

Estimating Binomial Proportion Confidence Interval with Zero Frequency Response using FREQ Procedure

Linga Reddy Baddam, Syneos Health®, Hyderabad, India

ABSTRACT

In the clinical industry, we often estimate binomial proportion confidence interval for the occurrences of the event using FREQ Procedure. There are cases where the frequency of an event will be zero (e.g., there are no patients with an event). In such a scenario, SAS by default computes the frequency, proportion and its confidence interval excluding zero frequency categories. Hence, the programmer would end up with unintended results. This paper will share tips on how to calculate binomial proportion, confidence interval and difference of proportion confidence interval with zero frequency event using FREQ procedure.

INTRODUCTION

In the day-to-day job, statistical programmers are highly accountable on quality and efficiency of reports for the given requirements by Statistician. For the given time constraints, it is difficult for them to get results from statistical procedure without understanding the consequences of how the data is prepared before passing to statistical procedure. This paper gives insights to statistical programmers, outlining step by step how to calculate binomial proportion confidence interval and difference of binomial proportion confidence interval, especially when frequency of event is zero (e.g., there are no patients with an event).

In clinical trials, the binomial test is used to make inferences about a proportion or response rate based on a series of independent observations, each resulting in one of two possible mutually exclusive outcomes. The outcomes can be response to treatment or no response, cure or no cure, survival or death, or in general, event or non-event.

For example, in the Clinical Trial Endpoint Tumor Response, usually we need to report response rate and its CI for Best Overall Response (CR/PR), Complete Response (CR), Partial Response (PR), Stable Disease (SD) and Progressive Disease (PD) in studies with more than one treatment arm, the comparison of response rates between treatment groups.

In SAS, the FREQ procedure can be used to obtain binomial proportion confidence interval and difference of proportion confidence interval.

Following are the general cases in which different ways we can pass data to PROC FREQ, which helps the reader to get fundamental understanding around PROC FREQ.

Of note, PROC FREQ can use either raw data or cell count data to produce frequency and cross tabulation tables.

CASE 1: Inputting RAW DATA or CASE-RECORD directly to PROC FREQ

Prerequisite: Raw data is to be prepared in such a way that has one record per one patient.

Example:

```
data raw;
input Subject $ Treatment Response $ @@;
datalines;
01 1 Yes 02 1 Yes 03 1 Yes 04 1 Yes 05 1 Yes
06 1 No 07 1 No 08 1 No 09 2 Yes 10 2 Yes
11 2 Yes 12 2 Yes 13 2 No 14 2 No 15 2 No
;
/*Response 1=CR, 2=Non CR; Treatment 1=Placebo,
2=Active*/
run;
```

	SUBJECT	TREATMENT	RESPONSE
1	01	1	Yes
2	02	1	Yes
3	03	1	Yes
4	04	1	Yes
5	05	1	Yes
6	06	1	No
7	07	1	No
8	08	1	No
9	09	2	Yes
10	10	2	Yes
11	11	2	Yes
12	12	2	Yes
13	13	2	No
14	14	2	No
15	15	2	No

Below is the SAS code to input the raw data to produce frequency and cross tabulation table:

```
proc freq data=raw ;
  tables treatment*response/nocol norow nopct;
run;
```

The FREQ Procedure

Table of Treatment by Response

Treatment	Response		Total
Frequency	No	Yes	
1	3	5	8
2	3	4	7
Total	6	9	15

CASE 2: Inputting CELL COUNT DATA directly to PROC FREQ

Prerequisite: The dataset is to be prepared in such a way to have the frequency counts of all possible combinations of treatment and response.

Example:

```
data Cellcounts;
input Treatment Response $ Count @@;
datalines;
1 Yes 5 1 No 3
2 Yes 4 2 No 3
;
```

	TREATMENT	RESPONSE	COUNT
1	1	Yes	5
2	1	No	3
3	2	Yes	4
4	2	No	3

Below is the SAS code to input the cell count data to produce the frequency and cross tabulation table:

```
proc freq data=cellcounts;
  weight count;
  tables treatment*response/nocol norow nopct;
run;
```

The FREQ Procedure

Table of Treatment by Response

Treatment	Response		Total
Frequency	No	Yes	
1	3	5	8
2	3	4	7
Total	6	9	15

Note: When we use the WEIGHT statement, PROC FREQ assumes that an observation represents n observations, where n is the value of a variable.

