Package ‘gvlma’

October 13, 2022

Type    Package
Title   Global Validation of Linear Models Assumptions
Version 1.0.0.3
Author  Edsel A. Pena <pena@stat.sc.edu> and Elizabeth H. Slate <eslate@fsu.edu>
Maintainer Elizabeth Slate <slate@stat.fsu.edu>
Depends R (>= 2.1.1)
License GPL
NeedsCompilation no
Repository CRAN
Date/Publication 2019-01-05 19:30:03 UTC

R topics documented:

gvlma-package .................................................. 2
CarMileageData .................................................. 3
deletion.gvlma .................................................. 4
display.delstats ............................................... 5
gvlma .......................................................... 6
plot.gvlma ..................................................... 8
plot.gvlmaDel ................................................. 10
summary.gvlma ............................................... 11
summary.gvlmaDel ............................................. 13
update.gvlma ................................................. 14

Index 16
Description

Perform a single global test to assess the linear model assumptions, as well as perform specific directional tests designed to detect skewness, kurtosis, a nonlinear link function, and heteroscedasticity.

Details

Package: gvlma
Type: Package
Version: 1.0
Date: 2006-06-07
License: GPL

The function gvlma will take either a linear models object or a formula and data set for a linear model (single response) and compute the global and directional tests for assessing modeling assumptions as described in the reference listed below. The function deletion.gvlma will compute the deletion (“leave-one-out”) global statistics described in that paper.

Author(s)

Slate, EH <slate@stat.fsu.edu> and Pena, EA <pena@stat.sc.edu>
Maintainer: Slate, EH <slate@stat.fsu.edu>

References


See Also

gvlma

Examples

```r
x1 <- rnorm(100,0,2)
x2 <- runif(100)
y <- 3*x1 -x2 + rnorm(100)
gvmodel <- gvlma(lm(y ~ x1 + x2))
plot(gvmodel)
summary(gvmodel)
gvmodel.del <- deletion.gvlma(gvmodel)
summary(gvmodel.del)
```
plot(gvmodel.del)

---

CarMileageData  Car Mileage Data Recorded at Each Gasoline Fill-Up

**Description**


**Usage**

```r
data(CarMileageData)
```

**Format**

A data frame with 205 observations on the following 7 variables.

- **Date**  Date of gasoline fill-up
- **Lag1Date**  Lagged gasoline fill-up date
- **NumDaysBetw**  Number of days since last gasoline fill-up
- **TotalMiles**  Current odometer reading
- **NumGallons**  Number of gallons to fill tank
- **MilesLastFill**  Miles driven since last fill-up
- **AveMilesGal**  Average miles per gallon achieved since last fill-up

**Details**

Many people routinely record data on automobile mileage performance at each gasoline fill-up. Prof. E. Pena generously contributed his data for this time period.

**Source**

These data were used in Example 1 of the publication “Global Validation of Linear Model Assumptions” by E. Pena and E. Slate, *Journal of the American Statistical Association, 101*(473):341-354, 2006. The data were recorded by Prof. E. Pena.

**Examples**

```r
data(CarMileageData)
plot(CarMileageData)
```
**deletion.gvlma**  
*Deletion Statistics for a Linear Model*

**Description**
Computes the deletion statistics (leave-one-out) for assessing unusual observations in a linear model.

**Usage**

```r
deletion.gvlma(gvlmaobj)
```

**Arguments**

- `gvlmaobj`  
  A `gvlma` object, as the result of a call to `gvlma`.

**Details**
Given a `gvlma` object, which contains in the component `GlobalTest` the test statistics and p-values for the global and directional tests to assess linear models assumptions, `deletion.gvlma` computes the leave-one-out global and directional statistics. The deletion statistics are reported as percent relative change from the corresponding statistic value based on the full data set.

**Value**
A dataframe is returned with variables `DeltaGlobalStat`, `GStatpvalue`, `DeltaStat1`, `Stat1pvalue`, `DeltaStat2`, `Stat2pvalue`, `DeltaStat3`, `Stat3pvalue`, `DeltaStat4`, and `Stat4pvalue`. Each “Delta” variable is the percent relative change in the statistic when the corresponding observation (row of the data frame) is dropped. Each “pvalue” variable is the p-value associated with the deletion statistic. (Note the p-value is NOT a change in the p-values for the full and leave-one-out statistic values.)

