

# Package: parasiteR (via r-universe)

May 13, 2026

**Type** Package

**Title** A Theoretical-Practical Approach to Parasitological Data Analysis

**Description** Standardizes and streamlines the processing of parasitological data by integrating descriptive analyses of parasite count distributions, automated calculation of parasitological indices and their dispersion measures, and intuitive visualizations for representing these metrics (Bush et al. 1997 <[doi:10.2307/3284227](https://doi.org/10.2307/3284227)>, Reiczigel et al. 2019 <[doi:10.1016/j.pt.2019.01.003](https://doi.org/10.1016/j.pt.2019.01.003)>).

**License** GPL (>= 3)

**RoxygenNote** 7.3.3

**Version** 1.0

**Encoding** UTF-8

**Imports** rlang, ggplot2, magrittr, dplyr, tidyr, stats, BlakerCI, boot, readr

**Depends** R (>= 3.5)

**LazyData** true

**NeedsCompilation** no

**Author** Exequiel Oscar Furlan [aut] (ORCID: <<https://orcid.org/0000-0002-5642-3956>>), Juan Manuel Cabrera [aut, cre, cph] (ORCID: <<https://orcid.org/0000-0003-4378-1495>>), Elisa Helman [aut] (ORCID: <<https://orcid.org/0000-0002-6663-0334>>)

**Maintainer** Juan Manuel Cabrera <[juan.cabrera@uner.edu.ar](mailto:juan.cabrera@uner.edu.ar)>

**Repository** <https://cran.r-universe.dev>

**Date/Publication** 2026-05-13 10:47:54 UTC

**RemoteUrl** <https://github.com/cran/parasiteR>

**RemoteRef** HEAD

**RemoteSha** 178e70b3d18ef5cbf2f49c33ad884924a493796c

## Contents

para_abundance_CI . . . . .	2
para_data . . . . .	5
para_descriptors . . . . .	6
para_explo_abund . . . . .	9
para_explo_prev . . . . .	10
para_intensity_CI . . . . .	12
para_plot_CI . . . . .	15
para_prevalence_CI . . . . .	16
para_read_data . . . . .	19

<b>Index</b>	<b>21</b>
--------------	-----------

---

para_abundance_CI	<i>Mean or median abundance estimation and confidence intervals</i>
-------------------	---

---

## Description

This function calculates point estimates and confidence intervals (CIs) for parasite abundance, using either the mean or the median as a measure of central tendency. Confidence intervals are estimated via a non-parametric bootstrap approach based on resampling (permutations) of the observed data. Specifically, the function implements bias-corrected and accelerated (BCa) bootstrap intervals, which adjust for both bias and skewness in the bootstrap distribution. This approach does not assume a specific underlying distribution and is particularly robust for overdispersed and zero-inflated parasitological data.

## Usage

```
para_abundance_CI(dataset, c_median = TRUE,
  sp_cols, group_vars = NULL, perm = 2000, decimal_places = 2,
  combine_ci = FALSE, conf_level = 0.95, verbose = FALSE)
```

## Arguments

dataset	Data frame with parasitological data.
c_median	Logical. If TRUE, the results will include the median as a central tendency of measure; if FALSE, the results will include the mean of the data.
sp_cols	Vector with the names of the columns containing abundance of parasites (taxa) to calculate the parasitological descriptors.
group_vars	Vector with the names of categorical variables used to define groups (e.g., "Sex", "Site"). Default = NULL.
perm	Number of permutations to perform for confidence interval estimation. Default = 2000.
decimal_places	Number of decimal places to include in the calculation. Default = 2.
combine_ci	Logical. If TRUE, the interval is expressed as a single column (min - max). If FALSE, the interval is split into separate lower and upper limit columns.

conf_level	Confidence level for the interval estimation (e.g., 0.95 for 95% CI).
verbose	A logical value indicating if progress messages should be given. Default = FALSE.

