Formulas and Steps for Statistical Calculations

Coefficient of Reliability

Scott's *pi*

Standard Deviation

One-Way Chi-Square

Contingency Analysis

Contingency Table

t-Test for Independent Means

Paired Sample *t*-Test

Correlation

COEFFICIENT OF RELIABILITY

The coefficient of reliability (*C.R.*) is a simple way to calculate coding reliability for unitizing or categorizing decisions. This formula bases reliability on the ratio of decisions coders agreed upon to the total number of coding decisions made by each coder.

$$C.R. = \frac{2M}{N_1 + N_2}$$

M = Number of coding decisions agreed upon

N = Total number of coding decisions made by each coder

Step 1: Identify the number of coding decisions made by coder number (N_1).

Step 2: Identify the number of coding decisions made by coder number (N_2).

Step 3: Identify the number of coding decisions agreed upon by the coders M.

Step 4: Compute *C.R.*

$$C.R. = \frac{2M}{N_1 + N_2}$$

$$C.R. = \frac{2(76)}{79 + 81} \quad C.R. = \frac{152}{160}$$

$$C.R. = .95 \text{ or } 95\%$$

SCOTT'S *pi*

Scott's *pi* is used as an index of categorizing reliability for content coding, accounting for the number of categories in the coding scheme and the frequency with which each category is used. This reliability formula can accommodate any number of coders and any number of categories. Moreover, it accounts for the rate of agreement that would occur by chance alone.

$$pi = \frac{\text{\% observed agreement} - \text{\% expected agreement}}{1 - \text{\% expected agreement}}$$

Step 1: Determine the coefficient of reliability (*C.R.*), which is the ratio of coding agreements to the total number of coding decisions, as demonstrated on page 354; this value is the percentage of observed agreement to be used in Scott's *pi*.

Step 2: Set up three columns as shown, and list the frequency with which each category was used in column 1.

Step 3: Determine the percentage of expected agreement by dividing the frequency for each category by the total number of codings. Enter in column 2.

Step 4: Compute the squared percentages of expected agreement for column 3 by squaring the values in column 2.

Step 5: Sum the values in column 3; this number is the percentage of expected agreement.

Step 6: Compute Scott's *pi*.

Category	*Column 1 Frequency n = 160*	*Column 2 Percentage of Expected Agreement*	*Column 3 Squared Percentage of Expected Agreement*
A	60	.375	.141
B	48	.300	.09
C	36	.225	.051
D	16	.100	.01
			.292

$$pi = \frac{\text{\% observed agreement} - \text{\% expected agreement}}{1 - \text{\% expected agreement}}$$

$$pi = \frac{.95 - .292}{1 - .292} \qquad pi = \frac{.658}{.708} \qquad pi = .93$$

STANDARD DEVIATION

Deviation score method:

$$SD = \sqrt{\frac{\Sigma x^2}{n}}$$

Step 1: Set up columns as shown, and list the scores in column 1 (can be in any order).

Step 2: Obtain n by counting the cases in the distribution.

Step 3: Find $\overline{X}$ for the distribution of scores; put this value in column 2.

Step 4: Obtain x (the deviation score) in column 3 by subtracting column 2 from column 1.

Step 5: Obtain x^2 (the squared deviation score) in column 4 by squaring the values in column 3.

Step 6: Obtain Σx^2 by adding the values in column 4.

Step 7: Enter Σx^2 and n in the formula and solve.

Column 1 *Scores* (X)	*Minus*	*Column 2* *Mean*	*Equals*	*Column 3* *Deviation* *Scores* (x)	*Column 4* *Squared* *Deviation* *Scores* (x^2)
15	–	9	=	6	36
12	–	9	=	3	9
10	–	9	=	1	1
9	–	9	=	0	0
7	–	9	=	−2	4
7	–	9	=	−2	4
3	–	9	=	−6	36
					90

$$SD = \sqrt{\frac{\Sigma x^2}{n}} \qquad SD = \sqrt{\frac{90}{7}} \qquad SD = \sqrt{12.86} \qquad SD = 3.59$$

Raw score method:

$$SD = \sqrt{\frac{\Sigma x^2}{n}}$$

Step 1: Set up two columns and list the scores in column 1 (can be in any order).

