"Comparison between two statistical tests of significance"

G.V.Lakshmi

Associate Professor of Statistics
Dhruva Intitute of Engineering&Technology
Hyderabad, Telangana, India

ABSTRACT: *When comparing two treatments or characters we cannot rely on just the numerical differences. This is because each group is represented by only a sample of observations and if other samples were drawn the numerical values would change .This paper describes the comparison between two statistical tests of significance using numerical calculations.*

KEY WORDS: *t-test, Mean difference, SED, Paired t-test, Unpaired t-test Standard Error etc*

I.   OBJECTIVE:

Statistical science provides an objective procedure, called a test of significance, for distinguishing whether the observed difference suggests any real difference between the two groups or the difference could be due to chance.

The 't' test of significance is one such statistical procedure. The procedure to carry out the 't' test:
[1]   't' value = Mean Difference / SED
[2]   Mean Difference = Mean of Treat.1 - Mean of Treat.2

Each mean value is obtained by dividing the sum of all observations by the Number of observations. SED refers to Standard Error of Difference between two treatment mean Values. We use   SED because we are testing the difference between two Mean values.

3. SED = Standard Error x $\{\ddot{O}\ 2.\ r\}$.

Where r is the # observations on each of the two treatments.

Procedure to carryout t test
To calculate SED:
Calculate Correction Factor (CF) for each treatment
CF = Square of (Sum of observations) / r

Calculate Sum of Squares (SS) for each Treat.
$$SS = \text{Sum of (squares of individual observations)}$$

Calculate Standard Error (SE) for each Treat.
$$SE = \ddot{O}\ SS\text{-}CF\ \raisebox{-0.3ex}{\ }\ r\text{-}1$$
Finally Calculate SED

$$SED = SE\ x\ \ddot{O}\ 2.\ r$$
Where 'Standard Error' is computed as the sq root of pooled variance of the two treatments, assuming that the variance of individual treatments are statistically equal. Thus the values for the Mean Difference and SED are calculated to arrive at the calculated 'value. Finally, the calculated 't' value is compared with the theoretical value from a 't' Table at 5% or 1% probability level for the given sample size. Based on the comparison of calculated 'value with the theoretical 'value from the table, we conclude:

• If the calculated 't value is greater than the theoretical 't value, then the difference between the two

treatments is significant. This means the difference is not likely due to chance but more likely due to a real difference between the two treatments.

- If the calculated' value is less than the theoretical' value, then the difference between the two treatments is not significant. This means the observed difference is more likely due to chance and we conclude that the two treatments are not different.

An Important Caution: While applying the' t' test, we should identify whether the two sets of values are independent or interdependent. If the sets of values are independent, then we consider the values as unpaired and carry out the't' test.

If there is any correspondence (or interdependence) between the individual values in the two sets, then the values should be paired and difference taken between individual set of values and analyzed directly.

If we ignore this interdependency, we may overestimate the error of difference and underestimate the significance. Let us study this important statistical procedure with the help of examples.

Example of Unpaired't' tests

Example 1: The following yields were recorded when two sorghum cultivars were tested at ten locations. Determine whether a true difference exists in the performance of these two cultivars:

- Assuming the samples is independent.
- Assuming the samples could be paired.

(a) Let us carry out an unpaired t test assuming the samples are independent.

Location	CSH 5 YIELD	SPH 224 (100 kg/ha)
1	41.5	26.1
2	40.2	24.0
3	42.5	25.0
4	42.6	21.3
5	41.0	26.2
6	32.2	40.8
7	31.9	37.3
8	36.0	41.7
9	36.8	32.2
10	32.1	38.9
⟨⟩ Sum	376.8	313.5
Mean	37.68	31.35
CF	14197.80	9828.23
Sum Sq.	14375.40	10367.21
SE	4.44	7.74

Step 1 :

Get the sum of the yields for each cultivar and then the mean yields.

Step 2 :

Calculate Correction Factor(CF):
= {Sq(376.8)/10}; {Sq(313.5)/10}

Step 3 :

Square each value to calculate the Sum of Squares for each cultivar.

Step 4 :

Calculate Standard Error (SE) : = Sq.root{(Sum Sq - CF) / (Obs. - 1)}
= sq.root{(14375.40-14197.80) / (10-1)};
= Sq.root {(10367.21-9828.23) / (10-1)}

Location	CSH 5 YIELD	SPH 224 (100 kg/ha)
1	41.5	26.1
2	40.2	24.0
3	42.5	25.0
4	42.6	21.3
5	41.0	26.2
6	32.2	40.8
7	31.9	37.3
8	36.0	41.7
9	36.8	32.2
10	32.1	38.9
◇ Sum	376.8	313.5
Mean	37.68	31.35
CF	14197.80	9828.23
Sum Sq.	14375.40	10367.21
SE	4.44	7.74
Pooled SE	6.09	

Step 5 :

We have two Standard Errors. Calculate the pooled SE. Pooled SE = SE1 + SE2 / 2

Calculate Standard Error of Diff.
SED = SE {Sq.root(2/No. obs)}
SED = 6.09 {Sq.root(2/10)}
SED = 2.72

Step 7 :

Calculate the 't' value
't' = Mean Difference / SED 't' = (37.68-31.35) / 2.72
't' = 2.32

Step 8 :

Earlier we found that calculated 't' = 2.32.

