



t – Test for unmatched samples

Tests the difference between two independent sets of measured (interval) data that are normally distributed.



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 similar Standard Deviation.



Question

- A student wanted to find out if two adjacent areas of soil (one treated with fertiliser and one untreated) produced different amounts of biomass after 3 months.
- She randomly sampled 11 plots using a quadrat and measured the dry biomass of each $\frac{1}{4}$ square metre above ground level. She applied a t-test to the data.



State the hypotheses

- Null hypothesis:
 - There **is no** significant difference between the biomass produced from the treated & untreated soils.
- Alternative hypothesis:
 - The biomass produced by the treated & untreated soils 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

	Biomass yield (gm ⁻²)										
Without fertiliser	55.2	58.4	60.2	62.3	64.0	65.4	66.9	68.4	70.8	72.3	75.3
With fertiliser	58.2	61.2	64.3	66.2	69.1	70.4	71.6	72.0	73.8	76.3	82.4

Both sets of samples were randomly chosen from different areas of soil.

Previous research has shown that biomass is normally distributed.

The ranges of these distributions are similar, indicating that the standard deviations probably are also.

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

Call one sample X and the other Y

Calculate the sample means

Biomass (gm ⁻²)	
Without fertiliser (X)	With fertiliser (Y)
55.2	58.2
58.4	61.2
60.2	64.3
62.3	66.2
64.0	69.1
65.4	70.4
66.9	71.6
68.4	72.0
70.8	73.8
72.3	76.3
75.3	82.4




Biomass (gm ⁻²)	
Without fertiliser (X)	With fertiliser (Y)
55.2	58.2
58.4	61.2
60.2	64.3
62.3	66.2
64.0	69.1
65.4	70.4
66.9	71.6
68.4	72.0
70.8	73.8
72.3	76.3
75.3	82.4
Mean	
65.38	69.59





Calculate squares of each sample & add them up

Without fertiliser (X)	X^2	With fertiliser (Y)	Y^2
55.2	3047.04	58.2	3387.24
58.4	3410.56	61.2	3745.44
60.2	3624.04	64.3	4134.49
62.3	3883.29	66.2	4382.44
64.0	4096.00	69.1	4774.81
65.4	4277.16	70.4	4956.16
66.9	4475.61	71.6	5126.56
68.4	4678.56	72.0	5184.00
70.8	5012.64	73.8	5446.44
72.3	5227.29	76.3	5821.69
75.3	5670.09	82.4	6789.76
65.38	47400.28	69.59	53749.03


$$\bar{x}$$


$$\sum x^2$$


$$\bar{y}$$


$$\sum y^2$$





Calculate the Variance (s_x^2) of sample x

$$s_x^2 = \left(\frac{\sum x^2}{n_x} \right) - \bar{x}^2$$

$$\begin{aligned} s_x^2 &= 47400.28/11 - 65.38^2 \\ &= 34.57 \end{aligned}$$

- Where:
 - n_x = the number of items in sample x
 - \bar{x}^2 = the square of the mean of x.



Now do the same for the Variance (s_y^2) of sample y

- Where:
 - n_y = the number of items in sample y
 - \bar{y}^2 = the square of the mean of y .

$$s_y^2 = 53749.03/11 - 69.59^2$$
$$= 43.51$$



So...

- We now have the variance of sample x:
 - $s_x^2 = 34.57$
- And the variance of sample y:
 - $s_y^2 = 43.51$

Now calculate the estimated pooled
variance

$$\hat{\sigma}^2 = \frac{n_x s_x^2 + n_y s_y^2}{n_x + n_y - 2}$$
$$= \frac{11 \times 34.57 + 11 \times 43.51}{11 + 11 - 2}$$

$$= 858.88 / 20$$

$$= 42.944$$





Calculate t using:

$$t = \frac{(\bar{x} - \bar{y})\sqrt{n_x n_y}}{\hat{\sigma}\sqrt{n_x + n_y}}$$

$$= \frac{(65.38 - 69.59)\sqrt{121}}{6.55\sqrt{22}}$$

$$= -46.31/30.72$$

$$t = -1.51$$

$$t = 1.51$$

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...

- $t_{\text{crit}} = 2.09$
- $t = 1.51$
- Cannot Reject H_0
 - There **is no** significant difference between the biomass produced from the treated & untreated soils.
 - The probability that the two samples are from the same population is greater than 5% - this is too high.



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