Package 'TeachHist'

January 20, 2025

Type Package

Title A Collection of Amended Histograms Designed for Teaching
Statistics

Version 0.2.1

Maintainer Carsten Lange `<clange@cpp.edu>`

Description Statistics students often have problems understanding
the relation between a random variable's true scale and its z-values.
To allow instructors to better better visualize histograms for these
students, the package provides histograms with two horizontal
axis containing z-values and the true scale of the variable.
The function TeachHistDens() provides a density histogram with two axis.
TeachHistCounts() and TeachHistRelFreq() are variations for count and
relative frequency histograms, respectively. TeachConfInterv() and
TeachHypTest() help instructors to visualize confidence levels
and the results of hypothesis tests.

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 3.5.0), ggplot2, dplyr

RoxygenNote 7.2.3

NeedsCompilation no

Author Carsten Lange [aut, cre]

Repository CRAN

Date/Publication 2023-10-27 05:50:02 UTC

Contents

TeachHist	*TeachHist: A Collection of Amended Histograms Designed for Teaching Statistics*

Description

Statistics students often have problems understanding the relation between a random variable's true scale and its z-values. To allow instructors to better visualize histograms for these students, the package provides histograms with two horizontal axis containing z-values and the true scale of the variable. The function TeachHistDens() provides a density histogram with two axis. TeachHistCounts() and TeachHistRelFreq() are variations for count and relative frequency histograms, respectively. TeachConfInterv() and TeachHypTest() help instructors to visualize confidence levels and the results of hypothesis tests.

Functions

- TeachHistDens provides a density histogram with two axis. One axis for the data and one for the z-value.

- TeachHistCounts is a variation of TeachHistDens() for count histograms.

- TeachHistRelFreq is a variation of TeachHistDens() for relative frequency histograms.

- TeachHistConfInterv helps to visualize confidence levels

- TeachHistHypTest helps to visualize the results of hypothesis tests.

Author and Maintainer

Carsten Lange (clange@cpp.edu)

TeachHistConfInterv	*Histogram to Visualize Confidence Intervalls*

Description

The TeachHistConfInterv function is a variation of the TeachHistDens function of this package. It generates a histogram with two horizontal axis. One axis displays z-values the other one the dimension of the displayed variable and it is optimized to visualize confidence intervals.

Usage

```
TeachHistConfInterv(
  SampleMean = 0,
  StandardError = 1,
  DegreeFreedom = NULL,
  Confidence = 0.95,
  TestType = "BothTails",
  IsSdEstimated = FALSE,
  NOfSimData = 10000,
  XAxisMax = 5,
  AxisFontSize = 10,
  LabelSize = 3,
  BinWidth = 0.5,
  PrintDensities = FALSE,
  PrintRelFreq = TRUE,
  PlotNormCurv = TRUE,
  RandVarName = "Means",
  SeedValue = NULL,
  Mean = NULL
)
```

Arguments

SampleMean	(default=0) Value for sample mean.
StandardError	(default=1) Standard error from sample.
DegreeFreedom	(default=NULL) required for t-test (when IsSdEstimated=TRUE)
Confidence	(default=0.95) Confidence level.
TestType	(default="BothTails") User can also choose TestType="LeftTail" or TestType="RightTail"
IsSdEstimated	(default=FALSE) standard deviation or standard error are known. Set to FALSE, if estimated.
NOfSimData	(default=10000) Number of simulated data points. Higher NOfSimData results in better precision but possibly not all points are considered.
XAxisMax	(default=5) X-axis ranges from negative to positive XAxisMax. If XAxisMax is not high enough, some of the simulated points might not be considered.
AxisFontSize	(default=10) Font size for axis labels.
LabelSize	(default=3) Size of the labels inside the diagram.
BinWidth	(default= 0.5) BinWidth expressed in standard deviations.
PrintDensities	(default=FALSE) Set PrintDensities=TRUE to print densities in the histogram.
PrintRelFreq	(default=TRUE) Set PrintRelFreq=FALSE to suppress the printing of relative frequencies in the histogram.
PlotNormCurv	(default=TRUE) If TRUE normal curve is plotted
RandVarName	(default="Means") Variable name that is plotted in histogram at x-axis.
SeedValue	(default=NULL) Can be any integer and makes simulated points reproducible.
Mean	(default=NULL) Value for sample mean. Depreciated. Use SampleMean instead.

