Package 'weightedZdiff'

August 18, 2020

Type Package

Title Calculation of z-Differences

Version 0.1.0

Author Tim Filla

Maintainer Tim Filla <tim.filla@hhu.de>

Description
Calculates z-differences (O.Kuss (2013) <doi:10.1016/j.jclinepi.2013.06.001>) for each variable scale (continuous, binary, ordinal and nominal) with or without weights (e.g. generated by propensity score methods).

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

Depends R (>= 2.10)

NeedsCompilation no

Repository CRAN

Date/Publication 2020-08-18 09:40:03 UTC

R topics documented:

`testdata` *Medical data of 5735 patients.*

Description

This dataset contains medical data of 5735 patients diamonds.

Usage

```
testdata
```

Format

A data frame with 5735 rows and 12 variables:

age patients age

ARF something

female is patient female (1) or not (0)

sepsis Sepsis Diagnosis

CHF Congestive Heart Failure

Cirr Cirrhosis

colcan Colon Cancer

Coma Coma

lungcan Lung cancer

MOSF Malignancy

treatment RHC (Swan-Ganz catheter)

meanbp1 Mean blood pressure ...

Source

http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/rhc.html

zdifference *zdifference for dataset*

Description

The function calculates the zdifferences for each variable in a dataset or each column in a matrix (depends on the format of your data). Furthermore the sum of the squared zdifferences is calculated. The variables are set into classes continuous, binary and nominal automatically by the following algorithm. If the variable has only 2 different values its treated as binary. If the variable has more then 9 observations or the class of the variable is factor its treated as nominal and otherwise continuous. The user can specify the type of every variable by hand.

Usage

```
zdifference(dataset,ref,weights=NULL,standard_weights=FALSE,na.rm=TRUE,
binary_variable=NULL,ordinal_variable=NULL,continuous_variable=NULL,nominal_variable=NULL,
r=2,var.est=FALSE,coefvar.est=FALSE,grad=1)
```

Arguments

dataset	An object of class data.frame or matrix, which contains the variables for which the zDifferences should be calculated and the reference variable in columns.
ref	The name of the reference variable, name must be in datasets' names.
weights	The name of the variable containing the weights for each observation, name must be in datasets' names.
standard_weights	Should the unweighted zdifferences be calculated or not.
na.rm	Should NAs be removed or not. If NAs exists in dataset and na.rm=FALSE then an error will occure.
binary_variable	optional: Names of binomial variables.
ordinal_variable	optional: Names of ordinal variables.
continuous_variable	optional: Names of continuous variables.
nominal_variable	optional: Names of nominal variables.
r	Number of digits to round the result.
var.est	Should the weighted z-Difference for the variances of continuous variables be reported (TRUE) or not (FALSE)
coefvar.est	Should the coefficient of variation for continuous variables be reported (TRUE) or not (FALSE)
grad	The Moments for which to calculate the weighted z-Difference for continuous variables. grad=2 means the first and second moments are calculated.

Author(s)

Tim Filla

References

For standard z-difference (unweighted) https://pubmed.ncbi.nlm.nih.gov/23972521/

Examples

```
data(testdata)
#new dataset
zdifference(testdata,"treatment",grad=2,continuous_variable=c("age","meanbp1"),
binary_variable=c("CHF","Cirr","colcan","Coma","lungcan","MOSF","sepsis","female","ARF"))
#generate iptw weights
p<-glm(treatment~.,data=testdata,family="binomial")$fitted.values
testdata$weights<-ifelse(testdata$treatment==0,1/(1-p),1/p)
zdifference(testdata,"treatment",weights="weights",grad=2,
continuous_variable=c("age","meanbp1"),binary_variable=c("CHF","Cirr",
"colcan","Coma","lungcan","MOSF","sepsis","female","ARF"),standard_weights=TRUE)
```

---

zdifference_binary *z-difference for binary variables*

Description

The function calculates the binary weighted z-Difference for a binary reference variable (ref) and a binary variable (x)

Usage

```
zdifference_binary(x,ref,w=NULL,na.rm=TRUE,r)
```

Arguments

x	The binary variable for which the weighted z-Difference should be calculated.
ref	The binary reference variable as a vector.
w	The weights to calculate the weighted binary z-Difference
na.rm	Should NAs be removed or not. If NAs exists in dataset and na.rm=FALSE then an error will occure.
r	digits to round the returned value, default is 2

Value

The function returns the calculated z-Difference as a numeric value.

