Package 'CADStat'

May 10, 2017

Version 3.0.8

Date 2017-05-09

Title Provides a GUI to Several Statistical Methods

Author Lester Yuan, Tom Stockton, Doug Bronson, Pasha Minallah, and Mark Fitzgerald

Maintainer David Rebhuhn <rebhuhn.david@epa.gov>

Depends R (>= 2.15.0), JGR (>= 1.7-10)

Imports rJava, XML, lattice, bio.infer, gdata, gmodels, quantreg, rpart, JavaGD, car, MASS,

Description Using Java GUI for R (JGR), CADStat provides a user interface for several statistical methods - scatterplot, boxplot, linear regression, generalized linear regression, quantile regression, conditional probability calculations, and regression trees.

License GPL (>= 2)

URL http://www.rforge.net/CADStat/

LazyData true

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2017-05-10 02:09:22 UTC

R topics documented:

1

---

bioinfer.JGR   *Dialog box for predicting env. conditions from biological observations*

Description

This module estimates environmental conditions from observations of benthic macroinvertebrates for a set of sites, by combining observations of each particular taxon with a taxon-environment relationship.

This calculation has three steps. First, observations of benthic macroinvertebrates are matched with standard taxonomic names, with a full taxonomic hierarchy. Second, operational taxonomic units (OTUs) are defined for the data set, so that each individual taxon is counted at a consistent taxonomic level. Third, observations are matched with existing taxon-environment relationships and an estimate of environmental conditions computed.

Details

User Interface

Select *Analysis Tools -> Predicting env. conditions from biological observations* from the menus.

The Data Directory can be specified to indicate the default location that you wish to search for input files and write output files. Unless the browse buttons for individual files are used to specify other directories, all input and output comes from this directory. Note: Changing the data directory does not change the working directory for R, only for the default locations of the input/output files for this module.

There are three options for starting points for running this module. By default, the entire process should be run. However, since the first two steps of the process may require some feedback, these steps can be skipped in subsequent runs to reduce user effort.

Run Entire Process: This option is the default setting. It begins with the benthic macroinvertebrate count file and runs all steps required for: merging taxonomy with a standard taxonomic hierarchy from the Integrated Taxonomic Information System (ITIS), generating OTUs corresponding to the coefficient file, and computing maximum likelihood estimates of environmental conditions. Output files include an merged taxonomic file, an OTU file, and a maximum likelihood file.

Start with Merge Taxonomic Merge File This option assumes that the tool has been run previously to merge taxonomy of the benthic macroinvertebrate count file with a standard taxonomic hierarchy. The benthic macroinvertebrate file is not necessary for input, as the merged taxonomic file is read as input for the starting point. Output files include an OTU file and a maximum likelihood file. This option is recommended only for experienced users.

Start with Operational Taxonomy File This option assumes that the tool has been run previously to merge taxonomy of the benthic macroinvertebrate count file with a standard taxonomic hierarchy and to generate OTUs. The benthic macroinvertebrate file and the taxonomic merge file are not necessary for input, as the OTU file is read as input for the starting point. Only one output file is generated: the maximum likelihood file. This option is recommended only for experienced users.

Input files

Coefficients file: This file is an R data file (e.g., .rda or .rdata) that contains the taxon-environment relationships. At the present time only one coefficient file is available: coef.west.wt.rda, which provides relationships for stream temperature and percent substrate sand/fines for macroinvertebrates of western U.S. streams. Select this file from the default directory that appears when the Browse button is selected.

Benthic Count File: This file contains the observations of benthic macroinvertebrate. The file should be supplied as tab-delimited text. Three columns in the following order are required: (1) Sample identifier, (2) taxon name, and (3) taxon abundance.

Taxonomic Merge File: If Run Entire Process is chosen, this file is the result of the merge between local macroinvertebrate observations and a standard taxonomic hierarchy. If **Start with Merge Taxonomic Merge File** is chosen, this file is the starting input file. The format of the file is tab-delimited text.

