

Package ‘ConfidenceEllipse’

June 1, 2025

Type Package

Title Computation of 2D and 3D Elliptical Joint Confidence Regions

Version 1.1.0

Maintainer Christian L. Goueguel <christian.goueguel@gmail.com>

Description Computing elliptical joint confidence regions at a specified confidence level. It provides the flexibility to estimate either classical or robust confidence regions, which can be visualized in 2D or 3D plots. The classical approach assumes normality and uses the mean and covariance matrix to define the confidence regions. Alternatively, the robustified version employs estimators like minimum covariance determinant (MCD) and M-estimator, making them less sensitive to outliers and departures from normality. Furthermore, the functions allow users to group the dataset based on categorical variables and estimate separate confidence regions for each group. This capability is particularly useful for exploring potential differences or similarities across subgroups within a dataset. Varmuza and Filzmoser (2009, ISBN:978-1-4200-5947-2). Johnson and Wichern (2007, ISBN:0-13-187715-1). Raymaekers and Rousseeuw (2019) <[DOI:10.1080/00401706.2019.1677270](https://doi.org/10.1080/00401706.2019.1677270)>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

Imports cellWise, dplyr, forcats, ggplot2, magrittr, pcaPP, purrr,
rgl, rlang, stats, tibble, tidyr, tidyselect

URL <https://christiangoueguel.github.io/ConfidenceEllipse/>,
<https://github.com/ChristianGoueguel/ConfidenceEllipse>

Suggests knitr, patchwork, rmarkdown, spelling, testthat (>= 3.0.0)

VignetteBuilder knitr

Depends R (>= 2.10)

Language en-US

Config/testthat/edition 3

BugReports <https://github.com/ChristianGoueguel/ConfidenceEllipse/issues>

NeedsCompilation no

Author Christian L. Goueguel [aut, cre] (ORCID:
<<https://orcid.org/0000-0003-0521-3446>>)

Repository CRAN

Date/Publication 2025-06-01 08:10:02 UTC

Contents

confidence_ellipse	2
confidence_ellipsoid	4
glass	5
Index	7

confidence_ellipse	<i>Confidence Ellipse Coordinates</i>
--------------------	---------------------------------------

Description

Compute the coordinate points of confidence ellipses at a specified confidence level.

Usage

```
confidence_ellipse(  
  .data,  
  x,  
  y,  
  .group_by = NULL,  
  conf_level = 0.95,  
  robust = FALSE,  
  distribution = "normal"  
)
```

Arguments

.data	data frame or tibble.
x	column name for the x-axis variable.
y	column name for the y-axis variable.
.group_by	column name for the grouping variable (NULL by default). Note that this grouping variable must be a factor.
conf_level	confidence level for the ellipse (0.95 by default).
robust	optional (FALSE by default). When set to TRUE, it indicates that robust estimation method is employed to calculate the coordinates of the ellipse. The location is estimated using a 1-step M-estimator with the biweight psi function, while the scale is estimated using the Minimum Covariance Determinant (MCD) estimator. This approach is more resistant to outliers and provides more reliable ellipse boundaries when the data contains extreme values or follows a non-normal distribution.

distribution optional ("normal" by default). The distribution used to calculate the quantile for the ellipse. It can be either "normal" or "hotelling".

Details

The function computes the coordinates of the confidence ellipse based on the specified confidence level and the provided data. It can handle both classical and robust estimation methods, and it supports grouping by a factor variable. The `distribution` parameter controls the statistical approach used for ellipse calculation. The "normal" option uses the chi-square distribution quantile, which is appropriate when working with very large samples. Whereas the "hotelling" option uses Hotelling's T^2 distribution quantile. This approach accounts for uncertainty in estimating both mean and covariance from sample data, producing larger ellipses that better reflect sampling uncertainty. This is statistically more rigorous for smaller sample sizes where parameter estimation uncertainty is higher.

The combination of `distribution = "hotelling"` and `robust = TRUE` offers the most conservative and statistically rigorous approach, particularly recommended for exploratory data analysis and when dealing with datasets that may not meet ideal statistical assumptions. For very large samples, the default settings (`distribution = "normal"`, `robust = FALSE`) may be sufficient, as the differences between methods diminish with increasing sample size.

Value

Data frame of the coordinates points.

