

The background is a complex, abstract composition of various green elements. It includes several 3D bar charts of different heights and widths, some with a gradient from light to dark green. Overlaid on these are numerous lines: some are straight and jagged, resembling line graphs; others are smooth and curved, suggesting statistical models or confidence intervals. The lines vary in opacity, with some being solid and others semi-transparent. The overall color palette is a range of greens, from pale lime to deep forest green, set against a white background.

Statistics

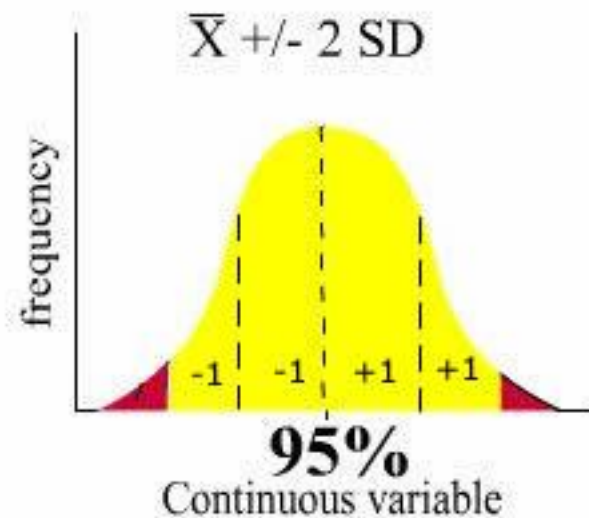
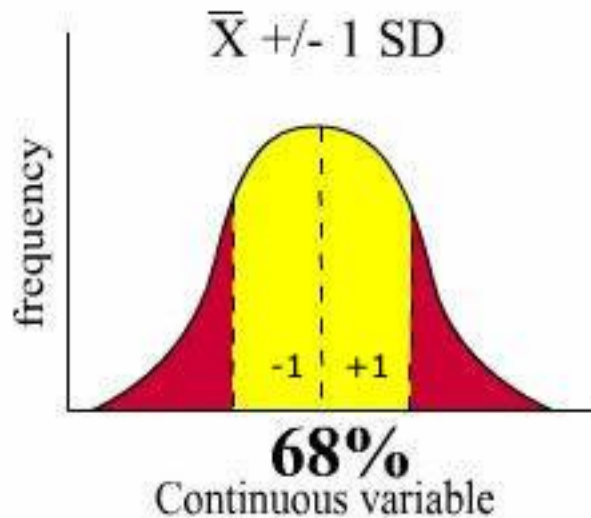
For biologists

Standard Deviation

Standard deviation shows us the spread of the values about the mean. Both of these graphs show normal distributions.

The standard deviation can be shown as error bars on graphs

Standard deviation as a measure of variation in samples



Hypothesis

The holly berries from *Ilex aquifolium* trees with variegated leaves are smaller than berries from a non variegated tree.



$$\sqrt{\frac{\sum(X - \bar{X})^2}{(n - 1)}}$$

where:

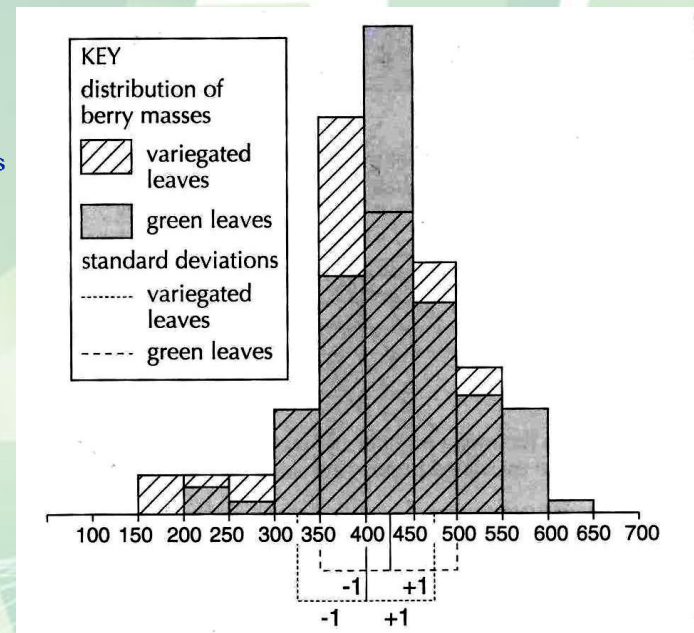
X = each score

\bar{X} = the mean or average

n = the number of values

Σ means we sum across the values

Tree	Mean berry mass \pm 1mg	Standard deviation mg
Green leaves	427	73
Variegated leaves	399	80



Null Hypothesis

The holly berries from *Ilex aquifolium* trees with variegated leaves are not smaller than berries from a non variegated tree.

This means that any differences are due to chance. In this instance, it would be reasonable to suggest that the null hypothesis is supported.