Operational Taxonomy File: If **Start with Operational Taxonomy File** is chosen, it is the starting input file. See above.

Output file

Maximum likelihood File: This file is the saved output of this module. It is a tab-delimited text file, with at least three columns: 1) Site name; 2) Maximum likelihood estimate of environmental conditions; 3) An indicator of whether or not multiple maximum likelihood estimates were found. **TRUE** here indicates that global convergence was not attained. The number of columns may increase, depending on how many environmental parameters are estimated simultaneously.

See Also

`get.taxonomic`, `get.otu`, `mlsolve`, `coef.west.wt`

---

bioinfer.new.JGR   *Dialog box for predicting env. conditions from biological observations*

Description

This module estimates environmental conditions from observations of benthic macroinvertebrates for a set of sites, by combining observations of each particular taxon with a taxon-environment relationship.

This calculation has three steps. First, observations of benthic macroinvertebrates are matched with standard taxonomic names, with a full taxonomic hierarchy. Second, operational taxonomic units (OTUs) are defined for the data set, so that each individual taxon is counted at a consistent taxonomic level. Third, observations are matched with existing taxon-environment relationships and an estimate of environmental conditions computed.

Details

User Interface

Select *Analysis Tools -> Predicting env. conditions from biological observations* from the menus.

The Data Directory can be specified to indicate the default location that you wish to search for input files and write output files. Unless the browse buttons for individual files are used to specify other directories, all input and output comes from this directory. Note: Changing the data directory does not change the working directory for R, only for the default locations of the input/output files for this module.

There are three options for starting points for running this module. By default, the entire process should be run. However, since the first two steps of the process may require some feedback, these steps can be skipped in subsequent runs to reduce user effort.

Run Entire Process: This option is the default setting. It begins with the benthic macroinvertebrate count file and runs all steps required for: merging taxonomy with a standard taxonomic hierarchy

from the Integrated Taxonomic Information System (ITIS), generating OTUs corresponding to the coefficient file, and computing maximum likelihood estimates of environmental conditions. Output files include an merged taxonomic file, an OTU file, and a maximum likelihood file.

Start with Merge Taxonomic Merge File This option assumes that the tool has been run previously to merge taxonomy of the benthic macroinvertebrate count file with a standard taxonomic hierarchy. The benthic macroinvertebrate file is not necessary for input, as the merged taxonomic file is read as input for the starting point. Output files include an OTU file and a maximum likelihood file. This option is recommended only for experienced users.

Start with Operational Taxonomy File This option assumes that the tool has been run previously to merge taxonomy of the benthic macroinvertebrate count file with a standard taxonomic hierarchy and to generate OTUs. The benthic macroinvertebrate file and the taxonomic merge file are not necessary for input, as the OTU file is read as input for the starting point. Only one output file is generated: the maximum likelihood file. This option is recommended only for experienced users.

Input files

Coefficients file: This file is an R data file (e.g., .rda or .rdata) that contains the taxon-environment relationships. At the present time only one coefficient file is available: coef.west.wt.rda, which provides relationships for stream temperature and percent substrate sand/fines for macroinvertebrates of western U.S. streams. Select this file from the default directory that appears when the Browse button is selected.

Benthic Count File: This file contains the observations of benthic macroinvertebrate. The file should be supplied as tab-delimited text. Three columns in the following order are required: (1) Sample identifier, (2) taxon name, and (3) taxon abundance.

Taxonomic Merge File: If Run Entire Process is chosen, this file is the result of the merge between local macroinvertebrate observations and a standard taxonomic hierarchy. If **Start with Merge Taxonomic Merge File** is chosen, this file is the starting input file. The format of the file is tab-delimited text.

Operational Taxonomy File: If **Start with Operational Taxonomy File** is chosen, it is the starting input file. See above.

