Research Methods: Appendices

- **Choosing a test** — *Identifying type of data obtained*
Selecting an appropriate test
Looking up values in the statistical tables
- **Appendix 1** — *Mann-Whitney U test*
- **Appendix 2** — *Sign test*
- **Appendix 3** — *Wilcoxon test*
- **Appendix 4** — *Spearman's rho test*
- **Appendix 5** — *Chi-squared test*

In order to use the statistical tables on the following pages, you first need to decide whether:

1. Your data is in numerical form, in which case it is suitable for quantitative analysis; otherwise, use qualitative analysis.
2. You have obtained nominal, ordinal, interval or ratio data.
3. Your data shows a difference between the two conditions (the experimental hypothesis) or not (the null hypothesis).
4. You can use parametric tests (i.e. if data is interval or ratio, normally distributed, and the variances in the two conditions are similar); otherwise non-parametric tests can be used. Non-parametric tests can be used in nearly all cases, and it is the most useful of these that are described in this book.

Once you have obtained your results, you can construct a table of frequencies, and decide which type of chart or graph you wish to use in order to present your data graphically in the clearest way possible.

The next step to take is to analyse your data, as follows:

1. Calculate measures of central tendency: mean, median and mode.
2. Calculate measures of dispersion: range, interquartile range, variation ratio and standard deviation.

You will then need to apply further statistical analysis using the statistical tests described in the Research methods: Data analysis chapter. Please refer to the worked examples for each of these and follow the step- by-step instructions in the main text.

HOW TO DECIDE WHICH TEST TO USE

The main purpose of these tests is to decide the probability of the null hypothesis being correct, and to evaluate its significance. Each test involves calculating your observed value

from your results, and then looking up the critical value in a table of values, to see whether your value is greater than, equal to, or less than the critical value. Use the appropriate column or table, depending on (a) whether you used a one- or two-tailed test and (b) which level of significance, or probability (p) you wish to check. If p is less than or equal to 0.05 or 5%, which is the standard probability of significance used by psychologists, the null hypothesis is rejected in favour of the experimental hypothesis. To see whether the findings are highly significant, look at whether the null hypothesis still holds true at $p = 0.01$, or 1%, or even $p = 0.001$, or 0.1%.

If your experimental hypothesis is directional (i.e. you predicted the direction of any effects), you need to use a one-tailed test; otherwise you have a non-directional hypothesis, in which case you need to use a two-tailed test.

If the design of your test of difference is independent, as long as the data are ordinal or interval, the Mann-Whitney U test can be used. If you have used a repeated measures or matched participants design, the sign test can be used, as long as the data are ordinal; or if the data are interval or ratio, the Wilcoxon matched pairs signed ranks test can be used. The latter is more sensitive than the signed test. The sign test provides us with a crude analysis, which is sufficient when data are ordinal, but when actual values are obtained (interval or ratio data) the Wilcoxon test will provide a more sophisticated analysis. Therefore, although it is possible to use the sign test for interval or ratio data, it would be best to limit its use to analysis of ordinal data.

If you manipulated the independent variable (some aspect of the situation), you need to use a test of difference (such as the Mann-Whitney U test, the sign test or the Wilcoxon matched pairs signed ranks test); otherwise, you need to use a test of correlation (such as Spearman's rho test, as long as the data are ordinal, interval or ratio) or a test of association (such as the chi-squared test, as long as the data are nominal).

HOW TO USE THE TABLES

Mann-Whitney U test Appendix 1, pages 4–5

In the Mann-Whitney U test, use the smaller value of U and U′ to look up the critical value of U for a one- or two-tailed test, as appropriate, at 0.05, initially (bottom table, page 5). If the tabled value is equal to or less than your value at that level, the null hypothesis is retained; if it is greater than your value, it is rejected and your experimental hypothesis is proved.

