

# Package ‘gausscov’

March 14, 2022

**Version** 0.1.6

**Date** 2022-03-14

**Title** The Gaussian Covariate Method for Variable Selection

**Author** Laurie Davies [aut, cre]

**Maintainer** Laurie Davies <laurie.davies@uni-due.de>

## Description

Given the standard linear model the traditional way of deciding whether to include the  $j$ th covariate is to apply the F-test to decide whether the corresponding beta coefficient is zero. The Gaussian covariate method is completely different. The question as to whether the beta coefficient is or is not zero is replaced by the question as to whether the covariate is better or worse than i.i.d. Gaussian noise. The P-value for the covariate is the probability that Gaussian noise is better. Surprisingly this can be given exactly and it is the same as the P-value for the classical model based on the F-distribution. The Gaussian covariate P-value is model free, it is the same for any data set. Using the idea it is possible to do covariate selection for a small number of covariates  $25$  by considering all subsets. Post selection inference causes no problems as the P-values hold whatever the data. The idea extends to stepwise regression again with exact probabilities. In the simplest version the only parameter is a specified cut-off P-value which can be interpreted as the probability of a false positive being included in the final selection. For more information see the web site below and the accompanying papers: L. Davies and L. Duembgen, “Covariate Selection Based on a Model-free Approach to Linear Regression with Exact Probabilities”, 2022, <[arxiv:2202.01553](https://arxiv.org/abs/2202.01553)>. L. Davies, “Linear Regression, Covariate Selection and the Failure of Modelling”, 2022, <[arXiv:2112.08738](https://arxiv.org/abs/2112.08738)>.

**LazyData** true

**License** GPL-3

**Depends** R ( $\geq 3.5.0$ ), stats

**Encoding** UTF-8

**RoxygenNote** 6.1.1

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2022-03-14 14:50:02 UTC

**R topics documented:**

abcq . . . . .	2
boston . . . . .	3
decode . . . . .	4
decomp . . . . .	4
f1st . . . . .	5
f2st . . . . .	6
f3st . . . . .	7
f3sti . . . . .	8
fasb . . . . .	9
fgeninter . . . . .	10
fgentrig . . . . .	10
fgr1st . . . . .	11
fgr2st . . . . .	12
fgrall . . . . .	13
flag . . . . .	13
fntp . . . . .	14
fpsired . . . . .	15
fpval . . . . .	16
fr1st . . . . .	17
frasb . . . . .	18
frpval . . . . .	19
fselect . . . . .	20
fundr . . . . .	20
fvauto . . . . .	21
leukemia . . . . .	21
mel-temp . . . . .	22
nufp . . . . .	22
redwine . . . . .	23
snspt . . . . .	24
<b>Index</b>	<b>25</b>

abcq

*American Business Cycle***Description**

The 22 variables are quarterly data from 1919-1941 of the variables GNP72, CPRATE, CORPYIELD, M1, M2, BASE, CSTOCK, WRICE67, PRODUR72, NONRES72, IRES72, DBUSI72, CDUR72, CNDUR72, XPT72, MPT72, GOVPUR72, NCSPDE72, NCSBS72, NCSCON72, CCSPDE72 and CCSBS72.

**Usage**

abcq

**Format**

A matrix of size 240 x 22

**Source**

<http://data.nber.org/data/abc/>

---

boston

*Boston data*

---

**Description**

This data set is part of the MASS package. The 14 columns are:

crim per capita crime rate by town

zn proportion of residential land zoned for lots over 25,000 sq.ft.

indus proportion of non-residential business acres per town

chas Charles River dummy variable (=1 if tract bounds river; 0 otherwise)

nox nitrogen oxides concentration (parts per 10 million)

rm average number of rooms per dwelling

age proportion of owner-occupied units built prior to 1940

dis weighted mean of distances to five Boston employment centres

rad index of accessibility to radial highways

tax full-value property-tax rate per \$10,000

ptration pupil-teacher ration by town

black  $100(Bk-0.63)^2$  where Bk is the proportion of blacks by town

lstat lower status of the population (percent)

medv median value of owner occupied homes in \$1000s.

