

Learning to Calculate Odds Ratios in R: A Step-by-Step Guide

Authored by
Mohammed loot

October 27, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Learning to Calculate Odds Ratios in R: A Step-by-Step Guide*.
PSYCHOLOGICAL STATISTICS. Retrieved from
<https://statistics.arabpsychology.com/?p=4404>

In the field of [statistics](#) and epidemiology, the [Odds Ratio](#) (OR) is an indispensable metric used to quantify the strength of association between a specific exposure and a given outcome. This measure fundamentally establishes the ratio of the odds of an event occurring in an exposed or **treatment group** compared to the odds of the event occurring in an unexposed or **control group**. The OR is particularly powerful for analyzing data derived from research designs like case-control studies, offering a clear interpretation of risk or impact factors.

The magnitude of the odds ratio provides immediate insight into the relationship being studied. An OR exceeding 1 indicates that the odds of the outcome are significantly higher in the exposed group relative to the unexposed group. Conversely, if the OR is less than 1, the exposure is associated with decreased odds of the outcome. Crucially, an OR exactly equal to 1 suggests a complete absence of association between the factor under investigation and the outcome event. By interpreting the odds ratio, researchers can effectively gauge the potential impact of various factors, informing critical decisions across public health, clinical research, and social sciences.

The calculation of the odds ratio is typically anchored in data organized within a [2x2 contingency table](#). This tabular structure serves as an efficient framework for capturing the frequencies of two categorical variables--exposure status and outcome status--thereby making the subsequent computation and rigorous interpretation of the odds ratio both straightforward and highly systematic. Its clear presentation allows for immediate visualization of event occurrences and non-occurrences across the different exposure categories.

Understanding the 2x2 Contingency Table

The 2x2 table, frequently referred to as a fourfold or **contingency table**, is a foundational instrument in the analysis of categorical data. It systematically displays the joint frequency distribution for two binary variables: one typically defining the exposure or intervention status, and the other defining the outcome or event status. This rigid, structured approach ensures that all observations are precisely categorized into one of four distinct cells, which is essential for accurate epidemiological calculations.

The standard architecture of the table utilizes the cell counts 'a', 'b', 'c', and 'd' to represent the joint frequencies:

	Event	No Event
Treatment	A	B
Control	C	D

Within this standard layout, 'a' denotes the number of individuals who were exposed and

experienced the outcome; 'b' represents those who were exposed but did not experience the outcome; 'c' tallies individuals who were unexposed but developed the outcome; and finally, 'd' counts those who were unexposed and remained free of the outcome. From these counts, the odds of the outcome in the exposed group are calculated as the ratio (a/b) , and the odds in the unexposed group are (c/d) . The **Odds Ratio** is then derived as the ratio of these two odds: $(a/b) / (c/d)$, which simplifies algebraically to the well-known formula: $(a \times d) / (b \times c)$. This elegant formula is the backbone of quantifying associations in many types of statistical studies.

Calculating Odds Ratios in R: Leveraging the **epitools** Package

When conducting statistical analyses within the **R** environment, the calculation of odds ratios is greatly streamlined by utilizing specialized packages. The **epitools** package is highly regarded in epidemiological analysis, providing a comprehensive and efficient set of tools tailored for this specific type of data. It includes the dedicated function, `oddsratio()`, which automates complex underlying computations, delivering clear and readily interpretable results.

To begin working with `oddsratio()`, you must first ensure the package is installed and loaded into your current **R** session. If **epitools** is not yet installed on your system, you must execute the command `install.packages("epitools")`. Once installed, the subsequent command `library(epitools)` makes all its included functions, including the essential `oddsratio()` function, accessible for use in your script or console.

The `oddsratio()` function primarily requires a 2x2 **matrix** or a standard table object as its input argument, representing the contingency counts. Upon execution, the function generates a multi-component output. Beyond the core odds ratio estimate, it provides crucial accompanying statistics, most notably the 95% **confidence interval** (C.I.) and various **p-values**. These statistics are fundamental for rigorously assessing both the precision of the estimate and the statistical significance of the observed association.

Practical Example: Evaluating a Training Program's Efficacy

To demonstrate the practical application of calculating an odds ratio using **R**, let us analyze a hypothetical study designed to compare the effectiveness of two different basketball training programs: a **new training program** versus an **old training program**. In this randomized study, 100 basketball players are recruited. Fifty participants are randomly allocated to the new program, and the remaining 50 are assigned to the established old program. The core objective is to determine if the new methodology alters a player's likelihood of achieving success on a standardized skills test compared to the traditional program.

Following the completion of their respective training regimens, all participants undergo the standardized skills assessment to measure their acquired proficiency. The central research

question revolves around whether participation in the new program significantly changes a player's odds of passing this test when contrasted against those who followed the old regimen. The outcomes are meticulously recorded, resulting in the categorical data needed for a 2x2 analysis.

The aggregated results from the skills test, categorized by the training program and the pass/fail outcome, are structured as follows:

	Passed	Failed
New Program	34	16
Old Program	39	11

Our primary analytical goal is to compute the [odds ratio](#). This calculation will provide a quantitative comparison of the odds of a player successfully passing the skills test if they utilized the new program versus if they utilized the old program. This metric is essential for understanding the relative efficacy and potential benefits of adopting the new training methodology.

