Wilcoxon & Kruskal-Wallis

Dr Azmi Mohd Tamil

Explore

- It is the first step in the analytic process
- to explore the characteristics of the data
- to screen for errors and correct them
- to look for distribution patterns - normal distribution or not
- May require transformation before further analysis using parametric methods
- Or may need analysis using non-parametric techniques

Choosing an appropriate method

- Number of groups of observations
- Independent or dependent groups of observations
- Type of data
- Distribution of data
- Objective of analysis

Nonparametric Test Procedures

- Statistic does not depend on population distribution
- Data may be nominally or ordinally scaled
  - Example: Male-female
- May involve population parameters such as median
- Based on analysis of ranks
- Example: Wilcoxon rank sum test

non-parametric tests

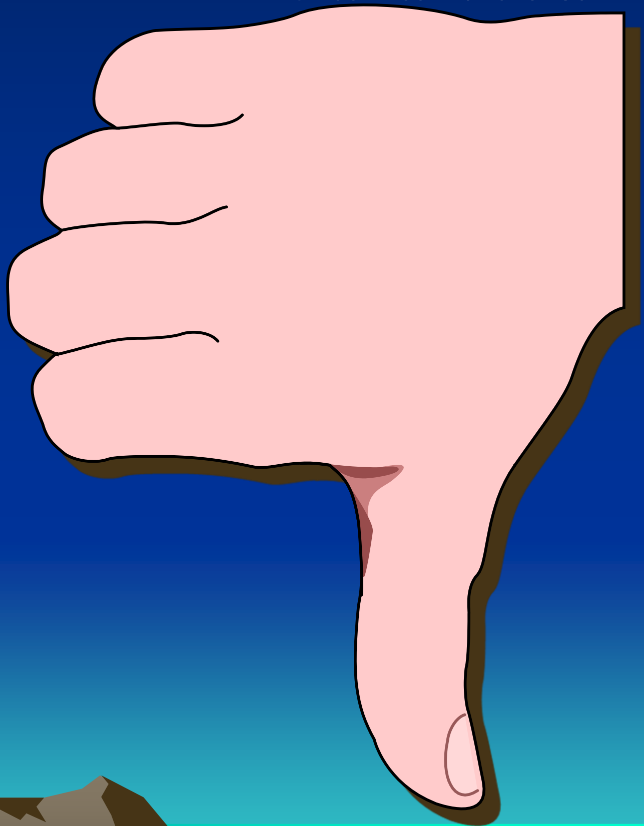
Variable 1	Variable 2	Criteria	Type of Test
Qualitative Dichotomus	Qualitative Dichotomus	Sample size < 20 or (< 40 but with at least one expected value < 5)	Fisher Test
Qualitative Dichotomus	Quantitative	Data not normally distributed	Wilcoxon Rank Sum Test or U Mann-Whitney Test
Qualitative Polinomial	Quantitative	Data not normally distributed	Kruskal-Wallis One Way ANOVA Test
Quantitative	Quantitative	Repeated measurement of the same individual & item	Wilcoxon Rank Sign Test
Continous or ordinal	Quantitative - continous	Data not normally distributed	Spearman/Kendall Rank Correlation

Advantages of Non-parametric Tests

- Used with all scales
- Easier to calculate
  - Developed before wide computer use
- Make fewer assumptions
- Need not involve population parameters
- Results may be as exact as parametric procedures

Disadvantages of Non-parametric Tests

- May waste information
  - If data permit using parametric procedures
  - Example: Converting data from ratio to ordinal scale
- Difficult to calculate by hand for large samples
- Tables not widely available



Mann-Whitney U Test/ Wilcoxon Rank Sum

- Non-parametric comparison of 2 groups
- Requires all the observations to be ranked as if they were from a single sample

Mann-Whitney U Test/ Wilcoxon Rank Sum

- Tests two independent population medians
- Non-parametric test procedure
- Assumptions
  - Ordinal, interval, or ratio scaled data
  - Population is nearly symmetrical
    - Bell-shaped, rectangular etc.
- Can use normal approximation if $n_i > 10$

Mann-Whitney U Test/ Wilcoxon Rank Sum Procedure

- Assign ranks, R_i, to the $n_1 + n_2$ sample observations
  - If unequal sample sizes, let n_1 refer to smaller-sized sample
  - Smallest value = rank of 1
  - Same value -> Average ties
- Sum the ranks, T_i, for each group
- Test statistic is T_1 (smallest group)

Example

- Comparing the blood glucose level between taxi drivers (code 3) and bus drivers (code 1)

nores	kerja	glu
234	1	124
243	1	141
244	1	93.6
410	3	139
508	3	104
821	3	105
829	3	96.2
832	3	95

Example step 2

- Arrange the blood glucose level in ascending order. Give rank from lowest to highest.
- If the same values, take the mean rank.

nores	kerja	glu	rank
244	1	93.6	1
832	3	95	2
829	3	96.2	3
508	3	104	4
821	3	105	5
234	1	124	6
410	3	139	7
243	1	141	8

Example step 3

- Arrange the blood glucose level in ascending order. Give rank from lowest to highest.
- If the same values, take the mean rank.

nores	kerja	glu	rank
244	1	93.6	1
832	3	95	2
829	3	96.2	3
508	3	104	4
821	3	105	5
234	1	124	6
410	3	139	7
243	1	141	8

Example step 3

- Total up the rank in the smaller group
  - Bus drivers; T = 6+8+1=15
  - Taxi drivers; T = 7+4+5+3+2=21
- Compare the result with the respective table at n1 and n2 ; 3, 5.
- T is between the two critical range (6 – 21). Only significant if T= or <6,
  or T= or > 21.
- Conclusion: p > 0.05; **Null Hypothesis NOT REJECTED**

Table A8 Critical ranges for the Wilcoxon rank sum test.

n_1, n_2 = sample sizes of two groups
T = sum of ranks in group with smaller sample size
Significant if T on boundaries or outside critical range

