Computation of CIs for Binomial proportions in SAS and its practical difficulties

Jose Abraham, Kreara Solutions Pvt. Ltd., Trivandrum, India
Binomial Proportions

- One of the most commonly used method for the analysis of categorical data
- Used in situations in which the outcomes of interest can be categorized into two, i.e. either a success or a failure
- Binomial proportion can either be the proportion of successes or the proportion of failures
CIs for Binomial Proportion

- An interval which indicates that for different samples, the proportion may vary within this limit.
- It uses the proportion estimated in a statistical sample and allows for sampling error.
How to compute the CIs?

■ Common methods
  ■ Normal Approximation method
  ■ Clopper Pearson (exact) method
  ■ Wilson’s Score method

■ How to compute exact CIs in SAS
  ■ PROC FREQ with a ‘binomial’ option in the TABLES statement will give the exact CIs for the binomial proportion.
Are you confident that CIs from PROC FREQ are always correct?

<table>
<thead>
<tr>
<th>Question</th>
<th>Person No.</th>
<th>Response</th>
<th>Codedval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>1</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Q1</td>
<td>2</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Q1</td>
<td>3</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Q1</td>
<td>4</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Q1</td>
<td>5</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Q1</td>
<td>6</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Q1</td>
<td>7</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Q1</td>
<td>8</td>
<td>No</td>
<td>0</td>
</tr>
</tbody>
</table>
What we are interested in?

- Proportion of people who are confident that the Binomial Proportion CIs from PROC FREQ are always correct.
  - Proportion = $\frac{5}{8} = 0.625$

- Confidence interval for this proportion.
SAS Program and output

```sas
proc freq data=resp;
    tables codedval/binomial;
run;
```

### Binomial Proportion

for codedval = 0

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion</td>
<td>0.3750</td>
</tr>
<tr>
<td>ASE</td>
<td>0.1712</td>
</tr>
<tr>
<td>95% Lower Conf Limit</td>
<td>0.0395</td>
</tr>
<tr>
<td>95% Upper Conf Limit</td>
<td>0.7105</td>
</tr>
</tbody>
</table>

### Exact Conf Limits

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>95% Lower Conf Limit</td>
<td>0.0852</td>
</tr>
<tr>
<td>95% Upper Conf Limit</td>
<td>0.7551</td>
</tr>
</tbody>
</table>
Problem 1

- By default, PROC FREQ computes the CI for the lowest level of the variable which we use.
  - Lowest response level is zero (i.e. No)
  - PROC FREQ computed the proportion of people who are not confident that the Binomial Proportion CIs from PROC FREQ are always correct

Proportion value (.375) and the exact CIs (0.0852, 0.7551) from PROC FREQ are not the intended ones.
Solution?

- Solution 1

- Change the level of the variable by changing the coded values

  - Changing the code ‘0’ to ‘2’ will make the level ‘1’ the lowest level and it will be taken automatically

```data
data resp1;
set resp;
if codedval=0 then codedval=2;
run;
```
Solution?

- Solution 2
- Use the (level= ) option in the tables statement.

```plaintext
proc freq data=resp;
  tables codedval/binomial(level=2);
run;
```
SAS Output

Proportion value (0.6250) and the exact CIs (0.2449, 0.9148) from PROC FREQ are correct
Are you confident that CIs from PROC FREQ are always correct?

<table>
<thead>
<tr>
<th>Question</th>
<th>Person No.</th>
<th>Response</th>
<th>Codedval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>1</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Q1</td>
<td>2</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Q1</td>
<td>3</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Q1</td>
<td>4</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Q1</td>
<td>5</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Q1</td>
<td>6</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Q1</td>
<td>7</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Q1</td>
<td>8</td>
<td>No</td>
<td>0</td>
</tr>
</tbody>
</table>
Problem 2

- Absence of a required response level (or when the proportion is zero)
  - Proportion of people who are confident that the Binomial Proportion CIs from PROC FREQ are always correct is zero
  - But we still need to find the CIs for this zero proportion.
  - Since the required level is absent, PROC FREQ computes the CIs for the other level (ie. No)
Proportion value (1.0000) and the exact CIs (0.6306, 1.0000) from PROC FREQ are not the intended ones.
Solution?

- Add records to the dataset and then to make use the ‘WEIGHT’ statement with zeroes option in PROC FREQ.

- A dataset (ques) with possible outcomes for the question.

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
<th>Codedval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Q1</td>
<td>Yes</td>
<td>1</td>
</tr>
</tbody>
</table>
Solution?

- Merge these two datasets and create ‘wgt’ variable and then use it in the weight statement along with the (level=) option.

```plaintext
data resp3;
   merge resp(in=a) ques(in=b);
   by question response;
   if b and not a then wgt=0;
   else wgt=1;
run;

proc freq data=resp3;
   tables codedval/binomial(level=2);
   weight wgt/zeroes;
run;
```
SAS Output

Proportion value (0.0000) and the exact CIs (0.0000, 0.3694) from PROC FREQ are correct.
Is it possible to compute these CIs without PROC FREQ?

- A formula that uses quantiles from the beta distribution can be directly used in a data step.

```plaintext
data clop_ci;
  set aenum;
  p=round ((x/n),.0001);
  if p=0 then CI_LOW=0;
  if p=1 then CI_HIGH=1;
  if p ne 0 then CI_LOW=round((1-betainv(.975,(n-x+1),x)),.0001);
  if p ne 1 then CI_HIGH=round((1-betainv(.025,(n-x),x+1)),.0001);
run;
```

- Where
  - ‘x’ is the number of success.
  - ‘n’ is the total number
- Values 0.975 (1-\(\alpha/2\)) and 0.025 (\(\alpha/2\)) may vary with the chosen significance level
Summary

- Use of PROC FREQ without appropriate options may not always give the desired confidence intervals.
- To correct the CI computations in PROC FREQ
  - Make use of appropriate statements and options in PROC FREQ
  - Recoding of values and addition of records sometimes needed
- Alternate methods are also available in SAS to compute the Binomial CIs
Thank you