69	32, 73	28, 77	10, 10	78, 132	71, 139	63, 147
7, 8	38, 74	34, 78	29, 83	10, 11	81, 139	74, 146	65, 155
7, 9	40, 79	35, 84	30, 89	10, 12	85, 145	76, 154	67, 163
7, 10	42, 84	37, 89	31, 95	10, 13	88, 152	79, 161	69, 171
7, 11	44, 89	38, 95	32, 101	10, 14	91, 159	81, 169	71, 179
7, 12	46, 94	40, 100	33, 107	10, 15	94, 166	84, 176	73, 187
7, 13	48, 99	41, 106	34, 113	10, 16	97, 173	86, 184	75, 195
7, 14	50, 104	43, 111	35, 119	10, 17	100, 180	89, 191	77, 203
7, 15	52, 109	44, 117	36, 125	10, 18	103, 187	92, 198	79, 211
7, 16	54, 114	46, 122	37, 131	10, 19	107, 193	94, 206	81, 219
7, 17	56, 119	47, 128	38, 137	10, 20	110, 200	97, 213	83, 227
7, 18	58, 124	49, 133	39, 143				
7, 19	60, 129	50, 139	41, 148	11, 11	96, 157	87, 166	78, 175
7, 20	62, 134	52, 144	42, 154	11, 12	99, 165	90, 174	81, 183
7, 21	64, 139	53, 150	43, 160	11, 13	103, 172	93, 182	83, 192
7, 22	66, 144	55, 155	44, 166	11, 14	106, 180	96, 190	85, 201
7, 23	68, 149	57, 160	45, 172	11, 15	110, 187	99, 198	87, 210
				11, 16	114, 194	102, 206	90, 218
8, 8	49, 87	43, 93	38, 98	11, 17	117, 202	105, 214	92, 227
8, 9	51, 93	45, 99	40, 104	11, 18	121, 209	108, 222	94, 236
8, 10	53, 99	47, 105	41, 111	11, 19	124, 217	111, 230	97, 244
8, 11	55, 105	49, 111	42, 118				
8, 12	58, 110	51, 117	43, 125	12, 12	115, 185	106, 194	95, 205
8, 13	60, 116	53, 123	45, 131	12, 13	119, 193	109, 203	98, 214
8, 14	63, 121	54, 130	46, 138	12, 14	123, 201	112, 212	100, 224
8, 15	65, 127	56, 136	47, 145	12, 15	127, 209	115, 221	103, 233
8, 16	67, 133	58, 142	49, 151	12, 16	131, 217	119, 229	105, 243
8, 17	70, 138	60, 148	50, 158	12, 17	135, 225	122, 238	108, 252
8, 18	72, 144	62, 154	51, 165	12, 18	139, 233	125, 247	111, 261
8, 19	74, 150	64, 160	53, 171				
8, 20	77, 155	66, 166	54, 178	13, 13	137, 214	125, 226	114, 237
8, 21	79, 161	68, 172	56, 184	13, 14	141, 223	129, 235	116, 248
8, 22	82, 166	70, 178	57, 191	13, 15	145, 232	133, 244	119, 258
				13, 16	150, 240	137, 253	122, 268
9, 9	63, 108	56, 115	50, 121	13, 17	154, 249	140, 263	125, 278
9, 10	65, 115	58, 122	52, 128				
9, 11	68, 121	61, 128	53, 136	14, 14	160, 246	147, 259	134, 272
9, 12	71, 127	63, 135	55, 143	14, 15	164, 256	151, 269	137, 283
9, 13	73, 134	65, 142	56, 151	14, 16	169, 265	155, 279	140, 294
9, 14	76, 140	67, 149	58, 158				
				15, 15	185, 280	171, 294	156, 309

Wilcoxon rank sum test: Procedure

A non-parametric alternative to the two-sample t-test

Situation:

Two independent samples
No assumptions about distribution

Hypotheses:

$$H_0 : F_{\text{smokers}} = F_{\text{non-smokers}}$$

$$H_1 : F_{\text{smokers}} \neq F_{\text{non-smokers}}$$

Test statistic:

