Statistical tests

Why do we use statistical tests?

Statistical tests are a tool for helping support sample data where sample sizes are small or any trends are unclear. They allow null hypotheses behind research questions to be accepted or rejected. Students must return to the data to discuss possible geographical explanations

Statistical test Determines the level of confidence in sample data (= significance level), by calculating a statistical value which is used to accept or reject a null hypothesis. The statistical value is compared to a critical value

Null hypothesis (H_0) There is NO significant difference / correlation / association i.e. there is no pattern in the data

Alternative hypothesis (H_1) There IS a significant difference / correlation / association i.e. there is a pattern in the data

Parametric Parametric statistical tests make the assumption that the distribution of data is normal. You can check this by drawing a size frequency histogram.

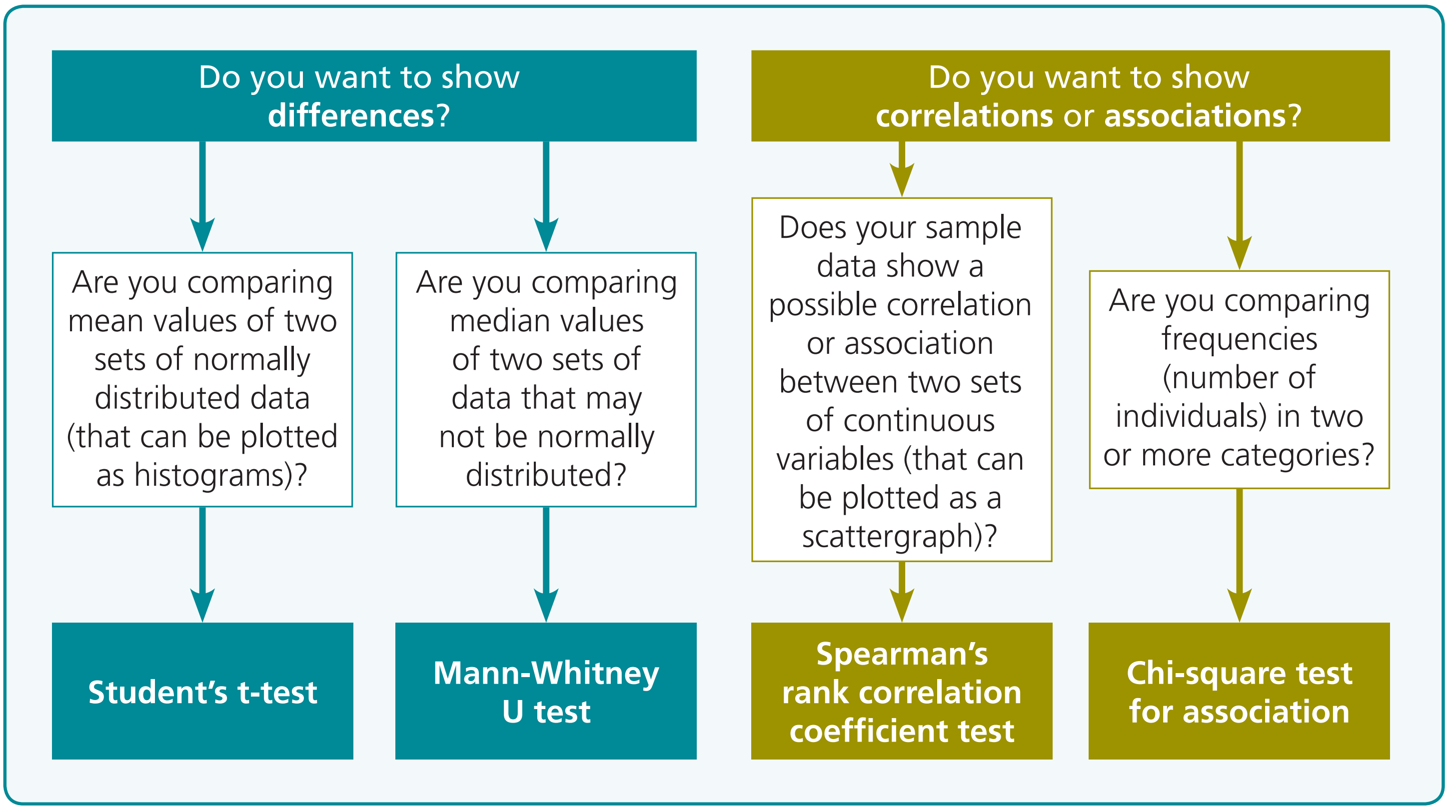
Non-parametric Non-parametric statistical tests make no assumptions about the distribution of the data