## Details

Parasite abundance is defined as the number of individuals of a given parasite taxon per host. For each taxon, abundance metrics are calculated based on the observed counts across hosts. The function reshapes the dataset into long format and computes abundance statistics for each parasite taxon and grouping combination (if specified). The following are estimated:

- A is the total parasite abundance
- nH is the number of hosts analyzed
- nH\_inf is the number of infected hosts

Depending on the argument `c_median`, the function calculates:

- Mean abundance MeanA: average number of parasites per host
- Median abundance MedA: median number of parasites per host

Confidence intervals are estimated using a non-parametric bootstrap approach. Specifically, bias-corrected and accelerated (BCa) bootstrap intervals are computed by resampling the observed abundance values with replacement a specified number of times `m`. This method adjusts for both bias and skewness in the bootstrap distribution. Statistical considerations: parasite abundance data are typically overdispersed and zero-inflated, making parametric assumptions inappropriate in many cases. The use of bootstrap methods allows robust estimation of confidence intervals without assuming normality. Mean abundance is sensitive to extreme values, whereas median abundance provides a more robust measure under highly skewed distributions. When sample size is small, bootstrap confidence intervals may be unstable or wide, and results should be interpreted with caution. The interpretation of results remains the responsibility of the user.

## Value

A data frame containing abundance estimates and confidence intervals for each parasite taxon, either globally or by group. The following variables are returned:

- nH: Number of hosts analyzed
- nH\_inf: Number of infected hosts
- A: Total parasite abundance
- MeanA: Mean parasite abundance
- MedA: Median parasite abundance
- Lower\_CI: Lower bound of the bootstrap confidence interval
- Upper\_CI: Upper bound of the bootstrap confidence interval
- CI: If `combine_ci = TRUE`, confidence interval expressed are store as a single column (Lower\_CI - Upper\_CI)
- Observation: Categorical description of the data context:



---

para_data	<i>Simulated parasite abundance data for multiple species across hosts and sites</i>
-----------	--

---

## Description

This dataset contains hypothetical generated parasite count data representing multiple parasite species infecting individual hosts across different sampling sites. Each row corresponds to a single sampling unit (i.e., an individual host), and parasite abundance is recorded as counts for each parasite species (Sp1–Sp4).

## Usage

```
para_data
```

## Format

```
## 'para_data' A list with 4 elements
```

- dataset A data frame with 81 rows and 6 columns:
  - Site: Factor or character. Sampling location where hosts were collected (sites A, B, C and D). Multiple hosts can belong to the same site.
  - Factor or character. Host species identifier. In this dataset, each site includes up to two host species (HostA, HostB), although some site–host combinations may be absent by design.
  - Sp1: Integer. Abundance (count) of parasite species 1 per host. Simulated using an aggregated (negative binomial) distribution across all sites.
  - Sp2: Integer. Abundance of parasite species 2 per host. Present only in Sites A and B; missing (NA) in Site C to represent non-analyzed combinations.
  - Sp3: Integer. Abundance of parasite species 3 per host. Designed to represent heterogeneous infection patterns: full infection in one host group, rare infection in another and absence elsewhere.
  - Sp4: Integer. Integer. Abundance of parasite species 4 per host. Includes several edge cases: only one host examined, no infected hosts, a single infected host, multiple infected hosts.
- factors\_v: A list of columns with factor values.
- num\_v: A list of columns with numeric values.
- summ: A summary of the loaded data. Check [summary](#).

## Details

The dataset was intentionally constructed to reproduce common scenarios encountered in parasitological studies, rather than to reflect a specific empirical system. These scenarios include:

- zero-inflated parasite distributions
- aggregated parasite abundances

- missing data (non-analyzed host–parasite combinations)
- rare infections (single infected host)
- absence of infection
- small sample sizes for specific host–site combinations

This structure allows testing and demonstrating the behavior of analytical functions under realistic and edge-case conditions.

---

para\_descriptors      *Parasitological descriptors and summary statistics*

---

### Description

Computes standard parasitological descriptors and classical summary statistics from parasite abundance data, optionally stratified by grouping variables.

### Usage

```
para_descriptors(dataset, sp_cols = NULL, group_vars = NULL,
  decimal_places = 2, verbose = FALSE)
```

### Arguments

dataset	Data frame with parasitic abundance data.
sp_cols	Vector with the names or indices of the species columns.
group_vars	Vector with the names of the categorical variables to consider (e.g., 'Sex', 'Site').
decimal_places	Number of decimal places to round the values.
verbose	A logical value indicating if progress messages should be given.