**Usage**

boston

**Format**

A 506 x 14 matrix.

**Source**

R package MASS [https://cran.r-project.org/web/packages/available\\_packages\\_by\\_name.html](https://cran.r-project.org/web/packages/available_packages_by_name.html)

**References**

MASS Support Functions and Datasets for Venables and Ripley's MASS

---

decode	<i>Decodes the number of a subset selected by fasb.R to give the covariates</i>
--------	---

---

**Description**

Decodes the number of a subset selected by fasb.R to give the covariates

**Usage**

```
decode(ns, k)
```

**Arguments**

ns	The number of the subset
k	The number of covariates

**Value**

ind	The list of covariates
set	A binary vector giving the covariates

**Examples**

```
a<- decode(19,8)
```

---

decomp	<i>Decomposes given coded interactions into their component parts</i>
--------	---

---

**Description**

decomposes a given interaction ic into its component parts

**Usage**

```
decomp(ic, k, ord, inc=0)
```

**Arguments**

ic	The numbers of the coded interactions
k	The number of covariates of x without intercept, number plus 1 if inr=T
ord	The order of the interactions
inc	The indices of the interaction covariates with no dummy covariates when all powers are calculated including dummy covariates. This is returned by fgeninter.R

**Value**

decomp The component parts of the interaction.

**Examples**

```
bosint<-fgeninter(boston[,1:13],3,4)
a<-decomp(100,14,3,inc=bosint[[2]])
```

---

 f1st

*Stepwise selection of covariates*


---

**Description**

Stepwise selection of covariates

**Usage**

```
f1st(y,x,p0=0.01,nu=1,kmn=0,kmx=0,mx=21,kex=0,sub=T,inr=T,xinr=F,qq=0)
```

**Arguments**

y	Dependent variable
x	Covariates
p0	The P-value cut-off
nu	The order statistic of Gaussian covariates used for comparison
kmn	The minimum number of included covariates irrespective of cut-off P-value
kmx	The maximum number of included covariates irrespective of cut-off P-value.
mx	The maximum number covariates for an all subset search
kex	The excluded covariates
sub	Logical if TRUE best subset selected
inr	Logical if TRUE include intercept if not present
xinr	Logical if TRUE intercept already present
qq	The number of covariates to choose from. If qq=0 the number of covariates of x is used.

**Value**

pv In order the included covariates, the regression coefficient values, the Gaussian P-values, the standard P-values and the proportional reduction in the sum of squared residuals due to this covariate

res The residuals

stp The in order stepwise P-values, sum of squared residuals and the proportional reduction in the sum of squared residuals due to this covariate.

**Examples**

```
data(boston)
bostint<-fgeninter(boston[,1:13],2)[[1]]
a<-f1st(boston[,14],bostint,kmn=10,sub=TRUE)
```

f2st

*Repeated stepwise selection of covariates***Description**

Repeated stepwise selection of covariates

**Usage**

```
f2st(y, x, p0=0.01, nu=1, kmn=0, kmx=0, kex=0, mx=21, lm=9^9,
sub=T, inr=T, xinr=F, qq=0)
```

**Arguments**

y	Dependent variable
x	Covariates
p0	The P-value cut-off
nu	The order statistic of Gaussian covariates used for comparison
kmn	The minimum number of included covariates irrespective of cut-off P-value
kmx	The maximum number of included covariates irrespective of cut-off P-value.
kex	The excluded covariates
mx	The maximum number of covariates for an all subset search
lm	The maximum number of linear approximations
sub	Logical if TRUE select the best subset
inr	Logical if TRUE include an intercept
xinr	Logical if TRUE intercept already included
qq	The number of covariates to choose from. If qq=0 the number of covariates of x is used.

**Value**

pv In order the linear approximation, the included covariates, the regression coefficient values, the Gaussian P-values, the standard P-values and the proportional reduction in the sum of squared residuals due to this covariate.

