Package 'HoRM'

October 12, 2022

Type Package

Title Supplemental Functions and Datasets for ``Handbook of Regression Methods''

Version 0.1.3

Date 2021-03-10

Depends R (>= 3.5.0)

Imports ggplot2, MASS, orthopolynom, quantmod, rsm, stats4

Description Supplement for the book ``Handbook of Regression Methods'' by D. S. Young. Some datasets used in the book are included and documented. Wrapper functions are included that simplify the examples in the textbook, such as code for constructing a regressogram and expanding ANOVA tables to reflect the total sum of squares.

URL https://github.com/dsy109/HoRM

License GPL (>= 2)

NeedsCompilation no

Author Derek S. Young [aut, cre]

Maintainer Derek S. Young <derek.young@uky.edu>

Repository CRAN

Date/Publication 2021-03-11 06:10:06 UTC

R topics documented:

HoRM-package   *Supplemental Functions and Datasets for "Handbook of Regression Methods"*

Description

Various wrapper functions and datasets to supplement examples for the book "Handbook of Regression Methods" by D. S. Young.

Details

Package:	HoRM
Type:	Package
Version:	0.1.3
Date:	2021-03-10
Imports:	ggplot2, MASS, orthopolynom, quantmod, rsm, stats4
License:	GPL (>= 2)

Author(s)

Derek S. Young, Ph.D.

Maintainer: Derek S. Young <derek.young@uky.edu>

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

---

amit	*Amitriptyline Dataset*

Description

This dataset is from a study on the side effects of amitriptyline, which is a drug some physicians prescribe as an antidepressant.

Usage

```
data(amit)
```

Format

This data frame consists of 7 variables on 17 subjects:

- TOT The subject's total TCAD plasma level.
- AMI The amount of amitriptyline present in the TCAD plasma level.
- GEN The subject's gender, where 0 is for a male subject and 1 is for a female subject.
- AMT Amount of the drug taken at the time of overdose.
- PR The subject's PR wave measurement.
- DIAP The subject's diastolic blood pressure.
- QRS The subject's QRS wave measurement.

Source

Rudorfer, M. V. (1982), Cardiovascular Changes and Plasma Drug Levels After Amitriptyline Overdose, *Journal of Toxicology - Clinical Toxicology*, **19**, 67–78.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

---

auditory   *Auditory Discrimination Dataset*

Description

This dataset is from a study to assess auditory differences between environmental sounds given several other factors.

Usage

```
data(auditory)
```

Format

This data frame consists of 3 variables on 20 subjects:

- pre.test The subject's pre-test score.
- gain The gain in auditory score between the pre- and post-test administration of the treatment.
- Culture A cultural status indicator of individuals, where 1 means the subject is from a "culturally-nondeprived" group and 0 means the subject is from a "culturally-deprived" group.

Source

Hendrix, L. J., Carter, M. W., and Scott, D. T. (1982), Covariance Analysis with Heterogeneity of Slopes in Fixed Models, *Biometrics*, **38**, 226–252.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

Auto   *Canadian Auto Insurance Dataset*

Description

This dataset is from The Statistical Unit of the Canadian Underwriters' Association collated automobile insurance policies (policy years 1956 and 1957) for private passenger automobile liability for non-farmers in Canada excluding those in the province of Saskatchewan.

Usage

```
data(Auto)
```

Format

This data frame consists of 20 categories (rows) and 6 variables (columns):

- `Merit` Merit rating of policyholder.
- `Class` Class rating of policyholder.
- `Insured` Earned car years under the policy.
- `Premiums` Earned premium at present rates (in 1000's of Canadian dollars).
- `Claims` Number of claims incurred.
- `Cost` Incurred losses (in 1000's of Canadian dollars).

Source

Bailey, R. A. and Simon, L. J. (1960), Two Studies in Automobile Insurance Ratemaking, *ASTIN Bulletin*, **1**, 192–217.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

BAC   *Blood Alcohol Concentration Dataset*

Description

This dataset is from a study to compare the blood alcohol concentration (BAC) of subjects using two different methods.