The **proportion** or **response** rate is defined as the number of successes divided by the number of trials.

In clinical words, the number of patients who responded (x) to an investigative treatment out of a total number of patients (n) studied.

Formula: Proportion or Response rate = x/n

OBTAINING BINOMIAL PROPORTION CI USING FREQ PROCEDURE

In SAS, the FREQ procedure can be used to obtain binomial proportion and its confidence interval. By default, PROC FREQ provides Wald and exact (Clopper-Pearson) for the binomial proportion. PROC FREQ also provides binomial proportion CI for Agresti-Coull, JEFFREYS and Wilson (Score) Confidence Limits when you request with CL = binomial option.

Following is the general scenario to obtain confidence interval for proportion or response rate. We will use the following sample data to obtain confidence interval for proportion:

```
data rawdata;
input Subject $ Treatment Response @@;
datalines;
01 1 1 02 1 1 03 1 1 04 1 1 05 1 1
06 1 2 07 1 2 08 1 2 09 2 1 10 2 1
11 2 1 12 2 1 13 2 2 14 2 2 15 2 2
;
```

*/*Response 1=CR, 2=Non CR; Treatment 1=Placebo, 2=Active*/*

```
run;
```

```

proc sort data=rawdata;
  by treatment;
run;

proc freq data=rawdata;
  by treatment;
  tables response/binomial(level="1") alpha=0.1;

  /*LEVEL= binomial option identifies the variable level for which to compute
  the proportion*/
  /*ALPHA= option determines the confidence level*/
  /*Example ALPHA=0.1 produces 100*(1-alpha)%=100*(1-0.1)% = 90% CI for
  binomial proportion*/
run;

```

The SAS System					The SAS System				
TREATMENT=1					TREATMENT=2				
The FREQ Procedure					The FREQ Procedure				
RESPONSE	Frequency	Percent	Cumulative Frequency	Cumulative Percent	RESPONSE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	5	62.50	5	62.50	1	4	57.14	4	57.14
2	3	37.50	8	100.00	2	3	42.86	7	100.00
Binomial Proportion RESPONSE = 1					Binomial Proportion RESPONSE = 1				
Wald CI					Exact or Clopper-Pearson CI				
Proportion	0.6250				Proportion	0.5714			
ASE	0.1712				ASE	0.1870			
90% Lower Conf Limit	0.3435				90% Lower Conf Limit	0.2638			
90% Upper Conf Limit	0.9065				90% Upper Conf Limit	0.8791			
Exact Conf Limits					Exact Conf Limits				
90% Lower Conf Limit	0.2892				90% Lower Conf Limit	0.2253			
90% Upper Conf Limit	0.8889				90% Upper Conf Limit	0.8712			
Test of H0: Proportion = 0.5					Test of H0: Proportion = 0.5				
ASE under H0	0.1768				ASE under H0	0.1890			
Z	0.7071				Z	0.3780			
One-sided Pr > Z	0.2398				One-sided Pr > Z	0.3527			
Two-sided Pr > Z	0.4795				Two-sided Pr > Z	0.7055			
Sample Size = 8					Sample Size = 7				

OUTPUT 1: Output from a FREQ procedure shows the binomial proportion and confidence interval.

There are cases where the frequency of an event is zero (e.g., there are no patients with an event). In such cases,

- I. If you use LEVEL binomial option, PROC FREQ throws an error in the LOG window.
- II. If you do not use LEVEL binomial option, PROC FREQ by default computes proportion and it's CI for next level (e.g., for patients without event).

