

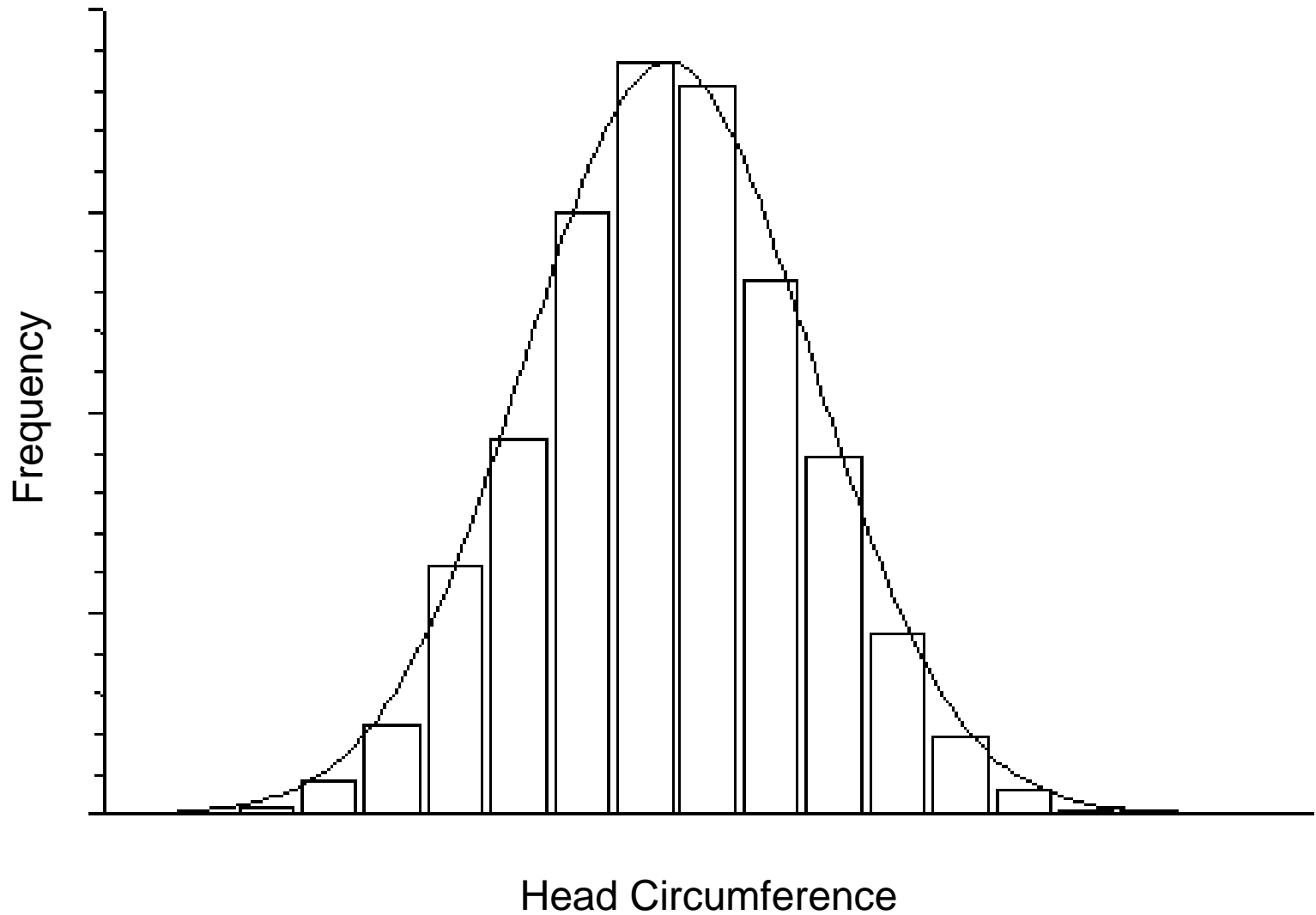


T – Test

Tests the difference between the means of two independent sets of measured data that are normally distributed.