Output file

Maximum likelihood File: This file is the saved output of this module. It is a tab-delimited text file, with at least three columns: 1) Site name; 2) Maximum likelihood estimate of environmental conditions; 3) An indicator of whether or not multiple maximum likelihood estimates were found. **TRUE** here indicates that global convergence was not attained. The number of columns may increase, depending on how many environmental parameters are estimated simultaneously.

See Also

get.taxonomic, get.otu, mlsolve, coef.west.wt

---

browseHelp *Launches help in the browser*

Description

Internal function

| buildresultsXML | *Prepare R output for XML presentation* |

Description

Internal function

| bxplot.JGR | *Dialog box for generating boxplots for data with factor subsets* |

Description

Creates boxplots for chosen subsets of the data. A boxplot displays distributional charactersitics (location and shape) of the data. Boxplots also provide a qualitative means of comparing several data sets.

Details

User Interface

Select *Graph -> Boxplot* from the menus. A dialog box will open.

Select the data set of interest from the pull-down menu, or browse for a tab-delimited text file.

Select the variable that you wish to plot from the pull-down menu, **Result**. By default, the module will produce two plots, one in the original units of Result, and one in log transformed units.

Text indicating sample size of the data set is included in the plot if **Sample Sizes** is selected. You can alter the labels for the plot by modifying **Plot Title** and **Result-Axis**.

By default, the labels on the x-axis are printed horizontally, but can be changed to vertical with **Rotate X Axis Labels**.

| CADStat.help | *Launch browser help files for CADStat* |

Description

Internal function

| CADStat.update | *Update CADStat files* |

Description

Internal function

---

cleanCADStatWorkspace   *Clean temporary files from CADStat workspace*

Description

Internal function

---

conditionalprob.JGR   *Dialog box for computing and plotting conditional probabilities*

Description

Computes and plots conditional probabilities of an event (e.g., an undesirable biological condition) given values of a another variable (e.g., a stressor).

Details

A conditional probability plot provides a graphical representation of the impact one variable can have on the probability of another occuring. One way to assess whether the probability of an event of interest occuring changes for different values of an explanatory variable is to consider how susets of the response change as a function of the explanatory variable.

The first variable is the response. It is this column that subsets will be taken from based on a user-specified conditioanl value. The condfitional value must be between the maximum and minimum values of the response variable. The second variable is the explanatory variable.

It is often the case in biological applications, that thresholds exist. Conditional probability plots are useful for qualitatively identifying thresholds when they may exist.Conditional probability plots are often used in the development of tolerance values. Once the empirical curve for conditional probability is generated, possible threshold levels for eliciting impact can be identified for eventual use in the development of stressor criteria. This threshold of impact is determined by the identification of a changepoint that separates the empirical curve into two groups so that the probability of impact is different for samples above and below the changepoint. There are three potential methods for the identification of thresholds of impact for the conditional probability empirical curves. These include: 1) change in curvature of fitted curve, 2) nonparametric deviance reduction, and 3)non-overlapping confidence intervals (CIS). For a detailed explanation of these methods see Paul and McDonald (Journal of the American Water Resources Association 41(5) p. 1211-1223 2005).

Once thresholds have been identified, it may be of interest to utilize additional analyses that will quantify the relationship between the subsetted data. Specifically, one may choose to compute the correlation between the variables in the conditional dataset or if a threshold exists it may be of interest to generate a regression line for the sets of probabilities that are both above and below a threshold.

User Interface

Select *Analysis Tools -> Conditional Probability* from the menus. A dialog box will open. Select the data set of interest from the pull-down menu, or browse for a tab-delimited text file. The Data

Subsetting tab can be used to select a subset of the data file by choosing a variable from the pull down menu and then selecting the levels of that variable to include. You can hold down the <CTRL> key to select several levels.

Select **Stressor** and **Response** variables from the pull-down menu. If each sample in your data set represents a different proportion of your population (e.g., data collected using a stratified random sampling design), you can Specify a **weighting variable** that provides the relative weight for each sample.

