

## Using Minitab for Paired Differences

On Minitab, you handle paired data by using the `let` to compute the differences, and then using the `tinterval` (to get a confidence interval for a mean) or `tttest` to test a hypothesis about a mean or (`sinterval` to get a confidence interval for the population median or `stest` to test a hypothesis about the population median. Here is the example from Andy Siegel in which chocolate desserts with and without nuts are being compared. Each participant tastes both versions, and rates them on a scale of 1 to 5. The data are paired by taster.

```
MTB > info
```

```
COLUMN      NAME      COUNT
C1          w/nuts      6
C2          w/o nuts     6
CONSTANTS USED: NONE
```

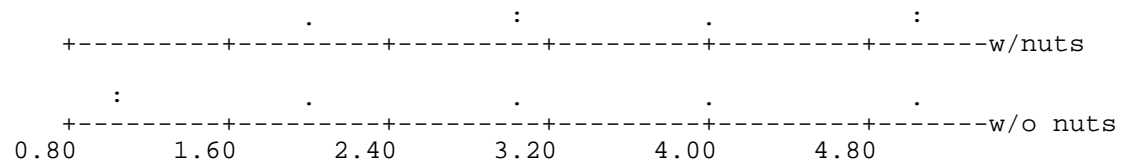
```
MTB > describe c1 c2
```

	N	MEAN	MEDIAN	TRMEAN	STDEV	SEMEAN
w/nuts	6	3.667	3.500	3.667	1.211	0.494
w/o nuts	6	2.667	2.500	2.667	1.633	0.667

	MIN	MAX	Q1	Q3
w/nuts	2.000	5.000	2.750	5.000
w/o nuts	1.000	5.000	1.000	4.250

```
MTB > dotplot c1 c2;
SUBC>same.
```



```
MTB > let c3=c1-c2
MTB > name c3 'diff.'
MTB > print c1-c3
```

ROW	w/nuts	w/o nuts	diff.
1	3	2	1
2	5	4	1
3	3	1	2
4	5	5	0
5	2	1	1
6	4	3	1

```

MTB > describe 'diff.'

          N      MEAN   MEDIAN   TRMEAN   STDEV   SEMEAN
diff.      6      1.000     1.000     1.000     0.632     0.258

          MIN      MAX      Q1      Q3
diff.      0.000     2.000     0.750     1.250

MTB > stem 'diff'
* ERROR * Undefined name or improper use of quote
* ERROR * 0 is an illegal number of arguments

MTB > stem 'diff.'

Stem-and-leaf of diff.      N = 6
Leaf Unit = 0.10

     1      0 0
    (4)      1 0000
     1      2 0

```

This output illustrates a disadvantage of addressing columns by name rather than by number. We tried to get a stem and leaf display of the differences, but left out the period. Minitab gave no display and pelted the user with error messages. If you address columns by name, you have to type the names *exactly* the same way every time. If you address the columns by number, you have to remember what data you put in which column. The choice is up to you. Since we can remember better than we can type, we usually address columns by number. Of course, one can always type `info` to see where things are, assuming things have been given names!

In the example above, we showed dotplots for each group, and also a single stem and leaf for the differences. In practice, we would only want the latter for paired data. Assuming the data really is paired, and that we have a random sample, the principal assumption is that the *differences* are more or less normally distributed. That certainly seems to be the case here. As always, we don't care too much whether it's mound-shaped, but we do want to watch for

1. long, fat tails
2. extreme skewness
3. or outliers.

If we see any signs of these, we need to reconsider using *t*. We could transform the data, use the bootstrap, or use an `sinterval` or `stest` for a median rather than a `tinterval` or `ttest` for a mean. Here we see no problems so we forge ahead.

```

MTB > tinterval 95% c3

          N      MEAN   STDEV   SE MEAN   95.0 PERCENT C.I.
diff.      6      1.000     0.632     0.258   ( 0.336, 1.664)

```

Since all the numbers in the confidence interval are positive, we can conclude that the nuts make a positive difference (or reject the hypothesis of no difference).

It is absolutely essential that you learn to distinguish between paired data and independent sample data. Below we have used the `twosamplet` command for independent samples on this same paired chocolate dessert data, WHICH WOULD **NOT** BE CORRECT! You can see that the results are nowhere near the correct confidence interval given by the `tinterval` command applied to the differences.

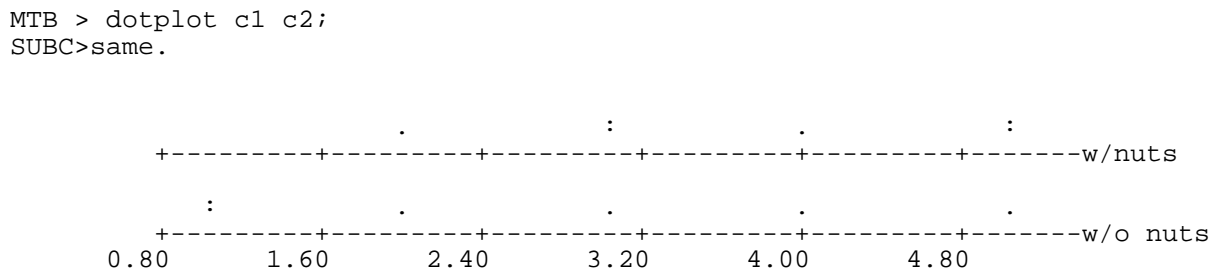
```
MTB > twosamplet 95% c1 c2

TWOSAMPLE T FOR w/nuts VS w/o nuts
      N      MEAN      STDEV      SE MEAN
w/nuts  6      3.67      1.21      0.49
w/o nuts 6      2.67      1.63      0.67

95 PCT CI FOR MU w/nuts - MU w/o nuts: (-0.88, 2.88)

TTEST MU w/nuts = MU w/o nuts (VS NE): T= 1.20  P=0.26  DF= 9
```

This procedure would be appropriate if you drew one sample of tasters to taste the nut-free dessert and a second, independently chosen sample of tasters to taste the version with nuts. We would check assumptions with separate displays for each group.



The principal assumption is that the **each group** is more or less normally distributed. There are no obvious signs of trouble here. As always, we don't care too much whether it's mound-shaped, but we do want to watch for

1. long, fat tails
2. extreme skewness
3. or outliers.