Critical values (from statistical tables) These depend on:

- **Probability levels** (p) A value of 0.05 (5% significance or 95% confidence) means there is a 0.05 probability that any difference, correlation or association could still have occurred by chance. Null hypotheses are rejected (significant) at probabilities less than or equal to 0.05
- **Degrees of freedom** The bigger the sample the more freedom there will be for the data to be statistically significant. Degrees of freedom are calculated slightly differently for different tests

Association Any type of relationship between two variables where as one variable changes then there is a corresponding change in the second variable

Correlation An association where the relationship between the two variables is linear



www.field-studies-council.org
© Field Studies Council 2018. This page may be copied for educational use only.

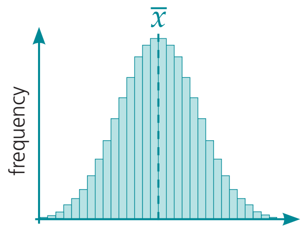
Measures of dispersion

Dispersion means the spread of data

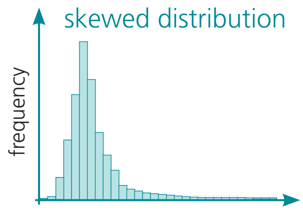
Range is the difference between the maximum and minimum values of a particular data set

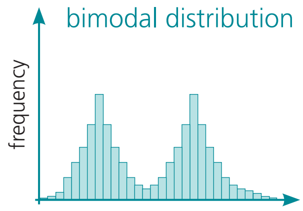
Normal distribution When measurements are plotted on a frequency histogram, they form a symmetrical bell-shaped curve

The mean ($\overline{x}$) in the middle. There is an equal number of smaller and larger values on either side



Non-normal distribution When measurements are plotted on a frequency histogram, they do not form a symmetrical bell-shaped curve





© Field Studies Council 2018. This page may be copied for educational use only.

Student's t-test

This test is used to determine whether the sample means of two data sets are significantly different

- data must be be normally distributed
- designed for small sample sizes ($n < 30$)

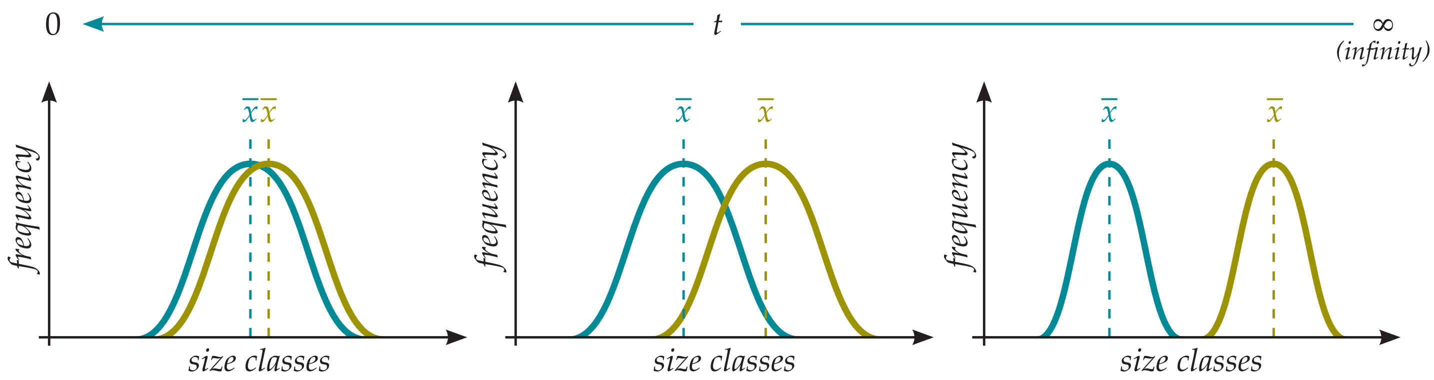
t varies from 0 to *infinity*

$t = 0$ means that the sample means ($\bar{x}$) of the two data sets are the same. The higher the t value, the greater the difference between the sample means of the two data sets