Follow algorithm:

1. Rank values from both groups together
2. T = rank sum in group with smaller sample size

→ Example: $T = 150.5$

Test decision for small sample size:

$150.5 < 164 \rightarrow$ reject H_0 at 5% level

n_0, n_1	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.001$
14, 14	160, 246	147, 259	134, 272
14, 15	164, 256	151, 269	137, 283
14, 16	169, 265	155, 279	140, 294
15, 15	185, 280	171, 294	156, 309

Wilcoxon rank sum test: Procedure (2)

A non-parametric alternative to the two-sample t-test

Test decision for large sample size:

$$z = \frac{R_1 - \frac{n_1(n_0 + n_1 + 1)}{2}}{\sqrt{\frac{n_1 n_0 (n_0 + n_1 + 1)}{12}}} = -2.597$$

z has Standard Normal Distribution \rightarrow compare with quantile $z_{1-\alpha/2}$ from SND

If $|z| > 1.96 \rightarrow$ reject H_0 at 5% level

It makes no difference which group is chosen for evaluation of rank sum:

$$\frac{R_1 - \frac{n_1(n_0 + n_1 + 1)}{2}}{\sqrt{\frac{n_1 n_0 (n_0 + n_1 + 1)}{12}}} = - \frac{R_0 - \frac{n_0(n_0 + n_1 + 1)}{2}}{\sqrt{\frac{n_1 n_0 (n_0 + n_1 + 1)}{12}}}$$

Wilcoxon rank sum test

What is a large sample size?

