

Paired 2 Sample t -Test

Suppose we are interested in the effect of different sampling strategies on the quality of data we recover from archaeological field surveys. For example, we would hypothesize that the width of survey transects between individuals effects the density estimate of artifacts per unit area, however, we wish to test this hypothesis formally.

To test this hypothesis we set up an experiment on a known productive archaeological landscape. We establish twenty five 50 meter by 50 meter plots, and walk them with a survey crew at two different transect widths, the first at 5 meter intervals, the second 10 meter intervals. Each time artifacts are encountered they are recorded, but left in place and the results are tabulated.

As this is a manipulated experiment we can no longer treat the two samples as independent, as we are interested in the mean difference between both methods on each plot. These samples are *dependent* and so we use a paired t -test. The paired test is similar to the independent 2 sample tests but with some important differences. As we are interested in the difference between the two techniques at each plot, we treat each sample plot as a pair of results (Y_{i5}, Y_{i10}) , and calculate the difference (d_i) , that is $d_{ij} = Y_{i5} - Y_{i10}$, where Y_{ij} = the recorded density of artifacts on plot i under treatment j . These differences give us a new distribution of twenty five values that has it's own sample statistics; we can then treat this new distribution as a 1 sample t -test, against the null hypothesis that the mean = zero. Therefore, let \bar{d} = the mean value of the differences between the two treatments. We state the null hypothesis at the $\alpha = 0.05$ level (95%):

$$H_o : \bar{d} = 0$$

$$H_A : \bar{d} \neq 0$$

The data is:

	5 meter	10 meter	Difference
1	9.998	15.384	-5.385
2	140.228	54.803	85.425
3	40.669	38.913	1.756
4	40.030	10.875	29.155
5	86.292	10.733	75.559
6	76.255	36.444	39.811
7	16.689	34.774	-18.085
8	115.963	49.252	66.712
9	161.497	51.759	109.737
10	29.529	0.643	28.886
11	37.765	0.908	36.857
12	16.919	114.969	-98.050
13	22.415	42.673	-20.259
14	4.496	43.370	-38.874
15	22.272	27.073	-4.801

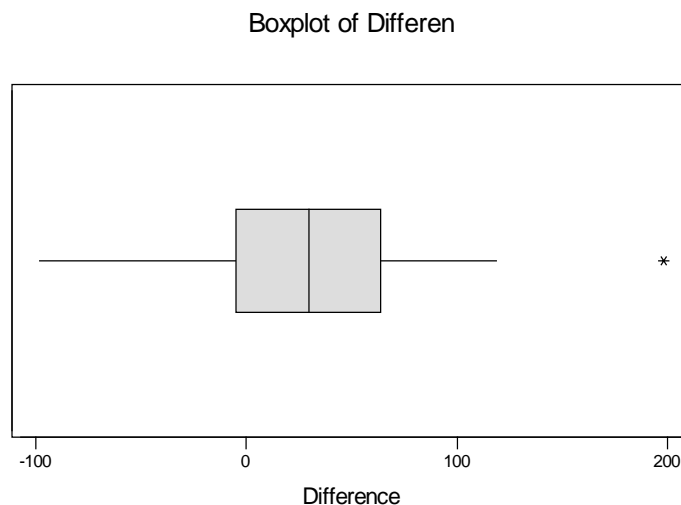
16	73.061	19.343	53.718
17	57.477	27.489	29.988
18	64.188	2.808	61.380
19	40.958	2.994	37.963
20	10.224	95.575	-85.351
21	91.245	53.564	37.682
22	38.836	32.265	6.571
23	160.985	42.102	118.883
24	1.452	3.544	-2.093
25	209.540	11.333	198.207

The important difference with the paired *t*-test is that the assumption of normality refers to the distribution of the differences, not the original data themselves. Therefore, when we run or calculate our descriptive statistics we are only concerned with the 1 sample of d_i 's.

The MINTAB output and boxplot for the descriptive statistics looks like:

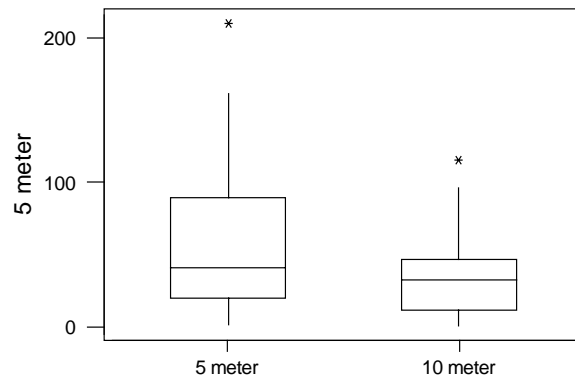
Descriptive Statistics							
Variable	N	Mean	Median	Tr Mean	StDev	SE Mean	
Differen	25	29.8	30.0	28.1	62.9	12.6	
Variable	Min	Max	Q1	Q3			
Differen	-98.0	198.2	-5.1	64.0			

And the boxplot:



Looking at the output, we see the mean is very close to the median, and the boxplot shows the distribution is pretty symmetrical, therefore we conclude it is normal and can proceed with the t -test.