Sign test Appendix 2, page 6

In the sign test, look up the critical value of S for a one- or two-tailed test, as appropriate, for N, the number of participants with differing scores, at 0.05, initially. If the tabled value is equal to or less than your value at that level, the null hypothesis is retained; if it is greater than your value, it is rejected and your experimental hypothesis is proved.

Wilcoxon test Appendix 3, page 7

In the Wilcoxon test, look up the critical value of T for a one- or two-tailed test, as appropriate, for N, the number of participants with differing scores, at 0.05, initially. If the tabled value is equal to or less than your value at that level, the null hypothesis is retained; if it is greater than your value, it is rejected and your experimental hypothesis is proved.

Spearman's rho test Appendix 4, page 8

In the Spearman's rho test, look up the critical value of r_S for a one- or two-tailed test, as appropriate, for N, the number of participants, at 0.05, initially. If the tabled value is greater than or equal to your value at that level, the null hypothesis is retained; if it is less than your value, it is rejected and your experimental hypothesis is proved.

Chi-squared test Appendix 5, page 9

In the chi-squared test, look up the critical value of chi-squared (also shown as χ^2) for a one- or two-tailed test, as appropriate, for df, the degrees of freedom, at 0.05, initially. If the tabled value is greater than or equal to your value at that level, the null hypothesis is retained; if it is less than your value, it is rejected and your experimental hypothesis is proved.

TIPS

Remember that decisions based on statistical tests are open to error, but if you follow the standard procedures outlined in the Research methods: Data analysis chapter the potential

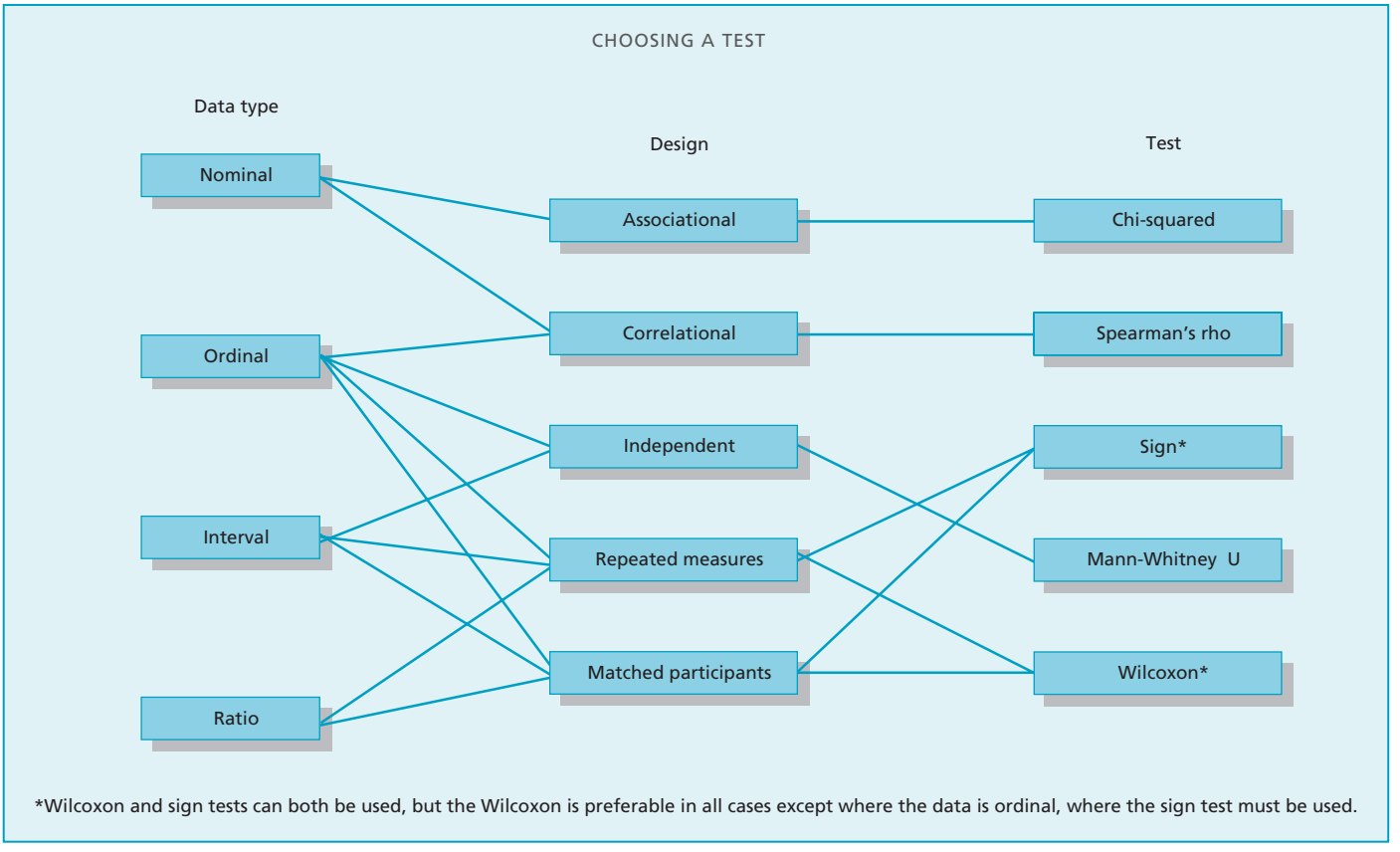
for errors can be minimised. Try to be as unbiased as possible, and try not to assume too much about the results in advance.