	One-sided P value		
	0.025	0.005	0.0005
	Two-sided P value		
n_1, n_2	0.05	0.01	0.001
2, 8	3, 19		
2, 9	3, 21		
2, 10	3, 23		
2, 11	4, 24		
2, 12	4, 26		
2, 13	4, 28		
2, 14	4, 30		
2, 15	4, 32		
2, 16	4, 34		
2, 17	5, 35		
2, 18	5, 37		
2, 19	5, 39	3, 41	
2, 20	5, 41	3, 43	
2, 21	6, 42	3, 45	
2, 22	6, 44	3, 47	
2, 23	6, 46	3, 49	
2, 24	6, 48	3, 51	
2, 25	6, 50	3, 53	
3, 5	6, 21		
3, 6	7, 23		
3, 7	7, 26		
3, 8	8, 28		
3, 9	8, 31	6, 33	
3, 10	9, 33	6, 36	
3, 11	9, 36	6, 39	
3, 12	10, 38	7, 41	
3, 13	10, 41	7, 44	
3, 14	11, 43	7, 47	
3, 15	11, 46	8, 49	
3, 16	12, 48	8, 52	
3, 17	12, 51	8, 55	
3, 18	13, 53	8, 58	
3, 19	13, 56	9, 60	
3, 20	14, 58	9, 63	
3, 21	14, 61	9, 66	6, 69
3, 22	15, 63	10, 68	6, 72
3, 23	15, 66	10, 71	6, 75
3, 24	16, 68	10, 74	6, 78
3, 25	16, 71	11, 76	6, 81
4, 4	10, 26		
4, 5	11, 29		
4, 6	12, 32	10, 34	
4, 7	13, 35	10, 38	
4, 8	14, 38	11, 41	
4, 9	15, 41	11, 45	

	One-sided P value		
	0.025	0.005	0.0005
	Two-sided P value		
n_1, n_2	0.05	0.01	0.001
4, 10	15, 45	12, 48	
4, 11	16, 48	12, 52	
4, 12	17, 51	13, 55	
4, 13	18, 54	14, 58	10, 62
4, 14	19, 57	14, 62	10, 66
4, 15	20, 60	15, 65	10, 70
4, 16	21, 63	15, 69	11, 73
4, 17	21, 67	16, 72	11, 77
4, 18	22, 70	16, 76	11, 81
4, 19	23, 73	17, 79	12, 84
4, 20	24, 76	18, 82	12, 88
4, 21	25, 79	18, 86	12, 92
4, 22	26, 82	19, 89	13, 95
4, 23	27, 85	19, 93	13, 99
4, 24	28, 88	20, 96	13, 103
4, 25	28, 92	20, 100	14, 106
5, 5	17, 38	15, 40	
5, 6	18, 42	16, 44	
5, 7	20, 45	17, 48	
5, 8	21, 49	17, 53	
5, 9	22, 53	18, 57	15, 60
5, 10	23, 57	19, 61	[illegible] 65
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
5, 14	28, 72	[illegible]	[illegible]
5, 15	29, 76	23, 82	[illegible]
5, 16	31, 79	24, 86	18, 92
5, 17	32, 83	25, 90	19, 96
5, 18	33, 87	26, 94	19, 101
5, 19	34, 91	27, 98	20, 105
5, 20	35, 95	28, 102	20, 110
5, 21	37, 98	29, 106	21, 114
5, 22	38, 102	29, 111	21, 119
5, 23	39, 106	30, 115	22, 123
5, 24	40, 110	31, 119	23, 127
5, 25	42, 113	32, 123	23, 132
6, 6	26, 52	23, 55	
6, 7	27, 57	24, 60	
6, 8	29, 61	25, 65	21, 69
6, 9	31, 65	26, 70	22, 74
6, 10	32, 70	27, 75	23, 79
6, 11	34, 74	28, 80	23, 85
6, 12	35, 79	30, 84	24, 90
6, 13	37, 83	31, 89	25, 95

Refer to Table A8.
Look at n1, n2 ; 3, 5.

For p=0.05, the critical range is < 6 or > 21. Only significant if; T < 6,
or T > 21.
Therefore p > 0.05

	Two-sided P value		
n_1, n_2	0.05	0.01	0.001
3, 5	6, 21		
3, 6	7, 23		
3, 7	7, 26		
3, 8	8, 28		
3, 9	8, 31	6, 33	

Appendix Table E

Table E Mann-Whitney test on unpaired samples: 5% and 1% levels of P
5% Critical points of rank sums

n1→ n2	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4			10											
5		6	11	17										
6		7	12	18	26									
7		7	13	20	27	36								
8	3	8	14	21	29	38	49							
9	3	8	15	22	31	40	51	63						
10	3	9	15	23	32	42	53	65	78					
11	4	9	16	24	34	44	55	68	81	96				
12	4	10	17	26	35	46	58	71	85	99	115			
13	4	10	18	27	37	48	60	73	88	103	119	137		
14	4	11	19	28	38	50	63	76	91	106	123	141	160	
15	4	11	20	29	40	52	65	79	94	110	127	145	164	185
16	4	12	21	31	42	54	67	82	97	114	131	150	169	
17	5	12	21	32	43	56	70	84	100	117	135	154		
18	5	13	22	33	45	58	72	87	103	121	139			
19	5	13	23	34	46	60	74	90	107	124				

SPSS Output

Ranks

	KERJA	N	Mean Rank	Sum of Ranks
GLU	1.00	3	5.00	15.00
	3.00	5	4.20	21.00
	Total	8		

Test Statistics[b]

	GLU
Mann-Whitney U	6.000
Wilcoxon W	21.000
Z	-.447
Asymp. Sig. (2-tailed)	.655
Exact Sig. [2*(1-tailed Sig.)]	.786[a]

a. Not corrected for ties.

b. Grouping Variable: KERJA

The only way for the result to be significant

Is for all the data of the smallest group to be at one end or the other.

Assume the results were like this

- The bus drivers all had lower blood glucose level compared to the taxi drivers.

nores	kerja	glu	rank
244	1	93.6	1
832	1	95	2
829	1	96.2	3
508	3	104	4
821	3	105	5
234	3	124	6
410	3	139	7
243	3	141	8

Now the result is significant

- Total up the rank in the smaller group
  - Bus drivers; T = 1+2+3=6
  - Taxi drivers; T = 4+5+6+7+8=30
- Compare the result with the respective table at n1 and n2 ; 3, 5.
- T is exactly the value of the lower critical range (6 – 21). Now significant since T= or < 6
- Conclusion: $p < 0.05$; **Null Hypothesis is REJECTED**

Table A8 Critical ranges for the Wilcoxon rank sum test.

