

## **Calculating the point estimate and confidence interval of Hodges-Lehmann's median using SAS<sup>®</sup> software**

Lingling Han, Merck & Co., Inc., North Wales, PA

### **ABSTRACT**

Nonparametric analysis methods are sometimes used to detect treatment differences of Pharmacokinetic variables in clinical trials. For example, Hodges-Lehmann's median analysis is recommended to calculate the treatment differences for  $T_{\max}$ . However, there is no existed SAS procedures can directly produce the Hodges-Lehmann's median estimate. This paper describes the procedure for calculating the point estimate and confidence interval of the Hodges-Lehmann's median differences for both independent and paired groups. SAS MACROs are provided for readers to use.

### **INTRODUCTION**

In clinical trials, Hodges-Lehmann's median methods are sometimes used to assess the differences between treatment groups. For example,  $T_{\max}$ , the time associated with the maximum plasma drug concentration, is recommended to use nonparametric confidence interval method by the European guidance on bioavailability and bioequivalence [1].

The point estimate of the median and the confidence interval of the median were developed by Hodges and Lehmann [2] and are described in detail in Lehmann [3]. The methodology to obtain point estimate and confidence interval for the median for both paired and independent two samples is provided below. However, there is no procedures currently exist in SAS software to produce the point estimate and the corresponding confidence interval.

Chris [4] used SQL procedure to calculate the estimate for the treatment difference for two independent groups. However, it is common to have paired groups in clinical trials, for example, the data from the crossover trials.

This paper modified Chris's code for the independent groups, and provided a point estimate and confidence interval for treatment difference from paired groups. SAS MACROs are provided for both methods, and ready for readers to use.

### **METHODS**

**a) Procedure for calculating the Hodges-Lehmann's median difference (paired groups)**

Assume that the data consist of  $N$  paired observations  $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ , where  $X$  and  $Y$  are correlated random variables, usually through a matched-pair design, for example, crossover designs. Define the random variable  $D = X - Y$  and let  $d_i = x_i - y_i$ ,  $i=1, 2, \dots, N$ , denote the  $N$  observed differences.

### 1. Point Estimate of the Median difference

The point estimate of the Hodges-Lehmann's median difference can be obtained as follows:

- Form all possible ordered pairs of differences  $(d_i, d_j)$  with  $i \leq j$ . There are  $N(N+1)/2$  such ordered pairs.
- For each of the above ordered pairs, compute the average value  $(d_i + d_j)/2$ .
- The point estimate of the median difference can be obtained by  $\text{med}_{i \leq j} [(d_i + d_j)/2]$ .

### 2. Confidence Interval for the Median difference

Let  $[\lambda^*, \lambda^*]$  be the exact  $100 \times (1-\alpha) \%$  confidence interval for the median difference. Let  $M = N(N+1)/2$  and let  $A_{[1]} \leq A_{[2]} \leq \dots \leq A_{[M]}$  be the  $M$  averages,  $(d_i + d_j)/2$  for all  $i \leq j$ , sorted in ascending order.

$\lambda_*$  and  $\lambda^*$  can be obtained based on the Wilcoxon Signed-Rank distribution. An asymptotic confidence interval can be computed by applying the normal approximation to the Wilcoxon Signed-Rank distribution. For example, to obtain the lower confidence bound  $\lambda_*$ , we use the following procedure.

- Find  $t_*$  such that  $\Phi \left[ \frac{t_* - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}} \right] = \alpha/2$ ,

- Set  $\lambda_* = A_{[i]}$ , where  $i = 1 + \lfloor t_* \rfloor$ , and  $\lfloor t_* \rfloor$  rounds  $t_*$  down to the nearest integer.

To obtain the upper confidence bound  $\lambda^*$ , a complimentary procedure is used.

