Dealing with Exact Confidence Interval of Binary Endpoint Data

Kamlesh Patel, Jigar Patel, Dilip Patel, Vaishali Patel

Rang Technologies Inc, Piscataway, NJ

Abstract

Programmers come across various binary data when working on data analysis. However, when working on binary endpoint, one of the challenges is getting correct confidence interval (CI) for proportion. In SAS, most of the programmers get confidence interval using PROC FREQ procedure. Under certain scenario, PROC FREQ calculates confidence interval for proportion of another group and hence, it turns out to be incorrect confidence interval for proportion of the group programmers have requested. Here, we will present a simple but accurate way to get confidence interval of proportion. We will present a macro to calculate confidence interval which is based on derivation as per statistical method in data step. This macro will help programmer to compact the code and avoid miscalculation of confidence interval.

Intro

- Binary endpoint:
- It is one type of categorical end point where there are only two possible outcomes.
  - Response vs No-Response
  - Yes vs No
  - 1 vs 0
- Clopper-Pearson (exact) 95% CI
  - PROC FREQ

```
*TO Get 95% CI;
proc freq data=DS1 ;
by treatment;
tables RESPONSE / binomial;
exact binomial;
run;
```

Challenges of a programmers

1. Calculating 95% CI for a level which is not expected.
2. When the proportion is zero of Yes (responders), it calculates for No (non-responders)

Solution

EXACT macro - This macro can calculate 95% CI for all types of data correctly.

*NOTE: Macro is provided in paper

Dealing with Exact Confidence Interval of Binary Endpoint Data

Kamlesh Patel, Jigar Patel, Dilip Patel, Vaishali Patel

Rang Technologies Inc, Piscataway, NJ

How to use the EXACT Macro

USAGE OF EXACT MACRO

1. Things to know before using exact macro.
2. Binomial Response Variable: *Please make sure SUCCESS/RESPONSE is 1.* (level for which you want 95% CI).
3. Required Macro Parameters: ds=, var= .
4. Final output dataset would be named - CI_n like CI_1, CI_2 etc.

Parameter	Is it req?	Default Value	Note
DS	Req	Last Created Dataset (_last_)	Name of Dataset (Req.), If Missing, it would take last created dataset.
VAR	Req		Binomial Variable (Req.) i.e. response outcome variable. NOTE: Success/Response should be 1. All other values are considered as No-Success/No-Response.
BYGRP	Optional		Name of BY Group -- > Columns in table like treatment group. e.g. BYGRP = TRT01P. If more than one group, then you can pass more than one variable name. E.g. BYGRP = TRT01P STAGE
OUT	Optional	1	Any number to make output dataset unique. E.g. out=1 would give dataset CI_1.
KEEP	Optional	N	Keep=N (Default) ---> Just keep final results variable and remove intermediate variables like p, n, x, CI_low, CI_high. Keep=Y ---> Keep all intermediate variables like p, n, x, CI_low, CI_high AND result variable. You can Keep=Y for debugging.
PDEC	Optional	1	Decimals of Proportion. Default is 1 decimal
DEC	Optional	1	Decimals of CI values. Default is 1 decimal.
CI	Optional	95	Confidence interval. Default is 95, but if you want to calculate CI for any other % like 99, 90 then you can pass parameter here.

Dealing with Exact Confidence Interval of Binary Endpoint Data

Kamlesh Patel, Jigar Patel, Dilip Patel, Vaishali Patel

Rang Technologies Inc, Piscataway, NJ

```
/*Usage 1: Simplest with only required parameters - DS and VAR*/
%exact (ds=ADEFF,
     var=ResponseN);
*Note:
1. It will consider all records of DS in one group and will calculate 95% CI.
2. Will use default values as provided.
3. Output dataset = CI_1 (as out=1 in default);
```

```
/*Usage 2: Get 95% CI BY Treatment */
%exact (ds=ADEFF,
        bygrp=treatment,        /*Put By Group here*/
        var=ResponseN,
        out=2);         /*To create output dataset CI_2 */
*Note:
1. It will get it BY TREATMENT and will calculate 95% CI for each treatment group.
2. Will use default values as provided.
3. Output dataset = CI_2 (as out=2);
```

```
/*Usage 3: Get 99% CI BY Treatment and by Stages - Two BY GROUP */
%exact (ds=ADEFF,
        bygrp=treatment stage, /*Put By Group and Stage here*/
        var=ResponseN,
        out=3,           /* To create output dataset CI_3 */
        CI=99);              /* To get 99% CI */
*Note:
1. It will get it 99% BY TREATMENT, BY STAGE and will calculate
   99% CI for each treatment group within each STAGE.
2. Will use default values as provided.
3. Output dataset = CI_3 (as out=3)
4. CI = 99 will calculate CI for 99% ;
```

```
/*Usage 4: Get 95% CI BY Treatment and by Stages and want to look at
           other variables for Debugging */
%exact (ds=ADEFF,
        bygrp=treatment stage, /*Put By Group and Stage here*/
        var=ResponseN,
        out=4,
        CI=95,          /* To get 95% CI */
        Keep=Y);            /*Keep other intermediate variable*/
*Note:
1. It will get it 95% BY TREATMENT, BY STAGE and will calculate 95%
   CI for each treatment group within each STAGE.
2. Will use default values as provided.
3. Output dataset = CI_4 (as out=4)
4. CI = 95 will calculate CI for 95% ;
```

```
/*Usage 5: Get 95% CI BY Treatment and by Stages  +
        2 Decimal of proportion and CI in output + No other variable in final DS */
%exact (ds=ADEFF,
        bygrp=treatment stage, /*Put By Group and Stage here*/
        var=ResponseN,
        out=5,
        CI=95,          /* To get 95% CI */
        Pdec=2,         /* To get 2 decimals in proportion */
        Dec=3,          /* To get 3 decimals in CI */
        Keep=N);
*Note:
1. It will get it 95% BY TREATMENT, BY STAGE and will calculate 95%
   CI for each treatment group within each STAGE.
2. Will use default values as provided.
3. Output dataset = CI_5 (as out=5)
4. CI = 95 will calculate CI for 95%
5. Decimal in final output would be 2 for Proportion (PDEC) and 3 Decimal for CI;
```