Step 2: Obtain n by counting the number of cases in the distribution.

Step 3: Obtain X^2 (the squared score) for column 2 for each score (X) by squaring the score in column 1.

Step 4: Obtain ΣX by adding the scores in column 1.

Step 5: Obtain ΣX^2 by adding the values in column 2.

Step 6: Obtain Σx^2 using the formula $\Sigma x^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$

Step 7: Enter Σx^2 and n in the formula for SD and solve.

Column 1 *Scores* (X)	*Column 2* *Squared Score* (X^2)
15	225
12	144
10	100
9	81
7	49
7	49
3	9
63	657

$$\Sigma x^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n} \qquad \Sigma x^2 = 657 - \frac{(63)^2}{7} \qquad \Sigma x^2 = 657 - \frac{3{,}969}{7}$$

$$\Sigma x^2 = 657 - 567 \qquad \Sigma x^2 = 90$$

$$SD = \sqrt{\frac{\Sigma x^2}{n}} \qquad SD = \sqrt{\frac{90}{7}} \qquad SD\sqrt{12.86} \qquad SD = 3.59$$

ONE-WAY CHI-SQUARE

$$\chi^2 = \Sigma \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}} \quad \text{or} \quad \chi^2 = \Sigma \frac{(Q - E)^2}{E}$$

Step 1: Graph the categories for the variable; set up five columns and list the observed frequency for each category or cell in column 1.

Step 2: Calculate the expected frequency for each category (unless previous research suggests otherwise, simply divide the number of cases by the number of categories); use this value as the expected frequency for each category; list in column 2.

Step 3: Obtain the value for column 3 for each cell by subtracting the expected frequency from the observed frequency.

Step 4: Obtain the value for column 4 for each cell by squaring the value found in step 3.

Step 5: Obtain the value for column 5 by dividing the value in column 4 by the expected frequency.

Step 6: Obtain χ^2 by summing the values in column 5.

Step 7: Calculate the degrees of freedom (number of categories minus 1).

Step 8: Identify critical value of χ^2 at .05 significance level corresponding to the degrees of freedom in Table B.1.

Step 9: Interpret χ^2: If calculated χ^2 meets or exceeds the critical value, results are statistically significant; accept the research hypothesis. If calculated χ^2 is less than the critical value, retain the null hypothesis.

	Column 1 *Observed* *Frequency*	*Minus*	*Column 2* *Expected* *Frequency*	*Equals*	*Column 3* (O − E)	*Column 4* $(O - E)^2$	*Column 5* $\frac{(O - E)^2}{E}$
Single	8	–	10	=	−2	4	.4
Dating	8	–	10	=	−2	4	.4
Married	14	–	10	=	4	16	1.6
							2.4

$\chi^2 = 2.4$

$df = (3 - 1) \qquad df = 2$

Critical $\chi^2 = 5.991$

Interpretation: Results are not statistically significant; retain the null hypothesis.

TABLE B.1 Critical Values of Chi-Square

	Significance Level		
Degrees of Freedom	*.10*	*.05*	*.01*
1	2.706	3.841	6.635
2	4.605	5.991	9.210
3	6.251	7.815	11.345
4	7.779	9.488	13.227
5	9.236	11.070	15.086
6	10.645	12.592	16.812
7	12.017	14.067	18.475
8	13.362	15.507	20.090
9	14.684	16.919	21.666
10	15.987	18.307	23.209
11	17.275	19.675	24.725
12	18.549	21.026	26.217
13	19.812	22.362	27.688
14	21.064	23.685	29.141
15	22.307	24.996	30.578
16	23.542	26.296	32.000
17	24.769	27.587	33.409
18	25.989	28.869	34.805
19	27.204	30.144	36.191
20	28.412	31.410	37.566

SOURCE: Adapted from *Statistical Tables for Biological, Agricultural, and Medical Research*, 6th edition, by R. A. Fisher and F. Yates, 1963. London: Hafner.

$$\chi^2 = \Sigma\frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}} \quad \text{or} \quad \chi^2 = \Sigma\frac{(O - E)^2}{E}$$

Step 1: Graph the categories for the variables in a contingency table (see next page).