- Find  $t^*$  such that  $1 - \Phi \left[ \frac{t^* - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}} \right] = \alpha/2$

- Set  $\lambda^* = A_{[i]}$ , where  $i = \lceil t^* \rceil$ , and  $\lceil t^* \rceil$  rounds  $t^*$  up to the nearest integer.

### **b) Procedure for calculating the Hodges-Lehmann's median difference (independent groups)**

Readers may refer Chris [4] for the methods for the independent groups.

## STEP-BY-STEP SAS CODE

### 1. Data step

It is assumed that the data are contained in one data set and are structured as two observations for each subject from the two corresponding treatments, say, *treatA* and *treatB*. So, in the code below, we use data *DSN* as an example, which has three variables, *subject*, *treatA*, *treatB*, where *treatA* and *treatB* are the response from the two treatments. If the data, let's call it *TEMP*, has three variables as *subject*, *treat*, and *response*, with *treat* has two values "A" and "B". We can use the following code to obtain a data similar to *DSN*.

```
data temp1(rename=(response=treatA)); set temp;
if treat="A";
run;
data temp2(rename=(response=treatB)); set temp;
if treat="B";
run;
data DSN; merge temp1 temp2;
drop treat;
run;
```

The first step is to calculate the differences between the two groups for each subject and record the sample size as a macro variable for later use.

```
data one; set dsn;
call sumput('size', trim(left(_N_)));
diff=treatA-treatB;
run;
```

### 2. Using ARRAY to form all possible ordered pairs of differences and average values.

The next step is to form all possible ordered pairs of differences and their average values. We can first transform the data then use ARRAY statement as described below:

```
proc transpose data=one out=two(drop=_name_);
var diff;
run;
data three; set two;
array x{&size} col1-col&size;
do i=1 to &size;
  do j= i to &size;
    stat=(x{i}+x{j})/2;
    output;
  end;
end;
```

### 3. Calculate the point estimate of the Hodges-Lehmann median difference

```

proc sort data=three;
by stat;
run;
proc means median noprint;
var stat;
output out=est(drop=_freq_ _type_) median=estimate;
run;

```

#### 4. Calculate the confidence interval of the Hodges-Lehmann median difference

```

data four;
  loword=1+int(&size*(&size+1)/4+probit(&alpha/2)*
             sqrt(&size*(&size+1)*(2*&size+1)/24));
  uppord= ceil( &size*(&size+1)/4+probit(1-&alpha/2)
             *sqrt(&size*(&size+1)*(2*&size+1)/24 ));
run;
data five; set three end=last;
if _N_=1 then set four;
retain lower upper;
  if _N_=loword then lower=stat;
  if _N_=uppord then upper=stat;
  if last then output;
keep lower upper;
run;

```

#### 5. Put the point estimate and confidence into one dataset

```

data HL_est; merge est five;
run;

```

## SAS MACRO CODES

In this section, we put the codes together into SAS MACROS, and readers can invoke the macro to get the results easily.

SAS Macro to calculate the point estimate and confidence interval of the Hodges-Lehmann's median (paired):

We assume data DSN has three variables: *subject*, *treatA*, and *treatB*, where *treatA* and *treatB* are the response for the two treatments.

```

%macro HL_paired(dsn=, trt1=, trt2=, alpha=);
data one; set &dsn;
call symput('size',trim(left(_N_)));
diff=&trt1-&trt2;
run;
proc transpose data=one out=two(drop=_name_);
var diff;
run;

data three; set two;
array x{&size} col1-col&size;

```

```

do i=1 to &size;
  do j= i to &size;
    stat=(x{i}+x{j})/2;
    output;
  end;
end;
proc sort data=three;
by stat;
run;
proc means median noprint;
var stat;
output out=est(drop=_freq_ _type_) median=estimate;
run;

data four;
  loword=1+int(&size*(&size+1)/4 + probit(&alpha/2)
    *sqrt(&size*(&size+1)*(2*&size+1)/24));
  uppord= ceil(&size*(&size+1)/4 + probit(1-&alpha/2)
    *sqrt(&size*(&size+1)*(2*&size+1)/24));
run;
data five; set three end=last;
if _N_=1 then set four;
retain lower upper;
  if _N_=loword then lower=stat;
  if _N_=uppord then upper=stat;
  if last then output;
  keep lower upper;
run;
data HL_est; merge est five;
run;
%mend HL_paired;