Assign a **Response Cutoff Value**. This value defines the threshold for defining your response of interest. For example, if your response is an IBI, the Response Cutoff Value would be the criterion value for your IBI below which the site is assess as biologically impaired. Selection of the **Cutoff Value Direction** depends on whether you expect the value of your response to increase or decrease with stress. For IBI, which decreases with stress, **Cutoff Value Direction** would be Less Than.

Probability Direction defines the direction in which you would like to assess the stressor level. For a stressor whose severity increases with increasing values (e.g., percent fines), you should select **Greater Than or Equal**. Other stressors (e.g., pH) increase in severity as their value decreases, in which case you would select **Less Than or Equal**.

Selecting **Bootstrap confidence intervals** directs the tool to repeatedly resample the data set to estimate confidence intervals on the estimate relationship.

You may change the axes labels and plot title by typing them in the **Plot Labels** dialog.

The output is a graph of the fitted probability (that the dependent variable is less than the conditional value) versus the independent variable.

---

condprob.fn   *Working function to generate conditional probability values*

Description

Internal function

---

condprob.impute   *End function for conditional probability*

Description

Internal function

`cor.JGR`   *Dialog box for computing a correlation matrix*

Description

Computes a correlation matrix and confidence intervals for the chosen variables. Possible computational methods are Pearson's r, Kendall's tau, and Spearman's rho.

Details

Correlation coefficients measure the strength of the linear relationship between two variables. This module produces a correlation matrix, which presents the correlation coefficients for several variables, along with confidence intervals.

Correlation coefficients are unitless and range between -1 and 1. A positive correlation implies as one variable increases, the other increases. A negative correlation implies as one variable increases, the other decreases. The closer the correlation is to -1 or 1, the stronger the relationship. If the correlation is near 0, then the variables are not related *linearly*. In all cases, a scatterplot should be inspected to help determine if a linear model is appropriate for the variables.

Estimation

This module offers 3 methods for computing a correlation coefficient. One is parametric, Pearson's r, and two are nonparametric, Kendall's tau and Spearman's rho.

Pearson's r is equal to the covariance of the two variables divided by the square root of the product of the variances. Since Pearson's r is based upon variances and covariances, it is sensitive to outliers. The nonparametric methods are based upon data ranks rather than data values so they are not sensitive to outliers. Also, Spearman's rho may be a better indicator of a non-linear relationship.

Plots

The user may select to have a scatterplot matrix created. This is a matrix of scatterplots between all pairs of variables. These plots are useful in determining the type and strength of the relationship between two variables.

User Interface

To perform a correlation analysis, select *Analysis Tools -> Correlation Analysis* from the menu. A dialog box will appear that allows the user to enter analysis parameters:

Provide the variables (column names) of interest.

Click the **Method** radio button to indicate the desired method.

Click the **Create Scatterplot Matrix** check box to indicate whether a scatterplot matrix should be produced.

Click the **Compute Confidence Intervals** check box to indicate whether confidence intervals for the correlation between all variable pairs should be computed.

If confidence intervals are desired, then enter the confidence level into the **Confidence Level** text box. The confidence level must be between 0 and 1.

datamerge.JGR	*Merge data dialog for JGR*

Description

Internal function

docbook	*General function for preparing output*

Description

Internal function

docbook.data.frame	*Prepare summary output from a data frame*

Description

Internal function

docbook.default	*General function for preparing output*

Description

Internal function

docbook.glm	*Prepare output from glm*

Description

Internal function

---

docbook.glm.summary *Prepare summary output from glm*

Description

Internal function

---

docbook.hmtestp *Prepare summary output from hmtest*

Description

Internal function

---

docbook.htest *Prepare summary output from htest*

Description

Internal function

---

docbook.lm *Prepare output from lm*

Description

Internal function

---

docbook.lm.summary *Prepare summary output from lm*

Description

Internal function

| docbook.try | *Try a docbook* |

Description

Internal function

| envdata.or | *Environmental data from western OR* |

Description

Environmental data for stream sampled in western Oregon.

Format

A data frame with 245 observations on the following 13 variables.