Details

The TeachHistConfInterv function supports confidence intervals based on the t-distribution (IsS-dEstimated=TRUE) and the z-distribution (IsSdEstimated=FALSE).

Value

Returns a histogram with confidence interval

Examples

```
TeachHistConfInterv()
TeachHistConfInterv(SampleMean=2, StandardError=0.2,  IsSdEstimated = FALSE)
TeachHistConfInterv(SampleMean=2, StandardError=0.2, DegreeFreedom = 35,  IsSdEstimated = TRUE)
```

TeachHistCounts   *Build a Histogram for Counts with Regular and z-Value Axis*

Description

#' The TeachCounts function is a variation of the TeachHistDens function of this package. Teach-HistCounts generates a histogram based on counts with two horizontal axis. One axis displays z-values the other one the dimension of the displayed variable.

Usage

```
TeachHistCounts(
  Mean = 0,
  Sd = 1,
  PlotData = dplyr::tibble(x = c(0)),
  VLine1 = NULL,
  VLine2 = NULL,
  BinWidth = 1,
  XAxisMax = 5,
  PrintZAxis = TRUE,
  AxisFontSize = 10,
  LabelSize = 3,
  NOfSimData = 1000,
  SeedValue = NULL,
  PrintCounts = TRUE
)
```

Arguments

Mean	(default=0) Mean for simulated data (ignored if PlotData is set).
Sd	(default=1) Standard Deviation for simulated data (ignored if PlotData is set).

PlotData	If PlotData is set to a Tibble with one column, the histogram will be based on the provided data. If PlotData is not set, the data for the histogram will be simulated based on the settings for Mean and Sd.
VLine1	(default= NA) A value for the simulated variable resulting in a vertical line at related x-axis position (can be se by user. E.g. lower boundary of conf intervall).
VLine2	(default=NA) A value for the simulated variable resulting in a vertical line at related x-axis position (can be se by user. E.g. upper boundary of conf intervall).
BinWidth	(default=1) Bin width expressed in standard deviations.
XAxisMax	(default=5) X-axis ranges from negative to positive XAxisMax. If XAxisMax is not high enough, some of the simulated points might not be considered.
PrintZAxis	(default=TRUE) TRUE adds a second axis with z values. FALSE suppresses z axis and plots only axis with original values.
AxisFontSize	(default=10) Font size for axis labels.
LabelSize	(default=3) Size of the labels inside the diagram.
NOfSimData	(default=1000) Number of simulated data points. Higher NOfSimData results in better precision but possibly not all points are considered.
SeedValue	(default=NULL) Can be any integer and makes simulated points reproducible.
PrintCounts	(default=TRUE) Set PrintCounts=FALSE to suppress the printing of counts in the histogram.

Details

There are two ways to generate the histogram.: 1) Simulated Data: Provide Mean and Standard Deviation and the histogram will be generated based on simulated data. 2) Provide Data as a Data Frame (Tibble) Provide data (a Tibble with one column) and a histogram will be generated based on the provided data. Histogram Design: The Histogram will have two axis. One axis is for the simulated random variable the other is the related Z-value (deviation from mean in standard deviations). The user can also define up to two reference lines that will be plotted as vertical lines in the plot (see VLine1 and VLine2).

Value

Returns a histogram

Examples

```
TeachHistCounts()
TeachHistCounts(70,2)
MyTibbleWithOneColumn=dplyr::tibble(x=rnorm(2000,20,2))
TeachHistCounts(PlotData=MyTibbleWithOneColumn)
```

TeachHistDens   *Build a Histogram with Regular and z-Value Axis*

Description

The TeachHistDens function is the main function of this package. It generates a density histogram
with two horizontal axis. One axis displays z-values the other one the dimension of the displayed
variable. Relative Frequencies are shown in the midde of each bar.