Usage

```
data(BAC)
```

Format

This data frame consists of 2 variables measured on 15 subjects:

- `breath` BAC obtained using the Breathalyzer Model 5000.
- `labtest` BAC based on a breath estimate in a laboratory.

Source

Krishnamoorthy, K., Kulkarni, P. M., and Mathew, T. (2001), Multiple Use One-Sided Hypotheses Testing in Univariate Linear Calibration, *Journal of Statistical Planning and Inference*, **93**, 211–223.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

---

cheese   *Cheese-Tasting Experiment Dataset*

Description

This dataset is from an experiment concerning the effect on taste of various cheese additives.

Usage

```
data(cheese)
```

Format

This data frame (36 rows by 3 columns) is a tabulation of the responses by 208 subjects to 4 different cheeses:

- `Cheese` The cheese additive used (four levels labeled A, B, C, and D).
- `Response` The response based on the 9-point hedonic scale.
- `N` The number of subjects who responded according to the value in `Response` for the cheese additive in `Cheese`.

Source

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, CRC Press.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

---

chem   *Odor Dataset*

Description

This dataset is from an experiment that was designed to determine the effects of three factors in reducing the unpleasant odor in a chemical product being sold for household use.

Usage

```
data(chem)
```

Format

This data frame consists of 4 variables (stored in coded form) at 15 design points:

- odor A measure of the chemical's odor.
- temp Temperature at time of measurement - uncoded units are 40, 80, and 120.
- ratio Gas-liquid ratio - uncoded units are 0.3, 0.5, and 0.7.
- height Packing height - uncoded units are 2, 4, and 6.

Source

John, P. W. (1971), *Statistical Design and Analysis of Experiments*, MacMillan Company.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

---

compasst   *Computer-Assisted Learning Dataset*

Description

This dataset is from a study of computer-assisted learning by students in an effort to assess the cost of computer time.

Usage

```
data(compasst)
```

Format

This data frame consists of 2 variables measured on 12 students:

- num.responses Total number of responses in completing a lesson.
- cost Cost of the computer time (in cents).

Source

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2005) *Applied Linear Statistical Models, Fifth Edition*, McGraw-Hill/Irwin.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

---

cracker   *Cracker Dataset*

Description

This dataset is from marketing research on the sales of crackers for a particular company.

Usage

```
data(cracker)
```

Format

This data frame consists of 15 stores, each receiving a particular promotional strategy, with the following 4 variables (columns):

- treat An indicator for which of the three marketing strategies (treatment) was employed.
- store The store number within the particular treatment.
- y The number of cases of the crackers sold during the promotional period.
- x The store's cracker sales during the preceding sales period.

Source

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2005) *Applied Linear Statistical Models, Fifth Edition*, McGraw-Hill/Irwin.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

---

credloss   *Credit Loss Dataset*

Description

This dataset consists of credit portfolio loss data that were extracted from the Altman-NYU Salomon Center Corporate Bond Default Database for the years 1982 through 2005.

Usage

```
data(credloss)
```

Format

This data frame consists of 5 variables over 24 years:

- year The year the statistics were collected.
- PD The probability of default.
- defs The number of defaults.
- LGD.mean The mean loss given default.
- LGD.vol A loss given default volatility measure.

Source

Bruche, M. and Gonzalez-Aguado, C. (2010), Recovery Rates, Default Probabilities, and the Credit Cycle, *Journal of Banking and Finance*, **34**, 754–764.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

---

fiber *Fiber Strength Dataset*

Description

This dataset is from a study about the strength of a particular type of fiber based on the amount of pressure applied.

Usage

```
data(fiber)
```

Format

This data frame consists of 30 samples with the following 2 variables measured:

- pressure The amount of water pressure applied (measured in bars); the unique levels are 60, 80, 100, 120, 150, and 200.
- tensile The tensile strength of fiber (measured in N/5 cm).