**Examples**

```
data(boston)
bostint<-fgeninter(boston[,1:13],2)[[1]]
a<-f2st(boston[,14],bostint,lm=3,sub=TRUE)
```

---

f3st *Stepwise selection of covariates*

---

### Description

Stepwise selection of covariates

### Usage

```
f3st(y, x, m, kexmx=100, p0=0.01, nu=1, kmn=0, kmx=0, mx=21, lm=1000, kex=0, sub=T, inr=T, xinr=F, qq=0)
```

### Arguments

y	Dependent variable
x	Covariates
m	The number of iterations
kexmx	The maximum number of covariates in an approximation
p0	The P-value cut-off
nu	The order statistic of Gaussian covariates used for comparison
kmn	The minimum number of included covariates irrespective of cut-off P-value
kmx	The maximum number of included covariates irrespective of cut-off P-value.
mx	The maximum number covariates for an all subset search
lm	The maximum number of approximations.
kex	The excluded covariates
sub	Logical if TRUE best subset selected
inr	Logical if TRUE include intercept if not present
xinr	Logical if TRUE intercept already present
qq	The number of covariates to choose from. If qq=0 the number of covariates of x is used.

### Value

covch The sum of squared residuals and the selected covariates ordered in increasing size of sum of squared residuals.

lai The number of rows of covch

### Examples

```
data(leukemia)
a<-f3st(leukemia[[1]],leukemia[[2]],m=2,kexmx=5,kmn=5,sub=TRUE)
```

f3sti

*Selection of covariates with given excluded covariates***Description**

Selection of covariates with given excluded covariates

**Usage**

```
f3sti(y,x,covch,ind,m,kexmx=100,p0=0.01,nu=1,kmn=0,kmx=0,
      mx=21,lm=1000,kex=0,sub=T,inr=T,xinr=F,qq=0,lm0=0)
```

**Arguments**

y	Dependent variable
x	Covariates
covch	Sum of squared residuals and selected covariates
ind	The excluded covariates
m	Number of iterations
kexmx	The maximum number of covariates in an approximation.
p0	The P-value cut-off
nu	The order statistic of Gaussian covariates used for comparison
kmn	The minimum number of included covariates irrespective of cut-off P-value
kmx	The maximum number of included covariates irrespective of cut-off P-value.
mx	The maximum number covariates for an all subset search
lm	The maximum number of approximations.
kex	The excluded covariates
sub	Logical if TRUE best subset selected
inr	Logical if TRUE include intercept if not present
xinr	Logical if TRUE intercept already present
qq	The number of covariates to choose from. If qq=0 the number of covariates of x is used.
lm0	The current number of approximations

**Value**

ind1 The excluded covariates

covch The sum of squared residuals and the selected covariates ordered in increasing size of sum of squared residuals

lm0 The current number of approximations.



**Examples**

```

data(leukemia)
covch=c(2.023725,1182,1219,2888,0)
covch<-matrix(covch,nrow=1,ncol=5)
ind<-c(1182,1219,2888)
ind<-matrix(ind,nrow=3,ncol=1)
m<-1
a<-f3sti(leukemia[[1]],leukemia[[2]],covch,ind,m)

```

---

 fasb

---

*Calculates all subsets where each included covariate is significant.*


---

**Description**

It sel =TRUE it calls fselect.R and removes all such subsets which are a subset of some other selected subset. The remaining ones are ordered according to the sum of squared residuals. Subsets can be decoded with decode.R.