**Author(s)**
Slate, EH <slate@stat.fsu.edu> and Pena, EA <pena@stat.sc.edu>.

**References**

**See Also**
`gvlma`
Examples

```r
data(CarMileageData)
CarModelAssess <- gvlma(NumGallons ~ MilesLastFill, data = CarMileageData)
CarModelDel <- deletion.gvlma(CarModelAssess)
CarModelDel
display.delstats
```

---

**display.delstats**

*Plot Deletion Statistics and Their P-Values for Assessment of Unusual Observations*

**Description**

Creates a graph of the p-values associated with the deletion statistics versus the deletion statistics with unusual observations highlighted. This function is called by `plot.gvlmaDel`.

**Usage**

```r
display.delstats(deletedStatvals, deletedpvals, nsd = 3,
                 TukeyStyle = TRUE, statname = "G", pointlabels)
```

**Arguments**

- `deletedStatvals` The vector of deletion statistics, with i-th entry defined as the percent relative change in the global test statistic when the i-th observation is removed from the analysis.
- `deletedpvals` The vector of p-values associated with the global test statistics, with i-th entry being the p-value for the global test statistic with observation i removed.
- `nsd` Parameter that governs which observations are deemed unusual. When `TukeyStyle = TRUE`, “control limits” are drawn nsd times the interquartile range beyond the quartiles for both the `deletedStatvals` and `deletedpvals`. When `TukeyStyle = FALSE`, “control limits” are drawn at nsd standard deviations away from the sample means. Observations beyond these “control limits” are marked and labeled using the text in `pointlabels`, if provided (else by observation number).
- `TukeyStyle` Controls how unusual observations are determined. If `TukeyStyle = TRUE` (default), then unusual observations are farther than nsd times the interquartile range from the quartiles (in either of the `deletedStatvals` and `deletedpvals` directions). If `TukeyStyle = FALSE`, then unusual observations are farther than nsd times the sample standard deviation from the sample mean.
- `statname` A string used to label the `deletedStatvals` axis of the plot. If missing, this label is determined from the variable name passed as the `deletedStatvals` argument, if possible; otherwise defaults to "Deleted statistics."
- `pointlabels` Character vector of same length as `deletedStatvals` and `deletedpvals` used for labelling unusual observations.
Details

Generally `display.delstats` is not called directly, but rather by the function `plot.gvlmaDel`. Plots the `deletedpvals` versus the `deletedStatvals` and adds “control limits” determined by the parameters `nsd` and `TukeyStyle`. Points outside the “control limits” (in either the `deletedStatval` or `deletedpval`) are labeled as unusual.

Value

A dataframe consisting of the unusual observations with variables `deletedStatval` and `deletedpval`.

Author(s)

Slate, EH <slate@stat.fsu.edu> and Pena, EA <pena@stat.sc.edu>.

References


See Also

gvlma

Examples

data(CarMileageData)
CarMileageAssess <- gvlma(NumGallons ~ MilesLastFill + NumDaysBetw, data = CarMileageData)
CarMileageDel <- deletion.gvlma(CarMileageAssess)
plot(CarMileageDel)
display.delstats(CarMileageDel$DeltaGlobalStat, CarMileageDel$GStatpvalue)
display.delstats(CarMileageDel$DeltaStat1, CarMileageDel$Stat1pvalue)

gvlma Create a Gvlma Object

Description

Top-level function for Global Validation of Linear Models Assumptions.

Usage

gvlma(x, data, alphalevel = 0.05, timeseq, ...)
gvlma.form(formula, data, alphalevel = 0.05, timeseq = 1:nrow(data), ...)
gvlma.lm(lmobj, alphalevel = 0.05, timeseq)
Arguments

x Either a formula, in which case gvlma.form will be called, or a linear models object, in which case gvlma.lm will be invoked.

formula A linear models formula interpretable within the dataframe data. Should have a single reponse variable.

lmobj An object resulting from a call to lm.

data Required if x is a formula, ignored if x is an lm object. A dataframe in which the variables in the formula x can be interpreted.

alphalevel Level of significance at which to perform the global and directional tests for linear models assumptions.

timeseq A vector of length the number of observations in the linear model that gives a "time ordering" for the observations. This time sequence is used in the heteroscedasticity test statistic. Defaults to 1:n where n is the number of observations in the linear model.

