5.2. Paired T-tests (Related Samples)

There are many experimental situations where each subject receives both treatments. For example, each subject in the earlier example could have been measured in the absence of the drug (control value) and after having received the drug (treatment value). The response times for the control and treatment groups would no longer be independent.

Subject	Control Value	Treatment Value
1	90	95
2	87	92
3	100	104
4	80	89
5	95	101
6	90	105

```
DATA PAIRED;
   INPUT CTIME TTIME;
   DIFF=TTIME-CTI
ME; DATALINES;
90 95
87 92
100 104
80 89
95 101
90 105
;

PROC TTEST DATA=PAIRED;
VAR DIFF;
RUN;

PROC  TTEST  DATA=PAIRED;
PAIRED TTIME*CTIME;
RUN;

PROC MEANS DATA=PAIRED N MEAN STDERR T PRT;
   TITLE 'PAIRED T-TEST EXAMPLE';
   VAR DIFF;
RUN;
```

*If we have more than one set of paired values, we can enter several pairs on one PAIRED statement. For example, if you have before1,after1,before 2, after 2 and want to compare before to after for each, you could write: paired before1*after1 before2*after2;

Output:

PAIRED T-TEST EXAMPLE
The TTEST Procedure
Difference: TTIME - CTIME

N	Mean	Std Dev	Std Err	Minimum	Maximum
6	7.3333	4.1312	1.6865	4.0000	15.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
7.3333	2.9979	11.6687	4.1312	2.5787	10.1322

DF	t Value	Pr > |t|
5	4.35	0.0074

In this example, the mean difference is 7.333, p value is 0.0074. We can state that response times are longer under the drug treatment compared to the control values.

***** After taking the paired differences – make sure you use one variable to subtract the other, consistently – the two population/sample problem is then reduced to a one population/one sample problem: on the population/sample of the paired differences**

$$\begin{cases} H_0: \mu_{male} = \mu_{female} \\ H_a: \mu_{male} > \mu_{female} \end{cases} \Leftrightarrow \begin{cases} H_0: \mu_{diff} = 0 \\ H_a: \mu_{diff} > 0 \end{cases}$$

$$\begin{cases} H_0: \mu_{male} = \mu_{female} \\ H_a: \mu_{male} < \mu_{female} \end{cases} \Leftrightarrow \begin{cases} H_0: \mu_{diff} = 0 \\ H_a: \mu_{diff} < 0 \end{cases}$$

$$\begin{cases} H_0: \mu_{male} = \mu_{female} \\ H_a: \mu_{male} \neq \mu_{female} \end{cases} \Leftrightarrow \begin{cases} H_0: \mu_{diff} = 0 \\ H_a: \mu_{diff} \neq 0 \end{cases}$$

SAS code:

```
data twins;
input IQmale IQfemale;
diff= IQmale-IQfemale;
datalines;
200 199
150 120
...
109 112
;
run;
```

```
proc univariate data=twins normal;
var diff;
run;
```

Example 1. To study the effectiveness of wall insulation in saving energy for home heating, the energy consumption (in MWh) for 5 houses in Bristol, England, was recorded for two winters; the first winter was before insulation and the second winter was after insulation:

House	1	2	3	4	5
Before	12.1	10.6	13.4	13.8	15.5
After	12.0	11.0	14.1	11.2	15.3

(a) Please provide a 95% confidence interval for the difference between the mean energy consumption before and after the wall insulation is installed. What assumptions are necessary for your inference?
(b) Can you conclude that there is a difference in mean energy consumption before and after the wall insulation is installed at the significance level 0.05? Please test it and evaluate the p-value of your test. What assumptions are necessary for your inference?
(c) Please write the SAS program to perform the test and examine the necessary assumptions given in (b).

SOLUTION: This is inference on two population means, paired samples.

(a). $\bar{d} = 0.36$, $S_d = 1.30$

CI: $0.36 \pm 2.776 \cdot \frac{1.30}{\sqrt{5}} = (-1.25, 1.97)$

(b). $H_0 : \mu_d = 0$, $H_a : \mu_d \neq 0$

(1) $t_0 = \frac{\bar{d}-0}{s_d/\sqrt{n}} = \frac{0.36-0}{1.30/\sqrt{5}} \approx 0.619$

$t_{n-1,\alpha/2} = t_{4,0.025} = 2.776$

Since $|t_0| \approx 0.619$ is smaller than $t_{4,0.025} = 2.776$, we cannot reject H_0.

(2) $p-value = 2 \cdot P(T \geq 0.619) \approx 0.57$

Assumptions for (a) and (b): the paired differences follow a normal distribution.

```
#Example 1 (c), in R #
Before=c(12.1,10.6,13.4,13.8,15.5)
After=c(12.0,11.0,14.1,11.2,15.3)
Diff=Before-After
shapiro.test(Diff)
#We can barely conclude Diff follows normal distribution since p=0.08496 is small but still >
0.05, while < 0.10 #
t.test(Diff)
t.test(Before, After, paired=TRUE)

#Wilcoxon Signed-Rank Test, based on the original data#
wilcox.test(Before, After, paired = TRUE)
wilcox.test(Before-After)
wilcox.test(Diff)
```

(c) The SAS program is:

```
Data energy;
Input before after @@;
Diff=before - after;
Datalines;
12.1 12.0 10.6 11.0 13.4 14.1 13.8 11.2 15.5 15.3
;
Run;
Proc univariate data = energy normal;
Var Diff;
Run;

*PROC TTEST DATA=energy;
*VAR DIFF;
*RUN;

*PROC TTEST DATA=energy; PAIRED before*after;
*RUN;

*PROC MEANS DATA=energy N MEAN STDERR T PRT; TITLE 'PAIRED T-TEST EXAMPLE';
*VAR DIFF;
*RUN;
```

Note 1: As you can see from the output below, the variable "Diff" is barely normal. In this situation, one can either use the t-test or the signed-rank test(*although not the best practice, but $\alpha = 0.05$ can be used as the significance level to judge the normality; although $\alpha = 0.1$ will be more secure.)

Note 2: Although several SAS procedures can generate the t-test, I would recommend you to master the proc univariate first because it has everything you will need: normality test, t-test and non-parametric tests.

SAS Output (selected):

```
                    The UNIVARIATE Procedure
                       Variable:  Diff


                   Tests for Location: Mu0=0

         Test           -Statistic-    -----p Value------

         Student's t    t  0.616851    Pr > |t|    0.5707
         Sign           M       0.5    Pr >= |M|   1.0000
         Signed Rank    S       0.5    Pr >= |S|   1.0000


                      Tests for Normality

  Test                  --Statistic---    -----p Value------

  Shapiro-Wilk          W      0.80253    Pr < W      0.0850
  Kolmogorov-Smirnov    D     0.348791    Pr > D      0.0442
  Cramer-von Mises      W-Sq   0.10017    Pr > W-Sq   0.0859
  nderson-Darling      A-Sq  0.551915   Pr > A-Sq   0.0779
```