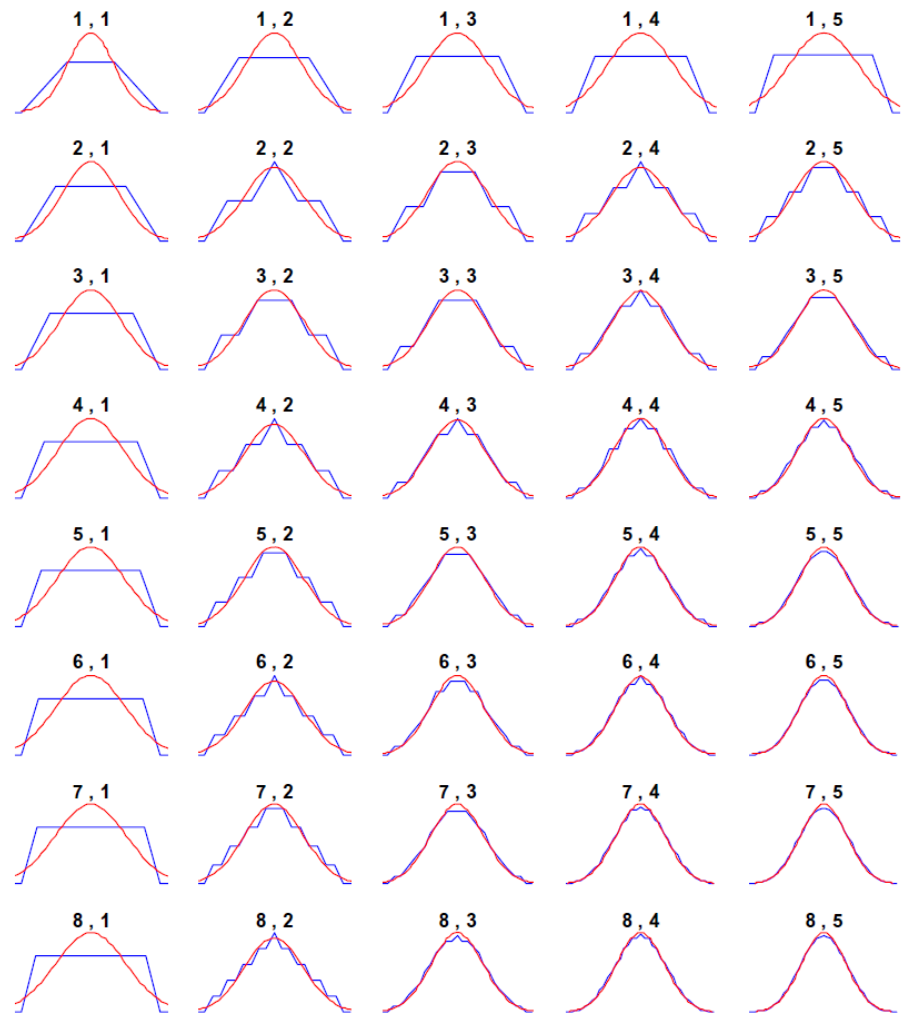


Figure 4. Exact distribution of the Wilcoxon rank-sum statistic for samples of sizes $n_1 = 1$ to 5 and $n_2 = 1$ to 8, and curve of the normal distribution with mean $(n_1 + n_2 + 1)/2$ and variance $n_1 n_2 (n_1 + n_2 + 1)/12$.

Wilcoxon rank sum test = Mann-Whitney U-test

$$U = R_1 - \frac{n_1(n_1 + 1)}{2} = 150.5 - 105 = 45.5$$

U: test statistic for Mann-Whitney U-test

- If distribution in the two groups differ only by shift in location, the Wilcoxon rank sum test tests for the difference of two medians.
- In this situation confidence interval for difference of medians:
Hodges-Lehmann estimate.
- If number of ties is large, then use correction for ties.

Wilcoxon rank sum test in SigmaPlot

A non-parametric alternative to the two-sample t-test

The screenshot displays the SigmaPlot software interface. The 'Statistics' menu is open, showing the 'Compare Two Groups' option selected. A submenu is visible with 't-test...' and 'Rank Sum Test...' options. The 'Rank Sum Test' option is highlighted. The 'Rank Sum Test 4 Report' window is open, showing the results of the Mann-Whitney Rank Sum Test. The report includes the data source, normality test results, a table of group statistics, the Mann-Whitney U statistic, and a conclusion.

Rank Sum Test 4 Report *

Mann-Whitney Rank Sum Test Saturday, December 10, 2011, 1:24:10 PM

Data source: Data 2 in Notebook1.JNB

Normality Test (Shapiro-Wilk) Failed ($P < 0.050$)

Group	N	Missing	Median	25%	75%
Col 2	15	0	3.610	3.410	3.890
Col 1	14	0	3.250	2.830	3.612

Mann-Whitney U Statistic= 45.500

T = 150.500 **n(small)**= 14 **n(big)**= 15 ($P = 0.010$)

The difference in the median values between the two groups is greater than would be expected by chance; there is a statistically significant difference ($P = 0.010$)

Data 2*

	1	2
1	2.3400	2.7100
2	2.3800	3.3100
3	2.7400	3.3600
4	2.8600	3.4100
5	2.9000	3.5100
6	3.1800	3.5400
7	3.2300	3.6000
8	3.2700	3.6100
9	3.4200	3.7000
10	3.5300	3.7300
11	3.6000	3.8300
12	3.6500	3.8900
13	3.6500	3.9900
14	3.6900	4.0800
15		4.1300

Wilcoxon rank sum test in Web tool

Mann-Whitney U Test Calculator

The value of U is 45.5.

You'll notice below that we have calculated a critical value for U based on alpha level and whether your hypothesis is one or two tailed. We have also calculated a value for Z and its associated *p*-value. Results in blue reach significance. Results in red do not.

Sample 1	Sample 2
2.34	2.71
2.38	3.31
2.74	3.36
2.86	3.41
2.90	3.51
3.18	3.54
3.23	3.60
3.27	3.61
3.42	3.70
3.53	3.73
3.60	3.83
3.65	3.89
3.65	3.99
3.69	4.08
	4.13

<https://www.socscistatistics.com/tests/mannwhitney/default2.aspx>

Significance Level:

☐ .01

☒ .05

1 or 2-tailed hypothesis?:

☐ One-tailed

☒ Two-tailed

The *U*-value is 45.5. The critical value of *U* at *p* < .05 is 59. Therefore, the result is significant at *p* < .05.