Ensure that you have not made errors of either Type I, which can be reduced by using a greater level of significance (e.g. $p = 0.01$, or 1%, or even $p = 0.001$, or 0.1%), or Type II, which can be reduced by using a lesser level of significance (e.g. $p = 0.10$, or 10%).

In the Mann-Whitney U test, remember that ties are possible—this reduces the accuracy, but has only a small effect unless there are several ties.

In the chi-squared test, do follow the rules on page 25 of the Research methods: Data analysis chapter to avoid incorrect use of this test.

The tests described in the Research methods: Data analysis chapter provide different levels of analysis, and they require a particular type of data. The following chart outlines the tests that can be used for different data types and experimental designs. Please note that this chart deals only with the statistical tests described in the Research methods: Data analysis chapter, even though other tests do exist.



*Wilcoxon and sign tests can both be used, but the Wilcoxon is preferable in all cases except where the data is ordinal, where the sign test must be used.

APPENDIX 1: MANN-WHITNEY U TEST

Critical values of U for a one-tailed test at 0.005; two-tailed test at 0.01*

	N_A																			
N_B	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	0
3	—	—	—	—	—	—	—	—	0	0	0	1	1	1	2	2	2	2	3	3
4	—	—	—	—	—	0	0	1	1	2	2	3	3	4	5	5	6	6	7	8
5	—	—	—	—	0	1	1	2	3	4	5	6	7	7	8	9	10	11	12	13
6	—	—	—	0	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17	18
7	—	—	—	0	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
8	—	—	—	1	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
9	—	—	0	1	3	5	7	9	11	13	16	18	20	22	24	27	29	31	33	36
10	—	—	0	2	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	42
11	—	—	0	2	5	7	10	13	16	18	21	24	27	30	33	36	39	42	45	48
12	—	—	1	3	6	9	12	15	18	21	24	27	31	34	37	41	44	47	51	54
13	—	—	1	3	7	10	13	17	20	24	27	31	34	38	42	45	49	53	56	60
14	—	—	1	4	7	11	15	18	22	26	30	34	38	42	46	50	54	58	63	67
15	—	—	2	5	8	12	16	20	24	29	33	37	42	46	51	55	60	64	69	73
16	—	—	2	5	9	13	18	22	27	31	36	41	45	50	55	60	65	70	74	79
17	—	—	2	6	10	15	19	24	29	34	39	44	49	54	60	65	70	75	81	86
18	—	—	2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81	87	92
19	—	0	3	7	12	17	22	28	33	39	45	51	56	63	69	74	81	87	93	99
20	—	0	3	8	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	105

*Dashes in the body of the table indicate that no decision is possible at the stated level of significance.
For any N_A and N_B the observed value of U is significant at a given level of significance if it is *equal* to or *less* than the critical values shown.