```

SAS Macro to calculate the point estimate and confidence interval of the Hodges-Lehman's median (independent):

We assume data DSN has three variables: *subject*, *treat*, and *response*, where *treat* has two values.

```

%macro HL_indep(dsn=, trt1=, trt2=, alpha=);
data treatA(rename=(response=resptra)); set &dsn;
if treat="&trt1";
run;
data treatB(rename=(response=resptra)); set &dsn;
if treat="&trt2";
run;

data treatA; set treatA;
call symput('size1',trim(left(_N_)));
run;
data treatB; set treatB;
call symput('size2',trim(left(_N_)));
run;

proc SQL;
create table all as

```

```

select treatA.resptrtA, treatB.resptrtB,
       (1) as merge
       from treatA, treatB
       quit;

data all; set all;
diff=resptrtA-resptrtB;
proc sort data=all;
by diff;
run;

Proc means data=all noprint;
var diff;
output out=est(drop=_freq_ _type_) median=estimate;
run;

data four;
  loword=&size1*&size2-ceil(probit(1-&alpha/2)
    *sqrt(&size1*&size2*(&size1+&size2+1)/12)+&size1*&size2/2);
  uppord=&size1*&size2-int(probit(&alpha/2)
    *sqrt(&size1*&size2*(&size1+&size2+1)/12)+&size1*&size2/2);
run;
data five; set all end=last;
if _N_=1 then set four;
retain lower upper;
  if _N_=loword then lower=diff;
  if _N_=uppord then upper=diff;
  if last then output;
  keep lower upper;
run;
data HL_est; merge est five;
proc print;
run;
%mend HL_indep;

```

## CONCLUSION

With several DATA steps, SAS PROC procedures, and ARRAY statement, this paper provided SAS MACROS to calculate point estimate and confidence interval for the Hodges-Lehmann's median differences. StatXact 7 Procs for SAS<sup>®</sup> users has a procedure available to calculate the point estimate and confidence interval for the Hodges-Lehmann's median differences. However, not everyone is familiar with this software, and SAS<sup>®</sup> is more popular software. Before a build-in SAS procedure to calculate the Hodges-Lehmann's median is available for us, it is very helpful and handy to have the macros provided in this paper. One can easily invoke the macros provided in this paper to obtain the point estimate and confidence interval for the Hodges-Lehmann's median difference.

## REFERENCES

1. EMEA/EUDRA (2001). Note for guidance on the investigation of bioavailability and bioequivalence.

2. Hodges, J. L and Lehmann E. L. (1963). *Estimates of location based on rank tests*, The Annals of Mathematical Statistics, 34: 598:611.
3. Lehmann E. L. (1975). *Nonparametrics: Statistical methods based on ranks*, Holden-Day, San Francisco.
4. Decker, C. *Calculating a nonparametric estimate and confidence interval using SAS<sup>®</sup> software*,
5. Cytel Software Corporation. (2005), *StatXact 7 Procs for SAS 9.x users*, user manual, Cambridge: Cytel Software Corporation.

## **TRADEMARKS**

SAS<sup>®</sup> and all other SAS institute Inc. product or service names are registered trademarks or trader SAS Institute Inc. in the USA and other countries.

## **CONTACT INFORMATION**

Your comments and questions are valued and encouraged, please contact the author at:

Lingling Han  
Merck & Co., Inc.  
351 N. Sumneytown Pike  
UG1D-44  
North Wales, PA  
19454  
Phone: 267-305-5506  
Email: [lingling\\_han@merck.com](mailto:lingling_han@merck.com)