Here is the sample data where frequency of an event is zero and its corresponding results:

```

data rawdata0;
  input Subject Treatment Response @@;
  datalines;
  01 1 1 02 1 1 03 1 1 04 1 1 05 1 1
  06 1 2 07 1 2 08 1 2 09 2 2 10 2 2
  11 2 2 12 2 2 13 2 2 14 2 2 15 2 2
  ;
  /*Response 1=CR, 2=Non CR; Treatment 1=Placebo,
  2=Active*/
run;

```

Subject	Treatment	Response
1	1	1
2	2	1
3	3	1
4	4	1
5	5	1
6	6	1
7	7	1
8	8	1
9	2	2
10	2	2
11	2	2
12	2	2
13	2	2
14	2	2
15	2	2

Frequency Table:

```
proc freq data= rawdata0;
  tables treatment*response/nocol norow nopct;
run;
```

The FREQ Procedure

Table of Treatment by Response

Treatment	Response		Total
Frequency	1	2	
1	5	3	8
2	0	7	7
Total	5	10	15

NOTE: If you observe here, treatment=2 has zero frequency for response=1. So going forward we will discuss how to calculate binomial proportion CI correctly to response=1 for treatment 2 as it has zero frequency.

i. Results with LEVEL binomial option

```
proc sort data=rawdata0;
  by treatment;
run;

proc freq data=rawdata0;
  by treatment;
  tables response/binomial(level="1") alpha=0.1;
  ***With LEVEL= binomial option***;
run;
```

SAS OUTPUT

The SAS System

TREATMENT=2

The FREQ Procedure

RESPONSE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2	7	100.00	7	100.00

SAS Log

```
157 ***With LEVEL= binomial option;
158 proc freq data=rawdata0;
159   by treatment;
160   tables response/binomial(level="1") alpha=0.1;
161 run;

ERROR: BINOMIAL LEVEL='1' is not found for variable Response.
NOTE: The above message was for the following by-group:
      Treatment=2
NOTE: The SAS System stopped processing this step because of errors.
NOTE: There were 15 observations read from the data set WORK.RAWDATA0.
NOTE: PROCEDURE FREQ used (Total process time):
      real time      0.12 seconds
      cpu time       0.01 seconds
```

If observed above, there is an error in the log, which explains there are no patients with event (e.g., response=1 for treatment 2). As a result, PROC FREQ has not calculated proportion and CI for response=1.

If we do not use the LEVEL option here, PROC FREQ gives proportion and it's CI for response=2 instead of response=1 without throwing an error. As a result, programmers may assume these results are accurate and put them in report for response=1.

ii. Results without LEVEL binomial option

```
proc freq data=rawdata0;
  by treatment;
  tables response/binomial alpha=0.1;
  ***Without LEVEL= binomial option;
run;
```

```

The SAS System
TREATMENT=1
The FREQ Procedure

RESPONSE   Frequency    Percent    Cumulative
Frequency    Cumulative
Percent
-----
1             5         62.50      5
2             3         37.50      8
-----

Binomial Proportion
RESPONSE = 1

Proportion          0.6250
ASE                  0.1712
90% Lower Conf Limit 0.3435
90% Upper Conf Limit 0.9065

Exact Conf Limits
90% Lower Conf Limit 0.2892
90% Upper Conf Limit 0.8889

Test of H0: Proportion = 0.5

ASE under H0        0.1768
Z                   0.7071
One-sided Pr > Z    0.2398
Two-sided Pr > |Z|  0.4795

Sample Size = 8

```

```

The SAS System
TREATMENT=2
The FREQ Procedure

RESPONSE   Frequency    Percent    Cumulative
Frequency    Cumulative
Percent
-----
2             7         100.00     7
-----

Binomial Proportion
RESPONSE = 2

Proportion          1.0000
ASE                  0.0000
90% Lower Conf Limit 1.0000
90% Upper Conf Limit 1.0000

Exact Conf Limits
90% Lower Conf Limit 0.6518
90% Upper Conf Limit 1.0000

Test of H0: Proportion = 0.5

ASE under H0        0.1890
Z                   2.6458
One-sided Pr > Z    0.0041
Two-sided Pr > |Z|  0.0082

Sample Size = 7

```

OUTPUT 2: Output from a FREQ procedure without ZEROS option.