### Details

The para\_descriptors function provides a practical and efficient way to estimate the main parasitological descriptors commonly used in ecological and parasitological studies. Calculations can be performed globally or at different hierarchical levels defined by grouping variables.

The function computes descriptors based on parasite abundance per sampling unit (e.g., host, site, or pooled hosts), following standard definitions:

- Prevalence (P): Proportion of infected hosts..
- Abundance (A): Total number of parasites recorded.
- Intensity (I): Number of parasites per infected host.

Statistical validity and sample size considerations: The estimation of summary statistics is subject to fundamental statistical constraints related to sample size and variability.

- **Host population (nH):** Summary statistics for abundance (e.g., mean, median, standard deviation) are only meaningful when more than one host is analyzed ( $nH > 1$ ). When only a single host is available ( $nH = 1$ ), no population of hosts exists and therefore no variability can be estimated. In such cases, the observed value is reported, but it should not be interpreted as a population-level descriptor. Summary statistics for abundance (e.g., mean, median, standard deviation) are only meaningful when more than one host is analyzed ( $nH > 1$ ). When only a single host is available ( $nH = 1$ ), no population of hosts exists and therefore no variability can be estimated. In such cases, the observed value is reported, but it should not be interpreted as a population-level descriptor.
- **Infected host population (nH\_inf):** Similarly, intensity-based descriptors are only meaningful when more than one infected host is available ( $nH\_inf > 1$ ). When only one infected host is present, the parasite count corresponds to a single observation rather than a distribution and summary statistics of intensity cannot be formally estimated.

These constraints reflect a fundamental principle: statistical descriptors require variability, and variability requires more than one observational unit. When this condition is not met, results should be interpreted cautiously, and no generalization beyond the observed case is justified. Handling of special cases: The function automatically adjusts calculations depending on data availability:

- When no data are available → results are reported as NA.
- When hosts are analyzed but none are infected → prevalence is 0 and intensity measures are not computed.
- When only one host or one infected host is available → corresponding summary statistics are not computed, and interpretation should be limited to the observed value.

The selection and interpretation of descriptors remain the responsibility of the user, particularly when working with small sample sizes.

## Value

A data frame containing the calculated parasitological descriptors for each parasite taxon, either globally or by group (if grouping variables are specified). The following variables are returned:

- nH: Number of hosts analyzed
- nH\_inf: Number of infected hosts
- A: Total parasite abundance
- min: Minimum parasite count
- max: Maximum parasite count
- P: Parasitic prevalence
- MeanA: Mean parasitic abundance
- MeanA\_sd: Standard deviation of mean parasite abundance
- A\_iqr: Interquartile range of mean parasite abundance
- MedA: Median parasite abundance
- MedA\_sd: Median absolute deviation of parasite abundance
- MeanI: Mean parasite intensity

- MeanI\_sd: Standard deviation of mean parasite intensity
- I\_iqr: Interquartile range of mean parasite intensity
- MedI: Median parasite intensity
- MedI\_sd: Median absolute deviation of parasite intensity
- Observation: Qualitative descriptor indicating data availability and sample structure for each hierarchical combination:
  - "Not analyzed": No valid observations are available for the given combination (all values are missing or the combination is absent in the dataset); therefore, no estimates can be computed.
  - "One host analyzed": Only a single host analyzed is available for the given combination; thus, no population-level inference is possible and statical summary measures are not estimated.
  - "No hosts infested": Hosts are present for the given combination, but none are infested (abundance = 0 for all observations); consequently, no statical summary measures of abundance or intensity can be estimated.
  - "One host infested": Only a single infested host is recorded for the given combination; therefore, no sample-based estimation of intensity or related summary measures is possible.
  - "Multiple hosts infested": More than one infested host is recorded for the given combination, allowing the estimation of summary measures.

### Author(s)

Juan Manuel Cabrera, Exequiel Furlan and Elisa Helman

### References

- Bush, A.O., Lafferty, K.D., Lotz, J.M., Shostak, A.W. (1997). Parasitology meets ecology on its own terms: Margolis revisited. *Journal of Parasitology*, 83(4), 575–583.
- Reiczigel, J., Marozzi, M., Fabian, I., Rózsa, L. (2019). Biostatistics for parasitologists – a primer to quantitative parasitology. *Trends in Parasitology*, 35(4), 277–281.

