General AUC calculated based on the trapezoidal rule

ABSTRACT

Generally, the trapezoidal is used to calculate the area under PK curve. Because the PK test data value is actual data, then all observation values should be positive. In practical application, we need to calculate the AUC using the derived data which include the positive values and negative values. For example, the change from baseline value is derived from the baseline value and the observation value, which have different sign. Therefore, this kind of data will not be directly used the trapezoidal rule to calculate the AUC. In this paper, we introduce a method that the adjacent original observations with different sign generate new dummy observations by the geometric triangle theory, to calculate the AUC use the original observations and the dummy observations by the trapezoidal rule.

Keywords

TRAPEZOIDAL RULE, DUMMY OBSERVATION, AUC

INTRODUCTION

The trapezoidal rule is widely used to calculate the area under the PK curve. The basic principle is that observations segment entire integration interval into multiple sub-intervals. Each sub-interval will form a closed area. The trapezoidal rule formula is as follow.
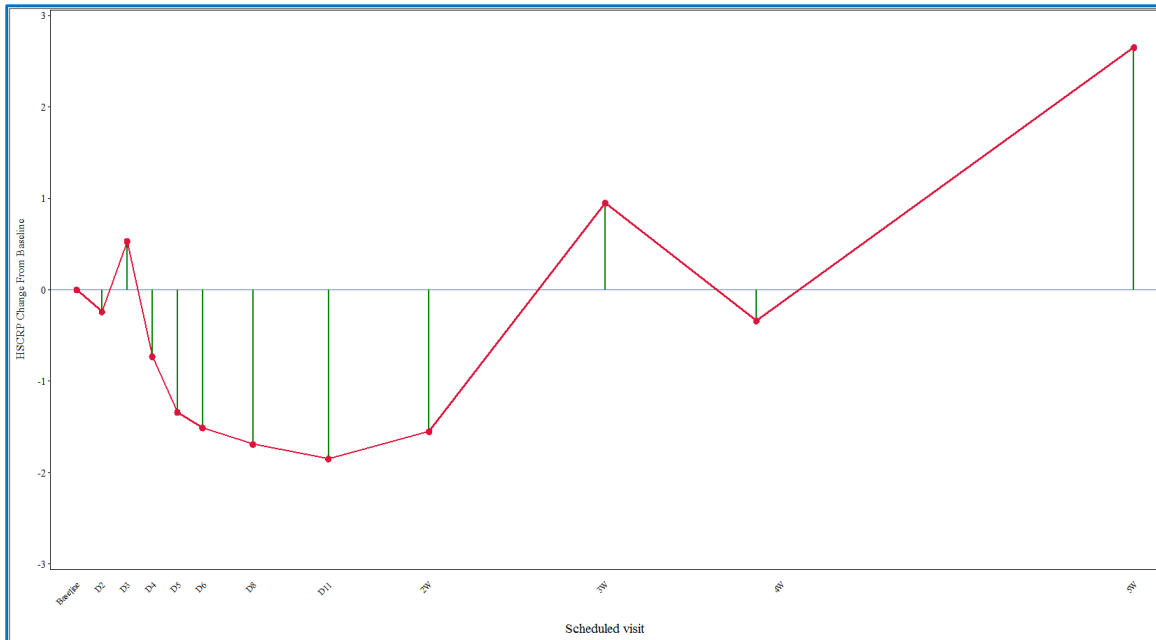
$$AUC_{(t_i - t_{i-1})} = (t_i - t_{i-1}) \times \frac{f(t_i) + f(t_{i-1})}{2}$$