STRM.ID Unique sample code

jday Sampling day

lon Longitude in decimal degrees

lat Latitude in decimal degrees

sed Percent sands and fines

elev Log-transformed elevation

slope Slope

area Log-transformed catchment area

sed.log Log-transformed percent sands and fines

temp Grab stream temperature

temp.avg Seven day average maximum temperature

elev.ut Elevation

sp Estimate of unit stream power

Source

Oregon Department of Environmental Quality

| genResultSpace | *Make new space for results* |

Description

Internal function

---

gisdt.subset	*Subset data frame*

Description

Internal function

---

glm.JGR	*Dialog box for fitting a generalized linear regression model*

Description

Fits a generalized linear regression model returning model coefficient estimates, confidence intervals, and diagnostic plots.

Details

User Interface

Select *Analysis Tools -> Linear Regression* from the menus. A dialog box will open. Select the data set of interest from the pull-down menu, or browse for a tab-delimited text file. The Data Subsetting tab can be used to select a subset of the data file by choosing a variable from the pull down menu and then selecting the levels of that variable to include. Hold down the <ctrl> key to add several levels.

Select a response variable (the variable whose value you wish to predict). An appropriate distribution for your response variable can be selected by clicking on a radio button. **Normal** distributions are most common. A **Poisson** distribution may be appropriate if your response variable reflects a count of some attribute (e.g., number of distinct taxa), and a **Binomial** distribution may be appropriate if you are modeling the probability of success and your response variable is a number of successes for a fixed number of trials (e.g., the presence of a species in a sample).

Select all explanatory variables you wish to include in the model (the variables used to predict the dependent variable. You can hold down the <CTRL> key to add multiple variables. Note: the dependent variable is in the list of possible independent variables, but it *should not* be selected as an independent variable.

By default, an intercept is included in the model., The intercept can be excluded by selecting **Remove Intercept**, in which case at least one independent variable must have been selected.

Confidence intervals for the regression coefficients are generated only if **Compute Confidence Intervals** is selected. The confidence level can be changed only after Compute Confidence Intervals is selected.

Several diagnostic plots are available for plotting. None are produced by default, but any subset of the five diagnostic plots can be selected for display.

Should you wish to save the linear regression as an R object, select **Save R Results?** and type a variable name in R Result Name.

The output is the regression coefficients, along with standard errors, t-statistics, and p-values. Confidence intervals and diagnostic plots are produced only if selected.

See Also

glm

| glm.pred.JGR | *Dialog box to apply predictions from a regression model to assess test sites* |

Description

The regression prediction tool provides a way to model natural variations in an environmental parameter (e.g., stream temperature) in reference (i.e., least-disturbed) sites. Then, one can use the reference site model to determine whether observations of that same parameter in test sites are within the range of variation defined by the reference sites.

The linear regression tool has more options available for diagnosing and assessing the fit of a regression model on the quality data. The regression prediction tool assumes that such diagnostics have been performed to define a regression model, and the goal is now to apply the model to new data.

Details

User Interface

Select *Analysis Tools -> Regression Prediction* from the menus. A dialog box will open. Select the data set of interest from the pull-down menu, or browse for a tab-delimited text file. The Data Subsetting tab can be used to select a subset of the data file by choosing a variable from the pull down menu and then selecting the levels of that variable to include. You can hold down the <CTRL> key to add several levels.

Select **Dependent** variable and **Independent** variables as in the linear regression tool.

By default, an intercept is included in the model. The intercept can be excluded by selecting **Remove Intercept** under the Analysis Options, in which case at least one independent variable must have been selected.

Select a Reference Variable: a variable that contains (character) values which can be used to identify which rows correspond to reference data. Once the Reference Variable is selected, select the levels of that variable that correspond to reference data. You can hold down the <CTRL> key to select multiple levels that indicate reference.