Source

Ndaro, M. S., Jin, X.-Y., Chen, T., and Yu, C.-W. (2007), Splitting of Islands-in-the-Sea Fibers (PA6/COPET) During Hydroentangling of Nonwovens, *Journal of Engineered Fibers and Fabrics*, **2**, 1–9.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

| fly | *Fruit Fly Dataset* |

Description

This dataset is from a study on the effects of temperature on development of the common fruit fly.

Usage

```
data(fly)
```

Format

This data frame consists of 23 batches with the following 8 variables measured:

- temp The experimental temperature (in degrees Celsius).

- exp.no The experiment number.

- duration The mean duration of the embryonic period (in hours).

- dur.var The standard deviation of the duration of the recorded embryonic period.

- batch The number of eggs in each batch.

- batch.sd The standard deviation of the number of eggs in each batch.

- egg.temp The temperature in which the eggs were laid (in degrees Celsius).

- egg.dur The duration the eggs remained in the given conditions (in hours).

Source

Powsner, L. (1935), The Effects of Temperature on the Durations of the Developmental Stages of *Drosophila Melanogaster*, *Physiological Zoology*, **8**, 474–520.

References

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, CRC Press.

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

---

gas   *Natural Gas Prices Dataset*

Description

This dataset is of the monthly observations of spot prices for natural gas from January 1988 to October 1991 for the states of Louisiana and Oklahoma.

Usage

```
data(gas)
```

Format

This data frame consists of a total of 46 (monthly) observations of spot prices for the 2 states stated above:

- OK Oklahoma spot prices for natural gas (dollars per million British thermal units).
- LA Louisiana spot prices for natural gas (dollars per million British thermal units).

Source

Wei, W. W. S. (2005), *Time Series Analysis: Univariate and Multivariate Methods*, Pearson.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

GOF.tests   *Summary of Goodness-of-Fit Tests*

Description

A function that reports the Pearson statistic, the deviance statistic, and their respective p-values for goodness-of-fit testing based on a linear regression fit (lm) or a generalized linear regression fit (glm).

Usage

```
GOF.tests(out)
```

Arguments

out   An object of class lm or glm.

Value

GOF.tests returns a data frame with rows corresponding to the goodness-of-fit test and columns corresponding to the respective test statistic and p-value.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

See Also

glm, lm

Examples

```
## Goodness-of-fit tests for the logistic regression fit to the
## menarche dataset.

data(menarche, package = "MASS")

glm.out = glm(cbind(Menarche, Total - Menarche) ~ Age,
              family = binomial, data = menarche)
GOF.tests(glm.out)
```

GRB   *Gamma-Ray Burst Dataset*

Description

This dataset consists of measurements on gamma-ray bursts, which are short, intense flashes of gamma-ray radiation that occur at (seemingly) random times and locations in space.

Usage

```
data(GRB)
```

Format

This data frame consists of 63 measurements of the following 2 variables:

- TIME Time of measurement (in seconds).
- FLUX X-ray flux measurement (in units of 10^(-11) erg/cm2/s, 2-10 keV).

Source

Blustin, A. J., Band, D., Barthelmy, S., Boyd, P., Capalbi, M., Holland, S. T., Marshall, F. E., Mason, K. O., Perri, M., Poole, T., Roming, P., Rosen, S., Schady, P., Still, M., Zhang, B., Angelini, L., Barbier, L., Beardmore, A., Breeveld, A., Burrows, D. N., Cummings, J. R., Canizzo, J., Campana, S., Chester M. M., Chincarini, G., Cominsky, L. R., Cucchiara, A., de Pasquale, M., Fenimore, E. E., Gehrels, N., Giommi, P., Goad, M., Gronwall, C., Grupe, D., Hill, J. E., Hinshaw, D., Hunsberger, S., Hurley K. C., Ivanushkina, M., Kennea, J. A., Krimm, H. A., Kumar, P., Landsman, W., La Parola, V., Markwardt, C. B., McGowan, K., Meszaros, P., Mineo, T., Moretti, A., Morgan, A., Nousek, J., O'Brien, P. T., Osborne, J. P., Page, K., Page, M. J., Palmer, D. M., Parsons, A. M., Rhoads, J., Romano, P., Sakamoto, T., Sato, G., Tagliaferri, G., Tueller, J., Wells, A. A. and White, N. E. (2006), Swift Panchromatic Observations of the Bright Gamma-Ray Burst GRB 050525a, *The Astrophysical Journal*, **637**, 901–913.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

---

hildreth.lu	*Hildreth-Lu Procedure*

Description

Returns the linear regression fit for a given level of rho using the Hildreth-Lu procedure.