... Additional arguments such as subset that are passed on to the call to lm when x is a formula. Note that weights, while being passed on to the call to lm, will not be used in any special way in the gvlma computations.

Details

gvlma is the top-level function to create a gvlma object for assessment of linear models assumptions.

Value

A gvlma object is returned. This is a list of class “gvlma” that contains all of the components returned by the call to lm for fitting the linear model, plus an additional component entitled “GlobalTest.” This new GlobalTest component is a list with the following components:

LevelOfSignificance The level of significance at which the decisions reported for the global and directional tests were made.

GlobalStat4 A list consisting of the components Value, pvalue and Decision containing the global test statistic value, associated p-value, and text phrase reporting the decision concerning appropriateness of the linear model assumptions.

DirectionalStat1 A list consisting of the Value, pvalue and Decision associated with the skewness directional test statistic.

DirectionalStat2 A list consisting of the Value, pvalue and Decision associated with the kurtosis directional test statistic.

DirectionalStat3 A list consisting of the Value, pvalue and Decision associated with the link function directional test statistic.

DirectionalStat4 A list consisting of the Value, pvalue and Decision associated with the heteroscedasticity directional test statistic.
timeseq  The ordering of the observations used when computing the heteroscedasticity directional statistic.
call     The call used to invoke gvlma.

Author(s)
Slate, EH <slate@stat.fsu.edu> and Pena, EA <pena@stat.sc.edu>.

References

See Also
plot.gvlma, deletion.gvlma, update.gvlma, lm

Examples
data(CarMileageData)
CarModelAssess <- gvlma(NumGallons ~ MilesLastFill + NumDaysBetw,
data = CarMileageData)
CarModelAssess
summary(CarModelAssess)
CarModel2 <- gvlma(lm(NumGallons ~ MilesLastFill + NumDaysBetw,
data = CarMileageData))
CarModel2
summary(CarModel2)
plot(CarModel2)

plot.gvlma  Various Plots for a Gvlma Object

Description
Diagnostic plots for a single-response gvlma linear model.

Usage
## S3 method for class 'gvlma'
plot(x, onepage = TRUE, ask = !onepage && prod(par("mfcol")) <
   ncol(model.matrix(x)) + 4 && dev.interactive(), ...)

Arguments
x         A gvlmaobj object.
onepage    If TRUE, all plots will be displayed in one page of graphs.
ask        If TRUE, user will be prompted before plots begin a new page.
...        Additional arguments that are ignored.
Details

A series of plots is generated for diagnostic assessment of a linear model for a single response variable. The plots are similar to those generated by \texttt{plot.lm}. The plots are (a) the response versus each of the predictors in the model, (b) the response versus the time sequence in the gvlma object (\texttt{gvlmaobj$GlobalTest$timeseq}), which is the time sequence used for computing the directional test statistic $S_2^2$, (c) the standardized residuals vs the fitted values, (d) a histogram of the standardized residuals, (e) a normal probability plot of the standardized residuals, and (f) a plot of the standardized residuals versus the time sequence.

Note that the standardized residuals here are computed as the raw residuals divided by the MLE of the error standard deviation (i.e. $\sqrt{\text{SSE}/n}$).

Value

No value is returned.

Note

The standardized residuals here are computed as the raw residuals divided by the MLE of the error standard deviation (i.e. $\sqrt{\text{SSE}/n}$).

Author(s)

Slate, EH <slate@stat.fsu.edu> and Pena, EA <pena@stat.sc.edu>.

References


See Also

gvlma

Examples

data(CarMileageData)
CarModelAssess <- gvlma(NumGallons ~ MilesLastFill + NumDaysBetw, data = CarMileageData)
plot(CarModelAssess)
par(mfrow=c(2,2))
plot(CarModelAssess, onepage = FALSE)
plot.gvlmaDel

Various Plots for a GvlmaDel Object

Description

Plots to display the behavior of the deletion statistics stored in a gvlmaDel object.