**Usage**

```
fasb(y,x,p0=0.01,q=-1,ind=0,sel=T,inr=T,xinr=F)
```

**Arguments**

y	The dependent variable
x	The covariates
p0	Cut-off p-value for significance
q	The number of covariates from which to choose. Equals number of covariates minus length of ind if q=-1.
ind	The indices of a subset of covariates for which all subsets are to be considered
sel	If TRUE calls fselect.R to removes all subsets of chosen sets
inr	If TRUE to include intercept
xinr	If TRUE intercept already included

**Value**

nv Coded List of subsets with number of covariates and sum of squared residuals

**Examples**

```

data(redwine)
nvv<-fasb(redwine[,12],redwine[,1:11])

```

---

**fgeninter**                      *Generation of interactions*


---

**Description**

Generates all interactions of degree at most ord excluding powers of 0-1 covariates

**Usage**

```
fgeninter(x,ord,inr=T,idv=0)
```

**Arguments**

x	Covariates
ord	Order of interactions
inr	Logical to include intercept
idv	List of 0-1 dummy covariates

**Value**

xx All interactions of order at most ord.

**Examples**

```
data(boston)
bostinter<-fgeninter(boston[,1:13],2)[[1]]
```

---

**fgentrig**                      *Generation of sine and cosine functions*


---

**Description**

Generates  $\sin(\pi*j*(1:n)/n)$  (odd) and  $\cos(\pi*j*(1:n)/n)$  (even) for  $j=1,\dots,m$  for a given sample size  $n$ .

**Usage**

```
fgentrig(n,m)
```

**Arguments**

n	Sample size
m	Maximum order of sine and cosine functions

**Value**

x The functions  $\sin(\pi*j*(1:n)/n)$  (odd) and  $\cos(\pi*j*(1:n)/n)$  (even) for  $j=1,\dots,m$ .

**Examples**

```
trig<-fgentrig(36,36)
```

---

fgr1st

*Calculates a dependence graph using Gaussian stepwise selection*


---

**Description**

Calculates an independence graph using Gaussian stepwise selection

**Usage**

```
fgr1st(x,p0=0.01,ind=0,nu=1,kmn=0,kmx=0,nedge=10^5,inr=T,xinr=F)
```

**Arguments**

x	The matrix of covariates
p0	Cut-off P-value
ind	Restricts the dependent nodes to this subset
nu	The order statistic of Gaussian covariates used for comparison.
kmn,	The minimum number selected variables for each node irrespective of cut-off P-value
kmx	The maximum number selected variables for each node irrespective of cut-off P-value
nedge	Maximum number of edges
inr	Logical, if TRUE include an intercept
xinr	Logical, if TRUE intercept already included

**Value**

ned Number of edges

edg List of edges together with P-values for each edge and proportional reduction of sum of squared residuals.

**Examples**

```
data(boston)
a<-fgr1st(boston[,1:13],ind=3:6)
```

fgr2st

*Calculates an independence graph using repeated stepwise selection***Description**

Calculates a dependency graph using repeated Gaussian stepwise selection

**Usage**

```
fgr2st(x, p0=0.01, ind=0, nu=1, kmn=0, kmx=0, nedge=10^5, inr=T, xinr=F)
```

**Arguments**

x	Matrix of covariates
p0	Cut-off P-value
ind	Restricts the dependent nodes to this subset
nu	The order statistic of Gaussian covariates used for comparison.
kmn	The minimum number of selected variables for each node irrespective of cut-off P-value
kmx	The maximum number of selected variables for each node irrespective of cut-off P-value
nedge	Maximum number of edges
inr	Logical, if TRUE include an intercept
xinr	Logical, if TRUE intercept already included

**Value**

ned Number of edges

edg List of edges giving nodes (covariates), the approximations for each node, the covariates in the approximation and the corresponding P-values.