Usage

```
hildreth.lu(y, x, rho)
```

Arguments

y	A vector of response values.
x	A vector of predictor values. Must be the same length as y.
rho	A value for the correlation assumed for the autoregressive structure of the errors.

Value

hildreth.lu returns an object of class lm using the transformed quantities calculated for the Hildreth-Lu procedure.

References

Hildreth, C. and Lu, J. Y. (1960), Demand Relations with Autocorrelated Disturbances, *Technical Bulletin 276*, Michigan State University Agricultural Experiment Station.

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

See Also

[lm](lm)

Examples

```
## Example using the natural gas dataset.

data(gas)

out.1 <- hildreth.lu(y = gas$OK, x = gas$LA, rho = 0.1)
out.2 <- hildreth.lu(y = gas$OK, x = gas$LA, rho = 0.5)

out.1
out.2
```

JamesBond   *James Bond Dataset*

Description

This dataset consists of various metrics pertaining to the officially-produced *James Bond* films.

Usage

```
data(JamesBond)
```

Format

This data frame consists of 18 variables measured on the 24 films:

- Year Year of the film's theatrical release.
- Movie Title of the movie.
- Bond Actor who played James Bond.
- US_Gross The film's U.S. gross (in U.S. dollars).
- US_Adj The film's 2013-adjusted U.S. gross (in 1000's of U.S. dollars).
- World_Gross The film's worldwide gross (in U.S. dollars).
- World_Adj The film's 2013-adjusted worldwide gross (in 1000's of U.S. dollars).
- Budget The film's budget (in U.S. dollars).
- Budget_Adj The film's 2013-adjusted budget (in 1000's of U.S. dollars).
- Film_Length Length of the theatrical release.
- Avg_User_IMDB The average user rating on IMDB (www.imdb.com).
- Avg_User_Rtn_Tom The average user rating on Rotten Tomatoes (www.rottentomatoes.com).
- Conquests The number of "conquests" by Bond in the film.

- Martinis The number of martinis Bond drank in the film.
- BJB The number of times Bond stated "Bond. James Bond." in the movie.
- Kills_Bond The number of people killed by Bond.
- Kills_Others The number of people killed in the film by people other than Bond.
- Top_100 An indicator where a value of 1 means the title song within the top 100 on the UK Singles Chart and the U.S. Billboard Hot 100 and a value of 0 means it did not.

Source

Young, D. S. (2014), Bond. James Bond. A Statistical Look at Cinema's Most Famous Spy, *CHANCE*, **27**(2), 21–27.

Young, D. S. (2019), Bond. James Bond. A Statistical Look at Cinema's Most Famous Spy (*The Best of CHANCE Issue*), *Chance*, **32**(1), 27–35.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

---

light *Light Dataset*

Description

This dataset is from an experiment where light was transmitted through a chemical solution and an optical reading was recorded.

Usage

```
data(light)
```

Format

This data frame consists of 2 variables measured on 12 different instances:

- reading The optical reading.
- concentration The concentration of the chemical.

Source

Graybill, F. A. and Iyer, H. K. (1994), *Regression Analysis: Concepts and Applications*, Duxbury Press.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

logdiag   *Diagnostic Measures of Certain Regression Estimates*

Description

A function for computing various residual-based and influence-based quantities from a linear regression fit using `lm` or a generalized linear regression fit using `glm`.

Usage

```
logdiag(out)
```

Arguments

out   An object of class `lm` or `glm`.