### Examples

```
gral_descriptor <- para_descriptors(para_data$dataset,
  sp_cols = c("Sp1", "Sp2", "Sp3", "Sp4"),
  group_vars = c("Site", "Sp_host"),
  decimal_places = 2,
  verbose = FALSE)
```

```
gral_descriptor
```

---

para\_explo\_abund      *Exploratory plots of parasite abundance distributions*

---

### Description

Generates exploratory visualizations of parasite abundance distributions across taxa and optional grouping variables. The function produces histograms combined with kernel density curves to facilitate the assessment of distributional patterns, including skewness, dispersion and zero inflation.

### Usage

```
para_explo_abund(dataset, sp_cols, group_vars = NULL,  
  bins = 30, n_col = NULL, verbose = FALSE)
```

### Arguments

dataset	Data frame containing parasite data.
sp_cols	Vector with the names of the columns containing parasite abundance (taxa) to be plotted.
group_vars	Vector with the names of categorical variables used to define groups (e.g., "Sex", "Site"). Default = NULL.
bins	Integer specifying the number of bins used in the histogram. Higher values provide finer resolution but may introduce noise, while lower values produce smoother but less detailed distributions. Default = 30.
n_col	Integer specifying the number of columns in the faceted plot layout. If NULL, the number of columns is determined automatically by ggplot2. Default = NULL.
verbose	A logical value indicating if progress messages should be given. Default = FALSE.

### Details

The function reshapes the input dataset into a long format, where parasite taxa are treated as a single variable and their abundances as observations. For each parasite taxon and combination of grouping variables (if provided), the function generates:

- A **histogram** representing the distribution of parasite abundance values
- A **kernel density curve** (when sufficient data are available), providing a smoothed approximation of the underlying distribution.

Both elements are scaled to represent density, allowing direct comparison between distributions. These plots are intended for exploratory purposes and should not be used as formal inference tools. Faceting is applied to display each taxon and grouping combination in separate panels. Special cases are handled as follows:

- When all abundance values for a given combination are zero, no histogram or density curve is drawn and a message is displayed indicating that the parasite was not recorded for that combination.

- When the number of observations is insufficient (less than 2), a message is displayed indicating that there is not enough data to compute a meaningful distribution.
- Density curves are only computed when there are more than two observations and non-zero variation.

All plots use independent scales (free scales) to better represent the variability within each facet.

### Value

A ggplot2 object containing the generated faceted plots. This object can be further customized using standard ggplot2 functions.

### Author(s)

Juan Manuel Cabrera, Exequiel Furlan and Elisa Helman

### Examples

```
#Species 1 and 2

para_explo_abund (para_data$dataset,
  sp_cols = c("Sp1", "Sp2"),
  group_vars = c("Site", "Sp_host"),
  bins = 30,
  n_col = 4,
  verbose = TRUE)

#Species 3 and 4

para_explo_abund (para_data$dataset,
  sp_cols = c("Sp3", "Sp4"),
  group_vars = c("Site", "Sp_host"),
  bins = 30,
  n_col = 4,
  verbose = TRUE)
```

---

para\_explo\_prev

*Exploratory plots of parasite prevalence*

---

### Description

Generates exploratory visualizations of parasite prevalence across taxa and optional grouping variables. The function produces stacked bar plots showing the proportion of infested and non-infested hosts, facilitating the assessment of prevalence patterns across hierarchical combinations.

### Usage

```
para_explo_prev(dataset, sp_cols, group_vars = NULL,
  n_col = NULL, verbose = FALSE)
```

## Arguments

dataset	Data frame containing parasite data.
sp_cols	Vector with the names of the columns containing parasite abundance (taxa) to be plotted.
group_vars	Vector with the names of categorical variables used to define groups (e.g., "Sex", "Site"). Default = NULL.
n_col	Integer specifying the number of columns in the faceted plot layout. If NULL, the number of columns is determined automatically by ggplot2. Default = NULL.
verbose	A logical value indicating if progress messages should be given. Default = FALSE.

## Details

The function reshapes the dataset into long format and calculates prevalence as the proportion of infested hosts (hosts with parasite counts > 0) relative to the number of analyzed hosts for each parasite taxon and grouping combination. For each combination, the function generates:

- The proportion of infested hosts.
- The proportion of non-infested hosts.

