

The path less trodden - PROC FREQ for ODDS RATIO

Manjusha Gondil, Inventiv International Pharma Services Pvt. Ltd., Pune, India
 Linga Reddy Baddam, Inventiv International Pharma Services Pvt. Ltd., Pune, India

ABSTRACT

In our clinical industry we conduct various types of simple/complex statistical analysis for data interpretation and Odds Ratio is one such extremely useful and simple to understand technique.

Odds Ratio is used for making the critical decision about which treatment is benefiting the trial subjects in a clinical trial. In statistics, the odds of an event occurring is the probability of the event divided by the probability of an event not occurring.

Historically we calculate odds ratio using complex stats procedure like PROC LOGISTICS, PROC GENMOD etc., but this paper gives us an idea about easier alternative of how to find odds ratio using simpler stats procedures like PROC FREQ.

Programmers who do not have background of statistics will find this simplified method of finding odds ratio very easy to code and understand.

This method allows anyone to program *in a much easier way*.

INTRODUCTION

There are numbers of statistics that are useful for making decisions and drawing conclusions and odds ratio is one of such versatile and robust statistic. An odds ratio (OR) is defined as the ratio of the odds of an event occurring in one group to the odds of it occurring in another group, or to a data-based estimate of that ratio.

This paper gives a brief idea how to find odds ratio using Proc freq with example:

	Placebo	Study treatment
Event happen	a	b
Event does not happen	c	d

Table 1

Here is the formula:

$$OR = (c \times b)/(a \times d)$$

The odds ratio can be any nonnegative number. When the row and column variables are independent, the true value of the odds ratio equals 1. An odds ratio greater than 1 indicates that the odds of a positive response are higher in row 1 than in row 2. Below is an example of how to find the odds ratio using both, the historical PROC LOGISTIC and the faster and easier PROC FREQ.

	Standard Treatment	New Drug
died	120	10
survived	205	98

Table 2

The odds ratio is a way of comparing whether the odds of a certain outcome is the same for two different groups .It is simply the ratio between the following two ratios: The ratio between standard treatment and the new drug for those who died, and the ratio between standard treatment and the new drug for those who survived. From the data in the table 1, it is calculated as follows:

$$OR = (205 \times 10) / (120 \times 98) = 0.1743$$

There are two ways to get Odds Ratios in SAS when there is one predictor and one outcome variable.

HISTORICAL APPROACH

Algorithm for PROC LOGISTIC:

1. Create the new dataset from our existing dataset.
2. Change our variables to have values of 1 and 0
 - If someone has died then we will have a value 1 in new variable "pat1" and if they survived variable will have a value of "0."
 - If someone is taking standard treatment, we will have a value 1 in new variable "drug1" and new drug will have a value of "0."

Sample Code:

```

data test;
  input pat $ drug $ count;
datalines;
d s 120
d n 10
s s 205
s n 98
;
run;

data test2;
  set test;
  if pat="d" then pat1=1; else pat1=0;
  if drug="s" then drug1=1; else drug1=0;
run;

proc logistic data=test2;
  weight count;
  model pat1=drug1;
run;
    
```

The LOGISTIC Procedure					
Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.2824	0.3320	47.2693	<.0001
drug2	1	-1.7469	0.3513	24.7257	<.0001
Odds Ratio Estimates					
Effect	Point Estimate	95% Wald Confidence Limits			
drug2	0.174	0.088	0.347		
Association of Predicted Probabilities and Observed Responses					
Percent Concordant	25.0	Somers' D	0.000		
Percent Discordant	25.0	Gamma	0.000		
Percent Tied	50.0	Tau-a	0.000		
Pairs	4	c	0.500		

When performing a logistic regression with PROC LOGISTIC, the "Odds Ratio Estimates" table contains the odds ratio estimates and the corresponding 95% Wald confidence intervals. Similarly using PROC GENMOD, the logistic regression can be performed to calculate the odds ratio using the ESTIMATE statement with the EXP option. Also a 95% confidence interval for the OR is calculated.

Now that we have seen examples of the standard PROC LOGISTIC and PROC GENMOD used to calculate Odds ratio let us have a look at how this differs from our suggested method of using PROC FREQ.

THE EASIER APPROACH

Algorithm for PROC FREQ:

1. TABLES tell SAS to construct a table with the two specified variables.
2. The chisq option requests all Chi-Square statistics.
3. The relrisk option gives you estimates of the odds ratio and relative risks for the two columns.

Sample Code:

```

data test;
  input pat $ drug $ count;
datalines;
d s 120
d n 10
s s 205
s n 98
;
run;

proc print data = test;
run;

proc freq data = test;
  weight count;
  tables pat * drug / chisq relrisk;
run;
    
```

```

                Fisher's Exact Test
-----
Cell (1,1) Frequency (F)          10
Left-sided Pr <= F                5.605E-09
Right-sided Pr >= F               1.0000

Table Probability (P)             4.668E-09
Two-sided Pr <= P                 7.284E-09
    
```

```

                The FREQ Procedure

                Statistics for Table of pat by drug

                Estimates of the Relative Risk (Row1/Row2)

Type of Study          Value          95% Confidence Limits
-----
Case-Control (Odds Ratio)  0.1743          0.0876          0.3470
Cohort (Col1 Risk)       0.2378          0.1283          0.4409
Cohort (Col2 Risk)       1.3644          1.2440          1.4963
    
```

INTERPRETATION:

As you can see, Odds ratios can be calculated with PROC FREQ by specifying the relrisk option in the TABLES statement.

The odds ratio value is then listed beside "Case-control" in the section labeled "Estimates of the Relative Risk (Row1/Row2)."

The confidence intervals are labeled "Confidence bounds" and are 95% confidence intervals by default (can be changed using the ALPHA= option in the TABLES statement).

When the odds of the first group experiencing the event are less than the odds of the second group, one must reverse the two columns so that the second group becomes the first and the first group becomes the second. This way it will be possible to interpret the difference because that reversal will calculate how many more times the second group experienced the event than the first.

CONCLUSION

Following are a few key differentiators in favor of PROC FREQ:

1. We do not need to create new dataset from existing dataset to pass through model.
2. No model required so programmers who do not have background of statistics will find this easier.
3. Execution time will be less if you have huge data.

PROC FREQ is a procedure that is widely used among SAS users and this paper establishes that it does not only provide cross-tabulation tables but is a powerful procedure that can also compute statistics like odds ratio.

Odds ratio is simple to calculate, easy to interpret, provides results upon which clinical decisions can be made.

ACKNOWLEDGMENTS

We would like to thank our manager Namrata Murthy and all our colleagues and managers for their invaluable ideas and suggestions for preparation of this paper.

REFERENCES

Biochemia Medica: The journal of Croatian Society of Medical Biochemistry and Laboratory Medicine.

CONTACT INFORMATION

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

Name: Manjusha Mohanrao Gondil
Enterprise: Inventive International Pharma Services Pvt. Ltd.
Address: Commerzone, Building 4, 6th Floor, Yerwada, Pune.
City, State ZIP: Pune, India – 411006
E-mail: Manjusha.mohanrao@inventivhealth.com

Name: Linga Reddy Baddam
Enterprise: Inventive International Pharma Services Pvt. Ltd.
Address: Commerzone, Building 4, 6th Floor, Yerwada, Pune.
City, State ZIP: Pune, India – 411006
E-mail: linga.reddy@inventivhealth.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

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