Critical values of U for a one-tailed test at 0.01; two-tailed test at 0.02*

	N_A																			
N_B	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2	—	—	—	—	—	—	—	—	—	—	—	—	0	0	0	0	0	0	1	1
3	—	—	—	—	—	—	0	0	1	1	1	2	2	2	3	3	4	4	4	5
4	—	—	—	—	0	1	1	2	3	3	4	5	5	6	7	7	8	9	9	10
5	—	—	—	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
6	—	—	—	1	2	3	4	6	7	8	9	11	12	13	15	16	18	19	20	22
7	—	—	0	1	3	4	6	7	9	11	12	14	16	17	19	21	23	24	26	28
8	—	—	0	2	4	6	7	9	11	13	15	17	20	22	24	26	28	30	32	34
9	—	—	1	3	5	7	9	11	14	16	18	21	23	26	28	31	33	36	38	40
10	—	—	1	3	6	8	11	13	16	19	22	24	27	30	33	36	38	41	44	47
11	—	—	1	4	7	9	12	15	18	22	25	28	31	34	37	41	44	47	50	53
12	—	—	2	5	8	11	14	17	21	24	28	31	35	38	42	46	49	53	56	60
13	—	0	2	5	9	12	16	20	23	27	31	35	39	43	47	51	55	59	63	67
14	—	0	2	6	10	13	17	22	26	30	34	38	43	47	51	56	60	65	69	73
15	—	0	3	7	11	15	19	24	28	33	37	42	47	51	56	61	66	70	75	80
16	—	0	3	7	12	16	21	26	31	36	41	46	51	56	61	66	71	76	82	87
17	—	0	4	8	13	18	23	28	33	38	44	49	55	60	66	71	77	82	88	93
18	—	0	4	9	14	19	24	30	36	41	47	53	59	65	70	76	82	88	94	100
19	—	1	4	9	15	20	26	32	38	44	50	56	63	69	75	82	88	94	101	107
20	—	1	5	10	16	22	28	34	40	47	53	60	67	73	80	87	93	100	107	114

*Dashes in the body of the table indicate that no decision is possible at the stated level of significance.
For any N_A and N_B the observed value of U is significant at a given level of significance if it is *equal* to or *less* than the critical values shown.

Critical values of U for a one-tailed test at 0.025; two-tailed test at 0.05*

	N_A																			
N_B	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2	—	—	—	—	—	—	—	0	0	0	0	1	1	1	1	1	2	2	2	2
3	—	—	—	—	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4	—	—	—	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5	—	—	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6	—	—	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7	—	—	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	—	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	—	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	—	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	—	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	—	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	—	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	—	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15	—	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	—	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	—	2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18	—	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	—	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	—	2	8	13	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

*Dashes in the body of the table indicate that no decision is possible at the stated level of significance.
For any N_A and N_B the observed value of U is significant at a given level of significance if it is *equal* to or *less* than the critical values shown.