**Examples**

```
data(redwine)
a<-fgr2st(redwine[,1:11], ind=4:8)
```

---

fgrall	<i>Calculates a dependence graph using Gaussian all subset selection</i>
--------	--

---

**Description**

Calculates an independence graph using Gaussian stepwise selection

**Usage**

```
fgrall(x,p0=0.01,kmx=0,mx=21,inr=T,xinr=F)
```

**Arguments**

x	The matrix of covariates
p0	Cut-off P-value
kmx	Maximum number included covariates for each node irrespective of cut-off P-value
mx	The maximum number of covariates.
inr	Logical, if TRUE include an intercept
xinr	Logical, if TRUE intercept already included

**Value**

ned Number of edges  
 edg List of edges with Gaussian P-value and percentage of sum of squared residuals explained by edge

**Examples**

```
data(boston)
a<-fgrall(redwine[,1:8])
```

---

flag	<i>Calculation of lagged covariates</i>
------	---

---

**Description**

Calculation of lagged covariates

**Usage**

```
flag(x,n,lag)
```

**Arguments**

x	The covariates
n	The sample size
lag	The maximum lag

**Value**

- y The first covariate of x without a lag, the dependent covariate.
- x1 The covariates with lags from 1 :lag starting with the first covariate.

**Examples**

```
data(abcq)
abcq1<-flag(abcq,240,16)
a<-f1st(abcq1[[1]],abcq1[[2]])
```

---

fnfp	<i>Estimates the number of false positives for given dimensions (n,k) and given order statistics nu</i>
------	---

---

**Description**

Interpolates using nufp or simulates the number of false positives for given dimensions (n,k) and given order statistics nu

**Usage**

```
fnfp(n,k,p0,nu,nufp,gr=F,nsim=0,kmx=0,idum=1)
```

**Arguments**

n	The dimension of dependent variable
k	The number of covariates
p0	Cut-off P-value
nu	The order statistic
gr	Logical, if TRUE then $p0 < p0/k$ as is the default for graphs
nufp	Requires a data set nufp.rda of previous simulations
nsim	Number of simulations
kmx	Maximum number of selected covariates, must be larger than nu, for example $nu+10$
idum	Seed for the random number generator

**Value**

enfp Estimated number of false positives.

mnfp Mean number of false positives when simulating.

hist Histogram of number of false positives when simulating

**Examples**

```
a<-fnfp(100,24,0.01,1:5,nufp,nsim=1000,kmx=10)
```

---

fpsired	<i>Calculates Hampel's redescending psi function</i>
---------	--

---

**Description**

Calculates Hampel's redescending psi function

**Usage**

```
fpsired(x,cnr)
```

**Arguments**

x	The point at which the psi function is evaluated
cnr	The parameters of Hampel's redescending psi function

**Value**

rpsi The value of the function

**Examples**

```
fpsired(1,c(1,3,5))
```

---

fpval	<i>Calculates the regression coefficients, the P-values and the standard P-values for the chosen subset ind</i>
-------	---

---

### Description

Calculates the regression coefficients, the P-values and the standard P-values for the chosen subset ind.

### Usage

```
fpval(y,x,ind,q=-1,nu=1,inr=T,xinr=F)
```

### Arguments

y	The dependent variable
x	The covariates
ind	The indices of the subset of the covariates whose P-values are required
q	The total number of covariates from which ind was chosen. If q=-1 the number of covariates of x minus length ind plus 1 is taken.
nu	The order statistic used to compute the P-values
inr	Logical If TRUE intercept to be included
xinr	If TRUE intercept already included

### Value

apv In order the subset ind, the regression coefficients, the Gaussian P-values, the standard P-values and the proportion of sum of squares explained.

res The residuals

### Examples

```
data(boston)
a<-fpval(boston[,14],boston[,1:13],c(1,2,4:6,8:13))
```



---

fr1st *Robust stepwise selection of covariates*

---

### Description

Robust stepwise selection of covariates

### Usage

```
fr1st(y,x,cn=1,cnr=c(1,3,5),p0=0.01,sg=0,nu=1,kmx=0,mx=21,kex=0,sub=T,inr=T,xinr=F,red=F)
```

### Arguments

y	Dependent variable
x	Covariates
cn	The constant for Huber's psi-function
cnr	The constants for Hampel's three part redescending psi function
p0	The P-value cut-off
sg	Scale value of residuals
nu	The order for calculating the P-value
kmx	The maximum number of included covariates
mx	The maximum number of included covariates if the option subset =TRUE is used
kex	The excluded covariates
sub	Logical, if TRUE best subset selected
inr	Logical TRUE to include intercept
xinr	Logical TRUE if intercept already included
red	Logical If true Hampel's three part redescending psi function