If observed above, we got the proportion and its CI for response=2 for treatment=2 as we have zero frequency for response=1. It's against our requirement and intentions, and this is where a programmer needs to be proactive and precautionary while sending data to PROC FREQ.

To avoid these kinds of situations, following are the steps that guide programmers to pre-process data properly before we pass on to PROC FREQ to correctly calculate.

1. Convert raw data or case-record data to cell count data (e.g., listing all possible combinations of data values along with the frequency counts).

Raw Data

	SUBJECT	TREATMEN	RESPONSE
1	1	1	1
2	2	1	1
3	3	1	1
4	4	1	1
5	5	1	1
6	6	1	2
7	7	1	2
8	8	1	2
9	9	2	2
10	10	2	2
11	11	2	2
12	12	2	2
13	13	2	2
14	14	2	2
15	15	2	2

Cell Count Data

	TREATMENT	RESPONSE	COUNT
1	1	1	5
2	1	2	3
3	2	2	7



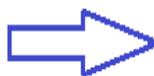
2. Update the data set created in above step 1 to add observation(s) for unavailable or missing variable level with zero frequency.

Cell Count Data

	TREATMENT	RESPONSE	COUNT
1	1	1	5
2	1	2	3
3	2	2	7

Cell Count Data with added zero frequency row

	TREATMENT	RESPONSE	COUNT
1	1	1	5
2	1	2	3
3	2	1	0
4	2	2	7



3. Add ZEROS option in the WEIGHT statement to correctly calculate proportion and its CI:

```
proc freq data=summary;
  by treatment;
  weight count/zeros;
  tables response/binomial(level="1") alpha=0.1;
  ***With LEVEL= binomial option;
run;
```

Here is the SAS program to calculate binomial proportion and its CI to get intended results using above 3 steps:

```
data rawdata0;
input Subject Treatment Response @@;
datalines;
01 1 1 02 1 1 03 1 1 04 1 1 05 1 1
06 1 2 07 1 2 08 1 2 09 2 2 10 2 2
11 2 2 12 2 2 13 2 2 14 2 2 15 2 2
;
/*Response 1=CR, 2=Non CR**; Treatment 1=Placebo, 2=Active*/
run;

/*Convert raw data to cell count data and add zero frequency for missing
response*/

proc format;
  value resp
    1="1"
    2="2";

  value trt
    1="1"
    2="2";
run;

proc means data=rawdata0 n completetypes nway;
  class treatment response/preloadfmt;
  var subject;
  output out=summary(drop=_type_ _freq_) n=count;
  format treatment trt. response resp.;
run;

*Getting Binomial Proportion and CI with Zeros option;
proc freq data=summary;
  by treatment;
  weight count/zeros;
  tables response/binomial(level="1") alpha=0.1;
run;
```

Final Results

The SAS System				
TREATMENT=1				
The FREQ Procedure				
RESPONSE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	5	62.50	5	62.50
2	3	37.50	8	100.00

The SAS System				
TREATMENT=2				
The FREQ Procedure				
RESPONSE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	0	0.00	0	0.00
2	7	100.00	7	100.00

Binomial Proportion	
RESPONSE = 1	
Proportion	0.6250
ASE	0.1712
90% Lower Conf Limit	0.3435
90% Upper Conf Limit	0.9065
Exact Conf Limits	
90% Lower Conf Limit	0.2892
90% Upper Conf Limit	0.8889
Test of H0: Proportion = 0.5	
ASE under H0	0.1768
Z	0.7071
One-sided Pr > Z	0.2398
Two-sided Pr > Z	0.4795
Sample Size	8

Binomial Proportion	
RESPONSE = 1	
Proportion	0.0000
ASE	0.0000
90% Lower Conf Limit	0.0000
90% Upper Conf Limit	0.0000
Exact Conf Limits	
90% Lower Conf Limit	0.0000
90% Upper Conf Limit	0.3482
Test of H0: Proportion = 0.5	
ASE under H0	0.1890
Z	-2.6458
One-sided Pr < Z	0.0041
Two-sided Pr > Z	0.0082
Sample Size	7

OUTPUT 3: Output from a FREQ procedure with ZEROS option.