Value

`logdiag` returns a data frame with the following columns:

`r.i`	The raw residuals.
`p.i`	The Pearson residuals.
`d.i`	The deviance residuals.
`stud.r.i`	The Studentized raw residuals.
`stud.p.i`	The Studentized Pearson residuals.
`stud.d.i`	The Studentized deviance residuals.
`h.ii`	The leverage values.
`C.i`	The Cook's distance value.
`C.i.bar`	The average Cook's distance value when omitting observation i.
`DFDEV`	The change in the deviance statistic when omitting observation i.
`DFCHI`	The change in the Pearson's chi-square statistic when omitting observation i.
`fit`	The estimated response (fitted) values.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

See Also

[glm](glm), [lm](lm)

Examples

```
## Diagnostic summaries for the logistic regression fit to the
## menarche dataset.

data(menarche, package = "MASS")

glm.out = glm(cbind(Menarche, Total - Menarche) ~ Age,
              family = binomial, data = menarche)
logdiag(glm.out)
```

---

poly2form   *Expands Design Matrix Based on Polynomials*

Description

This function takes a list of objects having class `polynomial`, evaluates each polynomial as a function of x, then returns the results in a matrix.

Usage

```
poly2form(poly.out, x)
```

Arguments

poly.out   A list whose objects are of class `polynomial`.

x   A vector of values for which each polynomial in `poly.out` is to be evaluated.

Value

`poly.out` returns a matrix whose columns are the evaluation of each polynomial in `poly.out` using x.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

See Also

legendre.polynomials

Examples

```
## Evaluating the order 5 Legendre polynomials.

require(orthopolynom)

px <- legendre.polynomials(n = 5, normalized = FALSE)
lx <- poly2form(poly.out = px, x = 1:10)

lx
```

---

power.b   *Power Functions for Tests of Simple Linear Regression Coefficients*

Description

A function to calculate the power of the t-tests corresponding to tests on the intercept and slope coefficients in the simple linear regression model.

Usage

```
power.b(x, y, alpha = 0.05, B0 = 0, B1 = 0)
```

Arguments

x	A vector of predictor values. Must be the same length as y.
y	A vector of response values. Must be the same length as x.
alpha	Significance level of the test. Default level is 0.05.
B0	Null value for the test about the intercept.
B1	Null value for the test about the slope.

Value

power.b returns a matrix with the noncentrality parameters and power levels for the corresponding t-tests.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

See Also

power.F

Examples

```
## Applied to the toy dataset.

data(toy)

power.b(x = toy$x, y = toy$y)
```

---

power.F   *Power Function for the General Linear F-Test*

Description

A function to calculate the power of the general linear F-test.

Usage

```
power.F(full, reduced, alpha = 0.05)
```

Arguments

full	The full model (specified in the alternative hypothesis) in the general linear F-test. This is an object of class lm.
reduced	The reduced model (specified in the null hypothesis) in the general linear F-test. This is an object of class lm.
alpha	Significance level of the test. Default level is 0.05.

Value

power.F returns a single value (saved as a matrix) with the power for the corresponding general linear F-test.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

See Also

anova, lm

Examples

```
## Applied to the toy dataset.

data(toy)

full <- lm(y~x, data = toy)
reduced <- lm(y~1, data = toy)
power.F(full = full, reduced = reduced, alpha = 0.05)
```

ppr_funs   *Ridge Functions for Projection Pursuit Regression*

Description

The portion of the plot.ppr code that computes the ridge traces for projection pursuit regression.

Usage

```
ppr_funs(obj)
```

Arguments

obj   A fit of class ppr as produced by the ppr function.

Details

This is just the segment of code in plot.ppr, which calculates the ridge traces.

Value

ppr_funs returns the evaluated ridge trace values based on output from the ppr function.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

See Also

ppr, plot.ppr

Examples

```
## Projection pursuit regression on the rock dataset.

data(rock)

ppr.out <- ppr(log(perm) ~ area + peri + shape,
               data = rock, nterms = 2, max.terms = 5)
obj <- ppr_funs(ppr.out)

obj
```

---

psi.andrew   *Andrew's Sine Function*

Description

Andrew's sine function for use when fitting a linear model by robust regression using an M-estimator.