Now compare Calculated 't' value with table 't' value for 18 df {(10-1)+(10-1)} at 5% probability (table 't' value = 2.10)

As 't'cal. is greater than 't'tab. We conclude that the difference in yield is significant.

Hence, CSH 5 yielded better than SPH 224.

(b) Now we will conduct a   paired 't' test assuming the samples are interdependent.

Location	CSH 5 YIELD	SPH 224 (100 kg/ha)	difference d
1	41.5	26.1	15.4
2	40.2	24.0	16.2
3	42.5	25.0	17.5
4	42.6	21.3	21.3
5	41.0	26.2	14.8
6	32.2	40.8	-8.6
7	31.9	37.3	-5.4
8	36.0	41.7	-5.7
9	36.8	32.2	4.6
10	32.1	38.9	-6.8
◇ Sum	63.3		
Mean	6.33		

1. 't' value = Mean Difference / SED

Mean Difference now refers to the mean of the differences between the individual pairs in the data. This Mean Difference is obtained by the sum of differences divided by the number of pairs of observations. Remember that the number of observations refers to the number of pairs of observations.

As we are testing only one mean the divisor will be SEM and not SED. SEM refers to Standard Error of a Mean, while SED refers to Standard. Error of Difference between TWO Mean values.

Example of Paired 't' test.

Location	CSH 5 YIELD	SPH 224 (100 kg/ha)	difference d
1	41.5	26.1	15.4
2	40.2	24.0	16.2
3	42.5	25.0	17.5
4	42.6	21.3	21.3
5	41.0	26.2	14.8
6	32.2	40.8	-8.6
7	31.9	37.3	-5.4
8	36.0	41.7	-5.7
9	36.8	32.2	4.6
10	32.1	38.9	-6.8
<> Sum	63.3		
Mean	6.33		

1. 't' value = Mean Difference / SED To calculate SEM:

a. Calculate correction factor (CF)
CF = sq (sum of diff.) / No. of obs.

b. Calculate the sum of squares(SS)
SS = Sum of(Sq of individual diff.)

c. Calculate the Standard Error(SE)
SE = sq.root{(SS-CF) / (No. of obs. - 1)} Ö SS-CF ¸ n-1

d. Finally calculate SEM
SEM = SE / {sq.root(No. of obs.)}

Location	CSH 5 YIELD	SPH 224 (100 kg/ha)	difference d
1	41.5	26.1	15.4
2	40.2	24.0	16.2
3	42.5	25.0	17.5
4	42.6	21.3	21.3
5	41.0	26.2	14.8
6	32.2	40.8	-8.6
7	31.9	37.3	-5.4
8	36.0	41.7	-5.7
9	36.8	32.2	4.6
10	32.1	38.9	-6.8
<> Sum	63.3		
Mean	6.33		

Step 1:

Calculate the yield difference at each location taking sign into account.

Step 2:

Calculate the sum of differences

Step 3:

Calculate the mean difference by diving the sum of differences by the number of pairs of observations.

Step 4:

Calculate the correction factor (CF).
= {Sq(63.3) / 10} = 400.69

Location	CSH 5 YIELD	SPH 224 (100 kg/ha)	difference d	Sq(diff.) sq(d)	
1	41.5	26.1	1.54	237.16	**Step 5:**
2	40.2	24.0	1.62	262.44	Calculate the squares of difference at each location and the sum of these squares of difference.
3	42.5	2.50	1.75	306.25	
4	42.6	21.3	2.13	453.69	**Step 6:**
5	41.0	26.2	1.48	219.04	Calculate the Standard Error (SE)
6	32.2	40.8	-0.86	73.96	
7	31.9	37.3	-0.54	29.16	=Sq root{(SS-CF)/(No. of Obs.-1)}
8	36.0	41.7	-0.57	32.49	=Sq root{(1681.59-400.69)/9}=1.193
9	36.8	32.2	0.46	21.16	
10	32.1	38.9	-0.68	46.24	**Step 7:**
Sum		63.3	1681.59		Calculate the Standard Error of Mean(SEM)
Mean		6.33			SE / {Sqroot(No. of Obs)}
CF		400.69			=11.93 / {Sqroot(10)} = 3.77
SE		11.93			
SEM		3.77			

Sum	63.3	1681.59
Mean	6.33	
CF	400.69	
SE	11.93	
SEM	3.77	

Step 9: Hence, we conclude that yield differences between CSH 5 and SPH are not really different

II.   CONCLUSION

In this paper I observed that there is no significant difference between two types of t-tests, after conducting various experiments using numerical calculations. Whether the sample is   Independent, or paired one can observe that the yield differences are approximately same. This concept has got many applications in Bio medical statistics to verify whether the drug is effective by adopting different types of t-tests.