Author(s)

Tim Filla

References

For standard z-difference (unweighted) https://pubmed.ncbi.nlm.nih.gov/23972521/

Examples

```
#generate the data. The weights are taken from uniform #distribution and the
#values of x are generated from a bernoulli distribution with
#success rate 0.3. The reference variable
#is chosen from a bernoulli distribution with success rate 0.8.
ref<-sample(0:1,1000,replace=TRUE,prob=c(0.2,0.8))
erg<-unlist(lapply(1:1000,function(z){
  w<-runif(1000)
  x<-rbinom(1000,1,0.3)
  zdifference_binary(x,ref,w)
}))
hist(erg,breaks=50,main="z-difference for continuous data")
plot(seq(0.005,0.97,0.01),quantile(erg,seq(0.005,0.97,0.01)),
type="l",lwd=3,xlab=c("quantile"),ylab=c("x-value"))
points(seq(0.005,0.97,0.01),qnorm(seq(0.005,0.97,0.01)),col="red",type="l",lwd=3,lty=2)
legend("topleft",legend=c("N(0,1) distribution","sample distribution"),lty=c(2,1),
lwd=c(3,3),col=c("red","black"),cex=1.3)
```

`zdifference_coefvar` *z-difference for the coefficient of vartiation for normal distributed variables.*

Description

The function calculates the coefficient of variation z-Difference for a binary reference variable (ref) and an ordinal variable (x)

Usage

```
zdifference_coefvar(x,ref,na.rm=TRUE,r=2)
```

Arguments

`x`	The variable for which the z-Difference should be calculated.
`ref`	The binary reference variable as a vector.
`na.rm`	Should NAs be removed or not. If NAs exists in dataset and na.rm=FALSE then an error will occure.
`r`	digits to round the returned value, default is 2

Value

The function returns the calculated z-Difference as a numeric value.

Author(s)

Tim Filla

References

https://pubmed.ncbi.nlm.nih.gov/23972521/

Examples

```
#generate the data.
#variable x has 5 different status with probability of
#beeing in status i is given by:0.1,0.2,0.3,0.3,0.1. #The reference variable
#is chosen from a bernoulli distribution with success #rate 0.8.
ref<-sample(0:1,1000,replace=TRUE,prob=c(0.2,0.8))
erg<-unlist(lapply(1:1000,function(z){
  w<-runif(1000)
  x<-rnorm(1000,25)
  zdifference_coefvar(x,ref)
}))
hist(erg,breaks=50,main="z-difference for continuous data")
plot(seq(0.005,0.97,0.01),quantile(erg,seq(0.005,0.97,0.01)),
type="l",lwd=3,xlab=c("quantile"),ylab=c("x-value"))
points(seq(0.005,0.97,0.01),qnorm(seq(0.005,0.97,0.01)),col="red",type="l",lwd=3,lty=2)
legend("topleft",legend=c("N(0,1) distribution","sample distribution"),lty=c(2,1),
lwd=c(3,3),col=c("red","black"),cex=1.3)
```

`zdifference_continuous` *z-difference for continuous variables.*

Description

The function calculates the continuous weighted z-Difference for a binary reference variable (ref) and a continuous variable (x)

Usage

```
zdifference_continuous(x, ref, w=NULL, na.rm = TRUE, r = 2)
```

Arguments

`x`	The continuous variable for which the weighted z-Difference should be calculated.
`ref`	The binary reference variable as a vector.
`w`	The weights to calculate the weighted continuous z-Difference
`na.rm`	Should NAs be removed or not. If NAs exists in dataset and na.rm=FALSE then an error will occure.
`r`	digits to round the returned value, default is 2

Value

The function returns the calculated z-Difference as a numeric value.

Author(s)

Tim Filla

References

For standard z-difference (unweighted) https://pubmed.ncbi.nlm.nih.gov/23972521/

Examples

```
#generate the data. The weights are taken from uniform distribution and the
#values of x are normal distributed with mean 45 and variance 9. The reference variable
#is chose from a bernoulli distribution with success rate 0.8.
ref<-sample(0:1,1000,replace=TRUE,prob=c(0.2,0.8))
erg<-unlist(lapply(1:1000,function(z){
  w<-runif(1000)
  x<-rnorm(1000,45,9)
  zdifference_continuous(x,ref,w)
}))
hist(erg,breaks=50,main="z-difference for continuous data")
plot(seq(0.005,0.97,0.01),quantile(erg,seq(0.005,0.97,0.01)),type="l",
lwd=3,xlab=c("quantile"),ylab=c("x-value"))
points(seq(0.005,0.97,0.01),qnorm(seq(0.005,0.97,0.01)),col="red",type="l",lwd=2,lty="dashed")
```

---

zdifference_nominal *z-difference for nominal variables.*

Description

The function calculates the nominal weighted z-Difference for a binary reference variable (ref) and a nominal variable (x)

Usage

```
zdifference_nominal(x,ref,w=NULL,na.rm=TRUE,norma=TRUE,r=2)
```

Arguments

x	The continuous variable for which the weighted z-Difference should be calculated.
ref	The binary reference variable as a vector.
w	The weights to calculate the weighted continuous z-Difference
na.rm	Should NAs be removed or not. If NAs exists in dataset and na.rm=FALSE then an error will occure.

norma	If norma = TRUE the weighted z-Difference has a standard Gaussian distribution. If norma = FALSE the resulting distribution is chi squared with #status -1 as degree of freedom.
r	digits to round the returned value, default is 2

Value

The function returns the calculated z-Difference as a numeric value.