Step 2: Set up five columns as shown, and list the observed frequency for each cell in column 1.

Step 3: Compute the row marginals for the contingency table by adding the observed frequencies across each row.

Step 4: Compute the column marginals for the contingency table by adding the observed frequencies down each column.

Step 5: Confirm that the row marginals and column marginals in the contingency table equal the grand sum, or n.

Step 6: Calculate the expected frequency for each cell by multiplying the row marginal by the column marginal for each respective cell; divide that value by the grand sum.

Step 7: List the expected frequency for each cell in column 2 and in the contingency table.

Step 8: Obtain the value for column 3 for each cell by subtracting the expected frequency from the observed frequency.

Step 9: Obtain the value for column 4 for each cell by squaring the value found in step 3.

Step 10: Obtain the value for column 5 by dividing the value in column 4 by the expected frequency.

Step 11: Obtain χ^2 by summing the values in column 5.

Step 12: Using the contingency table, calculate the degrees of freedom; multiply (number of columns minus 1) by (number of rows minus 1).

Step 13: Identify critical value of χ^2 at .05 significance level corresponding to the degrees of freedom (calculated in step 12) in Table B.1.

Step 14: Interpret χ^2: If calculated χ^2 meets or exceeds the critical value, the results are statistically significant; accept the research hypothesis. If calculated χ^2 is less than the critical value, retain the null hypothesis.

Contingency Table

	Females	*Males*	*Row Marginals*
Target of Sexual Harassment	$E = 25.15$ 44	$E = 30.85$ 12	56
Not a Target of Sexual Harassment	$E = 49.85$ 31	$E = 61.15$ 80	111
Column Marginals	75	92	**Grand sum = 167**

	Column 1 Observed Frequency	*Minus*	*Column 2 Expected Frequency*	*Equals*	*Column 3* (O − E)	*Column 4* $(O - E)^2$	*Column 5* $\frac{(O - E)^2}{E}$
Female Target	44	−	25.15	=	18.85	355.32	14.13
Male Target	12	−	30.85	=	−18.85	355.32	11.52
Female Not a Target	31	−	49.85	=	−18.85	355.32	7.13
Male Not a Target	80	−	61.15	=	18.85	355.32	5.81
							38.59

$\chi^2 = 38.59$

$df = (2 - 1) \times (2 - 1) \qquad df = 1 \times 1 \qquad df = 1$

Critical $\chi^2 = 3.841$

Interpretation: Results are statistically significant; accept the research hypothesis.

t-TEST FOR INDEPENDENT MEANS

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - 0}{s_{\overline{X}_1 - \overline{X}_2}}$$

Step 1: Set up four columns and list the dependent variable scores for one category of the independent variable in column 1; list dependent variable scores for the other category of the independent variable in column 3.

Step 2: Square each score by multiplying the score by itself to create values for column 2 (X_1^2) and column 4 (X_2^2).

Step 3: Sum each column to find values for ΣX_1, ΣX_1^2, ΣX_2, and ΣX_2^2.

Step 4: Calculate the mean scores for both column 1 ($\overline{X}_1$) and column 3 ($\overline{X}_2$).

Step 5: Calculate Σx_1^2 and Σx_2^2.

Step 6: Calculate s_p^2.

Step 7: Calculate $s_{\overline{X}_1 - \overline{X}_2}$.

Step 8: Calculate t.

Step 9: Calculate the degrees of freedom ($n_1 + n_2 - 2$).

Step 10: Identify critical value of t at .05 significance level corresponding to the degrees of freedom (calculated in step 9) in Table B.2.

Step 11: Interpret t: If calculated t meets or exceeds the critical value, the results are statistically significant; accept the research hypothesis. If calculated t is less than the critical value, retain the null hypothesis.

