

Package ‘regtomean’

July 15, 2025

Type Package

Title Regression Toward the Mean

Version 1.2.1

Imports formattable, effsize, plotrix, stats, ggplot2, htmlwidgets

Suggests testthat (>= 3.0.0)

Description In repeated measures studies with extreme large or small values it is common that the subjects measurements on average are closer to the mean of the basic population. Interpreting possible changes in the mean in such situations can lead to biased results since the values were not randomly selected, they come from truncated sampling. This method allows to estimate the range of means where treatment effects are likely to occur when regression toward the mean is present.

Ostermann, T., Willich, Stefan N. & Luedtke, Rainer. (2008). Regression toward the mean - a detection method for unknown population mean based on Mee and Chua's algorithm. BMC Medical Research Methodology.<doi:10.1186/1471-2288-8-52>.

Acknowledgments: We would like to acknowledge ``Lena Roth" and ``Nico Steckhan" for the package's initial updates (Q3 2024) and continued supervision and guidance. Both have contributed to discussing and integrating these methods into the package, ensuring they are up-to-date and contextually relevant.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

Depends R (>= 3.4.0)

NeedsCompilation no

Author Daniela Recchia [aut, cre],
Thomas Ostermann [ctb],
Julian Stein [ctb]

Maintainer Daniela Recchia <daniela.rodriguesrecchia@uni-wh.de>

Repository CRAN

Config/testthat/edition 3

Date/Publication 2025-07-15 14:10:02 UTC

Contents

cordata	2
language_test	3
meechua_eff.CI	4
meechua_plot	5
meechua_reg	6
plot_mu	8
plot_t	9
replicate_data	11

Index	13
--------------	-----------

cordata	<i>Correlation and Cohen's d effect sizes.</i>
---------	--

Description

This function calculates the correlation for the data and Cohen's d effect sizes, both based on pooled and on treatment standard deviations. It can optionally display the results in an HTML widget.

Usage

```
cordata(Before, After, within = TRUE, data = NULL)
```

Arguments

Before	a numeric vector giving the data values for the first (before) measure.
After	a numeric vector giving the data values for the second (after) measure.
within	A logical indicating whether the effect sizes should be computed based on paired samples (TRUE, default) or independent samples (FALSE).
data	an optional data frame containing the variables in the formula. By default the variables are taken from environment (formula).

Details

This function computes the correlation between two measures and calculates Cohen's d effect sizes using both pooled and treatment standard deviations.

- If `within = TRUE`, the effect sizes are computed assuming paired samples. - If `within = FALSE`, the effect sizes are computed assuming independent samples.

The results are returned as a data frame and also displayed in an HTML widget in the RStudio Viewer or default web browser.

Value

Return a table containing the correlation, effect size pooled and effect size based on treatment.

Author(s)

Daniela Recchia, Thomas Ostermann.

References

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). New York:Academic Press.

See Also

[cohen.d.cor](#)

Examples

```
cordata("Before", "After", data=language_test)
```

language_test

Language Test in High School

Description

A dataset with scores from 8 students who failed a high school test and could not get their diploma. They repeated the exam and got new scores.

Usage

```
data("language_test")
```

Format

A data frame with 8 observations on the following 9 variables.

Student a numeric vector

Before a numeric vector

After a numeric vector

‘Total N’ a numeric vector

Cross a numeric vector

‘Pre-treatment Mean’ a numeric vector

‘Pre-treatment Std’ a numeric vector

‘Post-treatment Mean’ a numeric vector

‘Post-treatment Std’ a numeric vector

Author(s)

Daniela Recchia, Thomas Ostermann.

Source

McClave, J.T; Dietrich, F.H.: "Statistics"; New York, Dellen Publishing; 1988.

Examples

```
data(language_test)
## maybe str(language_test) ; plot(language_test) ...
```

meechua_eff.CI	<i>Calculates and plots treatment and regression effects as also its p-values.</i>
----------------	--

Description

This function calculates and plots treatment and regression effects of both before and after measures as also its p-values.

Usage

```
meechua_eff.CI(x, n, se_after)
```

Arguments

x	a data frame containing the results from meechua_reg. It is stored as mod_coef.
n	the original sample size (number of observations) from data.
se_after	the estimated standard error from meechua_reg. It is stored as se_after.

Details

After performing the meechua_reg the model coefficients mod_coef as also its global variable se_after are used as input in this function to estimate treatment and regression effects.

Value

Two plots are performed, the first "Treatment Effect and p-value" and the second "Confidence Intervals" for mu.

Author(s)

Daniela Recchia, Thomas Ostermann

References

Ostermann, T., Willich, Stefan N. & Luedtke, Rainer. (2008). Regression toward the mean - a detection method for unknown population mean based on Mee and Chua's algorithm. BMC Medical Research Methodology.

