Package 'distWorkshop'

November 4, 2019

Title Distributional analyses of linguistic data in the GAMM framework

Version 1.0

Description Data sets and code for distributional analyses of linguistic data in the GAMM framework.

Depends R (>= 3.5.0), survival, mgcv, qgam, pammtools, RColorBrewer

Imports survival, mgcv, qgam, pammtools, RColorBrewer

License GNU General Public License v3.0

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1.9000

R topics documented:

`ld` *Lexical decision data*

Description

Lexical decision data from the British Lexicon Project (Keuleers et al., 2012)

Usage

```
ld
```

Format

An object of class `data.frame` with 25401 rows and 6 columns.

Source

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. Behavior Research Methods, 44(1), 287-304.

`nam` *Word naming data*

Description

Word naming data from the English Lexicon Project (Balota et al., 2007)

Usage

```
nam
```

Format

An object of class `data.frame` with 37107 rows and 6 columns.

Source

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. Behavior Research Methods, 39(3), 445-459.

`plotPAM` *Plot PAM*

Description

Plot the results of a PAM model (Bender & Scheipl, 2018)

Usage

```
plotPAM(model, data, predictor = "logFrequency", response = "RT",
  se = 2, area = FALSE, num_grid = 100,
  pallet = colorRampPalette(rev(brewer.pal(n = 7, name =
  "RdYlBu")))(500), levs = NA, rugx = TRUE, rugy = TRUE, main = NA,
  xlab = NA, ylab = NA, ...)
```

Arguments

`model`	A PAM model.
`data`	The data the PAM model was fit to. Needs to include the response variable in the task, as well as all predictors in these models. Note: this is the data frame in its raw format, not the data frame converted to the piece-wise exponential data format.
`predictor`	The predictor to be plotted. This predictor needs to be present in the fitted model, as well as in data.

response	The name of the response variable in data.
se	The number of standard errors that is used for the significance test. Default: 2 (i.e., 95% confidence intervals)
area	Should the significance of the effect at different predictor values be plotted. Default: FALSE.
pallet	A vector of color names that will be used for the contour plot.
levs	A vector of values at which the contour lines will be plotted. By default, these values are selected automatically
rugx	Should a rug be plotted for the x-axis? Default: TRUE
rugy	Should a rug be plotted for the y-axis? Default: TRUE

References

Bender, A. and Scheipl, F. (2018). pammtools: Piece-wise exponential additive mixed modeling tools. arXiv:1806.01042

Examples

```
# Remove outliers
ld = removeOutliers(ld, predictors)
ld = na.omit(ld)

# Prepare data in exponential data format
ld$status = 1
cut_points = as.numeric(quantile(ld$RT[which(ld$RT <= 1085 &
               ld$RT >= 500)],seq(0, 1, by = 0.02)))
ped = split_data(Surv(RT, status)~., data = ld, id = "id",
                  cut = cut_points)

# Run PAM (warning: computationally heavy)
pam_ld = gam(ped_status ~ s(tend) +
             s(logFrequency) + ti(tend, logFrequency) +
             s(Length) + ti(tend, Length) +
             s(logOLD20) + ti(tend, logOLD20) +
             s(SND20) + ti(tend, SND20),
             data = ped, offset = offset, family = poisson())

# Plot frequency effect
plotPAM(model = pam_ld, data = ld, predictor = "logFrequency")
```

---

plotQGAMs *Plot quantiles*

Description

Plot the results of a series of QGAM models (Fasiolo et al., 2017)

Usage

```
plotQGAMs(models, predictor, data, cols = c("#3C7A5B", "#468E6A",
  "#50A379", "#5AB788", "#65CC98"), se = 2, xlab = NA, ylab = NA,
  main = NA, ylim = NA, ...)
```

Arguments

models	A list of QGAM models as generated by the mqgam() function in the qgam package.
predictor	The predictor to be plotted. This predictor needs to be present in the fitted models, as well as in data.
data	The data the QGAM models were fit to. Needs to include the response variable, as well as all predictors in these models.
cols	A vector of colors. The lines corresponding to the quantiles will be plotted in these colors.
se	The number of standard errors for the confidence intervals. Default: 2 (i.e., 95% confidence intervals)

References

Fasiolo M., Goude Y., Nedellec R. and Wood S. N. (2017). Fast calibrated additive quantile regression. URL: https://arxiv.org/abs/1707.03307.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. Behavior Research Methods, 44(1), 287-304.

Examples

```
# Remove outliers from the ld data set, which contains lexical
decision latencies from the British Lexicon Project (Keuleers
et al, 2012)
ld = removeOutliers(ld, predictors)
ld = na.omit(ld)

# Tune learning rate for median
tune = tuneLearnFast(RT ~ s(logFrequency) + s(Length) +
                           s(logOLD20) + s(SND20),
                     data = ld, qu = 0.5)
sigpar = tune$lsig

# Define quantiles
quants = c(0.10,0.25,0.50,0.75,0.90)

# Run qgam models
qgams = mqgam(RT ~ s(logFrequency) + s(Length) + s(logOLD20) +
                   s(SND20),
                   data = ld, qu = quants, lsig = sigpar)

# Plot effect of frequency at quantiles
plotQGAMs(qgams, "logFrequency", ld)
```

---

pn *Picture naming data*

Description

Picture naming data (Balota et al., 2007)

Usage

```
pn
```

Format

An object of class `data.frame` with 484 rows and 6 columns.

Source

Székely, A., D'amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., ... & Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. Behavior Research Methods, Instruments, & Computers, 35(4), 621-633.

`removeOutliers` *Remove outliers*

Description

Remove outliers from a data frame

Usage

```
removeOutliers(data, predictors, sd = 3)
```

Arguments

`data`	A data frame.
`predictors`	A vector of the column names in data corresponding to the predictors that outliers should be removed for sd The number of standard deviations. Predictor values further than sd standard deviations from the predictor mean are removed. Default: 3.
`model`	A PAM model.

Examples

```
# Load data for the lexical decision latencies in the British Lexicon Project (Keuleers et al, 2012)
data(ld)

# Remove outliers
ld = removeOutliers.fnc(ld, c("logFrequency","Length","logOLD20","SND20"))
ld = na.omit(ld)
```

Index