Critical values of U for a one-tailed test at 0.05; two-tailed test at 0.10*

	N_A																			
N_B	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	0
2	—	—	—	—	0	0	0	1	1	1	1	2	2	2	3	3	3	4	4	4
3	—	—	0	0	1	2	2	3	3	4	5	5	6	7	7	8	9	9	10	11
4	—	—	0	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18
5	—	0	1	2	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25
6	—	0	2	3	5	7	8	10	12	14	16	17	19	21	23	25	26	28	30	32
7	—	0	2	4	6	8	11	13	15	17	19	21	24	26	28	30	33	35	37	39
8	—	1	3	5	8	10	13	15	18	20	23	26	28	31	33	36	39	41	44	47
9	—	1	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54
10	—	1	4	7	11	14	17	20	24	27	31	34	37	41	44	48	51	55	58	62
11	—	1	5	8	12	16	19	23	27	31	34	38	42	46	50	54	57	61	65	69
12	—	2	5	9	13	17	21	26	30	34	38	42	47	51	55	60	64	68	72	77
13	—	2	6	10	15	19	24	28	33	37	42	47	51	56	61	65	70	75	80	84
14	—	2	7	11	16	21	26	31	36	41	46	51	56	61	66	71	77	82	87	92
15	—	3	7	12	18	23	28	33	39	44	50	55	61	66	72	77	83	88	94	100
16	—	3	8	14	19	25	30	36	42	48	54	60	65	71	77	83	89	95	101	107
17	—	3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96	102	109	115
18	—	4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109	116	123
19	0	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	130
20	0	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138

*Dashes in the body of the table indicate that no decision is possible at the stated level of significance.
For any N_A and N_B the observed value of U is significant at a given level of significance if it is *equal* to or *less* than the critical values shown.

APPENDIX 2: SIGN TEST

	Level of significance for one-tailed test				
	0.05	0.025	0.01	0.005	0.0005
	Level of significance for two-tailed test				
N	0.10	0.05	0.02	0.01	0.001
5	0	—	—	—	—
6	0	0	—	—	—
7	0	0	0	—	—
8	1	0	0	0	—
9	1	1	0	0	—
10	1	1	0	0	—
11	2	1	1	0	0
12	2	2	1	1	0
13	3	2	1	1	0
14	3	2	2	1	0
15	3	3	2	2	1
16	4	3	2	2	1
17	4	4	3	2	1
18	5	4	3	3	1
19	5	4	4	3	2
20	5	5	4	3	2
25	7	7	6	5	4
30	10	9	8	7	5
35	12	11	10	9	7

Calculated S must be *equal* to or *less* than the table (critical) value for significance at the level shown.

Source: F. Clegg (1982), *Simple statistics*, Cambridge University Press.

APPENDIX 3: WILCOXON SIGNED RANKS TEST

	Levels of significance			
	One-tailed test			
	0.05	0.025	0.01	0.001
	Two-tailed test			
Sample size	0.1	0.05	0.02	0.002
$N = 5$	$T \leq 0$			
6	2	0		
7	3	2	0	
8	5	3	1	
9	8	5	3	
10	11	8	5	0
11	13	10	7	1
12	17	13	9	2
13	21	17	12	4
14	25	21	15	6
15	30	25	19	8
16	35	29	23	11
17	41	34	27	14
18	47	40	32	18
19	53	46	37	21
20	60	52	43	26
21	67	58	49	30
22	75	65	55	35
23	83	73	62	40
24	91	81	69	45
25	100	89	76	51
26	110	98	84	58
27	119	107	92	64
28	130	116	101	71
29	141	125	111	78
30	151	137	120	86
31	163	147	130	94
32	175	159	140	103
33	187	170	151	112

Calculated T must be *equal* to or *less* than the table (critical) value for significance at the level shown.

Source: From R. Meddis (1975), *Statistical handbook for non-statisticians*, London: McGraw-Hill.

APPENDIX 4: SPEARMAN'S RHO TEST

	Level of significance for two-tailed test			
	0.10	0.05	0.02	0.01
	Level of significance for one-tailed test			
	0.05	0.025	0.01	0.005
$N = 4$	1.000			
5	0.900	1.000	1.000	
6	0.829	0.886	0.943	1.000
7	0.714	0.786	0.893	0.929
8	0.643	0.738	0.833	0.881
9	0.600	0.700	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.536	0.618	0.709	0.755
12	0.503	0.587	0.671	0.727
13	0.484	0.560	0.648	0.703
14	0.464	0.538	0.566	0.675
15	0.443	0.521	0.604	0.654
16	0.429	0.503	0.582	0.635
17	0.414	0.485	0.566	0.615
18	0.401	0.472	0.550	0.600
19	0.391	0.460	0.535	0.584
20	0.380	0.447	0.520	0.570
21	0.370	0.435	0.508	0.556
22	0.361	0.425	0.496	0.544
23	0.353	0.415	0.486	0.532
24	0.344	0.406	0.476	0.521
25	0.337	0.398	0.466	0.511
26	0.331	0.390	0.457	0.501
27	0.324	0.382	0.448	0.491
28	0.317	0.375	0.440	0.483
29	0.312	0.368	0.433	0.475
30	0.306	0.362	0.425	0.467