Usage

```
psi.andrew(u, k=1.339, deriv=0)
```

Arguments

u   Numeric vector of evaluation points.

k   Tuning constant. The suggested default value is 1.339.

deriv   0 or 1: to compute values of this function or of its first derivative.

Value

`psi.andrew` returns a vector of points evaluated using Andrew's sine function.

References

Andrew, D. F. (1974), A Robust Method for Multiple Linear Regression, *Technometrics*, **16**, 523–531.

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

See Also

[rlm](rlm)

Examples

```
## Robust fit of the stackloss dataset.

require(MASS)

data(stackloss, package="datasets")

out <- rlm(stack.loss ~ ., data = stackloss,
           psi = psi.andrew)

out
```

---

reg.anova   *Expanded ANOVA Table*

Description

Calculate the ANOVA table for an object of class lm. The results are identical to those obtained from anova, but an extra line is included that prints the total degrees of freedom and the total sum of squares.

Usage

```
reg.anova(lm.out)
```

Arguments

lm.out   An object of class lm (i.e., the results from the linear model fitting routine such that the anova function can act upon).

Value

reg.anova returns exactly the same output as the anova function applied to an object of class lm, but includes an extra line that summarizes the total source of variability.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

See Also

anova, lm

Examples

```
## Applied to the toy dataset.

data(toy)

lm.out <- lm(y ~ x, data = toy)
anova(lm.out)
reg.anova(lm.out)
```

---

reg.manova   *Expanded MANOVA Table*

Description

Expands the MANOVA results from an object of class `summary.aov`. The results are identical to those obtained from `summary.aov`, but an extra line is included that prints the total degrees of freedom and the total sum of squares for each dimension of the response vector.

Usage

```
reg.manova(AOV.out)
```

Arguments

AOV.out   An object of class `summary.aov`.

Value

`AOV.out` returns exactly the same output as the `summary.aov` function, but includes an extra line that summarizes the total source of variability for each dimension of the response vector.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

See Also

anova, reg.anova, summary.aov

Examples

```
## Applied to the amit dataset.

data(amit)

fits <- manova(cbind(TOT, AMI) ~ ., data = amit)
out <- summary.aov(fits)
```

```
mvreg.out <- lapply(out, reg.manova)
mvreg.out
```

---

regressogram   *Regressogram*

Description

Computes and plots the regressogram for a single predictor and single response relationship. The regressogram is plotted using ggplot2.

Usage

```
regressogram(x, y, nbins = 10, show.bins = TRUE,
        show.means = TRUE, show.lines = TRUE,
             x.lab = "X", y.lab = "Y", main = "TITLE")
```

Arguments

x	A vector of predictor values for the data. Must be the same length as y.
y	A vector of response values for the data. Must be the same length as x.
nbins	How many bins to use construction of the regressogram.
show.bins	A logical argument specifying if dashed vertical lines should be drawn at the boundaries of the bins. Default is TRUE.
show.means	A logical argument specifying if a large point should be overlayed at the midpoint of each bin and the respective mean of the response values within that bin. Default is TRUE.
show.lines	A logical argument specifying if a line should be drawn connecting the points determined by show.means. Default is TRUE.
x.lab	Label for the x-axis.
y.lab	Label for the y-axis.
main	Title for the regressogram.

Value

regressogram returns a plotted regressogram using the ggplot2 package.

References

Wasserman, L. (2006), *All of Nonparametric Statistics*, Springer.

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

See Also

[hist](#)

Examples

```
## Regressogram for the natural gas dataset.

data(gas)

regressogram(x = gas$LA, y = gas$OK, nbins = 6, x.lab = "LA",
             y.lab = "OK", main = "Regressogram")
```

---

repair   *Computer Repair Dataset*

Description

This dataset is from a random sample of service call records for a computer repair company.