Author(s)

Tim Filla

Examples

```
#generate data. The weights are taken from uniform distribution and the
#values of x are generated from a multinomial distribution with success
#rate (0.2,0.2,0.3,0.15,0.15) for the five different status.The reference
#variable is chosen from a bernoulli distribution with success rate 0.8.
ref<-sample(1:0,1000,replace=TRUE,prob=c(0.2,0.8))
erg<-unlist(lapply(1:1000,function(z){
  w<-runif(1000)
  x<-sample(0:4,1000,replace=TRUE,prob=c(0.2,0.2,0.3,0.15,0.15))
  zdifference_nominal(x,ref,w,norma=TRUE)
}))
hist(erg,breaks=50,main="z-difference for nominal data")
plot(seq(0.005,0.97,0.01),quantile(erg,seq(0.005,0.97,0.01)),type="l",lwd=3)
points(seq(0.005,0.97,0.01),qnorm(seq(0.005,0.97,0.01)),col="red",type="l",lwd=2,lty="dashed")
```

---

zdifference_ordinal *weighted z-difference for ordinal variables*

Description

The function calculates the ordinal weighted z-Difference for a binary reference variable (ref) and an ordinal variable (x)

Usage

```
zdifference_ordinal(x,ref,w=NULL,na.rm=TRUE,r=10)
```

Arguments

x	The ordinal variable for which the weighted z-Difference should be calculated.
ref	The binary reference variable as a vector.
w	The weights to calculate the weighted ordinal z-Difference
r	digits to round the returned value, default is 2
na.rm	

Value

The function returns the calculated z-Difference as a numeric value.

Author(s)

Tim Filla

References

For standard z-difference (unweighted) https://pubmed.ncbi.nlm.nih.gov/23972521/

Examples

```
#generate the data. The weights are taken from uniform distribution and the
#variable x has 5 different status with probability of beeing in status i is
#given by:0.1,0.2,0.3,0.3,0.1. The reference variable
#is chosen from a bernoulli distribution with success rate 0.8.
ref<-sample(0:1,1000,replace=TRUE,prob=c(0.2,0.8))
erg<-unlist(lapply(1:1000,function(z){
  w<-runif(1000)
  x<-sample(1:5,1000,replace=TRUE,prob=c(0.1,0.2,0.3,0.3,0.1))
  zdifference_ordinal(x,ref,w)
}))
hist(erg,breaks=50,main="z-difference for continuous data")
plot(seq(0.005,0.97,0.01),quantile(erg,seq(0.005,0.97,0.01)),type="l",
lwd=3,xlab=c("quantile"),ylab=c("x-value"))
points(seq(0.005,0.97,0.01),qnorm(seq(0.005,0.97,0.01)),col="red",type="l",lwd=3,lty=2)
legend("topleft",legend=c("N(0,1) distribution","sample distribution"),lty=c(2,1),
lwd=c(3,3),col=c("red","black"),cex=1.3)
```

---

zdifference_var *z-difference for variance of continuous variable*

Description

The function calculates the weighted z-Difference for a continuous variable (x) with binaryreference variable (ref) a

Usage

```
zdifference_var(x,ref,w=NULL,na.rm=TRUE,r)
```

Arguments

x	The continuous variable for which the weighted z-Difference should be calculated.
ref	The binary reference variable as a vector.
w	The weights to calculate the weighted binary z-Difference

na.rm	Should NAs be removed or not. If NAs exists in dataset and na.rm=FALSE then an error will occure.
r	digits to round the returned value, default is 2

Value

The function returns the calculated z-Difference as a numeric value.

Author(s)

Tim Filla

Examples

```
#generate the data. The weights are taken from uniform distribution and the
#values of x are generated from a bernoulli distribution with success rate 0.3.
#The reference variable is chosen from a bernoulli distribution with success rate 0.8.
ref<-sample(0:1,1000,replace=TRUE,prob=c(0.2,0.8))
erg<-unlist(lapply(1:1000,function(z){
  w<-runif(1000)
  x<-rnorm(1000,1,0.3)
  zdifference_var(x,ref,w)
}))
hist(erg,breaks=50,main="z-difference for continuous data")
plot(seq(0.005,0.97,0.01),quantile(erg,seq(0.005,0.97,0.01)),type="l",
lwd=3,xlab=c("quantile"),ylab=c("x-value"))
points(seq(0.005,0.97,0.01),qnorm(seq(0.005,0.97,0.01)),col="red",type="l",lwd=3,lty=2)
legend("topleft",legend=c("N(0,1) distribution","sample distribution"),lty=c(2,1),
lwd=c(3,3),col=c("red","black"),cex=1.3)
```

Index