Faceting is applied to display each parasite taxon and grouping combination in separate panels. Special cases are handled as follows:

- When no observations are available (all values are missing or the combination is absent), a message is displayed indicating that the data were not analyzed.
- When only one host is available, a message is displayed indicating that the sample size is insufficient for prevalence estimation.
- When all observed values are zero, a message is displayed indicating that the parasite was not recorded for that combination.

All proportions are expressed on a 0–1 scale. These plots are intended for exploratory purposes and should not be used as formal inference tools.

## Value

A ggplot2 object containing the generated faceted stacked bar plots. This object can be further customized using standard ggplot2 functions.

## Author(s)

Juan Manuel Cabrera, Exequiel Furlan and Elisa Helman

## Examples

```
#Species 1 and 2
para_explo_prev(para_data$dataset,
                sp_cols = c("Sp1", "Sp2"),
```

```

group_vars = c("Site", "Sp_host"),
n_col = 4,
verbose = TRUE)

#Species 3 and 4

para_explo_prev(para_data$dataset,
  sp_cols = c("Sp3", "Sp4"),
  group_vars = c("Site", "Sp_host"),
  n_col = 4,
  verbose = TRUE)

```

---

para\_intensity\_CI      *Mean or median intensity estimation and confidence intervals*

---

## Description

This function calculates point estimates and confidence intervals (CIs) for parasite intensity, using either the mean or the median as a measure of central tendency. Confidence intervals are estimated via a non-parametric bootstrap approach based on resampling (permutations) of the observed data. Specifically, the function implements bias-corrected and accelerated (BCa) bootstrap intervals, which adjust for both bias and skewness in the bootstrap distribution. This approach does not assume a specific underlying distribution and is particularly robust for overdispersed and zero-inflated parasitological data.

## Usage

```

para_intensity_CI(dataset, c_median = TRUE, sp_cols, group_vars = NULL,
  perm = 2000, decimal_places = 2, combine_ci = FALSE,
  conf_level = 0.95, verbose = FALSE)

```

## Arguments

dataset	Data frame with parasitological data.
c_median	Logical. If TRUE, the results will include the median as a central tendency of measure; if FALSE, the results will include the mean of the data.
sp_cols	Vector with the names of the columns containing abundance of parasites (taxa) to calculate the parasitological descriptors.
group_vars	Vector with the names of categorical variables used to define groups (e.g., "Sex", "Site"). Default is NULL.
perm	Number of permutations to perform for confidence interval estimation. Default is 2000.
decimal_places	Number of decimal places to include in the calculation. Default is 2.
combine_ci	Logical. If TRUE, the interval is expressed as a single column (min - max). If FALSE, the interval is split into separate lower and upper limit columns

conf_level	Confidence level for interval estimation (e.g., 0.95 for 95% confidence intervals).
verbose	A logical value indicating if progress messages should be given.

## Details

Parasite intensity is defined as the number of individuals of a given parasite taxon per infested host. For each taxon, intensity metrics are calculated based only on hosts with parasite counts greater than zero. The function reshapes the dataset into long format and computes intensity statistics for each parasite taxon and grouping combination (if specified). The following are estimated:

- A is the total parasite abundance
- nH is the number of hosts analyzed
- nH\_inf is the number of infected hosts

Depending on the argument `c_median`, the function calculates:

- Mean intensity MeanI: average number of parasites per infested host
- Median intensity MedI: median number of parasites per infested host

Confidence intervals are estimated using a non-parametric bootstrap approach. Specifically, bias-corrected and accelerated (BCa) bootstrap intervals are computed by resampling the observed intensity values with replacement a specified number of times `perm`. This method adjusts for both bias and skewness in the bootstrap distribution. Statistical considerations: parasite intensity data are typically right-skewed and may exhibit high variability due to aggregation among hosts. The use of bootstrap methods allows robust estimation of confidence intervals without assuming normality. Mean intensity is sensitive to extreme values, whereas median intensity provides a more robust measure under highly skewed distributions. When the number of infested hosts is small, bootstrap confidence intervals may be unstable or wide, and results should be interpreted with caution. The interpretation of results remains the responsibility of the user.