Usage

```
TeachHistDens(
  Mean = 0,
  Sd = 1,
  PlotData = dplyr::tibble(x = c(0)),
  VLine1 = NULL,
  VLine2 = NULL,
  BinWidth = 1,
  XAxisMax = 5,
  PrintZAxis = TRUE,
  AxisFontSize = 10,
  LabelSize = 3,
  PlotNormCurv = TRUE,
  NOfSimData = 1000,
  SeedValue = NULL,
  PrintDensities = FALSE,
  PrintRelFreq = TRUE
)
```

Arguments

Mean	(default=0) Mean for simulated data (ignored if PlotData is set).
Sd	(default=1) Standard Deviation for simulated data (ignored if PlotData is set).
PlotData	If PlotData is set to a Tibble with one column, the histogram will be based on the provided data. If PlotData is not set, the data for the histogram will be simulated based on the settings for Mean and Sd.
VLine1	(default= NA) A value for the simulated variable resulting in a vertical line at related x-axis position (can be se by user. E.g. lower boundary of conf intervall).
VLine2	(default=NA) A value for the simulated variable resulting in a vertical line at related x-axis position (can be se by user. E.g. upper boundary of conf intervall).
BinWidth	(default=1) Bin width expressed in standard deviations.
XAxisMax	(default=5) X-axis ranges from negative to positive XAxisMax. If XAxisMax is not high enough, some of the simulated points might not be considered.
PrintZAxis	(default=TRUE) TRUE adds a second axis with z values. FALSE suppresses z axis and plots only axis with original values.

AxisFontSize	(default=10) Font size for axis labels.
LabelSize	(default=3) Size of the labels inside the diagram.
PlotNormCurv	(default=TRUE) If TRUE normal curve is plotted.
NOfSimData	(default=1000) Number of simulated data points. Higher NOfSimData results in better precision but possibly not all points are considered.
SeedValue	(default=NULL) Can be any integer and makes simulated points reproducible.
PrintDensities	(default=FALSE) Set PrintDensities=TRUE to print densities in the histogram.
PrintRelFreq	(default=TRUE) Set PrintRelFreq=FALSE to suppress the printing of relative frequencies in the histogram.

Details

There are two ways to generate the histogram.: 1) Simulated Data: Provide Mean and Standard Deviation and the histogram will be generated based on simulated data. 2) Provide Data as a Data Frame (Tibble) Provide data (a Tibble with one column) and a histogram will be generated based on the provided data. Histogram Design: The Histogram will have two axis. One axis is for the simulated random variable the other is the related Z-value (deviation from mean in standard deviations). The user can also define up to two reference lines that will be plotted as vertical lines in the plot (see VLine1 and VLine2).

Value

Returns a histogram

Examples

```
TeachHistDens()
TeachHistDens(70,2)
MyTibbleWithOneColumn=dplyr::tibble(x=rnorm(2000,20,2))
TeachHistDens(PlotData=MyTibbleWithOneColumn)
```

TeachHistHypTest	*Histogram to Visualize Hypothesis Tests*

Description

The TeachHistHypTest function is a variation of the TeachHistDens function of this package. It generates a histogram with two horizontal axis. One axis displays z-values the other one the dimension of the displayed variable and it is optimized to visualize hypothesis testing.

Usage

```
TeachHistHypTest(
  NullHyp,
  StandardError,
  SampleMean,
  DegreeFreedom = NULL,
  Confidence = 0.95,
  TestType = "BothTails",
  IsSdEstimated = FALSE,
  NOfSimData = 10000,
  XAxisMax = 5,
  AxisFontSize = 10,
  LabelSize = 3,
  BinWidth = 0.5,
  PrintDensities = FALSE,
  PrintRelFreq = TRUE,
  PlotNormCurv = TRUE,
  RandVarName = "Means",
  SeedValue = NULL
)
```

Arguments

NullHyp	required! Value for null hypotheses.
StandardError	required!
SampleMean	required! SampleMean is the mean from the research sample to support the alternative/research hypotheses.
DegreeFreedom	required for t-test (IsSdEstimated=TRUE)
Confidence	(default=0.950 Confidence level.
TestType	(default="BothTails") User can also choose TestType="LeftTail" or TestType="RightTail"
IsSdEstimated	(default=FALSE) standard deviation or standard error are known. Set to FALSE, if estimated.
NOfSimData	(default=10000) Number of simulated data points. Higher NOfSimData results in better precision but possibly not all points are considered.
XAxisMax	(default=5) X-axis ranges from negative to positive XAxisMax. If XAxisMax is not high enough, some of the simulated points might not be considered.
AxisFontSize	(default=10) Font size for axis labels.
LabelSize	(default=3) Size of the labels inside the diagram.
BinWidth	(default= 0.51) BinWidth expressed in standard deviations.
PrintDensities	(default=FALSE) Set PrintDensities=TRUE to print densities in the histogram.
PrintRelFreq	(default=TRUE) Set PrintRelFreq=FALSE to suppress the printing of relative frequencies in the histogram.
PlotNormCurv	(default=TRUE) If TRUE normal curve is plotted
RandVarName	(default="Means") Variable name that is plotted in histogram at x-axis.
SeedValue	(default=NULL) Can be any integer and makes simulated points reproducible.