Usage

```
data(repair)
```

Format

This data frame consists of a sample of 14 companies with the following 2 variables measured:

- `minutes` The length of service call (in minutes).
- `units` The number of components repaired or replaced during the service call.

Source

Chatterjee, S. and Hadi, A. S. (2012), *Regression Analysis by Example*, John Wiley and Sons, Inc.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

SSCP.fn   *Sums of Squares and Cross-Products Matrices for a MANOVA Table*

Description

Summarizes the MANOVA results based on the sum of squares and cross-products decomposition for the regression (SSCPR), the error (SSCPE), and the overall total (SSCPTO).

Usage

```
SSCP.fn(fits)
```

Arguments

fits   An object of class manova.

Value

SSCP.fn returns a list of length 3 with the SSCPR, SSCPE, and SSCPTO.

References

Johnson, R. A. and Wichern, D. W. (2007), *Applied Multivariate Statistical Analysis, Sixth Edition*, Pearson.

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

See Also

[manova](), [reg.manova]()

Examples

```
## Applied to the amit dataset.

data(amit)

fits <- manova(cbind(TOT, AMI) ~ ., data = amit)
SSCP.fn(fits = fits)
```

---

stock   *Google Stock Dataset*

Description

This dataset consists of the closing stock price of a share of Google stock during the trading days between February 7-th and July 7-th of 2005.

Usage

```
data(stock)
```

Format

This is an extensible time series (`xts`) object for the 105 trading days of interest:

- `GOOG.close` The closing stock price of a share of Google stock.

Source

Yahoo! Finance; accessed 01-26-2017.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

Examples

```
## Not run:
## How the data were accessed (1/26/17).

require(quantmod)

getSymbols("GOOG", src = "yahoo",
           from = "2005-02-07", to = "2005-07-07")
stock <- GOOG[,4]

## End(Not run)
```

---

tortoise   *Tortoise Eggs Dataset*

Description

This dataset is from a study on the number of eggs in female gopher tortoises in southern Florida.

Usage

```
data(tortoise)
```

Format

This data frame consists of 2 variables measured on 18 tortoises:

- length The carapace length (in millimeters).
- clutch The number of eggs (clutch size).

Source

Ashton, K. G., Burke, R. L., and Layne, J. N. (2007), Geographic Variation in Body and Clutch Size of Gopher Tortoises, *Copeia*, **2007**, 355–363.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

---

toy   *Toy Dataset*

Description

A made-up (toy) dataset.

Usage

```
data(toy)
```

Format

This data frame consists of 2 made-up variables for a sample of size 5:

- x The made-up x values.
- y The made-up y values.

Source

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

---

wood   *Pulp Property Dataset*

Description

This dataset is from a study about the pulp properties of wood density of the Australian blackwood tree.

Usage

```
data(wood)
```

Format

This data frame consists of 2 variables measured on 7 samples:

- pulp The percentage of pulp yield.
- Kappa The Kappa number, which is a measurement of standard potassium permanganate solution that the pulp will consume.

Source

Santos, A., Anjos, O., Amaral, M. E., Gil, N., Pereira, H., and Simoes, R. (2012), Influence on Pulping Yield and Pulp Properties of Wood Density of *Acacia melanoxylon*, *Journal of Wood Science*, **58**, 479–486.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

---

yarn   *Yarn Fiber Dataset*

Description

This dataset is from a mixture experiment regarding a fiber blend that is spun into yarn to make draperies.

Usage

```
data(yarn)
```

Format

This data frame consists of 4 variables measured at 15 design points for a {3,2} simplex lattice design:

- x1 The proportion of the polyethylene component.
- x2 The proportion of the polystyrene component.
- x3 The proportion of the polypropylene component.
- y The measurement of yarn elongation.

Source

Cornell, J. A. (2002), *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data, Third Edition*, John Wiley and Sons, Inc.

References

Young, D. S. (2017), *Handbook of Regression Methods*, CRC Press.

Index