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\bar{x}$ = mean value of x
s = standard deviation
n = sample size
(for data sets 1 and 2)

$$\textit{degrees of freedom} = (n_1 + n_2) - 2$$



Critical values from a statistical table		
Degrees of freedom	p = 0.05	p = 0.01
20	2.09	2.85
21	2.08	2.83
22	2.07	2.82
23	2.07	2.81
24	2.06	2.80
25	2.06	2.79
26	2.06	2.78
27	2.05	2.77
28	2.05	2.76
29	2.05	2.76
30	2.04	2.75
31	2.04	2.74
32	2.04	2.74
33	2.03	2.73
34	2.03	2.73
35	2.03	2.72

The null hypothesis is rejected at a stated probability value (e.g. p = 0.05) and for a given sample size (degrees of freedom) if calculated t is ***greater than or equal*** *to the critical value*

© Field Studies Council 2018. This page may be copied for educational use only.
FSC

Student's t-test

Carrying out the test

1. Make a null hypothesis (the statement we are hoping to disprove)
2. Count the number of values in each data set (n)
3. Calculate the mean ($\bar{x}$) for each data set
4. Calculate the standard deviation for each data set (use a calculator or spreadsheet)
5. Calculate t using the formula

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\bar{x}$ = mean value of x
s = standard deviation
n = sample size
(for data sets 1 and 2)

6. Test significance by comparing your calculated value of t against the critical value and conclude
7. Explain the result using geographical knowledge and understanding

Worked example

A student investigating the effectiveness of sea defences on a shingle beach compared the difference in sediment length on either side of a groyne.

Side	Sediment length (mm)										❷ n	❸ $\bar{x}$	❹ s
East	12	23	21	17	24	31	29	25	26	24	10	23.2	5.53
West	30	27	26	31	28	28	35	29	21	25	10	28.0	3.74

❶ Null hypothesis: 'There is no significant difference between sediment length on the eastern and western sides of a groyne.'

❺ $$t = \frac{|23.2 - 28.0|}{\sqrt{\frac{(5.53)^2}{10} + \frac{(3.74)^2}{10}}} \qquad t = \frac{4.8}{\sqrt{\frac{30.62}{10} + \frac{14.0}{10}}} \qquad t = \frac{4.8}{\sqrt{4.46}} \qquad t = 2.27$$

❻ The t value (2.27) is greater than critical value (2.10) at the 0.05 probability level with 18 degrees of freedom. The null hypothesis is rejected. Therefore there is a less than 0.05 probability that the difference between sediment length on the eastern and western sides of a groyne is due to chance.

© Field Studies Council 2018. This page may be copied for educational use only.

Mann Whitney U test

This test is used to determine whether the medians of two data sets are significantly different

- data do not need to be normally distributed
- both data sets need a sample size of more than 5 ($n > 5$)
- the data sets do not have to be the same size (n_1 can be different to n_2)

$$U_1 = n_1 \times n_2 + 0.5\, n_2\, (n_2 + 1) - \Sigma R_2$$
$$U_2 = n_1 \times n_2 + 0.5\, n_1\, (n_1 + 1) - \Sigma R_1$$

n_1 = sample size of the first data set
n_2 = sample size of the second data set
ΣR_1 = sum of the ranks of the first data set
ΣR_2 = sum of the ranks of the first data set