Choose a **Significance Level** for identification of anomalous data points in the new data. Note that the significance level corresponds to a level for a single prediction, not a family-wise significance level; no correction is performed for multiple comparisons when multiple predictions are being made. Keep in mind that some new data is expected to be flagged as significant (on average, about the number of predictions times the significance level should appear significant, depending on correlation).

The **ID Variable** allows for a label variable to be specified, so that points that differ significantly from reference can be identified on the plot. If no ID Variable is selected, then the row number will be used instead.

The output of the regression prediction tool is a table for the test (non-reference) data, including the observed and predicted values, the standard error associated with the prediction, a p-value for the prediction, an indication of whether the sample differs significantly from reference, an indication of whether the independent variables for the new data are in range of the model fit, and finally an indication of whether or not the normal approximation used to calculate the p-value is appropriate (for poisson and binomial regression only). Note: for binomial data, the observed value is converted from a count to a proportion (dependent variable divided by sample size). A plot of observed versus predicted values is also produced, to help assess the model fit and identify anomalous data points.

---

jgr.__removeMenu *Remove menus from JGR interface*

Description

Internal function

JGRMessageBox *Dialog box for predicting env. conditions from biological observations*

Description

A general Java message box contains information and error message

Usage

```
JGRMessageBox(w.title = "BiologicalInferences: Error", msg = "Hello!")
```

