EXAMPLES 4 — nonparametric difference tests

Many of these examples are also suitable for further practice with *t* tests.

Mann–Whitney *U* test

Find the value of *U* for each of the following pairs of groups of observations, and discover whether the difference between the groups is significant at the 0.05 level, two-tailed.

Q1.	Group A:	43	39	57	62						
	Group B:	51	63	70	55	59	66				
Q2.	Group A:	4.5	2.3	7.9	3.4	4.8	2.7	5.6	6.1	3.5	
	Group B:	3.5	4.9	1.1	2.5	2.3	4.1	0.7			
Q3.	Group A:	650	710	437	520	583	492	555			
	Group B:	573	617	648	861	732	689	741			
Q4.	Group A:	43	70	51	35	60	77	48	62	57	75
	Group B:	90	45	73	64	86	59	88	72	89	
Q5.	Group A:	48	60	75	86	79	39	52	75	93	57
		62	71	69	80	69	62	70			
	Group B:	54	93	82	67	81	77	91	79	63	74
		99	84	76	68	71	90				

Wilcoxon matched-pairs signed-rank test

In the following examples, find the significance level of the differences between the groups in (*a*) a one-tailed and (*b*) a two-tailed test. The groups are arranged in matched pairs, the members of each pair being shown one above the other.

Q6.	Group A:	4.5	2.3	7.9	6.8	5.3	6.2	5.7		
	Group B:	4.3	2.7	9.0	6.7	5.6	10.1	6.9		
Q7.	Group A:	127	163	149	101	137	125	141	142	133
	Group B:	135	170	181	111	151	120	138	153	140
Q8.	Group A:	5	3	7	11	9	4	3	2	
	Group B:	7	4	6	12	6	10	9	3	
Q9.	Group A:	14	17	19	25	33	15	17	19	23
	Group B:	11	17	15	26	19	14	13	20	18

Mixed examples

(These are all fictitious experiments!)

Q10. A traffic survey measures the speed of 15 cars chosen randomly each morning over a quarter-mile stretch of road. One ordinary Monday these were (in m.p.h.):

32	45	37	41	28	36	40	49	34	36	33	30	40	38	39

On the next Monday in another ordinary working week on which there were similar weather conditions, a ‘simulated accident scene’ was placed 50 yards before the start of the measurement area, and the speeds of fifteen cars measured were:

33	27	38	35	30	32	29	20	37	44	31	36	30	34	32

Did the simulated accident significantly reduce drivers' speeds?

Q11. In a reaction-time experiment, the stimulus to react to was a recorded voice, sometimes the same voice that had just given a 'ready' signal, and sometimes a different one. Twelve subjects' results were as follows (RTs in ms):

Subject	1	2	3	4	5	6	7	8	9	10	11	12
RT to same voice	302	287	350	296	411	337	326	343	315	371	299	316
RT to different voice	340	302	359	352	408	361	328	340	347	392	326	333

Is there a significant difference in RT between the two conditions?

Q12. Twelve student volunteers performed a card-sorting task: they sorted 250 cards on one day, 500 on the next day starting 20 min after having ingested a pharmacologically-active substance, and 250 on a third day. The table gives the number of errors in sorting they made on the second day, and the total errors on the first and third days. Does the substance have any effect on card-sorting accuracy?

Subject	1	2	3	4	5	6	7	8	9	10	11	12
Day 2	12	17	9	3	16	10	28	14	5	19	20	8
Days 1 & 3	16	16	11	5	10	13	36	11	8	11	20	14

Q13. A survey was conducted to determine people's opinions of selected foreign countries. The overall order of preference among the sample interviewed (starting with the most preferred) was Australia, Canada, Denmark, New Zealand, Holland, Germany, France, Zimbabwe, Spain, South Africa, Italy.

Was a significant preference shown for Commonwealth and ex-Commonwealth countries on the one hand over European countries on the other?

Q14. Twelve cod-graders grade the following numbers of cod per hour:

1382	1545	1106	1761	1560	1669	1292	1418	1477	1351	1523	1618

After a number of sessions working through the teaching programme *Defect Detection in White Fish Processing: Intermediate Level*, their cod grading rates were (taking the graders in the same order as above):

1390	1422	1119	1578	1553	1682	1101	1376	1468	1099	1478	1564

Has the teaching programme had any effect on their grading rates?

Q15. The short-term memory span for digits was measured for a number of students specializing in arts (A) and science (S) subjects. The table gives each student's mean span with his subject group:

A	A	S	A	S	S	S	S	A	S	A	A	S	S	A
5.8	7.3	7.1	6.9	8.2	5.9	6.4	6.8	7.7	6.0	6.3	5.2	6.2	6.6	7.4

A	S	S	S	A	S	A	A	S	S
6.5	7.0	7.2	6.1	7.9	7.4	7.0	6.2	6.4	8.0

Is there a significant difference between the digit spans of arts and science students?