Details

The TeachHistHypTest function upports t-test (IsSdEstimated=TRUE) and z-test (IsSdEstimated=FALSE). It also supports TestType="BothTails", TestType="LeftTail", and TestType="RightTail".

Value

Returns a histogram

Examples

```
TeachHistHypTest(70,2,77)
TeachHistHypTest(70,2,77, DegreeFreedom = 35,  IsSdEstimated = TRUE)
TeachHistHypTest(70,2,65, DegreeFreedom = 35,  IsSdEstimated = TRUE, TestType = "LeftTail")
```

TeachHistRelFreq	*Build a Histogram for Relative Frequencies with Regular and z-Value Axis*

Description

The TeachHistRelFreq function is a variation of the TeachHistDens function of this package. TeachHistRelFreq generates a relative frequency histogram with two horizontal axis. One axis displays z-values the other one the dimension of the displayed variable.

Usage

```
TeachHistRelFreq(
  Mean = 0,
  Sd = 1,
  PlotData = dplyr::tibble(x = c(0)),
  VLine1 = NULL,
  VLine2 = NULL,
  BinWidth = 1,
  XAxisMax = 5,
  PrintZAxis = TRUE,
  AxisFontSize = 10,
  LabelSize = 3,
  NOfSimData = 1000,
  SeedValue = NULL,
  PrintRelFreq = TRUE
)
```

Arguments

Mean	(default=0) Mean for simulated data (ignored if PlotData is set).
Sd	(default=1) Standard Deviation for simulated data (ignored if PlotData is set).

PlotData	If PlotData is set to a Tibble with one column, the histogram will be based on the provided data. If PlotData is not set, the data for the histogram will be simulated based on the settings for Mean and Sd.
VLine1	(default= NA) A value for the simulated variable resulting in a vertical line at related x-axis position (can be se by user. E.g. lower boundary of conf intervall).
VLine2	(default=NA) A value for the simulated variable resulting in a vertical line at related x-axis position (can be se by user. E.g. upper boundary of conf intervall).
BinWidth	(default=1) Bin width expressed in standard deviations.
XAxisMax	(default=5) X-axis ranges from negative to positive XAxisMax. If XAxisMax is not high enough, some of the simulated points might not be considered.
PrintZAxis	(default=TRUE) TRUE adds a second axis with z values. FALSE suppresses z axis and plots only axis with original values.
AxisFontSize	(default=10) Font size for axis labels.
LabelSize	(default=3) Size of the labels inside the diagram.
NOfSimData	(default=1000) Number of simulated data points. Higher NOfSimData results in better precision but possibly not all points are considered.
SeedValue	(default=NULL) Can be any integer and makes simulated points reproducible.
PrintRelFreq	(default=TRUE) Set PrintRelFreq=FALSE to suppress the printing of relative frequencies in the histogram.

Details

There are two ways to generate the histogram.: 1) Simulated Data: Provide Mean and Standard Deviation and the histogram will be generated based on simulated data. 2) Provide Data as a Data Frame (Tibble) Provide data (a Tibble with one column) and a histogram will be generated based on the provided data. Histogram Design: The Histogram will have two axis. One axis is for the simulated random variable the other is the related Z-value (deviation from mean in standard deviations). The user can also define up to two reference lines that will be plotted as vertical lines in the plot (see VLine1 and VLine2).

Value

Returns a histogram

Examples

```
TeachHistRelFreq()
TeachHistRelFreq(70,2)
MyTibbleWithOneColumn=dplyr::tibble(x=rnorm(2000,20,2))
TeachHistRelFreq(PlotData=MyTibbleWithOneColumn)
```

Index