Package ‘optimalAllocation’

June 17, 2013

Type Package

Title Optimal allocation for longitudinal studies with time-varying exposure

Version 1.0

Date 2013-06-11

Author Jose Barrera-Gómez and Xavier Basagaña

Maintainer Jose Barrera-Gómez <jbarrera@creal.cat>

Depends R (>= 2.14.0), tcltk

Description
This package computes the optimal combination of number of participants and number of repeated measurements in longitudinal studies with fixed follow-up duration such that maximizes the power to detect the effect under the alternative hypothesis without exceeding a fixed budget or minimizes the cost of the study while achieving a certain target power. Response variable is assumed to be continuous with damped exponential covariance structure and the exposure of interest is assumed to be binary which can be time-varying as well as its prevalence.

License GPL (>=2)

R topics documented:

---

optimalAllocation-package
Optimal combination of number of participants and number of repeated measurements in longitudinal studies with time-varying exposure.

Description

This function computes the optimal combination of number of participants and number of repeated measurements in longitudinal studies with fixed follow-up duration such that it maximizes the power to detect the effect under the alternative hypothesis without exceeding a fixed budget or it minimizes the cost of the study while achieving a certain target power. The response variable is assumed to be continuous and the exposure of interest is assumed to be binary and possibly time-varying.

Details

Package:	optimalAllocation
Type:	Package
Version:	1.0
Date:	2013-06-11
License:	GPL (>=2)

Author(s)

Jose Barrera-Gómez and Xavier Basagaña.

Maintainer: Jose Barrera-Gómez <jbarrera@creal.cat>

References

Barrera-Gómez J, Spiegelman D, Basagaña X. Optimal combination of number of participants and number of repeated measurements in longitudinal studies with time-varying exposure. Statistics in Medicine [Epub ahead of print].

OA *Optimal combination of number of participants and number of repeated measurements in longitudinal studies with time-varying exposure.*

Description

This function computes the optimal combination of number of participants and number of repeated measurements in longitudinal studies with fixed follow-up duration such that it maximizes the power to detect the effect under the alternative hypothesis without exceeding a fixed budget or it minimizes the cost of the study while achieving a certain target power. The response variable is assumed to be continuous and the exposure of interest is assumed to be binary and possibly time-varying.

Usage

```
OA(target, pattern, rMax = 20, theta = NULL, rho = NULL, sigma2, rhoe = NULL, pe0,
    per = NULL, piM = NULL, kappa = NULL, budget = NULL, c1, reqPower = NULL,
    beta, alpha = 0.05)


## Default S3 method:
```

```
OA(target, pattern, rMax = 20, theta = NULL, rho = NULL, sigma2, rhoe = NULL, pe0,
   per = NULL, piM = NULL, kappa = NULL, budget = NULL, c1, reqPower = NULL,
   beta, alpha = 0.05)
## S3 method for class 'OA'
print(x, ...)
## S3 method for class 'OA'
summary(object, ...)
## S3 method for class 'OA'
plot(x, Ncex = 0.7, ...)
```

Arguments

target	character: "maxPower" for maximizing the power or "minCost" for minimizing the financial cost of the study.
pattern	character: two possible response patterns under the alternative hypothesis are allowed. "CMD" assumes an acute and transient exposure effect while "LDD" assumes a cumulative exposure effect, i.e., an exposure-time interaction.
rMax	maximum value of the number of repeated measurements evaluated. See 'Details' section.
theta	damping parameter of the damped exponential covariance structure of the response. See 'Details' section.
rho	correlation between the first and the last measurements of the response. See 'Details' section.
sigma2	residual variance. See 'Details' section.
rhoe	intraclass correlation of the exposure. See 'Details' section.
pe0	exposure prevalence at the first measurement. See 'Details' section.
per	exposure prevalence at the last measurement. See 'Details' section.
piM	fraction of individuals lost at the end of the study. See 'Details' section.
kappa	ratio between the financial cost of the first measurement (including recruitment) and each of the subsequent ones.
budget	total budget for the study if target = "maxPower".
c1	cost of the first measurement (including recruitment).
reqPower	required power if target = "minCost".
beta	expected effect under the alternative hypothesis. See 'Details' section.
alpha	significance level.
x	object of class 'OA'.
object	object of class 'OA'.
Ncex	font size for N values in the plot. Default is 0.7.
...	further arguments for print, plot or summary functions.

Details

The response covariance structure is assumed damped exponential, whose covariance matrix has diagonal elements sigma^2 and off-diagonal [j, j'] elements, sigma2rho^[(|j'-j|/r)^theta], where sigma2 is the residual variance, rho is the correlation between the first and the last response measurements, theta in [0, 1] is the damping parameter and (r + 1) is the total number of measurements per participant. This covariance structure becomes compound symmetry if theta=0 and first order autoregressive if theta=1.

The intraclass correlation of exposure is defined as rhoe = (sum(SigmaE) - Tr(SigmaE))/(r Tr(SigmaE)) where sum() and Tr() denote the sum of the elements and the trace of a matrix respectively, and SigmaE is the covariance matrix of the exposure. rhoe can be interpreted as an index of similarity or agreement between each subject's exposure in the different time periods. If rhoe=1, the exposure is time-invariant.

The exposure prevalence is assumed to vary linearly from pe0 at the first measurement to per at the last measurement. If per is omitted, the exposure prevalence is assumed time-invariant.