See Also[meechua_reg](#)**Examples**

```
# Initialize environment explicitly
#regtomean_env <- new.env(parent = emptyenv())

# Generate example data
language_test <- data.frame(
  Before = rnorm(100, mean = 50, sd = 10),
  After = rnorm(100, mean = 55, sd = 10)
)

# Replicate data
mee_chua <- replicate_data(0, 100, "Before", "After", data = language_test)
mee_chua <- mee_chua[order(mee_chua$mu), ]

# Perform regression analysis and store results
results <- meechua_reg(mee_chua)
mod_coef <- results$mod_coef
se_after <- results$se_after

# Call meechua_eff.CI
meechua_eff.CI(mod_coef, 100, se_after)
```

`meechua_plot`*Plot models from meechua_reg*

Description

This functions plots all 4 diagnostics plots for each linear regression model: "Residuals vs Fitted", "Normal Q-Q", "Scale-Location" and "Residuals vs Leverage".

Usage

```
meechua_plot(models = NULL, env = regtomean_env)
```

Arguments

<code>models</code>	A list containing the estimated linear models, typically the output of <code>meechua_reg</code> . If <code>models</code> is <code>NULL</code> , the function attempts to retrieve the models from the specified environment (<code>env</code>).
<code>env</code>	An environment where the models are stored. The default is <code>regtomean_env</code> . This argument is used only if <code>models</code> is not explicitly provided.

Details

For each model from `models` 4 diagnostic plots are performed. For the first model the numbers 1 to 4 should be given, for the second model numbers from 8 to 12, and so on.

Value

Diagnostics plots for the set of models from `meechua_reg`.

Author(s)

Daniela Recchia, Thomas Ostermann.

References

Ostermann, T., Willich, Stefan N. & Luedtke, Rainer. (2008). Regression toward the mean - a detection method for unknown population mean based on Mee and Chua's algorithm. BMC Medical Research Methodology.

See Also

[plot.lm,meechua_reg](#)

Examples

```
# Generate example data
language_test <- data.frame(
  Before = rnorm(100, mean = 50, sd = 10),
  After = rnorm(100, mean = 55, sd = 10)
)

# Replicate data
mee_chua <- replicate_data(50, 60, "Before", "After", data = language_test)
mee_chua_sort <- mee_chua[order(mee_chua$mu), ]

# Perform regression analysis
results <- meechua_reg(mee_chua_sort)

# Plot models
meechua_plot(results$models)
```

`meechua_reg`

Fit linear models on the (replication) data.

Description

This function fit linear models for a subset of data frames.

Usage

```
meechua_reg(x)
```

Arguments

x Data to be used in the regression.

Details

The data used for the regression must be sorted by mu.

A set of linear models will be estimated and model coefficients are saved and stored in `mod_coef`.

The estimated standard error for the after measure is also stored in `se_after` to be used further in other functions.

Value

A table containing the estimations for each mu. The variables `models`, `mod_coef`, `se_after` are stored globally for further analysis if `to_global` is set to `TRUE`. In any case the values will be returned. The models are saved in an object called `mee_chua`, which is not automatically printed but is saved in the environment.

Author(s)

Daniela Recchia, Thomas Ostermann.

References

Ostermann, T., Willich, Stefan N. & Luedtke, Rainer. (2008). Regression toward the mean - a detection method for unknown population mean based on Mee and Chua's algorithm. *BMC Medical Research Methodology*.

See Also

[lm, dply](#)

Examples

```
# Generate example data
language_test <- data.frame(
  Before = rnorm(100, mean = 50, sd = 10),
  After = rnorm(100, mean = 55, sd = 10)
)

# Replicate data
replicate_data <- function(start, end, Before, After, data) {
  mu <- seq(start * 100, end * 100, by = (end - start))
  mu <- rep(mu, each = nrow(data))

  before <- data[[Before]] - mu / 100
  after <- data[[After]]
```

```

    mee_chua <- data.frame(mu = mu, before = before, after = after)
    return(mee_chua)
}

mee_chua <- replicate_data(0, 1, "Before", "After", data = language_test)
mee_chua <- mee_chua[order(mee_chua$mu), ] # Sortieren nach 'mu'

# Regression ausführen und Ergebnisse erhalten
reg_results <- mee_chua_reg(mee_chua)

# Zugriff auf Ergebnisse
mod_coef <- reg_results$mod_coef
se_after <- reg_results$se_after

# Anzeigen der Ergebnisse
print(mod_coef)
print(se_after)

```

plot_mu

Plot t-Statistics and p-Values for Intervention Impact

Description

Based on the data before and after the intervention and the regression models from the function `meechua_reg`, this function plots the t-statistics and p-values for a given range of μ to assess whether the intervention has a significant impact on the measurements, accounting for regression to the mean.

Usage

```
plot_mu(x, n, se_after, lower = F, alpha = 0.05)
```

Arguments

<code>x</code>	A data frame containing the results from <code>meechua_reg</code> . Specifically, this should be the <code>mod_coef</code> data frame obtained from <code>meechua_reg</code> .
<code>n</code>	The original sample size (number of observations) of the data.
<code>se_after</code>	The estimated standard error from <code>meechua_reg</code> . This should be the <code>se_after</code> vector obtained from <code>meechua_reg</code> .
<code>lower</code>	A boolean value specifying the direction of the one-sided tests. For <code>lower = FALSE</code> (the default), it tests whether the intervention is increasing the measurements. For <code>lower = TRUE</code> , it tests whether the second measurements are lower than expected.
<code>alpha</code>	Specifies the significance threshold for the p-values of the corresponding one-sided tests. The default is <code>alpha = 0.05</code> .