Q16. Two groups of subjects are shown an ambiguous figure, and the time taken until the first reversal of its appearance is measured for each subject. One group had previously seen the figure in a form strongly biased to show one of its alternative appearances; the other had no such pre-exposure (control group). The times to first reversal (in s) were:

Pre-exposure group	7.4	7.0	6.8	8.2	6.5	7.5	5.8	6.3	7.1	6.6
Control group	6.2	7.3	5.6	5.9	6.0	6.9	6.1	5.4		

Does pre-exposure to the biased figure lengthen the time to first reversal?

Q17. Twelve people are engaged in 'experimental conversation'. In the 'positive' condition they are 'reinforced' by an approving 'uh-huh' from the experimenter whenever they use the personal pronoun 'I'. In the 'negative' condition they are 'punished' by a disapproving 'huh' when they say 'I'. The rates of 'I' emission in the experiment are as follows (responses in a 10-min interval):

Subject	A	B	C	D	E	F	G	H	I	J	K	L
Pre-exposure group	17	62	20	11	31	25	15	38	47	22	26	8
Control group	14	68	19	3	27	26	9	22	40	19	20	11

Does reinforcement have the effect you would expect?

Q18. Two new-born bats are taken from each of a number of litters. One of each pair is kept in a cage, the other being allowed to live freely in the experimenter's office (despite protests from the occupants of nearby offices). After one month, their moth-catching abilities are tested in a standard Batman™ experimental chamber. The number of moths caught (out of a possible total of 25) are given below. Does experience in the first month of life have any effect on moth-catching ability in bats?

Litter number	1	2	3	4	5	6	7	8	9	10
Caged bat	8	16	0	10	6	12	8	2	15	9
Free-living bat	18	25	17	6	11	11	12	10	15	14

Q19. The following are the scores on the Seashore Test of Musical Aptitude of a number of 10-year-olds:

Right-handed children	28	54	37	102	66	30	41	56	34	72
Left-handed children	46	50	83	27	40	39	61	33	59	87

Do these data reveal a relationship between handedness and musical aptitude as measured by the Seashore Test?

Q20. In an experiment in which briefly-flashed letters were superimposed on either a random or a checkerboard black-and-white pattern, one subject gave the following results:

Letter	a	c	e	n	o	s	u	v	x	z
% correct recognitions:										
On random field	67	43	49	31	40	52	35	74	83	77
On checkerboard	79	51	58	28	44	52	28	87	90	81

Do the checkerboard and random fields have significantly different effects on the visibility of the letter?

Q21. The following were all Republican candidates in the electoral contests for various local offices in the city of Meltingpot, Ohio. **Elected:** Aaronson, Blomberg, Evans, Horsley, Jaspers, McTavish, O'Shaughnessy, Scorbini. **Defeated:** Neuhaus, Pickford, Rodsky, Toft, Verploot, Wilhelm, Young, Zotterman.

The ballot papers were organized alphabetically. Do these results show a relation between position on the ballot paper and electoral success?

Q22. A number of rats were assessed on the Nebraska Rodent Personality Scale, and the ten most introverted and the ten most extroverted were selected. They were trained to criterion on a discrimination task, and the number of trials required for extinction was then counted for each rat:

Introverted rats	23	18	107	16	35	40	28	21	46	21
Extroverted rats	62	17	33	25	38	19	44	29	80	36

Is there any connection shown between rate of extinction and the extroversion scale of the NRPS?

Q23. The crew of a radar station work four-hour shifts. The following are the numbers of guided missiles falsely reported by each operator in the first and last half-hours of her shift:

Operator	A	B	C	D	E	F	G	H	I	J	K	L
First half hour	6	3	0	2	4	3	8	5	0	1	7	2
Last half hour	5	8	3	4	2	7	12	9	2	0	5	8

Are operators significantly more prone to make false reports at either end of their shifts?

Q24. Sixteen subjects made settings of the same colour discrimination threshold on two successive days. The differences between the two settings made were as follows (in nanometres). Is there evidence of improvement (improvement = positive difference score)?

0.3 –0.6 1.2 2.3 –1.0 3.5 –2.0 1.1 0.8 1.4 2.7 –1.5 –2.6 2.4 3.1 1.9

Mann–Whitney test using a normal approximation

Q25. In a Mann–Whitney U test with $n_1 = 20$ and $n_2 = 60$ we find $U = 400$. What is (*a*) the one-tailed probability, and (*b*) the two-tailed probability of getting a value of U as extreme as this? (See instructions on the page of the *Tables and Formulae* booklet giving critical values of U.)