PharmaSUG 2013 - Paper CC28

How MEAN is T-test?

Naina Pandurangi, Inventiv Health Clinical, Mumbai, India

ABSTRACT

Statistical analysis is a blend of statistics and programming. Often the written description of the required analysis that is framed by the statistician may sound tricky to program and may make us fetch for the STAT procedures. But sometimes the underlying requirement can be easily derived using the regular basic non-STAT i.e. non-statistical SAS® procedures! One such case is where the t-test analysis is to be performed on one sample with the null hypothesis of population mean=0.

This paper is going to explicate some steps on the bridge between the STAT and non-STAT SAS procedures with the help of two situations where the requirement demands one sample t-test analysis and can be met using PROC MEANS.

INTRODUCTION

Analysis of statistics in a single sample is a vital section in clinical trials across all the phases. The pattern and behaviour of the drug is gauged by observing the results of the study parameter from one point in time to another. In other words, the study of change in the results is an integral part of such one sample analysis. When this analysis is based on t-test where one may want to study the confidence intervals about the mean change or compare the results across two time points as a pair, the procedure to be used to obtain the essential statistics need not be PROC TTEST always. If we read the given specification carefully and interpret after breaking it down we would learn that PROC MEANS can help us achieve our objective.

Let us see this with the help of the following examples.

EXAMPLE 1: CONFIDENCE INTERVAL (CI) OF MEAN CHANGE FROM BASELINE BASED ON STUDENT'S T-TEST.

Let us take a simple one sample data where we want to analyse the change in observations based on the Student's t-test. Consider the following dataset called DATA RES, where RES, BASE and CHG are the variables respectively showing the original results, the baseline and the change from baseline for each subject.

	SUBJECT	RES	BASE	CHG
1	1	70	40	30
2	2	65	39	26
3	3	71	36	35
4	4	100	38	62
5	5	81	44	37
6	6	88	34	54
7	7	101	50	51
8	8	92	46	46

Table 1: DATA RES dataset containing the subject wise original results, baseline, and change from baseline.

Our requirement is to find the CI of mean change based on t-test. Let us find this using the PROC TTEST first and then the PROC MEANS in SAS and compare the outputs.

METHOD 1: USING PROC TTEST

```
proc ttest data=res;
var chg;
run;
```

The SAS System

The TTEST Procedure

Variable: CHG

N	Mean	Std Dev	Std Err	Minimum	Maximum
8	42.6250	12.6031	4.4559	26.0000	62.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
42.6250	32.0885	53.1615	12.6031	8.3329	25.6508

DF	t Value	Pr > |t|
7	9.57	<.0001

OUTPUT 1.1: Output from PROC TTEST showing the t-test statistics for mean change

METHOD 2: USING PROC MEANS

```
proc means data=res lclm uclm;
var chg;
run;
```

The SAS System

The MEANS Procedure

Analysis Variable : CHG

Lower 95% CL for Mean	Upper 95% CL for Mean
32.0885064	53.1614936

OUTPUT 1.2: Output from PROC MEANS showing the 95% confidence limits for mean change

It is clear from the two outputs, OUTPUT1.1 and OUTPUT1.2, that the CI limits produced by both the procedures are identical.

Note: We have requested only the CI from MEANS procedure in the above example but other statistics seen in the t-test output also match the ones produced from MEANS procedure which can be seen when requested.

EXAMPLE 2: CI, T-STATISTIC AND P-VALUE OF WITHIN GROUP MEAN CHANGE USING PAIRED T-TEST.

By within group here we mean within individual group that is nothing but within one sample. Let us take a case where we need to analyse the results observed at two points of time as a pair. The following dataset DATA RES2 contains the results collected at two different time points, RES1 and RES2.

	SUBJECT	RES1	RES2
1	1	100	41
2	2	77	34
3	3	68	40
4	4	66	35
5	5	92	34
6	6	78	44
7	7	111	40
8	8	102	49
9	9	66	38
10	10	76	59

Table 2: DATA RES2 dataset containing independant results, RES1 and RES2, at two distinct time points for each subject.

Let us derive the CI, t-statistic and the p-value for within group mean change.

METHOD 1: USING PROC TTEST

```
proc ttest data=res2;
paired res1*res2;
run;
```

The SAS System

The TTEST Procedure

Difference: RES1 - RES2

N	Mean	Std Dev	Std Err	Minimum	Maximum
10	42.2000	17.3513	5.4870	17.0000	71.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
42.2000	29.7876	54.6124	17.3513	11.9348	31.6767

DF	t Value	Pr > |t|
9	7.69	<.0001

OUTPUT 2.1: Output from PROC TTEST showing the t-test statistics for difference in the two results

METHOD 2: USING PROC MEANS

As seen in the t-test output, OUTPUT2.1, the analysis is done on the difference of the variables mentioned in the PAIRED statement i.e. RES1 – RES2 in our case. So we need to create a similar variable first, say DIFF = RES1 – RES2 and then pass it to the MEANS procedure.

```
proc means data=res2 lclm uclm t prt;
var diff;
run;
```

The SAS System

The MEANS Procedure

Analysis Variable : DIFF

Lower 95% CL for Mean	Upper 95% CL for Mean	t Value	Pr > |t|
29.7876472	54.6123528	7.69	<.0001

OUTPUT 2.2: Output from PROC MEANS showing the 95% confidence limits, the t-test statistic and p-value for mean difference in the two results

As seen from the two outputs, OUTPUT2.1 and OUTPUT2.2, the statistics derived using both the methods are same.

Note: In paired t-test analysis the pair under analysis consists of independent results observed at two different time points, so the case considered in the first example can also be looked at as a pair i.e. a pair of baseline and post baseline results. So we can say that the first example is also a paired t-test analysis on the difference RES – BASE.

CONCLUSION

The relation between PROC MEANS and PROC TTEST is seen in a single sample case because of the same underlying null hypothesis - Ho: population mean is zero. So while working on one sample statistics around the mean, either procedure can be used to generate the statistics such as CI of the mean, t-statistic, t-test p-value because the statistical theories used in both procedures are the same.

ACKNOWLEDGMENTS

I thank all my colleagues who have helped me in understanding these statistical concepts and reviewed and provided useful comments on my paper. I am grateful to Dr. Prashant Kirkire, country manager, Inventiv Health Clinical, India, whose guidance and encouragement has helped me while writing this paper, Shuhong Zhao, Ph.D., Principal Biostatistician, Inventiv Health Clinical, who has provided valuable suggestions and inputs to my paper and my colleague Sandeep Sawant who has truly inspired me to make this happen.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Naina Pandurangi
Enterprise: Inventiv Health Clinical
Address: Marwah Centre , Ground Floor, B Wing, Krishnalal Marwah Marg,
Andheri (E) – 400 072.
City, State ZIP: Mumbai, MH, India
Work Phone: +91 9833915538
E-mail: nainap9@gmail.com