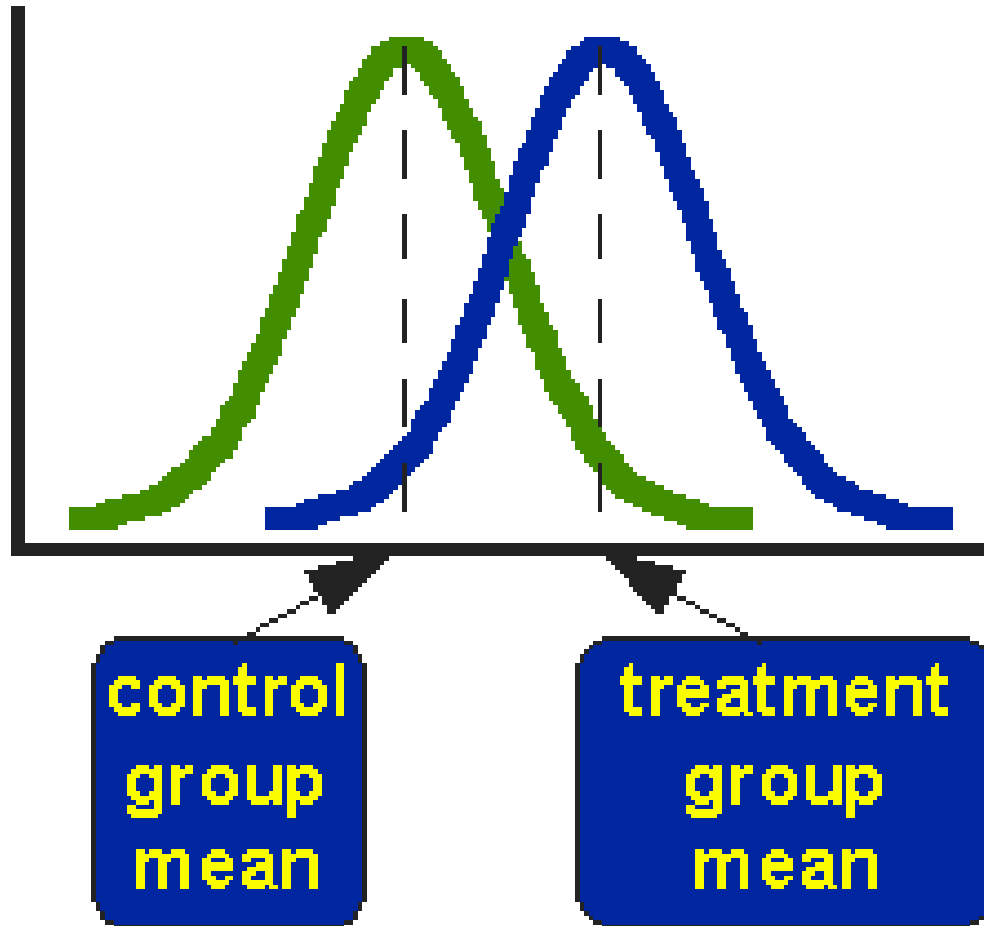


Year 13 Head Circumference data – NORMAL DISTRIBUTION





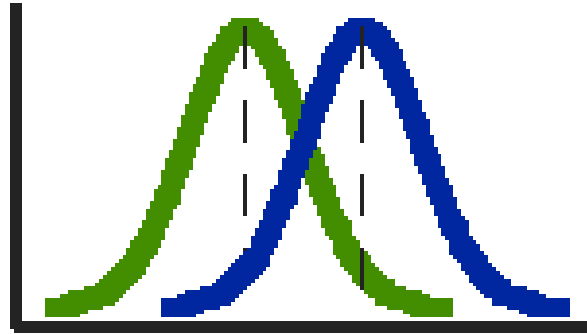
Are these means statistically different?



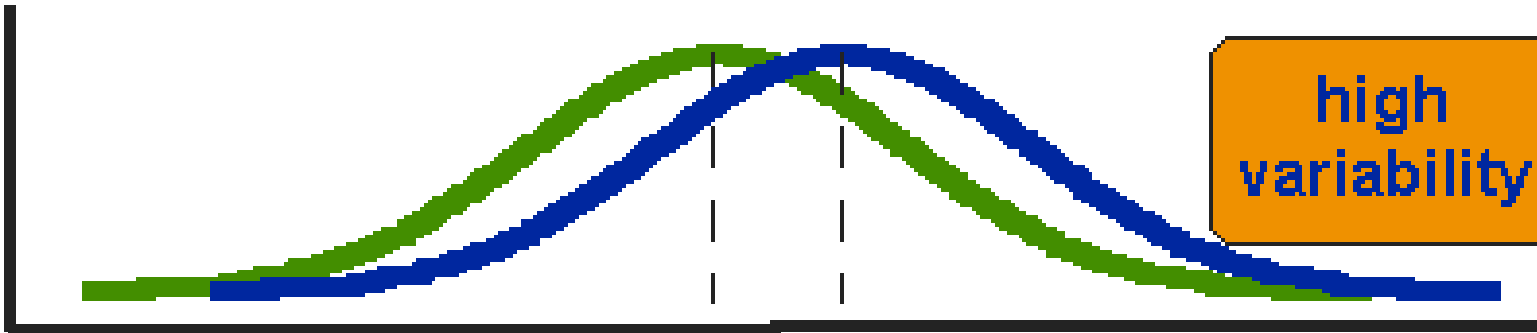


The t-test has to take account of the variation (standard deviation) of each set of data.