Reproduced from Table A7 of Cotton (1974) with permission of the author and publishers.

n_1, n_2 = sample sizes of two groups
T = sum of ranks in group with smaller sample size
Significant if T on boundaries or outside critical range

	One-sided P value		
	0.025	0.005	0.0005
	Two-sided P value		
n_1, n_2	0.05	0.01	0.001
2, 8	3, 19		
2, 9	3, 21		
2, 10	3, 23		
2, 11	4, 24		
2, 12	4, 26		
2, 13	4, 28		
2, 14	4, 30		
2, 15	4, 32		
2, 16	4, 34		
2, 17	5, 35		
2, 18	5, 37		
2, 19	5, 39	3, 41	
2, 20	5, 41	3, 43	
2, 21	6, 42	3, 45	
2, 22	6, 44	3, 47	
2, 23	6, 46	3, 49	
2, 24	6, 48	3, 51	
2, 25	6, 50	3, 53	
3, 5	6, 21		
3, 6	7, 23		
3, 7	7, 26		
3, 8	8, 28		
3, 9	8, 31	6, 33	
3, 10	9, 33	6, 36	
3, 11	9, 36	6, 39	
3, 12	10, 38	7, 41	
3, 13	10, 41	7, 44	
3, 14	11, 43	7, 47	
3, 15	11, 46	8, 49	
3, 16	12, 48	8, 52	
3, 17	12, 51	8, 55	
3, 18	13, 53	8, 58	
3, 19	13, 56	9, 60	
3, 20	14, 58	9, 63	
3, 21	14, 61	9, 66	6, 69
3, 22	15, 63	10, 68	6, 72
3, 23	15, 66	10, 71	6, 75
3, 24	16, 68	10, 74	6, 78
3, 25	19, 71	11, 76	6, 81
4, 4	10, 26		
4, 5	11, 29		
4, 6	12, 32	10, 34	
4, 7	13, 35	10, 38	
4, 8	14, 38	11, 41	
4, 9	15, 41	11, 45	
4, 10	15, 45	12, 48	
4, 11	16, 48	12, 52	
4, 12	17, 51	13, 55	
4, 13	18, 54	14, 58	10, 62
4, 14	19, 57	14, 62	10, 66
4, 15	20, 60	15, 65	10, 70
4, 16	21, 63	15, 69	11, 73
4, 17	21, 67	16, 72	11, 77
4, 18	22, 70	16, 76	11, 81
4, 19	23, 73	17, 79	12, 84
4, 20	24, 76	18, 82	12, 88
4, 21	25, 79	18, 86	12, 92
4, 22	26, 82	19, 89	13, 95
4, 23	27, 85	19, 93	13, 99
4, 24	28, 88	20, 96	13, 103
4, 25	28, 92	20, 100	14, 106
5, 5	17, 38	15, 40	
5, 6	18, 42	16, 44	
5, 7	20, 45	17, 48	
5, 8	21, 49	17, 53	
5, 9	22, 53	18, 57	15, 60
5, 10	23, 57	19, 61	[illegible], 65
[illegible]	[illegible]	[illegible]	[illegible]
5, 14	28, [illegible]	[illegible]	[illegible]
5, 15	29, 76	23, 82	[illegible]
5, 16	31, 79	24, 86	18, 92
5, 17	32, 83	25, 90	19, 96
5, 18	33, 87	26, 94	19, 101
5, 19	34, 91	27, 98	20, 105
5, 20	35, 95	28, 102	20, 110
5, 21	37, 98	29, 106	21, 114
5, 22	38, 102	29, 111	21, 119
5, 23	39, 106	30, 115	22, 123
5, 24	40, 110	31, 119	23, 127
5, 25	42, 113	32, 123	23, 132
6, 6	26, 52	23, 55	
6, 7	27, 57	24, 60	
6, 8	29, 61	25, 65	21, 69
6, 9	31, 65	26, 70	22, 74
6, 10	32, 70	27, 75	23, 79
6, 11	34, 74	28, 80	23, 85
6, 12	35, 79	30, 84	24, 90
6, 13	37, 83	31, 89	25, 95

Refer to Table A8.

Look at n1, n2 ; 3, 5.

For p=0.05, the critical range is ≤ 6 or ≥ 21. Only significant if; T ≤ 6,
or T ≥ 21.
Therefore $p < 0.05$

	Two-sided P value		
n_1, n_2	0.05	0.01	0.001
3, 5	6, 21		
3, 6	7, 23		
3, 7	7, 26		
3, 8	8, 28		
3, 9	8, 31	6, 33	

Let’s try it the other way!

- The bus drivers all had higher blood glucose level compared to the taxi drivers.

nores	kerja	glu	rank
244	3	93.6	1
832	3	95	2
829	3	96.2	3
508	3	104	4
821	3	105	5
234	1	124	6
410	1	139	7
243	1	141	8

Now the result is also significant

- Total up the rank in the smaller group
  - Bus drivers; T = 6+7+8=21
  - Taxi drivers; T = 1+2+3+4+5=15
- Compare the result with the respective table at n1 and n2 ; 3, 5.
- T is exactly the value of the upper critical range (6 – 21). Now significant since T= or > 21
- Conclusion: $p < 0.05$; **Null Hypothesis is REJECTED**