Column 1 *Married Men* (X_1)	*Column 2* X_1^2	*Column 3* *Married Women* (X_2)	*Column 4* X_2^2
19	361	16	256
24	576	22	484
21	441	20	400
23	529	15	225
14	196	13	169
15	225	16	256
17	289	14	196
16	256	18	324
19	361	15	225
21	441	20	400
24	576	17	289
23	529	18	324
$\Sigma X_1 = 236$	$\Sigma X_1^2 = 4{,}780$	$\Sigma X_2 = 204$	$\Sigma X_2^2 = 3{,}548$

(*Step 4*) $\overline{X}_1 = \dfrac{236}{12}$ $\quad \overline{X}_1 = 19.67$ $\quad \overline{X}_2 = \dfrac{204}{12}$ $\quad \overline{X}_2 = 17$

(*Step 5*) $\Sigma x_1^2 = \Sigma X_1^2 - \dfrac{(\Sigma X_1)^2}{n_1}$ $\quad \Sigma x_1^2 = 4{,}780 - \dfrac{(236)^2}{12}$

$\Sigma x_1^2 = 4{,}780 - \dfrac{55{,}696}{12}$ $\quad \Sigma x_1^2 = 4{,}780 - 4{,}641.33$ $\quad \Sigma x_1^2 = 138.67$

$\Sigma x_2^2 = \Sigma X_2^2 - \dfrac{(\Sigma X_2)^2}{n_2}$ $\quad \Sigma x_2^2 = 3{,}548 - \dfrac{(204)^2}{12}$ $\quad \Sigma x_2^2 = 3{,}548 - \dfrac{41{,}616}{12}$

$\Sigma x_2^2 = 3{,}548 - 3{,}468$ $\quad \Sigma x_2^2 = 80$

(*Step 6*) $s_p^2 = \dfrac{\Sigma x_1^2 + \Sigma x_2^2}{n_1 + n_2 - 2}$ $\quad s_p^2 = \dfrac{138.67 + 80}{12 + 12 - 2}$ $\quad s_p^2 = \dfrac{218.67}{22}$ $\quad s_p^2 = 9.94$

(*Step 7*) $s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}$ $\quad s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\dfrac{9.94}{12} + \dfrac{9.94}{12}}$

$s_{\overline{X}_1 - \overline{X}_2} = \sqrt{.83 + .83}$ $\quad s_{\overline{X}_1 - \overline{X}_2} = \sqrt{1.66}$ $\quad s_{\overline{X}_1 - \overline{X}_2} = 1.29$

(*Step 8*) $t = \dfrac{(\overline{X}_1 - \overline{X}_2) - 0}{s_{\overline{X}_1 - \overline{X}_2}}$ $\quad t = \dfrac{(19.67 - 17) - 0}{1.29}$ $\quad t = \dfrac{2.67}{1.29}$ $\quad t = 2.07$

(*Step 9*) $df = n_1 + n_2 - 2$ $\quad df = 12 + 12 - 2$ $\quad df = 22$

(*Step 10*) Critical $t = 1.717$

(*Step 11*) Interpretation: Results are statistically significant (for a one-tailed test); accept the research hypothesis.

PAIRED SAMPLE *t*-TEST

$$t = \frac{\overline{D} - \mu_{D(hyp)}}{s_{\overline{D}}}$$

Step 1: Set up four columns, and list the paired or matched scores for the dependent variable in column 1 and column 2.

Step 2: Compute the difference score for column 3 by subtracting X_2 from X_1.

Step 3: Compute the squared difference score for column 4 by squaring the value in column 3.

Step 4: Find the sum of the difference scores by adding the values in column 3.

Step 5: Find the sum of the squared difference scores by adding the values in column 4.

Step 6: Compute the average difference score ($\overline{D}$).

Step 7: Compute Σd^2.

Step 8: Compute s_D.

Step 9: Compute $s_{\overline{D}}$.

Step 10: Compute t.

Step 11: Calculate degrees of freedom (number of pairs minus 1).

Step 12: Identify critical value of t at .05 significance level corresponding to the degrees of freedom (calculated in step 11) in Table B.2.

Step 13: Interpret t: If calculated t meets or exceeds the critical value, the results are statistically significant; accept the research hypothesis. If calculated t is less than the critical value, retain the null hypothesis.