The sub-AUC value is summated as the AUC of entire interval

$$AUC_{(b-a)} = \sum_{t=a+1}^{b} AUC_{(t_i - t_{i-1})}$$
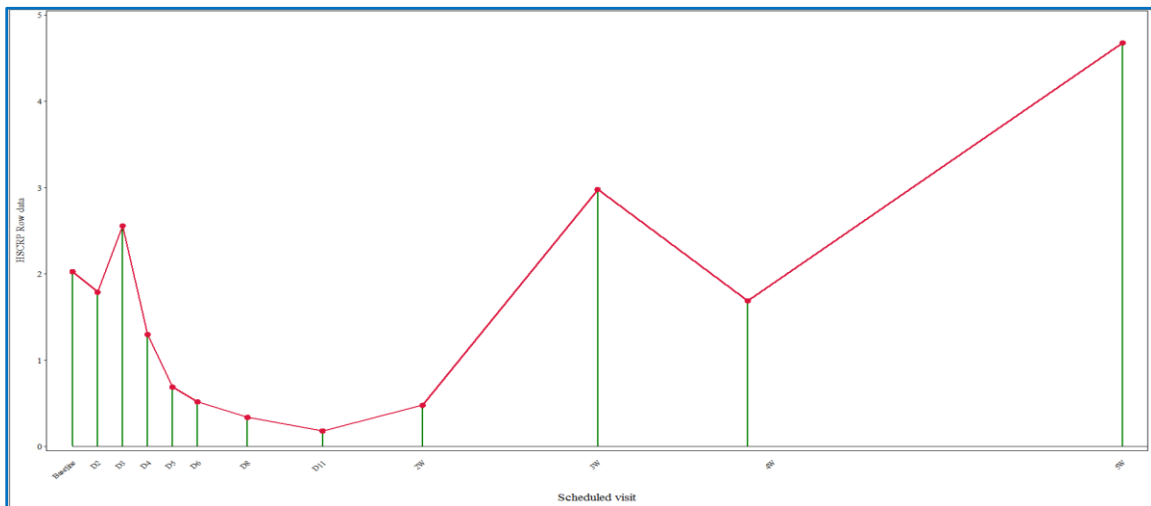
The limitation of this method is that all of observation should be same sign. If the sign of two adjacent observations are opposite. The AUC result of directly using trapezoidal rule will be wrong.

If we take the absolute value of each observation to calculate the sub-intervals AUC, The result is the half of rectangle areas which side length is the length of between adjacent observation and the value of adjacent observation.
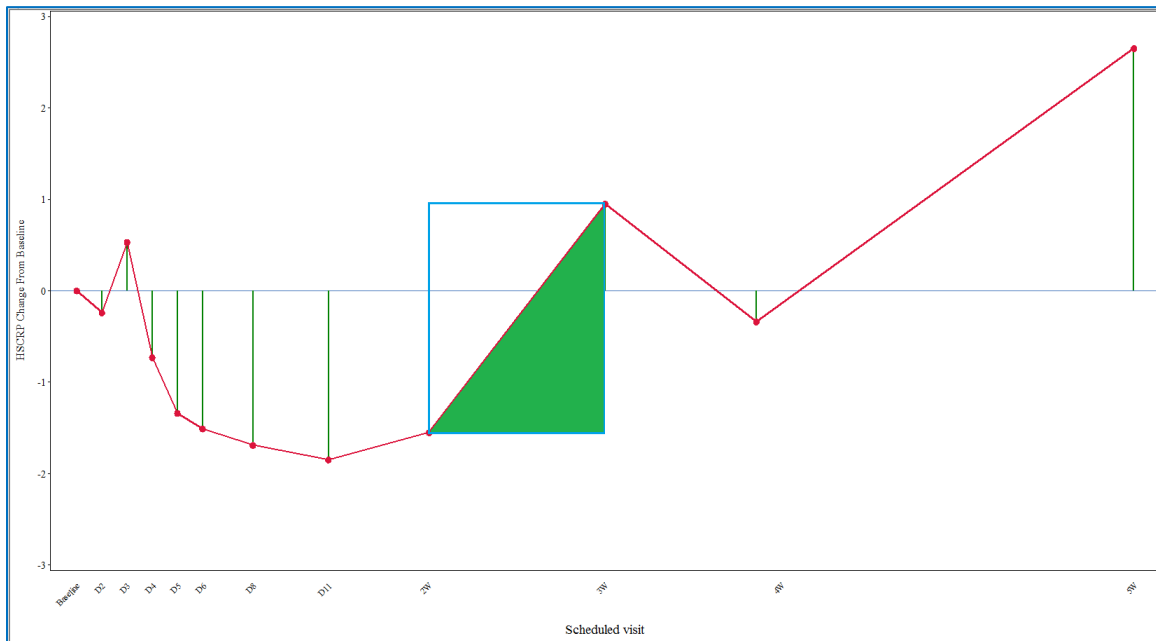
Thus, the trapezoidal rule will not be valid in this case.

To take the trapezoidal rule in the AUC calculation, firstly, you must divided whole interval into multiple closed sub-intervals .secondly, the line of between adjacent observations must only form one closed area. Because the line of adjacent observation and the reference line form two right triangles. In fact, the sub-interval AUC should be the area of two right triangles. For this reason, we can get the sub-AUC through calculate the sum of the two triangle areas use the triangle area formula

$$AUC_{(t_i - t_{i-1})} = S_{t_i} + S_{t_{i-1}}$$

$$S = \frac{1}{2} ah$$

We can get two heights from the original data, The key of this formula is how can get the length of lower side of the triangle. Assume the original data contains M observations and form N intersection with reference line. One of the intersection is set to t_x.