### Value

pv In order the subset ind, the regression coefficients, the Gaussian P-values, the standard P-values.  
 res The residuals  
 stpv The stepwise regression results: covariate, P-value and scale

### Examples

```
data(boston)
a<-fr1st(boston[,14],boston[,1:13],kex=7:8)
```

---

frsb	<i>Robust selection of covariates using Huber's psi-function or Hampel's redescending psi-function based on all subsets</i>
------	---

---

### Description

Calculates all possible subsets and selects those where each included covariate is significant using a robustified version of `fasb.R`

### Usage

```
frsb(y,x,cn=1,cnr=c(1,3,5),p0=0.01,q=-1,sg=0,ind=0,sel=T,inr=T,xinr=F,red=F)
```

### Arguments

<code>y</code>	The dependent variable
<code>x</code>	The covariates
<code>cn</code>	The constant for Huber's psi-function
<code>cnr</code>	The constants for for Hampel's three part redescending psi-function
<code>p0</code>	The P-value cut-off
<code>q</code>	If $q > 0$ the number of covariates from which <code>ind</code> was chosen
<code>sg</code>	The scale parameter
<code>ind</code>	The subset for which the results are required
<code>sel</code>	Logical, if TRUE removes all subsets of chosen sets
<code>inr</code>	Logical if TRUE include intercept
<code>xinr</code>	Logical If TRUE intercept included in <code>x</code>
<code>red</code>	Logical If true Hampel's three part redescending psi function

### Value

`nv` Coded list of subsets with number of covariates and scale ordered according to scale.

### Examples

```
data(boston)
a<-frsb(boston[,14],boston[,1:8])
ind<-decode(235,8)
```

---

frpval	<i>Robust regression using Huber's psi-function or Hampel's three part redescending psi-function providing P-values</i>
--------	---

---

**Description**

Robust regression using Huber's psi-function or Hampel's three part redescending psi-function providing P-values

**Usage**

```
frpval(y, x, ind, cn=1, cnr=c(1, 3, 5), sg=0, q=-1, scale=T, inr=T, xinr=F, red=F)
```

**Arguments**

y	Dependent variable
x	Covariates
ind	The subset of covariates for which the results are required
cn	Tuning constant for Huber's psi-function
cnr	Tuning constants for Hampel's three part redescending psi function
sg	Scale. If 0 the MAD is used
q	The number of covariates available. If q=-1 the covariates are used.
scale	Logical. If TRUE scale sg recalculated
inr	Logical, TRUE to include intercept
xinr	Logical TRUE if x already has intercept
red	Logical If TRUE Hampel's three part redescending psi function

**Value**

ppi In order the subset ind, the regression coefficients, the Gaussian P-values, the standard P-values  
 res Residuals  
 sg Scale  
 rho Sums of rho, psi and psi1 functions.

**Examples**

```
data(boston)
a<-frpval(boston[, 14], boston[, 1:13], 1:6)
```

---

fselect	<i>Selects the subsets specified by fasb.R and frasb.R.</i>
---------	---

---

**Description**

All subsets which are a subset of a specified subset are removed. The remaining subsets are ordered by the sum of squares of the residuals (fasb.R) or the scale (frasb.R)

**Usage**

```
fselect(nv, k)
```

**Arguments**

nv	The subsets specified by fasb.R or frasb.R
k	The variables

**Value**

ind The selected subsets.