Critical values from a statistical table

(p = 0.05)		n_1										
		5	6	7	8	9	10	11	12	13	14	15
n_2	5	2	3	5	6	7	8	9	11	12	13	14
	6		5	6	8	10	11	13	14	16	17	19
	7			8	10	12	14	16	18	20	22	24
	8				13	15	17	19	22	24	26	29
	9					17	20	23	26	28	31	34
	10						23	26	29	33	36	39
	11							30	33	37	40	44
	12								37	41	45	49
	13									45	50	54
	14										55	59
	15											64

*The null hypothesis is rejected at a stated probability value (e.g. p = 0.05) and for the given sample sizes (n_1 and n_2) if calculated U is **less than or equal** to the critical value*

www.field-studies-council.org
© Field Studies Council 2018. This page may be copied for educational use only.

Mann Whitney U test

Carrying out the test

1. Make a null hypothesis (the statement we are hoping to disprove)
2. Arrange the measurements from both data sets in order from left to right
3. Rank the measurements from lowest to highest as if they were one set of data (when two or more values are the same, a mean rank is calculated)
4. Calculate the sum (total) of the ranks ΣR_1 and ΣR_2
5. Calculate U_1 and U_2
6. Use the smaller U value as your statistic
7. Test significance by comparing your calculated value of U against the critical value and conclude
8. Explain the result using geographical knowledge and understanding

Worked example

A student has measured infiltration rates in two areas: coniferous and deciduous woodland.

	Infiltration rate (mm / hour)							
Coniferous	50	106	125	97	112	167	89	246
Deciduous	100	240	188	240	104	94	68	27

(1) Null hypothesis: 'There is no significant difference between infiltration rates in coniferous and deciduous woodland.'

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Coniferous (2)		50		89		97			106	112	125	167				246
(3) Rank (R_1)		2		4		6			9	10	11	12				16
Deciduous (2)	27		68		94		100	104					188	240	240	
(3) Rank (R_2)	1		3		5		7	8					13	14.5	14.5	

(4) Sum the ranks for each set of data

Site A: $\Sigma R_1 = 2 + 4 + 6 + 9 + 10 + 11 + 12 + 16 = 70$

Site B: $\Sigma R_2 = 1 + 3 + 5 + 7 + 9 + 13 + 14.5 + 14.5 = 66$

(5) Calculate U_1 and U_2

$U_1 = 8 \times 8 + 0.5 \times 8\ (8+1) - 70 = 30$

$U_2 = 8 \times 8 + 0.5 \times 8\ (8+1) - 66 = 34$

(6) Use the smaller U value as your statistic.

(7) The U value (30) is greater than the critical value (13) at the 0.05 probability level with sample sizes of $n_1 = 8$ and $n_2 = 8$. The null hypothesis is not rejected.

Therefore there is at least a 0.05 probability that the difference in infiltration rates between the two areas is due to chance.

www.field-studies-council.org
© Field Studies Council 2018. This page may be copied for educational use only.

Spearman's rank correlation coefficient test

This test is used to determine if there is a significant correlation between two variables

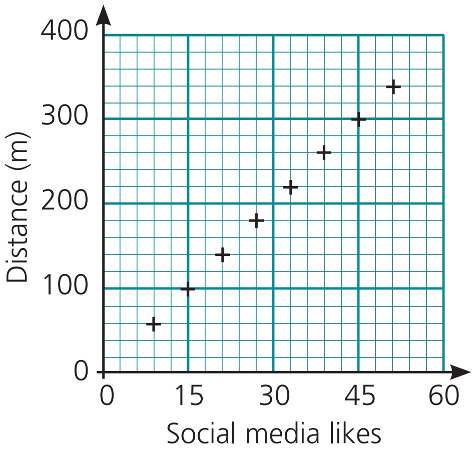
- number of pairs of data should be 10 or greater and 30 or less ($10 \leq n \leq 30$)