## Value

A data frame containing abundance estimates and confidence intervals for each parasite taxon, either globally or by group. The following variables are returned:

- nH: Number of hosts analyzed
- nH\_inf: Number of infected hosts
- A: Total parasite abundance
- MeanA: Mean parasite intensity
- MedA: Median parasite intensity
- Lower\_CI: Lower bound of the bootstrap confidence interval
- Upper\_CI: Upper bound of the bootstrap confidence interval
- CI: If `combine_ci = TRUE`, confidence interval expressed are store as a single column (Lower\_CI - Upper\_CI)
- Observation: Categorical description of the data context:



para\_plot\_CI

*Visualization of parasitological descriptor with confidence intervals***Description**

This function generates graphical representations of parasitological estimates (abundance, intensity, or prevalence) including their associated confidence intervals. It supports multiple input formats and automatically detects the response variable and confidence interval structure. The function allows flexible grouping, species filtering, and visualization either as faceted plots or separate panels. The function is designed to be compatible with outputs from different estimation functions within the package (e.g., [para\\_abundance\\_CI](#), [para\\_intensity\\_CI](#), [para\\_prevalence\\_CI](#)). Automatic detection of confidence intervals ensures flexibility across workflows. Interpretation of graphical outputs remains the responsibility of the user. It automatically detects:

**Usage**

```
para_plot_CI(para_data, group_vars, sp_cols = NULL, descriptor = NULL,
             lower_ci = NULL, upper_ci = NULL, point_color = "blue", line_size = 1,
             point_size = 3, n_cols = 1, include_zeros = TRUE, separate_plots = FALSE)
```

**Arguments**

para_data	Data frame containing parasitological descriptors and confidence intervals estimated with one of the following functions: <a href="#">para_abundance_CI</a> , <a href="#">para_intensity_CI</a> , <a href="#">para_prevalence_CI</a> .
group_vars	Character vector specifying the variable(s) to be used on the x-axis. Multiple variables will be combined.
sp_cols	Optional vector of parasite taxa to include in the plot. Default is NULL (all taxa are included).
descriptor	Name of the variable to be plotted on the y-axis. If NULL, the function automatically detects a suitable variable (e.g., prevalence, MeanA, MedA, MeanI, MedI).
lower_ci	Optional names of the columns containing the lower confidence. If NULL, the function automatically detects and extracts them. Default is NULL.
upper_ci	Optional names of the columns containing the upper confidence. If NULL, the function automatically detects and extracts them. Default is NULL.
point_color	Color of the points. Default is "blue".
line_size	Line width of the confidence interval bars. Default is 1.
point_size	Size of the points. Default is 3.
n_cols	Number of columns used in faceted plots. Default is 1.
include_zeros	Logical. If FALSE, zero values are excluded from the plot. Default is TRUE.
separate_plots	Logical. If TRUE, returns a list of plots (one per species). If FALSE, produces a faceted plot. Default is FALSE.

## Details

- The response variable to be plotted.
- The structure of confidence intervals, including:
  - Separate columns (Lower\_CI, Upper\_CI)
  - Method-specific intervals (e.g., exact or Blaker)
  - Combined intervals stored as a single character column (e.g., "min - max")

When multiple grouping variables are provided in `x_var`, they are combined into a single factor for visualization. Confidence intervals are displayed as vertical error bars, and point estimates are overlaid. When multiple parasite taxa are present, results are displayed using faceting or as separate plots.

## Value

A ggplot object or a list of ggplot objects representing the estimated values and their confidence intervals.

## Author(s)

Juan Manuel Cabrera, Exequiel Furlan and Elisa Helman

## References

- Bush, A.O., Lafferty, K.D., Lotz, J.M., Shostak, A.W. (1997). Parasitology meets ecology on its own terms: Margolis revisited. *Journal of Parasitology*, 83(4), 575–583.
- Reiczigel, J., Marozzi, M., Fabian, I., Rózsa, L. (2019). Biostatistics for parasitologists – a primer to quantitative parasitology. *Trends in Parasitology*, 35(4), 277–281.

---

para\_prevalence\_CI      *Parasite prevalence estimation and confidence intervals*

---

## Description

Estimates parasite prevalence and corresponding confidence intervals from parasite abundance data, optionally stratified by grouping variables. Two types of confidence intervals are provided: exact binomial intervals and Blaker intervals, allowing robust inference across a wide range of sample sizes and prevalence values.