For $n > 30$, the significance of r_s can be tested by using the formula:

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

and checking the value of *t*.

Calculated r_s must *equal* or *exceed* the table (critical) value for significance at the level shown.

APPENDIX 5: CHI-SQUARED TEST

	Level of significance for one-tailed test					
	0.10	0.05	0.025	0.01	0.005	0.0005
	Level of significance for two-tailed test					
df	0.20	0.10	0.05	0.02	0.01	0.001
1	1.64	2.71	3.84	5.41	6.64	10.83
2	3.22	4.60	5.99	7.82	9.21	13.82
3	4.64	6.25	7.82	9.84	11.34	16.27
4	5.99	7.78	9.49	11.67	13.28	18.46
5	7.29	9.24	11.07	13.39	15.09	20.52
6	8.56	10.64	12.59	15.03	16.81	22.46
7	9.80	12.02	14.07	16.62	18.48	24.32
8	11.03	13.36	15.51	18.17	20.09	26.12
9	12.24	14.68	16.92	19.68	21.67	27.88
10	13.44	15.99	18.31	21.16	23.21	29.59
11	14.63	17.28	19.68	22.62	24.72	31.26
12	15.81	18.55	21.03	24.05	26.22	32.91
13	16.98	19.81	22.36	25.47	27.69	34.53
14	18.15	21.06	23.68	26.87	29.14	36.12
15	19.31	22.31	25.00	28.26	30.58	37.70
16	20.46	23.54	26.30	29.63	32.00	39.29
17	21.62	24.77	27.59	31.00	33.41	40.75
18	22.76	25.99	28.87	32.35	34.80	42.31
19	23.90	27.20	30.14	33.69	36.19	43.82
20	25.04	28.41	31.41	35.02	37.57	45.32
21	26.17	29.62	32.67	36.34	38.93	46.80
22	27.30	30.81	33.92	37.66	40.29	48.27
23	28.43	32.01	35.17	38.97	41.64	49.73
24	29.55	33.20	36.42	40.27	42.98	51.18
25	30.68	34.38	37.65	41.57	44.31	52.62
26	31.80	35.56	38.88	42.86	45.64	54.05
27	32.91	36.74	40.11	44.14	46.96	55.48
28	34.03	37.92	41.34	45.42	48.28	56.89
29	35.14	39.09	42.69	46.69	49.59	58.30
30	36.25	40.26	43.77	43.49	50.89	59.70
32	38.47	42.59	46.19	50.49	53.49	62.49
34	40.68	44.90	48.60	53.00	56.06	65.25
36	42.88	47.21	51.00	55.49	58.62	67.99
38	45.08	49.51	53.38	57.97	61.16	70.70
40	47.27	51.81	55.76	60.44	63.69	73.40
44	51.64	56.37	60.48	65.34	68.71	78.75
48	55.99	60.91	65.17	70.20	73.68	84.04
52	60.33	65.42	69.83	75.02	78.62	89.27
56	64.66	69.92	74.47	79.82	83.51	94.46
60	68.97	74.40	79.08	84.58	88.38	99.61

Calculated value of χ^2 must *equal* or *exceed* the table (critical) value for significance at the level shown.

Abridged from R.A. Fisher and F. Yates (1974), *Statistical tables for biological, agricultural and medical research (6th Edn.)*, Harlow, UK: Addison Wesley Longman.