Through the similar triangle theorem,

$$\frac{t_x - t_{i-1}}{t_i - t_x} = \frac{|f(t_{i-1})|}{|f(t_i)|}$$

Derivation formula t_x which is the dummy observation

$$t_x = \frac{t_{i-1} \times |f(t_i)| + t_i \times |f(t_{i-1})|}{|f(t_{i-1})| + |f(t_i)|}$$

By the above formula, we can derive all dummy observation by the original observation value. The whole interval is split become M+N-1 sub-interval by the original observation and the dummy observation, at the same time, the line between the observation form M+N-1 closed area with reference line. The AUC is the sum of area of M+N-1 sub-interval.
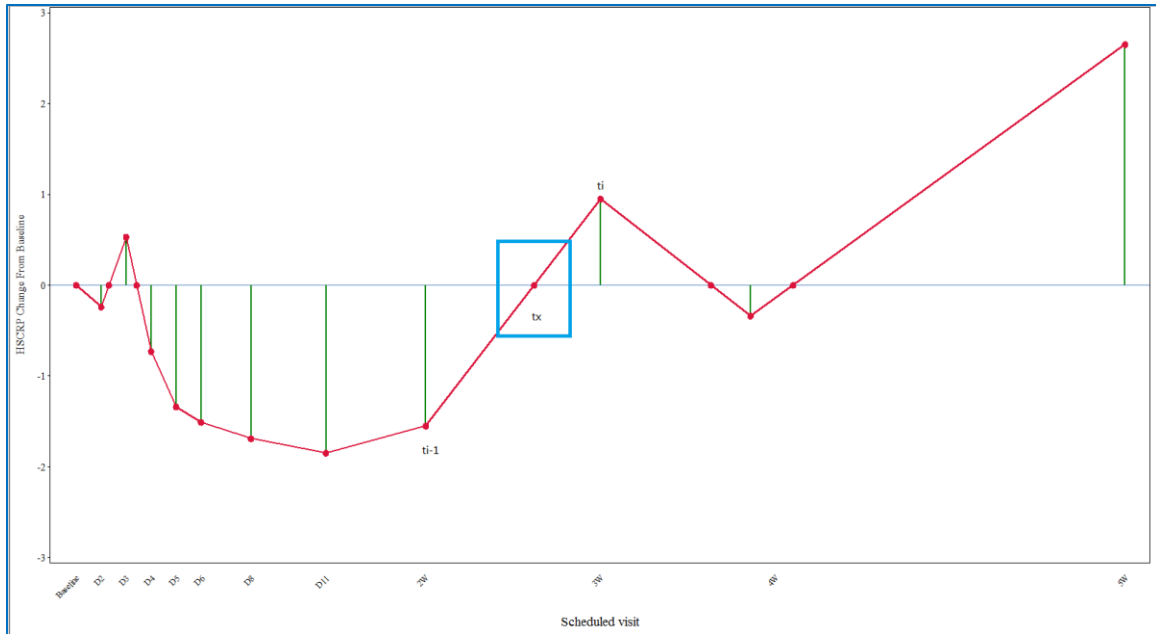
By introducing dummy observation, the adjacent observation only form one closed area, so it can be directly calculated area of each sub-interval according to the trapezoidal rule.

SAS MACRO

Input the original data, setting $f(t) = lbblchg$, $t = lbandy$.

	subjid	lbtestcd	lbstresn	lbblres	lbblchg	lbanldy
1	001	HSCRP	2.03	0	0	1
2	001	HSCRP	1.79	2.03	-0.24	2
3	001	HSCRP	2.56	2.03	0.53	3
4	001	HSCRP	1.3	2.03	-0.73	4
5	001	HSCRP	0.692	2.03	-1.338	5
6	001	HSCRP	0.52	2.03	-1.51	6
7	001	HSCRP	0.34	2.03	-1.69	8
8	001	HSCRP	0.182	2.03	-1.848	11
9	001	HSCRP	0.479	2.03	-1.551	15
10	001	HSCRP	2.98	2.03	0.95	22
11	001	HSCRP	1.69	2.03	-0.34	28
12	001	HSCRP	4.68	2.03	2.65	43

For calculated the final t_x, we must have the absolute value of adjacent observations and the value of previous adjacent observation.