It is clear from output 2 and output 3 that the ZEROS option of weight statement suppresses the PROC FREQ default nature of ignoring observations with zero frequency and produces the intended results.

OBTAINING DIFFERENCE OF BINOMIAL PROPORTION CONFIDENCE INTERVAL

The RISKDIFF option of TABLES statement in FREQ procedure provides binomial proportions (risks) and difference of binomial proportions (risk differences) for 2X2 contingency tables. This analysis is appropriate when comparing the proportion of some characteristic for two treatment groups.

When you request difference of binomial proportion CI using RISKDIFF option for 2X2 table, FREQ procedure by default gives Wald CI, Exact (Clopper-Pearson) CI for binomial proportion for row 1 (treatment 1), row 2 (treatment 2), total proportion and difference of proportion (treatment 1 – treatment 2).

In the 2X2 table, the row variable should be an exposure variable (e.g., treatment or dose and column variable should be response variable).

Below is the sample code to get difference of proportion confidence interval:

```
proc freq data=summary;
  weight count/zeros;
  tables treatment*response/riskdiff alpha=0.1;
run;
```

The SAS System						
The FREQ Procedure						
Statistics for Table of TREATMENT by RESPONSE						
Column 1 Risk Estimates						
	Risk	ASE	(Asymptotic) 90% Confidence Limits	(Exact) 90% Confidence Limits		
Row 1	0.6250	0.1712	0.3435	0.9065	0.2892	0.8889
Row 2	0.0000	0.0000	0.0000	0.0000	0.0000	0.3482
Total	0.3333	0.1217	0.1331	0.5335	0.1417	0.5774
Difference	0.6250	0.1712	0.3435	0.9065		
Difference is (Row 1 - Row 2)						
Column 2 Risk Estimates						
	Risk	ASE	(Asymptotic) 90% Confidence Limits	(Exact) 90% Confidence Limits		
Row 1	0.3750	0.1712	0.0935	0.6565	0.1111	0.7108
Row 2	1.0000	0.0000	1.0000	1.0000	0.6518	1.0000
Total	0.6667	0.1217	0.4665	0.8669	0.4226	0.8583
Difference	-0.6250	0.1712	-0.9065	-0.3435		
Difference is (Row 1 - Row 2)						
Sample Size = 15						

OUTPUT 4: Output from a FREQ procedure shows the binomial proportion CI and difference of proportion CI.

By default, FREQ procedure produces Wald CI, Exact (Clopper-Pearson) CI for binomial proportion (risk) for row 1, row 2, total proportion and difference of proportion (row 1 – row 2) for both the column 1 (response 1) and column 2 (response 2). We can control/restrict results using COLUMN=1 or 2 or BOTH options as required.

CONCLUSION

The objective and intentions of this paper is not to shed light on how PROC FREQ works, it is about how SAS programmers are to be proactive and precautionous in understanding sensitivity of data and further its consequences with respect to quality and efficiency of results. The author of this paper strongly believes that it's not the driving force of statistical procedure, but that the right understanding of the data gives quality and efficient results.

REFERENCES

SAS/STAT® 9.4 User's Guide, The FREQ Procedure.

ACKNOWLEDGMENTS

I take this opportunity to thank Rajani Dandamudi, Sudarshan Reddy, Purushotham Namburi, Krishna Gudipati and Brahma Reddy Kanala who have helped me in providing useful review comments on this paper. I am grateful to Vivek Dange at Syneos Health, whose support and guidance encouraged me to write this paper.

CONTACT INFORMATION

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

Linga Reddy Baddam
Syneos Health
2nd Floor, Block-1, DLF Cyber City, Gachibowli - 500032
Hyderabad, Telangana, India.
Work Phone: +91 40 496 52359
Email: linga.reddy@syneoshealth.com
Web: <http://www.syneoshealth.com/>

Any brand and product names are trademarks of their respective companies.