Hypothesis

Bank voles *Clethrionomys glareolus* on the Scottish Island of Raasay are larger than those on the Scottish mainland.



$$\sqrt{\frac{\Sigma(X - \bar{X})^2}{(n - 1)}}$$

where:

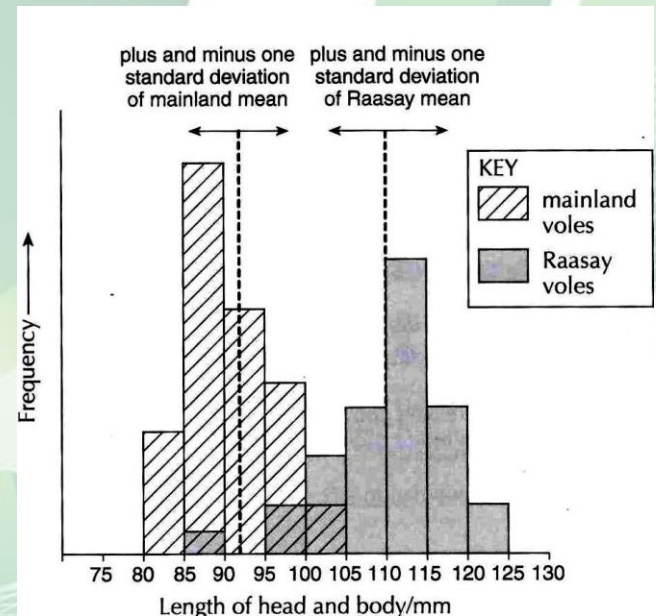
X = each score

\bar{X} = the mean or average

n = the number of values

Σ means we sum across the values

Vole population	Mean length \pm 1mm	Standard deviation mm
Mainland	82	5.2
Raasay	110	7.1



Null Hypothesis

The bank voles on Raasay are not larger than those on the mainland.

This means that any differences are due to chance. In this instance, it would be reasonable to suggest that the hypothesis is supported and there is a difference between the two populations. We can be almost certain about this!



T-test

A statistical test was required to determine the probability of the difference between two populations or two sets of results being due to chance.

Mr W.S. Gosset came up with the student's t-test:

KEY FACT

Here is the most usual version of the formula:

$$t = \frac{|\bar{x}_A - \bar{x}_B|}{\sqrt{\frac{(S_A)^2}{n_A} + \frac{(S_B)^2}{n_B}}}$$

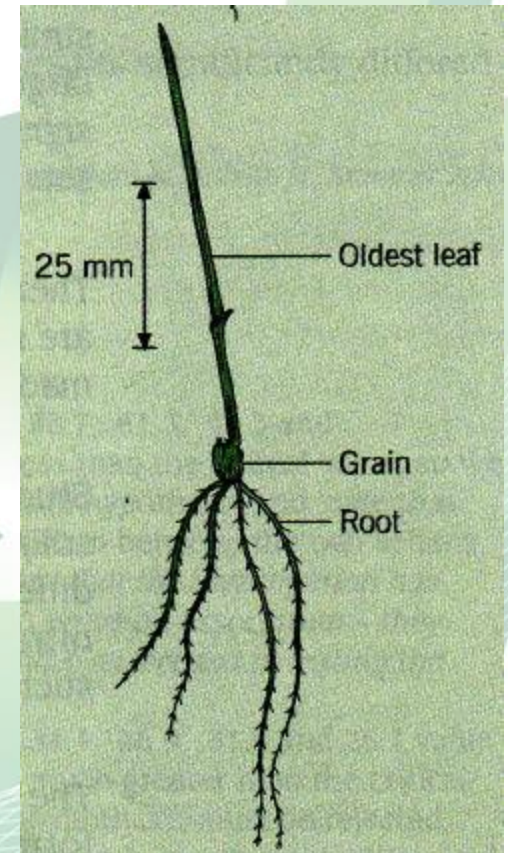
where $|\bar{x}_A - \bar{x}_B|$ is the difference in mean values of sample A and sample B. The vertical lines mean that the sign of the difference is not relevant. We just subtract the smaller from the larger number and ignore the sign.

$(S_A)^2$ and $(S_B)^2$ are the squares of the standard deviations of the samples and n_A and n_B are the sample sizes.

Example

Now let us deal with some data obtained by Open University students, who measured the lengths of leaves in 3-day germinated wheat seedlings that had been given different treatments. Batch A were grown from normal seeds and batch B from seeds that had been subjected to γ -radiation; here are their results:

	Normal, batch A	γ -irradiated, batch B
\bar{x} mean leaf length/mm	10.9	2.3
S standard deviation/mm	3.97	1.52
n sample size	15	15



Substitute these values into the equation:

$$t = \frac{|10.9 - 2.31|}{\sqrt{\left(\frac{15.76}{15}\right) + \left(\frac{2.31}{15}\right)}}$$
$$= \frac{8.6}{\sqrt{1.19}} = \frac{8.6}{1.09} = 7.89 \quad (\text{to 2 d.p.})$$

Degrees of freedom:

The degrees of freedom = (number in sample A – 1) + (number in sample B – 1)

In this case d.f. = (15-1) + (15-1) = 28

Now use the student t-test table

Degrees of freedom	Significance level					
	20% (0.20)	10% (0.10)	5% (0.05)	2% (0.02)	1% (0.01)	0.1% (0.001)
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.043	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.158	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

The t-test for matched and unmatched samples showing critical values of t at various significance levels.