Data Preparation and Odds Ratio Calculation in R

Before executing the `oddsratio()` function, the raw count data must be correctly structured into a 2x2 contingency **matrix** that **R** can process efficiently. The following R code illustrates the necessary steps for constructing this matrix, including assigning descriptive row and column names to ensure clarity and accurate interpretation of the output.

```
# Create a vector for program types (rows)
program <- c('New Program', 'Old Program')
# Create a vector for outcome types (columns)
outcome <- c('Pass', 'Fail')
# Populate the matrix with the observed counts: (New Program Pass, New Program Fail, Old
Program Pass, Old Program Fail)
data <- matrix(c(34, 16, 39, 11), nrow=2, ncol=2, byrow=TRUE)
# Assign meaningful row and column names to the matrix
dimnames(data) <- list('Program'=program, 'Outcome'=outcome)

# Display the created matrix to verify its structure and content
data
```

Outcome

Program Pass Fail

New Program 34 16

Old Program 39 11

With the data matrix successfully constructed and verified, the next critical step is the calculation of the odds ratio using **epitools**. It is imperative that the package has been loaded before running the analysis. The subsequent R commands detail the typical workflow: checking for installation, loading the library, and finally, invoking the `oddsratio()` function using our newly created data matrix as the primary input argument.

Install the epitools package if it is not already present in your R library

install.packages('epitools')

Load the epitools package into the current R session to access its functions

library(epitools)

Execute the oddsratio function on the prepared data matrix to calculate the odds ratio and related statistics

oddsratio(data)

\$measure

odds ratio with 95% C.I.

Program estimate lower upper

New Program 1.0000000 NA NA

Old Program 0.6045506 0.2395879 1.480143

\$p.value

two-sided

Program midp.exact fisher.exact chi.square

New Program NA NA NA

Old Program 0.271899 0.3678219 0.2600686

\$correction

FALSE

attr(,"method")

"median-unbiased estimate & mid-p exact CI"

Interpreting the Odds Ratio and Statistical Significance

The comprehensive output generated by the `oddsratio()` function provides all the necessary components for drawing conclusions: the odds ratio estimate, the 95% **confidence interval** (C.I.),

and several associated [p-values](#). In our training program example, the "New Program" acts as the reference category (with an implied OR of 1). The calculated odds ratio estimate for the "Old Program" relative to the new one is approximately **0.6045506**.

Since the odds ratio is less than 1, this result suggests that the odds of the desired outcome (passing the skills test) are lower for players in the "Old Program" group when compared to those in the "New Program" group. Specifically, an OR of 0.6045506 implies that the odds of passing the test using the old program are about 60.46% of the odds of passing with the new program. Alternatively, this can be interpreted as a reduction in odds: $(1 - 0.6045506) * 100\%$, indicating a nearly 39.55% decrease in the odds of success for players enrolled in the traditional program relative to the new one.

We must also critically examine the 95% confidence interval, which is reported as . This range indicates the precision of our estimate and, crucially, the range within which the true population [odds ratio](#) is likely located. The primary rule for interpreting the C.I. in this context is checking if the value 1 is included. If the interval encompasses 1, it signifies that we cannot rule out the possibility of no association between the exposure and the outcome, meaning the result is not [statistically significant](#) at the 0.05 alpha level. Since our interval clearly contains 1, this provides a strong initial indication that the observed difference is likely due to chance.

Finally, the output includes various p-values, such as the **midp.exact** (often preferred for small 2x2 tables), [fisher.exact](#), and [chi.square](#) results. The midp.exact p-value is approximately **0.271899**. Compared against the typical significance threshold ($\alpha = 0.05$), 0.271899 is substantially greater. Therefore, we conclude that the odds ratio is not [statistically significant](#). This means that while the new program appears to yield better results in our specific sample, the evidence is not strong enough to confidently generalize this difference to the broader population of basketball players. The observed variation in passing rates could plausibly be attributed to random sampling error.

Important Considerations for Odds Ratio Analysis

Although the [odds ratio](#) is a powerful measure in descriptive and inferential [statistics](#), its effective application demands careful consideration of context, design limitations, and underlying assumptions. A crucial distinction must be maintained between the odds ratio and the [relative risk](#) (or risk ratio). While these two metrics produce numerically similar results when the outcome event is rare (prevalence < 10%), they diverge significantly when the outcome is common. The OR is primarily suited for case-control studies, where relative risk cannot be directly computed, but it is also widely applicable in prospective cohort and cross-sectional designs.

Furthermore, the reliability of the calculated odds ratio depends on the validity of certain assumptions inherent to the study design. In randomized controlled trials, like the training program

example, comparability between groups is assumed at baseline due to the random assignment process. However, in observational research, the presence of confounding variables can introduce severe bias into the OR calculation. In such scenarios, statistical adjustment, often requiring multivariate analysis, becomes necessary to isolate the true association between exposure and outcome.

Finally, researchers must ensure a nuanced interpretation that integrates both statistical and practical significance. A statistically non-significant **odds ratio**, such as the one derived in our example, does not automatically equate to the absence of any real effect. Instead, it may signal that the study lacked sufficient **statistical power** to detect a true difference, or it could mean that the true effect size is negligible from a practical standpoint. Drawing comprehensive conclusions requires researchers to weigh the clinical or theoretical importance of the estimated OR alongside the evidence provided by the 95% **confidence interval** and the associated **p-value**.

Further Exploration

For individuals seeking to expand their knowledge of odds ratios, related epidemiological concepts, or advanced analytical implementations within the **R** environment, the following resources are highly recommended. These tutorials, documentation links, and further reading materials offer diverse perspectives, practical examples, and detailed technical explanations to solidify your understanding of these essential statistical measures and their effective deployment in data analysis.