

Package ‘stepjglm’

October 14, 2022

Title Variable Selection for Joint Modeling of Mean and Dispersion

Version 0.0.1

Description A Package for selecting variables for the joint modeling of mean and dispersion (including models for mixture experiments) based on hypothesis testing and the quality of model's fit. In each iteration of the selection process, a criterion for checking the goodness of fit is used as a filter for choosing the terms that will be evaluated by a hypothesis test. Pinto & Pereira (2021) <[arXiv:2109.07978](https://arxiv.org/abs/2109.07978)>.

Author Leandro A. Pereira [aut, cre],
Edmilson R. Pinto [aut]

Maintainer Leandro A. Pereira <leandro.ap@ufu.br>

License GPL-3

Depends R (>= 3.5.0), stats

LazyData true

RoxygenNote 7.1.2

Imports rsq

Encoding UTF-8

NeedsCompilation no

Repository CRAN

Date/Publication 2021-10-25 06:50:02 UTC

R topics documented:

bread_mixture	2
injection_molding	3
stepjglm	4

Index	8
--------------	----------

bread_mixture	<i>Bread-making problem data</i>
---------------	----------------------------------

Description

Data from a bread-making mixture experiment, to investigate and to value the final quality of flour.

Usage

```
data(bread_mixture)
```

Format

A data frame containing 90 rows and 6 variables.

The response variable was considered as the loaf volume after baking with target value of 530 ml.

Control variables:

- x_1 : Tjalve
- x_2 : Folke
- x_3 : HardRed Spring

Process variables:

- z_1 : mixing time
- z_2 : proofing (resting) time of the dough

Details

The bread-making problem, originally presented by Faergestad and Naes (1997), according to Naes et al. (1998), consisted of an experiment with three ingredients of mixture and two noise variables, and had as objective to investigate and to value the final quality of flour, composed by different mixtures of wheat flour, for production of bread.

References

Faergestad, E. M., Naes, T. (1997). Evaluation of baking quality of wheat flours: I: small scale straight dough baking test of heart bread with variable mixing time and proofing time. In: *Report MATFORSK*, As, Norway.

Naes, T., Faergestad, E. M., Cornell, J. A. (1998). A comparison of methods for analyzing data from a three component mixture experiment in the presence of variation created by two process variables, *Chemometrics and Intelligence Laboratory Systems*, v. 41, pp. 221-235.

Examples

```
data(bread_mixture)
head(bread_mixture)
```

injection_molding *Data from Injection molding experiment*

Description

The experiment was performed to study the influence of seven controllable factors and three noise factors on the mean value and the variation in the percentage of shrinkage of products made by injection molding.

Usage

```
data(injection_molding)
```

Format

A data frame containing 32 rows and 11 variables.

The responses were percentages of shrinkage of products made by injection molding (Y).

Controllable factors:

- A: cycle time
- B: mould temperature
- C: cavity thickness
- D: holding pressure
- E: injection speed
- F: holding time
- G: gate size

At each setting of the controllable factors, four observations were obtained from a $2^{(3-1)}$ fractional factorial with three noise factors:

- M: percentage regrind
- N: moisture content
- O: ambient temperature

Details

The data set considered is well known in the literature of industrial experiments and has been analyzed by several authors such as Engel (1992), Engel and Huele (1996) and Lee and Nelder (1998). The experiment was performed to study the influence of seven controllable factors and three noise factors on the mean value and the variation in the percentage of shrinkage of products made by injection molding. Noise factors are fixed during the experiment but are expected to vary randomly outside the experimental context.

The aim of the experiment was to determine the process parameter settings so that the shrinkage percentage was close to the target value and robust against environmental variations.

References

- Engel, J. (1992). Modeling variation in industrial experiments. *Applied Statistics*, 41, 579-593.
- Engel, J. and Huele, A. F. (1996). A generalized linear modeling approach to robust Design. *Technometrics*, 38, 365-373.
- Lee, Y. and Nelder, J.A. (1998). Generalized linear models for analysis of quality improvement experiments. *The Canadian Journal of Statistics*, 26, 95-105.

Examples

```
data(injection_molding)
head(injection_molding)
```

stepjglm	<i>Variable selection in joint modeling of mean and dispersion</i>
----------	--

Description

A Procedure for selecting variables in JMMD (including mixture models) based on hypothesis testing and the quality of the model's fit.

