Package ‘nullabor’

August 29, 2013

Version 0.2.1

Description Generate null datasets and null plots with ease.

Title Tools for graphical inference

Author Hadley Wickham <h.wickham@gmail.com>

Maintainer Hadley Wickham <h.wickham@gmail.com>

License GPL

Imports MASS, ggplot2, plyr

Suggests reshape2

LazyData true

Type Package

LazyLoad false

Collate 'encryption.r' 'lal.r' 'method-dist.r' 'method-model.r' 'method-permute.r' 'protocols.r'

Repository CRAN

Date/Publication 2012-07-23 18:06:20

NeedsCompilation no

R topics documented:

decrypt ......................................................... 2
lal ............................................................... 2
lineup .......................................................... 2
null_dist ......................................................... 3
null_lm .......................................................... 4
null_permute .................................................... 4
resid_boot ....................................................... 5
resid_pboot ..................................................... 5
resid_rotate ..................................................... 6
resid_sigma ...................................................... 6
rorschach ....................................................... 7
decrypt

Use decrypt to reveal the position of the real data.

Description

The real data position is encrypted by the lineup function, and writes this out as a text string. Decrypt, decrypts this text string to reveal which where the real data is.

Usage

decrypt(...) 

Arguments

... character vector to decrypt 

Examples

decrypt("0uXR2p rut L2O2")

lal Los Angeles Lakers play-by-play data.

Description

Play by play data from all games played by the Los Angeles Lakers in the 2008/2009 season.

lineup The line-up protocol.

Description

In this protocol the plot of the real data is embedded amongst a field of plots of data generated to be consistent with some null hypothesis. If the observer can pick the real data as different from the others, this lends weight to the statistical significance of the structure in the plot. The protocol is described in Buja, Cook, Hofmann, Lawrence, Lee, Swayne, Wickham (2009) Statistical inference for exploratory data analysis and model diagnostics, Phil. Trans. R. Soc. A, 367, 4361-4383.

Usage

lineup(method, true = NULL, n = 20, pos = sample(n, 1), samples = NULL)
Arguments

- **method**: method for generating null data sets
- **true**: true data set. If NULL, `find_plot_data` will attempt to extract it from the current ggplot2 plot.
- **n**: total number of samples to generate (including true data)
- **pos**: position of true data. Leave missing to pick position at random. Encrypted position will be printed on the command line, `decrypt` to understand.
- **samples**: samples generated under the null hypothesis. Only specify this if you don’t want lineup to generate the data for you.

Details

Generate n - 1 null datasets and randomly position the true data. If you pick the real data as being noticeably different, then you have formally established that it is different to with p-value 1/n.

Examples

```r
if (require("ggplot2")) {
  qplot(mpg, wt, data = mtcars) %+%
  lineup(null_permute("mpg"), mtcars) +
  facet_wrap(~ .sample)
  qplot(mpg, .sample, data = lineup(null_permute("cyl"), mtcars),
        colour = factor(cyl))
}
```

### null_dist

Generate null data with a specific distribution.

Description

Null hypothesis: variable has specified distribution

Usage

```r
null_dist(var, dist, params = NULL)
```

Arguments

- **var**: variable name
- **dist**: distribution name. One of: beta, cauchy, chi-squared, exponential, f, gamma, geometric, log-normal, lognormal, logistic, negative binomial, normal, poisson, t, weibull
- **params**: list of parameters of distribution. If NULL, will use `fitdistr` to estimate them.

Value

a function that given data generates a null data set. For use with `lineup` or `rorschach`
null_lm

Generate null data with null residuals from a model.

Description
Null hypothesis: variable is linear combination of predictors

Usage
null_lm(f, method = "rotate", ...)

Arguments
f
model specification formula, as defined by lm
method
method for generating null residuals. Built in methods "rotate", "pboot" and "boot" are defined by resid_rotate, resid_pboot and resid_boot respectively
...
other arguments passedd onto method.

Value
a function that given data generates a null data set. For use with lineup or rorschach

Examples
if (require("ggplot2") && require("reshape2")) {

  x <- lm(tip ~ total_bill, data = tips)
  tips.reg <- data.frame(tips, .resid = residuals(x), .fitted = fitted(x))
  qplot(total_bill, .resid, data = tips.reg) %+%
    lineup(null_lm(tip ~ total_bill, method = "rotate"), tips.reg) +
    facet_wrap(~ .sample)
}

null_permute

Generate null data by permuting a variable.

Description
Null hypothesis: variable is independent of others

Usage
null_permute(var)
**resid_boot**

**Arguments**

- **var**: name of variable to permute

**Value**

a function that given data generates a null data set. For use with *lineup* or *rorschach*

---

**resid_boot**

*Bootstrap residuals.*

---

**Description**

For use with *null_lm*

**Usage**

```r
resid_boot(model, data)
```

**Arguments**

- **model**: to extract residuals from
- **data**: used to fit model

---

**resid_pboot**

*Parametric bootstrap residuals.*

---

**Description**

For use with *null_lm*

**Usage**

```r
resid_pboot(model, data)
```

**Arguments**

- **model**: to extract residuals from
- **data**: used to fit model
resid_rotate  Rotation residuals.

Description

For use with null_lm

Usage

resid_rotate(model, data)

Arguments

model  to extract residuals from
data   used to fit model

resid_sigma  Residuals simulated by a normal model, with specified sigma

Description

For use with null_lm

Usage

resid_sigma(model, data, sigma = 1)

Arguments

model  to extract residuals from
data   used to fit model
sigma,  a specific sigma to model
Description

This protocol is used to calibrate the eyes for variation due to sampling. All plots are typically null data sets, data that is consistent with a null hypothesis. The protocol is described in Buja, Cook, Hofmann, Lawrence, Lee, Swayne, Wickham (2009) Statistical inference for exploratory data analysis and model diagnostics, Phil. Trans. R. Soc. A, 367, 4361-4383.

Usage

rorschach(method, true = NULL, n = 20, p = 0)

Arguments

  method    method for generating null data sets
  true      true data set. If NULL, find_plot_data will attempt to extract it from the current ggplot2 plot.
  n         total number of samples to generate (including true data)
  p         probability of including true data with null data.
Index

*Topic **datasets**
  lal, 2
  decrypt, 2, 3
  find_plot_data, 3, 7
  fitdistr, 3
  lal, 2
  lineup, 2, 3–5
  lm, 4
  null_dist, 3
  null_lm, 4, 5, 6
  null_permute, 4
  resid_boot, 4, 5
  resid_pboot, 4, 5
  resid_rotate, 4, 6
  resid_sigma, 6
  rorschach, 3–5, 7