Usage

```r
## S3 method for class 'gvlmaDel'
plot(x, which = 1:2, TukeyStyle = TRUE, ask = prod(par("mfcol")) < max(c(10, 5)[which]) && dev.interactive(), pointlabels, ...)
```

Arguments

- `x`: A gvlmaDel object.
- `which`: Vector indicating which, or both, of two types of plots to show.
- `TukeyStyle`: If TRUE, determine unusual observations in a robust way based on inter-quartile ranges, else based on standard deviations.
- `ask`: If TRUE, prompt the user before beginning a new page of graphs.
- `pointlabels`: A vector of length the number of observations in the linear model fit in the gvlmaDel object containing character strings to be used as labels for unusual points.
- `...`: Additional arguments that are ignored.

Details

If `which = 1`, each of the 5 deletion statistics (deletion global statistic and each of the 4 directional statistics) is plotted against the time sequence used for the 4th directional statistic (assessing heteroscedasticity).

If `which = 2`, the function `display.delstats` is called for each of the 5 deletion statistics. The argument `TukeyStyle` is passed directly to `display.delstats`. See the help for `display.delstats` for details.

If `which = c(1, 2)`, the default, then all 10 plots are generated.

The deletion statistics in the gvlmaDel object are the percent relative change when each observation, in turn, is omitted from the model fitting.

Value

No value is returned.

Author(s)

Slate, EH <slate@stat.fsu.edu> and Pena, EA <pena@stat.sc.edu>.
summary.gvlma

Print Basic Information for a Gvlma Object

Description

Prints the basic information for a gvlma object, which is the output object from the function gvlma.

Usage

## S3 method for class 'gvlma'
summary(object, ...)
## S3 method for class 'gvlma'
print(x, ...)
display.gvlmatests(gvlmaobj)

Arguments

x, object, gvlmaobj

An object resulting from a call to gvlma. It is a list containing the components of a call to lm plus an item with the name GlobalTest.

... Additional arguments that are passed to summary.lm.

References


See Also

gvlma, deletion.gvlma

Examples

data(CarMileageData)
CarModelAssess <- gvlma(NumGallons ~ MilesLastFill + NumDaysBetw, data = CarMileageData)
CarModelDel <- deletion.gvlma(CarModelAssess)
par(mfrow=c(1,1))
plot(CarModelDel)
par(mfrow=c(2,2))
plot(CarModelDel)
plot(CarModelDel, TukeyStyle = FALSE)
plot(CarModelDel, which = 2)
summary.gvlma

Details

print.gvlma invokes print on the lm object and then calls display.gvlmatests.

summary.gvlma invokes summary on the lm object with the additional...arguments and then calls display.gvlmatests.

display.gvlmatests provides the test statistics, p-values and decision (whether linear models assumptions are satisfied) for the global and directional tests associated with the gvlma object. The decision is reported at the level of significance used when the gvlma object was created. See the argument alphalevel to gvlma.

Value

The value returned invisibly is a dataframe with row names indicating the global test and the 4 directional tests. Variables are

- Value: Value of the test statistic.
- p-value: p-value associated with the test.
- Decision: Text string indicating whether the test statistic is significant at the significance level specified in the original call to gvlma.

Author(s)

Slate, EH <slate@stat.fsu.edu> and Pena, EA <pena@stat.sc.edu>.

References


See Also

gvlma, display.gvlmatests, summary.lm

Examples

data(CarMileageData)
CarModelAssess <- gvlma(NumGallons ~ MilesLastFill, data = CarMileageData)
CarModelAssess
summary(CarModelAssess)
Description

Summarize the test statistic values and p-values for assessing unusual observations using the global and directional test statistics that were computed in a gvlmaDel object resulting from a call to deletion.gvlma.

Usage

```r
## S3 method for class 'gvlmaDel'
summary(object, allstats = TRUE, ...)
## S3 method for class 'gvlmaDel'
print(x, ...)
```

Arguments

- `object, x` Object resulting from a call to deletion.gvlma, which takes a gvlma object and performs the leave-one-out analyses for assessment of the influence of each observation on the global and directional tests for linear model assumptions.
- `allstats` For summary.gvlmaDel, if allstats = TRUE (the default), then the summary statistics are provided for global test and all 4 directional test statistics. If summary.gvlmaDel is FALSE, then the summary is provided for the deletion global test statistics only.
- `...` Additional arguments that are ignored.