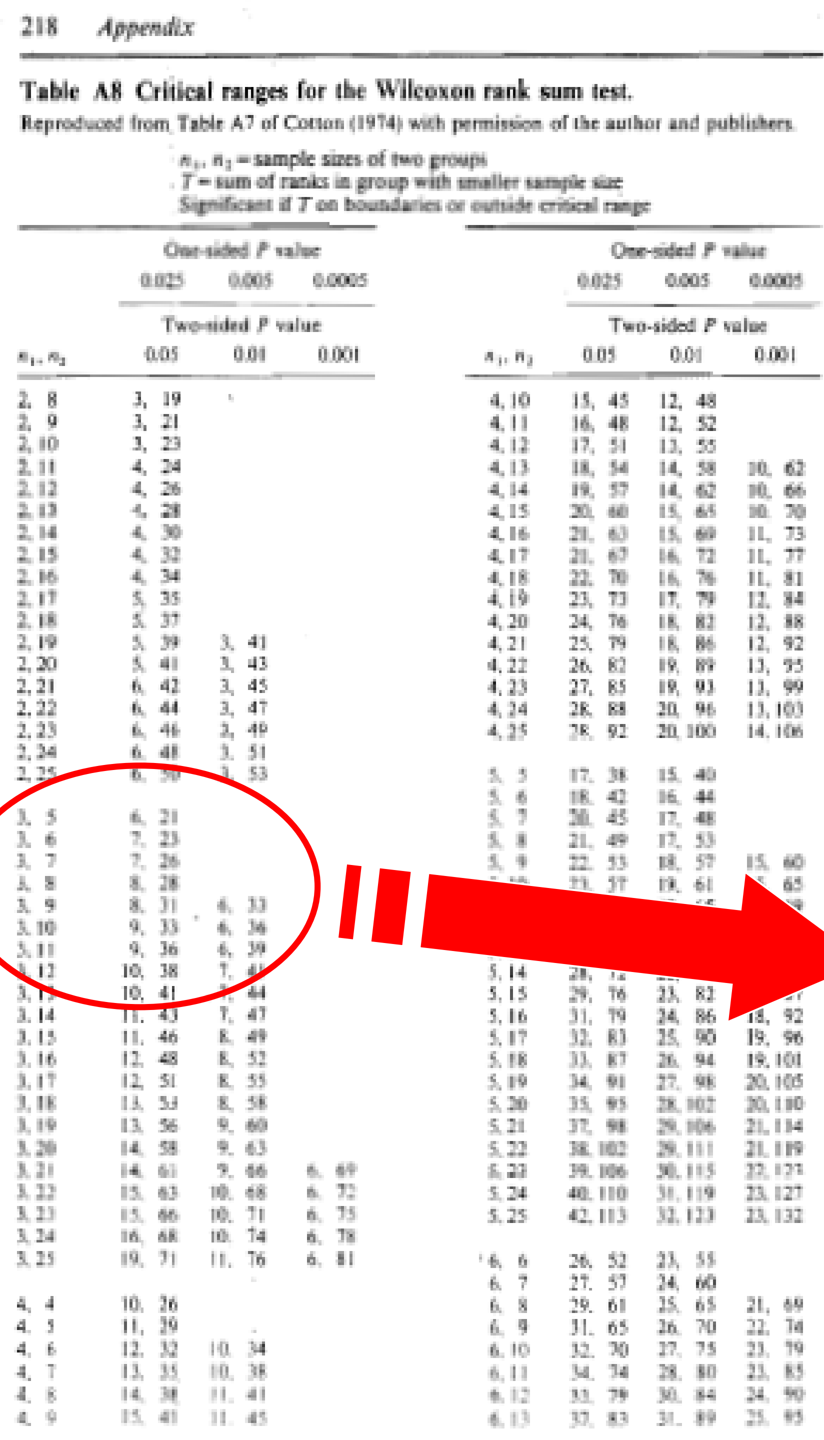


Table A8 Critical ranges for the Wilcoxon rank sum test.

Reproduced from Table A7 of Cotton (1974) with permission of the author and publishers.

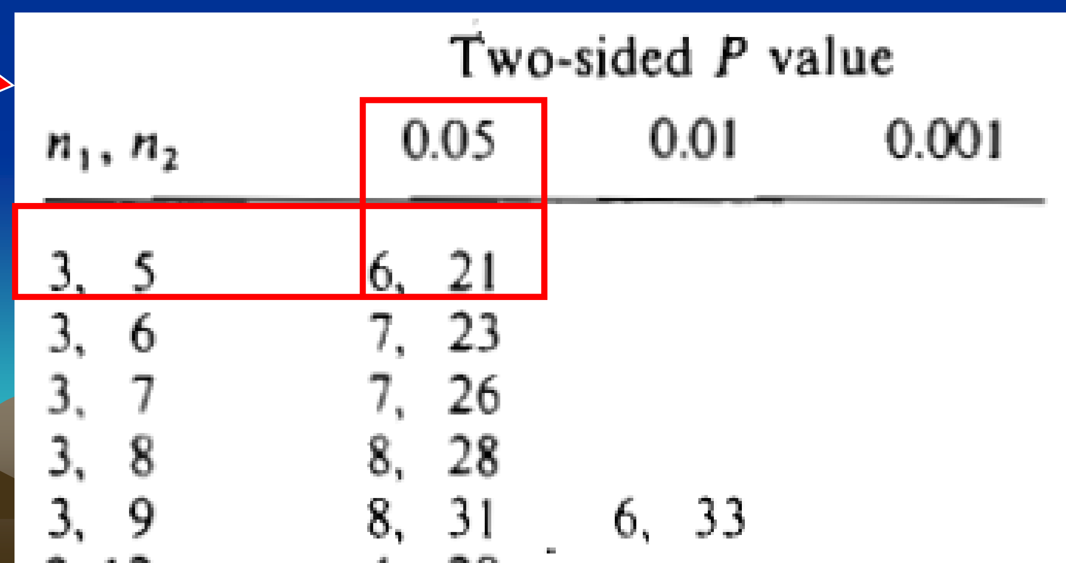
n_1, n_2 = sample sizes of two groups
T = sum of ranks in group with smaller sample size
Significant if T on boundaries or outside critical range

	One-sided P value 0.025	0.005	0.0005
n_1, n_2	Two-sided P value 0.05	0.01	0.001
2, 8	3, 19		
2, 9	3, 21		
2, 10	3, 23		
2, 11	4, 24		
2, 12	4, 26		
2, 13	4, 28		
2, 14	4, 30		
2, 15	4, 32		
2, 16	4, 34		
2, 17	5, 35		
2, 18	5, 37		
2, 19	5, 39	3, 41	
2, 20	5, 41	3, 43	
2, 21	6, 42	3, 45	
2, 22	6, 44	3, 47	
2, 23	6, 46	3, 49	
2, 24	6, 48	3, 51	
2, 25	6, 50	3, 53	
3, 5	6, 21		
3, 6	7, 23		
3, 7	7, 26		
3, 8	8, 28		
3, 9	8, 31	6, 33	
3, 10	9, 33	6, 36	
3, 11	9, 36	6, 39	
3, 12	10, 38	7, 41	
3, 13	10, 41	7, 44	
3, 14	11, 43	7, 47	
3, 15	11, 46	8, 49	
3, 16	12, 48	8, 52	
3, 17	12, 51	8, 55	
3, 18	13, 53	8, 58	
3, 19	13, 56	9, 60	
3, 20	14, 58	9, 63	
3, 21	14, 61	9, 66	6, 69
3, 22	15, 63	10, 68	6, 72
3, 23	15, 66	10, 71	6, 75
3, 24	16, 68	10, 74	6, 78
3, 25	16, 71	11, 76	6, 81
4, 4	10, 26		
4, 5	11, 29		
4, 6	12, 32	10, 34	
4, 7	13, 35	10, 38	
4, 8	14, 38	11, 41	
4, 9	15, 41	11, 45	