The *z*-score is 2.57497. The *p*-value is .01016. The result is significant at *p* < .05.

Wilcoxon rank sum test in Web tool

Sample 1	Sample 2	S1 Values	S1 Ranks	S2 Values	S2 Ranks
2.34	2.71	2.34	1	2.71	3
2.38	3.31	2.38	2	3.31	10
2.74	3.36	2.74	4	3.36	11
2.86	3.41	2.86	5	3.41	12
2.90	3.51	2.9	6	3.51	14
3.18	3.54	3.18	7	3.54	16
3.23	3.60	3.23	8	3.6	17.5
3.27	3.61	3.27	9	3.61	19
3.42	3.70	3.42	13	3.7	23
3.53	3.73	3.53	15	3.73	24
3.60	3.83	3.6	17.5	3.83	25
3.65	3.89	3.65	20.5	3.89	26
3.65	3.99	3.65	20.5	3.99	27
3.69	4.08	3.69	22	4.08	28
	4.13			4.13	29

Significance Level:

☐ 0.01

☒ 0.05

1 or 2-tailed hypothesis?:

☐ One-tailed

☒ Two-tailed

Result Details

Sample 1

Sum of ranks: 150.5

Mean of ranks: 10.75

Expected sum of ranks: 210

Expected mean of ranks: 15

U-value: 164.5

Expected *U*-value: 105

Sample 2

Sum of ranks: 284.5

Mean of ranks: 18.97

Expected sum of ranks: 225

Expected mean of ranks: 15

U-value: 45.5

Expected *U*-value: 105

Sample 1 & 2 Combined

Sum of ranks: 435

Mean of ranks: 15

Standard Deviation: 22.9129

Result 1 - *U*-value

The *U*-value is 45.5. The critical value of *U* at $p < .05$ is 59. Therefore, the result is significant at $p < .05$.

Result 2 - *Z*-ratio

The *Z*-Score is 2.57497. The *p*-value is .01016. The result is significant at $p < .05$.

Test Procedures for Quantitative Variables: Comparison of Means

Parametric test (assuming normal distribution)	<i>Non-parametric test</i>
<u>1 sample:</u> One-sample t-test	
<u>2 paired samples:</u> paired t-test	<i>Sign test, Wilcoxon signed rank test</i>
<u>2 independent samples:</u> t-test Welch-test	<i>Wilcoxon rank sum test (= Mann-Whitney U-test)</i>
<u>≥ 3 independent samples:</u> Analysis of Variance (ANOVA)	<i>Kruskal-Wallis-test</i>

Kruskal-Wallis test: Example

A non-parametric alternative to ANOVA

Effect of gene knock down on tumor volume (cf. ANOVA lecture, 10 Oct 2018):

Wild	knock.A	knock.B
26	50	20
46	75	50
63	60	6
20	72	48
61	120	72
56	80	60

Kruskal-Wallis test: Example

A non-parametric alternative to ANOVA

Effect of gene knock down on tumor volume:

Wild	rank	knock.A	rank	knock.B	rank
26	4	50	7.5	20	2.5
46	5	75	16	50	7.5
63	13	60	10.5	6	1
20	2.5	72	14.5	48	6
61	12	120	18	72	14.5
56	9	80	17	60	10.5

Kruskal-Wallis test: Example

A non-parametric alternative to ANOVA

Effect of gene knock down on tumor volume:

Wild	rank	knock.A	rank	knock.B	rank
26	4	50	7.5	20	2.5
46	5	75	16	50	7.5
63	13	60	10.5	6	1
20	2.5	72	14.5	48	6
61	12	120	18	72	14.5
56	9	80	17	60	10.5
Rank sum $R_1 = 45.5$		Rank sum $R_2 = 83.5$		Rank sum $R_3 = 42$	
$n_1 = 6$		$n_2 = 6$		$n_3 = 6$	
$\bar{R}_1 = R_1/n_1 = 7.583$		$\bar{R}_2 = R_2/n_2 = 13.912$		$\bar{R}_3 = R_3/n_3 = 7.0$	

Situation:

Three or more independent samples

No assumptions about distribution, but distributions have same shape

Hypotheses:

$$H_0 : F_1 = F_2 = \dots = F_k$$

H_1 : not all distributions are equal

Test statistic:

Follow algorithm:

1. Rank values from all groups together
2. Calculate in each group rank sum R_i and mean rank \bar{R}_i
3. Calculate test statistic

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k n_i \left(\bar{R}_i - \frac{N+1}{2} \right)^2 = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

If number of ties large, then use correction for ties.