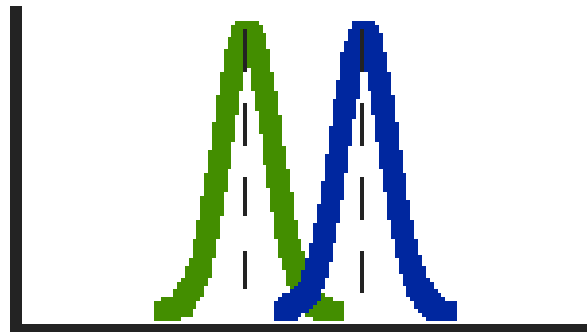
medium
variability



high
variability



low
variability



The difference between these means is identical, but are the differences significant??



So, to assess the significance we need to:

- Know the difference between the means.
- Know the variation in the data.

– Calculate the ratio of Difference : Variation

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

Difference between the two means.

Combined variation in the two data sets. (s = std dev)



What does this mean?

- If t is high, the difference between the means is more significant.
- If t is low, the variation in the data is too great to tell if the difference is real.



When used?

- With interval (accurately measured) data.
- With data that are unmatched.
 - There is a separate t-test for matched data.
- With normally distributed data.
 - A normality test should be done first.
- With samples that have a different Standard Deviation.



Question

- A student wanted to find out if buttercup plants grow taller if they are 1 metre away from a riverside or if they are 50 metres from the water.
- She systematically sampled 11 plants at each distance from a river using a transect and measured the height above ground level. She applied a t-test to the data.



State the hypotheses

- Null hypothesis:
 - There **is no** significant difference between the height of the plants at each distance from the water's edge.
- Alternative hypothesis:
 - The heights of plants at each site will be significantly different.
 - Note that in this case we are not predicting the direction of the difference, we are just saying they are different.



Tabulate the data

	Buttercup Plant height / cm										
1m from water	14.5	15.3	14.9	16.6	16.0	13.8	15.0	14.7	15.5	12.8	15.8
50m from water	11.7	9.6	10.4	6.6	8.2	10.9	10.6	9.2	14.1	11.3	12.5

Both sets of samples were systematically chosen from different areas of land.

Previous research has shown that height is normally distributed.

The ranges of these distributions are not similar, indicating that the standard deviations are probably not either.

A two tailed test is being used because H_A has no direction.



Calculate Mean & SD

	Buttercup Plant height / cm											Mean	SD
1m from water	14.5	15.3	14.9	16.6	16.0	13.8	15.0	14.7	15.5	12.8	15.8	15.0	1.1
50m from water	11.7	9.6	10.4	6.6	8.2	10.9	10.6	9.2	14.1	11.3	12.5	10.5	2.1

MS Excel will do this easily for you.



Calculate t using:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

$$= \frac{15.0 - 10.5}{\sqrt{(1.1^2/11 + 2.1^2/11)}}$$

$$= 4.5/\sqrt{0.51}$$

$$t = 6.30$$

If t is negative
then drop the sign
to make it
positive.

Calculate the Degrees of Freedom

- $\text{DoF} = n_x + n_y - 2$
- $\text{DoF} = 11 + 11 - 2 = 20$



Look in printed t-test tables

- Look for the critical value (t_{crit}) for the appropriate DoF.
- Reject the null hypothesis & accept the alternative hypothesis if t is greater than or equal to t_{crit} .

df	P = 0.05	P = 0.01	P = 0.001
1	12.71	63.66	636.61
2	4.30	9.92	31.60
3	3.18	5.84	12.92
4	2.78	4.60	8.61
5	2.57	4.03	6.87
6	2.45	3.71	5.96
7	2.36	3.50	5.41
8	2.31	3.36	5.04
9	2.26	3.25	4.78
10	2.23	3.17	4.59
11	2.20	3.11	4.44
12	2.18	3.05	4.32
13	2.16	3.01	4.22
14	2.14	2.98	4.14
15	2.13	2.95	4.07
16	2.12	2.92	4.02
17	2.11	2.90	3.97
18	2.10	2.88	3.92
19	2.09	2.86	3.88
20	2.09	2.85	3.85
21	2.08	2.83	3.82





So...

- $t_{\text{crit}} = 2.09$
- $t = 8.35$
- Reject H_0
 - There **is a** significant difference between the height of plants from the two areas.
 - The probability that the two samples are from the same population is less than 5% - this is good enough.



There is an easier way

- Enter your two samples into Excel.
- Enter the following formula into a cell:
`=TTEST(A3:A13,B3:B13, 1,2)`.
 - This gives you the probability of the two samples being the same – in this case 0.07 which is too high.

	A	B	C	D	E
1	Without fertiliser	With fertiliser			
2	(X)	(Y)			
3	55.2	58.2			
4	58.4	61.2		t test	0.07
5	60.2	64.3			
6	62.3	66.2			
7	64	69.1			
8	65.4	70.4			
9	66.9	71.6			
10	68.4	72			
11	70.8	73.8			
12	72.3	76.3			
13	75.3	82.4			

Click image to launch Excel