In fact, our parent distributions are both non-normal as a boxplot demonstrates, but as we are interested in a paired test, this is of no consequence.



Our degrees of freedom in this test are $\nu = 24$, so our $t_{CRIT}(\nu = 24, \alpha = 0.05) = 2.064$.

To calculate the t_{STAT} we plug our numbers into the equation as usual:

$$t_{STAT} = \frac{\bar{d} - 0}{se_d} = \frac{29.8 - 0}{12.6} = 2.37$$

As our $t_{STAT} > t_{CRIT}$ we reject the null hypothesis that there is no difference between the results of the sampling strategies in favor of the alternative.

To establish our confidence limits at the 95% level:

$$L_L = 29.8 - 2.064(12.6) = 3.79$$

$$L_U = 29.8 + 2.064(12.6) = 55.81$$

As these bounds are both positive, they do not encompass zero and we reject the null hypothesis at the 95% level in favor of the alternative. Notice that this is a two-tailed test, and that even though the alternative hypothesis states simply that there is difference between the two samples, we can tell that treatment 1 (5 meter transects) is significantly greater than treatment 2 (10 meter transects) as both the bounds are positive.

The procedure in MINITAB is a little different for the paired test.

MINITAB does not have a set function for a paired t -test, but we trick it into doing one for us. In this test we are interested in the differences between the two samples and so we calculate this distribution in either EXCEL or MINITAB and transfer the single column of data into MINITAB and run a 1 sample t -test on this data.

To calculate the confidence limits:

```
>STAT
  >BASIC STATISTICS
    >1 SAMPLE t
      >Click CONFIDENCE INTERVAL set at 95%
        >GRAPHS; select BOXPLOT
          >OKAY
            >OKAY
```

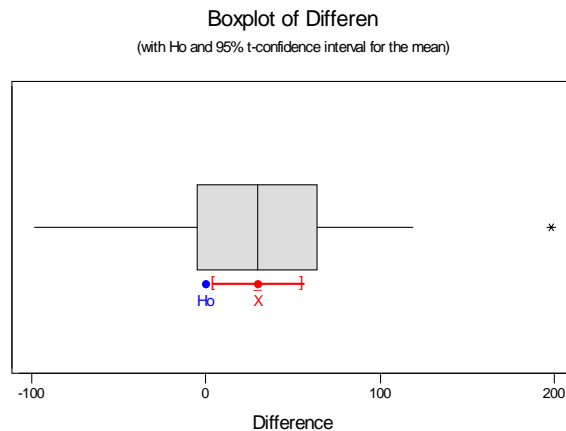
The output for this test looks like:

Confidence Intervals

Variable	N	Mean	StDev	SE Mean	95.0 % CI
Differen	25	29.8	62.9	12.6	(3.8, 55.8)

And checking our manual calculations, we see our confidence limits agree with MINITAB's if we were to round them to one decimal place.

The boxplot shows:



The boxplot shows us the confidence interval around \bar{d} (x on the boxplot) at the 95% level and the position of our null value, zero. As it falls outside the confidence interval we reject the null hypothesis in favor of the alternative.

For the actual t -test:

```
>STAT
  >BASIC STATISTICS
    >1 SAMPLE t
      >Click TEST MEAN, set at 0
        >ALTERNATIVE should be NOT EQUAL
          >OKAY
```

And the output is:

T-Test of the Mean

Test of mu = 0.0 vs mu not = 0.0

Variable	N	Mean	StDev	SE Mean	T	P
Differen	25	29.8	62.9	12.6	2.37	0.026

We find that the T value MINITAB gives us is the same as the one we calculated 2.37. The final $p = 0.026$, and as this is less than our stated $\alpha = 0.05$, we reject the null hypothesis of no difference in favor of the alternative. Now, notice that our $p = 0.026$, and as we are dealing with a two-tailed test, we might think we need to compare this to $\alpha/2 = 0.025$, in which case we would fail to reject the null hypothesis. However, t tables are (almost) always based on two-tailed tests and so our α level takes into account both sides of the distribution such that there is no need for us to divide α by 2; it is already done for us.