(vi) Conclusion: ... critical region, so we reject our null hypothesis of equal means.

18.4.4. Testing Hypotheses about Two Means with Paired Observations. In testing hypotheses about two means, we have used independent samples, but there are many situations in which the two samples are *not* independent. This happens when the observations are found a pairs as the two observations of a pair are related to each other. Pairing occurs either naturally or by design. Natural pairing occurs whenever measurement is taken on the same unit or individual at two different times. For examples, suppose 10 young recruits are given a strenuous physical training programme by the Army. Their weights are recorded before they begin and after they complete the training. The two observations obtained for each recruit, *i.e.* the *before-and-after* measurements constitute natural pairing. Observations are also paired to eliminate effects in which there is no interest. For example, suppose we wish to test which of two types (A or B) of fertilizers is the better. The two types of fertilizers are applied to a number of plots and the results are noted. Assuming that the two types are found significantly different, we may find that part of the difference may be due to the different types of soil or different weather conditions, etc. Thus the real difference between the fertilizers can be found only when the plots are paired according to the same types of soil or same weather conditions, etc. We eliminate the undesirable sources of variation to take the observations in pairs. This is pairing by design.

When the observations from two samples are paired either naturally or by design, we find the difference between two observations of each pair. Treating the differences as a random sample from a normal population with mean $\mu_D = \mu_1 - \mu_2$ and unknown standard deviation σ_D, we perform a one-sample t-test on them. This is called a *paired difference t-test* or a *paired t-test*.

Testing the hypothesis $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$ is equivalent to testing $H_0 : \mu_D = 0$ against $H_1 : \mu_D \neq 0$.

Let $d_i = x_{1i} - x_{2i}$ denote the difference between the two sample observations in the *ith* pair. Then the sample mean and standard deviation of the differences are

$$\bar{d} = \frac{\Sigma d_i}{n} \text{ and } s_d = \frac{\Sigma (d_i - \bar{d})^2}{n-1},$$

where n represents the number of pairs.

Assuming that (i) $d_1, d_2, \ldots, d_n$ is a random sample of differences and (ii) the differences are normally distributed, the test-statistic

$$t = \frac{\bar{d}}{s_d / \sqrt{n}},$$

follows a t-distribution with $\upsilon = n - 1$ degrees of freedom. The rest of the procedure for testing the null hypothesis $H_0 : \mu_D = 0$ is the same.

Example 18.8. Ten young recruits were put through a stenuous physical training programme by the Army. Their weights were recorded before and after the training with the following results:

Recruit	1	2	3	4	5	6	7	8	9	10
Weight *before*	125	195	160	171	140	201	170	176	195	139
Weight *after*	136	201	158	184	145	195	175	190	190	145

Using $\alpha = 0.05$, would you say that the programme affects the average weight of recruits? Assume the distribution of weights before and after to be approximately normal. (P.U., B.A/B.Sc. 1984)

The pairing was natural here, since two observations are made on the same recruit at two different times. The sample consists of 10 recuirts with two measurements on each.

The test is carried out as below:

(i) We state our null and alternative hypotheses as

$$H_0 : \mu_D = 0 \text{ and } H_1 : \mu_D \neq 0$$

(ii) The significance level is set at $\alpha = 0.05$.

(iii) The test-statistic under H_0 is

$$t = \frac{\bar{d}}{s_d / \sqrt{n}},$$

which has a t-distribution with $n - 1$ degrees of freedom.

(iv) The critical region is $|t| \geq t_{0.025,(9)} = 2.262$.

(v) Computations.

Recruit	Weight		Difference, d_i (after minus before)	d_i^2
	Before	After		
1	125	136	11	121
2	195	201	6	36
3	160	158	−2	4
4	171	184	13	169
5	140	145	5	25
6	201	195	−6	36
7	170	175	5	25
8	176	190	14	196
9	195	190	−5	25
10	139	145	6	36
Σ	1672	1719	47	673

Now $\bar{d} = \dfrac{\Sigma d_i}{n} = \dfrac{47}{10} = 4.7.$

$$s_d^2 = \frac{\Sigma(d_i - \bar{d})^2}{n-1} = \frac{1}{n-1}\left[\Sigma d_i^2 - \frac{(\Sigma d_i)^2}{n}\right]$$

$$= \frac{1}{9}\left[673 - \frac{(47)^2}{10}\right] = \frac{673 - 220.9}{9} = 50.23, \text{ so that}$$

$$s_d = \sqrt{50.23} = 7.09.$$

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{4.7}{7.09/\sqrt{10}} = \frac{(4.7)(3.16)}{7.09} = 2.09.$$

(vi) **Conclusion.** Since the calculated value of $t = 2.09$ does not fall in the critical region, so we accept H_0 and may conclude that the data do not provide sufficient evidence to indicate that the programme affects average weight.