	One-sided P value 0.025	0.005	0.0005
n_1, n_2	Two-sided P value 0.05	0.01	0.001
4, 10	15, 45	12, 48	
4, 11	16, 48	12, 52	
4, 12	17, 51	13, 55	
4, 13	18, 54	14, 58	10, 62
4, 14	19, 57	14, 62	10, 66
4, 15	20, 60	15, 65	10, 70
4, 16	21, 63	15, 69	11, 73
4, 17	21, 67	16, 72	11, 77
4, 18	22, 70	16, 76	11, 81
4, 19	23, 73	17, 79	12, 84
4, 20	24, 76	18, 82	12, 88
4, 21	25, 79	18, 86	12, 92
4, 22	26, 82	19, 89	13, 95
4, 23	27, 85	19, 93	13, 99
4, 24	28, 88	20, 96	13, 103
4, 25	28, 92	20, 100	14, 106
5, 5	17, 38	15, 40	
5, 6	18, 42	16, 44	
5, 7	20, 45	17, 48	
5, 8	21, 49	17, 53	
5, 9	22, 53	18, 57	15, 60
5, 10	23, 57	19, 61	[illegible], 65
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
5, 14	28, [illegible]	[illegible]	[illegible]
5, 15	29, 76	23, 82	[illegible]
5, 16	31, 79	24, 86	18, 92
5, 17	32, 83	25, 90	19, 96
5, 18	33, 87	26, 94	19, 101
5, 19	34, 91	27, 98	20, 105
5, 20	35, 95	28, 102	20, 110
5, 21	37, 98	29, 106	21, 114
5, 22	38, 102	29, 111	21, 119
5, 23	39, 106	30, 115	22, 123
5, 24	40, 110	31, 119	23, 127
5, 25	42, 113	32, 123	23, 132
6, 6	26, 52	23, 55	
6, 7	27, 57	24, 60	
6, 8	29, 61	25, 65	21, 69
6, 9	31, 65	26, 70	22, 74
6, 10	32, 70	27, 75	23, 79
6, 11	34, 74	28, 80	23, 85
6, 12	35, 79	30, 84	24, 90
6, 13	37, 83	31, 89	25, 95

Refer to Table A8.
Look at n1, n2 ; 3, 5.

For p=0.05, the critical range is $\leq$ 6 to $\geq$ 21. Only significant if; T $\leq$ 6,
or T $\geq$ 21.
Therefore $p < 0.05$



n_1, n_2	Two-sided P value 0.05	0.01	0.001
3, 5	6, 21		
3, 6	7, 23		
3, 7	7, 26		
3, 8	8, 28		
3, 9	8, 31	6, 33	

Kruskal-Wallis test

- When there is 3 or more independent groups of observation

Kruskal-Wallis Rank Test for *c* Medians

- Tests the equality of more than 2 (*c*) population medians
- Non-parametric test procedure
- Used to analyze completely randomized experimental designs
- Can use χ^2 distribution to approximate if each sample group size $n_j \geq 5$
  - Degrees of freedom = $c - 1$

Kruskal-Wallis Rank Test Assumptions

- Independent random samples are drawn
- Continuous dependent variable
- Ordinal, interval, or ratio scaled data
- Populations have same variability
- Populations have same shape

Kruskal-Wallis Rank Test Procedure

- Assign ranks, R_i , to the n combined observations
  - Smallest value = 1
  - Largest value = n
  - Average ties
- Test statistic

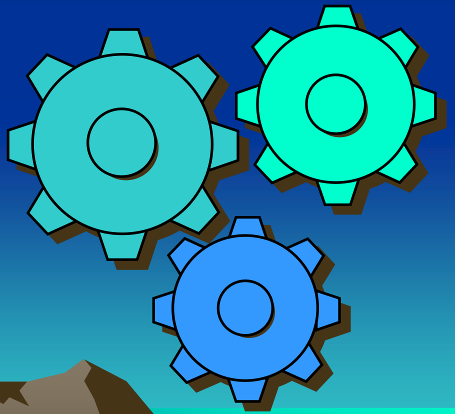
Squared total rank of each group

$$H = \frac{12}{n(n+1)} \sum \frac{T_i^2}{n_i} - 3(n+1)$$

Kruskal-Wallis Rank Test Example

As production manager, you want to see if 3 filling machines have different median filling times. You assign 15 similarly trained & experienced workers, 5 per machine, to the machines. At the .05 level, is there a difference in median filling times?

Mach1	Mach2	Mach3
25.40	23.40	20.00
26.31	21.80	22.20
24.10	23.50	19.75
23.74	22.75	20.60
25.10	21.60	20.40



Kruskal-Wallis Rank Test Solution

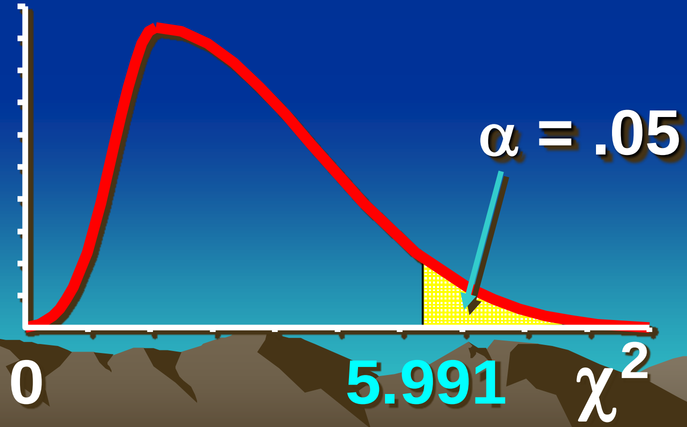
H_0: $M_1 = M_2 = M_3$

H_1: Not all equal

$\alpha = .05$

$df = c - 1 = 3 - 1 = 2$

Critical Value(s):



Test Statistic:

Decision:

Conclusion:

Obtaining Ranks Solution

Raw Data			Ranks		
Mach1	Mach2	Mach3	Mach1	Mach2	Mach3
25.40	23.40	20.00	14	9	2
26.31	21.80	22.20	15	6	7
24.10	23.50	19.75	12	10	1
23.74	22.75	20.60	11	8	4
25.10	21.60	20.40	13	5	3
			65	38	17

Test Statistic Solution

$$H = \left(\frac{12}{n \cdot (n+1)} \cdot \sum_{j=1}^{c} \frac{T_j^2}{n_j} \right) - 3 \cdot (n+1)$$

$$= \left(\frac{12}{(15) \cdot (16)} \cdot \left(\frac{(65)^2}{5} + \frac{(38)^2}{5} + \frac{(17)^2}{5} \right) \right) - 3 \cdot (16)$$

$$= \left(\frac{12}{240} \right) \cdot (1191.6) - 48$$

$$= 11.58$$

Mach1	Mach2	Mach3
14	9	2
15	6	7
12	10	1
11	8	4
13	5	3
65	38	17

Kruskal-Wallis Rank Test Solution

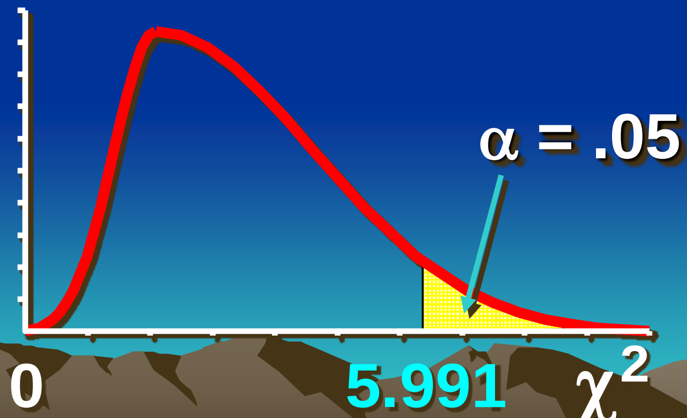
H_0: $M_1 = M_2 = M_3$

H_1: Not all equal

$\alpha = .05$

$df = c - 1 = 3 - 1 = 2$

Critical Value(s):