**Examples**

```
b<-fasb(redwine[,12],redwine[,1:5 ],sel=FALSE)[[1]]
a<-fselect(b,11)[[1]]
b[a,]
```

---

fundr	<i>Converts directed into an undirected graph</i>
-------	---

---

**Description**

Conversion of a directed graph into an undirected graph

**Usage**

```
fundr(gr)
```

**Arguments**

gr	A directed graph
----	------------------

**Value**

gr The undirected graph

**Examples**

```
data(boston)
grb<-fgr1st(boston[,1:13])
grbu<-fundr(grb[[2]][,1:2])
```

---

fvauto	<i>Vector autoregressive approximation</i>
--------	--

---

**Description**

Vector autoregressive approximation

**Usage**

```
fvauto(x, n, omx, p0=0.01)
```

**Arguments**

x	Variable
n	Sample size
omx	Maximum lag
p0	The P-value cut-off

**Value**

res The selected lagged variables for each variable

**Examples**

```
data(abcq)
a<-fvauto(abcq, 240, 10)
```

---

leukemia	<i>Leukemia data</i>
----------	----------------------

---

**Description**

Dataset of  $n = 72$  persons indicating presence or absence of leukemia and  $q = 3571$  gene expressions of the 72 persons

**Usage**

```
data(leukemia)
```

**Format**

itemleukemia[[1]]0-1 vector of length giving presence or absence of leukemia itemleukemia[[2]]72x3571 matrix giving the gene expressions of the 72 persons

**Source**

<http://stat.ethz.ch/~dettling/bagboost.html>

**References**

Boosting for tumor classification with gene expression data. Dettling, M. and Buehlmann, P. Bioinformatics, 2003,19(9):1061–1069.

---

mel-temp	<i>Melbourne minimum temperature</i>
----------	--------------------------------------

---

**Description**

The daily minimum temperature in Melbourne for the years 1981-1990.

**Usage**

mel\_temp

**Format**

A vector of length 3650

**Source**

<https://www.kaggle.com/paulbrabban/daily-minimum-temperatures-in-melbourne>

---

nufp	<i>nufp</i>
------	-------------

---

**Description**

This data set gives the results of simulations on the average number of false positives when the covariates are independent Gaussian noise. It is a 280 x 12 matrix. The first column is the number of covariates  $k$ , the second is the sample size  $n$ , columns 3:12 give the average number of false positives for the values of the order statistic  $\nu$  for  $\nu:1:10$ . The first 70 rows are for  $p_0=0.01$ , the rows 71:140  $p_0=0.05$ , the rows 141:210 for  $p_0=0.01/k$ , the rows 211:280 for  $p_0=0.05/k$

**Usage**

nufp

**Format**

A 280 x 12 matrix.

**Source**

Author's simulations

**References**

"Covariate Selection Based on a Model-free Approach to Linear Regression with Exact Probabilities", 2020, <arXiv:1906.01990>

---

redwine

*Redwine data*

---

**Description**

The subjective quality of wine on an integer scale from 1-10 (variable 12) together with 11 physicochemical properties

**Usage**

redwine

**Format**

A matrix of size 1599 x 12

**Source**

<https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/>

**References**

Modeling wine preferences by data mining from physicochemical properties, Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J., Decision Support Systems, Elsevier, 2009,47(4):547–553.

---

snspt

*Sunspot data*

---

**Description**

The average number of sunspots each month from January 1749 to January 2020: variable 1 year; variable 2, month; variable 3 number of sunspots.

**Usage**

snspt

**Format**

A matrix of size 3253 x 7

**Source**

WDC-SILSO, Royal Observatory of Belgium, Brussels



# Index

## \* datasets

- abcq, 2
- boston, 3
- leukemia, 21
- mel-temp, 22
- nufp, 22
- redwine, 23
- snspt, 24

abcq, 2

boston, 3

decode, 4

decomp, 4

f1st, 5

f2st, 6

f3st, 7

f3sti, 8

fasb, 9

fgeninter, 10

fgentrig, 10

fgr1st, 11

fgr2st, 12

fgrall, 13

flag, 13

fnfp, 14

fpsired, 15

fpval, 16

fr1st, 17

frasb, 18

frpval, 19

fselect, 20

fundr, 20

fvauto, 21

leukemia, 21

mel-temp, 22

mel\_temp (mel-temp), 22

nufp, 22

redwine, 23

snspt, 24