The dropout pattern is assumed to be monotone, i.e. losing one individual measurement implies losing all the subsequent measurements of that individual. No missing data at the first measurement is assumed.

If the response pattern is "CMD", beta can be interpreted as the expected difference in the mean of the response variable, at any time point, between exposed and non exposed. If the response pattern is "LDD", beta can be interpreted as the expected difference in the mean of the response variable between the worst exposure pattern (i.e., exposed at all measurements) and non-exposed, at the end of follow-up.

Note that the value of rMax exponentially increases the computational time.

Value

roptreal	real solution to the problem.
ropt	optimal number of repeated measurements (i.e., the optimal total number of measurements is ropt + 1) constrained to be not greater than rMax.
Nopt	optimal number of participants.
maxPower	maximized power if target = "maxPower".
minCost	minimized financial cost if target = "minCost".
cost	needed cost if target = "maxPower".
power	achieved power if target = "minCost".
sdBeta	standard error of beta.
f	data.frame with power and sample size calculations for each value of the number of repeated measurements from 1 to rMax.
parameters	input parameters.

References

Barrera-Gómez J, Spiegelman D, Basagaña X. Optimal combination of number of participants and number of repeated measurements in longitudinal studies with time-varying exposure. Statistics in Medicine [Epub ahead of print].

Examples

```
# Maximizing power under "CMD" pattern of the response:
ex1 <- OA(target = "maxPower", pattern = "CMD", rMax = 20, theta = 0.5, rho = 0.7,
          sigma2 = 1, rhoe = 0.2, pe0 = 0.2, per = 0.3, piM = 0.2, kappa = 3, budget = 40,
          c1 = 1, beta = -0.3, alpha = 0.05)
ex1
summary(ex1)
plot(ex1)
```

```
# Minimizing cost under "LDD" pattern:
ex2 <- OA(target = "minCost", pattern = "LDD", rMax = 20, theta = 0, rho = 0.6,
          sigma2 = 1, rhoe = 0.6, pe0 = 0.2, per = 0.2, piM = 0.2, kappa = 3,
          reqPower = 0.8, c1 = 1, beta = 0.8, alpha = 0.05)
ex2
summary(ex2)
plot(ex2)

# Particular cases: power and sample size calculations in a cross-sectional
# study (r = 0 and pattern = "CMD"):
# Power:
ex3 <- OA(target = "maxPower", pattern = "CMD", rMax = 0, sigma2 = 1, pe0 = 0.2,
          budget = 400, c1 = 1, beta = -0.3, alpha = 0.05)
ex3
summary(ex3)
plot(ex3)

# Sampe size:
ex4 <- OA(target = "minCost", pattern = "CMD", rMax = 0, sigma2 = 1, pe0 = 0.2,
          reqPower = 0.8, c1 = 2.5, beta = 0.8, alpha = 0.05)
ex4
summary(ex4)
plot(ex4)
```

---

plotExposedPeriods *Distribution of the number of exposed periods per participant.*

Description

This function plots the distribution of the number of exposed periods per participant, once the values of the exposure intraclass correlation, the exposure prevalence and the number of repeated measurements have been fixed and the exposure covariance structure is assumed to follow CS. This function is a tool to decide the exposure intraclass correlation value.

Usage

```
plotExposedPeriods(r, pe, rhoe, eps = 0.001, maxIter = 1000)
```

Arguments

r	number of repeated measurements, i.e., the total number of measurements is r + 1.
pe	exposure prevalence, assumed the same at all time point measurements.
rhoe	exposure intraclass correlation.
eps	precision in the results (relative error). Default is 0.001.
maxIter	maximum number of iterations for the computation of the distribution. Default is 1000.

References

Gange SJ, Generating Multivariate categorical variates using the iterative proportional fitting algorithm. The American Statistician 1995; 49(2): 134-138.

Barrera-Gómez J, Spiegelman D, Basagaña X. Optimal combination of number of participants and number of repeated measurements in longitudinal studies with time-varying exposure. Statistics in Medicine [Epub ahead of print].

See Also

plotExposedPeriodsInt

Examples

```
plotExposedPeriods(r = 3, pe = 0.25, rhoe = 0.5)
```

---

plotExposedPeriodsInt *Distribution of the number of exposed periods per participant (interactive).*

Description

This function performs an interactive plot to explore the distribution of the number of exposed periods per participant as a function of the exposure intraclass correlation, once the values of the exposure prevalence and the number of repeated measurements have been fixed and the exposure covariance structure is assumed to follow CS. This function is a tool to decide the exposure intraclass correlation value.

Usage

```
plotExposedPeriodsInt(r, pe)
```

Arguments

r	number of repeated measurements, i.e., the total number of measurements is r + 1.
pe	exposure prevalence, assumed the same at all time point measurements.

References

Gange SJ, Generating Multivariate categorical variates using the iterative proportional fitting algorithm. The American Statistician 1995; 49(2): 134-138.

Barrera-Gómez J, Spiegelman D, Basagaña X. Optimal combination of number of participants and number of repeated measurements in longitudinal studies with time-varying exposure. Statistics in Medicine [Epub ahead of print].

See Also

plotExposedPeriods

Examples

```
require(tcltk)
plotExposedPeriodsInt(r = 3, pe = 0.25)
```