Arguments

w.title	Title of the message box
msg	Message in the box

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

## The function is currently defined as
function (w.title = "BiologicalInferences: Error", msg = "Hello!")
{
    a <- .jnew("org.neptuneinc.cadstat.plots.BiologicalInferencesGeneralBox")
    .jcall(a, "Ljavax/swing/JFrame;", "getMyGUI", w.title, msg)
  }
```

`load.class`	*Load a new java object*

Description

Internal function

`load.menus`	*Load new CADStat menus*

Description

Internal function

`localJGRError`	*Output error from JGR*

Description

Internal function

`pca.fa.JGR`	*Dialog box for calculating a principle component analysis*

Description

This function receives inputs from the Java GUI for factor analysis/principle component analysis and sends these parameters to the FA/PCA function.

Details

User Interface

Select *Analysis Tools -> Factor Analysis* from the menus. A dialog box will open. Select the data set of interest from the pull-down menu, or browse for a tab-delimited text file. The Data Subsetting tab can be used to select a subset of the data file by choosing a variable from the pull down menu and then selecting the levels of that variable to include. You can hold down the <CTRL> key to select several levels.

Select the **No. of Factors** you wish to compute. Select all explanatory (**Variables**) you wish to include in the analysis. You can hold down the <CTRL> key to select several variables.

Output options include generating a **Scree Plot**, **Loading/Score Plot**, or using **Rotated Variables**.

Analysis options include centering all variables to zero and scaling all variables to a unit variance.

---

refids.or *Reference site identifications western OR*

Description

Sample identifiers for reference site in western Oregon

Format

A data frame with 245 observations on the following 13 variables.

SITE.ID Unique sample code

REF Flag for reference site

Source

Oregon Department of Environmental Quality

---

resample *Resample for cond prob*

Description

Internal function

---

rpart.JGR *Dialog box for classification and regression trees*

Description

Provides dialog box for fitting a classification and regression tree.

Details

User Interface

Select *Analysis Tools -> Tree Regression* from the menus. A dialog box will open. Select the data set of interest from the pull-down menu, or browse for a tab-delimited text file. The Data Subsetting tab can be used to select a subset of the data file by choosing a variable from the pull down menu and then selecting the levels of that variable to include. You can hold down the <CTRL> key to add several different levels.

Select a **Dependent** variable (the variable you wish to predict). Select all **Independent** variables you wish to include in the model (the variables used to predict the dependent variable). You may

hold down the <CTRL> key to several variables. Note: the dependent variable is in the list of possible independent variables, but it should not be selected as an independent variable, or a degenerate regression will be produced.

You can set further options for controlling the tree regression by setting the three analysis options.

The output is a graph containing the tree structure and its estimated regression coefficients.

See Also

rpart

---

rq.JGR   *Dialog box for calculating a quantile regression*

Description

This function receives inputs from the Java GUI for quantile regression and sends these parameters to the quantile regression function.

Details

User Interface

Select *Analysis Tools -> Quantile Regression* from the menus. A dialog box will open. Select the data set of interest from the pull-down menu, or browse for a tab-delimited text file. The Data Subsetting tab can be used to select a subset of the data file by choosing a variable from the pull down menu and then selecting the levels of that variable to include. You can hold down the <CTRL> key to select several levels.

Select a **Dependent** variable (the variable you wish to model). Select all explanatory (**Independent**) variables you wish to include in the model. You can hold down the <CTRL> key to select several variables. Note: the dependent variable is in the list of possible independent variables, but it should not be selected as an independent variable.

Type in the quantiles that you wish to estimate in the **Quantiles** dialog. You may type in up to 8 quantiles, all of which must be between 0 and 1. These must be typed from top to bottom: skipped lines will result in later values being ignored. Be sure that you hit <TAB> or <ENTER> after typing in your last quantile and before hitting Submit, or the last quantile will not register.

Confidence intervals for the regression coefficients are generated only if **Compute Confidence Intervals** is selected. The confidence level can be changed only after Compute Confidence Intervals is selected.

If **Regression Scatterplot** is selected, then one scatterplot is plotted for each independent variable, of the dependent versus independent variable with overlaid quantile regression fits. Confidence bounds on the fit are added to the plot if Coefficient Confidence Bands is chosen.

The output is the regression coefficients, with confidence bounds if selected, and scatterplots if selected.

See Also

[rq](rq)

`rq.plot.JGR`	*Plot quantile regression*

Description

Internal function

`scatterplot2`	*Draws a scatterplot based on GUI inputs*

Description

This function draws the scatterplot based on user selections from the GUI interface.

Details

Internal function.

`scatterplot2.JGR`	*Draws a scatterplot of bivariate data*

Description

Creates a scatterplot - a plot displaying the relationship between two variables. A linear regression line or loess smooth can be overlayed.

Details

A scatterplot provides a graphical means of determining the relationship between two variables. Important relationships that can be observed are linear, higher-order polynomial, exponential, or logarithmic. Along with the general pattern, it is possible to qualitatively determine the strength of the relationship. The "closer" or "tighter" the points are to the general shape, the stronger the relationship. A scatterplot is also a useful tool for identifying potential univariate or bivariate outliers (observations that are much larger or smaller than the other data points).

After gaining insight into the general relationship of the variables, a quantitative analysis method may be more easily chosen:

- Correlation analysis, which helps determine the strength of the linear relationship between two variables.
- Linear regression, which fits a linear model to two or more variables.
- Quantile regression, which fits a model to conditional quantiles of the dependent variable rather than to the conditional mean.

User Interface

To create a scatterplot, select "Scatterplot" from the "Analysis Tools" menu. A dialog box will appear that allows the user to enter plotting parameters. Select the data set of interest from the pull-down menu, or browse for a tab-delimited text file. The Data Subsetting tab can be used to select a subset of the data file by choosing a variable from the pull down menu and then selecting the levels of that variable to include. You can hold down the <CTRL> key to select multiple levels. A separate plot will be produced for each unique setting of the subsetting variables. Select a variable to plot on the x-axis and a different variable to plot on the y-axis. A regression fit can be overlaid on the scatterplot. If Linear Regression Line is selected, a straight line fit is drawn. If Loess Smooth is selected, a nonparametric loess fit is drawn. You may change the axes labels and plot title by typing them in the Labels dialog. The output is simply the scatterplot graph with any overlaid fits requested.

See Also

To learn about boxplots, see boxplot.

To learn about correlation analysis, see cor.

To learn about linear regression, see lm.

To view the help file for the general R plotting function, see plot.

---

setVisible	*Make something visible*

Description

Internal function

---

text.rpart.JGR	*Add text to rpart plot*

Description

Internal function

Index

21