1. Derivation four variable $|f(t_i)|$, $|f(t_{i-1})|$, $f(t_{i-1})$,t_{i-1}.

```
%macro deriveValue(dsin= lab_1
        ,dsout=lab_2
        ,var=lbblchg
        ,day=lbanldy
        );

proc sort data=&dsin out=data_1;
by subjid lbtestcd &day;
run;

data data_2;
set data_1;
by subjid lbtestcd &day;

    __day=&day;
    __value=&var;

    __pday=lag(__day);
    __pvalue=lag(__value);

    __abspvalue=abs(__pvalue);
    __absvalue=abs(__value);

    if first.lbtestcd then do;
        __pday=.;
        __pvalue=.;
    end;
run;
```

2. Using SAS LAG function get the previous value, $|f(t_{i-1})| = \_abspvalue$, $f(t_{i-1}) = \_pvalue$, $t_{i-1} = \_pday$ and derive the absolute value by current value $|f(t_i)| = \_absvalue$,

	subjid	lbtestcd	lbstresn	lbblres	lbblchg	lbanldy	__day	__value	__pday	__pvalue	__abspvalue	__absvalue
1	001	HSCRP	2.03	0	0	1	1	0	.	.	.	0
2	001	HSCRP	1.79	2.03	-0.24	2	2	-0.24	1	0	0	0.24
3	001	HSCRP	2.56	2.03	0.53	3	3	0.53	2	-0.24	0.24	0.53
4	001	HSCRP	1.3	2.03	-0.73	4	4	-0.73	3	0.53	0.53	0.73
5	001	HSCRP	0.692	2.03	-1.338	5	5	-1.338	4	-0.73	0.73	1.338
6	001	HSCRP	0.52	2.03	-1.51	6	6	-1.51	5	-1.338	1.338	1.51
7	001	HSCRP	0.34	2.03	-1.69	8	8	-1.69	6	-1.51	1.51	1.69
8	001	HSCRP	0.182	2.03	-1.848	11	11	-1.848	8	-1.69	1.69	1.848
9	001	HSCRP	0.479	2.03	-1.551	15	15	-1.551	11	-1.848	1.848	1.551
10	001	HSCRP	2.98	2.03	0.95	22	22	0.95	15	-1.551	1.551	0.95
11	001	HSCRP	1.69	2.03	-0.34	28	28	-0.34	22	0.95	0.95	0.34
12	001	HSCRP	4.68	2.03	2.65	43	43	2.65	28	-0.34	0.34	2.65

3. Through comparing $f(t_{i-1})$ and $f(t_i)$ whether have same sign to decide whether to derive the dummy observations。The dummy point t_x is calculated by the formula.

```
data data_3(drop=__:);
set data_2;
by subjid lbtestcd &day;

   DVALUE=__VALUE;
   DDAY=__DAY;
   output;

   if ((__value <0 and __pvalue>0) or (__value >0 and __pvalue<0)) and __pvalue ne . then do;

      __day=(__day*__abspvalue+__pday*__absvalue)/(__abspvalue+__absvalue);

   DVALUE=0;
   DDAY=__day;
   &var=lbblres;
   DFL='Y';
   output;
   end;
run;

proc sort data=data_3 out=&dsout;
by subjid lbtestcd &day dday;
run;
```

	subjid	lbtestcd	lbstresn	lbblres	lbblchg	lbanldy	DVALUE	DDAY	DFL
1	001	HSCRP	2.03	0	0	1	0	1	
2	001	HSCRP	1.79	2.03	-0.24	2	-0.24	2	
3	001	HSCRP	2.56	2.03	0.53	3	0.53	3	
4	001	HSCRP	2.56	2.03	2.03	3	0	2.3116883117	Y
5	001	HSCRP	1.3	2.03	-0.73	4	-0.73	4	
6	001	HSCRP	1.3	2.03	2.03	4	0	3.4206349206	Y
7	001	HSCRP	0.692	2.03	-1.338	5	-1.338	5	
8	001	HSCRP	0.52	2.03	-1.51	6	-1.51	6	
9	001	HSCRP	0.34	2.03	-1.69	8	-1.69	8	
10	001	HSCRP	0.182	2.03	-1.848	11	-1.848	11	
11	001	HSCRP	0.479	2.03	-1.551	15	-1.551	15	
12	001	HSCRP	2.98	2.03	0.95	22	0.95	22	
13	001	HSCRP	2.98	2.03	2.03	22	0	19.341063575	Y
14	001	HSCRP	1.69	2.03	-0.34	28	-0.34	28	
15	001	HSCRP	1.69	2.03	2.03	28	0	26.418604651	Y
16	001	HSCRP	4.68	2.03	2.65	43	2.65	43	
17	001	HSCRP	4.68	2.03	2.03	43	0	29.705685619	Y

CONCLUTSIONS

On the premise of not affect the final result, using the original observation to derive the dummy observation and using split the whole interval method to simple the procedure of calculation of the AUC. In order to easier calculate the different AUC. Eventually we modified the dataset to adapt to the trapezoidal rule.

CONTACT INFORMATION

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

Jianfeng Ye

Address 2/F, HNA Building, No. 108, Jianguo Road, Chaoyang District, 100022, Beijing, China

+8618710106985

yejianfeng186@163.com