

# How to Calculate Odds Ratios in SAS (With Example)

Authored by  
**stats writer**

November 19, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Calculate Odds Ratios in SAS (With Example)*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=96678>

**Odds ratios** are fundamental measures of association used extensively in fields like **epidemiology**, clinical research, and social sciences. They provide a precise quantification of the relationship between an exposure (such as a treatment, risk factor, or intervention) and a specific outcome (such as disease presence or successful completion of a task). Understanding how to calculate and interpret the OR is essential for assessing the strength of a potential risk factor or intervention effect.

The purpose of this comprehensive guide is to detail the methodology for calculating the Odds Ratio using **SAS**, one of the premier statistical software packages. We will walk through a practical example, providing clear, step-by-step instructions utilizing the powerful features of the SAS system. By the end of this tutorial, you will possess the requisite knowledge to accurately perform this critical statistical analysis and interpret the results effectively.

## Defining the Odds Ratio and Its Context

In statistical terms, the Odds Ratio (OR) represents the ratio of the odds of an event occurring in one group (typically the exposed or treatment group) compared to the odds of the event occurring in a reference group (typically the unexposed or control group). This metric is particularly valued because, unlike some other measures, it remains relatively consistent across various sampling designs, such as case-control studies. Furthermore, the OR is the natural output of logistic regression models, making it a cornerstone of modern statistical modeling.

The calculation of the OR is most frequently performed when analyzing discrete, categorical data summarized in a 2-by-2 contingency table. This table structure organizes the data based on two dichotomous variables: the exposure status and the outcome status. The four cells in the table (A, B, C, D) represent the frequencies of all possible combinations of exposure and outcome.

The standard 2-by-2 table format, which facilitates the straightforward calculation of the OR (often defined as the cross-product ratio,  $\$AD/BC\$$ ), is structured as follows:

	Event	No Event
Treatment	A	B
Control	C	D

## Utilizing PROC FREQ for Statistical Association Measures

To efficiently calculate the Odds Ratio and related measures within the SAS system, the **PROC FREQ** procedure is the definitive tool. **PROC FREQ** is designed specifically for analyzing one-way

to n-way frequency and cross-tabulation tables. By incorporating specific options within the procedure, SAS automatically calculates the OR, the associated standard error, tests for independence, and the corresponding confidence interval.

While the underlying arithmetic calculation for the OR is simple, leveraging **PROC FREQ** ensures that you receive the necessary statistical context. For rigorous analysis, it is essential not only to calculate the point estimate of the OR but also to determine its precision and whether the association is statistically meaningful. The required options, particularly **CHISQ** and **RELRISK**, provide this comprehensive output package.

## Case Study Setup: Comparing Athletic Training Programs

Consider a scenario involving performance enhancement among athletes. We aim to compare the effectiveness of a new training program against an established, old training program for basketball players. This constitutes a controlled comparison where the exposure is the training program type and the outcome is success on a skills test.

A total of 100 players participated in the study, equally divided: 50 players utilized the new training regimen, and the remaining 50 players followed the old regimen. At the conclusion of the intervention period, every player underwent a standardized skills test to determine if they successfully passed or failed. This setup perfectly aligns with the requirements of a 2x2 table analysis for calculating the OR.

The resulting categorical data summarizing the performance outcomes, based on the specific training program used, are aggregated into the following table. This table provides the raw frequency counts necessary to construct our 2-by-2 contingency table for the SAS analysis:

	Passed	Failed
New Program	34	16
Old Program	39	11

Our primary analytical objective is to calculate the Odds Ratio. Specifically, we want to quantify and compare the odds of a player successfully passing the skills test when participating in the **New Program** (our treatment group) relative to the odds of passing the skills test when participating in the **Old Program** (our reference group).

## Data Preparation and SAS Execution Syntax

Since our input is already summarized in frequency counts (aggregated data), we must structure our SAS data step to include variables for the outcome, the exposure, and the frequency count. The **WEIGHT** statement in **PROC FREQ** is mandatory when using aggregated data; it informs SAS that the value in the specified count variable represents multiple observations.

The code below first defines the dataset `my_data`, inputting the four cells of the contingency table. It then executes **PROC FREQ**. The **TABLES** statement specifies the cross-tabulation (program by result). The critical options **CHISQ** and **RELRISK** ensure that SAS computes the necessary inferential statistics, including the Chi-Square test and the Relative Risk measures alongside the Odds Ratio.