Author(s)

Christian L. Goueguel

References

- Raymaekers, J., Rousseeuw P.J. (2019). Fast robust correlation for high dimensional data. *Technometrics*, 63(2), 184-198.
- Brereton, R. G. (2016). Hotelling's T-squared distribution, its relationship to the F distribution and its use in multivariate space. *Journal of Chemometrics*, 30(1), 18–21.

Examples

```
# Data
data("glass", package = "ConfidenceEllipse")
# Confidence ellipse
ellipse <- confidence_ellipse(.data = glass, x = SiO2, y = Na2O)
ellipse_grp <- confidence_ellipse(
  .data = glass,
  x = SiO2,
  y = Na2O,
  .group_by = glassType
)
```

confidence_ellipsoid *Confidence Ellipsoid Coordinates*

Description

Compute the coordinate points of confidence ellipsoids at a specified confidence level.

Usage

```
confidence_ellipsoid(
  .data,
  x,
  y,
  z,
  .group_by = NULL,
  conf_level = 0.95,
  robust = FALSE,
  distribution = "normal"
)
```

Arguments

<code>.data</code>	data frame or tibble.
<code>x</code>	column name for the x-axis variable.
<code>y</code>	column name for the y-axis variable.
<code>z</code>	column name for the z-axis variable.
<code>.group_by</code>	column name for the grouping variable (NULL by default). Note that this grouping variable must be a factor.
<code>conf_level</code>	confidence level for the ellipsoid (0.95 by default).
<code>robust</code>	optional (FALSE by default). When set to TRUE, it indicates that robust estimation method is employed to calculate the coordinates of the ellipse. The location is estimated using a 1-step M-estimator with the biweight psi function, while the scale is estimated using the Minimum Covariance Determinant (MCD) estimator. This approach is more resistant to outliers and provides more reliable ellipse boundaries when the data contains extreme values or follows a non-normal distribution.
<code>distribution</code>	optional ("normal" by default). The distribution used to calculate the quantile for the ellipse. It can be either "normal" or "hotelling".

Details

The function computes the coordinates of the confidence ellipse based on the specified confidence level and the provided data. It can handle both classical and robust estimation methods, and it supports grouping by a factor variable. The `distribution` parameter controls the statistical approach used for ellipse calculation. The "normal" option uses the chi-square distribution quantile,

which is appropriate when working with very large samples. Whereas the "hotelling" option uses Hotelling's T^2 distribution quantile. This approach accounts for uncertainty in estimating both mean and covariance from sample data, producing larger ellipses that better reflect sampling uncertainty. This is statistically more rigorous for smaller sample sizes where parameter estimation uncertainty is higher.

The combination of `distribution = "hotelling"` and `robust = TRUE` offers the most conservative and statistically rigorous approach, particularly recommended for exploratory data analysis and when dealing with datasets that may not meet ideal statistical assumptions. For very large samples, the default settings (`distribution = "normal"`, `robust = FALSE`) may be sufficient, as the differences between methods diminish with increasing sample size.

Value

Data frame of the coordinates points.

Author(s)

Christian L. Goueguel

References

- Raymaekers, J., Rousseeuw P.J. (2019). Fast robust correlation for high dimensional data. *Technometrics*, 63(2), 184-198.
- Brereton, R. G. (2016). Hotelling's T-squared distribution, its relationship to the F distribution and its use in multivariate space. *Journal of Chemometrics*, 30(1), 18–21.

Examples

```
# Data
data("glass", package = "ConfidenceEllipse")
# Confidence ellipsoid
ellipsoid <- confidence_ellipsoid(.data = glass, x = SiO2, y = Na2O, z = Fe2O3)
ellipsoid_grp <- confidence_ellipsoid(
  .data = glass,
  x = SiO2,
  y = Na2O,
  z = Fe2O3,
  .group_by = glassType
)
```

glass

Glass Vessels Data

Description

The dataset is comprised of 13 different measurements for 180 archaeological glass vessels from different groups (Janssen, K.H.A., De Raedt, I., Schalm, O., Veeckman, J.: *Microchim. Acta* 15 (suppl.) (1998) 253-267. Compositions of 15th - 17th century archaeological glass vessels excavated in Antwerp.).

6

glass

Usage

`glass`

Format

Data frame of 180 rows and 14 columns.

Index

* **datasets**

glass, [5](#)

confidence_ellipse, [2](#)

confidence_ellipsoid, [4](#)

glass, [5](#)