r_s varies from $+1.0$ to -1.0

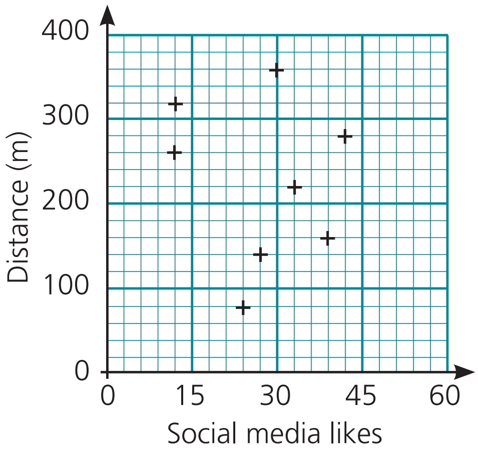
$$r_s = 1 - \left(\frac{6\Sigma D^2}{n(n^2-1)} \right)$$

D = difference between ranks
n = number of pairs of data (sample size)

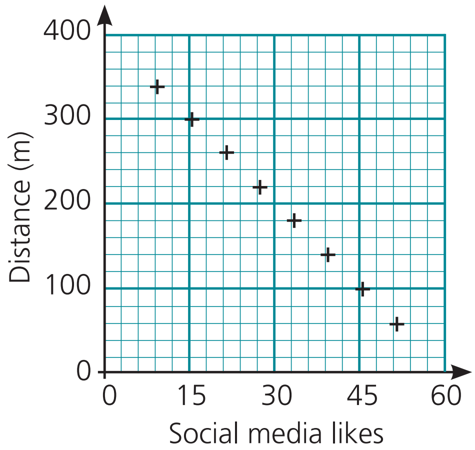
A larger sample size is preferable to a smaller sample size. As geographers, statistics allow us to quantify our confidence in the statements that we make based on our data. The larger the sample size used, the more confidently we can make these kinds of statements – this can be seen by looking at how the critical values for a given test change with increasing sample size.



Perfect positive correlation: $r_s = +1.0$



No correlation: $r_s = 0$



Perfect negative correlation: $r_s = -1.0$

Critical values from a statistical table		
Number of pairs (n)	p = 0.01	p = 0.05
10	0.794	0.648
11	0.755	0.618
12	0.727	0.587
13	0.703	0.560
14	0.679	0.538
15	0.654	0.521
16	0.635	0.503
17	0.618	0.488
18	0.600	0.472
19	0.584	0.460
20	0.570	0.47

Critical values from a statistical table		
Number of pairs (n)	p = 0.01	p = 0.05
21	0.556	0.436
22	0.544	0.425
23	0.532	0.416
24	0.521	0.407
25	0.511	0.398
26	0.501	0.390
27	0.492	0.383
28	0.483	0.375
29	0.475	0.368
30	0.467	0.362

*The null hypothesis is rejected at a stated probability value (e.g. p = 0.05) and for a given sample size (number of pairs) if calculated r_s is **greater than or equal** the critical value*

© Field Studies Council 2018. This page may be copied for educational use only.

Spearman's rank correlation coefficient test

Carrying out the test

1. Make a null hypothesis (the statement we are hoping to disprove)
2. Rank each variable separately
(when two or more values are the same, a mean rank is calculated)
3. Calculate the difference between the ranks (D)
4. Calculate the square of the difference (D^2)
5. Add together all the D^2 values (ΣD^2)
6. Calculate r_s
7. Test significance by comparing your calculated value of r_s against the critical value and conclude
8. Explain the result using geographical knowledge and understanding

Worked example

A student investigating the impact of a flagship scheme on an urban area, took a series of photos at increasing distances from the Library of Birmingham. The photos were uploaded to social media and the number of likes each received after 24 hours recorded.

Variable 1: distance (m)	Rank length (R_1) (2)	Variable 2: likes	Rank mass (R_2) (2)	Difference (D) (3)	D^2 (4)
0	1	55	15	-14	196
50	2	17	13	-11	121
100	3	22	14	-11	121
150	4	11	10.5	-6.5	42.25
200	5	6	7	-2	4
250	6	12	12	-6	36
300	7	7	8	-1	1
350	8	1	2	6	36
400	9	5	6	3	9
450	10	11	10.5	-0.5	0.25
500	11	9	9	2	4
550	12	2	4.5	7.5	56.25
600	13	1	2	11	121
650	14	2	4.5	9.5	90.25
700	15	1	2	13	169
				(5) ΣD^2	1007

1. Null hypothesis: 'There is no significant correlation between distance from a flagship scheme and visual appreciation (social media likes).'