Test Statistic:

$$H = 1158$$

Refer to Chi-Square table

Decision:

Reject at $\alpha = .05$

Conclusion:

There is evidence pop. medians are different

Table 3 : Percentage point of X^2

	P Value							
d.k.	0.5	0.25	0.1	0.05	0.025	0.01	0.005	0.001
1	0.45	1.32	2.71	3.84	5.02	6.63	7.88	10.83
2	1.39	2.77	4.61	5.99	7.38	9.21	10.60	13.82
3	2.37	4.11	6.25	7.81	9.35	11.34	12.84	16.27
4	3.36	5.39	7.78	9.49	11.14	13.28	14.86	18.47
5	4.35	[illegible]	9.24	11.07	12.83	15.09	[illegible]	20.52
6	5.35	7.84	10.64	[illegible]	[illegible]	16.81	18.55	22.46
7	6.35	9.04	12.02	14.07	16.01	18.48	20.28	24.32
8	7.34	10.22	13.36	15.51	[illegible]	20.09	21.96	26.13
9	8.34	11.39	14.68	16.92	[illegible]	21.67	23.59	27.88
10	9.34	12.55	15.99	18.31	[illegible]	[illegible]	25.19	29.59
11	10.34	13.70	17.28	19.68	21.92	[illegible]	26.76	31.26
12	11.34	14.85	18.55	21.03	23.34	[illegible]	28.30	32.91
13	12.34	15.98	19.81	22.36	24.74	[illegible]	29.69	34.53
14	13.34	17.12	21.06	23.68	26.12	[illegible]	31.32	36.12
15	14.34	18.25	22.31	25.00	27.49	[illegible]	32.80	37.70
16	15.34	19.37	23.54	26.30	28.85	[illegible]	[illegible]	39.25
17	16.34	20.49	24.77	27.59	30.19	[illegible]	[illegible]	40.79
18	17.34	21.60	25.99	28.87	31.53	34.81	[illegible]	42.31
19	18.34	22.72	27.20	30.14	32.85	36.19	[illegible]	43.82
20	19.34	23.83	28.41	31.41	34.17	37.57	[illegible]	45.32
21	20.34	24.93	29.62	32.67	35.48	38.93	[illegible]	46.80
22	21.34	26.04	30.81	33.92	36.78	40.29	[illegible]	48.27
23	22.34	27.14	32.01	35.17	38.08	41.64	[illegible]	[illegible]
24	23.34	28.24	33.20	36.42	39.36	42.98	[illegible]	[illegible]
25	24.34	29.34	34.38	37.65	40.65	44.31	46.93	[illegible]
26	25.34	30.43	35.56	38.89	41.92	45.64	48.29	[illegible]
27	26.34	31.53	36.74	40.11	43.19	46.96	49.64	[illegible]
28	27.34	32.62	37.92	41.34	44.46	48.28	50.99	[illegible]
29	28.34	33.71	39.09	42.56	45.72	49.59	52.34	[illegible]
30	29.34	34.80	40.26	43.77	46.98	50.89	53.67	[illegible]
40	39.34	45.62	51.81	55.76	59.34	63.69	66.77	[illegible]
50	49.33	56.33	63.17	67.50	71.42	76.15	79.49	[illegible]
60	59.33	66.98	74.40	79.08	83.30	88.38	91.95	[illegible]
70	69.33	77.58	85.53	90.53	95.02	100.43	104.22	112.32
80	79.33	88.13	96.58	101.88	106.63	112.33	116.32	124.84
90	89.33	98.65	107.57	113.15	118.14	124.12	128.30	137.21
100	99.33	109.14	118.50	124.34	129.56	135.81	140.17	149.45

Refer to Table 3.
Look at df = 2.
H = 11.58, larger than 10.60 (p=0.005) but smaller than 13.82 (p=0.001).
13.82>11.58>10.60
Therefore if H=11.58, 0.001<p<0.005.

d.k.	0.5	0.25	0.1	0.05	0.025	0.01	0.005	0.001
1	0.45	1.32	2.71	3.84	5.02	6.63	7.88	10.83
2	1.39	2.77	4.61	5.99	7.38	9.21	10.60	13.82
3	2.37	4.11	6.25	7.81	9.35	11.34	12.84	16.27
4	3.36	5.39	7.78	9.49	11.14	13.28	14.86	18.47

Distribution of μ						
	Probability					
d.f.	**0.5**	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
1	0.455	2.706	3.841	5.412	6.635	10.827
2	1.386	4.605	5.991	7.824	9.210	13.815
3	2.366	6.251	7.815	9.837	11.345	16.268
4	3.357	7.779	9.488	11.668	13.277	18.465
5	4.351	9.236	11.070	13.388	15.086	20.517

SPSS Output

Ranks

	MESIN	N	Mean Rank
MASA	1.00	5	13.00
	2.00	5	7.60
	3.00	5	3.40
	Total	15	

Test Statistics[a,b]

	masa
Chi-Square	11.580
df	2
Asymp. Sig.	.003

a. Kruskal Wallis Test

b. Grouping Variable: mesin

Wilcoxon Signed Rank Test

- Two groups of paired observations

Example

- Whether there is any difference of the systolic blood pressure taken at 2 different time for 36 patients.

nores	bps1	bps2
237	147	131
835	166	150
251	159	147
809	150	139
241	170	160
233	164	155
272	154	145
239	186	178
261	155	147
246	176	170
247	186	181
254	155	150
258	151	147
288	152	148
829	115	111
257	162	159