Pair	*Column 1* X_1	*Column 2* X_2	*Column 3* *Difference Score* (D)	*Column 4* *Difference Score* *Square* (D^2)
1	150	145	5	25
2	145	147	−2	4
3	188	141	47	2,209
4	125	120	5	25
5	135	128	7	49
6	171	176	−5	25
7	148	156	−8	64
8	168	147	21	441
9	143	124	19	361
10	131	154	−23	529
			$\Sigma D = 66$	$\Sigma D^2 = 3{,}732$

(*Step 6*) $\overline{D} = \frac{\Sigma D}{n} \qquad \overline{D} = \frac{66}{10} \qquad \overline{D} = 6.6$

(*Step 7*) $\Sigma d^2 = \Sigma D^2 - \frac{(\Sigma D)^2}{n} \qquad \Sigma d^2 = 3{,}732 - \frac{(66)^2}{10}$

$\Sigma d^2 = 3{,}732 - \frac{4{,}356}{10} \qquad \Sigma d^2 = 3{,}732 - 435.6 \qquad \Sigma d^2 = 3{,}296.4$

(*Step 8*) $s_D = \sqrt{\frac{\Sigma d^2}{n-1}} \qquad s_D = \sqrt{\frac{3{,}296.4}{9}} \qquad s_D = \sqrt{366.27} \qquad s_D = 19.14$

(*Step 9*) $s_{\overline{D}} = \frac{s_D}{\sqrt{n}} \qquad s_{\overline{D}} = \frac{19.14}{\sqrt{10}} \qquad s_{\overline{D}} = \frac{19.14}{3.16} \qquad s_{\overline{D}} = 6.06$

(*Step 10*) $t = \frac{\overline{D} - 0}{s_{\overline{D}}} \qquad t = \frac{6.6 - 0}{6.06} \qquad t = 1.09$

(*Step 11*) $df = n - 1 \qquad df = 10 - 1 \qquad df = 9$

(*Step 12*) t critical = 1.833

(*Step 13*) Interpretation: Results are not statistically significant (for a one-tailed test); retain the null hypothesis.

TABLE B.2 Critical Values of t

One-Tailed Tests →	.05	.025	.005
Two-Tailed Tests →	.10	.05	.01
df			
1	6.314	12.706	63.657
2	2.920	4.303	9.925
3	2.353	3.182	5.841
4	2.132	2.776	4.604
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.895	2.365	3.499
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
11	1.796	2.201	3.106
12	1.782	2.179	3.055
13	1.771	2.160	3.012
14	1.761	2.145	2.977
15	1.753	2.131	2.947
16	1.746	2.120	2.921
17	1.740	2.110	2.898
18	1.734	2.101	2.878
19	1.729	2.093	2.861
20	1.725	2.086	2.845
21	1.721	2.080	2.831
22	1.717	2.074	2.819
23	1.714	2.069	2.807
24	1.711	2.064	2.797
25	1.708	2.060	2.787
26	1.706	2.056	2.779
27	1.703	2.052	2.771
28	1.701	2.048	2.763
29	1.699	2.045	2.756
30	1.697	2.042	2.750
40	1.684	2.021	2.704
60	1.671	2.000	2.660
120	1.658	1.980	2.617
∞	1.645	1.960	2.576

- If the *df* for your test falls in between levels, use the level with fewer *df*.
- Use the significance level for one-tailed tests in the top row if the hypothesis predicted a specific directional difference.
- Use the significance level for two-tailed tests in the bottom row if the hypothesis predicted a difference without specifying the direction of the difference.

SOURCE: Adapted from *Statistical Tables for Biological, Agricultural, and Medical Research,* 6th edition, by R. A. Fisher and F. Yates, 1963. London: Hafner.

CORRELATION

$$r = \frac{\Sigma xy}{\sqrt{(\Sigma x_2)(\Sigma y_2)}}$$

Step 1: Set up five columns and list data for variable X in column 1 and data for variable Y in column 2.

Step 2: Compute X^2 for column 3 by squaring scores in column 1.

Step 3: Compute Y^2 for column 4 by squaring scores in column 2.

Step 4: Compute XY for column 5 by multiplying scores (X) in column 1 by scores (Y) in column 2.