Value

A list containing the most significant μ , t-statistic, p-value, and the range of μ for which the treatment impact is significant.

Author(s)

Julian Stein

References

Ostermann, T., Willich, S. N., & Luedtke, R. (2008). Regression toward the mean - a detection method for unknown population mean based on Mee and Chua's algorithm. BMC Medical Research Methodology.

Examples

```
data(language_test)
df <- replicate_data(0, 100, "Before", "After", data = language_test)
result <- meechua_reg(df)
plot_mu(result$mod_coef, n = 8, se_after = result$se_after)
```

plot_t

Plot Results for p-values and t-values

Description

This function plots the t-statistics and p-values for a range of μ values, based on the provided data and regression models. It helps visualize whether the intervention has a significant impact on the measurements, accounting for regression to the mean.

Usage

```
plot_t(
  mu_start,
  mu_end,
  n,
  y1_mean,
  y2_mean,
  y1_std,
  y2_std,
  cov,
  lower = F,
  alpha = 0.05,
  r_insteadof_cov = F
)
```

Arguments

mu_start	Numeric. The starting value of μ for the range of values to be plotted.
mu_end	Numeric. The ending value of μ for the range of values to be plotted.
n	Numeric. The original sample size (number of observations) of the data.
y1_mean	Numeric. The mean of the first measurement.
y2_mean	Numeric. The mean of the second measurement.
y1_std	Numeric. The standard deviation of the first measurement.
y2_std	Numeric. The standard deviation of the second measurement.
cov	Numeric. The covariance between the two measurements, or if r_insteadof_cov is TRUE, the correlation coefficient.
lower	Logical. If TRUE, the function tests whether the second measurements are lower than expected. If FALSE (the default), it tests whether the intervention is increasing the measurements.
alpha	Numeric. The significance threshold for the p-values of the one-sided tests. The default is 0.05.
r_insteadof_cov	Logical. If TRUE, cov is interpreted as the correlation coefficient instead of the covariance. Default is FALSE.

Value

A ggplot2 plot with two y-axes: one showing p-values and the other showing t-statistics. The function also prints key values including the most significant μ , the minimal p-value, and the range of μ where the treatment effect is significant.

Author(s)

Julian Stein

References

Ostermann, T., Willich, S. N., & Luedtke, R. (2008). Regression toward the mean - a detection method for unknown population mean based on Mee and Chua's algorithm. BMC Medical Research Methodology.

Examples

```
# Example usage of the plot_t function
plot_t(
  mu_start = 0, mu_end = 10, n = 50, y1_mean = 5,
  y2_mean = 5, y1_std = 2, y2_std = 2, cov = 0.5
)

plot_t(
  mu_start = 0, mu_end = 10, n = 50, y1_mean = 5,
  y2_mean = 5, y1_std = 2, y2_std = 2, cov = 0.5,
  lower = TRUE, alpha = 0.1
)
```

)

replicate_data	<i>Replicates before and after values 100 times.</i>
----------------	--

Description

This function replicates 100 times the "before" and "after" values, given a start and end reference for the population mean (μ).

Usage

```
replicate_data(start, end, Before, After, data)
```

Arguments

start	A numeric value specifying the start value for μ .
end	A numeric value specifying the end value for μ .
Before	A numeric vector giving the data values for the first ("before") measurement.
After	A numeric vector giving the data values for the second ("after") measurement.
data	An optional data frame containing the Before and After variables. If not provided, the Before and After vectors must be supplied directly.

Details

To overcome the limitations of Mee and Chua's test regarding the population mean (μ), this function performs a replication of the data over a specified range of values.

The replicated data is used for systematically estimating the unknown population mean (μ). Further analyses are based on this new dataset.

Value

A data frame containing the replicated dataset, which includes the columns μ , before, and after.

Author(s)

Daniela Recchia, Thomas Ostermann.

References

Ostermann, T., Willich, Stefan N., & Luedtke, Rainer. (2008). Regression toward the mean - a detection method for unknown population mean based on Mee and Chua's algorithm. *BMC Medical Research Methodology*.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute* (15: 246-263).

See Also[rep](#)**Examples**

```
# Example usage of replicate_data
replicate_data(0, 100, "Before", "After", data = language_test)
```

Index

* datasets

language_test, 3

cohen.d, 3

cor, 3

cordata, 2

dlply, 7

language_test, 3

lm, 7

meechua_eff.CI, 4

meechua_plot, 5

meechua_reg, 5, 6, 6

plot.lm, 6

plot_mu, 8

plot_t, 9

rep, 12

replicate_data, 11