Usage

```
stepjglm(model, alpha1, alpha2, datafram, family, lambda1=1, lambda2=1, startmod=1,
         interations=FALSE)
```

Arguments

model	an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted. if datafram is a mixture data, datafram doesn't contain the principal mixture components.
alpha1	significance level for testing add new terms on the mean models.
alpha2	significance level for testing add new terms on the dispersion models.
datafram	a data frame containing the data.
family	a character string naming a family function or the result of a call to a family function. For glm.fit only the third option is supported. (See family for details of family functions). Describe the family function for the mean model (families implemented by package stats). For the dispersion model, the Gamma family whit log link is assumed.
lambda1	some function of the sample size to calculate the \tilde{R}_m^2 (See Pinto and Pereira (in press) and Zhang (2017) for more details). If equal to 1 (default), uses the standard correction for the \tilde{R}_m^2 . If equal to "EAIC", uses the EAIC criterion.
lambda2	some function of the sample size to calculate the \tilde{R}_d^2 (See Pinto and Pereira (in press) and Zhang (2017) for more details). If equal to 1 (default), uses the standard correction for the \tilde{R}_d^2 . If equal to "AIC", uses the corrected AIC _c criterion.

startmod	if datafram is a mixture data, startmod is the principal mixture components, else, startmod must be equal to 1 (default).
interations	if TRUE shows the outputs of iterations procedure step by step. The default is FALSE.

Details

The function implements a method for selection of variables for both the mean and dispersion models in the JMMD introduced by Nelder and Lee (1991) considering the *Adjusted Quasi Extended Likelihood* introduced by Lee and Nelder (1998). The method is a procedure for selecting variables, based on hypothesis testing and the quality of the model's fit. A criterion for checking the goodness of fit is used, in each iteration of the selection process, as a filter for choosing the terms that will be evaluated by a hypothesis test. For more details on selection algorithms, see Pinto and Pereira (in press).

Value

model.mean	a glm object with the adjustments for the mean model.
model.disp	a glm object with the adjustments for the dispersion model.
EAIC	a numeric object containing the <i>Extended Akaike Information Criterion</i> . For details, see Wang and Zhang (2009).
EQD	a numeric object containing the <i>Extended Quasi Deviance</i> . For details, see Nelder and Lee (1991).
R2m	a numeric object containing the standard correction for the \tilde{R}_m^2 . For details, see Pinto and Pereira (in press).
R2d	a numeric object containing the standard correction for the \tilde{R}_d^2 . For details, see Pinto and Pereira (in press).

Author(s)

Leandro Alves Pereira, Edmilson Rodrigues Pinto.

References

- Hu, B. and Shao, J. (2008). Generalized linear model selection using R^2 . *Journal of Statistical Planning and Inference*, 138, 3705-3712.
- Lee, Y., Nelder, J. A. (1998). Generalized linear models for analysis of quality improvement experiments. *The Canadian Journal of Statistics*, v. 26, n. 1, pp. 95-105.
- Nelder, J. A., Lee, Y. (1991). Generalized linear models for the analysis of Taguchi-type experiments. *Applied Stochastic Models and Data Analysis*, v. 7, pp. 107-120.

Pinto, E. R., Pereira, L. A. (in press). On variable selection in joint modeling of mean and dispersion. *Brazilian Journal of Probability and Statistics*. Preprint at <https://arxiv.org/abs/2109.07978> (2021).

Wang, D. and Zhang, Z. (2009). Variable selection in joint generalized linear models. *Chinese Journal of Applied Probability and Statistics*, v. 25, pp.245-256.

Zhang, D. (2017). A coefficient of determination for generalized linear models. *The American Statistician*, v. 71, 310-316.

See Also

[glm](#)
[summary.glm](#)

Examples

```
# Application to the bread-making problem:

data(bread_mixture)

Form =
as.formula(y~ x1:x2+x1:x3+x2:x3+x1:x2:(x1-x2)+x1:x3:(x1-x3)+
  + x1:z1+x2:z1+x3:z1+x1:x2:z1
  + x1:x3:z1+x1:x2:(x1-x2):z1
  + x1:x3:(x1-x3):z1
  + x1:z2+x2:z2+x3:z2+x1:x2:z2
  + x1:x3:z2+x1:x2:(x1-x2):z2
  +x1:x3:(x1-x3):z2)

object=stepjglm(Form,0.1,0.1,bread_mixture,gaussian,sqrt(90),"AIC","-1+x1+x2+x3")

summary(object$modelo.mean)
summary(object$modelo.disp)

object$EAIC # Print the EAIC for the final model

# Application to the injection molding data:

form = as.formula(Y ~ A*M+A*N+A*O+B*M+B*N+B*O+C*M+C*N+C*O+D*M+D*N+D*O+
  E*M+E*N+E*O+F*M+F*N+F*O+G*M+G*N+G*O)

data(injection_molding)

obj.dt = stepjglm(form, 0.05,0.05,injection_molding,gaussian,sqrt(nrow(injection_molding)),"AIC")

summary(obj.dt$modelo.mean)
summary(obj.dt$modelo.disp)

obj.dt$EAIC # Print the EAIC for the final model
```

stepjglm

7

```
obj.dt$EQD # Print the EQD for the final model
obj.dt$R2m # Print the R2m for the final model
obj.dt$R2d # Print the R2d for the final model
```

Index

* datasets

bread_mixture, [2](#)

injection_molding, [3](#)

bread_mixture, [2](#)

glm, [6](#)

injection_molding, [3](#)

stepjglm, [4](#)

summary.glm, [6](#)