Step 5: Find ΣX, ΣY, ΣX^2, ΣY^2, and ΣXY by computing the sums for columns 1 through 5.

Step 6: Compute Σx_2.

Step 7: Compute Σy_2.

Step 8: Compute Σxy.

Step 9: Compute r.

Step 10: Calculate degrees of freedom (number of pairs minus 2).

Step 11: Identify critical value of r at .05 significance level corresponding to the degrees of freedom (calculated in step 10) in Table B.3.

Step 12: Interpret r: If calculated r meets or exceeds the critical value, the results are statistically significant; accept the research hypothesis. If calculated r is less than the critical value, retain the null hypothesis.

ID number	*Column 1* X	*Column 2* Y	*Column 3* X^2	*Column 4* Y^2	*Column 5* XY
101	14	15	196	225	210
102	12	10	144	100	120
103	8	12	64	144	96
104	6	7	36	49	42
105	10	10	100	100	100
	$\Sigma X = 50$	$\Sigma Y = 54$	$\Sigma X^2 = 540$	$\Sigma Y^2 = 618$	$\Sigma XY = 568$

(*Step 6*) $\Sigma x_2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$ $\quad \Sigma x_2 = 540 - \frac{(50)^2}{5}$ $\quad \Sigma x_2 = 540 - \frac{2{,}500}{5}$

$\Sigma x_2 = 540 - 500$ $\quad \Sigma x_2 = 40$

(*Step 7*) $\Sigma y_2 = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n}$ $\quad \Sigma y_2 = 618 - \frac{(54)^2}{5}$ $\quad \Sigma y_2 = 617 - \frac{2{,}916}{5}$

$\Sigma y_2 = 618 - 583.2$ $\quad \Sigma y_2 = 34.8$

(*Step 8*) $\Sigma xy = \Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n}$ $\quad \Sigma xy = 568 - \frac{(50)(54)}{5}$

$\Sigma xy = 568 - \frac{2{,}700}{5}$ $\quad \Sigma xy = 568 - 540$ $\quad \Sigma xy = 28$

(*Step 9*) $r = \frac{\Sigma xy}{\sqrt{(\Sigma x_2)(\Sigma y_2)}}$ $\quad r = \frac{28}{\sqrt{(40)(34.8)}}$ $\quad r = \frac{28}{\sqrt{1392}}$ $\quad r = \frac{28}{37.31}$

$r = .75$

(*Step 10*) $df = n - 2$ $\quad df = 5 - 2$ $\quad df = 3$

(*Step 11*) Critical $r = .805$ (one-tailed test)

(*Step 12*) Interpretation: Results are not statistically significant (for a one-tailed test); retain the null hypothesis.

TABLE B.3 Critical Values of *r*

One-Tailed Tests →	.05	.025	.005
Two-Tailed Tests →	.10	.05	.01
df			
1	.988	.997	.9999
2	.900	.950	.990
3	.805	.878	.959
4	.729	.811	.917
5	.669	.754	.874
6	.622	.707	.834
7	.582	.666	.798
8	.549	.632	.765
9	.521	.602	.735
10	.497	.576	.708
11	.476	.553	.684
12	.458	.532	.661
13	.441	.514	.641
14	.426	.497	.623
15	.412	.482	.606
16	.400	.468	.590
17	.389	.456	.575
18	.378	.444	.561
19	.369	.433	.549
20	.360	.423	.537
25	.323	.381	.487
30	.296	.349	.449
35	.275	.325	.418
40	.257	.304	.393
45	.243	.288	.372
50	.231	.273	.354
60	.211	.250	.325
70	.195	.232	.302
80	.183	.217	.283
90	.173	.205	.267
100	.164	.195	.254

- If the *df* for your test falls in between levels, use the level with fewer *df*.
- Use the significance level for one-tailed tests in the top row if the hypothesis predicted a specific directional relationship.
- Use the significance level for two-tailed tests in the bottom row if the hypothesis predicted a relationship without specifying the direction of the relationship.

SOURCE: Adapted from *Statistical Tables for Biological, Agricultural, and Medical Research*, 6th edition, by R. A. Fisher and F. Yates, 1963. London: Hafner.