6. $$r_s = 1 - \left(\frac{6 \times 1007}{15\,(15^2-1)}\right)$$

$$r_s = 1 - \left(\frac{6042}{3360}\right) \qquad r_s = -0.798$$

7. The r_s value (-0.798) is greater than the critical value (0.738) at the 0.05 probability level where $n = 8$. The null hypothesis is rejected.

Therefore there is a less than 0.05 probability that the correlation between the distance from the flagship scheme and visual appreciation is due to chance.

© Field Studies Council 2018. This page may be copied for educational use only.

Chi-square test for association

This is used to determine whether an observed distribution of a variable differs from a theoretical distribution or from a previous set of data

- data should be in the form of frequencies in a number of categories (i.e. nominal data)
- observations should be independent (i.e. one observation does not affect another)
- all expected frequencies should be larger in value than 5
- no more than 20 categories in total

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

O = observed values
E = expected values
Σ = sum of

Observed values (O) Frequencies in each category that are actually counted or observed

Expected values (E) Frequencies predicted in each category if the null hypothesis were true (e.g. a null hypothesis might predict an even spread of results across a series of categories)

$$degrees\ of\ freedom = (number\ of\ categories) - 1$$

Critical values from a statistical table		
Degrees of freedom	p = 0.01	p = 0.05
1	6.64	3.84
2	9.21	5.99
3	11.35	7.81
4	13.28	9.49
5	15.09	11.07
6	16.81	12.59
7	18.48	14.07
8	20.09	15.51
9	21.67	16.92
10	23.21	18.31
11	24.73	19.68
12	26.22	21.02
13	27.69	22.36
14	29.14	23.69
15	30.58	24.99
16	32.00	26.30
17	33.41	27.59
18	34.81	28.87
19	36.19	30.14
20	37.57	31.41

*The null hypothesis is rejected at a stated probability value (e.g. p = 0.05) and for a given size of data set (degrees of freedom) if calculated χ^2 is **greater than or equal** to the critical value*

© Field Studies Council 2018. This page may be copied for educational use only.

Chi-square test for association

Carrying out the test

1. Make a null hypothesis (the statement we are hoping to disprove)
2. Calculate the expected values

$$\textit{expected value} = \frac{\textit{row total} \times \textit{column total}}{\textit{grand total}}$$

3. Calculate the difference between each observed value its corresponding expected value $(O - E)$
4. Calculate the square of the difference $(O - E^2)$
5. Calculate the square of the difference divided by the expected value $(O - E)^2 \div E$
6. Calculate χ^2 by adding together all the $(O - E)^2 \div E$ values
7. Test significance by comparing your calculated value of χ^2 against the critical value and conclude
8. Explain the result using geographical knowledge and understanding

Worked example

A student investigating whether Gloucester services was more sustainable than its neighbouring Michaelwood services, asked whether there was an association between reusable coffee cup use and service station chosen. They tallied the total number of people leaving each service station in one hour (6 x 10-minute counts at hourly intervals throughout the day).

1 Null hypothesis: 'There is no significant association between service station and reusable cup use.'.

Service station	Observed values (O)	2 Expected values (E)	3 $(O - E)$	4 $(O - E)^2$	5 $(O - E)^2 \div E$
Gloucester	764	598	148	21 904	36.6
Michaelwood	432	598	-166	17 556	46.1
TOTAL	1196	1196			6 82.7

Calculated $\chi^2 = 82.7$

7 The χ^2 value (82.7) is greater than critical value (6.635)
at the 0.01 probability level with 1 degree of freedom. The null hypothesis is rejected.

Therefore there is a less than 0.01 probability that the association between service station and reusable cup use is due to chance.

www.field-studies-council.org
© Field Studies Council 2018. This page may be copied for educational use only.