**/\*create dataset: The dataset contains the outcome (result), the exposure (program), and the frequency count.\*/**

```
data my_data;
```

```
input result $ program $ count;
```

```
datalines;
```

```
Passed New 34
```

```
Passed Old 39
```

```
_Failed New 16
```

```
_Failed Old 11
```

```
;
```

```
run;
```

**/\*calculate odds ratio using PROC FREQ. The WEIGHT statement informs SAS that 'count' holds the frequencies.\*/**

```
proc freq data=my_data;
```

```
weight count;
```

```
tables program * result / chisq relrisk;
```

```
run;
```

## Interpreting the Frequency and Chi-Square Output

The first table generated by **PROC FREQ** is the detailed frequency and cross-tabulation table. This table confirms that the data has been read correctly, showing the absolute frequencies, the percentage of the total sample, and the conditional percentages (row and column percentages).

The output should match our input counts, confirming the total sample size (N=100) and the specific cell counts used for the OR calculation (e.g., 34 Passed/New Program, 11 Failed/Old Program).

### The FREQ Procedure

Frequency  
Percent  
Row Pct  
Col Pct

Table of program by result			
program	result		
	Passed	_Failed	Total
<b>New</b>	34	16	50
	34.00	16.00	50.00
	68.00	32.00	
	46.58	59.26	
<b>Old</b>	39	11	50
	39.00	11.00	50.00
	78.00	22.00	
	53.42	40.74	
<b>Total</b>	73	27	100
	73.00	27.00	100.00

Immediately following the frequency table, SAS provides the Chi-Square tests of independence (due to the CHISQ option). While not the OR itself, this P-value tells us if there is overall evidence of an association between the program used and the test result. A non-significant Chi-Square test often foreshadows a non-significant Odds Ratio, as they test related hypotheses about the relationship between the row and column variables.

### Analyzing Measures of Association: The Odds Ratio Value

The section labeled "Estimates of the Common Odds Ratio" provides the specific results derived from the RELRISK option. This output is crucial as it contains the point estimate for the OR and its associated confidence interval.

The output specifically identifies the structure of the ratio (Row 1 / Col 1), which means SAS is comparing the odds of being in the first row category (New Program) against the second row category (Old Program), and the odds of being in the first column category (Passed) against the second column category (Failed).

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	0.5994	0.2449	1.4666
Relative Risk (Column 1)	0.8718	0.6855	1.1088
Relative Risk (Column 2)	1.4545	0.7518	2.8144

Sample Size = 100

From this output, the point estimate for the Odds Ratio is **0.5994**. This value quantifies the observed relationship between the training program and the passing outcome in our sample.

### Interpreting the Practical Implications of the OR

An Odds Ratio of **0.5994** is less than 1.0, indicating that the odds of the outcome occurring (passing the test) are lower in the exposed group (New Program) compared to the reference group (Old Program). If the OR were greater than 1.0, it would imply increased odds of success with the new program.

Specifically, we can state that the odds of a player passing the skills test using the new program are approximately 0.60 times the odds of passing using the old program. To express this as a percentage change, we calculate  $(1 - \text{OR}) \times 100\%$ , which results in  $(1 - 0.5994) \times 100\% \approx 40.06\%$ . Therefore, the practical interpretation is that the odds of a player passing the test are lowered by about 40.06% when switching from the old program to the new one, based on our sample data.

### Evaluating Statistical Significance and Precision

The magnitude of the OR (0.5994) is important, but its reliability is determined by the **95% Confidence Limits**. These limits define the range of values within which the true population OR is likely to lie.

The 95% confidence interval for the Odds Ratio is reported as: .

To determine statistical significance, we must check whether the interval includes the null value of the Odds Ratio, which is 1.0 (representing no difference in odds between the groups). Since the calculated interval spans both values less than 1.0 and values greater than 1.0, it definitively contains 1.0. Consequently, we must conclude that the observed difference is **not statistically significant** at the 95% confidence level. Although the point estimate suggests the new program is

worse, we lack sufficient statistical evidence to confidently reject the null hypothesis that the two programs are equally effective in the broader population.

## Beyond the Odds Ratio: Relative Risk Comparison

While the Odds Ratio is the preferred measure in case-control studies or logistic modeling, **PROC FREQ** also provides the Relative Risk (RR), also known as the Risk Ratio, which is often more intuitive for experimental or cohort study designs. The RR compares the probability (risk) of the outcome occurring in the exposed group relative to the unexposed group, rather than comparing the odds.

For large populations and rare outcomes, the OR provides a close approximation of the RR. However, when outcomes are common (as passing the test is here), the OR tends to exaggerate the magnitude of the Relative Risk. Being able to access both measures through the **REL RISK** option in **PROC FREQ** allows the analyst to select the most appropriate measure for the study design.

The following resources can provide deeper insight into statistical procedures related to categorical data analysis in SAS:

Understanding when to use the Odds Ratio versus the Relative Risk based on study design.

Interpreting the results of the Mantel-Haenszel common odds estimate, particularly when dealing with stratified analyses.

Introduction to **PROC LOGISTIC** in SAS for multivariate categorical data analysis.