Step 2

- Calculate the difference between the two values.
- Rank them accordingly, ignoring + or -.
- Total up the + & - separately

nores	bps1	bps2	d	absd	rank
237	147	131	-16	16	32.5
835	166	150	-16	16	32.5
251	159	147	-12	12	31
809	150	139	-11	11	30
241	170	160	-10	10	29
233	164	155	-9	9	27.5
272	154	145	-9	9	27.5
239	186	178	-8	8	25
261	155	147	-8	8	25
246	176	170	-6	6	22
247	186	181	-5	5	18
254	155	150	-5	5	18
258	151	147	-4	4	14
288	152	148	-4	4	14
829	115	111	-4	4	14
257	162	159	-3	3	11.5
269	161	158	-3	3	11.5
273	138	136	-2	2	9
406	146	144	-2	2	9
232	164	163	-1	1	4
270	183	182	-1	1	4
252	144	144	0	T-	409
402	135	135	0		
409	144	144	0	T+	152
231	164	165	1	1	4
249	157	158	1	1	4
407	116	117	1	1	4
823	171	172	1	1	4
825	176	177	1	1	4
236	156	158	2	2	9
242	170	175	5	5	18
250	142	147	5	5	18
260	164	169	5	5	18
285	134	140	6	6	22
824	137	143	6	6	22

Step 3

- Total up the ranks of the positives and the negatives. These are T+ dan T-.
- T+ = 152 and T- = 409
- Take the smaller value i.e. 152 and refer to the appropriate table. Critical value for n = 33 (3 zero values so 36 - 3) for significance at 0.05 is 171. Therefore < critical range.
- Therefore : **Null hypothesis rejected.**
- *Conclusion: There is a sig difference of blood pressure measured at two different times. BP before rest is sig higher than after rest.*

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

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

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

	One-sided P value			
	0.05	0.025	0.01	0.005
	Two-sided P value			
N	0.1	0.05	0.02	0.01
5	1			
6	2	1		
7	4	2	0	
8	6	4	2	0
9	8	6	3	2
10	11	8	5	3
11	14	11	7	5
12	17	14	10	7
13	21	17	13	10
14	26	21	16	13
15	30	25	20	16
16	36	30	24	19
17	41	35	28	23
18	47	40	33	28
19	54	46	38	32
20	60	52	43	37
21	68	59	49	43
22	75	66	56	49
23	83	73	62	55
24	92	81	69	61
25	101	90	77	68
26	110	98	85	76
27	120	107	93	84
28	130	117	102	92
29	141	127	111	100
30	152	137	120	109
31	163	148	130	118
32	175	159	141	128
33	188	171	151	138
34	201	183	162	149
35	214	195	174	160
36	228	208	186	171
37	242	222	198	183
38	[illegible]	[illegible]	[illegible]	195
39	[illegible]	[illegible]	[illegible]	208
40	[illegible]	[illegible]	[illegible]	221
41	[illegible]	[illegible]	[illegible]	234
42	[illegible]	[illegible]	[illegible]	248
43	[illegible]	[illegible]	[illegible]	262
44	[illegible]	[illegible]	[illegible]	277
45	[illegible]	[illegible]	[illegible]	292
46	[illegible]	[illegible]	[illegible]	307
47	[illegible]	[illegible]	[illegible]	323
48	[illegible]	[illegible]	[illegible]	339
49	[illegible]	[illegible]	[illegible]	356
50	[illegible]	[illegible]	[illegible]	373

Refer to Table A7.
Look at n=33.

Take the smallest value T+=152. Critical value for n = 33 (3 zero values) for significance at 0.05 is 171. Therefore 152 < critical range; $0.02 < p < 0.05$

	Two-sided P value			
N	0.1	0.05	0.02	0.01
32	175	159	141	128
33	188	171	151	138
34	201	183	162	149

Table D Wilcoxon test on paired samples: 5% and 1% levels of P		
Number of pairs	**5% Level**	**1% Level**
7	2	0
8	2	0
9	6	2
10	8	3
11	11	5
12	14	7
13	17	10
14	21	13
15	25	16
16	30	19

SPSS Output

Ranks

		N	Mean Rank	Sum of Ranks
BPS2 - BPS1	Negative Ranks	21[a]	19.48	409.00
	Positive Ranks	12[b]	12.67	152.00
	Ties	3[c]		
	Total	36		

a. BPS2 < BPS1

b. BPS2 > BPS1

c. BPS2 = BPS1

Test Statistics[b]

	BPS2 - BPS1
Z	-2.300[a]
Asymp. Sig. (2-tailed)	.021

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Spearman/Kendall Correlation

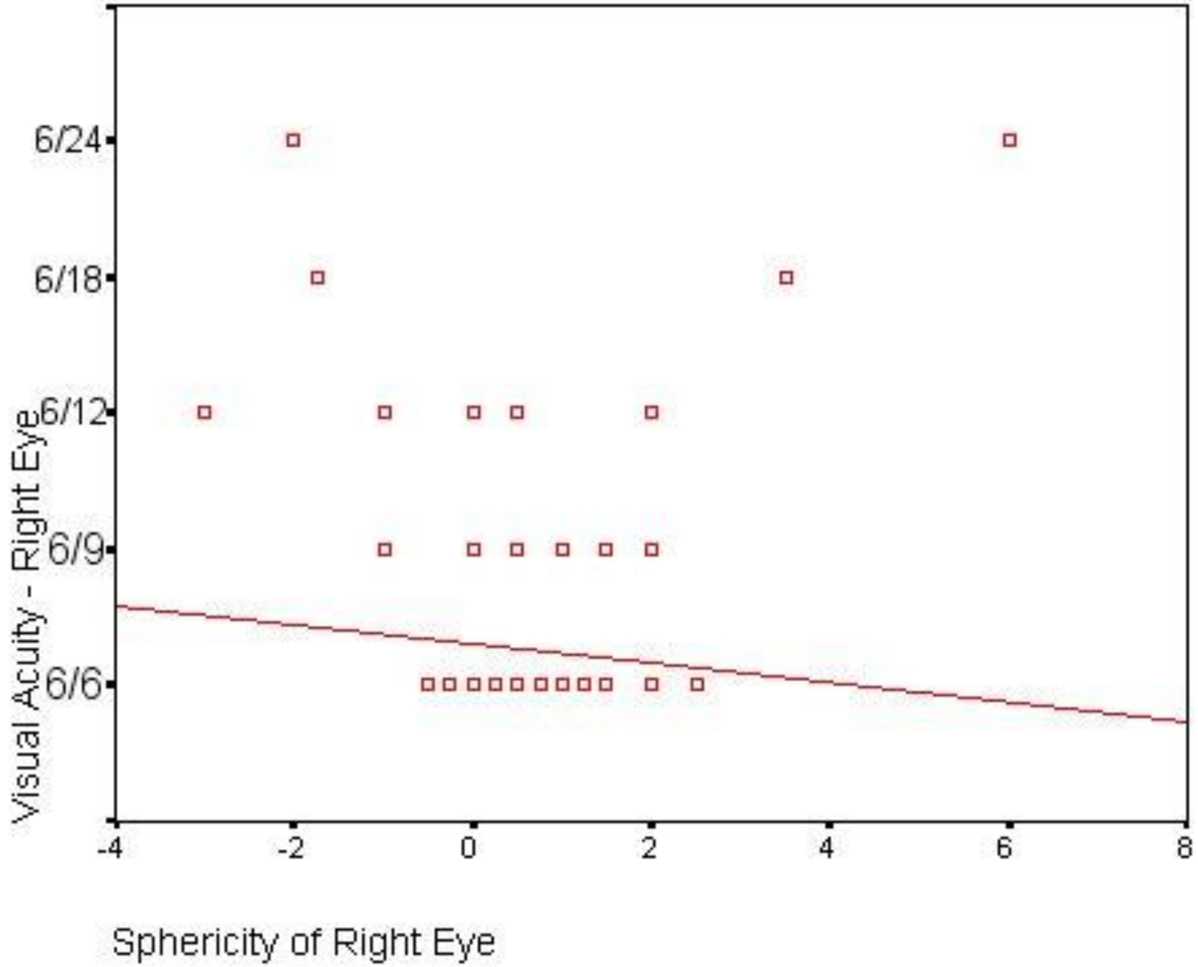
- To find correlation between a related pair of continuous data; or
- **Between 1 Continuous, 1 Categorical Variable (Ordinal)**
  - **e.g., association between Likert Scale on work satisfaction and work output.**