Details

The summary values are the min, first quartile, median, average, 3rd quartile and maximum of the deletion test statistic values and p-values. Additionally, observations and the corresponding deletion test statistic values and p-values for which the deletion test statistic value or its p-value is outside the outer fences (Q1 - 3*IQR, Q3 + 3*IQR) of the set of deletion statistics are reported.

print.gvlmaDel simply invokes summary.gvlmaDel with allstats = TRUE.

Value

A dataframe of dimension nobs x 5 is returned invisibly, where nobs is the number of observations in the linear model fit. The 5 columns are named DeltaGlobalStat, DeltaStat1, DeltaStat2, DeltaStat3, and DeltaStat4, indicating the deletion global test and the four deletion directional test statistics. Each entry in the dataframe is TRUE/FALSE, indicating whether the corresponding test statistic was unusual (i.e. beyond the outer fences) with respect to either its value or its p-value.

Author(s)

Slate, EH <slate@stat.fsu.edu> and Pena, EA <pena@stat.sc.edu>.
References


See Also

gvlma, deletion.gvlma

Examples

data(CarMileageData)
CarModelAssess <- gvlma(NumGallons ~ MilesLastFill, data = CarMileageData)
CarModelAssess
CarModelDel <- deletion.gvlma(CarModelAssess)
CarModelDel
summary(CarModelDel)
summary(CarModelDel, allstats = FALSE)

---

update.gvlma  
*Update a Gvlma Object*

Description

Update a gvlma object with changes to the linear model, the level of significance for global tests, or the time sequence used for the heteroscedasticity directional test.

Usage

```r
## S3 method for class 'gvlma'
update(object, formula, ...)
```

Arguments

- `object`  
  A gvlma object resulting from a call to gvlma.
- `formula`  
  (optional) A new formula describing the underlying linear model.
- `...`  
  Additional arguments to be changed from the original call to gvlma. These may include arguments to the lm function, such as subset, as well as the gvlma-specific arguments alphalevel and timeseq. Internal note: The function deletion.gvlma passes the argument warn = FALSE to suppress warnings about a time sequence of incorrect length.

Details

All arguments other than alphalevel and timeseq (and warn) are passed on to a call to update for the underlying linear model.

If alphalevel is specified, then subsequent displays of the global and directional test statistic decisions will be based on the new level of significance. If timeseq is specified, then the heteroscedasticity direction test, $S_4^*$, will be updated to use the new time sequence.
**Value**

A new gvlma object is returned.

**Author(s)**

Slate, EH <slate@stat.fsu.edu> and Pena, EA <pena@stat.sc.edu>.

**References**


**See Also**

gvlma.update.default

**Examples**

data(CarMileageData)
CarModelAssess <- gvlma(NumGallons ~ MilesLastFill + NumDaysBetw, 
data = CarMileageData)
CarModelAssess
summary(CarModelAssess)
CarModelNew <- update(CarModelAssess, alphalevel = 0.01)
CarModelNew
CarModelNew <- update(CarModelAssess, subset = -(1:10))
CarModelNew
summary(CarModelNew)
Index

* datasets
  CarMileageData, 3

* graphs
  display.delstats, 5
  plot.gvlma, 8
  plot.gvlmaDel, 10

* htest
  gvlma-package, 2

* methods
  update.gvlma, 14

* print
  summary.gvlma, 11
  summary.gvlmaDel, 13

* regression
  deletion.gvlma, 4
  gvlma, 6

CarMileageData, 3

destruction.gvlma, 4, 8, 11, 14
display.delstats, 5, 10
display.gvlmatests, 12
display.gvlmatests (summary.gvlma), 11

gvlma, 2, 4, 6, 9, 11, 12, 14, 15
gvlma-package, 2

lm, 8

plot.gvlma, 8, 8
plot.gvlmaDel, 10
print.gvlma (summary.gvlma), 11
print.gvlmaDel (summary.gvlmaDel), 13

summary.gvlma, 11
summary.gvlmaDel, 13
summary.lm, 12

update.default, 15
update.gvlma, 8, 14