Test decision:

H has χ^2_{k-1} -distribution: If $H > \chi^2_{k-1, 1-\alpha} \rightarrow$ reject H_0 at level α

Kruskal-Wallis test: Example calculation

A non-parametric alternative to ANOVA

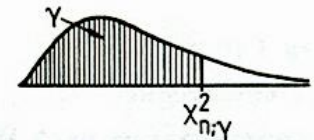
Test statistic:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k n_i \left(\bar{R}_i - \frac{N+1}{2} \right)^2 = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

$$= \frac{12}{18(18+1)} \frac{45.5^2 + 83.5^2 + 42^2}{6} - 3 \cdot (18+1) = 6.196$$

Test decision:

H has $\chi^2_{3-1} = \chi^2_2$ -distribution: $6.196 > 5.991 \rightarrow$ reject H_0 at 5% level



Tab. 4: Quantile $\chi^2_{n;\gamma}$ der χ^2 -Verteilung

γ	0,995	0,990	0,975	0,950	0,900	0,750	0,500	0,250	0,100	0,050	0,025	0,010	0,005
n													
1	7,879	6,635	5,024	3,841	2,706	1,323	0,455	0,102	$^{-2}1,58$	$^{-3}3,93$	$^{-4}9,82$	$^{-4}1,57$	$^{-5}3,93$
2	10,60	9,210	7,378	5,991	4,605	2,773	1,386	0,575	0,211	0,103	$^{-2}5,06$	$^{-2}2,01$	$^{-2}1,00$
3	12,84	11,34	9,348	7,815	6,251	4,108	2,366	1,213	0,584	0,352	0,216	0,115	$^{-2}7,17$
4	14,86	13,28	11,14	9,488	7,779	5,385	3,357	1,923	1,064	0,711	0,484	0,297	0,207
5	16,75	15,09	12,83	11,07	9,236	6,626	4,351	2,675	1,610	1,145	0,831	0,554	0,412

Kruskal-Wallis test in SigmaPlot

Kruskal-Wallis One Way Analysis of Variance on Ranks Monday, October 31, 2016, 5:17:55 PM

Data source: Data 1 in NotebookKW.JNB

Normality Test (Shapiro-Wilk): Passed ($P = 0.879$)

Equal Variance Test (Brown-Forsythe): Passed ($P = 0.926$)

Group	N	Missing	Median	25%	75%
wildtype	6	0	51.000	24.500	61.500
knock.A	6	0	73.500	57.500	90.000
knock.B	6	0	49.000	16.500	63.000

$H = 6.222$ with 2 degrees of freedom. ($P = 0.045$)

The differences in the median values among the treatment groups are greater than would be expected by chance; there is a statistically significant difference ($P = 0.045$)

To isolate the group or groups that differ from the others use a multiple comparison procedure.

All Pairwise Multiple Comparison Procedures (Tukey Test):

Comparison	Diff of Ranks	q	P	$P < 0.050$
knock.A vs knock.B	41.500	3.174	0.064	No
knock.A vs wildtype	38.000	2.906	0.099	Do Not Test
wildtype vs knock.B	3.500	0.268	0.980	Do Not Test

Note: The multiple comparisons on ranks do not include an adjustment for ties.

A result of "Do Not Test" occurs for a comparison when no significant difference is found between the two rank sums that enclose that comparison. For example, if you had four rank sums sorted in order, and found no significant difference between rank sums 4 vs. 2, then you would not test 4 vs. 3 and 3 vs. 2, but still test 4 vs. 1 and 3 vs. 1 (4 vs. 3 and 3 vs. 2 are enclosed by 4 vs. 2: 4 3 2 1). Note that not testing the enclosed rank sums is a procedural rule, and a result of Do Not Test should be treated as if there is no significant difference between the rank sums, even though one may appear to exist.