Reject the null hypothesis if your value of t is larger than the tabulated value at the chosen significance levels for the calculated number of degrees of freedom.

Table 2 Probability that chance alone above could produce the difference between observed and expected results

Degrees of freedom	Significance level				
	0.1	0.05	0.02	0.01	0.001
18	1.734	2.101	2.552	2.878	3.922
28	1.701	2.048	2.467	2.763	3.674
	<p>↑</p> <p>Regarded as <i>not significant</i> – any difference could be due to chance.</p>	<p>↑</p> <p>Difference so great that this could happen by chance only 5 times out of 100.</p>	<p>←↑→</p> <p>We can be <i>confident</i> that there is a highly significant difference.</p>		<p>←</p> <p>We can be <i>very confident</i> that there is a highly significant difference.</p>

Conclusion

So the probability of getting a value of t at least as large as 7.89 is less than 0.05 - in fact it is much less than 0.01. So it is extremely unlikely that the difference in these two sets of data could have arisen by chance. We can reject the null hypothesis and describe the difference in the means of A and B as being highly significant.

Test Yourself

Exercise 12.1.1

- 1 A market gardener was testing the effectiveness of plastic plant pots over clay pots. He used seed from a pure inbred line – so all seeds were the same genotype. He grew 10 plants in plastic pots and 10 plants in clay pots and observed how long it took before each reached a flowering stage suitable for sale. Here are the results.

	A – clay	B – plastic
Number (n)	10	10
Mean of time/days (\bar{x})	95	100
Standard deviation/days (S)	3.2	4.6

State a null hypothesis.

Calculate the value of t and compare it with the values in the table at the appropriate probability level and the correct number of degrees of freedom.

Task

Hypothesis 1: The needles on male yew trees are longer than on female yew trees

Hypothesis 2: The needles on female yew trees are longer than on male yew trees

Null hypothesis: There is no significant difference between the length of needle on male and female yew trees.

Task

Collect at least 30 needles (leaves) from the male (outside house 1) and at least 30 from the female (outside house 2). Measure the lengths to the nearest millimetre. Record the values in an Excel document. Carry out a t-test and comment on which hypothesis is supported.

Chi-squared test (χ^2)

The Chi-square test is intended to test how likely it is that an observed distribution is due to chance. It is also called a "goodness of fit" statistic, because it measures how well the observed distribution of data fits with the distribution that is expected if the variables are independent.

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

O = the frequencies observed

E = the frequencies expected

\sum = the 'sum of'

Chi-squared test (χ^2)

You have just returned from a 3 year stint in the jungles western Africa, where you studied the habitat selected by the native bee eaters (a family of birds that specialize in catching bees and wasps on the wing, taking them to a perch, bashing their stingers out, and devouring them. At a pinch, they will eat other flying or hopping insects, such as grasshoppers. Three habitats were available to the bee eaters:

Habitat	Jungle	Grassland	Fields
% Area	75	10	15
# birds	86	3	11

Hypothesis: The birds have a preference for certain habitats.

Null hypothesis: The birds do not have a preference for a certain habitat.

Calculate the value for χ^2

Table I. Critical Values of χ^2

df	LEVEL OF SIGNIFICANCE FOR TWO-TAILED TEST					
	.20	.10	.05	.02	.01	.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	47.96	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

Chi-squared test (χ^2)

What are the degrees of freedom?

Is the null hypothesis supported by the data (use the p=0.05 column)

Table I. Critical Values of χ^2

df	LEVEL OF SIGNIFICANCE FOR TWO-TAILED TEST					
	.20	.10	.05	.02	.01	.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	4.34	6.25	7.82	9.84	11.34	
4	5.99	7.78	9.49	11.67	13.28	
5	7.88	9.24	11.07	13.39	15.09	
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	47.96	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

Chi-squared test (χ^2)

$$\chi^2 = 7.580$$

Using $p = 0.05$, $\chi^2 < 5.99$ means that the null hypothesis is supported, $\chi^2 > 5.99$ means that the hypothesis is supported. The chance of the observed results differing from the expected results for the null hypothesis is less than 5%.

The statistical test indicates that the birds really do prefer the jungle.

Online statistical analysis

- <http://graphpad.com/quickcalcs/catMenu/>



Simpson's Diversity Index

G.3.1 Calculate the Simpson diversity index for two local communities.

$$D = \frac{N(N-1)}{\sum n(n-1)}$$

D = Diversity index

N= total number of organisms of all species found

n= number of individuals of a particular species

G.3.2 Analyse the biodiversity of the two local communities using the Simpson index.

Species	Number, n	$n(n-1)$
A	10	
B	6	
C	3	
D	4	
E	1	
	$N =$	$\Sigma n(n-1) =$

Complete the table above and use the formula to determine D

N.B. A greater number of species and greater evenness in the numbers of individual species means a higher diversity.