

Package ‘CATS’

August 5, 2013

Type Package

Title Joint power analysis for non-symmetric two-stage case-control designs for SNP data

Version 1.01

Date 2013-08-04

Author Anders Albrechtsen

Maintainer Anders Albrechtsen <albrecht@binf.ku.dk>

Description Joint power analysis for two-stage designs for SNP data. Generalized so that the ratios between the cases and controls, the allele frequencies and the relative risk can vary between stages

Reference Skol AD, Scott LJ, Abecasis GR, Boehnke M: Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 38: 209-213, 2006

License GPL

R topics documented:

CATS-package	1
cats	2
print.CATS	5
summary.CATS	5

CATS-package	<i>Power calculation for a joint analysis of a two-stage case control design for SNP data</i>
--------------	---

Description

Power calculation for a joint analysis of a two-stage case control design for SNP data.

Details

Package: CATS
Type: Package
Version: 1.0
Date: 2006-10-16
License: GPL

The function `cats` calculates the power

Author(s)

Anders Albrechtsen

Maintainer: Anders Albrechtsen <albrecht@binf.ku.dk>

References

Skol AD, Scott LJ, Abecasis GR, Boehnke M: Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 38: 209-213, 2006.

See Also

<http://www.sph.umich.edu/csg/abecasis/CaTS/>

Examples

```
#calculate the power under a multiplicative model using a two stage design and
#assuming a relative risk of 1.5
cats(freq=0.2,ncases=500,ncases2=500,ncontrols=1000,ncontrols2=1000,
risk=1.5,multiplicative=1)
```

<code>cats</code>	<i>Power calculation for a joint analysis of a two-stage case control design for SNP data</i>
-------------------	---

Description

Power calculation for a joint analysis of a two-stage case control design for SNP data.

Usage

```
cats(freq = 0.5, freq2 = -1, ncases = 500, ncontrols = 500, ncases2 = 500,
ncontrols2 = 500, risk = 1.5, risk2 = -1, pisamples = -1, prevalence = 0.1,
prevalence2 = -1, additive = 0, recessive = 0, dominant = 0,
multiplicative = 1, alpha = 1e-07, pimarkers = 0.00316)
```

Arguments

<code>freq</code>	numeric. The minor allele frequency (MAF) in the first stage
<code>freq2</code>	numeric. The MAF in the second stage, Optional, if -1 the same value as for the first stage is given
<code>ncases</code>	integer. The number of cases in the first stage
<code>ncontrols</code>	integer. The number of controls in the first stage
<code>ncases2</code>	integer. The number of cases in the second stage
<code>ncontrols2</code>	integer. The number of controls in the second stage
<code>risk</code>	numeric. The relative risk in the first stage
<code>risk2</code>	numeric. The relative risk in the second stage, Optional, if -1 the same value as for the first stage is given
<code>pisamples</code>	numeric. The weights used for the joint statistic. Optional. see details
<code>prevalence</code>	numeric. The prevalence of the disease in the population for the first stage
<code>prevalence2</code>	numeric. The prevalence of the disease in the population for the second stage, Optional, if -1 the same value as for the first stage is given
<code>additive</code>	boolean. if 1 an additive model is assumed
<code>recessive</code>	boolean. if 1 a recessive model is assumed
<code>dominant</code>	boolean. if 1 a dominant model is assumed
<code>multiplicative</code>	boolean. if 1 a multiplicative model is assumed
<code>alpha</code>	numeric. The significance threshold. Often the a threshold of 0.05 divided by the number of markers is chosen
<code>pimarkers</code>	numeric. The fraction of markers genotyped in the second stage

Details

These power analysis are based on Skol et al. 2006, But are generalized so that the ratio between cases and controls may vary between stages. Also the allele frequencies, disease prevalence and relative risk are also allowed to vary. The joint statistic $z_{\text{joint}} = z_1 \sqrt{\pi} + z_2 \sqrt{1-\pi}$ where z_1 is the z-score for the first stage and the weight π is calculated as $\pi = 1 / \text{var}(\hat{p}'_1 - \hat{p}_1) * (1 / \text{var}(\hat{p}'_1 - \hat{p}_1) + 1 / \text{var}(\hat{p}'_2 - \hat{p}_2))^{-1}$, where \hat{p}'_1 is the estimate of the allele frequency of the cases in the first stage. This is consistent with Skol et al 2006 when the ratios of cases and controls are the same in both stages. When this is not the case the weight π may vary slightly with different allele frequencies and different relative risks. For power calculations I would recommend calculating the weight at a likely scenario where there is about 80-90% power and fixing the weights at other scenarios (and the testing of the real data) to this weight. This can be done by assigning `pisample` to a value. In practice this will hardly affect the power.

Value

`P.one.study` The power if only one study was performed, NB! This is only a valid estimate if the relative risk and allele frequency is the same for both stages

<code>P.first.stage</code>	The power for a marker to proceed the the second stage
<code>P.rep.study</code>	The power of the study if based on replication and not a joint analysis
<code>P.joint.min</code>	The power of the joint analysis tp detect at least one susceptibility SNP assuming that five susceptibility SNPs exits
<code>P.joint</code>	The power of the joint analysis
<code>pi</code>	The weight used to calculate the joint statistic
<code>T.one.study</code>	Recommended thresholds for a one-stage study
<code>T.first.stage</code>	Recommended thresholds for the first stage in two-stage study
<code>T.second.stage.rep</code>	Recommended thresholds for the second stage in replication analysis
<code>T.second.stage.joint</code>	Recommended thresholds for the second stage in a joint analysis
<code>E.Disease.freq.cases1</code>	The expected disease allele frequency in stage 1 for cases
<code>E.Disease.freq.controls1</code>	The expected disease allele frequency in stage 1 for controls
<code>E.Disease.freq.cases2</code>	The expected disease allele frequency in stage 2 for cases
<code>E.Disease.freq.controls2</code>	The expected disease allele frequency in stage 2 for controls

Author(s)

Anders Albrechtsen

References

Skol AD, Scott LJ, Abecasis GR, Boehnke M: Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 38: 209-213, 2006.

See Also

<http://www.sph.umich.edu/csg/abecasis/CaTS/>

Examples

```
#calculate the power under a multiplicative model using a two stage design
#and assuming a relative risk of 1.5
cats(freq=0.2,ncases=500,ncases2=500,ncontrols=1000,ncontrols2=1000,
risk=1.5,multiplicative=1)

power.J<-c()
power.R<-c()
power.O<-c()
RR<-23:32/20
for(tal in 1:length(RR)) {
```

```

temp<-cats(risk=RR[tal])
power.J[tal]<-temp$P.joint
power.R[tal]<-temp$P.rep.study
power.O[tal]<-temp$P.one.study
}
plot(RR,power.J,type="b",lwd=2,ylab="Power")
lines(RR,power.R,lwd=2,col=2,type="b")
lines(RR,power.O,lwd=2,col=3,type="b")
legend(1.4,0.4,c("joint analysis","replication design"
,"one stage design"),col=1:3,lwd=3,bty="n")

```

print.CATS *prints the calculated power*

Description

prints the calculated power

Usage

```
print.CATS(x, ...)
```

Arguments

x	a CATS object
...	Dynamic parameter for the values of additional parameters for the summary method

See Also

cats

summary.CATS *summary of a CATS object*

Description

summary of a CATS object

Usage

```
summary.CATS(object, ...)
```

Arguments

`object` a CATS object
`...` Dynamic parameter for the values of additional parameters for the print method

See Also

`cats`

Examples

```
x<-cats()  
summary(x)
```