Analysis of pair-matched data

The method applies to cases where each subject or member of a group is observed twice (e.g., before and after certain interventions), or matched pairs are measured for the same continuous characteristic, such as blood pressure before and after each took an oral contraceptive ; the insulin level in the blood before and after some kind of nerve stimulation.

In another exercise, a popular application is an epidemiological design called a ***pair-matched case–control study***. In case–control studies, cases of a specific disease are ascertained as they arise from population-based registers or lists of hospital admissions, and controls are sampled either as disease-free individuals from the population at risk or as hospitalized patients having a diagnosis other than the one under investigation.

Data from matched or before-and-after experiments should never be considered as coming from ***two independent samples*****.** The procedure is to reduce the data to a one-sample problem by computing before-and-after (or case-and control) difference for each subject or pairs of matched subjects. By doing this with paired observations, we get a set of differences.

$\mu_d = 0$

$S^2_{\bar{d}} = \frac{S^2}{n}$

$t = \frac{\bar{d} - \mu d}{Sd/\sqrt{n}}$

Example:. Trace metals in drinking water affect the flavor of the water, and unusually high concentrations can pose a health hazard. The bellow data shows trace- metal concentrations (zinc, in mg/L) for both surface water and bottom water at six different river locations. Tested this at $\alpha = 0.05$?

Zinc concentration in bottom(mg /L):.

0.430 , 0.266 , 0.567 , 0.531 , 0.707 , 0.716

Zinc concentration in surface(mg /L):.

0.415 , 0.238 , 0.390 , 0.410 , 0.605 , 0.609

Solution:.

Location	Bottom	Surface	Difference di	di^2
1	0.430	0.415	0.015	0.000225
2	0.266	0.238	0.028	0.000784
3	0.567	0.390	0.177	0.031329
4	0.531	0.410	0.121	0.014641
5	0.707	0.605	0.102	0.010404
6	0.716	0.609	0.107	0.011449
Total			***0.550***	***0.068832***

$\bar{d}$ = average difference = $\frac{0.550}{6}$ = 0.0917 mg /L

$$S^2_d = \frac{\sum di^2 - \frac{(\sum di)^2}{n}}{n-1} = \frac{0.068832 - \frac{(0.550)^2}{6}}{6-1} = 0.00368$$

$$S_{\bar{d}} = \sqrt[2]{0.00368} = 0.06066$$

$$SE_{(d)} = \sqrt{\frac{S^2}{n}} = \frac{Sd}{\sqrt{n}} = \frac{0.06066}{\sqrt{6}} = 0.02476$$

$$t = \frac{\bar{d} - \mu d}{SE(d)} = \frac{0.0917}{0.02476} = 3.703$$

When the test is two – sides , using *t tabulated value* at α = 0.025 for 5 df. is 2.571 . Since

$t_{cal.}$ = 3.703 > $t_{tab.}$ = 2.571

H_0 is rejected if

- $t_{cal} \leq$ - tabulated value for Ha:$\mu_1 < \mu_2$

or **$t_{cal} \geq$ tabulated value for Ha:$\mu_1 > \mu_2$**

We conclude that the null hypothesis of no difference should be rejected at the 0.05 level; there is enough evidence to support the hypothesis of different mean zinc concentrations (two-sided alternative).

Example:. The systolic blood pressures of n=12 women between the ages of 20 and 35 were measured before and after administration of a newly developed oral contraceptive. Tested this data at $\alpha = 0.05$?

Systolic Blood Pressure (mmHg):.

Before/ 122, 126 , 132 , 120 , 142 , 130 , 142 , 137 , 128 , 132 , 128 , 129.

After/ 127 , 128 , 140 , 119 , 145 , 130 , 148 , 135 , 129 , 137 , 128 , 133.

Solution:.

Systolic blood pressure(mmHg) *Before*	Systolic blood pressure(mmHg) *After*	Difference d_i	d_i^2
122	127	5	25
126	128	2	4
132	140	8	64
120	119	-1	1
142	145	3	9
130	130	0	0
142	148	6	36
137	135	-2	4
128	129	1	1
132	137	5	25
128	128	0	0
129	133	4	16
Total		***31***	***185***

$$\bar{d} = \frac{\Sigma di}{n}$$

$$= \frac{31}{12} = 2.58 \text{ mmHg}$$

$$S^2_d = \frac{\sum di^2 - \frac{(\sum di)^2}{n}}{n-1}$$

$$= \frac{185 - \frac{(31)^2}{12}}{12-1} = 9.537$$

$$S_d = \sqrt{Sd^2}$$

$$= \sqrt{9.537} = 3.088$$

$$SE_{(d)} = \frac{Sd}{\sqrt{n}}$$

$$= \frac{3.088}{\sqrt{12}} = 0.891$$

$$t = \frac{\bar{d} - \mu d}{SE(d)}$$

$$= \frac{2.58}{0.891} = 2.895$$

When the test was two – sides we must divided α value by 2 , then t tabulated value at α = 0.025 and df. = 11 is 2.201 , Since

$t_{cal.}$ 2.895 > $t_{tab.}$ 2.201

We conclude that the null hypothesis of no blood pressure change should be rejected at the 0.05 level ; there is enough evidence to support the hypothesis of increased systolic blood pressure.