Summary Info

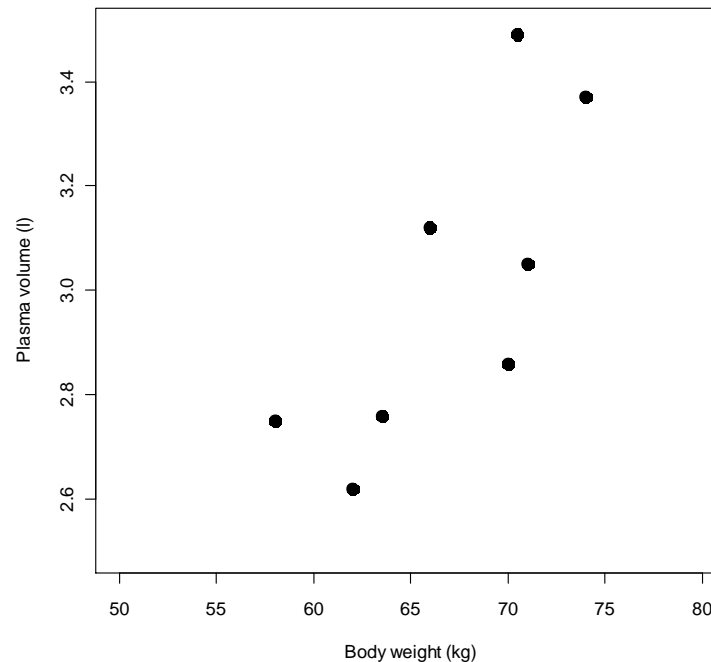
Created	10/31/2016 5:1...
Modified	10/31/2016 5:1...
Author	ZDV
Description	Report

Bivariate Analysis:

Investigating the association between two variables

A non-parametric alternative to Pearson correlation:

Spearman rank correlation

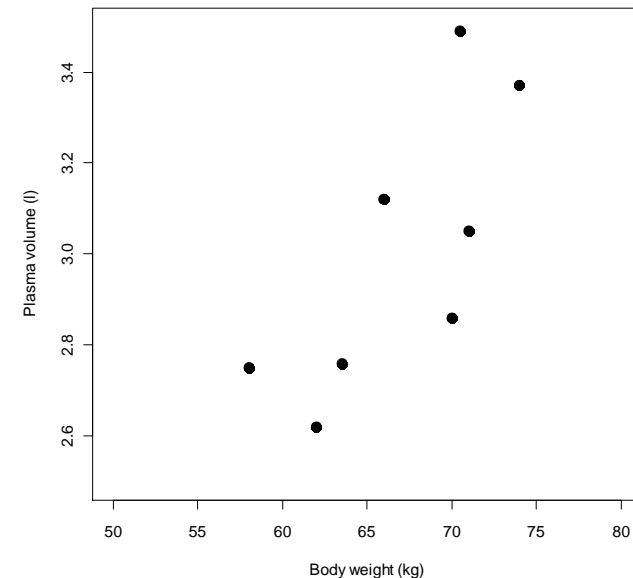


Spearman rank correlation: Example

A non-parametric alternative to Pearson correlation

Example: Plasma volume and body weight in eight healthy men:

Subject	Body weight (kg) value	Plasma volume (l) value
1	58.0	2.75
2	70.0	2.86
3	74.0	3.37
4	63.5	2.76
5	62.0	2.62
6	70.5	3.49
7	71.0	3.05
8	66.0	3.12



Spearman rank correlation

A non-parametric alternative to Pearson correlation

- Independently rank the values of X and Y
- Calculate Pearson's correlation for the ranks

Subject	Body weight (kg)		Plasma volume (l)	
	value	rank	value	rank
1	58.0	1	2.75	2
2	70.0	5	2.86	4
3	74.0	8	3.37	7
4	63.5	3	2.76	3
5	62.0	2	2.62	1
6	70.5	6	3.49	8
7	71.0	7	3.05	5
8	66.0	4	3.12	6