Spearman's rank correlation coefficient

- In statistics, **Spearman's rank correlation coefficient**, named for Charles Spearman and often denoted by the Greek letter ρ (rho), is a non-parametric measure of correlation – that is, it assesses how well an arbitrary monotonic function could describe the relationship between two variables, without making any assumptions about the frequency distribution of the variables. Unlike the Pearson product-moment correlation coefficient, it does not require the assumption that the relationship between the variables is linear, nor does it require the variables to be measured on interval scales; it can be used for variables measured at the ordinal level.

Example

- Correlation between sphericity and visual acuity.
- Sphericity of the eyeball is continuous data while visual acuity is ordinal data (6/6, 6/9, 6/12, 6/18, 6/24), therefore Spearman correlation is the most suitable.
- The Spearman rho correlation coefficient is -0.108 and p is 0.117. P is larger than 0.05, therefore there is no significant association between sphericity and visual acuity.



Correlations

			Visual Acuity - Right Eye	Sphericity of Right Eye
Spearman's rho	Visual Acuity - Right Eye	Correlation Coefficient	1.000	-.108
		Sig. (2-tailed)	.	.117
		N	215	211
	Sphericity of Right Eye	Correlation Coefficient	-.108	1.000
		Sig. (2-tailed)	.117	.
		N	211	211

Example 2

•- Correlation between glucose level and systolic blood pressure.
•Based on the data given, prepare the following table;
•For every variable, sort the data by rank. For ties, take the average.
•Calculate the difference of rank, d for every pair and square it. Take the total.
•Include the value into the following formula;

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

•$\sum d^2 = 4921.5$ n = 32
•Therefore $r_s = 1-((6*4921.5)/(32*(32^2-1)))$
= 0.097966.
This is the value of Spearman correlation coefficient (or ϒ).
•Compare the value against the Spearman table;
•p is larger than 0.05.
•Therefore there is no association between systolic BP and blood glucose level.

nores	glu	rank x	bps1	rank y	d	d2
231	123	23	164	25.5	-2.5	6.25
232	97	9	164	25.5	-16.5	272.25
233	325	32	164	25.5	6.5	42.25
234	124	24	118	7	17	289
235	107	12.5	126	8	4.5	20.25
236	95.7	8	156	20	-12	144
237	122	22	147	16	6	36
238	112	17	105	3	14	196
239	119	20	186	31.5	-11.5	132.25
240	132	25	112	5	20	400
241	105	11	170	28.5	-17.5	306.25
242	219	30	170	28.5	1.5	2.25
243	141	26	99	1.5	24.5	600.25
244	93.6	4	99	1.5	2.5	6.25
245	206	29	110	4	25	625
246	113	18.5	176	30	-11.5	132.25
247	167	28	186	31.5	-3.5	12.25
248	95.6	7	134	11	-4	16
249	108	14.5	157	21	-6.5	42.25
250	297	31	142	14	17	289
251	109	16	159	22	-6	36
252	100	10	144	15	-5	25
253	83.3	2	129	9	-7	49
254	145	27	155	18.5	8.5	72.25
255	90.2	3	140	13	-10	100
256	113	18.5	117	6	12.5	156.25
257	108	14.5	162	23	-8.5	72.25
258	121	21	151	17	4	16
259	94.5	6	137	12	-6	36
260	69.4	1	164	25.5	-24.5	600.25
261	94.2	5	155	18.5	-13.5	182.25
274	107	12.5	133	10	2.5	6.25
						4921.5

Spearman's table

- 0.097966 is the value of Spearman correlation coefficient (or ϒ).
- Compare the value against the Spearman table;
- p is larger than 0.05.
- Therefore there is no association between systolic BP and blood glucose level.

N (the number of pairs of scores):	0.05	0.02	0.01
5	1	1	
6	0.886	0.943	1
7	0.786	0.893	0.929
8	0.738	0.833	0.881
9	0.683	0.783	0.833
10	0.648	0.746	0.794
12	0.591	0.712	0.777
14	0.544	0.645	0.715
16	0.506	0.601	0.665
18	0.475	0.564	0.625
20	0.45	0.534	0.591
22	0.428	0.508	0.562
24	0.409	0.485	0.537
26	0.392	0.465	0.515
28	0.377	0.448	0.496
30	0.364	0.432	0.478

SPSS Output

Correlations

			GLU	BPS1
Spearman's rho	GLU	Correlation Coefficient	1.000	.097
		Sig. (2-tailed)	.	.599
		N	32	32
	BPS1	Correlation Coefficient	.097	1.000
		Sig. (2-tailed)	.599	.
		N	32	32

Presentation

- Never sufficient to present the results solely as a p value
- Median and quartiles should be given
- For small samples, the median and range can be given

Take Home Message

- Both parametric & non-parametric methods can be used for continuous data
- Parametric methods are preferred if all assumptions are met
- Reasons
  - Amenable to estimation and confidence interval
  - Readily extended to the more complicated data structures

Take Home Message

- We should carry out either parametric or non-parametric test on a set of data, not both.
- When there are doubts of the validity of the assumptions for the parametric method, a non parametric analysis is carried out
- If the assumptions are met, the two methods should give very similar answers
- If the answers differ, then the non-parametric method is more likely to be reliable