## Usage

```
para_prevalence_CI(dataset, sp_cols, group_vars = NULL, decimal_places = 2,  
conf_level = 0.95, output_type = "proportion", combine_ci = FALSE, verbose = FALSE)
```

### Arguments

dataset	Data frame with parasitological data.
sp_cols	Vector with the names of the columns containing abundance of parasites (taxa) to calculate the parasitological descriptors.
group_vars	Vector with the names of categorical variables used to define groups (e.g., "Sex", "Site"). Default is NULL.
decimal_places	Number of decimal places to round the values. Default is 2.
conf_level	Confidence level for interval estimation (e.g., 0.95 for 95% confidence intervals).
output_type	Format of the result: either "proportion" or "percentage". Default is "proportion".
combine_ci	Logical. If TRUE, the interval is expressed as a single column (min - max). If FALSE, the interval is split into separate lower and upper limit columns.
verbose	A logical value indicating if progress messages should be given. Default = FALSE

### Details

Prevalence is defined as the proportion of hosts infected with a given parasite taxon:

$$P = \frac{nH_{inf}}{nH}$$

where:

- $nH$  is the number of hosts analyzed (non-missing observations)
- $nH_{inf}$  is the number of infected hosts (abundance > 0)

The function reshapes the dataset into long format and computes prevalence for each parasite taxon and grouping combination (if specified). Two types of confidence intervals are calculated:

- **Exact (Clopper–Pearson) interval:** This is an exact binomial confidence interval, conservative but valid for all sample sizes, especially small samples or extreme prevalence values.
- **Blaker interval:** This interval is also exact but less conservative than Clopper–Pearson, providing shorter intervals while maintaining correct coverage.

Statistical considerations:

- Prevalence is a binomial proportion and can be estimated even for small sample sizes.
- However, when sample size is very small (e.g.,  $nH = 1$ ), the estimate lacks precision and confidence intervals become uninformative.
- When no infected hosts are observed ( $nH_{inf} = 0$ ), prevalence is 0, and confidence intervals reflect uncertainty around zero.

The interpretation of results, particularly under small sample sizes, remains the responsibility of the user.



```
combine_ci = TRUE,  
verbose = TRUE)  
  
prevalence_CI
```

---

para_read_data	<i>Read parasite data</i>
----------------	---------------------------

---

## Description

Load data from a .CSV file

## Usage

```
para_read_data(file_name, verbose = FALSE)
```

## Arguments

file_name	Name of .CSV table file.
verbose	A logical value indicating if progress messages should be given.

## Details

This package includes a specific function to import tables (.CSV files) into the R environment. Each row in the table should correspond to an individual host that was analyzed, while the columns may contain both quantitative and qualitative variables. Columns may represent two principal categories of variables:

- "Host-related variables": Encompassing metadata such as the site of specimen collection, host species, morphophysiological traits, applied experimental treatments, and other relevant descriptors.
- "Parasite-related variables": Denoting parasite abundance per host, typically structured across multiple columns corresponding to the finest available taxonomic resolution (e.g., species, genus, family, order).

Parasite abundance values must be encoded as non-negative integers. It is critical to distinguish between the following:

- 0: Represents a confirmed absence of the parasite in the host specimen.
- NA: Indicates that parasite detection or quantification was not feasible due to methodological or technical limitations.

**Value**

The function returns:

<code>dataset</code>	A table that can be used as input for other parasiteR functions.
<code>factors_v</code>	A list of columns with factor values.
<code>num_v</code>	A list of columns with numeric values.
<code>summ</code>	A summary of the loaded data. Check <code>summary()</code> function

**Author(s)**

Juan Manuel Cabrera, Exequiel Furlan and Elisa Helman

# Index

## \* datasets

- para\_data, [5](#)
  
- para\_abundance\_CI, [2](#), [15](#)
- para\_data, [5](#)
- para\_descriptors, [6](#)
- para\_explo\_abund, [9](#)
- para\_explo\_prev, [10](#)
- para\_intensity\_CI, [12](#), [15](#)
- para\_plot\_CI, [15](#)
- para\_prevalence\_CI, [15](#), [16](#)
- para\_read\_data, [19](#)
  
- summary, [5](#)