Spearman rank correlation: test

A non-parametric alternative to Pearson correlation

Significance test for Spearman rank correlation coefficient r_s :
Use test for Pearson correlation coefficient if $n \geq 10$, i.e.

$$t = r_s \sqrt{\frac{n-2}{1-r_s^2}}$$

and compare with t distribution, $n-2$ d.f.

$$r_s = 0.81$$

$$t = 3.378$$

$$t = 3.378 > 2.447 = t_{6,0.975}$$

→ reject H_0 at 5% significance level

Quantiles $t_{n,\gamma}$ of t-distribution

$\gamma \backslash n$	0,990	0,975	0,950	0,900
1	31,821	12,706	6,314	3,078
2	6,965	4,303	2,920	1,886
3	4,541	3,182	2,353	1,638
4	3,747	2,776	2,132	1,533
5	3,365	2,571	2,015	1,476
6	3,143	2,447	1,943	1,440
7	2,998	2,365	1,895	1,415

Spearman rank correlation in SigmaPlot

A non-parametric alternative to Pearson correlation

The screenshot displays the SigmaPlot software interface. The top menu bar includes Home, Worksheet, Create Graph, Graph Page, Analysis, Report, and ToolBox. The Analysis menu is open, showing options like Spearman Correlation, Sample size, Power, and Options. The Report menu is also open, showing options like Regression Wizard, Dynamic Fit Wizard, Global Curve Fit Wizard, Histogram, Smoothers, Plot Equation, and Linear. The Notebook Manager on the left shows a hierarchy: All Open Notebooks > Notebook1* > Section 1 > Data 1* > Spearman Correlation 1 > Spearman Correlation. The main window displays the 'Spearman Correlation 1 Report' for 'Data 1 in Notebook1'. The report includes the title 'Spearman Rank Order Correlation', the date 'Monday, November 09, 2015, 10:07:32 AM', and the data source. The cell contents are listed as 'Correlation Coefficient', 'P Value', and 'Number of Samples'. A table shows the correlation between 'body weight' and 'plasma volume' with a correlation coefficient of 0.810, a P value of 0.00960, and 8 samples. The report concludes with a paragraph explaining the interpretation of the correlation results.

Spearman Rank Order Correlation Monday, November 09, 2015, 10:07:32 AM

Data source: Data 1 in Notebook1

Cell Contents:
Correlation Coefficient
P Value
Number of Samples

	plasma volume
body weight	0.810
	0.00960
	8

plasma volume

The pair(s) of variables with positive correlation coefficients and P values below 0.050 tend to increase together. For the pairs with negative correlation coefficients and P values below 0.050, one variable tends to decrease while the other increases. For pairs with P values greater than 0.050, there is no significant relationship between the two variables.

Non-parametric methods

Advantages ...

- Don't require the assumption that data come from a Normal distribution.
- Robust to outliers.
- Robust against misspecification of distribution assumption.
- For large sample size almost as powerful as parametric methods.
- (Easy to calculate manually)

Disadvantages ...

- Less power than parametric tests when data come from Normal distribution.
- May have power 0 for small samples: e.g. two-sided Wilcoxon rank sum test for 3 vs. 3 samples at 5% level.
- Difficult to obtain confidence intervals.
- Not easily extended to regression models.

When to use non-parametric methods

- If no data transformation is available to transform data to normal distribution.
- In the presence of outliers.
- From formal perspective:
Automatic choice between parametric and non-parametric test, e.g. by Shapiro-Wilks test for normality, is not advised.
But from practical perspective: preliminary testing for normality does not seem to cause much harm (see Rochon et al. BMC Medical Research Methodology 2012, 12:81).
- Similar data should be evaluated with the same test.
- **Do not** run both parametric and non-parametric test and choose p-value you like!
- For large samples non-parametric nearly as powerful as parametric tests.

11 November

Diagnostic tests

Dr. Diana Tichy